Czech University of Life Sciences Prague
Department of Water Resources and Environmental modelling
Environmental Modelling

Modelling framework for representation of lumped hydrological models

Master thesis of:
Eleni Mickovska

Supervisor:
doc. Ing. Petr Maca, Ph.D.

30.03.2022

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

Eleni Mickovska

Environmental Modelling

Thesis title

Modelling framework for representation of lumped hydrological models

---

Objectives of thesis

The goal is to develop and test a modular modeling framework used for the description of water balance in the small and midsize catchments.

Methodology

Use a dHRUM – distributed Hydrological Response Unit Model – available at the Department as a framework for developing new hydrological model structures. The methodology consists of:

1. Develop additional modules and extend dHRUM modeling framework with new groundwater modules.
2. Extend current dHRUM framework using a new soil water modules.
3. Comparison of different structures of dHRUM models using CAMEL dataset.
4. Use the differential split sample test for the models comparison based on different objective functions.

The proposed extent of the thesis
standard

Keywords
runoff rainfall

Recommended information sources
BEVEN, K J. *Environmental modelling: an uncertain future? : an introduction to techniques for uncertainty estimation in environmental prediction.* London: Routledge, 2009. ISBN 978-0-415-45759-0.
BEVEN, K J. *Rainfall-runoff modelling : the primer.* Chichester: Wiley-Blackwell, 2012. ISBN 978-0-470-71459-1.

Expected date of thesis defence
2021/22 SS – FES

The Diploma Thesis Supervisor
doc. Ing. Petr Máca, Ph.D.

Supervising department
Department of Water Resources and Environmental Modeling

Electronic approval: 22. 3. 2022
prof. Ing. Martin Hanel, Ph.D.
Head of department

Electronic approval: 22. 3. 2022
prof. RNDr. Vladimír Bejček, CSc.
Dean

Prague on 30. 03. 2022

I hereby declare that I have independently elaborated the diploma/final thesis with the topic of: "Modelling framework for representation of lumped hydrological models" and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources. I am aware that my diploma thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendment of some acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis. I am aware that by submitting the diploma/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of some acts, as amended, regardless of the result of its defence. With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

Prague, 31.03.2022.

Eleni Mickovska

Author's Acknowledgement

I would like to thank doc. Ing Petr Máca, Ph.D. for his advice during my work on this thesis. I would also like to thank my family and friends for their constant support during the past two years.

Abstract

Hydrological modeling plays a central role in water resource management through access to risks and impacts of hydrological phenomena, where the model itself represents a simplified representation of a realworld system. The hydrological models are classified into many different groups depending on different factors. This study will be focusing on the performances of a conceptual lumped rainfall-runoff model. The main objective of this study is to extend and test an existing modular modelling framework, dHRUM which stands for Distributed Hydrological Response Unit Model, used for description of water balance in small and midsize catchments. This modelling framework was initially developed by the Department of Water Resources and Environmental Modelling at the Faculty of Environmental Sciences (Czech University of Life Sciences Prague) and furtherly extended with seven groundwater structures and three soil water structures. The analyzed time-series is the CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) dataset. For measuring the quality of the model's performance, calibration and validation were performed with the model using the differential split sample test, where calibration was applied for the dry period, and validation was applied for the wet period and vice versa. Three different single objective functions were used: KGE, NSE, MAE. The Differential Evolution algorithm was used as an optimization algorithm. The different structures were compared based on the three different goodness-of-fit criteria: KGE (Kling Gupta Efficiency), NSE (Nash Sutcliffe Efficiency), MAE (Mean Absolute Error) which were results from using the gof (goodness-of-fit) package in R. The results were very successfull when using the KGE objective function, as the validations were more successfull when the calibration was done on the wet period. An analysis of the different groundwater and soil-water structures were made, where 'superior' and 'non-superior' structures were chosen.

Keywords: Hydrological modelling, Differential split sample test, KGE, NSE, MAE, CAMELS dataset, rainfall-runoff modelling, DE, modelling framework

Abstract

Hydrologické modelování hraje ústřední roli v řízení vodních zdrojů prostřednictvím přístupu k rizikům a dopadům hydrologických jevů, kdy samotný model představuje zjednodušenou reprezentaci systému reálného světa. Hydrologické modely jsou klasifikovány do mnoha různých skupin v závislosti na různých faktorech. Tato studie se zaměří na výkony koncepčního modelu soustředěných srážek a odtoků. Hlavním cílem této studie je rozšířit a otestovat stávající modulární modelovací rámec dHRUM, což je zkratka pro Distributed Hydrological Response Unit Model, používaný pro popis vodní bilance v malých a středně velkých povodích. Tento modelovací rámec byl původně vyvinut Katedrou vodních zdrojů a environmentálního modelování Fakulty životního prostředí (ČZU) a dále rozšířen o sedm podzemních vodních struktur a tři půdní vodní stavby. Analyzovanou časovou řadou je datový soubor CAMELS (Catchment Attributes and MEteorology for Large-sample Studies). Pro měření kvality výkonu modelu byla s modelem provedena kalibrace a validace pomocí diferenciálního testu děleného vzorku, kde byla použita kalibrace pro suché období a validace byla aplikována na vlhké období a naopak. Byly použity tři různé jednoúčelové funkce: KGE, NSE, MAE. Algoritmus diferenciální evoluce byl použit jako optimalizační algoritmus. Různé struktury byly porovnány na základě tří různých kritérií dobré shody: KGE (účinnost Kling Gupta), NSE (účinnost Nash Sutcliffe), MAE (střední absolutní chyba), které byly výsledkem použití gof (dobrá shoda) balíček v R. Výsledky byly velmi úspěšné při použití objektivní funkce KGE, protože validace byly úspěšnější, když byla kalibrace provedena ve vlhkém období. Byla provedena analýza různých podzemních a půdně-vodních struktur, kde byly vybrány „nadřazené“ a „nadřazené“ struktury.

Klíčová slova: Hydrologické modelování, Diferenciální dělený vzorkový test, KGE, NSE, MAE, CAMELS dataset, modelování srážek a odtoků, DE, modelovací rámec

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The continuous movement of water on Earth and the atmosphere, commonly known as the natural water cycle, is a complex system consisting of many different processes. Climate change largely depends on the redistribution of solar energy which is directly influenced by the movement of water in the atmosphere and on land. Thus, understanding the movement of the water along with its many different processes is fundamental for hydrological modeling.

A hydrological system is a system consisting of different components and processes over a particular region, which are to be modeled by the modeler. The region is commonly known as watershed/drainage basin/catchment, defined as an area that contributes surface runoff to any point.

Hydrological modeling plays a central role in water resource management through access to risks and impacts of hydrological phenomena (Beven, 2006), where the model itself represents a simplified representation of a real world system (Sorooshian et al., 2008). This study will be focusing on the performances of a conceptual lumped rainfall-runoff model.

The main objective of this study is to extend and test an existing modular modelling framework, dHRUM, which stands for Distributed Hydrological Response Unit Model, used for description of water balance in small and midsize catchments. In this thesis the modelling framework was used for testing conceptual lumped rainfall-runoff models. This modelling framework was initially developed by the Department of Water Resources and Environmental Modelling at the Faculty of Environmental Sciences (Czech University of Life Sciences Prague).

The first objective of the thesis is the development and testing of seven groundwater models defined by Stoelzle et al. (2015). The second objective of this thesis is the development and testing of three soil water structures defined in Knoben et al. (2019a). The third objective of the study is to use the different model structures for testing the performance of the modelling framework, using

the CAMELS dataset. The fourth objective is to access the performances of the modeling framework by using different objective functions, different goodness-of-fit criteria, an optimization algorithm and the differential split sample test as defined by Klemeš (1986).

More information can be found in the Objectives and Literature Review parts of the thesis.

CHAPTER 2

Objectives

The main objective of this study is to extend and test an existing modular modelling framework used for description of water balance in small and midsize catchments. DHRUM (Distributed Hydrological response Unit Model) was used for the purposes of this study. DHRUM is a modelling framework developed by the Department of Water Resources and Environmental Modelling at the Faculty of Environmental Sciences (Czech University of Life Sciences Prague). The modelling framework was developed with R and Rcpp package, allowing an integration between R and C++. The main objectives of this study are:

- Extend the already existing dHRUM modelling framework by adding seven groundwater modules defined by Stoelzle et al. (2015)
- Extend the already existing dHRUM modelling framework by adding three soil water modules defined by Knoben et al. (2019a)
- Comparison of different structures of dHRUM models using the CAMELS dataset
- Use different techniques for the models comparison: differential split sample test, objective functions, goodness-of-fit criteria, exploratory data analysis.

The first objective is the development of 7 groundwater models defined in Stoelzle et al. (2015). The following perceptual groundwater models were implemented: Linear reservoir with leakage (LINLRES), Linear reservoir with a direct-by-pass (LINBYRES), Two serial linear reservoirs (LIN2SE), Two parallel linear reservoirs (LIN2PA), Non linear power law reservoirs (POWRES), Exponential reservoir (EXPRES), Linear reservoir with threshold - controlled increased storage outflow (FLEXRES).

Additionally, the linear reservoir (LINRES) was also included in the analysis as the eight model which was already defined in the framework.

The second objective is the development of 3 soil water models defined in Knoben et al. (2019a): Collie River Basin 2 (COLLIEV2), New Zealand (NEWZEALAND), GR4J.

Additionally, the probability distributed model (PDM) was used as a fourth modeling structure which was already defined in the framework.

The third objective of the study is to use the different model structures for testing the performance of the different lumped models by using the dHRUM modelling framework, using the CAMELS dataset, where only the nldas forcing data was used (North American Land Data Assimilation System). This objective is very closely connected to the fourth objective, where again the performance of the model was being accessed by:

- applying the differential split sample test, where calibration was applied on the dry period, validation was applied on the wet period, and vice versa
- applying three different single-objective functions: Kling Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE) and Mean Absolute Error (MAE)
- applying three different goodness-of-fit measures: Kling Gupta Efficiency (KGE), Nash-Sutcliffe Efficiency (NSE) and Mean Absolute Error (MAE)
- applying optimization algorithm used for searching the optimal parameters values by maximizing or minimizing the single objective functions using the previously mentioned goodness-of-fit measures
- applying exploratory data analysis.

CHAPTER 3

Literature review

Hydrological modeling plays a central role in water resource management through access to risks and impacts of hydrological phenomena (Beven, 2006).

A model is a simplified representation of a real world system (Sorooshian et al., 2008). Hydrological models represent a complex hydrological system over a region. There are different types of classification of hydrological models among which they can be recognized as physical models, conceptual models and empirical models (Devia et al., 2015). They are classified into many different groups depending on different factors. This study will be focusing on the performances of a conceptual lumped rainfall-runoff model.

Rainfall-runoff models have been generally used for research purposes. Rainfall-runoff models main focus is gaining knowledge about the movement of water in and out of the catchments (Beven, 2011). Conceptual rainfall-runoff models are used for forecasts such as streamflow forecast.

A very common idea among hydrologists which is continuously discussed is whether we should pursue an approach where "one model fits all" (Kavetski and Fenicia, 2011). This idea was based on the assumption that hydrological processes are the same everywhere (Perrin et al., 2003). So, the development of rainfall-runoff models was born from the idea that rainfall-runoff models could be applied to any catchment (Knoben et al., 2020). In contrast, Beven stands behind the idea of "uniqueness of place" (Beven, 2000) and he states that we don't have enough information of the fundamental processes undergoing a catchment which makes it unique. Given these two fundamentally different approaches to hydrological modelling into consideration, choosing the model for representation of the natural phenomena is the first critical step that each modeler has to make.

The Distributed Hydrological Response Unit Model (dHRUM) in this thesis is used as a framework for testing 32 different conceptual lumped rainfall-runoff models. The dHRUM model was at first inspired by the variable infiltration

capacity water balance model, VIC, which was built under the assumption that infiltration capacity, runoff and evapotranspiration are dependent on the catchment's properties such as vegetation and soil (Wood et al., 1992).

Many conceptual rainfall-runoff models were being developed due to the reason that these models are easy-to-use when it comes to runoff prediction in large-scale regions with sufficient observed streamflow data used for calibrating the model (Chiew, 2010). Our knowledge about how one model functions, had been extended by many hydrologists who have taken an active research into large-scale catchments (Addor et al., 2020; Coxon et al., 2019; Lane et al., 2019; Seiller et al., 2012).

In this study, the existing dHRUM was extended with development of multiple model structures, which contributed to creating a modeling framework for evaluating the performance of the model in large-scale catchments. This type of approach for hydrological modeling was motivated by many existent modeling frameworks that serve us as tools for comparing different modeling practices - "The diversity of hydrologic modeling approaches motivates our effort to develop a unified modeling framework to integrate and compare competing modeling approaches" - Clark et al. (2015).

There have been many modeling frameworks developed for the same purposes among which are the Structure for Unifying Multiple Modeling Alternatives SUMMA (Clark et al., 2015), Framework for Understanding Structural Errors FUSE (Saavedra et al., 2021), Modular Assessment of Rainfall–Runoff Models Toolbox MARRMoT (Knoben et al., 2019a), SuperflexPy (Dal Molin et al., 2021), Machine Learning Rainfall-Runoff Model Induction ML-RR-MI (Herath et al., 2021) and many others with the ability to choose different model structures for accessing the performance of the model.

Accessing the model performance by testing different model structures has been used for gaining knowledge to better understanding of the drainage area dynamics such as low-flows simulation, streamflow forecast, baseflow generation etc. (Fenicia et al., 2006; Nicolle et al., 2014; Staudinger et al., 2011; Stoelzle et al., 2015).

Many rainfall-runoff models were designed to represent the main hydrological processes in a catchment thus each one of them holds a different aspect about which are the main or dominant hydrological processes in a catchment, which is directly connected to the level of uncertainty depending on the choice of the model (Andréassian et al., 2009; Fenicia et al., 2008; Van Esse et al., 2013).

In this study, eight different groundwater storage structures were implemented and four soil water storage structures. DHRUM holds in total six storage structures: groundwater, soil, surface, vegetation, snow and interception storage. This

approach was inspired by Stoelzle et al. (2015) and his idea that developing various reliable storage-discharge relationships is crucial for understanding the underground processes of movement of water because it is directly connected to accessing the availability of water and the responsiveness to future changes.

In hydrological modeling, the performance of the model is accessed by calibration on the parameters descriptors of the system behavior with the aim to estimate the optimal model parameters in enabling the hydrological model to match the observations. With the rise of the computers and the digital era, hydrologists increasingly use automatic procedures for calibration. Different optimization techniques were applied by hydrologists for parameter calibrations of the model such as Genetic Algorithms (Duan et al., 1992; Liu et al., 2007), Particle Swarm Optimization (Gill et al., 2006) etc. In this study, the DE (Differential Evolution) algorithm was used. DE algorithm is a nature-inspired algorithm used for solving global optimizations problems (Ardia et al., 2011).

Optimization algorithms are used for searching the optimal parameters values by maximizing or minimizing the objective function using numerous goodness-of-fit measures. Goodness-of-fit measures such as Kling-Gupta Efficiency (KGE) (Gupta et al., 2009), Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970), Mean Annual Error (MAE), Weighted Sum of Squared Residual (WSSR), Root Mean Square Error (RMSE) are commonly used among modellers depending on the single objective function's purpose (Gupta et al., 2009; Knoben et al., 2019b; Moriasi et al., 2015; Razavi and Tolson, 2013; Ritter and Munoz-Carpena, 2013). In this study, KGE, NSE and MAE are used for the measure of fit of the dHRUM model.

So, accessing the model performance has been done by hydrologists in numerous ways among which (as we mentioned before) are building modelling frameworks for testing different modelling structures and using different optimization algorithms for the goodness-of-fit measures. Understanding the model's weaknesses and strengths is essential when building such models thus by using different conditions and approaches to access the model's performance lead us not only to better understanding the model, but also to better understanding the catchment hydrological characteristics (Knoben et al., 2020). So, in this study, for accessing the model performance the differential split sample test is used (Klemeš, 1986). The dataset was calibrated on the dry period in the given time-range, and validated on the wet period and vice versa. For the purposes of this study, the CAMELS dataset Newman et al. (2015) is used for calibrating and validating the model, described in the study area section of this study.

Only by extending our knowledge on the weaknesses of our model we will be able to build better models.

CHAPTER 4

Characteristics of study area

The analyzed time-series is the CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) dataset. The CAMELS dataset consists of 671 drainage areas in the contiguous United States (CONUS) where human activities have minimum impact (Newman et al., 2015). The drainage areas are represented by the following sets of attributes: land cover, soil, climate, streamflow, topography and geology. The large number of drainage areas which are evenly distributed, together with its landscape descriptors make this dataset 'a perfect' dataset for large-scale hydrological analysis.

The basin's metadata contains information for the basin physical characteristics such as: size of the basin, elevation, slope and forest fraction percentage, but also gauge information such as: latitude, longitude and the area of the basin. Elevations vary between 10 and 3,600 meters above sea level, while drainage areas vary between 5 and 26,000 square kilometers. The data set has been revised by Addor et al. (2017) with information on the basin's attributes and their interrelationships, based on the basin's topographic characteristics.
Additionally three different forcing data are included in the CAMELS dataset: Maurer, Nldas and Daymet. In this study, the Nldas forcing data was used (North American Land Data Assimilation System), where information about the minimum temperature, maximum temperature, precipitation, shortwave radiation, humidity, and snow water equivalent is contained, starting from 01.01.1980 to 31.12.2014. For the same time range, information for the daily discharges are contained in the CAMELS dataset from USGS (The United States Geological Survey).

For the purposes of this study, only the temperature and precipitation were used as inputs to the model. The temperature value was calculated as the mean value between the minimum and maximum temperature of the selected day. The daily discharge used for calibration was converted appropriately into millimeters per day. During the data cleaning part, fifty one basins were excluded due to the

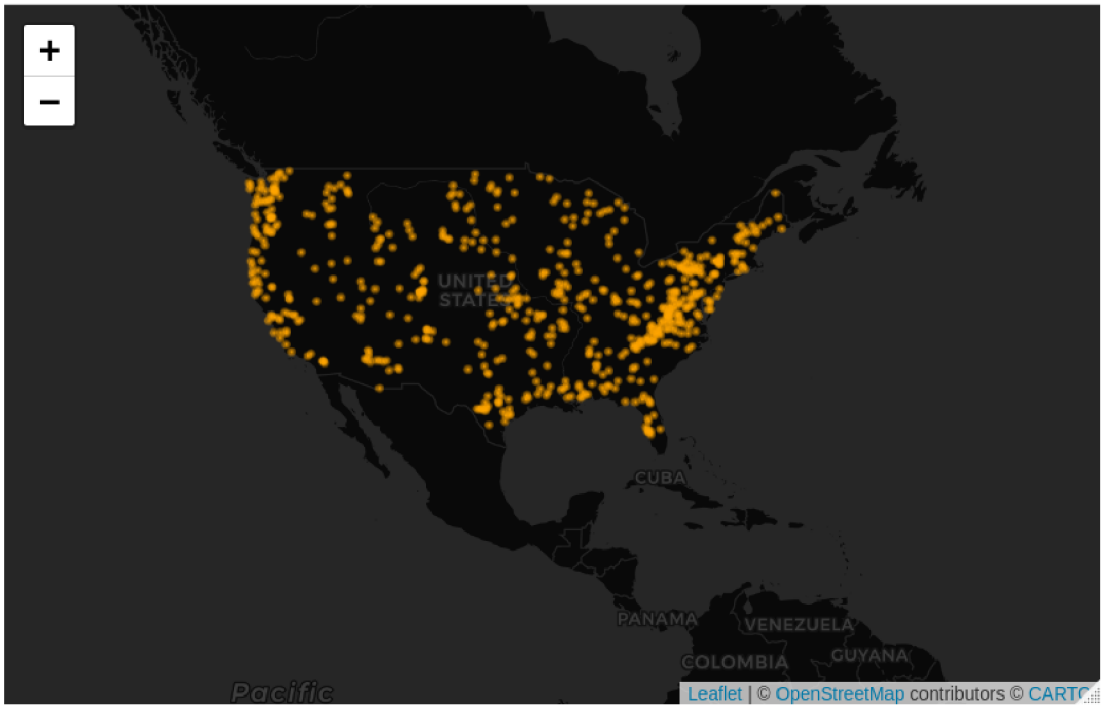

FIGURE 4.1: Chosen basins from the CAMEL dataset

missing data for the daily discharge in a total of more than five years from the thirty five years time range. For measuring the quality of the model's performance, calibration and validation were performed with the model using the differential split sample test (Klemeš, 1986), where calibration was applied for the dry period, and validation was applied for the wet period and vice versa. For calibrating the model on the dry period, an algorithm for finding the ten consecutive years with lowest precipitation was implemented, and accordingly the validation was performed on the years remaining. For calibrating the model on the wet period, an algorithm for finding the ten consecutive years with highest precipitation was implemented, and accordingly the validation was performed on the remaining years. It's crucial to take into consideration that the time period on which the calibration and validation were performed varies between different basins due to missing data, but no more than five years, giving the model at least twenty five years time range, and at most thirty years time range, from which calibration was always performed on a 10 years consecutive periods.

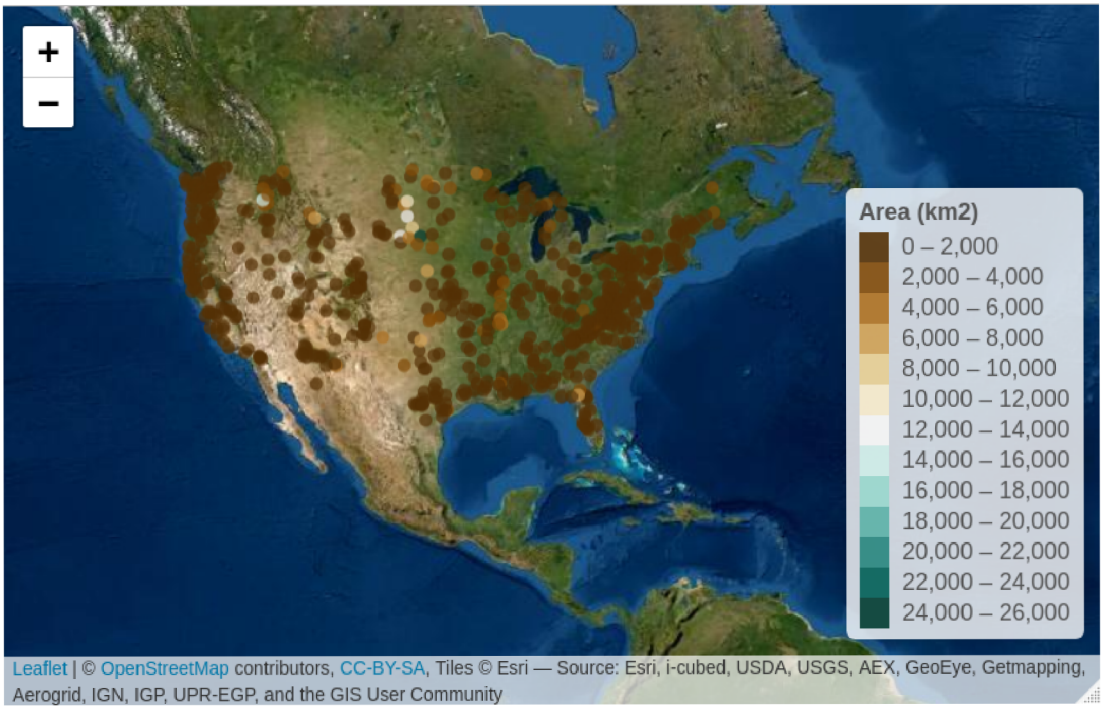

FIGURE 4.2: The area of the basins from CAMEL dataset

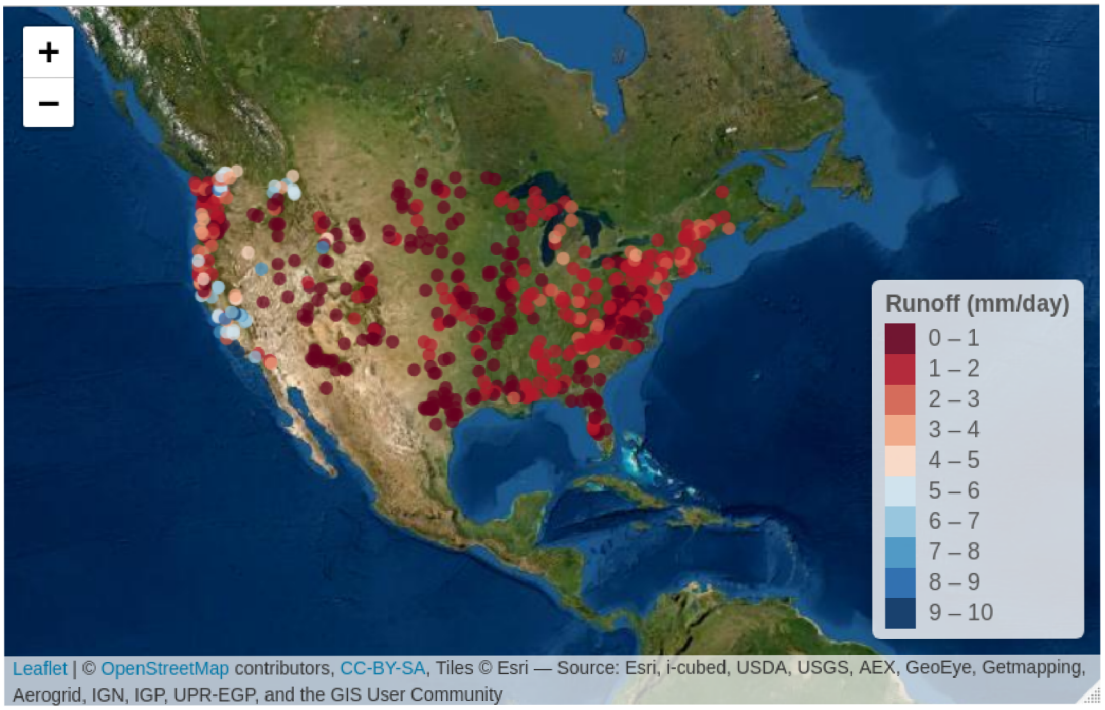

FIGURE 4.3: The runoff of the basins from the CAMELS dataset

CHAPTER 5

Methodology

In this section, each methodological step that was used for the purposes of the study will be described in details accordingly. The first section contains the description of the existing dHRUM model, along with all of its concepts described in detail. The second section contains the description of the different water balance structures which were implemented during this study. The third section contains detailed information about the calibration and evaluation steps using the CAMELS dataset, information about the basins that were chosen for this study, and the differential split sample test approach. The fourth section contains information about the three different objective functions used for the study. The fifth section contains information about the implemented data analysis and the wholesome assessment of the model.

5.1 Definition and description of dHRUM

As previously mentioned, for the purposes of this study the existing dHRUM model was used. Even though dHRUM stands for Distributed Hydrological Response Unit Model, this modelling framework can be used as a distributed, semi-distributed or lumped version, depending on the choice of the modeler. In this study, the lumped version was used. The distributed hydrological reponse unit model consists of six main storages: groundwater storage, soil water storage, surface retention storage, stem storage, snow storage and canopy storage. The models takes only the precipitation and the temperature as inputs, and gives us 22 outputs which are described in Table 5.1

The model also has 27 parameter inputs for calibration which are described in Table 5.2, of which 15 already existed in the original framework, and 12 more were added for the purposes of this study.

Table 5.1: Description of dHRUM outputs

	OUTPUTS	DEFINITION
1	Prec	Precipitation
2	Snow	Snow depth
3	PET	Potential Evapotranspiration
4	AET	Actual Evapotranspiration
5	Temp	Temperature
6	SteF	Stem flow
7	TroF	Through fall
8	SteS	Stem-storage
9	EvaC	Canopy Evaporation
10	EvaS	Stem-evaporation
11	EvbS	Bare soil Evapotranspiration
12	CanF	Canopy drainage
13	CanS	Canopy storage
14	IntS	Interception storage
15	GroS	Groundwater storage
16	SoiS	Soil storage
17	SurS	Surface retention
18	TotR	Total-runoff
19	DirR	Direct Runoff
20	Basf	Base-flow
21	Melt	Melting
22	Perc	Percolation

To get a better understanding of the structure of dHRUM, the following section will contain explanation of the water balance equations in each of the six storages: 3 inerception storages (canopy, stem and snow storage) and 3 linear storages for groundwater, soil and surface storage.

The dHRUM model follows the continuity equation for calculating the water balance:

$$\frac{dS}{dt} = I_{in} - O_{out} \tag{5.1}$$

where: $\frac{dS}{dt}$ is the change of state variable of particular accumulation space, I_{in} is the sum of input fluxes, O_{out} is the sum of output fluxes, The equation is dicretized on daily time step and solved using Euler method in each time-step. All units are in mm. The equation is upplied for updating all state variables: Snow, EvaC, SteS, SurS, SoiS, and GroS.

In the following section, different annotations are gonna be used for representing the state variables, each of them defined and described in details.

Potential Evapotranspiration

Table 5.2: Description of dHRUM input parameters

	INPUTS	DEFINITION
1	B_SOIL	Parameter controlling shape of Pareto distribution of soil storage [0.inf]
2	C_MAX	Max storage of storages distributed by Pareto distribution [0,inf]
3	B_EVAP	Parameter controlling soil evapotranspiration [0,inf]
4	SMAX	Max soil storage calculate using Cmax and b_soil
5	KS	Storage coefficient of groundwater storage [0,1]
6	KF	Storage coefficient of runoff response reservoirs [0,1]
7	ADIV	Divider of percolation into the direct runoff [0,1]
8	CDIV	Divider of gross rainfall as a Canopy input [0,1]
9	SDIV	Divider of gross rainfall as a Trunk input [0,1]
10	CAN_ST	The Max canopy storage [0,inf]
11	STEM_ST	The Max stem and trunk storage [0,inf]
12	CSDIV	The divider of canopy outflow to throughflow and stemflow storage [0,1]
13	TETR	The threshold temperature for determining snow [-inf,inf]
14	DDFA	The day degree model for snow melt [o, inf]
15	TMEL	The threshold temperature for determining melting process [-inf, inf]
16	RETCAP	The maximum capacity of surface retention [0, inf]
17	L	The amount of groundwater recharge removed from the linear reservoir [0,1]
18	D_BYPASS	The amount of groundwater recharge removed from the linear reservoir [0,1]
19	B_EXP	Power coefficient
20	KS2	Storage coefficient of groundwater storage [0,1]
21	THR	Threshold coefficient for threshold-controlled linear storage [0,inf]
22	ALPHA	Divider for two parallel linear reservoirs
23	CMIN	For pdm soil reservoir [0,inf]
24	FC	Field capacity [mm] [0,inf)
25	FOREST_FRACT	Forest fraction [0,1]
26	KF2	Storage coefficient of runoff response reservoirs [0,1]
27	KF_NONLIN	Runoff non-linearity parameter [-] [0,inf)

Before continuing to the water balance equations for each of the components, to get a better understanding of the model we will put focus on the calculation of the evapotranspiration in dHRUM. There are six methods for calculating the potential evapotranspiration: Oudin, Hamon, Thornthwaite, Blaney-Criddle, Jensen-Haise and McGuinnessBordne. For the purposes of this study, the Hamon potential evapotranspiration was used.

The Hamon potential evapotranspiration equations is implemented into dHRUM as:

$$PET = k(0.1651)(216.7)\frac{N}{12}\frac{e_s}{T + 273.3} \tag{5.2}$$

Description of Hamon PET formula:

PET - Potential Evapotranspiration $[\frac{mm}{day}]$

k - Coefficient of proportionality [-]

e_s - Saturation vapor pressure [millibars] which is equal to $6.108e^{\frac{17.27T}{T+273.3}}$

T - Monthly average temperature $[°C]$

N - Length of daytime calculated in units of 12 hours which is equal to $N = (\frac{24}{\pi})\omega$, where ω - sunset hour angle [radians]

$$\omega = \cos^{-1}[-\tan(\delta)\tan(\varphi)]$$

where, φ - the latitude [radians], δ - the declination [radians]

$$\delta = 0.409 \sin\left(\frac{2\pi}{365}\right) J - 1.39)$$

where, J is the Julian Day of the year.

A description of each of the storage dynamics of dHRUM will be described through the following subsections.

Canopy storage

The canopy storage is described with the following water balance equation:

$$\Delta W = CDIV(P + P_m) - E_c - R_c \tag{5.3}$$

where:
W - canopy intercepted water $[mm]$,
E_c - evaporation from canopy layer $[mm]$,
R_c - the overflow from canopy,
$CDIV$ - divider of gross rainfall as a canopy input,
P_m - snow melt $[mm]$,
P - precipitation $[mm]$.

Maximum evaporation from the canopy is produced when there is intercepted water, thus the maximum canopy evaporation E_c is calculated with:

$$E_c = \left(\frac{W_i}{W_{im}}\right)^{\frac{2}{3}} \tag{5.4}$$

where $W_{im} = CANST$ which is the maximum canopy storage defined within the interval [0,inf] $[mm]$, and the power of $\frac{2}{3}$ is described by Deardorff (1978). The canopy overflow is calculated as:

$$OF_{can} = CANS[i] - CAN_{ST} \tag{5.5}$$

where $CANS$ - the canopy storage, $CANST$ - the max canopy storage
Furthermore:

$$CanOut = \left(\frac{W_i}{W_{im}}\right) * E_c \tag{5.6}$$

so the total flow from canopy is the sum of $CanOut$ and OF_{can}.

Stem storage

The stem storage is described with the following water balance equation:

$$\Delta W = s(P + P_m) + (1 - c) * R_c - E_s - R_s \tag{5.7}$$

where:
W - stem intercepted water, $[mm]$
s - the divider of gross rainfall as a trunk input,
P - precipitation $[mm]$,
P_m - snow melt $[mm]$,
c - divider of gross rainfall as a Canopy input,
R_s - overflow from stem $[mm]$,
R_c - overflow from canopy $[mm]$,
E_s - evaporation from the stem layer $[mm]$.

The maximum stem evaporation (Es, mm) from each vegetation tile is calculated using the following formulation:

$$E_s = \left(\frac{W_i}{W_{im}}\right)^{\frac{2}{3}} \tag{5.8}$$

where $W_{im} = STEM_{ST}$ is the maximum stem storage defined within the interval [0,inf] in $[mm]$, and the power of $\frac{2}{3}$ is described by Deardorff (1978).
The stem overflow is calculated as:

$$StemOut = \left(\frac{W_i}{W_{im}}\right) * E_s \tag{5.9}$$

$$OF_{stem} = StemS[i] - STEM_{ST} \tag{5.10}$$

where $StemS$ is stem storage and $STEM_{ST}$ is the maximum stem storage. Then the total flow from the stem storage is the sum of $StemOut$ and OF_{stem}. Now, we get the total througflow from canopy and stem reservoirs in the following form:

$$TROF = cR_c + R_s \tag{5.11}$$

Snow storage
The snow storage is described with the following water balance equation:

$$\Delta S = P - R \tag{5.12}$$

where:
ΔS - the change in snow storage,
P - the precipitation as snow,
R - the snow melt calculated as:

$$MELT = DDFA * (TEMP - TMEL) \tag{5.13}$$

where $DDFA$ represents the day degree model for snow melt, $TEMP$ is the temperature at time t, $TMEL$ is the threshold temperature for determining melting processes.

Surface retention storage

The surface storage is described with the following water balance equation:

$$\Delta S = (1 - CDIV - SDIV)(P + P_m) + TROF - E_s - R \tag{5.14}$$

where:
ΔS is the change of the surface storage, $CDIV$ is the divider of gross rainfall as a canopy input within interval [0,1], $SDIV$ is the divider of gross rainfall as a trunk input within the interval [0,1], P is the precipitation, P_m is the precipitation as snow melt, E_s is the evaporation from the surface, $TROF$ is the through fall surface retention and R is the runoff. All units are calculated in $[mm]$. The actual evaporation is represented by the following equation:

$$AET = \left(\frac{W_i}{RETCAP}\right) * PET \tag{5.15}$$

where: W_i is the surface storage, $RETCAP$ is the maximum capacity of surface retention which is withing the interval [0, inf], PET is the potential evapotranspritaion calculated from Hamon method.

Soil water storage 1

The soil water storage is described with the following water balance equation:

$$\Delta S = P - E_s - R \tag{5.16}$$

where P - effective precipitation calculated from the surface storage, E_s - evaporation from the bare soil, R is the total overflow from the soil reservoir which is also referred to as percolation.

$$C = C_{max} * \left(1 - \left(1 - \frac{S1(t)}{S_{max}}\right)^{\frac{1}{B\_EVAP+1}}\right) \tag{5.17}$$

$$S_{max} = \frac{C_{max}}{b+1} \tag{5.18}$$

where C_{max} - the maximum storage capacity of the catchment, b - a dimensionless parameter, B_EVAP - parameter controlling soil evapotranspiration. So, the overflow from soil storage is calculated as:

$$OF1 = (C + PREF - C_{max}) \tag{5.19}$$

where C - critical storage capacity, $PREF$ - effective precipitation which is calculated from the surface reservoir, C_{max} - the maximum storage capacity of the catchment.

Furthermore, the infiltration is calculated as:

$$Infiltration = PREF - OF1 \tag{5.20}$$

The soil water depth is a sum of the $Infiltration$ and C. The soil buffer can be represented by the following formula (not affected by evapotranspiration)

$$SOIS = S_{max} * \left(1 - \left(1 - \frac{C}{C_{max}}\right)^{(B_{soil}+1)}\right) \tag{5.21}$$

where B_{soil} - a parameter controlling shape of Pareto distribution of soil storage within the interval [0,inf], C - the soil water depth, C_{max} - the maximum storage capacity, S_{max} - the maximum soil storage calculated using C_{max} and B_{soil}. The overflow can be calculated with the following equation:

$$OF2 = Infiltration - SOIS + SOIS(0) \tag{5.22}$$

The evaporation from the bare soil is calculated with the following equation:

$$E_{bs} = \left(1 - \left(\frac{Smax - SOIS}{Smax}\right)^{B_{evap}}\right) * PET \tag{5.23}$$

where $Smax$ is the maximum soil storage. Now, the critical storage capacity C*(t) can be calculated as the perculation (total overflow)

$$PERC = OF1 + OF2 \tag{5.24}$$

This soil water storage is defined as PDM which stands for Probability Distributed Model.

Groundwater storage 1

The groundwater storage is described with the following water balance equation:

$$\Delta S = (1 - ADIV)Perc - R \tag{5.25}$$

where:

ΔS - the change in the groundwater storage,

$Perc$ is the percolation from the groundwater reservoir,

R - the runoff as a base flow,

$ADIV$ - divider of percolation into the direct runoff and groundwater input. This groundwater storage is defined as LINRES which stands for linear reservoir.

5.2 Definition and description of the extended water-balance structures for groundwater and soil water storages in dHRUM

In the following section, the three additional soil water structures and the seven groundwater structures for describing the water balance will be described in details.

Groundwater storage 2: Linear reservoir with leakage (LINLRES)
The linear reservoir with leakage is described with the following water balance equation:

$$\Delta S = L + (1 - ADIV)Perc - R \tag{5.26}$$

where

$$R = S * KS \tag{5.27}$$

where:
ΔS - the change in the groundwater storage,
S - the current storage,
L - the leakage coefficient describing the amount of groundwater recharge removed from the linear reservoir in the interval [0,1],
$Perc$ - the percolation from the groundwater reservoir,
R - the runoff as a base flow, $ADIV$ - divider of percolation into the direct runoff and groundwater input,
KS - storage coefficient of groundwater storage.

Groundwater storage 3: Linear reservoir with a direct-by-pass (LINBYRES)
The linear reservoir with a direct-by-pass is described with the following water balance equation:

$$\Delta S = (1 - D\_BYPASS) * (1 - ADIV)Perc - R \tag{5.28}$$

where

$$R = S * KS + D\_BYPASS(1 - ADIV)Perc \tag{5.29}$$

where:
ΔS - the change in the groundwater storage,
S - the current storage,
D_BYPASS - the amount of groundwater recharge removed from the linear reservoir in the interval [0,1],
$Perc$ - the percolation from the groundwater reservoir,

R - the runoff as a base flow,
$ADIV$ - divider of percolation into the direct runoff and groundwater input, KS - storage coefficient of groundwater storage.

Groundwater storage 4: Two serial linear reservoirs (LIN2SE)
The two serial linear reservoirs are described with the following water balance equation:

$$\Delta S_1 = (1 - ADIV)Perc - R_1 \tag{5.30}$$

$$\Delta S_2 = R_1 - R_2 \tag{5.31}$$

where

$$R_1 = KsS1 \tag{5.32}$$

$$R_2 = Ks2S2, \quad \text{for} \quad Ks2 < Ks \tag{5.33}$$

where:
ΔS_1 - the change in the first serial linear reservoir,
ΔS_2 - the change in the second serial linear reservoir,
$S1$, $S2$ - the current storage in the first and second reservoir respectively,
$Perc$ - the percolation from both serial linear groundwater reservoirs,
R_1 - the runoff from the first serial linear reservoir as a base flow,
R_2 - the runoff from the second serial linear reservoir as a base flow,
$ADIV$ - divider of percolation into the direct runoff and groundwater input,
Ks - storage coefficient of groundwater storage in the first serial linear reservoir defined within the interval [0,1],
$Ks2$ - storage coefficient of groundwater storage in the second serial linear reservoir defined within the interval [0,1]

Groundwater storage 5: Two parallel linear reservoirs (LIN2PA)
The two parallel linear reservoirs are described with the following water balance equation:

$$\Delta S_1 = (1 - ADIV) * ALPHA * Perc - R_1 \tag{5.34}$$

$$\Delta S_2 = (1 - ADIV) * (1 - ALPHA) * Perc - R_2 \tag{5.35}$$

where,

$$R_1 = S1 * KS \tag{5.36}$$

$$R_2 = S2 * KS2 \tag{5.37}$$

$$R = R_1 + R_2 \tag{5.38}$$

where:
ΔS_1 - the change in the first parallel linear reservoir,
ΔS_2 - the change in the second parallel linear reservoir,
$S1$, $S2$ - the current storage in the first and second reservoir respectively,
$Perc$ - the percolation from both parallel linear groundwater reservoirs,
R_1 - the runoff from the first parallel linear reservoir as a base flow,
R_2 - the runoff from the second parallel linear reservoir as a base flow,
R - the total runoff from the groundwater reservoir,
$ADIV$ - divider of percolation into the direct runoff and groundwater input,
KS - storage coefficient of groundwater storage in the first parallel linear reservoir defined within the interval [0,1],
$KS2$ - storage coefficient of groundwater storage in the second parallel linear reservoir defined within the interval [0,1],
$ALPHA$ - the divider for two parallel linear reservoirs.

Groundwater storage 6: Non linear power law reservoir (POWRES)
The non linear power law reservoir is described with the following water balance equation:

$$\Delta S = (1 - ADIV) * Perc - R \tag{5.39}$$

$$R = S^{B\_EXP} * KS \tag{5.40}$$

where

$$\frac{1}{3} < B\_EXP < 1 \tag{5.41}$$

where:
ΔS - the change in the groundwater storage,
S - the current storage,
B_EXP - the power cofficient,
$Perc$ - the percolation from the groundwater reservoir,
R - the runoff as a base flow,
$ADIV$ - divider of percolation into the direct runoff and groundwater input,
KS - storage coefficient of groundwater storage.

Groundwater storage 7: Exponential reservoir (EXPRES)
The exponential reservoir is described with the following water balance equation:

$$\Delta S = (1 - ADIV) * Perc - R \tag{5.42}$$

$$R = KS * e^{\frac{S}{B\_EXP}} \tag{5.43}$$

where

$$B\_EXP \neq 0$$

where:
ΔS - the change in the groundwater storage,
S - the current storage,
B_EXP - the power cofficient,
$Perc$ - the percolation from the groundwater reservoir,
R - the runoff as a base flow,
$ADIV$ - divider of percolation into the direct runoff and groundwater input,
KS - storage coefficient of groundwater storage.

Groundwater storage 8: Linear reservoir with threshold controlled increased storage outflow (FLEXRES)
The FLEX reservoir is described with the following water balance equation:

$$\Delta S = (1 - ADIV) * Perc - R \tag{5.44}$$

$$R = \begin{cases} Ks\,S, & \text{for} \quad THR > S \\ \\ Ks2 * (S - THR) + Ks\,S & \text{for} \quad THR \leq S, \end{cases} \tag{5.45}$$

where:
ΔS - the change in the groundwater storage,
S - the current groundwater storage,
THR - the threshold cofficient within an interval [0,inf],
$Perc$ - the percolation from the groundwater reservoir,
R - the runoff as a base flow,
$ADIV$ - divider of percolation into the direct runoff and groundwater input,
Ks, $Ks2$ - storage coefficients of groundwater storage.

Soil water storage 2: Collie River Basin 2 (COLLIEV2)
This soil storage reservoir, known as Collie River Basin version 2, describes the bare soil evaporation and the vegetation evaporation. In our implementation,

they are distinct, but their sum is counted as the total evaporation. Also, in this model the total runoff is the sum of subsurface runoff, and the saturation excess from the surface runoff. The soil storage of type COLLIEV2 is represented by the following water balance equation:

$$\Delta S = P - E_b - E_v - R_{se} - R_{ss} \tag{5.46}$$

where

$$E_b = \frac{S}{SMAX} * (1 - FOREST\_FRACT) * PET \tag{5.47}$$

$$E_v = \begin{cases} FOREST\_FRACT * PET, & \text{if } S > FC \\ \frac{S}{FC} * FOREST\_FRACT * PET, & \text{otherwise} \end{cases} \tag{5.48}$$

$$R_{se} = \begin{cases} P, & \text{if } S > SMAX \\ 0, & \text{otherwise} \end{cases} \tag{5.49}$$

$$R_{ss} = \begin{cases} KF * (S - FC), & \text{if } S > FC \\ 0, & \text{otherwise} \end{cases} \tag{5.50}$$

where: ΔS - the change of soil storage,
P - effective precipitation,
E_b - evaporation from bare soil,
E_v - evaporation from vegetation,
R_{se} - saturation excess surface flow,
R_{ss} - subsurface runoff,
S - current soil storage,
$SMAX$ - maximum soil storage,
$FOREST\_FRACT$ - forest fraction defined within the interval [0,1],
FC - field capaity in mm within the interval [0,inf),
KF - storage coefficient of runoff response reservoir within the interval [0,1].

Soil water storage 3: New Zealand (NEWZEALAND)

This soil storage reservoir, known as New Zealand version 1, describes the bare soil evaporation and the vegetation evaporation. In our implementation, they are distinct, but their sum is counted as the total evaporation. Also, in this model the total runoff is the sum of subsurface runoff when soil moisture exceeds the field capacity, the saturation excess from the surface flow and the baseflow. The soil storage of type NEW_ZEALAND is represented by the following water balance equation:

$$\Delta S = P - E_b - E_v - R_{se} - R_{ss} - R_{bf}$$

where

$$E_{bs} = \frac{S}{SMAX} * (1 - FOREST_FRACT) * PET \tag{5.51}$$

$$E_v = \begin{cases} FOREST\_FRACT * PET, & \text{if } S > FC \\ \frac{S}{FC} * FOREST\_FRACT * PET, & \text{otherwise} \end{cases} \tag{5.52}$$

$$R_{se} = \begin{cases} P, & \text{if } S \geq SMAX \\ 0, & \text{otherwise} \end{cases} \tag{5.53}$$

$$R_{ss} = \begin{cases} (KF * (S - FC))^{KF\_NONLIN}, & \text{if } S \geq FC \\ 0, & \text{otherwise} \end{cases} \tag{5.54}$$

$$R_{bf} = KF2 * S \tag{5.55}$$

where:
ΔS - the change of soil storage,
P - effective precipitation,
E_b - evaporation from bare soil,
E_v - evaporation from vegetation,
R_{se} - saturation excess surface flow,
R_{ss} - subsurface runoff,
R_{bf} - baseflow,
S - current soil storage,
$SMAX$ - maximum soil storage,
$FOREST\_FRACT$ - forest fraction defined within the interval [0,1],
FC - field capaity in mm within the interval [0,inf),
KF - storage coefficient of runoff response reservoir within the interval [0,1].
KF_NONLIN - runoff non-linearity parameter [-] within the interval [0,inf),
$KF2$ - storage coefficient of runoff response within [0,1]

Soil water storage 4: a daily four-parameter rainfall-runoff model (GR4J)

This soil storage reservoir, known as GR4J, has two stores with four parameters. The reservoir works with an explicit time-step. The following equations are used for its implementation. The original model in Knoben et al. (2019a) have implemented equations from Santos et al. (2018) but have used the original unit hydrograph by Perrin et al. (2003). Here, we only use one store with one parameter, representing only the soil reservoir of the model.

The soil storage of type GR4J is represented by the following water balance

equation:

$$\Delta S = P_s - E_s - Perc \tag{5.56}$$

where

$$P_s = P_n\left(1 - \left(\frac{S}{SMAX}\right)^2\right) \tag{5.57}$$

where

$$P_n = \begin{cases} P - PET, & \text{if } P \geq PET \\ 0, & \text{otherwise} \end{cases} \tag{5.58}$$

$$E_s = E_n\left(2\frac{S}{SMAX} - \left(\frac{S}{SMAX}\right)^2\right) \tag{5.59}$$

where

$$E_n = \begin{cases} PET - P, & \text{if } PET > P \\ 0, & \text{otherwise} \end{cases} \tag{5.60}$$

$$Perc = \frac{SMAX^{-4}}{4} E_n * \left(\frac{4}{9}\right)^{-4} S^5 \tag{5.61}$$

where:
ΔS - the change of soil storage,
P_s - fraction of P_n, the net precipitation, diverted to soil moisture,
E_s - the fraction of E_n, the net evaporation, taken away from the soil moisture,
$Perc$ - percolation,
S - current soil storage,
$SMAX$ - maximum soil storage,
PET - potential evapotranspiration,
P - effective precipitation.

It is very important to note that during the calculation of evaporation and runoff in each time step, for each of the groundwater and soil water reservoir, a water balance equation was implemented. The water balance equation is a simple equation where we make sure that we are updating the storage correctly whenever we have different types of evaporations (vegetation or bare soil), or different types of runoff (surface, subsurface, baseflow), making sure our storage never becomes negative. If the amount of bare soil evaporation and vegetation evaporation are smaller than the current available storage, then we subtract them first, and then we proceed with calculating whether the subsurface, surface runoff and baseflow are smaller than the current available storage, and if they are, they are subtracted respectively from the current storage.

5.3 Usage of the differential split sample test

For accessing and testing the performance of the dHRUM modelling framework, the CAMELS dataset was used. As prevoously described in the study area section, only 620 out of 671 basins were included for the study, and the nldas forcing data was used. The decision to exclude 51 basins lies on the fact that many of the basins has missing or incomplete data. After the whole dataset was read in R, both the forcing data and the observations, a data analysis was performed for each of the basins during the time period between 01.01.1980 and 31.12.2014. While processing the data for each basin and each year, additional columns were added to a new table indicating: a negative discharge, a discharge with an NA value, missing dates (in such a way that the forcing data dates could not overlap with the dates of the observations). After processing this information, the results were summarized indicating the basins where there were more than 5 years of incomplete data. It was noticeable that for the year 2014, 411 out of 671 basins had missing data, so this year was excluded for all basins. Then, 51 basins in total had incomplete data for more than 4 years, so those basins were excluded from the study. The basins that were used for this study have no more than 4 years of missing data, which left us with at least 25 years and at most 30 years for calibration and validation, of which always the calibration was applied on a 10 year consecutive period, and the validation on the remaining years.

The differential split sample test was used as described in the literature review for choosing the data for calibration and the data for validation. A simple alghorithm was implemented to find the 10 consecutive years of each of the basins where the average annual precipitation was the lowest from all possible 10 consecutive years period given. Those data were used for calibrating the model on the dry period. The same alghorithm was used for finding the 10 consecutive years period for each of the basins where the annual average precipitation was the highest, and those data were used for calibrating the model on the wet period.

In conclusion, this whole study is performed separately on the data where the calibration was done for the dry period and on the data where the calibration was done on the wet period. By using this differential split sample test for calibration on dry and on wet period, we want to access the difference of the performance of the model depending on this factor. More information about the calibration and validation can be found in the next secion.

5.4 Calibration and validation of the model

In this study, 32 model structures were used for streamflow simulations for calibrating and evaluating the performances dHRUM. The study was done on 620 basins from the CAMELS dataset. The calibration was performed on a 10 year consecutive period for each of the 620 basins from the CAMELS dataset where the warming period was set to be the first year of each of the periods. The validation was performed on the remaining years as explained in the previous section and in the study area section. Three single-objective functions were used: KGE (Kling-Gupta Efficiency), NSE (Nash-Sutcliffe Efficiency) and MAE (Mean Absolute Error). Three goodness-of-fit criteria values were produced using the gof(goodness of fit) R library which is a function that returns a goodness-of-fit measures between the simulations and the observations, of which three values were chosen as such: KGE, NSE, MAE. In this section, the three objective functions will be furtherly described.

The goodness-of-fit of a parameter set depends on the different objective functions chosen when calibrating the model. This means that a parameter set can be a bad fit for the same objective function which can be a good fit for another parameter set, and vice versa.

Kling-Gupta efficiency (KGE)

The Kling-Gupta efficiency (KGE) was developed by Gupta et al. (2009). KGE is an objective function which in a balanced way combines three components of NSE of model errors more specifically the following three: bias, correlation, coefficients of variation. Those are correlation, bias, the ratio of variances). This objective function has become very popular among hydrologists through the years. It is represented by the following formula:

$$KGE = 1 - ED_s$$

$$ED_s = \sqrt{(s_1(r-1))^2 + (s_2(vr-1))^2 + (s_3(\beta-1))^2)}$$

$$\beta = u_s/u_o$$

$$\alpha = \sigma_s/\sigma_o$$

where:

ED_s - Euclidean distance in scaled space

r - the Pearson product-moment correlation coefficient

β - the ratio between the means of the simulated values and observed values

vr - variability ratio

α - ratio between the standard deviations of the simulated and observed values

γ - ratio between the coefficient of variations of simulated and observed values.

KGE values that are greater than -0.41 mean that the model is improving over the average flow reference value, even if it is negative Knoben et al. (2019b).

Nash-Sutcliffe efficiency (NSE)

Nash-Sutcliff efficiency (NSE) is a normalized statistic that calculates magnitude residual variance versus variance of measured data Nash and Sutcliffe (1970) Also an indicator for show how well the observed versus simulated data fit. While NSE = 1, it corresponds to a model that perfectly corresponds to the observed data, NSE = 0, corresponds to the predictions of the model as accurate as the average values of the observed data. To calibrate the NSE minimization model to 0 is used as the main fitness function for calibration, which is formulated below.

Nash-Sutcliffe Efficiency (NSE) is a normalized statistic that calculates the magnitude of the residual variance versus the variance of the measured data Nash and Sutcliffe (1970), and it is also an indicator of how well the observations fit the simulations. When NSE = 1, it means that the observations perfectlt fit the simulations, while when NSE = 0 means that the predictions of the model are accurate as the mean value of the observed data.

$$NSE = 1 - \frac{\sum_{n=1}^{M}(obs - sim)^2}{\sum_{n=1}^{M}(obs - mean(obs))^2}$$

where:

M - number of observation,

obs - observed values

sim - simulated values

Mean Absolute Error (MAE)

The mean absolute error (MAE) calculates the average absolute value of residual errors between the simulated and the observed flows. The closer it is to 0, the better it is.

$$MAE(\theta) = \frac{1}{N}\sum_{k=1}^{N}|y_k - y'_k(\theta)|$$

where:

y_k - the observed flow at time k

$y'_k(\theta)$ - the flow at time k estimated with the set of parameters θ

N - the number of time steps in the event.

MAE takes into account the accuracy of the simulations at for low flows Mediero et al. (2011).

5.5 Optimization Algorithm

The Differential Evolution algorithm was used in this study as an optimization algorithm for minimization of the objective functions. This algorithm is introduced by Storn and Price (1997). The Differential Evolution algorithm consists of three steps:

- 1. create a population of N in an m-dimensional space, randomly distributed.
- 2. replace the current population with a better fit new population
- 3. repeat step 2 until satisfactory results are obtained.

This algorithm had been explored and used my many as an algorithm for solving global optimization problems Ardia et al. (2011). The Differential Evolution algorithm (DE) in this study was used as a pre-build package in R Mullen et al. (2011). The optimization algorithms purpose is to find the best model parameter values based on minimizing or maximizing the objective function. In this study minimizing objective functions are measured.

CHAPTER 6

Results and Discussion

In this study, exploratory data analysis was performed. The aim of the exploratory data analysis was to try to understand the results in order to be able to describe the strengths and weaknesses of the models and propose next steps. The results from the three goodness-of-fit criteria generated with the gof package both for calibration and validation are represented in the following 6 rows: calib_KGE, valid_KGE, calib_NSE, valid_NSE, calib_MAE, valid_MAE.

6.1 Calibration on dry period, validation on wet period

The first step that needed to be done was to find the incomplete or noisy data, so the number of occurences of NA values were accessed for each of the resulting columns. After the resulting dataset from the calibration done on the dry period was processed in R and the is.na() function was applied to each column separately, NA values were noticed in the calib_KGE and valid_KGE columns. For calib_KGE there were 40 NA values out of 59520 values in total, and valid_KGE there were 35 NA values out of 59520 values in total. These values can be removed because they only represent 0,0013% of the dataset in the worst case scenario (if 75 rows in total were to be removed). After removing the NA values, the resulting dataset was left with 59446 records.

Next step was to understand the minimum and maximum values of each of the goodness-of-fit criteria. The decision behind this step was to understand the extreme high and low values that the model had produced, and where exactly it produced them.

It was observed that the maximum values for KGE and NSE both for calibration and validation were < 1, which means that there were no extreme high values

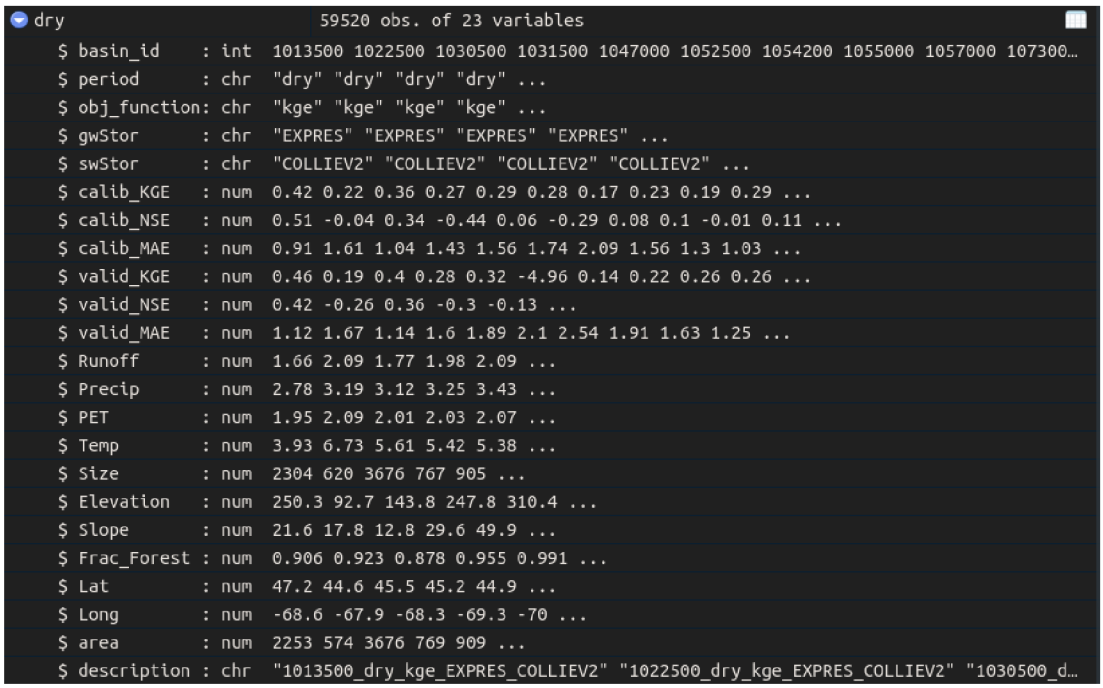

FIGURE 6.1: Overview of the results

Table 6.1: Min and max values from results on dry period

	min	max
calibKGE	-17.09	0.88
validKGE	-7.84	0.84
calibNSE	-347.41	0.84
validNSE	-6.12e+129	0.71
calibMAE	0.01	10.19
validMAE	0.01	3.45e+62

here. On the other hand, extreme low values exist, especially for NSE. In contrary, MAE has 0.01 as the minimum value, but very high maximum values.

To get a better overview of the distribution of each of these results, it has to be considered the total number of resulting records which is 59446. If these values were to be plotted, the distribution of the data would be distorted, so in order to be able to better understand the distribution of the data, the log() function was used. After using the log() function, almost half of the data was removed due to 'non-finite values'. So, it can be concluded that plotting the individual distributions of the results in a form of a histogram is not a very elegant solution while having so many values and outliers. The boxplot function could be used for visualizing the differences between the samples, the ranges, the medians and the outliers. In order to be able to proceed with the analysis, the boxplot() function was be used for better understanding of the results, thus setting the direction of the further analysis.

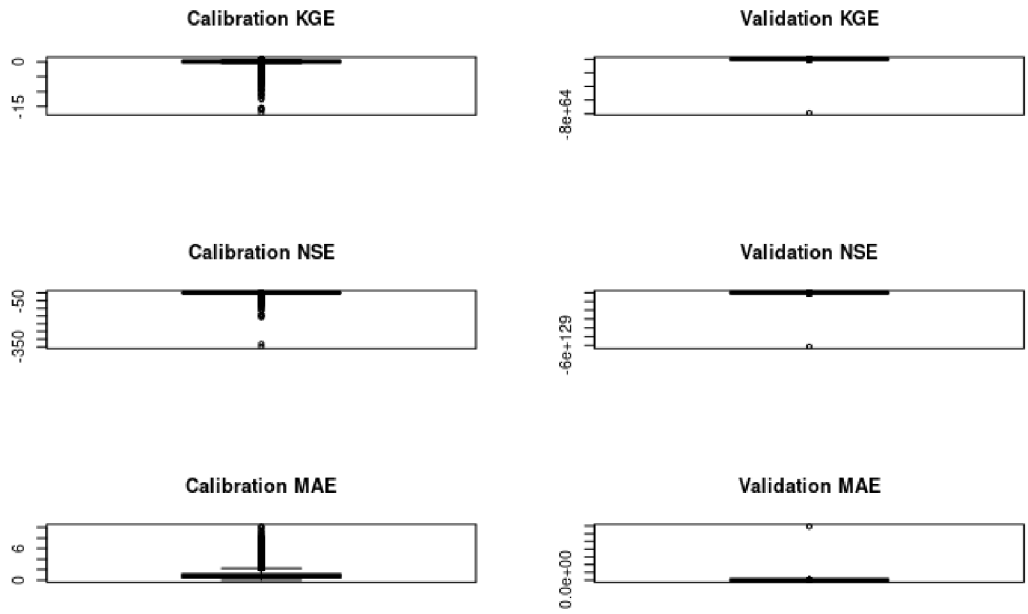

FIGURE 6.2: Boxplot results from goodness-of-fit values when calibration is done on dry period

The plotted outliers in Figure 6.2 in the case of KGE and NSE goodness-of-fit criteria are negative values. The outliers for calibration and validation from MAE goodness-of-fit criteria are positive values. There are 1265 outliers from the results from calibration on dry period produced from KGE as a goodness-of-fit criteria, 5171 outliers from the calibration done on the dry period produced from NSE as a goodness-of-fit criteria, and 4485 outliers from the calibration done on the dry period produced from MAE as a goodness-of-fit criteria, when using the boxplot() function for indicating the outliers. Even though this is a very small percentage of the resulting dataset, these ourliers contain important information about the model, so they will not be ommited from the analysis, but will be extracted later in order to gain some insights about when and where they occur.

Next step was to individually examine the results from the three different objective functions: KGE, NSE, MAE, where for each of them the model produced three goodness-of-fit criteria both for calibration and validation.

The KGE values in this study are considered as 'acceptable' values if they are within the ranges -0.41 to 1. There are many discussion about which values are considered as good values for KGE, in this study -0.41 is taken as a good value because it means that the model is better or improves upon the mean flow benchmark. The NSE values in this study are considered as acceptable values if they are within the ranges 0 to 1, where if they are smaller than 0.5 than they are considered as unsatisfactory values, if they are between 0.5 and 0.7 then they are

considered as satisfactory values, while the values between 0.7 and 1 are considered as good values. For MAE, the values which are considered as acceptable are values between 0 and 1, but what is important here to mention is that the values that are considered as good values that the model produced in this study don't have a fix range. Instead, for each of the basins, the MAE threshold is calculated from the observables taken as a quantile function with 0.2 quantile, and accordingly to each basin a different threshold value was assigned.
The exact amount of data which is within these acceptible values is still unknown, so the next step of the analysis was to see how much of the resulting data is within the acceptible range, to try and find a connection between the different structures tested with dHRUM depending on the previous conditions.

For KGE goodness-of-fit criteria, there are 56485 values out of all resulting values that are within the acceptible range. Respectively, for NSE there are 31252 which is almost 60% of the data within 0 and 1, and for MAE there are 1082 values whose MAE goodness-of-fit value was approximating the mae threshold for each basin at most for 0.5 and at least 0, the acceptible range, where 0 indicated the perfect fit. Now, a further analysis can be done on each of the calibrations separately for each objective function in order to try to better understand the results.

Summary of results (calibration on dry period)

In Figure 6.3 the results are grouped by their objective function and goodness-of-fit criteria, summarized by the median. When KGE is used as an objective function, best goodness-of-fit results are produced for KGE both for calibration and validation. However, NSE and MAE also produce satisfactory goodness-of-fit values when used KGE as an objectve function. When NSE is used as an objective function, results are better with NSE goodness-of-fit criteria, but also with KGE goodness-of-fit criteria. However, these results are not satisfactory since a good NSE value is considered above 0.5. When using the MAE as an objective function, the results are also not satisfactory, because the expected median threshold should be around 0.2, and the values produced here are above 0.5. A lot of outliers can be noticed for validation done with NSE objective function, and validation done with MAE objective function. The summary of the results from the calibration done on dry period can be found in Table 6.2

Next step was to examine the influence from the different groundwater and soil water structures used in the study. The results were grouped by the groundwater storage, soil water storage and the used objective function, summarized by their median value. There are 96 result rows in total (3 objective functions, 4 soil

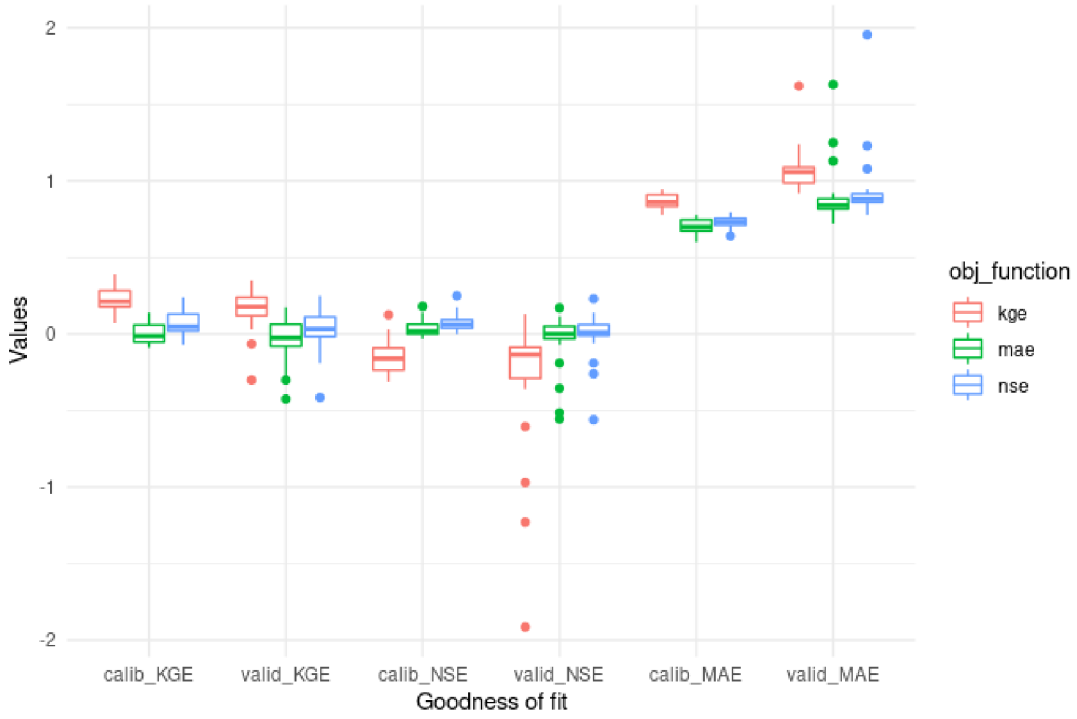

FIGURE 6.3: Boxplot results from dry period

Table 6.2: Summary of results (median) for calibration done on dry period

obj	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
kge	0.21	0.17	-0.16	-0.13	0.86	1.06
nse	0.04	0.03	0.06	-0.01	0.73	0.88
mae	-0.01	-0.02	0.01	0.0	0.7	0.84

water structures and 8 groundwater structures). The results done with KGE objective function are most satisfactory out of all three objective functions. Most dominant groundwater structures from the results with the KGE goodness-of-fit criteria (Table 6.3) which are satisfactory include the FLEXRES, LINRES and LINBYRES, while for soil are PDM, COLLIEV2 and NEWZEALAND. When KGE is used as an objective function, the NSE and MAE goodness-of-fit criteria values are not satisfactory.

From Table 6.4 and Table 6.5 it can be concluded, that the results are not satisfactory enough neither when NSE is used as an objective function, nor when MAE is used as an objective function. From 6.1 it is obvious that the model produced satisfactory results both for NSE and for MAE. The reason why this grouping is not showing good results for NSE and MAE is because of the many outliers that exist there as seen from Figure 6.3. Also, some other external factors may influence this, for an example the basin's physical characteristics. In order to better understand the results, they are plotted on a map, the goodness-of-fit results from the calibration for each of the objective functions (Figure 6.4). The

Table 6.3: Calibration done on dry period with KGE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
EXPRES	COLLIEV2	0.210	0.090	-0.295	-0.605	0.935	1.240
EXPRES	GR4J	0.080	-0.300	-0.310	-1.915	0.845	1.240
EXPRES	NEWZEALAND	0.200	-0.065	-0.285	-1.230	0.900	1.620
EXPRES	PDM	0.195	0.030	-0.170	-0.970	0.810	1.100
FLEXRES	COLLIEV2	0.340	0.280	-0.225	-0.255	0.900	1.090
FLEXRES	GR4J	0.280	0.230	-0.190	-0.210	0.855	1.060
FLEXRES	NEWZEALAND	0.320	0.240	-0.270	-0.360	0.870	1.065
FLEXRES	PDM	0.390	0.350	0.025	-0.040	0.780	0.945
LIN2PA	COLLIEV2	0.200	0.160	-0.130	-0.120	0.920	1.060
LIN2PA	GR4J	0.080	0.060	-0.160	-0.130	0.845	1.015
LIN2PA	NEWZEALAND	0.230	0.180	-0.270	-0.290	0.910	1.080
LIN2PA	PDM	0.170	0.160	-0.010	0.000	0.835	0.960
LIN2SE	COLLIEV2	0.200	0.170	-0.140	-0.120	0.920	1.090
LIN2SE	GR4J	0.070	0.050	-0.160	-0.140	0.860	1.000
LIN2SE	NEWZEALAND	0.230	0.180	-0.300	-0.300	0.915	1.090
LIN2SE	PDM	0.170	0.170	0.005	0.000	0.830	0.980
LINBYRES	COLLIEV2	0.280	0.240	-0.070	-0.080	0.880	1.040
LINBYRES	GR4J	0.150	0.130	-0.100	-0.090	0.810	0.980
LINBYRES	NEWZEALAND	0.320	0.270	-0.100	-0.140	0.870	1.040
LINBYRES	PDM	0.290	0.290	0.125	0.130	0.785	0.920
LINLRES	COLLIEV2	0.220	0.180	-0.160	-0.150	0.945	1.090
LINLRES	GR4J	0.100	0.080	-0.170	-0.160	0.860	1.010
LINLRES	NEWZEALAND	0.250	0.185	-0.290	-0.330	0.920	1.100
LINLRES	PDM	0.215	0.200	-0.010	-0.010	0.830	0.990
LINRES	COLLIEV2	0.310	0.260	-0.110	-0.115	0.895	1.070
LINRES	GR4J	0.190	0.160	-0.120	-0.110	0.860	1.000
LINRES	NEWZEALAND	0.300	0.240	-0.190	-0.230	0.850	1.090
LINRES	PDM	0.320	0.290	0.030	0.010	0.820	0.970
POWRES	COLLIEV2	0.200	0.170	-0.130	-0.100	0.920	1.070
POWRES	GR4J	0.080	0.060	-0.170	-0.130	0.845	0.970
POWRES	NEWZEALAND	0.240	0.180	-0.280	-0.290	0.910	1.095
POWRES	PDM	0.180	0.175	-0.020	-0.010	0.830	0.980

MAE goodness-of-fit values were calculated depending on each basin's threshold value, the MAE values was substituted from the threshold value for each basin, so a value which is closer to 0 but positive indicates that the value is not bigger than the threshold value and the closer it is to 0, the better the result. As seen from the map, such values are colored with a purple color, in the up right corner. The NSE values which are considered as satisfactory, are also colored with purple and pink color, very close to the border on the bottom right. In order to find the structures which dominate when the model performs good (but also when the model is not satisfactory enough), the appropriate ranges will be applied and the values within those ranges will be extracted from the results, in a way that the 'winning' and the 'bad' structures shall be discussed.

For NSE we consider values to be satisfactory if they are bigger than 0.5. This will be applied only to the calibration period for the sake of simplicity since the calibration and the validation within the appropriate objective functions are very similar. The total number of NSE values which satisfy the condition $>= 0.5$ are 1475 in total. The number of occurences for the objective functions are the following: NSE goodness-of-fit value with 611 values, KGE wth 345 and MAE with 519. The dominant groundwater structures are: LINBYRES with 335 values,

Table 6.4: Calibration done on dry period with NSE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
EXPRES	COLLIEV2	-0.010	-0.130	0.000	-0.190	0.795	1.230
EXPRES	GR4J	-0.070	-0.190	0.000	-0.260	0.760	1.080
EXPRES	NEWZEALAND	0.030	-0.415	0.030	-0.560	0.755	1.955
EXPRES	PDM	0.070	-0.015	0.070	-0.060	0.710	0.940
FLEXRES	COLLIEV2	0.170	0.150	0.060	0.020	0.740	0.915
FLEXRES	GR4J	0.080	0.090	0.060	0.020	0.700	0.870
FLEXRES	NEWZEALAND	0.160	0.125	0.050	-0.050	0.750	0.945
FLEXRES	PDM	0.240	0.250	0.175	0.120	0.650	0.795
LIN2PA	COLLIEV2	0.020	-0.010	0.040	0.010	0.770	0.920
LIN2PA	GR4J	-0.060	-0.070	0.010	-0.010	0.770	0.910
LIN2PA	NEWZEALAND	0.040	0.010	0.050	-0.010	0.740	0.880
LIN2PA	PDM	0.030	0.040	0.080	0.060	0.735	0.865
LIN2SE	COLLIEV2	0.015	-0.010	0.030	0.010	0.780	0.900
LIN2SE	GR4J	-0.040	-0.055	0.010	-0.010	0.755	0.890
LIN2SE	NEWZEALAND	0.030	0.010	0.060	-0.010	0.720	0.885
LIN2SE	PDM	0.060	0.050	0.090	0.060	0.730	0.870
LINBYRES	COLLIEV2	0.130	0.115	0.120	0.070	0.720	0.860
LINBYRES	GR4J	0.030	0.040	0.070	0.030	0.720	0.860
LINBYRES	NEWZEALAND	0.150	0.130	0.115	0.040	0.700	0.850
LINBYRES	PDM	0.180	0.180	0.250	0.230	0.695	0.815
LINLRES	COLLIEV2	0.050	0.020	0.040	0.010	0.775	0.920
LINLRES	GR4J	-0.040	-0.050	0.010	-0.010	0.755	0.880
LINLRES	NEWZEALAND	0.060	0.025	0.060	-0.010	0.740	0.875
LINLRES	PDM	0.060	0.070	0.100	0.080	0.730	0.860
LINRES	COLLIEV2	0.150	0.120	0.100	0.070	0.710	0.850
LINRES	GR4J	0.050	0.040	0.070	0.040	0.680	0.830
LINRES	NEWZEALAND	0.130	0.110	0.100	0.050	0.710	0.880
LINRES	PDM	0.180	0.180	0.170	0.140	0.640	0.780
POWRES	COLLIEV2	0.020	-0.020	0.040	0.010	0.770	0.920
POWRES	GR4J	-0.050	-0.060	0.010	-0.010	0.750	0.890
POWRES	NEWZEALAND	0.045	0.010	0.050	-0.010	0.720	0.880
POWRES	PDM	0.040	0.040	0.090	0.060	0.730	0.870

LINRES with 328 values, FLEXRES with 282 values, and a smaller portion for the rest of the groundwater structures. The dominant soil water structures are PDM with 660 values and COLLIEV2 with 333 values. Accoridngly, the same analysis for the number of occurences will be applied to the 'bad' structures, with one difference only: the values which are smaller than 0 will be taken into consideration, because those are the values which are outliers for NSE goodness-of-fit measure. For KGE as an objective functions, there are 13334 values which are smaller or equal to 0, after which follows the MAE objective function with 8370 values, and then NSE objective function with 6490. The dominant underperforming groundwater structures are EXPRES, POWRES, LIN2PA and LIN2SE with 4397, 3739, 3707 and 3714 accorgingly. The dominant underperforming soil water structures are GR4J with 8065, COLLIEV2 with 7514 and NEWZEALAND with 7674 values.

For MAE we consider values to be satisfactory if they are positive and bigger than 0, where 0 indicates that the calculated MAE goodness-of-fit value equals the threshold. The values between 0 and 0.5 were taken as such, in total 1082 values. Regarding the dominance of the objective functions, best performing is MAE with total of 438 values, followed by NSE 389 values and the rest with KGE. The dominant satisfactory groundwater structures are LINRES, FLEXRES,

Table 6.5: Calibration done on dry period with MAE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
EXPRES	COLLIEV2	-0.030	-0.300	-0.030	-0.555	0.780	1.250
EXPRES	GR4J	-0.070	-0.290	-0.020	-0.515	0.750	1.130
EXPRES	NEWZEALAND	-0.020	-0.425	-0.010	-0.355	0.710	1.630
EXPRES	PDM	0.030	-0.080	0.040	-0.190	0.695	0.920
FLEXRES	COLLIEV2	0.110	0.095	0.025	0.000	0.720	0.880
FLEXRES	GR4J	0.030	0.040	0.030	0.000	0.700	0.850
FLEXRES	NEWZEALAND	0.110	0.095	0.000	-0.070	0.690	0.880
FLEXRES	PDM	0.140	0.175	0.140	0.110	0.610	0.750
LIN2PA	COLLIEV2	-0.070	-0.090	0.010	0.000	0.760	0.890
LIN2PA	GR4J	-0.090	-0.085	0.000	-0.010	0.740	0.875
LIN2PA	NEWZEALAND	-0.020	-0.040	0.010	-0.030	0.670	0.810
LIN2PA	PDM	-0.010	-0.010	0.060	0.050	0.720	0.840
LIN2SE	COLLIEV2	-0.055	-0.075	0.010	0.000	0.760	0.890
LIN2SE	GR4J	-0.090	-0.090	0.000	-0.010	0.750	0.880
LIN2SE	NEWZEALAND	-0.010	-0.030	0.010	-0.030	0.670	0.810
LIN2SE	PDM	-0.010	-0.010	0.070	0.050	0.710	0.830
LINBYRES	COLLIEV2	0.080	0.070	0.080	0.060	0.700	0.835
LINBYRES	GR4J	-0.015	-0.010	0.040	0.030	0.700	0.820
LINBYRES	NEWZEALAND	0.100	0.070	0.060	0.010	0.645	0.785
LINBYRES	PDM	0.100	0.110	0.180	0.170	0.640	0.760
LINLRES	COLLIEV2	-0.040	-0.060	0.010	0.000	0.750	0.890
LINLRES	GR4J	-0.070	-0.060	0.000	-0.010	0.745	0.865
LINLRES	NEWZEALAND	-0.010	-0.030	0.000	-0.040	0.680	0.820
LINLRES	PDM	0.030	0.030	0.080	0.065	0.700	0.830
LINRES	COLLIEV2	0.050	0.060	0.070	0.060	0.655	0.800
LINRES	GR4J	-0.015	-0.020	0.040	0.030	0.675	0.810
LINRES	NEWZEALAND	0.100	0.080	0.050	0.020	0.660	0.840
LINRES	PDM	0.080	0.075	0.120	0.115	0.600	0.720
POWRES	COLLIEV2	-0.070	-0.080	0.010	0.000	0.750	0.890
POWRES	GR4J	-0.090	-0.080	0.000	-0.010	0.745	0.875
POWRES	NEWZEALAND	-0.050	-0.050	0.005	-0.030	0.685	0.825
POWRES	PDM	-0.020	-0.020	0.070	0.050	0.720	0.830

LINBYRES. The dominant soil water structures are NEWZEALAND and PDM. Accordingly, the same analysis could be applied for the 'bad' structures, which in this case we consider everything below 0 or greater than 0.5, in total 58364 values, which is 98% of our dataset.

However, this analysis of winning and lossing structures when NSE and MAE are used as objective functions cannot be taken for granted and considered as an appropriate analysis because the results for NSE and MAE goodness-of-fit values are not satisfactory enough. In both cases, around 95% of the resulting dataset has unsatisfactory goodness-of-fit measures.

6.2 Calibration on wet period, validation on dry period

The same analysis was applied for the calibration performed on the wet period, while the validation is performed on the dry period. First, the data from the calibration on the wet period is processed in R. There are 9 NA values for the KGE goodness-of-fit criteria on the calibration period, while 68 NA values for the KGE

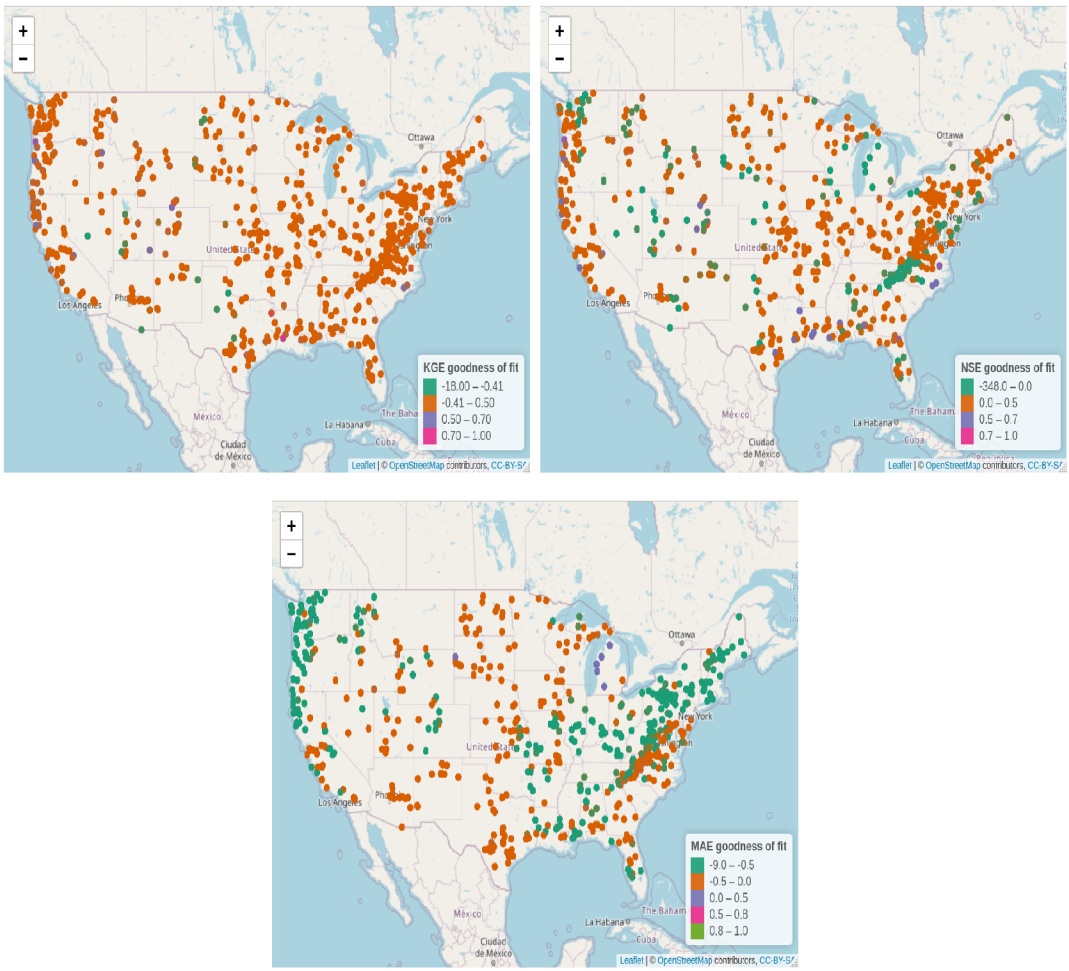

FIGURE 6.4: Distribution for goodness-of-fit values for calibration done with KGE, NSE, MAE objective functions on dry period

goodness-of-fit criteria on the validation period. As such, these values were removed from the resulting dataset, which left the dataset with 59451 records in total.

Summary of results (calibration on wet period)

The minimum and maximum values for each of the goodness-of-fit values is used in order to access the extreme high and low values. The result is represented in Table 6.6. It's noticeable that the results for the validation on the dry period, are better than the results for the validation on the wet period (Table 6.1), but on the other hand, the minimum values are lower or the same.

In this case also, there are extreme low values for KGE and NSE, while extremely high values for MAE. This can be further seen in the boxplot in Figure 6.5

In Figure 6.6 the results are grouped by their objective function and goodness-of-fit criteria, summarized by the median. When KGE is used as an objective

Table 6.6: Min and max values from results on wet period

	min	max
calibKGE	-43.22	0.89
validKGE	-1.187432e+85	0.82
calibNSE	-1876.23	0.85
validNSE	-1.386362e+170	0.73
calibMAE	0.01	8.93
validMAE	0.01	1.363783e+83

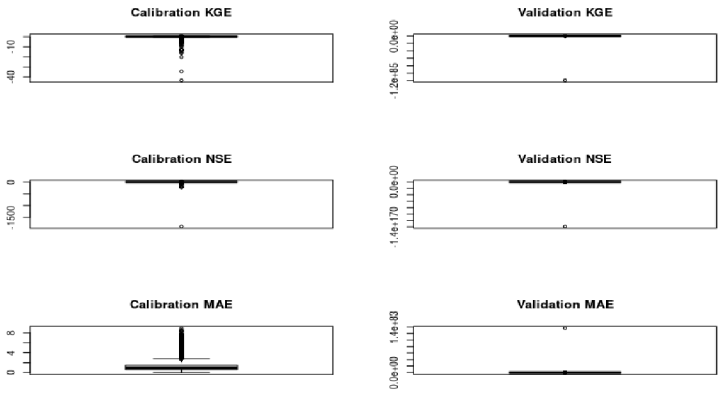

FIGURE 6.5: Boxplot results from goodness-of-fit values when calibration is done on wet period

function, best goodness-of-fit results are produced for KGE both for calibration and validation. Simimlar results where produced when calibration was done on the dry period: NSE and MAE also produce satisfactory goodness-of-fit values when used KGE as an objectve function, but the goodness-of-fit values when NSE or MAE is used as an objective function seem to be unsatisfactory. This will be again the subject of the analysis of the second part of the study.

Table 6.7: Summary of results (median) for calibration done on wet period

obj	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
kge	0.21	0.18	-0.14	-0.16	1.12	0.97
nse	0.045	0.035	0.04	0.015	0.93	0.8
mae	-0.005	-0.02	0.005	0.005	0.88	0.76

The summary of the results from the calibration done on the wet period can be found in Table 6.7. The combination of different groundwater and soil water structures along with the three different objective functions is the next subject of this analysis. As was done for the calibration on the dry period, the same summarization will be applied on the results produced on the calibration for the wet period. 96 results in total, indicating 3 objective functions, 8 groundwater structures and 4 soil water structures.

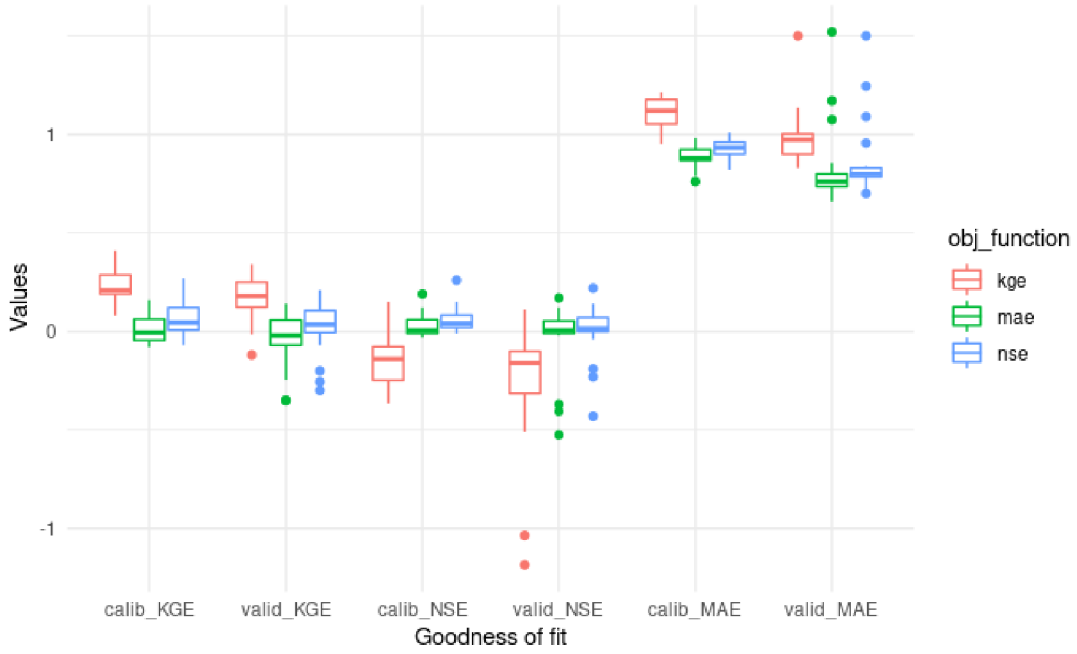

FIGURE 6.6: Boxplot results from wet period

In Table 6.8 the summarization by the median is performed on the different storages when used KGE as an objective function. It is noticeable that the KGE goodness-of-fit criteria values are higher than the ones from Table 6.3. For the combination of PDM as a soil water storage, along with the FLEXRES as a groundwater storage, the goodness-of-fit for KGE on the calibration is 0.41, while for validation is 0.34. Taken in consideration the huge amount of outliers, these results are satisfactory enough. The structures which have the highest KGE goodness-of-fit values are for groundwater: FLEXRES, LINBYRES, LINRES while for soil storage are PDM, NEWZEALAND and COLLIEV2. The same structures showed satisfactory results in Table 6.3. The results from the NSE and MAE goodness-of-fit values are still not wihin the satisfactory value ranges. Next step, is to do the same summarization when using NSE as an objective function.

Both in Table 6.9 and Table 6.10 are the summarized results by their median value for NSE and MAE objective functions for the different structures. The highest NSE goodness-of-fit value when used NSE as an objective function is 0.26 when LINBYRES is used as a groundwater reservoir and PDM is used as a soilwater reservoir. However this value is still unsatisfactory. When MAE is used as an objective function, the MAE goodness-of-fit for calibration period is best when the LINRES and PDM are used, respectively for groundwater and soil water storage, with a total value of 0.76, which is still far from the threshold of 0.18. Three different maps are plotted in Figure 6.7 from the calibration periods for each of the objective functions.

Table 6.8: Calibration done on dry period with MAE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
EXPRES	COLLIEV2	0.200	0.100	-0.295	-0.510	1.190	1.120
EXPRES	GR4J	0.080	-0.120	-0.320	-1.035	1.075	1.135
EXPRES	NEWZEALAND	0.190	-0.015	-0.365	-1.185	1.200	1.500
EXPRES	PDM	0.190	0.080	-0.170	-0.330	1.010	0.930
FLEXRES	COLLIEV2	0.320	0.280	-0.210	-0.220	1.190	1.000
FLEXRES	GR4J	0.290	0.230	-0.210	-0.200	1.100	0.950
FLEXRES	NEWZEALAND	0.300	0.250	-0.285	-0.310	1.170	0.980
FLEXRES	PDM	0.410	0.340	0.015	0.015	1.010	0.840
LIN2PA	COLLIEV2	0.190	0.180	-0.150	-0.160	1.180	1.000
LIN2PA	GR4J	0.080	0.060	-0.150	-0.130	1.090	0.905
LIN2PA	NEWZEALAND	0.230	0.190	-0.340	-0.355	1.200	1.040
LIN2PA	PDM	0.190	0.160	-0.010	0.000	1.030	0.880
LIN2SE	COLLIEV2	0.190	0.170	-0.130	-0.150	1.175	1.000
LIN2SE	GR4J	0.080	0.050	-0.140	-0.130	1.080	0.900
LIN2SE	NEWZEALAND	0.220	0.185	-0.300	-0.340	1.180	1.000
LIN2SE	PDM	0.200	0.160	-0.010	-0.010	1.050	0.870
LINBYRES	COLLIEV2	0.270	0.250	-0.070	-0.110	1.175	1.000
LINBYRES	GR4J	0.170	0.130	-0.080	-0.080	1.050	0.900
LINBYRES	NEWZEALAND	0.325	0.270	-0.130	-0.200	1.140	0.965
LINBYRES	PDM	0.320	0.290	0.150	0.110	0.950	0.830
LINLRES	COLLIEV2	0.220	0.200	-0.140	-0.170	1.180	1.015
LINLRES	GR4J	0.110	0.090	-0.180	-0.160	1.085	0.920
LINLRES	NEWZEALAND	0.240	0.200	-0.330	-0.370	1.160	1.030
LINLRES	PDM	0.230	0.200	-0.020	-0.030	1.055	0.900
LINRES	COLLIEV2	0.290	0.270	-0.130	-0.140	1.170	1.000
LINRES	GR4J	0.190	0.180	-0.140	-0.120	1.080	0.920
LINRES	NEWZEALAND	0.285	0.260	-0.235	-0.270	1.150	0.995
LINRES	PDM	0.340	0.300	0.040	0.010	1.030	0.870
POWRES	COLLIEV2	0.190	0.170	-0.140	-0.160	1.170	1.020
POWRES	GR4J	0.080	0.060	-0.140	-0.140	1.070	0.930
POWRES	NEWZEALAND	0.220	0.180	-0.330	-0.355	1.215	1.020
POWRES	PDM	0.200	0.170	-0.020	-0.030	1.050	0.890

When comparing Figure 6.7 and Figure 6.4, it is noticeable that KGE values contain very extreme low values when calibration is done on the wet period, but on the other hand there are better approximations with the high values than the calibration done on the dry period. For NSE, there are also more extreme lower values, but also more satisfactory high values, indicated with purple. The MAE value is calculated in the same way as described in the previous section, and it is noticeable that there are less satisfactory values in the calibration during the wet period than the calibration done on the dry period.

Next step is to find the dominating 'good' and 'bad' structures for each of the calibrations. The same ranges apply as described in the previous section. There are 1912 NSE values which are bigger or equal to 0.5, which is a higher number compared with the results for NSE satisfactory values from the calibration done on the dry period. Most frequent groundwater structures are LINBYRES, LINRES and FLEXRES with values of 418, 435 and 320 respectively while for soil water structures are PDM, COLLIEV2 and GR4J with values of 844, 436 and 369 respectively. NSE is the most dominant objective function with 792 values, then MAE with 642 structures and the last is KGE with 478 structures. Again, there are too many values considered as unsatisfactory, so any attempt for interpreting

Table 6.9: Calibration done on dry period with NSE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
LINBYRES	PDM	0.210	0.170	0.260	0.220	0.880	0.750
FLEXRES	PDM	0.270	0.210	0.150	0.140	0.845	0.720
LINRES	PDM	0.180	0.160	0.150	0.140	0.820	0.700
LINBYRES	COLLIEV2	0.120	0.100	0.120	0.090	0.930	0.800
LINBYRES	NEWZEALAND	0.160	0.140	0.100	0.060	0.890	0.760
LINRES	NEWZEALAND	0.110	0.125	0.090	0.070	0.935	0.785
LINRES	COLLIEV2	0.140	0.120	0.090	0.080	0.890	0.760
LINLRES	PDM	0.090	0.060	0.090	0.080	0.900	0.790
LIN2SE	PDM	0.070	0.050	0.080	0.060	0.900	0.780
LIN2PA	PDM	0.070	0.040	0.080	0.070	0.905	0.790
POWRES	PDM	0.070	0.040	0.080	0.070	0.900	0.780
EXPRES	PDM	0.060	-0.070	0.060	-0.040	0.930	0.955
FLEXRES	COLLIEV2	0.140	0.160	0.050	0.020	0.960	0.840
LINRES	GR4J	0.050	0.040	0.050	0.050	0.870	0.760
LINBYRES	GR4J	0.040	0.030	0.050	0.030	0.895	0.790
FLEXRES	GR4J	0.120	0.070	0.040	0.020	0.900	0.790
LINLRES	NEWZEALAND	0.040	0.040	0.040	0.010	0.930	0.790
FLEXRES	NEWZEALAND	0.130	0.130	0.030	0.000	1.000	0.830
POWRES	NEWZEALAND	0.030	0.030	0.030	0.010	0.935	0.800
LIN2SE	NEWZEALAND	0.020	0.030	0.030	0.010	0.940	0.810
LINLRES	COLLIEV2	0.030	0.020	0.030	0.020	0.980	0.840
LIN2PA	COLLIEV2	0.010	0.020	0.020	0.010	0.990	0.830
LIN2PA	NEWZEALAND	0.010	0.020	0.020	0.010	0.930	0.795
POWRES	COLLIEV2	0.010	0.010	0.020	0.010	0.990	0.830
LIN2SE	COLLIEV2	0.000	0.010	0.020	0.010	0.980	0.830
LINLRES	GR4J	-0.040	-0.050	0.010	0.000	0.940	0.810
EXPRES	NEWZEALAND	0.000	-0.255	0.010	-0.230	0.960	1.500
POWRES	GR4J	-0.060	-0.060	0.000	0.000	0.955	0.815
LIN2SE	GR4J	-0.055	-0.070	0.000	-0.010	0.950	0.815
LIN2PA	GR4J	-0.070	-0.070	0.000	-0.010	0.960	0.825
EXPRES	COLLIEV2	-0.010	-0.300	0.000	-0.430	1.010	1.245
EXPRES	GR4J	-0.070	-0.200	-0.010	-0.190	0.960	1.090

this results, cannot be valid.

For MAE, there are only 591 values which are considered as satisfactory values, which proves the previous assumption from the map analysis, that MAE does not perform good on the calibration on the wet period. Thus, any further analysis in this direction can only lead to wrong assumptions. In both cases for NSE and MAE, the goodness-of-fit values which are condsidered to be satisfactory values represent a very small portion of the resulting data, thus conclusions cannot be made.

Table 6.10: Calibration done on dry period with MAE, grouped by different storages, summarized on median value

gwStor	swStor	calib_KGE	valid_KGE	calib_NSE	valid_NSE	calib_MAE	valid_MAE
LINBYRES	PDM	0.150	0.125	0.190	0.170	0.795	0.670
FLEXRES	PDM	0.160	0.140	0.120	0.110	0.790	0.680
LINRES	PDM	0.105	0.080	0.120	0.120	0.760	0.660
LINBYRES	COLLIEV2	0.070	0.080	0.080	0.080	0.860	0.750
LINLRES	PDM	0.040	0.020	0.070	0.070	0.880	0.760
LINRES	COLLIEV2	0.040	0.050	0.060	0.060	0.830	0.720
LIN2PA	PDM	0.000	-0.020	0.060	0.060	0.880	0.770
LIN2SE	PDM	0.000	-0.020	0.060	0.060	0.890	0.770
POWRES	PDM	0.000	-0.020	0.060	0.050	0.880	0.760
LINBYRES	NEWZEALAND	0.090	0.100	0.045	0.030	0.830	0.700
LINRES	NEWZEALAND	0.090	0.090	0.040	0.050	0.870	0.735
LINBYRES	GR4J	-0.010	-0.020	0.040	0.030	0.860	0.750
LINRES	GR4J	-0.010	-0.015	0.030	0.030	0.830	0.715
EXPRES	PDM	0.060	-0.030	0.030	-0.010	0.890	0.855
FLEXRES	COLLIEV2	0.100	0.110	0.020	0.020	0.920	0.770
FLEXRES	GR4J	0.040	0.030	0.010	0.010	0.880	0.760
LINLRES	COLLIEV2	-0.040	-0.040	0.000	0.010	0.940	0.810
POWRES	COLLIEV2	-0.060	-0.060	0.000	0.000	0.950	0.800
LIN2PA	COLLIEV2	-0.070	-0.070	0.000	0.000	0.930	0.810
LIN2SE	COLLIEV2	-0.080	-0.070	0.000	0.000	0.940	0.810
FLEXRES	NEWZEALAND	0.090	0.120	-0.010	-0.020	0.920	0.760
LINLRES	NEWZEALAND	0.005	0.010	-0.010	-0.010	0.875	0.740
LIN2PA	NEWZEALAND	-0.040	-0.010	-0.010	-0.005	0.870	0.735
LIN2SE	NEWZEALAND	-0.030	-0.020	-0.010	0.000	0.880	0.745
POWRES	NEWZEALAND	-0.030	-0.020	-0.010	0.000	0.880	0.730
LINLRES	GR4J	-0.060	-0.070	-0.010	-0.010	0.920	0.790
LIN2SE	GR4J	-0.070	-0.085	-0.010	-0.010	0.930	0.800
LIN2PA	GR4J	-0.080	-0.090	-0.010	-0.010	0.920	0.800
POWRES	GR4J	-0.080	-0.095	-0.010	-0.010	0.930	0.795
EXPRES	COLLIEV2	-0.030	-0.350	-0.020	-0.370	0.980	1.170
EXPRES	NEWZEALAND	-0.030	-0.350	-0.020	-0.525	0.900	1.520
EXPRES	GR4J	-0.060	-0.245	-0.030	-0.405	0.950	1.075

6.3 Discussion

The overall performance of the structures differs. The 'winning' groundwater structures were FLEXRES, LINBYRES and LINRES, and the 'winning' soil water structure was PDM. This can be seen in the summary table Table 6.11, where all the results were grouped by the groundwater and soil water storage for both wet and dry period, summed by their median value, regardless of the choice of the objective function. The best performance was the structure where FLEXRES was used as a groundwater reservoir and PDM was used as a soil water reservoir. The FLEXRES is a groundwater linear reservoir where the storage outflow is controlled by a threshold. The reason behind the outperformance of this reservoir is because of its complexity, its structure means that groundwater storage can be controlled by multiple aquifers. Second in place comes the LINBYRES, a linear storage with a direct by pass, again in combination with PDM. Third in place comes the structure with LINRES and PDM reservoirs. What connects these three groundwater reservoirs is their ability to drain faster and more directly (Stoelzle et al., 2015). Also, whenever there is a recharge, the groundwater reservoir empties faster than initially estimated, which makes the storage-discharge relationship

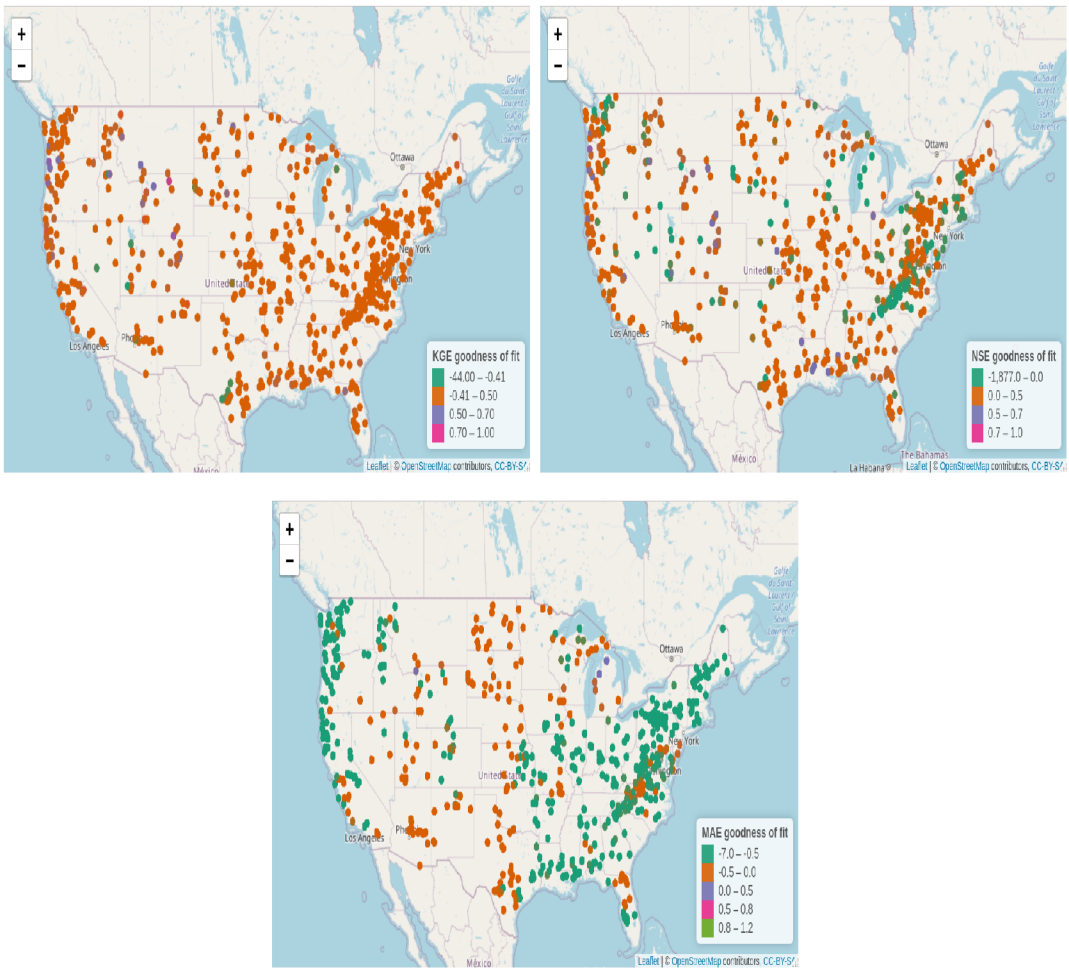

FIGURE 6.7: Distribution for goodness-of-fit values for calibration done with KGE, NSE, MAE objective functions on wet period

more linear (Fenicia et al., 2006) thus the reason why particularly these reservoirs had the best performance.

Regarding the groundwater storage, the findings about the superior models were also consistent with Stoelzle et al. (2015), where FLEXRES was defined to be the model with the highest performance, because this type of reservoir is designed with a threshold where as long as the linear reservoir is above this threshold, faster depletion will occur. LIN2SE was found to be the with the poorest performance. The influence of the different combinations of the groundwater storages can be seen in Table 6.12, where the results were grouped by the groundwater storage only, summed by their median value.

Regarding the soil water structures, the results were higher for PDM model, but also in the second place with the exact same performance is NEWZEALAND, which is then followed by COLLIEV2. The summary results where the grouping was performed on the soil water structures can be seen in Table 6.13. The PDM reservoir is a lumped model and it describes the spatial distribution (variability)

Table 6.11: Summary results of both periods, grouped by groundwater and soil water storage

gwStor	swStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
EXPRES	COLLIEV2	0.080	-0.05	0.95	-0.090	-0.46	1.200
EXPRES	GR4J	0.000	-0.05	0.89	-0.215	-0.56	1.120
EXPRES	NEWZEALAND	0.090	-0.04	0.88	-0.310	-0.78	1.600
EXPRES	PDM	0.110	0.01	0.84	-0.020	-0.17	0.950
FLEXRES	COLLIEV2	0.210	0.00	0.89	0.180	-0.02	0.910
FLEXRES	GR4J	0.160	0.01	0.85	0.130	-0.01	0.870
FLEXRES	NEWZEALAND	0.195	-0.03	0.89	0.160	-0.10	0.910
FLEXRES	**PDM**	**0.280**	**0.11**	**0.78**	**0.260**	**0.09**	**0.790**
LIN2PA	COLLIEV2	0.060	-0.01	0.91	0.050	-0.01	0.920
LIN2PA	GR4J	-0.020	-0.02	0.88	-0.030	-0.03	0.885
LIN2PA	NEWZEALAND	0.110	-0.01	0.87	0.080	-0.05	0.890
LIN2PA	PDM	0.080	0.05	0.85	0.070	0.04	0.850
LIN2SE	COLLIEV2	0.070	-0.01	0.92	0.050	-0.01	0.920
LIN2SE	GR4J	-0.010	-0.02	0.88	-0.020	-0.03	0.890
LIN2SE	NEWZEALAND	0.110	-0.01	0.87	0.080	-0.04	0.890
LIN2SE	PDM	0.090	0.05	0.84	0.075	0.04	0.850
LINBYRES	COLLIEV2	0.170	0.06	0.86	0.150	0.04	0.870
LINBYRES	GR4J	0.070	0.02	0.83	0.060	0.01	0.850
LINBYRES	NEWZEALAND	0.200	0.04	0.84	0.170	0.00	0.850
LINBYRES	**PDM**	**0.210**	**0.19**	**0.79**	**0.200**	**0.18**	**0.790**
LINLRES	COLLIEV2	0.090	-0.01	0.92	0.080	-0.01	0.930
LINLRES	GR4J	0.020	-0.02	0.87	0.000	-0.02	0.880
LINLRES	NEWZEALAND	0.120	-0.02	0.87	0.090	-0.07	0.880
LINLRES	PDM	0.120	0.06	0.84	0.110	0.05	0.860
LINRES	COLLIEV2	0.180	0.05	0.84	0.170	0.04	0.850
LINRES	GR4J	0.090	0.02	0.83	0.080	0.01	0.840
LINRES	NEWZEALAND	0.180	0.03	0.84	0.160	0.01	0.860
LINRES	**PDM**	**0.210**	**0.11**	**0.76**	**0.200**	**0.11**	**0.770**
POWRES	COLLIEV2	0.070	-0.01	0.92	0.050	-0.01	0.930
POWRES	GR4J	-0.010	-0.02	0.87	-0.025	-0.02	0.890
POWRES	NEWZEALAND	0.110	-0.01	0.87	0.080	-0.04	0.890
POWRES	PDM	0.090	0.05	0.85	0.070	0.04	0.860

Table 6.12: All results grouped by groundwater storage by median value

gwStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
EXPRES	0.07	-0.03	0.89	-0.135	-0.43	1.17
FLEXRES	**0.21**	**0.02**	**0.85**	**0.180**	**0.00**	**0.87**
LIN2PA	0.06	0.00	0.88	0.040	-0.01	0.89
LIN2SE	0.06	0.00	0.88	0.040	-0.01	0.89
LINBYRES	**0.17**	**0.07**	**0.83**	**0.150**	**0.05**	**0.84**
LINLRES	0.09	0.00	0.88	0.070	-0.01	0.89
LINRES	**0.16**	**0.05**	**0.82**	**0.150**	**0.04**	**0.83**
POWRES	0.06	0.00	0.88	0.040	-0.01	0.89

of the soil capacity, unlike the other three soil water storages where buckets are being used. It is noticeable that GR4J did not perform so well, unlike in Pagano et al. (2010) were GR4J outperformed every other model. This could be because the form in which the GR4J model in this study was implemented, was by taking only the soil water structure storage, which might have a huge affect on the performance of the model overall. So, this must not be overlooked and must be a subject for study for any future analysis in dHRUM.

As previously noted in the summary of the results section, the results where the KGE objective was used, were the most satisfactory ones, with 95% coverage of the resulting dataset, so taking this as a fact a summary of the results from

Table 6.13: All results grouped by soil water storage by median value

swStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
COLLIEV2	**0.11**	**0.00**	**0.90**	**0.09**	**-0.02**	**0.94**
GR4J	0.03	-0.01	0.86	0.00	-0.03	0.89
NEWZEALAND	**0.14**	**-0.01**	**0.86**	**0.09**	**-0.06**	**0.92**
PDM	**0.14**	**0.08**	**0.82**	**0.12**	**0.06**	**0.84**

both periods, grouped by their groundwater structure, and later by their soil water structure are shown in Table 6.14 and Table 6.15.

Table 6.14: Results obtained from using KGE as an objective function grouped by ground water storage by median value

gwStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
FLEXRES	0.33	-0.16	0.980	0.28	-0.190	0.990
LINRES	0.28	-0.10	0.970	0.24	-0.110	0.980
LINBYRES	0.27	-0.04	0.945	0.24	-0.060	0.960
LINLRES	0.20	-0.16	1.000	0.17	-0.160	1.000
LIN2PA	0.18	-0.13	1.000	0.15	-0.130	0.990
LIN2SE	0.18	-0.14	0.990	0.15	-0.130	0.990
POWRES	0.18	-0.14	1.000	0.15	-0.130	0.990
EXPRES	0.17	-0.27	0.990	-0.01	-0.805	1.195

Table 6.15: Results obtained from using KGE as an objective function grouped by soil water storage by median value

swStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
NEWZEALAND	0.25	-0.27	1.02	0.20	-0.33	1.08
PDM	0.25	0.00	0.92	0.22	-0.01	0.92
COLLIEV2	0.24	-0.15	1.05	0.20	-0.16	1.05
GR4J	0.13	-0.17	0.96	0.09	-0.17	0.98

The results are ordered in a descending order. The outliers were also included in all summary results. However, in the results grouped by both groundwater and soil water structures, only when KGE was used as an objective function, it is noticeable that GR4J performs good in combination with the FLEXRES (Table 6.16)

The structures with the lowest performance were indicated to be structures where GR4J was used as a soil-water storage in combination with the LIN2PA and LIN2SE reservoirs. Also, it is noticeable that the structure with lower performance were indicated to be also the POWRES and EXPRES especially when used with PDM, while when in combination with NEWZEALAND and COLLIEV2 that had a better performance.

Regarding the performance of the calibrations on the dry period versus the calibration on the wet period, the results in Table 6.17 were obtained, where all of the summary results were grouped by the period on their median value.

The results show that when calibration was done on the wet period, the results from the validation period were higher than the results from the validation when

Table 6.16: Results obtained from using KGE as an objective function grouped by soil water storage and groundwater storage by median value

swStor	swStor	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
FLEXRES	PDM	0.395	0.020	0.890	0.35	-0.010	0.890
LINRES	PDM	0.330	0.030	0.910	0.29	0.010	0.920
FLEXRES	COLLIEV2	0.330	-0.220	1.040	0.28	-0.240	1.050
LINBYRES	NEWZEALAND	0.320	-0.120	0.990	0.27	-0.180	1.005
LINBYRES	PDM	0.310	0.130	0.860	0.29	0.120	0.870
LINRES	COLLIEV2	0.300	-0.120	1.020	0.26	-0.120	1.035
FLEXRES	NEWZEALAND	0.300	-0.270	1.030	0.25	-0.330	1.020
LINRES	NEWZEALAND	0.290	-0.210	1.010	0.25	-0.250	1.030
FLEXRES	GR4J	0.280	-0.200	0.990	0.23	-0.200	0.990
LINBYRES	COLLIEV2	0.270	-0.070	1.020	0.24	-0.090	1.030
LINLRES	NEWZEALAND	0.250	-0.305	1.020	0.20	-0.350	1.060
LIN2SE	NEWZEALAND	0.230	-0.300	1.020	0.18	-0.330	1.050
POWRES	NEWZEALAND	0.230	-0.310	1.040	0.18	-0.320	1.060
LIN2PA	NEWZEALAND	0.230	-0.310	1.050	0.18	-0.320	1.060
LINLRES	PDM	0.220	-0.010	0.940	0.20	-0.020	0.950
LINLRES	COLLIEV2	0.220	-0.150	1.070	0.19	-0.160	1.060
EXPRES	NEWZEALAND	0.200	-0.330	1.050	-0.04	-1.210	1.560
POWRES	COLLIEV2	0.200	-0.140	1.055	0.17	-0.120	1.040
LIN2PA	COLLIEV2	0.200	-0.140	1.060	0.17	-0.130	1.030
EXPRES	COLLIEV2	0.200	-0.295	1.070	0.09	-0.540	1.170
LIN2SE	COLLIEV2	0.200	-0.140	1.070	0.17	-0.130	1.030
EXPRES	PDM	0.190	-0.170	0.910	0.07	-0.520	1.010
LIN2SE	PDM	0.190	0.000	0.925	0.16	0.000	0.930
LINRES	GR4J	0.190	-0.130	0.950	0.17	-0.120	0.970
LIN2PA	PDM	0.180	-0.010	0.930	0.16	0.000	0.920
POWRES	PDM	0.180	-0.020	0.930	0.17	-0.020	0.920
LINBYRES	GR4J	0.160	-0.090	0.920	0.13	-0.080	0.940
LINLRES	GR4J	0.110	-0.180	0.980	0.08	-0.160	0.960
EXPRES	GR4J	0.080	-0.315	0.950	-0.18	-1.395	1.170
POWRES	GR4J	0.080	-0.150	0.955	0.06	-0.140	0.950
LIN2PA	GR4J	0.080	-0.150	0.970	0.06	-0.130	0.970
LIN2SE	GR4J	0.070	-0.150	0.960	0.05	-0.130	0.955

Table 6.17: All results grouped by period by median value

calib_period	obj$_{function}$	calib_KGE	calib_NSE	calib_MAE	valid_KGE	valid_NSE	valid_MAE
dry	kge	0.22	-0.14	0.87	0.17	-0.15	1.05
wet	kge	0.22	-0.14	1.11	0.18	-0.16	0.97
dry	nse	0.06	0.06	0.73	0.03	0.02	0.90
wet	nse	0.06	0.05	0.93	0.04	0.03	0.82
wet	mae	0.01	0.02	0.89	-0.01	0.02	0.79
dry	mae	0.00	0.03	0.70	-0.02	0.00	0.86

the calibration was done on the dry period. For an example, as seen from the Table 6.11 when KGE was used as an objective function, the calibration produced the same goodness-of-fit KGE values (calib_KGE) for both periods, but the validation (valid_KGE) had a higher score for when the calibration was done on the wet period (0.18). Even when the calibration produced higher goodness-of-fit criteria for the dry period calibrations (for an example NSE objective function, NSE goodness-of-fit criteria - calib_NSE column), still the validation (valid_NSE column) produced higher values when the calibration was done for the wet period (0.03). Similar results regarding the usage of the split of the periods between wet and dry periods were found by Gao et al. (2018).

The influence of using the differential split sample test for separating the dry and the wet periods shows how these very contrasting periods influence the model

performance. However, it is important to note that in this study, distinguishing the dry from the wet period was done in a very specific way, where 10-years of consecutive periods were taken, and the average precipitation was measured and compared within all available 10 year consecutive periods. The wetter period had a higher performance, but also has higher extreme values (in this case the extreme values were negative values). This is due to overestimating and underestimating of the discharges because of the calibration done on the wet period (Coron et al., 2012). This tendency to overestimate and underestimate should be a concern of any future study proceeding in this direction, because it may greatly influence any analysis on climate impact or hydrologic predictions, since the error predictions could be very large (Merz et al., 2011).

These results suggest the need for more detailed future modelling which should be based on hydrological processes. Many models are similar in terms of efficiency, but what informaton is not visible here is the internal representation of the processes. This opens up a door for further investigations about dominant processes (Knoben et al., 2019a).

CHAPTER 7

Conclusion and contribution

The distributed hydrological response unit model was used as a modelling framework for testing lumped hydrological models in this study. The framework was extended by implementing additional seven groundwater structures and three soil water structures which gave us in total 32 lumped conceptual models for testing their modeling performance. The differential split sample test was implemented where calibration and validation was done on the wet and the dry period and vice versa using the CAMELS dataset and NLDAS data as forsing data.

The calibration was based on three different objective functions: KGE, NSE, MAE. The Differential Evolution algorithm was used as an optimization algorithm. Three goodness-of-fit criteria were chosen for evaluating the performance of the different lumped structures which were generated from the GOF (goodness of fit) package in R: KGE, NSE, MAE.

From the results, it can be concluded that satisfactory results were obtained only when KGE was used as an objective function.

The 'winning' groundwater structures were FLEXRES, LINBYRES and LINRES, and the 'winning' soil water structure were PDM, NEWZEALAND and COLLIEV2. In contrary, the LIN2PA, LIN2SE as groundwater structures and GR4J as soil water structure were found to have a poorer performance in this study. The structures with the lowest performance were indicated to be structures where GR4J was used as a soil-water storage in combination with the LIN2PA and LIN2SE reservoirs. From the results of this study it can be concluded that the structure with lower performance were indicated to be also the POWRES and EXPRES especially when used with PDM, while when in combination with NEWZEALAND and COLLIEV2 had a better performance.

The results also show that when calibration was done on the wet period, the validations of the model were better than the ones where calibration was done on the dry period.

Taking in consideration that the data analysis was performed only after all the lumped models were formulated and a proper calibration and validation was performed, along which the hypothesis of 'winning' and losing' structures emerged, this study only opens the door to more detailed research for the differences between these structures or any future structures that may be developed as part of the distributed hydrological reponse unit modelling framework.

Bibliography

Nans Addor, Andrew J Newman, Naoki Mizukami, and Martyn P Clark. The camels data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10):5293–5313, 2017.

Nans Addor, Hong X Do, Camila Alvarez-Garreton, Gemma Coxon, Keirnan Fowler, and Pablo A Mendoza. Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, 65(5):712–725, 2020.

Vazken Andréassian, C Perrin, L Berthet, N Le Moine, J Lerat, C Loumagne, Ludovic Oudin, T Mathevet, M-H Ramos, and A Valéry. Hess opinions" crash tests for a standardized evaluation of hydrological models". *Hydrology and Earth System Sciences*, 13(10):1757–1764, 2009.

David Ardia, Kris Boudt, Peter Carl, Katharine Mullen, and Brian G Peterson. Differential evolution with deoptim: an application to non-convex portfolio optimization. *the R Journal*, 3(1):27–34, 2011.

Keith Beven. A manifesto for the equifinality thesis. *Journal of hydrology*, 320 (1-2):18–36, 2006.

Keith J Beven. Uniqueness of place and process representations in hydrological modelling. *Hydrology and earth system sciences*, 4(2):203–213, 2000.

Keith J Beven. *Rainfall-runoff modelling: The primer*. John Wiley & Sons, 2011.

Francis HS Chiew. Lumped conceptual rainfall-runoff models and simple water balance methods: Overview and applications in ungauged and data limited regions. *Geography Compass*, 4(3):206–225, 2010.

Martyn P Clark, Bart Nijssen, Jessica D Lundquist, Dmitri Kavetski, David E Rupp, Ross A Woods, Jim E Freer, Ethan D Gutmann, Andrew W Wood, Levi D Brekke, et al. A unified approach for process-based hydrologic modeling: 1. modeling concept. *Water Resources Research*, 51(4):2498–2514, 2015.

Laurent Coron, Vazken Andréassian, Charles Perrin, Julien Lerat, Jai Vaze, Marie Bourqui, and Frederic Hendrickx. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 australian catchments. *Water Resources Research*, 48(5), 2012.

Gemma Coxon, Jim Freer, Rosanna Lane, Toby Dunne, Wouter JM Knoben, Nicholas JK Howden, Niall Quinn, Thorsten Wagener, and Ross Woods. Decipher v1: dynamic fluxes and connectivity for predictions of hydrology. *Geoscientific Model Development*, 12(6):2285–2306, 2019.

Marco Dal Molin, Dmitri Kavetski, and Fabrizio Fenicia. Superflexpy 1.3. 0: an open-source python framework for building, testing, and improving conceptual hydrological models. *Geoscientific Model Development*, 14(11):7047–7072, 2021.

James W Deardorff. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research: Oceans*, 83(C4):1889–1903, 1978.

Gayathri Devia, Ganasri Bigganahalli Puttaswamigowda, and G.S. Dwarakish. A review on hydrological models. *Aquatic Procedia*, 4:1001–1007, 12 2015. doi: 10.1016/j.aqpro.2015.02.126.

Qingyun Duan, Soroosh Sorooshian, and Vijai Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, 28 (4):1015–1031, 1992.

F Fenicia, HHG Savenije, P Matgen, and L Pfister. Is the groundwater reservoir linear? learning from data in hydrological modelling. *Hydrology and Earth System Sciences*, 10(1):139–150, 2006.

Fabrizio Fenicia, Hubert HG Savenije, Patrick Matgen, and Laurent Pfister. Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, 44(1), 2008.

Xin Gao, Xingwei Chen, Trent W Biggs, and Huaxia Yao. Separating wet and dry years to improve calibration of swat in barrett watershed, southern california. *Water*, 10(3):274, 2018.

M Kashif Gill, Yasir H Kaheil, Abedalrazq Khalil, Mac McKee, and Luis Bastidas. Multiobjective particle swarm optimization for parameter estimation in hydrology. *Water Resources Research*, 42(7), 2006.

Hoshin V Gupta, Harald Kling, Koray K Yilmaz, and Guillermo F Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2): 80–91, 2009.

Herath Mudiyanselage Viraj Vidura Herath, Jayashree Chadalawada, and Vladan Babovic. Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 25(8): 4373–4401, 2021.

Dmitri Kavetski and Fabrizio Fenicia. Elements of a flexible approach for conceptual hydrological modeling: 2. application and experimental insights. *Water Resources Research*, 47(11), 2011.

Vit Klemeš. Operational testing of hydrological simulation models. *Hydrological sciences journal*, 31(1):13–24, 1986.

Wouter JM Knoben, Jim E Freer, Keirnan JA Fowler, Murray C Peel, and Ross A Woods. Modular assessment of rainfall–runoff models toolbox (marrmot) v1. 2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, 12(6):2463–2480, 2019a.

Wouter JM Knoben, Jim E Freer, and Ross A Woods. Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10):4323–4331, 2019b.

Wouter JM Knoben, Jim E Freer, MC Peel, KJA Fowler, and Ross A Woods. A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, 56(9):e2019WR025975, 2020.

Rosanna A Lane, Gemma Coxon, Jim E Freer, Thorsten Wagener, Penny J Johnes, John P Bloomfield, Sheila Greene, Christopher JA Macleod, and Sim M Reaney. Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in great britain. *Hydrology and Earth System Sciences*, 23(10):4011–4032, 2019.

Shuming Liu, David Butler, Richard Brazier, Louise Heathwaite, and Soon-Thiam Khu. Using genetic algorithms to calibrate a water quality model. *Science of the Total Environment*, 374(2-3):260–272, 2007.

L Mediero, L Garrote, and FJ Martín-Carrasco. Probabilistic calibration of a distributed hydrological model for flood forecasting. *Hydrological sciences journal*, 56(7):1129–1149, 2011.

Ralf Merz, Juraj Parajka, and Günter Blöschl. Time stability of catchment model parameters: Implications for climate impact analyses. *Water resources research*, 47(2), 2011.

Daniel N Moriasi, Margaret W Gitau, Naresh Pai, and Prasad Daggupati. Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6):1763–1785, 2015.

Katharine Mullen, David Ardia, David L Gil, Donald Windover, and James Cline. Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

J Eamonn Nash and Jonh V Sutcliffe. River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.

AJ Newman, MP Clark, Kevin Sampson, Andrew Wood, LE Hay, A Bock, RJ Viger, D Blodgett, L Brekke, JR Arnold, et al. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous usa: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1):209–223, 2015.

Pierre Nicolle, Raji Pushpalatha, Charles Perrin, Didier François, Dominique Thiéry, Thibault Mathevet, Matthieu Le Lay, François Besson, J-M Soubeyroux, Christian Viel, et al. Benchmarking hydrological models for low-flow simulation and forecasting on french catchments. *Hydrology and Earth System Sciences*, 18(8):2829–2857, 2014.

Tom Pagano, Prasantha Hapuarachchi, and QJ Wang. Continuous rainfall-runoff model comparison and short-term daily streamflow forecast skill evaluation. 2010.

Charles Perrin, Claude Michel, and Vazken Andréassian. Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, 279(1-4): 275–289, 2003.

Saman Razavi and Bryan A Tolson. An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research*, 49(12):8418–8431, 2013.

Axel Ritter and Rafael Munoz-Carpena. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480:33–45, 2013.

Danny Saavedra, Pablo A Mendoza, Nans Addor, Harold Llauca, and Ximena Vargas. A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*, 2021.

Léonard Santos, Guillaume Thirel, and Charles Perrin. Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the gr4 model using state-space gr4 (version 1.0). *Geoscientific Model Development*, 11 (4):1591–1605, 2018.

G Seiller, F Anctil, and C Perrin. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrology and Earth System Sciences*, 16(4):1171–1189, 2012.

Soroosh Sorooshian, Kuolin Hsu, Erika Coppola, Barbara Tomassetti, Marco Verdecchia, and Guido Visconti. *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models*, volume 63. Springer, 01 2008. ISBN 978-3-540-77842-4. doi: 10.1007/978-3-540-77843-1.

Maria Staudinger, K Stahl, Jan Seibert, MP Clark, and LM Tallaksen. Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences*, 15(11):3447–3459, 2011.

Michael Stoelzle, Markus Weiler, Kerstin Stahl, Andreas Morhard, and Tobias Schuetz. Is there a superior conceptual groundwater model structure for baseflow simulation? *Hydrological processes*, 29(6):1301–1313, 2015.

Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

WR Van Esse, C Perrin, Martijn J Booij, Dionysius CM Augustijn, F Fenicia, D Kavetski, and F Lobligeois. The influence of conceptual model structure on model performance: a comparative study for 237 french catchments. *Hydrology and earth system sciences*, 17(10):4227–4239, 2013.

Eric F Wood, Dennis P Lettenmaier, and Valerie G Zartarian. A land-surface hydrology parameterization with subgrid variability for general circulation models. *Journal of Geophysical Research: Atmospheres*, 97(D3):2717–2728, 1992.