Added value of distribution in rainfall-runoff models for the Meuse basin

de Boer, Tanja

Important note
To cite this publication, please use the final published version (if applicable).
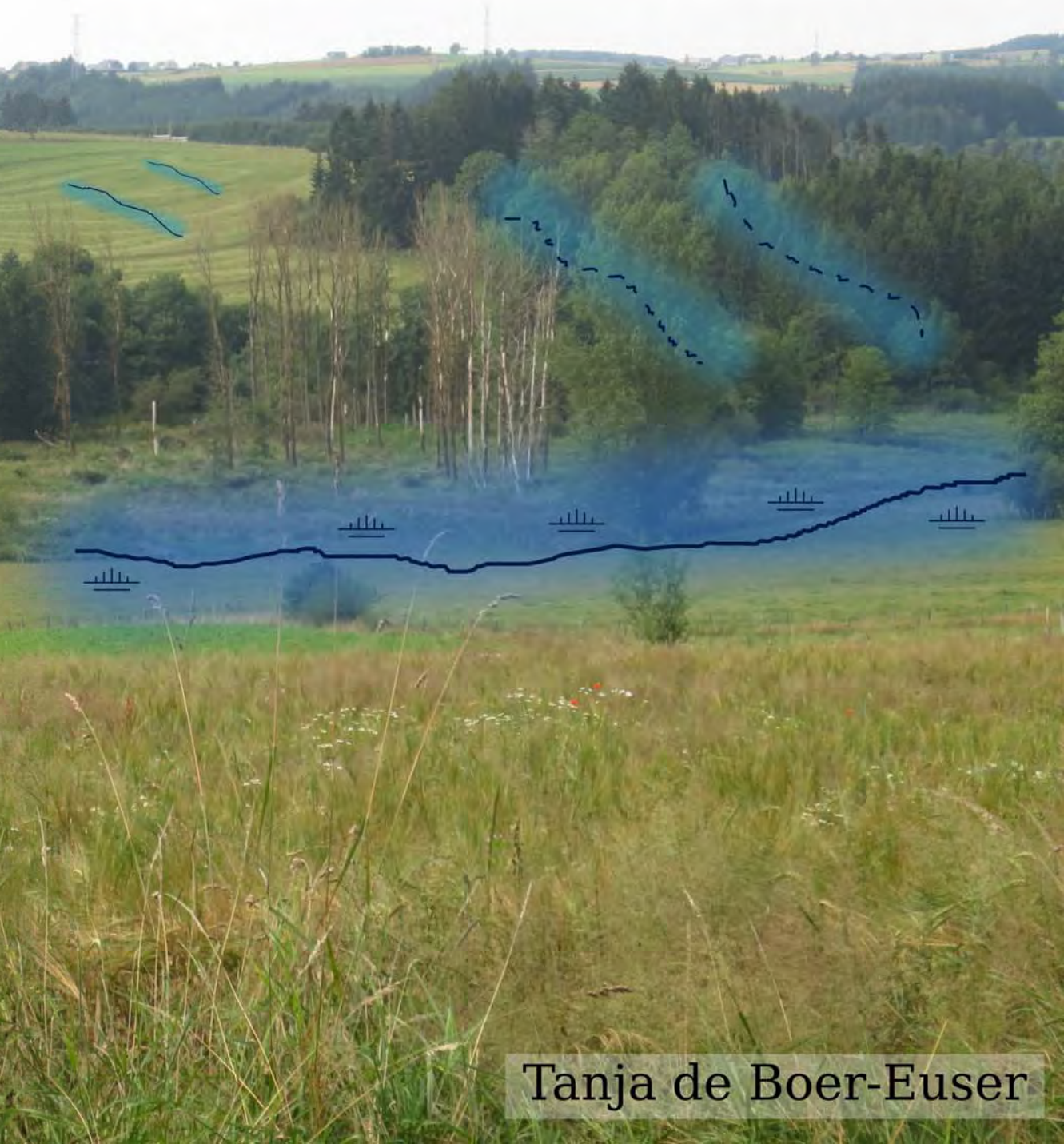Please check the document version above.

Added value of distribution
in rainfall-runoff models
for the Meuse basin

Tanja de Boer-Euser

ADDED VALUE OF DISTRIBUTION IN RAINFALL-RUNOFF MODELS FOR THE MEUSE BASIN

ADDED VALUE OF DISTRIBUTION IN RAINFALL-RUNOFF MODELS FOR THE MEUSE BASIN

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 10 januari 2017 om 12:30 uur

door

Tanja DE BOER-EUSER

civiel ingenieur
geboren te Maassluis, Nederland.

This dissertation has been approved by the

promotor: prof. dr. ir. H.H.G. Savenije
copromotor: dr. M. Hrachowitz

Composition of the doctoral committee:

Rector Magnificus,	voorzitter
Prof. dr. ir. H.H.G Savenije,	Delft University of Technology
Dr. M. Hrachowitz,	Delft University of Technology

Independent members:

Dr. H.K. McMillan,	San Diego State University
Prof. dr. ir. P. Willems,	University of Leuven
Prof. dr. ir. M.F.P. Bierkens,	Utrecht University
Prof. dr. W.G.M. Bastiaanssen,	Delft University of Technology
Prof. dr. ir. N.C. van de Giesen,	Delft University of Technology, reserve member

Other member:

Dr. ir. H.C. Winsemius,	Deltares

Dr. ir. H.C. Winsemius has contributed greatly to the preparation of this dissertation.

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

PREFACE

About a year after I started my PhD, my mother asked me whether I had imagined to end up doing the same as my father when starting my study civil engineering. In fact, by that time I could have never imagined it. Especially because the main thing I knew about my father's job was that it had to do with computers and that it was complicated. Although the final goal of hydrological modelling is certainly not to produce a lot of computer code, writing and debugging of some code is certainly important to get the results you would like to get. So, after almost five years of modelling research I understand what keeps you going after midnight when a computer does not do what you want it to do.

Having said that, I guess I have experienced the 20-80 rule in its fullest[1]. I have spent by far the largest trunk of time on solving errors, dealing with log-files on linux clusters, closing numerical(!) water balances and calibrating model configurations. On the other hand, most 'breakthroughs' happened when calculations on the linux cluster were not working, so I had to use some creativity and the calculation power of my laptop.

Breakthroughs which were in the end presented in peer review publications. An important step for a good publication is a thorough review process. A task which is taken free of charge and with best intentions by fellow researchers. Despite the hard work reviewers deliver, somehow a strong tendency exists in which reviews are submitted anonymously. Even positive and highly qualitative reviews are posted anonymously. I am really confused about a field in which we do not seem to dare to put our name on something we were asked to do, which is necessary to do and which we do with our best intentions.

By reading this, it might seem that the last four years went from fight to disappointment and back. This is certainly not true! I really enjoyed working on this piece of research, as I also really enjoyed all discussions about science and its side aspects. I hope you will enjoy reading (parts of) it and that the contents might turn out to be useful someday.

Tanja de Boer
Delft, October 2016

---

[1]for those who did not experience it yet: it assumes that 20% of the work is done in 80% of the time and vice versa

SUMMARY

Why do equal precipitation events not lead to equal discharge events across space and time? The easy answer would be because catchments are different, which then leads to the second question: Why do hydrologists often use the same rainfall-runoff model for different catchments? Probably because specifying and distributing hydrological processes across catchments is not straightforward. It requires catchment data and proper tools to evaluate the details and spatial representation of the modelled processes. However, making a model more specific and distributed can improve the performance and predictive power of the hydrological model. Therefore, this thesis evaluates the added value of including spatial characteristics in rainfall-runoff models.

Most model experiments in this thesis are carried out in the Ourthe catchment, a subcatchment of the Meuse basin. This catchment has a strong seasonal behaviour, responds quickly to precipitation and has a large influence on peak flows in the Meuse. It has a variety of landscapes, among which steep forested slopes and flat agricultural fields.

This thesis proposes a new evaluation framework (Framework to Assess Realism of Model structures (FARM)), based on different characteristics of the hydrograph (hydrological signatures). Key element of this framework is that it evaluates both performance (good reproduction of signatures) and consistency (reproduction of multiple signatures with the same parameter set). This framework is used together with various other model evaluation tools to evaluate models at three levels: internal model behaviour, model performance and consistency, and predictive power.

The root zone storage capacity (S_r) of vegetation is an important parameter in conceptual rainfall-runoff models. It largely determines the partitioning of precipitation into evaporation and discharge. Distribution of a climate derived S_r-value (i.e., based on precipitation and evaporation) was compared with S_r-values derived from soil samples in 32 New Zealand catchments. The comparison is based on spatial patterns and a model experiment. It is concluded that climate is a better estimator for S_r than soil, especially in wet catchments. Within the Meuse basin, climate derived S_r-values have been estimated as well; applying these newly derived storage estimates improved model results.

Two types of distribution have been tested for the Ourthe catchment: the distribution of meteorological forcing and the distribution of model structure. The distribution of forcing was based on spatially variable precipitation and potential evaporation. These were averaged at different levels within in the model, thereby creating four levels of model state distribution. The model structure was distributed by using two hydrological response units (HRUs), representing wetlands and hillslopes. Eventually, a lumped and a distributed model structure were compared, each with four levels of model state (forcing) distribution. From this, it is concluded that distribution of model structure is more

important than distribution of forcing. However, if the model structure is distributed, the forcing should be distributed as well.

Knowing that distribution of model structure is relevant, more detailed process conceptualisations have been tested for the Ourthe Orientale, a subcatchment of the Ourthe. An additional agricultural HRU was introduced for which Hortonian overland flow and frost in the topsoil are assumed to be relevant. In addition, a degree-day based snow module has been added to all HRUs. Adding these process conceptualisations improved the performance and consistency of the model on an event basis. However, the implemented processes and the related signatures are sensitive to errors in forcing and model outliers and should therefore be implemented carefully.

This thesis finishes with two explorative comparisons; one comparing the newly developed model of the Ourthe Orientale catchment with other catchments; the second between the newly developed model and other models, including the HBV configuration currently used for operational forecasting in the Meuse basin. These comparisons were carried out based on visual inspections of parts of the hydrograph. The results show that the newly developed model can be applied in neighbouring catchments with similar performance. The comparison with other models demonstrates that a very quick overland flow component and a parallel configuration of fast and slow runoff generating reservoirs is important to reproduce the dynamics of the hydrograph related to different time scales. Both aspects are included in the newly developed model. As a results, the newly developed model is better able to reproduce most of the dynamics of the hydrograph than the operational HBV configuration, used at the moment of writing.

Distribution and detailed process conceptualisation are very beneficial for rainfall-runoff modelling of the Ourthe catchment. However, they should be applied with care. Conceptual models are a strong simplification of reality. When confronting them only with discharge data, there is a risk of misinterpreting other hydrological processes.

This thesis suggests two possible opportunities to further improve conceptual models. First, catchment understanding could be increased by adding more physical meaning to the models, such as the climate derived root zone storage capacity. And second, remote sensing and plot scale data could be combined to link hydrological processes at different scales. In this way conceptual models can probably be used to get more insight into scaling issues, which occur when moving from hillslope to catchment scale.

SAMENVATTING

Waarom leiden dezelfde regenbuien niet altijd en overal tot dezelfde rivier afvoeren? Het makkelijke antwoord is omdat stroomgebieden van elkaar verschillen, wat leidt tot de volgende vraag: Waarom gebruiken hydrologen vaak hetzelfde regen-afvoer model voor verschillende stroomgebieden? Waarschijnlijk omdat het niet eenvoudig is om hydrologische processen te specificeren en toe te wijzen aan verschillende stroomgebieden. Echter, het meer specifiek en ruimtelijk gevarieerd maken van een model, kan de prestatie en het voorspellend vermogen van het hydrologische model vergroten. Daarom evalueert dit proefschrift de toegevoegde waarde van het opnemen van ruimtelijke karakteristieken in regen-afvoer modellen.

De meeste modelleerexperimenten in dit proefschrift zijn uitgevoerd voor het stroomgebied van de Ourthe, een substroomgebied van het Maas stroomgebied. Dit stroomgebied heeft een sterk seizoensgebonden karakter, reageert snel op neerslag en heeft veel invloed op de afvoer in de Maas. Het heeft een variëteit aan landschappen, waaronder steile beboste hellingen en vlakke akkers en weilanden.

Dit proefschrift stelt een nieuw evaluatiekader voor (Kader ter beoordeling van realisme van model structuren), gebaseerd op verschillende karakteristieken van de hydrograaf (hydrologische signaturen). Een essentieel onderdeel van dit kader is dat het zowel prestatie (het goed reproduceren van signaturen) als consistentie (het reproduceren van verschillende signaturen met dezelfde parameterset) beoordeelt. Dit kader is gebruikt samen met diverse andere model evaluatie technieken om modellen op drie niveaus te evalueren: intern model gedrag, model prestatie en consistentie, en voorspellend vermogen.

De bergingscapaciteit van de wortelzone (S_r) van vegetatie is een belangrijke parameter in conceptuele regen-afvoer modellen. Het bepaalt voor een groot deel de verdeling van neerslag naar verdamping en afvoer. Een ruimtelijk gevarieerde S_r-waarde afgeleid uit klimaatgegevens (i.e., neerslag en verdamping) is vergeleken met S_r-waarden afgeleid van bodem monsters voor 32 stroomgebieden in Nieuw Zeeland. De vergelijking is gebaseerd op ruimtelijke patronen en een modelleerexperiment. De conclusie is dat het klimaat een betere indicatie geeft voor S_r dan bodemeigenschappen, zeker in zeer natte stroomgebieden. De klimaat afgeleide S_r-waarden zijn ook vor een deel van het Maas stroomgebied berekend; het toepassen van deze klimaat afgeleide bergingscapaciteiten heeft de model resultaten verbeterd.

Vervolgens zijn twee mogelijkheden voor het toepassen van ruimtelijke variabiliteit getest voor het Ourthe stroomgebied: ruimtelijke variabiliteit in meteorologische invoer data en ruimtelijke variabiliteit in gemodelleerde afvoerprocessen. Toepassen van een ruimtelijke verdeling voor de meteorologische invoer data is gebaseerd op ruimtelijk variabele neerslag en potentiële verdamping. Deze zijn gemiddeld op vier verschillende niveaus in het model, zodat het effect van middeling op verschillende niveaus binnen het model getest kan worden. De gemodelleerde afvoerprocessen zijn ruimtelijk ver-

deeld door twee hydrologische eenheden (HE) te gebruiken, deze representeren valleien en hellingen. Uiteindelijk zijn een model structuur met en zonder ruimtelijke verdeling van gemodelleerde afvoerprocessen met elkaar vergeleken, allebei met de vier niveaus voor middeling van meteorologische invoer. Hieruit is geconcludeerd dat het ruimtelijk verdeeld toekennen van gemodelleerde afvoerprocessen belangrijker is dan het ruimtelijk verdeeld toekennen van meteorologische invoer data. Als er echter een ruimtelijke verdeling wordt toegepast voor de gemodelleerde processen, moet dit ook gebeuren voor de meteorologische invoer data.

Wetende dat het relevant is om een ruimtelijke verdeling toe te kennen aan de gemodelleerde afvoerprocessen, zijn gedetailleerdere conceptualisaties getest voor de Westelijke Ourthe, een substroomgebied van de Ourthe. Een extra hydrologische eenheid gericht op landbouw is geïntroduceerd, waarin oppervlakte afvoer ten gevolge van hoge neerslag intensiteit en ten gevolge van vorst in de toplaag van de bodem van belang zijn. Verder is een op temperatuur gebaseerde sneeuw module toegevoegd aan alle hydrologische eenheden. Door deze processen toe te voegen, is de prestatie en consistentie van de modellen verbeterd voor specifieke korte periodes. Echter, de geïmplementeerde processen en de bijbehorende signaturen zijn gevoelig voor fouten in invoer data en model uitschieters en moeten daarom zorgvuldig worden geïmplementeerd.

Deze thesis eindigt met twee verkennende vergelijkingen; in de eerste wordt gekeken hoe het nieuw ontwikkelde model voor de Westelijke Ourthe presteert in andere stroomgebieden; in de tweede wordt het nieuw ontwikkelde model vergeleken met andere modellen, waaronder de HBV configuratie die op dit moment wordt gebruikt voor operationele voorspellingen in het Maas stroomgebied. Deze vergelijkingen zijn uitgevoerd op basis van visuele inspecties van delen van de hydrograaf. De resultaten laten zien dat het nieuw ontwikkelde model vergelijkbaar presteert in naburige stroomgebieden. De vergelijking met andere modellen laat zien dat een zeer snelle oppervlakte afvoer component en een parallelle configuratie van snelle en langzame afvoerproductie belangrijk is om de dynamiek van verschillende tijdschalen in de hydrograaf te kunnen reproduceren. Beide aspecten zijn opgenomen in het nieuw ontwikkelde model. Dit resulteert erin dat het nieuw ontwikkelde model de meeste dynamiek van de hydrograaf beter kan reproduceren dan de operationele HBV configuratie, gebruikt op het moment van schrijven.

Ruimtelijk verdeelde en gedetailleerde proces conceptualisaties hebben een meerwaarde voor regen-afvoer modelering in het Ourthe stroomgebied. Echter, ze moeten zorgvuldig worden toegepast. Conceptuele modellen zijn sterke simplificaties van de werkelijkheid. Als je ze alleen confronteert met afvoer data, is er een risico dat andere hydrologische processen verkeerd worden geïnterpreteerd

Deze thesis doet een suggestie voor twee mogelijke kansen om conceptuele modellen verder te verbeteren. Ten eerste, zou het begrip van een stroomgebied vergroot kunnen worden door meer fysische processen toe te voegen aan de modellen, zoals de klimaat gebaseerde bergingscapaciteit van de wortelzone. Ten tweede, zouden satelliet data en puntmetingen gecombineerd kunnen worden om hydrologische processen op verschillende schalen aan elkaar te linken. Op deze manier zouden conceptuele modellen gebruikt kunnen worden om meer inzicht te krijgen in de schalingsvraagstukken die van belang zijn als je van de hellingschaal naar de stroomgebiedsschaa gaat.

LIST OF SYMBOLS

D	fraction of preferential recharge from S_r to S_s [-]
E_a	evaporation from agriculture reservoir [LT^{-1}]
E_c	performance combined from several metrics [-]
$E_{c,i}$	performance for an individual metric [-]
E_{NSE}	metric based on Nash Sutcliffe efficiency
E_{RE}	metric based on relative error
E_{RMSE}	metric based on root mean square error
E_{VE}	metric based on volume error
E_i	interception evaporation [LT^{-1}]
E_{pot}	potential evaporation [LT^{-1}]
$E_{pot,s}$	potential evaporation in case of snow cover [LT^{-1}]
E_r	evaporation from root zone reservoir (transpiration) [LT^{-1}]
E_w	sublimation of snow [LT^{-1}]
F_{AC}	signature representing coefficient of autocorrelation (lag = 24 h)
$F_{\text{AC,func}}$	signature representing coefficient of autocorrelation (1h < lag < 250h)
F_{DLD}	signature representing declining limb density
F_{FDC}	signature representing flow duration curve
$F_{\text{FDC,slp}}$	signature representing slope of normalised flow duration curve
F_{pks}	signature representing slope of peak distribution
F_{Q}	signature representing discharge
$F_{\text{Q,log}}$	signature representing the logaritm of the discharge
$F_{\text{Q,b}}$	signature representing yearly base flow
$F_{\text{Q,y}}$	signature representing average yearly discharge
F_{RC}	signature representing runoff coefficient
$F_{\text{RC,event}}$	signature representing event runoff coefficient
F_{RLD}	signature representing rising limb density
F_{SDL}	signature representing slope declining limb
F_{SRL}	signature representing slope rising limb
$F_{acc,fr}$	modelled accumulated frost [T][°C]
$F_{acc,fr0}$	max modelled accumulated frost resulting in $S_{a,min}$ [T][°C]
$F_{acc,fr1}$	min modelled accumulated frost resulting in $S_{a,max}$ [T][°C]
F_a	infiltration from S_a to S_r [LT^{-1}]
F_{dec}	decline of infiltration capacity [-]
F_{max}	maximum infiltration capacity [LT^{-1}]
F_t	reduction factor for $S_{a,max}$ [-]
I_{max}	maximum interception capacity [L]
K_c	coefficient for capillary rise [LT^{-1}]
$K_{f,a}$	coefficient for recession of $S_{f,a}$ runoff reservoir [T]
K_f	coefficient for recession of S_f runoff reservoir [T]

K_{FT}	coefficient for ratio between melt and frost speed of topsoil [-]
$K_{m,f}$	melt coefficient for melt of frozen topsoil [-]
$K_{m,P}$	multiplier for increased snow melt in case of liquid precipitation $[TL^{-1}]$
$K_{m,s}$	melt coefficient for snow melt $[LT^{-1}][°C^{-1}]$
K_p	coefficient for percolation $[LT^{-1}]$
K_s	coefficient for recession of slow runoff reservoir [T]
L_P	limit for potential evaporation [-]
P	precipitation $[LT^{-1}]$
P_s	solid precipitation $[LT^{-1}]$
Q_a	model flux representing hortonian ponding flow $[LT^{-1}]$
Q_c	model flux representing capilary rise $[LT^{-1}]$
$Q_{f,a}$	model flux representing hortonian overland flow $[LT^{-1}]$
Q_f	model flux representing fast runoff $[LT^{-1}]$
Q_i	model flux representing throughfall $[LT^{-1}]$
Q_p	model flux representing percolation $[LT^{-1}]$
Q_r	model flux representing preferential subsurface flow $[LT^{-1}]$
Q_s	model flux representing base flow $[LT^{-1}]$
Q_w	model flux representing snow melt $[LT^{-1}]$
$S_{a,max}$	maximum storage capacity in agricultural reservoir [L]
$S_{a,min}$	minumum storage capacity in agricultural reservoir (relative to $S_{a,max}$) [-]
S_a	model reservoir representing hortonian ponding [L]
$S_{f,a}$	model reservoir representing hortonian runoff generation [L]
S_f	model reservoir representing fast runoff generation [L]
S_i	model reservoir representing interception [L]
$S_{r,clm}$	climate derived root zone storage capacity [L]
$S_{r,max}$	maximum storage capacity in root zone storage [L]
$S_{r,soil}$	soil derived root zone storage capacity [L]
S_r	model reservoir representing root zone storage [L]
S_s	model reservoir representing groundwater (slow) runoff generation [L]
S_w	model reservoir representing snow storage [L]
T	transpiration $[LT^{-1}]$
T_a	air temperature [°C]
$T_{f,a}$	Base of lag function before $S_{f,a}$ [T]
T_f	Base of lag function before S_f [T]
T_m	threshold temperature for start snow melt [°C]
T_p	monthly averaged potential transpiration $[LT^{-1}]$
T_s	soil surface temperature [°C]
T_t	threshold temperature between liquid and solid precipitation [°C]
z	root depth [L]
α	albedo reflection coefficient [-]
β	coefficient determining storage discharge relation of S_r [-]
β_a	coefficient determining storage discharge relation of S_a [-]
$\Delta\theta_p$	water filled porosity at field capacity [-]
λ	latent heat of vaporisation from liquid to gaseous phase ($2.45*10^6$ Jkg^{-1})
λ_S	latent heat of vaporisation from solid to gaseous phase ($2.83*10^6$ Jkg^{-1})

LIST OF ABBREVIATIONS

CD	Cool-Dry (climate classification for S_r study in New Zealand)
CW	Cool-Wet (climate classification for S_r study in New Zealand)
FARM	Framework for Assessing Realism of Model structures
H	model configuration with only a Hillslope model element
HAND	Height Above the Nearest Drain
HOF	Hortonian Overland Flow
HRU	Hydrological Responds Unit
I	model state distribution level with only the reservoir representing Interception distributed
IR	model state distribution level with the reservoirs representing Interception and Root zone storage distributed
IRF	model state distribution level with the reservoirs representing Interception, Root zone storage and Fast runoff generation distributed
L	model state distribution level with all reservoirs Lumped
PC	Principal Component
PCA	Principal Component Analysis
SOF	Saturation Overland Flow
SSF	Subsurface Storm Flow
SWE	Snow Water Equivalent
WH	model configuration with a Wetland and Hillslope model element
WHS	model configuration with a Wetland and Hillslope and model element and a Snow module
WHP	model configuration with a Wetland, Hillslope and Plateau model element
WHPS	model configuration with a Wetland, Hillslope and Plateau model element and a Snow module
WHPF	model configuration with a Wetland, Hillslope and Plateau model element and a Frozen soil module
WHPSF	model configuration with a Wetland, Hillslope and Plateau model element and a Snow and Frozen soil module
WD	Warm-Dry (climate classification for S_r study in New Zealand)
WW	Warm-Wet (climate classification for S_r study in New Zealand)

CONTENTS

1

INTRODUCTION

It is particularly incumbent on those who never change their opinion, to be secure of judging properly at first.

Jane Austen (Pride and Prejudice)

1

1.1. Distribution in hydrological models?

1.1.1. Hydrological modelling

The origin of hydrological models depends on the definition used. If one defines a hydrological model as a (simplified) perception of the hydrological cycle (i.e., perceptual model; Beven, 2001), hydrological models are very old. In early writings philosophers already thought about the water cycle, but Leonardo Da Vinci (1452-1519) is named as one of the first hydrologists working with hypotheses and experiments to describe the hydrological cycle (Pfister et al., 2009). If one, however, defines a hydrological model as something being able to simulate or predict a hydrological process (i.e., conceptual or procedural model; Beven, 2001), hydrological models are younger. The work of Thomas Mulvaney (1822-1892) on the rational method can be seen as one of the first methods to simulate or predict discharge after a precipitation event (Stephenson, 1981).

The existence of two different definitions of a hydrological model already hints towards the double and interlinked purpose of model development. On one hand the more scientific purpose of increasing understanding of catchment functioning (e.g., Savenije, 2010; Zehe et al., 2013; Martínez-Carreras et al., 2015; Nippgen et al., 2015; Wrede et al., 2015; Hrachowitz et al., 2016), and on the other hand more practical purpose of making (reliable) discharge predictions or forecasts (e.g., Cloke and Pappenberger, 2009; Werner et al., 2013; Nicolle et al., 2014). Naturally, a good understanding of the catchment functioning is essential for making reliable predictions (e.g., Blöschl et al., 2013).

These two purposes, combined with large differences between catchments, have caused a wide variety of models and an almost 1:1 ratio of modellers to models. Models now come in many different versions: operational or scientific, physically based or conceptual, lumped or distributed, and in many more diversities. This abundance of models has advantages: different concepts can be tested for different areas and models can be selected based on the availability of data and required output. A disadvantage, however, is that new insights and developments are very scattered and difficult to combine (e.g., Weiler and Beven, 2015).

In addition to these general variations between models, there is also an often neglected contrast between operational and scientific models. Where scientific models are used for exploring catchment behaviour, operational models are used as a tool for decision making. Operational models set requirements for river routing, robustness, calibration, uncertainty analysis and bias correction, while these are less stringent for scientific models. On the other hand, the scientific models are essential to supply operational models with a good representation of the hydrological processes in a catchment. A combined effort may lead to sufficient understanding of the average catchment response, but the question remains how a catchment functions under extreme conditions and how it recovers after extreme events; under these conditions reliable forecasts are most relevant.

1.1.2. 'Distribution'

The Oxford Dictionary[1] gives the following definition of 'distribute': '*Give a share or a unit of (something) to each of a number of recipients*'. In hydrology 'distribute' or 'dis-

---

[1] Oxford Dictionary of English, second edition, Oxford University Press, 2003

tribution' has a lot of different definitions mainly focussing on spatial distribution. As a consequence, the methods of distribution need to be explicitly defined for each study. In this thesis I stick to the more general meaning, but applied to catchment hydrology: '*spatial distribution refers to any model or data selection that is not considered representative for a catchment as a whole*'.

This definition gives a lot of options to apply distribution in catchment modelling. In conceptual models there are mainly two options: spatial distribution of meteorological forcing (e.g., Oudin et al., 2004; Fenicia et al., 2008b; Lobligeois et al., 2014) and spatial distribution of dominant runoff processes. The latter again has many forms, for example incorporating different models, model concepts or model parametrisations for different parts of the catchment (Knudsen et al., 1986; Flügel, 1995; Beven and Freer, 2001; Uhlenbrook et al., 2004; Savenije, 2010; Hrachowitz et al., 2014). Especially distribution of runoff processes leads to including more specific process dynamics.

1.2. Process dynamics versus model complexity

Including more process dynamics, i.e. more detailed representation of runoff processes, in a model can lead to an improved model (e.g., Uhlenbrook et al., 2004; Clark et al., 2011; Brauer et al., 2014; Gharari et al., 2014a; Hrachowitz et al., 2014; Fenicia et al., 2016). However, when implemented carelessly, it quickly leads to a too complex model, of which the parameters are poorly identifiable and which will only function for the period it is calibrated on. Therefore, a balance should be found between the amount of represented process dynamics and the resulting model complexity (e.g., Perrin et al., 2001; Atkinson et al., 2002; Orth et al., 2015; Avanzi et al., 2016).

1.2.1. Creative use of data

Almost all hydrological modellers would answer 'more data' on the question what would be really helpful to improve their model. However, more data (i.e., measurements) is not always, or often not, available and if available, it is difficult to assess for which part of the study area it is representative. On the other hand, any form of distribution should be based on some form of data or expert knowledge: without known differences between two areas there is no point in using different models for those areas. In addition to data requirements for distribution, many catchment are ungauged (Hrachowitz et al., 2013a), meaning data is not at all available. These information requirements have led to various methods for creative use of available data sources.

In absence of ground measurements, satellites are increasingly used as a data source (Famiglietti et al., 2015). Estimates for all kinds of hydrologically interesting variables can be derived from different wavelengths measured by a large variety of satellites. Well known examples are land cover (LANDSAT, MODIS); soil moisture (SMOS/SMAP); vegetation parameters (MODIS) or snow cover extent (MODIS). These data cover large areas and spatial and temporal resolutions quickly increase. In addition to increased resolution, data becomes more and more available as complete products, enabling many people to use the data. Satellite data can among others be used to increase understanding of system functioning (e.g., Wang-Erlandsson et al., 2014; Humphrey et al., 2016; Simons et al., 2016), derive parameters (e.g., Winsemius et al., 2008; Gao et al., 2014b; Wang-

Erlandsson et al., 2016) or constrain and evaluate models (e.g., Gharari et al., 2014a; Sriwongsitanon et al., 2016).

However, satellite data is far from fully replacing ground data: ground measurements are still required to check satellite derived data. Serving as a support for satellite data is not or should not be the final destination of ground data, but until now ground data is often under exploited. Time series of ground data are often used for modelling and system understanding. However, by using a variety of hydrological signatures (i.e., specific characteristics of the data) a lot more information can be extracted from ground data (e.g., McMillan et al., 2011; Berghuijs et al., 2014). In some cases the hydrological signatures can even help to derive specific model parameters. (e.g., Fenicia et al., 2006).

1.2.2. MODEL CONCEPTUALISATION

When the available data gives reason to incorporate specific process dynamics, the model conceptualisation or model structure can be adapted. One of many methods to take into account specific process dynamics is by using stepwise model approaches and thus tuning your model structure to the dominant runoff process in the catchment (e.g., Fenicia et al., 2008b; Clark et al., 2008b; Fenicia et al., 2011; Clark et al., 2015). This works especially well in smaller areas with one or two strong dominant runoff processes (Fenicia et al., 2013).

When catchments become larger and more heterogeneous the output of the catchment generally becomes a mixture of different dominant processes. In these cases it is often not helpful any more to assume one or two dominant processes in the entire catchment. Instead, assigning different model structures to different parts of the catchment may be more suitable. Dominant processes can vary between different subcatchments, but it is more likely that they differ between different hydrological response units (HRUs). Thus, using spatial distribution of runoff processes based on HRUs may be beneficial to enhance the predictive power of a model (e.g., Hrachowitz et al., 2014; Fenicia et al., 2016).

Distribution based on HRUs requires division of the catchment into areas with comparable expected dominant runoff processes and selection of a model structure for each area. The division into HRUs and the selection of accompanying model structures are catchment dependent, because of the large variability of physical characteristics between catchments. However, many studies have shown the link between runoff processes and topographical indices, landscape elements or land cover (e.g., Beven and Kirkby, 1979; Rodhe and Seibert, 1999; Winter, 2001; Detty and McGuire, 2010; Savenije, 2010; Nobre et al., 2011), which can be helpful in delineating HRUs.

Sensibly delineating a catchment into HRUs and linking them to model conceptualisations is one way to include some physical meaning to conceptual models (e.g., Birkel et al., 2010). By increasing the physical basis of a conceptual model, the model is likely to be wider applicable and it can help to increase understanding of catchment functioning. Another option to add physical meaning is by linking a specific parameter to a physical catchment characteristic; for example, linking interception capacity to leaf area index (e.g., Bulcock and Jewitt, 2010; Wöhling et al., 2013) or linking root zone storage capacity to climate (Gao et al., 2014b; Wang-Erlandsson et al., 2016).

1.2.3. MODEL EVALUATION

Once distribution has been implemented in a model, the modeller needs to carefully check if the added process really adds value or only complexity. It is increasingly acknowledged that model evaluation based on single objective optimisation, often performed with Standard Least Squares optimisation, is insufficient to appropriately identify this added value. The use of hydrological signatures for (multi-objective) evaluation of the performance of hydrological models can give more information about the hydrological behaviour of the modelled catchments (e.g., Willems, 2009). The use of such hydrological signatures can therefore strengthen the link between a model and the underlying hydrological processes (e.g. Gupta et al., 2008; Yilmaz et al., 2008; Hingray et al., 2010; Wagener and Montanari, 2011).

In addition, the use of constraints based on expert knowledge (e.g., Gharari et al., 2014a; Nijzink et al., 2016) can help to narrow the parameter space to more realistic model realisations. During the entire evaluation process, it is important to assess whether the results represent possible catchment behaviour and really contain additional information and not only additional data (Das et al., 2008). This requires an adequate evaluation based on a broad set of diagnostic tools and performance metrics, which provide insight into different dominant behaviour of a catchment (Uhlenbrook et al., 2004; Clark et al., 2011; Gupta et al., 2008).

1.3. HYDROLOGICAL MODELLING IN THE MEUSE BASIN

The Meuse is a dominantly rain-fed river with quick hydrological response. Its catchment area is densely populated and the river is intensively used for, among others, navigation, cooling and drinking water intake. Therefore, pollution and flooding have large societal impacts, implying a need for reliable discharge predictions. To make reliable discharge predictions, the hydrological and hydraulic processes in the catchment need to be properly understood.

1.3.1. PREVIOUS MODELLING STUDIES

Due to the importance of reliable discharge predictions, a lot of hydrological modelling studies were performed for the Meuse basin. A part of the research focussed on (parametrisation of) the HBV-96 model, the rainfall-runoff model used by Rijkswaterstaat[2] at the moment of writing, partly under changing climate and land use conditions (Ashagrie et al., 2006; Tu, 2006; Booij, 2005). Other studies focussed more on other models to predict discharges (de Roo et al., 2000; van Deursen, 2000). In addition, de Wit et al. (2007) took into account the travel time of the flood waves in the tributaries.

1.3.2. WHAT IS NEW?

So, if so much work has already been done in the Meuse basin, what is the added value of another PhD thesis on modelling in Meuse basin? To start with, under extreme conditions the current operational model is not always able to give reliable predictions. This could for a large part be caused by the fact that the HBV model has been calibrated to

---

[2]Dutch public authority responsible for the design, construction, management and maintenance of the main infrastructure facilities

mimic the performance of specific areas, but that it has not been tailored to represent the dominant runoff processes in specific catchments. Parallel conceptual models based on a limited number of HRUs could help achieving this. Another cause of the limited performance of the current model is the absence of certain relevant runoff processes. Adding these processes, with their physical meaning, may improve the predictive power of the model.

Although this thesis focusses on the Meuse, it also tries to add to the general scientific discussion regarding the use and evaluation of distributed models in areas that are interesting from an operational forecasting point of view.

1.4. Research questions

Despite the diversity of studies carried out in the field of hydrological modelling, application of parallel landscape-based model structures combined with a range of model evaluation techniques did not yet receive full attention, especially not in the Meuse basin. Therefore this thesis tries to answer the question: 'What is the added value of distribution for rainfall-runoff modelling in the Meuse basin'. In order to answer this question, four other questions are relevant:

1. How can the results from distributed model experiments be evaluated?

2. How can the added value of distribution be tested?

3. What are the options to apply distribution and how useful are they?

4. How can the results be transferred to other catchments or entire basins?

The overall research question focusses on the Meuse basin; therefore, Chapter 2 will give a description of the Meuse basin including the Ourthe catchment, in which most of the model experiments have been carried out. The first two subquestions are strongly interlinked and are therefore combined into Chapter 3: this chapter gives a description of all model evaluation techniques used, together with the methods used for parameter selection.

The third question is investigated by a set of model experiments, which are described in Chapters 4 to 6. Chapter 4 has a special location among these chapters: it explores the option of using climate data to derive spatially variable root zone storage capacities, by using a set of New Zealand catchments. Chapters 5 and 6) adopt this experiment and investigate different distribution options in the Ourthe catchment: distribution of forcing and model structure (Ch. 5) and the implementation of more detailed runoff processes (Ch. 6).

Chapter 7 combines the results of the previous chapters and explores how the model persists when applied to neighbouring catchments or compared to other models, including the operational HBV configuration. To conclude, Chapter 8 summarises the results from all Chapters and discusses the implications, opportunities and consequences of the results. The research outline is schematised in Figure 1.1.
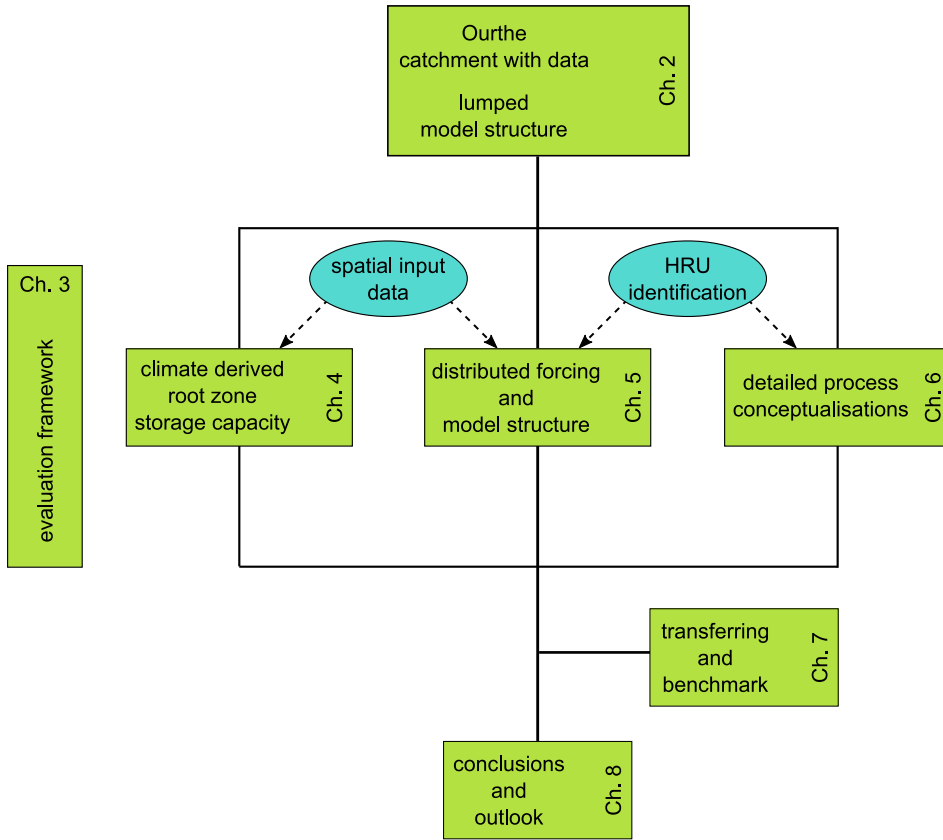
Figure 1.1: Schematic overview of the research presented in this thesis.

2

STUDY AREAS

*Van mijn gedrag zou het dus afhangen of de Nederlandsche universiteiten toen reeds voor
goed voor de vrouwen zouden worden opengesteld.
Thus, my behaviour would determine whether the Dutch universities would then forever
be opened to women.*

Aletta Jacobs (Herinneringen)

*The Meuse basin is the overall area of interest for this thesis, but the Ourthe and the Ourthe
Orientale were used for most of the model experiments. This chapter describes the catchment characteristics, hydrological response units and catchment response for a selection
of subcatchments of the Meuse basin. In addition, a set of New Zealand catchments was
used for two side experiments, the characteristics of these catchments are described in this
chapter as well.*

2

The Meuse basin is in principle the focus area of this thesis; however, the basin is relatively large for detailed model experiments, thus most model experiments were carried out in the Ourthe catchment and its subcatchments. This chapter describes all catchments used in this thesis, including those in New Zealand, which were used for the side studies in Chapters 3 and 4.

2.1. MEUSE BASIN

The Meuse basin (Fig. 2.1) was selected for this research, because of its importance for flood forecasting in the Netherlands. The response times of the subcatchments and the travel times through the rivers are relatively short, making proper discharge (and weather) predictions essential. The average precipitation ranges from 700 mmy^{-1} in the downstream part of the catchment to 1250 mmy^{-1} in the Ardennes (higher elevated part of eastern Belgium and Luxemburg); the average discharge in the Meuse is just over 300 mmy^{-1}. The discharge has a strong seasonal behaviour, which is caused by the seasonality in potential evaporation (de Wit et al., 2007). Liquid precipitation mainly dominates the runoff regime, but snow melt can have a large influence during some events.

   The topography of the Meuse basin varies throughout the catchment. The upstream part is mainly gently hilly with wide valleys. In the Ardennes the slopes are steeper and the valleys narrower, especially along the larger streams. The downstream part of the catchment consists mainly of the Belgium and Dutch lowlands (de Wit, 2008). The topography and stream network largely influence the catchment's behaviour under high flow conditions: floods in the Meuse are rather caused by the coincidence of peak flows from the different tributaries than by a flood peak travelling down the river (de Wit, 2008). There are even examples where there was a flood in the French part of the Meuse, but hardly an extreme discharge in the Dutch part (de Wit et al., 2007).

2.2. OURTHE CATCHMENT

2.2.1. WHY THE OURTHE CATCHMENT?

Within the Meuse basin, the Ourthe catchment (Fig. 2.2 and Tab. 2.1) is an interesting study area, as it is a mesoscale catchment of a size that is relevant for operational forecasting. The Ourthe contributes significantly to the total flow volume in the Meuse during floods, especially at the Dutch border. The Amblève and Vesdre catchments (located directly north ot the Ourthe catchment) have a large influence as well, but in these catchments the influence of artificial reservoirs is much larger than in the Ourthe. In addition, the large variability in topography and land use indicate that different runoff processes may be dominant in different areas of the Ourthe catchment. Finally, there is a clear spatial variability in precipitation (800-1250 mmy^{-1}), making this catchment interesting for investigating different distribution options.

2.2.2. CHARACTERISTICS OF OURTHE CATCHMENT

The catchment area of the Ourthe at Tabreux has a size of 1600km^2. The elevation ranges between 150m and 650m, with mild slopes in the upstream part of the catchment and steeper slopes along the main streams. The catchment responds quickly to precipita-
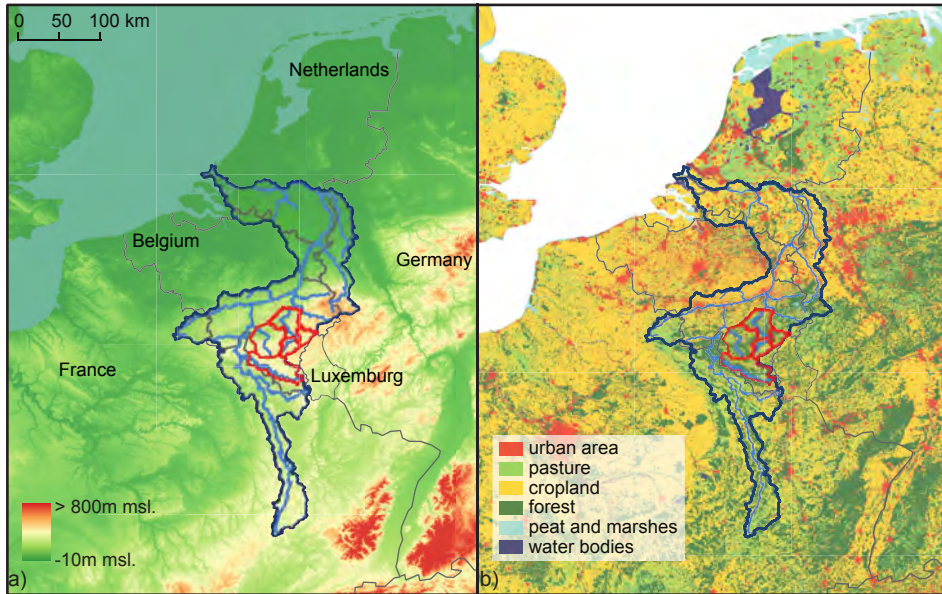
Figure 2.1: Meuse basin and river, with used subcatchments indicated; a) with elevation (Hy-droSHEDS, 2013); b) with land use (CORINE Land use map, European Environment Agency, 2006).

tion, due to shallow soils (Rakovec et al. (2012); the maximum travel time along the river is approximately 30 hours (Rakovec et al. (2012). The largest part of the catchment is used for agriculture (28% crops and 28% pasture), followed by forest cover (46%), while a small part of the catchment (6%) is built-up area (CORINE Land use map, European Environment Agency, 2006).

Distribution of model structure is based on hydrological response units (HRUs). In this thesis three different HRUs are used: wetlands, hillslopes and plateaus. During a field visit, it was observed that flat areas were present along most of the streams. For these areas it is expected that the groundwater levels are shallow and rise quickly during precipitation events; they are defined as *wetlands*. Flat areas are also located further away from and higher above the streams, these areas are mainly used for agriculture (i.e., crops and pasture). It is expected that these areas recharge to the groundwater and that they respond to precipitation via subsurface storm flow or Hortonian overland flow; they are defined as *plateaus*. The remainder of the catchment consists of (slightly) sloped forested areas. These areas are expected to recharge to the groundwater and respond to precipitation via subsurface storm flow; they are defined as *hillslopes*. The distribution of the different classes is shown in Figure 2.2b.

The distribution of the HRUs is based on a landscape classification derived from elevation (HydroSHEDS, 2013) and land use (CORINE Land use map, European Environment Agency, 2006) data. The classes are constructed based on thresholds for Height Above Nearest Drain (HAND; Rennó et al., 2008; Gharari et al., 2011; Nobre et al., 2011) and land use. Areas with a slope smaller than 0.13 and a HAND lower than 1 m, with

Table 2.1: Characteristics of the Ourthe catchment and its subcatchments Ourthe Orientale and Ourthe Occidentale.

	Ourthe (Tabreux)	Ourthe Orientale (Mabompré)	Ourthe Occidentale (Ortho)
location outflow	5°31'48"E, 50°26'24"N	50°8'24"N, 5°43'12"E	50°6'36"N, 5°39'36"E
catchment area (km^2)	1607	317	379
max elevation (m)	663	663	597
min elevation (m)	107	294	303
elevation range (m)	556	369	294
mean slope (-)	0.090	0.081	0.077
max slope (-)	0.75	0.62	0.58
max flow distance (km)	144	32	44
forest cover (%)	46	48	40
pasture cover (%)	21	20	23
urban cover (%)	6	5	4
crop cover (%)	27	27	33
wetland (%)	8	9	9
hillslope (%)	46	46	39
plateau (%)	46	45	52
mean annual precipitation (mmy^{-1})	1000	1080	1010
mean annual runoff (mmy^{-1})	460	480	500
mean annual temperature (°C)	9.6	9.1	9.3
mean annual pot evaporation (mmy^{-1})	730	710	720

a constant stream initialisation threshold of 0.09 km^2, were classified as wetlands and areas with crop or pasture land cover were classified as plateau. The remainder of the catchment was classified as hillslope. The percentages for each HRU for the Ourthe and its subcatchments can be found in Table 2.1. This classification of HRUs applies for the studies in Chapters 6 and 7. For Chapter 5 a slightly different classification was used: areas with a slope lower than 0.13 and a HAND lower than 5.9 m are classified as wetland, with a constant stream initialisation threshold of 1.8 km^2. The remaining area was classified as hillslope. These threshold values for slope and HAND are the same as those used by Gharari et al. (2011) for an adjacent catchment with similar climatic and geomorphic characteristics. The value for stream initialisation is larger than the value used by Gharari et al. (2011) to prevent arable land and forested area to be classified as wetland. Both methods for classification lead to approximately 10% of wetland in the Ourthe.
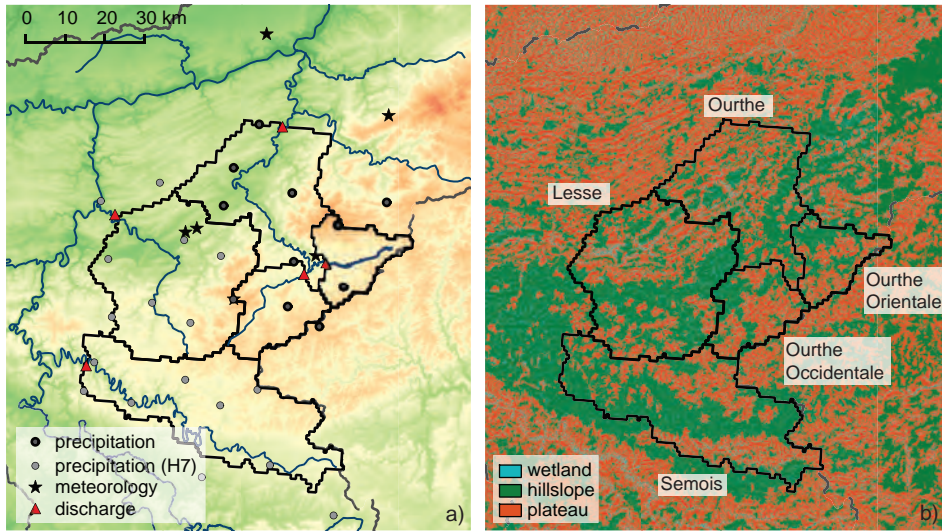
Figure 2.2: More detailed view of Ourthe and other used subcatchments. The dark dots indicate the precipitation gauges used for Chs. 5 and 6, the lighter ones are only used for the explorative comparison in Ch. 7. a) DEM (HydroSHEDS, 2013) with precipitation, meteorological and discharge gauges (the elevation scale is equal to the one in Figure 2.1); b) derived hydrological response units. Note that the wetlands are very narrow and generally coincide with the streams.

2.2.3. Subcatchments Ourthe

In addition to the entire Ourthe, also the two main tributaries of the Ourthe are used in this thesis: Ourthe Orientale (eastern side) and Ourthe Occidentale (western side). The Ourthe Orientale and Occidentale cover an area of 317km^2 and 379km^2 respectively, with an elevation difference of approximately 300 m. Both areas are characterised by gently sloped forests (39-46%) and flatter agricultural fields (45-52%).

Both catchments are hydrologically comparable: the Ourthe Orientale receives slightly more precipitation and produces slightly less runoff than the Ourthe Occidentale. Both catchment have a seasonal hydrological response, caused by the seasonality of the potential evaporation. Therefore, the evaporation is mainly energy constrained, but during warm summer periods the evaporation often shifts to being moisture constrained. The travel time through the river is approximately 10 hours for both catchments (Rakovec et al., 2012).

2.2.4. Response of Ourthe catchment

The annual average precipitation and potential evaporation in the Ourthe are approximately 900 mmy^{-1} and 700 mmy^{-1} respectively. These meteorological conditions result in an average annual runoff of 400 mmy^{-1}. The discharge varies between years and between seasons (Fig. 2.3); the former is mainly caused by variation in precipitation (Fig. 2.3a) and the latter by variations in potential evaporation (Fig. 2.3b).

Not only total yearly or monthly amounts vary, but also the discharge pattern differs
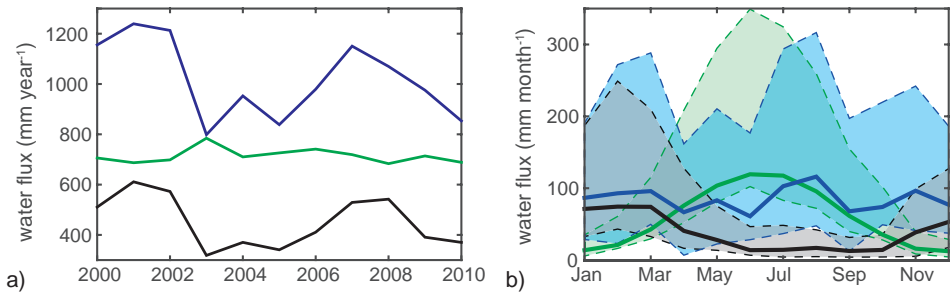
Figure 2.3: Patterns in precipitation (blue), potential evaporation (green) and discharge (black) for Ourthe catchment; a) yearly sum; b) monthly sums: the solid line gives the mean monthly sum, the shaded area the minimum and maximum monthly sum.

between years. Figure 2.4 shows discharge and precipitation for three years (2003, 2008 and 2009). Maximum (winter) discharges vary between 0.3 and 1 mmh^{-1}, as well as the discharge pattern in summer: discharges can be very low, but in case of high intensity precipitation events, very peaky as well. The bottom row shows the duration curves of discharge and precipitation, with the discharge being colour coded by season. These plots show again that peak flows can or mainly occur in winter (e.g., in 2003), or be spread over the entire year (e.g., in 2008).

2.2.5. DATA SERIES FOR MODELLING EXPERIMENTS

DATA AVAILABILITY

For this study, precipitation[1], potential evaporation[2] and discharge[1] data were used from 01 September 1999 - 01 May 2011. Hourly precipitation data were available for 28 stations in or within close proximity of the study catchments (dots in Fig. 2.2a), of which 11 stations (dark dots) are used for the majority of the model experiments. Hourly discharge data were available at the outlet of each subcatchment (triangles in Fig. 2.2a). Data for potential evaporation originated from six stations (stars in Fig. 2.2a), each with varying data availability. Potential evaporation was calculated with the Penman equation (Penman, 1948) on a daily basis and downscaled to hourly data with a sine function and day length derived from global radiation, via Equation 2.1. Gaps in the meteorological data were filled based on data of the remaining stations and correlations of meteorological variables between these stations. Each experiment uses different parts of this data set: specifics are included in each chapter.

$$E_{p,i} = \begin{cases} -A\cos(G(x+H)), & \text{if } R_{s,i} > 0 \\ 0, & \text{if } R_{s,i} < 0 \end{cases} \tag{2.1}$$

---

[1] Made available for this study by SPW-DGO2-Direction de la Gestion hydrologique intégrée, Bld du Nord 8 à B-5000 Namur, Belgique

[2] Required data was made available for this study by L'Institut Royal Météorologique, Belgique

with:

$$A \quad = \quad \frac{-E_p P}{\sin(P(D_e + H)) - \sin(P(D_s + H))}$$

$$G \quad = \quad \frac{2\pi}{D_l}$$

$$D_l \quad = \quad D_e - D_s + 1$$

$$H \quad = \quad D_l - 12$$

$x \quad = \quad$ hour of the day

$D_s \quad = \quad$ first hour of the day with $R_{s,i} > 0$

$D_e \quad = \quad$ last hour of the day with $R_{s,i} > 0$

$E_p \quad = \quad$ daily potential evaporation $(\text{mm}\,\text{d}^{-1})$

$R_{s,i} \quad = \quad$ hourly global radiation

$E_{p,i} \quad = \quad$ hourly potential evaporation $(\text{mm}\,\text{h}^{-1})$
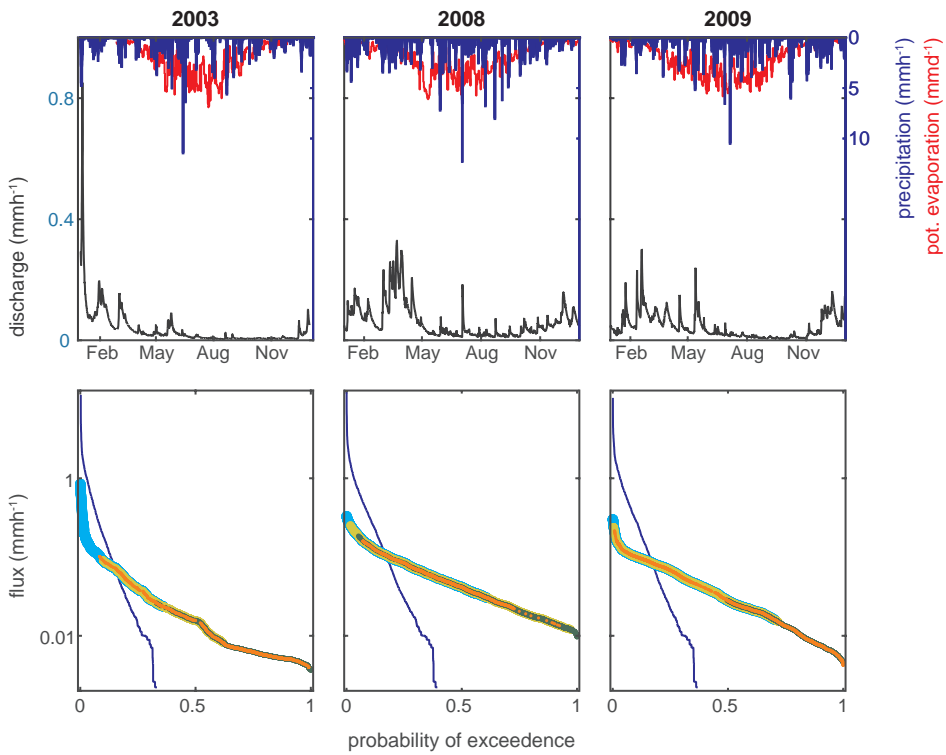

Figure 2.4: Discharge (black), precipitation (blue) and potential evaporation (red) for three characteristic years. Top: plotted versus time; bottom: plotted versus probability of exceedance. In the bottom row discharge is colour code for time of occurrence (blue: Jan-Mar, light green: Apr-Jun, dark green: Jul-Sep, orange: Oct-Dec).

2

To assess the influence of incorporating distributed forcing, a data set with distributed meteorolgical forcing was required. To this end, the catchment was divided into 'calculation cells', corresponding to the individual Thiessen polygons around the precipitation stations (dots in Fig. 2.2a). Distribution of potential evaporation data was first based on elevation; the catchment was divided into three (Ch. 5) or four (Ch. 6) elevation zones, each containing one or two meteorological stations. These elevation zones were then used to calculate the averaged potential evaporation for each calculation cell.

2.3. LESSE AND SEMOIS CATCHMENTS

The Lesse and Semois catchments are only used in Chapter 7 to investigate the implications of applying the derived results to other (nearby) catchments. The Lesse is very similar to the Ourthe, especially in terms of precipitation, discharge and topography. Only, the Lesse has less pasture and more forest land cover. The Semois differs more from the Ourthe: precipitation is about 25% higher in the Semois and the upstream parts of the Semois are relatively flat. Similarly to the Ourthe catchment, these two catchments have steep slopes along the larger streams.

2.4. CATCHMENTS FOR SIDE STUDIES

Two other sets of catchments have been used in addition to the Meuse and Ourthe catchments. These catchments were used for two specific analyses: development of FARM (Framework to Assess Realism of Model structures) and deriving root zone storage capacity from climate data. For the former, a small catchment was required, which are not gauged for longer periods in the Ourthe; for the latter a set of catchments with a strong climatic gradient was required.

2.4.1. EXPERIMENT FOR DEVELOPING FARM

Chapter 3 describes among others the development of a Framework to Assess Realism of Model structures (FARM). The aim of the development of FARM is to assess which model structure suits a catchment better. Assigning one model structure to the entire catchment is only feasible for small catchments. Therefore, two small headwater catchments have been selected for the FARM case study: the Maimai M8 catchment in New Zealand (0.038 km^2) and the Wollefsbach catchment in Luxembourg (4.6 km^2). The purpose of these catchments in this thesis is purely for illustration, therefore only the results for the Maimai are presented; the results for the Wollefsbach can be found in Euser et al. (2013).

2.4.2. MAIMAI

The Maimai M8 catchment is located in the northern part of New Zealand's South Island (Fig. 2.5a). The Maimai has short, steep slopes and shallow soils, where saturation seldom decreases below 90%. The yearly precipitation and discharge are approximately 2600 mm/year and 1550 mm/year, respectively. More information about this catchment and previous research is described in a review by McGlynn et al. (2002). The wet climate with little seasonality leads to a system with a limited number of hydrological regimes.
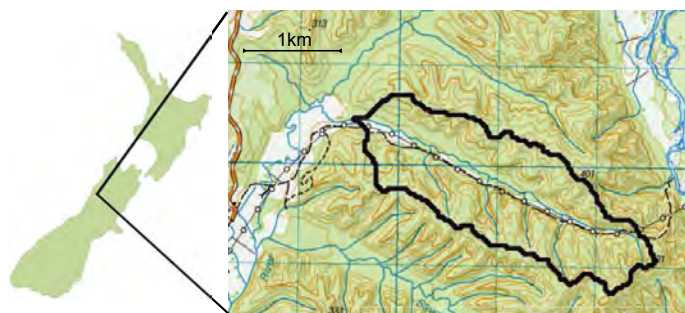
Figure 2.5: catchment used for developing FARM: Maimai in New Zealand.

The steep slopes together with the shallow, saturated soils and the impermeable subsurface lead to a quick response of the catchment (Vaché and McDonnell, 2006).

Hourly data of discharge, precipitation and potential evaporation from 01 January 1985 till 31 December 1987 were used[3]. The precipitation was measured with a recording gauge which is located inside the catchment. The potential evaporation was estimated as described by Rowe et al. (1994). The first year of the data was used as a warm-up period; the last two years were used for calibration, following the first year was used again as validation period.

2.4.3. EXPERIMENT ROOT ZONE STORAGE CAPACITY

Chapter 4 describes an experiment to derive root zone storage capacity from climate data. As the climatic variation within the Ourthe is small, 32 New Zealand catchments were used for this experiment. The combination of north-south oriented mountain range (Southern Alps) and prevailing westerly winds causes a strong climatic gradient over a distance of 200 km. Yearly precipitation ranges from less than 0.6 my^{-1} on the eastern (lee) side to more than 10 my^{-1} on the western (windward) side. Mean annual temperatures also vary across the country from 16°C in the north to 10°C in the south (NIWA, 2015). Before human colonisation the predominant land cover was indigenous forest; this forest is now confined to the mountain ranges, with pasture and crop land dominating elsewhere.

Table 2.2 and Figure 2.6 show the location and main characteristics of the 32 catchments[4]; they were mainly selected for variability in size (4th to 7th Strahler-order streams), climate and land cover. An example of differences in climate is shown in Figure 2.7a: for each climate category in Table 2.2 the average monthly precipitation and potential evaporation are shown. Catchments with more than 20 years of discharge data were selected from the set used by Booker and Woods (2014), containing catchments with limited human influence. Catchments with lake or glacial influence were not selected to prevent

---

[3]Made available for this study by John Payne and Lindsay Rowe (Landcare NZ) and professor Jeff McDonnell (University of Saskatchewan)

[4]Data of these catchments was made available for this study by NIWA and the NZ Regional Councils
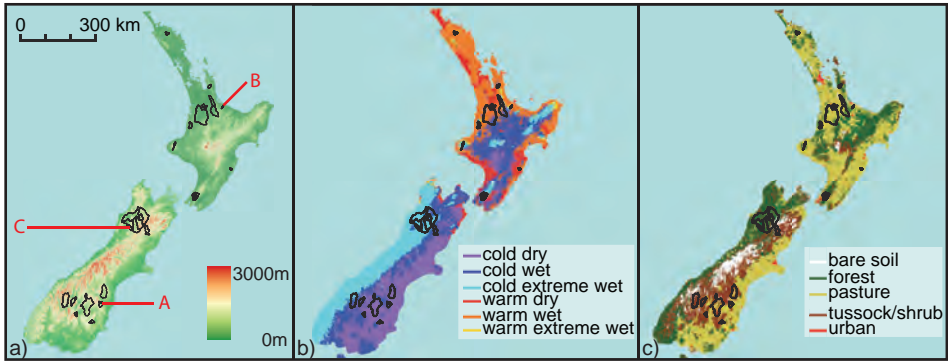
Figure 2.6: Catchments in New Zealand used for the root zone storage capacity analysis; each panel has a different background: (a) elevation (A, B and C indicate the catchments used in Figures 4.6 and 4.9); (b) climate; and (c) land cover.

Table 2.2: Number of selected gauges for combinations of climate[a] and land cover.

	Indigenous forest	Grasses[b]	$\overline{P}$ (my^{-1})	$\overline{E_p}$ (my^{-1})	$\overline{Q}$ (my^{-1})
warm-wet	1	6	1.8	0.9	1.2
warm-dry	0	1	1.1	0.9	0.4
cool-wet	14[c]	4	2.5	0.8	1.8
cool-dry	0	6	1.0	0.8	0.3

[a] warm: $\overline{T_{year}} > 12°C$, cool: $\overline{T_{year}} < 12°C$, dry: $\overline{P - E_p} < 0.5$ my^{-1}), wet: $\overline{P - E_p} > 0.5$ my^{-1} (this category contains areas classified as 'wet' and as 'extremely wet' in Fig. 2.6).

[b] This category contains both pasture and tussock grasses.

[c] This category contains one catchment with shrub land cover.

the effect of interannual storage changes. Finally, some nested catchments were specifically selected; these were used to investigate several possible methods to disaggregate the climate derived, catchment representative S_r to nested subcatchments.
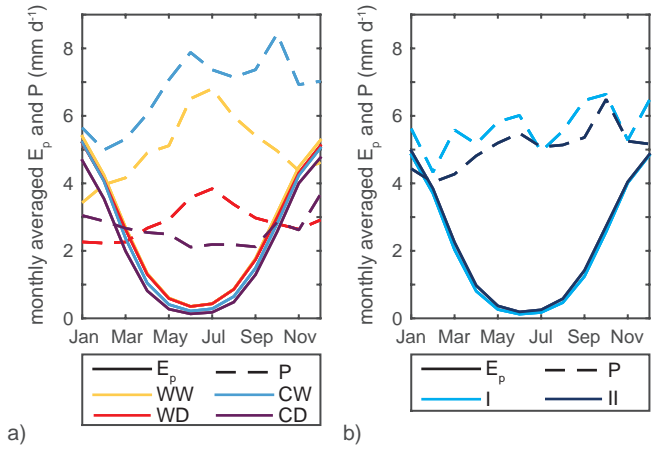
Figure 2.7: Monthly averaged precipitation (*P*; dashed lines) and potential evaporation (E_p; solid lines); a) Averaged values for each climate category in Tab. 2.2 (WW = warm-wet, WD = warm-dry, CW = cool-wet, CD = cool-dry); b) Values for catchments I and II, being classified as warm-wet with grass land cover, in Fig. 4.5.

3

MODEL CONFIGURATION

Le vrai est trop simple, il faut y arriver toujours par le compliqué.
The truth is too simple: one must always get there by a complicated route.

George Sand (Correspondence)

*Model **set-up**, **conditioning** and **evaluation** are essential elements of model experiments. At the same time they can be similar for a range of model experiments. This chapter discusses all three elements, to emphasise the importance of setting up model experiments systematically. The model **configurations** are based on a modular set-up, which enables adapting them to the expected runoff processes and the catchment. The main aim of the model **conditioning** is obtaining a set of behavioural model realisations, based on expert knowledge. The model **evaluation** is based on hydrological signatures and consists of a set of tools to evaluate different aspects of the model.*

3.1. INTRODUCTION

Incorporating more distribution in conceptual models comes with two important issues: 1) How can the results from distributed model experiments be evaluated? 2) How can the added value of distribution options be tested? Although these questions are strongly linked; they are not equal. The first one questions the amount of information needed to evaluate model results. Discharge is a very important source, but how much information does discharge contain regarding individual processes, as it is the result of the mixing of different processes. The second question assumes that the adapted model structure is evaluated as being reasonable and following this poses the question if the possible increase in performance is not solely a consequence of increased model complexity. Despite their different background, they can possibly be answered by fully exploiting discharge data and investigate the use of other data sources for the evaluation of individual processes.

Apart from these two questions, models need to be conditioned (or calibrated) before they can be evaluated. Conceptual models will not give you any output as long as no values are assigned to the model parameters. Thus, three elements are important before starting model comparison: 1) a general model set up; 2) guidelines for model conditioning and 3) evaluation tools. These three elements have a large overlap for the model experiments in this thesis; therefore, they are combined into this chapter.

3.2. MODEL SET-UP

The model set-up used in this thesis largely follows the modular set up used by Fenicia et al. (2008b, 2011). Each model structure consists of a set of reservoirs, closure relations (fluxes) and the accompanying parameters. Different hydrological processes are then represented by one or two reservoirs. This modular set-up enables incorporating different processes for different catchments. The basic model configuration is shown in Figure 3.1: this configuration was used as a starting point for the described model experiments. It consists of four reservoirs: an interception (S_i), an root zone (S_r), a fast runoff generating (S_f) and a slow runoff generating reservoir(S_s). The root zone storage is the reservoir from which transpiration is modelled; other models also use 'unstaturated zone' or 'soil moisture storage' for this reservoir. A non-linear power function is used to compute the outflow from S_u and a lag function is implemented before S_f. For a more detailed description of the different reservoirs, I refer to Gao et al. (2014a) and Fenicia et al. (2008b). A brief comparison between lumped model structures (results not presented) showed that this model structure is able to give a good overall performance for the Ourthe catchment.

In addition to configuring model structures for specific catchments, the modular set-up can also be used to configure model structures for specific areas within a catchment. The use of different model configurations can be used for catchments where multiple hydrological response units (HRUs) can be identified. In this way not only the model parameters can be varied between HRUs, but also the entire model configuration and thus the modelled processes.

The model structures belonging to different HRUs in a catchment need to be connected. This can be done in two ways: by connecting them in series or by connecting

them in parallel. In case of a serial connection, lateral fluxes are exchanged between HRUs within a grid, while in case of a parallel connection the lateral exchanges only take place through the groundwater or the stream network. Most approaches that use different model structures for different HRUs connect these (partly) in series (e.g., Flügel, 1995; Seibert et al., 2003; Uhlenbrook et al., 2004; Fenicia et al., 2008a). However, here model structures belonging to different HRUs are connected in parallel (parallel modelling), which limits model complexity and assures that different units can be treated as separated lumped model structures (e.g., Savenije, 2010; Gao et al., 2014a), provided that the model structures are arranged according to their hydrological connectivity.

3.3. MODEL CONDITIONING

3.3.1. PARAMETER SELECTION

Before the different model configurations can be compared in a model experiment, parameter values have to be selected which result in behavioural model realisations. For short term forecasting applications selecting an optimal parameter set is important to make the most reliable predictions. However, large computational efforts are required to determine a (mathematical) optimal parameter set. Besides, the mathematical optimum is often not the same as the hydrological optimum (Beven, 2006; Andréassian et al., 2012), which is hard to obtain due to the complexity of runoff processes. On the other hand, for studies analysing suitable model structures for a certain catchment, finding the hydrological or mathematical optimum is not necessary, but rather a set of adequately performing parameter sets is useful.

The purpose of such studies, including the model experiments in this thesis, is namely to compare a set of model structures for one or more catchments. Thus, the behaviour of the model structure is important under a range of behavioural parameter sets for a given catchment. The definition of what is behavioural can change per model experiment, but in this thesis it is defined as fulfilling all relevant constraints. Here, constraints are relationships between parameters and fluxes, based on expert knowledge, which describe the expected model behaviour (e.g., Gao et al., 2014a; Gharari et al., 2014b; Hrachowitz et al., 2014); only if a model realisation satisfies all constraints, the realisation is assumed
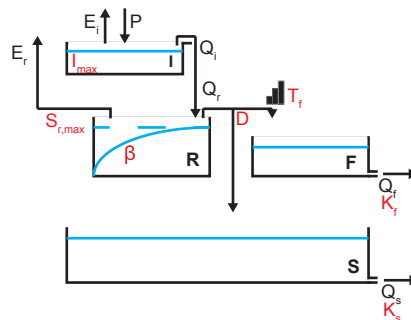

Figure 3.1: Basic model configuration with four reservoirs and seven parameters. In a preliminary analysis it was found that this model structure gives reasonable results for the Ourthe catchment.

to be behavioural, and therefore, useful for further analysis. This method for the selection of model realisations is similar to the regularisation methods used in environment system modelling (e.g., Maneta and Wallender, 2013; Eshagh et al., 2013).

For most model experiments a wide prior parameter range was selected based on expert knowledge and narrowed down with a preliminary sampling from this prior range. From the remaining parameter range, parameter sets were selected that satisfied all imposed parameter constraints for a specific model structure. These parameter sets do not necessarily contain the mathematical or hydrological optimum. Instead, they satisfy all posed constraints and are therefore a good population for further analysis.

3.3.2. MODEL CONSTRAINTS

In this thesis two types of constraints were used: parameter and process constraints (e.g., Gharari et al., 2014b). Parameter constraints were applied to the parameter values before the model run, and process constraints were applied to the modelled fluxes after the model run. By applying these constraints, the model behaviour is forced to follow the modeller's perception and thus can be expected to give more plausible results as the feasible parameter space is reduced. The applied constraints are comparable to those used by (Gharari et al., 2014a) and summarised in Table 3.1; the constraints were not the same for each experiment, thus Table 3.1 also indicates which constraints were used for which model configurations in which chapters. The selected values for each constraint are specified in the relevant chapters, together with an indicating of the most influential constraints.

One parameter constraint could be applied to all model structures, namely the recession coefficient K_s of the slow reservoir. K_s was considered to be fixed, and determined using a Master Recession Curve (e.g., Lamb and Beven, 1997; Fenicia et al., 2006; Yadav et al., 2007; Hrachowitz et al., 2011). Additional parameter constraints were applied to model structures with multiple HRUs. These additional parameter constraints describe the difference in behaviour between the HRUs.

More process constraints can be applied to model structures with more HRUs incorporated as well. This thesis applies four kinds of process constraints; values may vary per case study and are therefore specified in Chapters 5 and 6. First, the runoff coefficient (F_{RC}) of the entire catchment; the modelled F_{RC} should not deviate too much from the observed value. This constraint was used for the entire calibration period and for periods of low and high flow; Second, boundaries for yearly actual evaporation are derived from the Budyko framework (Budyko, 1974). Third, boundaries for runoff coefficients were applied to the relation between HRUs as well. Fourth, the Hortonian overland flow module appeared to be very sensitive, leading to incidentally very high peaks; therefore, a maximum was set for summer discharge.

The model structures with multiple HRUs included have more calibration parameters; however, additional constraints can be applied as well, which partly compensate for the additional complexity, by narrowing the feasible parameter space.

Table 3.1: Parameter and processes constraints used in this thesis.

	Model configuration	Chapter
Parameter constraints		
$I_{max;W}$[a]$< I_{max;P} < I_{max;H}$	WH-WHPSF[b]	5,6
$S_{r;W} < S_{r;H} = S_{r;P}$	WH-WHPSF	5,6
$\beta_W = \beta_{a;P} < \beta_H = \beta_P$	WH-WHPSF	5,6
$L_W < L_P < L_H$	WH-WHPSF	6
$K_{f;W} = K_{fa;P} < K_{f;H} = K_{f;P}$	WH-WHPSF	5,6
$D_H < D_P$	WHP-WHPSF	6
$K_{p;H} < K_{p;P}$	WHP-WHPSF	6
$K_s = 2500$ (h)	all	5,6
Process constraints		
min/max for $F_{RC;tot}$[c]	all	5,6
min/max for $F_{RC;low}$[d]	all	5,6
min/max $F_{RC;high}$[e]	all	5,6
min/max for E_{tot}[c]	WH-WHPSF	6
$F_{RC;f,W} > F_{RC;f,H}$	WH-WHPSF	6
$F_{RC;fa,P} > F_{RC;f,P}$ (if $Q_{fa;P} > 0.001$)	WHP-WHPSF	6
$\sum(Q_{f,H;high} * H_\%) > \sum(Q_{f,W;high} * W_\%)$	WH	6
$\sum(Q_{f,P;high} * P_\%) > \sum(Q_{f,W;high} * W_\%)$	WHP-WHPSF	6
max Q_{low}	WHP-WHPSF	6

[a] The subscripts 'W', 'H' and 'P' refer to the HRUs: wetland, hillslop and plateau.

[b] only the simplest and most detailed configuration are indicated.

[c] Applied for the entire modelled period.   [d] Applied for the low flow period, the specific time spans are selected from the calibration data.   [e] For the high flow period, the specific time spans are selected from the calibration data.

3.4. MODEL EVALUATION

Model evaluation has many different aspects and can be carried out in many different forms. In this thesis, it is split into three levels: internal model behaviour, model performance and consistency and predictive power. All three levels contain different tools; not all tools are useful for all model experiments. However, the model experiments in Chapters 5 and 6 contain at least one tool from each level. Below all tools are described which are used in one of the model experiments of this thesis. Key elements in these evaluation tools are the assessment of performance and consistency, so these terms are first explained.

3.4.1. PERFORMANCE AND CONSISTENCY

The *performance* of a model structure for a certain catchment is determined by its ability to reproduce a certain hydrological behaviour or signature. This can be measured with the best score for an evaluation criterion, which describes this hydrological signature,

and by the range of values covered by the evaluation criterion (belonging to all the parameter sets from the posterior distribution). In this thesis, the overall performance (E_c) of a model structure is combined from the performance for each evaluation criterion with Equation 3.1 (Hrachowitz et al., 2014).

$$E_c = \sqrt{\sum_{i=1}^{n} E_{c,i}^2}$$ (3.1)

The *consistency* of a model structure for a certain catchment is determined by the number of evaluation criteria, describing different hydrological signatures, that have their best performance for a specific parameter set. The consistency of model structures can vary gradually between fully consistent and fully inconsistent. It is important to have insight into the consistency of model structures for two reasons: first, a high consistency means that the model is capable of reproducing several hydrological signatures with the same parameter set, implying a better representation of real world processes (i.e. the model can reproduce different, ideally contrasting, aspects of the hydrograph). Second, a highly consistent model is thus expected to behave comparably in the calibration and validation period (Kirchner, 2006; Fenicia et al., 2007), satisfying a split-sample test (Klemeš, 1986), and would therefore have a reduced predictive uncertainty.

The consistency and performance of a model structure can be determined independently, but are both important for the evaluation of the model structures (Wagener et al., 2003). Only a model with high performance and high consistency may be considered a suitable hypothesis for a certain catchment and, therefore, points towards a high degree of realism. In reality all signatures occur simultaneously. Hence, a model that is able to reproduce all selected signatures to a high degree with the same parameter set has a higher degree of realism than a model structure that is not able to do that.

However, it is possible that, for a certain model structure, the degree of performance is different from the degree of consistency. The consequences for different combinations of the degree of consistency and performance are shown in Figure 3.2. For an inconsistently good model structure, signatures are reproduced well, but not with the same parameter set. For a consistently poor model structure, signatures are not represented correctly, although the model is consistent. So, a high degree of consistency only gives extra value in the evaluation process when it is combined with a high performance.

3.4.2. INTERNAL MODEL BEHAVIOUR
DISTRIBUTION OF MODEL STATES
A model configuration with the highest model state distribution level (interception, root zone, fast runoff and slow runoff reservoirs distributed) can be used to assess the spatial variability of the fluxes from each reservoir. A significant remaining spatial variability of a flux leaving a reservoir gives reason to distribute the reservoir (e.g., Lobligeois et al., 2014). To assess the spatial variability of a given flux, the flux in a specific calculation cell was subtracted from the spatial average of all calculation cells in the catchment, for each cell and for each time step in the calibration and validation period. Subsequently, the cumulative distributions of these differences were compared for each flux.
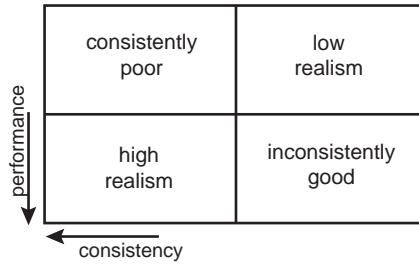
Figure 3.2: Consequences for model structures for different combinations of performance and consistency, under the condition that the uncertainty of the input data is limited. The use of signatures for the evaluation of performance and consistency limits the influence of input uncertainty.

CONTRIBUTION OF DIFFERENT FLUXES

The effect of distribution of model structure on model behaviour was tested by comparing the different contributions to the discharge for different model configurations. The contributions of different runoff fluxes to the total discharge were visually compared. With this method, the composition of the discharge can be compared for different model structures, and thus the influence of model structure distribution on the contribution of different processes.

EVALUATION OF MODEL STATES

In addition to runoff fluxes, model states tell a lot about the model behaviour as well. The dynamics of two model states (snow reservoir and accumulated frost) were compared with additional data sources. Modelled snow water equivalent was compared with MODIS/Terra daily snow cover MOD10A1 (Hall et al., 2006) with a binary approach: the MOD10A1 data set was only used to indicate presence or non-presence of snow cover. Accumulated frost (in the topsoil) was compared with temperature measurements of the soil surface. When the accumulated frost (sum of daily temperatures with a maximum of zero) is negative it is expected that the observed surface temperature is below 0°C as well.

3.4.3. MODEL PERFORMANCE AND CONSISTENCY

HYDROGRAPH COMPARISON

Evaluation criteria based on signatures are a valuable tool to compare models structures. However, some observed features are difficult to capture in a signature. Therefore, a visual hydrograph comparison can give additional insight into model behaviour. In this thesis the observed hydrograph is compared with a single model realisation from different model structures. From this comparison possible differences between the model configurations can be identified for specific events or marked periods within the season. Furthermore, the shape of the modelled hydrograph can be compared for different model configurations.

3

FARM

To evaluate the reproduced dynamics of the modelled hydrographs more quantitatively, the FARM analysis (Sect. 3.5) was used. A requirement for applying FARM is selecting a set of hydrological signatures. For each model experiment in this thesis a different subset of signatures was selected from Table 3.3. Categories for line thickness and colour coding were defined per experiment and are thus described in the relevant chapters.

3.4.4. PREDICTIVE POWER

SCATTERPLOTS

Scatterplots can be constructed from two evaluation criteria for multiple model realisations. They often reveal the possible trade-off between the two evaluation criteria and when carried out for different periods they can reveal a difference in performance between these periods as well. Here, Scatterplots were constructed for the score on two evaluation criteria for each model configuration, for all behavioural model realisations.

PERFORMANCE AND CONSISTENCY

With all the tools described above it is difficult to incorporate all model configurations, model realisations and evaluation criteria in one figure, while such a summarising figure can give more insight. Therefore, the results (performance and consistency) from the FARM analysis were quantified and compared for all model configurations, for the calibration and validation periods. Additionally, in Chapter 6 a larger catchment is included as well to not only carry out a validation in time, but also in space.

For the quantification, the 10th-percentile of the Euclidean distances (Eq. 3.1) was used as performance measure. For consistency, the sum of the weighted angles of all vectors was used, from the first three principal components (PC). By doing this, correlated vectors on a PC with a high explained variance will largely influence the consistency positively. For both measures, a smaller value means better performance or more consistency.

3.5. FRAMEWORK TO ASSESS REALISM OF MODEL STRUCTURES

As described in Section 3.4, the Framework to assess realism of model structures (FARM) is an important evaluation tool used in this thesis. Below the outline of the framework will be described together with a test case in the Maimai in New Zealand.

3.5.1. DESCRIPTION OF THE FRAMEWORK

GENERAL OUTLINE

FARM touches on three main elements: model structures, hydrological signatures and the principal component analysis (PCA). Figure 3.3 describes how these elements interact in the general framework. The PCA is the general part of this framework; therefore, it will be described first. The model structures and hydrological signatures depend on the specific study the framework will be used for. An overview of all signatures used in this thesis can be found in Section 3.6. Therefore, they are described separately for each case study. The framework consists of the following steps:

1. selection of a catchment and gathering of hydrological process knowledge;

2. definition of hydrological signatures;

3. definition of evaluation criteria to assess the models' ability to reproduce the hydrological signatures;

4. selection of a set of plausible model structures based on the assumed hydrological processes;

5. derivation of a posterior parameter distribution for the selected model structures and catchments (calibration);

6. random sampling of N parameter sets from the derived posterior parameter distribution and calculation of the evaluation criteria for the modelled hydrographs;

7. principal component analysis for each combination of catchment and model structure; and

8. assessment of relative performance and consistency for each combination of catchment and model structure.

PRINCIPAL COMPONENT ANALYSIS (PCA)

A principal component analysis (PCA) is a statistical tool which can be used to reduce the dimensions of a multivariate problem. For a PCA the eigenvectors of a covariance matrix are determined. For many data sets most of the variance is described in the direction of a limited number of eigenvalues. By transforming the original axes towards the eigenvalues (principal components (PCs)), the original variable can be expressed in terms of
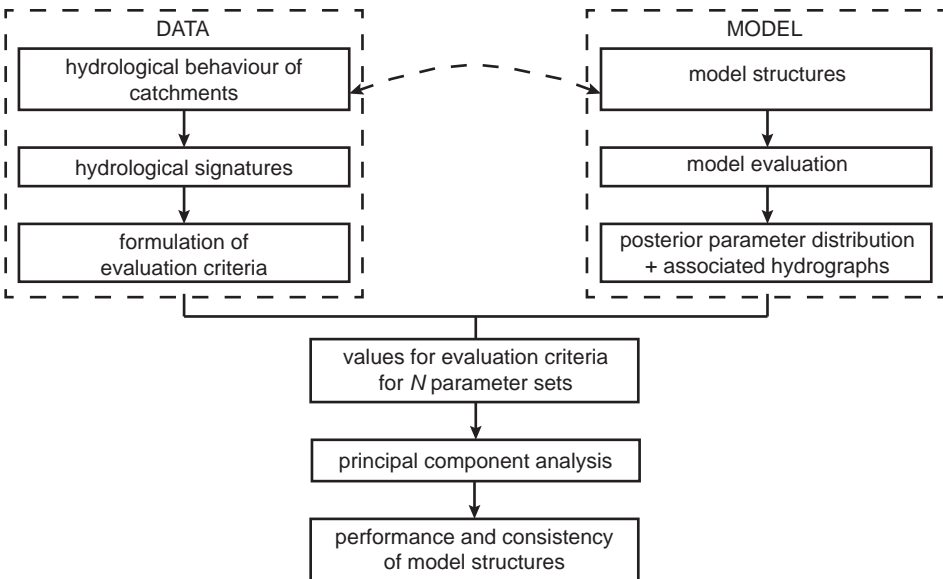

Figure 3.3: Schematic overview of FARM to compare the performance and consistency of model structures with respect to hydrological signatures.
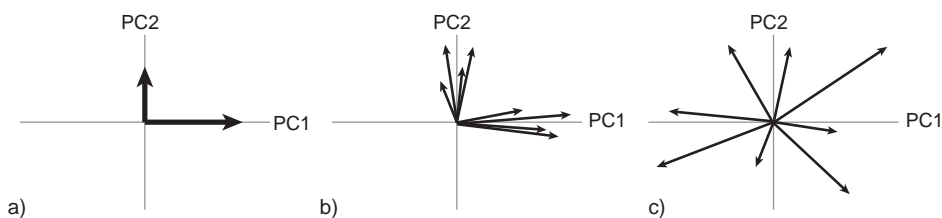
Figure 3.4: Illustration of possible configurations for the PCA diagram: each vector represents an evaluation criterion. The axes are formed by the first two principal components (PCs). **(a)** represents a fully consistent model structure, **(c)** a fully inconsistent model structure.

the PCs (the variables have a certain loading on the PCs). More detailed descriptions on the principles of a PCA can be found in literature about multivariate analysis (e.g. Krzanowski, 1988; Härdle and Simar, 2003). In Appendix A an example can be found explaining the use of PCA for FARM. Note that here the vectors of the loadings are referred to as 'vectors' thereafter.

Input for PCAs   For FARM PCA is used to explore the correlation structure between different evaluation criteria. A PCA is performed for each model structure in each catchment for N parameter sets. Here N is the number of parameter sets needed to reach convergence (see Sect. 3.5.2). The parameter sets are randomly sampled from a derived posterior parameter distribution. For these N samples all the evaluation criteria for the selected signatures are calculated (see Fig. 3.3); these values form the input to the PCA. Note that the model calibration strategy remains the choice of the modeller.

For a PCA it is assumed that the input data are generated from a normal distribution (Johnson and Wichern, 1998). Normality is especially important for the marginal distributions. Multivariate normality is of less importance if the PCA is used for dimension reduction, and thus as a mere descriptive tool as is the case with FARM (Jolliffe, 1986). If the marginals are not normally distributed, the values for the evaluation criteria have to be transformed to a normal distribution. This transformation could for example be done with a normal quantile transformation (Weerts et al., 2011; Montanari and Brath, 2004).

Interpretation of PCAs   The PCA represents two model characteristics: the performance and the consistency. Here, the performance is expressed in three indicative categories (see Sect. 3.5.2); they are presented by the thickness of the vectors in the PCA diagram (see for example the results of the Maimai in Fig. 3.6). As all model realisations are used, the range (difference between 10th and 90th percentile) in performance is also important. This range is presented with colours, which represent different classes as well: green for a small range, orange for a intermediate range and red for a large range. Note that, for each study, specific values for the categories (performance and range in performance) should be defined.

The degree of consistency is presented by the configuration of the vectors in the PCA. When a model structure is able to simulate different signatures well with the same

parameter set, the corresponding evaluation criteria should be directly correlated. In other words, a better performance on one evaluation criterion also means a better performance on another evaluation criterion, leading to a high consistency. For the PCA this results in the vectors, representing the evaluation criteria, pointing in the same direction. When evaluation criteria are inversely correlated, it means that a parameter set with a better performance for one criterion leads to a inferior performance for another. It is assumed that the signatures used for FARM are constructed to reflect different aspects of the hydrograph and, therefore, are not correlated by construction. The diagram which is the result of the PCA can be characterised by three general types of configurations (Fig. 3.4):

1. All evaluation criteria are or completely directly correlated or uncorrelated ('L-shaped diagram'). (Fig. 3.4a). When this is the case, the model is fully consistent, which would be the case for a hypothetical 'perfect' model.

2. The evaluation criteria have their longest distance in the same direction on one of the two principal components and are therefore all either directly correlated or uncorrelated ('L-shaped' diagram) (Fig. 3.4b). This configuration is the more practical version of the previous one and has a high consistency.

3. The evaluation criteria show a 'star-shaped' diagram and some evaluation criteria are uncorrelated, while others are inversely correlated (Fig. 3.4c). In this case the model is inconsistent.

The configurations in Figure 3.4 are basic configurations. In case of deviations from these basic configurations, three measures are important for interpretation of the PCA diagrams; these three are listed below. These measures, especially the first two, can be determined qualitatively and quantitatively; both will be used in this thesis. Quantitative assessment is based on the sum of the first three principal components and the variance explained per principal component.

- Spreading on PC1 or PC2 (x- or y-axis): PC1 always represents a larger part of the explained variance in the data, so a spread or inversely correlated evaluation criteria on PC1 determine the consistency to a larger extent than inversely correlated evaluation criteria on PC2.

- Length of the vectors: the longer a vector is, the higher the loadings, and thus the more influence the vector has on the total analysis. An inversely correlated vector which is relatively small influences the consistency less than an inversely correlated vector which is relatively long.

- Inversely correlated thick vectors: a thick vector means that there is a parameter set for which the signature can be modelled well; a thin vector indicates poorer model performance. So, inversely correlated thick vectors indicate that inconsistency is the main problem, while inversely correlated thin vectors indicate that performance is still the main problem.

Note that a PCA only shows the *relative* similarities and differences within the data used for the PCA; therefore, the absolute values on PC1 and PC2 and the individual di-

rection of the vectors are of no importance. Thus, the interpretation of the PCA diagrams is mainly carried out in a qualitative manner. This way of interpretation enables focussing on specific relations between evaluation criteria, giving insight in model deficiencies. However, when dealing with a large number of model structures or catchments, a more quantitative interpretation is necessary. In this thesis this is done by summing the weighted angles of all vectors, from the first three principal components (PC). By doing this, correlated vectors on a PC with a high explained variance will largely influence the consistency positively.

3.5.2. ILLUSTRATION OF FARM

Before applying FARM in the remainder of this thesis, an example for Maimai in New Zealand is shown below as illustration. Chapter 2 contains a description of this catchment. The case study uses model specifics (including model structures) which are only used here, for the Ourthe and Meuse different model configurations are used.

MODEL SPECIFICS

PCA   The model posterior parameter distributions were determined with Bayesian inference, using a heteroscedastic error model based on the weighted least squares (WLS) scheme (Thyer et al., 2009) and non-informative prior parameter distributions. A total of 1000 random samples were drawn from the posterior distributions, and all evaluation criteria were calculated for each random sample. The evaluation criteria distributions were then transformed to normal distributions with a normal quantile transformation (Weerts et al., 2011; Montanari and Brath, 2004). The transformed criteria were subsequently used as input for the PCAs.

The three indicative performance categories for this case study are defined as follows. For the illustration of FARM no colour coding was used for the range in performance for each evaluation criterion.

- High (continuous and very bold vectors), when the maximum value for the evaluation criterion is higher than 0.8 and 90 % of the values for the evaluation criterion are higher than 0.65.

- Medium (dashed and bold vectors), when is the maximum value for the evaluation criterion is higher than 0.4 and 90 % of the values for the evaluation criterion are higher than 0.3.

- Low (dotted and thin vectors), for all other cases.

Hydrological signatures   A set of signatures was selected from Table 3.3 for this illustration: F_{AC}, $F_{AC,low}$, F_{pks}, $F_{pks,low}$, F_{RLD} (evaluated with E_{RE}) and F_{FDC}, $F_{FDC,low}$, $F_{FDC,high}$, F_Q, F_{Qlog} (evaluated with E_{NSE}). In the abbreviations 'low' refers to the low flow period or lower part of the flow duration curve; 'high' refers to the high flow period or upper part of the flow duration curve. A low flow period might not be very relevant for very wet Maimai catchment, but this is an inheritance of the original case study in which both Maimai and Wollefsbach were considered.
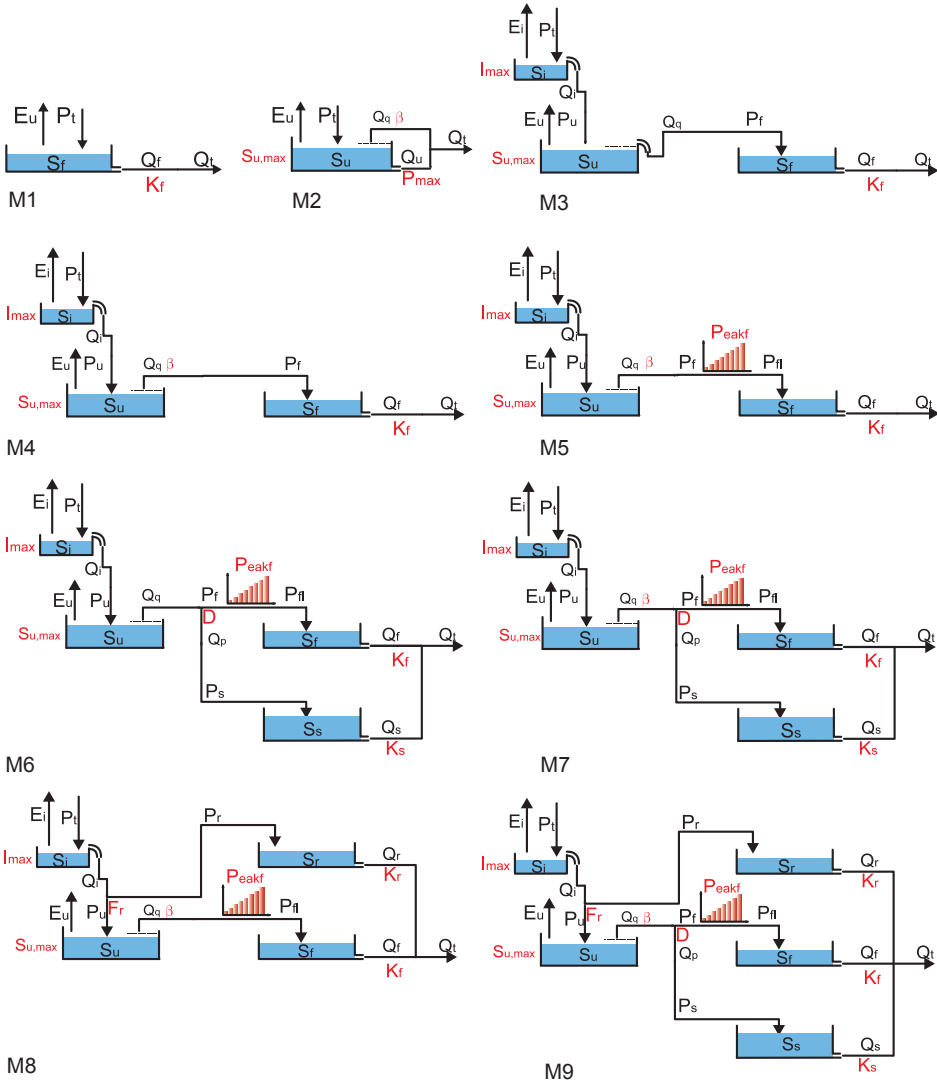
Figure 3.5: Conceptual configurations of the flexible model structures used for illustration of FARM.

Model structures   For this study nine flexible model structures are tested, and their performance and consistency are compared with 2 (fixed) benchmark models: GR4H (an hourly version of GR4J, Perrin et al., 2003) and a modified version of the HBV model (Lindström et al., 1997). The main adaptation on the HBV model is that river routing is not included (D. Kavetski, personal communication, 2012), because it is not considered as a crucial process due to the small size of the catchments. These benchmark models are mainly selected because they are widely used for hydrological modelling.

The nine flexible model structures have been configured with the SUPERFLEX framework (Fenicia et al., 2011). Model structures built with the SUPERFLEX framework consist of reservoir elements, lag function elements and junction elements. The created model structures (M1 to M9; see also Fig. 3.5) differ in the number of reservoirs (1 to 5), the number of fluxes (3 to 10) and the number of parameters (1 to 9). The selection of the model structures is mainly based on the model structures used by Kavetski and Fenicia (2011) and on experiences of previous modelling exercises. A discussion of processes represented by the model structures can be found in Kavetski and Fenicia (2011).

The model conditioning is done with Bayesian inference, as described by Kavetski and Fenicia (2011). The applied error model is based on weighted least squares. For the quasi-Newton parameter optimisation, 20 multi-starts are used. During the Markov chain Monte Carlo (MCMC) sampling, 5000 parameter sets were generated.

Plausibility checks   In this case study 1000 parameter sets, randomly drawn from the posterior distribution, are used to construct the PCA. To investigate whether this number is sufficient for stable PCA patterns, the sensitivity to the number of parameter sets was tested. To test the sensitivity of the PCA, it is important to know if the PCA is ergodic. When this is the case, there is a convergence to a stationary measure when enough samples are taken into account; this convergence is independent of the initial conditions (Descombes, 2012). To test whether the PCA is ergodic and to test if 1000 parameter sets are sufficient, a PCA was also performed with 500 and 200 parameter sets. When the differences between the diagrams with 200 and 500 parameter sets are larger than between the diagrams for 500 and 1000 parameter sets, it is an indication of convergence and ergodicity can be assumed.

In addition to the sensitivity to the number of parameter sets, the obtained results can also be evaluated for an independent test period. It may be expected that a consistent model structure behaves similarly in the calibration and validation period, as it is assumed to capture the dominant processes better than an inconsistent model (c.f., Seibert, 2000). Therefore, the model structures are run for an independent test period of one year with the parameter sets derived during the calibration. Both the performance and consistency are compared for the calibration and validation period.

Results
The PCA results for all model structures are shown in Figure 3.6. The PCA results are based on the covariance matrix of the evaluation criteria. All the model structures developed with the flexible framework except M8 show a very small difference in their maximum Nash-Sutcliffe efficiency; M3 to M5 even have an equal maximum Nash-Sutcliffe efficiency. However, the consistency differs between the model structures. M1 and M3
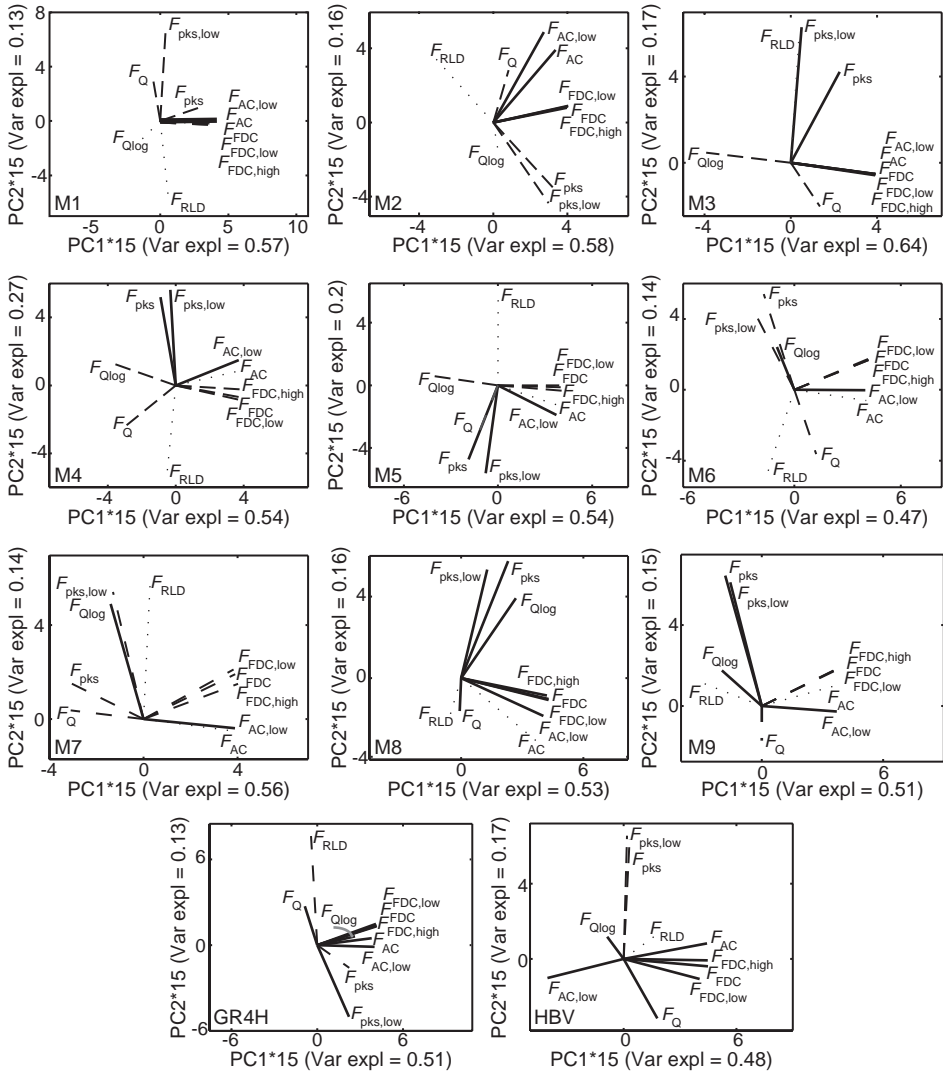
Figure 3.6: Results for PCA for the Maimai catchment. Each figure represents one of the model structures. The figures are based on 1000 parameter sets. The principal components are dimensionless, because the ratios of specific signatures of the modelled and observed hydrographs are used to construct the evaluation criteria and these ratios are dimensionless. The total variance explained by these figures is the sum of the explained variance per PC.
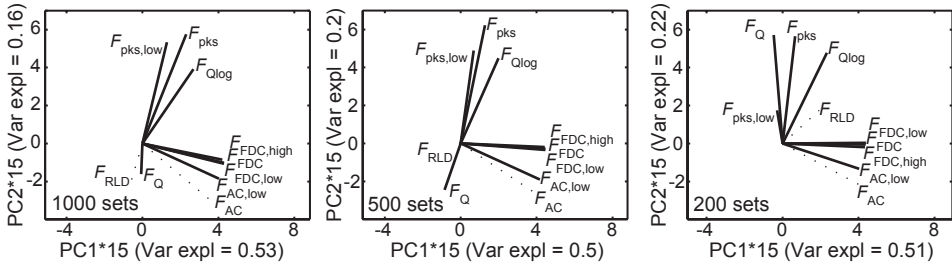
Figure 3.7: Result PCA for M8 for different number of parameter sets: 200(left)/500(middle)/1000(right). The difference between the diagrams with 1000 and 500 parameter sets is smaller than the difference between the diagrams with 500 and 200 parameter sets. The principal components are dimensionless.

show a comparatively high degree of consistency, i.e. a low spread of the vectors. For M1 the variance explained by PC2 is small compared to PC1; therefore, the spreading on PC2 has a minor influence. The evaluation criteria for M3 almost show an L-shape (see Sect. 3.5.1), and only $F_{\text{Qlog,NSE}}$ is inversely correlated. Model structures M4 to M7 are much less consistent. Model structure M8 behaves differently from model structures M1 to M7: it has a relatively high maximum Nash-Sutcliffe efficiency and a high performance for the other evaluation criteria; the diagram for M8 really shows an L-shaped configuration. The slight rotation of the L-shape indicates that the consistency can still be improved, as there are evaluation criteria on PC2 which are inversely correlated. However, as both the explained variance of PC2 and the loading of the vertical vectors on PC2 are low, this rotation decreases the consistency to a very limited degree.

Another interesting aspect is the high performance for most evaluation criteria for the HBV model, but a relatively low consistency. For the HBV model some evaluation criteria are inversely correlated on PC1, and the variance explained by PC2 is relatively high. GR4H has a high performance for most evaluation criteria, like the HBV model, but is more consistent than the HBV model, as the evaluation criteria are mainly inversely correlated on PC2, thus being of limited importance.

Plausibility of results   Figure 3.7 shows the PCA diagrams for M8 for 200, 500 and 1000 parameter sets. It can be seen that the difference between selecting 1000 and 500 parameter sets is smaller than the difference between selecting 500 and 200 parameter sets. This sensitivity analysis is performed for all the model structures, and the results are compared visually. Convergence is present to a varying degree for all model structures. Model structures with a higher performance and consistency and the model structures with less complexity exhibit larger convergence. However, these are not always the model structures with a more constrained posterior parameter distribution. In general, the convergence for all model structures shows that ergodicity can be assumed and that the use of 1000 parameter sets is sufficient to have an indication of consistency of the evaluated model structures in this study.

In Figure 3.8 an example is given to show the differences between two model structures with a more (M8) and a less (M7) comparable behaviour between the calibration
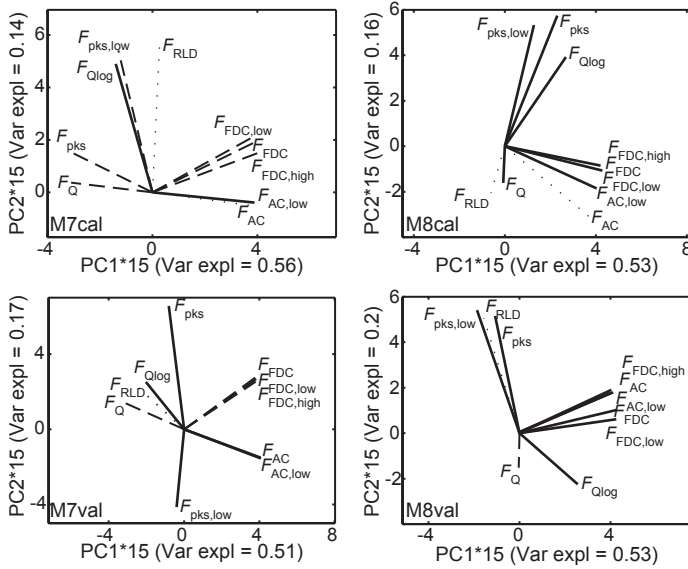
Figure 3.8: PCA diagrams for M7 (left) and M8 (right) for both the calibration (top) and validation (bottom) period. The principal components are dimensionless. M8 shows a higher consistency for the calibration period and a more consistent behaviour between the calibration and validation period. Presented results are for the Maimai catchment.

and validation period for the Maimai catchment. A summary of the results is presented in Table 3.2. The model structures in this table are ordered by consistency for the calibration period. It can be seen that both the performance and consistency changed between the calibration and validation period. Model structures with a low consistency in the calibration period have slightly larger changes for the validation period.

DISCUSSION

Applicability Comparing model structures based on both performance and consistency has advantages with respect to a comparison based on either performance or consistency. This can especially be seen for M8, M3, GR4H and HBV. Their performance is more or less equal, but their consistency is not. The results also show that consistency on itself does not give useful information about a model structure; for example M1. Rather, for model structures with a high performance, the degree of consistency gives useful information about the suitability for a certain catchment.

A model structure that suits a certain catchment is more likely to represent the dominant processes that actually occur in the catchment than model structures that are less suited for the catchment. Therefore, the model structure inhibits an indication of dominant processes in a catchment. However, when the hydrograph does not contain information about certain processes, these processes will not be taken into account when only the hydrograph is used as signature input. In that case, auxiliary data sources are required to reveal these processes (e.g., Vaché and McDonnell, 2006; Son and Sivapalan,

Table 3.2: Summary of differences between the PCA graphs for the calibration and the independent test period for the Maimai catchment[a].

	Performance Original[b]	Consistency Original[c]	Performance Validation[b]	Consistency Change
HBV	7	low	7	config. changed
M7	2	low	4 (+2)	3 $E_{c,i}$ changed[d]
M6	2	low	5 (+3)	2 $E_{c,i}$ changed
M9	4	low	5 (+1)	2 $E_{c,i}$ changed
M4	3	low	2 (−1)	small differences
M5	3	low	2 (−1)	small differences
M1	1	middle	2 (+1)	2 $E_{c,i}$ changed
M3	6	middle	5 (−1)	1 $E_{c,i}$ changed
GR4H	8	middle	9 (+1)	1 $E_{c,i}$ changed
M2	5	middle	5	small differences
M8	8	high	5 (−3)	1 $E_{c,i}$ changed

[a] The model structures are ordered by consistency in the calibration period. 1, 2 or 3 '$E_{c,i}$ *changed*' in the last column means the configuration of the PCA diagram of the calibration and validation are equal, but 1, 2 or 3 vectors have a different direction and/or length. '*config. changed*' means that the relative direction of almost all vectors changed. [b] The number of signatures in performance category high (thick vectors) is taken as a measure. [c] Based on the configurations in Fig. 3.4. [d] E = evaluation criterion.

2007; Fenicia et al., 2010; Hrachowitz et al., 2013b; Birkel et al., 2010). When additional data sources give extra information, it is expected that the evaluation criteria belonging to the extra hydrological signatures are uncorrelated with the evaluation criteria from the discharge data.

In addition, poor performance and poor consistency of a certain model structure can be an indicator for the absence of certain runoff processes in the catchment. This can for example be seen for M6, M7 and M9, which have a relatively low performance and consistency. These are the only flexible model structures with a groundwater reservoir, so possibly a groundwater reservoir is not important or incorrectly represented. This is also in accordance with the site description of the Maimai catchment: it has shallow soils and (almost) impermeable subsurface layers. The performance and consistency of M8 are very good; M8 has a riparian zone reservoir, which probably fits well with the almost year-round saturated soils of the Maimai catchment.

It should be noted that FARM is meant to indicate which model structures have a higher performance and consistency than others. However, data errors can influence the performance and consistency of a model, and this possible influence is not explicitly included in FARM (Bárdossy and Singh, 2008). The influence of these errors will be different for different signatures. On the other hand, by using signatures, mainly the dynamics of the measured and observed hydrograph are taken into account. These dynamics are more likely to represent catchment behaviour and to be less sensitive to small measurement errors than evaluation criteria that compare each point of the hydrograph

individually.

Using the framework   The use of PCAs for model evaluation also has limitations. The main limitation may be the low variance explained by the first two principal components as obtained in this study. For most model structures the variance explained is below 80 %. More reliable diagrams would therefore also incorporate the third principal component; however, a 3-D graph is more difficult to visualise and interpret than a 2-D graph. A more quantitative interpretation enables including more principal components, as shown in Sections 5.3.3 and 6.3.3.

There are two situations related to a low explained variance, which should be kept in mind when interpreting the PCA diagrams.

- Consistent configuration with low variance explained: the higher principal components (PC3 and higher) explain a smaller amount of variance each; this variance can reduce the high consistency, but will not make the model fully inconsistent.

- Inconsistent configuration with low variance explained: the first two principal components already show inconsistency. The variance explained by the higher principal components is lower, so they are unlikely to change a diagram from inconsistent to consistent.

In addition to this limitation, also three other aspects influence the usefulness of the framework. These include the selection of hydrological signatures and the sometimes different PCA results for calibration and validation periods. First, the hydrological signatures – selecting different signatures from different data sources – result in testing different aspects, which leads to different results. The selection of the signatures is subjective and influences the results. For this framework a good approach would be to start with many signatures for a catchment and test which signatures are directly correlated. The signatures that are strongly directly correlated with another signature for each model structure can be omitted.

The second is the different PCA results for the calibration and validation period for some model structures. In Sect. 3.5.2 it is shown that generally the model structures with a higher consistency behave more similarly in the calibration and validation period. However, this does not hold for all model structures, and the similarity between the calibration and validation period can be influenced by the length of the used time series as well. Therefore, before selecting a model structure which seems to have a very high consistency and performance, it may be beneficial to test the performance and consistency on a different time period.

3.6. Hydrological signatures

3.6.1. Signatures as evaluation criteria

The model structures are evaluated with evaluation criteria based on hydrological signatures. These signatures can be derived from the observed hydrograph, for example the flow duration curve (Westerberg et al., 2011) or the autocorrelation coefficient (e.g., Winsemius et al., 2009). However, these signatures can in principle also be derived from other data sources, for example groundwater levels, tracer data or satellite data. Note

that the 'more independent' the selected signatures are (i.e. reflecting contrasting parts of the hydrograph), the higher the significance for the model evaluation.

Most signatures are represented by a single value for the observed and a single value for each modelled hydrograph. A possibility to formulate the evaluation criterion (E) in that case is shown in Equation 3.2. Only the value for the signature of the modelled hydrograph changes per parameter set; the value for the observed hydrograph is the same for each parameter set. By dividing the modelled value by the observed value, the relative deviation of the modelled from the observed value can be obtained. The absolute value and '1 −' the ratio are required to obtain the same result (E) for the same absolute deviation of the modelled value above or below the observed value. Note that data errors are not explicitly accounted for in the calculation of the signatures and evaluation criteria; therefore, they should be applied with care in case of comparing results from different data sources.

$$E_{\text{RE}} = \left| 1 - \frac{F(Q_{\text{mod}})}{F(Q_{\text{obs}})} \right|, \tag{3.2}$$

with $F(Q_{\text{mod}})$ the value of the hydrological signature for the modelled hydrograph and $F(Q_{\text{obs}})$ the value of the hydrological signature for the observed hydrograph.

Equation 3.2 cannot be used if the signature is represented by a time series (e.g., the discharge itself or the flow duration curve). Therefore, three other metrics were used for evaluating these signatures: Root mean square error (E_{RMSE}), Nash-Sutcliffe efficiency (E_{NSE}) (Nash and Sutcliffe, 1970) and volume error (E_{VE}) (Criss and Winston, 2008). The relative error (E_{RE}) (Eq. 3.2) and RMSE give lower values for higher performances. For parts of the evaluation it is convenient to link a better performance to a higher value for the evaluation criterion (see also Sect. 3.5). In which case for example the formulation in Eq. (3.3) can be used.

$$E_{\text{PCA}} = 1 - E \tag{3.3}$$

3.6.2. DESCRIPTION OF SIGNATURES

The different experiments in this thesis use a selection of signatures. The selection depends on the elements in the model structures that were changed and thus needed evaluation. Signatures were calculated for the entire modelled period and in some cases for high and low flow periods as well. Below a short description is given of all signatures used in this thesis; Table 3.3 shows in addition the abbreviation and a sketch for each signature.

DISCHARGE

The most straightforward and at the same time most integrated signature is the observed discharge (F_Q) itself. In this thesis F_Q is used in combination with NSE and VE to assess the overall performance of different model structures. To assess overall performance of lower discharge, also the log of the discharge ($F_{Q,\text{log}}$) was used. In Chapter 4 the yearly discharge ($F_{Q,\text{y}}$) and the yearly base flow ($F_{Q,\text{b}}$) are used to assess water balance issues and interseasonal storages.

FLOW DURATION CURVE

Flow duration curves give the probability of exceedance for each observed (or modelled) discharge. Flow duration curves (F_{FDC}) are often used in model evaluation; either by using the entire curve or by using parts of it (e.g., Yadav et al., 2007; Yilmaz et al., 2008; Blazkova and Beven, 2009; Westerberg et al., 2011). The main difference between using F_Q and F_{FDC} is the timing of events: for F_Q this is very important, while for F_{FDC} this is neglected. In this thesis the lower and higher part of the flow duration curve are used in addition to the entire curve. Chapter 4 also uses the slope of the normalised flow duration curve ($F_{FDC,slp}$) to compare catchments.

COEFFICIENT OF AUTOCORRELATION

The coefficient of autocorrelation is a measure for the smoothness of a hydrograph: a high value means a small difference between two consecutive points (Winsemius et al., 2009). This thesis uses the coefficient of autocorrelation in two forms; first with a fixed lag of 24 hours (F_{AC}) and second with a varying lag of 0 - 250 hours ($F_{AC,func}$).

RISING LIMB DENSITY

Like the autocorrelation, the rising limb density (F_{RLD}) provides an indication of the smoothness of the hydrograph, but the RLD is averaged over the total period and is completely independent of the flow volume (Shamir et al., 2005). This signature is calculated by dividing the number of peaks by the total time the hydrograph is rising. Therefore, the RLD is the inverse of the mean time to peak (Shamir et al., 2005; Yadav et al., 2007).

DECLINING LIMB DENSITY

The declining limb density (F_{DLD}) is the counterpart of the rising limb density: it represents the time the hydrograph is declining with respect to the number of peaks. Due to hysteresis the rising and declining limb vary in steepness, thus they will provide different information about the catchment behaviour.

SLOPES OF RISING AND DECLINING LIMBS

The slopes of the rising (F_{SRL}) and declining (F_{SDL}) limbs provide information about the flashiness of the catchment. Therefore, they are relevant signatures to evaluate the representation of very fast processes in model structures. Again, due to hysteresis the slopes of the rising and declining limb will provide different information about the catchment behaviour.

PEAK DISTRIBUTION

The peak distribution (F_{pks}) shows the variability in observed or modelled peak discharges. Peak discharges are discharges at a time step of which both the previous and the following time step have a lower discharge. From these peak discharges a flow duration curve is constructed. If a model is better able to reproduce the peak distribution, it is more likely that it can reproduce the different responses of the catchment simultaneously.

RUNOFF COEFFICIENT

The runoff coefficient (F_{RC}) is a very integrated signature as well. However, it is a very simple signature to check if the water balance closes; with this purpose it was used in Chapter 4. In addition, the event runoff coefficient ($F_{RC,event}$) gives more insight into the catchment behaviour under different (or similar) precipitation conditions (e.g., McMillan et al., 2013)

Table 3.3: Overview of all signatures used in model experiments.[a]
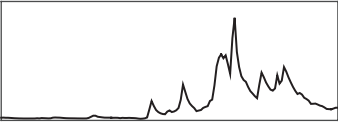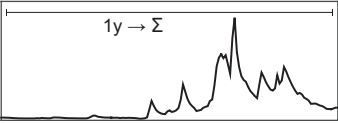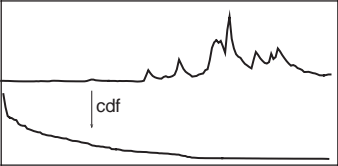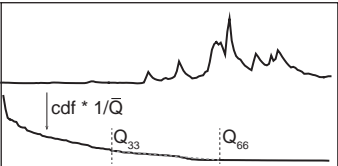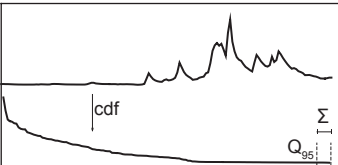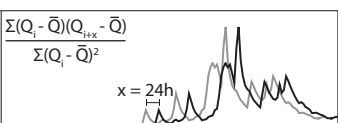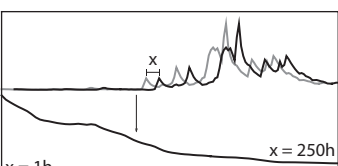
Signature	Abbreviation	Sketch
discharge	$F_Q, F_{Q,log}$	
yearly discharge	$F_{Q,y}$	
flow duration curve (e.g., Westerberg et al., 2011)	F_{FDC}	
slope of normalised Flow Duration Curve (e.g., Müller and Thompson, 2016)	$F_{FDC,slp}$	
yearly base flow	$F_{Q,b}$	
coefficient of autocorrelation (lag = 24 h) (e.g., Winsemius et al., 2009)	F_{AC}	
coefficient of autocorrelation (1h < lag < 250h) (e.g., Hrachowitz et al., 2014)	$F_{AC,func}$	

Table continues on next page

Table 3.3 – continued from previous page

Signature	Abbreviation	Sketch
rising limb density (e.g., Shamir et al., 2005)	F_{RLD}	
slope rising limb	F_{SRL}	
declining limb density (e.g., Shamir et al., 2005)	F_{DLD}	
slope declining limb	F_{SDL}	
slope of peak distribution	F_{pks}	
runoff coefficient	F_{RC}	
event runoff coefficient (e.g., McMillan et al., 2013)	$F_{\text{RC,event}}$	

[a] The signatures are evaluated using different metrics in different chapters; the metrics are specified in each chapter.

4

ROOT ZONE STORAGE CAPACITY DERIVED FROM CLIMATE DATA

Change happens by listening and then starting a dialogue with the people who are doing something you don't believe is right.

Jane Goodall

Root zone storage capacity ($S_{r,max}$) is an important variable for hydrology and climate studies, as it strongly influences the hydrological functioning of a catchment and, via evaporation, the local climate. Despite its importance, it remains difficult to obtain a well-founded catchment representative estimate. This chapter tests the hypothesis that vegetation adapts its $S_{r,max}$ to create a buffer large enough to sustain the plant during drought conditions of a certain critical strength. Following this method, $S_{r,max}$ can be estimated from precipitation and evaporative demand data. The results of this 'climate based method' are compared with traditional estimates from soil data for 32 catchments in New Zealand. The results show that the variability of root zone storage capacity among catchments is larger for the climate based method than for the soil based method. Using a model experiment it is shown that the climate derived $S_{r,max}$ can better reproduce hydrological regime signatures for humid catchments; for more arid catchments the soil and climate methods perform similarly.

4.1. INTRODUCTION

Applying distribution in hydrological models requires additional data or creative use of existing data. An example of the latter is estimating specific parameters directly from data, and thus bypassing 'conventional calibration'. This chapter describes a method to estimate root zone storage capacity (S_r) from climate data.

Root zone storage capacity is an important hydrological descriptor, which strongly influences the hydrological functioning of a catchment. Root zone storage capacity can be understood as a volume of water per unit area within reach of plant roots for transpiration; outside this area of influence water flows are largely controlled by gravity-induced gradients. In models, $S_{r,max}$ is frequently used as a parameter representing catchment storage capacity in the dynamic part of the unsaturated zone. $S_{r,max}$ controls water partitioning between evaporation and drainage, and thus the (long term) water balance of a catchment (Field et al., 1992; Zhang et al., 2001). Understanding the water balance and the associated dynamics of different storage components over time is essential to understand the hydrological functioning of a catchment, necessary for robust predictions of discharge and evaporation. Apart from hydrological purposes, accurate estimation of soil water storage and water fluxes (e.g., evaporation and discharge) is of critical importance for climate (e.g., Kleidon and Heimann, 2000; Dirmeyer, 2011; Orth and Seneviratne, 2014) and ecological models (e.g., Liancourt et al., 2012; Zelikova et al., 2015) as well.

Despite the importance of $S_{r,max}$, it is difficult to obtain well-founded catchment representative estimates. Although soil and plant root properties can be observed at the point scale, it remains problematic to integrate these measurements to the catchment scale due to their spatially heterogeneous character (e.g., Crow et al., 2012). Even if the soils would be completely homogeneous, it is not necessarily clear how to map measured soil properties to model parameters, including $S_{r,max}$. Therefore, it is unknown whether point observations allow for an adequate representation of catchment representative $S_{r,max}$. Another common method to estimate $S_{r,max}$ is by calibration, preferably using expert knowledge or additional data to guide or constrain the calibration (e.g., Winsemius et al., 2008; Gharari et al., 2014a), which has the advantage that catchment representative $S_{r,max}$ is directly estimated. On the other hand, even constrained calibration is subject to parameter sensitivity (i.e., equifinality; Beven, 2006), making it again difficult to assess if the derived value is a plausible representation of the catchment representative $S_{r,max}$.

Following from the above, catchment understanding and flux modelling may be improved by further independent information on the amount of water which is (or can be) accessed by vegetation. The accessible water amount is not necessarily related to root depth, but rather to root density, i.e., the pore volume within the area of influence of the roots (Schenk and Jackson, 2002; Gentine et al., 2012; Cassiani et al., 2015; Tron et al., 2015; Brunner et al., 2015). To ensure long-term survival, vegetation adapts to its environment (Eagleson, 1982; Canadell et al., 1996; Sampson and Allen, 1999; Troch et al., 2009). Vegetation attempts to balance the resources invested in above surface growth with the resources necessary to create a root zone storage capacity large enough to buffer hydrological variability and to provide sufficient water for survival. This is an example of the widely acknowledged interaction between climate and vegetation (Milly,

1994; Rodriguez-Iturbe, 2000; Schymanski et al., 2008). It is likely that the required $S_{r,max}$ is strongly dependent on climate (Kleidon and Heimann, 1998; Guswa, 2008; Donohue et al., 2012; Gentine et al., 2012), i.e., precipitation and evaporative demand, but that the dependence on soil type is much weaker: in a specific climate vegetation needs a certain amount of water, irrespective of the soil type it is growing on. Two similar plants in the same climate, but on different soil type, might develop a different root structure, but they will require the same amount of buffer capacity to survive (cf. Camporese et al., 2015).

Climatic variability is characterized by higher frequency temporal dynamics than the formation process of soils. In turn, it is plausible that the medium-term dynamics of root growth have the same time scale as climatic variability (e.g., Sivandran and Bras, 2013); thus, when $S_{r,max}$ depends on climate rather than on soil, this would mean that $S_{r,max}$ should be dynamic at time scales of climatic variability. This supports the results of various hydrological modelling studies that have also shown the need for a variable reflecting medium-term dynamics (e.g., Wagener et al., 2003; Fenicia et al., 2009). Beekman et al. (2014) showed that a model based on soil data underestimates the evaporation of a forest on sandy soil and overestimates the evaporation of crops on clay soil during a very dry summer in the Netherlands. Although the modelled evaporation above the forest was almost zero, the vegetation survived the drought. Hence we may assume that the vegetation did develop a root system which could adapt to the climatic variability and thus created a buffer large enough to bridge the dry summer (cf. Vico et al., 2015), even on the sandy soil. This follows the line of argument set by Gimbel et al. (2016), who concluded that medium- to long-term climatic conditions of an area were more important than short-term antecedent soil moisture for the system behaviour under drought conditions.

Kleidon and Heimann (1998) showed that root depth is strongly related to climate, especially to the difference in precipitation and potential evaporation. Following on this, Gao et al. (2014b) recently successfully demonstrated for 400 catchments in the USA that catchment representative root zone storage capacities estimated from climate are strongly correlated with estimates derived from the calibration of a hydrological model. These studies indicate that climate information contains at least a certain level of information on root zone dynamics and thus on the influence of vegetation on the partitioning of water fluxes. Based on the results of Gao et al. (2014b) and following the arguments above, this chapter tests the hypothesis that climate is a more suitable estimator for the catchment-scale root zone storage capacity $S_{r,max}$ than observation inferred soil characteristics.

This chapter uses data from New Zealand and the New Zealand national model, thus the model set-up and conditioning are different from those described in Chapter 3.

4.2. STUDY AREAS

As described in Chapter 2, New Zealand has a strong climate gradient, which is very useful when comparing the influence of climate. 32 catchments are used for an overall comparison between climate and soil derived $S_{r,max}$-values. Additionally, some analyses were carried out for a subset of catchments only. Using a selection of catchments makes it possible to point out some effects in more detail. The catchments in this subset are Otekaieke at Stockbridge (A), Raparapahoe at drop structure (B) and Inangahua at Blacks

Table 4.1: Characteristics of sub set of catchments.

	A	B	C
aridity index ($\overline{E_p}/\overline{P}$)	0.67	0.40	0.32
long-term averaged runoff coefficient	0.42	0.48	0.84
mean annual precipitation (m y^{-1})	1.1	2.4	2.6
mean annual pot. evaporation (m y^{-1})	0.74	0.97	0.82
seasonality precipitation	0.16	0.15	0.11
seasonality pot. evaporation	0.72	0.61	0.69
number of months between peak P and E_p	0	7	3
dominant climate[a]	CD	WW	CXW
dominant land cover[b]	P	P	F
derived $S_{r,soil}$ (m)	0.16	0.15	0.17
derived $S_{r,clm}$ (m)	0.14	0.31	0.02

[a] CD = cool-dry; WW = warm-wet; CXW = cool-extremely wet.   [b] P = pasture; F = forest.

point (C). The catchments vary in aridity, land cover and size, refer to Table 4.1 for more details and Figure 2.6 for locations.

Daily discharge data were available from flow gauges; daily precipitation and potential evaporation data were available from the Virtual Climate Station Network (VCSN; Tait et al., 2006, 2012), which contains interpolated data at a 0.05° grid at daily time steps. The VCSN is comprised from national rain gauge data; mountainous areas have a lower gauge density, leading to less reliable estimates at these locations. Potential evaporation was calculated with the Priestley Taylor method (Priestley and Taylor, 1972). Catchment average precipitation and potential evaporation estimates were used from 1972 to 2012. The VCSN precipitation data were corrected for spatial bias, using an analysis of the long-term water balance: mean annual modelled runoff was compared with observed mean annual runoff and errors were assumed to be mainly caused by inaccuracies in precipitation measurements (Woods et al., 2006). The available discharge data varied per catchment, both in length and in period; however, for each catchment at least 20 complete years were available.

4.3. Methods

4.3.1. Comparison of soil- and climate-derived root zone storage capacity

Soil and climate derived $S_{r,max}$-values ($S_{r,soil}$ and $S_{r,clm}$ respectively) were calculated for the 32 selected catchments (description in Sect. 4.3.2). The soil and climate methods were compared based on these derived values and based on model results obtained with these derived values.

Comparison based on derived $S_{r,max}$-values

Comparison of the soil and climate derived $S_{r,max}$-values was based on spatial patterns and scatterplots. Catchment representative $S_{r,max}$-values were compared based on their

location in New Zealand. In addition, they were compared with a scatterplot, in which the catchment average runoff coefficient was used as an explanatory factor.

COMPARISON BASED ON MODEL RESULTS

Soil and climate derived $S_{r,max}$-values are based on different methods, but the values should be similar when assuming that they both represent the catchment representative root zone storage capacity. Although, it may not be possible to determine which $S_{r,max}$-estimate is closer to the true catchment value, it can be tested which estimate is more suitable for a modelling concept of root zone storage capacity. This second comparison was made using the hydrological model TopNet, which was run with each $S_{r,max}$-estimate in turn. The modelled and observed discharges were compared using hydrological signatures, as described in Section 3.6.

TopNet is a distributed conceptual model covering all third order subcatchments in New Zealand. The national version is not calibrated, but parameters are estimated for each third order subcatchment from available topography, land use and soil data (Clark et al., 2008a). The model conceptualisation consists of five storage reservoirs and the closure relations between these storages (Fig. 4.1; refer to Clark et al. (2008a) for model equations and parameter explanations). Distributed calculations for all storages and fluxes are carried out for all third order subcatchments (approx. 10 km^2): the underlying first order streams are defined with an upstream area of 0.14 km^2. Discharges from each third order subcatchment are then routed along the river network to calculate the discharge of the higher order catchments. Using TopNet with climate derived $S_{r,max}$ requires a transfer of $S_{r,clm}$ to a model parameter; $S_{r,clm}$ is transferred to the soil porosity corresponding to plant available water ($\Delta\theta_p$) (for details see Sect. 4.3.2).

TopNet simulates third order subcatchments, but the catchment representative $S_{r,clm}$ values were derived for fourth to seventh order catchments; therefore, the catchment representative $S_{r,clm}$ needs to be disaggregated to these third order subcatchments. In addition to the 40 years of catchment average data for the 32 catchments, 10 years of daily precipitation data and estimates of long term mean annual discharge (Woods et al., 2006) were also available for each third order subcatchment. From the results presented in Section 4.4 it follows that the catchment representative $S_{r,clm}$ shows a linear relation with the runoff coefficient (Fig. 4.2a), where the runoff coefficient is tightly linked to the aridity index, as illustrated by the Budyko framework (Budyko, 1974) and therefore indicates the wet- or dryness of the catchment. A strong relationship between these two variables is expected because a smaller $S_{r,max}$-estimate decreases the water available for transpiration, and therefore increases the runoff coefficient, as illustrated by previous studies (e.g., Donohue et al., 2012; Gao et al., 2014b). This dependency was used to proportionally disaggregate the catchment representative values to the third order subcatchments. For the disaggregation a linear relation between $S_{r,clm}$ and the runoff coefficient (F_{RC}) was assumed, while preserving the catchment representative $S_{r,clm}$; zero storage was set as boundary condition for F_{RC} equals 1. In case of nested catchments, disaggregation was carried out from upstream to downstream, preserving the values assigned to the third order subcatchments in the nested catchments. Figures 4.2b,c show an example of the disaggregation for the Buller catchment; it can be seen that the main catchment (I) spans all the $S_{r,clm}$ values occurring in the third order subcatchments. Disaggregation was not necessary for $S_{r,soil}$, as these values were already estimated for
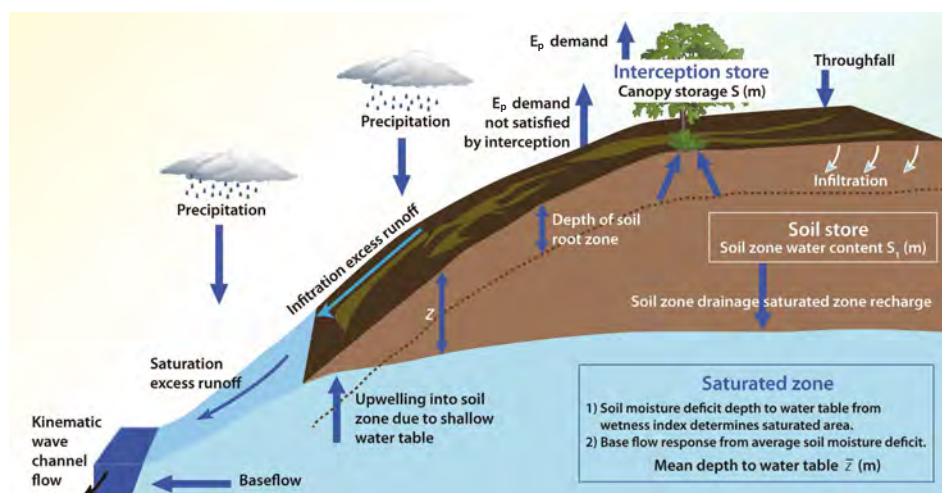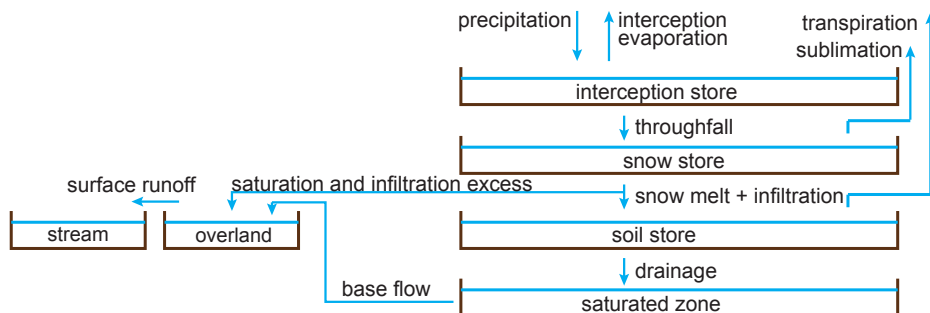
Figure 4.1: a) Perceptual representation of TopNet (adapted from Bandaragoda et al. (2004); ©Elsevier Ltd.); b) conceptual representation of TopNet.

first order subcatchments and were averaged to third order subcatchments according to catchment area.

Changing the root zone storage capacity in the model can influence the modelled catchment response at an event time scale and at longer, e.g. yearly, time scales. For instance at the event time scale, a smaller $S_{r,max}$ causes a quick reduction of soil moisture deficits during events, leading to faster connectivity, as processes such as preferential flow are activated. As an effect of this the shapes of the peaks (effect 1a) and the event runoff coefficient (effect 1b) will change. At the yearly time scale, however, a smaller $S_{r,max}$ decreases the storage and buffer capacity. This in turn affects the partitioning between discharge and evaporation (effect 2a), the flow variability between years (effect 2b) and the runoff volume during the dry season (effect 2c). These five effects (1a/1b/2a/2b/2c) were used to compare the discharges modelled with $S_{r,soil}$ and $S_{r,clm}$. For each effect one or more hydrological signatures were selected from Table 3.3 and
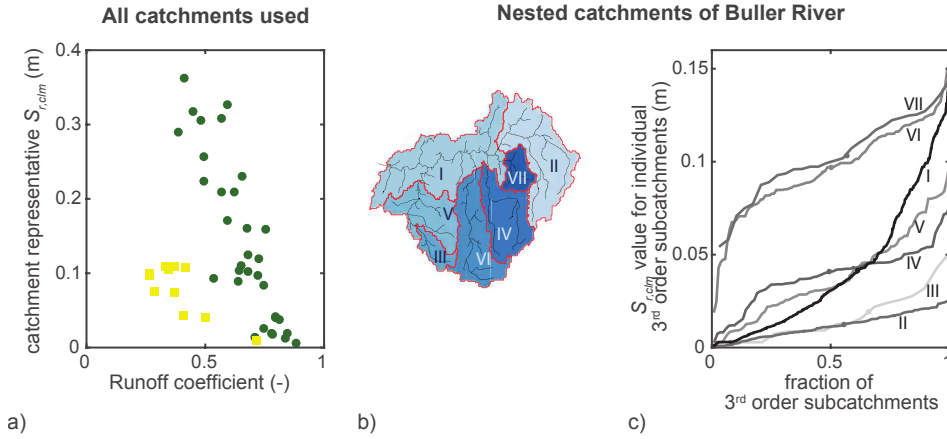
Figure 4.2: Summary of disaggregation method. a) Relation between runoff coefficient and catchment representative $S_{r,clm}$ for catchments where dormancy is not dominant (green circles) and where dormancy is dominant (yellow squares); b) Different nested catchments of Buller catchment; c) Cumulative distribution of $S_{r,clm}$ values of third order subcatchments for each nested basin in (b), characters refer to different nested catchments in (b); markers indicate the catchment representative value. Note that not all third order subcatchments have an equal catchment area; thus, the catchment representative value is not always the average of the maximum and minimum value of the third order subcatchments.

linked the to different effects in Table 4.2. Both modelled discharges were evaluated on their ability to reproduce these observed signatures. In addition to the described signatures also discharge was used. The discharge was evaluated with Nash-Sutcliffe Efficiency (E_{NSE}) and Volume Error (E_{VE}). All signatures were combined into an integrated performance measure by using the Euclidean distance to the 'perfect model', i.e. metrics being zero, (Eq. 3.1).

4.3.2. DERIVATION OF $S_{r,max}$-VALUES

CLIMATE DERIVED ROOT ZONE STORAGE CAPACITY

For the climate derived root zone storage capacity ($S_{r,clm}$) it was assumed that transpiration will deplete the root zone storage during dry spells. Following, to estimate the root zone storage capacity it was assumed that vegetation reserves a storage large enough to overcome a dry spell with a certain return period. To estimate the required annual storages a simplified water balance model was introduced; one with only an interception and root zone storage reservoir (e.g., Fenicia et al., 2008b) and only one parameter (max. interception storage capacity) and no further closing relations. The root zone storage reservoir has zero moisture deficit at the beginning of the simulation (i.e., end of the wet period) and the deficit increases when transpiration exceeds net precipitation ($P - E_i$) (Fig. 4.3a,b); any excess precipitation is assumed to runoff or recharge. The simulation was carried out for each catchment for the entire length of the precipitation series (1972-2012) on a daily basis. By doing this simulation, the yearly maximum deficits can be

Table 4.2: Overview of used signatures[a].

Signature	Metric	Effect
Event time scale		
median slope rising limb	E_{RE}	1a
median slope declining limb	E_{RE}	1a
slope of peak distribution	E_{RE}	1a
mean event runoff coefficient[b]	E_{RE}	1b
std event runoff coefficient	E_{RE}	1b
correlation coefficient auto correlation	E_{RE}	1a
Yearly time scale		
average yearly base flow (lowest 5% of flow)	E_{RE}	2c
slope of normalised flow duration curve	E_{RE}	2c
runoff coefficient	E_{RE}	2a
std yearly discharge	E_{RE}	2b
discharge	E_{NSE}	NA
discharge	E_{VE}	NA

[a] The process numbers indicate for which runoff process a signature is selected: the signatures do not only evaluate this process, but have a strong focus towards this process (E_{RE} = relative error, Eq. 3.2), E_{NSE} = Nash-Sutcliffe efficiency, E_{VE} = volume error).   [b] Events were defined as in McMillan et al. (2011).

determined, which are equivalent to the root zone storage capacities required to maintain a sufficient supply of water to vegetation during dry periods and to thereby fulfil the evaporative demand in the individual years.

Following the daily simulation, the Gumbel extreme value distribution was used to standardise the results from different catchments (Gumbel, 1935). The maximum moisture deficits of the individual years are used as input into the extreme value distribution (Fig. 4.3c). From this distribution, the root zone storage capacities can be estimated which are necessary for vegetation to bridge dry spells with specific return periods. The results from Gao et al. (2014b) suggest that many ecosystems tend to develop root zone storage capacities large enough to survive dry spells with return periods between 10 to 20 years; it is likely that grasses are adapted to lower return periods. Therefore, the analysis here was based on $S_{r,max}$-values belonging to dry spells with return periods of 10 years, Section 4.5.1 shows a sensitivity analysis regarding the chosen return period.

An estimate for *actual* transpiration (T) is required before the analysis described above can be carried out. Transpiration is estimated based on the long-term water balance and estimates of potential evaporation (Fig. 4.3d). The long term average transpiration ($\overline{T}$) is derived from the water balance:

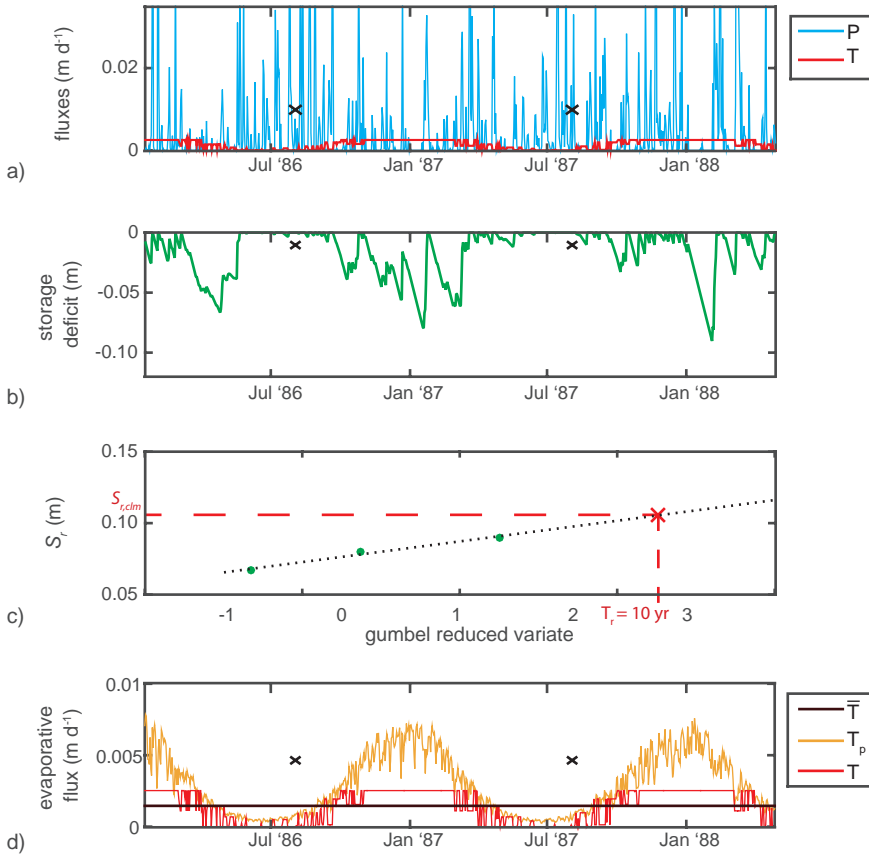$$\overline{T} = \overline{P - E_i} - \overline{Q}, \tag{4.1}$$

Figure 4.3: Method for climate derived root zone storage capacity; a) precipitation and estimated transpiration; b) calculated soil moisture deficit; c) use of yearly minimum storage in Gumbel distribution; d) use of potential transpiration to estimate actual transpiration. Black crosses indicate the start of a simulation year.

where P is precipitation, E_i is interception evaporation and $\overline{Q}$ is long term averaged observed discharge. Interception evaporation is determined by simulating an interception reservoir, accounting for interception storage, effective precipitation (throughfall) and interception evaporation. Maximum interception storages were taken from the Top-Net configuration, they depend on land cover and range from 0.5 mm for pasture to 1.9 mm for forest. The influence of these values was evaluated in a sensitivity analysis (Sect. 4.5.1). The required root zone storage capacity was set to zero when discharge equals net precipitation ($P-E_i$), which would theoretically occur in the absence of evaporative demand or when no more storage is available. For estimating the transpiration only those hydrological years were used, which had complete discharge data.

Transpiration is not constant over the year; therefore, potential transpiration ($T_p = \overline{E_p} - E_i$) is used to add seasonality to the long term average transpiration ($\overline{T}$). When $\overline{T}$ ex-
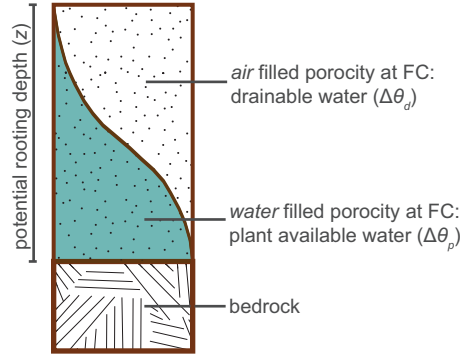
Figure 4.4: Method for soil derived root zone storage capacity (FC = field capacity).

4

ceeds daily T_p (in winter), T equals T_p (Fig. 4.3d). By doing this, some evaporative energy of $\overline{T}$ is not assigned to T in the winter period, so to close the water balance, this energy is equally redistributed over the months in which T_p exceeds $\overline{T}$ (in summer). Seasonality in transpiration is not only caused by seasonality in potential transpiration, but also by vegetation going into a state of dormancy. The latter is assumed to occur in catchments classified as dry and with a pasture land cover. During the summer months (December till March), the grass in these catchments turns yellow and does not transpire. Therefore, transpiration is set to zero during the summer months for these catchments. The influence of this dormancy on $S_{r,clm}$ is evaluated with a sensitivity analysis (Sect. 4.5.1).

Soil derived root zone storage capacity

Existing, national estimates of soil derived root zone storage capacity ($S_{r,soil}$) are available for the study catchments, based on the available storage between wilting point and field capacity. The soil moisture depth corresponding to wilting point and field capacity was derived from field measurements: water and air filled porosity at field capacity and potential rooting depth (Fig. 4.4) were observed for different soils and locations in New Zealand. The characteristics of different soils were used to estimate the soil moisture depth at field capacity and wilting point for all first order catchments in New Zealand (Newsome et al., 2000; Webb and Wilson, 1995). These values are currently used for the uncalibrated and distributed version of the national hydrological model TopNet. Specifically, $S_{r,soil}$ is here inferred from the model parameters in TopNet representing the potential rooting depth (z) and the fraction of plant available water ($\Delta\theta_p$, water filled porosity at field capacity), by using Equation 4.2.

$$S_{r,max} = z * \Delta\theta_p \tag{4.2}$$

4.4. Results

4.4.1. Comparison of $S_{r,max}$-values

Figure 4.5 shows the comparison between $S_{r,clm}$ and $S_{r,soil}$. It can be seen that $S_{r,clm}$ is lower than $S_{r,soil}$ for the areas classified as 'wet' or 'extremely wet' (Fig. 2.6). For the
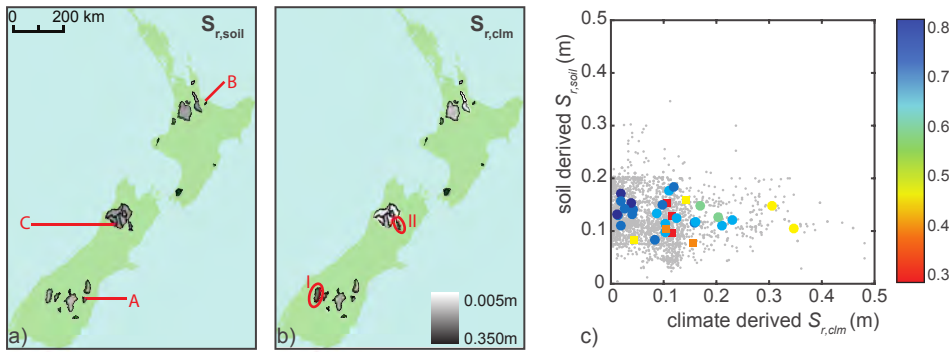
Figure 4.5: Catchment representative $S_{r,max}$-values for selected catchments A, B and C (see Tab. 4.1 for more details); a) soil derived $S_{r,soil}$; b) climate derived $S_{r,clm}$; c) scatterplot, the catchment representative $S_{r,max}$-values from (a) and (b) are colour coded by runoff coefficient (circles = no dormancy, squares = dormancy), grey dots show $S_{r,max}$-values per third order subcatchment.

dryer areas the values are more comparable, but with slightly higher values for $S_{r,clm}$.

The same interpretation follows from Figure 4.5c, which shows a scatterplot between $S_{r,clm}$ and $S_{r,soil}$ for catchment representative values (the ones shown in Fig. 4.5a,b), and for values for each subcatchment of at least third order (the individual subcatchments in TopNet). It can be seen that $S_{r,clm}$ is correlated with the runoff coefficients and that the range in $S_{r,clm}$ is larger than the range in $S_{r,soil}$. The relation between $S_{r,clm}$ and runoff coefficient is less pronounced for catchments where grass dormancy is important. The latter is probably an artefact of the dormancy assumption: transpiration is set to zero in summer, meaning that transpiration must be higher in winter to close the water balance. However, this is not incorporated in the analysis (see also Sect. 4.5.1).

Although Figures 4.2 and 4.5c show that the runoff coefficient has a strong control on $S_{r,clm}$, the figures also show it is not the only controlling factor. Other factors can be interstorm duration, seasonality and yearly precipitation depth (Gao et al., 2014b). An example can be found on the South Island: $S_{r,clm}$ for catchments I and II (indicated with red circles in Fig. 4.5) is higher than the values for the dry catchments and extremely wet catchments, while they have a runoff coefficient larger than the dry and smaller than the extremely wet catchments. A possible reason for their high $S_{r,clm}$ is that although they are classified as wet, their precipitation surplus is much smaller than for the average wet catchments (Fig. 2.7). But because they are classified as wet, the vegetation is assumed not to undergo dormancy and therefore to require sufficient soil water to maintain transpiration through the summer (Wang-Erlandsson et al., 2016) (see also Sect. 4.5.1).

4.4.2. COMPARISON OF MODEL RESULTS

HYDROGRAPHS

Figure 4.6 shows the observed and modelled hydrographs for three contrasting catchments: one where soil has more explanatory power (i.e., using $S_{r,soil}$ gives more accurate results) (a), one where soil and climate give similar results (b) and one where climate has more explanatory power (c). Details about these catchments, including derived $S_{r,max}$-
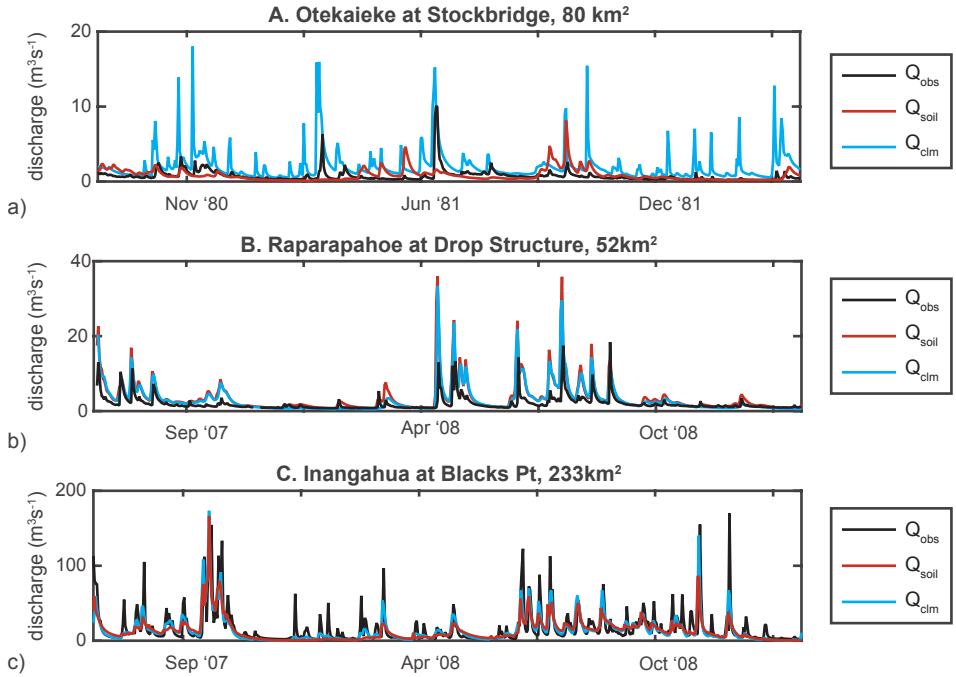
Figure 4.6: Observed and modelled discharge for both methods for three catchments; a) $S_{r,soil}$ better (Otekaieke at Stockbridge, Otago); b) $S_{r,soil}$ and $S_{r,clm}$ in balance (Raparapahoe at above drop structure, Bay of Plenty); c) $S_{r,clm}$ better (Inangahua at Blacks pt, West Coast). For details about the catchments refer to Tab. 4.1.

values can be found in Table 4.1. The figure indicates that replacing $S_{r,soil}$ by $S_{r,clm}$ can have different effects: $S_{r,soil}$ is similar for all three catchments while $S_{r,clm}$ strongly varies. For the top panel a small change in $S_{r,max}$ causes considerable changes in the responsiveness of the catchment, while for the middle panel even a doubled $S_{r,max}$ has only a limited influence on the modelled discharge. In the lower panel, in contrast, a decrease in $S_{r,max}$ leads to a flashier flow; however, to match the responsiveness of the observed flow, $S_{r,clm}$ should decrease even more.

Although the results in Figure 4.6 show reasonable matches between the modelled and observed flow, they also contain some clear shortcomings. Figure 4.6 for example shows that because the model is not calibrated or tailored for the specific catchments some observed features are still poorly reproduced by the model. Replacing $S_{r,soil}$ by $S_{r,clm}$ does not change the reproduction of these features, which indicates that these features are dominated by other (poorly identified) parameters (Clark et al., 2008a; Booker and Woods, 2014).

SPECIFIC SIGNATURES
Figure 4.7 shows the values of two signatures for all catchments for the observed and the two modelled cases. The catchments are ordered by increasing observed runoff coeffi-
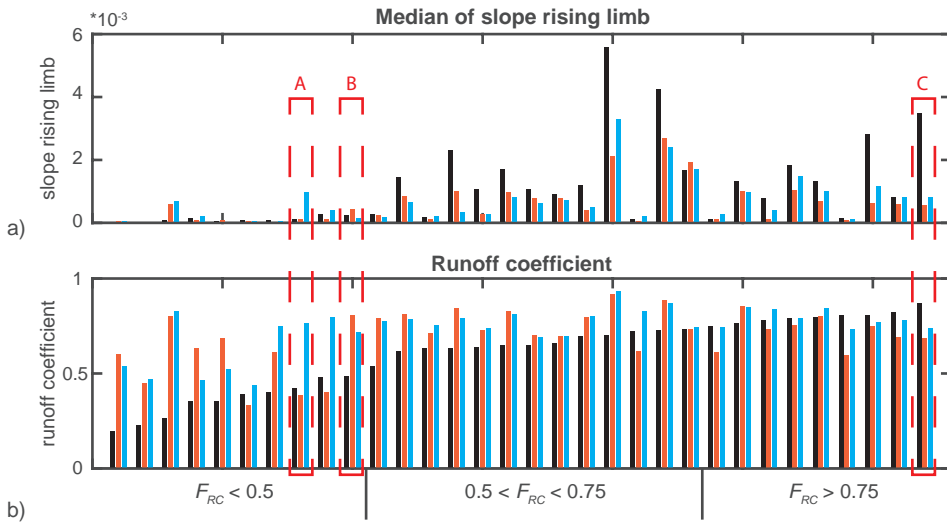
Figure 4.7: Reproduction of a selection of the used signatures, both panels are ordered by increasing runoff coefficient; red rectangles indicate catchments presented in Fig. 4.6 (black (first) bar = observed, red (second) bar = soil, blue (third) bar = climate); a) Median of slope of rising limb; b) runoff coefficient.

cient in both panels; it can be seen that high runoff coefficients do not always coincide with high slopes of the rising limb, although generally, the catchments with lower runoff coefficients have lower slopes for their rising limbs. The signature values for the two modelled cases are either both too high or both too low for the majority of the catchments and for both signatures, demonstrating that the value of $S_{r,max}$ is only part of the control on these signatures. Despite this, the model with $S_{r,clm}$ can better reproduce these two signatures for a slight majority of the catchments. Further it follows from the graph that the model is better able to reproduce the variability in slope of rising limbs than the variability in runoff coefficients, irrespective of using soil or climate data to estimate $S_{r,max}$.

COMBINATION OF SIGNATURES

By combining the results of all signatures and catchments, an overall comparison between soil and climate derived $S_{r,max}$-values can be made. Figure 4.8 shows whether a signature (rows) can be better reproduced with $S_{r,soil}$ (red squares) or $S_{r,clm}$ (blue dots) for each catchment (columns). The shading of the symbols indicate the difference between $S_{r,clm}$ and $S_{r,soil}$; the absence of a symbol indicates no significant difference between $S_{r,clm}$ and $S_{r,soil}$. It can be seen that the differences between $S_{r,soil}$ and $S_{r,clm}$ are largest for dryer catchments (left side); differences are smallest for the intermediate catchments (F_{RC} between 0.5 and 0.75). Further, the figure shows that for the dry and intermediate catchments $S_{r,soil}$ and $S_{r,clm}$ perform equally well (57 red vs. 63 blue and 45 red vs. 45 blue respectively), while for the wet catchments $S_{r,clm}$ strongly outperforms $S_{r,soil}$ (20 red vs. 62 blue). Regarding the signatures, Figure 4.8 shows that the differences

between $S_{r,clm}$ and $S_{r,soil}$ are larger for those focusing on the event time scale (lower part of graph) and thus the shape of the peaks, i.e., the short term memory of the system. The differences are smaller for the signatures focusing on the longer time scale (upper part of graph). A reason for this could be that the root zone storage is more important for the high frequency processes, which have a stronger existence in the event time scale (i.e., overland flow) than in the yearly time scale (i.e., base flow) (Oudin et al., 2004).

The top row in Figure 4.8b shows the Euclidean distance (Eq. 3.1) as an overall performance measure, integrating all signatures for each catchment. The overall performance exhibits the same pattern as the individual signatures: while either climate- or soil-derived $S_{r,max}$ may perform better for the drier catchments, depending on the specific catchment, little differences were found for the intermediate catchments. In contrast, the results strongly suggest that climate has considerably more explanatory power for the wet catchments. The median difference in Euclidean distance to the perfect model is 0.11 in favour of $S_{r,soil}$, 0.1 in favour of $S_{r,clm}$ and 0.05 in favour of $S_{r,clm}$ for low, middle and high F_{RC} respectively. Figure 4.8a further underlines that the higher variability of $S_{r,clm}$ and thus in particular the very low values derived for wet catchments contribute to consistently higher model skill in these catchments. More generally, it can be seen that $S_{r,clm}$ outperforms $S_{r,soil}$ when the first is larger for the drier catchments and when $S_{r,clm}$ is lower for the wetter catchments. We suggest that this is because the soil-derived values have similar ranges of magnitude for all catchments, and do not, at least at the time scales of hydrological interest, have a relationship with climate. However, for wet catchments the precipitation deficits, if any, are smaller during the drier summer period than for dry catchments (Fig. 2.7) Thus, vegetation does not need to make use of the full soil depth, and therefore the soil-derived values are too large to be used in a hydrological model. In such wet catchments, climate-derived values result in improved model performance.

4.5. DISCUSSION

4.5.1. APPLICATION OF CLIMATE METHOD

By using the described climate method I hypothesise that vegetation adapts its rooting system according to the storage required by the evaporative demand, leading to smaller storage capacities in wet areas and larger capacities in dry areas. This is hypothesized because in wetter areas the periods are generally shorter in which the evaporative demand exceeds the precipitation and characterized by a smaller precipitation deficit than for drier areas. This hypothesis would also imply that the medium-term dynamics of climatic variability can be found in the medium-term dynamics of root development, and thus in root zone storage capacities. The results show that the differences in $S_{r,clm}$ between catchments are larger than the differences in $S_{r,soil}$ between catchments and that a model with $S_{r,clm}$ has a higher performance for a majority of the catchments. To apply the climate based method, several assumptions need to be made which may influence the derived root zone storage capacity, concerning grass dormancy, the return period, and interception storage. Figure 4.9 shows a sensitivity analysis regarding these assumptions for a subset of catchments containing dry and wet catchments (Tab. 4.1). The influence of dormancy was not tested for the extremely wet catchment, as it is not
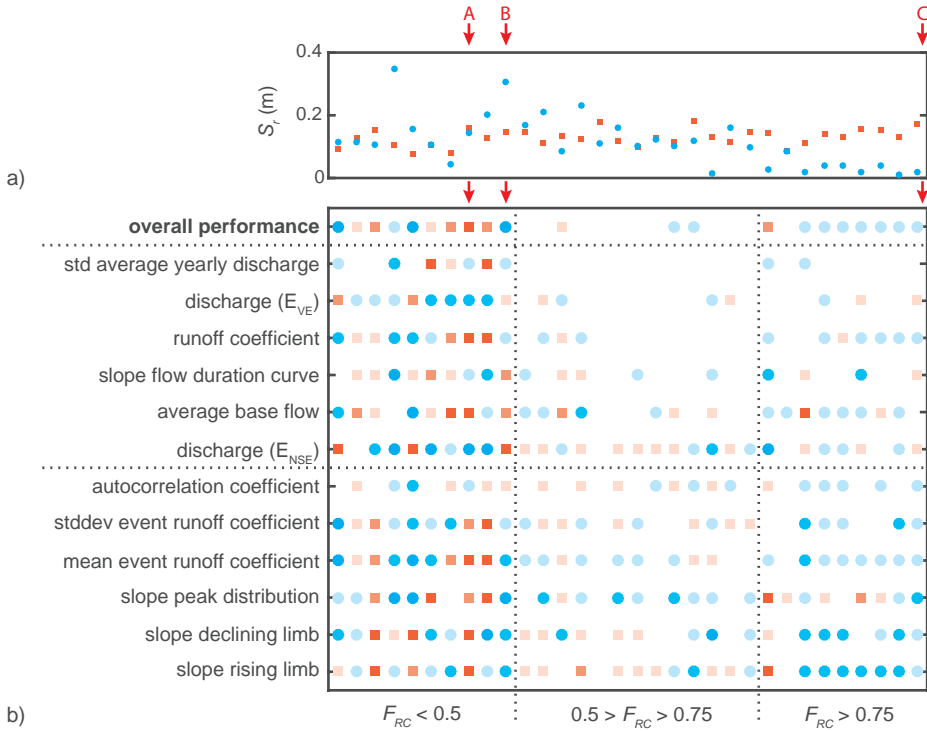
Figure 4.8: a) Soil- (red squares) and climate- (blue circles) derived $S_{r,max}$, catchments are ordered by increasing runoff coefficient; b) Overview of signature scores: blue circles and red squares indicate that a specific signature can be better reproduced with $S_{r,clm}$ respectively $S_{r,soil}$ for a specific catchment. The shading of the symbols from dark to light indicates the degree of performance difference by using $S_{r,clm}$ and $S_{r,soil}$ (differences in performance of > 0.5 (darkest), 0.2-0.5 (middle), 0.03-0.2 (lightest), < 0.03 (no symbol)). Signatures on top focus more on long term water balance, while signatures at the bottom focus more on peak shape and height. The arrows indicate the catchments used in Fig. 4.6.

relevant under such conditions.

Figure 4.9 shows that the assumption regarding grass dormancy has the largest influence on the derived $S_{r,clm}$. In the current experiment, the effect of dormancy is applied by setting transpiration to zero instantaneously from December to March for dry catchments with grass land cover. In contrast to the redistributed 'winter evaporative energy' (Sect. 4.3.2), this energy is not redistributed in the remainder of the year. The large uncertainty in $S_{r,clm}$ in these catchments therefore may explain the lower performance for these catchments: often the modelled discharge is too high and too responsive. This indicates that the derived $S_{r,clm}$ is too small and that setting transpiration to zero during dry summers probably results in underestimation of transpiration. Another reason for the inferior results in dry catchments is likely to be that grass does not go into dormancy instantaneously, but rather gradually (e.g., Ofir and Kigel, 1999). In addition, a

longer rain event in summer can lead to partial activation of transpiration of grass again. If these effects are reflected in the transpiration, larger values of $S_{r,clm}$ will be derived for these catchments. Instead of applying dormancy by setting transpiration to zero, it can also be taken into account by using a shorter return period (i.e., two to three years) for grasses (Wang-Erlandsson et al., 2016) and let the transpiration follow the potential transpiration.

For this experiment a return period of 10 years was chosen, following Gao et al. (2014b) and resulting in values in the same range as the soil derived values. Figure 4.9 shows that especially for dryer areas $S_{r,max}$ strongly depends on the selected return period, while this relation is much weaker for wet catchments. This makes it possible to derive more stable values of $S_{r,clm}$ for wet catchments than for dry catchments, which is likely to result in better model results. On the other hand, Figure 4.8 shows that $S_{r,clm}$ should probably be slightly smaller than the 10 year value for wet catchments (shorter return period) and larger for dry catchment (longer return period). The reason for the former may be that in wetter areas it is more likely that water is available at depths that can be reached by plant roots within a feasible period of time.

The influence on $S_{r,clm}$ of the chosen maximum interception storage is relatively small (Fig. 4.9), especially for the dryer catchments (grey lines coincide with coloured lines). For the wetter catchments the maximum interception storage has a larger influence, but still a smaller influence than dormancy or return period in dryer catchments. Surprisingly, the derived $S_{r,clm}$ values for a higher and lower maximum interception storage are both higher than the $S_{r,clm}$ value used during the study in case of no dormancy. For the lower interception capacity this is as expected; for the higher interception capacity it is caused by the decrease of potential transpiration.

4.5.2. INFLUENCE OF DATA QUALITY

$S_{r,soil}$ and $S_{r,clm}$ are both based on a set of data; errors in these data sources can have different effects on the derived values and the comparison between the two. One of the main data sources for $S_{r,clm}$ is precipitation. An underestimation of the precipitation, which is more likely to occur in the wet mountainous areas, would lead to an underestimation of $\overline{T}$. The influence on the required storage, on the other hand, will be smaller, as both precipitation and transpiration are smaller.

The quality of the soil data to a large extent depends on the spatial resolution of the measurements. The soil data was not originally collected for country-wide rainfall-runoff predictions, but rather as supporting information for agricultural practices, as in most places around the world. Therefore, it can be expected that the derived $S_{r,soil}$ is more reliable for the flatter agricultural areas, where it is easier to collect data, than for the mountainous forested areas. In the latter areas the climate method currently strongly outperforms the soil method; thus, this could be due to either a better representation by the climate method (with stable values in wet areas) or poorer data quality for the soil method. Irrespective of the reason, this performance difference clearly highlights the value of the climate method: with less (field) effort conceptually more adequate estimates can be achieved.
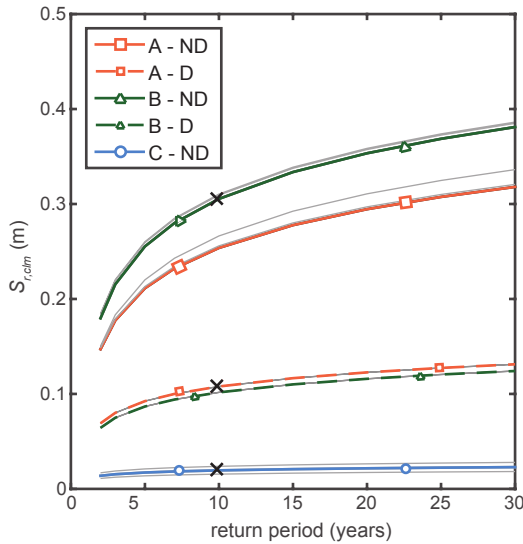
Figure 4.9: Sensitivity analysis for the subset of catchments: A is a dry, B a wet and C an extremely wet catchment (for more details see Tab. 4.1). Coloured lines show the $S_{r,clm}$ values for the interception capacity used during the entire analysis, the grey lines show the $S_{r,clm}$ values for an interception capacity which is 0.5 mm smaller or larger. The black crosses indicate the $S_{r,clm}$ values used in the analyses (ND = without dormancy, D = with dormancy).

4.5.3. MODEL EFFECTS AND IMPLICATIONS

Not only values of storage capacities, but also model results were compared for this experiment. In general, the outcome of the comparison of model results depends on the selected models and on how the newly derived values are incorporated in the model. Here, the uncalibrated version of TopNet was used, implying that parameters were estimated based on country-wide observed data. An uncalibrated model has the disadvantage that it does not perform very well in all catchments (Booker and Woods, 2014); the used model even already showed better performances in wetter areas as shown by McMillan et al. (2016). On the other hand, the advantage is that the other model parameters are not tuned towards a specific $S_{r,soil}$, therefore replacing only $S_{r,max}$, creates a more equal comparison than would be the case for a calibrated model. It should be noted that the long term average observed discharge was used to derive $S_{r,clm}$; the modelled discharges were then compared to the same discharge observations, leading to a small dependency between $S_{r,clm}$ and the observed discharge.

4.5.4. APPLICABILITY IN UNGAUGED CATCHMENTS

Understanding the hydrological behaviour of a catchment is important, both if the catchment is well gauged (like those used for this study), or if it is poorly gauged (like the majority of catchments worldwide; Hrachowitz et al., 2013a). The climate method described here does considerably improve understanding even of ungauged catchments. Although no discharge data is available for these catchments, estimates of $S_{r,max}$ can be obtained

from precipitation and evaporation data readily available worldwide from remote sensing products. This information is becoming more widely available and Wang-Erlandsson et al. (2016) have shown that when using these products very plausible worldwide estimates for $S_{r,max}$ can be derived.

4.5.5. Variables influencing $S_{r,max}$

This study compares the relative influence of soil and climate on the root zone storage capacity. However, more factors influence $S_{r,max}$ in addition to soil and climate. These factors are plant physiology, nutrient availability, plant competition and alternative plant survival strategies (e.g., a cactus plant stores water in its body systems); these factors can have large influences and in some cases overrule the influence of soil and climate. It is worth noting that any of these other factors have to be measured at point scale and therefore create the need for upscaling, which is not necessary for the climate method.

Although the described climate method is an engineering approach, it is based on the principle of co-evolution: vegetation strives to create an optimal environment and if this is not possible, the vegetation may not grow at the specific location. This notably implies that the suggested climate method is probably less suitable in areas where humans continually influence the vegetation, such as managed areas with (annual) crops.

4.6. Applicability of results in Meuse basin

The experiment described in this chapter uses 32 catchments in New Zealand, instead of the subcatchments of the Meuse which are used in the other chapters. This is because the climate gradient is much larger in New Zealand, making the comparison of the soil and climate method more interesting.

Following the results from this chapter, however, the climate derived root zone storage capacity can be used in the Meuse basin as well. Figure 4.10 shows $S_{r,clm}$-values for five different subcatchments of the Meuse for different return periods. Clear differences can be seen between the catchments, despite the small climate gradient, which is probably caused by the different land covers and thus evaporation patterns in the subcatchments.

One additional element is not yet accounted for in the New Zealand experiment: using $S_{r,clm}$ in combination with a model structures with multiple HRUs included. Different $S_{r,max}$-values for different HRUs is one of the most important constraints between HRUs (Sect. 3.3.2); up till now only (sub)catchment averages are determined, without taking into account different HRUs. There are roughly two options to take differences in $S_{r,max}$ between HRUs into account. First, using different $S_{r,max}$-values for different vegetation types, as demonstrated by Wang-Erlandsson et al. (2016). This option requires more or less homogeneous vegetation per HRU. Second, the $S_{r,max}$ for wetlands can be fixed to a small value (considering its quick saturation); following, from the wetland and catchment average $S_{r,max}$-values, the value for the remainder (hillslope/plateau) of the catchment can be determined. A consequence of this option is that only two different $S_{r,max}$-values can be assigned to the catchment. Probably, a combination of both methods would lead to the best results.
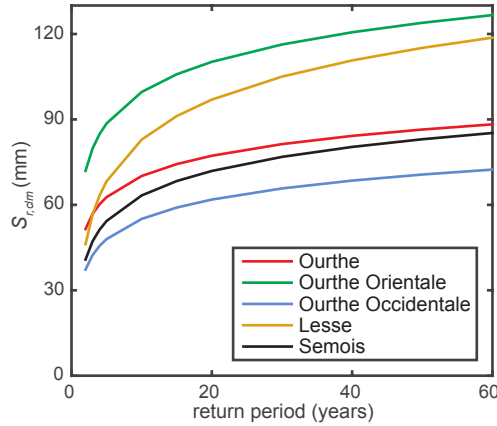
Figure 4.10: $S_{r,clm}$ for five subcatchments of the Meuse basin for different return periods.

4.7. CONCLUSIONS

This experiment shows a comparison between soil and climate derived root zone storage capacity ($S_{r,max}$). $S_{r,max}$-values from climate and soil were compared directly as well as via hydrological model results for 32 contrasting catchments in New Zealand. The key findings are that climate-derived $S_{r,max}$-values on balance outperform soil-derived values for most natural catchments, based on multiple metrics. For drier catchments the differences in model results with $S_{r,max}$-estimates from soil and climate are larger than for wetter catchments. In wetter catchments, climate-derived $S_{r,max}$-values clearly outperform soil-derived values, despite smaller absolute differences in performance. Thus, we can conclude that climate data has a higher explanatory power for $S_{r,max}$ than soil data, this higher explanatory power allows for taking into account the medium term development of catchment vegetation. Combining the medium term dynamics and the easier accessibility of data, makes the climate derived $S_{r,max}$ a valuable addition to hydrological and climate models and opens doors to evaluate changes in catchment response within a changing climate.

5

DISTRIBUTION OF FORCING AND MODEL STRUCTURE

In science, we must be interested in things, not in persons.

Marie Curie, quoted by Eve Curie Labouisse (Madame Curie: A Biography)

This chapter compares the value of incorporating both spatially distributed meteorological forcing and spatially distributed model conceptualisations, as a first step in determining the appropriate level of distribution. Distributed forcing data were used to create a spatial distribution of model states. The model experiment consists of eight different configurations: a lumped and distributed model structure, each with four levels of model state distribution. The results show that the distributed model structure can in general better reproduce the dynamics of the hydrograph, and furthermore, that the differences in performance and consistency between calibration and validation are smallest for the distributed model structure with distributed model states. It can be concluded that the positive effect of incorporating a distributed model structure is larger than that of incorporating distributed model states.

5.1. Introduction

As briefly noted in Chapter 1, distribution in hydrological models can roughly be applied in two ways: distribution of the meteorological forcing and distribution of the model. Distribution of the latter can refer to both the model conceptualisation as the model parameters.

Many studies have compared the relative effect on model performance of distributed forcing and distributed parametrisations, where both distributions were based on sub-catchments (e.g., Boyle et al., 2001; Ajami et al., 2004; Zhang et al., 2004; Carpenter and Georgakakos, 2006; Das et al., 2008; Kling and Gupta, 2009). Distributed parametrisations have often been based or linked to detailed physical catchment properties, like soil and vegetation. Although the results vary between the different studies, most studies have shown that distribution of forcing has a larger effect on model performance than distribution of parameters (Boyle et al., 2001; Ajami et al., 2004; Andréassian et al., 2004). With the exception of Kling and Gupta (2009), who showed that the noise in distributed parameter values was caused by physical properties, rather than by spatially variable meteorological forcing. A limiting factor of the distribution of parameters is the increased degrees of freedom, making it difficult to meaningfully calibrate the model, a problem called equifinality (Beven, 2006).

Before more detailed options of forcing or model distribution can be explored, it is relevant to compare the effect of distributed forcing with the effect of parallel model structures based on HRUs. Therefore, the objective of this chapter is to test and compare the relative importance of distributing meteorological forcing and distributing model structures for improving model performance and consistency, using the diagnostic tools described in Chapter 3.

5.2. Model experiment

5.2.1. Input data

For this experiment the entire Ourthe catchment was used; ten precipitation gauges were used (all dark dots in Fig. 2.2 except st. Hubert). The meteorological stations at Rochefort, Nadrin, Bierset and st. Hubert were used to derive potential evaporation; elevation zones were used to distribute the meteorological data over the ten calculation cells (i.e., Thiessen polygons around precipitation gauges). The period 01 September 1999 - 01 July 2000 was used as model spin-up period, followed by a calibration period 01 July 2000 - 01 July 2002 and a validation period 01 July 2002 - 01 July 2005.

5.2.2. Model configurations

Eight model configurations were tested in this experiment, composed of a lumped and a distributed model structure and with different options for model state distributions (Tab. 5.1). The model structures are based on two dominant processes: saturation overland flow (SOF) and subsurface storm flow (SSF). Where the first one is assumed to be dominant in the wetland HRU and the latter in the hillslopes HRU (Sect. 2.2.2). The lumped model structure takes only hillslopes (H) into account, the distributed model structure both wetlands and hillslopes (WH). Note that hereafter the term 'model structure' is used for the set of conceptual reservoirs and fluxes, and the term 'model config-

Table 5.1: Different configurations for model experiments, with the accompanying abbreviations.

| Model states | Model structure[a] | |
	Lumped	Distributed
lumped	*H-L*[b]	*WH-L*
interception reservoir distributed	*H-I*	*WH-I*
interception and root zone reservoirs distributed	*H-IR*	*WH-IR*
interception and root zone and fast runoff reservoirs distributed	*H-IRF*	*WH-IRF*

[a] The lumped model structure is presented in Fig. 5.1a and the distributed model structure in Fig. 5.1c.

[b] The character before the hyphen gives the model structure: lumped (hillslope; 'H') and distributed (wetland and hillslope; 'WH'); the characters after the hyphen give the model state distribution level (L/I/IR/IRF).

uration' for the combination of a model structure with a model state distribution level. This model approach is similar to the FLEX-Topo approach in Savenije (2010); therefore the name 'FLEX-Topo' is used for the distributed model developed in this thesis as well.

SATURATION OVERLAND FLOW (SOF)

Saturation overland flow was conceptualised with two reservoirs: a root zone (S_r) and a fast runoff generating (S_f) reservoir. Transpiration takes place from the root zone and the fast runoff reservoir has a short response time to create a fast response to precipitation. During drier periods the root zone is supplied with water from the groundwater reservoir (S_s) via capillary rise. The saturation overland flow is combined with an interception routine (S_i): excess precipitation overflows to the root zone and the interception reservoir evaporates with potential rate. Potential evaporation for transpiration is the potential evaporation decreased by the interception evaporation.

SUBSURFACE STORM FLOW (SSF)

Subsurface storm flow was conceptualised with the same two reservoirs as saturation overland flow. The two main differences between the two conceptualisations are the response time of the fast runoff reservoir and the interaction with the groundwater. The response time for SSF is larger than for SOF and the root zone reservoir is not supplied with water *from* the groundwater reservoir, but supplies water *to* the groundwater reservoir via preferential recharge and percolation. SSF is combined with an interception reservoir as well.

MODEL STRUCTURES

Figure 5.1 shows the conceptualisation of the lumped (a) and the distributed (c) model structure. The lumped model structure is equal to the one shown in Figure 3.1 and contains one model element (hillslope). The distributed model structure contains two parallel model elements (wetland and hillslope). For this experiment the HRUs hillslope and plateau are combined into 'hillslope'. Plateau is not taken into account separately, as this
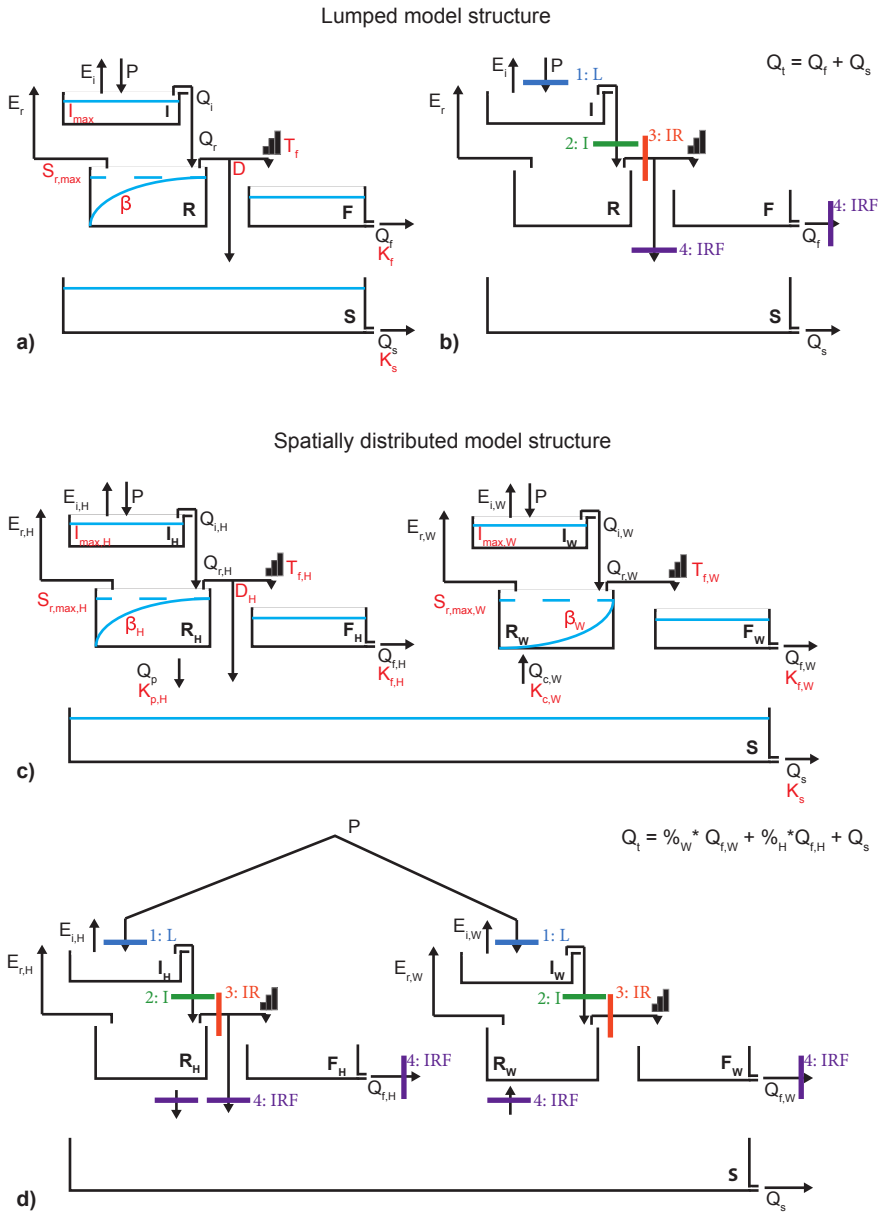
Figure 5.1: a) Lumped model structure. The red characters refer to parameters (total of 7) and the black lines refer to fluxes. b) Example for lumped model structure for distributed model states: coloured lines indicated locations where fluxes were averaged to create lumped and distributed model states. c) Distributed model structure: the right part shows the reservoirs for the wetland class and the left part shows the reservoirs for the hillslope class. Both parts are connected via the groundwater reservoir (S_S). The red characters refer to parameters (total of 13) and the black lines refer to fluxes. d) Example for distributed model structure for distributed model states: coloured lines indicated locations where fluxes were averaged to create lumped and distributed model states.

would lead to incorporating detailed processes, which are not necessary for a comparison between forcing and model structure distribution. The model elements for wetland and hillslope are connected by a combined ground water reservoir. The model equations and parameters for both model structures can be found in Tables B.1 till B.4.

Travel time through the river is incorporated in addition to the lag function for both model structures. Travel time through the river is accounted for by delaying fast runoff (Q_f) according to the estimated travel time along the flow paths in the catchments. Travel time through the river is estimated based on the averaged flow velocity in the river, which is derived from the maximum travel time through the catchment and assumed to be $1 \, \text{ms}^{-1}$. The flow path length is averaged over the entire catchment for the model state distribution levels with lumped forcing (L), the interception reservoir distributed (I) and the interception and root zone reservoirs distributed (IR). The flow path length is averaged for each calculation cell for the model state distribution level with the interception, root zone and fast runoff reservoirs distributed (IRF), as for IRF different values for Q_f were available per calculation cell.

MODEL STATE DISTRIBUTION

Four different levels of model state distribution were applied in the upper three reservoirs of the model structures, ranging from all being lumped to all being distributed. Distribution of model states here means a distributed accounting of the state variables, representing the storage in different model reservoirs. Parameters were not distributed between calculation cells, as this was not warranted by data, and previous studies suggested that distribution of forcing was more important (Ajami et al., 2004; Andréassian et al., 2004; Boyle et al., 2001).

Model states were distributed in four steps, each step with increasing distribution. To create lumped and distributed model states, the incoming fluxes were averaged at different locations in the model structure (Fig. 5.1b,d). The coloured lines in Figure 5.1b,d show the locations where fluxes were averaged over the catchment, according to the surface area of the calculation cells. This resulted in four distribution levels for the model states: all reservoirs lumped; interception reservoir distributed; interception and root zone reservoirs distributed; interception, root zone and fast runoff reservoirs distributed. These four distribution levels for model states were applied to both model structures (Tab. 5.1; Fig. 5.1b,d). The slow runoff reservoir was not distributed as it was assumed that the spatial variability in this reservoir was very small; this will be elaborated further in Section 5.3.1.

5.2.3. MODEL CONDITIONING AND EVALUATION

Model conditioning and evaluation are carried out with the methods and tools described in Chapter 3. Some of these methods and tools need some further specification, which can be found below.

MODEL CONSTRAINTS

The five parameter and five process constraints used for this experiment are summarized in Table 5.2. All constraints were selected from those presented in Table 3.1. For the different model configurations 1000-1250 model realisations were kept after applying all

Table 5.2: Parameter and process constraints for the lumped (H) and distributed (WH) model structure.

Parameter constraints		Process constraints	
$S_{r,max;W} < S_{r,max;H}$	WH	F_{RC}[a]$ < 0.52$[b]	L/D
$\beta_W > \beta_H$	WH	$F_{RC} > 0.4$[b]	H/WH
$I_{max;W} < I_{max;H}$	WH	$F_{RC,low} < 0.3$[c]	H/WH
$K_{f;W} < K_{f;H}$	WH	$F_{RC,high} < 0.97$[d]	H/WH
$K_s = 2500$ (h)	WH	$F_{RC,high} > 0.6$[d]	H/WH

[a] F_{RC} = modelled runoff coefficient.   [b] Applied for the entire modelled period.   [c] Applied for the low flow period (selected from the calibration period).   [d] Applied for the high flow period (from December till March).

Table 5.3: Overview of used signatures for model evaluation and period of flow to which they are applied.

Signature	Metric[a]	Period
discharge ($F_{Q,NSE}$)	E_{NSE}	Total, high
log of discharge	E_{NSE}	Total, low
discharge ($F_{Q,VE}$)	E_{VE}	Total, high
coefficient of autocorrelation (lag = 1 day)	E_{RE}	Total, low
coefficient of autocorrelation (1h < lag < 250h)	E_{NSE}	Total, high, low
rising limb density	E_{RE}	Total
declining limb density	E_{RE}	Total, high
peak distribution	E_{RE}	Total
flow duration curve	E_{NSE}	Total, lowest 15%

[a] The observed and measured signatures are compared with three different metrics: Nash-Sutcliffe Efficiency (E_{NSE}), Volume Error (E_{VE}), and Relative Error (E_{RE}).

constraints. The constraints regarding the maximum runoff coefficient over the total period and over the dry period turned out to have the largest filtering capacity.

HYDROLOGICAL SIGNATURES

For this study, nine signatures were selected (Tabs. 3.3 and 5.3), each focussing on a different part of the hydrograph. All these signatures were calculated for the entire period; in addition, a selection of signatures is also calculated for the low (June - August) or high (December - March) flow period. Nash-Sutcliffe Efficiency (E_{NSE}; Nash and Sutcliffe, 1970)), Volume Error (E_{VE}; Criss and Winston, 2008) and the Relative Error (E_{RE}; Eq. 3.2) were used to compare the modelled and observed signatures; for each signature a different evaluation metric was selected.
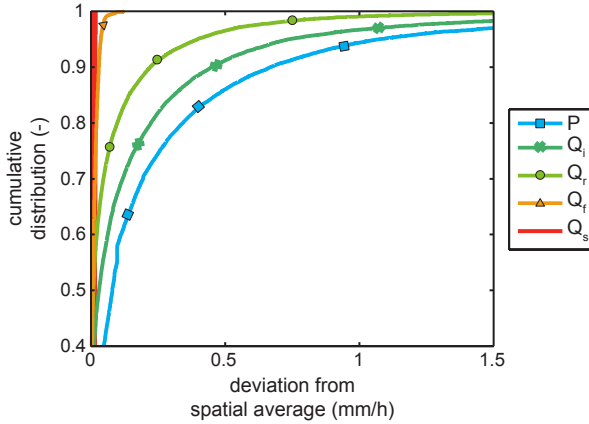
Figure 5.2: Sensitivity of different reservoirs for variability in precipitation data. The line corresponding to Q_s overlaps with the y-axis.

DETAILS FOR FARM ANALYSIS

In addition to the selected signatures, two other specifications are required for the FARM analysis: definition of the categories for the performance and for the range in performance. For the performance three categories were defined: high (solid lines): $E_{c,i,10} > 0.85$; middle (dashed lines): $0.55 < E_{c,i,10} < 0.85$; low (dotted lines): $E_{c,i,10} < 0.55$. Three categories were defined for the range (difference between 10th and 90th percentile) in performance as well: small (green): $E_{c,i,10} - F_{c,i,90} < 0.15$; middle (orange): $0.15 < F_{c,i,10} - F_{c,i,90} < 0.5$; large (red): $F_{c,i,10} - F_{c,i,90} > 0.5$).

5.3. RESULTS

Both the lumped and the distributed model structure were run with the four options for model state distribution. Tables B.2 till B.4 present the posterior parameter ranges of the model realisations which satisfied all parameter and process constraints.

5.3.1. INTERNAL MODEL BEHAVIOUR

DISTRIBUTION OF MODEL STATES

Figure 5.2 represents the results for the sensitivity of each reservoir to variability in precipitation. Each line shows the cumulative distribution for a flux leaving one of the model reservoirs. For the computation of the deviation from the spatial average, all ten calculation cells are taken into account. The graphs start at 0.4, as for 40 % of the time steps no precipitation was recorded for all stations. The lines in Figure 5.2 show that the variability, i.e. the deviation from the spatial average, in outgoing fluxes decreases for reservoirs lower in the model structure, which is caused by the filtering mechanism of the different reservoirs (e.g., Oudin et al., 2004). The variability in P, Q_i and Q_u is much larger than the variability in Q_f and Q_s, indicating that distribution of the first two reservoirs (interception and root zone reservoir) is more important than distribution of the lower two reservoirs (fast and slow runoff reservoirs).

CONTRIBUTION OF DIFFERENT FLUXES

Figure 5.3 shows the contribution of different fluxes to the total discharge, indicating the difference in behaviour between the lumped and distributed model structure. For this comparison, a winter and a summer period in the validation period were used. The contributions in time clearly show that for the distributed model structure the sharp peaks are generated by the wetlands, both in summer and winter. This leads to higher variability in the dynamics for both the hillslope and wetland response compared to the quick response from the lumped model structure. The quick wetland response was expected, as these areas are quickly saturated and can, therefore, generate quick and sharp peaks.

The main difference for groundwater is the contribution in summer: this is higher for the lumped model structure than for the distributed model structure, while it was expected that during low flows almost all discharge comes from the groundwater, as during this period the hillslopes are expected to be less connected. The large contribution from the hillslope during low flow conditions indicates that the distributed model structure still misses an important runoff process, such as Hortonian overland flow. This will be further elaborated in Section 6.3.1.

5.3.2. MODEL PERFORMANCE AND CONSISTENCY

HYDROGRAPHS

Figure 5.4 shows the hydrographs for the observed discharge together with the modelled discharge from H-L, WH-L and WH-IRF (Tab. 5.1), for the same periods as in Figure 5.3. The hydrographs are among the best performing realisations for each model configuration, based on $F_{Q,NSE}$, $F_{Q,log}$ and visual inspection. Small differences can be better visualised when single hydrographs are compared; therefore, only one model realisation is presented for this analysis; the analyses in Sections 5.3.2 and 5.3.3 contain all feasible model realisations, where the analysis in Section 5.3.2 shows more quantitative results for the reproduction of the dynamics in the hydrograph.

The three model configurations presented in Figure 5.4 show all a good overall performance ($F_{Q,NSE}$ of 0.83-0.85 and $F_{Q,log}$ of 0.62-0.88), but still there are some differences between the configurations. First, the dynamics of the discharge are better reproduced by the distributed model structure (blue line) than by the lumped model structure (purple line) for these realisations. For example, the shape of the recession and the peak distribution is better. In January 2004 (grey rectangle) the peak heights are equal for the three consecutive peaks for the distributed model structure and increases in time for the lumped model structure; for the observed discharge (red line) these peaks are of equal height. Furthermore, at the end of summer, the peaks of the lumped model structure are too low. This is slightly better for the distributed model structure as it better captures the dynamics during the wetting-up of the system.

While the peaks of the lumped model structure are not always high enough for these model realisations, the peaks of the distributed model structure with lumped forcing are sometimes too high and often too sharp. When the model states are distributed up to the fast runoff reservoir (green line), the shapes of the peaks become better, due to attenuation. This attenuation is caused by the distributed travel time through the river which is applied for this configuration. In addition to the shape of the peaks, WH-IRF also gives some more response during the first large peak at the end of the summer (grey
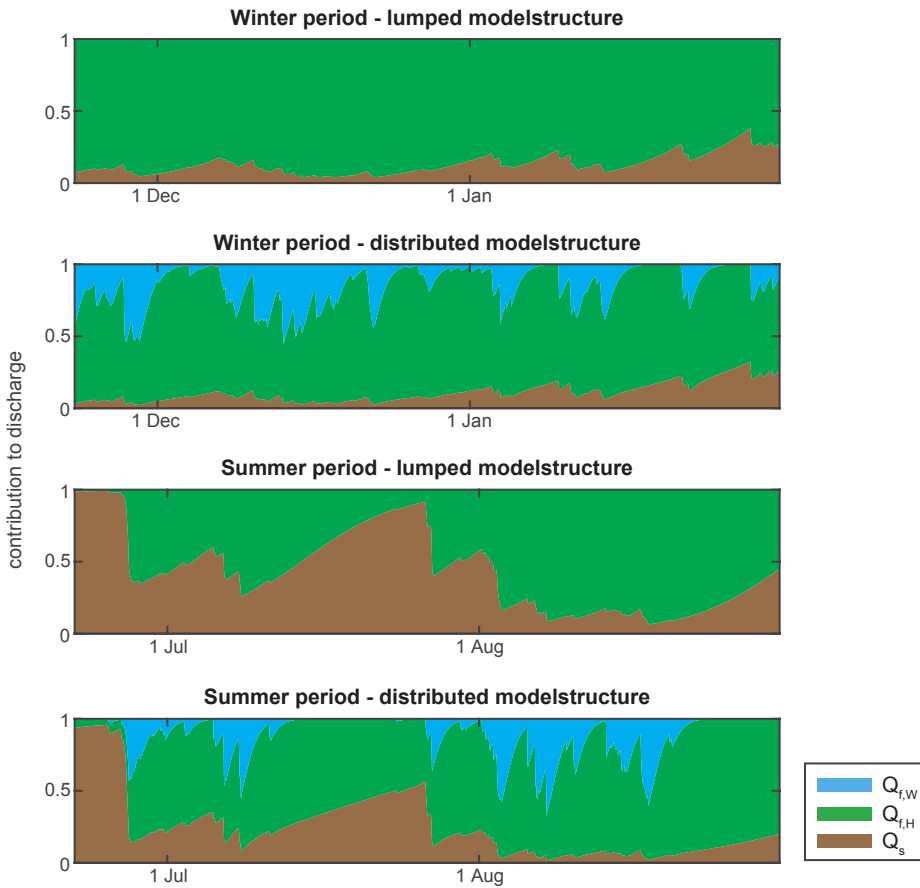
Figure 5.3: Contribution of the different fluxes to the total discharge, for a winter (23 December 2003 - 28 February 2004) and summer (5 July - 10 September 2004) period during validation for H-L and WH-L. For H-L the fractions of $Q_{f,H}$ ($= Q_f$) and Q_s are shown and for WH-L the fractions of $Q_{f,H}$, $Q_{f,W}$ and Q_s are shown.

rectangle). The reproduction of this peak is still very limited, but distribution of model states has a clear effect. The limited reproduction of this peak can probably be linked to overestimation of actual evaporation during the summer period or a missing runoff process, as the volume of observed precipitation is larger than the volume of the observed discharge for this event. This is further elaborated in Section 6.3.2.

FARM
Figure 5.5 shows results of the FARM analysis, which is used to assess both performance and consistency of the model configurations, based on all model realisations from the validation period. The figure shows that performance and consistency of WH-IRF are highest, indicating that the dynamics of the observed hydrograph can be better reproduced. The higher number of solid-line vectors indicates that the performance for WH-
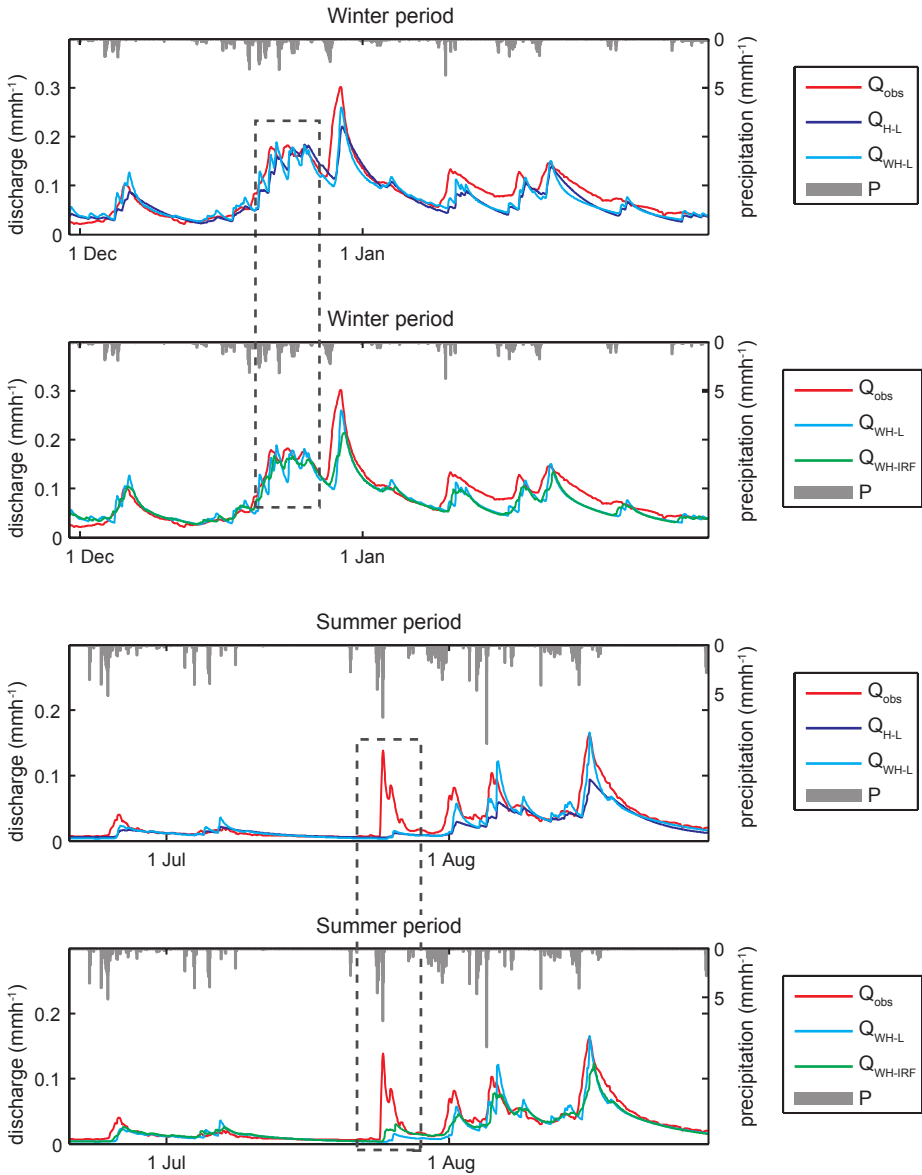
Figure 5.4: Hydrographs for the same periods as shown in Fig. 5.3, for H-L, WH-L and WH-IRF. Per model only one hydrograph is presented, to visualise specific differences between the model configurations. The selected hydrographs are among the best performing for each model configuration. The grey dotted rectangles show time spans with main differences between the model configurations. Note that the discharge scale is different for the summer and winter period, while the precipitation scale is equal.

IRF is higher than for H-L and WH-L, i.e., different signatures can be better reproduced. In addition, the ranges in performance are smallest for WH-IRF: ten green vectors, versus eight and five for WH-L and H-L respectively. The consistency is less straightforward to determine: H-L has fewer vectors with opposite directions, indicating increased consistency, compared with WH-L and WH-IRF. However, H-L has some opposite long red vectors, which reduce consistency, while this is less pronounced for WH-IRF. In addition, WH-IRF has three vectors with a large loading on PC3 (not shown), two of which are in the same direction, indicating improved consistency.

Besides a global evaluation, more specific model behaviour can also be derived from Figure 5.5. For example, the F_{AC} and $F_{AC,func}$ (4 and 5) represents the flashiness and therewith partly the response time of the catchment. For H-L, these signatures for the low and high flow period are only partly correlated with those for the total period, while for WH-IRF they are very well correlated. For WH-L, on the other hand, F_{AC} for the total period (4) and the low flow period (4L) are inversely correlated, meaning that both cannot be reproduced well with the same parameter set. Probably the autocorrelation signatures can be better reproduced by the distributed model structure, as this model structure can build its response up from two separate fast components (saturation overland flow and subsurface storm flow), which accommodate two time scales within the hydrological response. The inverse correlation of 4 and 4L for WH-L indicates that the combination of model state and model structure distribution is important to reproduce the flashiness of the catchment. Another example is the peak distribution (8), this signature is inversely correlated with most of the other signatures for all three configurations. This indicates that both model structures do not represent the right processes to reproduce the overall peak distribution. The limitations of certain model configurations pointed out by the FARM analysis are difficult to generalise, but can be used as a basis for further model improvement.

5.3.3. PREDICTIVE POWER

SCATTERPLOTS

Figure 5.6 shows the scores for $1 - F_{Q,NSE}$ and $1 - F_{Q,log}$ for all feasible model realisations: the top and bottom row show the configurations with a lumped model structure and a distributed model structure respectively. These two evaluation criteria have been selected as they represent the overall performance for both high and low flows.

The figure shows that the differences between the model configurations are small for the calibration period (grey dots), but large for the validation period (black dots). For the configurations WH-I, WH-IR and WH-IRF the performance during the validation period is much more comparable with the calibration period than for the other configurations. This implies that the distributed model configurations have a higher predictive power than the lumped configurations. The distributed model structure with all model states lumped, performs especially poor for low flow during the validation period, which indicates the importance of distribution of model states in case of a distributed model structure. On the other hand, the difference between WH-I, WH-IR, WH-IRF is very small, both in the calibration and the validation period. This indicates that the additional distribution of the root zone and fast runoff reservoirs are of minor importance compared to distribution of the interception reservoir.

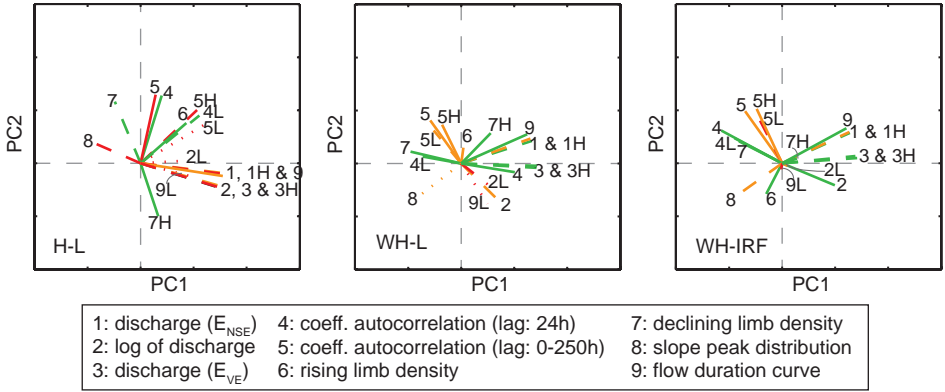| 1: discharge (E_{NSE})	4: coeff. autocorrelation (lag: 24h)	7: declining limb density
2: log of discharge	5: coeff. autocorrelation (lag: 0-250h)	8: slope peak distribution
3: discharge (E_{VE})	6: rising limb density	9: flow duration curve

Figure 5.5: PCA for H-L, WH-L and WH-IRF for the validation period. The lengths of the vectors are corrected for variance explained per PC, so the weight of both PCs shows equal. 'L' or 'H' combined with the number of a signature, indicate the low and high flow period. The line style of the vector indicates the 10th-percentile for the evaluation criterion: a solid vector indicates the best performance, followed by a dashed and a dotted vector. The line colour indicates the difference in performance between the 10th and 90th percentile: green indicates a small, orange a middle and red a large range.

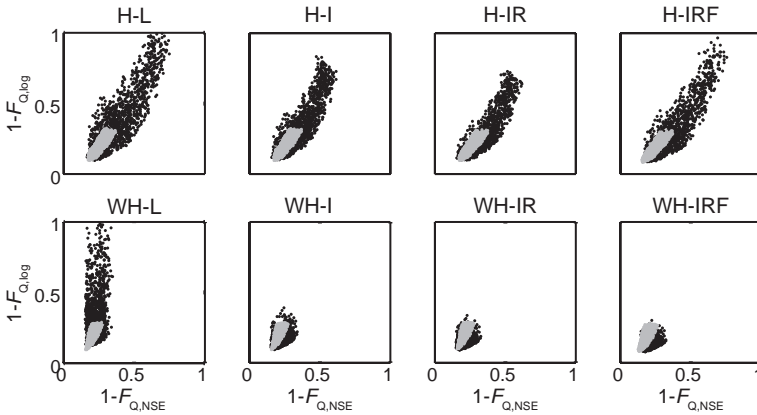Figure 5.6: Scatter plots of $1-F_{Q,NSE}$ vs. $1-F_{Q,log}$ for all model configurations; each dot represents a model realisation which satisfies all imposed constraints. The origin of the graphs represents the best performance. The top row shows results for the lumped model structure and the bottom row for the distributed model structure (grey dots = calibration period, black dots = validation period).
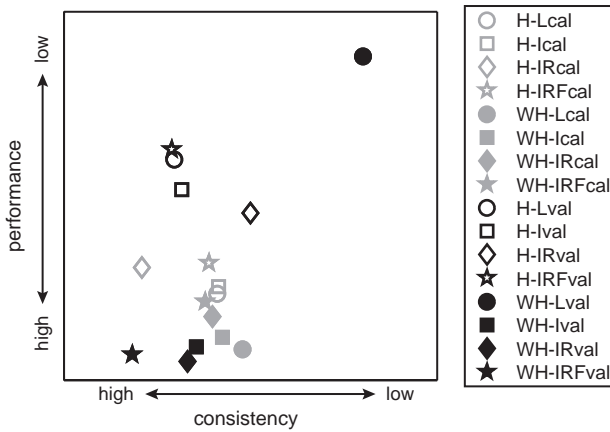
Figure 5.7: Overall performance and consistency of all model configurations. A low value for consistency means a consistent model and a small Euclidean distance means a high performance. The open symbols show the lumped model configurations, the solid the distributed model configurations; grey symbols are for the calibration period, black symbols for the validation period.

PERFORMANCE AND CONSISTENCY

To make an overall comparison, the performance and consistency derived from the FARM analysis were compared for all model configurations. Figure 5.7 shows for each configuration a measure for performance and for consistency. For performance the 10th-percentile of the Euclidean distance (Eq. 3.1) of all evaluation criteria (Tab. 5.3) for all realisations was used. For consistency the sum of the weighted angles of all signatures for the first three principal components was used. Smaller values mean a better performance or consistency.

The performance for the distributed model structures is generally better, and the consistency of these structures even improves during the validation period. The latter has probably to do with that the hydrological and meteorological conditions occurring in the validation period are easier to model. For the calibration period (grey symbols), the configurations are clustered and the differences are small; the performance of the configurations with a distributed model structure (closed symbols) is slightly better. However, H-IR has the most consistent configuration during calibration. Turning to the validation period (black symbols), a clear difference between lumped and distributed configurations emerges: the performance and consistency of most distributed model structures except for WH-L clearly improve, while the performance of the lumped model structures (open symbols) decreases.

5.4. DISCUSSION

The results presented in the previous section show that both distribution of model structure and distribution of model states (up to a certain degree) are important to increase model performance and consistency. The model structures and parametrisations used are likely to be non optimal for this catchment, but as the basis of both the lumped and

distributed model structure is the same, the investigation of the effect of different distribution levels is possible.

5.4.1. Distribution of model structure

Figures 5.4, 5.6 and 5.7 show in general better performance for the distributed model structure. A possible reason for this improvement could potentially be the result of the increased degrees of freedom of the distributed model structure. As discussed (see Fig. 5.4 and 5.5), the distributed model structure can reproduce the dynamics of the observed hydrograph better than the lumped model structure. In addition, Figures 5.6 and 5.7 show that the performance of the distributed model structure is stable or increases during the validation period for all model realisations, while the performance of the lumped model structure decreases for a large part of the model realisations. Therefore, with the use of constraints, the additional complexity of the distributed model structure is warranted to increase model realism and predictive power.

For this experiment, the catchment was only split into two landscape classes. The amount of classes and the accompanying model structures may differ per catchment. What is key in the method is to define the landscapes (or hydrological response units) that correspond with the dominant hydrological processes. In this case two response units already show an improvement compared to the lumped model structure. Chapter 6 investigates the influence of more response units.

This experiment also shows that a parallel arrangement of landscapes provides plausible results, whereas most other modelling approaches arrange hydrological response units in a serial configuration (e.g., Flügel, 1995; Fenicia et al., 2008a; Seibert et al., 2003; Uhlenbrook et al., 2004). The reason why parallel modelling is performing well in this catchment is because the subsurface stream lines do not cross between landscapes and surface processes connect directly to the stream network.

5.4.2. Distribution of model states

The results in Figures 5.4 and 5.6 indicate that the distribution of model states is important. However, it is only important up to a certain degree: for the lumped model structure IR seems to be the most optimal distribution level, while for the distributed model structure IR and IRF give the same results. This follows from the analysis of the internal behaviour (Fig. 5.2) as well: the spatial variability in the interception and root zone reservoirs is much larger than that in the fast runoff reservoir. In addition to model state distribution, distribution of the travel time through the river has a large effect on the shapes of the peaks. Thus, distribution of the fast runoff reservoir does not really add value, but using distributed travel times through the river for Q_f has a positive effect on the modelled hydrographs.

Figure 5.7 shows mainly clusters of symbols for the lumped or the distributed model structure, especially for the validation period. This clustering indicates that for the Ourthe catchment the effect of distributed model states is smaller than the effect of distributed model structure, which is in line with what Atkinson et al. (2003) found for the Mahuranghi catchment in New Zealand. However, if the model structure is distributed, distribution of forcing becomes more relevant. This is corroborated by the fact that if the model structure represents more detailed process dynamics, the dynamics in and

distribution of the forcing becomes more important. If this comparison would be made in a larger catchment with a stronger climate gradient, probably both the importance of model state distribution and process distribution would increase (e.g. Gao et al. (2014a)).

5.4.3. FARM AS EVALUATION TOOL

Part of the model evaluation is based on the FARM framework, with the advantages that multiple model realisations and multiple hydrological signatures can be easily taken into account. Furthermore, signatures which cannot be reproduced with the same parameter set can be visualised. However, disadvantages are, for example that the 2D-plot does not always give a complete overview (e.g. PC3 is very important for WH-IRF; Fig. 5.5) and results are sensitive to the selected signatures.

Because the FARM framework provides more than only one performance measure, I believe that it is a useful additional evaluation tool for model realism. More standard performance measures do often not give sufficient information about model behaviour. Evaluation of both consistency and performance enables a better assessment of the model's behaviour. For this experiment, the visual inspection of hydrographs and different landscape contributions to the discharge appeared to be very useful as well, thus the assessment of a suitable distribution level requires different model evaluation techniques.

5.5. CONCLUSIONS

This chapter compared the effects of incorporating distributed model states and a distributed model structure, based on landscape heterogeneity. For each model configuration, contributions to the discharge were investigated and performance and consistency were studied. The key findings are that a distributed model structure can better reproduce the dynamics of the hydrograph than a lumped model structure in the Ourthe catchment. In addition, the performance and consistency of the distributed model structure with distributed model states increase in the validation period, while for the lumped model structure the performance decreases during validation. However, the distributed model structure performs poorly during the validation period if model states are not distributed. Therefore, combining the distribution of model states and model structure is important. Although the parameter sets and the model structures are probably not the best ones for this catchment, it can be concluded that incorporating a distributed model structure has the largest effect on the model performance and consistency for the Ourthe catchment. Using different levels of model state distribution mainly affects smaller details of the modelled discharge, such as the shapes of the peaks. Although this study was performed in one catchment, we expect that distribution of model structure is beneficial for catchments were different dominant runoff processes can be identified, but the conceptualisation of the different parallel components should be adapted to the specific catchment. This experiment, and earlier work by Gharari et al. (2014a) and Gao et al. (2014a), indicate that parallel modelling of different landscape classes performs very well and that serial modelling of distributed hydrological response units is not required.

6

EFFECT OF MORE DETAILED PROCESSES ON GENERATED RUNOFF

People do not like to think. If one thinks, one must reach conclusions. Conclusions are not always pleasant.

Hellen Keller (Hellen Keller: selected writings)

Chapter 5 presented a model configuration with a wetland and hillslope HRU. This model still has some shortcomings, while the agricultural land use and snow processes were neglected. For this experiment the relevance of different process conceptualisations for snow and agricultural influence are tested stepwise. Among the conceptualisations are Hortonian overland flow in summer and winter, reduced infiltration capacity due to a partly frozen topsoil and the relative importance of rainfall and snow melt. The results show that these processes can make a large difference on event basis, especially the Hortonian overland flow in summer and the combination of snow melt and rain on frozen soil in winter.

6.1. Introduction

Chapter 5 described a case study comparing the relative importance of distributing forcing data and model structure. It was shown that the distribution of model structure was more important; however, the FLEX-Topo configuration with wetland and hillslope HRUs incorporated showed some shortcomings as well. For example the reproduction of the peak distribution and the saturation phase in autumn. With respect to the peak distribution, the model structure could not capture the highest peaks, both in summer and in winter.

On the other hand, Chapter 2 described an HRU with agricultural influence as well. These kinds of HRUs can be identified in many catchments, especially in meso- and large-scale catchments. In large parts of western Europe the land use consists of agricultural areas, either being pasture or crops (CORINE Land use map, European Environment Agency, 2006). In addition, many studies have shown that Hortonian overland flow is likely to occur on agricultural fields (e.g., Fiener et al., 2011; Boardman and Vandaele, 2016; Saffarpour et al., 2016; Ritsema et al., 1996). Further, agricultural fields are expected to behave differently under snow and frost conditions due to limited isolation of the vegetation.

The effect of agriculture has been extensively studied and modelled on plot scale or in small catchments (e.g., Moussa et al., 2002; Ferreira et al., 2015; Peñuela et al., 2016). However, the upscaling of Hortonian overland flow mechanisms can be especially difficult, due to the heterogeneity of the events and reinfiltration processes (e.g, Cerdan et al., 2004; Vannametee et al., 2013). Some conceptual models started with including agricultural processes on larger scales (e.g., Brauer et al., 2014), but it is not yet a common process in meso-scale conceptual models. For freezing of the topsoil something ar is the case: this is investigated in detail in alpine or permafrost areas (e.g., Cherkauer and Lettenmaier, 1999; Li et al., 2010; Koren et al., 2014), but hardly in a simple form in a conceptual model in temperate climates.

Thus, this chapter investigates the effect of representation of more detailed processes like Hortonian overland flow, snow and frost. The experiment aims to construct simple conceptualisations that can be applied in a mesoscale catchment and are relevant for the temperate climate in the Meuse basin. Following it is investigated what the effect is of adding these event scale processes and how the effects can be evaluated.

6.2. Model experiment

6.2.1. Input data

For this experiment the Ourthe Orientale was used, as more detailed process representations are more easily to test in smaller basins, because the influence of mixing of large tributaries is smaller and forcing heterogeneity is generally smaller. However, it is assumed that the relevant model scale for the Ourthe and Ourthe Orientale is similar. The three precipitation gauges in or closest to the Ourthe Orientale were used (Rachamps-Noville, Tailles and Vielsalm, Fig. 2.2). Meteorological data was used from Humain, Spa and st. Hubert; elevation zones were used to distribute the meteorological data over the three calculation cells (i.e., Thiessen polygons around precipitation gauges). For the calculation of the potential evaporation a distinction was made between snow and

no snow conditions by changing the latent heat of vaporisation (λ) and the albedo (α). $\lambda = 2.45 * 10^6$ Jkg^{-1} and $\alpha = 0.25$ for no snow conditions and $\lambda = 2.83 * 10^6$ Jkg^{-1} and $\alpha = 0.5$ for snow conditions. The albedo values were derived from the MODIS MOD10A1 product (Hall et al., 2006).

The periods 1 July 2006 till 1 July 2007 and 1 July 2002 till 1 July 2003 were used as model spin-up periods, followed by a calibration period from 1 July 2007 till 1 July 2010 and a validation period from 1 July 2003 till 1 July 2007. Following the results from Chapter 5, only the groundwater reservoir was lumped for this experiment.

The results for the Ourthe Orientale are transferred to the entire Ourthe catchment as well. For the Ourthe eleven precipitation gauges were used (dark dots in Fig. 2.2) and in addition the meteorological station at Bierset was used as well. The model was only run for the Ourthe for the validation period.

6.2.2. MODEL CONFIGURATIONS
The influence of snow and agriculture on discharge was tested with a series of model structures with different process representation. Each model structure consists of two or three HRUs; five different process conceptualisations, applicable to one ore two HRUs, are combined into six different model structures (Tab. 6.1). The stepwise approach of adding process conceptualisations gives the opportunity to evaluate the influences of each conceptualisation separately. Figure 6.1 gives an overview of the different conceptualisations.

Saturation overland flow (SOF) and subsurface storm flow (SSF) are conceptualised similar as in Chapter 5 and therefore not discussed in detail in this chapter again.

HORTONIAN OVERLAND FLOW (HOF)
Hortonian overland flow, also called infiltration excess overland flow, is conceptualised with two reservoirs: a ponding reservoir and a runoff generating reservoir. When the ponding reservoir is filled up, runoff is being generated. The runoff generating reservoir is included to slightly smooth the precipitation signal; due to the size of the catchment, even very quick runoff does not create discharge immediately. The response time of the runoff generation reservoir is very short, to create a quick response to precipitation. The HOF conceptualisation is located between the interception reservoir and the root zone of the subsurface storm flow conceptualisation.

Table 6.1: Overview of different model conceptualisations used in this model experiment.

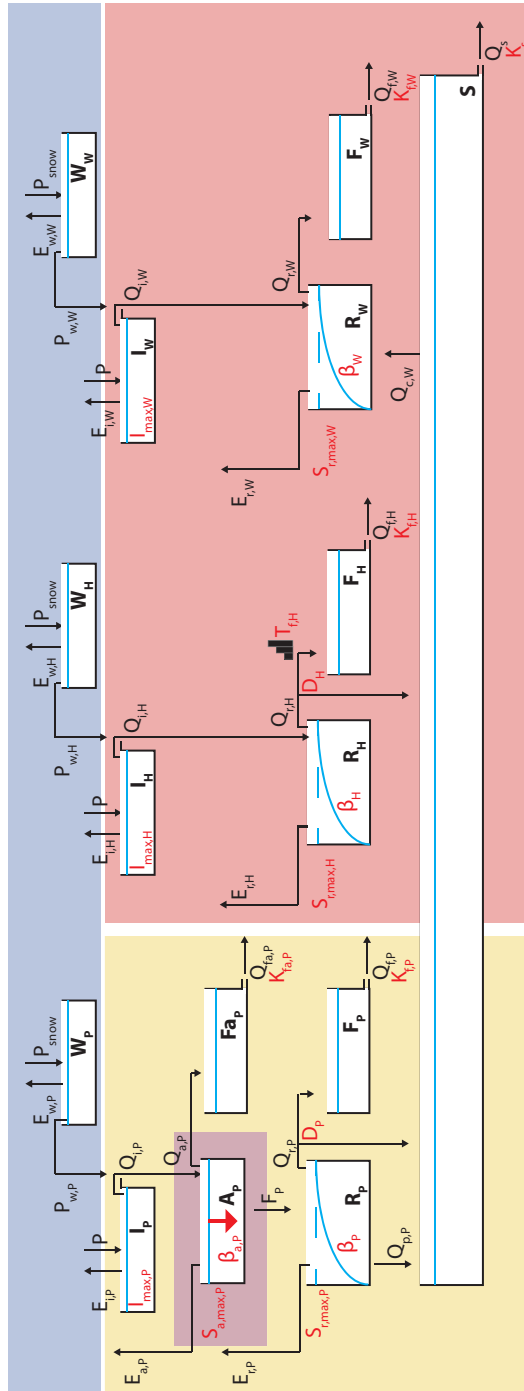| Name	HRUs	SOF	SSF	HOF	Snow	Frost
		Process conceptualisations				
WH	wetland, hillslope	x	x			
WHS	wetland, hillslope	x	x		x	
WHP	wetland, hillslope, plateau	x	x	x		
WHPS	wetland, hillslope, plateau	x	x	x	x	
WHPF	wetland, hillslope, plateau	x	x	x		x
WHPSF	wetland, hillslope, plateau	x	x	x	x	x

Figure 6.1: Model structure with all process conceptualisations. Colours indicate the different process conceptualisations, where wetland and hillslope are both coloured red to indicate that this model configuration from Ch. 5 serves as a benchmark for this model experiment.
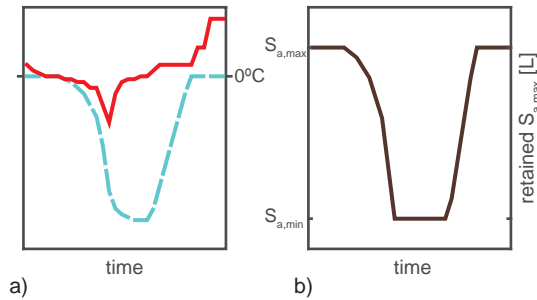
Figure 6.2: Schematic representation of the influence of frost in the topsoil on Hortonian storage; a) air temperature (red) and modelled frost accumulation (blue dashed); b) retained $S_{a,max}$ (mm).

SNOW

Snow is conceptualised like most degree day models: precipitation is stored as snow when the air temperature is below a certain threshold. Once the temperature rises above another threshold, the accumulated snow will melt with a specified melting rate. In the studied areas snow generally occurs during short periods (in the order of days to weeks) and the end of a snow period often coincides with a rain event. Therefore, the melt co-efficient was said to increase with precipitation. The same snow module is used for all HRUs.

FROZEN SOIL (FROST)

During winter periods the temperature can be below or around zero for a number of weeks, causing the topsoil to freeze. This is especially likely to happen on the plateaus, where there is a limited insulating vegetation layer and absence of deep preferential cracks. It is assumed that if the topsoil is (partly) frozen, Hortonian overland flow occurs quicker. This is conceptualised by decreasing the storage capacity of the agricultural reservoir. A smaller storage capacity will already cause overland flow for precipitation or snow melt events with lower intensities. The storage capacity is linearly decreased with a modelled frost accumulation, which is the sum of the hourly temperatures with a maximum of zero (Fig. 6.2).

6.2.3. MODEL CONDITIONING AND EVALUATION

Model conditioning and evaluation are again carried out with the methods and tools described in Chapter 3. Some of these methods and tools need some further specification, which can be found below.

NARROWING PARAMETER SPACE

The parameter space for this experiment was narrowed stepwise. First, $S_{r,max}$ was derived from climate data, as described in Section 4.6. Second, the space of the relevant parameters for WH was narrowed based on a visual inspection of the modelled and observed hydrographs. Following, the space of the parameters for the HOF, snow and frost modules was narrowed on an event basis.

Table 6.2: Parameter and process constraints for all model configurations.

Parameter constraints	Process constraints
$I_{max;W} < I_{max;P} < I_{max;H}$	F_{RC}[a] < 0.55[b]
$S_{r,max;W} < S_{r,max;H} = S_{r,max;P}$	$F_{RC} > 0.4$[b]
$\beta_W = \beta_{a;P} < \beta_H = \beta_P$	$F_{RC,low} < 0.27$[c]
$L_W < L_P < L_H$	$F_{RC,low} > 0.12$[c]
$K_{f;W} = K_{fa;P} < K_{f;H} = K_{f;P}$	$F_{RC,high} < 0.84$[c]
$D_H < D_P$	$F_{RC,high} > 0.69$[c]
$K_{p;H} < K_{p;P}$	$E_{tot} < 616$ mmy^{-1}
$K_s = 2500$ (h)	$E_{tot} > 504$ mmy^{-1}
	$F_{RC;f,W} > F_{RC;f,H}$
	$F_{RC;fa,P} > F_{RC;f,P}$[d]
	$\sum(Q_{f,H;high} * H_\%) > \sum(Q_{f,W;high} * W_\%)$[e]
	$\sum(Q_{f,P;high} * P_\%) > \sum(Q_{f,W;high} * W_\%)$[d]
	$Q_{low} < 0.28$ mmh^{-1}

[a] F_{RC} = modelled runoff coefficient.   [b] Applied for the entire modelled period.
[c] Applied for the low and high flow period (selected from calibration period).
[d] Only for configurations with plateau included.   [e] Only for configurations with plateau not included.

For HOF the summer of 2008 was used, as this summer had many high intensity precipitation events. For snow and frost all three winters in the calibration period were used: each period was split into events with solid precipitation, liquid precipitation and snow melt based on discharge, precipitation and temperature. In addition, for each event it was determined if it was likely or not that the topsoil was frozen. This information was then used together with snow cover data from the MODIS MOD10A1 data set (Hall et al., 2006) to narrow the parameter space for the snow and frost module.

For the frost module it turned out that especially the amount of frost accumulation and the melting rate were important for simulating the discharge response. For the snow module it turned out that a larger melt coefficient in case of liquid precipitation indeed helped in simulating the occurrence of snow cover in the MODIS data.

USED CONSTRAINTS

The eight parameter and thirteen process constraints used for this experiment are summarized in Table 6.2. All constraints were selected from those presented in Table 3.1. The constraints regarding the maximum runoff coefficient over the total period and over the dry period and the maximum yearly evaporation turned out to have the largest filtering capacity.

USED SIGNATURES

For this experiment, again a set of signatures was selected (Tabs. 3.3 and 6.3), each of them focussing on winter (snow) events or summer (high intensity) events. Root mean square error (RMSE) and the Relative Error (E_{RE}; Eq. 3.2) were used to compare the modelled and observed signatures; different evaluation metrics were selected for different

Table 6.3: Overview of used signatures for model evaluation and period of flow to which they are applied.

Signature	Metric[a]	Period
event runoff coefficient	E_{RMSE}	total, high[b]
mean high intensity event runoff coefficient[d]	E_{RE}	total
peak distribution	E_{RMSE}	low[c]
coefficient of autocorrelation (1h < lag < 250h)	E_{RMSE}	low[c]
declining limb density	E_{RE}	high[b]
slope declining limb for low temperatures[e]	E_{RMSE}	total

[a] The observed and measured signatures are compared with two different metrics: root mean square error (E_{RMSE}) and relative error (E_{RE}).   [b] December till March was used as high flow period. [c] June till August was used as low flow period.   [d] High intensity precipitation events mainly occur during the low flow period.   [e] Events with at least 10% of precipitation during temperatures below 1°C; these mainly occur during the high flow period.

signatures. Precipitation events were defined with the method described by McMillan et al. (2011), with values for x_{inter}, i_{storm}, x_{storm}, i_{inter} being 0.03 mmh^{-1}, 0.2 mmh^{-1}, 24 h, 30 h and 1 mmh^{-1}, 2 mmh^{-1}, 5 h, 8 h for 'general' events and for high intensity events respectively. Discharge events started at the same time as precipitation events and ended 6 days (general events) and 1 day (high intensity events) after the precipitation event or when the next precipitation event started.

DETAILS FOR FARM ANALYSIS

In addition to the selected signatures, two other specifications are required for the FARM analysis: definition of the categories for the performance and for the range in performance. For the performance three categories were defined: high (solid lines): $E_{c,i,10} > 0.9$; middle (dashed lines): $0.9 < E_{c,i,10} < 0.8$; low (dotted lines): $E_{c,i,10} < 0.8$. Three categories were defined for the range (difference between 10th and 90th percentile) in performance as well: small (green): $E_{c,i,10} - E_{c,i,90} < 0.1$; middle (orange): $0.1 < E_{c,i,10} - E_{c,i,90} < 0.2$; large (red): $E_{c,i,10} - E_{c,i,90} > 0.2$).

6.3. RESULTS

All six model configurations were run for the calibration and validation period for the same 2000 parameter sets. The model results from the calibration period were filtered with the process constraints, resulting in approximately 300 behavioural sets for model configurations with plateaus included and more than 1000 for those without plateaus included. For an equal comparison 300 model realisations were randomly drawn from the behavioural ones for WH and WHS. The posterior parameter ranges are presented in Tables B.2 till B.4. When in the analysis single model realisations are used, the same parameter set is selected for all model configurations.

Comparing the model structures only with NSE of the discharge and NSE of the log of the discharge shows that overall the model structures perform equally well. Scores range

between 0.75 and 0.9 for all model realisations for all model structures.

6.3.1. Internal model behaviour

Contribution of different fluxes

Figure 6.3 shows the contribution of different runoff components to the discharge for three model configurations for a summer and a winter period. For the summer period (left) it can be seen that the fast contribution of the hillslope has decreased between WH and WHP, especially because the overland flow from the plateaus (HOF) makes up the main part of the sharp peaks. It can also be seen that the contribution of the wetland is limited in summer, probably because this area dries out as well.

For the winter period it can be seen that the differences between WH and WHP are relatively small: the fast response for WH consists of only hillslope and for WHP it consists of a fast contribution from both hillslope and plateau. The overland flow on the plateaus is negligible. However, looking at WHPSF the influence of the frozen soil can be seen. The moments when the modelled accumulated frost deviated from zero (Fig. 6.4), there is a clear contribution of the overland flow from the plateaus. Except for these event, the sharper peaks in winter have a larger contribution from the wetlands. Thus, higher peaks in summer and during frost conditions seems to be generated by Hortonian overland flow mechanisms, while under other winter conditions they seem to be generated by saturation overland flow mechanisms.

Evaluation of states

Figure 6.4 shows two panels with modelled and observed variables other than discharge regarding the snow and frost processes for the winter period presented in Figure 6.3. Figure 6.4a shows the modelled snow water equivalent (SWE) together with the snow cover extent from the MOD10A1 product. Because the MOD10A1 snow data has a lot of missing values, it is mainly used in a binary way: snow or no snow. The figure shows consistency for the period in March, especially the end of the snow cover period seems to coincide very well. For the period at the end of January the MOD10A1 data gives a significant snow cover, while there is no modelled SWE. This is probably caused by the undercatch of snow, especially because the observed air temperatures were low in that period. It could also be that a dense cloud cover was identified as snow cover.

Figure 6.4b shows the modelled frost accumulation (F_{acc}) together with the daily minimum soil surface temperature (T_s). The frost accumulation was modelled based on air temperature, so comparing it with T_s indicates whether the modelled values are reasonable. The figures shows a quick transition in T_s at the end of March: temperatures rise clearly above zero. At the same time F_{acc} has fully melted. For the preceding period it can be seen that T_s drops a bit quicker than F_{acc}. Overall, it can be seen that there is a reasonable match between T_s and F_{acc}, and thus that air temperature can be used to estimate F_{acc}.

6.3.2. Model performance and consistency

Hydrograph comparison

Figure 6.5 shows the observed and modelled hydrographs for most of the model configurations for the same periods as in Figure 6.3. The presented model configurations
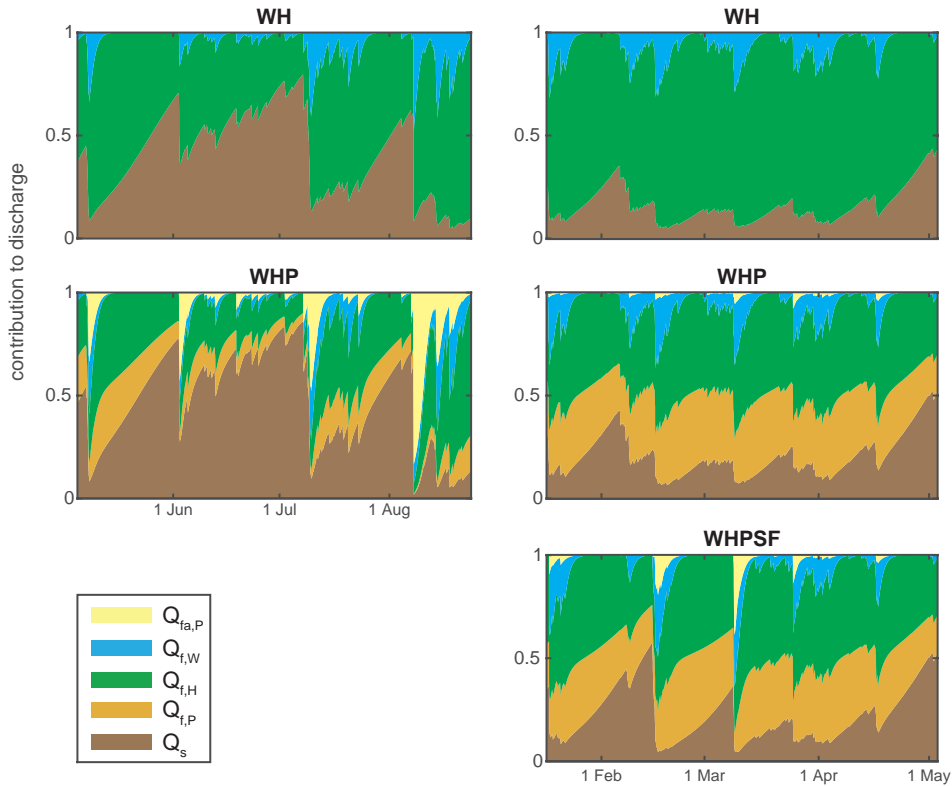
Figure 6.3: Contributions to the discharge for different model configuration for the summer period 5 May to 24 August 2004 (left) and the winter period 17 January to 3 May 2006 (right).

are among the best performing ones. Figure 6.5a shows a summer period: in the hydrographs the effect of HOF can be clearly seen as well, both for the first peak, as for the one in early August. The peak in August also shows the sensitivity of the HOF conceptualisation to input data: the event was probably very local and resulted in a too high modelled peak. On the other hand, it can be seen that the water has left the system very quickly as well, which is not the case for the WH configuration. Therefore, the HOF seems a reasonable process to incorporate.

Figures 6.5b and 6.5c show a winter period: Figure 6.5b shows the positive effect of incorporating a snow module (light blue line), both for the highest peak half of March as for the period following the highest peak. The effect of HOF without a frozen topsoil (cf. WH and WHP) is very small for a winter period, consequently the difference between WHS and WHPS is relatively small as well. Figure 6.5c shows the additional effect of a frozen topsoil: this effect is smaller than that of a snow module, but still it increases the peak height. All model configuration are not able to model the correct peak volume, which could be explained by the possible discrepancy in Figure 6.4a between modelled and observed snow cover.
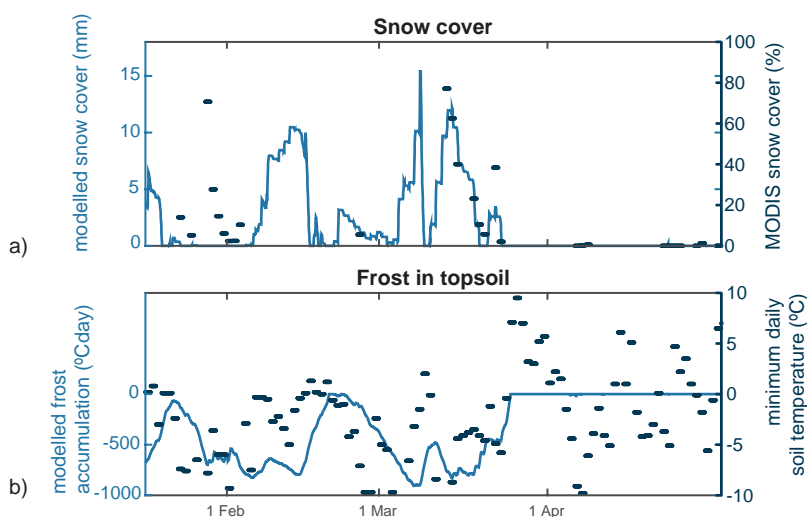
Figure 6.4: Evaluation of snow and frost processes for the calculation cell belonging to Tailles for the same winter period as in Fig. 6.3; a) modelled snow cover (line) and MOD10A1 snow cover fraction (dots) ; b) modelled frost accumulation (line) and minimum daily soil temperature (dots).

6

FARM

Figure 6.6 shows the plots of the FARM analysis for WH, WHP and WHPSF for the calibration and validation period. In general it can be seen that WHPSF has more solid and green vectors than the other two, indicating a higher performance and consistency. However, the individual signatures reveal more about the model behaviour. For example for the signatures focussing on the summer low flow period (3 and 5): when including HOF, 5 ($F_{ACfunc,low}$) is no longer directly correlated with 3 (F_{RCintH}). For the calibration period 5 is inversely correlated, for the validation period uncorrelated. Figures 6.3 and 6.5 indicate the usefulness of a HOF conceptualisation. Combining the analyses indicates that the HOF conceptualisation is useful, but might need some improvement.

The signatures focussing on the winter high flow period (6 and 7, $F_{DLD,high}$ and F_{SDL}) show the effect of the snow module: when the snow module is incorporated (WHPSF), 6 and 7 are directly correlated, while without the snow module (WH and WHP), 6 and 7 are inversely correlated, both for the calibration and validation period, meaning they cannot be reproduced well with the same parameter set. The other configurations with a snow module are not shown here, but have 6 and 7 directly correlated for the calibration and validation period as well.

6.3.3. PREDICTIVE POWER

SCATTERPLOT

Figure 6.7 shows the scatterplots for all model configurations for two signatures: one focussing on the high intensity summer events ($F_{RC,intH}$), the other on the winter events ($F_{DLD,high}$). The grey dots show the scores for the calibration period, the black dots for
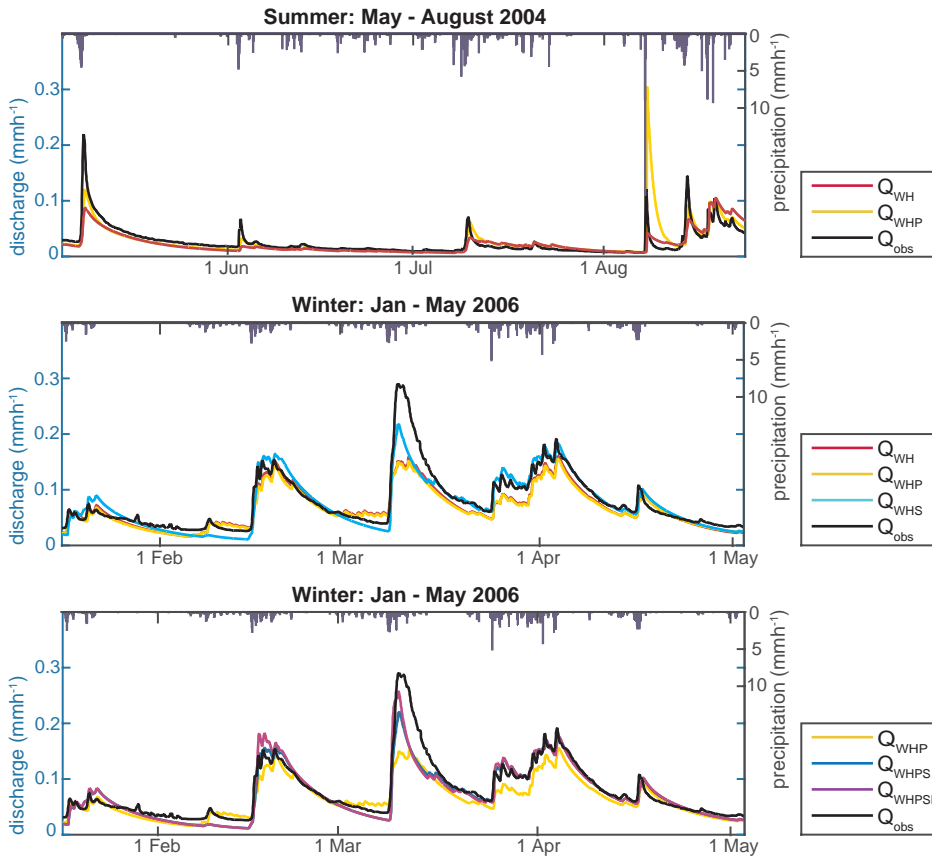
Figure 6.5: Modelled and observed hydrograph for the same summer and winter period as in Fig. 6.3.

the validation period, with zero being the best performance. Like in Figure 6.6, it can be seen that the snow module is essential for reproducing $F_{\mathrm{DLD,high}}$: not only is the range in performance larger for the calibration period for the model configurations without snow (top row), the performances in the validation period are also lower than in the calibration period. This is not the case for the model configurations with snow (bottom row). Thus, with respect to $F_{\mathrm{DLD,high}}$, the model configurations with a snow model are more robust for transferring between different periods.

The performance of $F_{\mathrm{RC,intH}}$ can be linked to the presence of a HOF conceptualisation on the plateaus. For the model configurations without HOF (left column), some of the performances deteriorate in the validation period, while this is hardly the case for the model configurations with HOF (middle and right column). This finding supports again that including a HOF conceptualisation can be beneficial to the model performance; however, the conceptualisation might need some improvements as discussed in Section 6.3.2.

1: event runoff coefficient
2: event runoff coefficient (high)
3: high intensity event runoff coefficent
4: slope peak distribution (low)

5: coeff. autocorrelation, lag: 0-250h (low)
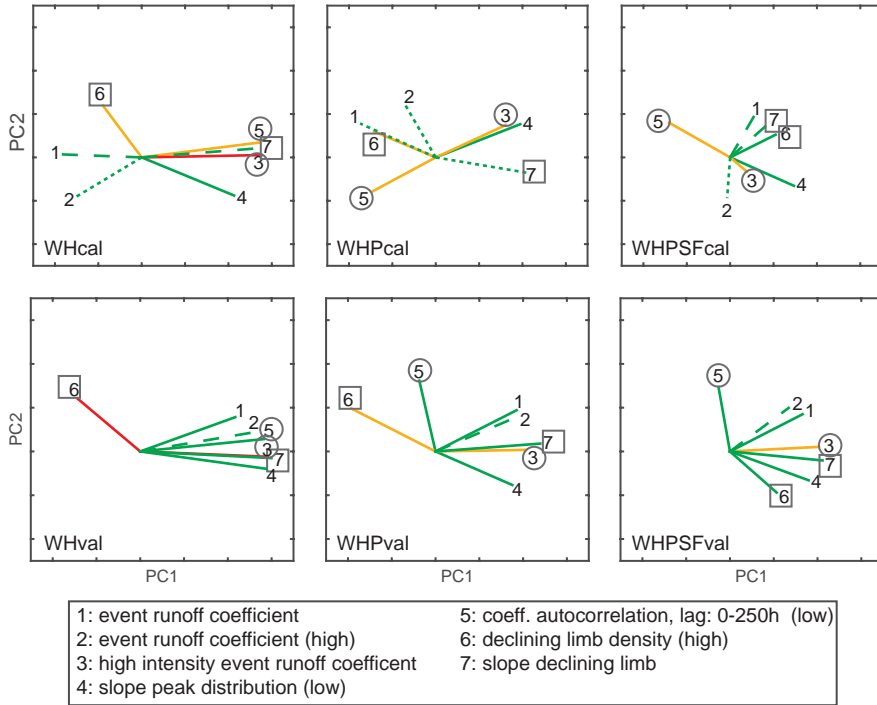6: declining limb density (high)
7: slope declining limb

Figure 6.6: Plots for the FARM analysis for WH, WHP and WHPSF for the calibration (top) and validation (bottom) period. The grey circle indicates the signatures focussing on the summer low flow period, the grey squares indicates the signatures focussing on the winter high flow period. The line style of the vector indicates the 10th-percentile for the evaluation criterion: a solid vector indicates the best performance, followed by a dashed and a dotted vector. The line colour indicates the difference in performance between the 10th and 90th percentile: green indicates a small, orange a middle and red a large range.

PERFORMANCE AND CONSISTENCY

Figure 6.8 shows the overall performance and consistency for the Ourthe Orientale (a) and a comparison between the Ourthe Orientale and Ourthe (b). For the Ourthe Orientale the largest difference can be observed between the calibration and validation period: the latter has a higher performance and consistency. This is probably caused by the easier to model meteorological conditions in the validation period. Further a distinction is visible for the calibration between the model structures without and with a snow module: the latter has a higher performance and consistency. Apart from these clear differences, the differences between the model structures are small, again highlighting that it is difficult to evaluate the effect of detailed processes which mainly have an effect at the event time scale.

Figure 6.8b shows the performance and consistency for only two model structures: WH and WHPSF. For each model structure two options for the Ourthe are presented: one with the same S_r-values as for the Ourthe Orientale (black symbols) and one with
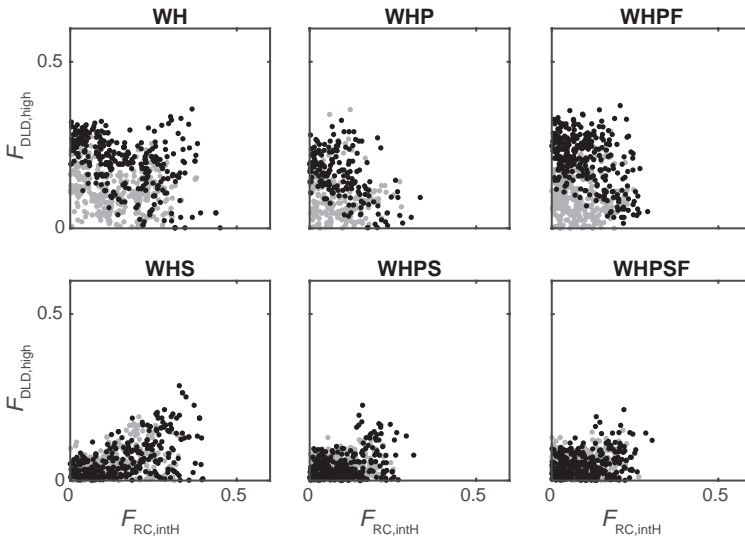
Figure 6.7: Scatterplots for all model structures for the high intensity event runoff coefficient ($F_{RC,intH}$) and the declining limb density for the high flow period ($F_{DLD,high}$) (grey dots = calibration period, black dots = validation period).

the S_r-value derived for the Ourthe (see also Fig. 4.10) (blue symbols). It can be seen that WHPSF is better transferable than WH. Further, it turns out that changing the S_r-value is very useful: the effect is even larger than that of applying a different model structure in the Ourthe Orientale.

6.4. DISCUSSION

Section 6.3 described the model results for six model configurations. In general, the results show that including HOF, snow and frost can improve the reproduction of certain events, without deteriorating other events. However, some aspects might influence the presented results and are thus discussed below.

6.4.1. DATA QUALITY

Data quality is a very often returning topic among hydrological modellers: the model results can only be as good as the input data. In other words: if you feed rubbish to your model, rubbish will come out, no matter how good the model is. However, I find it too easy to give data all the blame: even with imperfect data it is possible to compare models. A good example can be seen in Figure 6.5a for the very local precipitation event in August. WHP gives a too high response, but gets rid of the water very quickly; WH on the other hand accumulates the additional precipitation, overestimating the saturation of the catchment.

The same can be observed for the undercatch of solid precipitation (e.g., Figs. 6.4 and 6.5). Although the model configurations with a snow module could not reproduce the
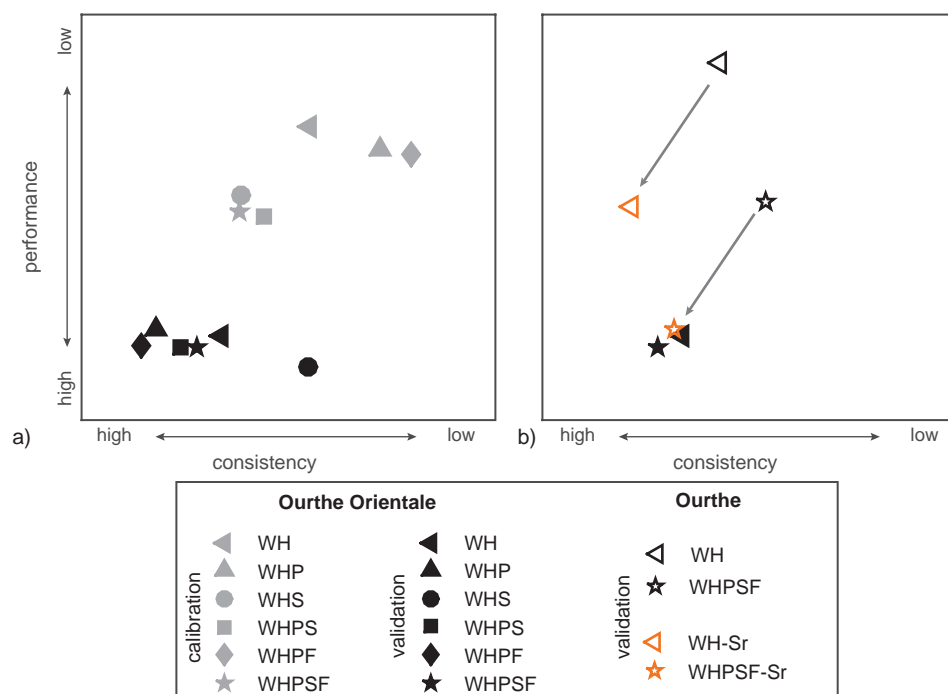
Figure 6.8: Performance and consistency of all model configurations; a) for the Ourthe Orientale; b) for the Ourthe and Ourthe Orientale; the arrows indicate the change in performance and consistency when changing the value for S_r

entire peak volume, they showed clearly better performances than those without snow. Thus, even with data errors, better or worse conceptualisations can be discriminated. But one has to keep in mind that models with more detailed process representation, can be more sensitive to data errors.

6.4.2. Signature construction

Hydrological signatures are the basis for half of the evaluation tools used in this chapter. The selection of signatures is a subjective process, which thus influences the results. However, if signatures are selected which are specifically relevant for the model adaptations under consideration, they can make model evaluation more meaningful.

It should be noted though, that it can be very difficult to fully capture the characteristic of the hydrograph under consideration in a signature. For example, due to the combination of snow melt and frozen soil, it is expected that the rising limb of the hydrograph becomes steeper and more similar to the observed slope. However, identifying this specific part of the hydrograph can hardly be automated, making it unsuitable as hydrological signature.

6.4.3. PROCESS CONCEPTUALISATIONS

In this experiment three process conceptualisations were added: HOF, snow and frost. For the combination of snow and frost one possible option is evaluated. During the calibration processes small differences have been experimented with, but these differences were not systematically tested. The result that this combination has a positive effect on the model results does not mean that it is reality or that is the only combination leading to a positive effect. On the other hand, incorporating a detailed, data intensive snow model might not be feasible if the influence of snow is only dominant during one or two weeks per year. A conceptualisation with a positive effect during multiple events and no significant negative effects, might under these conditions be reasonable to incorporate.

Something similar can be argued for the conceptualisation of the Hortonian overland flow. The results clearly show that a process with a short time scale and depending on precipitation intensities has a positive effect on the model results. On the other hand, currently the response time of the Hortonian runoff reservoir is in the order of 0.5-2 days, which might be too long for a fast process like Hortonian overland flow. In addition, by including the agricultural ponding reservoir, the infiltration capacity of the soil is indirectly incorporated. Adding more physical meaning to the HOF conceptualisation, can probably further improve the model results. For example by removing the runoff reservoir and making the overland flow directly related to estimated infiltration capacities.

6.4.4. MODELLING CATCHMENT SATURATION

One of the problems with the WH configuration from Chapter 5 was the limited simulation of the saturation of the catchment in autumn. It was expected that adding processes in this experiment would help to better simulate catchment saturation. However, this is not really the case: especially after very dry periods (i.e., transpiration being moisture constraint instead of energy constraint), the runoff response of the model is much higher than that of the catchment, which can probably be explained by the root zone storage filling up too quickly.

Up till now I can think of mainly two processes which would prevent the root zone storage filling up very quickly. First, additional preferential flow paths to either deeper aquifers or neighbouring catchments are activated under high infiltration rates (i.e, more water being stored in the root zone storage). However I am not sure where and when this water contributes to the discharge again and why this additional preferential flow would mainly increase under high infiltration rates. Second, following Chapter 4, vegetation could temporarily increase the storage capacity under drier conditions: with a larger root zone storage capacity, the root zone storage fills up slower. However, for this option to have effect, the vegetation seems to create a large amount of temporal additional storage capacity. NDII data as a proxy for water stress in vegetation (Sriwongsitanon et al., 2016) could help to identify possible water stress during moisture constraint evaporation and thus the plausibility whether vegetation creates an additional buffer.

6.5. CONCLUSIONS

The model experiment in this chapter evaluated different process conceptualisations to include the effect of snow and agriculture in the FLEX-Topo configuration from Chapter

5. The results show that including more detailed processes and process conceptualisations regarding Hortonian overland, snow and frost of the topsoil can have large effect on an event basis. When integrating the effects over the entire validation or calibration period, the effect becomes much smaller. Although the adaptations have a clear effect at the event scale, the additional effect of each added process is smaller and more difficult to evaluate. Thus, it seems that with these additional process conceptualisations we are on the edge of gaining significant effects.

6

7

RESULTS FROM OURTHE USED IN MEUSE BASIN

We especially need imagination in science. It is not all mathematics, nor all logic, but it is somewhat beauty and poetry.

Maria Mitchell (Maria Mitchell, life, letters, and journals)

The previous chapters focussed on comparing model structures for the Ourthe or the Ourthe Orientale and not on the effect of transferring the derived model structures to neighbouring catchments. In this chapter three explorative comparisons are described which aim to put the previous results into a wider perspective. FLEX-Topo is compared between the Ourthe and other catchments, it is compared with other models, and it is compared with the operational HBV configuration for periods that were not part of the calibration. In general FLEX-Topo gives good results compared to other models, including the operational HBV configuration.

7.1. Introduction

The previous chapters focussed on comparing model structures for the Ourthe and the Ourthe Orientale. This chapter aims to explore the implications and possibilities when generalizing the results obtained earlier. Both with respect to neighbouring catchments as well as in comparison with other models. This chapter is not meant to give a full evaluation of FLEX-Topo in different catchments and compared to other models. It is rather a first step in investigating the use of FLEX-Topo for wider applications.

A comparison between a large set of models and a large set of catchments is difficult to structure. Therefore, the comparison in this chapter is split into three parts: FLEX-Topo in neighbouring catchments, FLEX-Topo and other models and FLEX-Topo and HBV. The last comparison is made because the discharge predictions at the Dutch border are made with the HBV model at the moment of writing. The exact set-up slightly differs between the three comparisons, partly because not all three originate from the same experiment and partly to simplify the comparison. Thus, for each comparison the set-up is described before the results are discussed.

The forcing used for all comparisons is based on a gridded data set with a spatial resolution of 0.0091°. For the precipitation the grid was overlaid with a Thiessen network around the 28 precipitation stations in Figure 2.2. Daily minimum and maximum temperatures from the E-OBS open dataset (0.25° x 0.25° resolution; Haylock et al., 2008) were disaggregated to hourly values using radiation data at Maastricht (KNMI[1]) and Equation 2.1. Daily potential evaporation was derived with the Hargreaves formula (Hargreaves and Samani, 1985) and disaggregated to hourly values using a sine function during the day and no evaporation at night.

7.2. FLEX-Topo in neighbouring catchments

7.2.1. Comparison set-up

For the evaluation of FLEX-Topo in neighbouring catchments the WHPSF configuration (Tab. 6.1) was used. However, the configuration was used in a gridded set-up, instead of using the Thiessen polygons as calculation cells like in Chapters 5 and 6. A gridded set-up was used because the operational forecast system works with a gridded set-up and because the preparation of the forcing data was already done for the intercomparison study (Sect. 7.3). The consequence of the gridded set-up is that the catchment outline and relative contributions of each rain gauge slightly changed compared to the set-up with the calculation cells (Fig. 7.1).

The calibrated model for the Ourthe Orientale from Chapter 6 was transferred to the Ourthe Occidentale, Ourthe, Lesse and Semois. The root zone storage capacity (Fig. 4.10) and percentages for each HRU were changed between the catchments: the remaining parameters were kept constant. The models were run for the period which was used for validation in Chapter 6: 1 July 2003 to 1 July 2007, with 1 July 2002 to 1 July 2003 as spin up period.

---

[1]http://www.knmi.nl/nederland-nu/klimatologie/uurgegevens, visited 14-12-2012
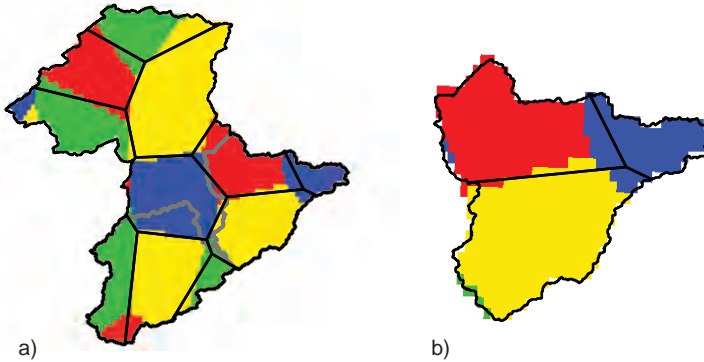
Figure 7.1: Differences in catchment outline and contributions of individual Thiessen polygons between the set-up with calculation cells used in Chs. 5 and 6 (black lines) and the gridded set-up used in this chapter (coloured polygons). Note: the colours of the polygons are only meant to indicate their outline, they do not have any other meaning. a) Ourthe, b) Ourthe Orientale.

7.2.2. RESULTS AND DISCUSSION

Figures 7.2 and 7.3 show the results for FLEX-Topo for the five catchments for the same periods as in Figure 6.5. When comparing these figures with Figure 6.5, it can be seen that the results for the Ourthe Orientale are slightly different. This is caused by the differences in catchment outline and relative contributions of the rain gauges. The differences in the winter period are also partly caused by the different temperature forcing, this can especially influence the onset of snow melt.

When comparing the catchments, it can be seen that the discharge (and precipitation) in the Semois are much higher than in the other four catchments. In addition, the discharge pattern in the Semois is different from that of the other catchments, especially in summer. Further, it can be seen that the model performs best for the Ourthe and the Lesse, whereas it was not specifically calibrated for these catchments.

Looking in detail to the summer period (Fig. 7.2), it can be seen that the shape of the first peak of the season is captured for all catchments. Only for the Ourthe Occidentale it is too low, probably because the groundwater contribution in this catchment is too low. Further it can be seen that there was a very local precipitation event in the Ourthe Oriental, causing a large peak in discharge. This peak propagates into the Ourthe, but is much smaller in the Ourthe Occidentale and the Lesse and does not occur in the Semois. Finally, it can be seen for the Semois that during the recession the model responds quicker to precipitation than in reality. A reason for this could be that the Hortonian overland flow module is calibrated in the Ourthe Orientale, which has mainly the same soils as the Ourthe Occidentale, Ourthe and Lesse. The Semois on the other hand has soils which are slightly more permeable (de Wit, 2008) in the upstream part of the catchment; this might require a recalibration of the Hortonian overland flow module.

Looking in detail to the winter period (Fig. 7.3), it can be seen that the Ourthe Orientale and Ourthe Occidentale overestimate the first peaks and underestimate the declining limb of the highest peak. However, for the entire Ourthe these effects are much

smaller; it is likely that this is an effect of the averaging within in the catchment: on a larger scale the effectof local precipitation or snow melt events is smaller. In addition, it can be seen that for the Semois the model responds too quickly again in February and that snow melts too quickly at the end of March. The former could again be linked to the calibration of the Hortonian overland flow, the latter could be linked to the calibration of the snow module. The Semois has more north and south facing slopes than the other catchments, which might lead to a different melt temperature or melt coefficient.

7.3. FLEX-Topo and other models

7.3.1. Comparison set-up

The comparison of FLEX-Topo with other models originates from an international inter-comparison study for the Meuse basin. Eight international research groups participated in this model intercomparison study using one or several rainfall-runoff models. A total of eleven models were used, the FLEX-Topo model, six other independent models and four models from the SUPERFLEX framework (Fenicia et al., 2011). A modelling protocol was prescribed to enable a comparison of the results. The protocol for the modelling involved a split-sample calibration (Jan 2004 till Dec 2007) and validation (Jan 2001 till Dec 2003) using a common dataset for the Ourthe catchment and a validation in space for the nested catchments Ourthe Orientale and Ourthe Occidentale and for the neighbouring Lesse and Semois catchments (Tab. 2.1). The models were calibrated based on NSE of the discharge and of the log of the discharge. The details and results of this study are described in de Boer-Euser et al. (2016); here, a selection of the results is presented for the Ourthe to show a first benchmark of FLEX-Topo.

Each modelling group provided results as described above. A variety of models was used, including lumped, semi- and fully distributed models. All models are conceptual, but with a varying degree of complexity and are used by institutes or universities working in the Meuse basin. Figure 7.4 depicts the main fluxes and storages of the models. The term 'very quick runoff' is used for a faster process than 'fast runoff'. For FLEX-Topo the WHPS configuration (Tab. 6.1) was used instead of the WHPSF configuration, as the intercomparison study started before the WHPSF configuration was available.

7.3.2. Results and discussion

All models gave a good performance, based on general metrics. In addition the model results were compared based on statistics (cumulative discharges, empirical extreme value analyses and flow duration curves) and based on hydrographs for specific events. These additional evaluations revealed differences and similarities between the models. Two analyses showed clear differences between the models: the lowest 20% of the flow duration curve and the hydrographs of the 2008 summer. These two analyses are shown in Figures 7.5 and 7.6.

Low flows

The lowest 20% of the observed and modelled flow duration curves was used to analyse low flow behaviour. Discharges during the summer recession periods are generally low (ranging between 0.004 mm h^{-1} and 0.015 mm h^{-1} for the lowest 20%) compared to averaged flow (0.05 mm h^{-1}). The influence of a groundwater reservoir on the modelled
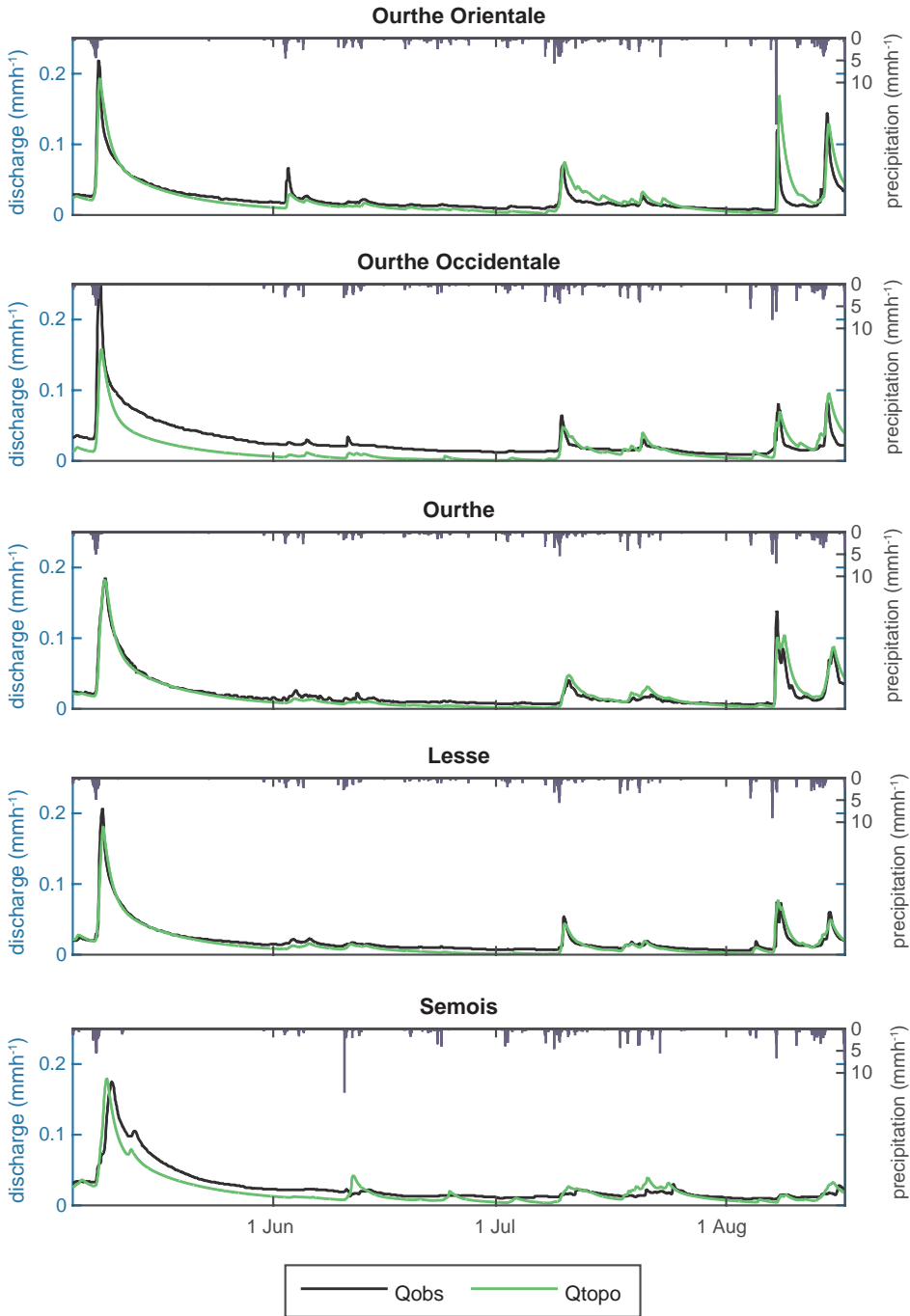
Figure 7.2: Modelled (green) and observed (black) discharge for the summer period 5 May to 24 August in 2004 for five catchments in the Meuse basin.
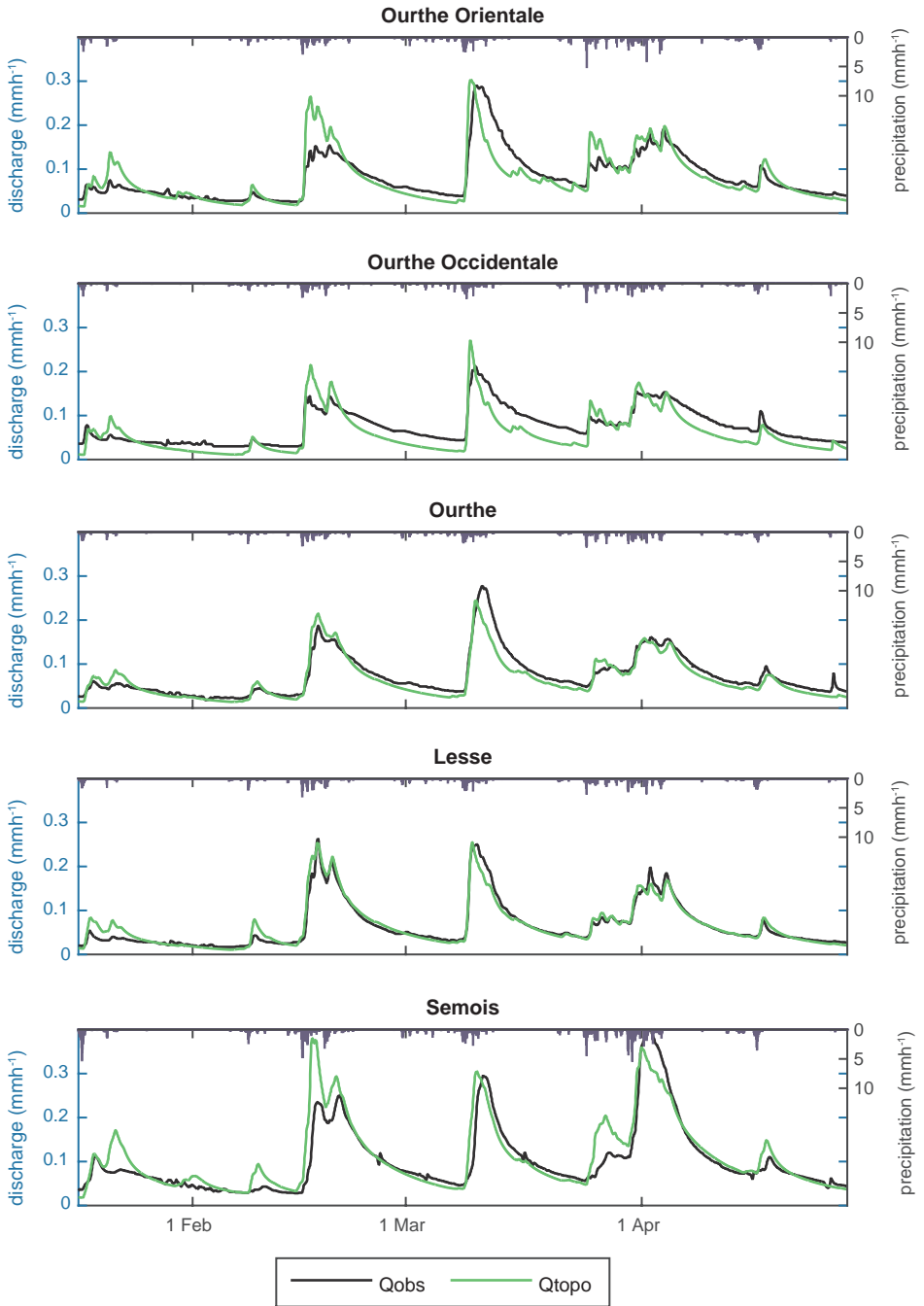
Figure 7.3: Modelled (green) and observed (black) discharge for the winter period 17 January to 3 May 2006 for five catchments in the Meuse basin.
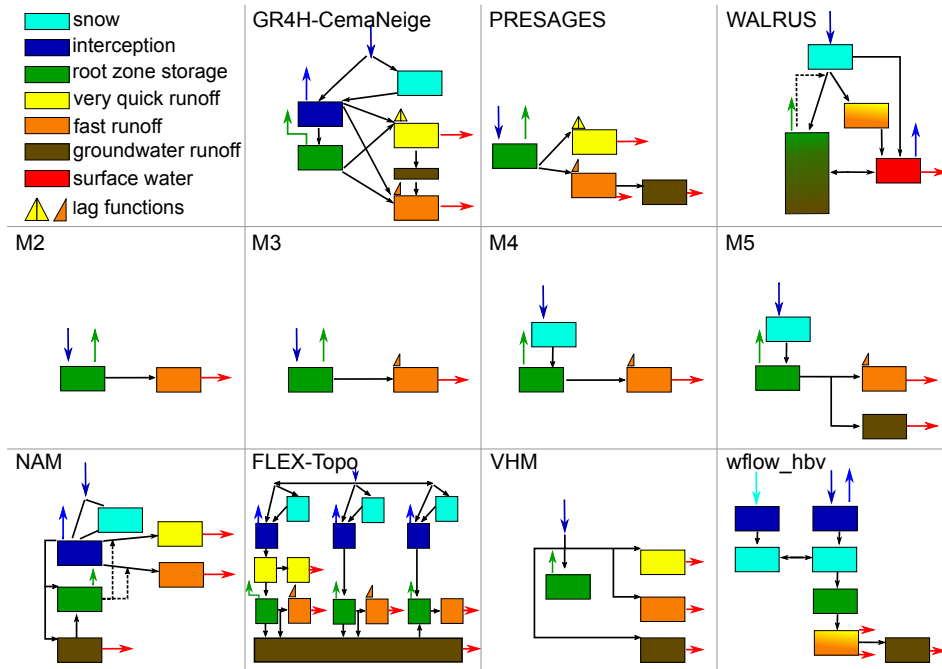
Figure 7.4: Model structures used in the intercomparison study in the subcatchments of the Meuse. Note: the model structures are simplified to only highlight main differences and similarities.

discharge is significant as the flow duration curves illustrate (Fig. 7.5). Adding a groundwater reservoir improves the simulation of the low flows: the only difference between M4 and M5 is the groundwater reservoir and where the low flows are simulated too low by M4, they are properly simulated by M5. This indicates that during the high flow period in winter water is stored in the catchment, which is released during the low flow period in summer.

In addition, a difference can also be seen for the configuration of the groundwater reservoir. M5, NAM, FLEX-Topo and VHM have their groundwater reservoir configured in parallel to the fast runoff reservoir. The other models have it configured in series or interacting with the fast runoff reservoir (GR4H, PRESAGES, WALRUS, wflow_hbv). The parallel groundwater reservoir can better simulate the low flows. On one hand this indicates the importance of preferential recharge in the catchment, on the other hand it indicates the existence of different time scales in the catchment response. With a parallel groundwater reservoir, the time scales for runoff generation are decoupled, with a serial or interactive groundwater reservoir they are connected. The different results for the low flow period indicate that these time scales are relatively independent in the studied catchments.

The performance of the models during the 2008 summer was analysed by plotting the hydrographs, as shown in Figure 7.6. Although total precipitation amounts during this summer were not higher than in other years, the precipitation intensities were. The antecedent root zone storages before the events can be expected to be low, due to high transpiration rates in summer. While most models are not able to capture the summer peaks, VHM and FLEX-Topo are able to simulate the dynamics well, although FLEX-Topo overestimates the summer peaks. As shown in Figure 7.4, VHM and FLEX-Topo are the only ones with a very quick runoff component that is not influenced by the root zone storage. Hence, it appears that this component is essential for simulating short, intense summer events. Models with a very quick runoff component which is influenced by the root zone storage (GR4H, presages, WALRUS, NAM and wflow_hbv) and models with a very quick runoff component following the root zone storage do perform better than models where the very quick runoff component is entirely lacking (M2 to M5).

A possible reason for these differences is the high intensity character of the summer precipitation events. These events are likely to cause infiltration excess overland flow, i.e., precipitation intensity being higher than infiltration capacity of the soil. Under dry conditions the infiltration capacity of the soil is assumed to be disconnected to the saturation of the soil. Thus, linking the very quick flow mechanism to the storage in the root zone causes a damping of the generated peak flows. A very quick flow component influenced by the root zone may not respond as fast as necessary to the precipitation; however, it enables to model to respond quicker than without a very quick flow component at all. This results indicate that infiltration excess runoff occurs on a event basins in the studied catchment and that it is relatively disconnected from the storage in the root zone.

These two examples show that differences between models are mainly important under drier conditions: under wetter conditions the model results are more comparable. The second example shows again that the Hortonian overland flow component which was included in FLEX-Topo in Chapter 6 is very important for high intensity summer events. The side effect that this process is very sensitive to data handling and data errors (Sect. 6.4 and 7.2) only means that this process should be included and calibrated carefully, not that it should be left out.

7.4. FLEX-Topo and HBV

7.4.1. Comparison set-up

The final comparison in this chapter is between the WHPSF configuration of FLEX-Topo and the HBV configuration used for operational forecasting in the Meuse basin, which is the benchmark model for this entire research. The operational version of HBV is slightly different from the version used in the model intercomparison (Sect. 7.3): the operational version has a non-linear fast runoff reservoir, instead of a combined fast and very quick runoff reservoir (yellow/orange reservoir in Fig. 7.4). In addition, a different calibration procedure was used, so the parameters are slightly different as well.

For both models the gridded set-up was used, to diminish the influence of differences in forcing data and spatial model resolution. The models were run in validation
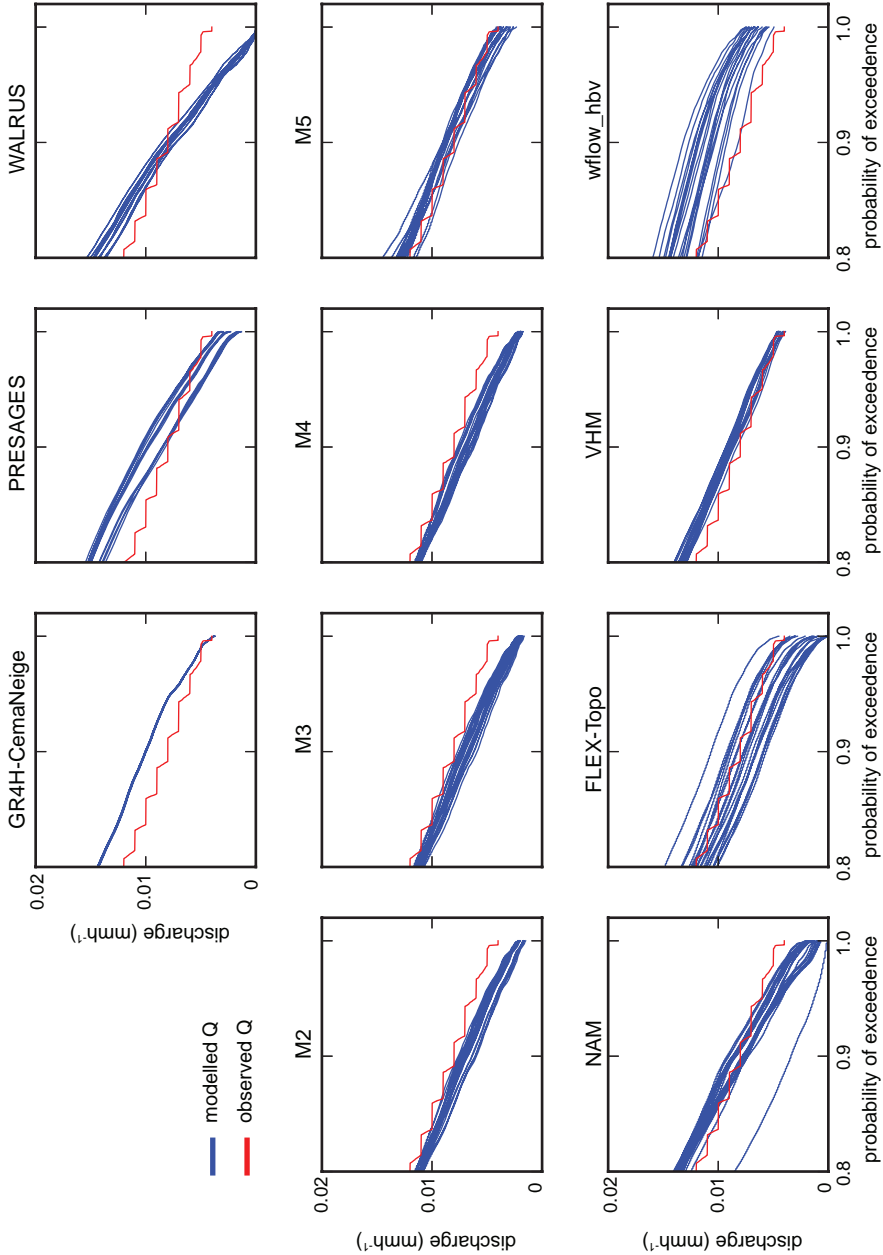
Figure 7.5: Model results for low flows for all model structures used in the model intercomparison study. The Flow duration curves are based on the entire modelled period (Jan 2001 - Dec 2010).
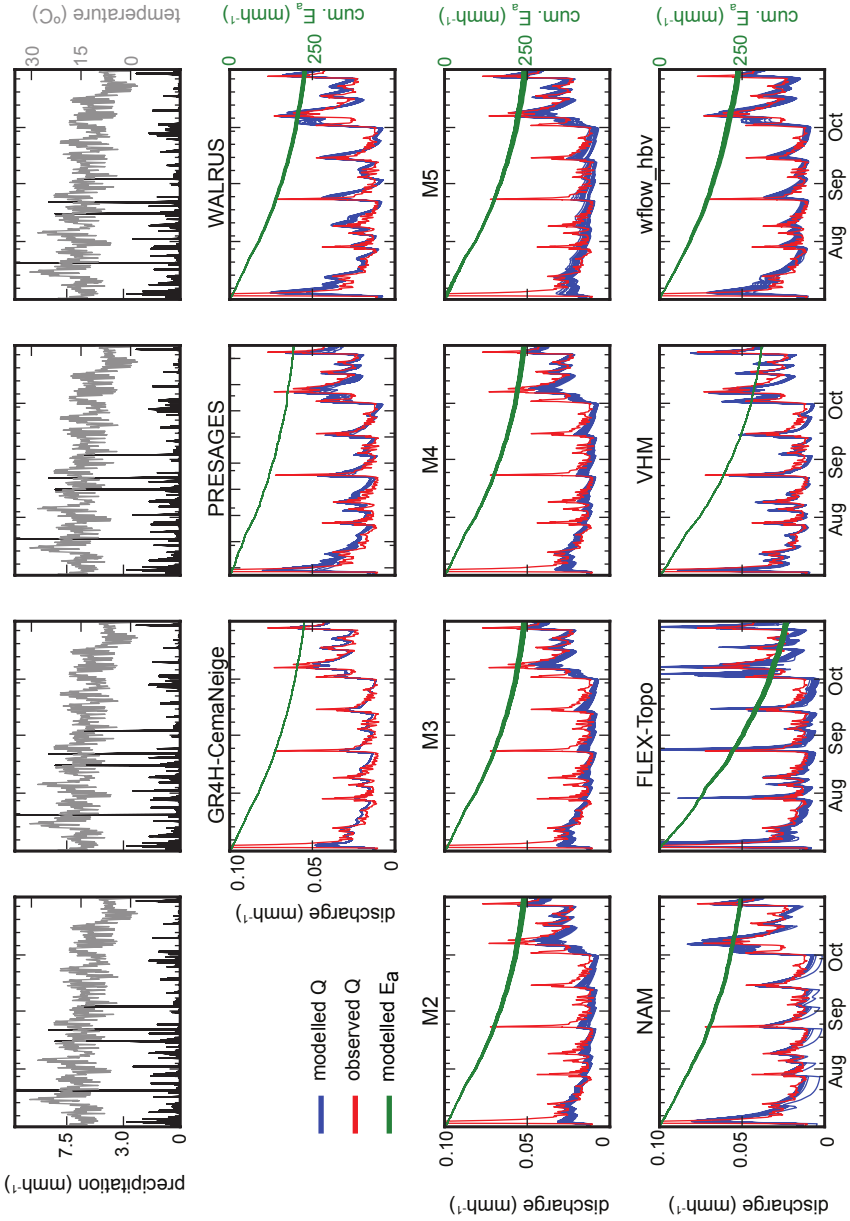
Figure 7.6: Model results for summer period 1 July to 1 November 2008 for all model structures used in the model intercomparison study.

mode from 1 January 2000 till 1 May 2011 and the modelled hydrographs were compared for four events.

7.4.2. RESULTS AND DISCUSSION

Figure 7.7 shows the hydrographs of four events: two winter periods (2003 and 2011) and two summer periods (2004 and 2008). In general, FLEX-Topo is better able to represent the dynamics of the hydrographs. For example for the summer periods: the observed hydrograph shows very sharp peaks as a consequence of high intensity precipitation events. HBV is more or less able to capture the height of the peaks, but the recession of the peaks is too slow and the base flow is too high. This is probably because FLEX-Topo has Hortonian overland flow incorporated and HBV not. In the 2011 winter a quick snow melt period followed a longer period with snow accumulation, causing a flood event in the lower Meuse. During this event, HBV is not able to turn the snow cover into runoff quick enough, while FLEX-Topo seems to be able to do this, because of the frozen soil module.

On the other hand, in a winter without significant snow cover (2003), HBV is better able to model the peak flows. This is probably caused by the non-linear fast runoff reservoir, which can very quickly turn precipitation into runoff, especially if the calibration period included some of the highest events. However, as it only works under certain conditions, it is not a very thrustworthy representation of the catchment behaviour. It would be very interesting to investigate which runoff mechanisms can be described so well with this non-linear runoff reservoir. These processes could then be included in for example FLEX-Topo.

7

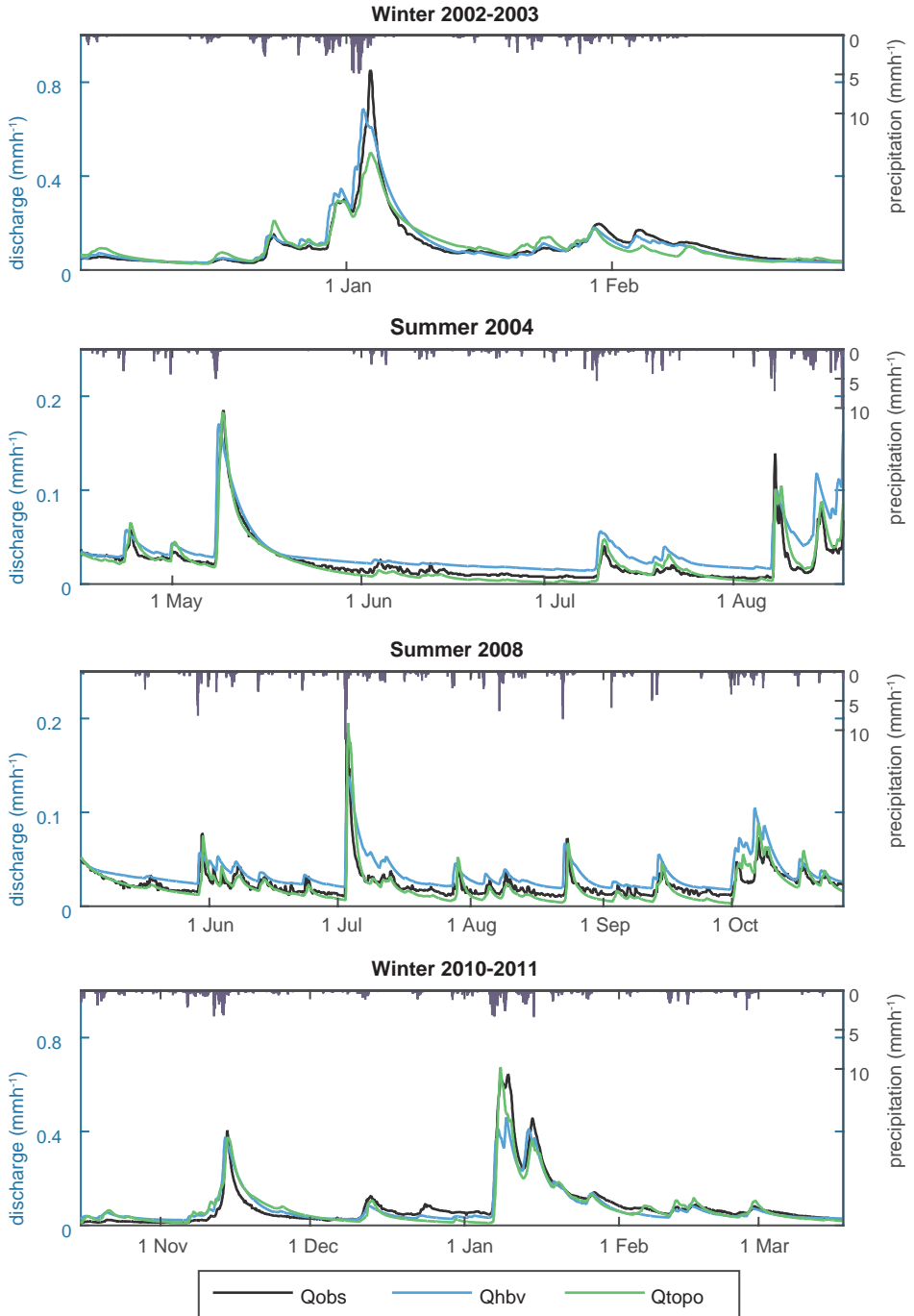Figure 7.7: Observed (black) and modelled discharges for two winter and two summer events. Modelled discharges are for the gridded set-up of FLEX-Topo (WHPSF) (green) and the operational HBV configuration (blue)

8

CONCLUSIONS AND OUTLOOK

εἰδέναι μὲν μηδὲν πλὴν αὐτὸ τοῦτο εἰδέναι
he knew nothing except that he knew that very fact

Socrates, quoted by Diogénes Laértios (Lives and Opinions of Eminent Philosophers)

8.1. THE VALUE OF DISTRIBUTION

The goal of this thesis is to obtain more insight in the added value of applying distribution in rainfall-runoff models, with specific focus on the Meuse basin. The effect of distribution is generally difficult to evaluate. Therefore, this thesis also aimed to set-up and use a strong evaluation framework and systematic model experiments to analyse various distribution options for a set of subcatchments of the Meuse basin.

A NEW EVALUATION FRAMEWORK FOR RAINFALL-RUNOFF MODELS

This thesis presents a new framework for the evaluation of rainfall-runoff models. An important starting point in this framework is that the added value of distribution in rainfall-runoff models cannot be assessed by only using standard overall performance metrics (i.e., NSE, RMSE, MAE). Instead, using a set of hydrological signatures can reveal more about model performance. However, a good performance is not very helpful if the consistency (i.e., the ability of the model to reproduce a set of signature with the same parameter set) is low. This framework therefore takes both performance and consistency into account. Signatures derived from independent data sources can reveal even more about the performance and consistency of model structures. With this evaluation framework in place, different distribution options have been explored, ranging from broad characteristics to specific runoff processes.

ROOT ZONE STORAGE CAPACITY FROM CLIMATE DATA

The root zone storage capacity (S_r) is one of the most important parameters in rainfall-runoff models. This thesis tests the influence of climate and soil characteristics on the root zone storage capacity. It is concluded that climate is a better estimator for S_r than soil, especially in wet areas. Added advantage is that climate data is more easily available, and in larger detail, than soil data, while using climate data also enables models to adapt to changing climate conditions. Furthermore, it is found that the climate derived S_r differs indeed between subcatchments of the Meuse. Following this approach, deriving specific S_r-values for the Ourthe gave better performance and consistency than transferring the S_r-value derived for the Ourthe Orientale. These results support the value of distributing this parameter among subcatchments.

DISTRIBUTING FORCING DATA AND MODEL STRUCTURE

Zooming in on the Ourthe, first the relative effect of distributed forcing and model structure was tested. Distribution of model structure is based on hydrological response units (HRUs) linked to dominant runoff processes. For the Ourthe, distribution of the model structure has a larger and more positive effect on performance and consistency than distribution of forcing data. This means that a parallel configuration of HRUs works in this area. Although the distribution of forcing data has a smaller influence, it appears to be essential when distributing the model structure. In other words: when distributing the model structure, one should distribute the forcing data as well to obtain good results. Distribution of the groundwater reservoir does not influence the results, because of the long time scales related to this reservoir.

INCLUDING MORE DETAILED PROCESS CONCEPTUALISATIONS

Knowing that it makes sense to distribute a model structure based on HRUs, the influence of more detailed process conceptualisations was tested: snow melt, Hortonian overland flow and frost in the topsoil. The latter two are especially important on agricultural fields and, so far, rarely applied in conceptual models. It is concluded that at an event time scale, incorporating these processes improves both model performance and consistency. It should be noted though, that incorporating these processes should be done cautiously, because they are more sensitive to calibration, data errors and signature selection.

BENCHMARK OF THE NEW FLEX-TOPO MODEL

After testing detailed process conceptualisations in FLEX-Topo, the value of FLEX-Topo was explored in other catchments and compared to other models, including the model used for operational forecasts in the Meuse basin at the moment of writing. FLEX-Topo could be transferred to other catchment without much change in performance; only the Hortonian overland flow module might require recalibration. The comparison with other models shows the importance of including a very fast runoff component and a parallel configuration between the fast and slow runoff generating reservoirs, to capture the different dynamics of the hydrograph. These two elements are included in FLEX-Topo. The explorative comparison further showed some shortcomings for FLEX-Topo, for example during winters without significant snow and during the saturation of the catchment after dry periods. The study did not focus on the former and no effective options were found for the latter. Overall, FLEX-Topo is much more able to reflect the dynamics of the hydrograph than the HBV model configuration.

Distribution and detailed process conceptualisations are very beneficial for a rainfall-runoff model of the Ourthe catchment. However, they should be applied with care. Each intermediate step should be specifically evaluated in order to prevent adding model elements that only increase complexity without additional value.

8.2. IMPLICATIONS

When working with conceptual models you often come across the question whether you know what is actually happening. To be honest, often you don't know. In many cases it is very likely that the model is not representing what is actually happening, simply because reality is too much simplified in a conceptual model. The additional question is than whether our conceptual models represent the emergent behaviour at the scale we are modelling, or that they are actually only representing the relevant time scales. Having this in mind, it is difficult to distinguish the effect of small adaptations to a model from the general background noise.

The focus of model evaluation on discharge data slightly adds to this problem: does discharge data contain enough information to identify the complete range of hydrological processes? Probably not, but visual inspections of the hydrograph and the use of hydrological signatures help to filter patterns from the discharge data. However, this increases the risk to see what we want to see.

Eventually, model conceptualisation and evaluation mainly comes down to data availability and data handling. The effect of data handling especially emerges when incorporating detailed process conceptualisations and focussing on the event time scale (e.g., Chs. 6 and 7. So, it is important to evaluate your data carefully and investigate how sensitive your model results are, especially when making forecasts. The limited variety of data, on the other hand, can make you short-sighted, and thus preventing to see the larger picture.

8.3. OPPORTUNITIES

Instead of arguing why a lack of data limits obtaining good and useful model results, we can also look at what we can achieve with the available data. By combining data and expert knowledge we often know more about a catchment than we think (e.g., Hrachowitz et al., 2014)). If we, in addition, divide our catchment (and model) sensibly into a small number of HRUs, we can make the model behaviour more specific, without making our models unnecessarily complex (e.g., Chs. 5 and 6). Conceptual models can in this way, tell us something about the response time scales of a catchment. Due to their ability to incorporate threshold behaviour, they can be applied at slightly different scales as well.

If a model is conceptual, it does not mean that it cannot contain any physical meaning. The climate-derived root zone storage capacity is a very relevant example. By arguing that this conceptual reservoir should be large enough to meet the evaporative demand, it can actually be derived from commonly available data. When we evaluate the parameters and reservoirs of our conceptual models in such a way, we can increase their physical basis. This can increase the applicability of our conceptual models, both in time and in space.

Finally, the increasing availability and diversity of remote sensing data can help us with the scaling issues. We now have data from two sides of the spectrum: both at small (plot) scale as at large (integrated) scale. These data can help us to better understand our hydrological systems at different scales. Furthermore, it can help strengthen the link between different scales and how they influence each other. This link may be very important for solving the transfer and scaling issues, which are relevant when moving from plot to catchment scale.

8.4. FANTASIES

Objectivity, systematic experiments and transparency are naturally linked to science. However, creativity and fantasies are a maybe small, but important part of science as well (Savenije, 2009). Without crazy ideas new research projects would not be initiated.

As you might have understood from above, my scientific fantasies as a hydrologist are strongly related to data availability. Specifically focussing on the question, which data would be needed to better identify or reject a specific hydrological process. Hoping that in the end we are able to filter the relevant system at the relevant scale and identify parameters and processes at this scale (e.g., Kirchner, 2006).

I consider three types of data sets especially valuable: components of evaporation, root zone storage dynamics and groundwater dynamics. These data sets are not very exotic and do not include variables which are not yet measured. However, they are hardly

available at meso- or large-scale catchments, while heterogeneity and mixing of hydrological processes is more relevant for these catchments, compared to small experimental catchments or hillslopes.

Components of evaporation   Evaporation is one of the largest hydrological fluxes, but up to now we only have scattered point measurements or model outputs. Having an evaporation data set covering an entire catchment, would enable us to check the water balance of catchments more thoroughly. In addition, a division of total evaporation into transpiration, canopy interception, ground interception, soil evaporation and open water evaporation would enable us to investigate the behaviour of vegetation under wetter and drier conditions and account separately for the different evaporation components in our models. This in turn would add more physical meaning to our conceptual models.

Root zone storage dynamics   Runoff generation is mainly determined by the antecedent storage and the precipitation event. Currently we estimate antecedent storage from model states or point observations of soil moisture. Having a root zone storage data set covering the entire catchment, would enable us to better estimate the (antecedent) storage in the catchment. From there the response of the catchment to a certain precipitation event can be determined. The different responses of the different HRUs are partly caused by differences in root zone storage (capacity). Thus, knowing the spatial variability in root zone storage would strengthen the hypothesis about the contribution of different HRUs. In addition, available root zone storage can give insight in the behaviour of vegetation under wetter and drier conditions, when combined with transpiration data.

Groundwater dynamics   Often we assume that base flow originates from groundwater storage. However, the groundwater storage in our conceptual model often is a linear reservoir, while in reality the subsurface pathways are much more complicated. Having a groundwater data set covering the dynamics in an entire catchment, would help us obtain more insight into the storage-discharge relation during low flow conditions and thus whether our linear reservoir assumption can sustain. Further, it would help us to identify which specific areas recharge at which moments to the groundwater. This can for example be useful to further investigate the dynamics of different HRUs or the infiltration patterns during snow and frost conditions. For the previous two data sets I have some (very) conceptual idea of how to obtain these data and how they would look like; on the other hand, the subsurface contains many hidden processes influencing groundwater dynamics. These hidden processes make it for me very difficult to even imagine how a data set of groundwater dynamics for an entire catchment would look like.

8

The need or wish for new data often comes from model results that cannot be explained with the available data. So, iteratively modelling and measuring might be a good option to obtain the data most useful to support our models. This approach is already used in smaller catchments. In larger catchments this requires a different approach and probably more planning and resources, but in the end, I believe it can help to better understand our catchments and thus make better models.

And of course, everyone hopes that at some point in time part of your fantasies become reality.

8

A

ILLUSTRATION **PCA** FOR **FARM**

INTRODUCTION

This appendix gives a synthetic example of the use of a principal component analysis for FARM. For FARM multiple evaluation criteria are used; however, for this example only two evaluation criteria are used, to be able to visualise the results. In this example two cases will be discussed:

1. two directly correlated evaluation criteria and
2. two inversely correlated evaluation criteria.

BASIC PRINCIPLES OF PCA

The PCA applied for FARM consists of several steps, which are listed below.

- The original data with values for evaluation criterion 1 (E_1) and evaluation criterion 2 (E_2) (first column of Fig. A.1) are obtained.
- The covariance matrix of the evaluation criteria is calculated (Tab. A.1).
- Calculation of the eigenvalues and eigenvectors of the covariance matrix (Tab. A.2) results in as many eigenvectors as evaluation criteria. The eigenvectors are orthogonal and the eigenvector with the largest eigenvalue describes the largest amount of variance in the data. The eigenvectors can be expressed in terms of E_1 and E_2 (second column of Fig. A.1).
- Selection of the amount of principal components (PCs) (the eigenvalues) that are taken into account is done based on the variance explained by each PC. The explained variance per PC is the eigenvalue of that PC divided by the sum of all eigenvalues. In case of two evaluation criteria, all PCs can be presented in a 2-D graph, as there are only 2 PCs.
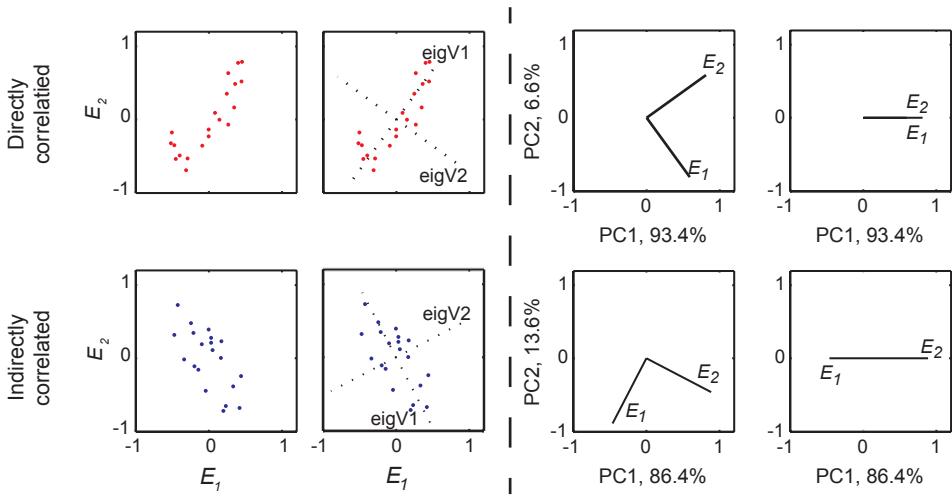
Figure A.1: Example showing the basic principles of principal component analysis and the application of PCA for FARM. top row: case 1; bottom row: case 2; first column: original data for E_1 and E_2; second column: original data with eigenvectors ('eigV', dotted); third column: Es expressed in terms of PC1 and PC2; fourth column: Es expressed in terms of PC1.

- Expression of evaluation criteria in terms of the principal components (in this case: PC1 and PC2). The third column in Figure A.1 shows the loadings of both Es on PC1 and PC2.

- The relative direction of the vectors can be used to identify the consistency. In Figure A.1(third column) the relative direction of the vectors in both cases seems similar. However, for case 1 the vectors have an opposite loading on PC2, which represents a very small amount of the variance. For case 2 the vectors have an opposite loading on PC1, which describes the largest amount of the variance. Therefore, case 1 has a much higher consistency than case 2 (of course, for two Es this can be easily deduced from the original data as well).

REDUCTION OF DIMENSIONS

In the step-wise approach described above, both PCs are kept. However, as can be seen in Figure A.1 PC1 describes a much larger part of the variance than PC2; thus, PC2 can be disregarded. The result of disregarding PC2 is shown in the last column of Figure A.1. A reduction of the dimensions leads for case 1 to two vectors in the same direction, while for case 2 it leads to two vectors with exactly opposite directions. This is because for case 1 the vectors had an opposite loading on PC2 and for case 2 an opposite loading on PC1.

Table A.1: Covariance matrix, left: case 1; right: case 2.

Directly correlated Es		Inversely correlated Es	
0.124	0.125	0.070	−0.070
0.125	0.128	−0.070	0.095

Table A.2: Eigenvalues and eigenvectors, left: case 1; right: case 2.

Directly correlated Es		Inversely correlated Es	
		eigenvalues	
0.0004	0.251	0.011	0.154
		eigenvectors	
−0.714	0.700	−0.768	−0.641
0.700	0.714	−0.641	0.768

B

MODEL EQUATIONS AND PARAMETERS

This appendix contains the model equations (Tab. B.1) and model parameters (Tab. B.2 to B.4) used in the FLEX-Topo configurations in Chapters 5, 6 and 7.

Table B.1: Equations for all structures; fluxes are calculated based on previous time step (including inflow of current time step) and corrected according to ratio in case of negative storage. The numbers refer to the chapter in which the equations are used. S represents a state (mm), Q a discharge (mmh^{-1}) and E an evaporative flux (mmh^{-1}); the subscripts indicate the reservoir to which the state or flux applies, except for subscripts p and c, which represent percolation and capillary rise. P and E_{pot} indicate precipitation (mmh^{-1}) and potential evaporation (mmh^{-1}), respectively. The used parameters are described in Tables B.2 to B.4.

Equation	Wetland[a]	Hillslope	Plateau
Snow (S_w)			
$\dfrac{\mathrm{d}S_w}{\mathrm{d}t} = P_s - Q_w - E_i$	6[b]	6	6
$Q_w = \begin{cases} K_{m,s,P}(T_a - T_m), & \text{if } S_w > 0 \\ 0 & \text{else} \end{cases}$	6	6	6
$E_w = \begin{cases} E_{pot,s}, & \text{if } S_w > 0 \\ 0 & \text{else} \end{cases}$	6	6	6
$K_{m,s,P} = \begin{cases} K_{m,s}K_{m,P}P, & \text{if } P \geq 1 \\ K_{m,s} & \text{else} \end{cases}$	6	6	6
Interception (S_i)			
$\dfrac{\mathrm{d}S_i}{\mathrm{d}t} = P - Q_i - E_i$	5,6	5,6,5L[c]	6
$Q_i = \begin{cases} S_i + P - I_{max}, & \text{if } S_i + P > I_{max} \\ 0, & \text{else} \end{cases}$	5,6	5,6,5L	6
$E_i = \begin{cases} E_{pot}, & \text{if } S_i > 0 \\ 0 & \text{else} \end{cases}$	5,6	5,6,5L	6
$Q_i = Q_i + Q_w$	6	6	6
Agriculture zone (S_a)			
$\dfrac{\mathrm{d}S_a}{\mathrm{d}t} = Q_i - E_a - Q_a - F_a$			6
$E_a = \begin{cases} (E_{pot} - E_i)(\frac{S_a}{S_{a,max,fr}L_P}), & \text{if } S_a < S_{a,max,fr}L_P \\ 1, & \text{else} \end{cases}$			6
			6
$Q_a = \begin{cases} S_a + Q_i - S_{a,max,fr}, & \text{if } S_a + Q_i > S_{a,max,fr} \\ Q_i(1 - (1 - \overline{S_a})^\beta), & \text{else} \end{cases}$			6
$F_a = F_{max} \exp^{-F_{dec}\overline{S_a}}$			6
$\overline{S_a} = \dfrac{S_a}{S_{a,max,fr}}$			6
$S_{a,max,fr} = S_{a,max}$			6[d]
$S_{a,max,fr} = S_{a,max}F_t$			6[e]

Table continues on next page

Table B.1 – continued from previous page

Equation	Wetland	Hillslope	Plateau
$F_t = \begin{cases} S_{a,min}, & \text{if } F_{acc,fr} < F_{acc,fr0} \\ \frac{F_{acc,fr}}{F_{acc,fr1}-F_{acc,fr0}} - \frac{F_{acc,fr0}}{F_{acc,fr1}-F_{acc,fr0}}, & \text{if } F_{acc,fr0} \leq F_{acc,fr} \leq F_{acc,fr1} \\ 1, & \text{if } F_{acc,fr} > F_{acc,fr1} \end{cases}$			6
$\frac{\mathrm{d}F_{acc,fr}}{\mathrm{d}t} = \begin{cases} T_a K_{m,f}, & \text{if } T_a < 0 \\ T_a K_{FT} K_{m,f}, & \text{if } T_a > 0 \end{cases}$			6

B

Agriculture fast runoff ($S_{f,a}$)

Equation	Wetland	Hillslope	Plateau
$\frac{\mathrm{d}S_{f,a}}{\mathrm{d}t} = Q_{a,lag} - Q_{f,a}$			6
$Q_{a,lag} = Q_a *_1^{T_{f,a}} T_{f,a,lag}$			6
$T_{f,lag} = \frac{2}{T_{f,a}} - \frac{2i}{T_{f,a}(T_{f,a}+1)}$			6
$Q_{f,a} = \frac{S_{f,a}}{K_{f,a}}$			6

Root zone (S_r)

Equation	Wetland	Hillslope	Plateau
$\frac{\mathrm{d}S_r}{\mathrm{d}t} = Q_i - E_r - Q_r$		5L	
$\frac{\mathrm{d}S_r}{\mathrm{d}t} = Q_i - E_r - Q_r - Q_p + Q_c$	5,6	5,6	6
$E_r = E_{pot}\left(1 - \frac{0.01\overline{S_i}}{\overline{S_i}+0.01}\right)\frac{1.01 * \overline{S_r}}{\overline{S_r}+0.01}$ **f**		5,5L	
$E_r = \begin{cases} E_{pot}, & \text{if } S_i > 0 \\ 0 & \text{else} \end{cases}$	5		
$E_r = \begin{cases} (E_{pot}-E_i)(\frac{S_r}{S_{r,max}L_P}, & \text{if } S_r < S_{r,max}L_P \\ 1, & \text{else} \end{cases}$	6	6	
$E_r = \begin{cases} (E_{pot}-E_i-E_a)(\frac{S_r}{S_{r,max}L_P}, & \text{if } S_r < S_{r,max}L_P \\ 1, & \text{else} \end{cases}$			6
$Q_r = Q_i \left(\frac{S_r + Q_i}{S_{r,max}}\right)^{\beta}$ **f**	5	5,5L	
$Q_r = \begin{cases} S_r + Q_i - S_{r,max}, & \text{if } S_r + Q_i > S_{r,max} \\ Q_i(1-(1-\overline{S_r})^{\beta}), & \text{else} \end{cases}$	6	6	
$Q_r = \begin{cases} S_r + F_a - S_{r,max}, & \text{if } S_r + F_a > S_{r,max} \\ F_a(1-(1-\overline{S_r})^{\beta}), & \text{else} \end{cases}$			6
$Q_p = K_p * \overline{S_r}$	5,6	5,6	6
$Q_c = \begin{cases} K_c(1-\overline{S_r}), & \text{if } S_s > 0 \\ 0 & \text{else} \end{cases}$	5,6	5,6	6

Table continues on next page

B

Table B.1 – continued from previous page

Equation	Wetland	Hillslope	Plateau
$\overline{S_i} = \dfrac{S_i}{I_{max}}$	5,6	5,6,5L	6
$\overline{S_r} = \dfrac{S_r + Q_i}{S_{r,max}}$	5,6	5,6,5L	6
Fast runoff (S_f)			
$\dfrac{\mathrm{d}S_f}{\mathrm{d}t} = Q_{r,lag} - Q_f$	5,6	5,6,5L	6
$Q_{r,lag} = Q_{f,in} *_1^{T_f} T_{f,lag}$	5,6	5,6,5L	6
$Q_{f,in} = (1 - D) Q_r$	5,6	5,6,5L	6
$T_{f,lag} = \dfrac{2}{T_f} - \dfrac{2i}{T_f(T_f + 1)}$	5,6	5,6,5L	6
$Q_f = \dfrac{S_f}{K_f}$	5,6	5,6,5L	6
Slow runoff (S_s)			
$\dfrac{\mathrm{d}S_s}{\mathrm{d}t} = DQ_r - Q_s$		5L	
$\dfrac{\mathrm{d}S_s}{\mathrm{d}t} = DQ_r + Q_p - Q_c - Q_s$	5,6	5,6	6
$Q_s = \dfrac{S_s}{K_s}$	5,6	5,6,5L	6

[a] The numbers indicate in which chapter the equation is used.   [b] The model equations for Chs. 6 and 7 are the same.   [c] '5L' refers to the lumped model structure in Ch. 5.

[d] For model configurations without frost in the topsoil.

[e] For model configurations with frost in the topsoil.   [f] After Fenicia et al. (2011).

B

Table B.2: Model parameters and derived values for the wetland HRU for the different versions of FLEX-Topo

Parameter	Symbol	Unit	Ch. 5	Wetland Ch. 6	Ch. 7[a]
maximum interception capacity	I_{max}	mm	0.8-1.2	0.8-1.2	1.0
minumum storage capacity in agricultural reservoir (relative to $S_{a,max}$)	$S_{a,min}$	-	NA	NA	NA
maximum storage capacity in agricultural reservoir	$S_{a,max}$	mm	NA	NA	NA
maximum storage capacity in root zone storage	$S_{r,max}$	mm	70-150	60-90	65
coefficient determining storage discharge relation of S_r	β	-	4.0-5.5	0.10-0.30	0.26
maximum infiltration capacity	F_{max}	mmh^{-1}	NA	NA	NA
decline of infiltration capacity	F_{dec}	-	NA	NA	NA
fraction of preferential recharge from S_r to S_s	D	-	0	0	0
coefficient for recession of fast runoff reservoir	K_f	h	16.7-25.0	5.0-90	25.1
coefficient for recession of HOF runoff reservoir	$K_{f,a}$	h	NA	NA	NA
base of lag function before S_f	T_f	h	1	1	1
coefficient for capillary rise	K_c	mmh^{-1}	0.08-0.15	0.0001-0.0400	0.012
coefficient for percolation	K_p	mmh^{-1}	0	0	0
limit for potential evaporation	LP	-	x	0.10-0.53	0.2
coefficient determining storage discharge relation of S_a	β_a	-	NA	NA	NA
threshold temperature between liquid and solid precipitation[d]	T_t	°C	x	1.8-2.0	1.9
threshold temperature for start snow melt[d]	T_m	°C	x	2.2-2.5	2.2
melt coefficient for snow melt[d]	$K_{m,s}$	mm °C^{-1}h^{-1}	x	0.1-0.18	0.18
coefficient for recession of slow runoff reservoir	K_s	h	2500	2500	2500

[a] Ch. 7 only uses 1 model realisation. [b] x = not used: parameter can be used for this HRU, but is not selected. [c] NA = not applicable: parameter is not relevant for this HRU. [d] Only for configurations with snow module.

Table B.3: Model parameters and derived values for the hillslope HRU for the different versions of FLEX-Topo

Parameter	Symbol	Unit	Lumped Ch. 5	Ch. 5	Hillslope Ch. 6	Ch. 7[a]
maximum interception capacity	I_{max}	mm	1.0-2.0	1.8-2.2	1.8-2.5	2.3
minumum storage capacity in agricultural reservoir (relative to $S_{a,max}$)	$S_{a,min}$	-	NA[c]	NA	NA	NA
maximum storage capacity in agricultural reservoir	$S_{a,max}$	mm	NA	NA	NA	NA
maximum storage capacity in root zone storage	$S_{r,max}$	mm	150-500	150-350	112-141	117
coefficient determining storage discharge relation of S_r	β	-	1.0-3.0	1.0-1.3	0.2-0.4	0.28
maximum infiltration capacity	F_{max}	mmh^{-1}	NA	NA	NA	NA
decline of infiltration capacity	F_{dec}	-	NA	NA	NA	NA
fraction of preferential recharge from S_r to S_s	D	-	0.20-0.30	0.15-0.30	0.1-0.4	0.17
coefficient for recession of fast runoff reservoir	K_f	h	100-200	111-200	111-500	196
coefficient for recession of HOF runoff reservoir	$K_{f,a}$	h	NA	NA	NA	NA
base of lag function before S_f	T_f	h	2-3	2-3	2-3	2
coefficient for capillary rise	K_c	mmh^{-1}	x[b]	0	0	0
coefficient for percolation	K_p	mmh^{-1}	x	0.001-0.0082	0.0001-0.0050	0.0008
limit for potential evaporation	LP	-	x	x	0.13-0.6	0.35
coefficient determining storage discharge relation of S_a	β_a	-	NA	NA	NA	NA
threshold temperature between liquid and solid precipitation[d]	T_t	°C	x	x	1.8-2.0	1.9
threshold temperature for start snow melt[d]	T_m	°C	x	x	2.2-2.5	2.2
melt coefficient for snow melt[d]	$K_{m,s}$	mm°C^{-1}h^{-1}	x	x	0.1-0.18	0.18
coefficient for recession of slow runoff reservoir	K_s	h	2500	2500	2500	2500

[a] Ch. 7 only uses 1 model realisation.   [b] x = not used: parameter can be used for this HRU, but is not selected.   [c] NA = not applicable: parameter is not relevant for this HRU.
[d] Only for configurations with snow module.

B

Table B.4: Model parameters and derived values for the plateau HRU for the different versions of FLEX-Topo

Parameter	Symbol	Unit	Plateau Ch. 6	Ch. 7[a]
maximum interception capacity	I_{max}	mm	0.8-2.4	2.3
minumum storage capacity in agricultural reservoir (relative to $S_{a,max}$)[b]	$S_{a,min}$	-	0.15-0.21	0.19
maximum storage capacity in agricultural reservoir	$S_{a,max}$	mm	40-50	42
maximum storage capacity in root zone storage	$S_{r,max}$	mm	112-141	117
coefficient determining storage discharge relation of S_r	β	-	0.20-0.40	0.27
maximum infiltration capacity	F_{max}	mmh^{-1}	1.20-1.70	1.34
decline of infiltration capacity	F_{dec}	-	0.1-0.4	0.2
fraction of preferential recharge from S_r to S_s	D	-	0.20-0.50	0.31
coefficient for recession of fast runoff reservoir	K_f	h	111-500	196
coefficient for recession of HOF runoff reservoir	$K_{f,a}$	h	5.0-90	25.1
base of lag function before S_f	T_f	h	2	2
coefficient for capillary rise	K_c	mmh^{-1}	0	0
coefficient for percolation	K_p	mmh^{-1}	0.0001-0.0050	0.0008
limit for potential evaporation	LP	-	0.11-0.56	0.21
coefficient determining storage discharge relation of S_a	β_a	-	0.10-0.30	0.26
threshold temperature between liquid and solid precipitation[c]	T_t	°C	1.8-2.0	1.9
threshold temperature for start snow melt[c]	T_m	°C	2.2-2.5	2.2
melt coefficient for snow melt[c]	$K_{m,s}$	mm°C^{-1} h^{-1}	0.1-0.18	0.18
coefficient for recession of slow runoff reservoir	K_s	h	2500	2500

[a] Ch. 7 only uses 1 model realisation.   [b] Only for configurations with frozen topsoil.   [c] Only for configurations with snow module.

B

Table B.5: Model parameters with fixed values for all model experiments and HRUs

Parameter	Symbol	Unit	Value
max modelled accumulated frost resulting in $S_{a,min}$	$F_{acc,fr0}$	°Cd	-3
min modelled accumulated frost resulting in $S_{a,max}$	$F_{acc,fr1}$	°Cd	0
melt coefficient for melt of frozen topsoil	$K_{m,f}$	-	1
coefficient for ratio between melt and frost speed of topsoil	K_{FT}	-	2
multiplier for increased snow melt in case of liquid precipitation	$K_{m,P}$	hmm^{-1}	1

REFERENCES

Ajami, N., Gupta, H., Wagener, T., Sorooshian, S., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. Journal of Hydrology 298 (1-4), 112–135.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., 2012. All that glitters is not gold: the case of calibrating hydrological models. Hydrological Processes 26 (14), 2206–2210.

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C., Loumagne, C., 2004. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds. Water Resources Research 40 (5), W05209.

Ashagrie, A. G., de Laat, P. J. M., de Wit, M. J. M., Tu, M., Uhlenbrook, S., 2006. Detecting the influence of land use changes on discharges and floods in the meuse river basin – the predictive power of a ninety-year rainfall-runoff relation? Hydrology and Earth System Sciences 10, 691–701.

Atkinson, S. E., Sivapalan, M., Viney, N. R., Woods, R. A., 2003. Predicting space-time variability of hourly streamflow and the role of climate seasonality: Mahurangi Catchment, New Zealand. Hydrological Processes 17 (11), 2171–2193.

Atkinson, S. E., Woods, R. A., Sivapalan, M., 2002. Climate and landscape controls on water balance model complexity over changing timescales. Water Resources Research 38 (12), 1314.

Avanzi, F., De Michele, C., Morin, S., Carmagnola, C. M., Ghezzi, A., Lejeune, Y., 2016. Model complexity and data requirements in snow hydrology: seeking a balance in practical applications: Complexity vs. Data Requirements in Snow Hydrology Models. Hydrological Processes 30, 2106–2118.

Bandaragoda, C., Tarboton, D. G., Woods, R., 2004. Application of TOPNET in the distributed model intercomparison project. Journal of Hydrology 298 (1-4), 178–201.

Bárdossy, A., Singh, S. K., 2008. Robust estimation of hydrological model parameters. Hydrology and Earth System Sciences 12, 1273–1283.

Beekman, W., Caljé, R., Schaars, F., Heijkers, J., 2014. Vergelijking van enkele schattingsmethoden voor de actuele verdamping. Stromingen 20, 227–239.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., Savenije, H. H. G., 2014. Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. Water Resources Research 50 (7), 5638–5661.

Beven, K., 2001. Rainfall-Runoff Modelling: The Primer. John Wiley & Sons, Chichester, West Sussex, England.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320 (1-2), 18–36.

Beven, K. J., Freer, J., 2001. A dynamic TOPMODEL. Hydrological Processes 15 (10), 1993–2011.

Beven, K. J., Kirkby, M. J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24 (1), 43–69.

Birkel, C., Dunn, S. M., Tetzlaff, D., Soulsby, C., 2010. Assessing the value of high-resolution isotope tracer data in the stepwise development of a lumped conceptual rainfall-runoff model. Hydrological Processes 24 (16), 2335–2348.

Blazkova, S., Beven, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. Water Resources Research 45 (12), W00B16.

Blöschl, G., Sivapalan, M., Wagener, T., 2013. Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales. Cambridge University Press.

Boardman, J., Vandaele, K., 2016. Effect of the spatial organization of land use on muddy flooding from cultivated catchments and recommendations for the adoption of control measures. Earth Surface Processes and Landforms 41 (3), 336–343.

Booij, M. J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. Journal of Hydrology 303, 176–198.

Booker, D., Woods, R., 2014. Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. Journal of Hydrology 508, 227–239.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., Smith, M., 2001. Toward improved streamflow forecasts: value of semidistributed modeling. Water Resources Research 37 (11), 2749–2759.

Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., 2014. The Wageningen Lowland Runoff Simulator (WALRUS): a lumped rainfall–runoff model for catchments with shallow groundwater. Geoscientific Model Development 7 (5), 2313–2332.

Brunner, I., Herzog, C., Dawes, M. A., Arend, M., Sperisen, C., 2015. How tree roots respond to drought. Frontiers in Plant Science 6, 547.

Budyko, M., 1974. Climate and Life. Academic Press New York.

Bulcock, H. H., Jewitt, G. P. W., 2010. Spatial mapping of leaf area index using hyperspectral remote sensing for hydrological applications with a particular focus on canopy interception. Hydrology and Earth System Sciences 14 (2), 383–392.

Camporese, M., Daly, E., Paniconi, C., 2015. Catchment-scale Richards equation-based modeling of evapotranspiration via boundary condition switching and root water uptake schemes. Water Resources Research 51, 5756–5771.

Canadell, J., Jackson, R. B., Ehleringer, J. B., Mooney, H. A., Sala, O. E., Schulze, E.-D., 1996. Maximum rooting depth of vegetation types at the global scale. Oecologia 108 (4), 583–595.

Carpenter, T. M., Georgakakos, K. P., 2006. Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. Journal of Hydrology 329 (1-2), 174–185.

Cassiani, G., Boaga, J., Vanella, D., Perri, M. T., Consoli, S., 2015. Monitoring and modelling of soil–plant interactions: the joint use of ERT, sap flow and eddy covariance data to characterize the volume of an orange tree root zone. Hydrology and Earth System Sciences 19 (5), 2213–2225.

Cerdan, O., Le Bissonnais, Y., Govers, G., Lecomte, V., van Oost, K., Couturier, A., King, C., Dubreuil, N., 2004. Scale effect on runoff from experimental plots to catchments in

agricultural areas in Normandy. Journal of Hydrology 299 (1-2), 4–14.

Cherkauer, K. A., Lettenmaier, D. P., 1999. Hydrologic effects of frozen soils in the upper Mississippi River basin. Journal of Geophysical Research 104 (D16), 19599.

Clark, M. P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research 47 (9), W09301.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., Marks, D. G., 2015. A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies: A unified approach for process-based hydrologic modeling. Water Resources Research 51 (4), 2515–2542.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., Uddstrom, M. J., 2008a. Hydrological data assimilation with the ensemble kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Advances in Water Resources 31 (10), 1309–1324.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., Hay, L. E., 2008b. Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research 44, W00B02.

Cloke, H., Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375 (3-4), 613–626.

Criss, R. E., Winston, W. E., Jul. 2008. Do Nash values have value? Discussion and alternate proposals. Hydrological Processes 22 (14), 2723–2725.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., Walker, J. P., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Reviews of Geophysics 50 (2), RG2002.

Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. Journal of Hydrology 356 (1-2), 106–118.

de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G., Willems, P., 2016. Looking beyond general metrics for model comparison - lessons from an international model intercomparison study. Hydrology and Earth System Sciences Discussions, 1–29.

de Roo, A., Odijk, M., Schmuck, G., E. Koster, A. L., 2000. Assessing the effects of land use changes on floods in the meuse and oder catchment. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere 26, 593–599.

de Wit, M. J. M., 2008. Van regen tot Maas: grensoverschrijdend waterbeheer in droge en natte tijden. Veen Magazines, Diemen.

de Wit, M. J. M., Peeters, H. A., Gastaud, P. H., Dewil, P., Maeghe, K., Baumgart, J., 2007. Floods in the meuse basin: Event descriptions and an international view on ongoing measures. International Journal of River Basin Management 5, 279–292.

Descombes, X., 2012. Stochastic geometry for image analysis. ISTE Ltd, London.

Detty, J. M., McGuire, K. J., 2010. Topographic controls on shallow groundwater dynamics: implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment. Hydrological Processes 24 (16), 2222–2236.

Dirmeyer, P. A., 2011. The terrestrial segment of soil moisture-climate coupling. Geophysical Research Letters 38, L16702.

Donohue, R. J., Roderick, M. L., McVicar, T. R., 2012. Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model. Journal of Hydrology 436-437, 35–50.

Eagleson, P. S., 1982. Ecological optimality in water-limited natural soil-vegetation systems: 1. Theory and hypothesis. Water Resources Research 18 (2), 325–340.

Eshagh, M., Lemoine, J.-M., Gegout, P., Biancale, R., 2013. On regularized time varying gravity field models based on grace data and their comparison with hydrological models. Acta Geophysica 61 (1), 1–17.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., Savenije, H. H. G., 2013. A framework to assess the realism of model structures using hydrological signatures. Hydrology and Earth System Sciences 17 (5), 1893–1912.

Famiglietti, J. S., Cazenave, A., Eicker, A., Reager, J. T., Rodell, M., Velicogna, I., 2015. Satellites provide the big picture. Science 349 (6249), 684–685.

Fenicia, F., Kavetski, D., Savenije, H. H. G., 2010. Assessing the impact of mixing assumptions on the estimation of streamwater mean residence time. Hydrological Processes 24, 1730–1741.

Fenicia, F., Kavetski, D., Savenije, H. H. G., 2011. Elements of a flexible approach for conceptual hydrological modeling: 1. motivation and theoretical development. Water Resources Research 47, W11510x.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., Freer, J., 2013. Catchment properties, function, and conceptual model representation: is there a correspondence? Hydrological Processes 28 (4), 2451–2467.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Pfister, L., 2016. From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. Water Resources Research 52 (2), 954–989.

Fenicia, F., McDonnell, J. J., Savenije, H. H. G., 2008a. Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research 44 (6), W06419.

Fenicia, F., Savenije, H. H. G., Avdeeva, Y., 2009. Anomaly in the rainfall-runoff behaviour of the meuse catchment. climate, land-use or land-use management. Hydrology and Earth System Sciences 13, 1727–1737.

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

Fenicia, F., Savenije, H. H. G., Matgen, P., Pfister, L., 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling. Water Resources Research 43 (3), W03434.

Fenicia, F., Savenije, H. H. G., Matgen, P., Pfister, L., 2008b. Understanding catchment behavior through stepwise model concept improvement. Water Resources Research 44 (1), W01402.

Ferreira, C., Walsh, R., Steenhuis, T., Shakesby, R., Nunes, J., Coelho, C., Ferreira, A., 2015. Spatiotemporal variability of hydrologic soil properties and the implications for overland flow and land management in a peri-urban Mediterranean catchment. Journal of

Hydrology 525, 249–263.

Field, C. B., Chapin, F. S., Matson, P. A., Mooney, H. A., 1992. Responses of terrestrial ecosystems to the changing atmosphere: a resource-based approach. Annual Review of Ecology, Evolution and Systematics 23, 201–235.

Fiener, P., Auerswald, K., Van Oost, K., 2011. Spatio-temporal patterns in land use and management affecting surface runoff response of agricultural catchments—A review. Earth-Science Reviews 106 (1-2), 92–104.

Flügel, W.-A., 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using prms/mms in the drainage basin of the river brol, germany. Hydrological Processes 9 (3-4), 423–436.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., Savenije, H. H. G., 2014a. Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the upper heihe, china. Hydrology and Earth System Sciences 18 (5), 1895–1915.

Gao, H., Hrachowitz, M., Schymanski, S. J., Fenicia, F., Sriwongsitanon, N., Savenije, H. H. G., 2014b. Climate controls how ecosystems size the root zone storage capacity at catchment scale: Root zone storage capacity in catchments. Geophysical Research Letters 41 (22), 7916–7923.

Gentine, P., D'Odorico, P., Lintner, B. R., Sivandran, G., Salvucci, G., 2012. Interdependence of climate, soil, and vegetation as constrained by the Budyko curve. Geophysical Research Letters 39 (19), L19404.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., Savenije, H. H. G., 2014a. Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. Hydrology and Earth System Sciences 18 (12), 4839–4859.

Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H. H. G., 2011. Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central european meso-scale catchment. Hydrology and Earth System Sciences 15 (11), 3275–3291.

Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. V., Savenije, H. H. G., 2014b. A constraint-based search algorithm for parameter identification of environmental models. Hydrology and Earth System Sciences 18 (12), 4861–4870.

Gimbel, K., Puhlmann, H., Weiler, M., 2016. Does drought alter hydrological functions in forest soils? An infiltration experiment. Hydrology and Earth System Sciences 20, 1301–1317.

Gumbel, E., 1935. Les valeurs extrêmes des distributions statistiques. Annales de l'I. H. P. 5 (2), 115–158.

Gupta, H. V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes 22 (18), 3802–3813.

Guswa, A. J., 2008. The influence of climate on root depth: A carbon cost-benefit analysis. Water Resources Research 44 (2), W02427.

Hall, D., Salomonson, V., Riggs, G., 2006. MODIS/Terra Snow Cover Daily L3 global 500m grid, version 5. Tech. Rep. http://dx.doi.org/10.5067/63NQASRDPDB0, accessed: 11-02-2016, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA.

Härdle, W., Simar, L., 2003. Applied multivariate statistical analysis. Springer-Verlag Berlin Heidelberg.

Hargreaves, G., Samani, Z., 1985. Reference Crop Evapotranspiration from Temperature. Applied engineering in agriculture 1 (2), 96–99.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. Journal of Geophysical Research 113, D20119.

Hingray, B., Schaefli, B., Mezghani, A., Hamdi, Y., 2010. Signature-based model calibration for hydrological prediction in mesoscale alpine catchments. Hydrological Sciences Journal 55 (6), 1002–1016.

Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J., Ruiz, L., van der Velde, Y., Wade, A. J., 2016. Transit times-the link between hydrology and water quality at the catchment scale: Linking hydrology and transit times. Wiley Interdisciplinary Reviews: Water 3, 629–657.

Hrachowitz, M., Bohte, R., Mul, M. L., Bogaard, T. A., Savenije, H. H. G., Uhlenbrook, S., 2011. On the value of combined event runoff and tracer analysis to improve understanding of catchment functioning in a data-scarce semi-arid area. Hydrology and Earth System Sciences 15 (6), 2007–2024.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., Gascuel-Odoux, C., 2014. Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. Water Resources Research 50, 7445–7469.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., M.P., C., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., Cudennec, C., 2013a. A decade of Predictions in Ungauged Basins (PUB) - a review. Hydrological Sciences Journal 58 (6), 1198–1255.

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., Soulsby, C., 2013b. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? Hydrology and Earth System Sciences 17 (2), 533–564.

Humphrey, V., Gudmundsson, L., Seneviratne, S. I., 2016. Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes. Surveys in Geophysics 37 (2), 357–395.

HydroSHEDS, June 2013. 3sec grid: Conditioned dem.
    URL http://hydrosheds.cr.usgs.gov/dataavail.php

Johnson, R. A., Wichern, D. W., 1998. Applied multivariate statistical analysis. Prentice-Hall, Inc., Upper Saddle River.

Jolliffe, I. T., 1986. Principal Component Analysis. Springer-Verlag, New York Inc.

Kavetski, D., Fenicia, F., 2011. Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. Water Resources Research 47 (11), W11511.

Kirchner, J. W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42 (3), W03S04.

Kleidon, A., Heimann, M., 1998. A method of determining rooting depth from a terrestrial biosphere model and its impacts on the global water and carbon cycle. Global Change Biology 4 (3), 275–286.

Kleidon, A., Heimann, M., 2000. Assessing the role of deep rooted vegetation in the climate system with model simulations: mechanism, comparison to observations and implications for Amazonian deforestation. Climate Dynamics 16 (2-3), 183–199.

Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal 31 (1), 13–24.

Kling, H., Gupta, H., 2009. On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability. Journal of Hydrology 373 (3-4), 337–351.

Knudsen, J., Thomsen, A., Refsgaard, J., 1986. Watbal a semi-distributed, physically based hydrological modelling system. Nordic Hydrology 17 (4-5), 347–362.

Koren, V., Smith, M., Cui, Z., 2014. Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process. Journal of Hydrology 519, 3475–3491.

Krzanowski, W. J., 1988. Principles of multivariate analysis: a user's perspective. No. 3 in Oxford statistical science series. Clarendon Press ; Oxford University Press, Oxford [Oxfordshire] : New York.

Lamb, R., Beven, K., 1997. Using interactive recession curve analysis to specify a general catchment storage model. Hydrology and Earth System Sciences 1 (1), 101–113.

Li, Q., Sun, S., Xue, Y., 2010. Analyses and development of a hierarchy of frozen soil models for cold region study. Journal of Geophysical Research 115, D03107.

Liancourt, P., Sharkhuu, A., Ariuntsetseg, L., Boldgiv, B., Helliker, B. R., Plante, A. F., Petraitis, P. S., Casper, B. B., 2012. Temporal and spatial variation in how vegetation alters the soil moisture response to climate manipulation. Plant Soil 351 (1-2), 249–261.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed hbv-96 hydrological model. Journal of Hydrology 201, 272–288.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? an evaluation using 3620 flood events. Hydrology and Earth System Sciences 18 (2), 575–594.

Maneta, M., Wallender, W., 2013. Pilot-point based multi-objective calibration in a surface-subsurface distributed hydrological model. Hydrological Sciences Journal 58 (2), 390–407.

Martínez-Carreras, N., Wetzel, C. E., Frentress, J., Ector, L., McDonnell, J. J., Hoffmann, L., Pfister, L., 2015. Hydrological connectivity inferred from diatom transport through the riparian-stream system. Hydrology and Earth System Sciences 19 (7), 3133–3151.

McGlynn, B. L., McDonnel, J. J., Brammer, D. D., 2002. A review of the evolving perceptual model of hillslope flowpaths at the Maimai catchments, New Zealand. Journal of Hydrology 257 (1-4), 1–26.

McMillan, H., Booker, D., Cattoën, C., 2016. Validation of a national hydrological model. Journal of Hydrology 541, 800–815.

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., Rupp, D. E., 2013. Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2

catchment. Hydrological Processes 28 (18), 4896–4913.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., Woods, R. A., 2011. Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. Hydrological Processes 25 (4), 511–522.

Milly, P., 1994. Climate, interseasonal storage of soil water, and the annual water balance. Advances in Water Resources 17 (1-2), 19–24.

Müller, M. F., Thompson, S. E., 2016. Comparing statistical and process-based flow duration curve models in ungauged basins and changing rain regimes. Hydrology and Earth System Sciences 20 (2), 669–683.

Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 40 (1), W01106.

Moussa, R., Voltz, M., Andrieux, P., 2002. Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events. Hydrological Processes 16 (2), 393–412.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology 10 (3), 282–290.

Newsome, P., Wilde, R., Willoughby, E., 2000. Land resource information system spatial data layers. Tech. rep., Landcare Research NZ Ltd, Palmerston North, New Zealand.

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F., Soubeyroux, J.-M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B., Morice, E., 2014. Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. Hydrology and Earth System Sciences 18 (8), 2829–2857.

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., Hrachowitz, M., 2016. The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. Hydrology and Earth System Sciences 20 (3), 1151–1176.

Nippgen, F., McGlynn, B. L., Emanuel, R. E., 2015. The spatial and temporal evolution of contributing areas. Water Resources Research 51 (6), 4550–4573.

NIWA, 2015. Climate summaries. http://www.niwa.co.nz/education-and-training/schools/resources/climate/summary. Retrieved 06.07.2015.

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S., 2011. Height Above the Nearest Drainage – a hydrologically relevant new terrain model. Journal of Hydrology 404 (1-2), 13–29.

Ofir, M., Kigel, J., Apr. 1999. Photothermal control of the imposition of summer dormancy in Poabulbosa, a perennial grass geophyte. Physiologia Plantarum 105 (4), 633–640.

Orth, R., Seneviratne, S. I., 2014. Using soil moisture forecasts for sub-seasonal summer temperature predictions in Europe. Climate Dynamics 43 (12), 3403–3418.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., Zappa, M., 2015. Does model performance improve with complexity? A case study with three hydrological models. Journal of Hydrology 523, 147–159.

Oudin, L., Andréassian, V., Perrin, C., Anctil, F., 2004. Locating the sources of low-pass behavior within rainfall-runoff models. Water Resources Research 40 (11), W11101.

Penman, H., 1948. Natural evaporation from open water, bare soil and grass (193), 120–146.

Perrin, C., Michel, C., Andréassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology 242 (3-4), 275–301.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology 279, 275–289.

Peñuela, A., Darboux, F., Javaux, M., Bielders, C. L., 2016. Evolution of overland flow connectivity in bare agricultural plots. Earth Surface Processes and Landforms 41, 1595–1613.

Pfister, L., Savenije, H., Fenicia, F., 2009. Leonardo Da Vinci's Water Theory - On the origin and fate of water, iahs special publication 9 Edition. IAHS Press, Wallingford, Oxfordshire, UK.

Priestley, C. H. B., Taylor, R. J., 1972. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Monthly Weather Review 100 (2), 81–92.

Rakovec, O., Hazenberg, P., Torfs, P. J. J. F., Weerts, A. H., Uijlenhoet, R., 2012. Generating spatial precipitation ensembles: impact of temporal correlation structure. Hydrology and Earth System Sciences 16 (9), 3419–3434.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. a. V., Hodnett, M. G., Tomasella, J., Waterloo, M. J., 2008. HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in amazonia. Remote Sensing of Environment 112 (9), 3469–3481.

Ritsema, C. J., Stolte, J., Oostindie, K., Van Den Elsen, E., Van Dijk, P. M., 1996. Measuring and modelling of soil water dynamics and runoff generation in an agricultural loessial hillslope. Hydrological Processes 10 (8), 1081–1089.

Rodhe, A., Seibert, J., 1999. Wetland occurrence in relation to topography: a test of topographic indices as moisture indicators. Agricultural and Forest Meteorology 98-99, 325–340.

Rodriguez-Iturbe, I., 2000. Ecohydrology: A hydrologic perspective of climate-soil-vegetation dynamies. Water Resources Research 36 (1), 3–9.

Rowe, L. K., Pearce, A. J., O'Loughlin, C. L., 1994. Hydrology and related changes after harvesting native forest catchments and establishing pinus radiata plantations. Part 1. Introduction to study. Hydrological Processes 8 (3), 263–279.

Saffarpour, S., Western, A. W., Adams, R., McDonnell, J. J., 2016. Multiple runoff processes and multiple thresholds control agricultural runoff generation. Hydrology and Earth System Sciences Discussions, 1–31.

Sampson, D., Allen, H., 1999. Regional influences of soil available water-holding capacity and climate, and leaf area index on simulated loblolly pine productivity. Forest Ecology and Management 124 (1), 1–12.

Savenije, H. H. G., 2009. HESS Opinions "The art of hydrology". Hydrology and Earth System Sciences 13 (2), 157–161.

Savenije, H. H. G., 2010. Topography driven conceptual modelling (FLEX-topo). Hydrology and Earth System Sciences 14, 2681–2692.

Schenk, H. J., Jackson, R. B., 2002. Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. Journal of

Ecology 90 (3), 480–494.

Schymanski, S. J., Sivapalan, M., Roderick, M. L., Beringer, J., Hutley, L. B., 2008. An optimality-based model of the coupled soil moisture and root dynamics. Hydrology and Earth System Sciences 12 (3), 913–932.

Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrology and Earth System Sciences 4, 215–224.

Seibert, J., Rodhe, A., Bishop, K., 2003. Simulating interactions between saturated and unsaturated storage in a conceptual runoff model. Hydrological Processes 17 (2), 379–390.

Shamir, E., Imam, B., Morin, E., Gupta, H. V., Sorooshian, S., 2005. The role of hydrograph indices in parameter estimation of rainfall-runoff models. Hydrological Processes 19 (11), 2187–2207.

Simons, G., Bastiaanssen, W., Ngô, L., Hain, C., Anderson, M., Senay, G., 2016. Integrating Global Satellite-Derived Data Products as a Pre-Analysis for Hydrological Modelling Studies: A Case Study for the Red River Basin. Remote Sensing 8 (4), 279.

Sivandran, G., Bras, R. L., 2013. Dynamic root distributions in ecohydrological modeling: A case study at Walnut Gulch Experimental Watershed: Root Distributions in Ecohydrological Modeling. Water Resources Research 49 (6), 3292–3305.

Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. Water Resources Research 43 (1), W01415.

Sriwongsitanon, N., Gao, H., Savenije, H. H. G., Maekan, E., Saengsawang, S., Thianpopirug, S., 2016. Comparing the Normalized Difference Infrared Index (NDII) with root zone storage in a lumped conceptual model. Hydrology and Earth System Sciences 20 (8), 3361–3377.

Stephenson, D., 1981. Stormwater hydrology and drainage. No. v. 14 in Developments in water science. Elsevier Scientific Pub. Co., Amsterdam.

Tait, A., Henderson, R., Turner, R., Zheng, X., 2006. Thin plate smoothing spline interpolation of daily rainfall for new zealand using a climatological rainfall surface. Int. J. Climatol. 26 (14), 2097–2115.

Tait, A., Sturman, J., Clark, M., 2012. An assessment of the accuracy of interpolated daily rainfall for new zealand. Journal of Hydrology (NZ) 51 (1), 25–44.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using bayesian total error analysis. Water Resources Research 45, W00B14.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C., Brooks, P. D., Gupta, H., Huxman, T., 2009. Climate and vegetation water use efficiency at catchment scales. Hydrological Processes 23 (16), 2409–2414.

Tron, S., Perona, P., Gorla, L., Schwarz, M., Laio, F., Ridolfi, L., 2015. The signature of randomness in riparian plant root distributions. Geophysical Research Letters 42, 7098–7106.

Tu, M., 2006. Assessment of the effects of climate variability and land use change on the hydrology of the meuse river basin. Ph.D. thesis, Vrije Universiteit Amsterdam and UNESCO-IHE Institute for Water Education, Delft.

Uhlenbrook, S., Roser, S., Tilch, N., 2004. Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model. Journal of Hydrology 291 (3-4), 278–296.

Vaché, K. B., McDonnell, J. J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resources Research 42 (2), W02409.

van Deursen, W., 2000. Meuseflow 2.1 laagwaterstudies maastroomgebied. Tech. rep., RIZA project RI-2988A.

Vannametee, E., Karssenberg, D., Hendriks, M. R., Bierkens, M. F. P., 2013. Hortonian runoff closure relations for geomorphologic response units: evaluation against field data. Hydrology and Earth System Sciences 17 (7), 2981–3004.

Vico, G., Thompson, S. E., Manzoni, S., Molini, A., Albertson, J. D., Almeida-Cortez, J. S., Fay, P. A., Feng, X., Guswa, A. J., Liu, H., Wilson, T. G., Porporato, A., 2015. Climatic, ecophysiological, and phenological controls on plant ecohydrological strategies in seasonally dry ecosystems. Ecohydrology 8 (4), 660–681.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., Gupta, H. V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrological Processes 17 (2), 455–476.

Wagener, T., Montanari, A., 2011. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resources Research 47, W06301.

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., Savenije, H. H. G., 2016. Global root zone storage capacity from satellite-based evaporation. Hydrology and Earth System Sciences, 1459–1481.

Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J., Savenije, H. H. G., 2014. Contrasting roles of interception and transpiration in the hydrological cycle - Part 1: Temporal characteristics over land. Earth System Dynamics 5 (2), 441–469.

Webb, T., Wilson, A., 1995. A manual of land characteristics for evaluation of rural land. Vol. Landcare Research science series; ISSN 1 172-269X; no. 10. Manaaki Whenua Press, New Zealand.

Weerts, A. H., Winsemius, H. C., Verkade, J. S., 2011. Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). Hydrology and Earth System Sciences 15 (1), 255–265.

Weiler, M., Beven, K., 2015. Do we need a Community Hydrological Model? Water Resources Research 51, 7777–7784.

Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., Heynert, K., 2013. The Delft-FEWS flow forecasting system. Environmental Modelling & Software 40, 65–77.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., Xu, C.-Y., 2011. Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences 15 (7), 2205–2227.

Wöhling, T., Samaniego, L., Kumar, R., 2013. Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. Environmental Earth Sciences 69 (2), 453–468.

Willems, P., 2009. A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. Environmental Modelling & Software 24 (3), 311–321.

Winsemius, H. C., Savenije, H. H. G., Bastiaanssen, W. G. M., 2008. Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins? Hydrology and Earth System Sciences 12 (6), 1403–1413.

Winsemius, H. C., Schaefli, B., Montanari, A., Savenije, H. H. G., 2009. On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. Water Resources Research 45 (12), W12422.

Winter, T. C., 2001. The concept of hydrologic landscapes. Journal of the American Water Resources Association 37 (2), 335–349.

Woods, R., Hendrikx, J., Henderson, R., Tait, A., 2006. Estimating mean flow of new zealand rivers. Journal of Hydrology (NZ) 45 (2), 95–110.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., Pfister, L., 2015. Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments. Hydrological Processes 29 (12), 2731–2750.

Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources 30 (8), 1756–1774.

Yilmaz, K. K., Gupta, H. V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resources Research 44 (9), W09417.

Zehe, E., Ehret, U., Blume, T., Kleidon, A., Scherer, U., Westhoff, M., 2013. A thermodynamic approach to link self-organization, preferential flow and rainfall-runoff behaviour. Hydrology and Earth System Sciences 17 (11), 4297–4322.

Zelikova, T. J., Williams, D. G., Hoenigman, R., Blumenthal, D. M., Morgan, J. A., Pendall, E., 2015. Seasonality of soil moisture mediates responses of ecosystem phenology to elevated CO_2 and warming in a semi-arid grassland. J. Ecol. 103, 1119–1130.

Zhang, L., Dawes, W. R., Walker, G. R., 2001. Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resources Research 37 (3), 701–708.

Zhang, Z., Koren, V., Smith, M., Reed, S., Wang, D., 2004. Use of next generation weather radar data and basin disaggregation to improve continuous hydrograph simulations. Journal of Hydrologic Engineering 9 (2), 103–115.

ACKNOWLEDGEMENTS

CURRICULUM VITÆ

Tanja DE BOER-EUSER

28-01-1987   Born in Maassluis, the Netherlands.

EDUCATION

2005-2009   BSc Civil Engineering
                 Delft University of Technology

2009-2012   MSc Civil Engineering
                 Delft University of Technology

2016   PhD. Hydrology
                 Delft University of Technology
                 Thesis: The added value of distribution in rainfall-runoff
                                   models for the Meuse basin
                 Promotor: Prof. dr. ir. H.H.G. Savenije

EXPERIENCE

2008-2012   Student assistant
                 Water management department, Delft University of Technology

Jan-Apr 2015   Visiting researcher
                 National Institute of Water and Atmospheric Research, Christchurch, New Zealand

LIST OF PUBLICATIONS

de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G., and Willems, P.: *Looking beyond general metrics for model comparison – lessons from an international model intercomparison study*, Hydrology and Earth System Sciences Discussions, in review, (2016).

de Boer-Euser, T., McMillan, H.K., Hrachowitz, M., Winsemius, H.C. and Savenije, H.H.G., *The effect of soil and climate on root zone storage capacity*, Water Resources Research **52** (2016).

Euser, T., Hrachowitz, M., Winsemius, H.C. and Savenije, H.H.G., *The effect of forcing and landscape distribution on performance and consistency of model structures*, Hydrological Processes **29**, 3727–3743 (2015).

Hrachowitz, M., Fovet, O., Ruiz, L., **Euser, T.**, Gharari, S., Nijzink, R., Freer, J., Savenije, H.H.G. and Gascuel-Odoux, C., *Process consistency in models: The importance of system signatures, expert knowledge, and process complexity*, Water Resources Research **50**, 7445–7469 (2014).

Euser, T., Luxemburg, W.M.J., Everson, C.S., Mengistu, M.G., Clulow, A.D., and Bastiaanssen, W.G.M., *A new method to measure Bowen ratios using high-resolution vertical dry and wet bulb temperature profiles*, Hydrology and Earth System Science **18**, 2021-2032 (2014).

Euser, T., Winsemius, H.C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H.H.G., *A framework to assess the realism of model structures using hydrological signatures*, Hydrology and Earth System Science **17**, 1893-1912 (2013).

Conceptual rainfall-runoff models are useful tools for predicting river discharges, both in cases of floods and droughts. They can also help to increase our understanding of catchment behaviour. Adding spatial information, such as variability in meteorological forcing or in runoff processes, could further improve these models, but also increases their complexity. Therefore, it is essential to carefully select the spatial information that is most useful and establish methods to assess their added value. This thesis focusses on these issues using a case study in the Meuse and Ourthe catchments.