UNIVERSITY OF BELGRADE

FACULTY OF CIVIL ENGINEERING

Andrijana Dj. Todorović

IMPACT OF CALIBRATION PERIOD ON PARAMETER ESTIMATES IN THE CONCEPTUAL HYDROLOGIC MODELS OF VARIOUS STRUCTURES

Doctoral Dissertation

Belgrade, 2015

УНИВЕРЗИТЕТ У БЕОГРАДУ

ГРАЂЕВИНСКИ ФАКУЛТЕТ

Андријана Ђ. Тодоровић

УТИЦАЈ ПЕРИОДА КАЛИБРАЦИЈЕ НА ОЦЕНЕ ПАРАМЕТАРА КОНЦЕПТУАЛНИХ ХИДРОЛОШКИХ МОДЕЛА РАЗЛИЧИТИХ СТРУКТУРА

Докторска дисертација

Београд, 2015.

INFORMATION ON THE SUPERVISORS AND THE EXAMINING COMMITTEE

Supervisors:

Associate Professor Dr Jasna Plavšić, University of Belgrade, Faculty of Civil Engineering

Associate Professor Dr Miloš Stanić, University of Belgrade, Faculty of Civil Engineering

Examining Committee members:

Associate Professor Dr Jovan Despotović, University of Belgrade, Faculty of Civil Engineering

Assistant Professor Dr Nemanja Branisavljević, University of Belgrade, Faculty of Civil Engineering

Assistant Professor Dr Borislava Blagojević, University of Niš, Faculty of Civil Engineering and Architecture

Date of Thesis defence:_____

ПОДАЦИ О МЕНТОРУ И ЧЛАНОВИМА КОМИСИЈЕ

Ментори:

в. проф. др Јасна Плавшић, дипл. грађ. инж., Грађевински факултет Универзитета у Београду

в. проф. др Милош Станић, дипл. грађ. инж., Грађевински факултет Универзитета у Београду

Чланови комисије:

в. проф. др Јован Деспотовић, дипл. грађ. инж., Грађевински факултет Универзитета у Београду

доц. др Немања Бранисављевић, дипл. грађ. инж., Грађевински факултет Универзитета у Београду

доц. др Борислава Благојевић, дипл. грађ. инж., Грађевинско-архитектонски факултет Универзитета у Нишу

Датум одбране докторске дисертације:_____

ЗАХВАЛНИЦА

Захваљујем се менторима в. проф. др Јасни Плавшић и в. проф. др Милошу Станићу на саветима и помоћи при изради ове тезе.

Посебно се захваљујем в. проф. др Милошу Станићу и колеги Жељку Василићу, који су прилагодили претходну верзију модела 3DNet-Catch за потребе овог истраживања.

Такође се захваљујем и члановима комисије: доц. др Борислави Благојевић, в. проф. др Јовану Деспотовићу и доц. др Немањи Бранисављевићу који су својим коментарима учинили ову тезу бољом.

Подаци за симулације добијени су од Републичког Хидрометеоролошког Завода. Захваљујем се и проф. Весни Ристић са Рударско-геолошког факултета и Николи Златановићу који су ми уступили податке о сливовима, као и Пеђи за помоћ у анализи података.

Велику захвалност дугујем и свим својим колегама студентима докторских студија, посебно Ањи и Буду, и „цимеркама“ Љиљи и Зорани на њиховој подршци. Захваљујем се и доц. др Драгутину Павловићу, који је преузео део мојих обавеза у настави.

Посебну, највећу захвалност дугујем мојој породици, на њиховом стрпљењу и безусловној подршци. Хвала вам!

У Београду, 2015.

Аутор

Doctoral Dissertation title: IMPACT OF THE CALIBRATION PERIOD ON PARAMETER ESTIMATES IN CONCEPTUAL HYDROLOGIC MODELS

ABSTRACT

Conceptual hydrologic models are commonly applied for flow forecasting, estimation of design flows and assessment of climate change impact on water resources. Therefore, reliability of hydrologic simulations obtained by employing these models is crucial. However, these simulations are fraught with uncertainties, which stem, *inter alia*, from parameter estimates. The parameter estimates are affected by data errors, objective functions and optimisation algorithm employed for model calibration, but also by properties of the calibration period. Namely, model calibration over different periods may result in quite different parameter estimates because parameter optimality does not hold outside the calibration period. This temporal variability of optimal parameter estimates yields deterioration in model performance outside the calibration period. Therefore, variability of optimal parameter estimates is major issue when it comes to application of hydrologic models, because these models are primarily used for runoff simulations outside the calibration period.

In this Thesis temporal variability in parameters of the 3DNet-Catch model is analysed. The AMALGAM algorithm, aimed at multi-objective optimisation, is applied for model calibration. The model is calibrated in dynamic manner, over all 1- to 25-year long calibration periods, with one water year prior to every calibration aimed at model warm-up. Prior ranges of the parameters and settings for the optimisation algorithms (e.g. population size, mutation probability, etc.) are kept constant through all simulations for given catchment. The analysis of temporal variability in model parameters is based on the non-dominated, or Pareto-optimal sets, which are selected subsequent to the optimisation of the initially sampled population of parameter sets. Impact of combination of objective functions used for model calibration and model structural complexity on temporal variability in the Pareto-optimal parameters is also examined in this research. To isolate temporal variability in parameters from anthropogenic effects (e.g. urbanisation or river engineering works) three catchments that have not undergone human-induced changes are considered in this research: the Kolubara River catchment upstream of the Slovac

stream gauge, the Toplica River catchment upstream of the Doljevac stream gauge and the Mlava River catchment upstream of the Veliko Selo stream gauge.

The results reveal considerable temporal variability in the Pareto-optimal parameters. The variability is somewhat lower in the parameters to which the model is sensitive, although strong correlation between parameter sensitivity and temporal variability has not been detected. Also, correlation between parameter estimates and hydro-meteorological characteristics of the calibration period is shown rather weak. Temporal variability in the Pareto-optimal parameter persists regardless of the objective functions used for model calibration, although an increase in the number of objective functions appears to lead to more consistent parameter estimates and better model performance. Comparison among different versions of the 3DNet-Catch model suggests application of spatially-distributed parameters and reduction in number of free model parameters (parameters that are to be estimated in the model calibration). Spatial distribution of the parameters has to be accompanied by the temporal data resolution, whereas reduction in the number of free model parameters has to be supported by the results of the sensitivity analysis. Additionally, empirical cumulative distribution functions derived from all Pareto-optimal parameters obtained over all calibration periods are shown different from the distribution functions obtained from the Pareto-optimal parameters optimised over the full record period. This result indicates that dynamic model calibration enables extraction of more information form the observations available, and assumedly it could result in more robust parameter estimates and consequently to more reliable simulations outside the calibration period. As for model performance, the results indicate that overall water balance can be properly simulated by the model regardless of the calibration period, while model's ability to reproduce dynamic in catchment response exponentially decreases with the length of the calibration period.

A comprehensive analysis of temporal variability in the Pareto-optimal parameters and model performance is presented in this Thesis, along with the recommendations for model calibration and development in order to obtain more consistent parameter estimates and the model performance. Recommendations for further research are also presented.

Key words: conceptual hydrologic models; the 3DNet-Catch; multi-objective calibration; the AMALGAM; calibration period; parameter temporal variability; consistency in parameter estimates.

Research area: Civil Engineering

Specific research area: Hydrology

UDK: 624/556(043.3)

Наслов докторске дисертације: УТИЦАЈ ПЕРИОДА КАЛИБРАЦИЈЕ НА ОЦЕНЕ ПАРАМЕТАРА КОНЦЕПТУАЛНИХ ХИДРОЛОШКИХ МОДЕЛА РАЗЛИЧИТИХ СТРУКТУРА

РЕЗИМЕ

Концептуални хидролошки модели су нашли широку примену у изради хидролошких прогноза и предикција, и у анализи утицаја климатских промена на водне ресурсе. Стога је поузданост симулација добијених применом ових модела веома важна. Међутим, у хидролошким симулацијама постоје неизвесности, које потичу и од оцена параметара модела. На оцене параметара модела утичу грешке у подацима, избор критеријумских функција и оптимизационог алгоритма, али и карактеристике калибрационог периода. Наиме, калибрација модела током различитих периода даће различите оцене параметара, зато што параметри који су оптимални током једног периода не морају бити оптимални изван њега. Последица варијабилности оптималних параметара у времену је и лошија ефикасност модела тј. мање поуздане симулације ван калибрационог периода. Имајући у виду да се хидролошки модели користе за хидролошке симулације ван калибрационог периода, за њихову примену кључно је изучавање променљивости оптималних параметара модела током времена.

У овој дисертацији анализиран је утицај калибрационог периода на оцене параметара хидролошког модела 3DNet-Catch. За калибрацију модела коришћен је савремени алгоритам за вишекритеријумску оптимизацију AMALGAM, који представља комбинацију неколико глобалних оптимизационих алгоритама. Хидролошки модел је калибрисан на свим периодима дужине од једне до 25 узастопних хидролошких година, уз једну хидролошку годину намењену „загревању“ модела. Овакав приступ је назван „динамичка“ калибрација модела. Почетни опсези параметара, као и подешавања за оптимизациони алгоритам (нпр. број чланова популације, вероватноћа мутације и др.) исти су за све калибрационе периоде за разматрани слив. Након оптимизације параметара издвојена су тзв. међусобно недоминантна решења (Парето оптимални скупови параметара или скупови са Парето фронта), на основу којих је вршена анализа променљивости

оптималних параметара у времену. У овом истраживању анализиран је утицај избора критеријумских функција и комплексности структуре модела на варијабилност Парето оптималних параметара у времену. Како би се искључио утицај вештачких промена на сливу на оцене параметара (нпр. урбанизација или формирање акумулација), овај приступ је примењен на три природна слива у Србији: слив реке Колубаре до водомерне станице (в.с.) Словац, слив реке Топлице до в.с. Дољевац и слив реке Млаве до в.с. Велико Село.

Резултати су показали изразиту варијабилност Парето оптималних параметара. Осетљиви параметри (односно они који знатно утичу на ефикасност модела) нешто мање варирају, мада зависност између осетљивости параметара и њихове варијабилности у времену није утврђена. Такође, корелисаност између оцена параметара и хидро-метеоролошких карактеристика датог калибрационог периода је веома слаба. Варијабилност параметара у времену постоји за све анализиране комбинације критеријумских функција. Међутим, свеобухватна анализа резултата указује да калибрација модела са више критеријумских функција даје конзистентније оцене параметара и ефикасније моделе. Поређењем различитих структура модела 3DNet-Catch предност се даје структурама са просторно дистрибуираним параметрима и са мањим бројем параметара модела. Наравно, смањење броја параметара који се калибришу мора се оправдати анализом осетљивости, а просторна резолуција у дистрибуираном моделу временском резолуцијом улазних података. Један од резултата овог истраживања су и емпиријске расподеле Парето оптималних параметара добијених из свих калибрационих периода. Ове расподеле могу знатно да одступају од расподела Парето оптималних параметара који су добијени калибрацијом модела за цео период, што указује на могућу бољу искоришћеност информација уколико се модел динамички калибрише. Резултати су такође показали да модел може да репродукује биланс вода на сливу без обзира на калибрациони период, док слагање симулираних и осмотрених хидрограма експоненцијално опада са дужином калибрационог периода.

У овој дисертацији приказана је свеобухватна анализа временске променљивости оцена параметара које су добијене вишекритеријумском калибрацијом модела, као и анализа ефикасности модела. Допринос овог истраживања су препоруке за

вишекритеријумску калибрацију и развој модела у циљу постизања што конзистентнијих оцена параметара које би водиле поузданијим хидролошким симулацијама ван калибрационог периода. На крају су дати и предлози за даља истраживања.

Кључне речи: концептуални хидролошки модели; 3DNet-Catch; вишекритеријумска калибрација; AMALGAM; калибрациони период; временска променљивост оцена параметара; конзистентност оцене параметара.

Научна област: грађевинарство

Ужа научна област: хидрологија

UDK: 624/556(043.3)

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Hydrologic cycle and runoff generation mechanisms

Water circulation near the surface of the Earth[1], i.e. hydrologic cycle, consists of many processes, which are illustrated in Figure 1.

Water evaporates form the surface water bodies (oceans, lakes, etc.), soil and canopy. Evaporation is driven by the Sun radiation (Chow et al., 1988; Shaw, 2005). The water vapour condenses in the atmosphere, and precipitates on the land and surface water bodies. The precipitation that reaches the land is partly intercepted by the vegetative cover (up to 30% of precipitation may be intercepted, Beven, 2001b), one part comprises depression storage, while the remaining amount of infiltrate into the unsaturated soil layer, or becomes an overland flow. Interception capacity changes over the growing season, and it also depends on the vegetation type, precipitation and wind conditions, and vegetation age (Jovanović & Radić, 1990; Fenicia et al., 2009). Partition between infiltration and overland flow depends on the land use type, soil type and soil moisture conditions. For example, infiltration rates increase after prolonged dry periods (due to soil crusting, which enables preferential macropore infiltration) and after freezing of dry soil, as opposed to saturated soils, frozen wet soil or crusted soil, as shown in Figure 2 (Beven, 2001b; Stähli & Hayashi, 2015). Preferential infiltration through macropores is higher than the infiltration through soil matrix (Weiler et al., 2005). Precipitation onto the saturated soil results in quick surface runoff (saturated overland flow). Soil saturation may occur if there is a soil layer of low permeability (e.g. with high clay content) underneath the surface soil layer (perched subsurface stormflow). Infiltration also depends on the precipitation intensity: high rainfall intensities that exceed soil infiltration capacity, results in prompt overland flow (Hortonian overland flow, infiltration-excess runoff generation) (Leibundgut et al., 2001). Runoff dependence on the soil moisture and rainfall intensity is illustrated in Figure 3.

Part of the infiltrated water percolates deeper (groundwater recharge) and groundwater eventually seeps into effluent streams (subsurface flow). Considering catchment area,

[1] 15 km up in the atmosphere and 1 km down in the lithosphere (Chow et al., 1988).

total runoff at the catchment outlet is a sum of the overland (surface, direct, fast) and subsurface runoff. Amount of water that comprises the overland flow is determined by the precipitation within a catchment area, while the subsurface flow may originate beyond the waterdivide.

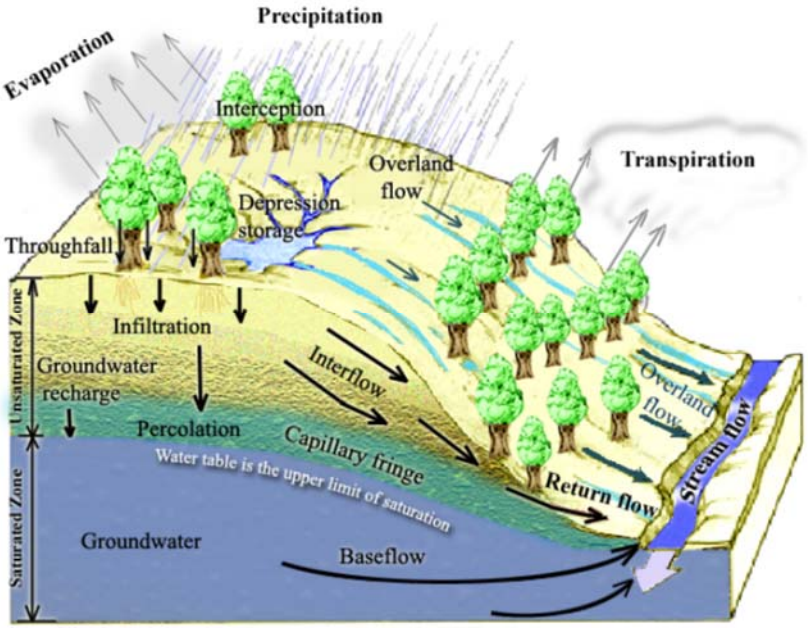


Figure 1. Hydrologic cycle (Tarboton, 2003).

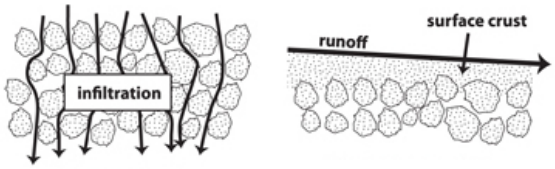


Figure 2. Soil crusting impact on the infiltration rate and surface runoff.

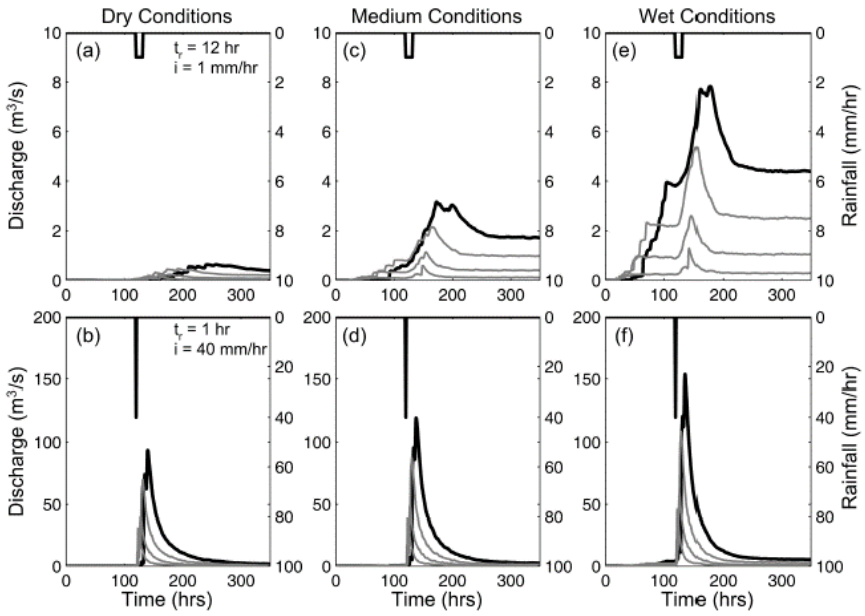


Figure 3. Runoff dependence on the soil moisture conditions and rainfall intensity (Vivoni et al., 2007).

1.2. Rainfall-runoff modelling

Hydrologic (rainfall-runoff, deterministic) models are convenient tools to provide hydrologic predictions. As such, they have wide range of applications in design of water infrastructure and water resources management (e.g. water supply, hydropower generation, irrigation, flood forecasting, etc.) (Blasone, 2007; Pechlivanidis et al., 2011).

There are many "types" of hydrologic models according to numerous classifications. Essentially hydrologic models enable estimation of flows, soil moisture and other hydrologic variables over some period, for the given input data (usually precipitation and temperature). According to Beven (2001b), development of a hydrologic model consists of the following stages:

(1) *Perceptual model*: identification of the hydrological processes that are in control of runoff generation in a particular catchment.

(2) *Conceptual model*: mathematical description of the identified processes. The results of this stage often includes partial differential equations.

(3) *Numerical model*: numerical approximation of the equations of the conceptual model.

(4) *Procedural model* is a code to be run on a computer, and it is based on the numerical model. It should be verified to assure that the numerical model equations are properly reproduced within the code (Beven & Young, 2013).

(5) *Estimation of model parameters* that do not represent directly measurable variables is necessary, to achieve a satisfactory degree of agreement between the simulated and corresponding observed variables according to some goodness-of-fit measure(s). The parameters that have to be estimated are referred to as the free model parameters. Methods for parameter estimations are elaborated in chapter 1.3.

(6) *Model evaluation* (*validation*): a calibrated model should be applied for runoff simulations over an independent period to test its robustness.

Hydrologic models are commonly classified according to their structure and the manner of treating spatial variability of the catchment properties (Chow et al., 1988).

Sivapalan et al., (2003) distinguish between upward (bottom-up) and downward (top-down) approach. The former implies a perceptual model based on the processes identified at small scale (e.g. hillslope). These processes are scaled-up, considering possible interactions among the processes at the catchment scale (Beven, 2001a). In the latter approach the processes at the catchment scale are inferred from the observed catchment responses.

The bottom-up approach is adopted in the physically-based (or white-box) models. These models are usually comprised of partial differential equations describing processes of runoff generation. These equations are applied to the computational grid. An issue about this model type is the scale-up of the processes itself. Namely, the processes that are important on small scales do not necessarily have to be important at the catchment scale (for example, heterogeneity of the soil properties may be averaged out at the catchment scale), and vice-versa: key processes at the catchment scale may not be captured at the hillslope scale (e.g. large-scale lateral subsurface pipe flow) (Sivapalan et al., 2003). The parameters of these models carry a physical connotation (such as the saturated hydraulic conductivity) and theoretically could be inferred *a priori* (chapter 1.3).

The conceptual (or grey-box) models are based on the top-down approach. These models imply approximating the runoff cycle components by the reservoirs (e.g. canopy or soil reservoirs). The mass conservation equations holds for the reservoirs, whereas the fluxes among the reservoirs are approximated by empirical relations (e.g. linear or non-linear reservoir equations) (Shaw, 2005). There are two implications of such a setup: (1) conceptual models are less data demanding than the physically-based models, and (2) some of the model parameters do not represent physically meaningful, measurable variables (e.g. reservoir coefficient or non-linearity coefficient of a non-liner reservoir), so they have to be estimated from the observations (Ebel & Loague, 2006; Todini, 2007; Pechlivanidis et al., 2011). These models may vary in complexity to considerable extent, but their complexity should be justified by the available data (Sivapalan et al., 2003; Wagener et al., 2003; Ebel & Loague, 2006). Clark et al. (2008) developed the FUSE methodology, which enables development of new model structures, by combining the building blocks of the existing ones (e.g. modules for simulation of *ET*, saturated soil zone storage, overlandflow, etc.). Box and Jenkins (1970) introduced principle of parsimony, meaning that simpler models with fewer parameters are preferred over the more complex ones. It was demonstrated by van Esse et al. (2013) that the models with simple structure may perform as satisfactorily as more complex models. Thorough review on the conceptual models is presented by Daniel et al. (2011).

Beven and Young (2013) refer to both physically-based and conceptual models as "deductive", since their model structure is defined prior to model application.

Further, there are data-driven (black-box or empirical) models that do not rely on description of the runoff generation processes. These models are based on the functional relationship between the input (i.e. meteorological forcing) and the output (e.g. flow), where these relationships do not carry any physical meaning. Since the structure of these models is identified based on the observed data, Beven and Young (2013) refer to these models as "inductive". For example, in the neural network models the number of layers, number of neurons and type of transfer functions are determined from data.

Regarding treatment of spatial variability, the parameter sets in the *lumped* models apply to an entire catchment, and the model forcing datasets (precipitation and temperature) are spatially averaged (e.g. Chow et al., 1988; Shaw, 2005). Consequently, these models

result in the spatially averaged values of the simulated variables and in the integrated catchment response (Yilmaz et al., 2010). The model parameters are usually estimated only against the flows observed at the catchment outlet. The models that include a single parameter set for an entire catchment but spatially distributed input forcing have been known earlier as the semi-distributed models, but more recently are termed the semi-lumped models (Ajami et al., 2004; Khakbaz et al., 2012; Schaefli et al., 2014). In the fully-distributed models simulations are performed on a grid (which may be composed of irregular cells), resulting in spatially distributed simulated variables. Every computational cell may be assigned a different parameter set, and the observations at many interior points in a catchment may be used for parameter estimation: for example, nested stream gauges, groundwater levels across the catchment and maps of soil moisture from remote sensing sources (Ajami et al., 2004; Ivanov et al., 2004). Since this approach results in large number of parameters to be estimated, the regularisation techniques are employed for calibration of distributed models (chapter 1.3.5). Recalling the definition of physically-based models, these models are at the same time spatially distributed.

As a compromise between the lumped and the fully-distributed models, semi-distributed models emerged. These models imply model application to the individual sub-catchments, where different input data and parameter sets are appointed to each sub-catchment (Schumann, 1993). The semi-distributed models are less data demanding compared to the fully-distributed ones. They also enable estimating the parameters against the flows observed at the nested stream gauges (Ajami et al., 2004; Khakbaz et al., 2012).

Hydrologic models can be applied for simulations of isolated events or for continuous simulations (e.g. Pechlivanidis et al., 2011). The former implies simulation of hydrographs caused by a single precipitation event, while the latter includes simulation of direct runoff over precipitation events and simulation of baseflows between them. Models aimed at continuous hydrologic simulations involve more processes (for example, evapotranspiration, which is commonly neglected in the event-based modelling). Event based modelling is hampered by the subjective nature of streamflow partitioning into the direct runoff and baseflow (Boughton & Droop, 2003).

1.3. Estimation of hydrologic model parameters

Model parameters enable a model of predefined structure to be adjusted for simulations of runoff from a particular catchment. The parameters can be estimated *a priori* or *a posteriori*: *a priori* parameter assessment is based on the catchment properties, while *a posteriori* estimation implies parameter conditioning on the observations (Beven 2001a; Blasone 2007; Sorooshian et al., 2008; Yilmaz et al., 2010). Only directly measurable parameters (so called "physical parameters", such as catchment area, share of urbanised area in the total catchment area, slope of a stream section, etc.) can be estimated *a priori*. However, majority of the parameters, especially those of the conceptual models, are not directly measurable due to simplification of the highly complex runoff generation processes (e.g. linear outflow equations) and due to spatial aggregation of the processes (e.g. Gupta et al., 2005; Renard et al., 2010). These parameters are named "process" or "conceptual" parameters, and they are estimated indirectly (e.g. Vrugt et al., 2003; Gupta et al., 2005; Vrugt et al., 2006; Blasone 2007).

1.3.1. Approaches to parameter estimation

There are two approaches to parameter estimation (Figure 4): the frequentists and the Bayesian approach (Daniel et al., 2011). In the former approach the parameters are assumed to have optimal values that result in minimum discrepancy between the observed and simulated variables (point estimates of the parameters). This approach is referred to as "model calibration", "parameter optimisation" or "inverse modelling" (Blasone 2007). The latter approach is grounded in the Bayes theorem: the parameters are considered random variables with their posterior probability distribution function (pdf), inferred from the prior distribution and likelihood of the simulated variables for the given observations (Montgomery & Runger, 2003). Both approaches rely on the maximum likelihood theory. The objective functions for parameter optimisation stem from the maximum likelihood estimator (Sorooshian et al., 1983), while the likelihood function in the Bayesian approach is a generalisation of the maximum likelihood estimation method (Kottegoda & Rosso, 2008). Razavi et al. (2010) entitled the former approach "optimisation-based" calibration, and the latter one – "uncertainty-based" calibration, which suggests application of these methods for parameter uncertainty assessment.

Both approaches require flow observations in a catchment. For ungauged catchments model parameters are inferred from corresponding parameter estimates at adjacent gauged catchments and their properties. Namely, empirical relations between the parameter estimates and catchments' characteristics are established and applied to estimate values of the parameters for the ungauged catchments; this approach is known as the regional approach (e.g. Gupta et al., 2005; Yilmaz et al., 2010).

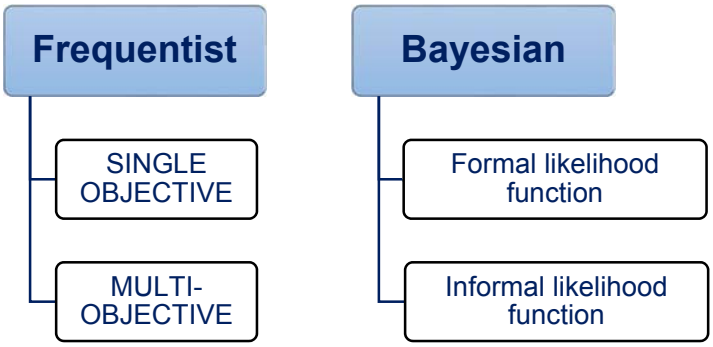


Figure 4. Approaches to estimation of hydrologic model parameters.

Optimisation-based calibration

Optimisation-based calibration implies adjustment of the parameter values to achieve the best possible fit between the simulated and observed variables over some period (parameter optimisation) (e.g. Gupta et al., 2005; Yilmaz et al., 2010). The models are commonly calibrated against the observed flows, although groundwater levels or conservative tracer concentrations may be used as well (e.g. Leibundgut et al., 2001; Seibert 2003).

Calibration is referred to as "inverse modelling" because the input and output are known, while the model parameters are to be inferred (Sorooshian et al., 2008). The initial conditions (e.g. soil wetness or canopy reservoir storage) are assumed. Impact of the erroneous assumptions about the initial states is mitigated by excluding the first part of the simulation period (warm-up or spin-up period) from process of parameter assessment – calculation of the objective function(s).

Agreement between the simulated and observed variables is quantified by means of the goodness-of-fit measures (model efficiency or performance measure), which in context

of the parameter optimisation become objective functions of the optimisation algorithms (e.g. Yilmaz et al., 2008). A goodness-of-fit measure is basically an aggregate statistic of the residuals (differences between the simulated and observed variables), $e(t)$:

$$e(t) = y(t) - \hat{y}(t) \tag{1.3.1}$$

$$f(\theta) = f(e(t)) \tag{1.3.2}$$

where $y(t)$ denotes observed, $\hat{y}(t)$ simulated hydrologic variable (e.g. flow) and $f(\theta)$ goodness-of-fit measure. One of the commonly used objective function is Mean square error, *MSE*:

$$MSE = \mathrm{E}\left[\left(y - \hat{y}\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2 \tag{1.3.3}$$

In the model calibration, *MSE* is to be minimised with respect to the model parameters θ:

$$\min_{\theta \in \Theta} MSE\left(y, \hat{y} \middle| x, \theta\right) \tag{1.3.4}$$

where Θ denotes plausible ranges of the parameters θ. Simulated variable $\hat{y}(t)$ is conditioned on the model input x and the parameter set θ.

An overview of the objective functions most frequently used for hydrologic model calibration is given in chapter 1.3.3.

Calibration can be performed manually or automatically by employing an optimisation algorithm.

– ***Manual ("trial and error") calibration*** is performed by an expert hydrologist and, as such, inevitably involves subjectivity (e.g. Vrugt et al., 2003; Yilmaz et al., 2010). Agreement between the simulated and observed variables is estimated subjectively, based on visual inspection of the results (e.g. hydrographs) (Boyle et al., 2000). Also, the values of optimised parameters heavily depend on the hydrologist's experience, thus two hydrologists may come up with quite different parameter estimates (e.g. Vrugt et al., 2003; Yilmaz et al., 2010). This procedure is time and labour consuming, and it becomes virtually inapplicable to highly parameterised, complex models (e.g. Lindstrom 1997;

Gupta et al., 2005; Yilmaz et al., 2008). For these reasons, manual calibration has been replaced by automatic one (Liu & Gupta, 2007).

– ***Automatic calibration.*** In addition to calibration data, an automatic model calibration requires prior ranges or distributions of model parameters, objective function(s) (chapter 1.3.3), an optimisation algorithm (chapter 1.3.4) and the convergence criteria (Figure 5). Parameter values are sought within the pre-specified ranges aiming at minimisation (or maximisation) of particular objective function(s). The optimisation procedure lasts until convergence criterion is fulfilled. Convergence criteria may be expressed as the maximum number of iterations, or the minimum relative change of objective function(s) or parameter estimates between consecutive iterations (Madsen 2003; Blasone et al., 2007; Blasone 2007). Two types of automatic calibration can be distinguished depending on whether the model is calibrated with respect to one or more objective functions: single- objective calibration and multi-objective one. Since the multi-objective calibration is employed in this research, particulars of this approach are discussed in detail in chapter 1.3.2.

An automatic model calibration can result in the unrealistic parameter estimates because the parameters are adjusted to obtain the best-fit model, regardless of their hydrologic connotation (Yilmaz et al., 2008). This problem can be approached through the plausible prior parameter ranges. For example, the SCS Curve number *CN* can take values from approximately zero to one hundred. Yet, the initial range of this parameter should be reduced based on the land use and soil types in the catchment (for example, *CN* equal to 95 for a natural catchment is questionable).

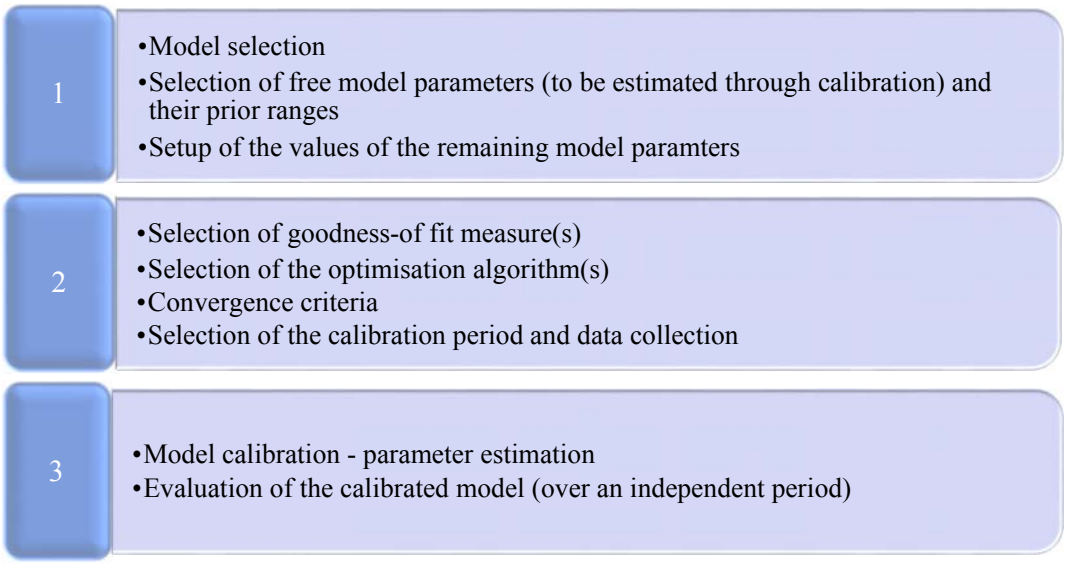


Figure 5. Overview of the steps of automatic model calibration procedure.

After the parameters are estimated, the obtained model is tested over an independent period. Poor model performance outside the calibration period generally indicates overcalibration, i.e. a large prediction uncertainty (e.g. Schoups et al., 2008; Remesan and Mathew 2013). It is a consequence of noise in the observed data, low information content of the observations or high model complexity, which is not supported by the available data (Andréassian et al., 2012; Remesan and Mathew 2013).

Uncertainty-based calibration (Bayesian approach)

Uncertainty-based calibration aims at identification of the parameter posterior pdf, based on the prior distribution and the likelihood function (e.g. Liu & Gupta, 2007; Kottegoda and Rosso 2008). According to Bayes equation, the parameter posterior pdf, $p(\theta \mid D)$, is estimated as following (e.g. Kuczera & Parent, 1998; Renard et al., 2010):

$$p(\theta|D) = \frac{p(D|\theta)\cdot p(\theta)}{\int p(\theta|D)\cdot p(\theta)\,\mathrm{d}\theta} \propto p(D|\theta)\cdot p(\theta) \qquad (1.3.5)$$

where D stands for the data set, $p(D \mid \theta)$ is the likelihood, and $p(\theta)$ is the parameter prior distribution. The prior distribution enables a hydrologist to incorporate their knowledge

on the parameter values into the calibration procedure by imposing a particular distribution. The parameter prior distribution is commonly assumed to be uniform.

According to Todini (2007), certainty about a parameter is represented by the Dirac delta function at the parameter value, total uncertainty is represented by the [-∞, +∞] interval, and the parameter uncertainty (i.e. partial knowledge about the parameter value) is represented by pdf. A peakier pdf indicates smaller variability and less uncertainty. Hence, the parameter posterior distribution, $p(\theta \mid D)$, enables quantification of the parameter uncertainty (Kuczera & Parent, 1998) and this approach is therefore referred to as the uncertainty-based calibration (Razavi et al., 2010). Prior parameter distribution is updated with the new observations (Engeland et al., 2006). There are two types of these methods, depending on the likelihood function: formal and informal ones (Engeland et al., 2006; Vrugt et al., 2008; Jin et al., 2010; Sadegh and Vrugt 2013; Shafii et al., 2014).

– In the ***formal Bayesian methods*** the likelihood functions are based on the stochastic properties of the residuals (e.g. for autocorrelated residuals with homoskedastic variance[2], an AR(1) error model can be used to derive the likelihood function). Additionally, these functions can include various sources of uncertainty: for example, multiplicative error term can be used to correct input rainfall data (Kuczera et al., 2006; McMillan et al., 2011). Despite being statistically sound, these likelihood functions rely on strong assumptions on the residual stochastic properties (Vrugt et al., 2008). The SCEM (*Shuffled Complex Evolution Metropolis algorithm*) and DREAM (*DiffeRential Evolution Adaptive Metropolis*) are some formal Bayesian methods commonly employed in hydrological modelling (Vrugt, 2003; Vrugt et al., 2008).

– Application of the ***informal Bayesian methods*** does not rely on derivation of the likelihood functions that reflect various sources of uncertainty. For example, in the most frequently used informal Bayesian method, the Generalised Likelihood Uncertainty estimation (GLUE[3]) introduced by Beven and Binly (1992), the likelihood functions are

[2] Homoskedasticity is defined in relation to regression analysis as independence of the residuals on the magnitude of the independent variable (as opposed to heteroskedasticity). Homoskedasticity or homogeneity of variance means that all sequences of the series have the same variance.

[3] The GLUE method is based on the Regional Sensitivity Analysis (RSA) by Spear and Hornberger (1980).

based on the model performance measures (less formal likelihood). Therefore, no strong assumptions on the stochastic properties of the residuals are required. However, all sources of uncertainty are lumped together into the parameter uncertainty, resulting in smoother posterior parameter pdfs (Vrugt et al., 2008). Montanari (2005), Mantovan and Todini (2006) and Todini (2007) argued about disadvantages of the GLUE method. First, they pointed out the subjectivity in selection of the behavioural (good, acceptable) parameter sets that could significantly affect the estimated parameter uncertainty. Second, they pointed out the improper likelihood functions, which do not ensure equivalence between batch and sequential learning incoherency in learning. However, due to simplicity of the GLUE method, it has been widely applied and reported in the literature (e.g. Campling et al., 2002; Choi & Beven, 2007; Coron et al., 2012; de Vos et al., 2010; Fenicia et al., 2008; Muñoz et al., 2014).

Comparison between the formal (DREAM) and informal (GLUE) Bayesian methods is reported in several papers (e.g. Vrugt et al., 2008; Jin et al., 2010), while Nott et al. (2012) and Sadegh and Vrugt (2013) compared GLUE with the approximate Bayesian computation (ABC) methods. Vrugt et al. (2008) demonstrated that the formal Bayesian approach (DREAM) results in considerably peakier posterior pdfs of the HYMOD parameters, higher percentage of the observations encompassed by ensemble band, narrower spread of the band and better overall performance over calibration and evaluation periods, than the GLUE method. Sadegh and Vrugt (2013) argued that the GLUE is type of the ABC method, what is suported by similar results of these two approaches.

1.3.2. Issues concerning parameter estimation

Hydrologic model calibration is rather difficult task, regardless of the approach to parameter estimation. Goal is to obtain robust parameter estimates that result in strong performance over the calibration and evaluation periods (parameter transferability in time) (Krauße et al., 2012). Some issues about parameter estimation in hydrological modelling are briefly discussed in this chapter.

Stochastic nature of the residuals

Objective functions are derived from the maximum likelihood estimator, with certain assumptions on the stochastic properties of the residuals. For example, if the residuals are assumed to be independent and normally distributed with constant variance (homoskedasticity) and if the sample size is sufficiently large, the maximum likelihood estimator reduces to the mean squared error (*MSE*) (e.g. Sorooshian et al., 1983; Gupta et al., 2005; Schoups et al., 2008). Instead of *MSE,* its root square value is frequently used (*RMSE*), which is expressed in the same unit as the simulated variable. *RMSE* is sensitive to extreme values (i.e. it primarily depends on the model efficiency in the high flow domain). Therefore, modified versions of this objective function with transformed flows are used for model calibration (Fenicia et al., 2007). Other objective functions based on the squared difference between simulated and observed flows are also introduced and applied, depending primarily on the modelling purpose (chapter 1.3.3).

However, when it comes to hydrologic modelling, such the assumptions about the residuals (normal distribution, randomness and homoskedasticity) are not usually valid (e.g. Gupta et al., 2005; Schoups and Vrugt 2010). Ignorance of the stochastic nature of the residuals leads to the parameter estimates that do not result in optimal model performance in the calibration and evaluation periods and significantly vary with the calibration period (Sorooshian et al., 1983). Sorooshian et al. (1983) compared the results obtained with *HMLE* (Maximum Likelihood Estimator for Heteroskedastic Error Case) and with *AMLE* (Maximum Likelihood Estimator for Autocorrelated Error Case). Parameter estimates obtained using *HMLE* resulted in higher model efficiency, indicating that heteroskedasticity[4] in the residuals is more important to be recognised in an objective function than autocorrelation. In addition to *HMLE*, various monotonic (e.g. logarithmic, root, reciprocal, Box-Cox) transformations can be applied to the variables to stabilise heteroskedasticity. The objective functions are calculated from the transformed series (Sorooshian et al., 1983; Yilmaz et al., 2010).

[4] Heteroskedasticity is defined in relation to regression analysis as dependence of the residuals on the magnitude of the independent variable. Here, heteroskedasticity is dependence of the residuals on the flows. Usually the absolute residual values increase with the increasing flow magnitude and therefore residuals in high flows have greater impact on the value of the objective function.

The Box-Cox transformation of flows y reads (Box & Tiao, 1973):

$$\tau(y,\lambda)=\begin{cases}\left(y^{\lambda}-1\right)/\lambda & ,\ \lambda\neq 0\\ \ln\left(y\right) & ,\ \lambda=0\end{cases} \tag{1.3.6}$$

where λ is a parameter to be estimated from the data.

Response surface

Another major concern about hydrologic model calibration is the response surface or the fitness landscape. The response surface is a hypersurface defined by the values of model parameters and objective function. For example, axes of the response surface in Figure 6 represent values of two model parameters, while the contour lines denote values of the objective function (in this example, the Nash-Sutcliffe efficiency).

Duan et al. (1992) highlighted several features of the response surfaces of hydrologic models:

(1) The response surface does not take convex shape with a single optimum. On the contrary, there are numerous regions of attraction spread throughout the parameter space.

(2) There are a lot of local optima within each basin of attraction which can be rather remote from the global optimum region (Figure 6).

(3) Response surface contains numerous ridges, which indicate non-linear interaction among model parameters (e.g. Kavetski & Clark, 2010).

(4) The response surface is not smooth: there are discontinuities in the first and second derivatives of the response surface with respect to the parameters.

(5) The parameter sensitivity, represented by the local slope of the response surface, varies across the response surface. Flat areas of the response surface indicate low parameter sensitivity (parameters that have negligible influence on model output). Such behaviour may be detected in the vicinity of the optima.

The aforementioned properties of the response surface impose severe limitations to the optimisation algorithms (chapter 1.3.4).

Multiple optima and ridges in the response surface indicate that different parameter sets, which may be scattered throughout the parameter hyperspace, result in approximately equal value of the objective function. Despite similar performance in the calibration period, these sets may result in quite different predictions in an independent period (Seibert, 1997). This is referred to as *equifinality* among different parameter sets (Beven and Binly, 1992). Parameter interaction (i.e. correlation between the parameters) is represented by the ridges of the response surface and means that the effects due to changes in one parameter can be compensated by changes in another parameter.

Discontinuities in the derivatives of the response surface may be attributed to the threshold processes in the model, to the objective functions or to the numerical scheme implemented within the model (e.g. Sorooshian et al., 1983; Kavetski et al., 2006; Kavetski & Clark, 2010). Derivatives of the response surface also reveal local parameter sensitivity. The parameter sensitivity signifies how changes in the parameter affect model output and the response surface. Small changes in sensitive parameters result in considerable changes of the objective function, and vice-versa. Low parameter sensitivity may be due to poor parameter identifiability, which means that the optimum parameter values cannot be inferred because a flat response surface indicates similar model performance. Poor parameter identifiability can be attributed to parameter interdependence, parameter insensitivity, data noninformativeness (there are no hydrologic conditions required to activate the parameter), inadequacy of the criterion (the objective function does not enable sufficient extraction of information from the data), or inadequacy of model structure (e.g. Sorooshian et al., 1983; Yapo et al., 1996; Beven, 2001b; Abebe et al., 2010). Conversely, well identified parameters converge to a narrow interval and they are considered well defined within the model structure.

Correlation among the model parameters produces ridges in the response surface. It may lead to poor parameter identifiability and hinder parameter optimisation and sensitivity analysis. Blasone at al. 2007 considered that weaker correlation among the parameters and absence of the ridges in the response surface indicate better model parameterisation.

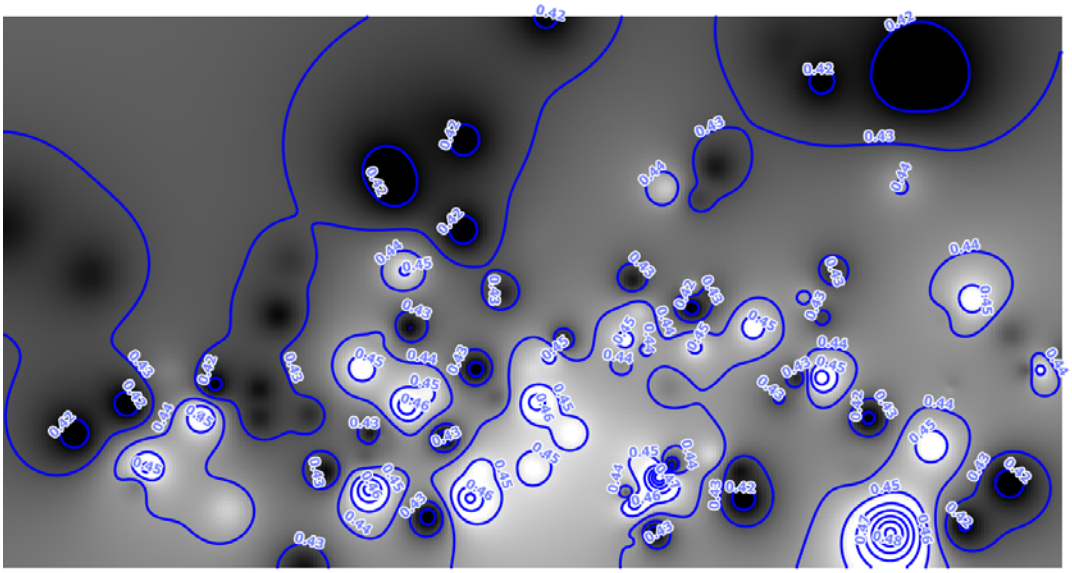


Figure 6. Response surface of the 3DNet-Catch model for the Kolubara River catchment (calibration in the 1988-2013 period): the Nash-Sutcliffe efficiency *NSE* as a function of two model parameters (precipitation gradient α and filtration coefficient K_d).

Ill-posedness of the calibration problem – model overparameterisation

Model overparameterisation signifies discrepancy between a great number of free model parameters and the number of the observed variables (usually, there is just one output variable – flow). This is a "one-to many" mapping problem, so many different parameter sets may result in similar simulated hydrographs and in similar values of the objective function(s) (Sorooshian et al., 2008). This is an ill-posed problem also called the *parameter equifinality* (Ebel and Loague 2006).

The discrepancy between dimensionality of the parameter space and the number of observed variables may be mitigated by (Blasone, 2007):

– ***Sensitivity analysis.*** The results of a sensitivity analysis indicate how changes in some parameters affect model output and model performance (Marino et al., 2008). It is commonly employed to detect the most influential (or sensitive) parameters, allowing to exclude insensitive ones from the calibration procedure. In this way, ill-posedness of the calibration is mitigated (Blasone, 2007; Muleta, 2012). Sensitivity analysis may be performed to quantify impact of input data and its uncertainty, boundary conditions, etc.

A review of the techniques used for sensitivity analysis is given by Razavi and Gupta (2015).

– Subsequent to the sensitivity analysis, ***two-step calibration*** may be performed. In the first step only the sensitive parameters are optimised. In the second step values of these parameters are set to the optimum values, and the remaining parameters are optimised. An issue related to this approach are interactions among parameters, which are neglected if all parameters are not optimised simultaneously (Blasone, 2007).

– ***Regularisation techniques***, which are commonly employed to reduce parameter dimensionality in the distributed models. These techniques are briefly discussed in chapters 1.3.5 and 2.3.2.

– ***Linking the parameter values to physically-based limitations.*** Schoups et al. (2008) have shown that if such limitations are imposed on the model (e.g. head-flow relations), an increased number of model parameters does not lead to an increased uncertainty of model predictions.

Irrespective of the technique applied, hydrological models are generally overparameterised, which means that there is no single optimal parameter set, but several acceptable sets (chapters 1.3.1 and 1.3.5). Therefore, "point" estimates of flows or other hydrologic variables are replaced by an ensemble of simulations produced with all acceptable parameter sets (Beven & Young, 2013).

Selection of the calibration period

The model parameters should be calibrated over a period that is sufficiently long so that it contains enough information on the catchment responses to enable "excitation" of all processes included in the model (Sorooshian et al., 1983; Beven 2001a; Wagener et al., 2003).

Wagener et al. (2003) demonstrated that information content for the identification of model parameters varies in time. For example, a parameter that represents fast overland flow can be identified during peak flow periods, whereas the highest information content for base flow takes place over prolonged droughts. Juston et al. (2009) calibrated the

model against the "informed-observer" data, sampled once per month or once per week, so that the information-rich events are captured. Performance of the model calibrated in this manner was slightly lower compared to the efficiency of model calibrated over the full record period, which means that most of the information in the full record period is redundant and that a model could be calibrated using discontinuous observations (also Perrin et al., 2007 and Kim & Kaluarachchi, 2010). Generally, the calibration period should include wet periods, as they contain more information and result in better identification of model parameters (e.g. Sorooshian et al., 1983; Yapo et al., 1996; Vrugt et al., 2006; Melsen et al., 2014). For example, Kim and Kaluarachchi (2010) demonstrated that a period comprising at least 36 wet months results in reliable model calibration. Xia et al. (2004) showed that different lengths of the calibration period are required to obtain consistent parameter estimates (for example, soil-moisture related parameter requires longest calibration period). Perrin et al. (2007) suggested that a parsimonious model requires fewer data and shorter calibration period for parameter optimisation.

There are numerous recommendations in the literature regarding length of the calibration period. Merz et al. (2011) considered that 5 years are the shortest calibration period sufficient for proper model calibration with the global SCE-UA optimisation algorithm (Shuffled Complex Evolution, Yapo et al., 1996). Yapo et al. (1996) calibrated a conceptual hydrologic model over 1-, 3-, 5-, 8- and 11-year long calibration periods by employing the SCE-UA. Their results suggested that at least 8 years is necessary to obtain parameter estimates that do not vary with the calibration period. Kim and Kaluarachchi (2010) demonstrated that model calibration with a global optimiser over periods longer than 10 years does not lead to any further reduction in parameter uncertainty.

Vrugt et al. (2006) calibrated the Sacramento SMA Model using the SCEM-UA global optimization algorithm over periods of increasing number of years. The posterior parameter pdfs indicated that 2-3 years of calibration suffice for obtaining stable pdfs. However, parameters obtained over longer calibration period resulted in more consistent performance. Wriedt and Rode (2006) demonstrated that calibration periods up to two years contain enough information to optimise sensitive model parameters using the GLUE method. Brigode et al. (2013) calibrated parsimonious models over the 3-year long periods using both the DREAM and GLUE methods.

As for the model performance, Luo et al. (2012) suggested that a longer calibration period (e.g. longer than 20 years) does not necessarily lead to an improvement of model performance. On the other hand, models calibrated in shorter periods perform well over the calibration period, but can result in a considerable decrease in model efficiency over the evaluation periods. Coron et al. (2014) demonstrated that models calibrated in the full record period reproduce water balance in sub-periods equally good as the models calibrated in any 10-year long sub-period. Romanowicz et al. (2013) compared model calibration over periods of increasing length to the problem of smoothing in statistics since the model performance is averaged over various hydrologic responses.

Clearly, there is a wide range of recommended lengths of calibration period in the literature. Recommendations depend on the model, parameter estimation method and catchment properties. Regardless of the record period length, it should be split into periods aimed for model calibration and evaluation. This imposes additional restrictions on the selection of the calibration period. Generally, as much information as possible should be used for model calibration, assuming that observations are reliable and diverse hydrologic responses are included in the calibration period.

Model overcalibration

Andréassian et al. (2012) identified two problems in the process of model calibration: miscalibration and overcalibration. Model miscalibration means that an optimisation algorithm failed to detect global optimum in a calibration period. On the other hand, model overcalibration means that the model does not perform well with optimal parameters outside the calibration period.

Model overcalibration is a major issue for model application. Merz et al. (2011) and Thirel et al. (2014) pointed out that it is crucial to understand the reasons for such model behaviour, especially if the model is to be used for simulations under conditions outside the range of the observed ones (for example for hydrologic forecasting and in climate change impact studies).

Impact of the calibration period on the parameter estimates and model performance, and model overcalibration problems are discussed in detail in chapter 1.5.

1.3.3. Objective functions

The objective functions in hydrological modelling are optimised with respect to model parameters in order to obtain the best possible agreement between the observed and simulated hydrologic variables (usually flows). Table 1 lists the objective functions commonly applied for hydrologic model calibration and evaluation.

Some objective functions indicate systematic errors (under- and over-estimation) or dynamic errors (timing). For example, presence of bias indicates under- and over-estimation of flows or runoff volume. Various hydrographs may result in the same bias because this performance measure is insensitive to dynamics of the simulated response. On the other hand, low coefficient of correlation indicates only dynamic errors; therefore it could take a maximum value even if the simulated flows were negative because it does not recognise bias (Criss and Winston 2008). However, majority of the objective functions reflect both types of error (Krause et al., 2005). For example, Gupta et al. (2009) separated the Nash-Sutcliffe efficiency *NSE* in two parts: ratio between the mean simulated and observed flows, which indicates bias, and the correlation coefficient, which quantifies the dynamic error.

Moriasi et al. (2007) categorised the most frequently used objective functions into regression-based, dimensionless and error indices. The first group of objective functions is comprised of the correlation coefficient, linear regression slope and interception. Dimensionless indices provide relative estimation of model efficiency and include e.g. *NSE*, index of agreement *d*, etc. Error indices are based on the mean square error, *MSE*.

Criss and Winston (2008) analysed ability of several objective functions to capture errors in timing and proportional increase / decrease of a hydrograph. They suggested that some objective functions do not properly reflect these errors, and proposed the volume error *VE*.

It has been recognised that the objective functions based on squared residuals (such as *RMSE* or *NSE*) are sensitive to outliers. The values of such the objective functions are principally determined by the model efficiency in high flow domain (e.g. Legates & McCabe, 1999; Krause et al., 2005). To improve robustness of the *NSE*, various modifications have been proposed in the literature. For example, in order to reduce

sensitivity to high flows, *NSE* can be calculated from the logarithms of flows, square root of flows or their reciprocal values[5] (Oudin et al., 2006; de Vos and Gupta 2010; Pokhrel et al., 2012; Seiller et al., 2012; Thirel et al., 2014). Lindstrom (1997) introduced a penalty to *NSE* in order to reduce *NSE* due to the runoff volume error. For balanced representation of systematic and dynamic errors in *NSE*, Gupta et al. (2009) proposed the *KGE* efficiency measure. Legates and McCabe (1999) suggested a general form of *NSE*, which enhances sensitivity to low flows. *NSE* can also be calculated for the flow duration curves. To cope with heteroskedasiticity in the residuals, Sorooshian et al. (1983) introduced *HMLE*.

The objective functions can be used as the evaluation criteria as well. This means that these functions are not included in model calibration, but they are employed to measure model performance instead. In addition to the objective function, Euser et al., (2013) proposed several "signatures" to test the realism of a hydrologic models, such as autocorrelation in the flow time series, rising limb density or peak distribution.

Further, Crochemore et al. (2015) studied the agreement between objective functions and expert judgement on model performance by conducting a survey among the hydrologic modellers. They revealed that the objective functions based on the squared or absolute error corroborate expert judgement about high flows. As for low flows, objective functions based on the log-transformed flows best reflect the expert judgment.

None of the objective functions is sufficiently versatile to reflect all aspects of agreement between simulated and observed flows. Model calibration should therefore employ several complementary performance criteria (e.g. Gupta et al., 1998; Moriasi et al., 2007). Recommendations on the acceptable values of *NSE* and flow bias are presented by Moriasi et al. (2007).

[5] Reciprocal values are calculated as $(1/(Q+\varepsilon))$, where ε is small constant (usually one per cent of mean flow value) to avoid dividing by zero (Thirel et al., 2014).

Table 1. An overview of the most frequently used objective functions for hydrologic model calibration against observed flows.

Objective function	Equation	Dimension	Target value	Comments and references
Relative flow bias	$bias_{\mathrm{rel}} = \frac{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)}{\sum_{i=1}^{n} Q_{\mathrm{obs},i}} \cdot 100$	%	0	Insensitive to dynamic errors. Bias which is not normalised is expressed in units of flow or runoff.
Coefficient of determination R^2	$R^2 = \left(\frac{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right)\left(Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim},i}\right)}{\sqrt{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right)^2}\ \sqrt{\sum_{i=1}^{n}\left(Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim},i}\right)^2}} \right)^2$	-	1	Insensitive to differences between simulated and observed flows (bias).
Mean absolute error *MAE*	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right\rvert$	m^3/s	0	*MAE* lower than one half of standard deviation of the observed flows is considered low. *MAE* is less sensitive to outliers than *RMSE*, therefore it is preferred over RMSE if outliers are present in the flow series (Legates & McCabe, 1999).
Root mean square error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2}$	m^3/s	0	Appropriate when data errors are uncorrelated and homoscedastic (Gupta et al., 1998; Romanowicz et al., 2013).
Transformed Root mean square error for high flows	$RMSE_{\mathrm{HF}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2 \cdot w_{\mathrm{HF},i}}$ $w_{\mathrm{HF},i} = \left(\frac{Q_{\mathrm{obs},i}}{Q_{\mathrm{obs,max}}}\right)^2$	m^3/s	0	Flow weighting puts more emphasis to high flows (Fenicia et al., 2007).

Table 1 (continued). An overview of the most frequently used objective functions for hydrologic model calibration against observed flows.

Objective function	Equation	Dimension	Target value	Comments and references
Transformed Root mean square error for low flows	$RMSE_{\mathrm{LF}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2 \cdot w_{\mathrm{LF},i}}$ $w_{\mathrm{LF},i} = \left(\frac{Q_{\mathrm{obs,max}} - Q_{\mathrm{obs},i}}{Q_{\mathrm{obs,max}}}\right)^2$	m^3/s	0	Flow weighting puts more emphasis to low flows (Fenicia et al., 2007).
Nash-Sutcliffe efficiency coefficient *NSE*	$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2}{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right)}$	-	1	Negative values indicate that mean value of the observed flows is better predictor than the model. *NSE* is rather sensitive towards high flows due to square values of the differences. *NSE* can be calculated using transformed flows (e.g. log-transformed or reciprocal values of flows). *NSE* can take low values if the observed flows exhibit small variability (Criss & Winston, 2008).
Linström measure *LM*	$LM = NSE - w\frac{\left\lvert\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)\right\rvert}{\sum_{i=1}^{n} Q_{\mathrm{obs},i}}$	-	1	*LM* is obtained by modifying NSE to account for error in simulated runoff volume. Value of *w* is commonly set to 0.1 (Lindstrom, 1997)
Kling-Gupta efficiency *KGE*	$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$ $r = \frac{\sum\left(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right)\left(Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim}}\right)}{\sqrt{\sum\left(Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right)^2 \sum\left(Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim}}\right)^2}}$ $\alpha = \frac{\hat{S}_{Q_{\mathrm{sim}}}}{\hat{S}_{Q_{\mathrm{obs}}}}\ ;\ \beta = \frac{\bar{Q}_{\mathrm{sim}}}{\bar{Q}_{\mathrm{obs}}}$	-	1	*KGE* is obtained by balancing model performance in reproducing mean flows and flow variability and linear correlation between observed and simulated flows (Gupta et al., 2009).

Table 1 (continued). An overview of the most frequently used objective functions for hydrologic model calibration against observed flows.

Objective function	Equation	Dimension	Target value	Comments and references
Index of agreement *d*	$d = 1 - \frac{\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2}{\sum_{i=1}^{n}\left(\left\|Q_{\mathrm{obs},i} - \bar{Q}_{\mathrm{obs}}\right\| + \left\|Q_{\mathrm{sim},i} - \bar{Q}_{\mathrm{sim}}\right\|\right)^2}$	-	1	Poor model performance may yield high values of this index (e.g. over 0.7) (Krause et al., 2005).
Volume error *VE*	$VE = 1 - \frac{\left\|\sum_{i=1}^{n}\left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)\right\|}{\sum_{i=1}^{n} Q_{\mathrm{obs},i}}$	-	1	*VE* denotes the flow volume common to the simulated and observed hydrograph and its complement denotes volume mismatch (Criss & Winston, 2008).
Maximum Likelihood Estimator for Heteroskedastic Error Case *HMLE*	$HMLE = \frac{\frac{1}{n}\sum_{i=1}^{n} w_i \cdot \left(Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}\right)^2}{\left(\prod_{i=1}^{n} w_i\right)^{1/n}}$ $w_i = Q_{\mathrm{obs},i}^{2(\lambda-1)}$	-	0	*HMLE* is calculated from the flows that are transformed applyign Box-Cox transformation (Box & Tiao, 1973). λ is paramter of the Box-Cox transformation that has to be estimated in the calibration along with the free model paramters (Sorooshian et al. 1983).

1.3.4. Optimisation methods used for hydrologic model calibration

Automatic model calibration implies parameter optimisation by employing a numerical optimisation procedure, which may be local or global. Prior to the optimisation procedure, the ranges of free model parameters should be restricted to a plausible parameter space Θ thus creating a constrained optimisation problem (Vanrolleghem, 2010).

Local optimisation methods start from a randomly sampled parameter set. These methods include derivative-based (gradient) and derivative-free (direct) methods. Gradient-based methods rely on the first- (e.g. steepest descent) or second-derivatives (e.g. Gauss-Newton algorithms) of the response surface with respect to the model parameters (Yilmaz et al., 2010). These methods can locate the optimum, provided that the response surface has convex or concave shape. For example, Duan et al. (1992) argued that the optimisation algorithms for the rainfall-runoff model calibration must be able to avoid trapping in the local optima regions and should not therefore rely on the derivatives of the response surface. Direct methods (e.g. Simplex or Pattern Search methods) explore the response surface in a systematic manner without calculating its derivatives. Since these methods are prone to trapping in a local optimum region, it is recommendable to repeat the optimisation procedure with different initial sets. Due to complex nature of the response function in hydrological modelling (chapter 1.3.2), local optimisation methods are not considered sufficiently robust for hydrologic model calibration because they are likely to fail in finding the optimal parameter set (Duan et al., 1992; Gupta et al., 2005; Yilmaz et al., 2010).

Nevertheless, some researchers believe that local optimisation methods could be applied for reliable hydrologic model calibration in case of smoother response surfaces. For example, smoother response surface can be obtained if the model is calibrated with HMLE as the objective function since this measure recognises heteroskedastic nature of the residuals (Sorooshian et al., 1983). Kavetski et al. (2006) advocated smoothing of the thresholds in a model to obtain smoother response surface and thus enable application of the gradient-based optimisation methods.

Global optimisation methods work with a large number of sampled initial parameter sets, which are iteratively moved towards the optimum regions (Figure 7). In this way, probability of the algorithm being trapped in a local optimum region is significantly decreased. Considering features of response surfaces in hydrological modelling, global optimisation methods have replaced local ones in the past few decades (e.g. Gupta et al., 2005).

Some combinations of the global and local methods also have been used for hydrologic model calibration (e.g. Seibert, 2000). Global algorithms narrow down the search to an optimum region, and the results of global optimisation are used as the starting point for the local ones.

Details on the global optimisation methods may be found in Weise (2009).

It has been recognised that a single global optimisation method cannot be efficient[6] in various optimisation problems. Vrugt and Robinson (2007) presented the AMALGAM, which contains several global optimisation methods, aiming at more effective and efficient parameter optimisation. The AMALGAM is employed in this research and it is therefore it is elaborated in chapter 2.2.

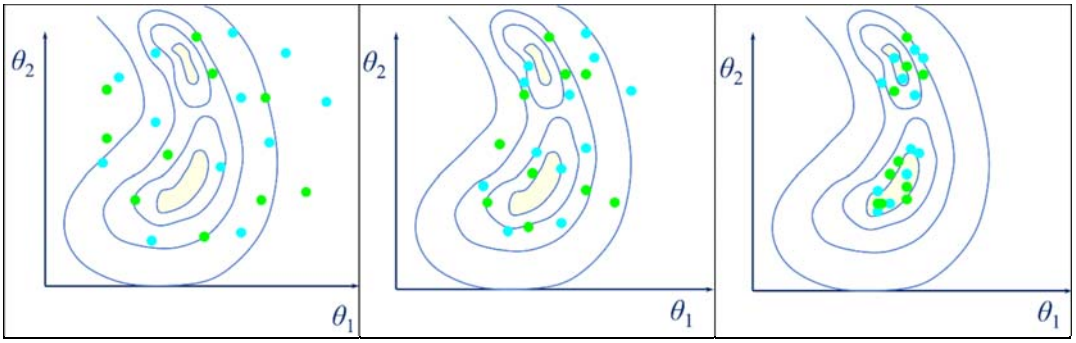


Figure 7. Global optimisation: multiple initial points in the parameter space (reproduced form the lectures by John Doherty, given in Belgrade in September 2013).

[6] Optimisation algorithms should be consistent, effective and efficient. Consistency implies algorithm's ability to locate the same optimum region of an optimisation problem in different repeated simulations; effectiveness refers to the probability of locating optimum region, and efficiency to the convergence speed (Duan et al., 1992; Blasone, 2007).

1.3.5. Multi-objective calibration

Multi-objective model calibration is performed with respect to two or more objective functions, thus enabling extraction of more information from the data (Gupta, et al., 1998; Wagener et al., 2003; Krauße et al., 2012; Shafii et al., 2014).

There are three types of multi-objective calibration (Madsen 2003):

(1) ***Multi-variable calibration*** involves several observed variables (e.g. flows, groundwater levels, concentration of tracers).

(2) ***Multi-site calibration*** relies on the data observed at several locations (e.g. flows at nested stream gauges in the catchment or groundwater levels at multiple wells)[7].

(3) ***Multi-response calibration*** implies parameter optimisation with respect to several objective functions, based on one observed variable. Multi-response calibration is employed in this research and it is briefly discussed here.

Multi-response calibration

Hydrologic models contain a considerable number of free parameters to be inferred from a single observed variable (i.e. flow). Ill-posedness of the calibration (chapter 1.3.2) can be mitigated if several objective functions are used. Single objective function cannot capture all aspects of model performance, but including several complementary objective functions increases extraction of information from the data. For example, *NSE* is aimed at assessment of the model performance in high-flow domain, but *NSE* calculated with the log-transformed flows quantifies model efficiency in the low-flow domain. Hence, some researchers consider manual calibration as the multi-objective calibration (Gupta, et al., 1998; Vanrolleghem 2010) because the modellers are usually looking at different aspects of agreement between the observed and the simulated hydrographs.

Multi-response calibration is defined as the optimisation of a set of *m* objective functions:

[7] Multi-site calibration cannot be applied with the semi- or fully-distributed hydrologic models.

$$F(\theta) = \left[f_1(\theta), f_2(\theta), ..., f_m(\theta) \right] \quad (1.3.7)$$

where f_k (k = 1, 2, ..., m) are the individual objective functions and θ is a parameter set.

One parameter set cannot be optimal according to several objective functions. Multi-response calibration therefore yields several parameter sets, which represent a trade-off among the objective functions. For example, one parameter set, θ_1, would reflect some aspect of the hydrograph (e.g. flow volume) better than other parameter sets, but θ_1 would not accurately quantify other aspects of model performance (e.g. model's ability to reproduce peak flows). These parameter sets are referred to as "non-dominated" or "Pareto-optimal", which means that one objective function cannot be further improved without deteriorating other objective functions. In other words, it is not possible to find a Pareto-optimal set θ_j such that $f_k(\theta_j) < f_k(\theta_i)$, $\forall k \in \{1, 2, ..., m\}$, where θ_i denotes other Pareto-optimal sets (Gupta et al., 1998). Figure 8A shows the Pareto-optimal parameter sets (of two parameters θ_1 and θ_2) which include the best values of two objective functions (A and B) and trade-off sets between them. The non-dominated sets make the so called Pareto front (Figure 8B) if the parameters are optimised according to two objective functions, or the Pareto surface if more objectives are used in calibration.

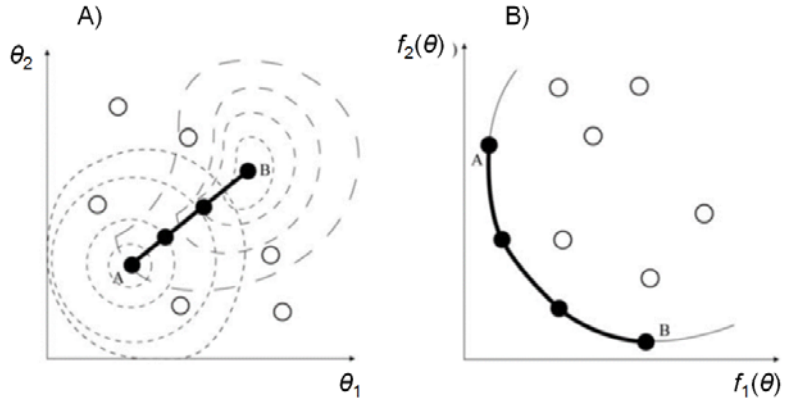


Figure 8. Pareto-optimal parameters in A) parameter space, and B) in the space of objective functions (reproduced from Blasone, 2007).

There are two ways to obtain optimum sets according to m objective functions: (1) to aggregate all objective function into a single, composite one (classical approach), or (2)

to optimise parameters according to all objective functions simultaneously. The former approach implies assigning weights to each objective function, such that all weights sum up to one. In this way, multi-objective optimisation is converted to a single objective optimisation problem. However, this approach inevitably involves subjectivity in assessment of the weights (Gupta et al., 1998). If the weights are systematically varied, the Pareto front may be obtained. In the latter approach the parameter sets are optimised and ranked according to the values of the objective functions and selected according to the definition of the Pareto-optimal solutions.

Efstratiadis and Koutsoyiannis (2010) discussed possibilities to recognise model issues from the properties of the Pareto front. For example, the Pareto fronts which resemble right angle indicate a significant trade-off between two objective functions and it is difficult to reach good values of both objective functions simultaneously because the parameters are rather sensitive to both of them (also Madsen, 2003).

Multi-objective vs. uncertainty-based calibration

It is important to distinguish between the multi-objective and uncertainty-based calibration, and between the Pareto-optimal and the GLUE behaviour solutions.

Uncertainty-based calibration is statistically grounded approach that aims at deriving posterior parameter pdf, thereby quantifying the parameter uncertainty from various sources (chapter 1.3.1). Multi-objective calibration does not rely on the Bayes theorem, nor it accounts for different sources of uncertainty such as the input data or model structure (Liu and Gupta 2007; Matott et al., 2009). However, some researchers (e.g. Engeland et al., 2006, Blasone, 2007 and Dotto et al., 2012) considered this approach relevant to estimate the parameter uncertainty.

Difference between the GLUE behavioural and Pareto-optimal solutions is illustrated in Figure 9. These sets may overlap, but generally they will not be identical (e.g. Gupta et al., 1998; Efstratiadis and Koutsoyiannis 2010). Namely, not all Pareto sets are GLUE behavioural, and vice-versa: some Pareto-optimal sets are not behavioural and they would be discarded within the GLUE method.

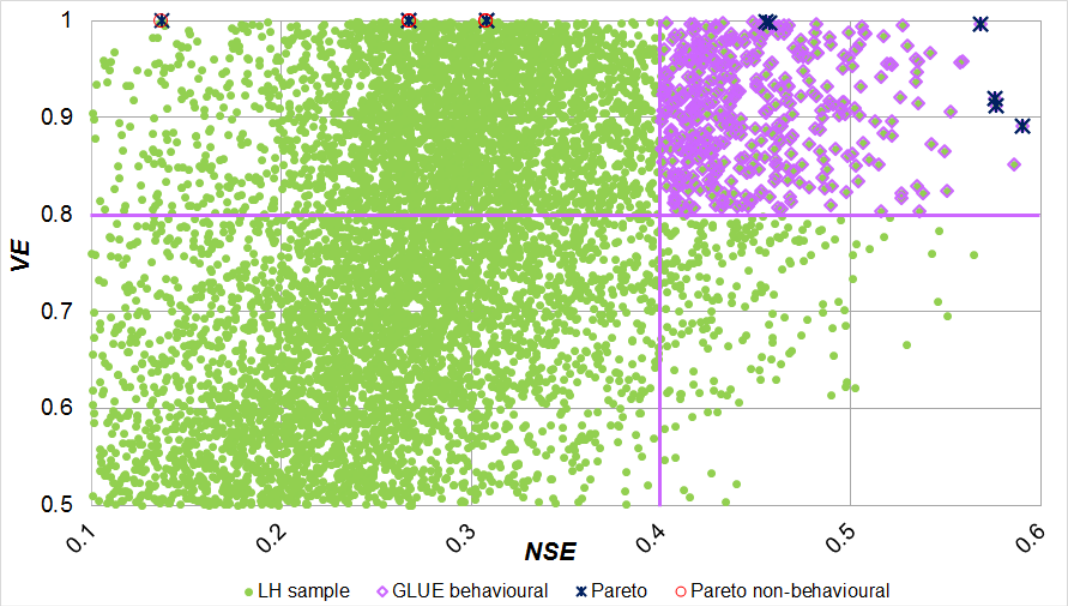


Figure 9. Pareto-optimal and GLUE-behavioural parameter sets of the 3DNet-Catch model for the Mlava River catchment (1988-2013).

Further details on the multi-objective calibration may be found in the literature (e.g. Yapo et al., 1998; Gupta et al., 1998; Madsen 2003; Vrugt et al., 2003; Engeland et al., 2006; Efstratiadis and Koutsoyiannis 2010; Shafii et al., 2014).

1.3.6. Calibration of distributed hydrologic models

Fully-distributed models contain a large number of free model parameters (chapter 1.2) and discrepancy between number of parameters and observed variables is substantial (ill-posednsess of the calibration problem), so calibration of these models is challenging.

Even in the physically-based models, whose parameters could (theoretically) be inferred *a priori* from the catchment properties (such as land use, soil types, topographic and geologic information etc.), still require calibration ("fine-tuning"). Regardless of fine spatial discretisation applied, the model cannot capture natural heterogeneity of the catchment properties, and the catchment properties may be averaged in the scale-up (e.g. Beven 2001a, Beven 2001b; Tucker et al., 2001; Blöschl and Zehe 2005; Gupta et al.,

2005; Schuol and Abbaspour 2006; Yilmaz et al., 2008). For example, value of the hydraulic conductivity inferred experimentally from the soil samples may considerably differ from the value that would result in the best model performance (Beven, 2001b).

To mitigate ill-posednsess of the calibration problem, various regularisation techniques[8] have been proposed in the literature. There are two approaches to mitigate calibration ill-posedness:

(1) Introducing additional information to the calibration procedure.

(2) Reduction of the dimensionality of the parameter space.

In the first approach, physically-based distributed models are commonly calibrated using the observations from the nested stream gauges or wells[9].

Reduced number of free parameters is easily achieved by detecting the areas (cells) that exhibit hydrologically similar behaviour and assigning a unique parameter set to each of them. These areas are called Hydrologic Response Units – HRUs (Beven, 2001b). HRUs are identified based on catchment topography, land use, vegetation or soil types, etc. Identification of HRUs can be facilitated by the Geographic Information System (GIS).

Commonly used regularisation techniques are based on the *a priori* parameter fields and the super-parameters that are optimised to achieve best fit to the observations. The super-parameters alter the entire parameter field, i.e. its spatial distribution. Location, variance or the entire parameter spatial distribution can be altered depending on the type of the super-parameter. At least one super-parameter is assigned to a free model parameter.

The simplest regularisation technique is based on scalar multipliers:

$$\hat{\theta}_{i,j} = m_i \cdot \theta_{\mathrm{PRIOR},i,j} \tag{1.3.8}$$

[8] Regularisation techniques are aimed at stabilisation of an ill-posed optimisation problem (Yilmaz et al., 2010; Pokhrel & Gupta, 2010; Pokhrel et al., 2012).

[9] Seibert (2000) calibrated the conceptual HBV model against observations in the wells by employing an auxiliary variable that links soil storage and groundwater levels, since the HBV model is not aimed at simulation of groundwater levels.

where $\hat{\theta}_{i,j}$ stands for the estimate of the i^{th} parameter (out of N_p free model parameters) in the j^{th} computational cell (out of N_g), m_i denotes the superparameter of the i^{th} parameter (scalar multiplier) and $\theta_{PRIOR,i,j}$ is the preset value of the i^{th} parameter in the j^{th} computational cell. This transformation changes both mean and variance of the parameter field. It contains N_p super-parameters to be optimised, thereby reducing the number of free parameters by N_g times (Pokhrel & Gupta, 2010).

A more complex regularisation techniques may also be applied in the following manner:

$$\hat{\theta}_{i,j} = m_i \cdot \left(\theta_{PRIOR,i,j} - \overline{\theta}_{PRIOR,i}\right) + a_i \tag{1.3.9}$$

$$\hat{\theta}_{i,j} = m_i \cdot \left(\theta_{PRIOR,i,j}\right)^{b_i} + a_i \tag{1.3.10}$$

where $\overline{\theta}_{PRIOR,i}$ is mean value of the i^{th} parameter in the catchment, a_i is an additive term and b_i is the power term of the i^{th} parameter. The linear transformation implies 2 N_p super-parameters, and the non-linear transformation implies 3 N_p super-parameters. Pokhrel and Gupta (2010) compared several regularisation techniques and obtained the best model performance with the regularisation given in equation 1.3.9.

In three aforementioned regularisation techniques additional constraints should be imposed to assure that transformed parameters take plausible values. A regularisation technique that enables non-linear transformation of the parameter field without imposing these constrains, presented by Yilmaz et al. (2008), is employed in this research and it is described in detail in chapter 2.3.2.

Other commonly applied regularisation techniques are the Tikhonov regularisation, which is based on the modification of the objective function to introduce penalty for the parameters that departure from the prior values, and the Singular Value Decomposition, which relies on the Principal Component Analysis (Yilmaz et al., 2010).

1.4. Uncertainties in rainfall-runoff models

Incomplete knowledge of hydrologic variables or models gives rise to uncertainty (Todini, 2007). Uncertainty can be considered as either aleatory or epistemic (Beven & Young, 2013; Beven, 2009; Blasone, 2007). The former is due to natural randomness of hydrologic and meteorological variables, and it is irreducible. This uncertainty can be represented by a probability function. The epistemic uncertainty stems from the limited knowledge, and it could be mitigated to a certain extent by enhancing the understanding of hydrologic processes, or by new measurements. These two types of uncertainties are not mutually exclusive (Beven and Young, 2013). For example, epistemic component of rainfall observations arise from negligence in the rainfall spatial heterogeneity, whereas the aleatory one steams from gaging errors, impact of wind, etc. Therefore, some uncertainty in hydrologic simulations always remains regardless of the model reliability (Gupta et al., 2005).

In hydrological modelling uncertainties stem from the input data, hydrologic model structure and parameters (Renard et al., 2010).

Uncertainty in input data is related to measurement errors or inadequate spatial and temporal resolution of the data. For example, Bardossy and Das (2008) showed that the number of rain gauges included in model calibration and consequently the precipitation spatial distribution significantly affect hydrologic model performance. Input-related uncertainties also stem from the rating curves, since flows are usually estimated from observed river stage using rating curves.

Model induces uncertainty on three levels: the perceptual, conceptual and numerical models. Perceptual model does not necessarily have to include all processes that participate in runoff generation at a particular catchment. Also, a conceptual model commonly represents approximate mathematical description of perceived processes: for example, effective precipitation depends on soil moisture, and vice-versa, which is linearized in the conceptual model. Numerical models are approximations of the partial differential equations of the conceptual model, and therefore introduce additional uncertainty. To account for uncertainty due to model structure, numerous multi-model combinations have been developed recently (Ajami et al., 2006; Clark et al., 2008; Fenicia et al., 2007; Li & Sankarasubramanian, 2012).

Model parameters are information integrators, meaning that they are affected by various sources of uncertainties: data errors, model structural inadequacies, lack of robustness of model calibration method (Todini, 2007, Figure 10). Parameter estimates also depend on the calibration period, i.e. calibration over different periods would result in different parameter estimates. Therefore, calibration period is also a source of uncertainty (Deletic et al., 2012). There are also secondary sources of uncertainty, such as the parameters' dependence on the state variables that are simulated using these parameters (e.g. soil moisture), but these dependencies are commonly neglected in hydrological modelling (Abebe et al., 2010).

According to Mantovan and Todini (2006) and Todini (2007), the parameters are considered as "dummy", uncertain quantities, which reflect various sources of uncertainty (e.g. input data or model structure). Marginalisation of their posterior pdf that represents parameter uncertainty, i.e. its integration over the entire feasible parameter space in every time step yields predictive uncertainty in that step (e.g. uncertainty in the simulated flow).

Parameter posterior pdf can be inferred following the uncertainty-based calibration methods, outlined in chapter 1.3. An alternative to these Bayesian (Monte Carlo, probabilistic) methods are local deterministic methods, such as the Taylor series expansion methods. In this approach higher order terms are discarded from the expanded Taylor series. These methods result in statistical moments of parameters rather than posterior pdf. To calculate derivatives of the model output (or an objective function), numerical differentiation is commonly applied. Nevertheless, these methods are based on the assumption of linearity of model response with respect to the model parameters, which is not valid in hydrological modelling. Therefore these methods have been replaced by the Bayesian ones (Kuczera & Parent, 1998; Vrugt et al., 2006). As pointed out in chapter 1.3.5, some researchers estimated parameter uncertainty by employing multi-objective calibration, which is not founded on the Bayesian statistics. An overview of the methods for estimation of the uncertainty in hydrologic modes is given by Matott et al. (2009).

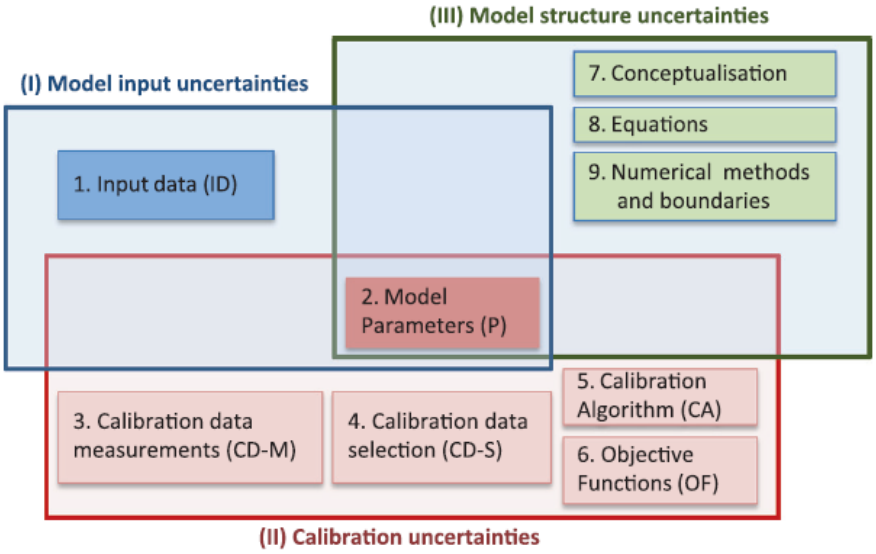


Figure 10. The key sources of parameter uncertainties in automatic model calibration (Deletic et al., 2012).

1.5. Hydrologic model transferability in time

As discussed in previous chapters, estimation of hydrologic model parameters remains a challenging task despite the robust optimisation algorithms and available computational resources. An issue about the rainfall-runoff models is deterioration in model performance and in reliability of the simulated hydrologic variables outside the calibration period. One of the reasons for such model behaviour may be the fact that the optimality of the parameter sets does not hold outside the calibration period. This undesired property of hydrologic models imposes constraints on their applicability.

Research on hydrologic model transferability in time has been conducted by analysing: (1) model performance over different periods, and (2) parameter variability in time, i.e. with the calibration period. The results of these analyses are thoroughly reviewed and presented in chapter 1.5.1. Attempts to improve model transferability in time that have been made so far are elaborated on in chapter 1.5.2.

1.5.1. Assessment of the consistency in model performance and model parameter estimates

Assessment of consistency in model performance

A framework for assessment of model performance consistency was established by Klemeš (1986) and it is outlined in Table 2. Model robustness is commonly assessed using Split Sample Test (*SST*) or Differential Split Sample Test (*DSST*). *DSST* is considered more robust compared to the Split Sample Test one (*SST*) and more suitable if temporal transferability of a model is to be evaluated (e.g. Klemeš, 1986; Seibert, 2003; Thirel et al., 2014; Refsgaard et al., 2014). However, a significant decrease in model performance in *DSST* has been reported in the literature (e.g. Klemeš, 1986; Seibert, 2003; Vaze et al., 2010; Li et al., 2012). Details on application of *DSST* are elaborated by Thirel et al. (2014).

Table 2. Parameter transferability tests (Klemeš, 1986)

TEST	APPLICATION
Split Sample Test (*SST*)	Model calibration and evaluation over two independent, but climatically similar periods.
Differential Split Sample Test (*DSST*)	Climatical transposability: model calibration over a period, and evaluation in period markedly different from the calibration one (e.g. dry-wet, cold-warm). If a model is intended to simulate runoff under dry conditions, it should be calibrated over wet and evaluated on dry periods, and vice-versa (Brigode et al., 2013).
Proximity Basin Test	Geographical transposability – model calibration against observations at one, and evaluation against observations at the other, similar catchment.
Proximity Basin Differential Split Sample Test	Evaluation of model overall transposability.

To estimate model ability to reproduce peak flows over periods wetter than the calibration one, Seibert (2003) calibrated the HBV model within the GLUE framework over dry years (lower peak flows) and evaluated on years with higher peak flows. The results indicated systematic underestimation of peak flows.

To increase validity of their conclusions, many researchers performed model evaluation using data from numerous catchments and /or different hydrologic models.

Le Lay et al. (2007) evaluated lumped GR4J model over various sub-periods, sampled from the full record period. They assumed that if the model parameters properly reflect catchment behaviour, the model efficiency over sub-periods with similar characteristics should be consistent and have minimum deviation from the model efficiency over the entire period. They calibrated the model and calculated variances of model efficiency over 200 sub-periods. Models calibrated in two sub-periods consisting of consecutive years with similar runoff coefficients (detected by applying Hubert's segmentation method) resulted in negligible deviation from the mean model performance, as opposed to the models calibrated in the periods with highest and lowest precipitation rates. They concluded that similarity of the simulation periods in terms of precipitation depths is crucial for consistent model performance.

Vaze et al. (2010) calibrated four lumped conceptual hydrologic models (SIMHYD, Sacramento, SMARG and IHACRES) over 10, 20, 30 and 40 consecutive driest and wettest years for 61 catchments. Every model was evaluated over the complementary periods. The results revealed a drop in model performance in the evaluation periods (*NSE* and flow bias), which increases in magnitude with the difference in annual precipitation depths. Provided that the evaluation periods are up to 15% drier or up to 20% wetter than the calibration one, the decrease in model performance is acceptable (flow bias is smaller than 20%). Model evaluation over periods drier than the calibration one resulted in larger decrease in model performance than other way round. Models calibrated over short periods performed poorly over long ones (even if the annual precipitation amounts are similar), and vice-versa. Differences among the results obtained by alternative model structures were subtle.

Merz et al. (2011) calibrated the HBV model over six 5-year long periods, and evaluated every parameter set over the remaining five periods. The results clearly indicated that bias in the simulated flow volume increases with the time lag between the calibration and an evaluation period.

Li et al. (2012) calibrated two lumped, conceptual models (DWBM and SIMHYD) for 30 catchments over two wettest and two driest periods, selected according to annual

precipitation. Models were calibrated using the GLUE method and *NSE* as the objective function. Each model was evaluated over the remaining three periods. The best models' performance was obtained in wet calibration period, whereas the lowest efficiency was obtained in dry period with models that were calibrated over wet period.

Luo et al. (2012) calibrated lumped conceptual SYMHID model for 12 catchments following four different calibration strategies. These strategies include: (1) periods of various length, (2) full record period, (3) periods of various climatic characteristics, and (4) monthly-based calibration. Periods with different climatic characteristics were selected according to the annual precipitation depths (wet, normal and dry years) and ENSO index (El Niño, La Niña and neutral years). The results revealed that "analogue" calibration strategies did not lead to any improvement in model performance over evaluation periods at most of the catchments. Model calibration on monthly basis was shown to enhance model performance at the catchments with distinct seasonality, provided that no significant shifts in seasonality occur.

Seiller et al. (2012) conducted the *DSST* with an ensemble of twenty lumped, conceptual model. Four 5-year long periods were selected according to the precipitation depths and temperature: dry/warm (HC), dry/cold (DC), humid/warm (HW) and humid/cold (HC). Transferability of the ensemble was higher than the transferability of the individual models, although some of the individual models outperformed the ensemble in some periods (e.g. GR4J yielded higher *NSE* value than the ensemble on HC-DW test).

Brigode et al. (2013) calibrated two lumped, conceptual hydrologic models (GR4J and TOPMO) for 89 catchments. The models were calibrated over three 3-year long periods (wet, intermediate and dry) which were selected according to the aridity index, and in the full record period by employing the DREAM algorithm (Vrugt et al., 2008). All models were evaluated over the driest period. The model calibrated over the wet period resulted in the highest decrease in model performance over the evaluation period. Simpler (GR4J) model slightly outperformed the TOPMO over the evaluation period in terms of both *NSE* and flow bias. They compared evaluation performance of individual optimum sets and an ensemble sampled form the posterior pdf, both being obtained over the full record period. The results of the simpler model were almost identical, whereas TOPMO ensemble outperformed individual optimum sets.

In order to further enhance robustness of the model transferability assessment, Coron et al. (2012) proposed the Generalised Split Sample Test (*GSST*). In *GSST* the model was calibrated over all 10-year long moving periods, shifted successively by one year. The obtained sets were tested over all remaining non-overlapping periods, resulting in increased number of *SST*s. They analysed model performance over evaluation periods along with the meteorological characteristics of the calibration and evaluation periods: precipitation depths, temperatures and potential evapotranspiration rates (*PET*). The results revealed that the drop in evaluation efficiency increases with the difference between precipitation in the calibration and evaluation periods, but no correlation with the differences in temperature or *PET* rates was detected. This lack of correlation was attributed to the fact that the catchments considered were water-limited. Model performance was represented by the ratio of a composite objective function obtained in the calibration and evaluation periods. The flow bias is affected by rainfall depths and *PET* rates, which was confirmed at almost all catchments considered.

This research was further extended by Coron et al. (2014) who calibrated three models of increasing complexity in the same manner (10-year sliding windows) for 20 catchments with the Kling-Gupta efficiency (*KGE*) as the objective function. After simulating the flow with all parameter sets obtained in calibration, 10-year moving averages of the simulated and observed flows are compared. When plotted, the ratios $\frac{\hat{Q}_{10\,years}}{Q_{10\,years}}$ for all parameter sets constituted nearly parallel curves shifted along the ordinate. The simplest model resulted in larger vertical spread in the curves.

Since consistent model performance has become major issue in hydrology, Thirel et al. (2014) presented a framework for assessment of model performance outside calibration period. They suggested that the model should be calibrated over the full record period and over five distinct sub-periods of equal length, resulting in six calibrated models. Every model should be evaluated over the remaining periods. Various graphs aimed at facilitating effective and adequate representation of the results are suggested in the paper. The methodology presented was followed by e.g. Li et al. (2014) and Magand et al. (2014).

In the literature reviewed, the models were tested using flow observations. However, model robustness can also be evaluated by analysing the model performance in simulating variables against which the model was not calibrated (e.g. ground water levels) or at different sites (e.g. simulation of flows at nested stream gauges) (Muleta, 2012; Seibert, 2003).

Assessment of consistency in the parameter estimates

The parameter estimates are affected by the properties of the period that they were calibrated over. Therefore, model calibration in different periods yield different parameter estimates. If model calibration over different periods yields approximately the same parameter values or posterior pdfs, it is referred to as the parameter consistency (Vrugt et al. 2006), parameter stability (e.g. Niel et al., 2003; Merz et al., 2011; Brigode et al., 2013), parameter sensitivity to calibration period (Yapo et al., 1996; Singh & Bárdossy, 2012) or uncertainty due to calibration period (Deletic et al., 2012). Assuming that model calibration yields optimal parameters in that period, parameter variability with the calibration period may be considered equivalent to the parameter temporal variability.

Consistency in parameter estimates is quite important since it warrants model transferability in time (extrapolation), i.e. model ability to properly reproduce catchments' behaviour outside the calibration period (e.g. Seibert 2003, Hartmann and Bardossy 2005). Andréassian et al. (2012) distinguished between "hydrologic optima" and "mathematical optima". The latter term denotes the optimal parameter sets for the given the objective function, optimisation method and calibration period, while the "hydrologically optimal" parameter sets result in high model performance within and outside the calibration period.

Along with analyses of model performance over various periods, consistency in parameter estimates, posterior pdfs, identifiability and sensitivity have been examined.

One of the first attempts to investigate temporal parameter variability was made by Wolf and Ostrowski (1982). They calibrated a model in each month over 10 years for 3 catchments and analysed intra-annual parameter distributions. They demonstrated resemblance among intra-annual distributions of the surface and subsurface reservoir

coefficients, but no systematic periodicity was detected. More importantly, they indicated that the data errors result in the parameter variability (uncertainty) having the same order of magnitude as the parameter temporal variability.

Wagener et al. (2003) pointed out that residual aggregation in time leads to loss of valuable information in the observed data. They proposed methodology for detection of "high information content" periods, entitled DYNIA (DYNamic Identifiability Analysis). This methodology is based on the Regional Sensitivity Analysis (RSA by Spear and Hornberger, 1980), which is applied in dynamic manner over 11- to 101-days long moving windows. In this way posterior parameter pdfs are obtained for all windows. They used a parsimonious model (RRM) with 5 free parameters. The results indicated that the posterior pdfs (measured by the 90% confidence bounds of the posterior pdfs) and the parameter optimum values (pdf modal value) vary in time. They related parameter identifiability with the narrow and peaky posterior pdf. Their results suggest that information required for identification of linear reservoir coefficient for direct runoff simulation are contained within peak flow periods, while linear reservoir coefficient for baseflow can be inferred in prolonged dry periods. Wriedt and Rode (2006) employed the DYNIA method with 101 days long time frames. They analysed the pdfs with respect to the magnitude of observed flows over the corresponding time window. They demonstrated that e.g. interflow-related parameter is identifiable within low flow domain since its posterior pdf becomes narrower with flow decrease. Abebe et al. (2010) applied the DYNIA method using the HBV model and fine temporal resolution data. Their results supported the previous findings: optimal parameter values (posterior pdfs' peaks) and uncertainty bounds varied in time.

Niel et al. (2003) assumed that if precipitation, flow and runoff coefficient (annual) time series were stationary model calibration over a sub-period would result in similar parameter estimates, and vice-versa. The proposed method consists of two steps: (1) detection of break points in the time series (by the Pettit test) aiming at detection of stationary periods, and (2) model calibration over different (contrasted) periods. Consistency in the parameter estimates is assessed by comparing the confidence regions of parameters optimised over contrasted calibration periods. If the confidence regions overlapped, the parameters were considered consistent. Conversely, if the regions of parameters were disjoint, the parameters were considered inconsistent. The method was

applied on 17 catchments in West Africa known for significant drop in water yield after 1970s, with the parsimonious GR2M model. The results revealed that non-stationarity in the time series does not necessarily imply inconsistency in model parameters. This research was further extended by Le Lay et al. (2007) who analysed parameter consistency by employing two methods. The first method implied: (1) sampling of 100 sub-periods form the full record period and selection of 100 sub-periods that are complements to the sampled ones, and (2) model calibration over these periods. Absolute differences between the parameters obtained over a sub-period and its complement were calculated and the pdf of differences was derived for each model parameter. The differences obtained for two stationary periods in terms of runoff coefficient and two contrasted periods regarding precipitation rates were near to the pdf modal value for all parameters. The second method involved derivation of posterior cdfs of the behavioural parameters using the GLUE method for the contrasted periods and implementation of the χ^2 test to examine whether the cdfs for contrasted periods were significantly different. Three out of four model parameters were significantly different for contrasted period, while the fourth one was insensitive over all periods.

Merz et al. (2011) examined the long-term trends in the HBV model parameters by calibrating the model over 5 consecutive years for 273 Austrian catchments. They indicated that some parameters of the soil and snow routines exhibit trends, but the correlations between the parameters and climatic variables (e.g. temperature) were catchment specific. This research was extended by Osuch et al. (2014) who calibrated the HBV model by employing the SCEM method and quantified correlations between parameter estimates and climatic indices in terms of the Pearson and weighted Pearson correlation coefficients. They revealed the surprisingly strong correlations between some parameters of the HBV model and climatic indices. For example, maximum soil storage was correlated to precipitation depths and standard deviation of precipitation; reduction factor for *PET* and percolation rate were correlated to mean *PET* and its standard deviation. Similarly to Merz et al. (2011), their results were catchment specific. Results of Li et al. (2014) are consistent with the results presented in these two papers in terms of the parameter variability with the calibration period: soil-related parameters exhibit largest variability, whereas the coefficient of linear reservoir of upper soil zone varies slightly.

Li et al. (2012) calibrated two lumped, conceptual models by employing the GLUE method with composite objective function over two wettest and driest periods for 30 catchments. The χ^2 test was carried out to compare posterior parameter pdfs obtained over the wettest and driest periods. The pdfs of all parameters were significantly different at least for 10% of the catchments. The soil-related parameters were found to be the most sensitive to the calibration period since these posterior pdfs are significantly different in over a half of the catchments considered.

Luo et al. (2012) calibrated the SYMHID following four different calibration strategies, as elaborated in previous subchapter. Estimates of majority of parameters were shown to vary in time. The parameter variability depends on the calibration strategy. The greatest parameter variation was obtained with monthly-based calibration. The interflow-related parameter exhibited slightest variability.

Sieber and Uhlenbrook (2005) analysed change in parameter sensitivity in time (two consecutive rainfall events). They quantified parameter sensitivity in terms of the standardised regression coefficients (chapter 2.3.2) and by employing the Regional Sensitivity Analysis (RSA). The results of both approaches revealed considerable temporal changes in sensitivity in most of the parameters. For example, sensitivity of some parameters abruptly increased or decreased over the precipitation events.

Sorooshian et al. (1983) argued that some parameters should vary in time due to seasonality in hydrologic cycle and long-term changes, such as urbanisation or deforestation. Merz et al. (2011) consider that the parameter variability may be due to the "secondary" processes which are not explicitly simulated by a hydrologic model (e.g. variable infiltration rates due to soil freezing or cracking (Beven 2001; Tian et al. 2012), or variable evapotranspiration due to vegetation aging (Fenicia et al. 2009).

1.5.2. Improvement of consistency in model performance and parameter estimates

To improve consistency in hydrologic models performance several approaches have been proposed in the literature: increasing parameters temporal transferability, ensemble model weighting, time variable parameterisations and enhancement of model structure.

Model calibration improvement

Hartmann and Bardossy (2005) proposed a linear combination of Nash-Sutcliffe efficiency coefficients (*NSE*) calculated not only for daily flows, but also for flows averaged over longer periods (e.g. weeks, months, seasons, years) and for the transformed flows (e.g. square root transformation). They carried out *DSST* with calibration over wet period and evaluation in dry one to appraise several calibration strategies – combinations of flow series according to which *NSE* was calculated. Parameter estimates obtained with *NSE* with daily and annual flows resulted in high model performance in terms of smaller decrease in *NSE* and flow bias in the evaluation period.

Gharari et al. (2013) advocated multi-objective calibration over several sub-periods of equal length, resulting in several Pareto fronts. They assumed that minimisation of the Euclidian distance to all sub-period Pareto fronts would result in the Pareto sets (so-called Minimum Distance Pareto Front – MDPF) that would perform consistently. The performance of MDPF over the sub-periods of the short testing period was almost as good as the performance of the Pareto fronts obtained over each sub-period. The MDPF performance over the long testing period was consistent, although suboptimal in some years compared to the Pareto front obtained over the full calibration period.

Conditional parameterisations

To obtain more consistent model performance some researchers applied time variable parameterisations (e.g. on monthly or seasonal basis). To obtain these conditional parameterisations, the model parameters are optimised in various climatic conditions (e.g. wet or dry periods).

Fenicia et al. (2009) assumed that changes in catchment properties would reflect in changes in the model parameters. They tried to explain a rainfall-runoff anomaly in the Meuse catchment behaviour (i.e. decrease in runoff from 1930 to 1965) by varying the model parameters in time. They calibrated the conceptual FLEX model with 10 free parameters using the GLUE method over consecutive 4-year long periods by employing three calibration strategies. All model parameters were allowed to vary with the calibration periods in the first strategy, while in the second and in the third strategy the number of such parameters was reduced to five and two, respectively. The results

indicated that two time variable parameters could explain the anomaly: namely, time to peak and the parameter relating changes in forest transpiration to the forest age. They attributed the decrease in the former parameter to the catchment urbanisation and river engineering works. Variability in the latter parameter was attributed to forest rotation i.e. changing age of the forests and consequently *ET*.

Muleta (2012) carried out a sensitivity analysis (SA) of the SWAT model parameters over wet and dry seasons and in entire calibration period. The wet and dry seasons were selected according to mean monthly runoff. The SA revealed that sensitivity of some parameters related to soil conductivity, evaporation and interception capacity changes between wet and dry seasons. He optimised the principal model parameters and obtained two version of the model. The first version comprised temporally invariant parameters, while the parameters of the second one varied over the seasons. Two versions of the model were evaluated by conducting *SST*. The model with varying parameters outperformed its counterpart in most of the evaluation periods.

Choi and Beven (2007) calibrated the TOPMODEL using the GLUE framework and various objective functions. Behavioural parameter sets were updated according to model performance over the years after the calibration period (globally conditioned models). There were numerous behavioural parameter sets in individual years, but only a few sets were behavioural over the full record period. To account for seasonal shifts in runoff generations mechanism, they calibrated the model in a dynamic manner (multi-period conditioned models) over 15 fuzzy clusters of time. The clusters were sampled according to precipitation, precipitation variance, maximum daily precipitation and *PET*. Behavioural parameters' posterior pdfs varied considerably over the clusters and none of the parameter sets was behavioural over all clusters. Minimum number of the behavioural sets was obtained over dry clusters due to poor model performance in dry periods, which was attributed to the model structural deficiencies. In the evaluation period, the multi-period conditioned model resulted in significantly higher percentage of flow observations within the prediction band than the globally conditioned one.

Zhang et al. (2011) calculated six aridity indices for each water year of the hydrologic record. They performed the principal component analysis (PCA) of the indices to reduce redundancy in data since all aridity indices are based on daily temperatures. The fuzzy C-

means clustering method was applied to the principal components resulting in five clusters. Every year was assigned to a particular cluster and split into the warm and cold seasons. Distributed SWAT model was calibrated in every season over all clusters (i.e. ten model calibrations) by employing the SCE calibration algorithm. The number of free parameters was reduced after the sensitivity analysis prior to the model calibration. The results in the calibration and evaluation periods were compared to the results of the model calibrated in the full record period. The "multi-period" model outperformed the "single-period" model in both periods in terms of *NSE* and flow bias. In addition, "multi-period" model resulted in narrower prediction intervals and in larger percentage of observation encompassed by the prediction band.

Model ensemble and model averaging

Oudin et al. (2006) applied dynamic weighting of two model parameterisations obtained with *NSE* calculated with flows and log-transformed flows. . They examined four different weighting strategies: (1) equal weights, (2) sinusoidal weights, (3) weight that is equal to normalised soil moisture (form 0 to 1) and its complement, and (4) weights calculated using the nonlinear functions of simulated soil moisture. The fourth weighting strategy resulted in the highest model performance.

Weighting of the outputs from different hydrologic models within Hierarchical Mixtures of Experts framework (HME) is employed by Marshall et al. (2007). HME is based on their individual models and gating functions that control weighting, i.e. probability of using the individual models. The gating function relates probability of using a model with the predictor variables, such as antecedent precipitation. HME allows that model with the same structure but different parameters have different weights – probabilities. Marshall et al. (2007) used HME with parsimonious models (3 free parameters) and simple gating functions. The results obtained by employing HME with three models outperformed those of the single model.

Hsu et al. (2009) applied Bayesian model averaging in a dynamic manner. Namely, probability of each version of the ARX model was conditioned on the model performance

over previous computational time step. Model ensembles obtained in this way outperformed individual models in calibration and evaluation periods.

Model structure improvement

Time variability of model parameters is assumed by de Vos et al. (2010) to be due to model structural inadequacy. These authors calibrated the lumped conceptual model HyMod in (1) single- (traditional calibration) and multi-objective manner over entire calibration period, and (2) over 12 clusters of time (dynamic calibration). The clusters are selected according to daily precipitation, 10-day moving average of precipitation and soil moisture simulated by the GR4J model. They successively improved the model structure by introducing a parameter for correcting the observed precipitation rates, upgrading linear reservoirs to the nonlinear ones, and introducing the routing function to the model. The corrections to the model are made so that traditionally calibrated model performs as well as the dynamically calibrated one.

Efstratiadis et al. (2014) enhanced the lumped hydrologic DM0 model to account for catchment urbanisation. They proposed two alternatives: (1) the liner reservoir coefficient for direct runoff simulation which was proportional to the share of urbanised areas (model DM1), and (2) application of a distributed version of the model DM2 which involved Hydrologic Response Units (HRUs). In the distributed model the catchment is delineated in two HRUs. One HRU included urbanised and the other non-urbanised areas in the catchment. Different parameter sets are assigned to the HRUs. They tested the models following the protocol presented by Thirel et al. (2014). The performance of the models DM0 and DM1 was similar, while the distributed model DM2 performed considerably better.

1.5.3. Model transferability in time and assessment of the climate change impact on water resources

The model ability to reproduce catchment behaviour under various climatic conditions is very important if the model is to be used for climate change (CC) impact on water resources (Hartmann and Bardossy 2005; Wilby 2005; Vaze et al., 2010; Peel and Bloschl 2011; Luo et al., 2012; Brigode et al., 2013; Thirel et al., 2014). Apart from the projected increase in temperature, issues in hydrologic model application stem from the unknown precipitation intensities and patterns (Steenbergen and Willems 2012) and from the unknown vegetation response to the enhanced CO_2 concentration (Vaze et al., 2010). These changes may also affect rainfall-runoff relations (Li et al., 2014).

Few attempts have been made to estimate uncertainty in hydrologic projections under climate change due to the parameters of rainfall-runoff models. As Jiang et al. (2007) and Bastola et al. (2011) pointed out, these uncertainties have not been sufficiently investigated and that further research in this domain is needed.

Wilby (2005) calibrated the lumped and semi-distributed version of the CATCHMOD hydrologic model in the wettest and the driest year, in the year that was considered analogue to the conditions projected for 2050s, and in the full hydrologic record period using the GLUE method. The results suggested that (1) the parameter estimates and identifiability vary with the calibration period, (2) the model calibrated in dry years results in poor performance, and (3) flow projections with semi-distributed models are less sensitive to the calibration period. He recommended that *DSST* should be conducted prior to model application to CC impact assessment.

Brigode et al. (2013) calibrated two hydrologic models over the full record period, and in the wettest, intermediate and dry 3-year periods selected according to the aridity index. They demonstrated that the flow projections depend on the model calibration period, either when a single optimal parameter set or an ensemble of parameter sets is used.

Magand et al. (2014) calibrated semi-distributed CLSM model in a multi-objective manner over consecutive 9-year periods and in the full record period. They used only one Pareto-optimal parameter set from each calibration period to obtain hydrologic

projections. Although these sets had similar performance in the *DSST*, projected flows and *ET* rates were quite different.

1.6. Research aims and objectives

1.6.1. Conclusions from the literature review

The results presented in the papers reviewed in the previous chapter can be summarised as follows:

– Parameters identifiability, posterior pdfs or optimised values vary with the calibration period.

– Consequently, model performance decreases outside of the calibration period. Larger differences between calibration and evaluation periods (in terms of meteorological characteristics, primarily precipitation) lead to greater decrease in model efficiency.

– Soil- (infiltration), snow-, and vegetation-related (*PET*) parameters are proven to be sensitive to selection of the calibration period (Wilby 2005; Fenicia et al., 2009; Merz et al., 2011; Li et al., 2012; Luo et al., 2012). Variability in parameters of the response routines may also be detected if the catchment has been urbanised or if the river engineering works have been implemented (Fenicia et al., 2009).

– Climatic non-stationarity (trends or jumps in e.g. precipitation and temperature time series) does not necessarily imply inconsistency in parameter estimates (Niel et al., 2003, Le Lay et al., 2007). Generally, strong correlation between model parameters and climatic variables has not been found, although the results are catchment specific.

– Values of the optimised parameters depend on the objective function(s) used.

– Inclusion of the parameter temporal variability in the modelling procedure (e.g. assigning different parameter sets to distinct clustered periods, dynamic weighting of parameter sets, etc.) is shown to yield better model performance over calibration and evaluation periods. However, these approaches have not been widely applied.

The literature review also reveals several gaps in the existing research in parameter variability with the calibration period:

– Variability in parameter estimates obtained by multi-objective model calibration has not been examined.

– Consequently, impact of selection of the objective functions on parameter consistency has not been analysed.

– The analysis of calibration period length on parameter variability should be extended, especially when it comes to parameter estimation in the multi-objective framework. Presumably, inclusion of more periods of varying length in the analysis could reveal some patterns in the parameter variability (e.g. correlation to meteorological characteristics of a calibration period).

– Impact of the model structural complexity on the variability in model parameters with calibration period has not been sufficiently explored. The variability in spatially distributed models has not been explored heretofore.

1.6.2. Specific aims and hypotheses

Considering wide practical application of hydrologic models, it is quite important to analyse sensitivity of parameter estimates to the calibration period. Goal of this research is to further examine consistency in conceptual hydrologic model parameter estimates. To this end the following will be analysed:

– *Temporal variability in optimal parameters obtained by multi-objective model calibration i.e. Pareto-optimal parameter set.* Which Pareto-optimal parameters are the most variable with (sensitive to) the calibration period? What consequences for model performance may arise from such variability?

– *Possible causes of the parameter estimates' variability with the calibration period.* Are there any patterns in the parameters' variability with the calibration period (e.g. length of the calibration period or its hydro-meteorological characteristics)?

– *Influence of the selection of objective functions on variability in Pareto-optimal parameters*. Does selection of the objective functions or increase in their number affects variability in the parameter estimates and, if so, which combination of objective functions results in the lowest variability?

– *Impact of the model structural complexity and spatial distribution on the optimal parameters' variability*. Does an increased model structural complexity (free model parameters), or spatial distribution of the parameters, affect the variability of Pareto-optimal parameters with the calibration period?

Therefore, the hypothesis to be evaluated in the research are as follows:

(1) Hydrologic model parameters depend on the calibration period, i.e. different calibration periods yield different estimates of the same parameter;

(2) Variability of the optimised parameters may be explained by the variation in the meteorological properties;

(3) Values of Pareto-optimal parameters and their sensitivity to the calibration period depend on the objective functions used in calibration;

(4) Variability of the Pareto-optimal parameters with the calibration period depends on the hydrologic model structural complexity.

Data (measurement) errors, and data (spatial and temporal) resolution are known to affect parameter estimates (e.g. Yapo et al., 1996; Gupta et al., 1998). However, in this research it is assumed that these errors do not affect sensitivity of the parameters to calibration period, thus this aspect is not considered.

1.6.3. Thesis outline

In this research, temporal variability of the Pareto-optimal parameter sets, i.e. their variability with the calibration period is examined. It is implicitly assumed that the Pareto sets reflect optimal parameters over given period. To obtain Pareto-optimal sets, novel 3DNet-Catch hydrologic model is calibrated in dynamic fashion by employing the AMALGAM algorithm. The model calibration results are analysed to test the hypotheses formulated in chapter 1.6.2.

The methodology employed in this research is presented in **chapter 2**.

The novel conceptual distributed hydrologic model 3DNet-Catch is presented in **chapter 2.1.** Model routines, along with their parameters, are described in detail.

Alternative model structures (three semi-lumped and a distributed one) are also elaborated on in this chapter.

The AMALAGAM, algorithm aimed at multi-objective model calibration, is briefly described in **chapter 2.2.**

The model setup (including the regularisation method for calibration of distributed version of the model), sensitivity analysis and evaluation of the 3DNet-Catch model are given in **chapter 2.3.**

Dynamic calibration procedure, adopted in this research, is presented in **chapter 2.4**, while the methods, used for analysis of the results, are outlined in **chapter 2.5**.

The methodology presented is applied to three (relatively) unchanged catchments in Serbia, namely the Kolubara River, Toplica River and Mlava River catchments. These catchments are described in **chapter 2.6**.

The results are presented and discussed in **chapter 3.**

Conclusions and recommendations for further research are given in **chapter 4**.

2. METHODOLOGY

2.1. The 3DNet-Catch conceptual hydrological model

The 3DNet-Catch is a conceptual, fully-distributed hydrological model aimed at continuous hydrologic simulations. The model comprises routines for vertical water balance simulation and runoff routing to the catchment outlet (horizontal water balance). Vertical water balance is simulated by employing the vegetation, snow and soil routines (Figure 11). The equations of these routines are applied to every cell of the computational grid. A grid cell is referred to as *Hydrologic Response Unit* – HRU[10]. Simulated runoff, which consists of the surface flow, fast shallow aquifer response and baseflow, is transformed through linear and nonlinear reservoir of the response routine. Neither lateral surface nor subsurface flow among HRUs is simulated, but from a HRU to the catchment outlet. Optionally, the surface runoff outlet does not have to coincide with the baseflow outlet, what is specified by the user. This option is rather convenient for karst catchments, enabling "soft" data (Seibert & McDonnell, 2002) on groundwater flow to be incorporated in the model. In addition, flow propagation along river reaches is simulated by employing the flow routing routine.

2.1.1. Model description: equations and model parameters

Model equations and parameters are presented in this chapter. All state variables are estimated at the end of a computational time step (denoted by subscript (i)), while the fluxes represent mean values over the time step (denoted by i). All water balance equations of the interception, snow and soil routines refer to the unit area of a catchment.

[10] HRUs are comprised of points that exhibit hydrologically similar behaviour (Beven, 2001b).

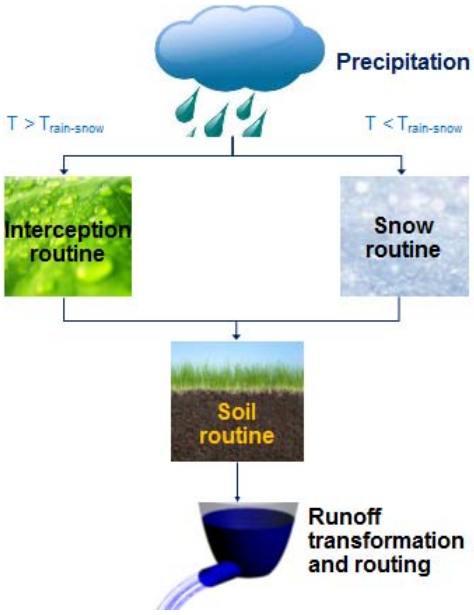


Figure 11. Routines of the 3DNet-Catch model.

Interception routine

Interception of precipitation depends primarily on the type of vegetative cover (Jovanović & Radić, 1990; Musy & Higy, 2011). The vegetative cover in the 3DNet-Catch model is represented by a single reservoir (canopy or interception storage) with maximum capacity equal to *CAN* (Figure 12). Maximum capacity of the canopy storage varies over the growing season along with the leaf development, which is quantified in terms of the Leaf Area Index (*LAI*):

$$CAN_{(i)} = CAN_{\max} \frac{LAI_{(i)}}{LAI_{\max}} \tag{2.1.1}$$

$CAN_{(i)}$ and $LAI_{(i)}$ denote capacity of the canopy reservoir and the value of the Leaf Area Index in the i^{th} time step, respectively. Correspondingly, $CAN_{\max}$ and $LAI_{\max}$ represent maximum capacity of the canopy reservoir and maximum value of the Leaf Area Index in the growing season. *LAI* values can be introduced to the model as input time series, or they can be calculated as a sine curve over the growing season.

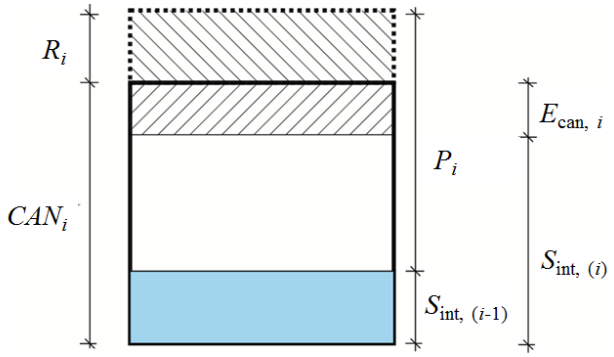


Figure 12. The 3DNet-Catch model: the canopy reservoir.

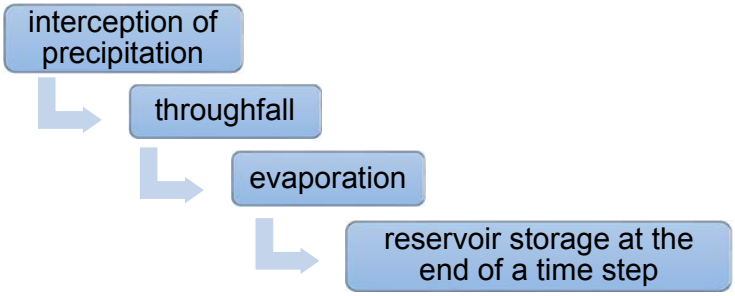


Figure 13. The 3DNet-Catch model: processes of the interception routine.

Water balance of the canopy reservoir consists of precipitation as the input, and throughfall and evaporation as the reservoir output (Figure 12). These processes are simulated according to the scheme given in Figure 13 by employing the following equations:

– ***Throughfall*** over the i^{th} time step, R_i:

$$R_i = \min\left[\max\left(0;\ S_{\text{int},(i-1)} + P_i - CAN_{(i)}\right);\ P_i\right] \tag{2.1.2}$$

In the above equation $S_{int,\ (i\text{-}1)}$ denotes the canopy reservoir storage at the end of the previous time step, P_i is observed precipitation depth in current, i^{th}, time step and $CAN_{(i)}$ is the capacity of the canopy reservoir in current time step.

– Canopy storage after interception, $S_{int,(i)}^{*}$:

$$S_{\text{int},(i)}^{*} = S_{\text{int},(i-1)} + P_i - R_i \tag{2.1.3}$$

– ***Evaporation*** from the canopy reservoir over the i^{th} time step, $E_{can,\ i}$:

$$E_{\mathrm{can},i} = \min\left(S_{\mathrm{int},(i)}^{*};\ PET_i \right) \tag{2.1.4}$$

$E_{can,i}$ denotes actual evaporation form the canopy reservoir, and PET_i is potential evapotranspiration over the i^{th} time step.

– The canopy reservoir storage at the end of the i^{th} time step, $S_{int,(i)}$:

$$S_{\mathrm{int},(i)} = S_{\mathrm{int},(i)}^{*} - E_{\mathrm{can},i} \tag{2.1.5}$$

Variables, parameters and initial conditions of this routine are given in Table 3.

Table 3. Overview of the state and dependent variables, fluxes, free parameters and initial conditions of the interception routine.

State variables		
S_{int}	Canopy reservoir storage	[mm]
Dependent variables		
CAN	Current maximum capacity of the vegetation reservoir	[mm]
LAI	Current Leaf Area Index value	[m^2m^{-2}]
Fluxes		
P	Total precipitation depth over a time step	[mm Δt^{-1}]
R	Throughfall over a time step	[mm Δt^{-1}]
E_{can}	Evaporation from the canopy reservoir	[mm Δt^{-1}]
Parameters		
CAN_{max}	Maximum interception reservoir capacity	[mm]
LAI_{max}	Maximum Leaf Area Index value	[m^2m^{-2}]
Initial conditions		
$S_{int\ (i=0)}$	Amount of water in the interception reservoir at the beginning of a simulation	[mm]

Snow routine

Precipitation that occurs at air temperature below $T_{S\text{-}R}$ (threshold temperature) is considered snow; otherwise it is treated as rainfall. Mixture of rainfall and snow (sleet) is not recognised in the model.

This routine of the 3DNet-Catch model is similar to the snow routine of the SWAT model (Neitsch et al., 2011). It is based on the degree-day method, which is preferred over the energy-balance methods because of modest data requirements (only air temperature is required), overall satisfactory performance and computational simplicity (He et al., 2014).

According to the degree-day method, snow ablation is proportional to the difference between air temperature and temperature at which snow melts (e.g. Bergström et al., 1992; Beven, 2001b; Anderson, 2006; Seibert and Vis, 2012, He et al., 2014), which is usually a free model parameter. Water balance in the snowpack (Figure 14) consists of the following components: interception of precipitation, snow sublimation and snow melt (Neitsch et al., 2011). Snowmelt refreezing and meltwater retainment by the snowpack are not taken into account in the model. The processes of this routine are simulated according to scheme given in Figure 15. Water balance components of the snowpack routine are expressed in millimetres of water equivalent.

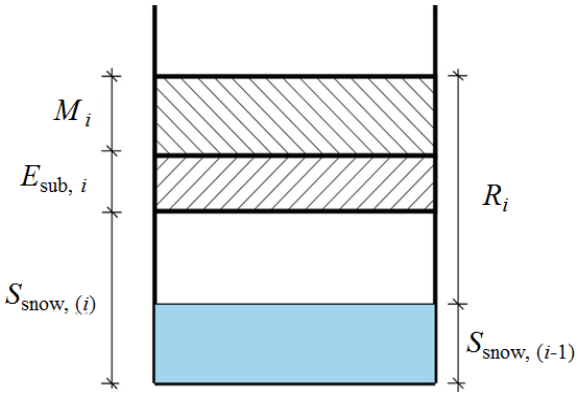


Figure 14. The 3DNet-Catch model: the snowpack reservoir.

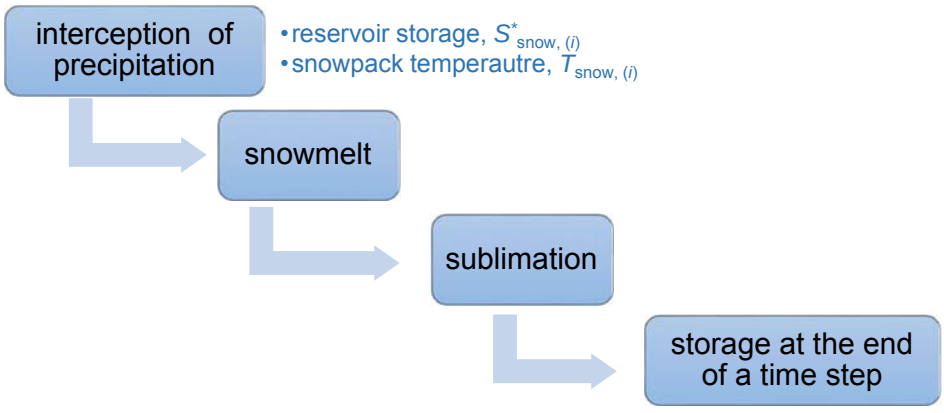


Figure 15. The 3DNet-Catch model: processes of the snow routine.

The governing equations of the snowpack routine are:

– The snowpack storage after intercepting the precipitation P_i , $S_{snow,(i)}{}^*$:

$$S_{snow,(i)}{}^* = S_{snow,\ (i-1)} + P_i \qquad (2.1.6)$$

where $S_{\text{snow, }(i\text{-}1)}$ is the snowpack storage at the end of previous time step, and P_i denotes total observed precipitation depth over current time step. The amount of precipitation intercepted by vegetation is assumed negligible because of leaf abscission (prevailing deciduous forests are assumed). Therefore, the interception routine is not invoked at temperatures below $T_{S\text{-}R}$ (as shown in Figure 11) and total precipitation depth appears in the balance equation instead of throughfall.

– Temperature of the snowpack at the end of current time step, $T_{\text{snow},(i)}$, is obtained by weighting the snowpack temperature at the end of previous time step, $T_{\text{snow},(i\text{-}1)\text{ and}}$ mean air temperature in the current time step (Beven 2001a; Neitsch et al., 2011):

$$T_{\text{snow},(i)} = \left(1-\lambda_{\text{snow}}\right)\cdot T_{\text{snow, }(i-1)} + T_{a,i}\,\lambda_{\text{snow}} \tag{2.1.7}$$

where $T_{a,i}$ denotes mean air temperature over a time step and λ_{snow} is the snowpack temperature lag factor which takes value between 0 and 1 (Zhang et al., 2009). Larger values of this factor imply greater influence of air temperature (Neitsch et al., 2011). The impact of the air temperature is inversely proportional to the snowpack thickness, and therefor smaller values of the λ_{snow} factor correspond to thicker snowpack, and vice-versa (Melloh, 1999).

– ***Snow melt*** in the i^{th} time step, M_i:

$$M_i = \min\left\{\max\left[0;\ b_{\text{melt}}\cdot snow_{\text{cov},i}\cdot\left(\frac{T_{\text{snow},i}+T_{a,i}}{2}-T_{\text{melt}}\right)\right];\ S_{\text{snow},i}^{*}\right\} \tag{2.1.8}$$

where T_{melt} is the threshold temperature at which snow ablation begins, and it is free model parameter. The value of this parameter should be set depending on what T_a in equation 2.1.8 stands for, either maximum or mean daily temperature. If T_a stands for mean daily temperature, T_{melt} should approximately be 0°C, otherwise, if maximum daily temperature is used, the value of T_{melt} should be somewhat larger, up to 4.4 °C (U.S. A.C.E., 1994).

The b_{melt} parameter is the melt (degree-day) factor [mm°C^{-1}day^{-1}] and it shapes the relation between air temperature and snow ablation. This relation is highly nonlinear, so the values of b_{melt} vary in time (Hock, 2003; He et al., 2014). For example, b_{melt} increases vastly in rainy conditions (rain-on-snow events) (Melloh, 1999; Hock, 2003). Also b_{melt}

increases over the melting season due to increasing albedo (Beven, 2001b; Anderson, 2006; Neitsch et al., 2011). In addition to climate conditions, the melt factor heavily depends on catchment properties. For example, b_{melt} increases with elevation, and depends on land use in the catchment (e.g. b_{melt} takes smaller values in forest-prevailed areas, and larger values in the urbanised areas) (Neitsch et al., 2011; He et al., 2014). Correlation of b_{melt} with net shortwave radiation, wind velocity, vapour pressure, insolation, albedo, terrain elevation, aspect and shading has been reported in the literature (Hock, 2003).

In the 3DNet-Catch model b_{melt} is assumed to vary seasonally, from the 21st of December (minimum value) to the 21st of June (maximum value), according to the sine curve (Braun et al., 1993; Hock, 2003; Anderson, 2006):

$$b_{\text{melt},i} = \frac{b_{\text{melt},6} + b_{\text{melt},12}}{2} + \frac{b_{\text{melt},6} - b_{\text{melt},12}}{2} \sin\left(\frac{2\pi}{365}\left(D_{n,i} - 81\right)\right) \tag{2.1.9}$$

where D_n denotes ordinal number of day in a year, while $b_{\text{melt, 6}}$ and $b_{\text{melt, 12}}$ are free model parameters.

Variations in b_{melt} due to other factors (primarily an increase over rainy periods) are not modelled in the 3DNet-Catch.

– The $snow_{\text{cov}}$ variable represents share of the catchment area covered with snow over the i^{th} time step, and it is estimated as:

$$snow_{\text{cov},i} = \min\left(\frac{S_{\text{snow},i}^{*}}{S_{\text{snow, 100}}};\ 1\right) \tag{2.1.10}$$

The snowpack thickness is rarely uniform over a catchment due to topography, wind drift, vegetation, aspect, etc. (Beven, 2001b; Anderson, 2006; Neitsch et al., 2011), meaning that not the entire catchment area is necessarily covered in snow. To account for this, the $S_{\text{snow,100}}$ parameter is introduced. This parameter is a threshold value of the snowpack storage (expressed in millimetres of water equivalent) at which the entire area of a catchment is certainly covered in snow (Neitsch et al., 2011).

– The snowpack storage after snow ablation, $S_{\text{snow},(i)}^{**}$:

$$S_{\text{snow},(i)}^{**} = S_{\text{snow},(i)}^{*} - M_i \tag{2.1.11}$$

– ***Sublimation*** from the snowpack over current time step, $E_{\text{sub},i}$:

$$E_{\text{sub},i} = \min\left(S_{\text{snow},i}^{**};\ PET_i \cdot \text{cov}_{soil,i}\right) \tag{2.1.12}$$

where cov_{soil} is the soil cover index that quantifies the share of bare soil (not covered with vegetation) in the catchment area. Soil cover index equal to 1 implies bare soil. If $S_{\text{snow},i}^{**}$ exceeds 0.5 mm, the value of $\text{cov}_{soil,i}$ is set to 0.5 (Neitsch et al., 2011). Otherwise, it is calculated based on the *LAI* value:

$$\text{cov}_{soil,i} = \exp\left(-0.4 \cdot LAI_i\right) \tag{2.1.13}$$

– The storage of the snow reservoir at the end of the current time step $S_{\text{snow},(i)}$:

$$S_{\text{snow},(i)} = S_{\text{snow},(i)}^{**} - E_{sub,i} \tag{2.1.14}$$

The variables, parameters and initial conditions of the snow routine are given in Table 4.

Table 4. Overview of the state and dependent variables, free parameters and initial conditions of the snow routine.

State variables		
$S_{\text{snow},(i)}$	Snowpack storage expressed in mm of water equivalent	[mm]
$T_{\text{snow},(i)}$	Temperature of the snowpack	[°C]
Dependent variables		
b_{melt}	Melt (degree-day) factor	[mm°C^{-1}day^{-1}]
cov_{soil}	Soil cover index	[-]
Fluxes		
P	Total precipitation depth over a time step	[mm Δt^{-1}]
M	Snow melt over a time step (in mm of water equivalent)	[mm Δt^{-1}]
E_{sub}	Snowpack sublimation over a time step (water equivalent)	[mm Δt^{-1}]
Parameters		
$T_{\text{S-R}}$	Boundary temperature	[°C]
$S_{\text{snow},100}$	Threshold snowpack storage at which the entire catchment is covered in snow	[mm]
λ	Snowpack temperature lag factor	[-]
T_{melt}	Threshold temperature at which snowmelt begins	[°C]
$b_{\text{melt},6}$	Melt factor on 21st of June	[mm°C^{-1}day^{-1}]
$b_{\text{melt},12}$	Melt factor on 21st of December	[mm°C^{-1}day^{-1}]
Initial conditions		
$S_{\text{snow},(i=0)}$	The snowpack storage at the beginning of simulation	[mm]
$T_{\text{snow}(i=0)}$	Temperature of the snowpack at the beginning of a simulation	[°C]

Soil moisture routine

The soil column is represented in the 3DNet-Catch by a surface layer and N_l sub-surface ones (Figure 16). The processes simulated within the soil routine are evapotranspiration, surface runoff and percolation. Surface runoff, which consists of the infiltration excess overland flow and saturation excess flow, is generated in the surface soil layer. As for *ET*, water evaporates from the surface layer, whereas transpiration takes place in the sub-surface ones (Figure 16). Water percolates from every layer into a deeper one, and, eventually into the nonlinear groundwater reservoir.

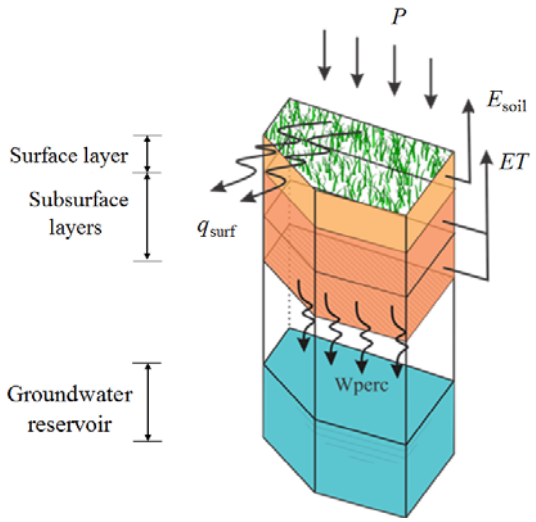


Figure 16. The 3DNet-Catch model: the soil column representation in the soil moisture routine (surface and subsurface soil layers and the groundwater reservoir).

(1) WATER BALANCE OF THE SURFACE SOIL LAYER:

Water balance of the surface soil layer consists of precipitation (throughfall or sum of precipitation and snowmelt), surface runoff, percolation into deeper soil layers and soil evaporation, as shown in Figure 16 and in Figure 17.

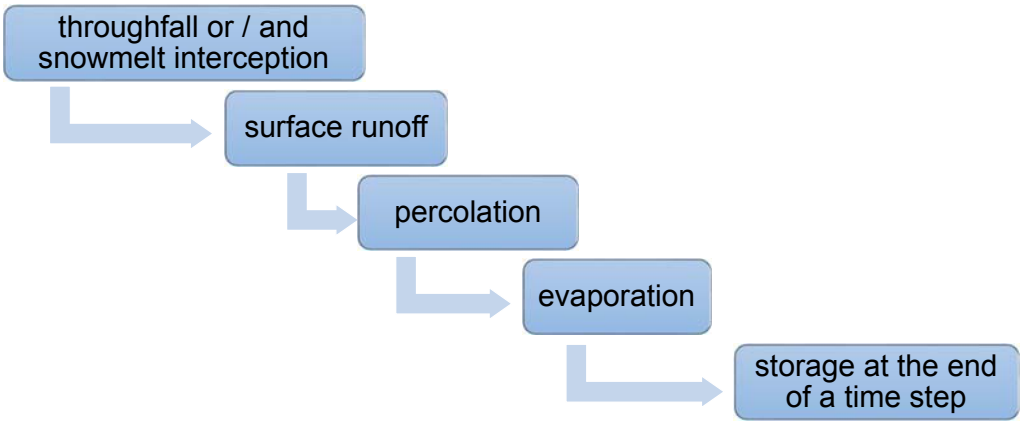


Figure 17. The 3DNet-Catch model: processes of the soil routine – surface soil layer.

– ***Surface runoff*** (q^*_{surf}) is estimated by applying the SCS method (e.g. Chow et al., 1988; Maidment, 1993; Yu, 1998; Beven, 2001b; Neitsch et al., 2011):

$$q^*_{\text{surf},i} = \begin{cases} \dfrac{(P_i - I_a)^2}{P_i - I_a + S_i} & ; \quad P_i > I_a \\ 0 & ; \quad P_i \le I_a \end{cases} \qquad (2.1.15)$$

where I_a stands for initial abstraction, S is potential soil retention in a time step (mm) and P_i denotes equivalent precipitation: throughfall or precipitation and snowmelt over a time step, depending on the meteorological conditions (Schaefli et al., 2014).

Initial abstraction I_a varies with the precipitation amount (e.g. Jovanović and Radić, 1990; White et al., 2009), and it is commonly assumed to be 0.2 of potential soil retention, S (Chow et al., 1988; U.S. A.C.E., 1994). Since I_a encompasses, *inter alia*, rainfall interception by the vegetative cover, which is explicitly simulated in the 3DNet-Catch, value of I_a is reduced by the simulated interception:

$$I_{a,i} = \max\left(0;\ I_{a\_REL}\, S_i - \left(P_i - R_i\right)\right) \qquad (2.1.16)$$

I_{a_REL} denotes assumed initial abstraction value (dimensionless, as multiplier of S) and it is a free model parameter. Difference (P_i – R_i) represents the amount of precipitation intercepted by vegetation.

Maximum potential soil retention S is related to the Curve Number (CN):

$$S = 25.4\left(\frac{1000}{CN} - 10\right) \tag{2.1.17}$$

The value of the *CN* depends on the land use type, soil properties, antecedent soil moisture, slope of the area, etc. (e.g. Chow et al., 1988; Maidment, 1993; Beven, 2001b; Neitsch et al., 2011). *CN* values that may be found in the literature are estimated for the 5% slope areas and average antecedent soil moisture conditions (AMC II). Hence, two corrections to *CN* value are made in the model to account for actual terrain slope and soil wetness (Neitsch et al., 2011):

- Correction to account for actual terrain slope, *ST* [-]:

$$S_{2s} = S \cdot \left[1.1 - \frac{ST}{ST + \exp(3.7 + 0.02117ST)}\right] \tag{2.1.18}$$

$$CN_{2S} = 1000 \cdot \left[\frac{S_{2S}}{25.4} + 10\right]^{-1} \tag{2.1.19}$$

- Correction to account for actual soil moisture conditions:

$$CN_1 = CN_{2s} - 20 \cdot \frac{100 - CN_{2s}}{100 - CN_{2s} + \exp\left(2.533 - 0.0636 \cdot \left(100 - CN_{2s}\right)\right)} \tag{2.1.20}$$

$$CN_3 = CN_{2s} \exp\left(0.00673\left(100 - CN_{2s}\right)\right) \tag{2.1.21}$$

where CN_1 corresponds to the minimum soil wetness (permanent wilting point) and CN_3 corresponds to maximum soil wetness (Neitsch et al., 2011; Zhang and Shuster, 2014).

Maximum and minimum potential retentions are calculated using the obtained *CN* values:

$$S_{\max} = 25.4\left(\frac{1000}{CN_1} - 10\right) \tag{2.1.22}$$

$$S_{\min} = 25.4\left(\frac{1000}{CN_3} - 10\right) \tag{2.1.23}$$

The SCS method is aimed at event-based modelling. To enable continuous runoff simulations, potential soil retention *S* should vary with the soil moisture content (Neitsch et al., 2011):

$$S_i = \sum_{j=1}^{N_l+1} s_j \left(STO_j - SW_{j,(i-1)} \right) \tag{2.1.24}$$

where $SW_{j,(i\text{-}1)}$ denotes storage of the j^{th} soil layer (j = 1 refers to the surface layer) at the end of previous time step, STO_j is maximum storage capacity of the j^{th} layer, i.e. the product of the soil layer thickness (D_j) and its effective porosity (p_j), both being free model parameters. Share of the j^{th} soil layer in the active soil layer, s_j (Figure 18), are estimated as following:

$$s_j = \left\{ \begin{array}{l} 0 \quad ; S_{\max} \le D_1 \left(p_1 - w_{WP,1} \right) \\ \min \left(1; \; \dfrac{S_{\max} - \sum_{j=1}^{N_l} D_j \left(p_j - w_{WP,j} \right)}{D_j} \right) ; \; S_{\max} > D_1 \left(p_1 - w_{WP,1} \right) \end{array} \right\} \tag{2.1.25}$$

where $w_{WP,1}$ stands for the wetness at permanent wilting point, which is minimum volumetric soil water content at which plant would not wilt (Shaw, 2005) and it is commonly estimated at pressure of -15 bars (Campling et al., 2002; White et al., 2009; Scorza Júnior and Silva, 2011; Diallo and Mariko, 2013; Yang and You, 2013). In the 3DNet-Catch, permanent wilting point w_{WP} is a free model parameter that may vary with the soil layer.

Thickness of the active soil layer (layer that determines surface runoff) is then:

$$D_{sum\_SCS} = \sum_{j=1}^{N_l+1} s_j \, D_j \tag{2.1.26}$$

Storage of the surface soil layer after surface runoff has taken place is:

$$SW_{1,(i)}^{*} = SW_{1,\,(i-1)} + (P_i - q^{*}_{surf,i}) \tag{2.1.27}$$

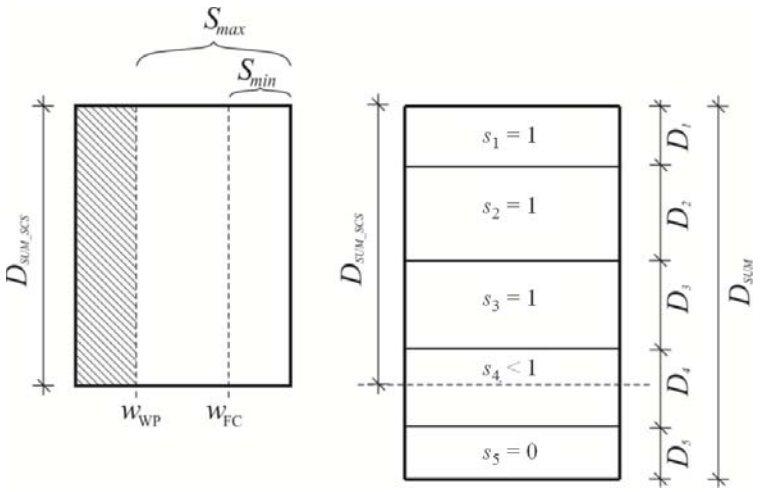


Figure 18. Thickness of the active soil layer.

– ***Percolation* (W_{perc})** is simulated assuming that water percolates into deeper soil layers until the residual wetness (w_{res}) is reached (Figure 19). At soil wetness less than w_{res}, water can be removed from the surface soil layer only by evaporation.

Percolation rate depends on the soil moisture (the largest rate occurs at soil saturation) and it is calculated as:

$$W_{perc1,i} = \begin{cases} K_{sat,1}\,\Delta T_{sat,i} + \left(STO_1 - WP_1\right)\left(S_{r1,i} - \left(S_{r1,i}^{(1-n)} + \dfrac{K_{sat,1}}{STO_1 - WP_1}\left(n_1 - 1\right)\Delta T_{unsat,i}\right)^{\frac{1}{1-n_1}}\right); SW_{1,(i)}{}^{*} > W_{res} \\ 0 \qquad\qquad ; SW_{1,(i)}{}^{*} \le W_{res} \end{cases} \tag{2.1.28}$$

Index 1 in the previous equation means that variables and parameters refer to the surface soil layer. The first term in the right-hand side of the equation quantifies percolation from the saturated soil layer. K_{sat} stands vertical permeability (hydraulic conductivity) at soil saturation, and it is a free model parameter. ΔT_{sat} denotes duration of percolation form the saturated layer (Figure 19):

$$\Delta T_{sat,i} = \begin{cases} \min\left(\dfrac{SW_{1,(i)}^{*} - STO_1}{K_{sat,1}};\ \Delta t \right) & ;\ SW_{1,(i)}^{*} \geq STO_1 \\ 0 & ;\ SW_{1,(i)}^{*} < STO_1 \end{cases} \tag{2.1.29}$$

The remaining term in equation 2.1.28 denotes percolation from the unsaturated soil layer, which is obtained by solving Richard's equation for vertical direction, with saturated hydraulic conductivity being estimated using van Genuchten equation[11]. The residual water content θ_r in the van Genuchten equation is assumed equal to the wetness at permanent wilting point, $w_{WP,1}$ (van Genuchten, 1980).

The residual soil wetness (W_{res}), which limits percolation, is estimated based on the soil wetness at field capacity (w_{FC}) and at permanent wilting point:

$$w_{res} = \frac{w_{FC,1} - w_{WP,1}}{p - w_{WP,1}} \tag{2.1.30}$$

The wetness at field capacity (w_{FC}), i.e. at the equilibrium of the capillary and gravitation forces, is estimated at 0.33 bar (Yang and You, 2013; Diallo and Mariko, 2013) In the 3DNet-Catch w_{FC} is a free model parameter.

$$W_{res} = w_{res} \cdot D_1 \tag{2.1.31}$$

In equation 2.1.28, $S_{r1,i}$ stands for the ratio between the available and maximum amount of water for percolation in a time step:

$$S_{r1,i} = \min\left(\frac{SW_{1,(i)}^{*} - WP_1}{STO_1 - WP_1};\ 1 \right) \tag{2.1.32}$$

where WP_1 is estimated from the wetness at wilting point and the layer thickness:

$$WP_1 = w_{wp} \cdot D_1 \tag{2.1.33}$$

[11] Simulation of flow in the unsaturated zone is carried out in similar manner in the MIKE SHE model (Madsen, 2003).

Parameter n_1 in equation 2.1.28 is van Genuchten pore-size distribution index, and it is greater than 1 (van Genuchten, 1980; Schaap et al., 2001). The values of this dimensionless parameter decrease with increase of clay and silt content in the soil (finer soil textures) and increase with the sand content (Schaap at al., 2001; Porêbska et al., 2006). However, no systematic change with soil depth has been proven (Porêbska et al., 2006; Scorza Júnior and Silva, 2011). This parameter can take values up to 10 (van Genuchten, 1980; Yang & You, 2013), although values greater than 2 are not recommended for finer soils (Schaap at al., 2001; Durner and Fluhler, 2005).

ΔT_{unsat} is the complement of ΔT_{sat} to the length of simulation time step:

$$\Delta T_{unsat,i} = \Delta t - \Delta T_{sat,i} \tag{2.1.34}$$

After percolation has taken place, the surface layer storage is calculated as the following:

$$SW_{1,(i)}^{**} = SW_{1,(i)}^{*} - W_{\text{perc }1,i} \tag{2.1.35}$$

If simulated storage $SW_{1,(i)}^{**}$ exceeds maximum capacity STO_1 of the layer, storage is set to the value of STO_1 and the excess water is added to surface runoff q^{*}_{surf}, mimicking saturation excess runoff generation mechanism (Beven, 2001b):

$$SW_{1,(i)}^{***} = \begin{cases} SW_{1,(i)}^{**}, \; q_{\text{surf},i}^{**} = q_{\text{surf},i}^{*} & ; SW_{1,(i)}^{**} \le STO_1 \\ STO_1, \; q_{\text{surf},i}^{**} = q_{\text{surf},i}^{*} + \left(SW_{1,(i)}^{**} - STO_1 \right) & ; SW_{1,(i)}^{**} > STO_1 \end{cases} \tag{2.1.36}$$

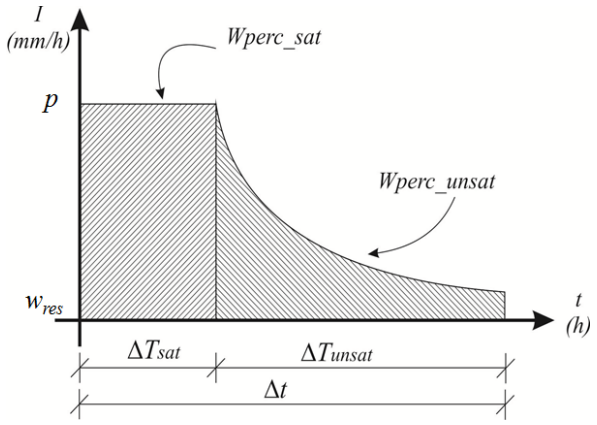


Figure 19. Percolation from the soil.

– ***Evaporation*** from the surface soil layer E_{soil} over the i^{th} time step is calculated using:

$$E^{*}_{\text{soil},i} = PET_i^{*} \, \text{cov}_{soil,i} \tag{2.1.37}$$

The soil cover index, cov_{soil}, is estimated based on the *LAI* value (equation 2.1.13), while PET_i^* is potential *ET* after evaporation form the canopy or snowpack reservoir occurred:

$$PET_i^{*} = PET_i - E_{\text{can},i} - E_{\text{sub},i} \tag{2.1.38}$$

Additionally, actual evaporation is limited by available amount of water in the surface layer. Therefore, $E_{\text{soil},i}^{*}$ is reduced if the water content is below FC_1 (Neitsch et al., 2011):

$$E_{\text{soil},i}^{***} = \begin{cases} E_{\text{soil},i}^{**} \cdot \exp\left(\dfrac{2.5 \, (SW_{1,(i)}^{***} - FC_1)}{FC_1 - WP_1} \right) ; & SW_{1,(i)}^{***} < FC_1 \\ E_{\text{soil},i}^{**} ; & SW_{1,(i)}^{***} > FC_1 \end{cases} \tag{2.1.39}$$

FC_1 is the storage (in millimetres per unit area) of the surface soil layer at field capacity:

$$FC_1 = w_{FC,1} \, D_1 \tag{2.1.40}$$

Maximum amount of water that can evaporate from the soil layer is limited to 80% of the amount of water available to plants[12] (Neitsch et al., 2011):

$$E_{\text{soil},i} = \min\left(E_{\text{soil},i}^{***} ; \; 0.8 \left(SW_{1,(i)}^{***} - WP_1 \right) \right) \tag{2.1.41}$$

Having the actual evaporation from the surface layer calculated, storage of the surface soil layer at the end of current time step $SW_{(i)}$ is estimated as:

$$SW_{(i)} = SW_{1,(i)}^{***} - E_{\text{soil},i}^{***} \tag{2.1.42}$$

(2) WATER BALANCE OF THE SUBSURFACE SOIL LAYERS

Water balance of a sub-surface layer is comprised of percolation from the upper soil layer, transpiration and percolation into deeper soil layer, if any, or into the groundwater non-liner reservoir, according to Figure 20.

[12] In this model, this value is set to 80%, although it may be a free parameter (as in e.g. the HBV model).

– A sub-surface layer storage after percolation form upper layer $W_{perc\,(l\text{-}1),i}$:

$$SW_{l,(i)}^{*} = SW_{l,(i-1)} + W_{perc\,(l-1),\,i} \tag{2.1.43}$$

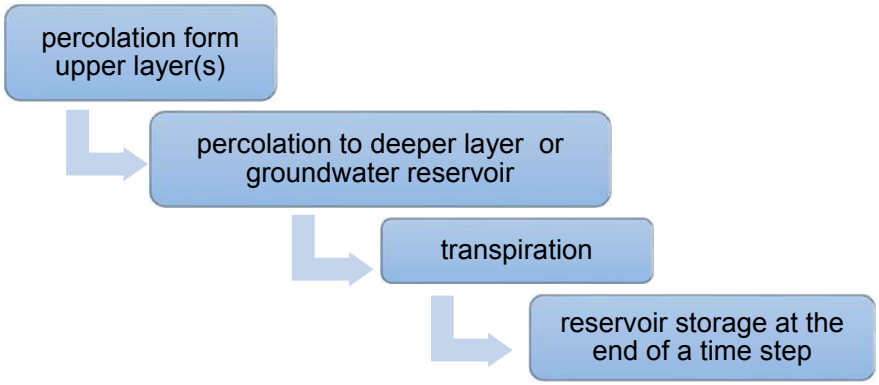


Figure 20. The 3DNet-Catch model: processes of the soil routine – subsurface layer.

– ***Percolation from the l^{th} sub-surface layer*** is simulated in the same manner as the percolation form the surface soil layer:

$$W_{perc,l,i} = \begin{cases} K_{sat,l}\,\Delta T_{sat,i} + \left(STO_l - WP_l\right)\left(S_{r,l,i} - \left(S_{r,l,i}^{(1-n_l)} + \dfrac{K_{sat,l}}{STO_l - WP_l}\left(n_l - 1\right)\Delta T_{unsat,i}\right)^{\frac{1}{1-n_l}}\right) & ;\ SW_{l,(i)}^{*} > FC_l \\ 0 & ;\ SW_{l,(i)}^{*} \le FC_l \end{cases} \tag{2.1.44}$$

Sub-surface layer storage after percolation into deeper layer is:

$$SW_{l,(i)}^{**} = SW_{l,(i)}^{*} - W_{perc,l,i} \tag{2.1.45}$$

If excess water occurs in the sub-surface layers, it is added to surface runoff, similarly to modelling of the surface layer balance:

$$SW_{l,(i)}^{***} = \begin{cases} SW_{l,(i)}^{**}\ ,\ q_{\mathrm{surf},i} = q_{\mathrm{surf},i}^{**} & ;\ SW_{l,(i)}^{**} \le STO_l \\ STO_l\ \ ,\ q_{\mathrm{surf},i} = q_{\mathrm{surf},i}^{**} + \left(SW_{l,(i)}^{**} - STO_l\right) & ;\ SW_{l,(i)}^{**} > STO_l \end{cases} \tag{2.1.46}$$

– ***Transpiration*** (water uptake by plants) is a difference between *PET* and actual evaporation from the surface layer over the i^{th} time step:

$$E_{t,\,i} = PET_i^{*} - E_{\mathrm{soil},i} \tag{2.1.47}$$

It is assumed that water uptake from l^{th} sub-surface later is proportional to its share in the total sub-surface thickness:

$$W_{up,l,i}{}^{*} = \frac{E_{t,i}}{\sum_{j=2}^{N_l} D_l} D_l \tag{2.1.48}$$

Transpiration is limited by available water within the layer, so the following limitations are imposed (Neitsch et al., 2011):

$$W_{up,l,i}{}^{**} = \begin{cases} W_{up,l,i}{}^{*} \exp\left(5\left(\dfrac{SW_{l,(i)}{}^{**} - WP_l}{0.25\left(FC_l - WP_l\right)} - 1\right)\right) & ;\ \left(SW_{l,(i)}{}^{**} - WP_l\right) \le 0.25\left(FC_l - WP_l\right) \\ W_{up,l,i}{}^{*} & ;\ \left(SW_{l,(i)}{}^{**} - WP_l\right) > 0.25\left(FC_l - WP_l\right) \end{cases} \tag{2.1.49}$$

$$W_{up,l,i}{}^{***} = \min\left(W_{up,l,i}{}^{**};\ \left(SW_{l,(i)}{}^{**} - WP_l\right);\ 0\right) \tag{2.1.50}$$

Storage in the l^{th} sub-surface layer at the end of i^{th} simulation time step is:

$$SW_{l,(i)} = SW_{l,(i)}{}^{**} - W_{up,l,i}{}^{***} \tag{2.1.51}$$

Features of the soil routine are presented in Table 5.

Table 5. Overview of the state and dependant variables, fluxes, parameters and initial conditions of soil routine.

State variables		
SW_l	Storage of a layer (in millimetres per unit area)	[mm]
Dependent variables		
I_a	Initial abstraction	[mm]
CN_{2S}	Value of the CN corrected with respect to slope of the area	[-]
CN_1	Value of the CN at permanent wilting point	[-]
CN_3	Value of the CN at field capacity	[-]
S_{max}	Potential retention at permanent wilting point	[mm]
S_{min}	Potential retention at field capacity	[mm]
D_{sum_SCS}	Thickness of the active soil layer	[mm]
WP_l	The l^{th} layer storage at permanent wilting point	[mm]
FC_l	The l^{th} layer storage at field capacity	[mm]
$S_{r,\,l}$	Ratio between available and maximum amount of water to percolate	[mm]
Fluxes		
P_i	Equivalent precipitation	[mm Δt^{-1}]
q_{surf}	Surface runoff	[mm Δt^{-1}]
$W_{perc,l}$	Percolation into l^{th} layer from the $(l-1)^{st}$ one ($l>1$)	[mm Δt^{-1}]
E_{soil}	Evaporation form the surface layer	[mm Δt^{-1}]
$W_{up,l}$	Water uptake (transpiration) from l^{th} sub-surface layer ($l>1$)	[mm Δt^{-1}]
Parameters		
CN	Curve number	[-]
$I_{a\_relative}$	Initial abstraction coefficient	[-]
D_{SURF}	Thickness of the soil layer	[m]
p_{SURF}	Effective porosity of the surface layer	[-]
$w_{WP,SURF}$	Permanent wilting point of the surface layer	[-]
$w_{FC,SURF}$	Soil wetness of the surface layer at field capacity	[-]
n_{SURF}	Pore-size distribution index of the surface layer	[-]
$K_{sat,surf}$	Saturated hydraulic conductivity of the surface layer	[m Δt^{-1}]
N_l	Number of sub-surface layers	[-]
$D_{SUB\text{-}SURF,l}$	Thickness of l^{th} sub-surface layer ($l>1$)	[m]
$p_{\,SUB\text{-}SURF,l}$	Effective porosity of the l^{th} sub-surface layer ($l>1$)	[-]
wp_l	Permanent wilting point of the l^{th} sub-surface layer ($l>1$)	[-]
$w_{FC,l}$	Soil wetness of the l^{th} layer at field capacity	[-]
n_l	Pore-size distribution index of the l^{th} sub-surface layer ($l>1$)	[-]
$K_{sat,l}$	Saturated hydraulic conductivity of the l^{th} sub-surface layer	[m Δt^{-1}]
Initial conditions		
$SW_{l\,(i=0)}$	The storage of every soil layer at the beginning of a simulation	[mm]

Response routine

Simulated surface runoff and percolation from the deepest soil layer are routed to the catchment outlet by applying both linear and nonlinear outflow equations. Surface runoff resulting in direct flow is routed through a linear reservoir. Water from the deepest soil layer percolates into a nonlinear groundwater reservoir, and certain amount of that water (below threshold value S_{max}) is transformed by the nonlinear outflow equation resulting in baseflow. The amount of water that exceeds the threshold value S_{max} is routed through a liner reservoir (Figure 21), resulting in fast groundwater discharge.

The routing equations are solved analytically in the 3DNet-Catch model instead of numerically since several problems are associated with numerical solutions. The explicit numerical schemes are proven to cause non-smoothness of the response surface, which significantly makes model calibration more complex (e.g. Kavetski et al., 2006; Kavetski & Clark, 2010). Implicit schemes, on the other hand, are unconditionally stable, but they require iterative solving procedure, leading to an increased computational burden and more time consuming simulations (Hirsch, 2007).

As the analytical solutions of the integrals over time are derived, state variables are denoted as functions of time t. For the sake of consistency, the fluxes over a time step are denoted with $(t \rightarrow t+\Delta t)$ in the following text.

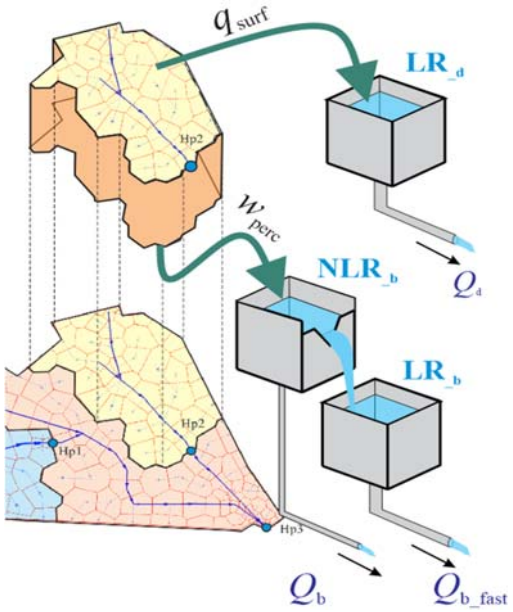


Figure 21. The 3DNet-Catch: runoff routing.

(1) ***Direct flow Q_d.*** Surface runoff q_{surf} is routed through a linear reservoir. Surface runoff from the drainage area is inflow to the reservoir, and direct flow is the outflow. The linear reservoir differential equation is:

$$\frac{d\,S_d(t)}{dt} = Q_{\text{surf}}(t) - Q_d(t) \tag{2.1.52}$$

where Q_{surf} is average surface runoff from the drainage area A over a time step Δt, assuming that surface runoff from the unit area q_{surf} is constant over the computational time step:

$$Q_{\text{surf},t\to t+\Delta t} = \frac{q_{\text{surf},t\to t+\Delta t}\,A}{\Delta t} \tag{2.1.53}$$

Direct flow is obtained from the linear outflow equation:

$$Q_d(t) = \frac{1}{K_d} S_d(t) \tag{2.1.54}$$

where K_d stands for the linear reservoir coefficient. Substituting equation 2.1.54 in 2.1.52 yields a first-order inhomogeneous ordinary differential equation (ODE):

$$\frac{d\,S_d(t)}{dt} + \frac{1}{K_d} S_d(t) = Q_{\text{surf}} \tag{2.1.55}$$

Multiplying all terms in the previous equation with the integration factor e^{t/K_d} yields:

$$\frac{d}{dt}\left[S_d \exp\left(\frac{t}{K_d}\right)\right] = \frac{1}{K_d} Q_{\text{surf}} \exp\left(\frac{t}{K_d}\right) \tag{2.1.56}$$

Direct flow volume V_d in a time step is obtained by integrating flow over a time step:

$$V_d(t) = \int_{t=0}^{t=\Delta t} \frac{S_d(t)}{K_d} = \int_{t=0}^{t=\Delta t} \left[S_{d,0} \exp\left(\frac{t}{K_d}\right) + Q_{\text{surf}}\,K_d\left(1-\exp\left(\frac{t}{K_d}\right)\right)\right] dt \tag{2.1.57}$$

The direct flow volume (outflow from the reservoir) within a time step is then:

$$V_{d,\,t\to t+\Delta t} = S_{d,0}\,\exp\left(1-\frac{\Delta t}{K_d}\right) + Q_{\text{surf}}\,\Delta t\left(1-\frac{1-\exp\left(\frac{-\Delta t}{K_d}\right)}{\Delta t / K_d}\right) \tag{2.1.58}$$

where $S_{d,0}$ denotes storage of the liner reservoir at the beginning of a time step. Based on the V_d, mean direct flow in a time step and storage of the reservoir at the end of a time step are simulated:

$$Q_{d,\,t\to t+\Delta t} = \frac{V_{d,\,t\to t+\Delta t}}{\Delta t} \tag{2.1.59}$$

$$S_{d,\,t+\Delta t} = S_{d,t} - Q_{d,t\to t+\Delta t} + Q_{\text{surf},\,t\to t+\Delta t} \tag{2.1.60}$$

Two previous equations may be written using condensed notation:

$$Q_{d,\,i} = \frac{V_{d,\,i}}{\Delta t} \tag{2.1.61}$$

$$S_{d,\,(i)} = S_{d,\,(i-1)} - Q_{d,i} + Q_{\text{surf},\,i} \tag{2.1.62}$$

Subscript i refers to mean flux over i^{th} time step (i.e. mean direct flow over a time step), while (i) and (i-1) imply values of the state variables at the end of the current and previous time step, respectively (e.g. storage at the beginning of the i^{th} time step).

The K_d coefficient may be estimated from the average time of concentration, T_c:

$$K_d = \left(1-e^{-\bar{b}}\right) \tag{2.1.63}$$

$$\bar{b} = \frac{1}{N_{\text{HRU}}}\sum_{j=1}^{N_{\text{HRU}}}\left(1-e^{-\Delta t/T_{c,j}}\right) \tag{2.1.64}$$

Subscript j refers to a HRU, and N_{HRU} is the number of hydrologic response units draining into the outlet (linear reservoir).

Time of concentration consists of overland flow duration (sheet flow and shallow concentrated flow, to a watercourse) and open duration of the channel flow, which are estimated by applying the following equations (Maidment, 1993; Wanielista et al., 1997):

$$T_{of} = \frac{L_{slope}\, n^{0.6}}{18\ I_{slope}} \tag{2.1.65}$$

$$T_{ch} = \frac{0.62\, L_{ch}\ n_{ch}^{\ 0.75}}{A^{0.125}\ I_{ch}^{\ 0.375}} \tag{2.1.66}$$

where L denotes the length of flow (surface, L_{slope}, and channel, L_{ch}), while I denotes (dimensionless) slope (of the area, I_{slope}, and riverbed, I_{ch}). These variables represent topographic properties of a catchment, which can be estimated *a priori*. Manning's roughness coefficient, n, for the catchment surface and the riverbed has to be inferred through calibration.

The number of the linear reservoirs can also be a free model parameter. Cascade of several identical reservoirs (i.e. the Nash model, Figure 22) introduces flexibility to the model in terms of peak flow attenuation and delay.

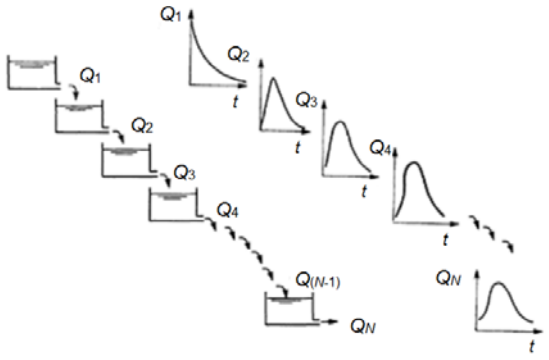


Figure 22. The Nash cascade of linear reservoirs (reproduced from Shaw, 2005).

(2) ***Baseflow*** **(Q_b)** is obtained by routing water percolating from the deepest sub-surface soil layer through a nonlinear reservoir (Figure 21), with the nonlinearity coefficient c and threshold $S_{\max}$:

$$Q_b(t) = \frac{1}{K_{\mathrm{NLR\_b}}}\left(S_b(t)\right)^c \tag{2.1.67}$$

If $c = 1$, the previous equation reduces to the linear outflow equation. $K_{\mathrm{NLR\_}b}$ represents a the nonlinear reservoir coefficient, and it is a free model parameter.

Combining the nonlinear outflow equation with the reservoir water balance equation yields an nonhomogeneous, nonlinear first-order ODE. Assuming that the inflow to the reservoir (V_{wperc}) occurs at the beginning of a time step (in discrete manner) reduces the equation to the homogenous one:

$$\frac{d\,S_b(t)}{dt} = -\frac{1}{K_{\mathrm{NLR\_}b}}\left(S_b(t)\right)^c \tag{2.1.68}$$

Integration of the previous equation over a time step results in the base flow volume V_b (Todini, 1996):

$$V_{b,\,t\to t+\Delta t} = S_{b,0}\left(1-\left(1-\frac{(1-c)\;Q_{b,0}\;\Delta t}{S_{b,0}}\right)^{\frac{1}{1-c}}\right) \tag{2.1.69}$$

where $S_{b,\,0}$ and $Q_{b,\,0}$ denote the reservoir storage and baseflow at the beginning of a current time step, respectively. The reservoir storage at the beginning of a time step is sum of the reservoir storage at the end of the previous time step and volume of percolation from the deepest soil layer in the current time step:

$$S_{b,0}{}^{*} = S_{b,\,t} + V_{Wperc,\,t\to t+\Delta t} \tag{2.1.70}$$

If the reservoir storage $S_{b,0}{}^{*}$exceeds threshold $S_{\max}$, the reservoir storage is corrected:

$$S_{b,0} = \max\left(S_{b,0}{}^{*};\, S_{\max}\right) \tag{2.1.71}$$

The threshold value $S_{\max}$ is calculated as:

$$S_{\max} = s_{\max}\,A \tag{2.1.72}$$

where $s_{\max}$ is a free model parameter and A is size of the drainage area.

The amount of water exceeding the threshold $S_{\max}$ comprises the volume of fast groundwater discharge in current time step (Figure 21):

$$V_{b\_\mathrm{fast\_LR},\,t\to\Delta t} = \max\left(0;\;\left(S_{b,0}{}^{*} - S_{\max}\right)\right) \tag{2.1.73}$$

Baseflow at the beginning of a computational step $Q_{b,\,0}$ is calculated based on the corrected reservoir storage $S_{b,\,0}$:

$$Q_{b,0} = B \left(\frac{S_{b,0}}{S_{\max}} \right)^{c} \tag{2.1.74}$$

Variable B denotes maximum baseflow value, and it is calculated following:

$$B = q_d \, A \tag{2.1.75}$$

where q_d denotes maximum specific baseflow yield (in L / s / km^2), which is a maximum baseflow rate per unit area, and is a free model parameter. Maximum baseflow (B) will occur if the reservoir storage $S_{b,0}$ is equal to threshold $S_{\max}$.

Mean baseflow over a computational time step is obtained by dividing volume of the baseflow (equation 2.1.69) by the length of the time step:

$$Q_{b,\, t \to t+\Delta t} = \frac{V_{b,\, t \to t+\Delta t}}{\Delta t} \tag{2.1.76}$$

Using the condensed notation, the previous equation may be written as:

$$Q_{b,\, i} = \frac{V_{b,\, i}}{\Delta t} \tag{2.1.77}$$

(3) ***Fast groundwater discharge – shallow aquifer response*** **(Q_{b_fast})** is simulated by transforming the excess water from the nonlinear groundwater reservoir (equation 2.1.72) through a linear one. Unlike linear reservoir for direct runoff, the inflow to the linear reservoir for interflow is added in a discrete manner (at the beginning of a computational time step). Therefore, the balance equation for the reservoir is reduced to the homogenous ODE:

$$\frac{dS_{b\_\text{fast}}(t)}{dt} = -Q_{b\_\text{fast}}(t) \tag{2.1.78}$$

Substituting the linear reservoir equation into the previous one yields:

$$\frac{d\, S_{b\_\text{fast}}(t)}{dt} + \frac{1}{K_{b\_\text{fast}}} S_{b\_\text{fast}}(t) = 0 \tag{2.1.79}$$

Solving the above equation for Q_{b_fast} and integrating over a time step results in the fast groundwater volume in a time step:

$$V_{b\_\text{fast},t\to t+\Delta t} = \int_{t=0}^{t=\Delta \text{t}} Q_{b\_\text{fast}}\, dt = S_{b\_\text{fast},\, t=0} \exp\left(- \frac{\Delta \text{t}}{K_{b\_\text{fast}}} \right) \tag{2.1.80}$$

Water balance of the reservoir at the end of a time step is estimated from:

$$S_{b\_\text{fast},\, t+\Delta t} = S_{b\_\text{fast},\, t} - V_{b\_\text{fast},\, t\to t+\Delta t} + V_{b\_\text{fast\_LR},\, t\to t+\Delta t} \tag{2.1.81}$$

where $V_{b\_\text{fast\_LR}}$ denotes the volume of water exceeding the threshold of the nonlinear reservoir, that is, an inflow to the linear reservoir of the fast groundwater response.

Fast groundwater discharge in a time step is a ratio of volume of the fast groundwater outflow to the length of a time step:

$$Q_{b\_\text{fast},t\to t+\Delta t} = \frac{V_{b\_\text{fast},t\to t+\Delta t}}{\Delta \text{t}} \tag{2.1.82}$$

Applying the condensed notations, the previous equation reads:

$$Q_{b\_\text{fast},i} = \frac{V_{b\_\text{fast},\, i}}{\Delta \text{t}} \tag{2.1.83}$$

Total simulated flow at the outlet of the catchment is the sum of the direct runoff, baseflow and fast groundwater discharge:

$$Q_i = Q_{d,\, i} + Q_{b,\, i} + Q_{b\_\text{fast},\, i} \tag{2.1.84}$$

Calculated balance of the reservoir at the end of current time step is equal to the water balance at the beginning of next time step. Reservoir states at the beginning of a simulation must be imposed on the model (initial conditions). Variables, parameters and initial conditions of the response routine are given in Table 6.

Table 6. Overview of the state and dependent variables, fluxes, free parameters and initial conditions of the response routine.

State variables		
S_{LR_d}	The surface runoff linear reservoir storage	[m^3]
S_{LR_b}	The fast groundwater linear reservoir storage	[m^3]
S_{NLR_b}	The baseflow nonlinear reservoir storage	[m^3]
Dependent variables		
S_{max}	Threshold of the non-liner baseflow reservoir	[m^3]
B	Maximum baseflow	[m^3/s]
Q_d	Direct flow	[m^3/s]
Q_{b_fast}	Fast groundwater response	[m^3/s]
Q_b	Baseflow	[m^3/s]
Parameters		
K_d	Linear reservoir coefficient for direct flow	[s]
q_d	Maximum specific baseflow yield	[$Ls^{-1}ha^{-1}$]
c	Non-linearity coefficient for baseflow simulation	[-]
$K_{gw\text{-}fast}$	Linear reservoir coefficient – fast groundwater response	[s]
s_{max}	Threshold of the non-liner baseflow reservoir per unit area	[mm]
Initial conditions		
$S_{d,\,(i=0)}$	State of the direct runoff linear reservoir at the beginning of a simulation	[m^3]
$S_{b\_fast,\,(i=0)}$	State of the fast groundwater response linear reservoir at the beginning of a simulation	[m^3]
$S_{b,\,(i=0)}$	State of the baseflow nonlinear reservoir at the beginning of a simulation	[m^3]

Flow routing routine

Flood routing is simulated in the 3DNet-Catch with a linear reservoir model. This model enables peak attenuation (due to friction), but the backwater effect cannot be simulated (Beven, 2005).

This model is based on the assumption that river reach behaves like a linear reservoir, so that the volume of the outflow from the reservoir (i.e. downstream node of the reach) in a time step can be estimated from:

$$V_{\text{down},\,t\to t+\Delta t} = S_t \exp\left(1-\frac{\Delta t}{K}\right) + Q_{\text{up},\,t\to t+\Delta t}\,\Delta t\left(1-\frac{1-\exp\left(\frac{-\Delta t}{K}\right)}{\Delta t/K}\right) \qquad (2.1.85)$$

where S_t is the reservoir storage at the beginning of current (and at the end of previous) time step, $Q_{\text{up},\, t \to t+\Delta t}$ is mean flow at the upstream node of the reach in current time step, K is the coefficient of the linear reservoir. The reservoir coefficient depends on the reach length, slope and roughness quantified via Manning's coefficient (Pedersen et al., 1980):

$$K = \frac{L n^{0.75}}{1.76 Q^{0.25} I^{0.375}} \tag{2.1.86}$$

Here L denotes length of the reach (km), n is the Manning roughness coefficient ($m^{-1/3}s$), Q is mean flow (m^3/s) and I is the slope of the reach (per cent).

The reservoir storage at the end of current time step and mean flow at the downstream node of the reach are calculated from the estimated volume:

$$S_{t+\Delta t} = S_t + V_{\text{up},\, t \to t+\Delta t} - V_{\text{down},\, t \to t+\Delta t} \tag{2.1.87}$$

$$Q_{\text{down},\, t \to t+\Delta t} = \frac{V_{\text{down},\, t \to t+\Delta t}}{\Delta t} \tag{2.1.88}$$

Applying condensed notation, previous equation reads:

$$Q_{\text{down},\, i} = \frac{V_{\text{down},\, i}}{\Delta t} \tag{2.1.89}$$

Computational grid – Hydrologic Response Units

Catchment delineation in the 3DNet-Catch model is based on the TIN (triangular irregular network) elevation model of a catchment. Triangles are generated by the Delaunay triangulation on given control points (contour maps or sampled elevation datasets, and stream gauge coordinates), according to specified maximum value of the circumcircle diameter (Figure 23).

TIN generation is followed by formation of Thiessen polygons (Voronoi diagram), as illustrated in Figure 23. Flow directions or flowpaths are determined along the steepest decent of a polygon (Figure 24). This approach is also applied in the tRIBS model (Ivanov et al., 2004; Vivoni et al., 2007).

Flow directions comprise a graph over which the Priority First Search algorithm (PFS) is applied. The algorithm starts from the polygon that contains a stream gauge (or other computational node) and propagates towards upstream polygons with the steepest slope (Figure 25), thus enabling catchment delineation.

The delineation results are stored in a database. For example, a sub-catchment (drainage area of a stream gauge) is an attribute in the database and it is assigned to each Thiessen polygon, i.e. surface runoff generated in the polygon will be routed towards the assigned stream gauge. Optionally, other drainage outlet may be attributed to the polygon for subsurface runoff, which is particularly convenient for karst catchments.

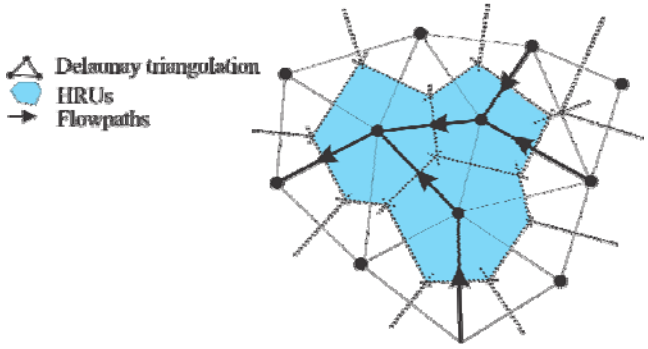


Figure 23. Development of HRUs from the TIN.

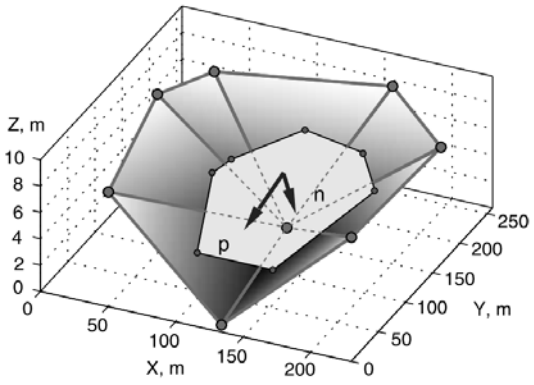


Figure 24. Flow direction from a Thiessen polygon (Ivanov et al., 2004). Grey areas represent triangles of TIN and the white one represents a HRU.

In the 3DNet-Catch model generated Thiessen polygons are considered Hydrologic Response Units (HRUs). All points within Thiessen polygons i.e. HRUs are assumed to exhibit hydrologically similar behaviour due to their geographical proximity.

Every HRU should be assigned unique parameter set based on the prevailing vegetative cover, soil and land use types in the polygon. However, if there are several polygons (HRUs) with similar properties, the same parameter set may be attributed to all of them, resulting in a significant decrease in dimensionality of the parameter estimation problem (chapter 2.3).

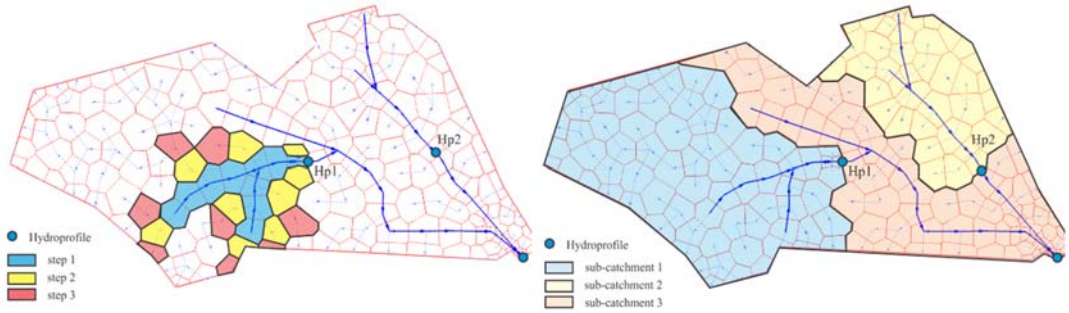


Figure 25. Catchment delineation in the 3DNet-Catch model: the PFS propagation algorithm.

2.1.2. Alternative model structures

In this research four versions of the 3DNet-Catch model are considered: three semi-lumped and a fully-distributed one. For the sake of simplicity, the semi-lumped versions are named SIMPLE, BASIC and FULL, implying increasing model complexity. The distributed version of the model stems from the BASIC version of the model.

The FULL version of the model

The FULL version of the model includes several subsurface soil layers and linear reservoirs for surface runoff routing (Nash cascade), where the number of sub-surface soil layers and the linear reservoirs are free model parameters. Since some parameters of the surface and subsurface soil layers are assumed equal, this version of the model contains 27 free parameters.

The structure of this model is presented in Figure 26.

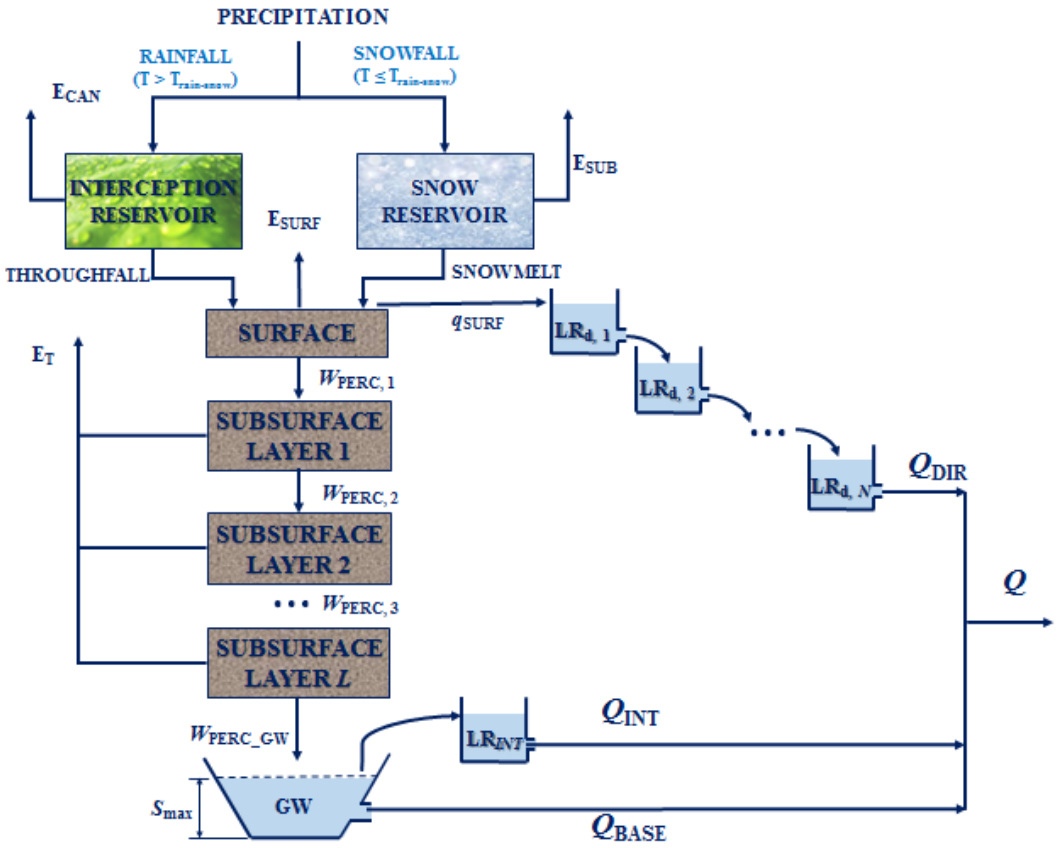


Figure 26. The FULL version of the 3DNet-Catch model.

The BASIC version of the model

In the BASIC version of the model, number of both subsurface layers and liner reservoirs for direct runoff simulation are set to 1 and the base temperature (T_{melt}) is assumed equal to the threshold temperature ($T_{S\text{-}R}$). In this way, the number of free parameters is reduced to 24 (some parameters of the surface and subsurface soil layers are equal).

This version of the model is presented in Figure 27.

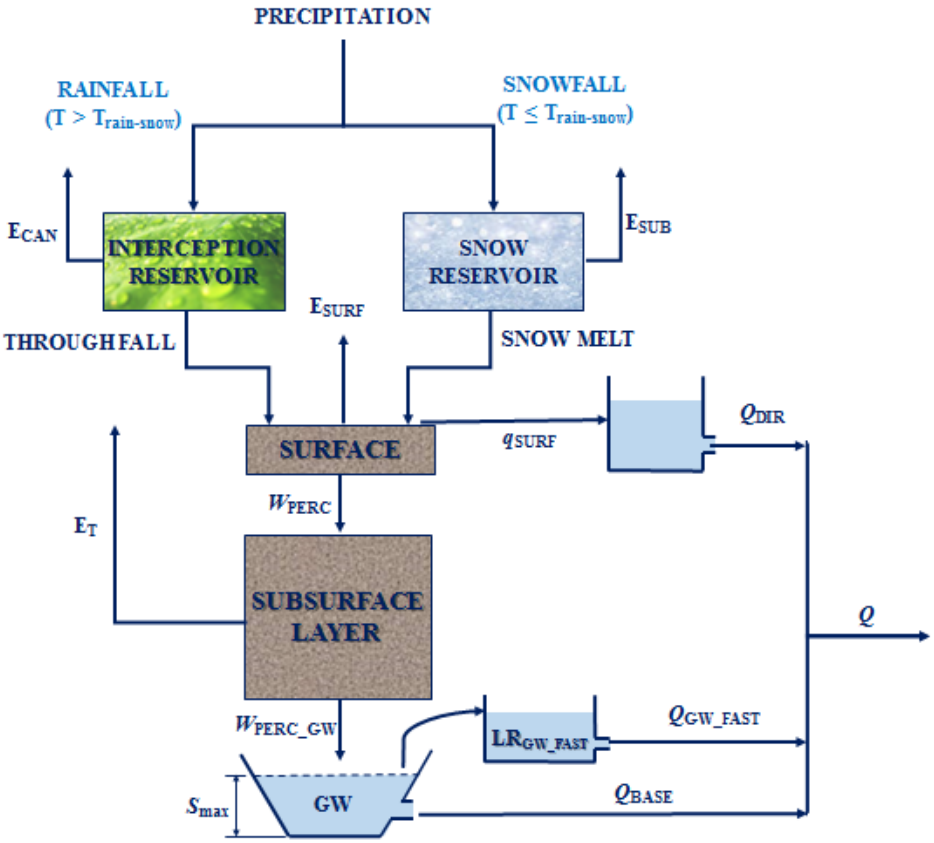


Figure 27. The BASIC version of the 3DNet-Cath model.

The SIMPLE version of the model

In the SIMPLE version of the 3DNet-Catch model, the parameters of the snow routine, except for $T_{S\text{-}R}$, are not optimised, since they are shown to be insensitive, i.e. do not affect model performance significantly (chapter 3.1). The values of these parameters are set to median values of the corresponding Pareto-optimal parameters, obtained over the full hydrologic record period. Thus, the number of free parameters in this version of the model is reduced to 20.

The distributed version of the model

The distributed model version is based on the BASIC model. Considering coarse temporal resolution of the available data and goals this research[13], elevation zones are assumed HRUs. Different parameters sets are assigned to each zone. However, majority of parameters are kept lumped due to limited information on the catchments. Spatially distributed parameters and regularisation method for calibration of the distributed version of the model are given in chapter 2.3.2.

In all versions of the model, seasonal variability in the *LAI* and meltrate coefficients (parameter b_{melt}) is retained. These parameters are aimed at modelling the processes that exhibit strong seasonality and therefore the parameter values should vary in time accordingly. The model setup is described in detail in chapter 2.3.1.

2.1.3. Model input data

In order to delineate a catchment, digital terrain model (DTM), stream network and catchment divide are required. Data on the land use type, vegetation or soil types can facilitate establishing prior ranges of some parameters (e.g. *CN*, CAN_{max}, soil-related parameters).

Data required for a model run include:

– Precipitation depths [mm Δt^{-1}],

– Potential evapotranspiration [mm Δt^{-1}],

– Temperature [°C],

– Observed flows [m^3/s].

[13] In this research impact of model structure on the Pareto-optimaltimal parameter temporal variability is analysed (chapter 1.6). Therefore, the results of the distributed model version are compared to the results of the semi-lumped one.

Input time series should be (dis)aggregated to match the computational time step (e.g. mean daily temperatures or flows).

Semi-lumped versions of the model. Since topography of the catchments considered in this research (chapter 2.6) considerably varies in elevation, model forcings (such as precipitation or temperature) are adjusted for elevation and different input vectors are estimated for every elevation zone of a catchment.

In this research, every catchment is divided into an arbitrary number of elevation zones of approximately equal spans. Each elevation zone is represented by its mean elevation, total area and mean slope. Precipitation depths and temperature are estimated for every zone, depending on the difference between the mean zone elevation and the reference altitude z_{MS}. The reference altitude z_{MS} is assessed as the weighted mean elevation of the meteorological stations following the methodology presented by Panagoulia (1995):

$$z_{\mathrm{MS}} = \sum_{i=1}^{N_s} \omega_i \; z_i \tag{2.1.90}$$

where N_s is number of meteorological stations, z_i and ω_i are the elevation and the weight of the i^{th} meteorological station, respectively. Station weights are obtained by applying the Thiessen polygon method.

In general, temperature exhibits a rather constant lapse rate i.e. decrease with elevation, while the increase in precipitation depths lessens with the elevation (Bardossy & Das, 2008; Hundecha & Bárdossy, 2004). However, the constant gradients of both variables with elevation are adopted in this research.

Mean precipitation depth $\bar{P}$ in an elevation zone is calculated according to:

$$\bar{P} = P_{\mathrm{MS}} \left(1 + \frac{\alpha \left(z_c - z_{\mathrm{MS}} \right)}{100 \cdot 100} \right) \tag{2.1.91}$$

where P_{MS} stands for mean catchment precipitation depth estimated by employing the Thiessen polygon weighting method, without any adjustment for elevation and α represents increase in precipitation (in per-cent) per 100 m of elevation increase (similar to e.g. P_{CALT} parameter of the HBV-light model, Seibert and Vis 2012). The value of α

can be estimated based on the slope of a linear regression between annual precipitation depths and meteorological station elevations. Some recommendations on increase in precipitation with elevation may be found in the literature. For example, Uhlenbrook et al. (2000) estimated an increase of 6% / 100 m, while Seibert and Vis (2012) recommended increase of 10% / 100 m for simulations.

Similarly, mean temperature in an elevation zone is calculated as:

$$\overline{T} = T_{\mathrm{MS}} + \frac{z_{\mathrm{MS}} - z_c}{100} T_{\mathrm{lapse}} \tag{2.1.92}$$

where T_{MS} is the mean catchment temperature calculated by applying the Thiessen polygon method. T_{lapse} is a temperature lapse rate (in °C/100 m). The value of this parameter is commonly assumed to be approximately -0.6 °C/100 m (e.g. U.S. A.C.E., 1994; Uhlenbrook et al., 2000; Seibert and Vis, 2012).

In this research α and T_{lapse} are free model parameters to be estimated in the calibration procedure. Their prior ranges are assessed for every catchment according to long-term observations at the meteorological stations.

The *PET* time series can be calculated externally and introduced into the model as the input time series, or within the model following the Hargreaves method (Hargreaves and Samani, 1982; Lu et al., 2005, Oudin et al., 2005; Trajkovic and Kolakovic, 2009; Tabari et al., 2011). To account for changes with elevation, *PET* rates are estimated for every elevation zone independently, using the obtained mean zone temperatures. Since only the temperature data were available for *PET* assessment, the *PET* rates had to be calculated by some of the temperature- or radiation-based methods, which have modest data requirement (Maidment, 1993).

Oudin et al. (2005) examined influence of the method for *PET* assessment on performance of hydrologic models. They simulated runoff at a lot of catchments using 27 methods for *PET* estimation and four lumped, conceptual hydrologic models. They demonstrated that the use of the temperature- or radiation-based methods may result in the same model performance as the use of more complex methods (e.g. Penman-Monteith). For example, models that used the McGuinness, Jensen-Haise (radiation-based) or Hamon methods

(temperature-based) outperformed models that used the Penman-Monteith method both in calibration and validation periods. Considering the results of Lu et al. (2005), Trajkovic and Kolakovic (2009), Rao et al. (2011) and Tabari et al. (2011), the Hamon method is selected for use in this study. Application of this method for hydrologic modelling purposes is reported in the literature by Fenicia et al. (2008), Gharari et al. (2012); Gharari et al. (2013) and Osuch et al. (2014).

According to the Hamon method (Hamon, 1961), daily *PET* rates are calculated for every elevation zone as:

$$PET = \left(\frac{DL}{12}\right)^2 \exp\left(\frac{T_a}{16}\right) \tag{2.1.93}$$

where T_a is a mean daily temperature in an elevation zone and DL is a daytime length (time from sunrise to sunset, in h day^{-1}), which depends on latitude φ and declination of the Sun δ, both in expressed in radians (Spitters et al., 1986):

$$DL = \frac{24}{\pi} \arccos\left(-\tan(\varphi)\,\tan(\delta)\right) \tag{2.1.94}$$

$$\delta = 0.4093 \sin\left(\frac{2\pi}{365}\left(D_n + 284\right)\right) \tag{2.1.95}$$

where D_n denotes a day of a year.

In the semi-lumped versions of the 3DNet-Catch model vertical water balance (surface runoff and percolation into the groundwater reservoir) is simulated independently for every elevation zone by using a single parameter set common to all zones, with precipitation, temperature and *PET* rates estimated for each particular zone. Simulated surface runoff and percolation generated in individual zones are summed and routed through the reservoirs at the catchment outlet.

Distributed model version. Generally, precipitation depths and temperatures are estimated in the 3DNet-Catch model for every HRU by employing the inverse-distance weighting method (Figure 28). Precipitation and temperature data are estimated based on the observations from up to 4 nearest meteorological stations.

In this research, however, the catchment elevation zones are considered HRUs. The input data obtained for semi-lumped versions of the model are therefore used in the distributed version as well.

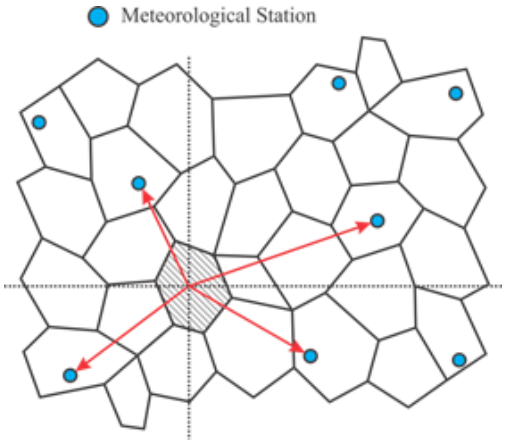


Figure 28. The inverse-distance weighting for estimating mean precipitation depths or temperature for a HRU in the 3DNet-Catch model.

2.2. Model multi-objective calibration with the AMALGAM algorithm

Properties of a response surface in hydrologic modelling impose difficulties on parameter optimisation. As elaborated in chapter 1.3.2, global optimisation algorithms are generally considered capable of coping with these difficulties (e.g. Yapo et al., 1996; Vrugt et al., 2009). However, it has been argued in the literature that a single optimisation algorithm cannot be efficient at various optimisation problems, i.e. different optimisation algorithms perform better for specific optimisation problem (Vrugt & Robinson, 2007; Vrugt et al.,c 2009). Hence, employing several optimisation algorithms in the calibration procedure is expected to locate global optimum basins of the response surface more efficiently and effectively. Therefore, an algorithm that combines several global optimisation algorithms, namely AMALGAM – *A MultiAlgorithm Genetically Adaptive Multiobjective*, is used for model calibration in this research.

The AMALGAM employs several global optimisation algorithms (mostly operators for population evolution) simultaneously, so every algorithm is in control of a certain number of (initial) parameter sets. The number of sets allocated to each optimisation algorithm is

altered dynamically in the calibration procedure, so that the algorithms with the highest reproductive success in previous iterations are allowed to generate more offspring in the current iteration. This peculiarity of the AMALGAM algorithm makes it superior compared to the individual global optimisation algorithms, especially in presence of multidimensionality of the optimisation problem (large number of free model parameters) (Vrugt et al. 2009).

The AMALGAM starts with the initial population sampling, which is based on the Latin Hypercube method – *LHS* (Vrugt and Robinson, 2007). The *LHS* is a type of stratified Monte Carlo sampling method without replacement (Keramat & Kielbasa, 1997; Marino et al. 2008). The parameter prior range is split into *N* non-overlapping intervals of equal probability (i.e. width of an interval amounts 1/*N* if the uniform probability is assumed). A random value of the (uniform) cumulative distribution function (cdf) is sampled from every interval and the corresponding value of the parameter is assessed from the sampled cdf values. In this way, *N* values of every model parameter are obtained. This procedure is repeated for all model parameters and the sampled parameters are combined together into *N* different (initial) parameter sets. Being computationally cheap (Gentle, 2003; Keramat & Kielbasa, 1997; Sieber & Uhlenbrook, 2005), the *LHS* is rather convenient for calibration of hydrologic models, which are known for multidimensionality. For example, the required number of model runs for assessment of a parameter uncertainty bounds is reduced by 90% if the *LHS* is applied compared to random sampling (Sieber & Uhlenbrook, 2005).

The number of parameter sets to be assigned to an optimisation algorithm i in current iteration (generation) t, N_i^t, depends on the number of sets allocated to i^{th} algorithm in the previous iteration, $N_i^{(t-1)}$, and the number of sets which are generated by i^{th} algorithm in previous iteration and participate in the current generation, P_i^t (Yilmaz et al., 2010):

$$N_i^t = N \frac{P_i^t}{N_i^{(t-1)}} \left[\sum_{i=1}^{q} \frac{P_i^t}{N_i^{(t-1)}} \right]^{-1} \tag{2.2.1}$$

Ratio between P_i^t and $N_i^{(t-1)}$ represents reproductive success of the i^{th} algorithm. N denotes total number of parameter sets (size of population) and q denotes number of optimisation algorithms employed within the AMALGAM.

Change in the number of parameter sets allocated to individual optimisation algorithms throughout parameter optimisation procedure is illustrated in Figure 29. In this example, the majority of the sets is in control of the GA (Genetic Algorithm) and DE (Differential Evolution) algorithms.

To prevent some optimisation algorithms from inactivating and, consequently, to preserve population diversity, a minimum number of parameter sets is allocated to an optimisation algorithm regardless of its reproductive success (Vrugt et al. 2009). For example, in Figure 29 it is shown that the AMS (Adaptive Metropolis Search) and PSO (Particle Swarm Optimiser) algorithms are assigned 5% of the total number of parameter sets, while majority of the set is evolved by the GA and DE algorithms[14].

After evolving assigned parameter sets and commutating their fitness, parent and offspring sets are merged into population of size 2N and ranked according to the values of the objective functions. The best ranked set remains in the new population, while the remaining N-1 sets are selected according to values of the objective functions and crowding distance. The crowding distance denotes Euclidian distance of a parameter set to the remaining sets of the Pareto front. Namely, remaining parameter sets are sorted into several Pareto fronts, where the first Pareto front contains non-dominated sets, the second one non-dominated sets of the remaining sets, and so forth. Selected sets are appended to the new population based on the rank of the Pareto front, and the crowding distance. This means that after inclusion of members of the first Pareto front, the members of Pareto front of a lower rank (i.e. second, third, etc.) are appended to the new population until it reaches size of N. If all sets of the p^{th} Pareto front cannot be included into the new population, the sets with larger crowding distance are preferred to preserve population diversity (Vrugt et al., 2009). If an optimised parameter estimate is outside the prior range, it is set equal to minimum or maximum parameter value, depending on whether lower or

[14] These two optimisation algorithms (GA and DE) are in control of the evolution of most parameter sets of the 3DNet-Catch model.

upper limit on the parameter value is crossed. Once again, different number of sets is assigned to every optimisation algorithm, and the sets of new population are optimised in the next iteration.

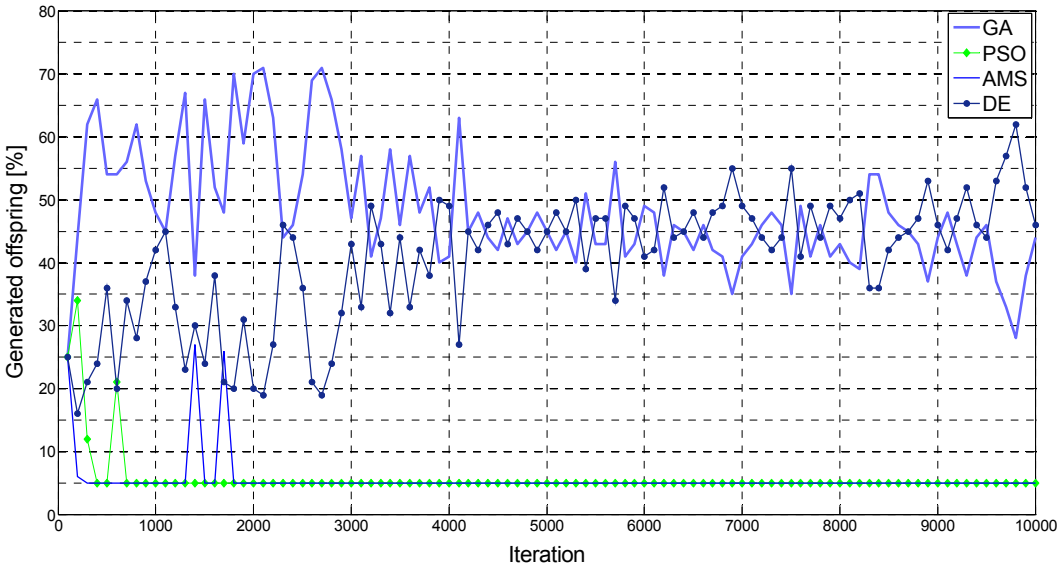


Figure 29. Percentage of offspring (parameter sets) generated by four different optimisation algorithms: the semi-lumped BASIC version of the model, the Kolubara River catchment, calibration in the 1955-2013 period.

The procedure for parameter optimisation reported above is repeated until one of the convergence criteria is met: minimum relative change between consecutive values of the objective function(s) or parameter estimates, or maximum number of iterations (Madsen 2003; Blasone et al. 2007; Blasone 2007). In this research, the latter convergence criterion is adopted (maximum number of iterations is set to 20.000, Table 7). This criterion is selected for two reasons: (1) constrains on computational time and resources, and (2) in this way, all calibrations are performed under same conditions. Selected number of iterations is supported by the:

– Results reported in the literature: e.g. Zhang et al. (2009) adopted 10.000 iterations for the same population size to calibrate the SWAT model, while Reed et al. (2013) demonstrated that 20.000 iterations given population size of 100 with the AMALAGAM resulted in effective optimisation of the HBV model;

– Analysis of change in objective function (*NSE*) with increasing number of iterations (Figure 30): the objective function reaches steady state after 10.000 iterations. Another objective function, *VE*, reaches steady state after approximately 500 iterations (not shown here).

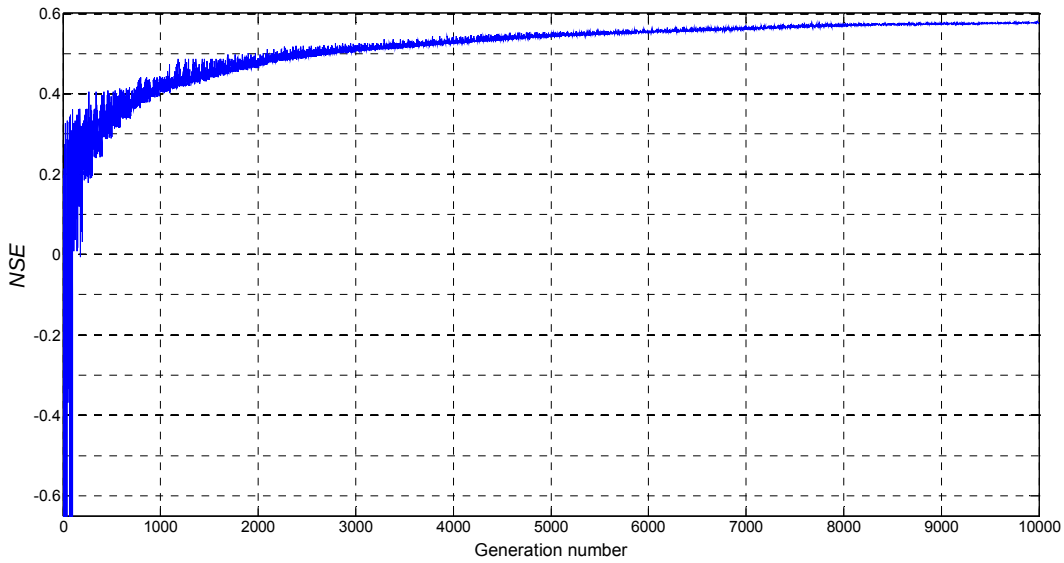


Figure 30. *NSE* values versus optimisation runs (offspring generations): the semi-lumped BASIC version of the model, the Kolubara River catchment, 1955-2013.

The MATLAB code for the AMALGAM algorithm is available from Washington University web site[15].

This version of the AMALGAM contains four optimisation algorithms: Non-dominated Sorting Genetic Algorithm (NSGA-II), Differential Evolution (*DE*), Particle Swarm Optimisation (*PSO*) and Adaptive Metropolis Search (*AMS*). These algorithms are briefly described in the remaining of this chapter.

[15] http://www.hydro.washington.edu/pub/blivneh/CONUS/misc/tools.uw.electric/MATLAB-Code-AMALGAM-Sequential-V1.2/

2.2.1. Optimisation algorithms in the AMALGAM employed in this research

Non-dominated Sorting Genetic Algorithm (NSGA-II)

Genetic algorithms are based on three principles of the evolution theory: selection, crossover (recombination) and mutation. Uniform crossover operation results in offspring with genes containing sequences form both parents, and (polynomial) mutation results in new allelic material, what is rather important to preserve population diversity and prevent the algorithm of trapping in a local optimum (Vrugt et al., 2009; Weise, 2009). Parameters of this algorithm are given in Table 7.

In the AMALGAM, the mutation factor p_m is equal to the reciprocal value of the number of free parameters (Vrugt et al., 2009). Since the 3DNet-Catch contains considerable number of free parameters, in this research p_m is set to 0.1 in order to preserve diversity in the parameter sets.

Differential Evolution

Differential evolution implies recombination of existing parameter sets x to generate offspring as follows (Vrugt et al., 2009):

$$u_k^{(t+1)} = x_{r_1}^{(t)} + F\left(x_{r_2}^{(t)} - x_{r_3}^{(t)} \right) \tag{2.2.2}$$

$$x_k^{(t+1)} = \begin{cases} u_k^{(t+1)}, & U \le CR \\ x_k^{(t)}, & U > CR \end{cases} \tag{2.2.3}$$

In the first equation, mutant vector u_k is calculated based on parameter sets x with randomly selected indices r_1, r_2 and r_3 such that $r_1 \neq r_2 \neq r_3 \neq k$; F denotes the mutation scaling factor which determines the level of combination between x_{r_2} and x_{r_3}. In the second equation, U is a random number in the [0, 1] interval and CR is the crossover constant which controls the probability of the mutant vector contributing to the offspring (similar to the crossover probability of the *NSGA*-II). Two parameters of the DE algorithm used in this research are specified in Table 7.

Particle Swarm Optimisation

Particle Swarm Optimisation emanate from the swam behaviour of a flock of birds or insects (Vrugt et al., 2009; Weise, 2009). In this algorithm a swarm of particles in the parameter space is simulated (Weise, 2009). Each particle (parameter set) is defined by its current position, $x_k^{(t)}$, and its velocity, $v_k^{(t)}$ (these values are randomly initialised). The particles change their position and velocity as follows (Vrugt et al., 2009):

$$v_k^{(t+1)} = \varphi \cdot v_k^{(t+1)} + c_1 r_1 \left(x_{\text{best}}^{(t)} - x_k^{(t)} \right) + c_2 r_2 \left(p_{\text{best}}^{(t)} - x_k^{(t)} \right) \tag{2.2.4}$$

$$x_k^{(t+1)} = x_k^{(t)} + v_k^{(t+1)} \tag{2.2.5}$$

where φ, c_1 and c_2 denote inertia factor, and cognitive and social factors of the particle, respectively. Values of these parameters used in this research are specified in Table 7.

Adaptive Metropolis Search

Adaptive Metropolis search is based on the random walk Metropolis-Hastings algorithm, which implies that a set x can be described by a target distribution π (Haario et al. 2001). This algorithm consists of the following steps:

(1) Initial sampling from the parameter space in order to obtain initial sample, x_0.

(2) Sampling a candidate point y from a proposal distribution $q_t\left(\cdot \middle| x_0\right)$. The proposal distribution is normal, with mean at current point x_t and covariance that, unlike classical Metropolis-Hasting algorithm, depends on all previous states.

(3) Calculation of the probability of acceptance of the sampled candidate point, α, which depends on the probability density π()::

$$\alpha\left(x_{t-1}, y\right) = \min\left(1, \frac{\pi(y)}{\pi\left(x_{t-1}\right)}\right) \tag{2.2.6}$$

(4) Calculation of the covariance matrix and the proposal distribution.

Table 7. Parameters of the AMALGAM, GA, DE, PSO and AMS algorithms

AMALGAM	
Number of optimisation algorithms q	4
Size of the population	100
Number of iterations	20.000
Minimum percentage of sets allocated to an optimisation algorithm	5
NSGA II	
Crossover probability p_c	0.9
Crossover distribution index η_c	20
Mutation probability p_m	0.1
Mutation distribution index η_m	20
DE	
Scaling factor F	From uniform distribution [0.6, 1]
Crossover constant CR	From uniform distribution [0.2, 0.6]
PSO	
Inertia factor φ	0.5 + U [0, 1] / 2
Weight for cognitive factor of particle c_1	1.5
Weight for social factor of particle c_2	1

2.3. Runoff modelling using the 3DNet-Catch model

2.3.1. The 3DNet-Catch model setup

As discussed in chapter 1.3.2, setting the prior ranges of the parameters is quite important for proper automatic calibration of hydrologic model. In this research prior ranges of some parameters are estimated based on information on the catchments considered (e.g. land use types or topography) and on the recommendations found in the literature. However, ranges of some parameters are inferred based on the results of the preliminary model runs, by comparing different aspects of simulated and observed hydrographs ("trial-and-error").

To avoid physically unrealistic parameter estimates, some parameters are calibrated in relative terms – as multipliers. For example, the melt factor in June should be greater than the melt factor in December, or porosity should be greater than soil wetness at the filed capacity, which, again, should be larger than the wetness at permanent wilting point. It is convenient to impose these limitations by expressing some parameters as the multipliers of other parameters: for example, wetness at the wilting point as multiplier of the porosity, or melt factor in December is a multiplier (smaller than 1) of the melt factor in June.

Some soil-related parameters may exhibit change with increasing soil depth. However, observed changes are not easily transferable to model parameters (e.g. Porêbska et al., 2006; Scorza Júnior & Silva, 2011). Therefore, most of the parameters of the soil routine are assumed equal for surface and subsurface layers (porosity, wetnesses at the wilting point and field capacity, pore-size distribution index). An exception is made for the hydraulic conductivity, since its value (exponentially) decreases with the soil depth (e.g. Beven, 1982; Ivanov et al., 2004). In this research, hydraulic conductivity for the sub-surface layers are calibrated as ratio to the sampled hydraulic conductivity of the surface soil layer.

Since hydraulic conductivity takes rather small values, common logarithm of its value is optimised. In this way, parameter space is better explored and under-sampling is prevented (Marino et al., 2008).

The setup of the semi-lumped versions of the model is presented in Table 8 through Table 12, while the prior parameter ranges for all model versions are given in Appendix A.

Impact of each model parameter on simulated hydrographs and runoff volume is illustrated in Appendix B and briefly outlined in these tables. Namely, parameter impact on simulated hydrographs is not straightforward and it may depend on values of other parameters. Therefore, parameter impact outlined in Table 8 through Table 12 is merely a summary of the hydrographs presented in Appendix B.

Table 8. Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: precipitation and temperature gradients with elevation

Parameter	Comment	Parameter impact on simulated flows	References
Precipitation gradient with elevation α [% / 100 m]	Estimated based on the long-term observations and altitude of meteorological stations used for runoff simulations at a particular catchment.	This parameter significantly affects simulated hydrographs and flow volume. Increase in α results in larger flow volume and considerably higher flows.	
Temperature gradient with elevation (lapse rate) T_{lapse} [°C / 100 m]	Estimated based on the long-term observations and altitude of meteorological stations used for runoff simulations at a particular catchment.	Impact of this parameter is highest in the snow-melt season. Small values of the lapse rate imply more uniform temperatures in the catchment. In the snow-melt season this means leads to somewhat delayed flood waves.	

Table 9. Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: the interception routine

Parameter	Comment	Parameter impact on simulated flows	References
Maximum interception reservoir capacity CAN_{max} [mm]	Estimated based on the recommendations for particular types of vegetation, and share of that vegetation type in the catchment area.	Large values of this parameter imply higher interception capacity and, hence lower peak flows. This parameter does not affect peak timing. Generally, impact of this parameter is low.	Jovanović and Radić (1990); Breuer et al. (2003)
Maximum value of the Leaf Area Index LAI_{max} [m^2m^{-2}]	*LAI* varies according to sine curve over the growing season, while outside growing season *LAI* is set to zero (chapter 2.1.1). The range of LAI_{max} is estimated after recommendations for vegetation types and their share in the catchments.	In the growing season small values of LAI_{max} result is small interception capacity and large runoff. Impact of this parameter on simulated flow volume and hydrograph shape is considerable.	Eschenbach and Kappen (1996); Breuer et al. (2003); Asner et al. 2008; Scurlock and Hicke (2008); He et al. (2014)

Table 10. Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: the snow routine

Parameter	Comment	Parameter impact on simulated flows	References
Boundary temperature, $T_{S\text{-}R}$ [°C] and base temperature for snowmelt T_{melt} [°C]	The ranges of these parameters are adopted from the literature and adjusted in preliminary simulations. Except for the FULL version of the model these parameters are assumed equal.	This parameter affects both peak magnitude and timing. Small values of temperature for snowmelt result in soon and rapid snowmelt, and therefore higher peak flows. Small values of $T_{S\text{-}R}$ imply rainfall and, consequently, more dynamic catchment response over periods with temperatures close to 0°C.	Anderson (2006); Feiccabrino and Lundberg (2008)
Threshold depth of snow (as water equivalent) above which the entire area is covered in snow $S_{snow,100}$ [mm]	Prior range of this parameter is assumed and tested in preliminary simulations.	Smaller values of $S_{snow,100}$ result in faster snowmelt, leading to higher peak flows which occur sooner compared to high values of $S_{snow,100}$. This parameter affect magnitude and timing of peak flows, although its impact is modest.	
Snowpack temperature lag factor λ[-]	The prior range of this parameter is set to be equal to its feasible range.	Values of λ close to 1 result in faster melt of the snowpack, which means that high flows due to snowmelt occur sooner compared to low values of λ. Impact of this parameter is marginal.	Zhang et al. (2009); Neitsch et al. (2011)
Melt factor on the 21st of June $b_{melt,6}$ [mm°C^{-1}day^{-1}]	The prior range of this parameter is set to be equal to its feasible range.	Small values of the melt factor result in more uniform snow melt in time, and therefore more uniform flows in snowmelt periods. However, impact of this parameter is marginal.	Neitsch et al. (2011)
Melt factor on the 21st of December $b_{melt,12}$	This parameter is calibrated in relative terms, as the percentage of the sampled value of $b_{melt,6}$. The ranges are selected not to transcend ranges recommended in the literature.	Lower values of this parameter, which is calibrated in relative terms, indicate higher seasonal variation in melt factor. Impact of this parameter in low.	Neitsch et al. (2011)

Table 11. Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: the soil routine

Parameter	Comment	Parameter impact on simulated flows	References
Curve number *CN* [-]	Reduced prior ranges *CN* values are estimated after land use types and hydrologic soil types, their share at the catchment area. In the semi-lumped versions of the model CN values are corrected to account for actual terrain slope of each elevation zone.	Higher values of *CN* result in increase in direct runoff and reduction in baseflow. Generally, larger *CN* values result in higher peaks flows and slightly lower baseflow.	Chow et al. (1988); Maidment (1993); Jovanović et al. (2013); Laura et al. (2011)
Initial abstraction I_{a_rel} [-]	The value of this parameter is approximately 0.2. Prior range of this parameter is adjusted in preliminary simulations.	Smaller values of the initial abstraction result in higher peak flows of individual flood waves.	
Surface layer thickness D_{surf} [mm]	The prior range of this parameter is set to be equal to its feasible range.	Larger surface layer thickness implies higher capacity of soil storage, which leads to decrease in peak flows.	Ogée and Brunet (2002)
Effective porosity [-]	The ranges of these parameters are adopted from the literature and adjusted in preliminary simulations.	Similarly to D_{surf}, larger values of soil porosity leads to higher capacity of the soil storage, and reduction In peak flows. Impact of this parameter is significant.	Rawls et al. (1982); Ivanov et al. (2004); Saxton and Rawls (2006); Scorza et al. (2011)
Saturated hydraulic conductivity of the surface soil layer K_{surf} [$m \cdot s^{-1}$]	Values of saturated hydraulic conductivity are rather small, thus calibration is performed on logarithms of the coefficients to prevent under-sampling, that is to better explore entire parameter space (Marino et al. 2008). Range of this parameter is adopted from the literature, and enlarged to certain extent to account for preferential flows.	If saturated hydraulic conductivity is set to minimum value direct runoff prevails over baseflow, and vice versa. This parameter considerably affects simulated hydrograph.	Beven (1982); Rawls et al. (1982); Ogée and Brunet (2002); Ivanov et al. (2004); Scorza Júnior and Silva (2011); Mathias et al. (2015)
Soil wetness at permanent wilting point w_{wp} [-]	This parameter is calibrated in relative terms, as ratio to sampled value of porosity. The ranges of the ratio as selected not to exceed feasible range of wetness at wilting point.	Small values of the wetness at wilting point imply larger soil storage, and therefore lower peaks flows.	Ogée and Brunet (2002); Saxton and Rawls (2006); Scorza Júnior and Silva (2011); Pavelková et al. (2012); Diallo and Mariko (2013); Singh (2013); Yang and You (2013); Mathias et al. (2015)

(continued). Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: the soil routine

Parameter	Comment	Parameter impact on simulated flows	References
Soil wetness at field capacity w_{fc} [-]	Wetness at field capacity is calibrated in relative terms, as ratio to the difference between porosity and wetness at permanent wilting point. In this way, order on the values of the porosity, w_{fc} and w_{wp} is imposed. The ranges for this parameter are set keeping in mind its values recommended in the literature.	Values of this (relative) parameter close to zero mean that the wetness of field capacity approaches to the wetness of wilting point and that more water percolates to the non-linear groundwater reservoir. This leads to higher baseflow.	Diallo and Mariko (2013); Singh (2013); Mathias et al. (2015)
Pore size distribution index n [-]	The initial range of this parameter is inferred by the recommendations in the literature and preliminary simulations.	Impact of this parameter depends on the sampled values of wetnesses at wilting point and field capacity. If the values of these parameters are small (as in the example in Appendix B), increase in n leads to decrease in percolation and consequently to decrease in baseflow.	Schaap et al. (2001); Porêbska et al. (2006); Yang and You (2013)
Number of sub-surface layers N_l	This parameter is free only in the FULL version of the model, while in the remaining versions its value is set to one. Since N_l can take only integer values, sampled values are rounded towards smaller integer value.	Increase in the number of sub-surface soil layers results in delay of baseflow. This delay increases with the thickness of the soil layers.	
Thickness of a sub-surface layer $D_{sub\text{-}surf}$ [mm]	Prior range is inferred based on expected soil thickness, on the results of preliminary simulations and recommendations in the literature for other models.	Rather small values of this parameter imply negligible capacity of the soil storage and baseflow. Impact of this parameter is considerable.	Ogée and Brunet (2002); Schaefli et al. (2014)
Saturated hydraulic conductivity of the sub-surface layers $K_{sub\text{-}surf}$ [mm*day^{-1}]	This parameter is calibrated relative to the saturated hydraulic conductivity of the surface layer as follows: $K_{sub\text{-}surf} = 10^{theta} \cdot K_{surf}$	Small values of this parameter result in derease of soil permebility and baseflow, and vice-versa. This parameter significantly affects simulated hydographs.	Beven (1982); Rawls et al. (1982)

Table 12. Calibration of the of the semi-lumped 3DNet-Catch model and prior parameter ranges: the response routine

Parameter	Comment	Parameter impact on simulated flows	References
Linear reservoir coeff. for direct flow K_d [days]	Range of this parameter is estimated based on the catchment time of concentration (Kirpich, SCS and Manning equations).	Small values of K_d imply prompt direct runoff without any attenuation of peak flows, and vice-versa: increase in K_d leads to mitigated flood waves. This parameter considerably affects hydrographs.	(*Urban Hydrology for Small Watersheds Technical Release 55*, (1986); Wanielista et al. (1997)
Number of linear reservoirs N_{LR} [-]	This parameter is optimised only in the FULL version of the model, while in the remaining versions its value is set to 1. N_{LR} can take only integer values, thus sampled values are rounded toward smaller integer value.	Higher number of the linear reservoir results in delayed flood waves and attenuation of peak flows.	
Fast groundwater response reservoir coeff. $K_{gw\text{-}fast}$ [days]	This parameter is calibrated relative to the coefficient of the direct runoff reservoir (as a multiplier of the K_d). The parameter range is inferred from preliminary simulations.	Impact of this parameter depends on the amount of fast groundwater discharge, i.e. on other baseflow-related parameters (e.g. s_{max} or q_d). Smaller values of $K_{gw\text{-}fast}$ result in faster response i.e. in steeper recessions. Generally, impact of this parameter is limited.	
Maximum specific baseflow yield q_d [L s^{-1} km^{-2}]	The ranges of this parameter are set after 25th percentile of the flows observed in July (minimum) and 75th percentile of the flows observed in March or April (maximum).	Large values of this parameter result in higher baseflow and slightly slower recessions. Impact of q_d is marginal.	
Non-linearity coefficient for baseflow simulation c [-]	Ranges of the parameter are estimated based on the preliminary simulations.	Larger values of c lead to steeper recessions and decrease in (minimum) baseflow.	
Threshold of the non-liner baseflow reservoir per unit area s_{max} [mm]	Ranges of this parameter are assumed after and the results of the preliminary simulations.	Smaller values of s_{max} result in baseflow increase (it approaches to the maximum value determined from q_d).	

2.3.2. Calibration of the distributed version of the model

In this research the regularisation method presented by Yilmaz et al. (2008) is applied to calibrate distributed version of the model. This method is based on a nonlinear transformation of the parameter field, and it is rather convenient for two reasons: (1) it requires only one superparameter per free model parameter, and (2) it keeps optimised parameters within predefined bound, without imposing additional limitations.

An application of this method starts with defining prior parameter values for every computational cell or HRU, $\theta_{p,i}$ (p refers to a model parameter, while i denotes HRU) and feasible ranges for every free model parameter ($\theta_{p,\min}$ and $\theta_{p,\max}$). An optimised parameter is calculated as follows:

$$\theta_{p,i} = \theta_{p,\min} + \left(\theta_{p,\max} - \theta_{p,\min}\right)\left(\frac{\overline{\theta}_{p,i} - \theta_{p,\min}}{\theta_{p,\max} - \theta_{p,\min}}\right)^{\alpha} \tag{2.3.1}$$

where $\overline{\theta}_{p,i}$ denotes prior value of the p^{th} parameter for the i^{th} HRU, and α is calculated as:

$$\alpha = \frac{\log_{10}\left(1 - \frac{2-\beta}{2}\right)}{\log_{10} 0.5} \tag{2.3.2}$$

where β stands for the superparameter, which can take value in the [0, 2] interval. In this way, parameters are prevented from exceeding the imposed prior range (Yilmaz et al., 2008):

$$\begin{aligned} &\beta \to 0 \quad \Rightarrow \quad \alpha \to \infty \quad \Rightarrow \quad \theta_{p,i} \to \theta_{p,\min} \\ &\beta = 1 \quad \Rightarrow \quad \alpha = 1 \quad \Rightarrow \quad \theta_{p,i} = \overline{\theta}_{p,i} \\ &\beta = 2 \quad \Rightarrow \quad \alpha = 0 \quad \Rightarrow \quad \theta_{p,i} = \theta_{p,\max} \end{aligned} \tag{2.3.3}$$

Due to limited information on the catchments (chapter 2.6), majority of the model parameters are assumed spatially uniform. Parameters that are spatially distributed are those related to land use or vegetation types within a HRU ($CAN_{\max}$, $LAI_{\max}$ and CN) or related to its elevation (λ – snowpack is assumed thicker at the higher altitudes, and α – change in precipitation with elevation is not linear). Prior values of the parameters are

adopted from the semi-lumped version of the model calibrated over the full hydrologic record period as the mean value across the optimised sets. These are presented in the Appendix A. Prior values of five spatially distributed parameters are slightly corrected to account for different land use and vegetation types, and different elevation of the HRUs. Ranges of these parameters are also given in the Appendix A, while their prior and posterior spatial distributions are presented in Appendix L.

2.3.3. Sensitivity analysis and correlation among the parameters

In this research, the parameter sensitivity is estimated by employing the regression based sensitivity analysis (Christiaens & Feyen, 2002; Sieber & Uhlenbrook, 2005; Marino et al., 2008; Mishra, 2009; Pan et al. 2011). This approach relies on the multiple regression between the parameters and the model outputs (so called regression metamodel). For example, Sieber and Uhlenbrook (2005) derived regression metamodels between the parameters and the simulated runoff time series, while Christiaens and Feyen (2002) established regression metamodels between the parameters and several simulated state variables, such as peak flows, average baseflow, average soil water content and groundwater levels. Sieber and Uhlenbrook (2005) demonstrated that the results of this method corroborate the results obtained from the Regional Sensitivity Analysis (RSA), which is a commonly applied global sensitivity method.

The parameter sensitivity is quantified here with respect to two objective functions: the Nash-Sutcliffe efficiency coefficient (*NSE*) and volume error (*VE*). In this way, parameters important for reproducing of dynamics of a catchment response and overall water balance can be detected.

The parameter sensitivity is quantified in terms of standardised regression coefficients (SRC), obtained from the standardised linear regression model (metamodel) as follows (e.g. Christiaens & Feyen, 2002; Pan et al. 2011):

$$\frac{\hat{y}_j - \overline{y}}{s_y} = \sum_i^N \hat{b}_i \frac{s_{\theta_i}}{s_y} \frac{\left(\theta_{j,i} - \overline{\theta}_i\right)}{s_{\theta_i}} \tag{2.3.4}$$

In the above equation, y is the output (i.e. the objective functions *NSE* and *VE*) and the left-hand side of the equation represents the standardised output with respect to its mean value $\overline{y}$ and standard deviation s_y. Model parameters θ_i are standardised in the same manner (last term in right-hand side of the previous equation). The remaining terms in the right-hand side of the equation represent standardised regression coefficients of every model parameter:

$$SRC_i = \hat{b}_i \frac{s_{\theta_i}}{s_y} \tag{2.3.5}$$

where $\hat{b}_i$ is the regression coefficient estimate for the i^{th} parameter in the metamodel.

The standardised regression coefficients (*SRC*s) are aimed to quantify uncertainty in the output variables (here objective functions) due to model parameters. A *SRC* may take values form -1 to 1, with higher absolute values indicating higher parameter sensitivity. The sign of the coefficient is irrelevant for the sensitivity analysis (Sieber & Uhlenbrook, 2005; Pan et al. 2011).

This method is based on the assumption that the model parameters are not correlated; parameter correlation is not explicitly accounted for in estimation of the linear metamodel coefficients. Therefore, correlation among model parameters is to be examined prior to the sensitivity analysis.

The correlation coefficients among the parameters are calculated here from best ("behavioural") 100 parameter sets out of 25.000 sampled ones, according to both objective functions. Correlation among parameters of the 3DNet-Catch model is quantified in terms of both Pearson (following Blasone et al., 2007; Foglia et al. 2009 and Dotto et al. 2012) and Spearman rank correlation coefficients. The former reveals a linear relationship and the latter reveals a monotonic relationship among model parameters (Kottegoda & Rosso, 2008).

2.3.4. Evaluation of the 3DNet-Catch model performance

To test the robustness of the 3DNet-Catch model (semi-lumped, BASIC version), the Split Sample (*SST*) and Differential Split Sample test (*DSST,* Table 2) are used. For the *SST*, the model is calibrated and evaluated over 5-year long periods with similar annual precipitation depths, while for the *DSST* the model is calibrated over five wettest years, and evaluated over five driest ones. This setup of the *DSST* is reported to lead to the greatest reduction in model performance (Li et al. 2012; Vaze et al. 2010; Brigode et al. 2013) and it is therefore selected to test the model robustness and transferability. The calibration and evaluation periods considered in both tests are given in Table 13 along with mean annual precipitation depths over each simulation period. The model robustness is estimated in terms of the model ability to reproduce the overall water balance and dynamics of catchment response. In both tests, the model is calibrated using the AMALGAM algorithm with 100 parameter sets and 20.000 iterations, and *NSE* and *VE* as objective functions. The simulations start with the beginning of a water year, and one water year prior to each simulations is intended for model warm-up.

Table 13. Calibration and evaluation periods in the *SST* and *DSST* at three catchments. Values in parenthesis denote mean annual precipitation depths in the given period.

CATCHMENT	*SST*		*DSST*	
	CALIBRATION	EVALUATION	CALIBRATION	EVALUATION
Kolubara	1980 – 1985 (790)	2001-2006 (790)	1974 – 1979 (886.4)	1989 – 1994 (647.5)
Toplica	1984 – 1989 (643.6)	1999 – 2004 (643.6)	2005 – 2010 (746.8)	1989 – 1994 (547.4)
Mlava	1999 – 2004 (655.3)	2006 – 2011 (655.8)	2001 – 2006 (734.8)	1989 – 1994 (561.9)

In addition, the BASIC model version is calibrated to simulate runoff from the Kolubara River catchment at the Slovac stream gauge and at from the Toplica River catchment at the Doljevac stream gauge. The models for these catchments are calibrated over the 1996-2009 period employing the AMALGAM algorithm with *NSE* as the objective function, while the difference between mean annual observed and simulated flow volume is used as an evaluation criterion. The models are evaluated over 1985-1996 period. These

catchments and periods are selected following Langholt et al. (2013), who calibrated HBV model (e.g. Bergström et al. 1992) for the same catchments using *NSE* as the model performance measure. In this way, the 3DNet-Catch model is evaluated using the HBV as a reference model. Runoff at the Toplica River catchment is simulated using precipitation and temperature data observed at the Kuršumlija and Niš meteorological stations (following Langholt et al., 2013).

These simulations start with the beginning of a water year, and the first water year is intended for model warm-up (assumedly, the same holds for the HVB simulations presented in Langholt et al., 2013). Population of 100 parameter sets is optimised using the AMALGAM, and maximum number of function evaluations of 25.000 is set as the convergence criterion.

The catchments considered are described in detail in chapter 2.6.

2.4. Dynamic multi-objective model calibration

As outlined in chapter 1.6.2, aim of this research is further analysis of:

(1) Variability with time in optimal parameters obtained by multi-objective model calibration.

(2) Patterns of the variability (e.g. correlation between the parameters and hydro-meteorological characteristic of a calibration period or its length).

(3) Influence of the objective functions, used for model calibration on temporal variability in the Pareto-optimal parameters.

(4) Influence of model structural complexity and spatial distribution of the parameters on temporal variability in the Pareto-optimal parameters.

Methods used in these analyses are described in detail in the remaining of this chapter.

2.4.1. Temporal variability in the Pareto-optimal parameter sets

To explore variability of the Pareto-optimal parameters in time, i.e. to explore their sensitivity to calibration period, semi-lumped BASIC version of the 3DNet-Catch model (chapter 2.1.2) is calibrated in dynamic fashion over 1 to 25 consecutive water years (water year starts on 1st October and ends on 30th September). Start of every calibration period is shifted by one water year (similar to Coron et al. 2012), so there is an overlap between consecutive periods longer than one year (Figure 31). One water year of model warm-up precedes every simulation. The model is run with daily time step.

Model parameters are optimised by employing the AMALGAM with respect to two objective functions: Nash-Sutcliffe efficiency coefficient (*NSE*) and volume error (*VE*). Every calibration is performed using the same prior ranges of model parameters and the same parameters of the AMALGAM and the optimisation algorithms included. Convergence criterion for parameter optimisation is maximum number of iterations, which is kept constant in all calibrations (as elaborated in chapter 2.2), resulting in different size of the Pareto front obtained over calibration periods. Only Pareto sets that result in *NSE* greater than 0.3 are retained for the analysis.

Along with the objective functions, evaluation criteria are calculated for every calibration: Nash-Sutcliffe efficiency coefficient based on log-transformed flows ($NSE_{\log Q}$), Kling-Gupta efficiency (*KGE*) and coefficient of determination (R^2).

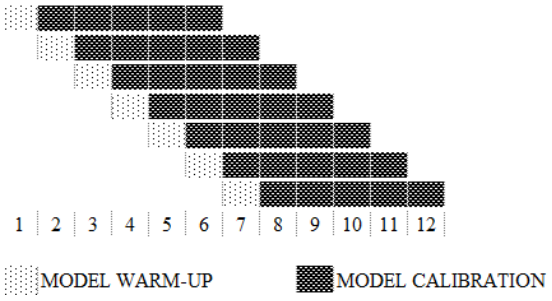


Figure 31. Five-year long calibration periods: model warm-up (light hatch) and consecutive calibration periods (dark hatch). Values on the abscissa denote water years.

2.4.2. Parameter temporal variability and hydro-meteorological characteristics of the calibration period

According to some researchers (e.g. Merz et al., 2011; Gharari et al., 2013), the variability in model parameters could be caused by presence of "secondary" processes, which are not explicitly simulated by a hydrologic model, such as variable infiltration rates due to soil freezing or cracking (Beven, 2001b; Tian et al., 2012), or variable evapotranspiration due to vegetation aging (Fenicia et al., 2009). These "secondary" processes may be linked to hydro-meteorological characteristics: for example, change in infiltration due to soil freezing of cracking may be related to the e.g. antecedent temperature and / or precipitation conditions.

Therefore, presence of correlation between the Pareto-optimal parameters and some meteorological variables is sought. Hydro-meteorological indices considered in this research are adopted after recommendations in the literature (e.g. Choi and Beven, 2007; Merz et al., 2011; Osuch et al., 2014), and presented in Table 14. The indices are estimated for every calibration period.

Correlation between the indices and Pareto-optimal parameters is quantified in terms of Pearson (Osuch et al., 2014) and Spearman (Merz et al., 2011) correlation coefficients. Considered indices are known to be correlated (e.g. mean and maximum precipitation depths), therefore principal component analysis (PCA) should be performed prior to correlation assessment. However, aim of this analysis is not to derive regression models, but to inspect for period characteristics that Pareto-optimal parameters may be sensitive to. Higher correlation coefficient indicates higher importance of a particular hydro-meteorological characteristic (according to Christiaens & Feyen (2002), correlation coefficients may be used to estimate sensitivity). Additionally, impact of a particular hydro-meteorological characteristic is quantified in terms of variable importance, which is obtained by applying bootstrap aggregating ("tree bagging") metamodel with 200 decision trees[16].

[16] The number of trees is selected after decrease in model error.

Table 14. Hydro-meteorological indices considered in the analysis of the parameter variability

Meteorological variable	Description
Precipitation-related indices	
P_{mean_daily} [mm]	Mean daily precipitation depth over a calibration period
P_{mean_rainy} [mm]	Mean daily precipitation depth during rainy days in a calibration period. A day is considered to be rainy if daily precipitation depth exceeds 0.1 mm.
P_{max} [mm]	Maximum daily precipitation depth in a calibration period
Std*P* [mm]	Standard deviation of daily precipitation depths in a calibration period
API30 [mm]	Mean 30-days antecedent precipitation index. *API* is calculated by applying the equation (Berthet et al., 2009; Kohler & Linsley, 1951; Raghunath, 2006): $API(t) = \sum_{i=0}^{t} K^{(t-i)} P_i$ Parameter *t* denotes the number of days in a period that *API* is estimated over (30 days), while *K* may take value from 0.85 to 0.98. *API*30 is estimated for every day of a calibration period with $K = 0.9$, and the *API*30 value is obtained by averaging *API*30 over an entire calibration period. The same method is applied for estimating other antecedent indices.
API5 [mm]	Mean 5-day antecedent precipitation index in a calibration period
N_{rainy_days} [-]	Number of rainy days in a calibration period, normalised with respect to the calibration period length
Temperature-related indices	
T_{mean_daily} [°C]	Average mean daily temperature in a calibration period
T_{min} [°C]	Minimum mean daily temperature in a calibration period
T_{max} [°C]	Maximum mean daily temperature in a calibration period
Std*T* [°C]	Standard deviation of mean daily temperature in a calibration period
ATI5 [°C]	Mean 5-days antecedent temperature index over a calibration period
ATI30 [°C]	Mean 30-days antecedent temperature index over a calibration period
N_{ice_days} [-]	Number of ice days in a calibration period, normalised with respect to the calibration period length. * As sub-daily data are not available, in this research a day is considered an ice day if mean daily temperature does not exceed 0°C.
PET_{mean_daily} [mm]	Mean daily *PET* rate over a calibration period
Hydrologic variables	
Q_{mean} [m^3 / s]	Mean daily observed flow over a calibration period

2.4.3. Impact of the objective functions on temporal variability in the Pareto-optimal parameters

To assess impact of the combination of objective functions used for model calibration on the parameter temporal variability, semi-lumped BASIC version of the model is calibrated using several different combinations of the objective functions (hereafter

referred to as calibration strategies). The calibration strategies considered are listed in Table 15, while the definitions of particular objective functions are given in the chapter 1.3.3. Note that the number of strategies is limited, because the multi-objective calibration requires mutually "conflicting" objective functions (chapter 1.3.5) and not highly correlated ones.

The model is calibrated over all overlapping 5-year long periods, with one water year of model warm-up prior to every simulation. Prior parameter ranges and the AMALGAM parameters are kept constant, regardless of the calibration strategy employed.

The parameter identifiability and model performance for given different calibration strategies are evaluated along with the consistency in parameter estimates.

Table 15. Calibration strategies considered in this research

Calibration strategy	Number of obj. funct.	Objective functions used for the model calibration
1	2	Nash-Sutcliffe for flows and volume error
2	2	Nash-Sutcliffe for flows and log-transformed flows
3	2	Kling-Gupta efficiency and volume error
4	2	Coefficient of determination and volume error
5	2	Root mean square error based on high and low flows (Fenicia et al. 2007)
6	2	Heteroskedastic maximum likelihood estimator and root mean square error
7	3	Nash-Sutcliffe for flows and log-transformed flows, and volume error

2.4.4. Impact of the model structure on temporal variability of the Pareto-optimal parameters

To enable analysis of the model structural complexity impact on the consistency in parameter estimates, four versions of the 3DNet-Catch model are developed and presented in chapter 2.1.2.

The models are calibrated over 5-year long overlapping periods, with one water year of model warm-up preceding every calibration period. The model is calibrated using *NSE* and *VE* as objective functions and using the same AMALGAM parameters for all model structure versions. Calibration of the semi-lumped model versions is described in chapter 2.3.1, while the regularisation method applied for calibration of the fully-distributed model version is elaborated in chapter 2.3.2. Prior parameter ranges in this analysis vary

with the model structure, and they are specified in Appendix A. As for the distributed model version, the spatial parameter fields are represented by super-parameters, which are optimised. Consequently, the consistency analysis of this model version is based on the optimised super-parameters.

In addition to the parameter variability with the calibration period, parameter identifiability and model performance are analysed as well.

2.5. Assessment of temporal consistency in parameter estimates and in the model performance

Dynamic multi-objective model calibration results in an ensemble of the Pareto-optimal parameter sets for each calibration period. As a result of the adopted convergence criterion for the AMALGAM algorithm, number of the Pareto sets in the ensemble varies with the calibration period.

Distribution of the Pareto-optimal parameter values describes the parameter variability for each parameter and each calibration period (or uncertainty due to calibration period)[17]. Central tendency and dispersion measures of this distribution are analysed. The median is preferred over the arithmetic mean as the central tendency measure due to its resistance to presence of outliers (Kottegoda & Rosso, 2008), while the parameter dispersion in the calibration period is quantified in terms of the information content (*IC*) value. The latter is estimated following the approach presented by Wagener et al. (2003):

$$IC = 1 - \left[\hat{\theta}_{norm,\,97.5} - \hat{\theta}_{norm,\,2.5} \right] \tag{2.5.1}$$

where $\hat{\theta}_{norm,\,97.5}$ and $\hat{\theta}_{norm,\,2.5}$ denote 2.5th and 97.5th percentiles of the distribution of the normalised Pareto-optimal parameters $\hat{\theta}_{norm}$, respectively. This statistic also enables quantifying the parameter identifiability: the narrower the optimised parameter range, the larger is the *IC* value and the parameter identification is better.

[17] Here, the term “uncertainty” is used in a broader sense.

The parameter values may differ for several orders of magnitude (e.g. sub-surface soil layer thickness in millimetres and effective porosity in fraction between 0 and 1). The normalisation enables comparison among different parameters because the normalised parameters take values from 0 to 1 regardless of their prior ranges (Vrugt et al., 2006; Luo et al., 2012). The optimised parameters θ_{PARETO} are therefore normalised with respect to the lower and upper bounds θ_{MIN} and θ_{MAX} of the prior parameter range:

$$\hat{\theta}_{norm} = \frac{\hat{\theta}_{\mathrm{PARETO}} - \theta_{\mathrm{MIN}}}{\theta_{\mathrm{MAX}} - \theta_{\mathrm{MIN}}} \tag{2.5.2}$$

During calibration, the parameter values are sampled from the uniform probability distribution with bounds θ_{MIN} and θ_{MAX}.

To illustrate the overall sensitivity of the Pareto-optimal parameters to calibration period and the changes in parameter identifiability, medians and *IC* statistic of the normalised parameters are presented in multi-temporal graphs (e.g. Hannaford et al. 2013).

Temporal parameter variability is quantified in terms of standard deviation $S_{\mathrm{t},\,i}$ of the ensemble medians $\mathrm{Me}_j(\theta_i)$, where j denotes calibration period ($j = 1, 2, ..., N_{\mathrm{cal}}$) from N_{cal} calibration periods, and i refers to the i^{th} model parameter. On the other hand, standard deviation $S_{\mathrm{u\_prior},\,i}$ of all initially sampled values of the i^{th} parameter from the prior uniform distribution is (e.g. Kottegoda and Rosso 2008):

$$S_{\mathrm{u\_prior},\,i} = \frac{\theta_{\mathrm{MAX},\,i} - \theta_{\mathrm{MIN},\,i}}{\sqrt{12}} \tag{2.5.3}$$

This standard deviation describes initial variability of a parameter. If the optimised parameters significantly vary with the calibration period, standard deviation S_{t} of the temporal parameter variability is expected to exceed the initial variability and vice versa. Therefore, parameter temporal consistency is estimated in terms of ratio of these two standard deviations $S_{t,\,i}$, and $S_{\mathrm{u\_prior},\,i}$:

$$t_{\theta_i} = \frac{S_{\mathrm{t},\,i}}{S_{\mathrm{u\_prior},\,i}} \tag{2.5.4}$$

This ratio enables estimating the parameter variability in time compared to its initial uncertainty (Vrugt et al., 2008). Smaller ratio indicates more consistent parameter estimate. Values greater than one suggest that the uncertainty due to calibration period

exceeds initial uncertainty, i.e. that the parameter is rather sensitive to calibration period. The ratios estimated for the calibration periods of increasing lengths are used to inspect whether an increase in the calibration period length leads to more consistent parameter estimates.

Additionally, parameter variability with calibration period is quantified in terms of standard deviation of the median values of the normalised Pareto-optimal parameters, $S_{t,\,norm}$, obtained from all calibration periods of given length. The values of $S_{t,\,norm}$, calculated for periods of increasing length indicate whether the parameter sensitivity decreases with an increase of the calibration period length.

Along with the parameter estimates and *IC* statistic variability in model performance is analysed. Model performance is quantified in terms of medians of the objective functions and evaluation criteria (chapter 2.4) obtained from the Pareto-optimal ensemble. In addition, performance of the Pareto-optimal ensembles is quantified in terms of *p*-factor and *r*-factor. The former represents per centage of observations within the 95% prediction band (95PPU), while the later quantifies relative width of the 95PPU (Schuol and Abbaspour, 2006; Yang et al., 2008; Zhang et al., 2011):

$$r\_\text{factor} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{\text{SIM},97.5\%} - Q_{\text{SIM},2.5\%}\right)}{\sigma_{Q_{\text{OBS}}}} \tag{2.5.5}$$

At any point in time 95PPU is calculated as a difference between the predicted variables (simulated flows) corresponding to 2.5^{th} and 97.5^{th}, respectively. Target value of *p*-factor is one, whereas *r*-factor should approach zero (Zhang et al., 2011). Bastola et al. (2011) referred to the *p*-factor as the "count efficiency", .

Correlation between hydro-meteorological indices (Table 14) and median values and *IC* statistic of the Pareto-optimal parameters is quantified in terms of the Pearson and Spearman correlation coefficients (chapter 2.4) which are calculated according to all calibration periods.

Impact of the objective functions on the variability in Pareto-optimal parameters, the *IC* statistic and overall model performance is analysed by calibrating the model over 5-year

long periods using different combinations of objective functions (chapter 2.3.1). Parameter variability is quantified in terms of $S_{t,\,i}$ / $S_{\mathrm{u\_prior}}$ for each calibration strategy. Along with this ratio, mean values of the *IC* statistic and model performance measures over all 5-year long calibration periods are calculated. Model performance is quantified in terms of mean number of the Pareto sets, median values of *NSE*, *VE* and $NSE_{\log Q}$ and *p*- and *r*-factors.

Impact of the model structural complexity on consistency in parameter estimates, values of *IC* statistic and model performance is assessed analogously with the impact of the objective functions.

2.6. Catchments and data

The methodology outlined in the previous chapters does not discriminate between parameter variability stemming from the properties of the calibration period and variability due to anthropogenic effects (e.g. changes in land use type, deforestation or afforestation, river training measures, etc.). To isolate variability with the calibration period, only catchments that have not undergone human-induced changes are considered in this research.

Since daily data (precipitation depths, temperatures and flows) are made available for purposes of this research, areas of the catchments considered should be sufficiently large to enable hydrologic simulations using daily time step. On the other hand, catchment area should be sufficiently small to allow approximation of the catchment response by semi-lumped models.

Three catchments in Serbia are found to meet these requirements: the Toplica, Kolubara and Mlava River catchments (Figure 32 and Table 16). Selected rivers belong to the Danube River basin. Stream gauges at these rivers are selected according to length of the hydrologic record periods and reliability of the observations. Selected stream gauges are presented in Table 16.

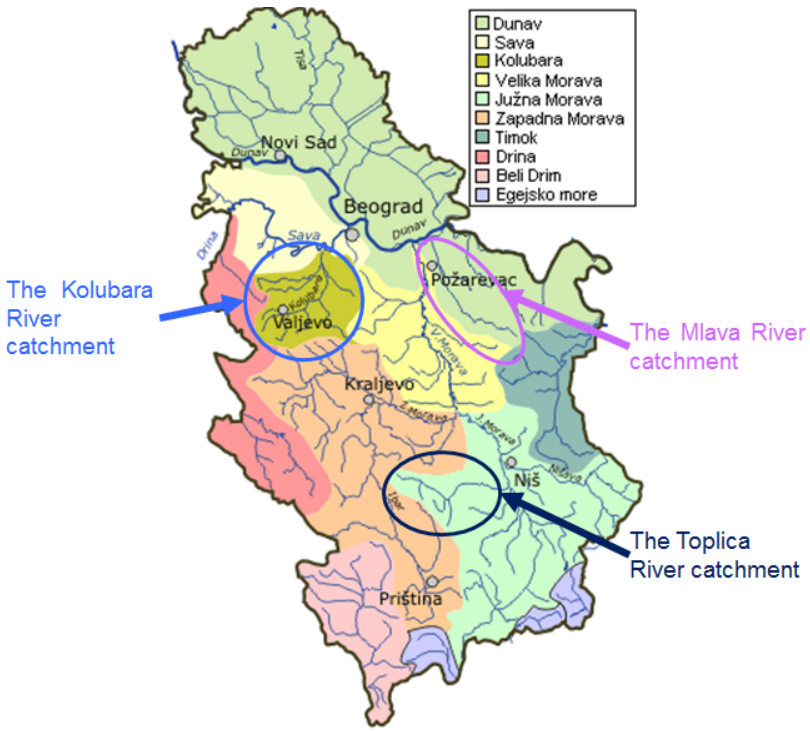


Figure 32. Major catchments in Serbia (Republic Hydrometeorological Service of Serbia). Three catchments considered in the research are highlighted.

Table 16. Properties of the stream gauges considered in this research

River	Stream gauge	Drainage area [km^2]	Distance from the confluence [km]	Beginning of the record period
Kolubara	Slovac	995	88	1954
Toplica	Doljevac	2052	2.5	1950
Mlava	Veliko Selo	1277	48.2	1986

2.6.1. The Toplica River catchment upstream of the Doljevac stream gauge

Catchment description

The Toplica River drains into the Juzna Morava River. The catchment of 2052 km^2 is situated in the southern Serbia. Topography of this catchment ranges in elevation from 193 to 1996 m.a.s.l., with mean catchment elevation of 621.82 m.a.s.l. (Figure 33 and Figure 34). Forests are dominant land use type at the catchment, while less than 1% of the total area is urbanised (Figure 35). Prevailing soil types in the catchment (Figure 36)

are smonitza soils (hydrologic soil group D) and acid, brown and podzolic soils (hydrologic soil group B). Remaining soils also belong to hydrologic soil group B (Djorković, 1984). Sandstones prevail, and there is no karst in the catchment (Figure 37).

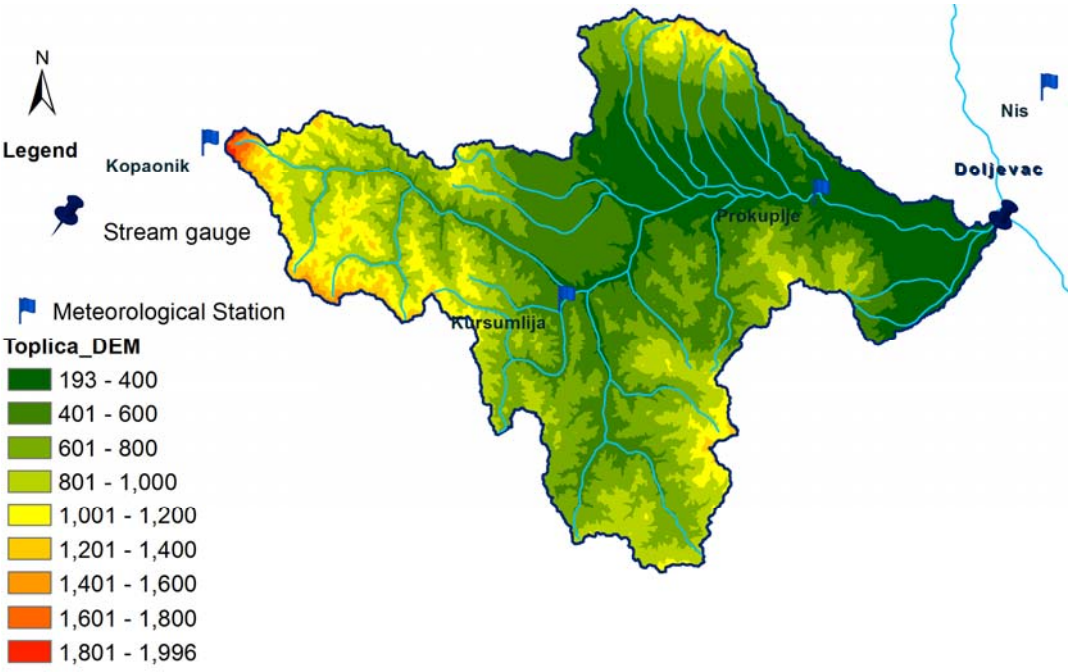


Figure 33. Topographic map of the Toplica River catchment.

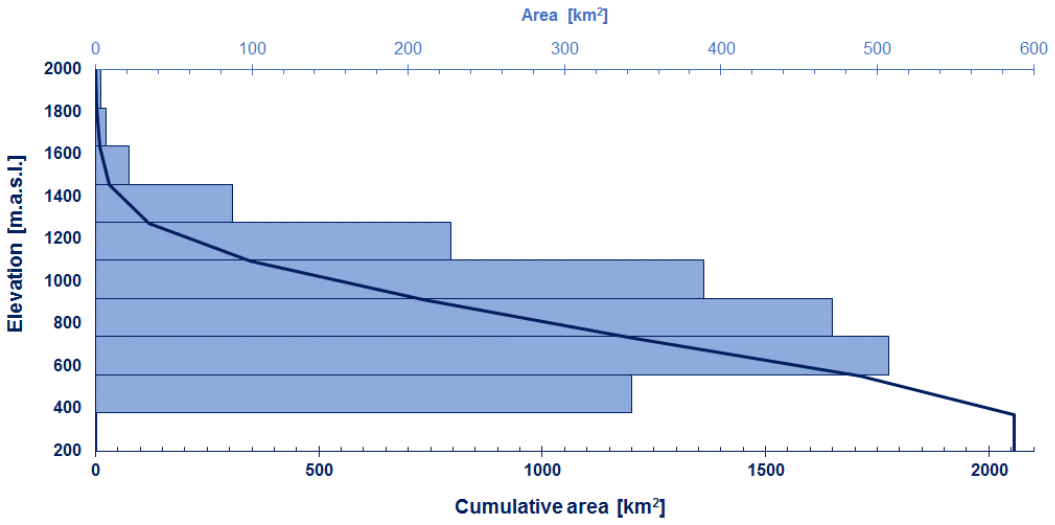


Figure 34. Hypsometric curve for the Toplica River catchment.

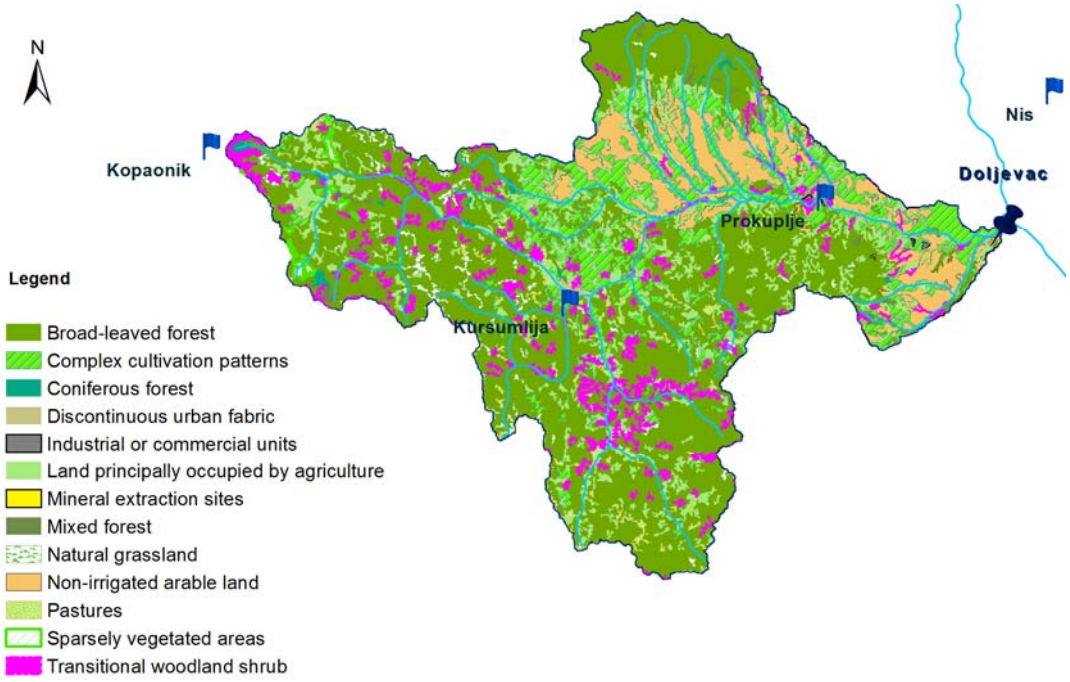


Figure 35. Land use types in the Toplica River catchment (CORINE 2006).

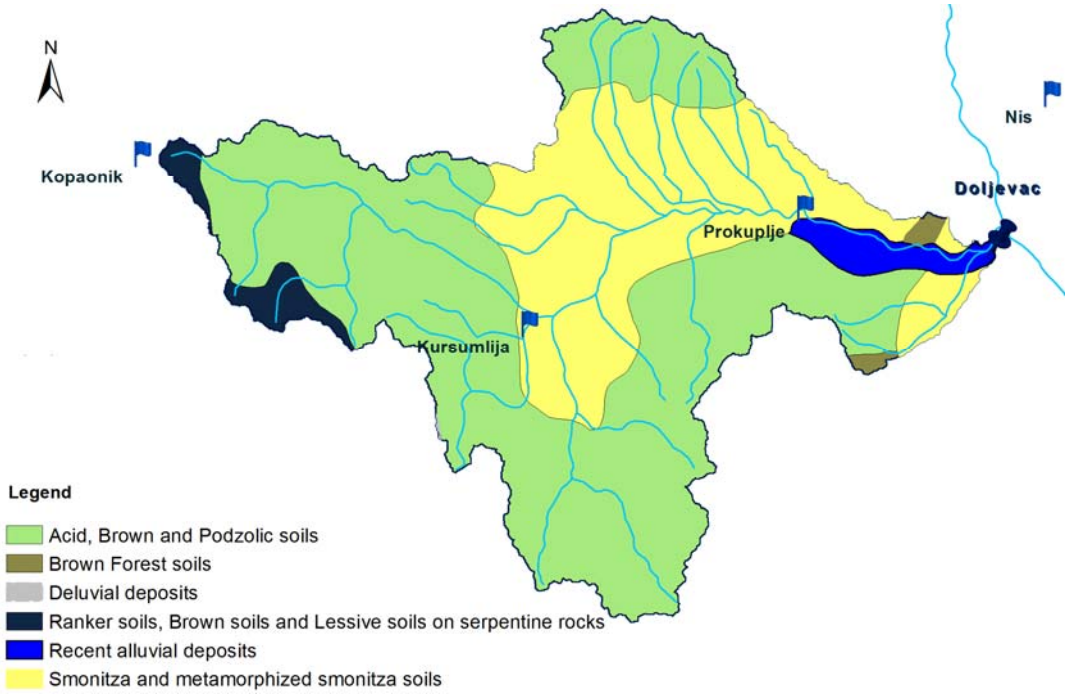


Figure 36. Soil types in the Toplica River catchment.

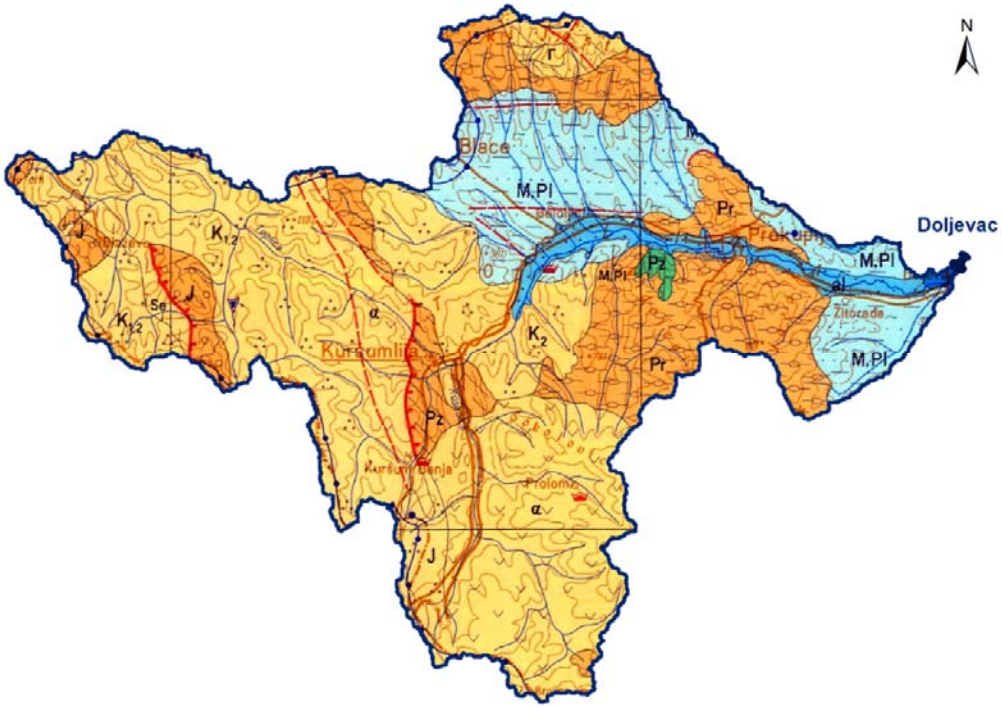


Figure 37. Geological structure of the Toplica River catchment: al – alluvium; J – serpentinites; $K_{1,2}$ – siltstone, sandstone and marl; K_2 – siltstone, marl and marly limestone; M,Pl – upper Miocene and Lower Pliocene clastics; Pr – magmatites; Pz – marbles.

Data

Gauging stations used for simulation of runoff from this catchment (Figure 33) are presented in Table 17. Elevation of stream gauges refers to the zero datum of the staff gauge. Meteorological stations that are not situated within the catchment area are shaded.

Due to numerous gaps in the record from the Kopaonik meteorological gauge, observations prior to 1980 are discarded. Hence, only observations made from 1980 to 2013 are used for runoff simulations (water years: 1st of October 1980 to 30th of September 2013). There are gaps in precipitation (2.3% of the observations) and temperature data (2.8% of the observations) at the Prokuplje station. Missing data are estimated using multiple linear regression and observations from the Niš and Kuršumlija stations (correlation coefficient is 0.65 for precipitation depths and 0.99 for temperatures).

Mean catchment precipitation depths, temperatures and reference altitude of the meteorological stations are estimated by employing a weighting method (chapter 2.1.2). Thiessen polygons and weights of the stations are given in Figure 38 and in Table 18. Reference altitude is the mean catchment elevation of 488.9 m.a.s.l.

Table 17. Hydro-meteorological stations available for runoff simulation at the Toplica River catchment (RHMSS).

Station	Variable	Elevation [m.a.s.l.]	Latitude	Longitude	Record period	Mean observed value (1980-2013)
Doljevac	Q	190.41	43 ° 11 ’	21 ° 49 ’	1954-2013	8.77 [m^3/s]
Kopaonik	P, T	1711	43 ° 17 ’	20 ° 48 ’	1967-2013	977.9 [mm] 3.7 [°C]
Kuršumlija	P, T	383	43 ° 08 ’	21 ° 16 ’	1961-2013	631.1 [mm] 10.4 [°C]
Prokuplje	P, T	266	43 ° 14 ’	21 ° 36 ’	1951-2013	549.9 [mm] 10.9 [°C]
Niš	P, T	204	43 ° 20 ’	21 ° 54 ’	1947-2013	576.6 [mm] 11.9 [°C]

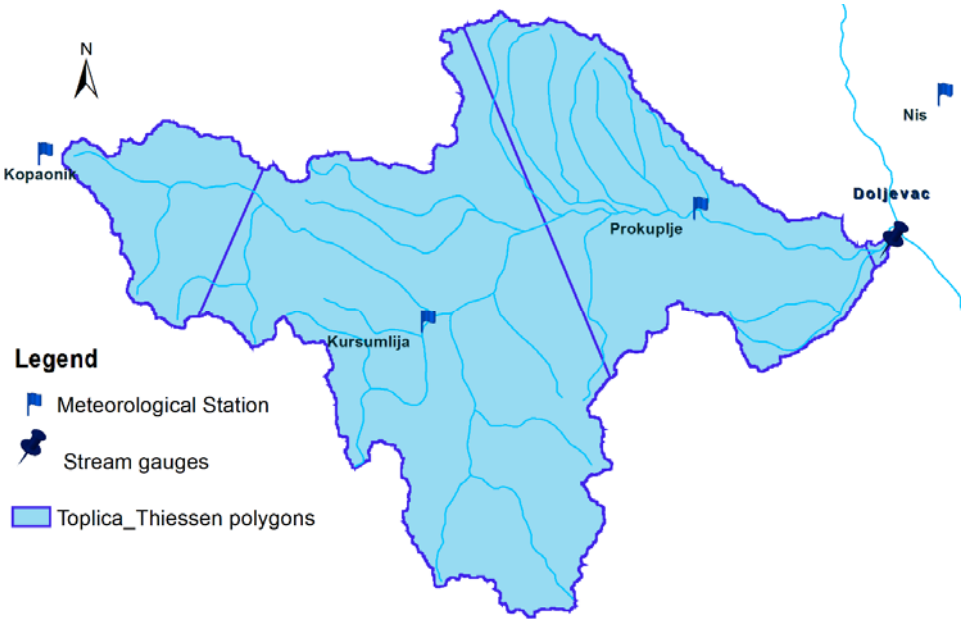


Figure 38. Thiessen polygons for meteorological stations in the Toplica River catchment.

Table 18. Weights of the meteorological stations in the Toplica River catchment.

Meteorological station	Area [km^2]	weight ω
Kopaonik	223.87	0.109
Kuršumlija	1158.92	0.563
Prokuplje	671.24	0.326
Niš	4.36	0.002

Hydrometeorological regime

Annual precipitation depths and mean annual temperature are presented in Figure 39. Mean annual flows and estimated annual runoff coefficients (ratio of total annual runoff to annual precipitation depth) are given in Figure 40.

The observed series are tested for presence of trend. Pearson and Spearman correlation coefficients are calculated, along with the *p* values of the regression slopes (probabilities of the statistics[18]) and presented in Table 19. The values of *p* less than 0.025 or greater than 0.975 (two-sided test, 95% confidence interval) indicate statistically significant trend in series. The results indicate an increasing trend in mean annual temperatures (shaded cells) and absence of statistically significant trends in other series.

The long term mean flow at the Doljevac stream gauge is 8.77 m^3/s (Table 17) and mean water yield of the catchment amounts to 4.27 L s^{-1} km^{-2}. According to the flow duration curve (Figure 41) median flow is 5.3 m^3/s. Intra-annual distribution of flows (Figure 42) shows distinct seasonality: the highest monthly flows (~20 m^3/s) occur in April (snowmelt and rainfall on the saturated soil), while low flows (~3 m^3/s) are observed in the late summer and early autumn. High flows exhibit wider dispersion around expected values (50th percentile). Annual runoff coefficient in the simulation period varies from 0.11 (in 1994) to 0.35 (in 2006), as shown in Figure 40. Mean annual runoff coefficient in this period amounts 0.21.

[18] As for Pearson correlation coefficient, *p* value is obtained as the probability of *t* statistic (linear regression slope test), which can be approximated by Student distribution.

Intra-annual distributions of precipitation and temperature are presented in Figure 43 and Figure 44, respectively. Temperature follows a distinct pattern, with the highest temperatures being observed in August, and lowest in January and February. Unlike high temperatures, low temperatures deviate significantly from the median values. Minimum monthly precipitation depths are observed in September (~40 mm) and maximum during June and July (somewhat greater than 60 mm). Extreme values widely diverge from the median values and show no clear pattern.

Precipitation and temperature gradients with elevation are presented in Figure 45 and Figure 46, respectively. These gradients are estimated for every year without gaps in the observations. The gradients are assessed according to annual precipitation depths and mean annual temperatures observed at each meteorological station, and the station altitude. Temperature decreases with elevation for approximately 0.5°C / 100 m, while precipitation increases with elevation from ~2.5% to ~6.5% per 100 m. In this catchment annual precipitation depths and mean annual temperatures have rather strong correlation to elevation (correlation coefficient exceeds 0.8).

Table 19. Trends in annual precipitation depths, mean annual temperatures and flows, and annual runoff coefficients in the Toplica River catchment (1980-2013).

VARIABLE	Pearson		Spearman	
	rho	*p* value	rho	*p* value
Precipitation [mm]	0.14	0.43	0.17	0.34
Temperature [°C]	0.58	$3 \cdot 10^{-4}$	0.55	$8 \cdot 10^{-4}$
Flow [m^3 /s]	-0.09	0.59	-0.12	0.48
Runoff coefficient [/]	-0.22	0.21	-0.18	0.31

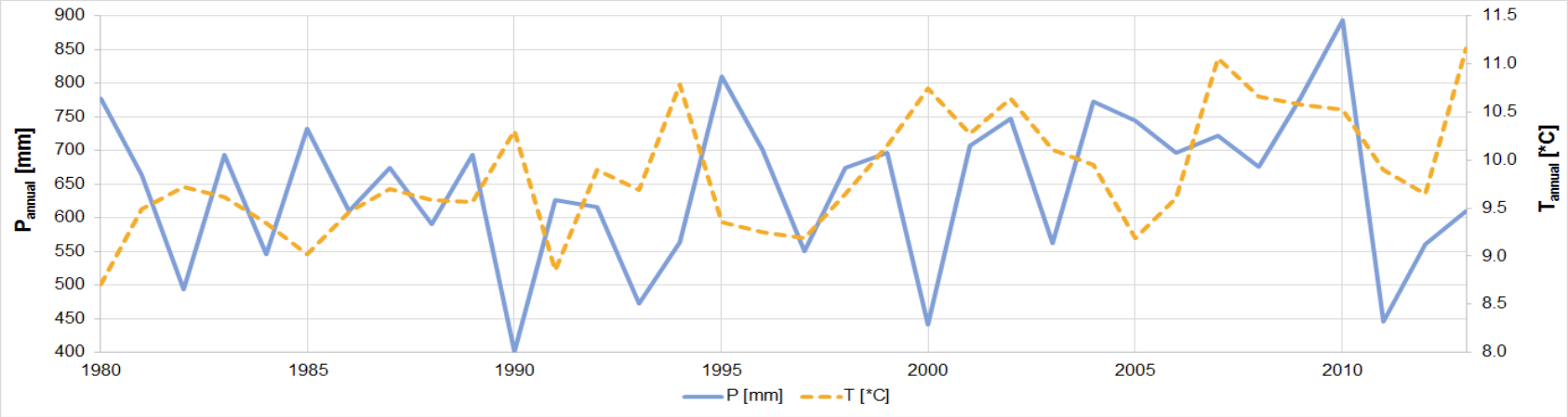


Figure 39. Annual precipitation depths and mean annual temperatures at the Toplica river catchment.

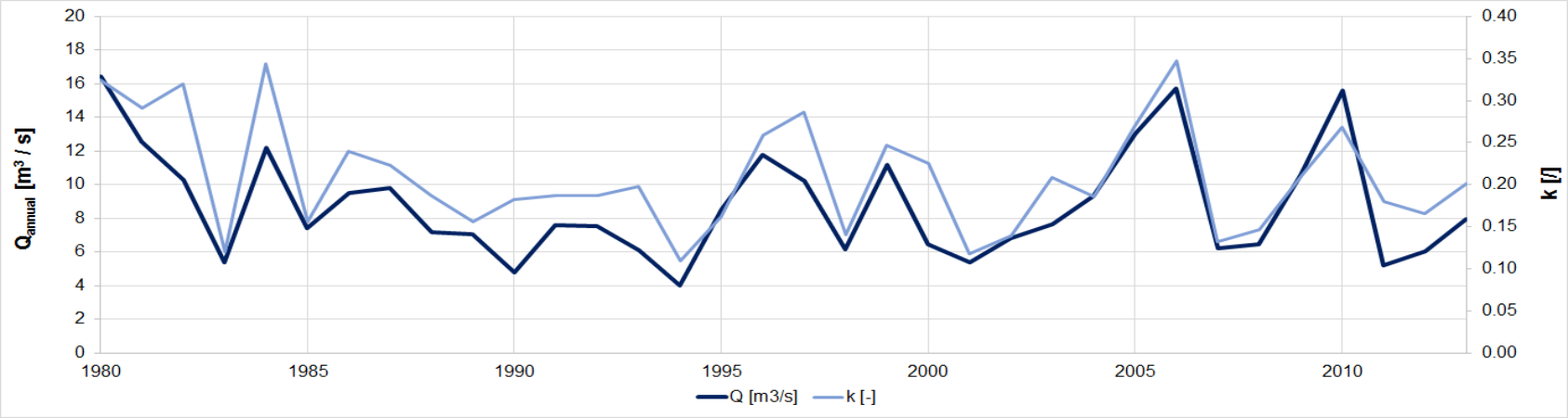


Figure 40. Mean annual flows and annual runoff coefficient at the Doljevac stream gauge.

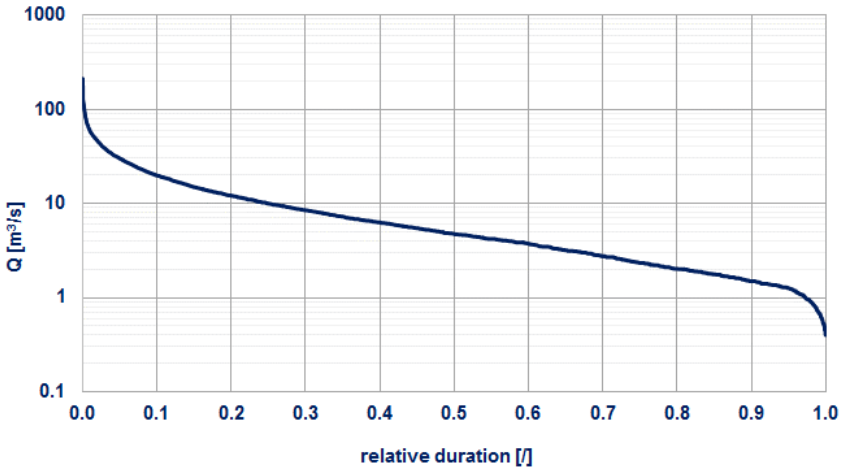


Figure 41. Flow duration curve derived from observed daily flows at the Doljevac stream gauge from 1980 to 2013.

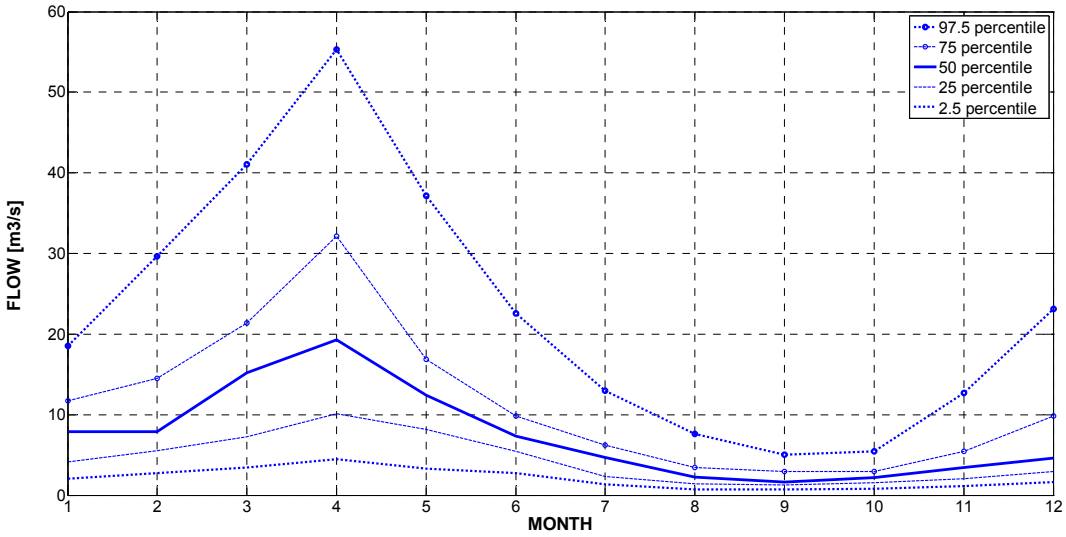


Figure 42. Intra-annual distribution of flows observed at the Doljevac stream gauge.

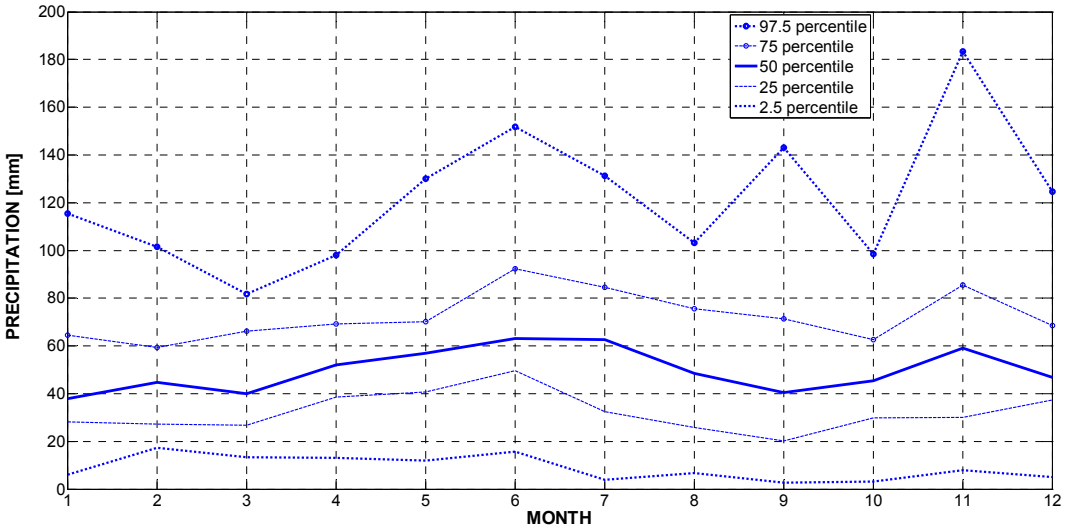


Figure 43. Mean monthly precipitation depths in the Toplica River catchment.

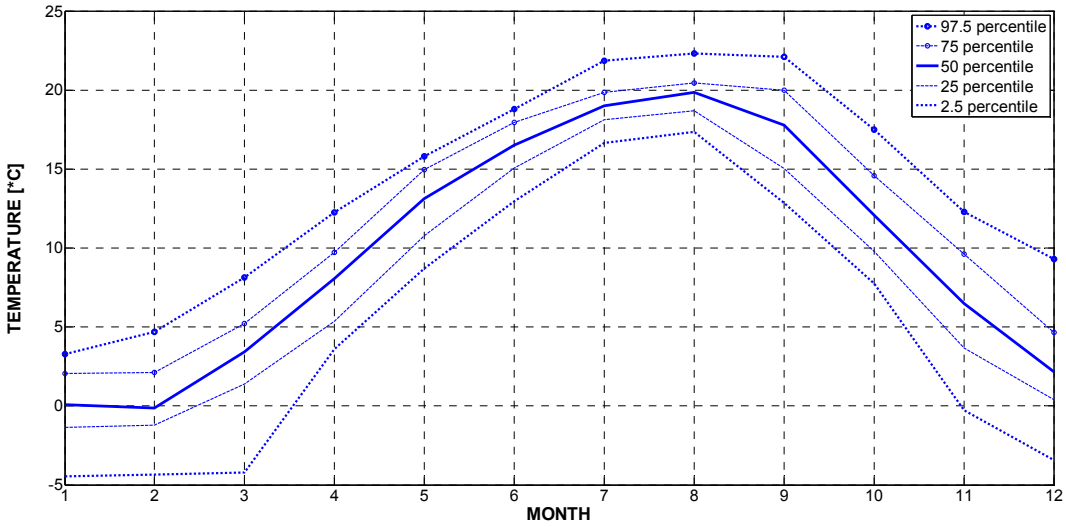


Figure 44. Mean monthly temperatures in the Toplica River catchment.

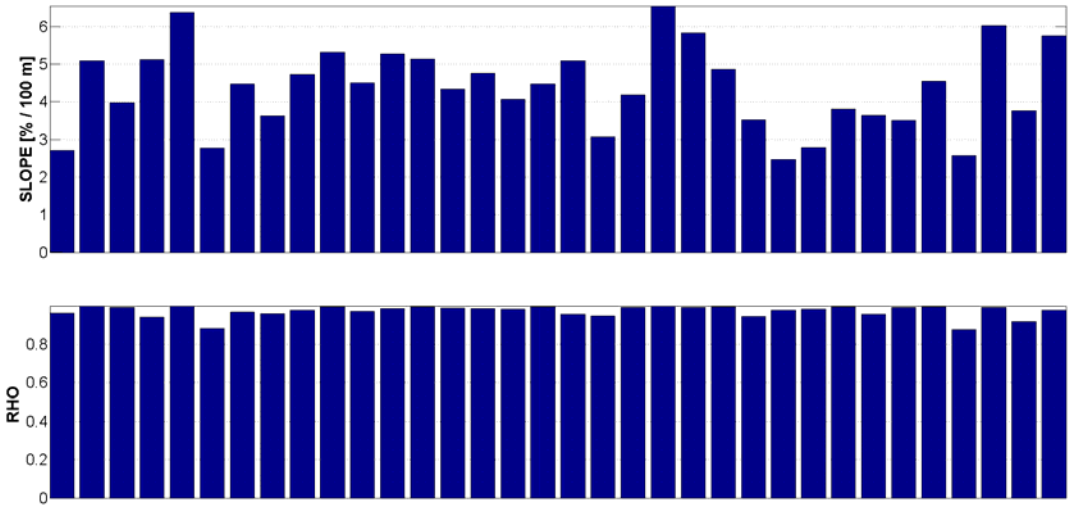


Figure 45. Precipitation gradient with elevation in the Toplica River catchment: slope of the linear regression and correlation coefficient in various water years.

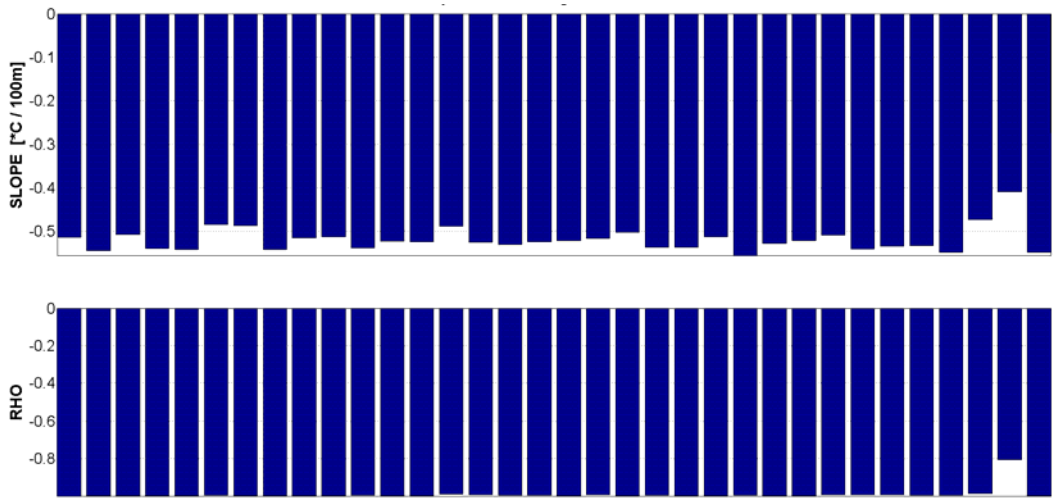


Figure 46. Temperature gradient with elevation in the Toplica River catchment: slope of the linear regression and correlation coefficient in various water years.

2.6.2. The Mlava River catchment upstream of the Veliko Selo stream gauge

Catchment description

The Mlava River drains into the Danube. The catchment area of 1255 km^2 ranges in elevation from 93 m.a.s.l. to 1333 m.a.s.l. (Figure 47), with mean catchment elevation of 366.33 m.a.s.l. (Figure 48). As shown in Figure 49, less than 2.5% of the catchment area is urbanised, while the forests and agricultural land prevail. Soil types at this catchment are presented in Figure 50. With exception of smonitza soils, which belong to hydrologic soil group D, soil types in the catchment are of hydrologic group B (Djorković, 1984). Geological structure of the catchment (Figure 51) indicates presence of karst in the upper (southern) parts of the catchment.

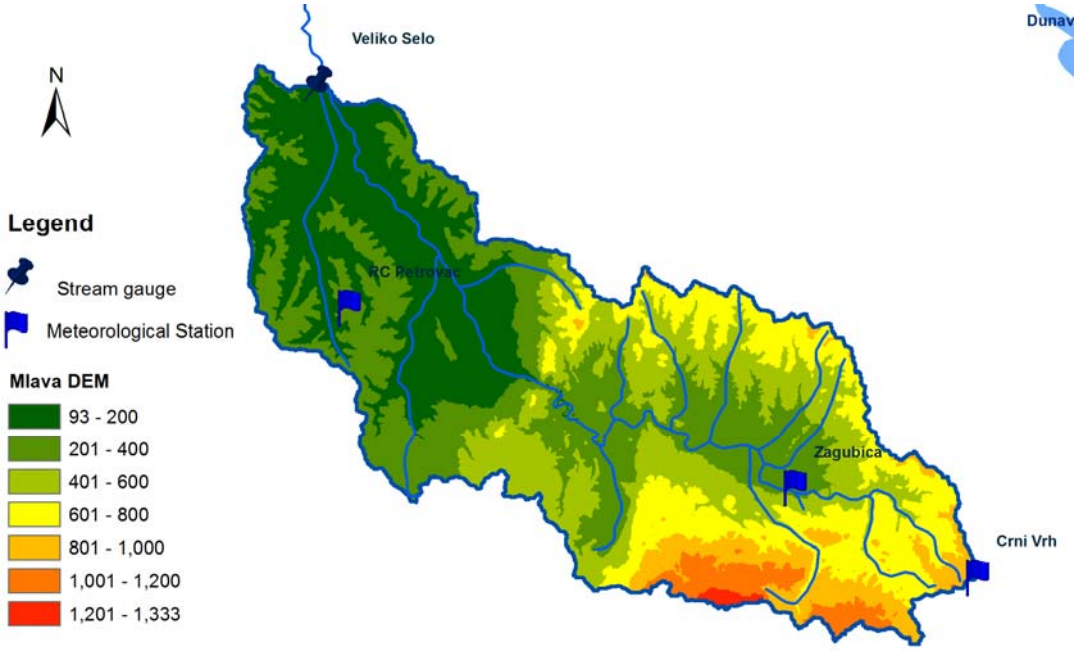


Figure 47. Topographic map of the Mlava River catchment.

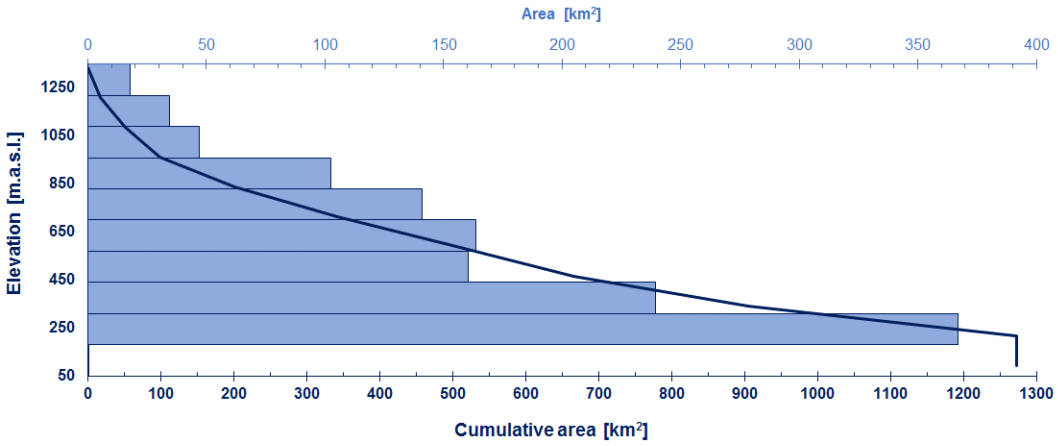


Figure 48. Hypsometric curve for the Mlava River catchment.

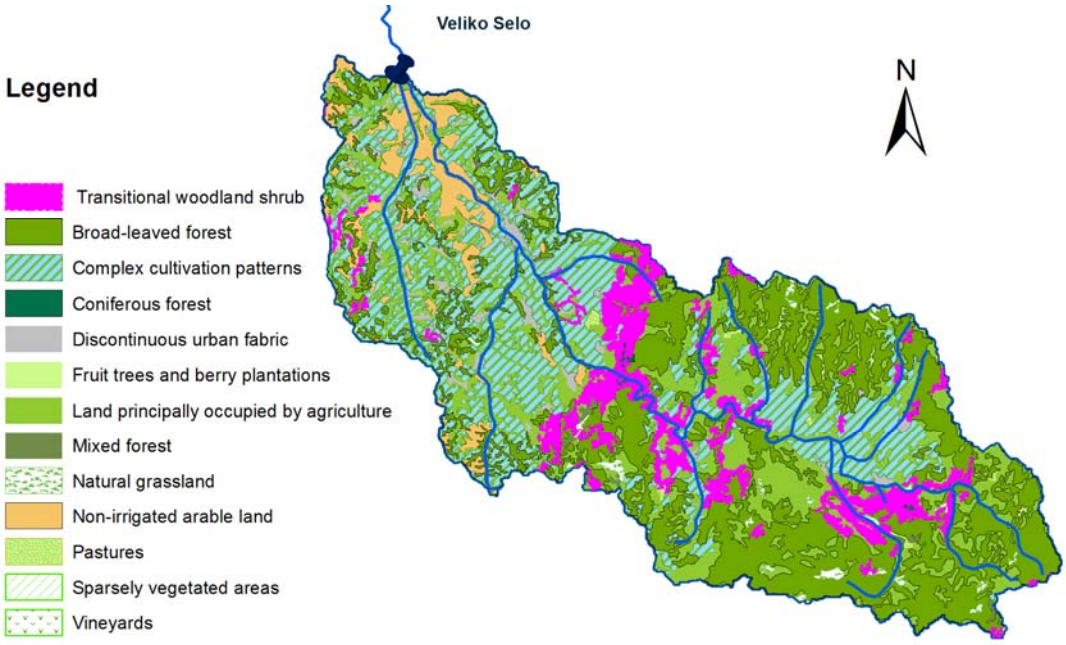


Figure 49. Land use types in the Mlava River catchment (CORINE 2006).

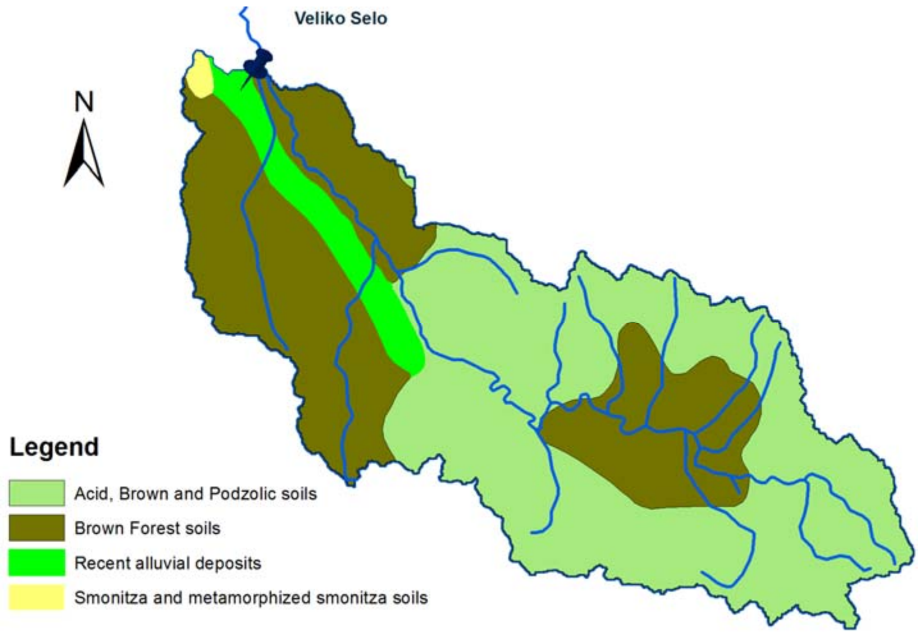


Figure 50. Soil types in the Mlava River catchment.

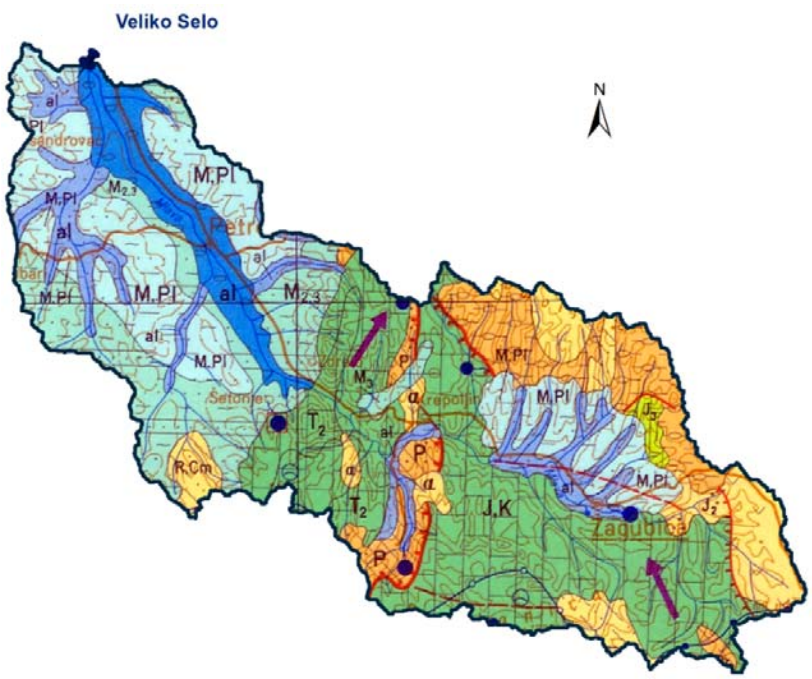


Figure 51. Geological structure of the Mlava River catchment: al – alluvium; J_2 – clastic and carbonate rocks; J_3 – massive, most frequently reef limestone; J, K – limestone; $M_{2,3}$ – sandstones. sands and clays (Middle-Upper Miocene); M_3 – sandstones, sands and clays (Upper Miocene); M, Pl – limnic sediments; P – red sandstones; T_2 – limestone (Middle Triassic).

Data

Properties of the gauging stations used for runoff simulation at the Mlava river catchment (Figure 47) are given in Table 20. Although observations from the Smederevska Palanka and Veliko Gradiste meteorological gauges are not used for runoff simulations, they facilitated assessing precipitation and temperature gradients with elevation.

There are few gaps in the observations (0.3% of missing data in daily precipitation series and in mean daily temperatures form the Pertovac meteorological station). These gaps are filled using (multiple) linear regression (correlation coefficient 0.6 for precipitation depths, and 0.98 for temperatures). Thus, full hydrologic record period is used for runoff simulations (1st of October 1987 to 30th of September 2013).

Mean precipitation depths and temperatures over the catchment are estimated form the observations form the Crni Vrh, Zagubica and RC Petrovac meteorological gauges (Figure 52, Table 20). Reference altitude of these gauges (z_{MS}) amounts 346.9 m.a.s.l.

Table 20. Hydro-meteorological stations available for runoff simulation at the Mlava River catchment (RHMSS).

Station	Variable	Elevation [m.a.s.l.]	Latitude	Longitude	Available data form	Mean observed value (1987-2013)
Veliko Selo	Q	92.55	44 ° 30 ’	21 ° 18 ’	1987	7.5 [m^3/s]
RC Petrovac	P, T	282	44 ° 20 ’	21 ° 20 ’	1972	688.2 [mm] 11.8 [°C]
Zagubica	P, T	314	44 ° 12 ’	21 ° 47 ’	1972	614.8 [mm] 10.3 [°C]
Crni Vrh	P, T	1027	44 ° 08 ’	21 ° 58 ’	1981	780 [mm] 6.9 [°C]
Smederevska Palanka	*P, T*	*121*	*44 ° 22 ’*	*20 ° 57 ’*	*1985*	*611 [mm]* *11.8 [°C]*
Veliko Gradiste	*P, T*	*80*	*44 ° 45 ’*	*21 ° 31 ’*	*1985*	*659.7 [mm]* *11.6 [°C]*

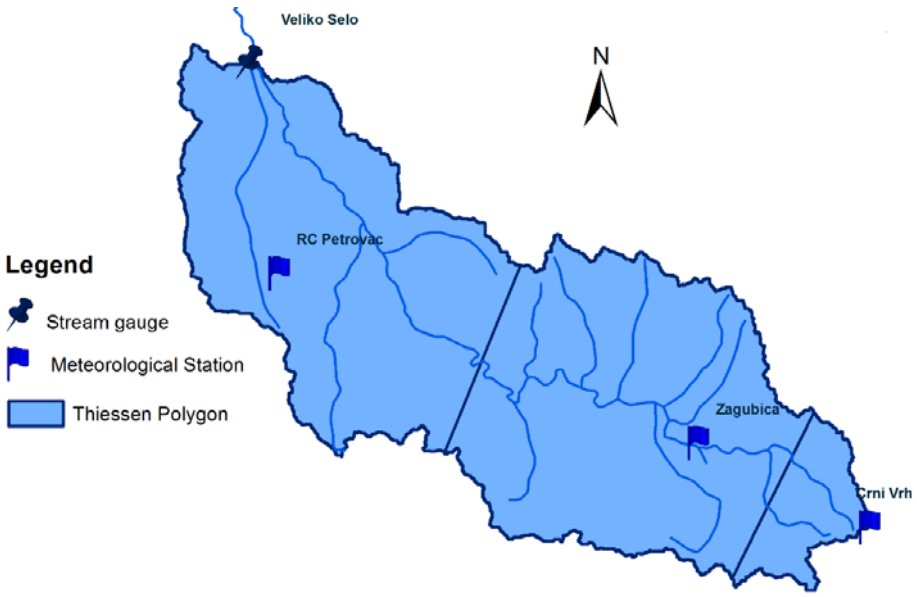


Figure 52. Meteorological stations and Thiessen polygons in the Mlava River catchment.

Table 21. Weights of the meteorological stations in the Mlava River catchment.

Meteorological station	Area [km^2]	weight ω
RC Petrovac	592.3	0.464
Zagubica	599.9	0.469
Crni Vrh	85.6	0.067

Hydrometeorological regime

The analysis of (hydro) meteorological regime at the catchment, with exception of changes in precipitation and temperatures with elevation, is based on the observations from three stations that are used for runoff simulations.

Annual precipitation depths and mean annual temperatures at the catchment are presented in Figure 53. Mean annual flows and estimated runoff coefficients are given in Figure 54. Mean value of the annual runoff coefficient amounts to 0.28, although there are significant variations (from 0.13 to 0.51).

These series are tested for presence of trend. Pearson and Spearman correlation coefficients, along with the *p* values of the regression slopes are given in Table 22.

Statistically significant trend is detected only in mean annual temperatures (95% significance level).

Mean flow at the Veliko Selo stream gauge is 7.5 m^3/s (Table 20), that is, specific water yield at the catchment amounts to 5.9 L s^{-1} km^{-2}. Median daily flow is 4.1 m^3/s (Figure 55). Intra-annual distribution of flows (Figure 56) exhibits clear seasonality: the highest flows occur in April (somewhat less than 15 m^3/s), while the lowest flows occur in September and October. There are significant departures of 97.5th percentile form the expected values in the low-flow period, unlike remaining percentiles.

Intra-annual distributions of precipitation and temperature are presented in Figure 57 and Figure 58, respectively. There is distinct seasonality in temperatures: highest temperatures are observed in August, and lowest in January. The largest dispersion is observed during December and March. Highest precipitation depth are observed in June (~70 mm) and the lowest in January (~35 mm). Maximum values (97.5th percentile) substantially deviate from the expected ones, without revealing any clear pattern.

Precipitation and temperature gradients with elevation are given in Figure 59 and Figure 60, respectively. These gradients are estimated from the observations from five meteorological stations (Table 20). The precipitation or temperatures gradient with elevation are calculated only over the years with complete records. Decrease in mean annual temperature with elevation (from 0.47 to 0.68 °C / 100 m) is nearly constant in time. As for annual precipitation depths, no apparent pattern emerges. Namely, precipitation decrease with elevation is detected in some years. This may be attributed to topographic impact on precipitation – significant heterogeneity in precipitation depths, or poor data quality.

These gradients are important for assessment of prior values of corresponding parameters of the 3DNet-Catch model (α and T_{lapse}). As for prior ranges of the parameter α, only gradients with high values of the Pearson correlation coefficient are considered.

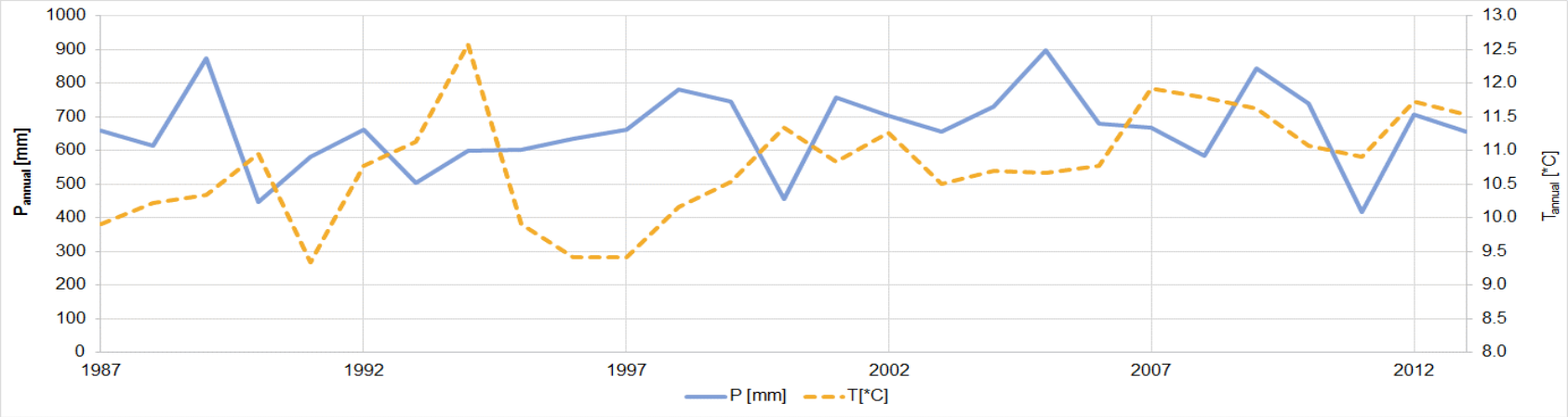


Figure 53. Annual precipitation and mean annual temperatures in the Mlava river catchment.

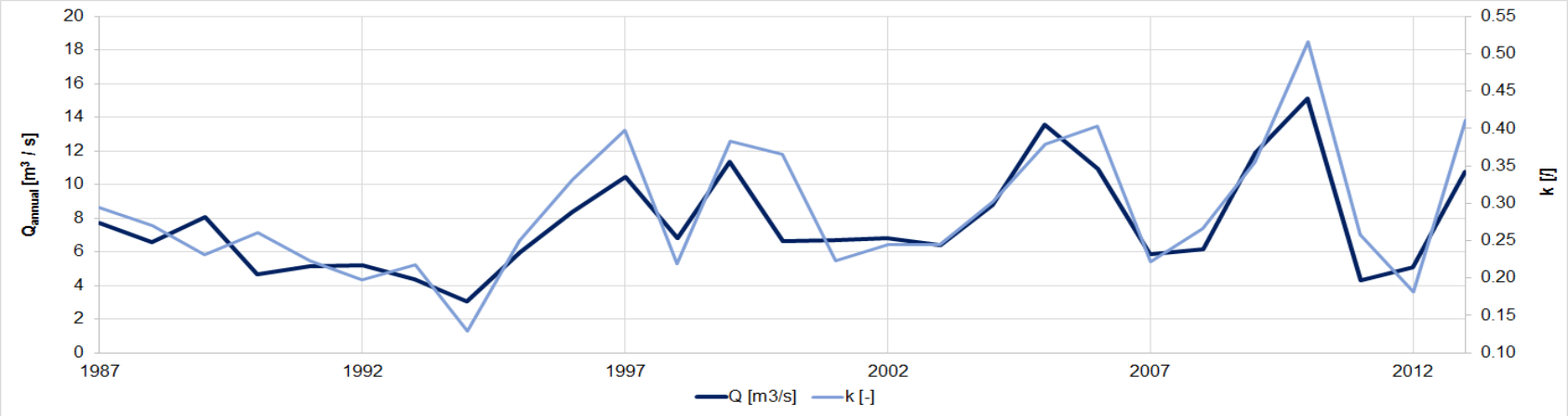


Figure 54. Mean annual flows and annual runoff coefficients at the Veliko Selo stream gauge.

Table 22. Trends in annual precipitation depths, mean annual temperatures and flows, and annual runoff coefficients at the Mlava river catchment (1987-2013).

VARIABLE	Pearson		Spearman	
	rho	*p* value	rho	*p* value
Precipitation [mm]	0.16	0.419	0.24	0.232
Temperature [°C]	0.51	$6 \cdot 10^{-3}$	0.57	$2 \cdot 10^{-3}$
Flow [m^3 /s]	0.38	0.054	0.27	0.179
Runoff coefficient [/]	0.37	0.06	0.31	0.12

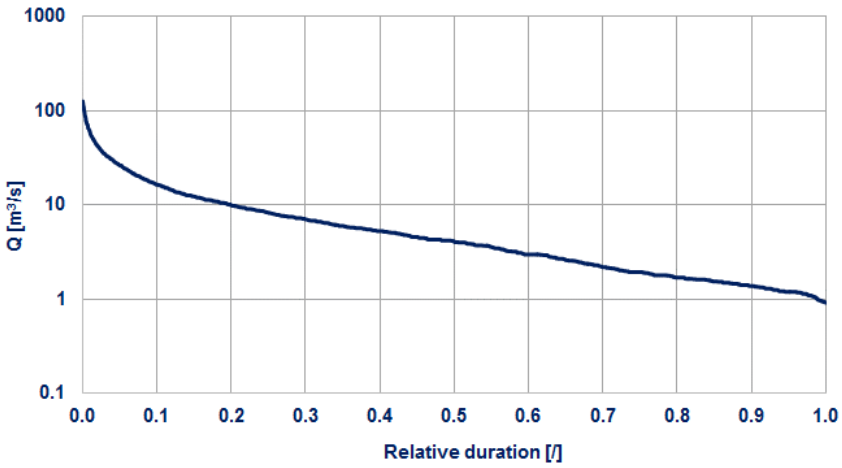


Figure 55. Flow duration curve derived from observed daily flows at the Veliko Selo stream gauge.

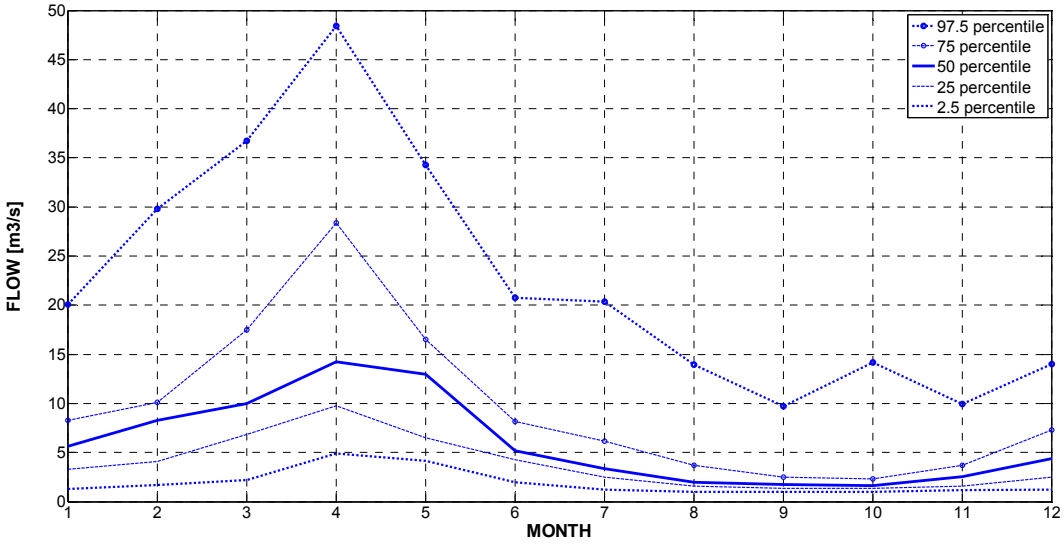


Figure 56. Intra-annual distribution of flows observed at the Veliko Selo stream gauge.

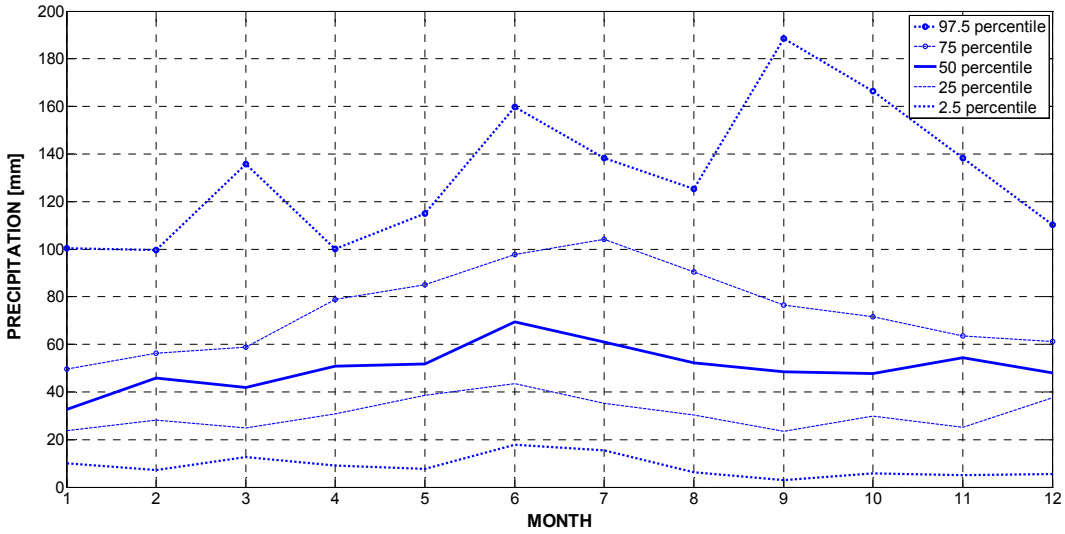


Figure 57. Total monthly precipitation depths in the Mlava River catchment.

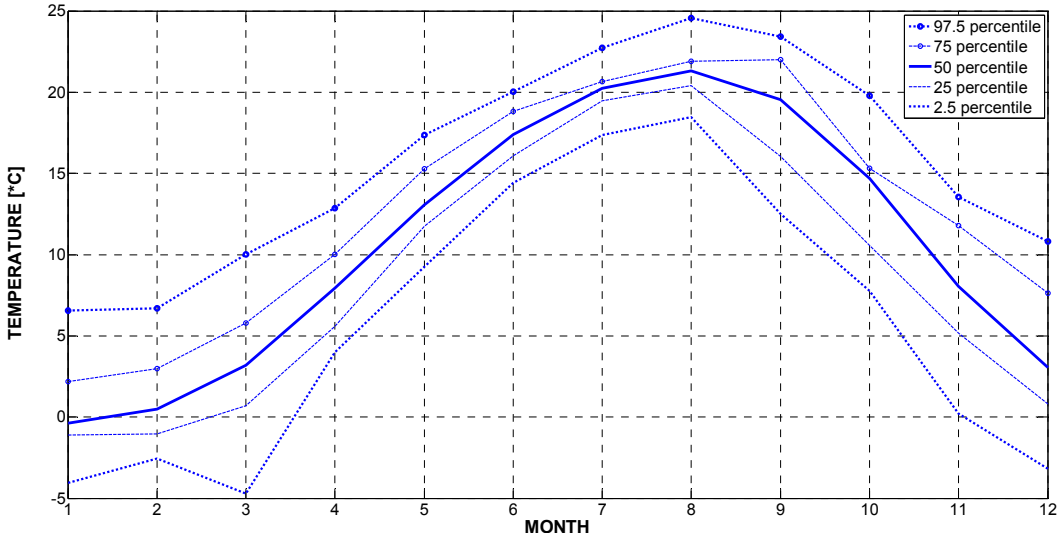


Figure 58. Mean monthly temperatures in the Mlava River catchment.

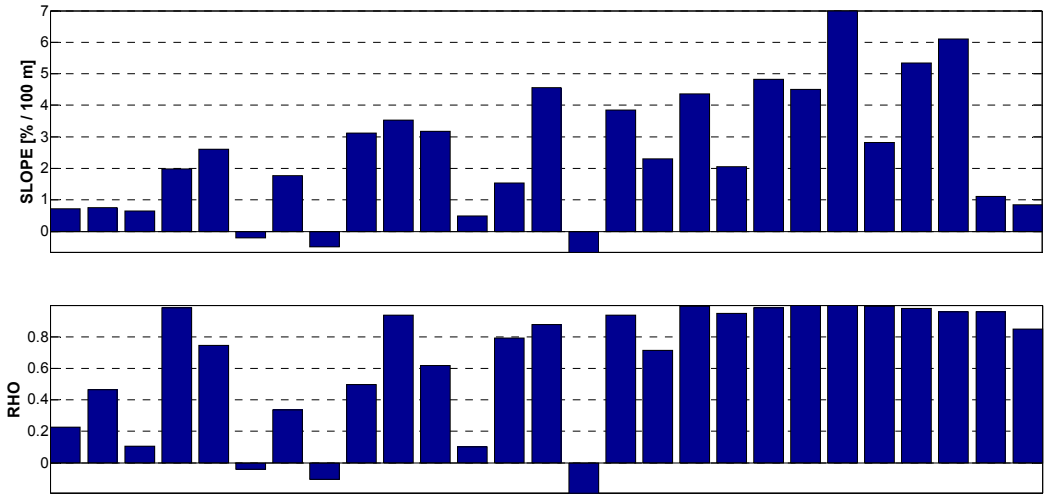


Figure 59. Precipitation gradients with elevation in the Mlava River catchment: slope of the liner regression and correlation coefficient in various water years.

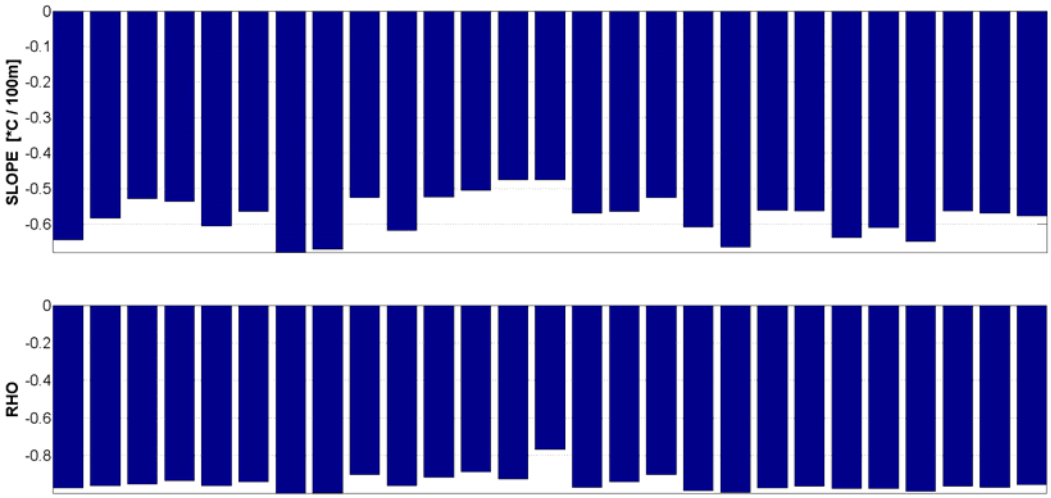


Figure 60. Temperature gradients with elevation in the Mlava River catchment: slope of the liner regression and correlation coefficient in various water years.

2.6.3. The Kolubara River catchment upstream of the Slovac stream gauge

Catchment description

The Kolubara River is a tributary of the Sava River. Entire catchment area amounts to 3639 km^2, but area upstream of the Slovac stream gauge is 995 km^2. Elevation ranges from 122 to 1331 m.a.s.l. (Figure 61) and mean elevation is 444.9 m.a.s.l. (Figure 62). Less than 1.5% of the catchment area is urbanised, while forests and agricultural land prevail (Figure 63). Parapodzol and parapodzoloic soils (hydrologic soil group C) are dominant in the catchment (Figure 64). Acid brown and podzolic soils (group B) and smonitza (group D, Djorković 1984) are present to a lesser extent. There is karst in the southern part of the catchment (Figure 65, Dimitrijević et al. 1975).

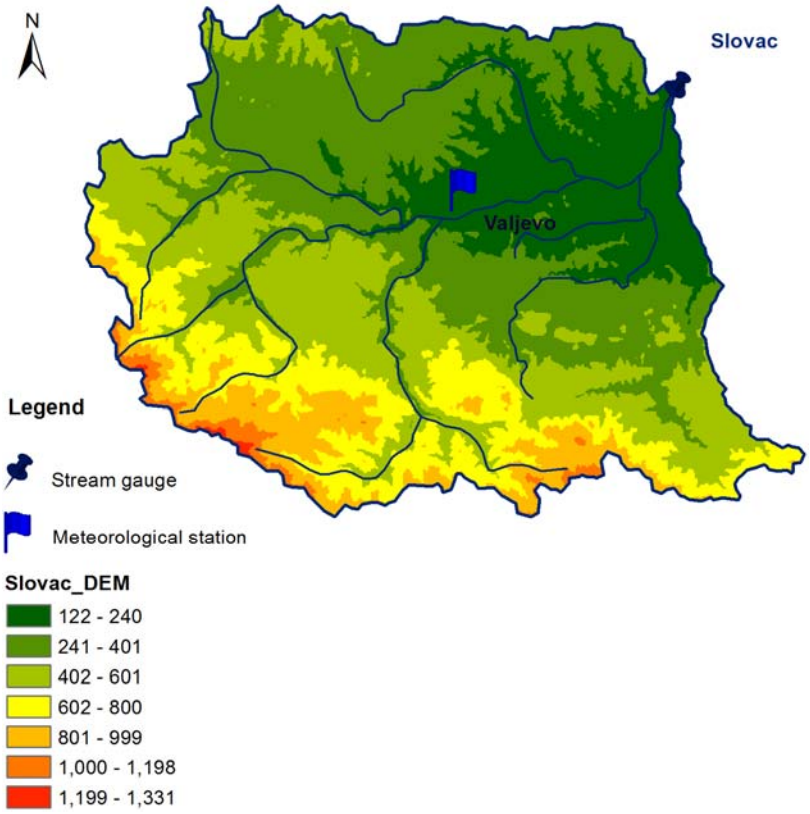


Figure 61. Topographic map of the Kolubara River catchment.

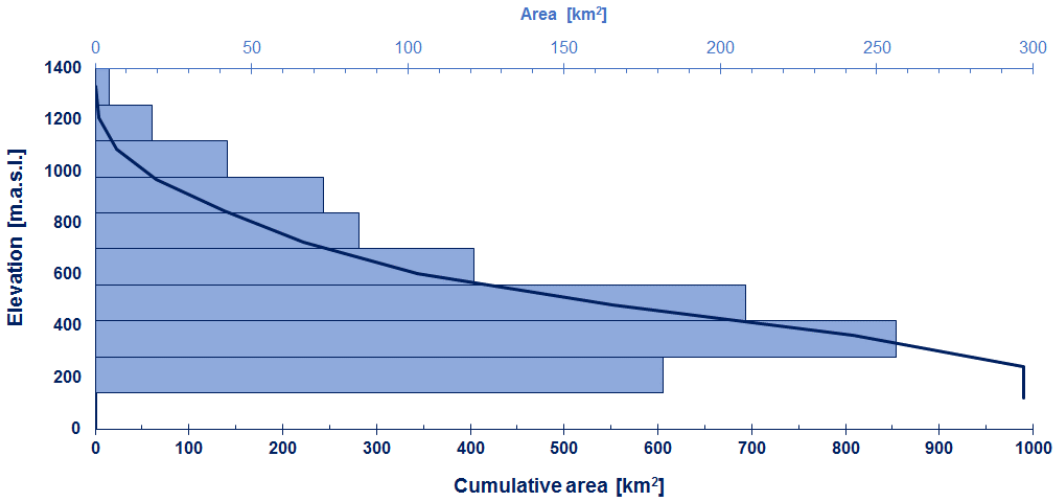


Figure 62. Hypsometric curve for the Kolubara River catchment.

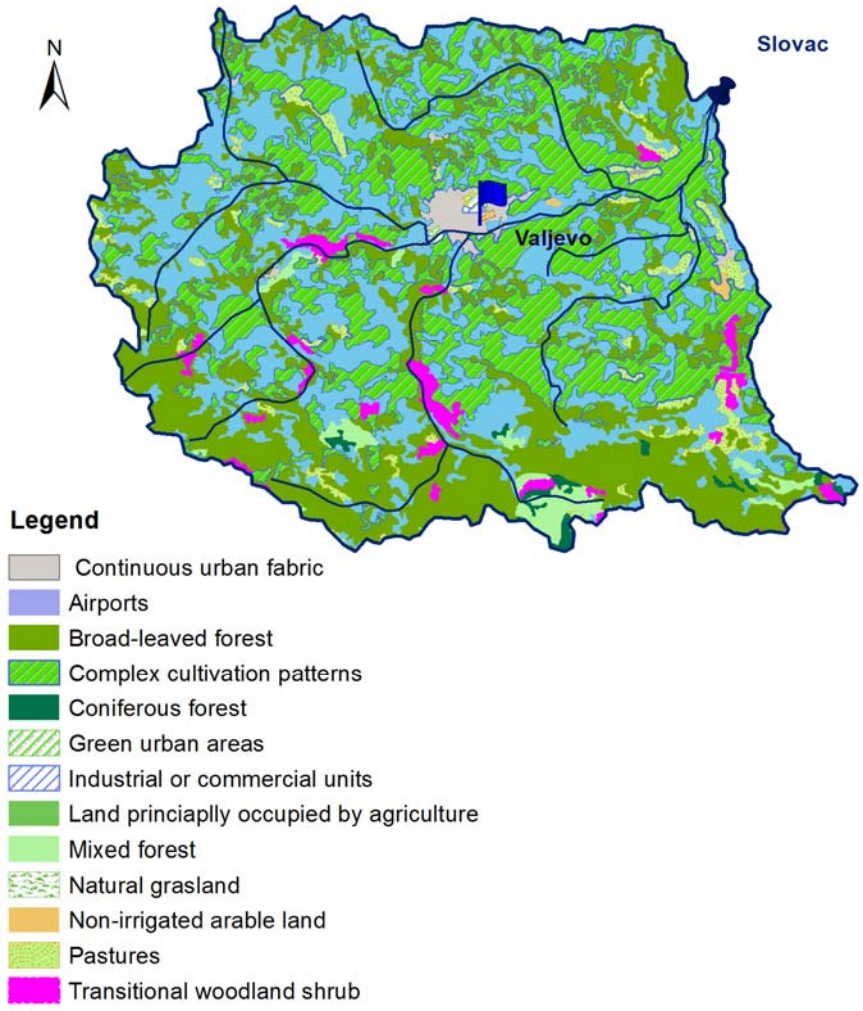


Figure 63. Land use types in the Kolubara River catchment (CORINE 2006).

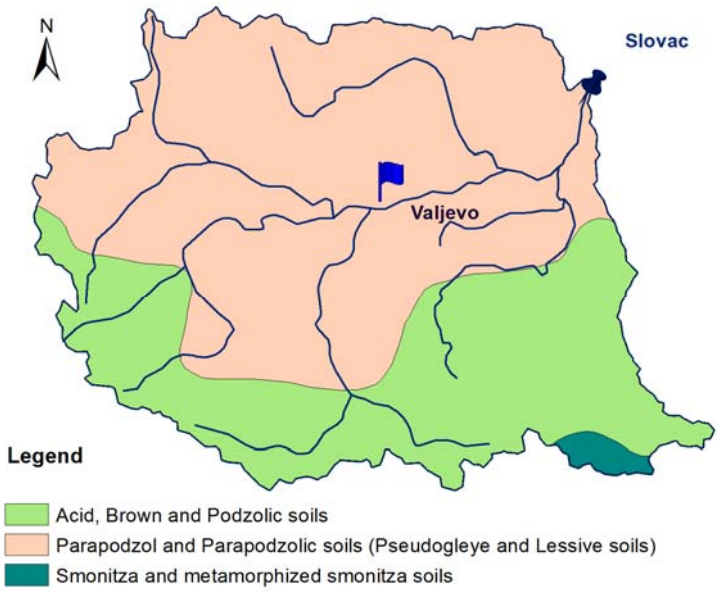


Figure 64. Soil types in the Kolubara River catchment.

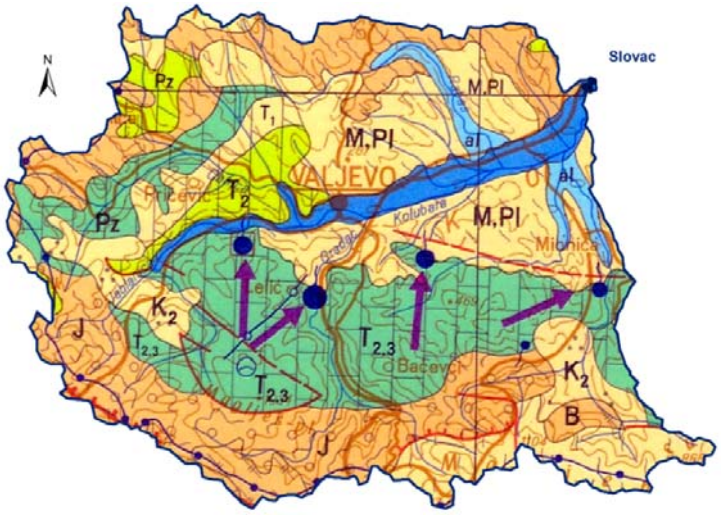


Figure 65. Geological structure data of the Kolubara River catchment: al – alluvium; J – serpentinites; K_2 – limestone with marls; M, Pl – marls, clays, bituminous clays and gravels; Pz – amphibolite, schists, granite-gneiss, phyllite-ricaschists and calcschists; T1 – limestone, quartz-conglomerate, quartz-sandstone and quartzite; $T_{2, 3}$ – limestone (upper and middle Triassic).

Data

Only observations from the Valjevo meteorological station are used for runoff simulations at this catchment (Table 23). There are few gaps in daily precipitation time series (0.7% of the observations, from 1988 to 1990). Missing values are assessed using linear regression and data from Valjevo RC rain gauge (correlation coefficient of 0.85). In addition to considerable length of the record period (60 water years), flow observations are considered of good quality. Namely, Slovac cross section has been relatively stable for decades (small fluctuations in the cross-section geometry due to riverbed erosion or deposition), resulting in reliable rating curve.

Table 23. Hydro-meteorological stations available for runoff simulation at the Kolubara River catchment (RHMSS).

Station	Variable	Elevation [m.a.s.l.]	Latitude	Longitude	Available data form	Mean observed value (1954-2013)
Slovac	Q	121.59	44 º 22 ’	20 º 13 ’	1954	9.8 [m^3/s]
Valjevo	P, T	176	44 º 17 ’	19 º 55 ’	1951	790.1 [mm] 11.3 [°C]

Hydrometeorological regime

Annual precipitation depths and mean annual temperatures are presented in Figure 66. Mean annual flows and estimated annual runoff coefficients are given in Figure 67.

These series are tested for trend. The *p*-values of the regression slopes coefficients (Table 24) suggest presence of statistically significant increasing trend in temperature at the 95% significance level (shaded cells in the table) and absence of the statistically significant trend in precipitation and flows.

Mean annual flow at the Slovac stream gauge is 9.8 m^3/s (Table 23), i.e. specific water yield from the catchment amounts to 9.87 L s^{-1} km^{-2}. According to the flow duration curve (Figure 68), median observed daily flow is 5.7 m^3/s. Intra-annual distribution of flows (Figure 69) exhibits seasonality in all percentiles considered. High flows are observed from February to April (~13 m^3/s) and low flows from August to October (~3 m^3/s). Annual runoff coefficient varies from 0.24 to 0.66, with mean value of 0.39.

Intra-annual distributions of monthly precipitation depths and mean monthly temperature are presented in Figure 70 and Figure 71, respectively. Highest precipitation depths and deviations from expected values are observed in the summer (e.g., mean precipitation depth in June amounts to ~80 mm) and lowest precipitation are observed in February (~45 mm). Mean monthly temperatures exhibit pronounced seasonality: maximum temperatures are observed in August (~23 C°) and the lowest in January (~1C°). Largest departures form mean values (up to 8°C) are observed in winter.

Precipitation gradient with elevation is estimated using data from four additions rain gauges over 19 years (1969-1974; 1979-1984; 2005-2011). These rain gauges are: Koceljeva (130 m.a.s.l.), Ljig (150 m.a.s.l.), Breždje (340 m.a.s.l.) and Majinović (400 m.a.s.l.) (Todorović and Plavšić, 2014). The gradient varies substantially (Figure 72), which may indicate significant spatial heterogeneity of precipitation depths or poor quality of the precipitation data. Only gradient values with high positive correlation coefficient are taken into account to estimate prior ranges of parameter α. As temperature data from other meteorological stations have not been available, initial range for T_{lapse} is adopted based on the temperature gradients estimated for other two catchments.

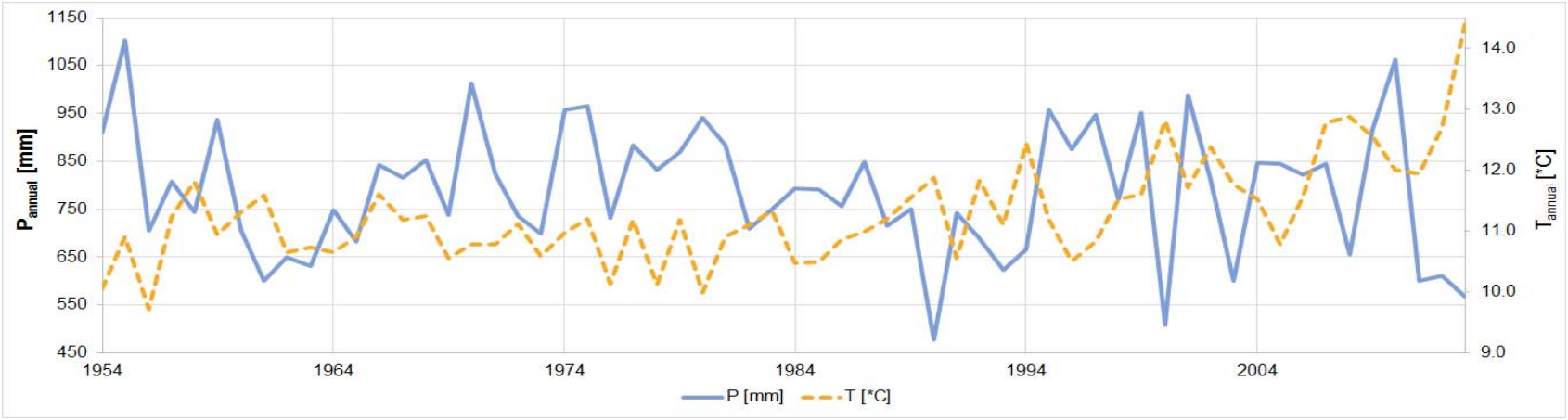


Figure 66. Annual precipitation depths and mean annual temperatures in the Kolubara river catchment.

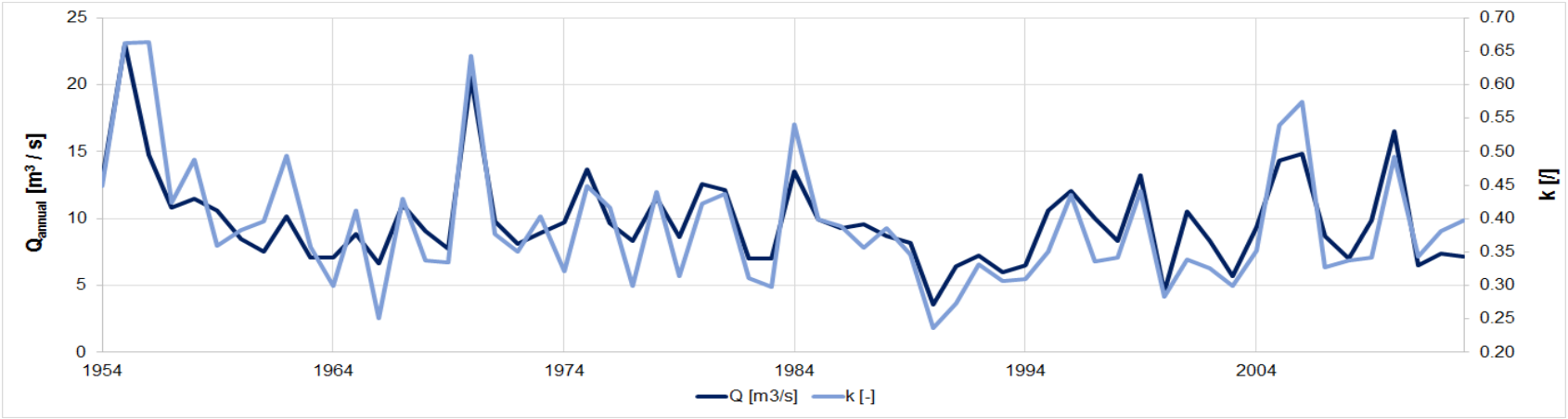


Figure 67. Mean annual flows and annual runoff coefficients at the Slovac stream gauge.

Table 24. Trends in annual precipitation depths, mean annual temperatures and flows, and annual runoff coefficients in the Kolubara river catchment (1954-2013).

VARIABLE	Pearson		Spearman	
	rho	*p* value	rho	*p* value
Precipitation [mm]	-0.11	0.42	-0.07	0.62
Temperature [◦C]	0.61	$2.6 \cdot 10^{-7}$	0.55	$4.5 \cdot 10^{-6}$
Flow [m^3 /s]	-0.22	0.09	-0.22	0.09
Runoff coefficient [/]	-0.26	0.04	-0.25	0.05

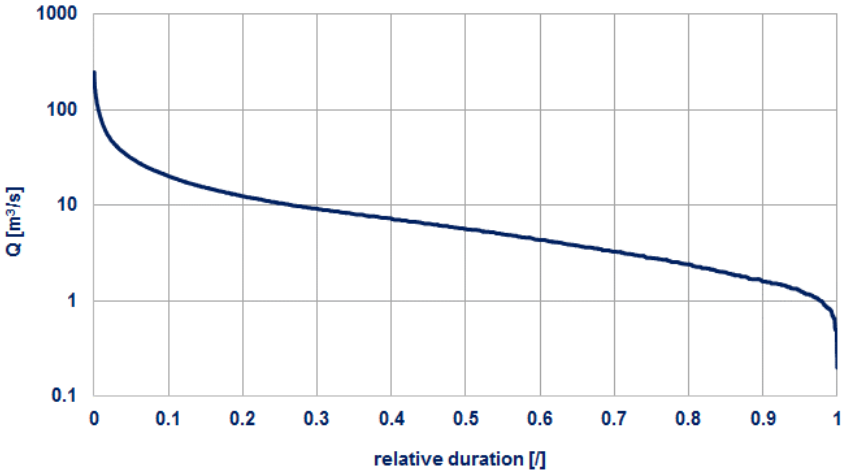


Figure 68. Flow duration curve derived from observed flows at the Slovac stream gauge.

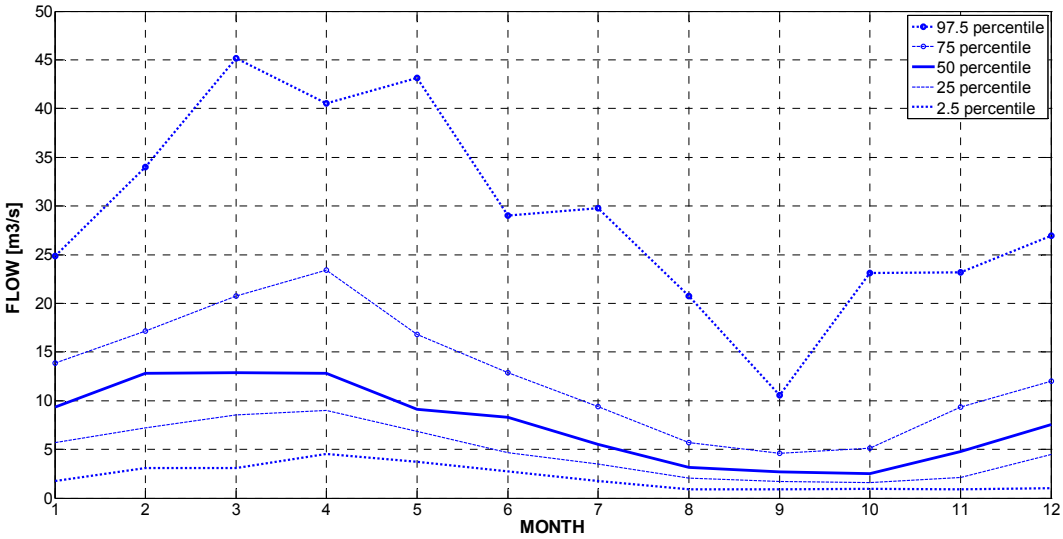


Figure 69. Intra-annual distribution of flows observed at the Slovac stream gauge.

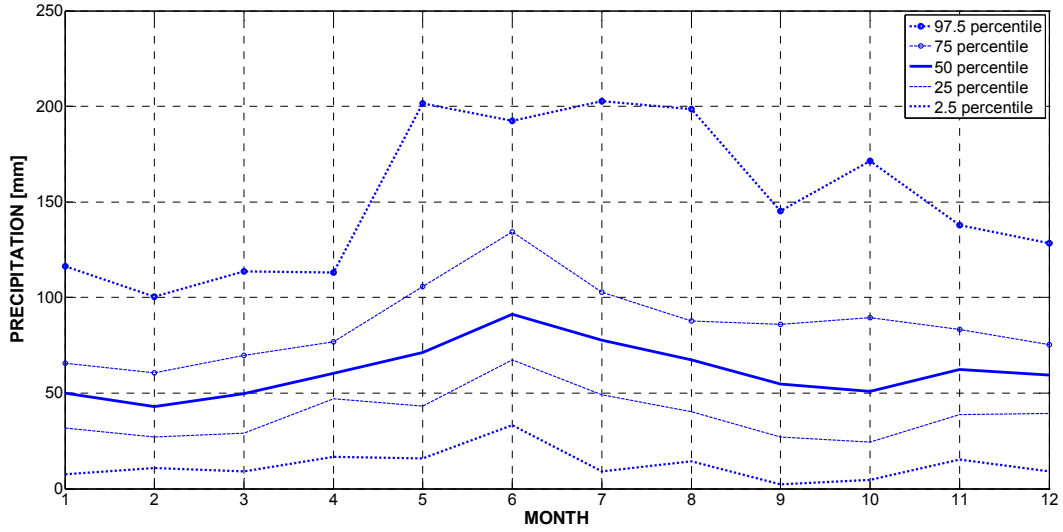


Figure 70. Intra-annual distribution of precipitation at the Valjevo meteorological station.

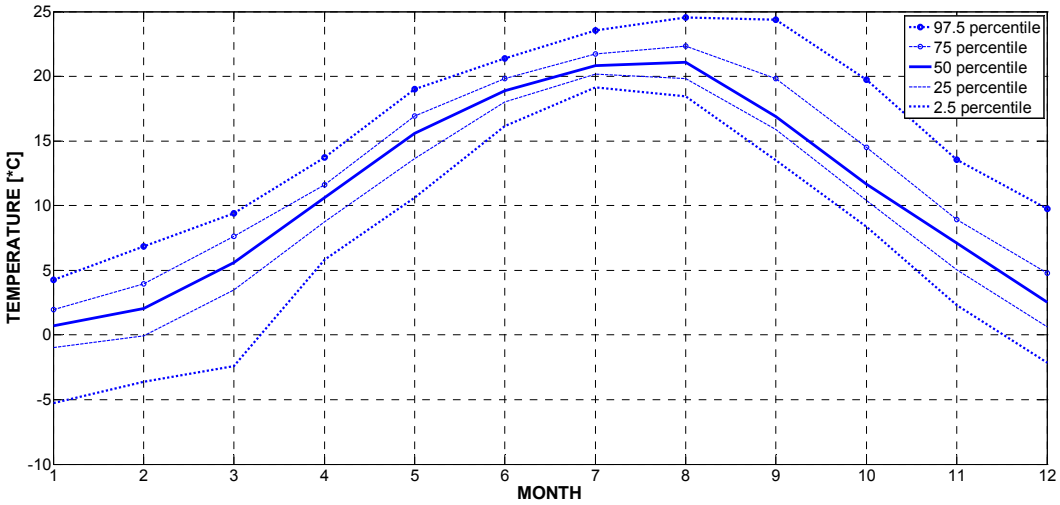


Figure 71. Intra-annual distribution of precipitation at the Valjevo meteorological station.

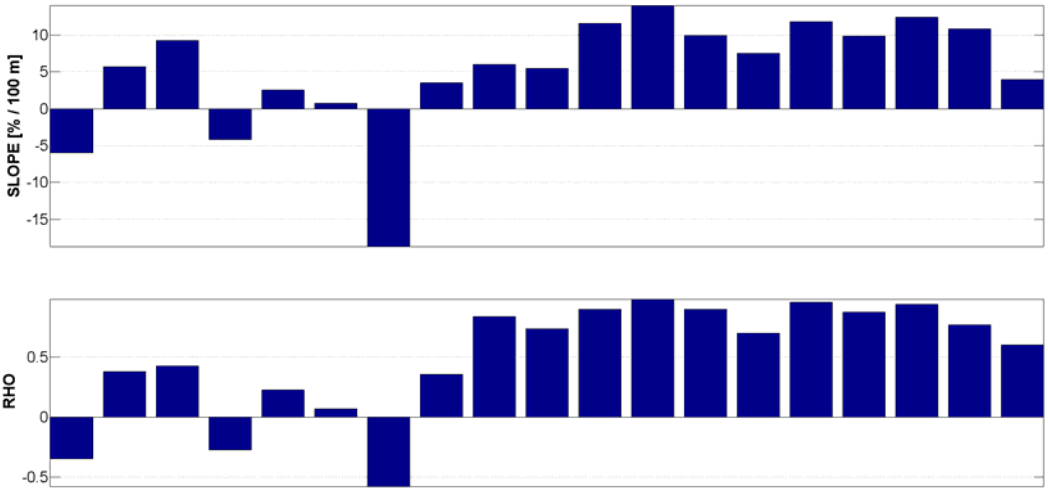


Figure 72. Precipitation gradients with elevation in the Kolubara River catchment: slope of the linear regression and the correlation coefficient in various water years.

The catchments considered are shown to be unaltered in terms of land use types. Since no river training measures have been imposed to the watercourses flow observations are homogenous, what is confirmed by the linear slope test. However, increasing trend in temperatures is detected for all three catchments.

3. RESULTS AND DISCUSSION

3.1. Application of the 3DNet-Catch hydrologic model

3.1.1. Sensitivity analysis and correlation among parameters

In this research, parameter sensitivity is estimated by employing regression based sensitivity analysis, which is based on the multiple regression between the parameters and two objective function values (chapter 2.3.3), namely Nash-Sutcliffe efficiency, *NSE* and the volume error, *VE*. However, validity of the linear metamodel(s) has to be confirmed prior to application for the sensitivity analysis. Goodness-of-fit of the metamodel is quantified in terms of coefficient of determination, while maximum variance inflation factors (VIF_{MAX}) are estimated to test the metamodel for multicollinearity. According to Christiaens and Feyen (2002), if VIF_{MAX} exceeds 10, linear regression metamodel should be discarded. As for determination coefficient, Pan et al. (2011) do not recommend application of the linear metamodel if R^2 is less than 0.7.

Values of these measures are given in Table 25. All metamodels based on *VE* meet criteria for both measures, while *NSE*-related metamodels result in a small R^2 value. Despite relatively low R^2 values for the *NSE*-based metamodels, they are nevertheless kept in the analysis for two reasons: firstly, *NSE* depends on model's ability to reproduce dynamics of the catchment response, and it is crucial to detect parameters in control of this aspect of the model; secondly, the aim of this sensitivity analysis is to identify the most sensitive parameters, and not to accurately estimate the standardised regression coefficients (*SRC*s) *per se*. Thus, the *NSE*-based metamodel is considered eligible for purposes of such sensitivity analysis.

Table 25. Coefficients of determination and maximum variance inflation factors for regression between the model performance measures and the parameters

Measure	CATCHMENT		
	Kolubara	Toplica	Mlava
R^2 – *NSE*	0.34	0.56	0.48
R^2 – *VE*	0.7	0.89	0.82
VIF_{MAX}	1.26	1	1.46

In addition to the validity of the linear metamodels, correlation among the model parameters has to be examined prior to the sensitivity analysis. According to Christiaens and Feyen (2002), if the strong correlation among model parameters is present, *SRC*s obtained for different model parameters cannot be mutually compared. Pearson and Spearman correlation coefficients among model parameters (one hundred best parameter sets out of 25.000 sampled ones, *NSE* values range from 0.41 to 0.62) are given in Figure 73. These graphs demonstrate a lack of especially linear correlation) among one hundred behavioural parameters. Spearman correlation coefficients tend to be somewhat larger than the Pearson ones, although they also take rather small values. The results of the sensitivity analysis may therefore be considered reliable.

The results of the sensitivity analysis, i.e. *SRC*s with respect to *NSE* and *VE*, are presented in Figure 74 and Figure 75, respectively. Their values are given in the Appendix C along with the correlation coefficients (*LCC*). According to Christiaens and Feyen (2002) ratios between *LCC* and *SRC* that are approximately equal to 1 indicate absence of strong linear correlations among model parameter. As shown in Appendix C, these ratios are approximately one for majority of the model parameters. The results of the sensitivity analysis obtained for three catchments are quite consistent and they suggest the following:

– Parameter α is indicated as the most sensitive parameter, and soil porosity, thickness and hydraulic conductivity of the subsurface soil layer ($D_{\text{sub-surf}}$ and $K_{\text{sub-surf}}$) and maximum Leaf Area Index ($LAI_{\max}$) are indicated as highly sensitive with respect to both objective functions.

– Other soil-related parameters, such as thickness and hydraulic conductivity of the surface soil layer, wetnesses at permanent wilting point and at field capacity, and pore-size distribution index are also shown to be sensitive with respect to both objective functions but to lesser extent compared to the above.

– Another parameter of the interception routine $CAN_{\max}$ is proven to be relatively insensitive with respect to both objective functions.

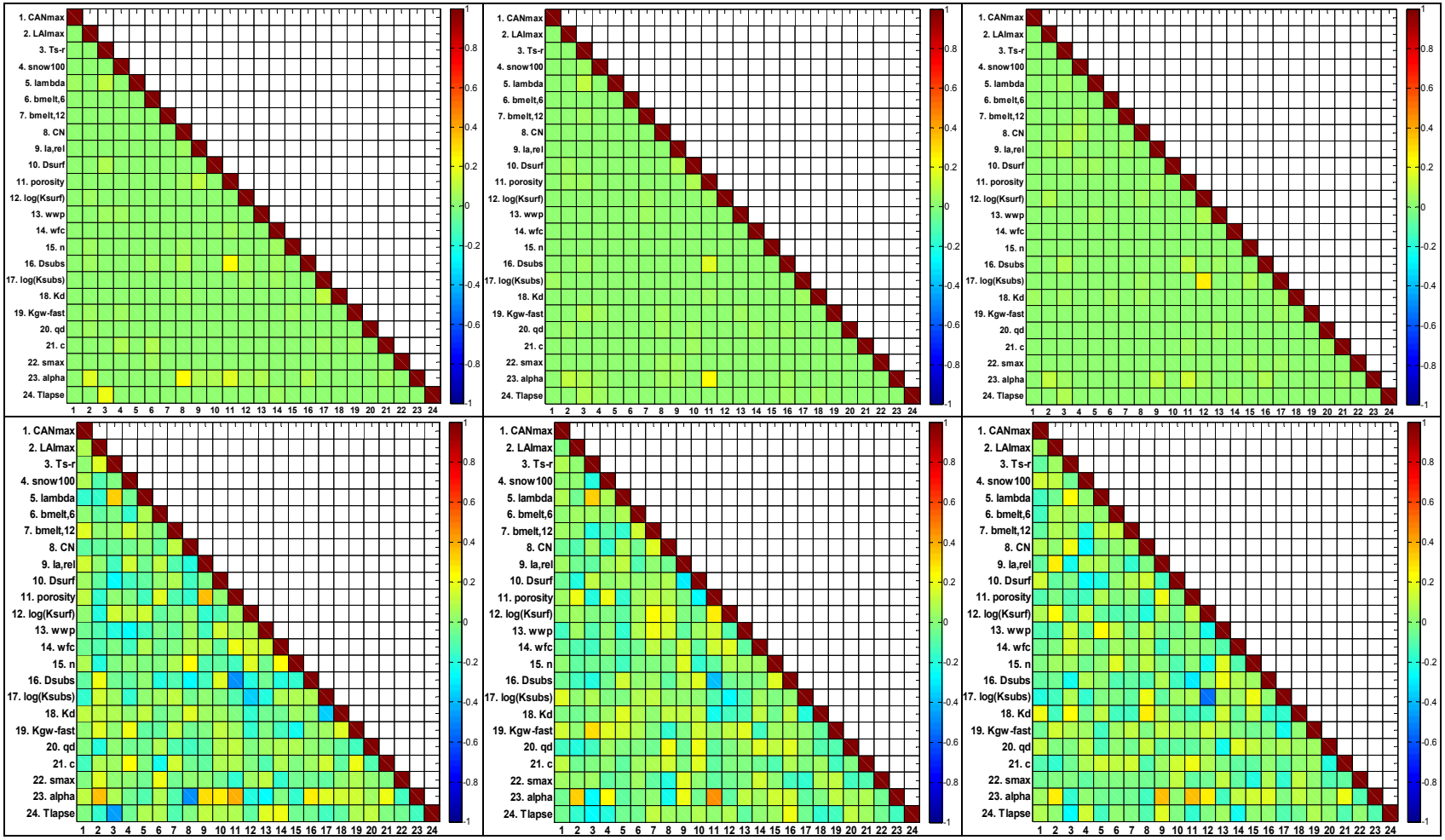


Figure 73. Pearson (upper panels) and Spearman correlation coefficients (lower panels) among 100 behavioural parameter sets for the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).

– The linear reservoirs coefficients (especially K_d) are indicated as fairly sensitive with respect to *NSE* and insensitive with respect to *VE*. Having in mind that the reservoir coefficients are in control of hydrologic response dynamics rather than the overall water balance, such outcome of the sensitivity analysis is expected. The SCS curve number (*CN*) displays some sensitivity with respect to *NSE*. The parameter *CN* determines the effective precipitation and, consequently, amount of direct runoff, thus implicitly affecting dynamics of catchment response. However, the sensitivity of this parameter is considerably smaller than the sensitivity of the reservoir coefficients.

– Parameters of the snow routine are shown to be insensitive. However, parameter sensitivity varies in time (e.g. Sieber & Uhlenbrook, 2005; Muleta, 2012) so these results should be interpreted with caution. Snow-related parameters can exhibit higher sensitivity in the snowmelt season, as shown in Figure 76.

– Other parameters, such as initial abstraction $I_{a,rel}$, or baseflow-related parameters of the response routine (q_d, c, s_{max}) exhibit negligible sensitivity. The lack of sensitivity in the latter parameters with respect to *VE* may be attributed to the fact that these parameters are not in control of water balance, but dynamics of the baseflow. As for *NSE*, it is primarily determined by agreement between simulated and observed high flows. However, sensitivity of these parameters with respect to another objective function that is less biased to high flows may be higher, since the results of the sensitivity analysis heavily depend on the objective function against which the sensitivity is estimated. For example, Figure 77 presents *SRC*s estimated according to *NSE* values obtained from the log-transformed flows ($NSE_{\log Q}$). The results reveal more uniform *SRC* values, implying that these parameters may not be insensitive when it comes to simulations of low flows. However, validity of the $NSE_{\log Q}$ linear metamodel is questionable (R^2=0.002), so these results are not taken into account in this research.

Despite the overall consistency, the results are somewhat catchments specific. For example, correlation of parameters λ and $b_{melt,\,6}$ with *NSE* is statistically significant only for the Kolubara River catchment, or correlation between $I_{a,rel}$ and both objective functions is significant only for the Mlava River catchment. These discrepancies are

negligible, because the sensitivity of those parameters is generally rather small for all catchments.

Generally, the results of the sensitivity analysis clearly indicate five parameters as the most sensitive: namely, α, $D_{\text{sub-surface}}$, *porosity*, K_{d} and LAI_{max} (e.g. these five parameters explain 51% of variations in *NSE* at the Toplica River catchment in 1988-2013). The results of the sensitivity analysis are supported by the hydrographs presented in Appendix B, which are obtained using the minimum and maximum values of the plausible parameter ranges. Sensitive parameters substantially affect hydrographs, while the impact of the insensitive parameters is marginal.

Correlation among parameters that are sampled for purposes of the sensitivity analysis is has proven to be weak; however, these results should be interpreted with caution. Namely, the parameter sets, which the sensitivity analysis is based on, are randomly sampled, and therefore scattered all over the parameter hyper-space. However, when it comes to the optimised sets (including the Pareto-optimal parameters), they converge to a narrow "basin" of the response surface. Due to specific properties of such regions (e.g. Duan et al., 1992), these parameters may be expected to exhibit stronger correlations (Vrugt et al., 2006). For example, correlation coefficients among optimised model parameters (100 sampled parameter sets, 20.000 generation, with *NSE* and *VE* objective functions) obtained over the same period (1988-2013) are presented in Figure 78. These results indicate much stronger correlations for the Kolubara River catchment, while the correlation among parameters which refer to the Toplica River catchment remained weak. Some of these correlation coefficients exceed 0.75, indicating strong correlation (Blasone et al., 2007). High correlation coefficients may be found even among the most sensitive model parameters. These results suggest that the analysis of the Pareto-optimal parameter has to be conducted bearing in mind possible interactions among them.

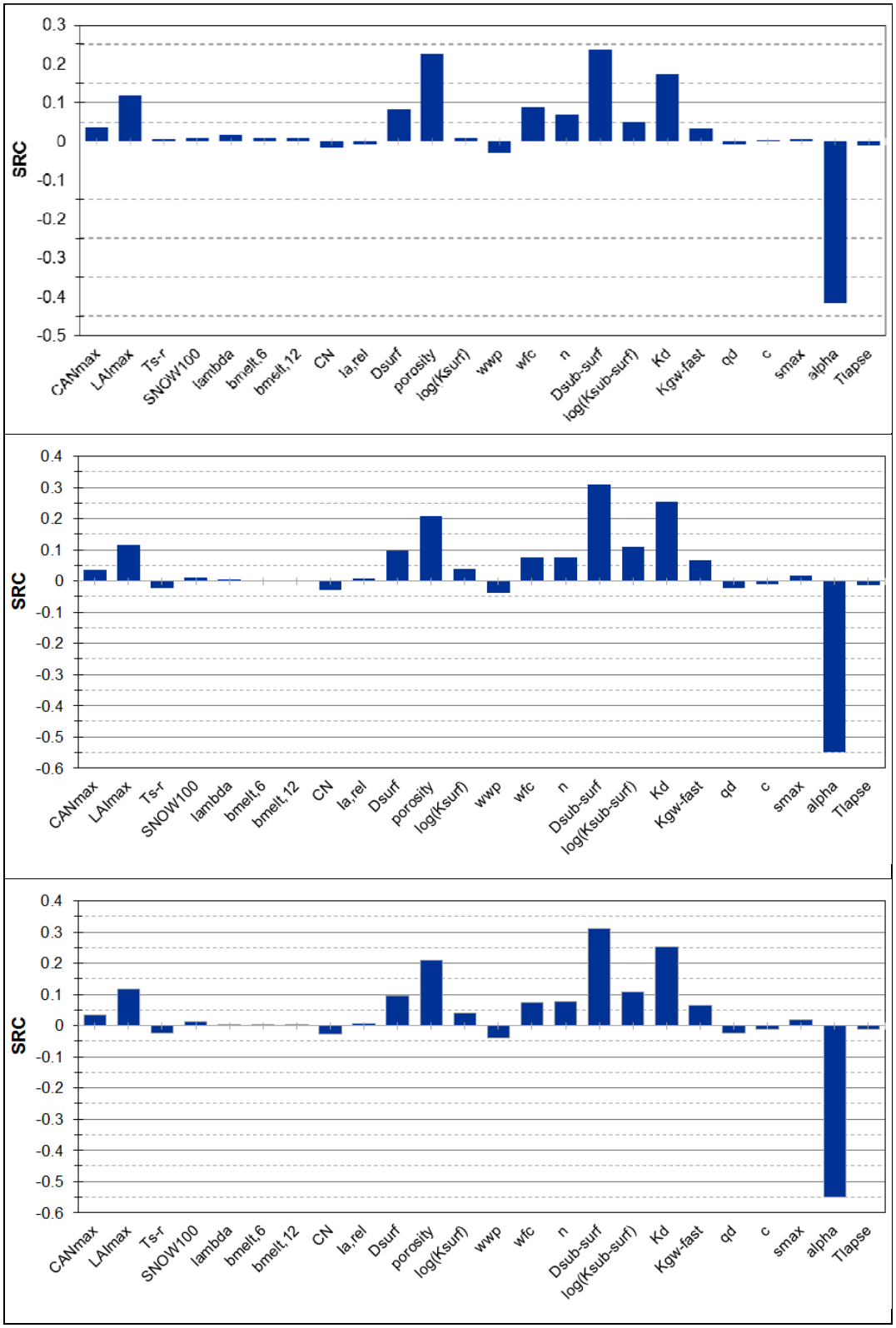


Figure 74. *SRC* in the regression between the model parameters and *NSE*: the Kolubara (upper panel), Toplica (mid panel) and Mlava (lower panel) River catchments.

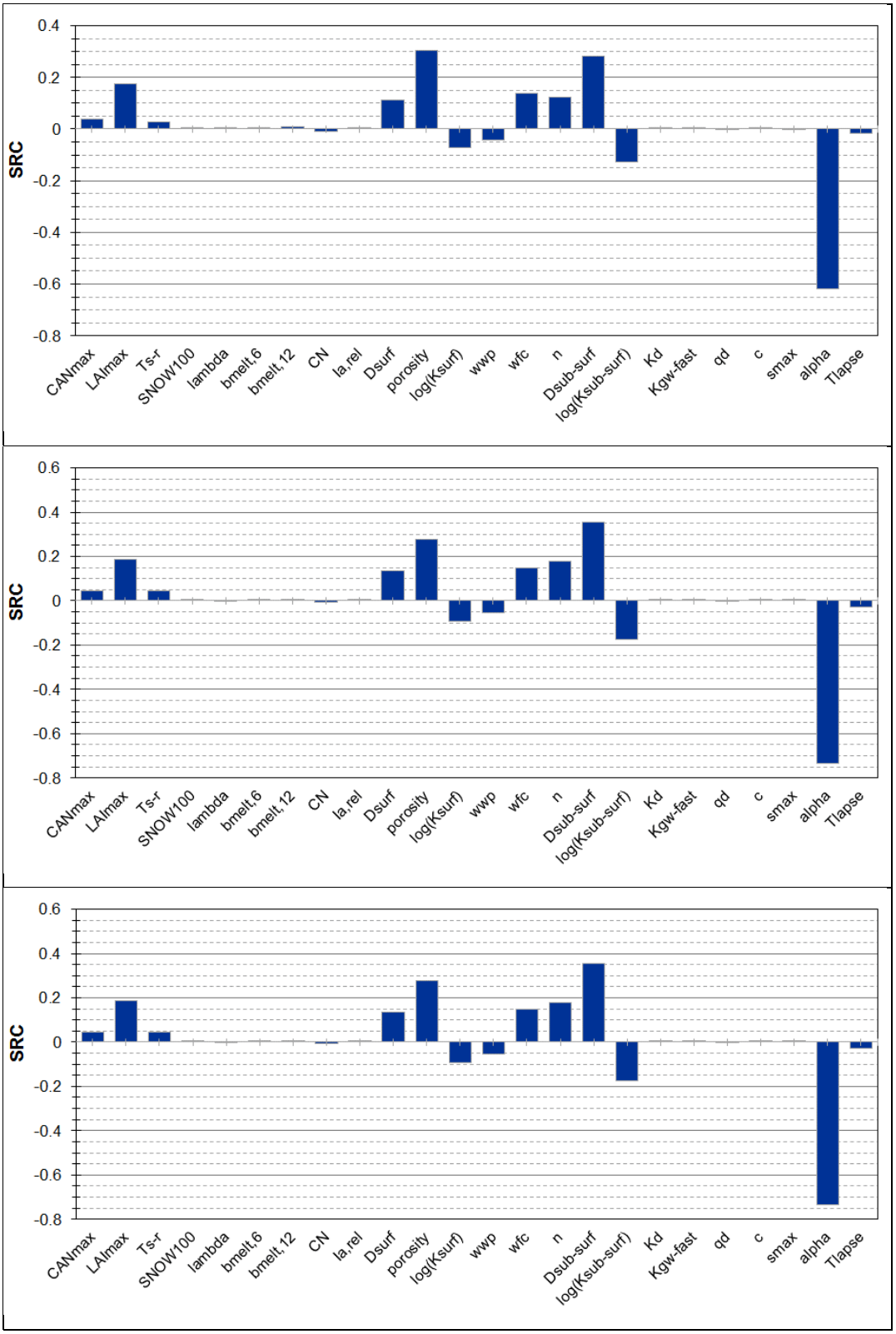


Figure 75. *SRC* in the regression between the model parameters and *VE*: the Kolubara (upper panel), Toplica (mid panel) and Mlava (lower panel) River catchments.

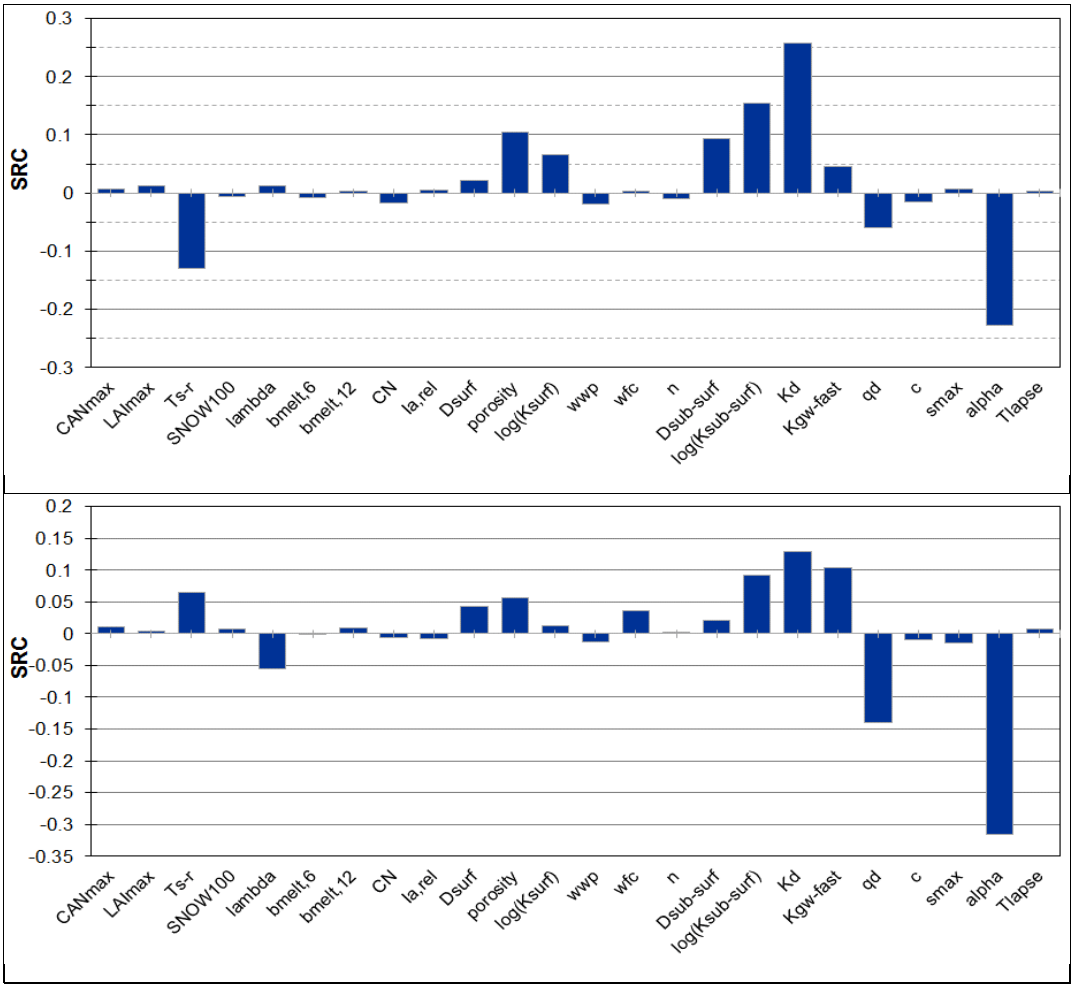


Figure 76. *SRC* in the regression between model parameters and *NSE* (top panel) and *VE* (bottom panel) from 1st December 1962 to 31st March 1963 for the Kolubara catchment.

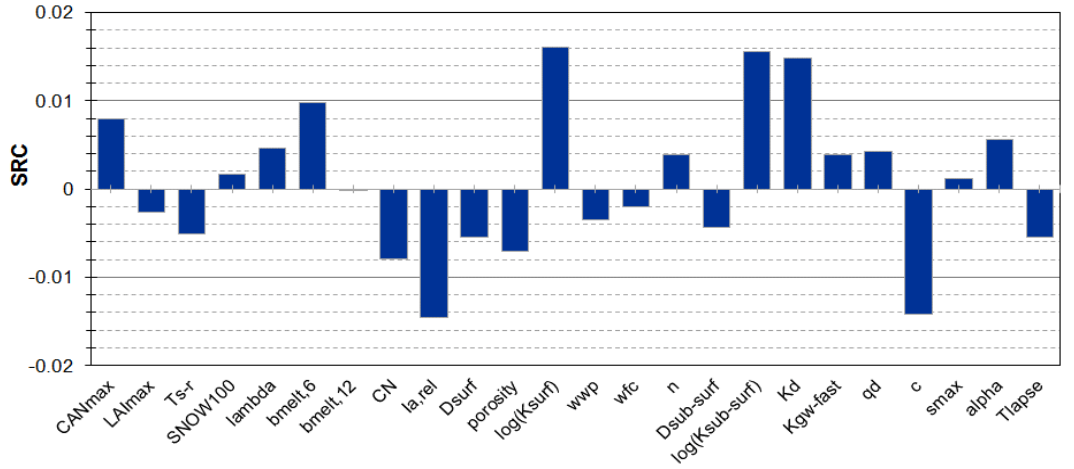


Figure 77. *SRC* in the regression between the model parameters and *NSE* based on the log-transformed flows for the Mlava River catchment.

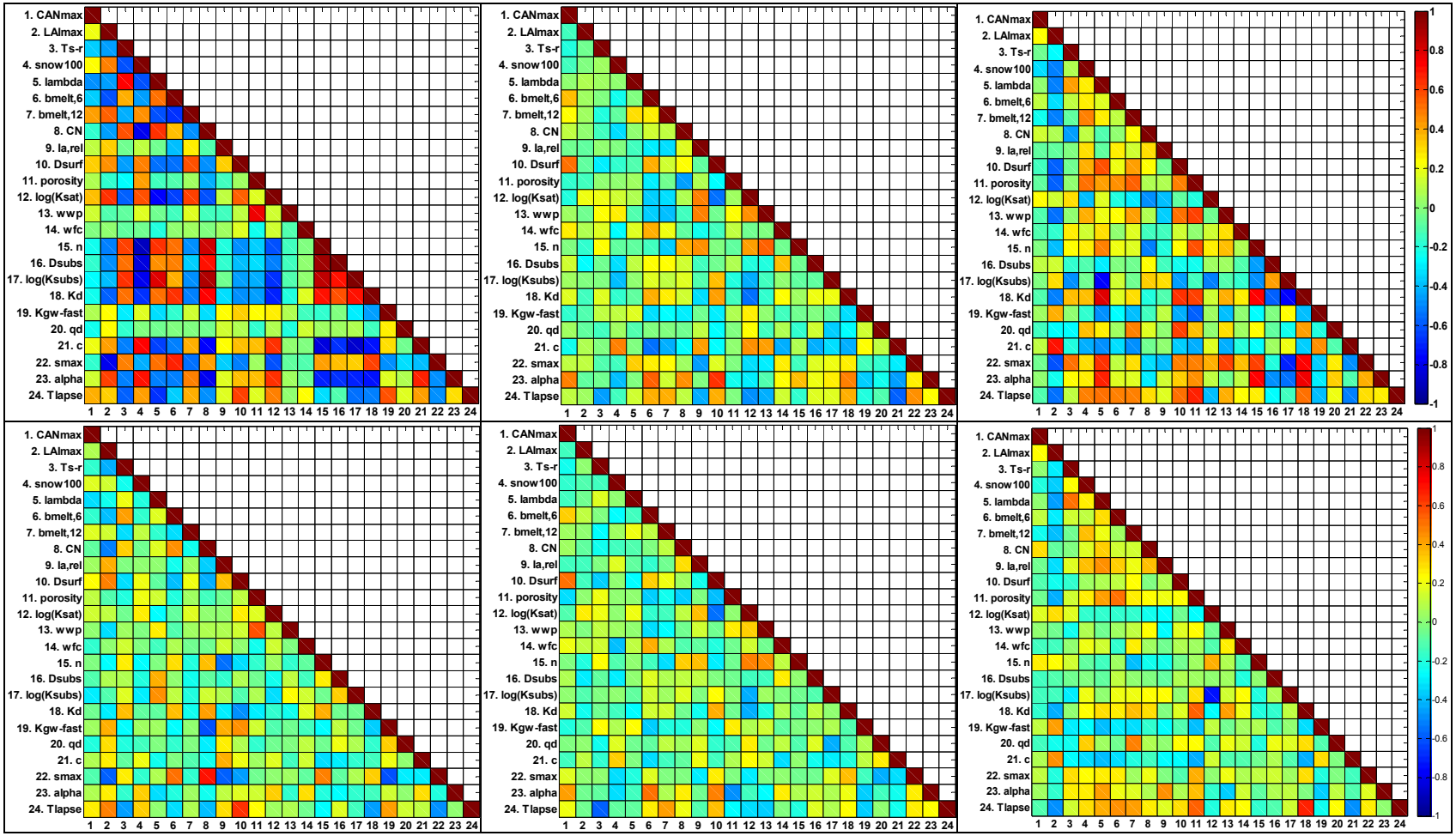


Figure 78. Pearson (upper panels) and Spearman correlation coefficients (lower panels) among the optimised model parameters for the Kolubara (left panels), Toplica (mid panels) and Mlava River (right panels).

3.1.2. Evaluation of the 3DNet-Catch model performance

Split sample (*SST*) and Differential Split Sample tests (*DSST*) are used to estimate robustness of semi-lumped BASIC version of the model. The model is calibrated in multi-objective manner and evaluated as described in chapter 2.3.4. The Pareto fronts obtained in the calibration periods and the model performance in the evaluation ones are presented in Figure 79.

The results demonstrate ability of the model to entirely reproduce runoff volume over the calibration periods. However, *NSE* values differ depending on the calibration period (average – *SST*, or wet period – *DSST*): for the Kolubara River catchment *NSE* values are higher over the average period, for the Toplica River catchment it is other way round, whereas for the Mlava River catchment *NSE* values are approximately the same.

Model performance deteriorated over the evaluation periods, except for the Toplica River catchment: namely, model efficiency over the evaluation period is higher than in the calibration one (in the *SST*). Drop in model performance is more pronounced in the results of *DSST*, except for the Kolubara River catchment (negligible decrease in *NSE* values). Largest decrease in *NSE* values and the smallest decrease in *VE* at the same time is observed for the Mlava river catchment (drop in *NSE* of 0.6 and 5.1% in bias). For the Kolubara and Toplica River catchments *NSE* values in evaluation periods (*DSST*) remain acceptable according to Moriasi et al. (2007) (exceed 0.35 on average), but there is a marked bias in the simulated runoff volume (~25%). Simulated hydrographs in the evaluation periods (*SST* for the Toplica, and *DSST* for the Mlava River catchment) are presented in Figure 80. The decrease in model performance generally corroborates the results of Vaze et al. (2010) (i.e. decrease of 0.3 in *NSE* values and up to 40% in bias for 30% smaller rainfall depths).

The BASIC version of the model is calibrated in 1996-2009 and evaluated over 1985-1996 against the observed flows at the Slovac and Doljevac stream gauges (chapter 2.3.4) and the results are compared to those obtained using the HBV model. Maximum, mean and minimum values of the performance measures calculated with 100 optimised parameter sets are presented in Table 26 along with the results of the HBV model reported

by Langholt et al. (2013), which are obtained by Norwegian Resources and Energy Directorate (NVE) and Republic Hydrometeorological Service of Serbia (RHMSS). Therefore every result in Table 26 presented by Langholt et al. (2013) is represented by two values: the first one is obtained by NVE, and the second one is obtained by RHMSS.

Performance of the 3DNet-Catch in the calibration period is similar for both catchments. In the evaluation period for the Kolubara River catchment, the 3DNet-Catch yields considerably higher *NSE* values, but larger bias in runoff volume. For the Toplica River catchment the HBV model resulted in higher *NSE* values, but slightly larger bias in estimated runoff volume. In addition, rather high values of the *NSE* for log-transformed flows obtained by the 3DNet-Catch model at the Toplica River catchment should be emphasised: $NSE_{\log Q}$ is 0.79 in the calibration period, and exceeds 0.7 in the evaluation period. On the whole, these results suggest that the semi-lumped version of the 3DNet-Catch is comparable to the HBV model.

The flow duration curves (observed and simulated) over the calibration period are presented in Figure 81. Flow duration curves (FDCs) show that the model reproduced the Toplica River catchment behaviour satisfactorily, while there are some discrepancies between FDCs of the simulated and observed flows at the Slovac stream gauge, which are particularly pronounced in the low flow domain.

In these tests the semi-lumped version of the model is analysed, because that version of the model is mainly used in this research. The fully-distributed version of the model is to be tested in the further research, but it is expected that it would yield higher model efficiency.

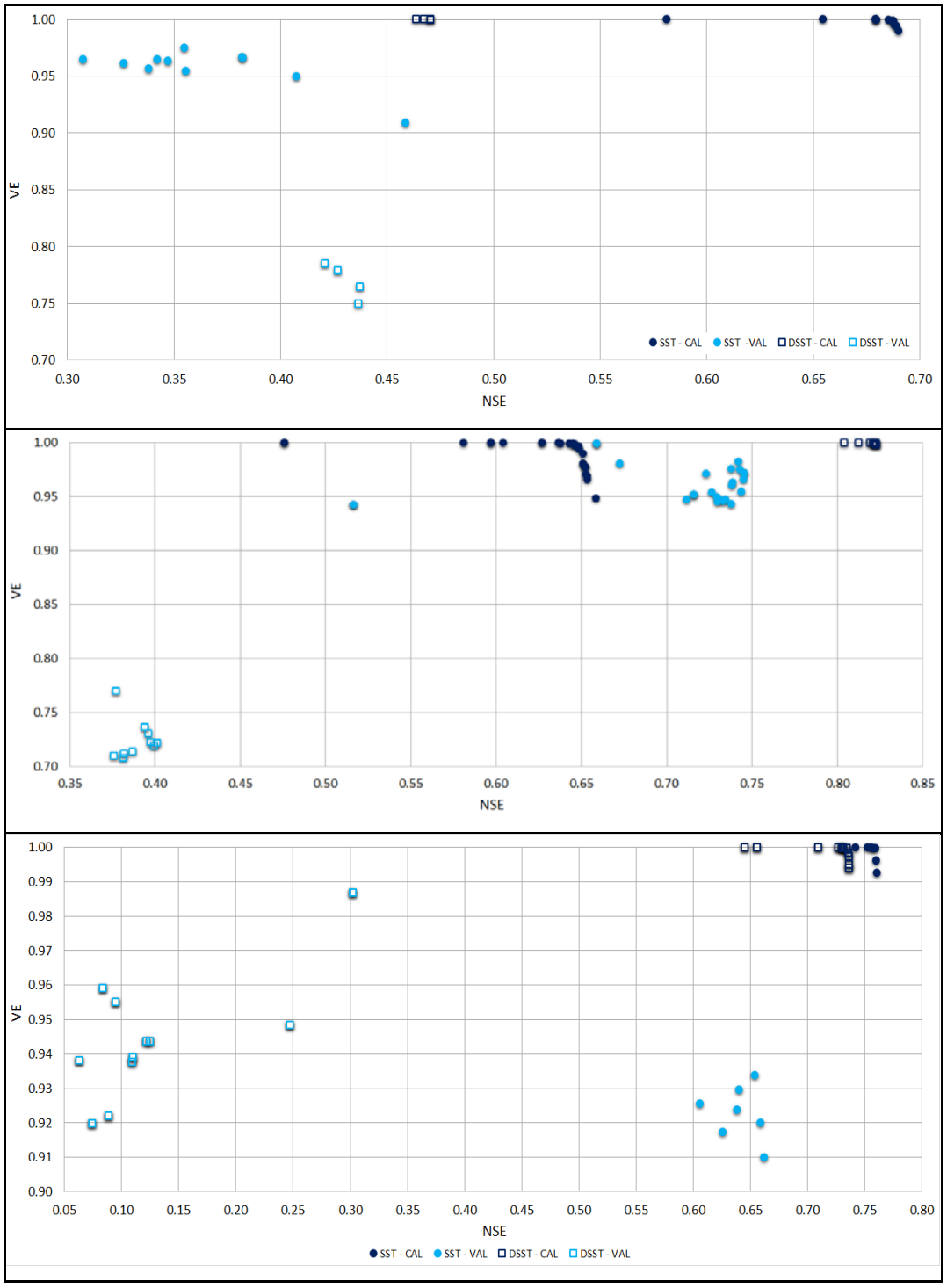


Figure 79. Results of the SST and DSST: the Kolubara (upper panel), Toplica (mid panel) and Mlava (lower panel) River catchments.

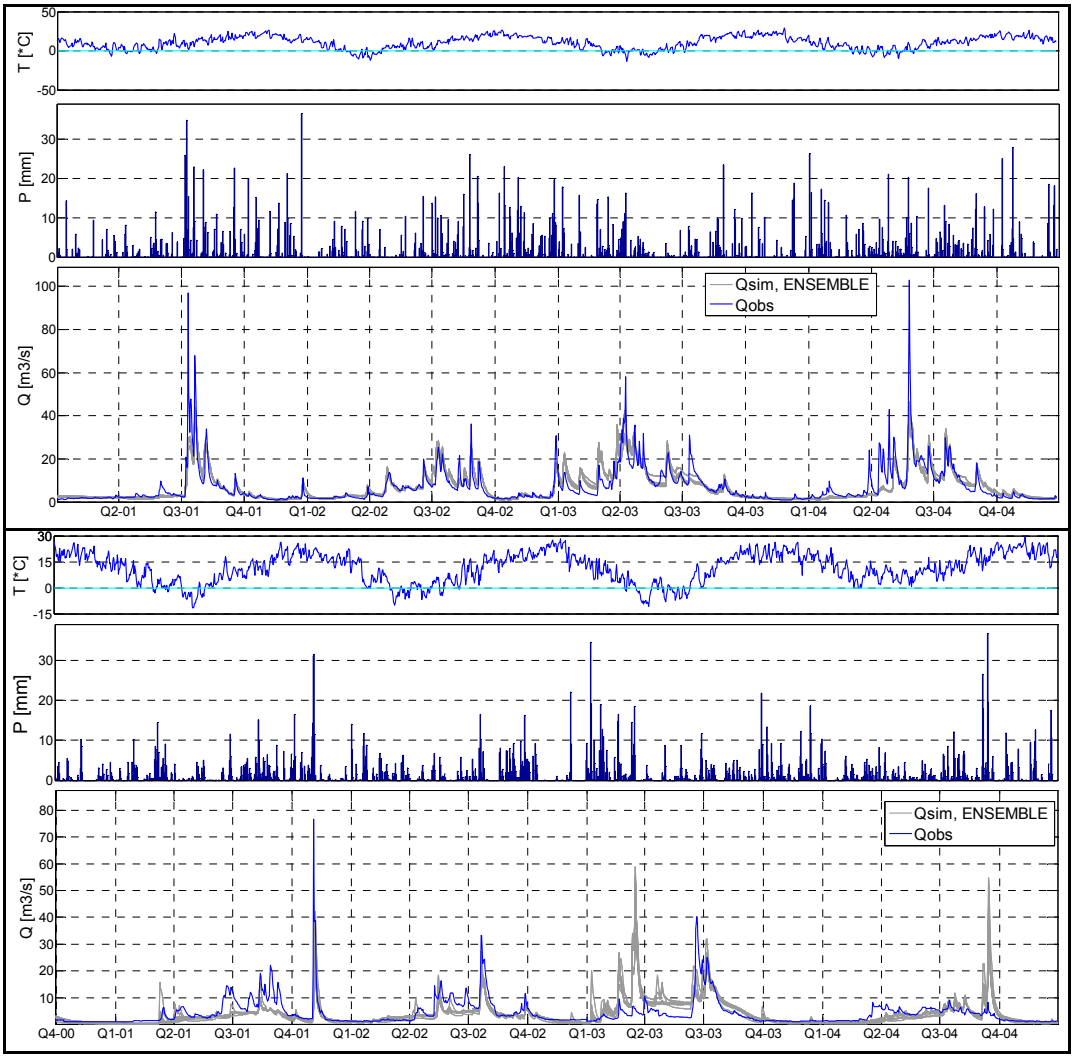


Figure 80. Simulations over the evaluation periods: *SST* – the Toplca River catchment (top panel) and *DSST* – the Mlava River catchment (bottom panel).

Table 26. Performance of the 3DNet-Catch and HBV models in the calibration and evaluation periods

Catchment	Calibration (1996-2010)				Evaluation (1986-1995)			
	3DNet-Catch		HBV		3DNet-Catch		HBV	
	NSE	Bias [mm / year]	*NSE*[19]	Bias [mm / year]	*NSE*	Bias [mm / year]	*NSE*	Bias [mm / year]
The Kolubara River	0.601 0.5998 0.5995	-0.8 (-0.25%) -7.9 (-2.6%) -16.6 (-5.5 %)	0.56 0.61	0 0	0.63 0.625 0.617	-24 (-10.6%) -31.6 (-13.7%) -38.5 (-16.9 %)	0.41 0.54	11.3 (5%) 12.3 (5.4%)
The Toplica River	0.719 0.7183 0.7182	-1.3 (-1%) -2.9 (-2.2%) -4 (-3%)	0.71 0.74	0 0	0.57 0.56 0.55	-16.7 (-16.6%) -18.6 (-18.5%) -19.9 (-19.8%)	0.58 0.68	-18.3 (-18.3%) 22.8 (23%)

A visual inspection of the simulated hydrographs along with the precipitation and temperature data reveals that finer temporal data resolution could contribute to more accurate flow simulations. High sensitivity of parameter α suggest that precipitation is crucial for model efficiency. Also, finer temporal resoulution of temperature observations would improve model performance in winter seasons. For example, prompt catchment response to precipitation during a period with mean daily temperaures below zero indicated in Figure 82 (highlighted with the rectangel) suggests that the temprature may have exceeded 0°C in that day and that precipitation may have been rainfall rather than the snowall. The model uderestimates the observed flood wave even if the value of the highest melt factor is adopted (dashed hydrograph in Figure 82).

[19] The first (top) values are obtained by NVE, while the second (bottom) ones are obtained by RHMSS (Langsholt et al., 2013).

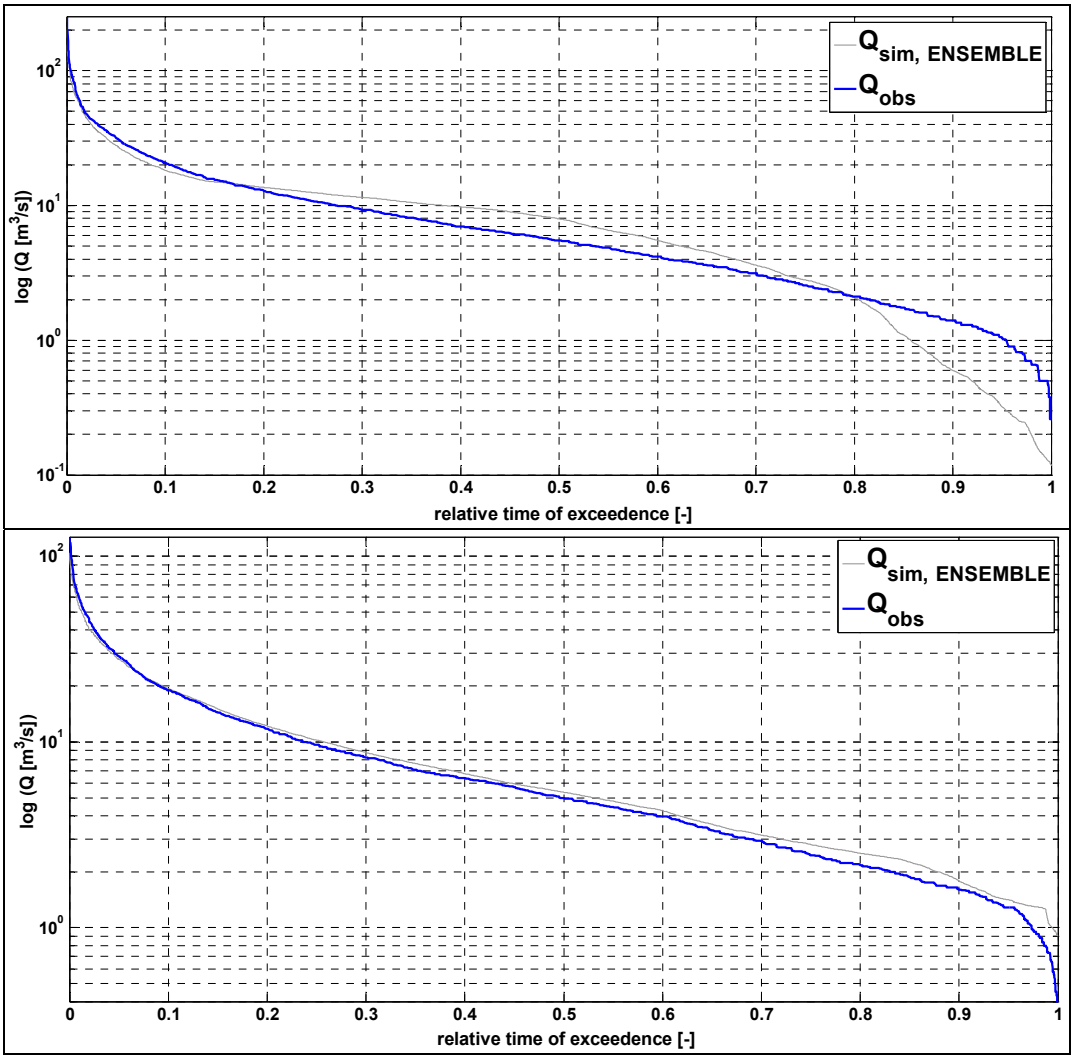


Figure 81. Flow duration curves of the simulated and observed flows at the Slovac (upper) and at the Doljevac stream gauge (bottom panel) over the calibration period (1996-2009).

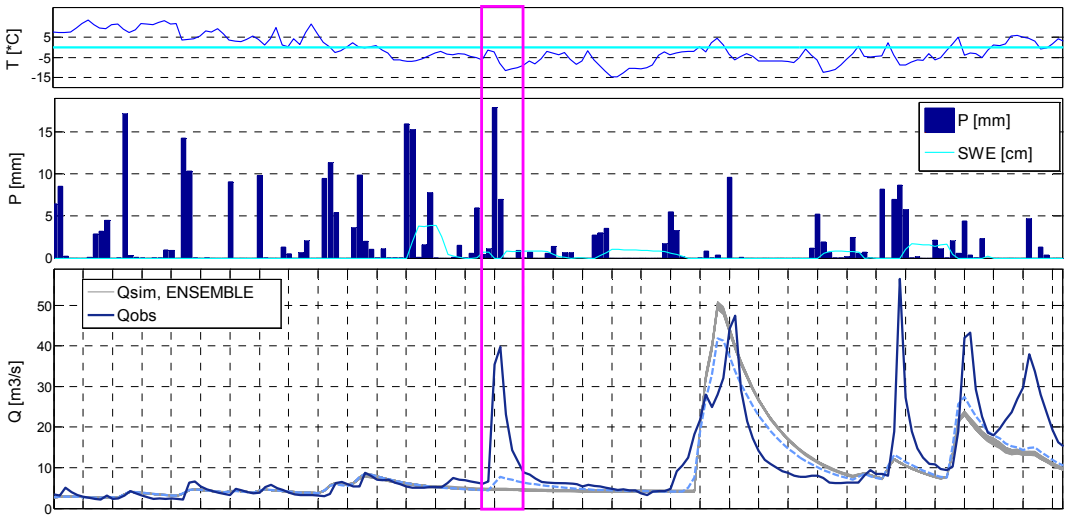


Figure 82. Snowmelt: the Toplica River catchment, winter 1999-2000.

3.2. Results of dynamic multi-objective calibration

3.2.1. Temporal variability in the model parameters

Variability of the normalised Pareto-optimal parameters of semi-lumped BASIC version of the model with the calibration period is presented in the multi-temporal graphs in Appendix D, which shows median value of the normalised Pareto-optimal parameters in every calibration period and *IC* statistic. Normalised parameters are preferred for these graphs, as they are insensitive to the prior parameter ranges, while the *IC* statistic provides insight into the spread of the Pareto-optimal parameters obtained over a period.

Median values of two Pareto-optimal parameters are presented in Figure 83. The abscissa values in these graphs denote start water year, while the ordinate values denote end water year of a calibration period. Diagonal cells in the graphs refer to 1-year long calibration period, while cell in the top left corner of the graphs refer to the 25-year calibration period (1988-2013).

These graphs reveal considerable variability with the calibration period in all parameters. This variability appears to be smaller for sensitive parameters such as α, $D_{\text{sub-surf}}$, K_{d} or *porosity*. Changes in the median parameter values between the periods can be abrupt for longer calibration periods as well as for shorter ones. For example, in the right-hand side

panel in Figure 83 median value of parameter *n* obtained over 1992-2008 is substantially larger than the values from all "surrounding" periods despite the overlap among the periods. Such the "chess-board" patterns emerge in the most insensitive parameters (e.g. baseflow-related and parameters of the snow routine).

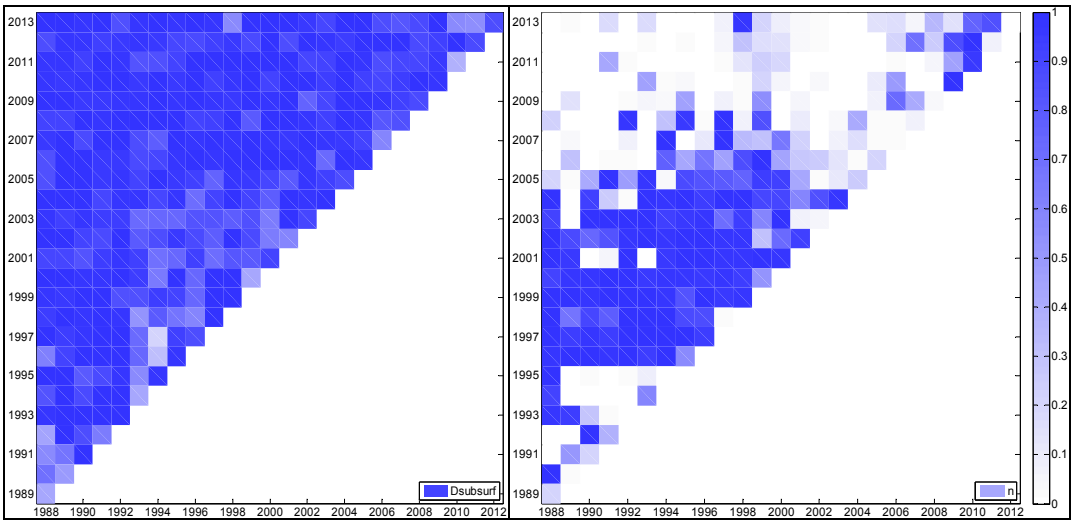


Figure 83. Multi-temporal graphs of the medians of the least ($D_{\text{sub-surf}}$) and the most variable (n) parameter of the 3DNet-Catch model for the Mlava river catchment. Abscissa values denote stat year and the ordinate values denote end year of a calibration period.

Parameter variability with the calibration period is quantified in terms of: (1) standard deviations calculated in respect to medians of the normalised Pareto-optimal parameters obtained over all calibration periods $S_{\text{t, norm}}$, and (2) ratios between S_{t} and the standard deviation obtained from the uniform prior distributions $S_{\text{u\_prior}}$ (Figure 84). The S_{t} diagram resembles the graph of S_{t} / $S_{\text{u\_prior}}$ given in Figure 84, but ordinate values vary between 0.1 and 0.45 (not shown here).

Despite being catchment specific, the results indicate maximum canopy storage (CAN_{max}), some parameters of the snow routine (SNOW_{100}, $b_{\text{melt, 6}}$ and $b_{\text{melt, 12}}$) and the soil routine (CN, $I_{\text{a,rel}}$, w_{wp}, w_{fc}, n) and baseflow-related parameters as rather sensitive to the calibration period. Temporal variability in these parameters exceeds their initial uncertainty, which is rather high is some parameters (e.g. SNOW_{100} or n). The smallest

variability is exhibited by parameters K_d, $D_{sub\text{-}surf}$, α, $T_{s\text{-}r}$ and λ. Former three parameters are sensitive parameters (Figure 74 and Figure 75, chapter 3.1).

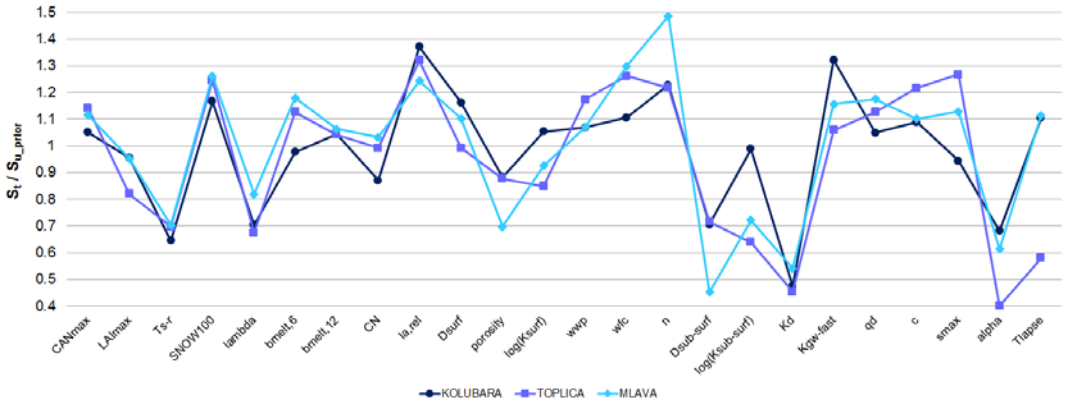


Figure 84. Ratio S_t / S_{u_prior} obtained from all calibration periods.

Although the variability of the sensitive parameters with calibration period appears to be low, no strong correlation between these two properties could be detected. This correlation for the Kolubara River catchment is illustrated in Figure 85 and similar results are obtained for other two catchments. For example, λ and $T_{S\text{-}R}$, both of which are insensitive, exhibit relatively low variability with the calibration period (Figure 84).

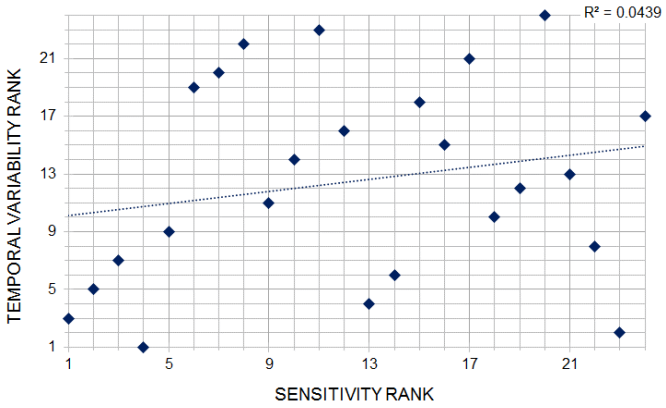


Figure 85. Rank correlation between parameter sensitivity and variability with the calibration period: the Kolubara River catchment.

To examine how the parameter variability with the calibration period depends on the period length, median values of the normalised Pareto-optimal parameters obtained from

all periods are plotted against the period length (Figure 86, Appendix E). Greater width of these box-plots indicates larger variability. It is expected that temporal variability in parameters decreases with the calibration period length due to the overlap between consecutive periods (e.g. the overlap in two consecutive 25-year long periods amounts to 24 years, i.e. 96%).

It appears that parameter variability tends to decrease with the length of the calibration period in some parameters (e.g. K_d, $\log_{10}$ ($K_{\text{sub-surf}}$) or $D_{\text{sub-surf}}$) although no clear pattern could be detected in majority of the parameters. For example, parameter $D_{\text{sub-surf}}$ in left-hand side panel in Figure 83 takes larger values over longer calibration periods and differences in colour of adjacent cells are slight, as opposed to cells that denote 1-year long calibration periods. As for parameter K_d, median values tend to converge to median value of the set optimised over the full record period for all three catchments (left panel in Figure 86, Appendix E). On the other hand, median values of some other parameters (such as $b_{\text{melt, 6}}$, $\log_{10}$ (K_{surf}), w_{fc} and s_{max}) significantly deviate from the medians of the Pareto set obtained over full record period.

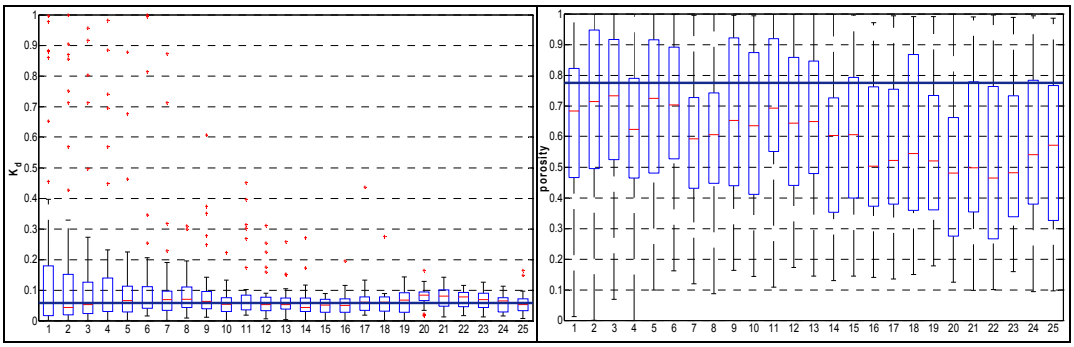


Figure 86. Box plots of the median normalised Pareto-optimal parameters from all calibration periods of length specified on the abscissa for the Kolubara river catchment: linear reservoir coefficient K_{d} (left panel) and *porosity*. Thick lines denote median value of Pareto-optimal parameters obtained over the full hydrologic record period.

Degree of dependence between the parameter sensitivity to calibration period and the period length is also quantified in terms of ratios between S_{t} obtained over 5- and 10-year long calibration periods (Appendix F) and S_{t} obtained over 1-year long periods (Figure 87 and Figure 88, respectively). Values of the ratios greater than 1 denote larger variability over longer calibration periods. This suggests inconsistency in parameter

estimation. Such variability is detected in few mainly insensitive parameters, such as w_{fc} or s_{max} (exception is *porosity*), although the results are catchment specific. The ratios $S_{(10\ years)} / S_{(1\ year)}$ are not always larger than $S_{(5\ years)} / S_{(1\ year)}$ (e.g. for LAI_{max} obtained in the Toplica River catchment), which is also an evidence that longer calibration period does not necessarily lead to more consistent parameter estimation.

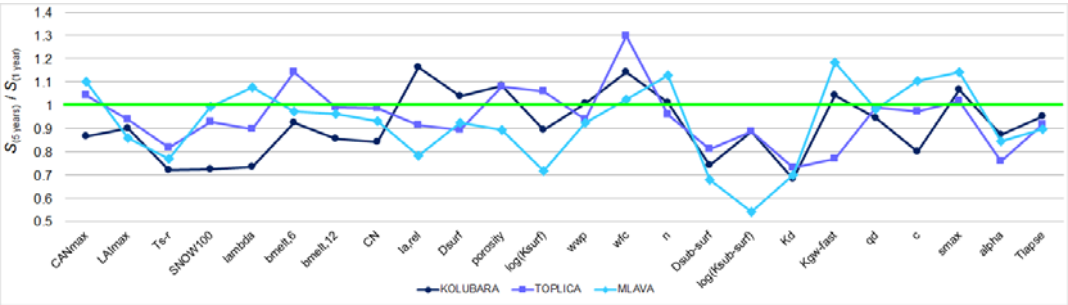


Figure 87. Ratio of S_t obtained over the 5-year long calibration periods to S_t obtained from the 1-year long periods.

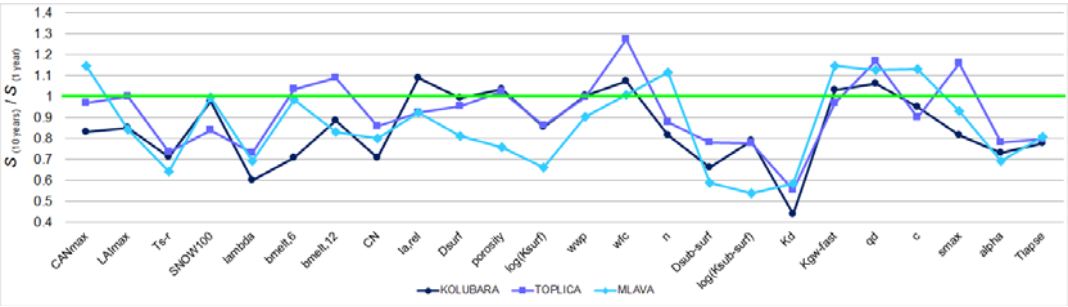


Figure 88. Ratio of S_t obtained over the 10-year long calibration periods to S_t obtained from the 1-year long periods.

In addition, empirical cumulative distribution functions (*ecdf*s) are derived from all Pareto sets obtained over calibration periods of given lengths (1, 5, 10, 15, 20 and 25 years) and from the Pareto-optimal parameter from all calibration periods. As such, these *ecdf*s reflect parameter uncertainty due to calibration period. These *ecdf*s are shown in Appendix G along with the *ecdf*s obtained for the full record period ($ecdf_{FRP}$). Behaviour of the *ecdf*s is consistent with the behaviour exhibited by the median values of Pareto-optimal parameters: (1) *ecdf*s depend on the length of the calibration period, and (2) *ecdf*s may considerably deviate from the $ecdf_{FRP}$ (Figure 89). For example, *ecdf*s obtained over

1-year long calibration periods are more similar to *ecdf*s of 20- and 25-years than to e.g. *ecdf*s of 15-years; i.e. there is no regular change in *ecdf*s with calibration period (left-hand side panel in Figure 89). The *ecdf*s of parameter K_d are found most consistent (right-hand side panel in Figure 89). Also, many *ecdf*s resemble the uniform distribution (especially *ecdf*s achieved over all calibration periods, e.g. $b_{melt,\ 12}$) which may be the consequence of both, parameter insensitivity (considerable variability within a Pareto set, i.e. parameter uncertainty) and variability with the calibration period.

Generally the $ecdf_{FRP}$ are significantly narrower than the *ecdf*s and the reason for such behaviour is the way the Pareto sets are obtained. Namely, $ecdf_{FRP}$ are obtained from the Pareto sets optimised over the full hydrologic record period, which means that these sets are adjusted to result in optimal average performance over the full record period, but they would probably not yield high model efficiency in different sub-periods. By averaging the model performance over the entire calibration period a significant amount of information form the data available is lost. The *ecdf*s are obtained from the Pareto-optimal set in various sub-periods, which enables extraction of more information from the available observations. The *ecdf*s obtained over short calibration periods represent parameters that are optimised to reproduce a catchment response in few events, while the *ecdf*s achieved over long calibration period describe parameter sets that have been exposed to wide variety of a catchment's responses and thus parameters sampled form these *ecdf*s may be considered more robust. Further research is required to test these hypotheses.

The results presented so far suggest high sensitivity of the Pareto-optimal parameters to the calibration period. To estimate effects of such variability on model performance, the following statistics are estimated:

$$S_{NSE_i} \propto SRC_{NSE,i} \frac{S_{t,i}}{S_{\text{u\_prior},\ i}} \tag{3.2.1}$$

$$S_{VE_i} \propto SRC_{VE,i} \frac{S_{t,i}}{S_{\text{u\_prior},\ i}} \tag{3.2.2}$$

where SRC_i stands for the standardised regression coefficient of the i^{th} parameter given in the equation 2.3.5 for two objective functions (*NSE* and *VE*), S_t is the standard deviation

of the i^{th} parameter calculated from all calibration periods and S_{u_prior} denotes standard deviation of the prior uniform distribution of the i^{th} parameter. Previous two equations stand for standardised regression coefficients, with the standard deviation of sample parameters being substituted with standard deviation of median parameter values estimated over all calibration periods. This modification is assumed to enable assessment of possible consequences of the parameter variability with the calibration period to model performance (objective functions). Implicit assumption in these equations is that there are no strong correlations among the model parameters.

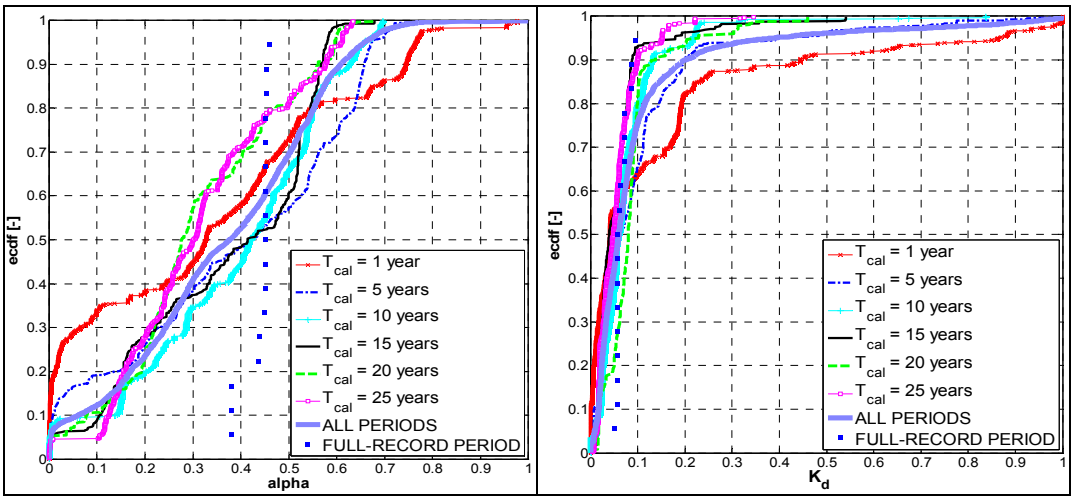


Figure 89. Empirical cumulative distribution functions of the precipitation gradient α (left panels) and K_d parameters (right panel): the Kolubara River catchment.

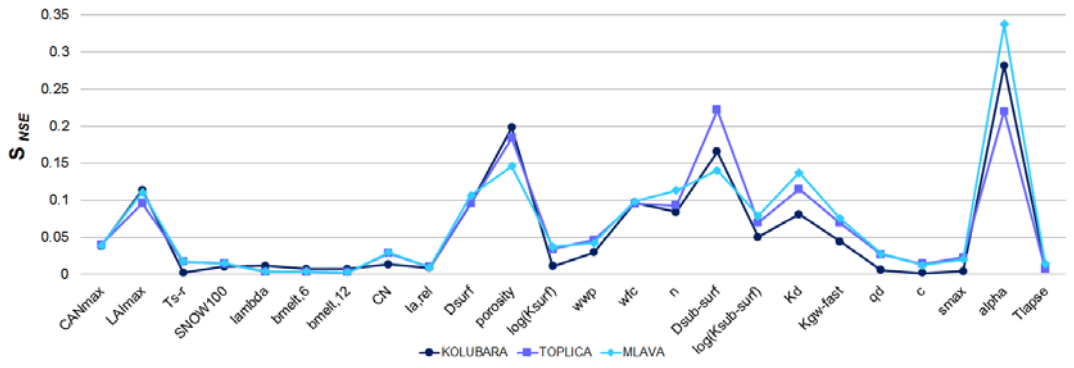


Figure 90. Statistic S_{NSE} .

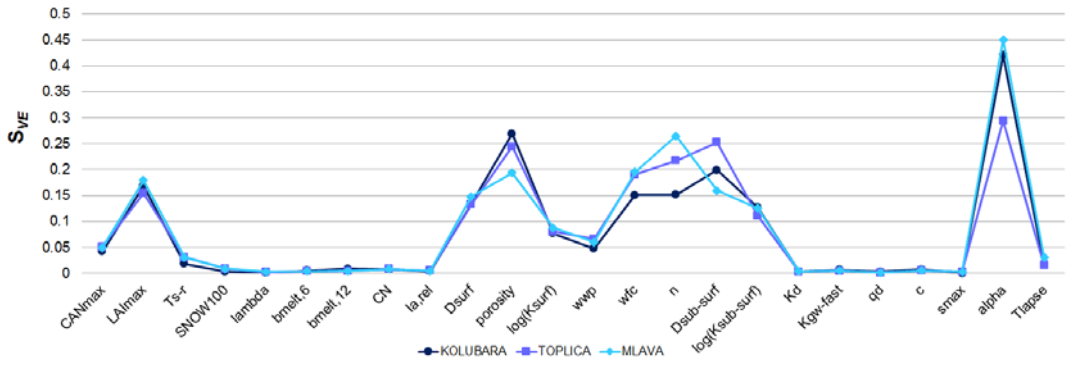


Figure 91. Statistic S_{VE}.

Values of these statistics are presented in Figure 90 and Figure 91, respectively. The results are catchment specific, although they all reveal that the variability of α, LAI_{max} and parameters of the soil-routine could strongly affect model performance. This hypothesis is tested by simulating runoff in 5 driest years in the Kolubara River catchment with parameters optimised in this period, with parameters optimised over 5 wettest years (*DSST*) and with three combinations of these parameters sets:

(1) *IS_VE*_unchanged: parameters optimised in the driest 5-year long period with low values of S_{VE} are kept, while the remaining parameters are substituted with median values optimised in the wettest 5-years.

(2) *S_VE*_unchanged: parameters with high values of S_{VE} are kept (CAN_{max}, LAI_{max}, most parameters of the soil routine and precipitation gradient α) and the remaining parameters are replaced with median values obtained over the wettest period.

(3) *S_NSE_VE*_unchanged: in addition to previous, K_d is kept (high value of *S-NSE* statistic).

The results of this analysis are presented in Figure 92. If parameters with low S_{VE}, optimised over the driest five year, are kept the results are slightly better compared to the results of the *DSST*. Second group of parameters sets resulted in higher values of *VE*, but without any improvement of *NSE* values. The third group of parameters resulted in better overall performance, which is still weaker than the performance of the Pareto-optimal sets. These results confirm the parameter sensitivity to the calibration period. Discrepancy

between performance of the Pareto and *S_NSE_VE_*unchanged sets may be attributed to somewhat higher sensitivities of replaced the parameters over this particular period (here *CN*) and thus, drop in the values of the objective functions. This difference in model performance may also be caused possible interactions among model parameters, which is neglected in this analysis.

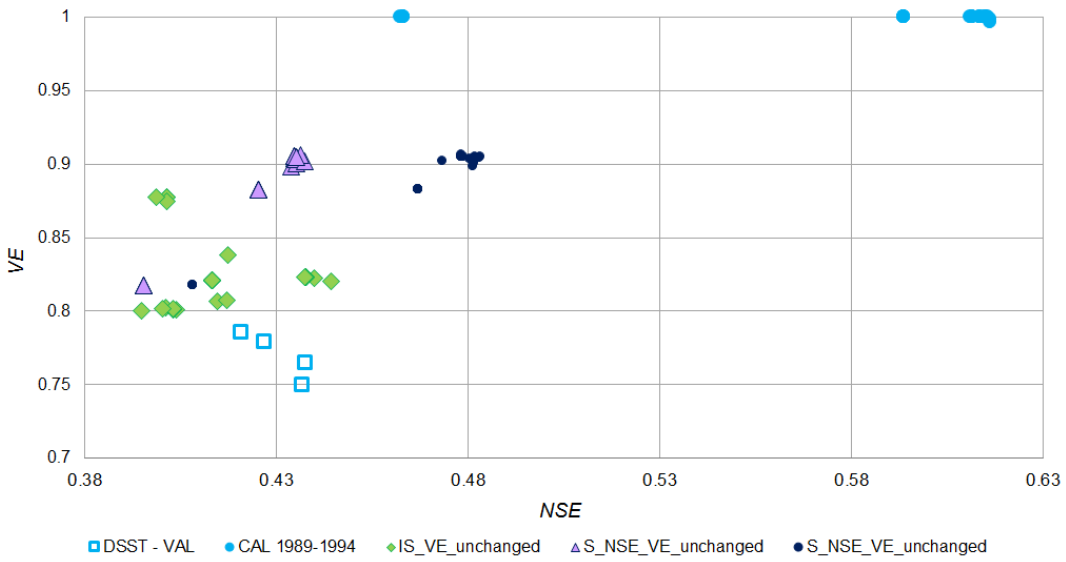


Figure 92. Model performance over five driest water years (1989-1994) given different parameter sets: the Kolubara River catchment.

The reason for variability in Pareto-optimal parameters with time may be the variability of the process represented by the parameter. For example, estimated parameter α, which is shown to be rather important for model performance, varies extensively with the calibration period as presented in Figure 93. Since the increase in precipitation with elevation also varies in time, as shown in Figure 45, Figure 59 and Figure 72, variability in this parameter seems to be inevitable. However, this variability may decrease if data with finer resolution were used (as discussed in chapter 2.6) or if a different (nonlinear) change of precipitation increase with elevation were employed.

This variability with the calibration is also illustrated in Figure 94, which shows that the "contour lines" of the response surface obtained over two different 25-year long periods mismatch, i.e. the response surface changes dynamically with new observations.

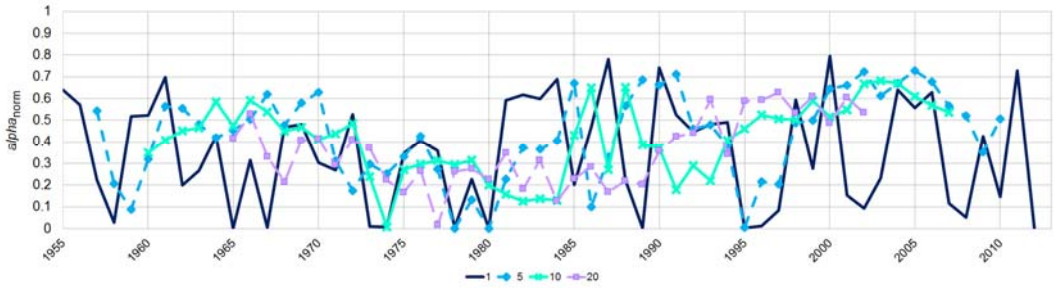


Figure 93. Variability in normalised parameter α optimised over different periods: the Kolubara River catchment.

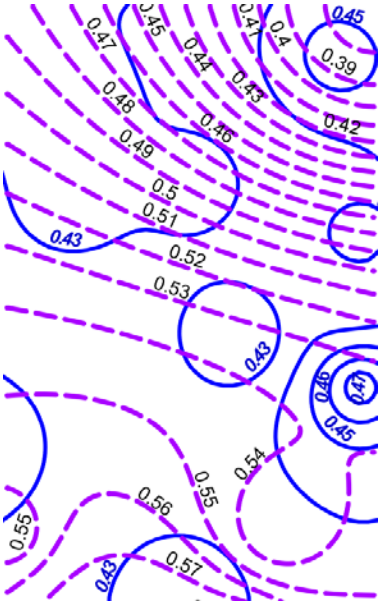


Figure 94. Response surface of the 3DNet-Catch model for the Kolubara River catchment: the Nash-Sutcliffe efficiency *NSE* as a function of two model parameters (precipitation gradient α and filtration coefficient K_d) in the 1988-2013 period (solid contour lines) and in 1957-1982 (dashed contour lines).

Parameter identifiability

Parameter identifiability is quantified in terms of the *IC* statistic, which is presented in the multi-temporal graphs in Appendix D. Mean *IC* for the model parameters, averaged over all calibration periods, is given in Figure 95. The results suggest a high identifiability of all parameters, especially the sensitive ones. This may be attributed to both, model parameterisation and robustness of the AMALGAM. The largest spread in the Pareto sets is found in the snowmelt factors.

Variation of the *IC* is quantified by its standard deviation and coefficient of variation, and presented in Figure 96. Low variability of the *IC* indicates narrow ranges of Pareto-optimal parameters in most calibration periods. Such behaviour is exhibited by the most sensitive parameters (*alpha*, K_d, $D_{sub\text{-}surf}$, *porosity*) and by some less sensitive parameters: $T_{s\text{-}r}$ and log ($K_{sub\text{-}surf}$). These results are supported by the multi-temporal graphs in Appendix D. No apparent correlation between the *IC* and the calibration period length could be detected (Figure 97, Appendix H).

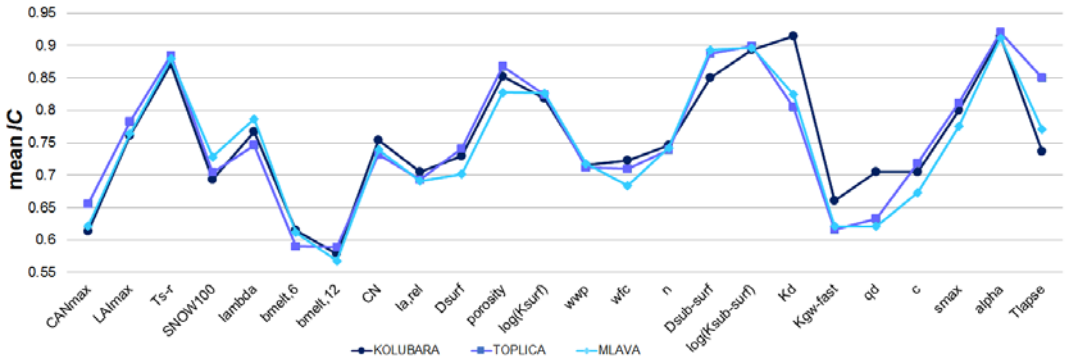


Figure 95. Mean *IC* statistic for the model parameters obtained over all calibration periods.

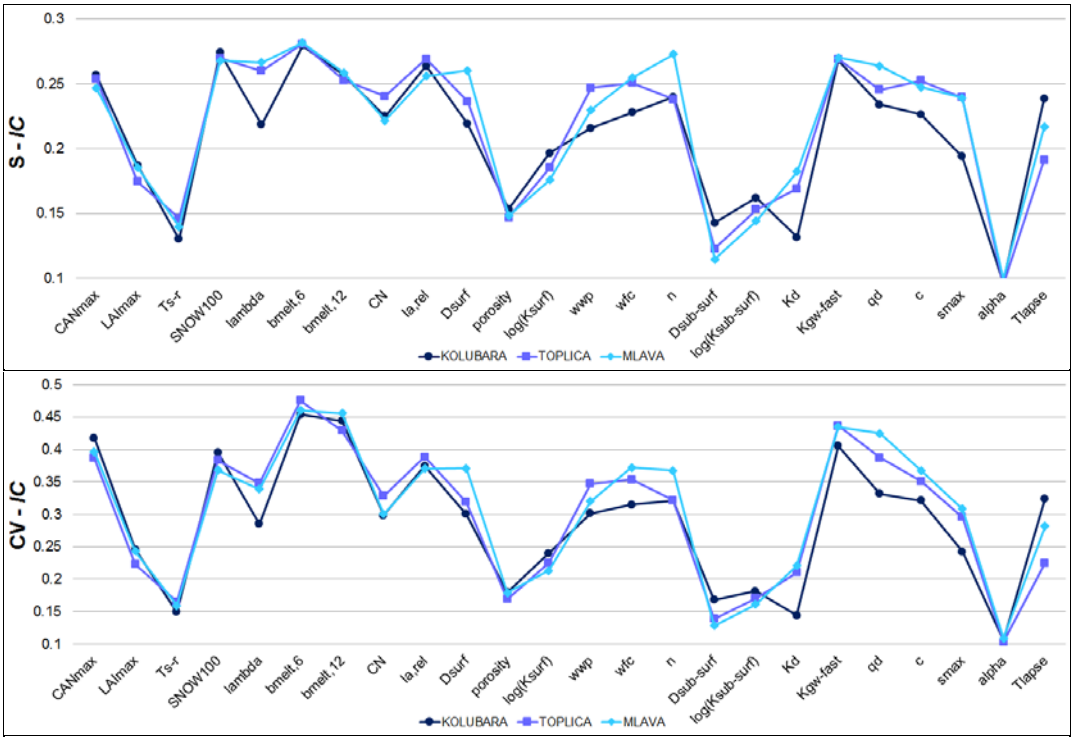


Figure 96. Standard deviation (top panel) and coefficient of variation (bottom panel) of the *IC* statistic for the model parameters.

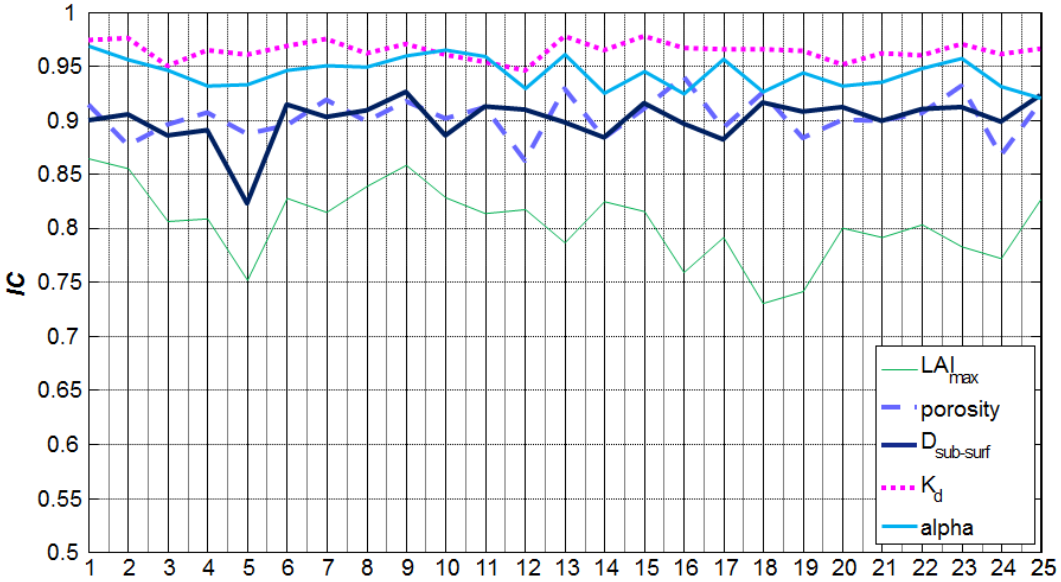


Figure 97. Variation of the *IC* statistic with the length of the calibration period (abscissa values): the Kolubara River catchment.

Model performance

Model performance over different calibration periods is illustrated in multi-temporal graphs in Appendix I. The results reveal that less than 30% of parameter sets converge to Pareto front. Nevertheless, the result are period-specific. Seldom all parameter sets become non-dominated (the Pareto-optimal sets), but significant pattern of such periods emerges.

Performance of a Pareto-ensemble (hydrographs obtained with the Pareto-optimal parameter sets) is quantified in terms of the *p*-factor and the *r*-factor (chapter 2.5). Both performance measures have small values for all catchments and in all calibration periods. These results indicate rather narrow ensemble band (*r*-factor) and a few observations are encompassed by the ensemble despite a satisfactory resemblance between the simulated and observed hydrographs. It again may be a result of the robustness of the AMALGAM, meaning that all sets converge to quite a narrow optimum region and thus resulting in similar response hydrographs (and flow duration curves as illustrated in Figure 81). Small values of the *r*-factor are desirable, but this does not hold for values of the *p*-factor. Poor performance in terms of the *p*-factor may be the consequence of small values of the *r*-factor. Also, these results may indicate that the parameter uncertainty[20] is not the prevailing one and / or that multi-objective calibration with the AMALGAM algorithm underestimates parameter uncertainty.

Maximum, mean and minimum values of two objective functions obtained from all calibration periods are presented in the Appendix I. Model performance with respect to *VE* is significantly better compared to the performance with respect to *NSE*. This indicates model capability for overall water balance simulation, but its modest ability to reproduce dynamics of these catchments with the coarse resolution of input data.

[20] Multi-objective calibration is not a genuine method for assessment of parameter uncertainty (chapters 1.3 and 1.4). Here, the 2.5-97.5 percentile parameter interval obtained from the Pareto-optimal sets is used to represent the parameter uncertainty.

Mean *NSE* [21] was found to decrease approximately logarithmically with calibration period length (Figure 98). A slight increase in *NSE* in periods longer than 20 years in length for Kolubara and Mlava can be observed in Figure 98. Changes in *VE* with calibration period length (Figure 99) are not distinct, although a slight increase with the period length can be detected. Variability in *VE* is smaller than variability in *NSE*. The values of *VE* are generally very high (even the minimum *VE* of Pareto-optimal sets). Mean differences between the simulated and observed annual runoff are illustrated in Figure 100. The graph shows that the smallest differences are obtained over 19-year calibration periods and longer, and that the model tends to slightly underestimate runoff rather than other way round. The results obtained for the Mlava River catchment have a slight departure from the results for the remaining two catchments, which may be attributed to short observation period (e.g. there is only one 25-year long calibration period). Generally, model performance for the Toplica River catchment is slightly better than at the remaining catchments.

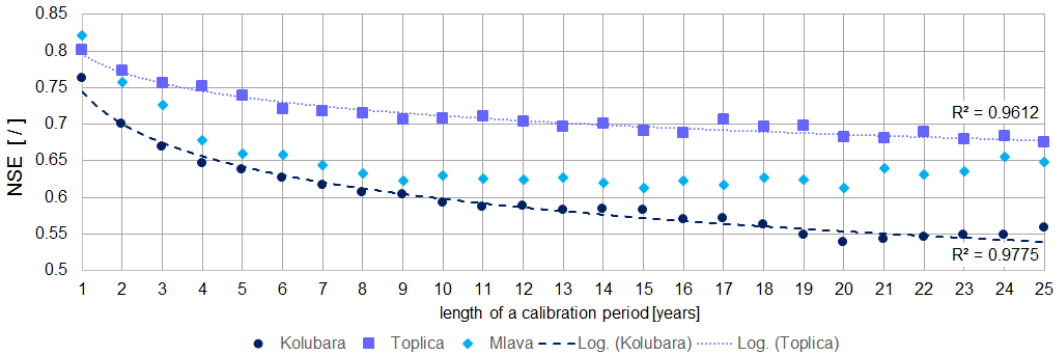


Figure 98. Mean *NSE* values against the length of a calibration period.

Model performance is also quantified with respect to other three evaluation measures: *NSE* obtained from log-transformed flows ($NSE_{\log Q}$), Kling-Gupta efficiency (*KGE*) and coefficient of determination (R^2). $NSE_{\log Q}$ varies substantially with the calibration period without following any regular pattern. The reason for such behaviour may be the fact that model efficiency in low-flow domain was not included in the calibration procedure, i.e.

[21] Mean *NSE* values are calculated according to the median *NSE* values in every calibration period, as described in chapter 2.5.

the parameters are not optimised to accurately reproduce baseflows. The remaining two performance measures take fairly high values and slightly decrease with the length of calibration period, similarly to *NSE*. Considering similarities between these measures and *NSE* (chapter 1.3.3), such result is expected.

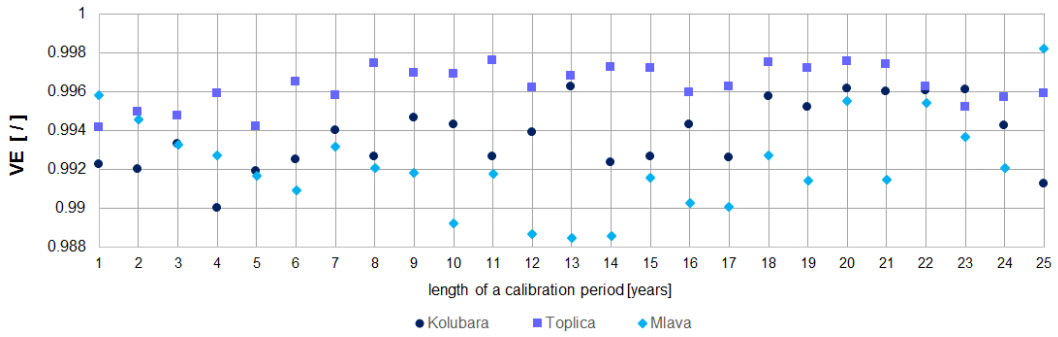


Figure 99. Mean *VE* values against the length of calibration period.

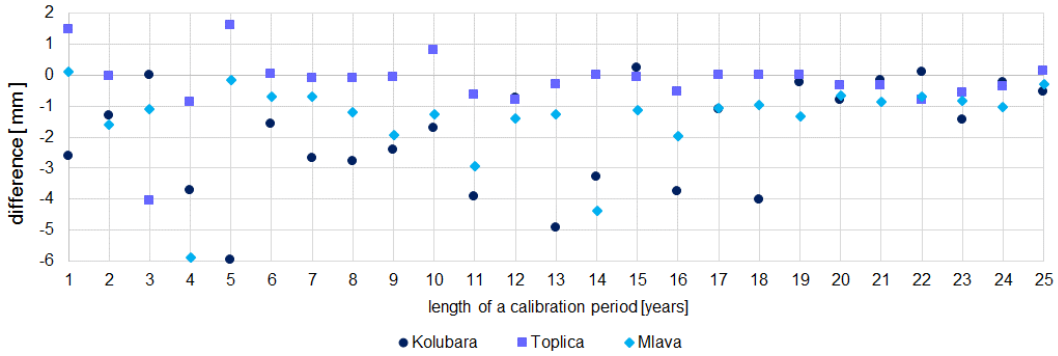


Figure 100. Mean difference between observed and simulated runoff against the length of a calibration period.

3.2.2. Parameter temporal variability and hydro-meteorological characteristics

Hydro-meteorological indices considered in this analysis (Table 14) are illustrated in multi-temporal graphs in Appendix J. Unlike the Pareto-optimal parameters, no abrupt change across the calibration periods can be detected in the indices, except for the minimum or maximum values of precipitation and temperature.

Strength of the relations between the optimised model parameters (medians of the Pareto parameters) and the hydro-meteorological indices are quantified in terms of the Pearson and Spearman correlation coefficients and as variable importance in the "tree bagging" metamodel (chapter 2.4).

To illustrate relation between Pareto-optimal parameters and hydro-meteorological characteristics of the calibration period, values of the normalised Pareto-optimal parameter ($D_{\text{sub-surf}}$) and API 30 index obtained over 5-year long calibration periods for the Kolubara River catchment are illustrated in Figure 101. Resemblance between these variables is small.

The behaviour illustrated in Figure 101 is confirmed by Pearson correlation coefficients between median parameter values and hydro-meteorological indices. The correlation coefficients between these values are given in the top panels of Figure 102 through Figure 104, while their statistical significance is illustrated in the bottom panels: white fields denote statistically significant correlations at 95% significance level. The Spearman correlation coefficients are presented in Appendix K and exhibit a similar pattern as the Pearson correlation coefficients. Correlation is quite weak for all catchments or combinations of the parameters and hydro-meteorological indices. Statistically significant correlations at the 95% significance level considerably vary for three catchments. Similar results are obtained for the *IC* statistic and the median values of the objective functions (Appendix K).

As for variable importance (Appendix K), majority relations between the Pareto-optimal parameters and the indices are weak. For example, parameter α is sensitive to some precipitation related indices in the Kolubara and Mlava River catchments, but insensitive in the Toplica River catchment. Such the contrasting results are consistent with weak correlations detected.

Additionally, no resemblance between multi-temporal graphs of the model parameters (Appendix E) and multi-temporal graphs of the hydro-meteorological indices (Appendix J) can be observed, which is consistent with these results.

The results of this analysis suggest a lack of relationship between the optimal model parameters and hydro-meteorological conditions in a catchment over some period. Consequently, the indices used to describe hydro-meteorological characteristics over a

calibration period cannot be used for conditioning of the Pareto-optimal parameters (i.e. to select different parameter values based on the hydro-meteorological characteristics of the simulation period).

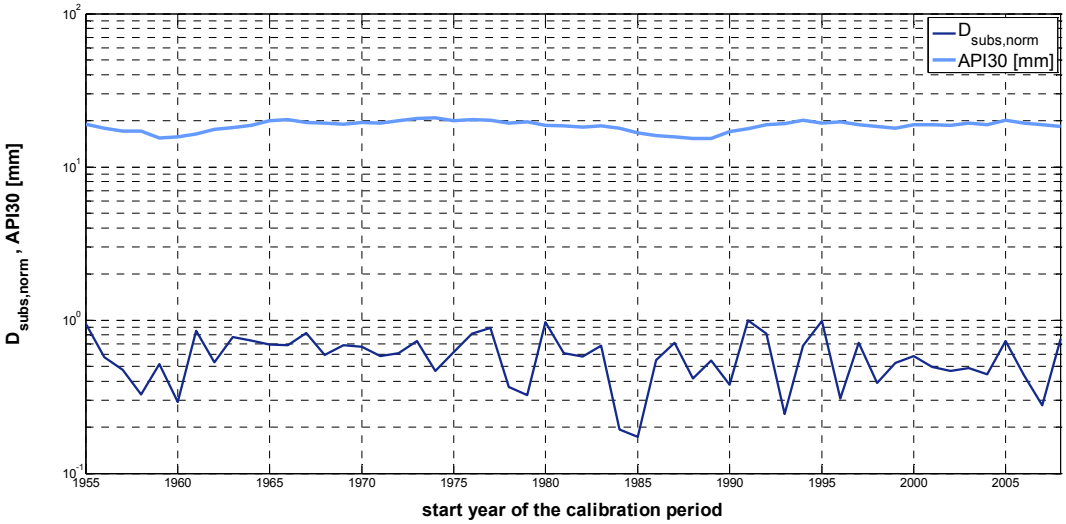


Figure 101. Median values of the normalised Pareto-optimal parameter $D_{\text{sub-surf}}$ and values of API 30 estimated over 5-year long calibration periods: the Kolubara River catchment.

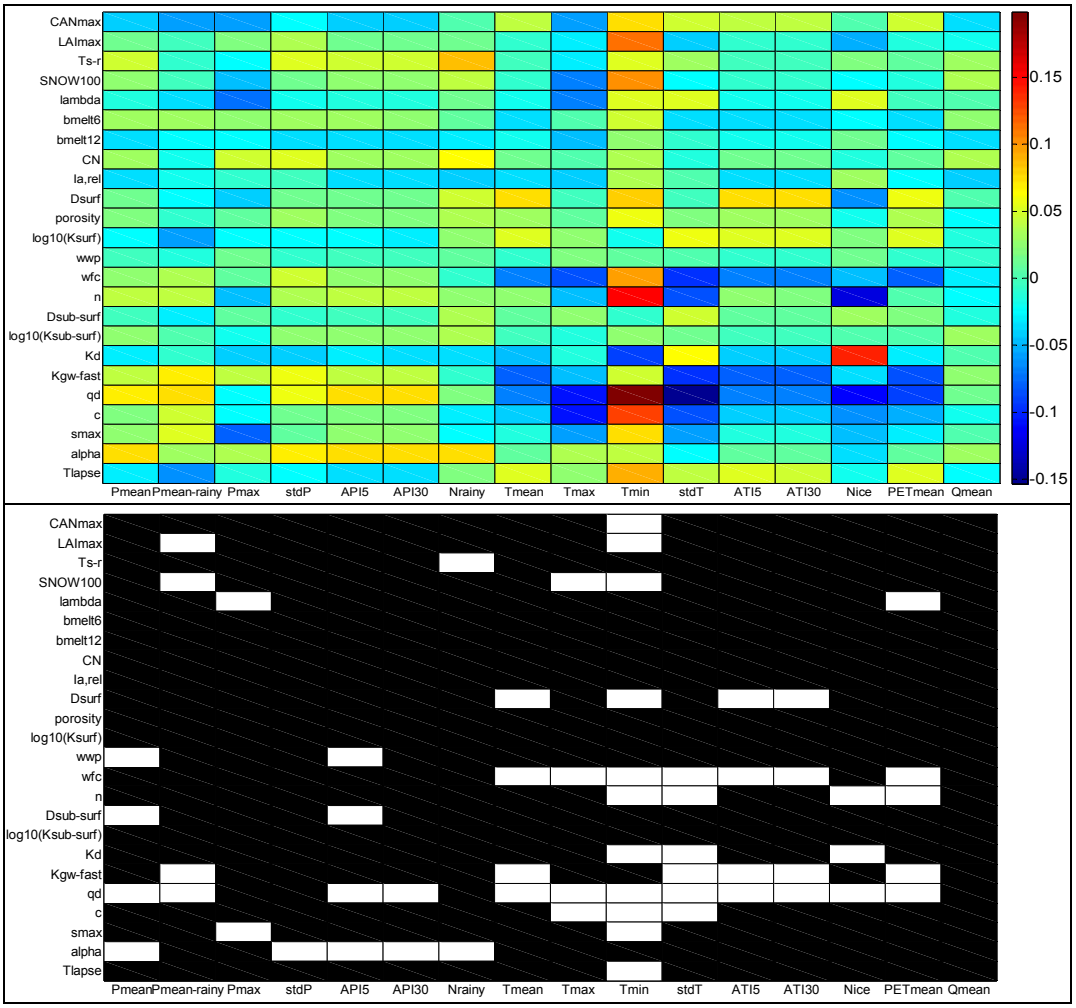


Figure 102. Pearson correlation coefficients between median parameter values and hydro-meteorological indices (top panels) and statistical significance at 95% significance level (bottom panel): the Kolubara River catchment.

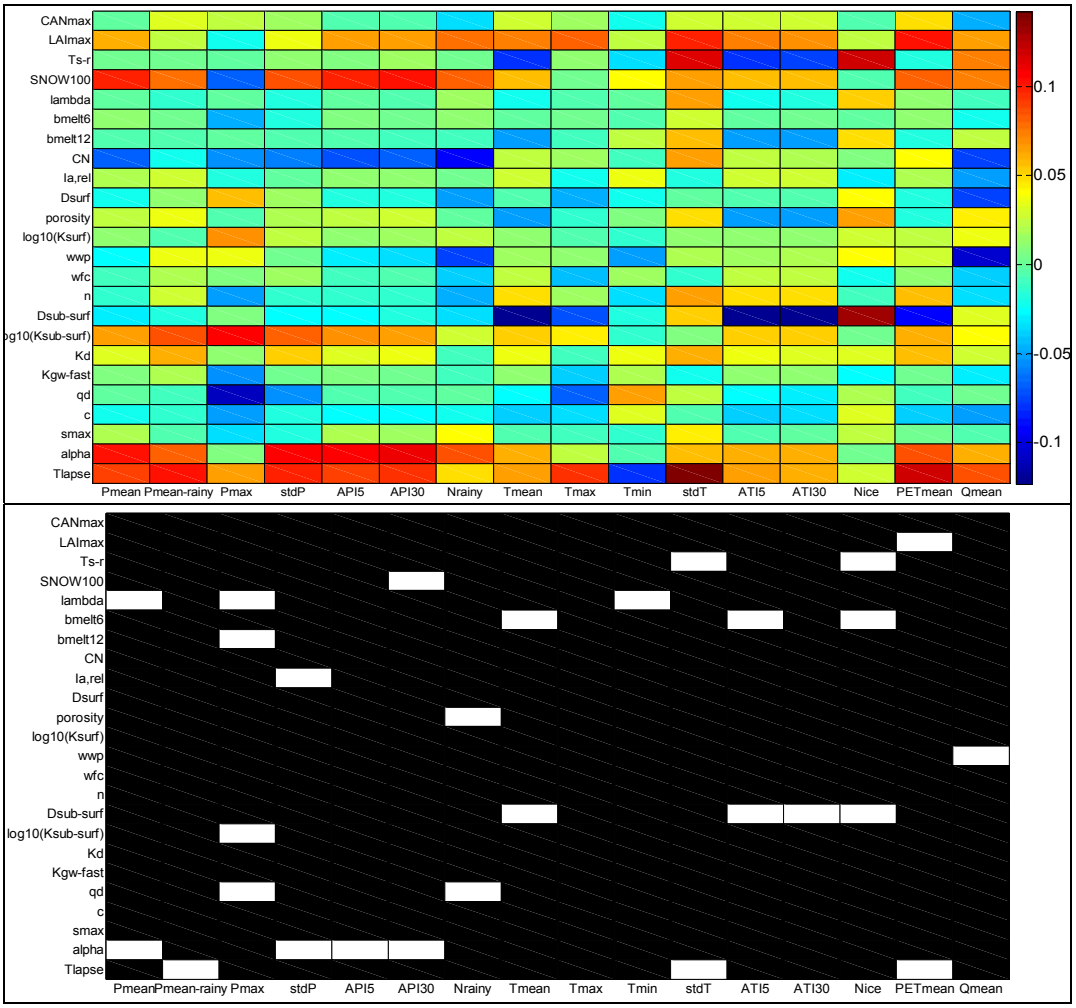


Figure 103. Pearson correlation coefficients between median parameter values and hydro-meteorological indices (top panels) and statistical significance at 95% significance level (bottom panel): the Toplica River catchment.

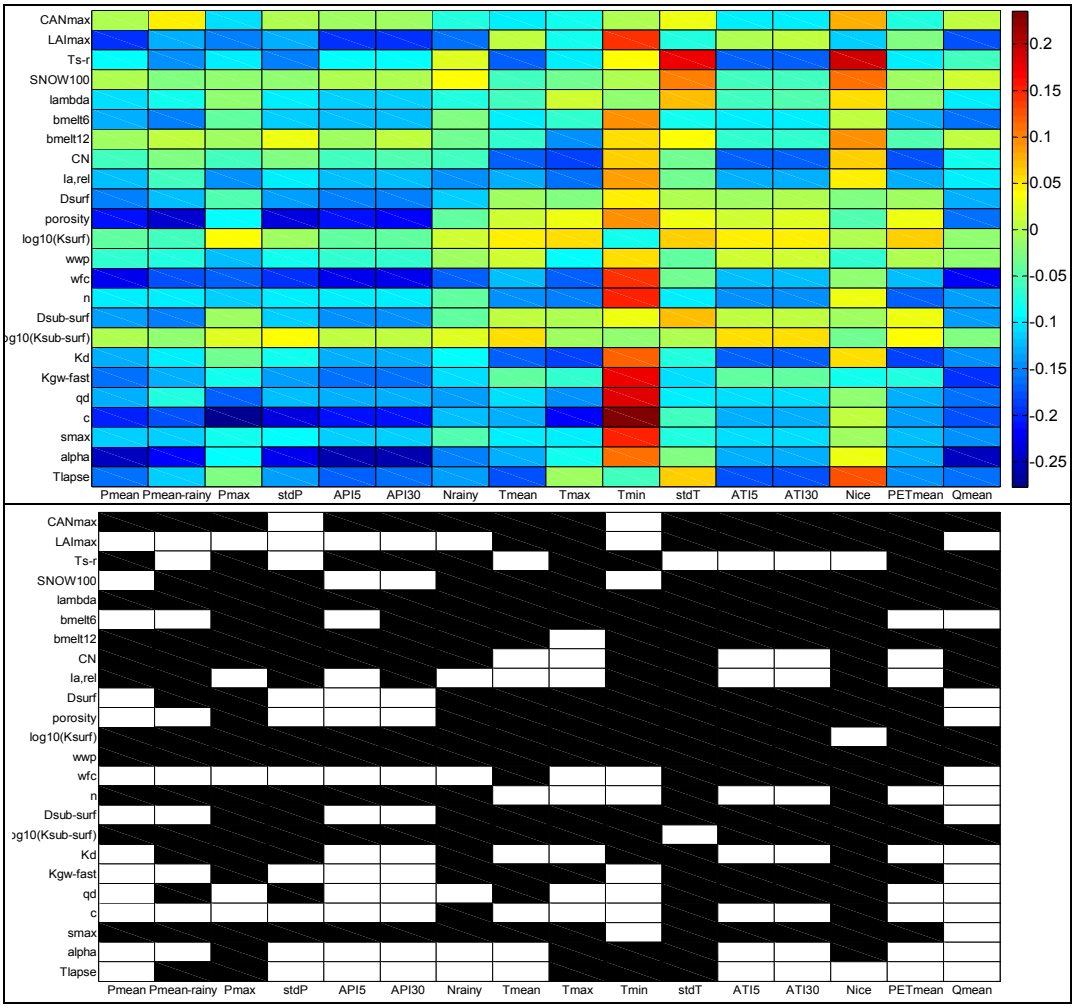


Figure 104. Pearson correlation coefficients between median parameter values and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Mlava River catchment.

3.2.3. Impact of the objective functions

Impact of the combination of objective functions (calibration strategy) on parameter temporal variability is assessed by analysing seven different strategies outlined in Table 15. As different combinations of the objective functions yield different Pareto-optimal parameters (as illustrated in Figure 105), aim of this analysis is to examine whether some calibration strategies lead to more consistent parameter estimation.

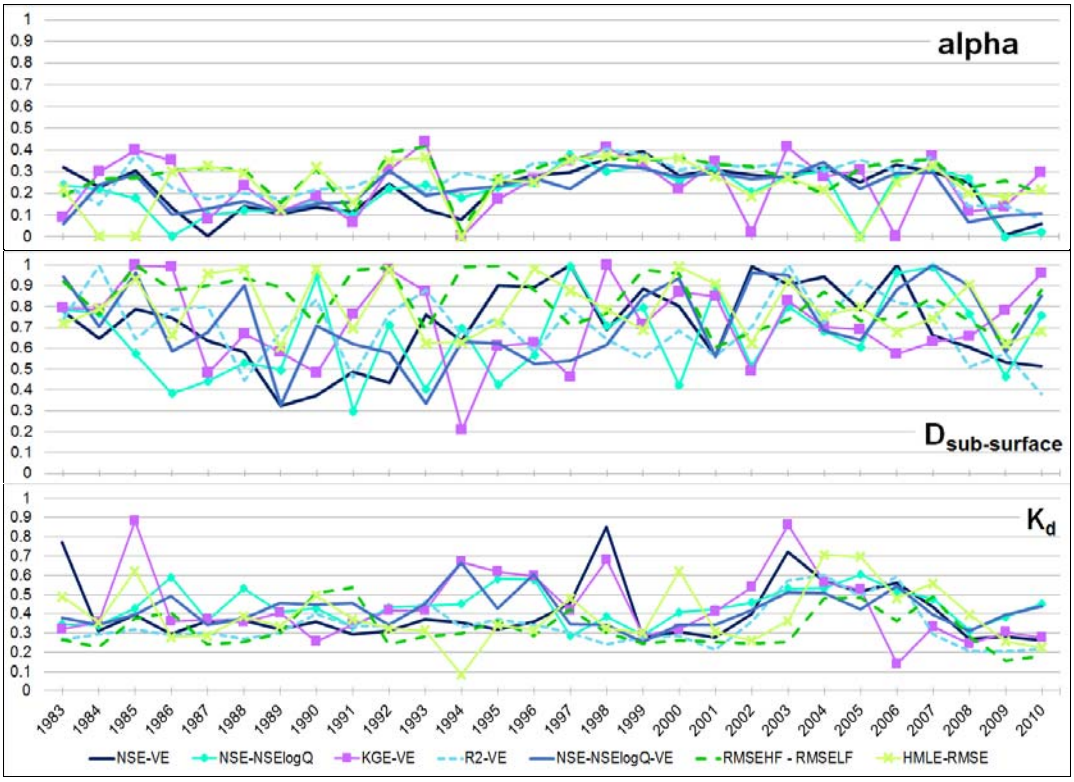


Figure 105. Median values of the normalised Pareto-optimal parameters (*alpha*, $D_{\text{sub-surface}}$ and K_d) obtained for the Toplica River catchment for different calibration strategies.

Parameter variability with the calibration period is quantified in terms of the ratio S_t / S_{u_prior} (Figure 106) with S_t being calculated over all 5-year long calibration periods. Note that a single set of prior ranges is used for each catchment in all simulations with the semi-lumped BASIC version of the model. Therefore, S_{u_prior} calculated for one parameter (and for one catchment) holds for all calibration strategies.

The results reveal that parameter variability with calibration period exists regardless of the objective functions used. Differences in this variability among different calibration strategies are relatively small. The values of S_t / S_{u_prior} averaged over the entire set are ranked and presented in Table 27: lower ranks in imply lower variability with the calibration period. The results are catchment specific, but in general the R^2-*VE* strategy results in lower parameter variability while the *HMLE-RMSE* strategy yields higher parameter variability. The *NSE-NSE*$_{logQ}$ strategy and the combination of three objective functions also result in low parameter variability, whereas *NSE-VE* strategy results in high parameter sensitivity to calibration period.

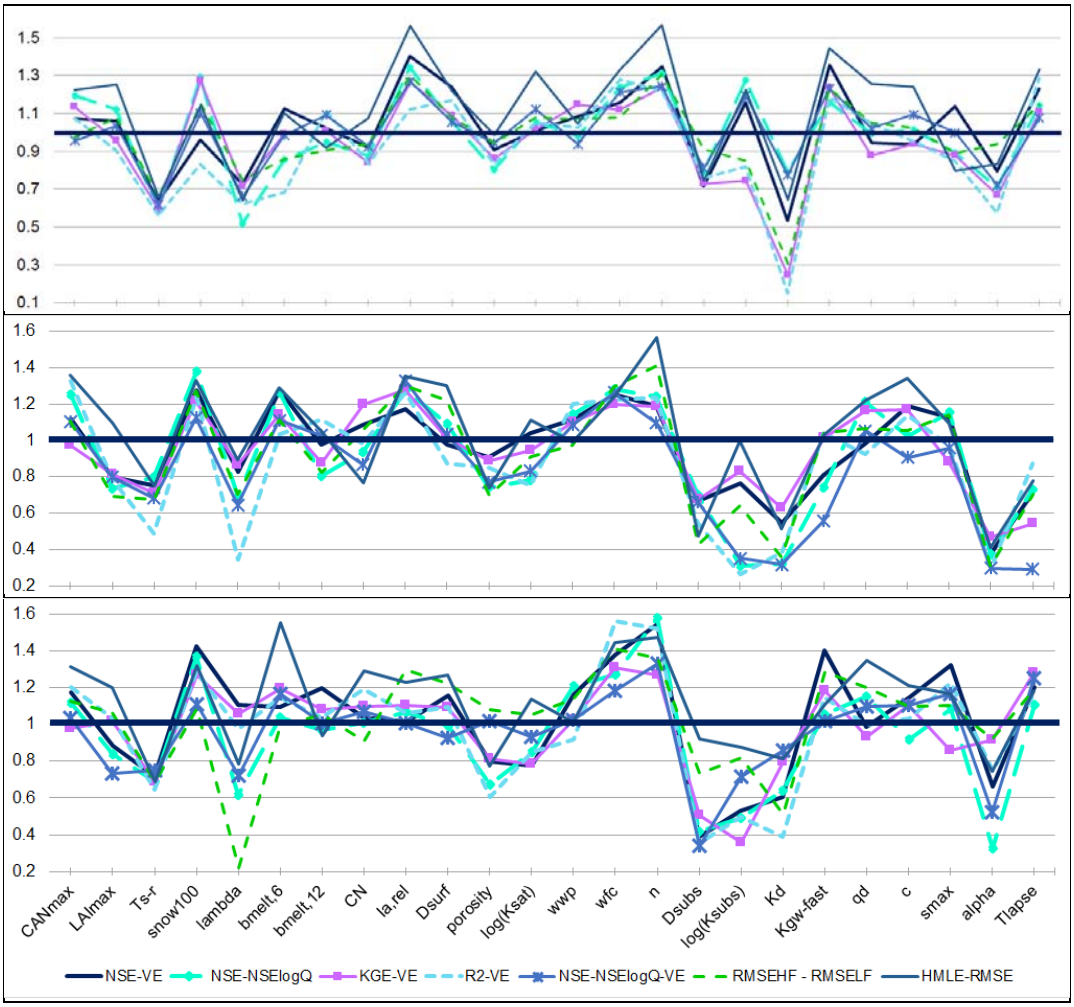


Figure 106. Ratio S_t / S_{u_prior} obtained from all 5-year long calibration periods: the Kolubara (top), Toplica (mid) and Mlava River catchments (bottom panel).

Table 27. Ranks of the average S_t / S_{u_prior} and mean *IC* statistic for different calibration strategies

Objective functions	Kolubara		Toplica		Mlava	
	S_t / S_{u_prior}	Mean *IC*	S_t / S_{u_prior}	Mean *IC*	S_t / S_{u_prior}	Mean *IC*
NSE-VE	6	5	6	4	6	3
NSE-NSE$_{logQ}$	4	2	3	3	1	4
KGE-VE	2	4	5	6	4	5
R^2*-VE*	1	3	2	2	3	2
RMSE$_{HF}$ - *RMSE*$_{LF}$	3	7	4	7	5	7
HMLE-RMSE	7	1	7	1	7	1
NSE-NSE$_{logQ}$*-VE*	5	6	1	5	2	6

Values of the *IC* statistic averaged over all 5-year long calibration periods are presented in Figure 107, while the *IC* ranks are given in Table 27 where lower ranks imply higher values of the *IC* statistics, i.e. narrower ranges of the Pareto-optimal parameters. The ranks are obtained from the *IC* values averaged over the parameter entire set. The *HMLE-RMSE* strategy results in rather narrow ranges of the optimised parameters (exceptionally high values of the *IC* statistic), as opposed to the *RMSE*$_{HF}$ -*RMSE*$_{LF}$ strategy. The R^2-*VE* and *NSE-NSE*$_{logQ}$ strategies also yield high *IC* values. Other calibration strategies result in similar ranks of *IC* values.

Despite resemblance among the lines in Figure 107 there are dissimilarities in the ranks of *IC* values for individual parameters. For example, highest *IC* value (highest parameter identifiability) is obtained for α with the *KGE-VE* strategy and for K_d with the *RMSE*$_{HF}$-*RMSE*$_{LF}$ strategy for the Kolubara catchment (top panel of Figure 107). This result confirm that different (combinations of) objective functions result in different parameter identifiability, i.e. with some strategies a parameter can be well identified, whereas some other strategies may result in wide posterior parameter distribution.

Model performance is evaluated for considered calibration strategies and the results are presented in Appendix L. Strategies *RMSE*$_{HF}$-*RMSE*$_{LF}$, *NSE-NSE*$_{logQ}$ and *NSE* - *NSE*$_{logQ}$ - *VE* result in the greatest number of non-dominated solutions, as opposed to *NSE-VE* or *KGE-VE*. Model efficiency quantified in terms of *NSE, VE* and ratios of *p*- to *r*-factor is presented in Figure 108. These performance measures are obtained by averaging the medians form all 5-year long calibration period. All strategies yield high values of *VE*, with *RMSE*$_{HF}$ -*RMSE*$_{LF}$ resulting in slightly lower *VE* value. Lower *NSE* values are

obtained from the model calibrated using the R^2-*VE* strategy, despite the fact that coefficient of determination is contained within *NSE* (Gupta, et al. 2009). Values of the *p*- and *r*-factors are low regardless of the calibration strategy. Greater values of the ratio of these factors indicate higher percentage of the observation encompassed by the simulated ensemble, given the same width of the ensemble band. Higher values of the ratio are obtained by the model calibrated using strategies *NSE* - $NSE_{\log Q}$, *NSE* - $NSE_{\log Q}$–*VE* and *HMLE-RMSE*. The *HMLE-RMSE* strategy results in exceptionally low values of the *r*-factor and consequently in high value of the ratio.

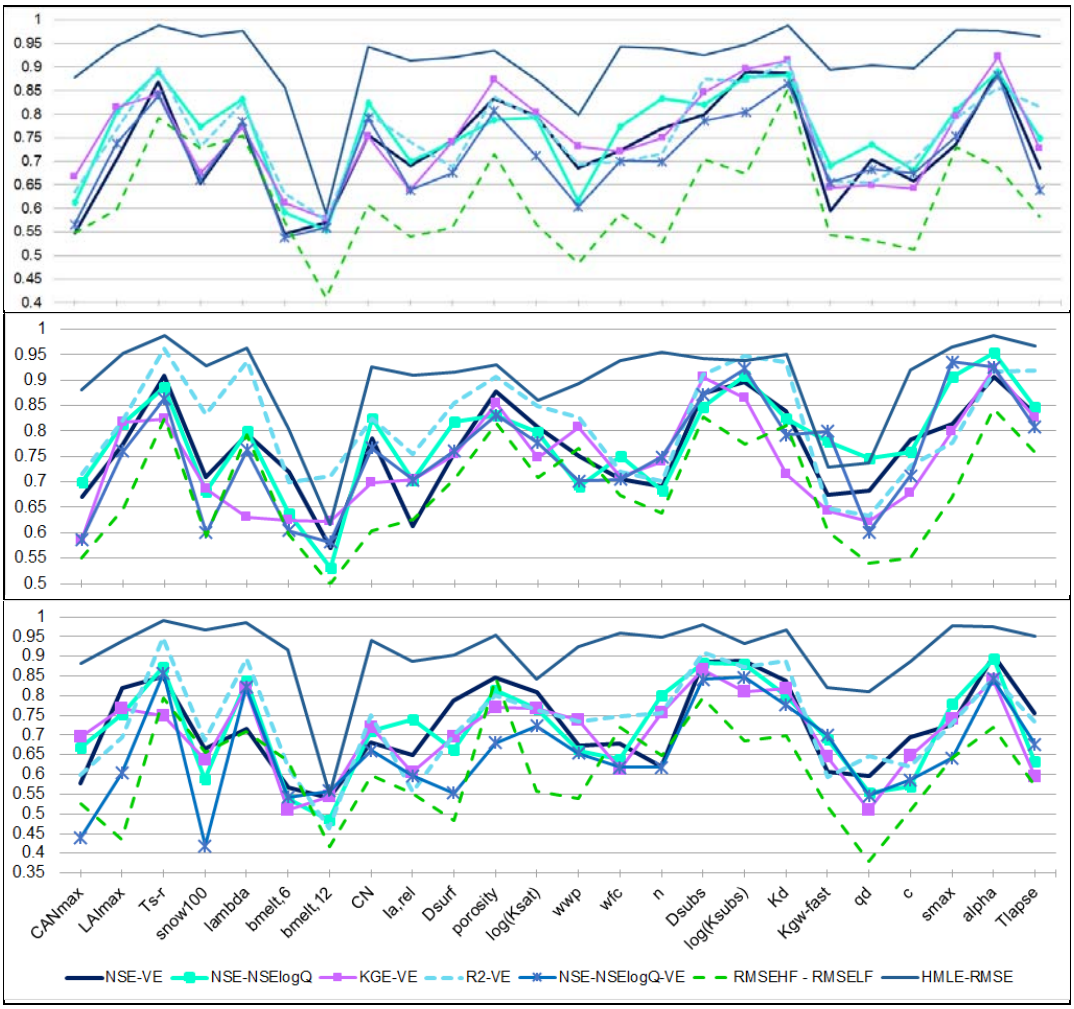


Figure 107. Values of the *IC* statistic of the Pareto-optimal parameters for different calibration strategies: the Kolubara (top), Toplica (mid) and the Mlava River catchments (bottom panel).

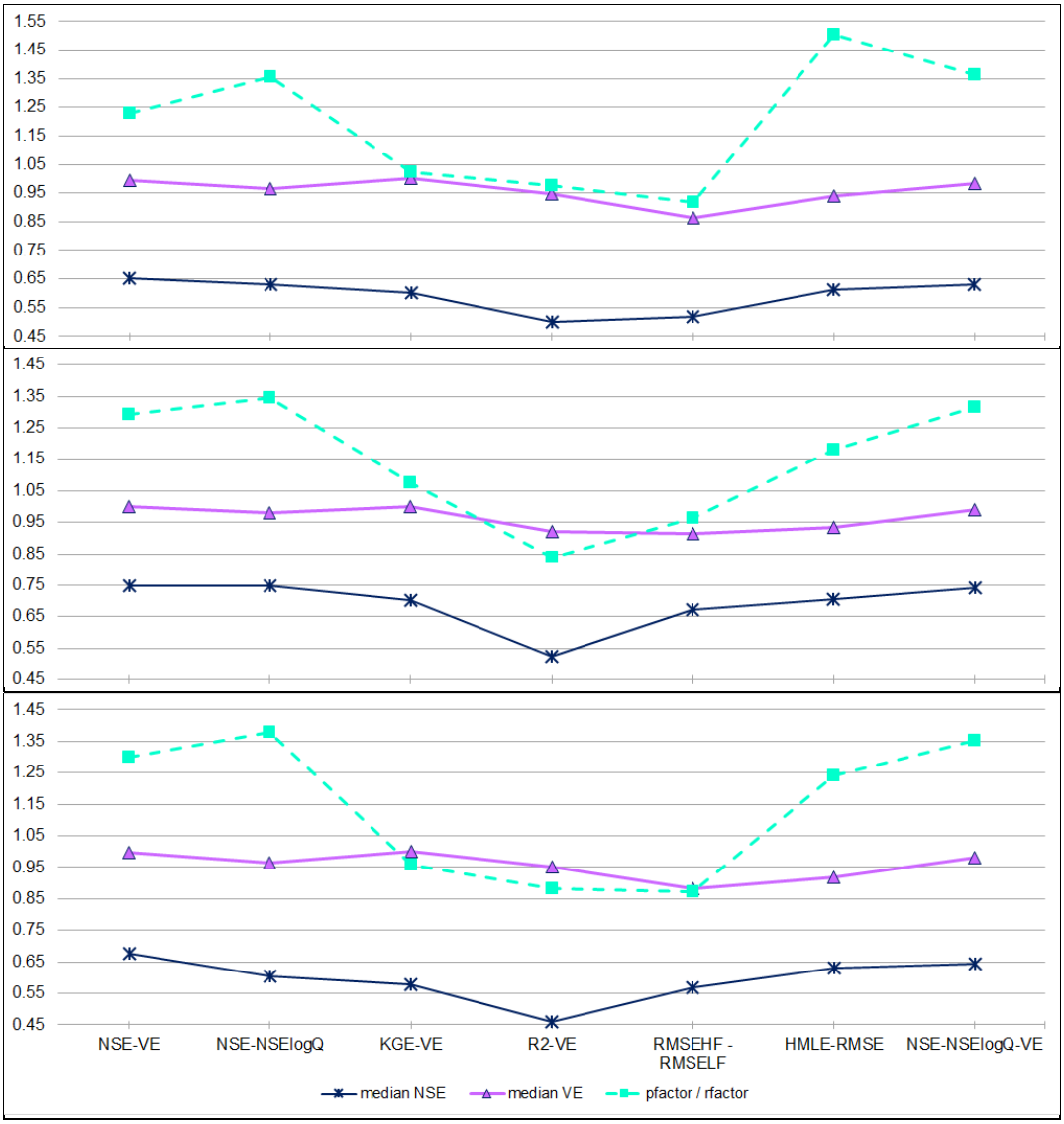


Figure 108. Model performance for different calibration strategies: the Kolubara (top), Toplica (mid) and the Mlava River catchments (bottom panel).

Generally, differences among different combinations of the objective functions in the calibration strategies are not pronounced. Considering both consistency in parameter estimates and model performance, strategies *NSE-VE*, *NSE - NSE_{logQ}*, *HMLE-RMSE* and *NSE - NSE_{logQ}–VE* seem to be somewhat advantage in comparison to other analysed strategies. Combination of three objective functions is quite appealing, not only because it results in satisfactory model performance and large number of Pareto sets, but also

because more aspects of the simulated hydrographs are taken into account with three objective functions.

3.2.4. Impact of the model structure

Parameter variability for different versions of the 3DNet-Catch model is quantified in terms of the ratio S_t / S_{u_prior} (Figure 109), where S_t is calculated based on all 5-year long calibration periods. Variability of the entire parameter set, quantified in terms of S_t / S_{u_prior} and averaged over the entire parameter set, is ranked and presented in Table 28. Lower ranks imply lower parameter variability with the calibration period.

These results reveal small differences in parameter variability for different model structures. The distributed model version results in somewhat more consistent parameter estimates, as opposed to the FULL version (that has the largest number of parameters with S_t / S_{u_prior} values greater than 1). This is detected for all catchments. The most sensitive parameters (α, $D_{sub\text{-}surface}$, *porosity*, K_d, LAI_{max}) tend to result in slightest variability in the BASIC and SIMPLE model versions. This means that inconsistency in parameter estimates may be increased by model overparameterisation. Sensitivity of the spatially distributed parameters to calibration period is generally lower than of the corresponding parameters of the BASIC model version, with exception of the LAI_{max} parameter.

Values of the *IC* statistic are presented in Figure 110 and ranks of the *IC* statistic averaged over the entire set are given in Table 28 . Lower ranks denote higher *IC* values.

The SIMPLE model version yields narrow ranges of the Pareto-optimal parameters over a calibration period and the distributed model version yields the widest ranges. However, identifiability of the spatially distributed parameters tends to be better than the lumped ones. For example, CAN_{max} and α are better identified for all catchments; however, LAI_{max} is better identified in the BASIC model version. Interestingly, for the Toplica and Mlava

catchments the distributed model version results in the widest range of the parameter $D_{sub\text{-}surface}$, which is quite a sensitive model parameter[22].

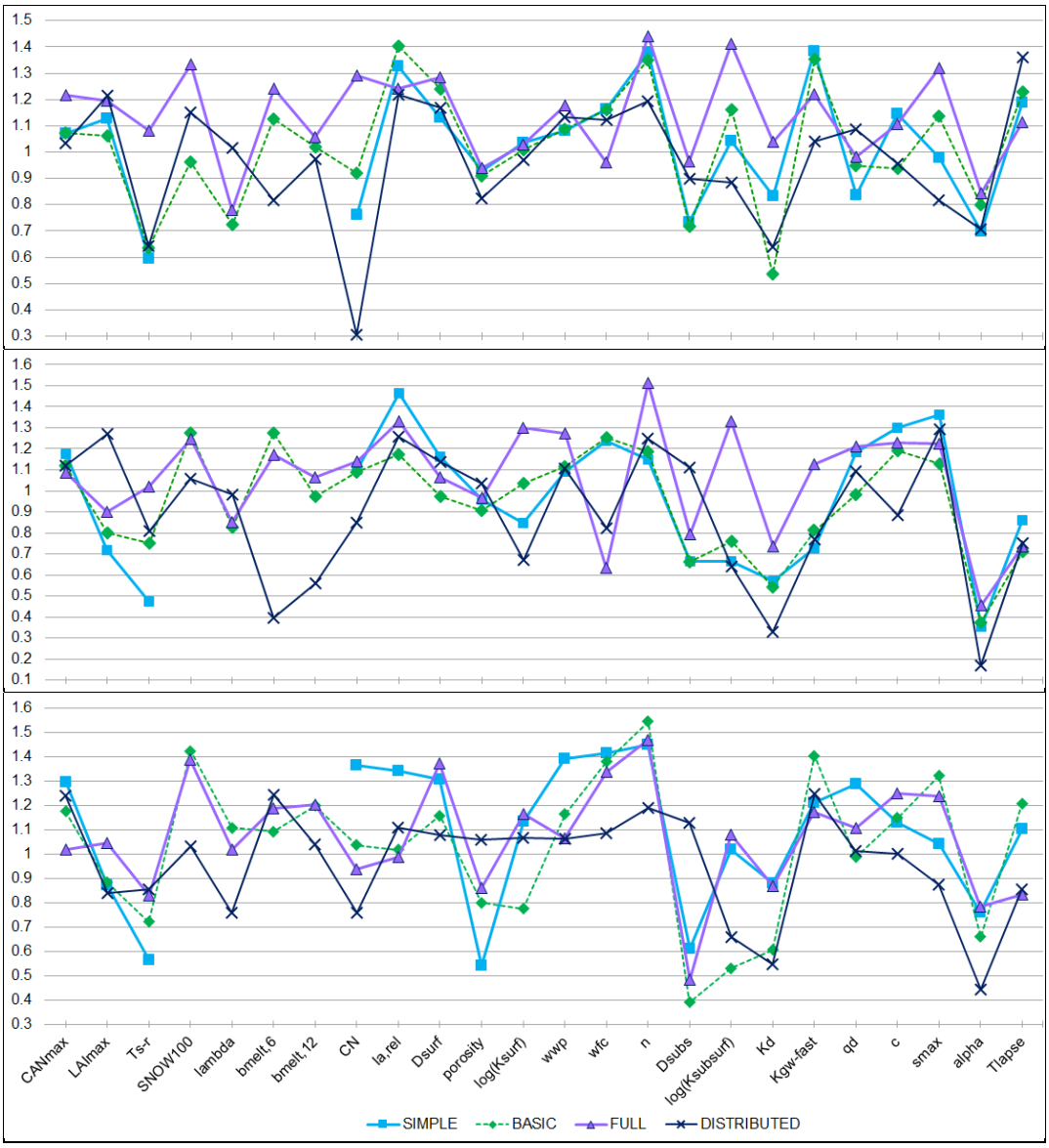


Figure 109. Ratios S_t / S_{u_prior} obtained over 5-year long calibration periods for different versions of the 3DNet-Catch model: the Kolubara (top panel), the Toplica (mid panel) and the Mlava River catchment (bottom panel).

[22] Sensitivity analysis of the SIMPLE, FULL and the distributed model versions was not conducted. It was assumed that parameter sensitivity of the BASIC model holds for corresponding parameters of the other model versions.

Table 28. Ranks of the average S_t / S_{u_prior} and average *IC* statistic for different versions of the 3DNet-Catch model

Version of the model	Kolubara		Toplica		Mlava	
	S_t / S_{u_prior}	Mean *IC*	S_t / S_{u_prior}	Mean *IC*	S_t / S_{u_prior} θ	Mean *IC*
SIMPLE	3	2	2	1	4	1
BASIC	2	4	3	2	2	3
FULL	4	1	4	3	3	2
DISTRIBUTED	1	3	1	4	1	4

Efficiency of these model versions is given in Appendix L. The median *NSE* values, median *VE* values and ratios of *p*- to *r*-factors, all of which are averaged over all 5-year long calibration periods, are presented in Figure 111. Overall, differences between the model versions are minor. However, the distributed model version is the most efficient for two catchments. Higher values of ratios between mean *p*- and *r*-factors are obtained by the SIMPLE and BASIC model version, although all versions result in very small values of both statistics (the SIMPLE model version results in the smallest values of *r*-factor). The Pareto fronts obtained over one calibration period for the Toplica River catchment are illustrated in Figure 112. These results confirm that a more complex model version does not necessarily result in better model performance.

Despite being catchment specific, the results of this analysis point to the following:

– Model overparameterisation (large number of free parameters) may lead to higher sensitivity of the model parameters to the calibration period, while spatial distribution of the parameters could contribute to more consistent parameter estimates.

– None of the structures is proven to be superior in terms of model efficiency. This implies that simpler model structures may perform quite satisfactory (Figure 112). In addition, the distributed version of the model may (slightly) outperform the (semi-)lumped one. Therefore application of distributed models appears to be advantageous over application of the lumped, heavily parameterised models. Model simplification by reducing the number of free parameters has to be supported by the results of a sensitivity analysis.

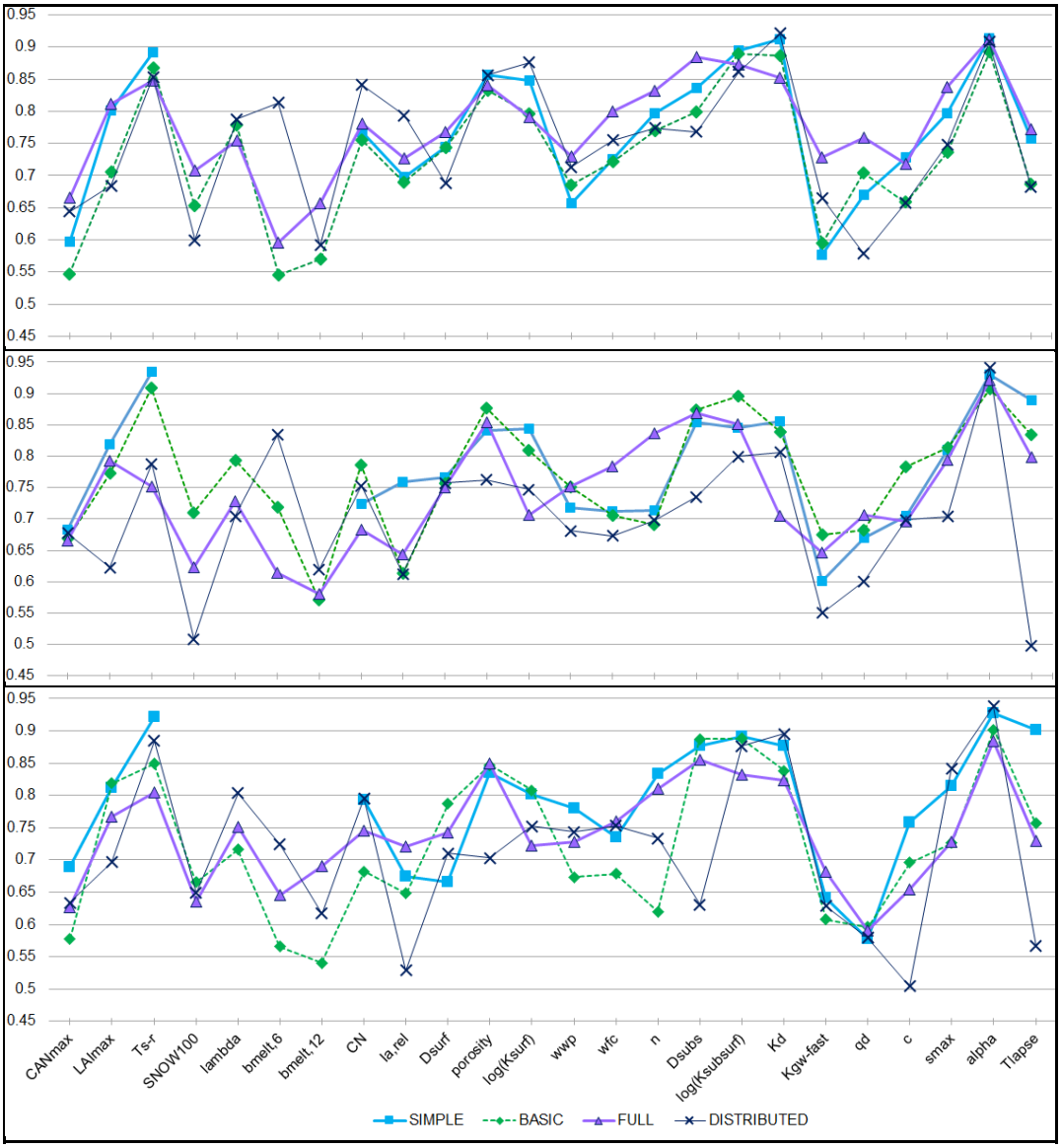


Figure 110. Mean *IC* values of the Pareto-optimal parameters obtained over 5-year long calibration periods for different versions of the 3DNet-Catch model: the Kolubara (top panel), the Toplica (mid panel) and the Mlava River catchments (bottom panel).

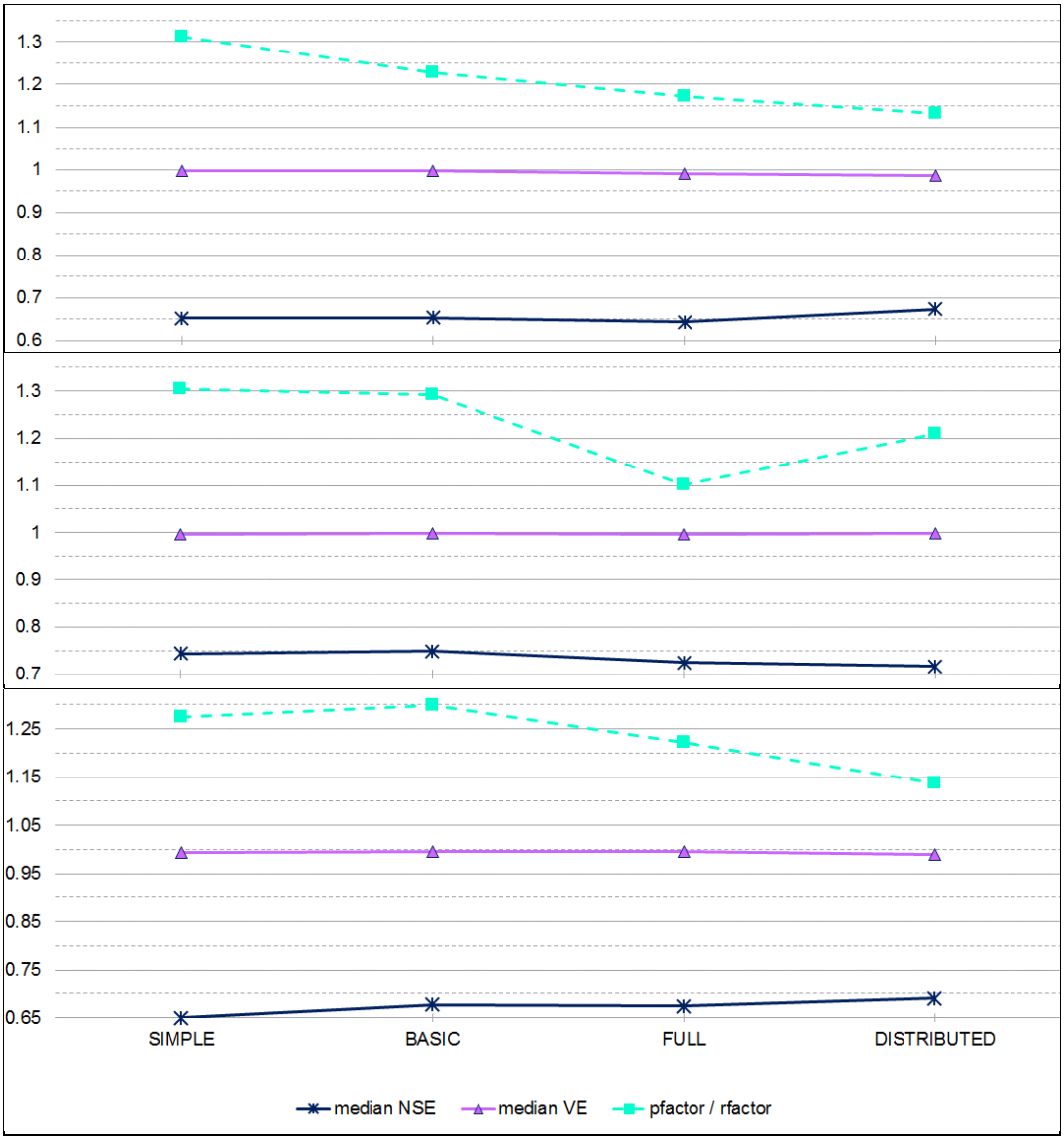


Figure 111. Model performance for different model structure: the Kolubara (top), Toplica (mid) and the Mlava River catchment (bottom panel).

Although the results presented suggest application of the distributed models, it should be noted that application of these models is conditioned on the available data: finer spatial resolution should be accompanied by finer temporal resolution (e.g. runoff modelling on the 1 m by 1 m grid and with daily data has no sense). Also, transformation of spatial data into spatial distribution of the model parameters inevitably involves subjectivity.

It should be emphasized that the distributed model version is applied in this research with rather coarse spatial resolution and a limited number of spatially distributed parameters. Further research and data with finer resolution are required to assess robustness of the distributed models more accurately.

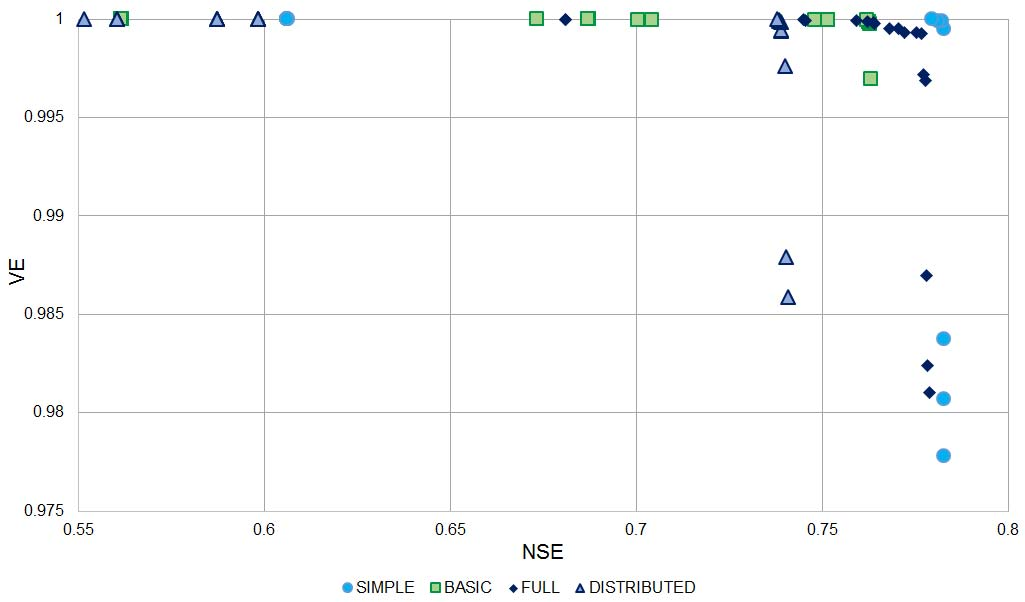


Figure 112. Pareto fronts of the SIMPLE, BASIC, FULL and distributed versions of the model obtained by calibration in 2007-2012 for the Toplica River catchment.

4. CONCLUDING REMARKS

Conceptual hydrologic models are used for flow forecasting, estimation of design flows or assessment of the climate change impact on water resources. Therefore, reliability of flows simulated by these models is crucial. Prior to application of a conceptual hydrologic model, its parameters have to be inferred through model calibration, and the model should be applied for runoff simulations over an independent period in order to be evaluated. The obtained parameter values depend on the method for parameter estimation (e.g. optimisation algorithm), objective function(s), predefined parameter ranges, data quality and properties of the period the parameters are being estimated over. Once calibrated, the hydrologic models are assumed to be capable of reproducing catchment behaviour in any period. However, a decrease in model performance and consequently in reliability of the simulated variables outside the calibration period has been repeatedly reported in the literature (e.g. presence of bias in simulated runoff volume, weaker correlation between the simulated and observed flows). This is a major issue when it comes to application of hydrologic models, because these models are primarily used for runoff simulations outside calibration period.

Reasons for such model behaviour are numerous, such as: model overfitting to noisy data in the calibration period, ineffective parameter optimisation, model structural inadequacies or non-stationarity in catchment response. A decrease in model performance may also be due to the fact that optimality of model parameters does not hold outside the calibration period. Bearing in mind wide practical application of the conceptual hydrologic models, change of optimal parameter sets in time (with the calibration period) needs to be explored.

To this end, in this Thesis a novel fully-distributed hydrologic model, entitled 3DNet-Catch, is calibrated over moving 1-year to 25-year long calibration periods ("bootstrapping" of the calibration periods) within a multi-objective framework employing the AMALGAM optimisation algorithm. Each calibration period is shifted by one water year from the previous period, resulting in the overlap between the consecutive periods. The analysis of the parameter variability in time is based on the Pareto-optimal parameters obtained over different calibration periods. The analysis is performed using

the observations from three relatively unchanged catchments in Serbia: the Kolubara, the Toplica and the Mlava River catchments. In this way the effects of the human-induced changes on parameter variability are excluded.

For purposes of this research, four alternative model structures were developed: three semi-lumped (SIMPLE, BASIC and FULL) and the distributed version, which is based on the BASIC model version. However, majority of the simulations are carried out with the BASIC model version.

Prior to the dynamic model calibration, the regression-based parameter sensitivity analysis is carried out. Goal of the sensitivity analysis is to detect the model parameters that are important for reproducing the overall water balance and dynamics of catchment response. The results indicate that the precipitation gradient with elevation is the most sensitive parameter of the 3DNet-Catch model for these catchments. Some soil-related parameters, such as porosity and subsurface soil layer thickness, are very important for maintaining water balance, while the linear reservoir coefficient for surface runoff routing K_d is important for reproducing the catchment dynamics. The parameters of the snow routine and the baseflow-related parameters are less sensitive for these catchments.

The 3DNet-Catch model is evaluated by conducting the split sample test and differential split sample test, and also by using the HVB model as the reference model. The results confirmed that the 3DNet-Catch can simulate the overall water balance very well, although some bias may be expected if the model is evaluated over a climatically contrasted period. The results also confirm that semi-lumped BASIC version of the 3DNet-Catch is comparable to the HBV model in terms of reproducing the overall water balance and catchment dynamics.

The simulated hydrographs and the estimated Pareto-optimal parameters confirm the robustness of the AMALGAM algorithm. Namely, the parameter sets converge to rather small optimal region of the response surface, resulting in narrow ranges of the Pareto-optimal parameters and consequently narrow prediction band with the Pareto-optimal ensemble.

Temporal variability of the model parameters and model performance. The results of the dynamic model calibration reveal a surprisingly high temporal variability of the Pareto-optimal parameters, which does not follow any clear pattern. This variability

appears to be smaller for the sensitive model parameters, although some parameters of the snow routine that are shown insensitive, also exhibit low variability with the calibration period. The variability in some parameters slightly decreases with the length of the calibration period, but no apparent pattern emerges. For example, a considerable variability can be detected even among the consecutive overlapping 25-year long calibration periods in some mainly insensitive parameters.

Based on all Pareto solutions obtained over different calibration periods, empirical cumulative distribution functions (*ecdf*s) of the model parameters are derived. The *ecdf*s of some insensitive parameters resemble the uniform distribution, thus indicating a great parameter uncertainty due to calibration period. However, the *ecdf*s of most parameters deviate from the uniform distribution, which means that the initial uncertainty represented by wide predefined parameter ranges is reduced. Interestingly, these *ecdf*s of all Pareto-optimal parameter values from different calibration periods may also deviate from the *ecdf*s of the Pareto sets obtained over the full hydrologic record period ($ecdf_{FRP}$). The parameters sampled from the *ecdf*s are expected to outperform the sets sampled from $ecdf_{FRP}$ over an independent period. Namely, $ecdf_{FRP}$ is expected to underestimate uncertainty due to aggregating the model performance over long calibration period into a few objective functions and thereby losing a significant amount of information, while dynamic model calibration enables extraction of more information form the observations available. However, this is beyond the scope of this Thesis and further research is required to test this hypothesis. If this hypothesis would turn out plausible, the multi-temporal dynamic calibration could be used to obtain more robust parameter estimates and reliable simulations over an independent period.

Parameter identifiability, which is rather high for the many model parameters (i.e. narrow optimal ranges), also varies in time regardless of the calibration period and without following any regular pattern.

Model performance also varies in time. Considering the model ability to reproduce overall water balance, rather high values of *VE* and a negligible flow bias are obtained over all calibration periods, regardless of their length. Interestingly, difference between the observed and simulated runoff is approximately constant in the calibration periods of 20 years or longer. Concerning *NSE*, it decreases almost exponentially with the

calibration period length, although a slight increase may be detected over the calibration periods of 20 years or longer. Such a behaviour is also detected in *KGE* and R^2 evaluation measures. This demonstrates a non-linearity in catchment behaviour that cannot be properly reproduced by the model with temporally invariant parameter sets, i.e. single parameter set cannot simulate such a wide range of catchment responses.

Temporal variability of the model parameters and hydro-meteorological characteristics. Attempts to correlate the parameter variability and the hydro-meteorological characteristics of the calibration period are made in this research. However, weak and mainly statistically insignificant correlations are obtained with contrasted results among the catchments. Such the results suggest that relationship between hydro-meteorological conditions in a catchment and parameter estimates is not a straightforward one, and that the values of the Pareto-optimal parameters cannot be conditioned on the indices analysed in this research.

Impact of the calibration strategy (combination of objective functions) on temporal parameter variability. Another goal of this study was to analyse whether more consistent parameter estimates may be obtained by using certain combinations of the objective functions (i.e. different calibration strategies). The differences in the results among 7 calibration strategies analysed are small. Calibration strategy with coefficient of determination consistently yielded lower parameter temporal variability, but poorer model performance. Considering various aspects of the results altogether suggests that application of calibration strategy that involves more objective functions may result in more consistent parameter estimates and higher model efficiency.

Impact of the model structure on temporal parameter variability. None of the model structures is found superior because the differences among the model versions are slight, especially in terms of model performance. Application of the overparameterised FULL model version results in increased parameter variability with calibration period without any improvement in model performance. The distributed model version yielded the most consistent parameter estimates, although the BASIC and SIMPLE model versions

resulted in lowest variability in the most sensitive model parameters. Simpler model structures perform satisfactory, although reduction the number of free parameters has to be supported by the results of the sensitivity analysis. The results of this analysis encourage application of models with fewer but spatially distributed parameters.

Application of the methodology and further research. The methodology reported in this Thesis could be used for studies of model structure evaluation, because more suitable structures for a particular catchment are assumed to yield less variable parameters. Consequently, the results of such an analysis may be used for model improvement: high variability in parameters of a model routine indicate that the routine should be reparametrised. Evaluation of already calibrated models over the periods selected according to the methodology presented in this Thesis would be thorough and more robust than commonly conducted evaluation with the *SST*. Also, calibration strategy may be tested in this manner: the strategies that consistently yield high model performance and / or consistent parameter estimates should be preferred.

The empirical cumulative distribution functions derived from the Pareto parameter sets obtained over all calibration periods are assumed to contain more information regarding catchment response; hence, they are expected to yield more reliable simulations over an independent period. Such an approach to model calibration is to be tested in further research. Additionally, this methodology should be applied with data of finer temporal and spatial resolution (preferably data form experimental catchments) which would certainly yield more reliable results and perhaps reveal some patterns in temporal variability of the Pareto-optimal parameters. The distributed models should be applied with more spatially distributed parameters and, considering that precipitation data are crucial for model performance, with more methods for spatial interpolation of precipitation. It is important to emphasise that parameter variability was assessed relative to the initial parameter uncertainty, which was quite large. Further research is required to examine whether narrower prior ranges of the parameters could yield more consistent parameter estimates, and how experience of a hydrologist, reflected by the prior parameter distribution, affects parameter temporal variability. In this research, the models are calibrated solely against the observed flows: inclusion of different variables in the

model calibration (i.e. multi-variable and / or multi-site calibration) is required to test the hypothesis that utilisation of more objective functions yields more consistent parameter estimates. Further research is also needed to examine the parameter variability on the sub-annual time scales. Dynamic parameter estimation presented in this research should be applied with other methods for parameter estimation and uncertainty assessment, and also with other hydrologic models.

REFERENCES

Abebe, N. A., Ogden, F. L., & Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: Implications for parameter estimation. *Journal of Hydrology*, *389*(3-4), 301–310. doi:10.1016/j.jhydrol.2010.06.007

Ajami, N., Duan, Q., Gao, X., & Sorooshian, S. (2006). Multimodel Combination Techniques for Analysis of Hydrological Simulations : Application to Distributed Model Intercomparison Project Results. *Journal of Hydrometeorology*, *7*(4), 755–768. doi:http://dx.doi.org/10.1175/JHM519.1

Anderson, E. (2006). *Snow Accumulation and Ablation Model – SNOW-17*. Retrieved from http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.-H., Oudin, L., Mathevet, T., … Berthet, L. (2012). All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes*, *26*(14), 2206–2210. doi:10.1002/hyp.9264

Asner, G. P., Scurlock, J. M. O., & Hicke, J. A. (2008). Global synthesis of leaf area index observations : implications for ecological and remote sensing studies. *Global Ecology and Biogeography*, *12*(2003), 191–205.

Bardossy, A., & Das, T. (2008). Influence of rainfall observation network on model calibration and application. *Hydrology and Earth System Sciences*, *12*, 77–89.

Bastola, S., Murphy, C., & Sweeney, J. (2011). The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Advances in Water Resources*, *34*(5), 562–576. doi:10.1016/j.advwatres.2011.01.008

Bergström, S., Harlin, J., & Lindström, G. (1992). Spillway design floods in Sweden: I. New guidelines. *Hydrological Sciences Journal*, *37*(5), 505–519. doi:10.1080/02626669209492615

Berthet, L., Andreassian, V., Perrin, C., & Javelle, P. (2009). How crucial is it to account for the antecedent moisture conditions in flood forecasting ? Comparison of event-based and continuous approaches on 178 catchments. *Hydrology and Earth System Sciences*, *13*(6), 819–831.

Beven, K. (1982). On subsurface stormf low: an analysis of response times. *Hydrological Sciences - Journal Des Sciences Hydrologiques*, *4*, 505–521. Retrieved from https://www.itia.ntua.gr/hsj/27/hysj_27_04_0505.pdf

Beven, K. (2001a). How far can we go in distributed hydrological modelling ? *Hydrology and Earth System Sciences*, *5*(1), 1–12.

Beven, K. (2001b). *Rainfall-Runoff Modelling - The Primer*. West Sussex: John Wiley & Sons.

Beven, K. (2006). A Manifesto for the Equifinality Thesis. *Journal of Hydrology*, *320*, 18–36.

Beven, K. (2009). *Environmental Modelling: An Uncertain Future?* New York: Routledge.

Beven, K., & Binly, A. (1992). The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes*, *6*(3), 279–298.

Beven, K. J. (2005). Rainfall-runoff Modeling: Introduction. In M. Anderson & J. McDonnell (Eds.), *Encyclopedia of Hydrological Sciences*. John Wiley & Sons Ltd.

Beven, K., & Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. *Water Resources Research*, *49*, 1–7. doi:10.1002/wrcr.20393

Blasone, R.-S. (2007). *Parameter Estimation and Uncertainty Assessment in Hydrological Modelling*. Technical University of Denmark.

Blasone, R.-S., Madsen, H., & Rosbjerg, D. (2007). Parameter estimation in distributed hydrological modelling: comparison of global and local optimisation techniques. *Nordic Hydrology*, *38*(4-5), 451. doi:10.2166/nh.2007.024

Boughton, W., & Droop, O. (2003). Continuous simulation for design flood estimation—a review. *Environmental Modelling & Software*, *18*(4), 309–318. doi:10.1016/S1364-8152(03)00004-5

Box, G. E. P., & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco.: Holden- Day.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *36*(12), 3663–3674. doi:10.1029/2000WR900207

Braun, L. N., Grabs, W., & Rana, B. (1993). Application of a conceptual precipitation-runoff model in the Langtang Khola basin, Nepal Himalaya. In G. J. Young (Ed.), *Snow and Glacier Hydrology, Kathmandu Symposium 1992* (pp. 221–237). IAHS Publ. no. 218.

Breuer, L., Eckhardt, K., & Frede, H.-G. (2003). Plant parameter values for models in temperate climates. *Ecological Modelling*, *169*(2-3), 237–293. doi:10.1016/S0304-3800(03)00274-6

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

Campling, P., Gobin, A., Beven, K., & Feyen, J. (2002). Rainfall-runoff modelling of a humid tropical catchment: the TOPMODEL approach. *Hydrological Processes*, *16*(2), 231–253. doi:10.1002/hyp.341

Choi, H. T., & Beven, K. (2007). Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology*, *332*(3-4), 316–336. doi:10.1016/j.jhydrol.2006.07.012

Chow, V. Te, Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill Book Company.

Christiaens, K., & Feyen, J. (2002). Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model. *Water Resources Research*, *38*(9), WR000478, 8–1 – 8–15. doi:10.1029/2001WR000478

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. a., Vrugt, J. a., Gupta, H. V., … Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*, W0B02, 1–14. doi:10.1029/2007WR006735

CORINE 2006. (n.d.). Retrieved from http://www.eea.europa.eu/data-and-maps/figures/corine-land-cover-types-2006

Cornell University. (2010). Soil and Water Management. Retrieved from http://nrcca.cals.cornell.edu/soil/CA2/CA0212.1-3.php

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research*, *48*(5), W05552, 1–17. doi:10.1029/2011WR011721

Coron, L., V., A., C., P., M., B., & F., H. (2014). On the lack of robustness of hydrologic models regarding water balance simulation : a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. *Hydrology and Earth System Sciences*, *18*(2), 727–746. doi:10.5194/hess-18-727-2014

Criss, R. E., & Winston, W. E. (2008). Do Nash values have value ? Discussion and alternate proposals. *Hydrological Processes*, *22*, 2723–2725. doi:10.1002/hyp

Daniel, E. B., Camp, J. V, Leboeuf, E. J., Penrod, J. R., Dobbins, J. P., & Abkowitz, M. D. (2011). Watershed Modeling and its Applications : A State-of-the-Art Review. *The Open Hydrology Journal*, *5*, 26–50. Retrieved from http://benthamopen.com/tohydj/articles/V005/26TOHYDJ.pdf

De Vos, N. J., Rientjes, T. H. M., & Gupta, H. V. (2010). Diagnostic evaluation of conceptual rainfall – runoff models using temporal clustering. *Hydrological Processes*, *24*(20), 2840–2850. doi:10.1002/hyp.7698

Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., … Tait, S. (2012). Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth*, *42-44*, 3–10. doi:10.1016/j.pce.2011.04.007

Diallo, D., & Mariko, A. (2013). Field capacity (FC) and permanent wilty point (PWP) of clay soils developed on Quaternary alluvium in Niger River loop (Mali). *International Journal of Engineering Research and Applications (IJERA)*, *3*(1), 1085–1089.

Dimitrijević, M., Karamata, S., Sikošek, B., & Veselinović, D. (1975). *Osnovna Geološka karta 1:100000*. Beograd.

Djorković, M. (1984). Odredjivanje hidrološke grupe zemljišta pri definisanju oticanja u metodi SCS. *Vodoprivreda*, *87*(1), 57–60.

Dotto, C. B. S., Mannina, G., Kleidorfer, M., Vezzaro, L., Henrichs, M., McCarthy, D. T., … Deletic, A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*, *46*(8), 2545–2558. doi:10.1016/j.watres.2012.02.009

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

Durner, W., & Fluhler, H. (2005). Soil Hydraulic Proprties. In M. G. . Anderson & J. J. McDonnell (Eds.), *Encyclopedia of Hydrological Sciences* (pp. 1103–1119). John Wiley & Sons Ltd.

Ebel, B. A., & Loague, K. (2006). Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrological Processes*, *20*(13), 2887–2900. doi:10.1002/hyp.6388

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multiobjective calibration approaches in hydrological modelling : A review. *Hydrological Sciences Journal*, *55*(1), 58–78. doi:10.1080/02626660903526292

Efstratiadis, A., Nalbantis, I., & Koutsoyiannis, D. (2014). Hydrological modelling of temporally-varying catchments: facets of change and the value of information. *Hydrological Sciences Journal*, 1–32. doi:10.1080/02626667.2014.982123

Engeland, K., Braud, I., Gottschalk, L., & Leblois, E. (2006). Multi-objective regional modelling. *Journal of Hydrology*, *327*, 339–351. doi:10.1016/j.jhydrol.2005.11.022

Eschenbach, C., & Kappen, L. (1996). Leaf area index determination in an alder forest : a comparison of three methods. *Journal of Experimental Botany*, *47*(302), 1457–1462.

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. doi:10.5194/hess-17-1893-2013

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*(9), 1727–1737. doi:10.5194/hess-13-1727-2009

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resources Research*, *44*(1), W01402, 1–13. doi:10.1029/2006WR005563

Fenicia, F., Solomatine, D. P., Savenije, H. H. G., & Matgen, P. (2007). Soft combination of local models in a multi-objective framework. *Hydrology and Earth System Sciences*, *11*(6), 275–301.

Foglia, L., Hill, M. C., Mehl, S. W., & Burlando, P. (2009). Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resources Research*, *45*, 1–18. doi:10.1029/2008WR007255

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods*. (J. Chambers, W. Eddy, W. Hardle, S. Sheather, & L. Tiemey, Eds.) (Second Edi.). Springer.

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2012). Moving beyond traditional model calibration or how to better identify realistic model parameters: sub-period calibration. *Hydrology and Earth System Sciences Discussions*, *9*(2), 1885–1918. doi:10.5194/hessd-9-1885-2012

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences*, *17*(1), 149–161. doi:10.5194/hess-17-149-2013

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763.

Gupta, H. V., Beven, K. J., & Wagener, T. (2005). Model Calibration and Uncertainty Estimation. In M. Anderson (Ed.), *Encyclopedia of Hydrological Sciences* (pp. 1–20). John Wiley & Sons Ltd.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. doi:10.1016/j.jhydrol.2009.08.003

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernouli*, *7*(2), 223–242. doi:10.2307/3318737

Hamon, W. R. (1961). Estimating potential evaporation. In *Proceedings of the American Society of Civil Engineers, Division, J.o.H.* (pp. 107–120).

Hannaford, J., Buys, G., Stahl, K., & Tallaksen, L. M. (2013). The influence of decadal-scale variability on trends in long European streamflow records. *Hydrology and Earth System Sciences*, *17*(7), 2717–2733. doi:10.5194/hess-17-2717-2013

Hartmann, G., & Bardossy, A. (2005). Investigation of the transferability of hydrological models and a method to improve model calibration. *Advances in Geosciences*, *5*, 83–87.

He, Z. H., Parajka, J., Tian, F. Q., & Blöschl, G. (2014). Estimating degree-day factors from MODIS for snowmelt runoff modeling. *Hydrology and Earth System Sciences*, *18*(12), 4773–4789. doi:10.5194/hess-18-4773-2014

Hirsch, C. (2007). *Numerical Computation of Internal and External Flows* (second edi.). John Wiley & Sons Ltd.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*(1-4), 104–115. doi:10.1016/S0022-1694(03)00257-9

Hsu, K., Moradkhani, H., & Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. *Water Resources Research*, *45*(12), n/a–n/a. doi:10.1029/2008WR006824

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1-4), 281–295. doi:10.1016/j.jhydrol.2004.01.002

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., & Entekhabi, D. (2004). Catchment hydrologic response with a fully distributed triangulated irregular network model. *Water Resources Research*, *40*, W11102, 1–23. doi:10.1029/2004WR003218

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

Jin, X., Xu, C.-Y., Zhang, Q., & Singh, V. P. (2010). Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology*, *383*(3-4), 147–155. doi:10.1016/j.jhydrol.2009.12.028

Jovanović, M., Pavić, D., Mesaroš, M., Tankov, U., Pantelić, M., Armenski, T., … Crnojević, V. (2013). Water shortage and drought monitoring in Bačka region (Vojvodina , North Serbia) – setting-up measurement stations network. *Geographica Pannonica*, *17*(4), 114–124.

Jovanović, S., & Radić, Z. (1990). *Parametarska hidrologija*. Građevinski fakultet u Beogradu.

Juston, J., Seibert, J., & Johansson, P. (2009). Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. *Hydrological Processes*, *3109*(September), 3093–3109. doi:10.1002/hyp

K. 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. doi:10.1016/j.jhydrol.2004.03.033

Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. *Water Resources Research*, *46*(10), W10511, 1–27. doi:10.1029/2009WR008896

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*, *320*(1-2), 173–186. doi:10.1016/j.jhydrol.2005.07.012

Keramat, M., & Kielbasa, R. (1997). Latin Hypercube Sampling Monte Carlo Estimation of Average Quality Index for Integrated Circuits. *Analog Integrated Circuits And Signal Processing*, *1-2*, 131–142.

Khakbaz, B., Imam, B., Hsu, K., & Sorooshian, S. (2012). From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. *Journal of Hydrology*, *418-419*, 61–77. doi:10.1016/j.jhydrol.2009.02.021

Kim, U., & Kaluarachchi, J. J. (2010). Hydrologic model calibration using discontinuous data : an example from the upper Blue Nile River Basin of Ethiopia. *Hydrological Processes*, *3717*(September 2009), 3705–3717. doi:10.1002/hyp

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

Kohler, M. A., & Linsley, R. K. (1951). *Predicting the runoff from storm event*. Washington, DC. Retrieved from http://docs.lib.noaa.gov/rescue/wb_researchpapers/QC852U55no34.pdf

Kottegoda, N. T., & Rosso, R. (2008). *Applied Statistics for Civil and Environmental Engineers* (Second Edi.). Oxford, United Kingdom: Blackwell Publishing.

Krause, P., Boyle, D. P., & Base, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97.

Krauße, T., Cullmann, J., Saile, P., & Schmitz, G. H. (2012). Robust multi-objective calibration strategies – possibilities for improving flood forecasting. *Hydrology and Earth System Sciences*, *16*(10), 3579–3606. doi:10.5194/hess-16-3579-2012

Kuczera, G., Kavetski, D., Franks, S., & Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, *331*(1-2), 161–177. doi:10.1016/j.jhydrol.2006.05.010

Kuczera, G., & Parent, E. (1998). Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, *211*(1-4), 69–85. doi:10.1016/S0022-1694(98)00198-X

Langsholt, E., Lawrence, D., Wong, W. K., Andjelic, M., Ivkovic, M., & Vujadinovic, M. (2013). *Effects of climate change in the Kolubara and Toplica catchments, Serbia*. Oslo, Norway.

Laura, M. I., Keri, A. a., & Rusu, T. (2011). Soil Conservation Service Curve Number Method for Surface Runoff Estimation Using GIS Techniques , in Rosia Poieni Mining Area (Romania). *ProEnvironment*, *4*, 240–246.

Le Lay, M., Galle, S., Saulnier, G. M., & Braud, I. (2007). Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa. *Water Resources Research*, *43*(7), W07420, 1–10. doi:10.1029/2006WR005257

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. doi:10.1029/1998WR900018

Leibundgut, C., Uhlenbrook, S., & McDonnell, J. (Eds.). (2001). Runoff Generation and Implications for River Basin Modelling. In *Runoff Generation and Implications for River Basin Modelling*. Institut für Hydrologie, Universität Freiburg.

Li, C. Z., Zhang, L., Wang, H., Zhang, Y. Q., Yu, F. L., & Yan, D. H. (2012). The transferability of hydrological models under nonstationary climatic conditions. *Hydrology and Earth System Sciences*, *16*(4), 1239–1254. doi:10.5194/hess-16-1239-2012

Li, H., Beldring, S., & Xu, C.-Y. (2014). Stability of model performance and parameter values on two catchments facing changes in climatic conditions. *Hydrological Sciences Journal*, (October), 141023055853009. doi:10.1080/02626667.2014.978333

Li, W., & Sankarasubramanian, A. (2012). Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination. *Water Resources Research*, *48*(12), W12516, 1–17. doi:10.1029/2011WR011380

Lindstrom, G. (1997). A Simple Automatic Calibration Routine for the HBV Model. *Nordic Hydrology*, *28*(3), 153–168. doi:10.2166/nh.1997.009

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), W07401, 1–18. doi:10.1029/2006WR005756

Lu, J., Sun, G., Mcnulty, S. G., & Amatya, D. M. (2005). A Comparison of Six Potential Evapotranspiration Methods for Regional Use in the Southeastern United States. *Journal of the American Water Resources Association*, *41*(3), 621–633. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2005.tb03759.x/pdf

Luo, J., Wang, E., Shen, S., Zheng, H., & Zhang, Y. (2012). Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity. *Hydrological Processes*, *26*(26), 3953–3961. doi:10.1002/hyp.8420

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, *26*(2), 205–216. doi:10.1016/S0309-1708(02)00092-1

Magand, C., Ducharne, A., Le Moine, N., & Brigode, P. (2014). Parameter transferability under changing climate: case study with a land surface model in the Durance watershed, France. *Hydrological Sciences Journal*. doi:10.1080/02626667.2014.993643

Maidment, D. R. (1993). *Handbook of Hydrology*. McGraw-Hill Book Company.

Mantovan, P., & Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology*, *330*(1-2), 368–381. doi:10.1016/j.jhydrol.2006.04.046

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, *254*, 178–196. doi:10.1016/j.jtbi.2008.04.011

Marshall, L., Nott, D., & Sharma, A. (2007). Towards dynamic catchment modelling : a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes*, *861*(October 2006), 847–861. doi:10.1002/hyp

Mathias, S. A., Skaggs, T. H., Quinn, S. A., Egan, S. N. C., Finch, L. E., & Oldham, C. D. (2015). Water resources research. *Water Resources Research*, *51*(1), 506 – 523. doi:10.1016/0022-1694(68)90080-2

Matott, L. S., Babendreier, J. E., & Purucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resources Research*, *45*, W06421, 1–14. doi:10.1029/2008WR007301

McMillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, *400*(1-2), 83–94. doi:10.1016/j.jhydrol.2011.01.026

Melloh, R. A. (1999). *A Synopsis and Comparison of Selected Snowmelt Algorithms*. Hanover, New Hampshire. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a366395.pdf

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2), W02531, 1–17. doi:10.1029/2010WR009505

Montanari, A. (2005). Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, *41*, 1–13. doi:10.1029/2004WR003826

Montgomery, D. C., & Runger, G. C. (2003). *Applied Statistics and Probability for Engineers* (Third Edit.). John Wiley & Sons.

Moriasi, D. N., Arnold, J. G., Liew, M. W. Van, Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *American Society of Agricultural and Biological Engineers*, *50*(3), 885–900. Retrieved from http://swat.tamu.edu/media/1312/moriasimodeleval.pdf

Muleta, M. K. (2012). Improving Model Performance Using Season-Based Evaluation. *Journal of Hydrologic Engineering*, *17*(1), 191–200.

Muñoz, E., Rivera, D., Vergara, F., Tume, P., & Arumí, J. L. (2014). Identifiability analysis: towards constrained equifinality and reduced uncertainty in a conceptual model. *Hydrological Sciences Journal*, *59*(9), 1690–1703. doi:10.1080/02626667.2014.892205

Musy, A., & Higy, C. (2011). *Hydrology A Sicience of Nature*. CRC Press, Science Publishers.

Neitsch, S. L., Arnold, J. G., & Williams, J. R. (2011). *Soil & Water Assessment Tool Theoretical Documentation*.

Niel, H., Paturel, J.-E., & Servat, E. (2003). Study of parameter stability of a lumped hydrologic model in a context of climatic variability. *Journal of Hydrology*, *278*(1-4), 213–230. doi:10.1016/S0022-1694(03)00158-6

Nott, D. J., Marshall, L., & Brown, J. (2012). Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection? *Water Resources Research*, *48*(12), n/a–n/a. doi:10.1029/2011WR011128

Ogée, J., & Brunet, Y. (2002). A forest floor model for heat and moisture including a litter layer. *Journal of Hydrology*, *255*, 212–233. doi:10.1016/S0022-1694(01)00515-7

Osuch, M., Romanowicz, R. J., & Booij, M. J. (2014). The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. *Hydrological Sciences Journal*, 1–48. doi:10.1080/02626667.2014.967694

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., & Michel, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, *42*(7), W07410, 1–10. doi:10.1029/2005WR004636

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1-4), 290–306. doi:10.1016/j.jhydrol.2004.08.026

Pan, F., Zhu, J., Ye, M., Pachepsky, Y. a., & Wu, Y. S. (2011). Sensitivity analysis of unsaturated flow and contaminant transport with correlated parameters. *Journal of Hydrology*, *397*(3-4), 238–249. doi:10.1016/j.jhydrol.2010.11.045

Panagoulia, D. (1995). Assessment of daily catchment precipitation in mountainous regions for climate change interpretation. *Hydrological Sciences Journal*, *40*(3), 331–350. doi:10.1080/02626669509491419

Pavelková, H., Dohnal, M., & Vogel, T. (2012). HILLSLOPE RUNOFF GENERATION – COMPARING DIFFERENT MODELING APPROACHES. *Journal of Hydrology and Hydromechanics*, *60*, 73–86. doi:10.2478/v10098-012-0007-2

Pedersen, J. T., Peters, J. C., & Helweg, O. J. (1980). *Hydrographs by Single Linear Reservoir Model*. *Journal of Hydraulics Division, ASCE* (Vol. 106). US Army Corps of Engineers. Retrieved from http://www.hec.usace.army.mil/publications/TechnicalPapers/TP-74.pdf

Peel, M. C., & Bloschl, G. (2011). Hydrological modelling in a changing world. *Progress in Physical Geography*, *35*(2), 249–261. doi:10.1177/0309133311402550

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., & Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological Sciences Journal*, *52*(1), 131–151. doi:10.1623/hysj.52.1.131

Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, *46*(1), 1–17. doi:10.1029/2009WR008066

Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, *418-419*, 49–60. doi:10.1016/j.jhydrol.2008.12.004

Porêbska, D., Lamorski, K., & Walczak, R. T. (2006). Relationship between van Genuchten ' s parameters of the retention curve equation and physical properties of soil solid phase. *International Agrophysics*, *20*, 153–159.

Raghunath, H. M. (2006). *Hydrology: Principles, Analysis, Design* (Second Edi.). NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS.

Rawls, W. J., Brakensiek, D. L., & Saxton, K. E. (1982). Estimation of Soil Water Properties. *Transactions of the ASAE*, *25*(5), 1316–1320 & 1328.

Razavi, S., & Gupta, H. V. (2015). Water resources research. *Water Resources Research*, *51*(4), 1–23. doi:10.1016/0022-1694(68)90080-2

Razavi, S., Tolson, B. a., Matott, L. S., Thomson, N. R., MacLean, A., & Seglenieks, F. R. (2010). Reducing the computational cost of automatic calibration through model preemption. *Water Resources Research*, *46*(11), 1–17. doi:10.1029/2009WR008957

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, *51*, 438–456. doi:10.1016/j.advwatres.2012.01.005

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., … Christensen, J. H. (2014). A framework for testing the ability of models to project climate change and its impacts. *Climatic Change*, *122*, 271–282. doi:10.1007/s10584-013-0990-2

Remesan, R., & Mathew, J. (2013). Hydroinformatics and Data-Based Modelling Issues in Hydrology. In *Hydrological Data Driven Modelling: A Case Study Approach (Earth Systems Data and Models)*. Springer.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5), W05521,1–22. doi:10.1029/2009WR008328

RHMSS. (n.d.). Catchments of major rivers in Serbia. Retrieved from http://www.hidmet.gov.rs

Romanowicz, R. J., Osuch, M., & Grabowiecka, M. (2013). On the Choice of Calibration Periods and Objective Functions: A Practical Guide to Model Parameter Identification. *Acta Geophysica*, *61*(6), 1477–1503. doi:DOl: 10.2478/s11600-013-0157-6

Sadegh, M., & Vrugt, J. A. (2013). Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches. *Hydrology and Earth System Sciences Discussions*, *10*(4), 4739–4797. doi:10.5194/hessd-10-4739-2013

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578.

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). Rosetta: a Computer Program for Estimating Soil Hydraulic Parameters With Hierarchical Pedotransfer Functions. *Journal of Hydrology*, *251*(3-4), 163–176. doi:10.1016/S0022-1694(01)00466-8

Schaefli, B., Nicótina, L., Imfeld, C., Da Ronco, P., Bertuzzo, E., & Rinaldo, A. (2014). SEHR-ECHO v1.0: a Spatially Explicit Hydrologic Response model for ecohydrologic applications. *Geoscientific Model Development*, *7*(6), 2733–2746. doi:10.5194/gmd-7-2733-2014

Schoups, G., van de Giesen, N. C., & Savenije, H. H. G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research*, *44*(12), W00B03, 1–14. doi:10.1029/2008WR006836

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, *46*(10), W10531, 1–17. doi:10.1029/2009WR008933

Schumann, a. H. (1993). Development of conceptual semi-distributed hydrological models and estimation of their parameters with the aid of GIS. *Hydrological Sciences Journal*, *38*(6), 519–528. doi:10.1080/02626669309492702

Scorza Júnior, R. P., & Silva, J. P. Da. (2011). Sensibility analysis of the pearl model for pesticide leaching in the State of Mato Grosso do Sul, Brazil. *Engenharia Agrícola*, *31*, 965–973. doi:10.1590/S0100-69162011000500014

Seibert, J. (1997). Estimation of Parameter Uncertainty in the HBV Model. *Nordic Hydrology*, *28*(4/5), 247–262.

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Science*, *4*(2), 215–224. doi:10.5194/hess-4-215-2000

Seibert, J. (2003). Reliability of Model Predictions Outside Calibration Conditions. *Nordic Hydrology*, *34*(5), 477–492.

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, *38*(11), 23, 1–14. doi:10.1029/2001WR000978

Seibert, J., & Vis, M. J. P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, *16*(9), 3315–3325. doi:10.5194/hess-16-3315-2012

Seiller, G., Anctil, F., & Perrin, C. (2012). Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrology and Earth System Sciences*, *16*(4), 1171–1189. doi:10.5194/hess-16-1171-2012

Shafii, M., Tolson, B., & Matott, L. S. (2014). Uncertainty-based multi-criteria calibration of rainfall-runoff models: a comparative study. *Stochastic Environmental Research and Risk Assessment*, *28*, 1493–1510. doi:10.1007/s00477-014-0855-x

Shaw, E. M. (2005). *Hydrololgy in Practice* (3rd Editio.). Taylor & Francis.

Sieber, A., & Uhlenbrook, S. (2005). Sensitivity analyses of a distributed catchment model to verify the model structure. *Journal of Hydrology*, *310*(1-4), 216–235. doi:10.1016/j.jhydrol.2005.01.004

Singh, R. (2013). *AN UNCERTAINTY FRAMEWORK FOR HYDROLOGIC PROJECTIONS IN GAUGED AND UNGAUGED BASINS UNDER NON-STATIONARY CLIMATE CONDITIONS*.

Singh, S. K., & Bárdossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources*, *38*, 81–91. doi:10.1016/j.advwatres.2011.12.006

Sivapalan, M., Bloeschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*(11), 2101–2111. doi:10.1002/hyp.1425

Sorooshian, S., Gupta, V. K., & Fulton, J. L. (1983). Evaluation of Maximum Likelihood Parameter Estimation Techniques for Conceptual Rainfall-Runoff Models: Influence of Calibration Data Variability and length on Model Credibility. *Water Resources Research*, *19*(1), 251–259.

Sorooshian, S., Hsu, K.-L., Coppola, E., Tomassetti, B., Verdecchia, M., Visconti, G., & Editors. (2008). Hydrological Modelling and the Water Cycle. In V. P. Singh, M. Anderson, L. Bengtsson, J. F. Cruise, U. C. Kothyari, S. E. Serrano, … W. G. Strupczewski (Eds.), *Water Science and Technology Library, Volume 63*. Springer.

Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peel inlet, II, identification of critical uncertianties via generalized sensitivity analysis. *Water Resources Research*, *14*, 43–49.

Spitters, C. J. T., Toussaint, H. a. J. M., & Goudriaan, J. (1986). Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis Part I. Components of incoming radiation. *Agricultural and Forest Meteorology*, *38*(1-3), 217–229. doi:10.1016/0168-1923(86)90060-2

Steenbergen, N. Van, & Willems, P. (2012). Method for testing the accuracy of rainfall–runoff models in predicting peak flow changes due to rainfall changes, in a climate changing context. *Journal of Hydrology*, *414-415*, 425–434. doi:10.1016/j.jhydrol.2011.11.017

Tabari, H., Grismer, M. E., & Trajkovic, S. (2011). Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrigation Science*, *31*(2), 107–117. doi:10.1007/s00271-011-0295-z

Tarboton, D. (2003). *Rainfall-runoff processes*. *Utah State University*. Utah State University. Retrieved from http://ceefsl.engr.usu.edu/cee6400/TarbotonRainfallRunoffProcesses.pdf

Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P., … Vaze, J. (2014). Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal*, (October), 141017020632007. doi:10.1080/02626667.2014.967248

Tian, F., Li, H., & Sivapalan, M. (2012). Model diagnostic analysis of seasonal switching of runoff generation mechanisms in the Blue River basin, Oklahoma. *Journal of Hydrology*, *418-419*, 136–149. doi:10.1016/j.jhydrol.2010.03.011

Todini, E. (1996). The ARNO rainfall-runoff. *Journal of Hydrology*, *175*(1-4), 339–382.

Todini, E. (2007). Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences*, *11*(1), 468–482. doi:10.5194/hess-11-468-2007

Todorović, A., & Plavšić, J. (2014). Mogućnost primene modela HEC-HMS za kontinualne hidrološke simulacije. *Vodoprivreda*, *46*(1-6), 117–128.

Trajkovic, S., & Kolakovic, S. (2009). Evaluation of Reference Evapotranspiration Equations Under Humid Conditions. *Water Resources Management*, *23*(14), 3057–3067. doi:10.1007/s11269-009-9423-4

U.S. A.C.E. (1994). *Engineering and Design FLOOD-RUNOFF ANALYSIS*. (W. D. Brown, Ed.) (Manual No.). Washington, DC: U.S. Army Corps of Engineers. doi:EM 1110-2-1417

Uhlenbrook, S., Seibert, J. A. N., Leibundgut, C., & Rodhe, A. (2000). Prediction uncertainty of conceptual rainfall- runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences*, *44*(5), 779–797.

Urban Hydrology for Small Watersheds Technical Release 55. (1986). Washington, DC. Retrieved from http://www.cpesc.org/reference/tr55.pdf

Van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., & Lobligeois, F. (2013). The influence of conceptual model structure on model performance: a comparative study for 237 French catchments. *Hydrology and Earth System Sciences*, *17*(10), 4227–4239. doi:10.5194/hess-17-4227-2013

Van Genuchten, M. T. (1980). A Closed-form Equation for Predicting Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, *44*(5), 892–898. doi:doi:10.2136/sssaj1980.03615995004400050002x

Vanrolleghem, P. (2010). *Modelling Aspects of Water Framework Implementation*. IWA Publishing.

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

Vivoni, E. R., Entekhabi, D., Bras, R. L., & Ivanov, V. Y. (2007). Controls on runoff generation and scale-dependence in a distributed hydrologic model. *Hydrology and Earth System Sciences*, *11*(5), 1683–1701.

Vrugt, J. A., Braak, C. J. F., Gupta, H. V., & Robinson, B. a. (2008). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, *23*(7), 1011–1026. doi:10.1007/s00477-008-0274-y

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., & Sorooshian, S. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, *39*(8), SWC 5 1–19. doi:10.1029/2002WR001746

Vrugt, J. A., Gupta, H. V., Dekker, S. C., Sorooshian, S., Wagener, T., & Bouten, W. (2006). Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model. *Journal of Hydrology*, *325*(1-4), 288–307. doi:10.1016/j.jhydrol.2005.10.041

Vrugt, J. A., & Robinson, B. A. (2007). Improved evolutionary optimization from genetically adaptive multimethod search. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 104, pp. 708–11). doi:10.1073/pnas.0610471104

Vrugt, J. A., Robinson, B. A., & Hyman, J. M. (2009). Self-Adaptive Multimethod Search for Global Optimization in Real-Parameter Spaces. *IEEE Transactions on Evolutionary Computation*, *13*(2), 243–259.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. a. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12), W00B09, 1–15. doi:10.1029/2007WR006720

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. doi:10.1002/hyp.1135

Wanielista, M., Kersten, R., & Eaglin, R. (1997). *Hydrology: water quantity and quality control* (2nd Editio.). John Wiley & Sons Ltd.

Weiler, M., Mcdonnell, J. J., van Meerveld, i. T., & Uchida, T. (2005). Subsurface Stormflow. In *Encyclopedia of Hydrological Sciences* (pp. 1–14).

Weise, T. (2009). *Global Optimization Algorithms - Theory and Application*. Retrieved from http://www.it-weise.de/

White, E. D., Feyereisen, G. W., Veith, T. L., & Bosch, D. D. (2009). iImproving daily water yield estimates in the little river watershed: SWAT adjustments. *American Society of Agricultural and Biological Engineers*, *52*(1), 69–79.

Wilby, R. L. (2005). Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes*, *19*(16), 3201–3219. doi:10.1002/hyp.5819

Wolf, U., & Ostrowski, M. W. (1982). Estimation of Time Variant Hydrological Parameters. In *Proceedings of the International Symposium on Hydrological Research Basins*. Bern.

Wriedt, G., & Rode, M. (2006). Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. *Advances in Geosciences*, *9*, 145–150.

Xia, Y., Yang, Z.-L., Jackson, C., Stoffa, P. L., & Sen, M. K. (2004). Impacts of data length on optimal parameter and uncertainty estimation of a land surface model. *Journal of Geophysical Research*, *109*(D7), D07101, 1–13. doi:10.1029/2003JD004419

Yang, J., Reichert, P., Abbaspour, K. C., Xia, J., & Yang, H. (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology*, *358*(1-2), 1–23. doi:10.1016/j.jhydrol.2008.05.012

Yang, X., & You, X. (2013). Estimating Parameters of Van Genuchten Model for Soil Water Retention Curve by Intelligent Algorithms. *Applied Mathematics and Information Sciences*, *7*(5), 1977–1983. doi:10.12785/amis/070537

Yapo, P., Gupta, H. V., & Sorooshian, S. (1996). Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology*, *181*(1-4), 23–48.

Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, *204*(1-4), 83–97. doi:10.1016/S0022-1694(97)00107-8

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, 1–18. doi:10.1029/2007WR006716

Yilmaz, K. K., Vrugt, J. A., Gupta, H. V., & Sorooshian, S. (2010). Model Calibration in Watershed Hydrology. In B. Sivakumar & R. Berndtsson (Eds.), *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting* (pp. 53–105). Singapore: World Scientific Publishing.

Zhang, H., Huang, G. H., Wang, D., & Zhang, X. (2011). Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. *Advances in Water Resources*, *34*(10), 1292–1303. doi:10.1016/j.advwatres.2011.06.005

Zhang, X., Srinivasan, R., & Liew, M. Van. (2009). On the use of multi-algorithm , genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrological Processes*, *24*(8), 955–969. doi:10.1002/hyp

Zhang, Y., & Shuster, W. (2014). The Comparative Accuracy of Two Hydrologic Models in Simulating Warm-Season Runoff for Two Small, Hillslope Catchments. *JAWRA Journal of the American Water Resources Association*, *50*(2), 434–447. doi:10.1111/jawr.12135

APPENDICES

APPENDIX A. Prior ranges of the parameters for the Kolubara, Toplica, Mlava and River catchments

APPENDIX B. Impact of parameters of the 3DNet-Catch model on hydrograph – semi-lumped BASIC version of the model, the Mlava River catchment from 1st October 1989 to 30th September 1990

APPENDIX C. Sensitivity analysis: semi-lumped BASIC version of the 3DNet-Catch model

APPENDIX D. Temporal variability in median values of normalised Pareto-optimal parameters and information content (the IC statistic)

APPENDIX E. Median values of normalised Pareto-optimal parameters against the length of the calibration period: semi-lumped BASIC version of the model

APPENDIX F. Values of S_t obtained over calibration periods of various lengths

APPENDIX G. Empirical cumulative distribution functions of normalised Pareto-optimal parameters

APPENDIX H. Parameter identifiability over calibration period of different lengths

APPENDIX I. Model performance over different calibration periods: semi-lumped BASIC version of the model

APPENDIX J. Temporal variability in hydro-meteorological characteristics in the catchments

APPENDIX K. Relationship between the Pareto-optimal parameters and hydro-meteorological characteristics of the calibration period

APPENDIX L. Model performance for different calibration strategies and for different model structures

APPENDIX A. Prior ranges of the parameters for the Kolubara, Toplica, Mlava and River catchments

A. Table. 1. Prior ranges of the precipitation gradient with elevation and temperature lapse rate: semi-lumped versions of the model

Catchment	Kolubara			Toplica			Mlava		
Version of the 3DNet-Catch model	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL
Precipitation gradient with elevation α [% / 100 m]		0 – 25			0 – 15			0 – 20	
Lapse rate T_{LAPSE} [°C / 100 m]		0.45 – 0.75			0.4 – 0.8			0.45 – 0.8	

A. Table. 2. Prior ranges of the parameters of the interception routine: semi-lumped versions of the model

Catchment	Kolubara			Toplica			Mlava		
Version of the 3DNet-Catch model	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL
Maximum interception reservoir capacity CAN_{max} [mm]		1 – 8			1 – 9			1 – 8	
Maximum value of the Leaf Area Index LAI_{max} [m^2m^{-2}]					0.5 – 12				

A. Table. 3. Prior ranges of the parameters of the snow routine: semi-lumped versions of the model

Catchment	Kolubara			Toplica			Mlava		
Version of the 3DNet-Catch model	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL
Boundary temperature $T_{S\text{-}R}$ [°C]		(-3) – 3			(-3.5) – 3			(-4) – 3	
Threshold depth of snow (as water equivalent) above which the entire area is covered in snow $S_{snow,100}$ [mm]	147.3	1-150		149.1	1-150		123.7	1-150	
Snowpack temperature lag factor λ [-]	0.24	0 – 1		0.14	0 – 1		0.07	0 – 1	
Snowmelt temperature T_{melt} [°C]	-	-	(-3) – 3	-	-	(-3) – 3	-	-	(-3) – 3
Melt factor on the 21st of June $b_{melt,6}$ [mm°C^{-1}day^{-1}]	1.68	1.4 – 8		1.65	1.4 – 8		1.66	1.4 – 8	
Melt factor on the 21st of December, $b_{melt,12}$ – ratio to $b_{melt,6}$ [-]	0.85	0.05 – 0.95		0.74	0.05 – 0.95		0.21	0.05 – 0.95	

A. Table. 4. Prior ranges of the parameters of the soil routine: semi-lumped versions of the model

Catchment	Kolubara			Toplica			Mlava		
Version of the 3DNet-Catch model	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL
Curve number *CN* [-]		55 – 85			60 – 85			50 – 85	
Initial abstraction I_{a_rel} [-]		0.1 – 0.35			0.1 – 0.3			0.1 – 0.35	
Surface layer thickness D_{surf} [mm]		5 – 110			5 – 100			20 – 100	
Effective porosity [-]		0.05 – 0.65			0.1 – 0.5			0.2 – 0.4	
Common logarithm of saturated hydraulic conductivity of the surface soil layer K_{surf} [m / s]					(-5.5) – (-4)				
Wetness at permanent wilting point w_{wp} – ratio to effective porosity [-]		0.01 – 0.3			0.05 – 0.3			0.01 – 0.3	
Wetness at field capacity w_{fc} – ratio to (porosity-w_{wp}) [-]		0 – 0.45			0 – 0.4			0 – 0.4	
Pore size distribution index *n* [-]		1.1 – 5.5			1.1 – 6			1.1 – 6	
Number of sub-surface layers N_l	1	1	1 – 5	1	1	1 – 5	1	1	1 – 5
Thickness of a sub-surface layer $D_{sub-surf}$ [mm]		15 – 1500			15 – 1500			15 – 1000	
$K_{sub-surf}$ – common logarithm of ratio to the K_{surf} [-]					(-3.5) – 0				

A. Table. 5. Prior ranges of the parameters of the response routine: semi-lumped versions of the model

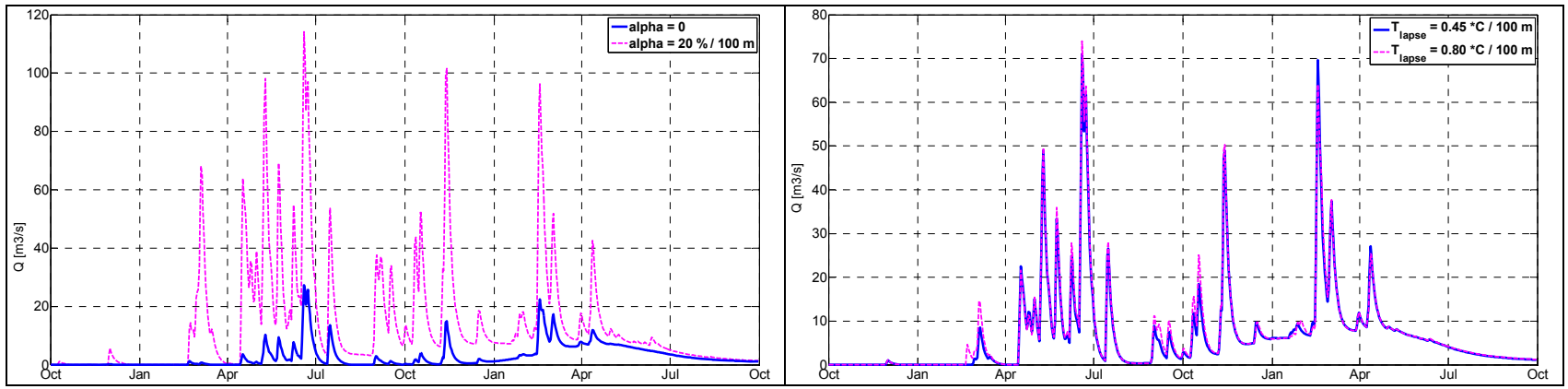
Catchment	Kolubara			Toplica			Mlava		
Version of the 3DNet-Catch model	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL	SIMPLE	BASIC	FULL
Linear reservoir coeff. for direct flow K_d [days]		0.1 – 20			0.15 – 15			0.25 – 15	
Number of the linear reservoirs N_{LR} [-]	1	1	1 – 10	1	1	1 – 10	1	1	1 – 10
Fast groundwater response reservoir coeff. $K_{gw\text{-}fast}$ – ratio to K_d [-]		1.05 – 60			1.01 – 60			1.01 – 50	
Maximum specific baseflow yield q_d [L / s / km^2]		0.01 – 0.45			0.05 – 0.2			0.075 – 0.225	
Non-linearity coefficient for baseflow simulation c [-]		1.01 – 60			1.05 – 30			1.03 – 30	
Threshold of the non-liner baseflow reservoir per unit area s_{max} [mm]		1 – 550			1 – 300			1 – 300	

A. Table. 6. Prior values of parameters of the distributed version of the model

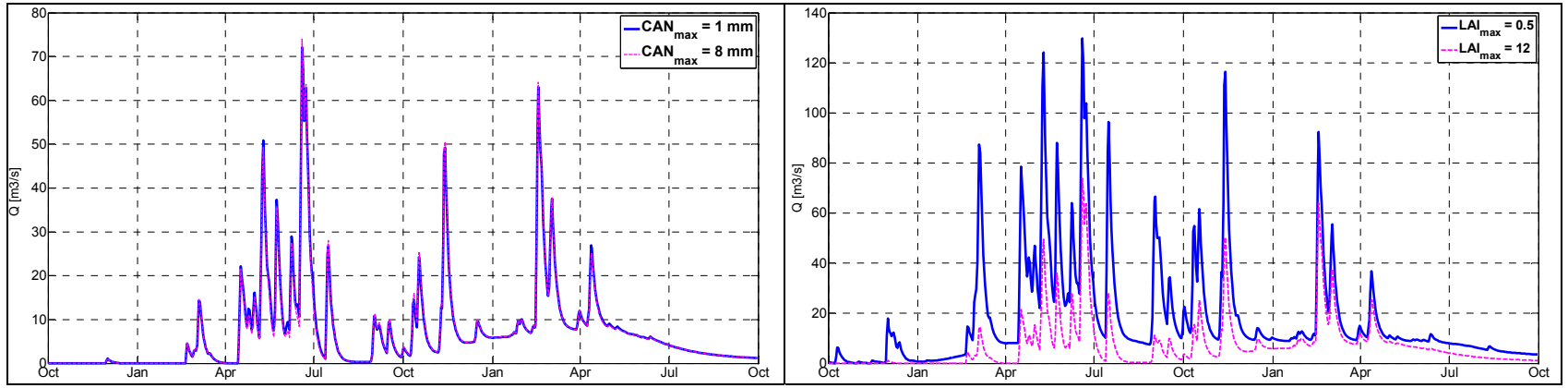
Catchment	Kolubara	Toplica	Mlava
Precipitation gradient with elevation α [% / 100 m]	10-12.5	1-9	10-12
Lapse rate T_{LAPSE} [°C / 100 m]	0.67	0.8	0.8
Maximum interception reservoir capacity CAN_{max} [mm]	2.2-2.53	1.13-2.6	2.48-3.21
Maximum value of the Leaf Area Index LAI_{max} [m^2m^{-2}]	9.54-10.11	8.48-9.93	9.63-10.77
Boundary temperature $T_{S\text{-}R}$ [°C]	-1.7	-3.3	-1.3
Threshold depth of snow (as water equivalent) above which the entire area is covered in snow $S_{snow,100}$ [mm]	147.3	149.1	123.7
Snowpack temperature lag factor λ [-]	0.05-0.3	0.05-0.25	0.07-0.25
Melt factor on the 21st of June $b_{melt,6}$ [$mm°C^{-1}day^{-1}$]	1.68	1.65	1.66
Melt factor on the 21st of December $b_{melt,12}$ – ratio to $b_{melt,6}$ [-]	0.85	0.74	0.21
Curve number *CN* [-]	64.3-83.6	64-83.2	64.3-72.3
Initial abstraction I_{a_rel} [-]	0.16	0.28	0.05
Surface layer thickness D_{surf} [mm]	70.7	25.7	26.9
Effective porosity [-]	0.35	0.44	0.38
Common logarithm of saturated hydraulic conductivity of the surface soil layer K_{surf} [$mm \cdot day^{-1}$]	-4.72	-5.34	-5.44
Wetness at permanent wilting point w_{wp} – ratio to eff. porosity [-]	0.24	0.06	0.14
Wetness at field capacity w_{fc} – ratio to the (porosity- w_{wp}) [-]	0.27	0.11	0.13
Pore size distribution index *n* [-]	1.3	1.7	1.1
Thickness of a sub-surface layer $D_{sub\text{-}surf}$ [mm]	1107.8	1383.6	984.1
$K_{sub\text{-}surf}$ – common logarithm of the ratio to K_{surf} [-]	-3.03	-3.25	-2.7
Linear reservoir coeff. for direct flow K_d [days]	1	7.46	4.31
Fast groundwater response reservoir coeff. $K_{gw\text{-}fast}$ – ratio to K_d [-]	1.08	40.8	31
Maximum specific baseflow yield q_d [$L / s / km^2$]	0.34	0.06	0.08
Non-linearity coefficient for baseflow simulation *c* [-]	1.13	1.08	29.6
Threshold of the non-liner baseflow reservoir per unit area s_{max} [mm]	1.35	220.4	277

[*] *Highlighted cells denote spatially distributed parameters.*

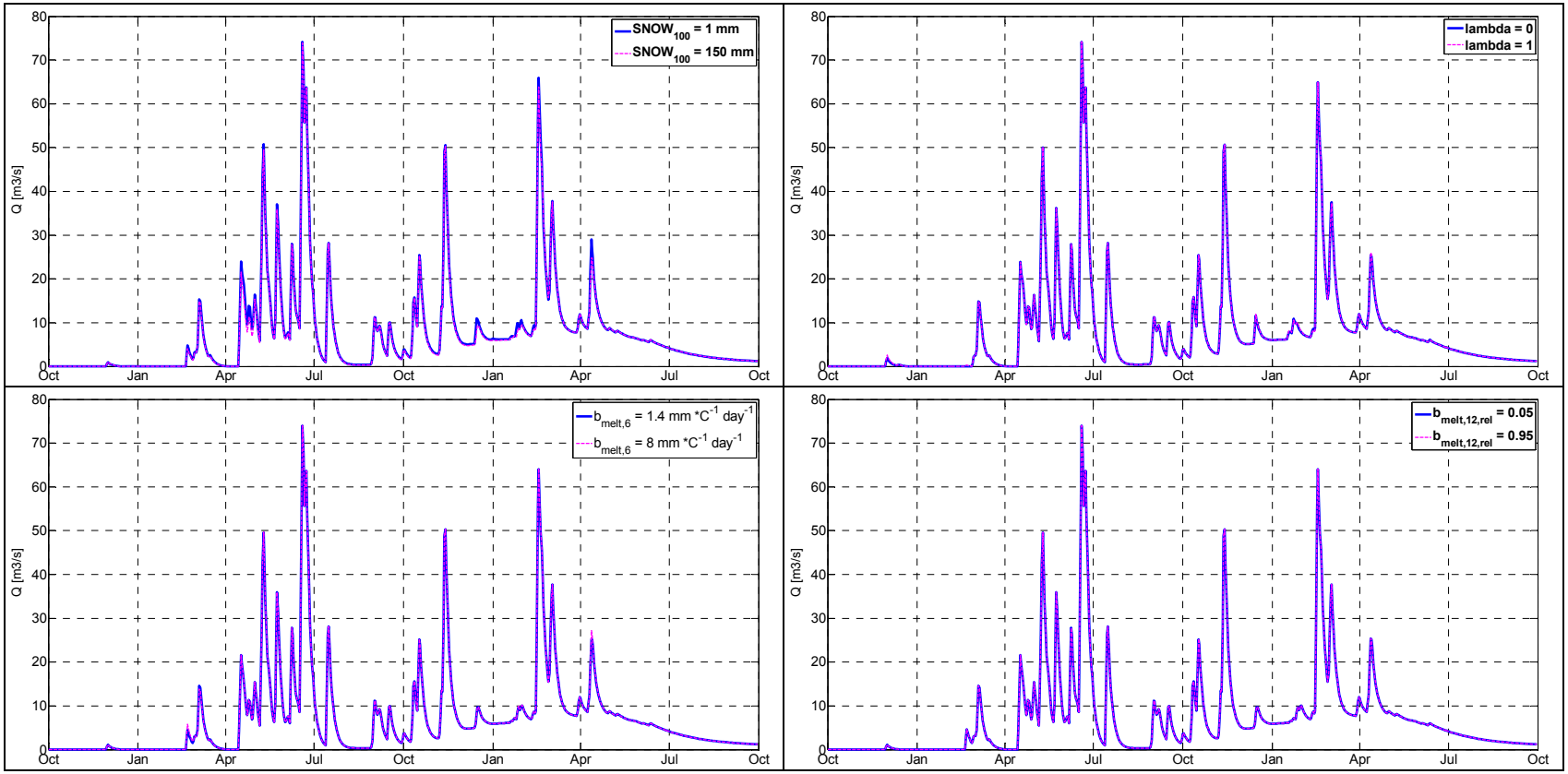
APPENDIX B. Impact of parameters of the 3DNet-Catch model on hydrograph – semi-lumped BASIC version of the model, the Mlava River catchment from 1st October 1989 to 30th September 1990



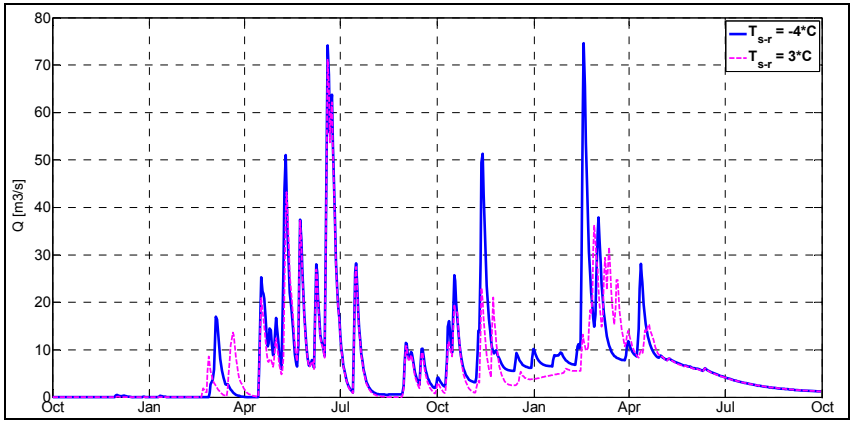
A. Figure 1. Precipitation gradient with elevation and lapse rate.



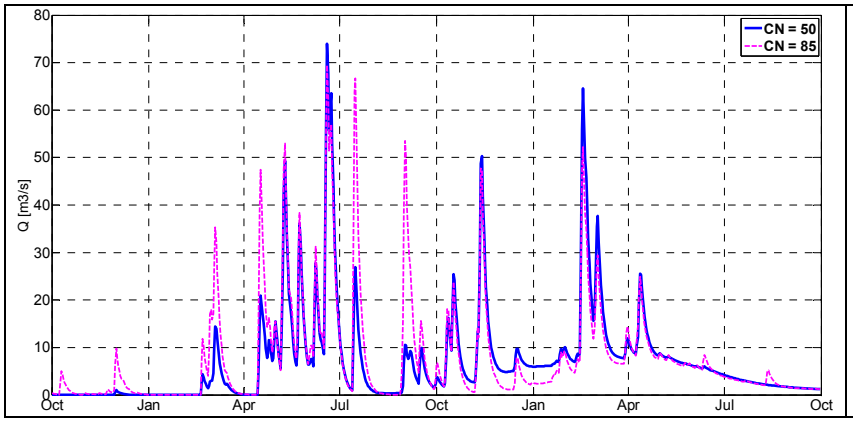
A. Figure 2. The parameters of the interception routine.

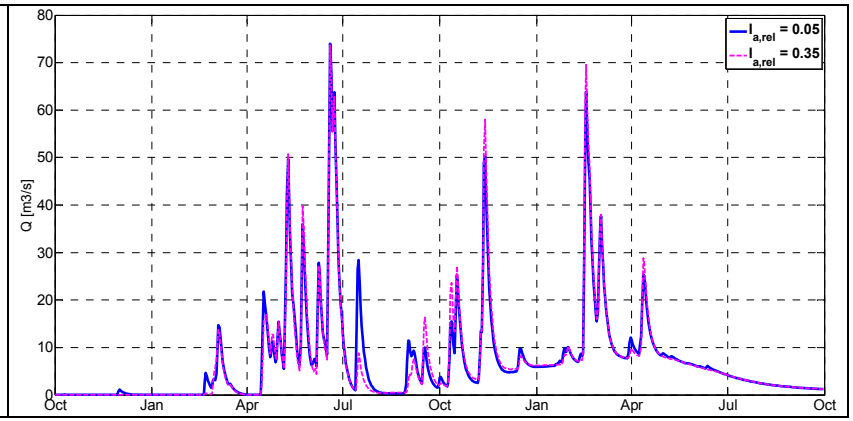


A. Figure 3. The parameters of the snow routine.

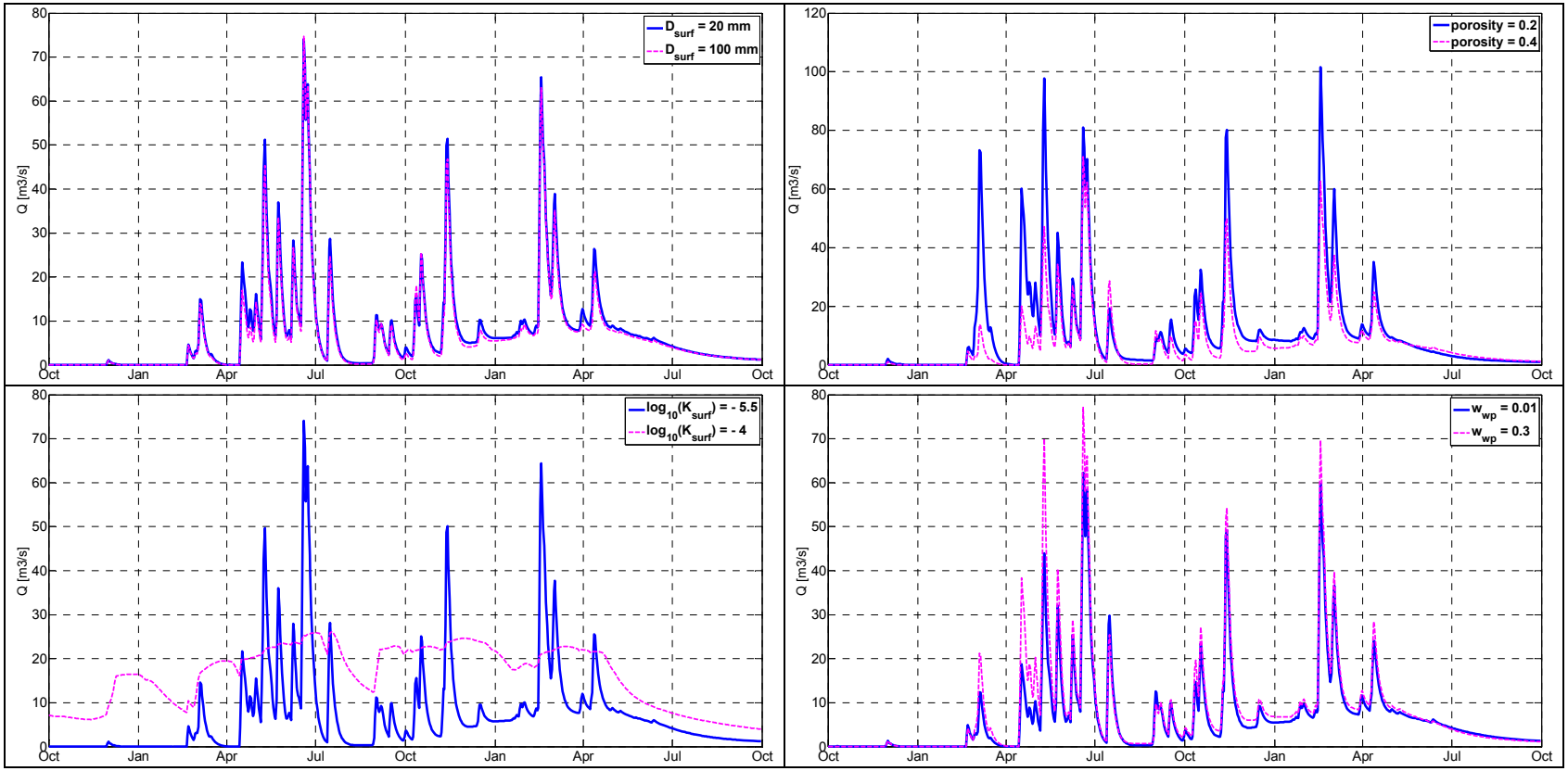


A. Figure 3 (continued). The parameters of the snow routine.

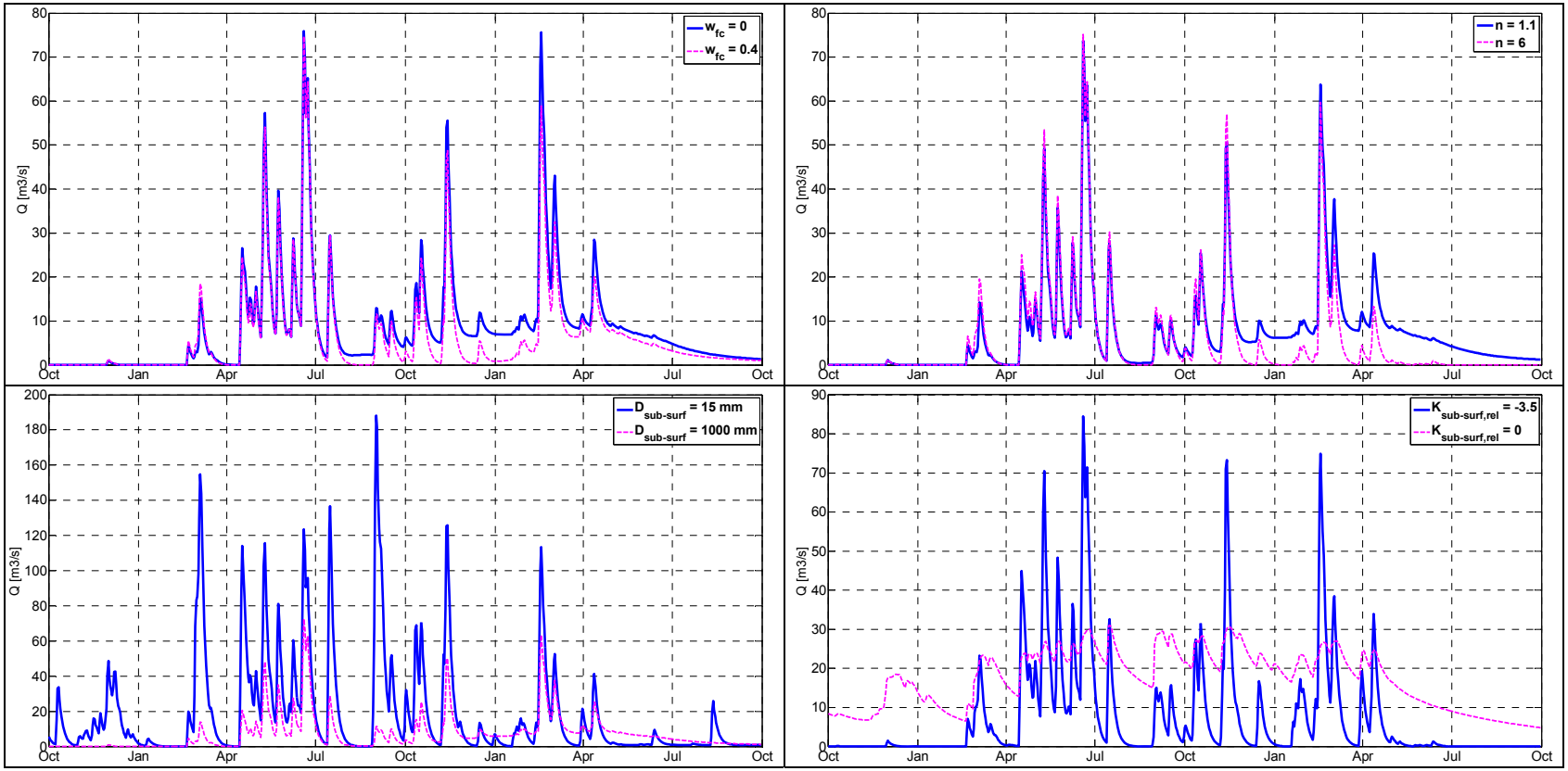




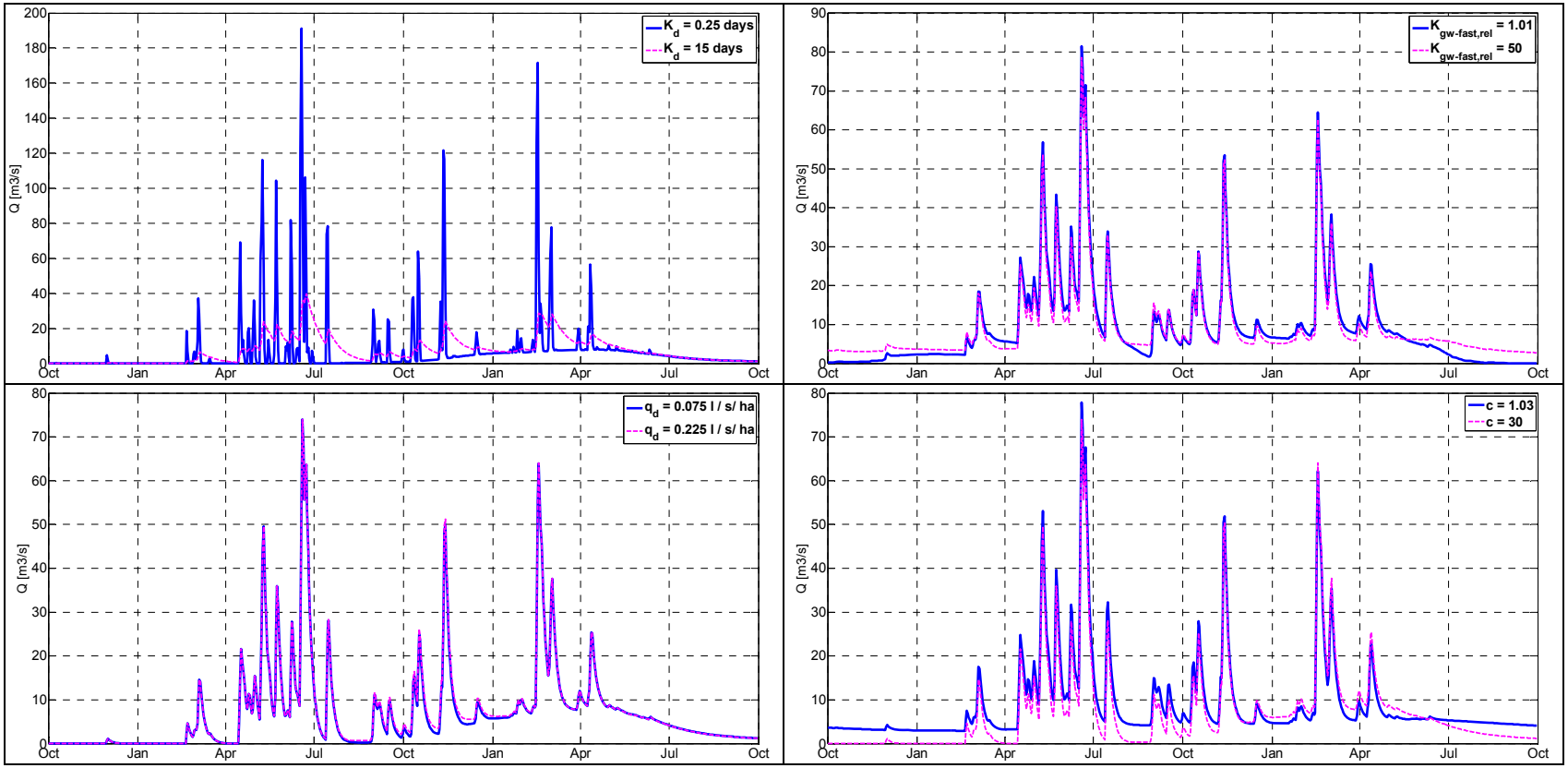
A. Figure 4. The parameters of the soil routine.



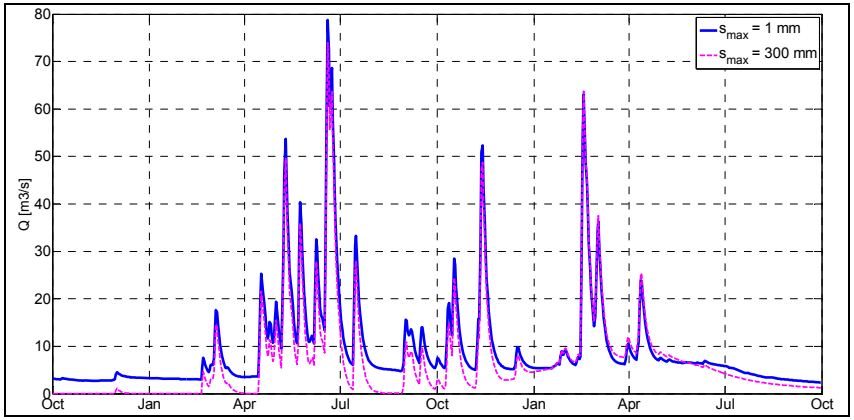
A. Figure 4 (continued). The parameters of the soil routine.



A. Figure 4 (continued). The parameters of the soil routine.



A. Figure 5. The parameters of the response routine.



A. Figure 5 (continued). The parameters of the response routine.

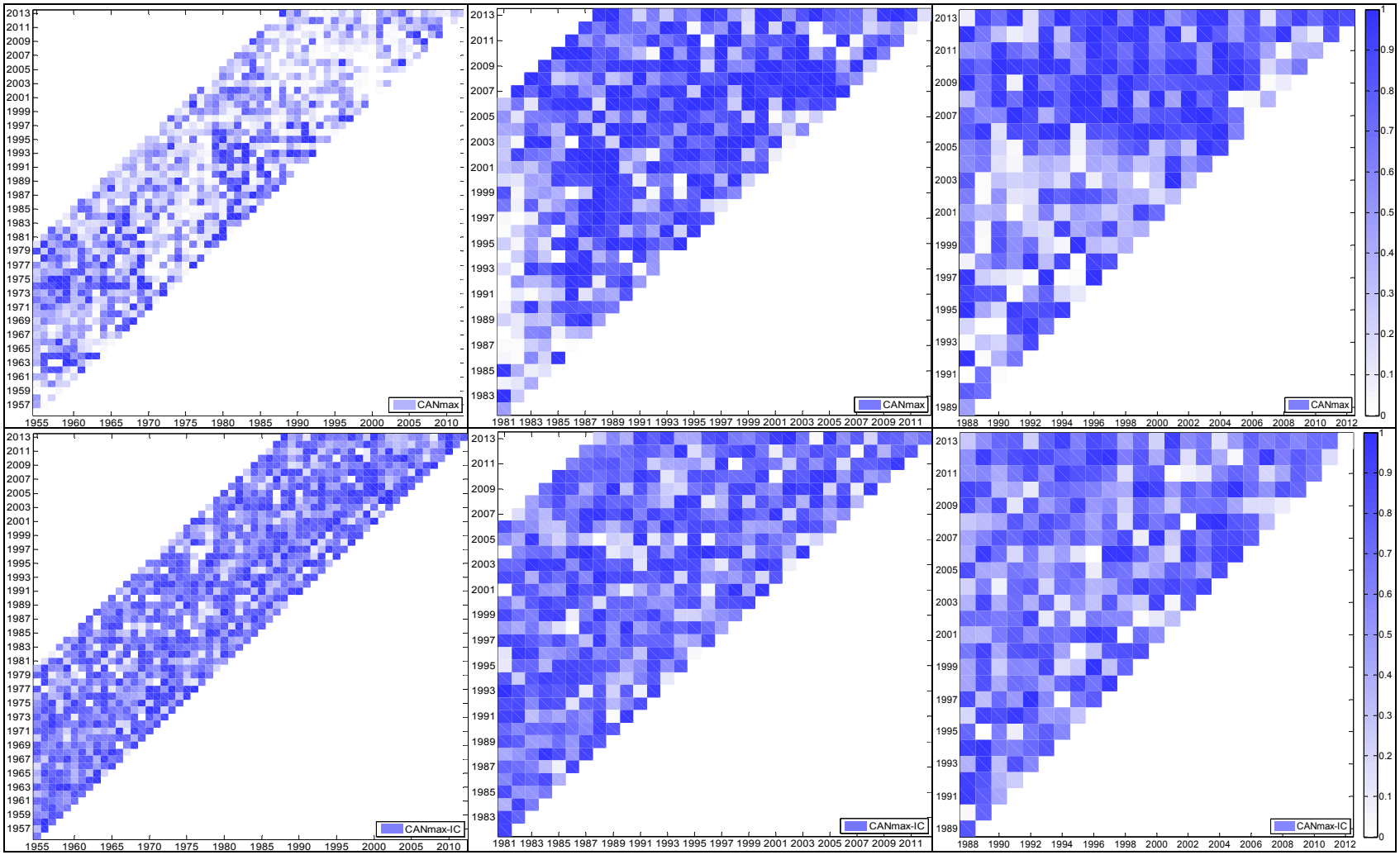
APPENDIX C. Sensitivity analysis: semi-lumped BASIC version of the 3DNet-Catch model

Catchment	Output	Measure	CAN_{max}	LAI_{max}	$T_{s\text{-}r}$	$SNOW_{100}$	lambda	$b_{melt,6}$	$b_{melt,12}$	CN	$I_{a,rel}$	D_{surf}	porosity	$\log(K_{surf})$	w_{wp}	w_{fc}	n	$D_{sub\text{-}surf}$	$\log(K_{sub\text{-}surf})$	K_d	$K_{gw\text{-}fast}$	q_d	c	s_{max}	alpha	T_{lapse}
Kolubara	*NSE*	SRC	0.04	0.12	0.00	0.01	0.02	0.01	0.01	-0.02	-0.01	0.08	0.22	0.01	-0.03	0.09	0.07	0.23	0.05	0.17	0.03	-0.005	0.001	0.004	-0.41	-0.01
		LCC	0.04	0.12	0.000	0.01	0.02	0.01	0.01	-0.02	-0.001	0.09	0.23	0.01	-0.03	0.09	0.07	0.23	0.05	0.17	0.03	-0.002	-0.003	0.01	-0.41	-0.01
		|LCC/SRC|	0.97	1.00	0.07	1.01	0.96	1.94	1.10	1.02	0.16	1.04	1.00	0.89	1.17	1.00	0.96	1.00	1.00	0.97	0.98	0.29	2.45	1.83	1.00	0.56
	VE	SRC	0.04	0.18	0.03	0.002	0.001	0.01	0.01	-0.01	0.00	0.11	0.30	-0.07	-0.04	0.14	0.12	0.28	-0.13	0.01	0.01	-0.003	0.01	0.000	-0.62	-0.02
		LCC	0.04	0.17	0.03	0.002	0.00	0.01	0.01	-0.01	0.01	0.12	0.30	-0.07	-0.05	0.14	0.12	0.28	-0.13	-0.001	0.01	0.004	0.001	0.004	-0.62	-0.01
		|LCC/SRC|	1.01	0.97	0.91	0.74	4.18	2.59	0.79	0.77	4.61	1.04	1.00	1.02	1.14	1.02	0.98	1.01	1.00	0.08	1.00	1.56	0.21	22.10	1.00	0.63
Toplica	*NSE*	SRC	0.04	0.12	-0.02	0.01	0.00	0.00	0.003	-0.03	0.007	0.10	0.21	0.04	-0.04	0.08	0.08	0.31	0.11	0.25	0.07	-0.02	-0.01	0.02	-0.55	-0.01
		LCC	0.03	0.12	-0.03	0.01	0.004	0.01	0.004	-0.03	0.015	0.10	0.21	0.04	-0.05	0.07	0.07	0.31	0.11	0.25	0.06	-0.02	-0.02	0.02	-0.55	-0.01
		|LCC/SRC|	0.94	1.01	1.19	1.06	0.80	3.68	1.54	1.04	2.03	1.04	1.01	0.92	1.15	0.98	0.95	1.00	1.02	0.97	0.97	0.73	1.52	1.26	1.00	0.56
	VE	SRC	0.04	0.19	0.04	0.007	0.00	0.003	0.004	-0.01	0.004	0.13	0.28	-0.09	-0.06	0.15	0.18	0.35	-0.17	0.01	0.00	-0.001	0.004	0.002	-0.73	-0.03
		LCC	0.05	0.18	0.04	0.007	0.00	0.01	0.00	-0.01	0.016	0.14	0.28	-0.10	-0.06	0.15	0.17	0.36	-0.17	0.00	0.00	0.01	-0.003	0.007	-0.73	-0.02
		|LCC/SRC|	1.02	0.97	0.93	0.97	0.07	3.82	0.56	0.64	4.21	1.05	1.00	1.03	1.13	1.02	0.98	1.01	1.00	0.44	0.96	6.09	0.59	3.26	1.00	0.71
Mlava	*NSE*	SRC	0.04	0.12	-0.02	0.01	0.00	0.003	0.003	-0.029	0.01	0.10	0.21	0.04	-0.04	0.08	0.08	0.31	0.11	0.25	0.07	-0.02	-0.01	0.02	-0.55	-0.013
		LCC	0.03	0.12	-0.03	0.01	0.00	0.01	0.004	-0.03	0.015	0.10	0.21	0.04	-0.05	0.07	0.07	0.31	0.11	0.25	0.06	-0.02	-0.02	0.02	-0.55	-0.007
		|LCC/SRC|	0.94	1.01	1.19	1.065	0.80	3.68	1.54	1.04	2.03	1.04	1.01	0.92	1.15	0.98	0.95	1.00	1.02	0.97	0.97	0.73	1.52	1.26	1.00	0.56
	VE	SRC	0.04	0.19	0.04	0.007	-0.004	0.003	0.00	-0.008	0.004	0.13	0.28	-0.09	-0.06	0.15	0.18	0.35	-0.17	0.007	0.00	-0.001	0.004	0.002	-0.73	-0.03
		LCC	0.05	0.18	0.04	0.007	0.000	0.01	0.00	-0.005	0.016	0.14	0.28	-0.10	-0.06	0.15	0.17	0.36	-0.17	0.00	0.004	0.008	0.00	0.007	-0.73	-0.02
		|LCC/SRC|	1.02	0.97	0.93	0.97	0.07	3.82	0.56	0.64	4.21	1.05	1.00	1.03	1.13	1.02	0.98	1.01	1.00	0.44	0.96	6.09	0.59	3.26	1.00	0.71

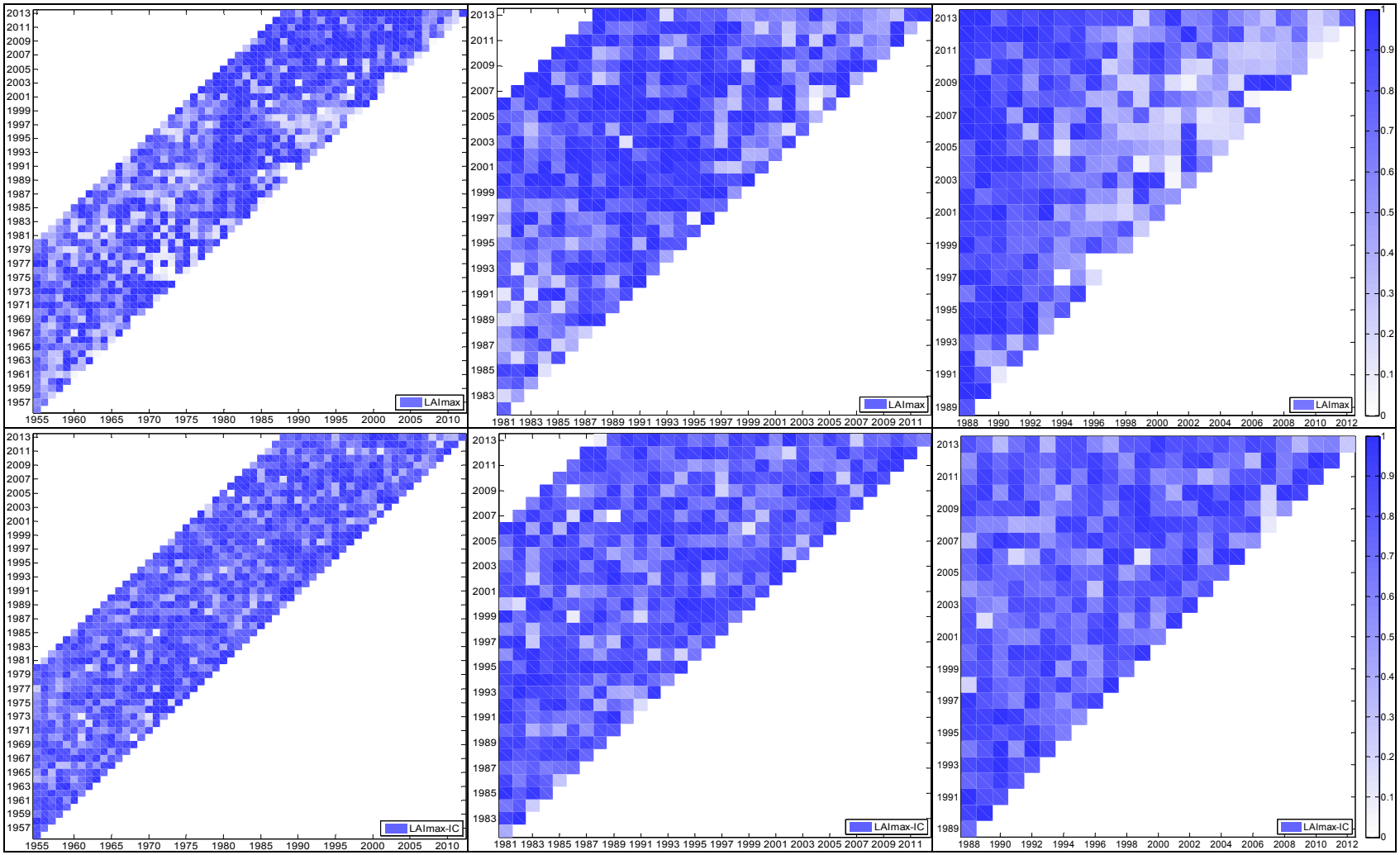
[*] *Statistically significant correlations at 95% significance level are highlighted.*

APPENDIX D. Temporal variability in median values of normalised Pareto-optimal parameters and information content (the IC statistic)

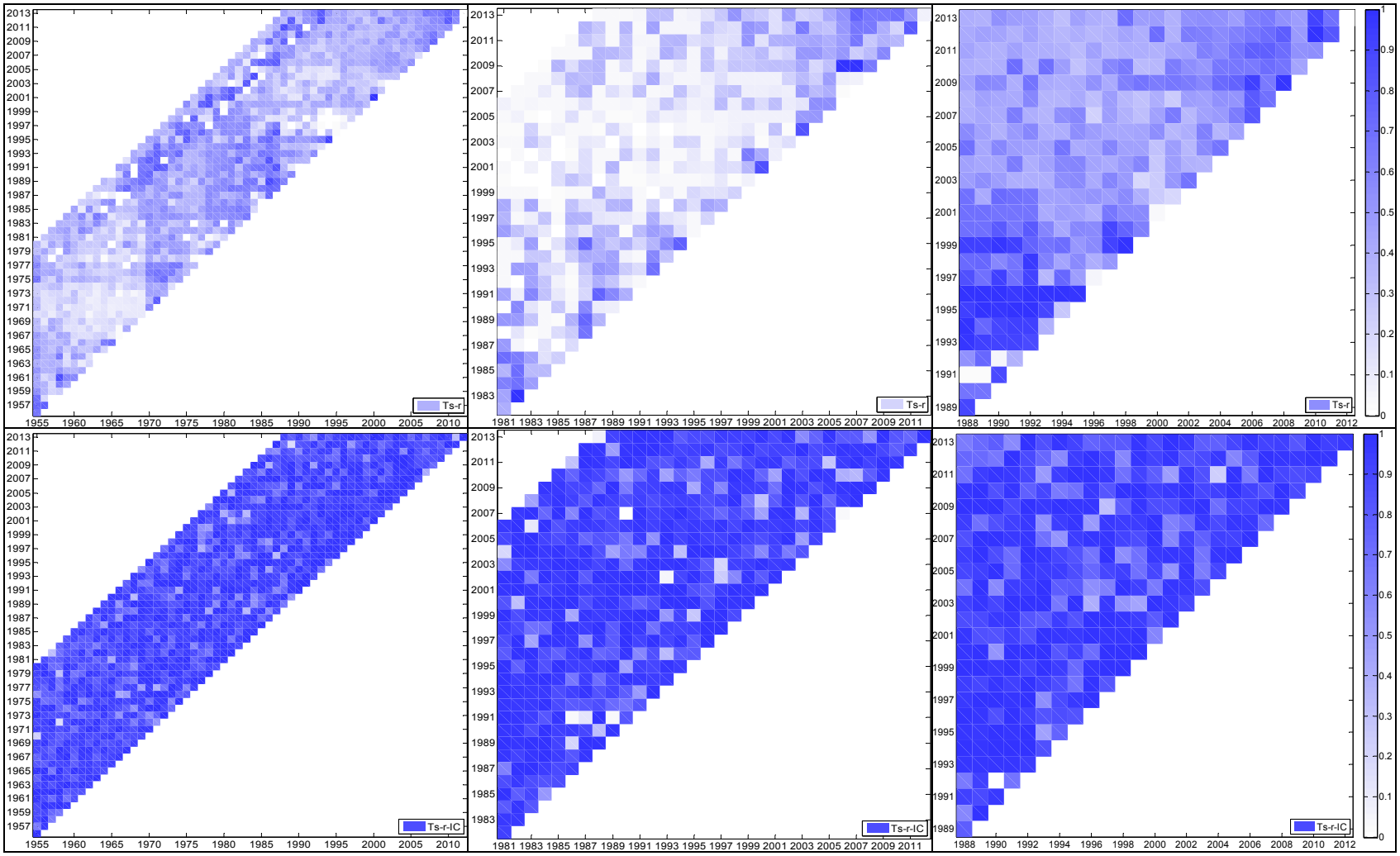
Median values of the Pareto-optimal parameters (top panels) and values of the *IC* statistic (bottom panels) obtained over every calibration period are presented in A. Figure 6 through A. Figure 29. Left panels of these figures refer to the Kolubara River catchment, mid panels to the Toplica River catchment and right panels to the Mlava River catchment.



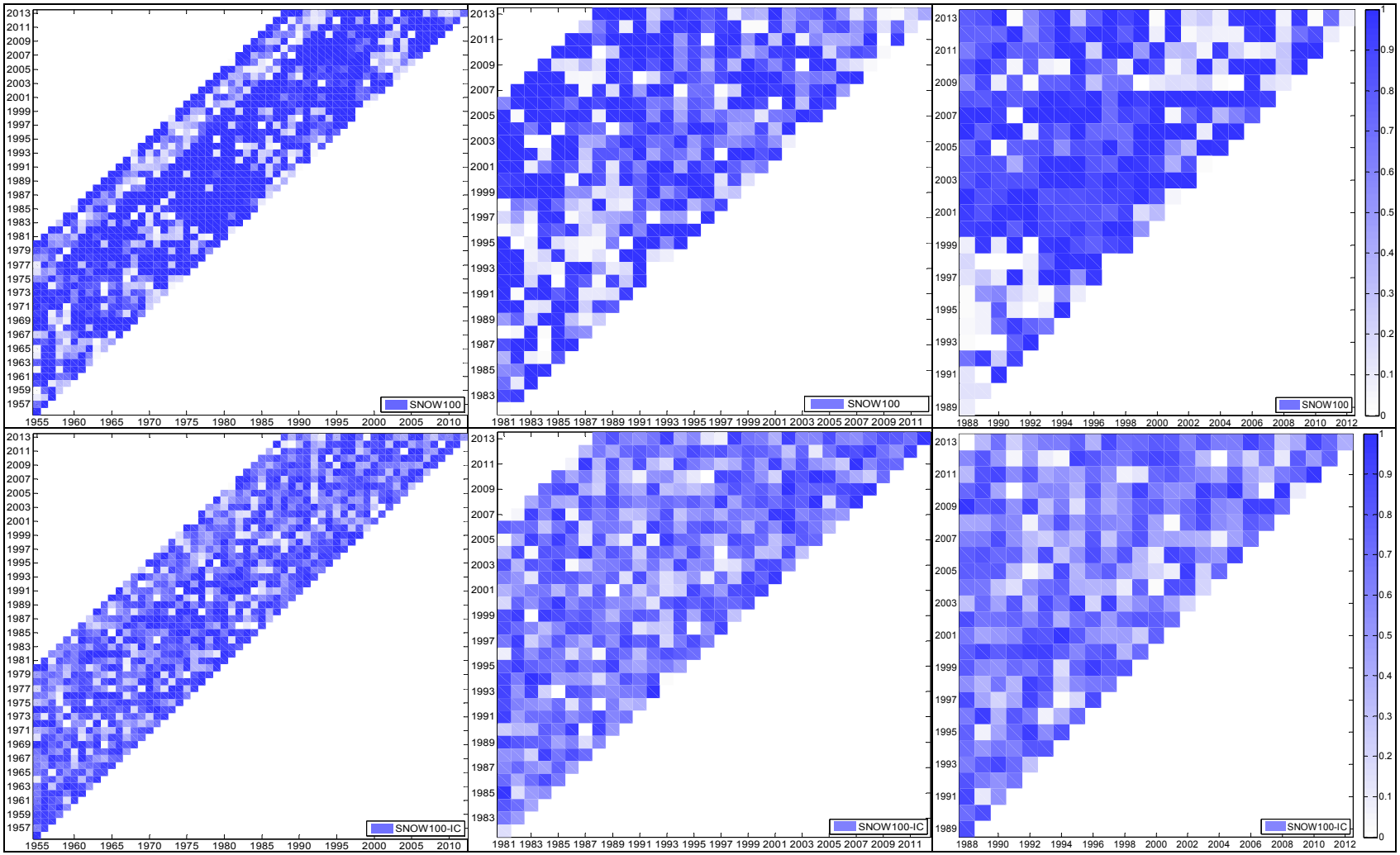
A. Figure 6. CAN_{max}: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



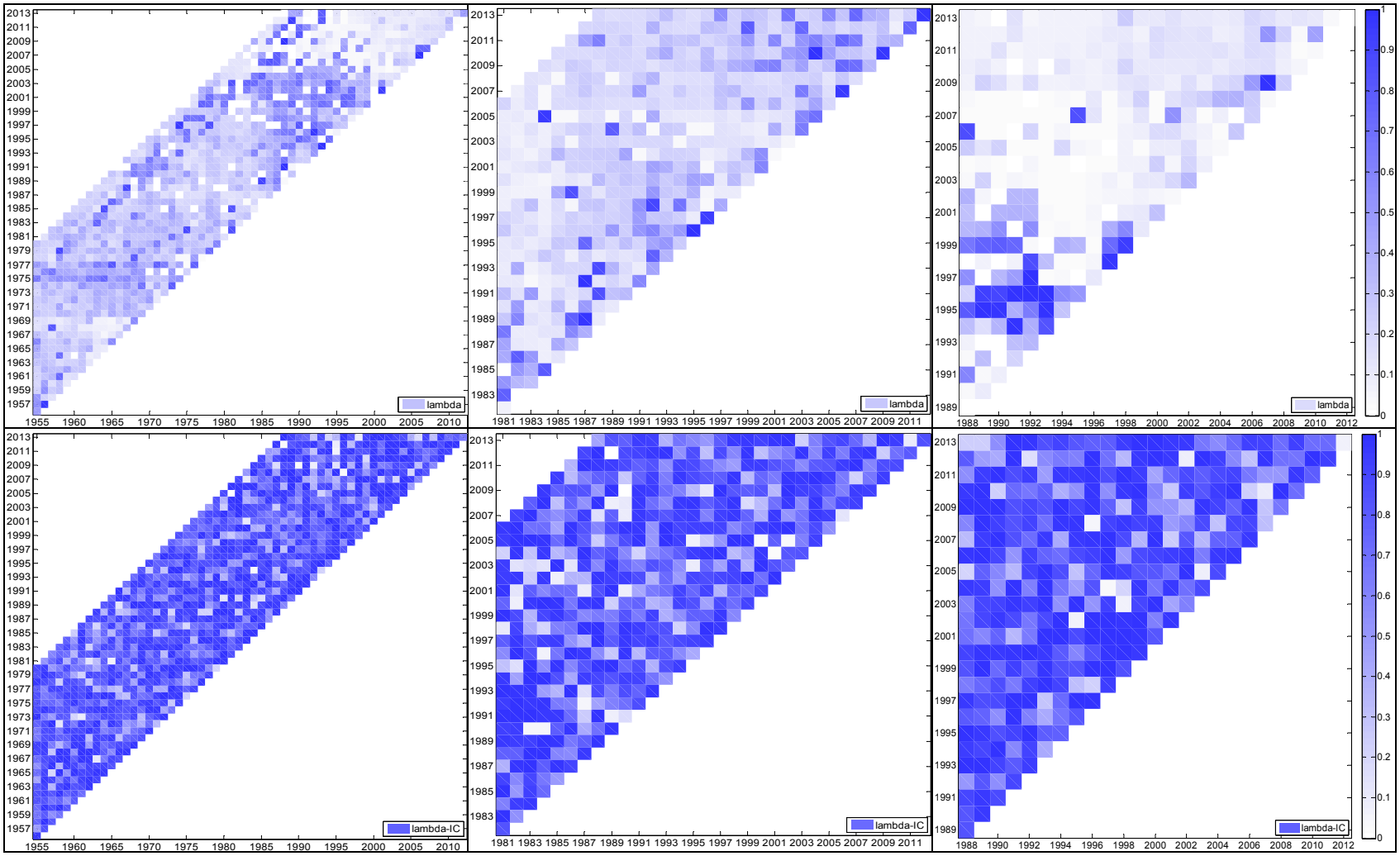
A. Figure 7. LAI_{max}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



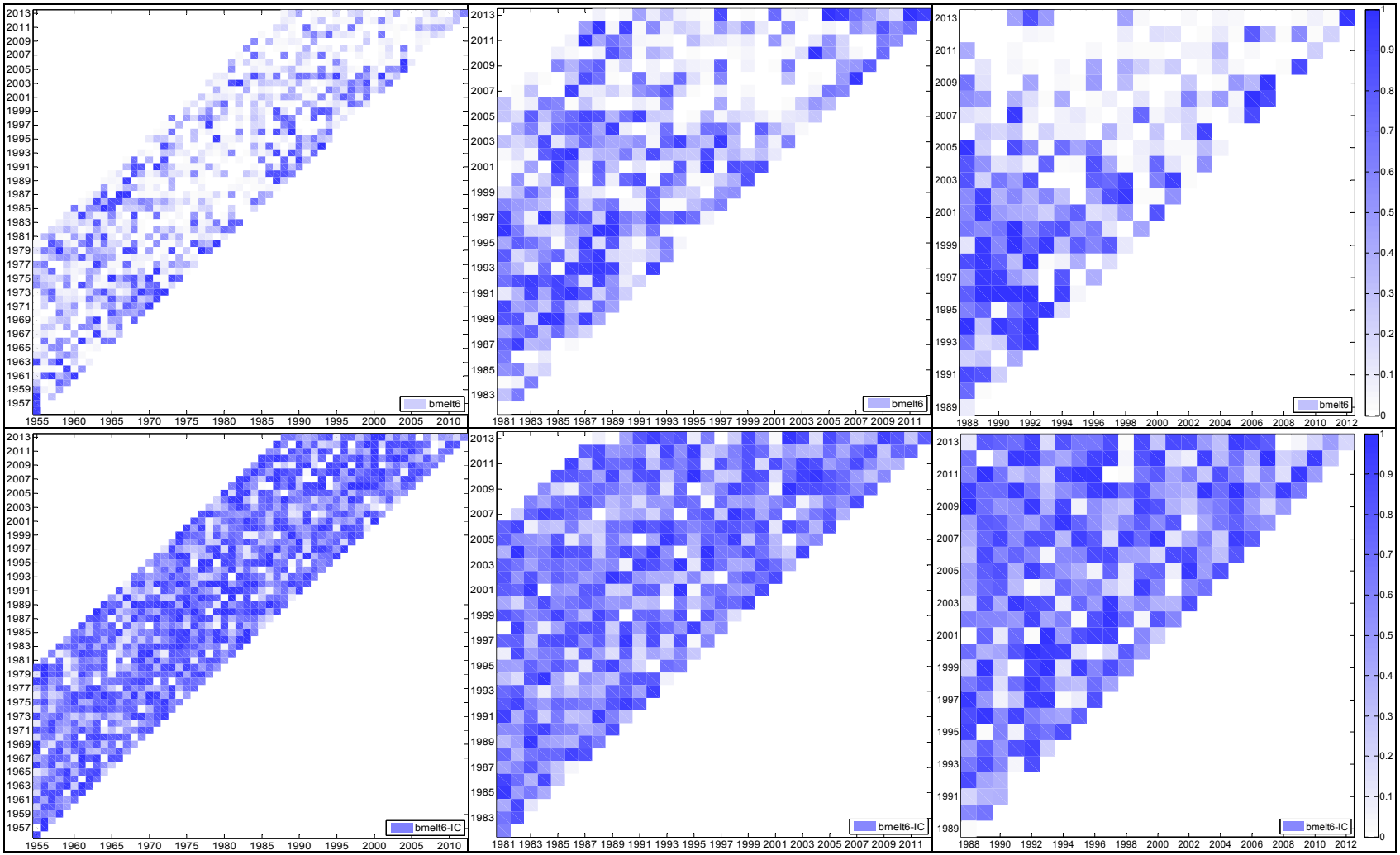
A. Figure 8. $T_{S\text{-}R}$: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



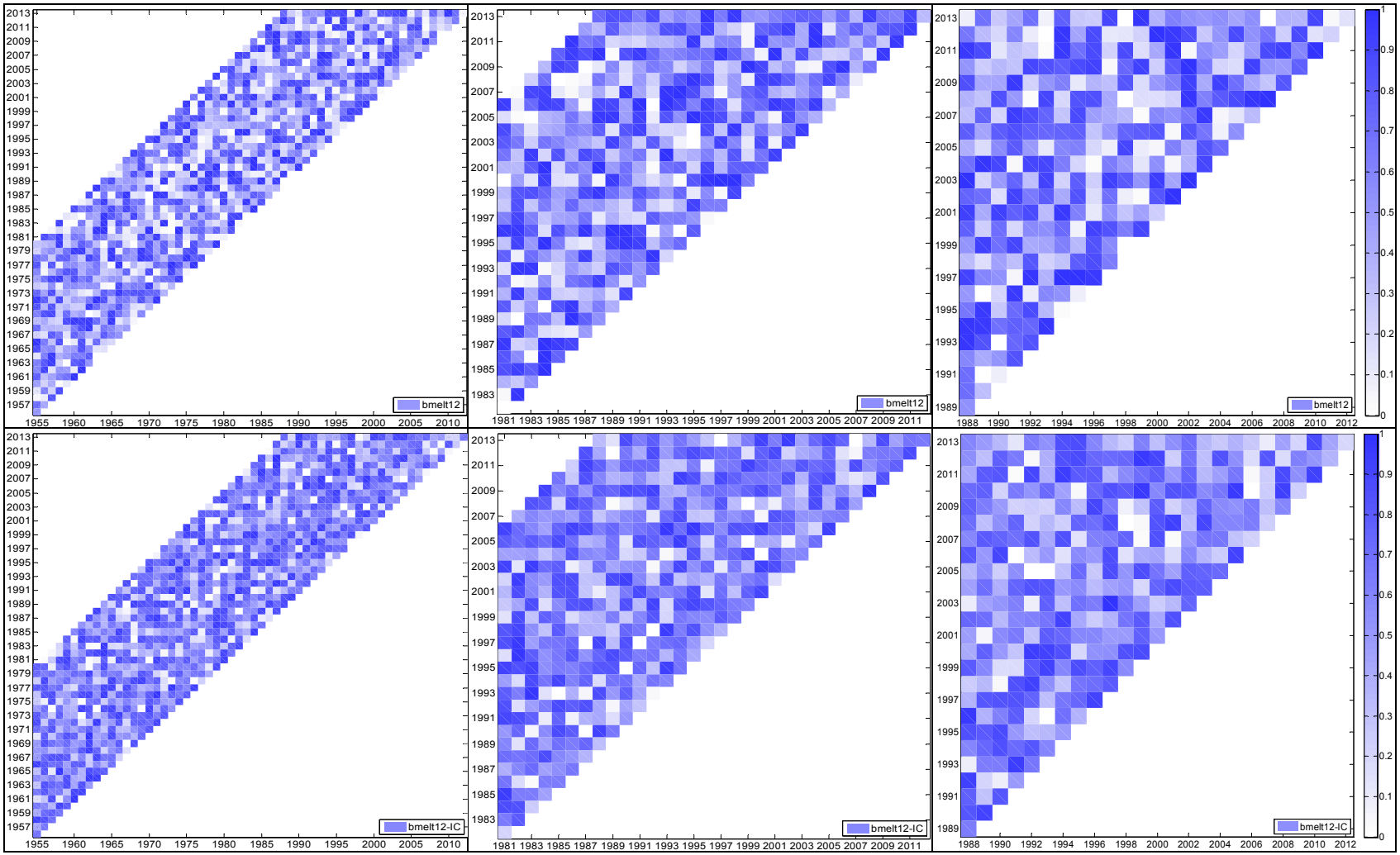
A. Figure 9. $SNOW_{100}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



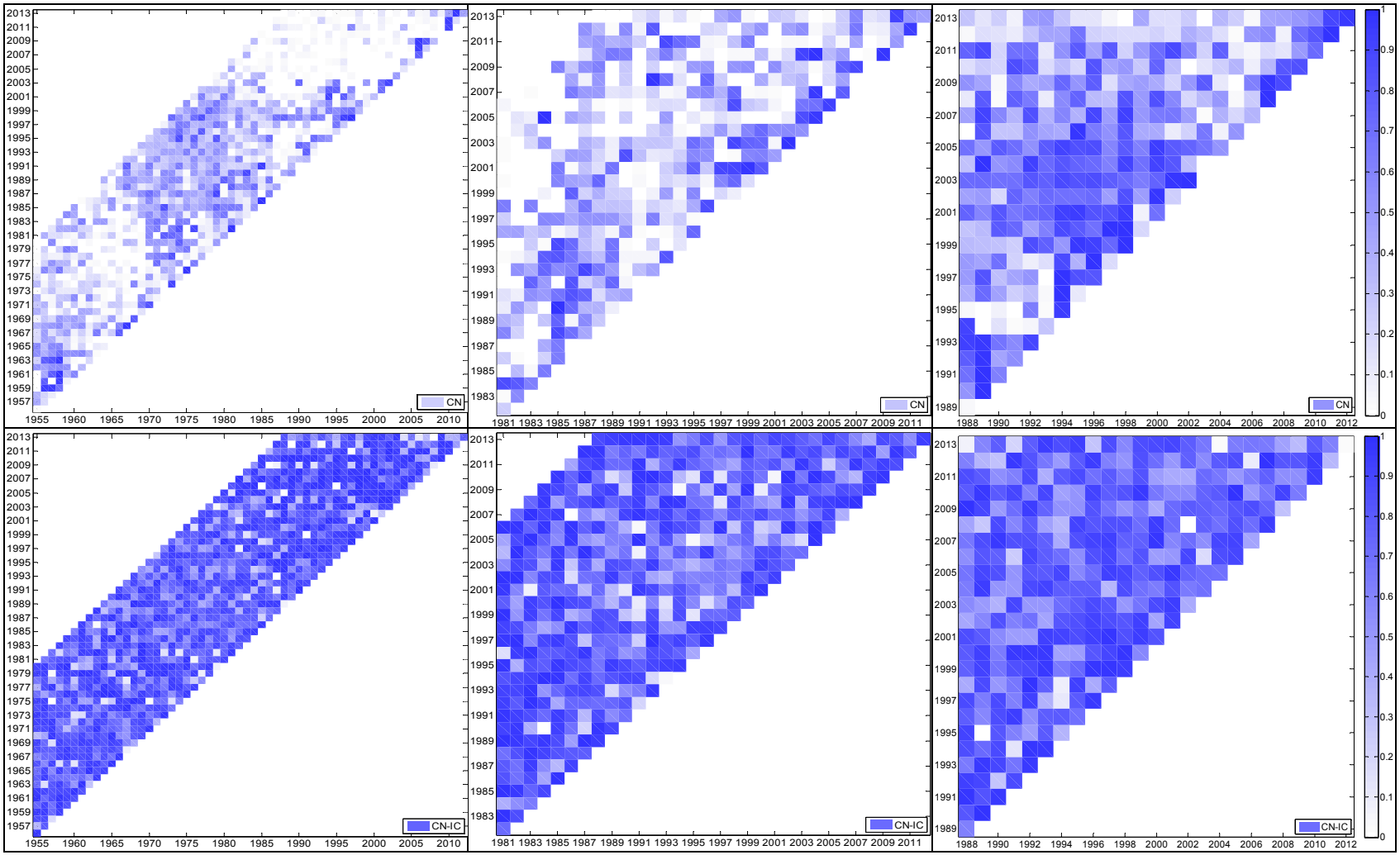
A. Figure 10. λ: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



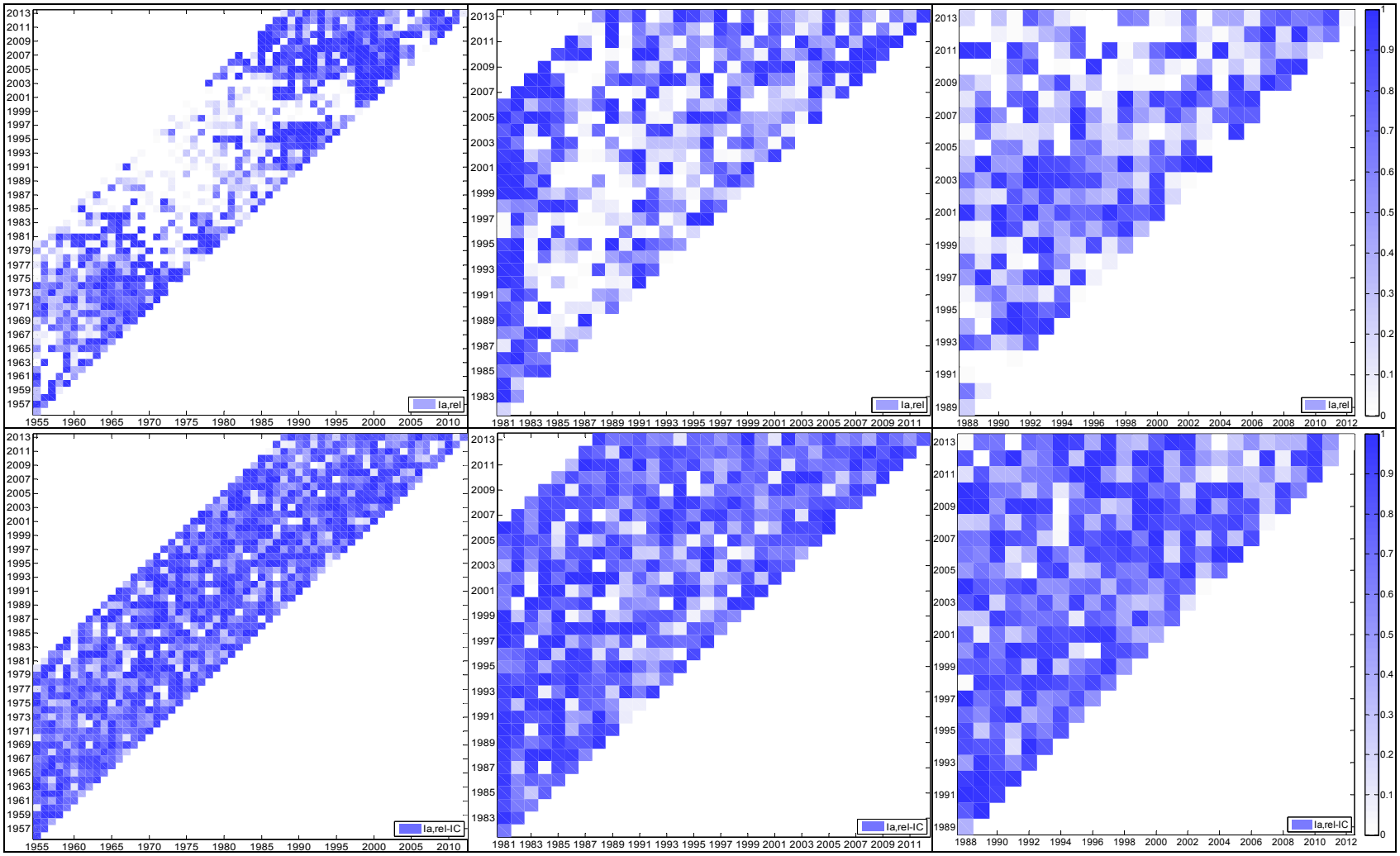
A. Figure 11. $b_{\text{melt, 6}}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



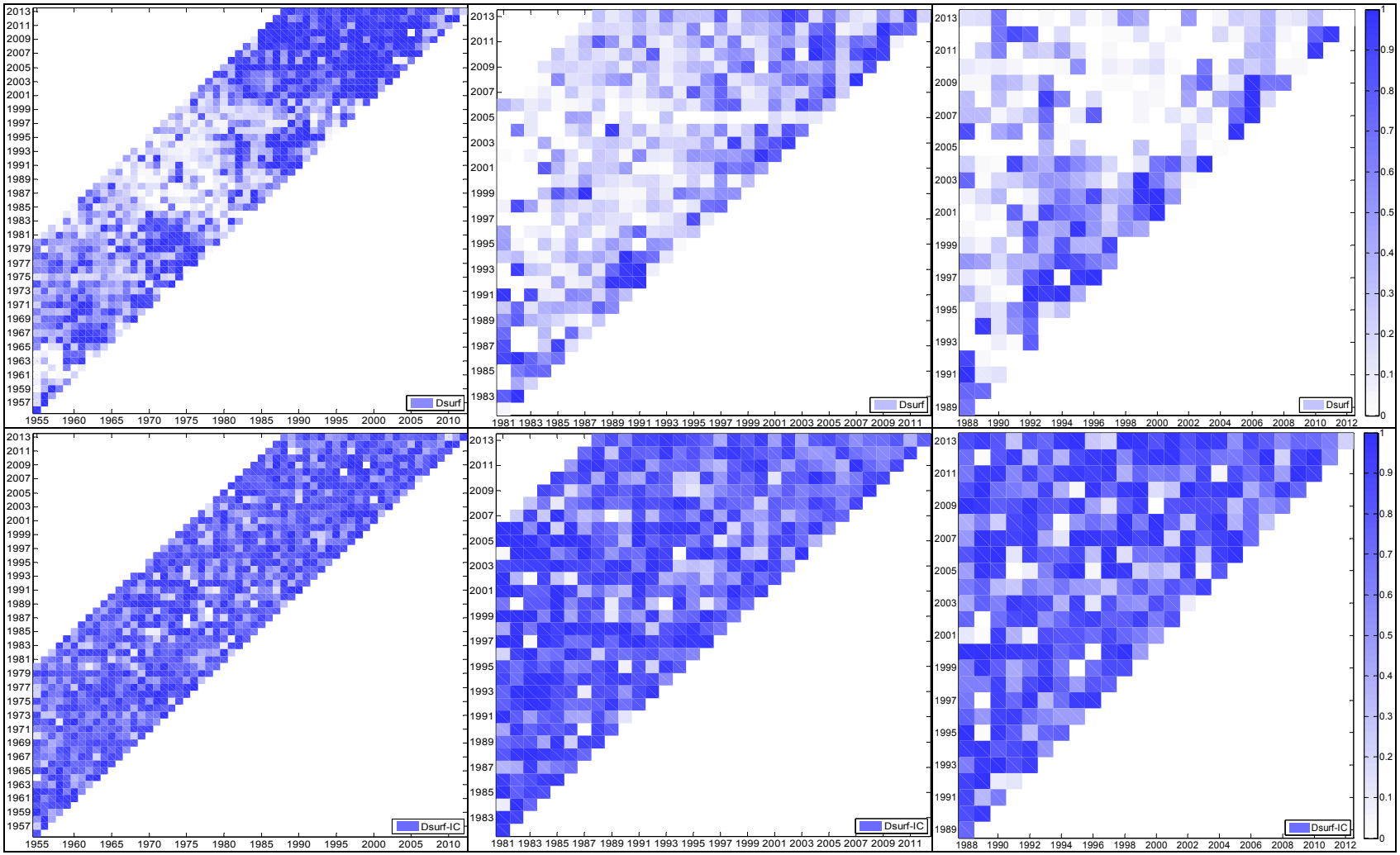
A. Figure 12. $b_{melt,12}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



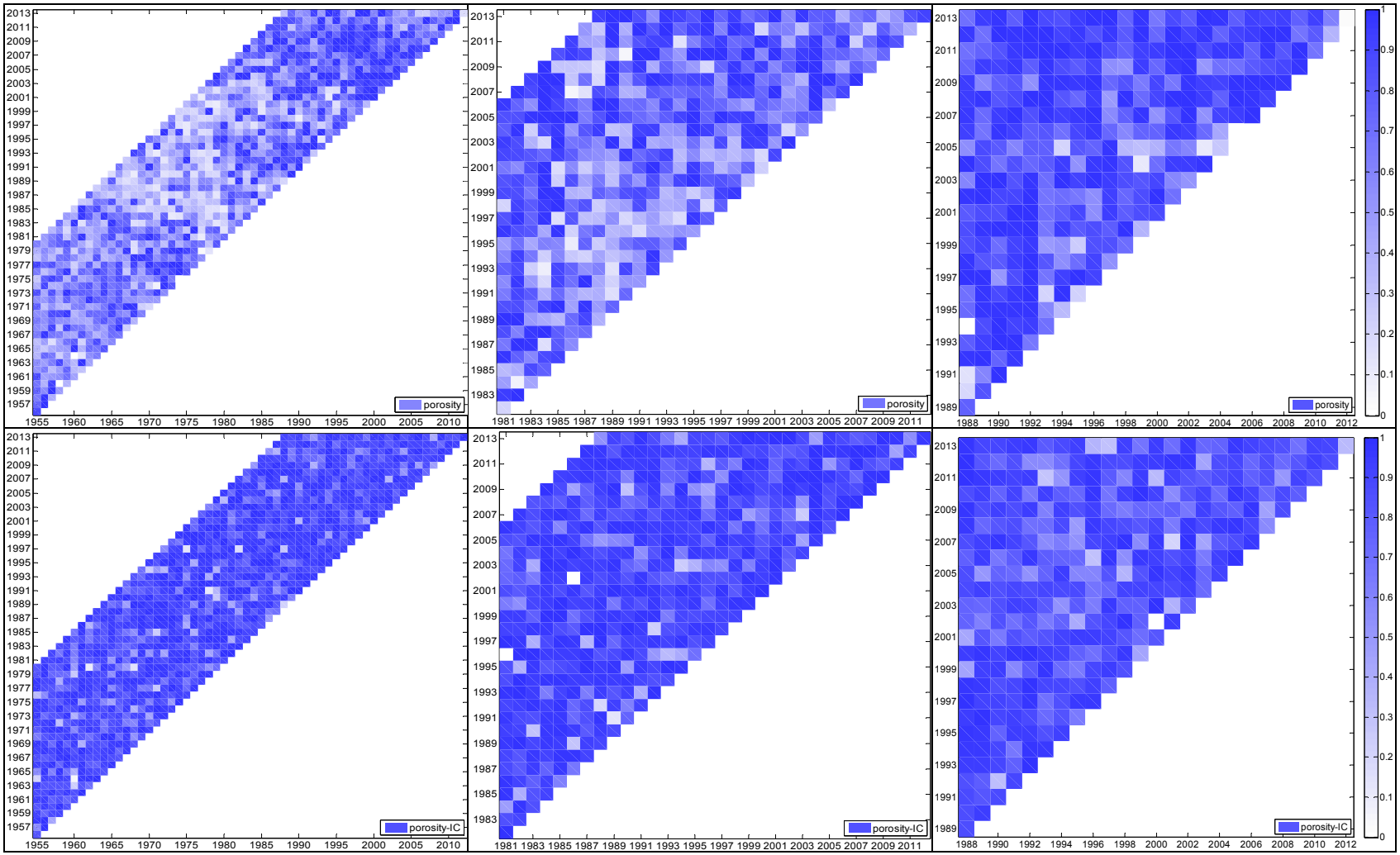
A. Figure 13. *CN*: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



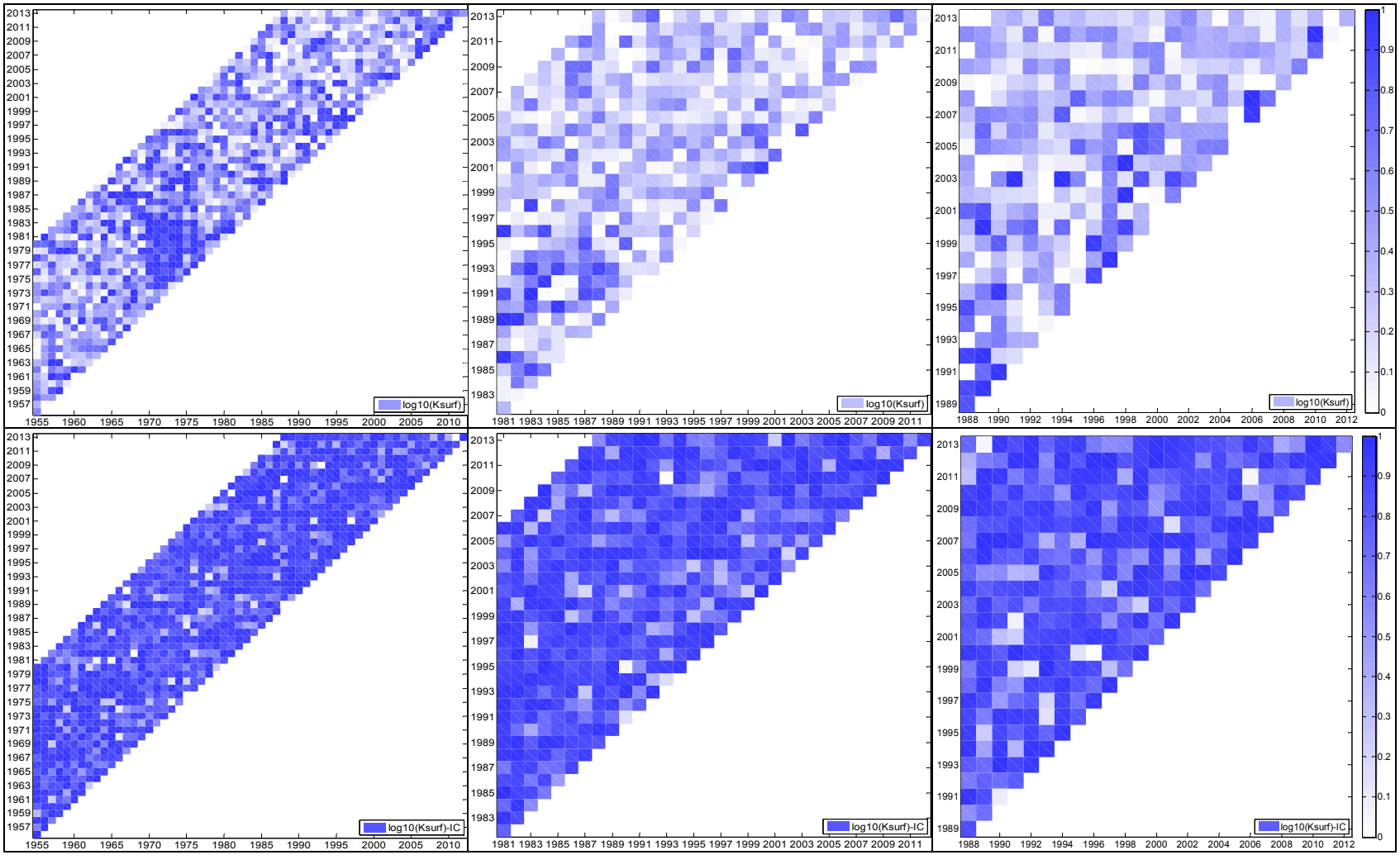
A. Figure 14. $I_{a,rel}$: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



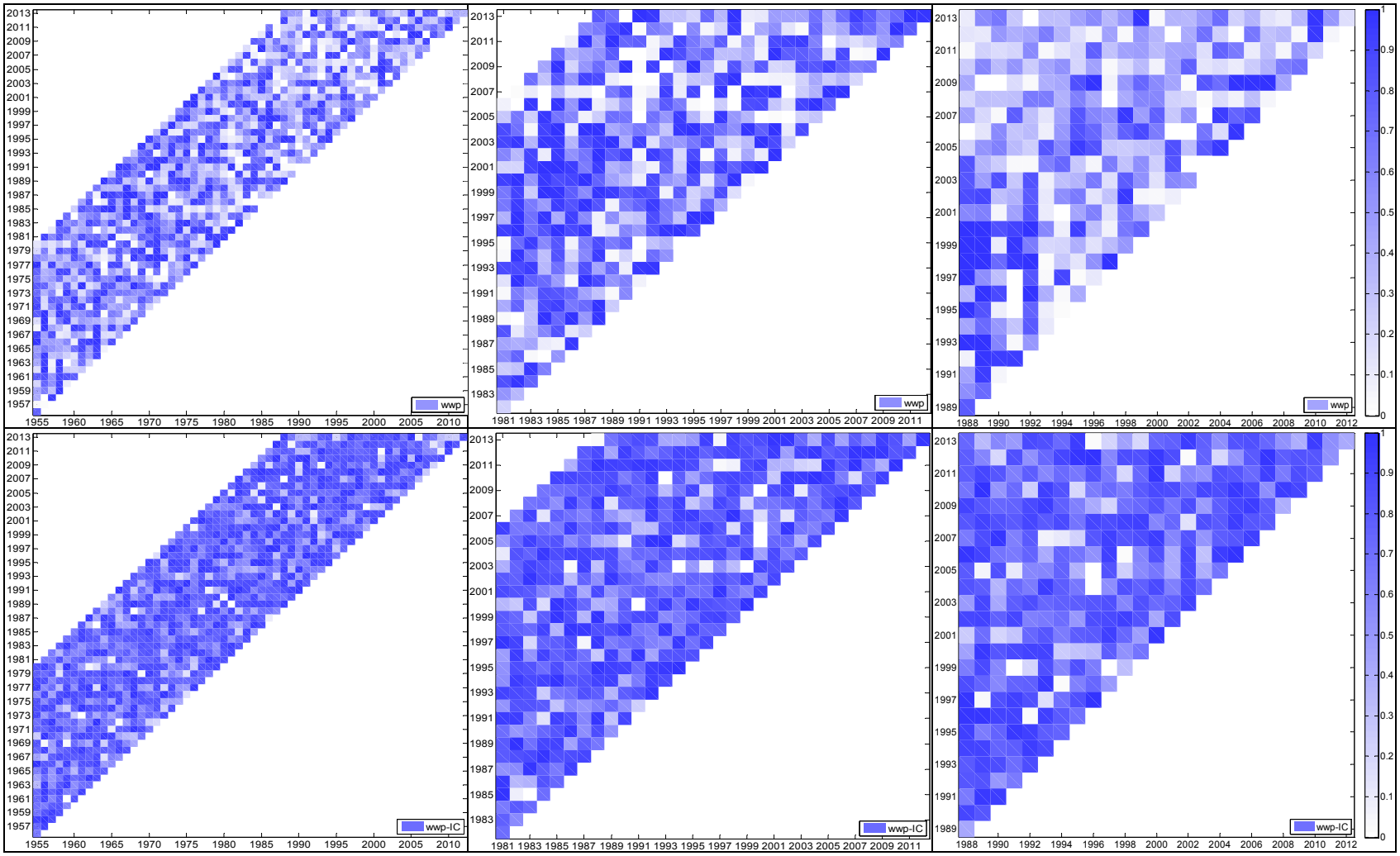
A. Figure 15. D_{surf}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



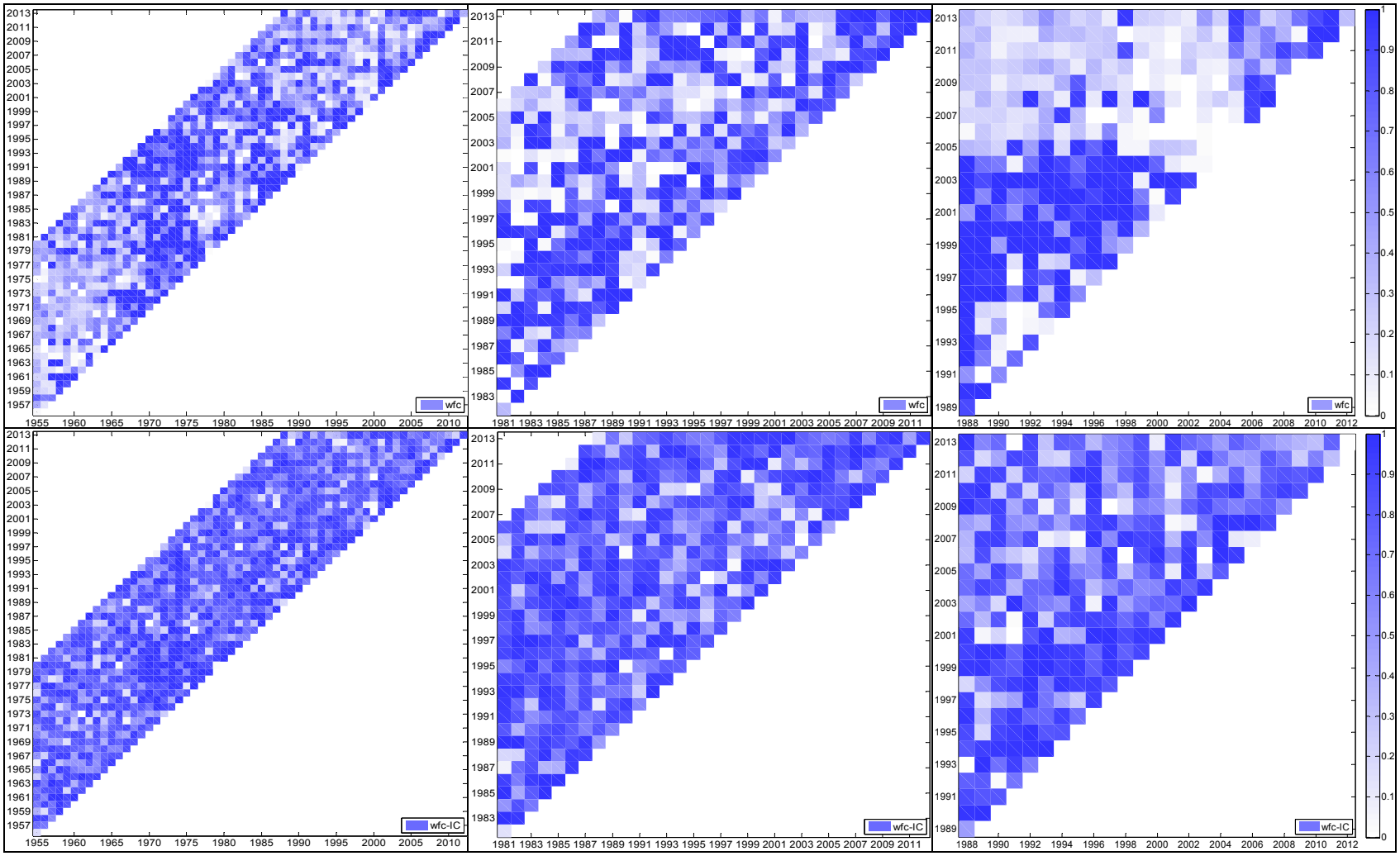
A. Figure 16. *porosity*: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



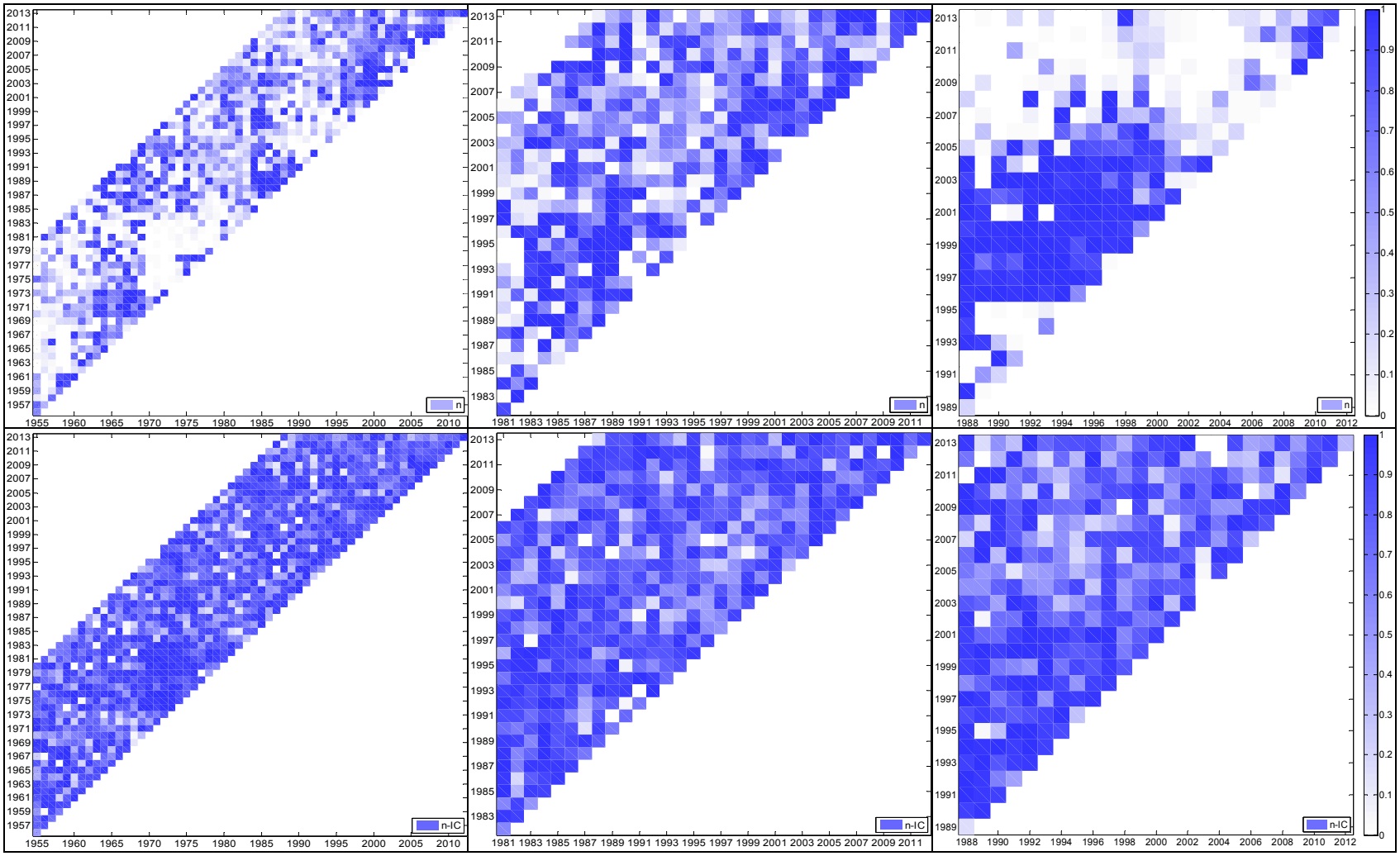
A. Figure 17. K_{surf}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



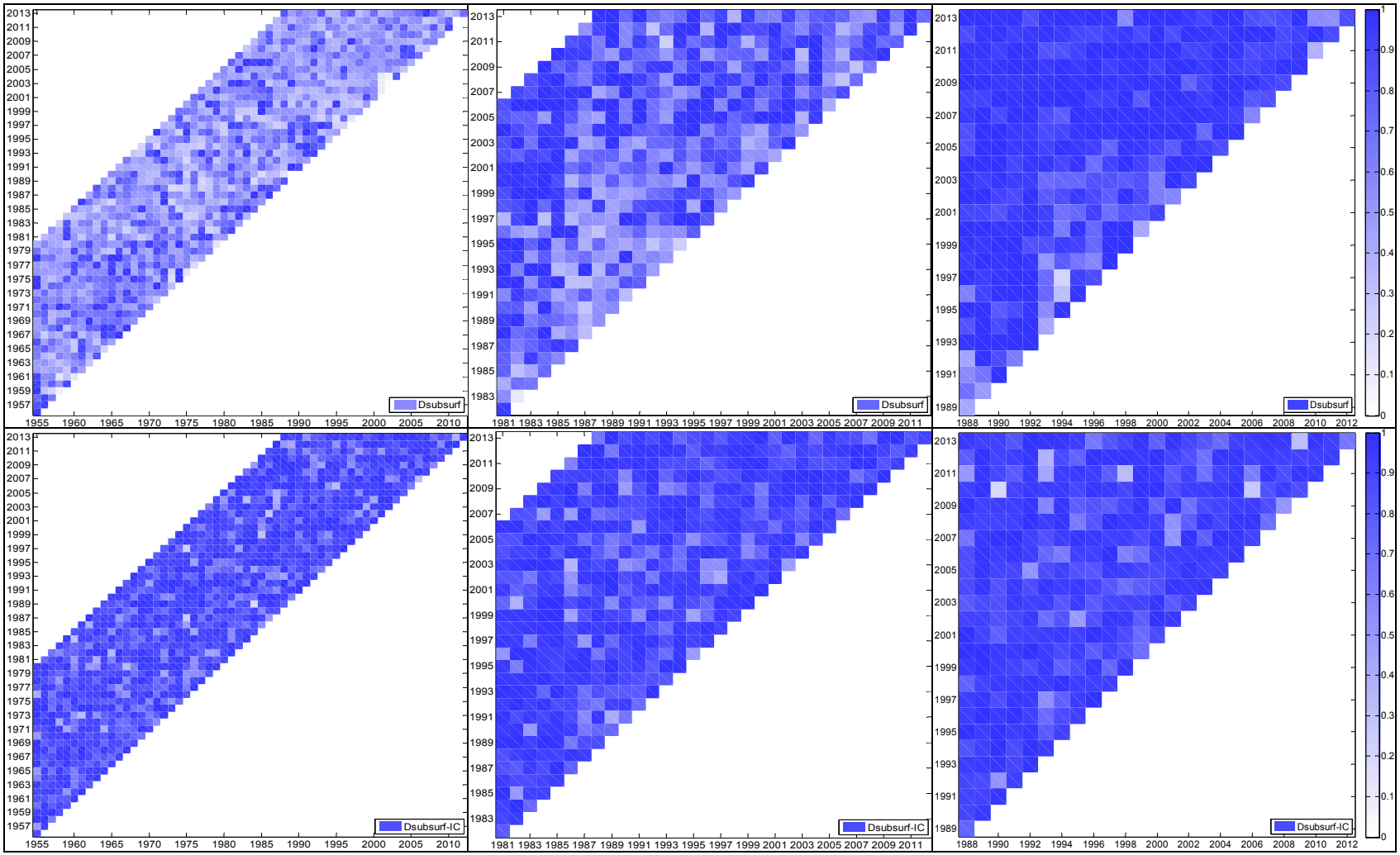
A. Figure 18. w_{wp}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



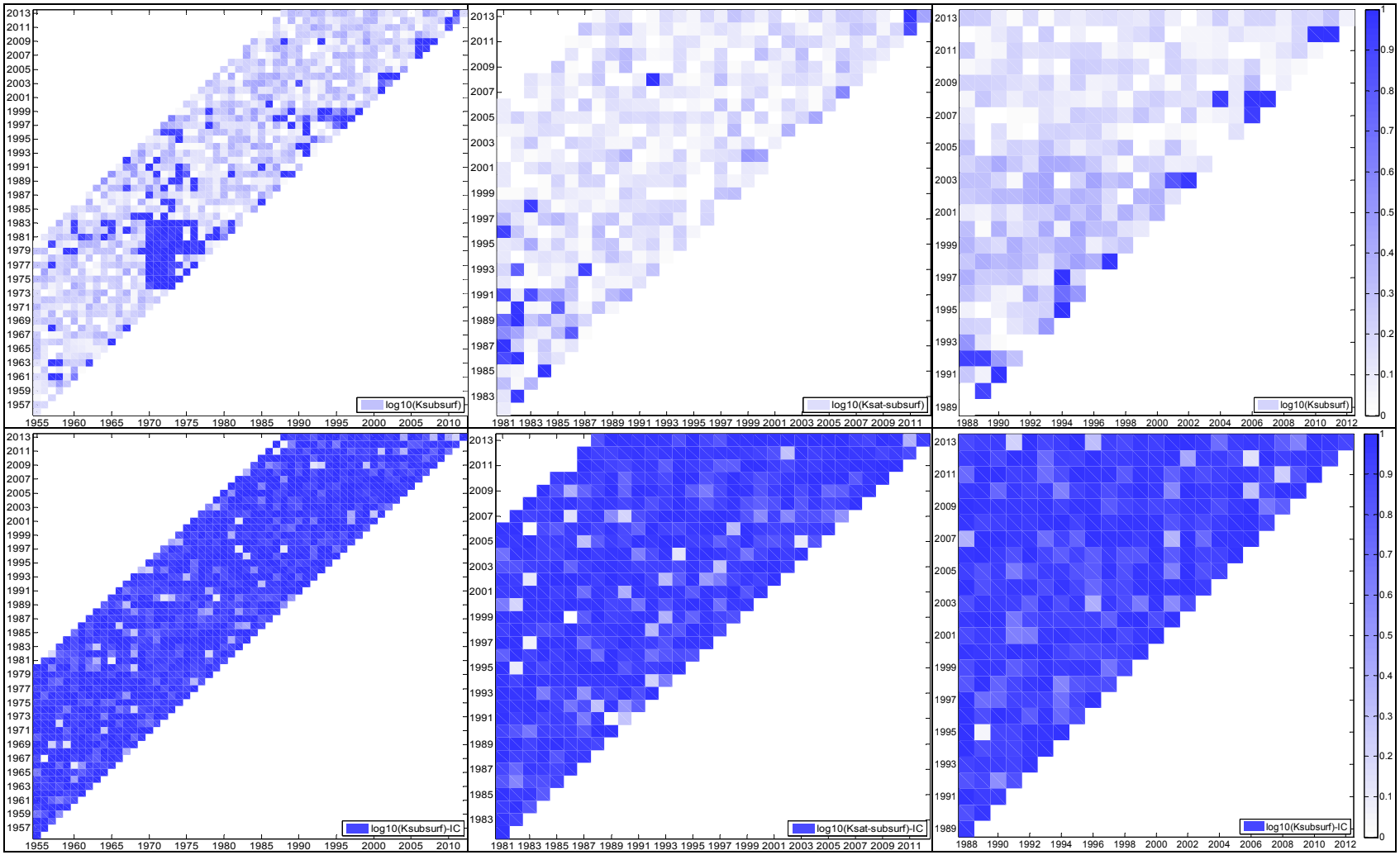
A. Figure 19. w_{fc}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



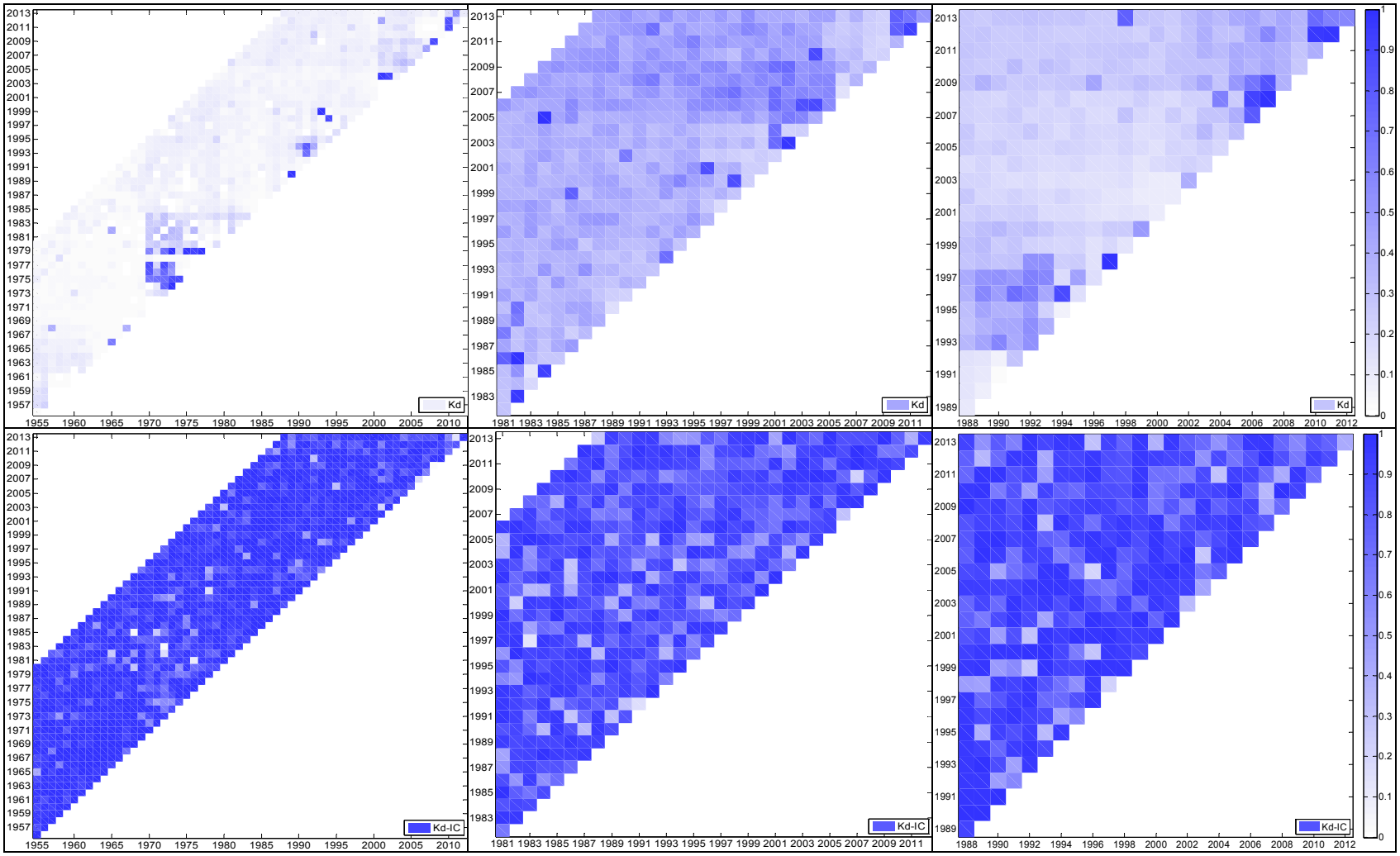
A. Figure 20. *n*: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



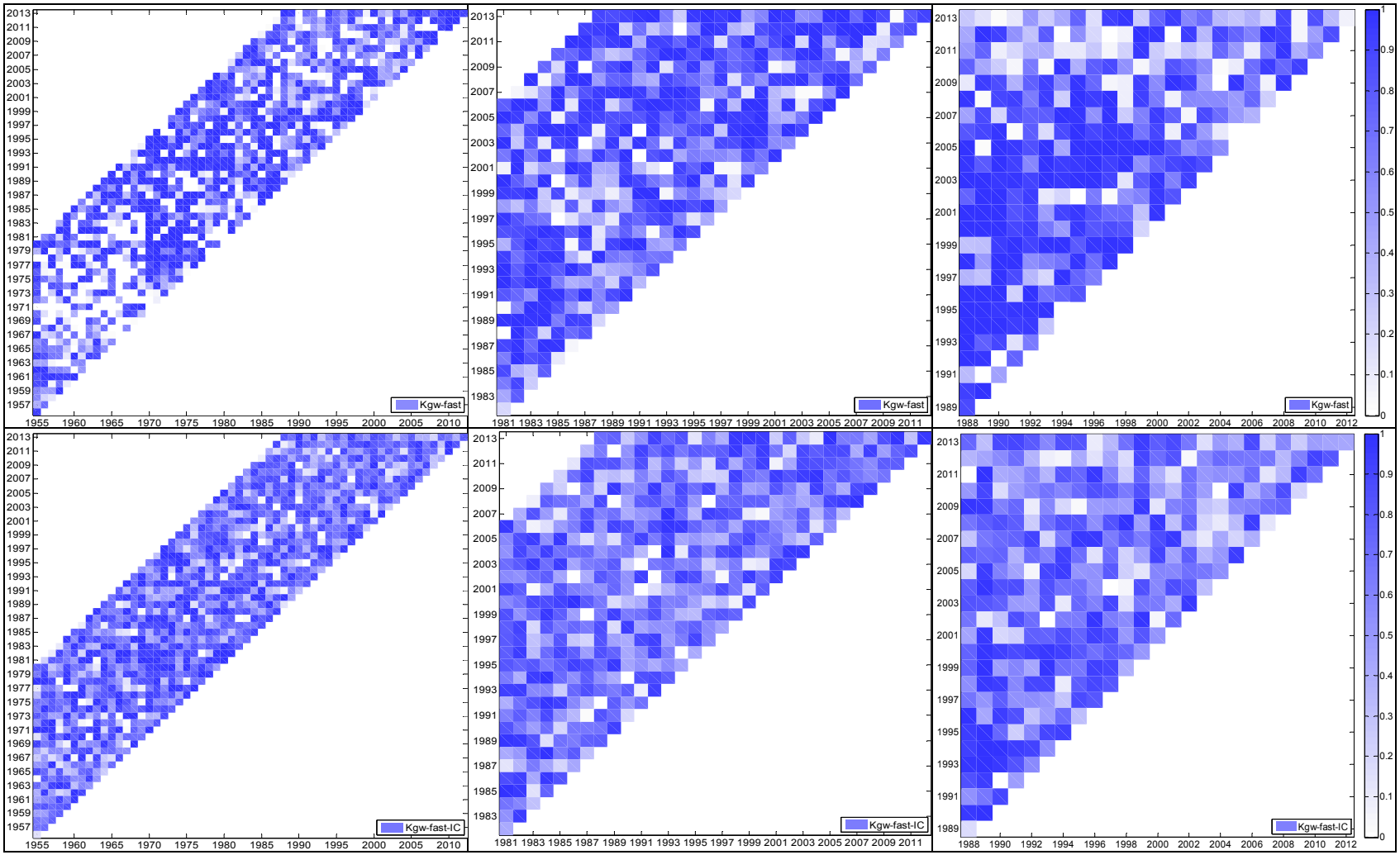
A. Figure 21. $D_{\text{sub-surf}}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



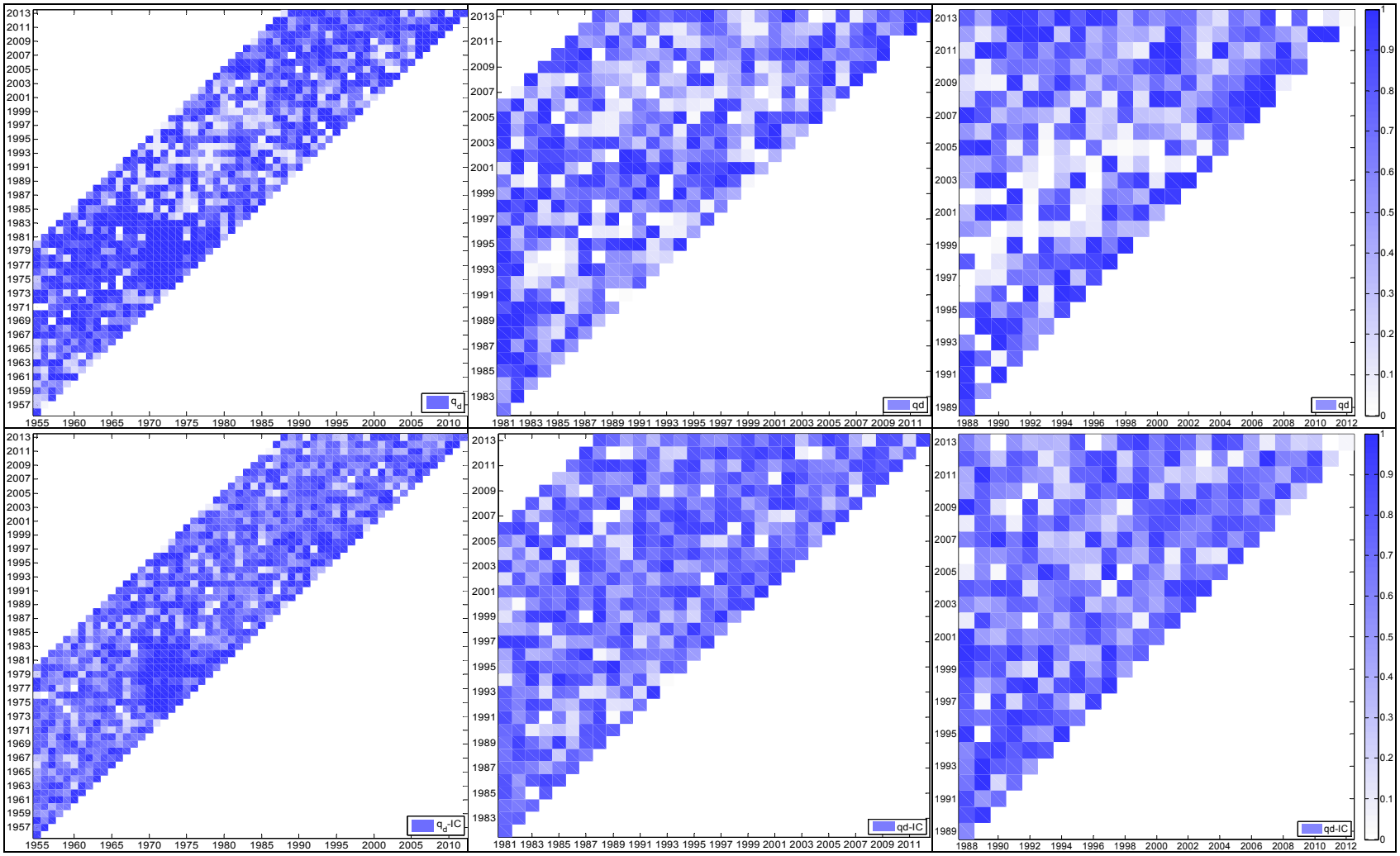
A. Figure 22. $K_{\text{sub-surf}}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



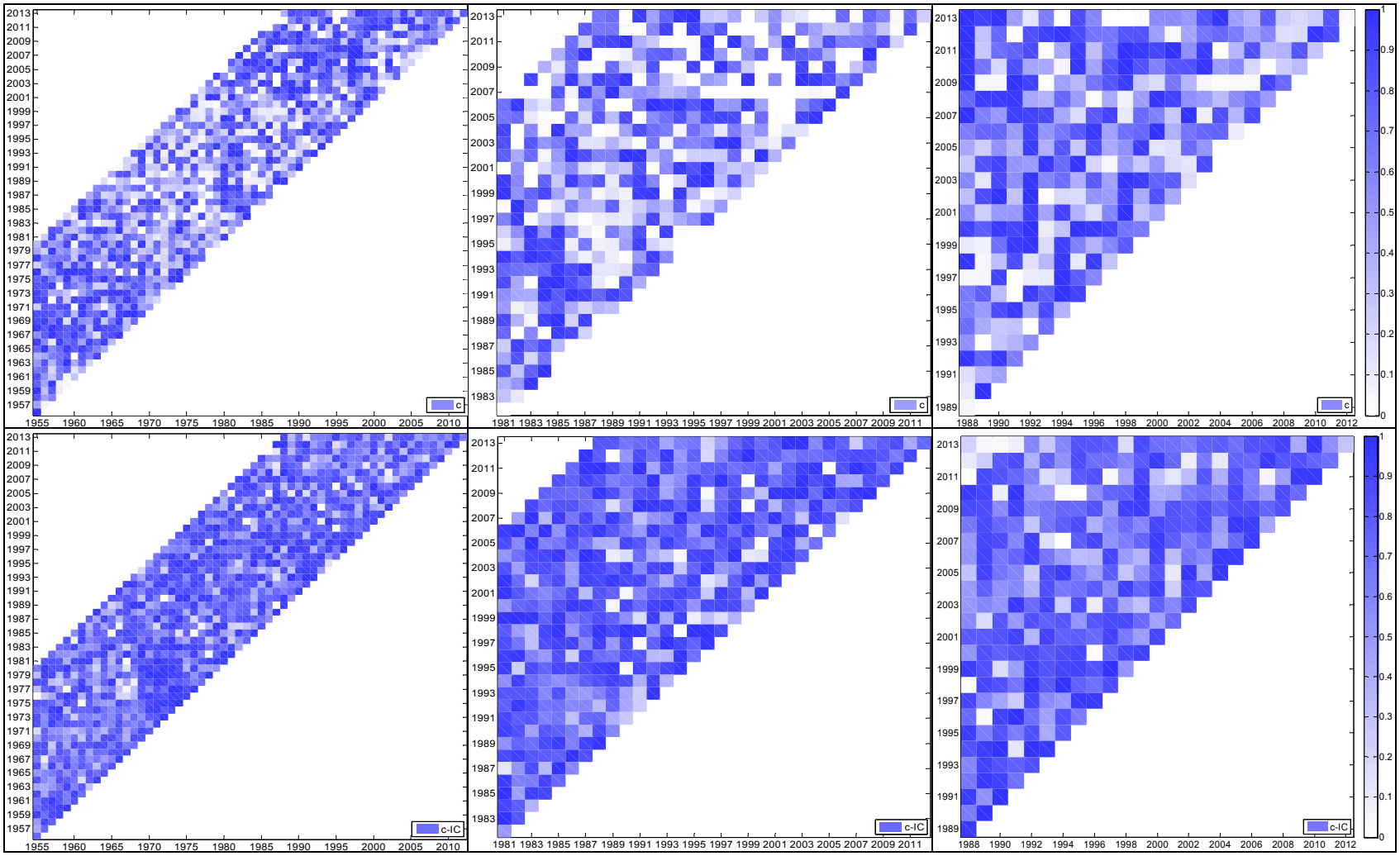
A. Figure 23. K_d: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



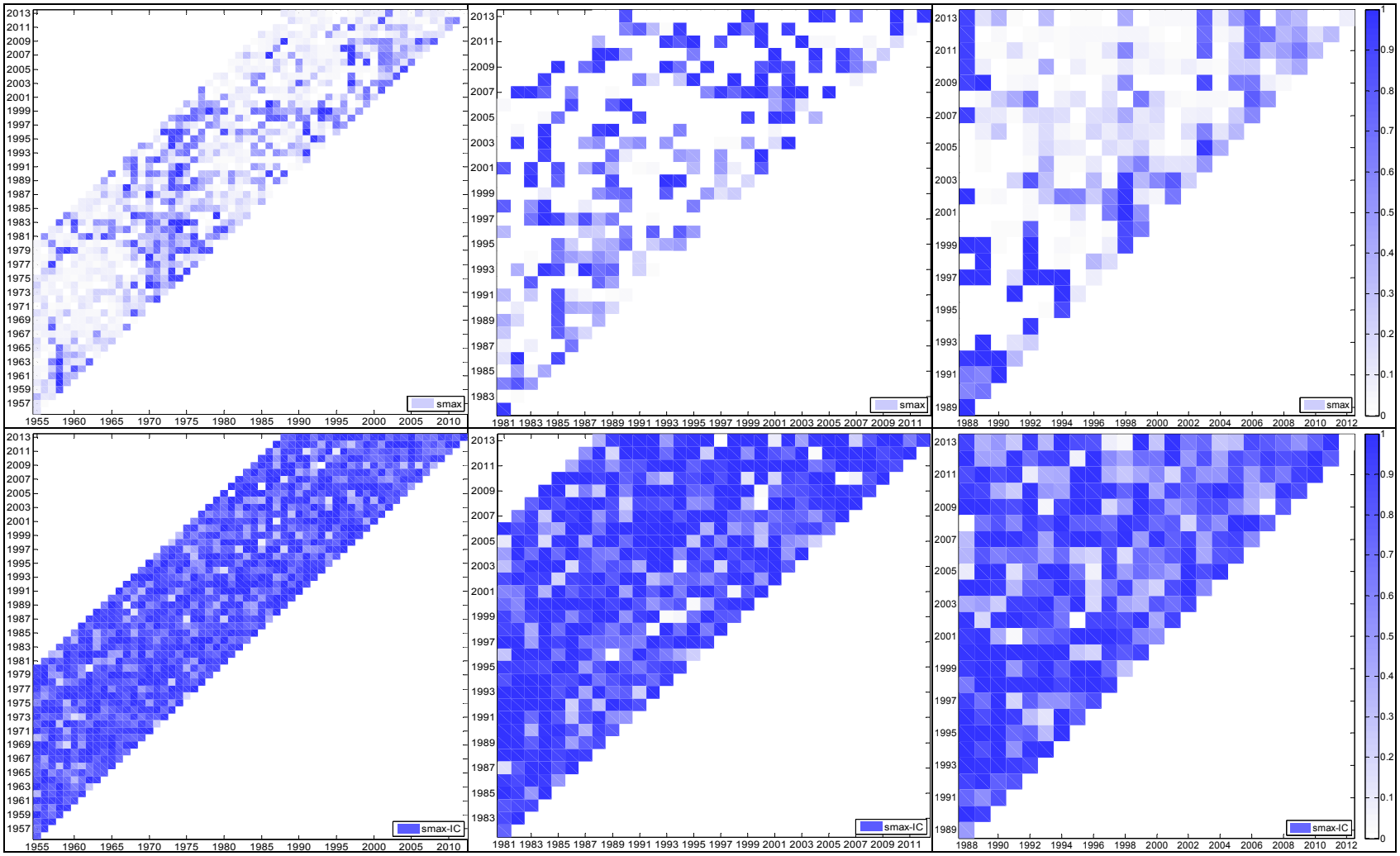
A. Figure 24. $K_{\text{gw-fast}}$: Median parameter values (top) and *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



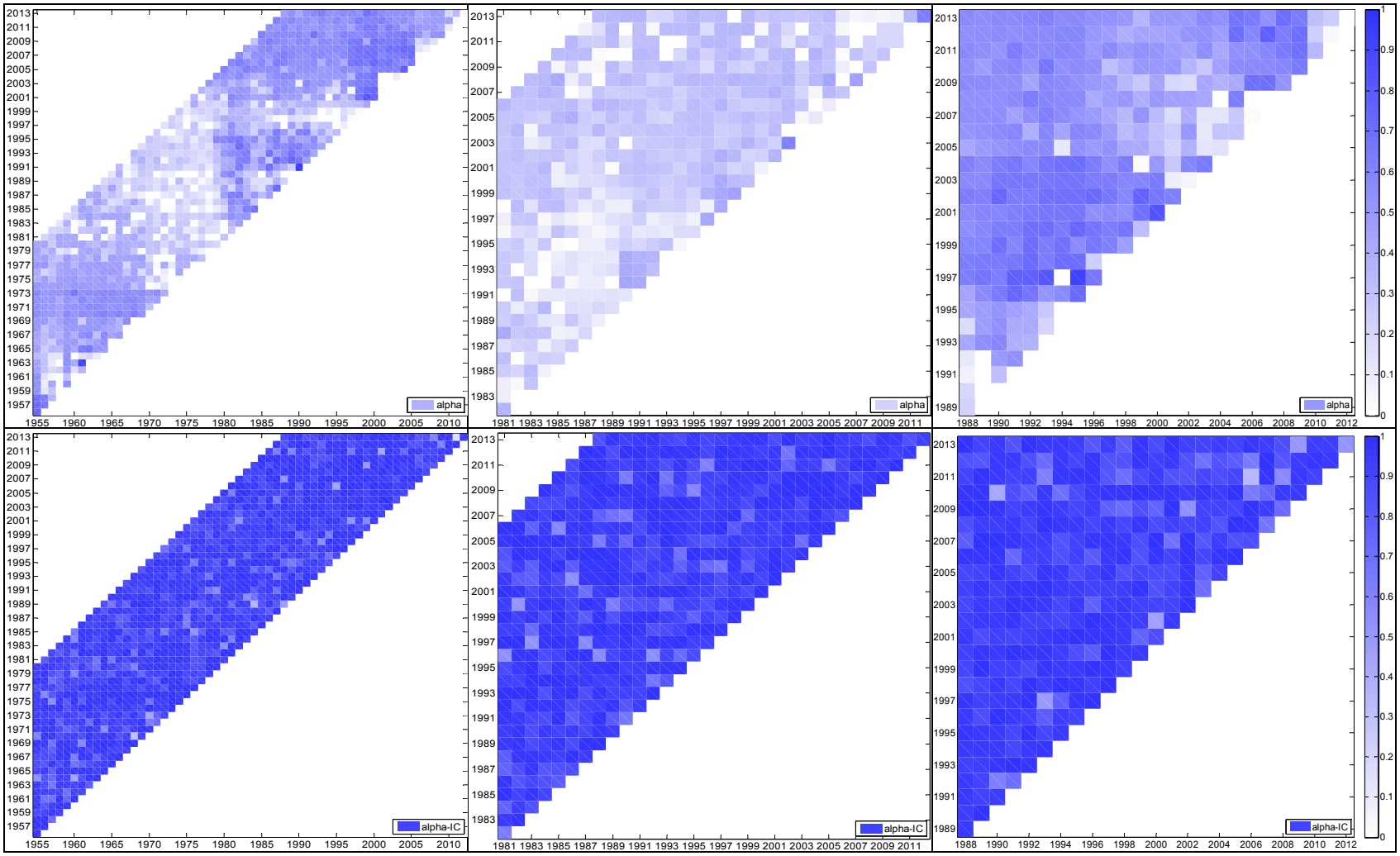
A. Figure 25. q_d: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



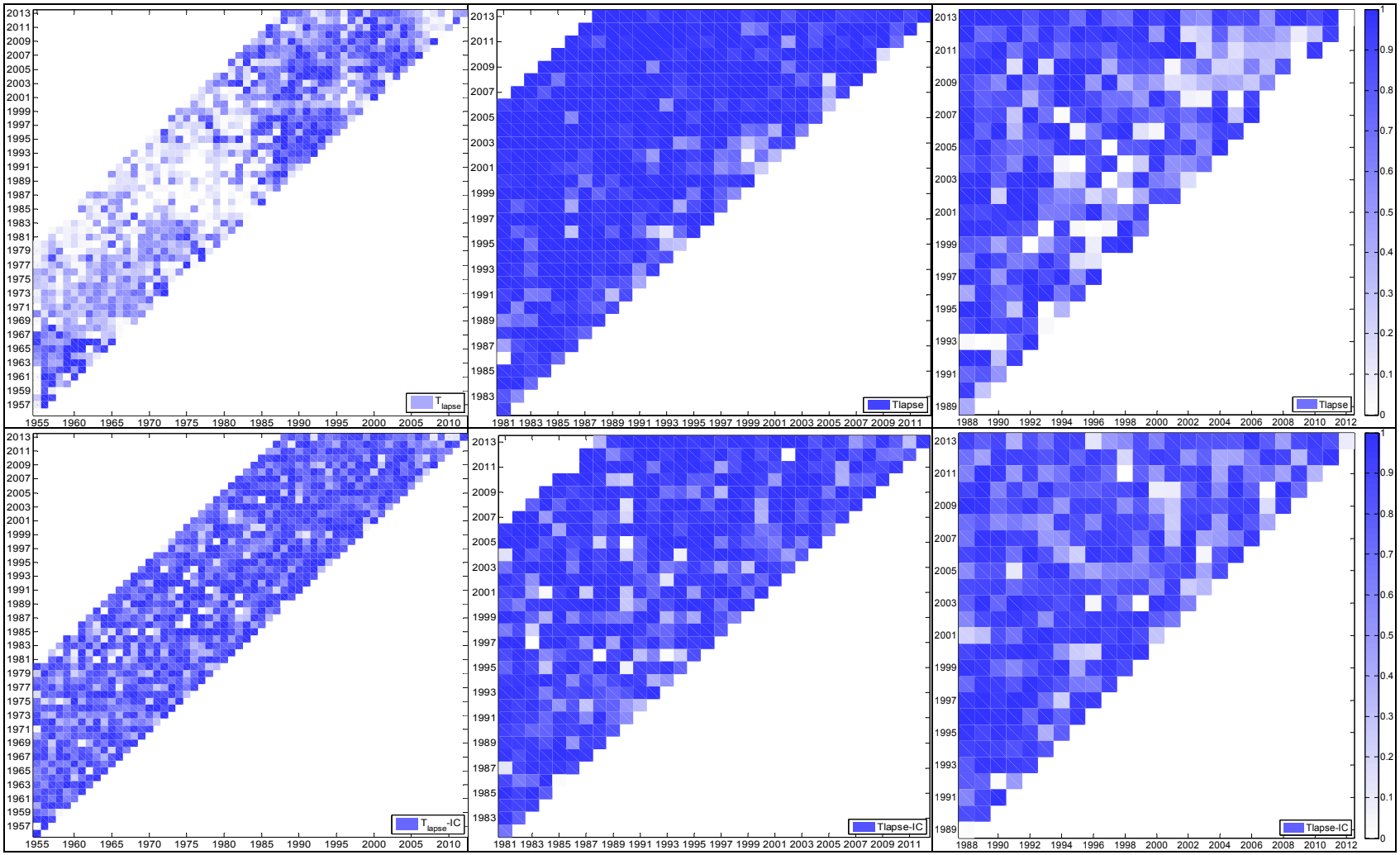
A. Figure 26. *c*: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



A. Figure 27. s_{max}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



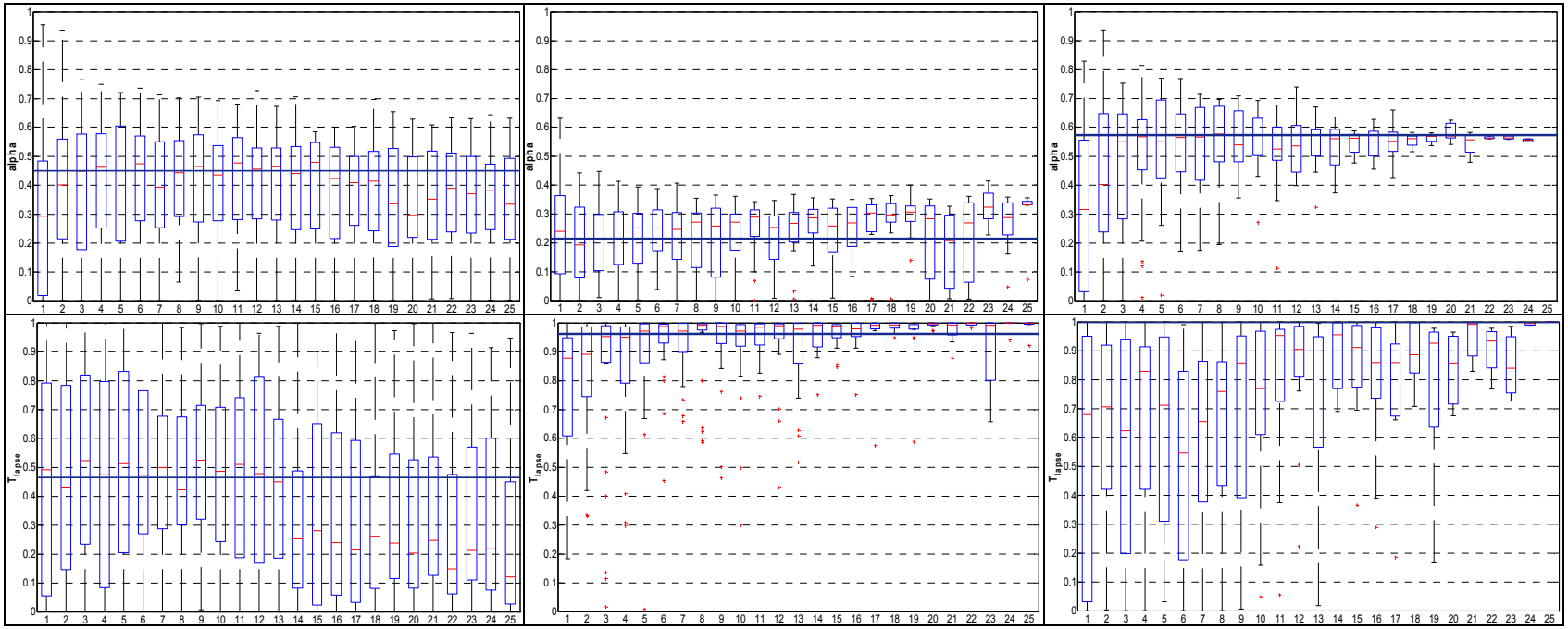
A. Figure 28. α: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.



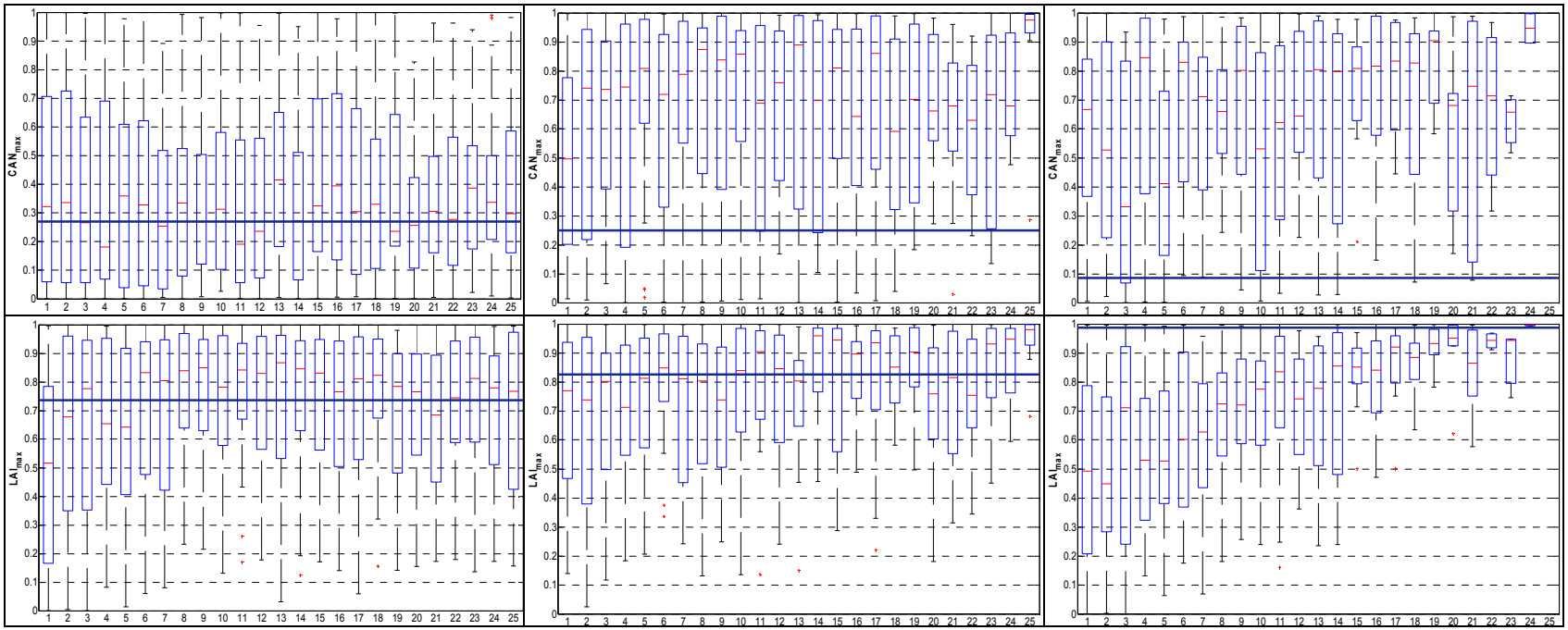
A. Figure 29. T_{lapse}: Median parameter values (top) and the *IC* statistic (bottom panels): the Kolubara, Toplica and Mlava River catchments.

APPENDIX E. Median values of normalised Pareto-optimal parameters against the length of the calibration period: semi-lumped BASIC version of the model

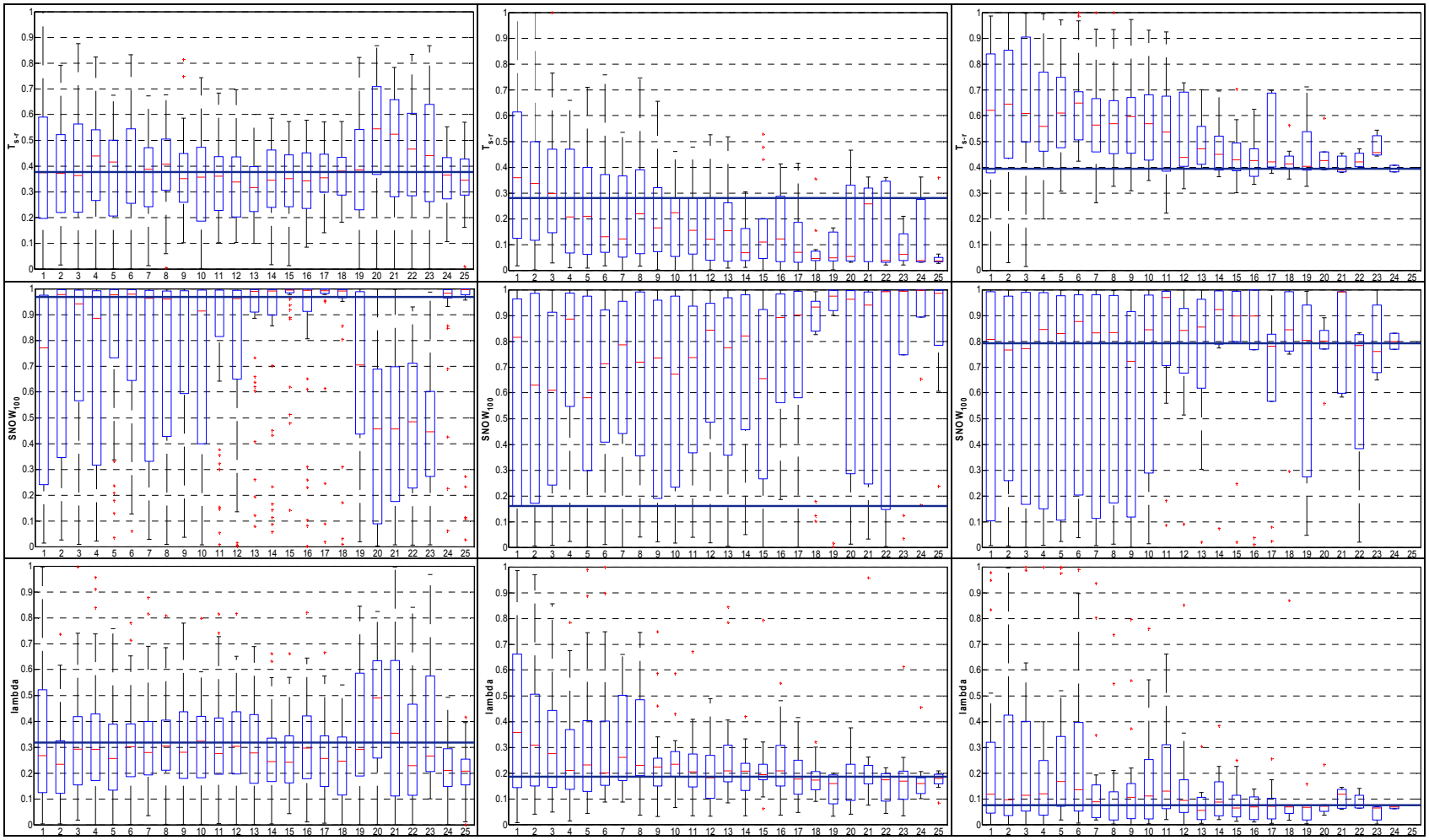
Box plots denote median values of normalised Pareto-optimal parameters, obtained over all calibration periods of given length (abscissa values). Thick lines denote median value of the Pareto-optimal parameters obtained over the full hydrologic record period.



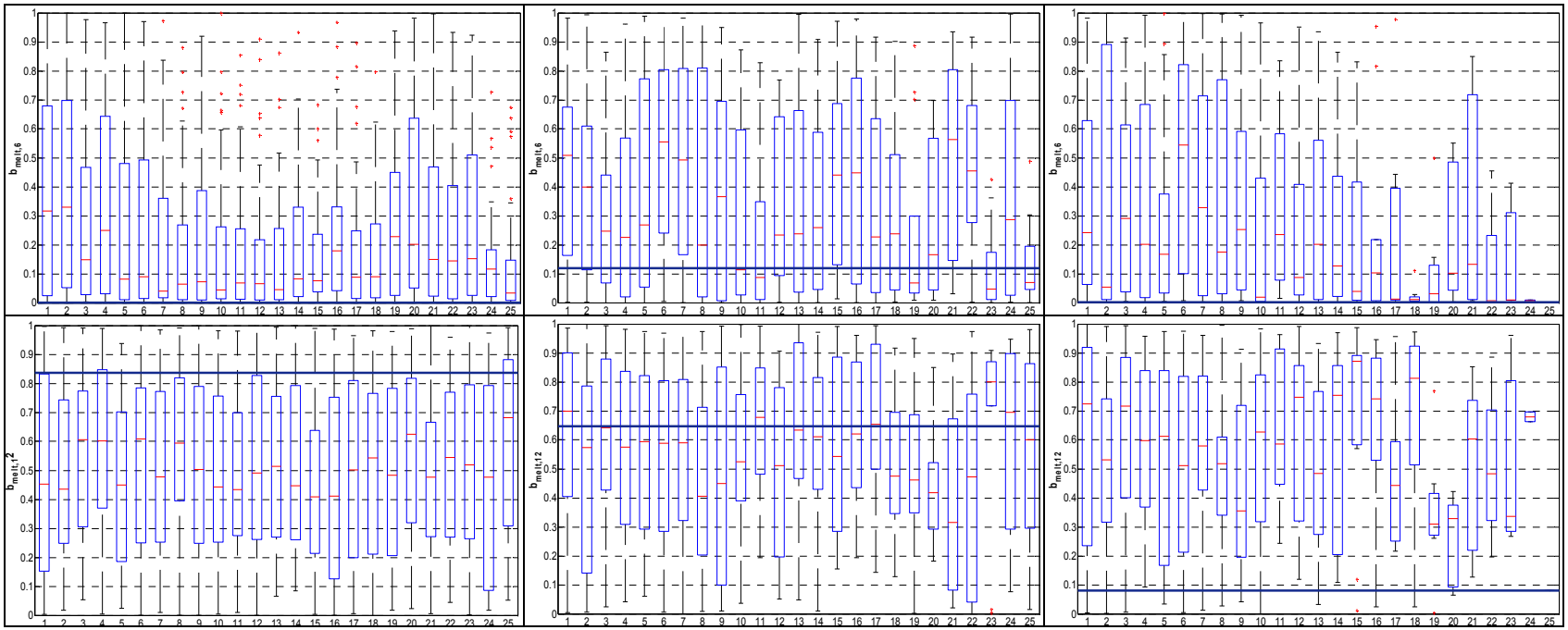
A. Figure 30. Median values of normalised Pareto-optimal precipitation gradients with elevation and lapse rates: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



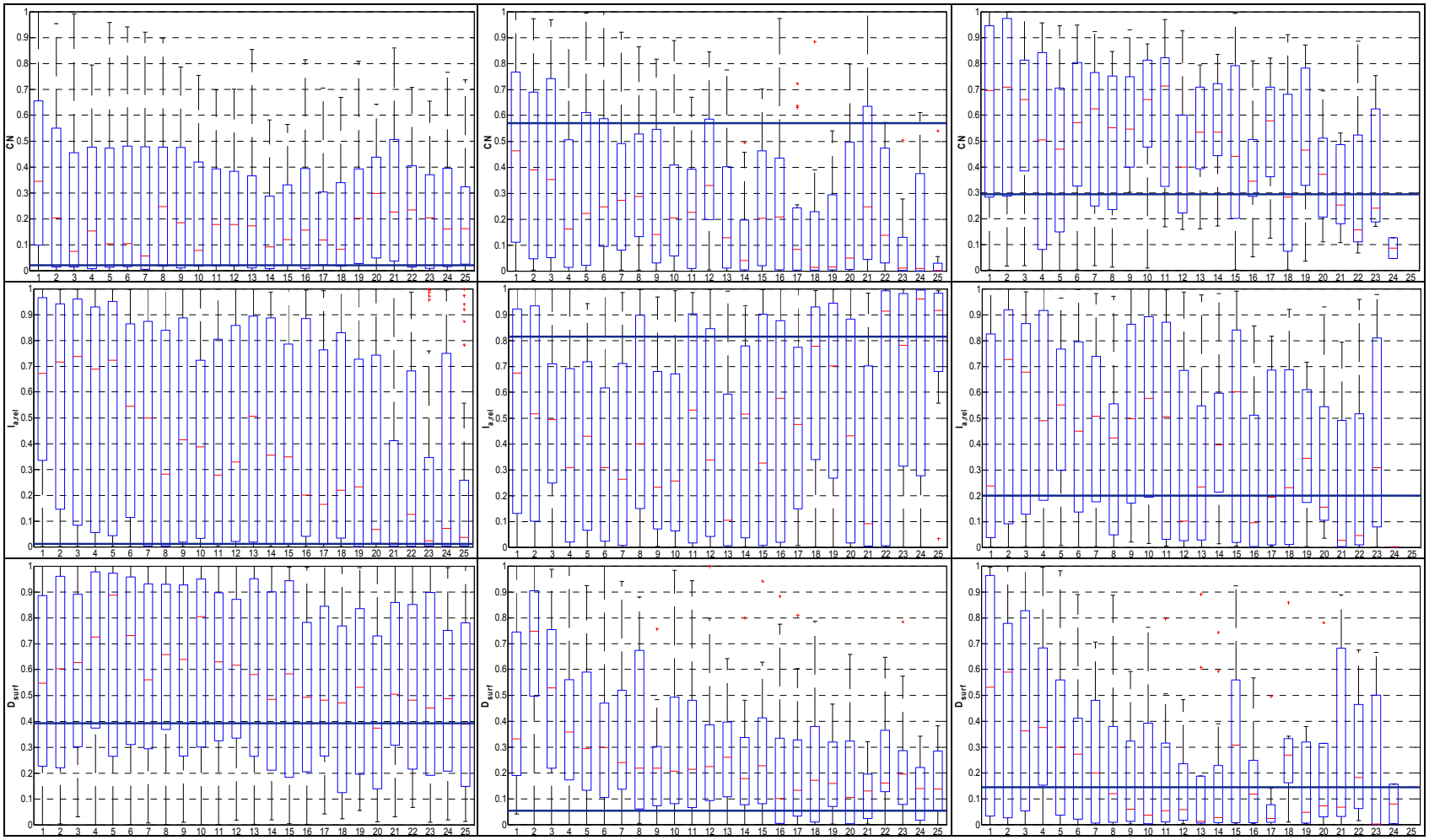
A. Figure 31. Median values of normalised Pareto-optimal parameters of the interception routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



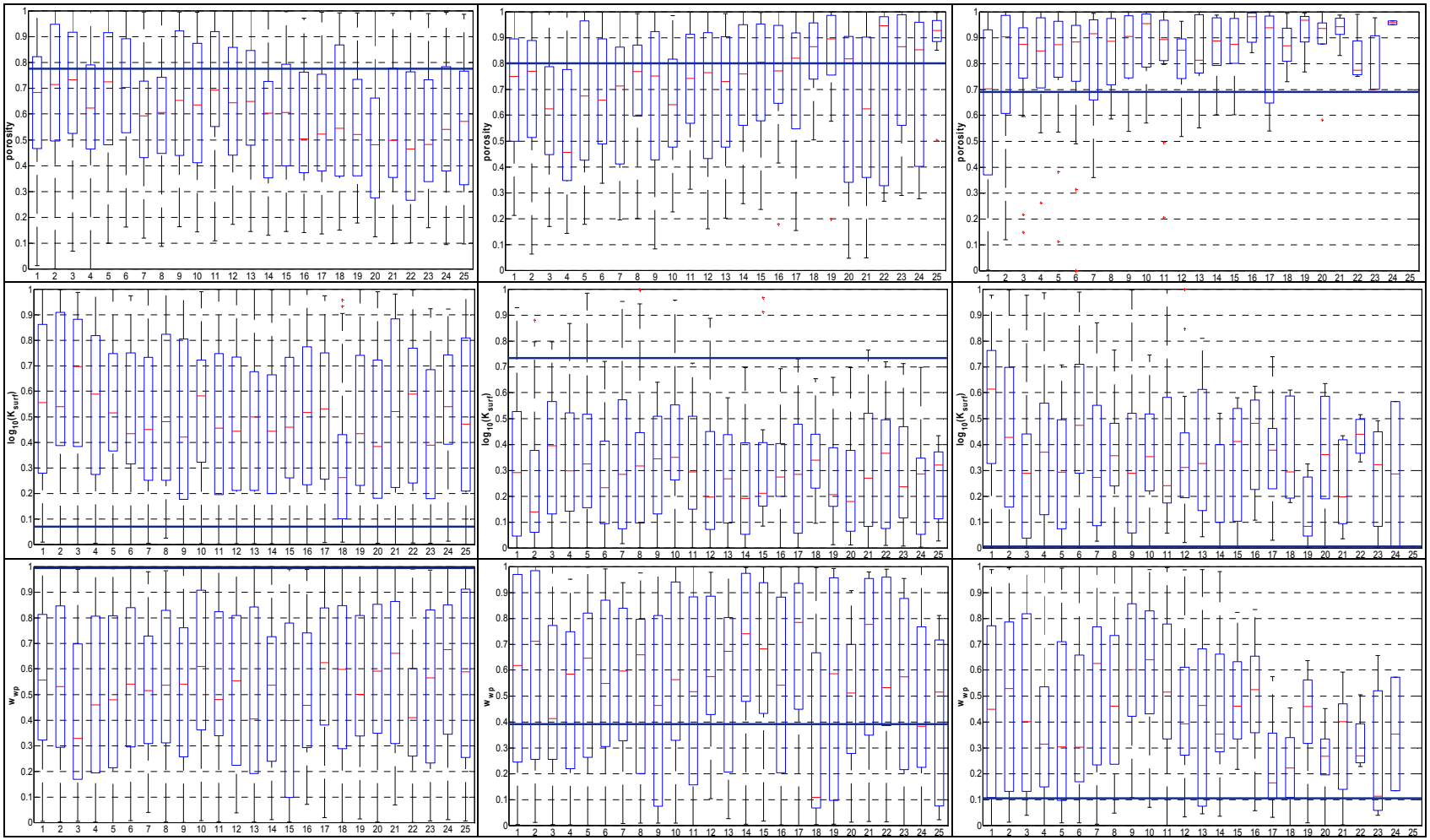
A. Figure 32. Median values of normalised Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



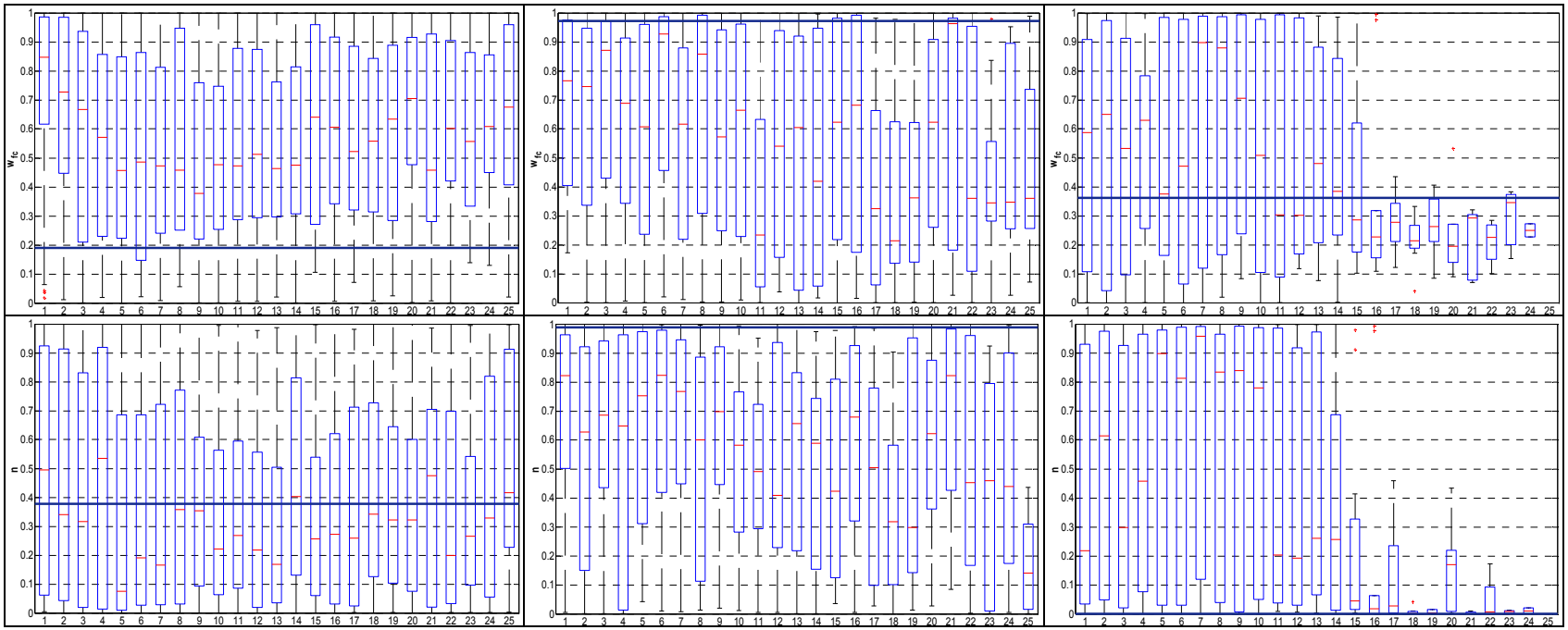
A. Figure 32 (continued). Median values of normalised Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



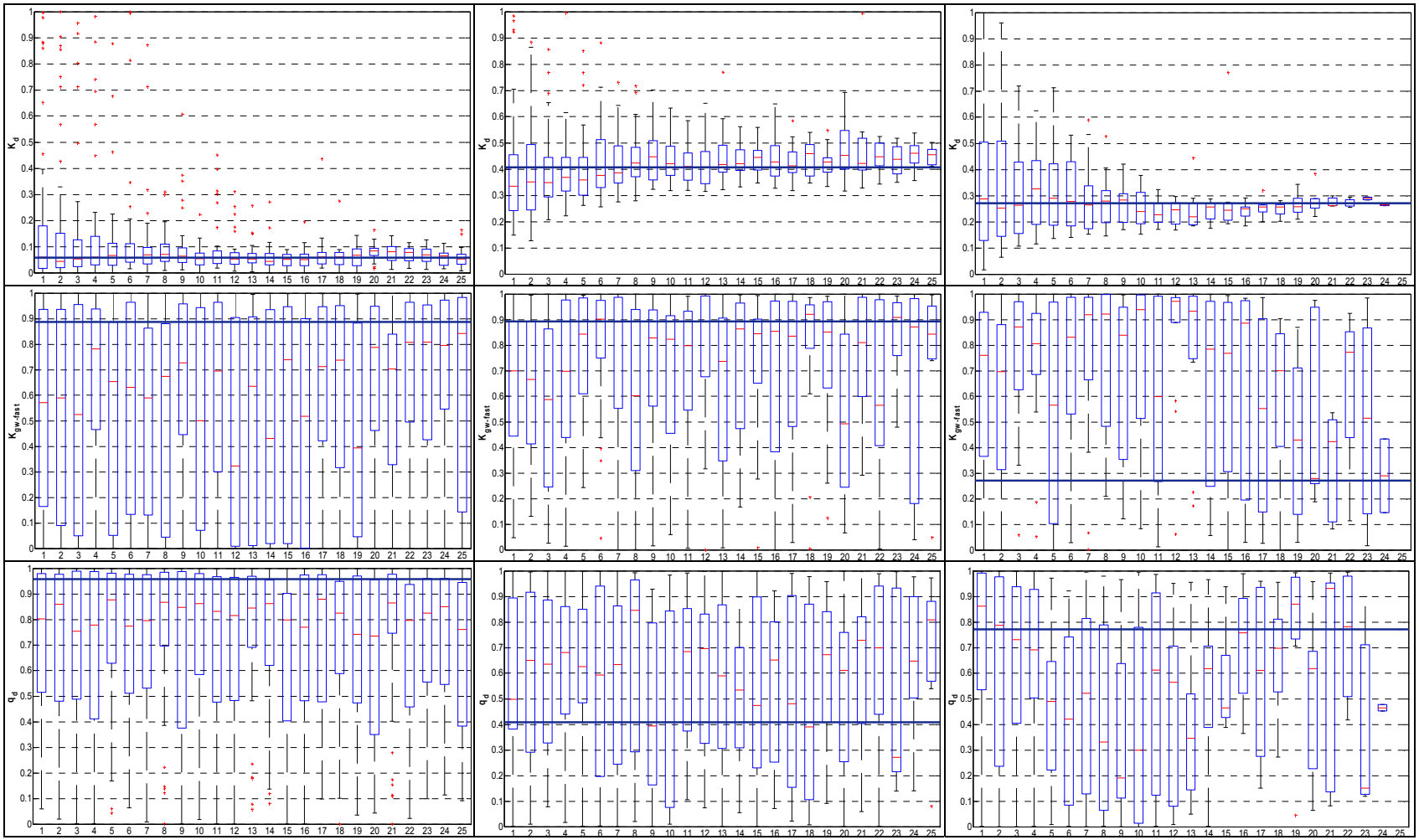
A. Figure 33. Median values of normalised Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



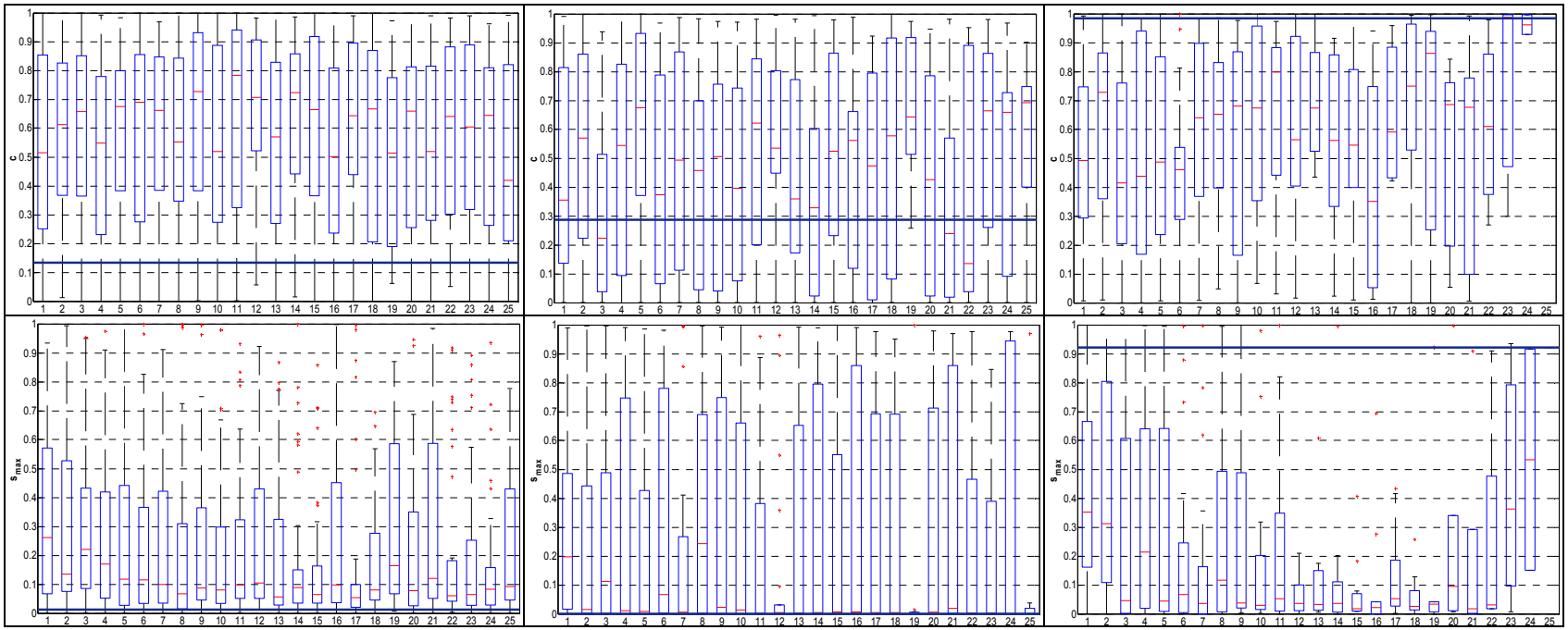
A. Figure 33 (continued). Median values of normalised Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 33 (continued). Median values of normalised Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 34. Median values of normalised Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 34 (continued). Median values of normalised Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

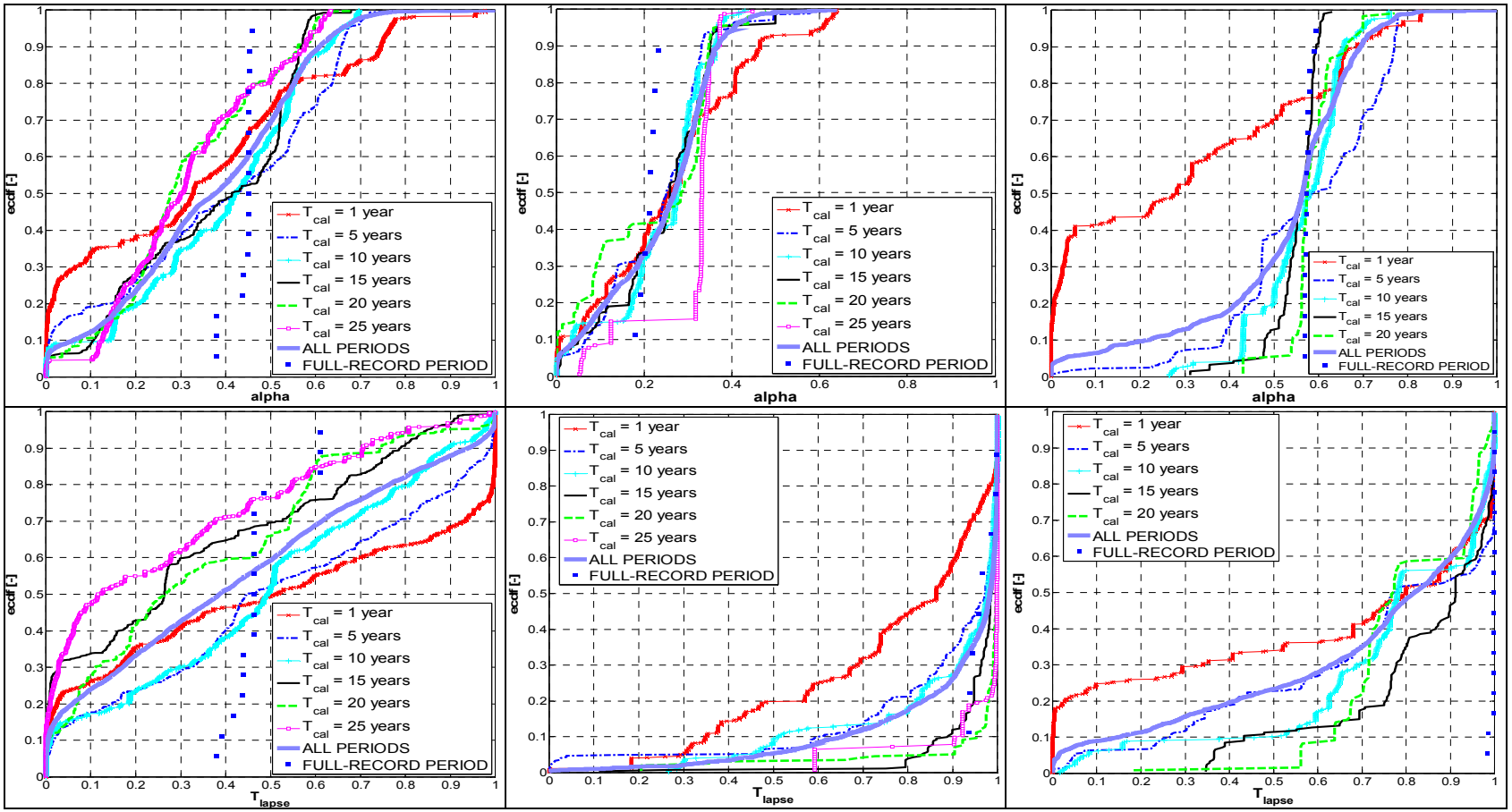
APPENDIX F. Values of S_t obtained over calibration periods of various lengths

A. table. 7. Values of S_t are calculated over 1-, 5-, 10-, 15-, 20- and 25-year long calibration periods

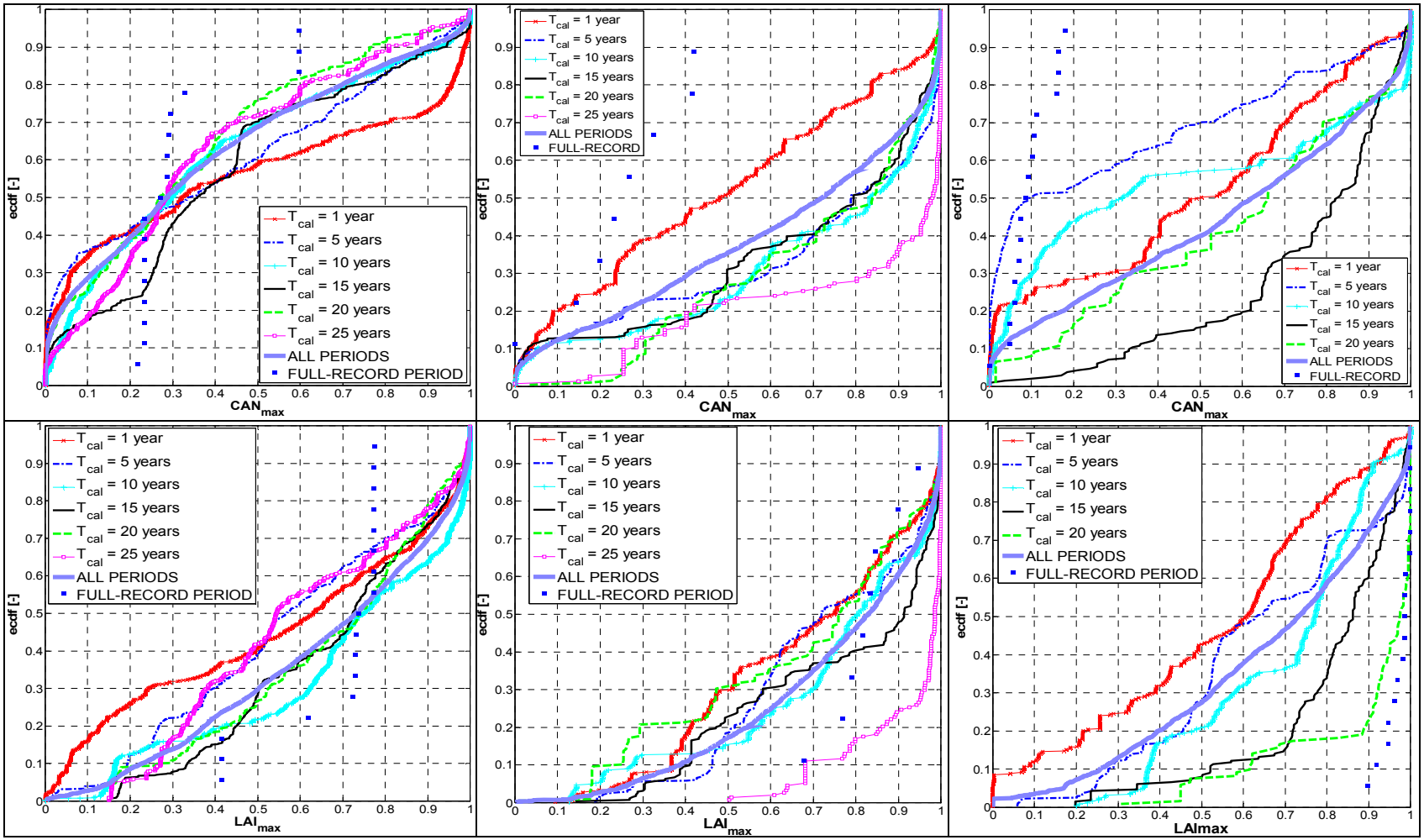
CATCHMENT	CALIBRATION PERIOD [years]	CAN_{max}	LAI_{max}	$T_{s\text{-}r}$	$SNOW_{100}$	lambda	$b_{melt,6}$	$b_{melt,12}$	CN	$I_{a,rel}$	D_{surf}	porosity	log(K_{surf})	w_{wp}	w_{fc}	n	$D_{sub\text{-}surf}$	log($K_{sub\text{-}surf}$)	K_d	$K_{gw\text{-}fast}$	q_d	c	S_{max}	alpha	T_{lapse}
KOLUBARA	1	0.358	0.342	0.260	0.374	0.281	0.343	0.347	0.327	0.347	0.343	0.245	0.329	0.307	0.297	0.395	0.295	0.376	0.262	0.380	0.285	0.331	0.302	0.271	0.369
	5	0.310	0.308	0.187	0.270	0.207	0.318	0.297	0.275	0.403	0.356	0.266	0.294	0.309	0.339	0.399	0.219	0.334	0.180	0.396	0.269	0.265	0.323	0.237	0.352
	10	0.297	0.290	0.185	0.365	0.169	0.242	0.307	0.231	0.378	0.340	0.253	0.280	0.308	0.319	0.322	0.195	0.296	0.115	0.392	0.303	0.315	0.246	0.198	0.287
	15	0.310	0.253	0.143	0.252	0.146	0.181	0.294	0.196	0.394	0.365	0.245	0.285	0.338	0.319	0.341	0.203	0.239	0.112	0.403	0.310	0.303	0.203	0.193	0.315
	20	0.256	0.243	0.234	0.375	0.248	0.302	0.304	0.220	0.392	0.361	0.259	0.290	0.308	0.312	0.345	0.196	0.220	0.109	0.360	0.295	0.316	0.236	0.188	0.300
	25	0.277	0.265	0.157	0.358	0.165	0.204	0.312	0.204	0.396	0.353	0.250	0.289	0.325	0.310	0.354	0.202	0.197	0.112	0.404	0.277	0.318	0.213	0.189	0.326
TOPLICA	1	0.321	0.251	0.286	0.400	0.319	0.317	0.293	0.341	0.395	0.327	0.236	0.278	0.341	0.280	0.348	0.236	0.275	0.242	0.292	0.310	0.363	0.328	0.173	0.253
	5	0.334	0.236	0.234	0.371	0.287	0.363	0.290	0.336	0.362	0.292	0.255	0.295	0.320	0.364	0.334	0.191	0.243	0.177	0.225	0.307	0.353	0.333	0.132	0.232
	10	0.310	0.251	0.211	0.335	0.233	0.328	0.319	0.292	0.364	0.312	0.241	0.239	0.339	0.356	0.304	0.184	0.214	0.134	0.283	0.362	0.327	0.379	0.135	0.201
	15	0.323	0.255	0.217	0.342	0.249	0.302	0.325	0.285	0.383	0.296	0.206	0.236	0.320	0.384	0.323	0.184	0.248	0.125	0.229	0.330	0.327	0.353	0.128	0.150
	20	0.286	0.250	0.213	0.365	0.241	0.278	0.295	0.297	0.361	0.295	0.257	0.197	0.302	0.353	0.299	0.198	0.169	0.136	0.282	0.299	0.348	0.373	0.145	0.150
	25	0.304	0.251	0.221	0.340	0.233	0.274	0.319	0.303	0.342	0.274	0.213	0.182	0.313	0.341	0.328	0.201	0.174	0.132	0.231	0.277	0.318	0.337	0.141	0.150
MLAVA	1	0.300	0.315	0.305	0.422	0.286	0.349	0.361	0.345	0.399	0.396	0.312	0.330	0.352	0.390	0.401	0.218	0.407	0.301	0.335	0.322	0.299	0.308	0.271	0.418
	5	0.330	0.270	0.235	0.419	0.307	0.341	0.348	0.321	0.314	0.366	0.278	0.237	0.326	0.399	0.452	0.148	0.221	0.211	0.396	0.316	0.331	0.352	0.229	0.375
	10	0.344	0.265	0.196	0.420	0.198	0.344	0.301	0.276	0.368	0.320	0.236	0.218	0.318	0.393	0.446	0.129	0.219	0.176	0.384	0.363	0.338	0.286	0.188	0.338
	15	0.225	0.133	0.112	0.333	0.081	0.284	0.339	0.349	0.413	0.342	0.123	0.228	0.198	0.322	0.372	0.125	0.108	0.165	0.383	0.203	0.284	0.122	0.040	0.192
	20	0.252	0.163	0.111	0.212	0.071	0.219	0.310	0.238	0.363	0.282	0.130	0.247	0.205	0.251	0.292	0.123	0.100	0.156	0.349	0.274	0.329	0.301	0.048	0.130

** Values of S_t are calculated form median values of normalised Pareto-optimal parameters obtained for all calibration periods of given length.*

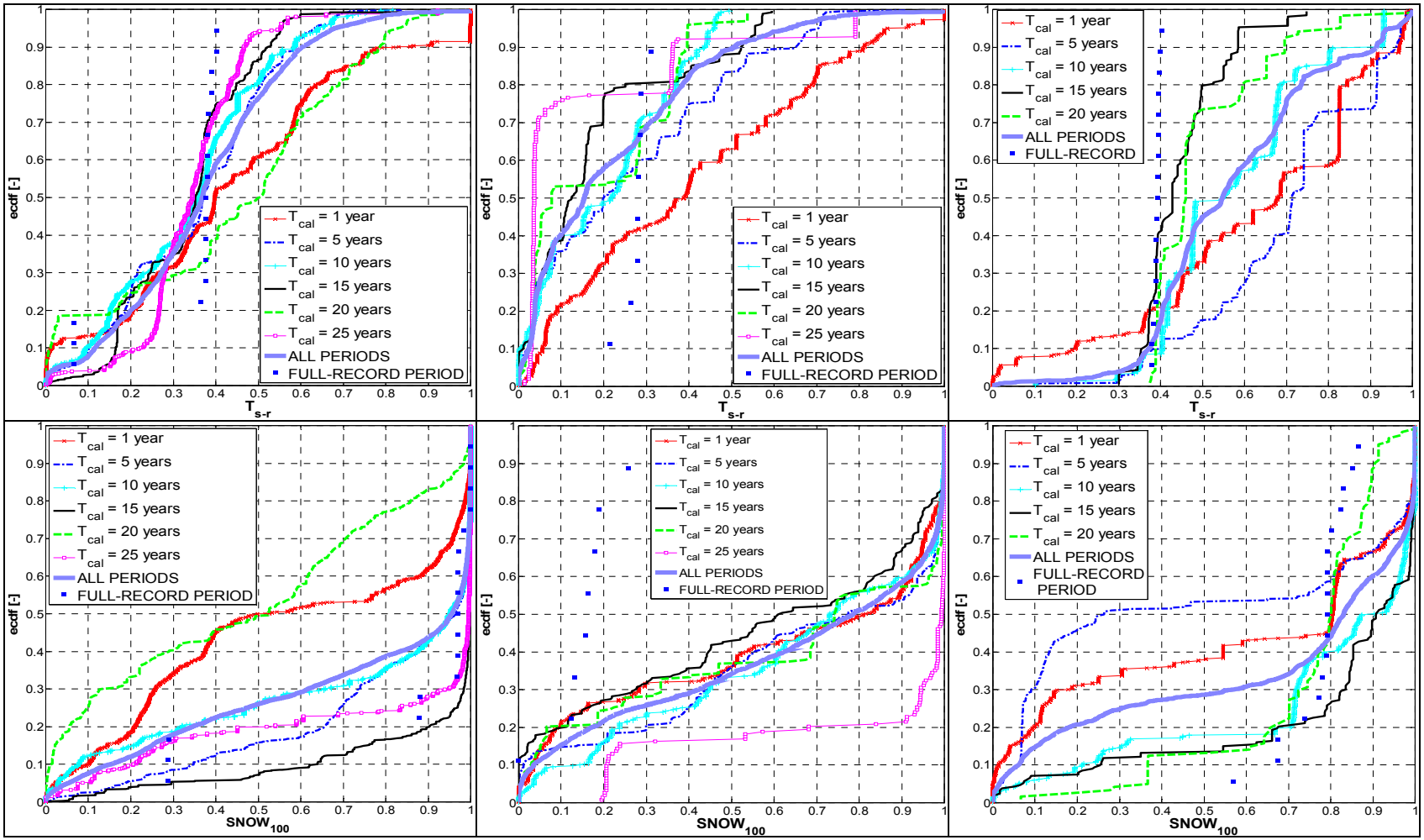
APPENDIX G. Empirical cumulative distribution functions of normalised Pareto-optimal parameters



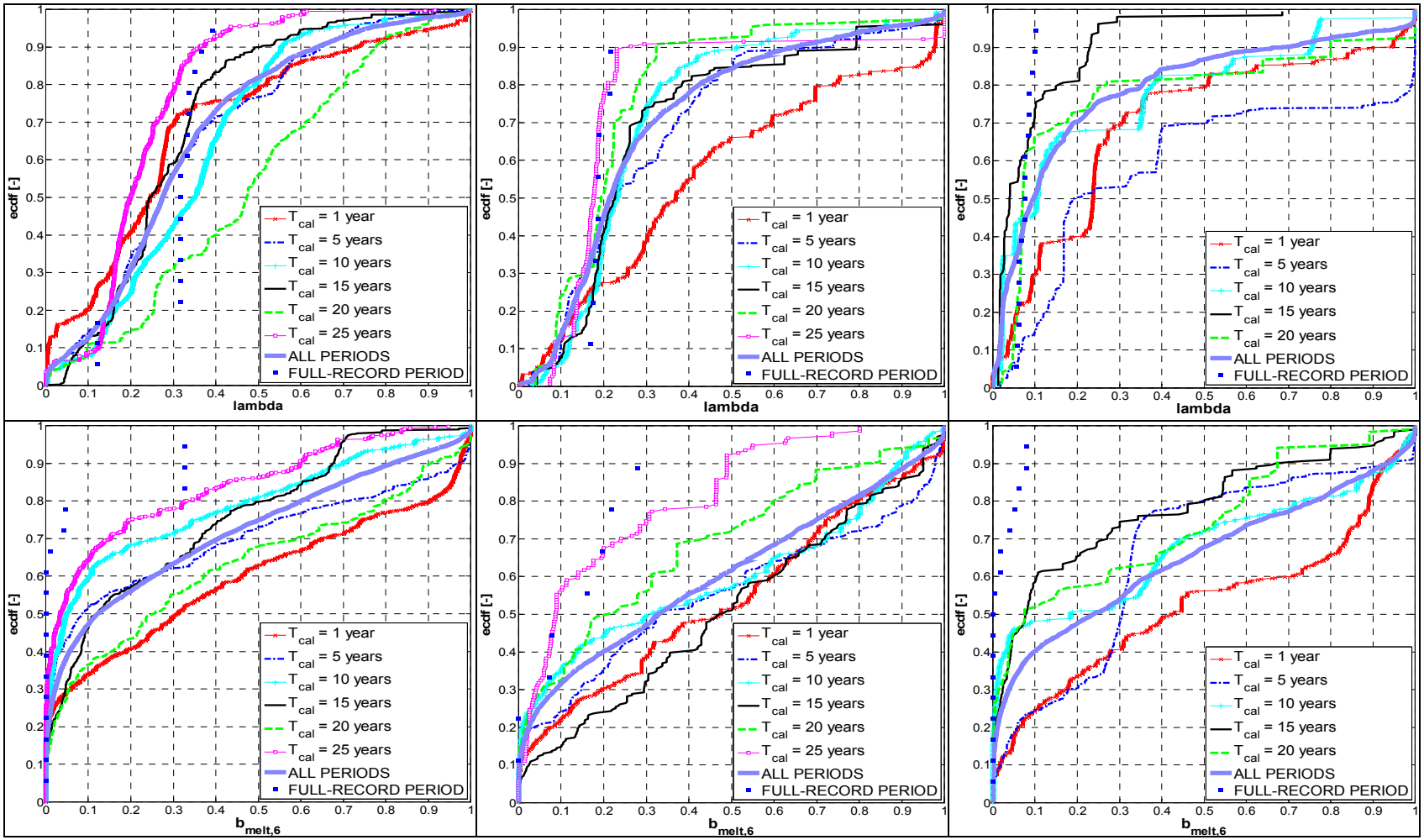
A. Figure 35. Empirical cumulative distribution functions of the Pareto-optimal precipitation gradients with elevation and lapse rates: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



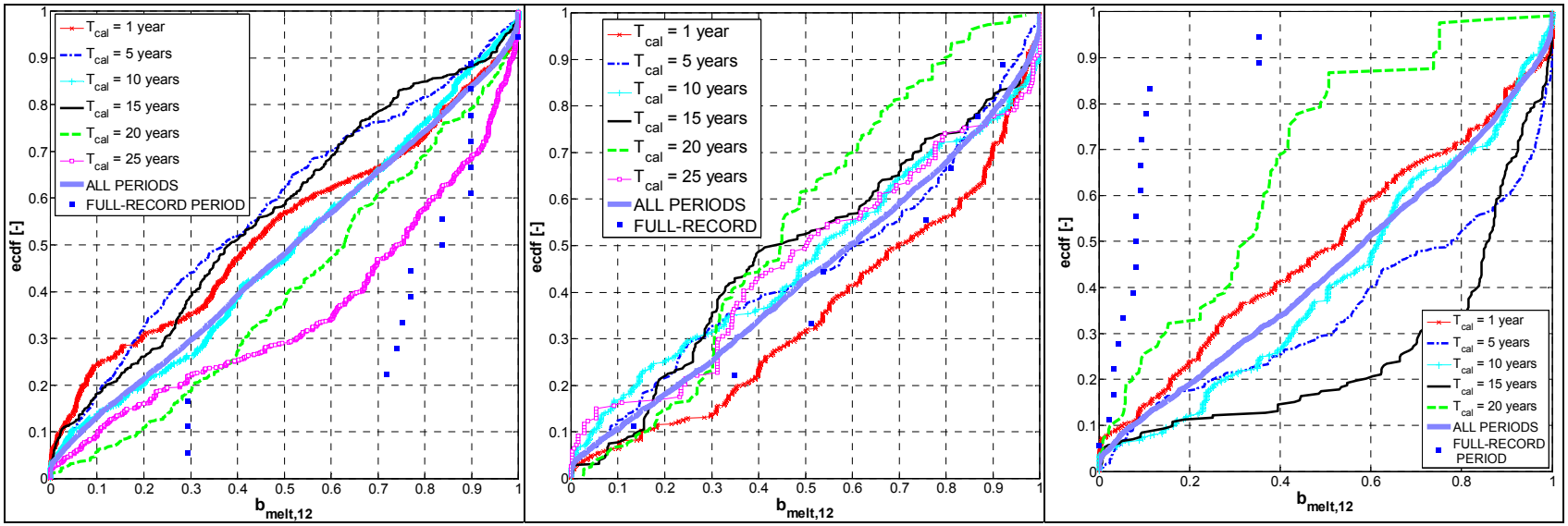
A. Figure 36. Empirical cumulative distribution functions of the Pareto-optimal parameters of the interception routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



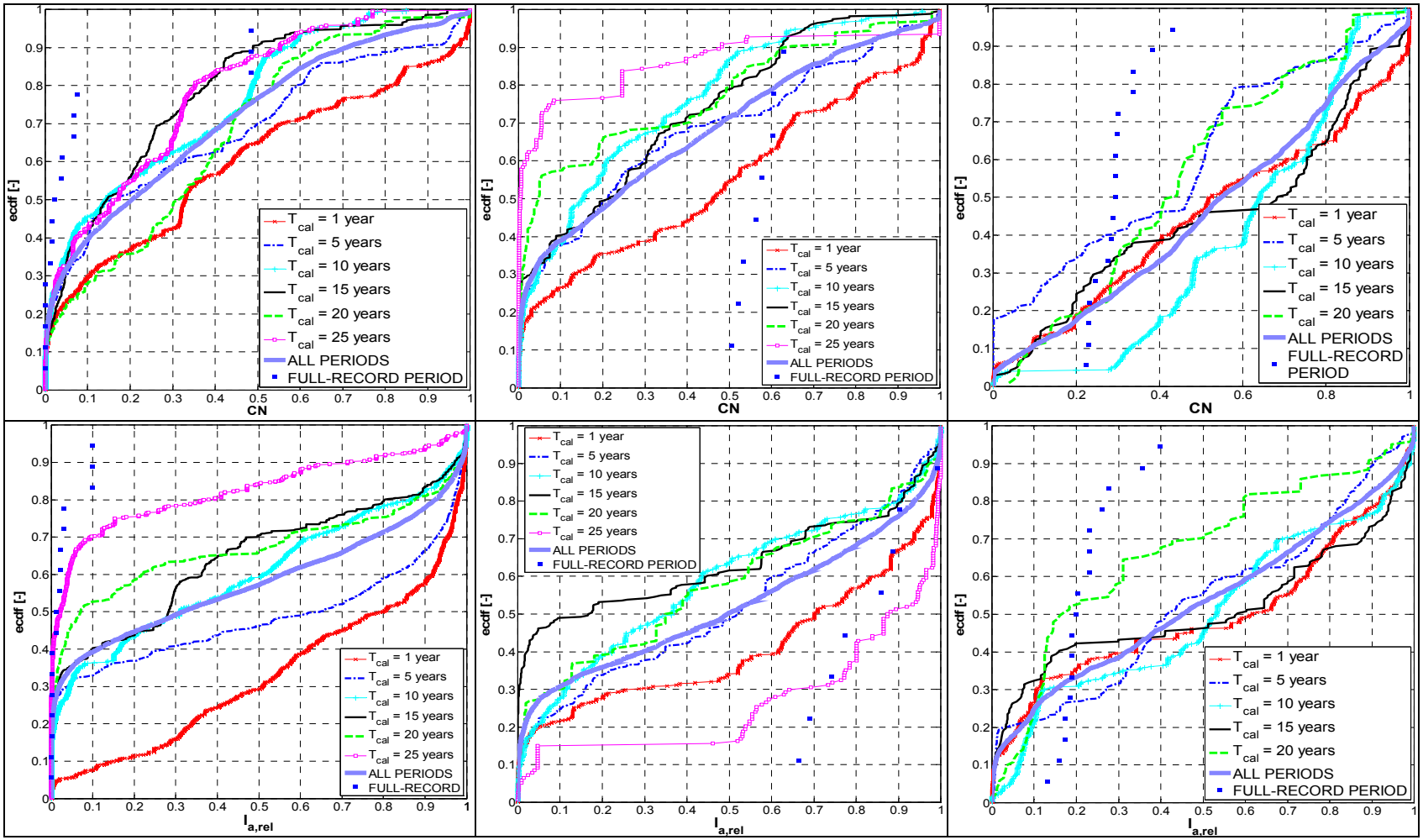
A. Figure 37. Empirical cumulative distribution functions of the Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



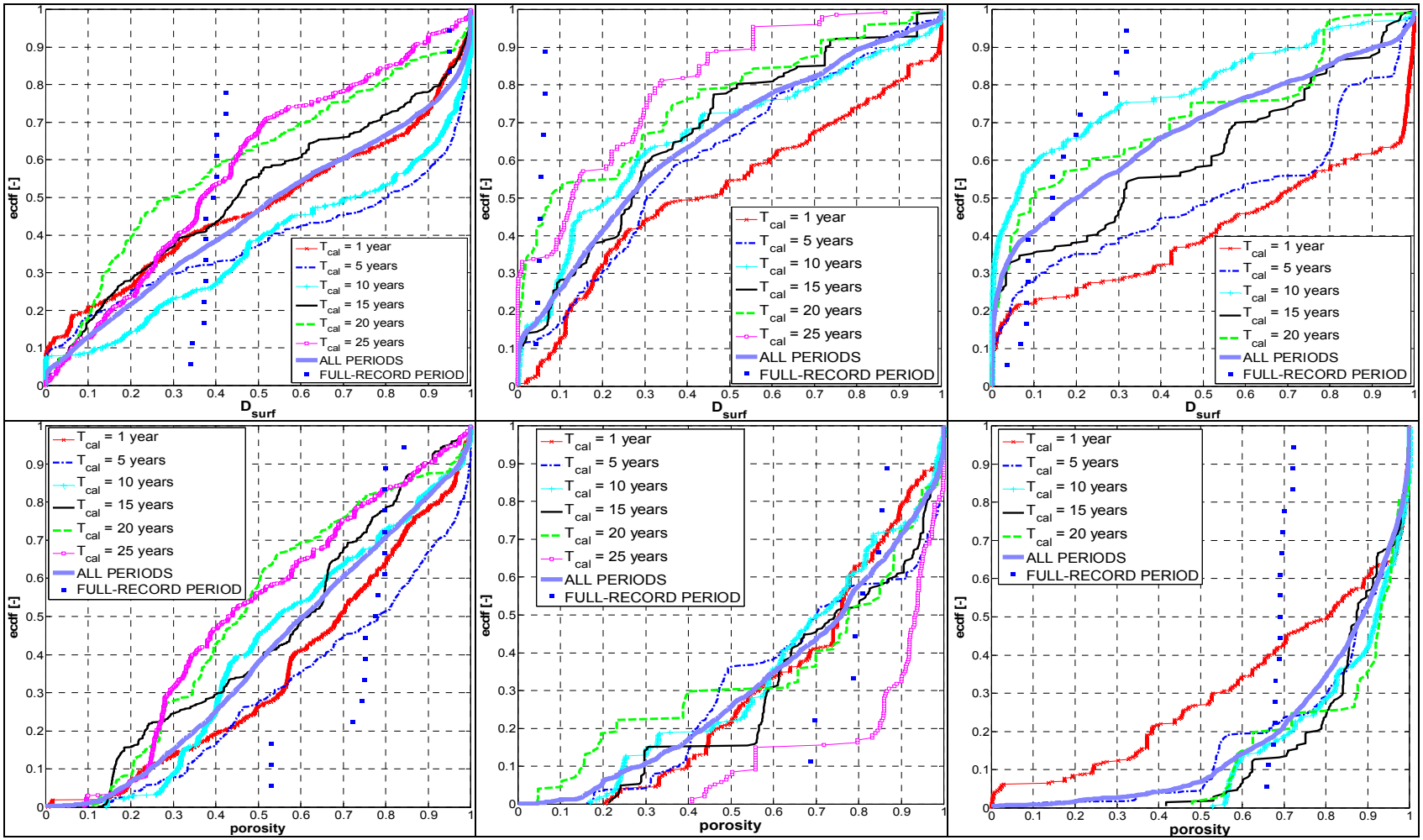
A. Figure 37 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



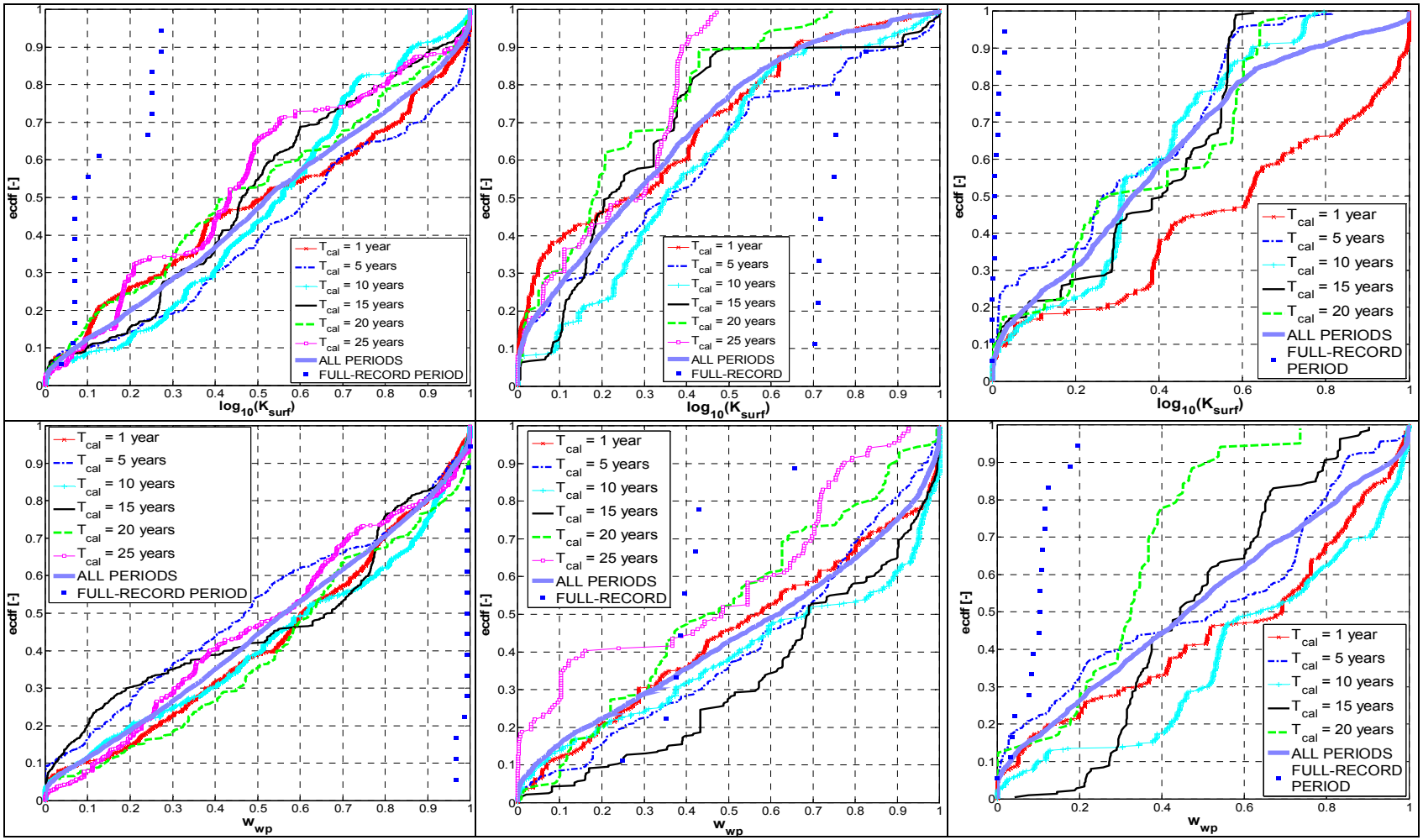
A. Figure 37 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



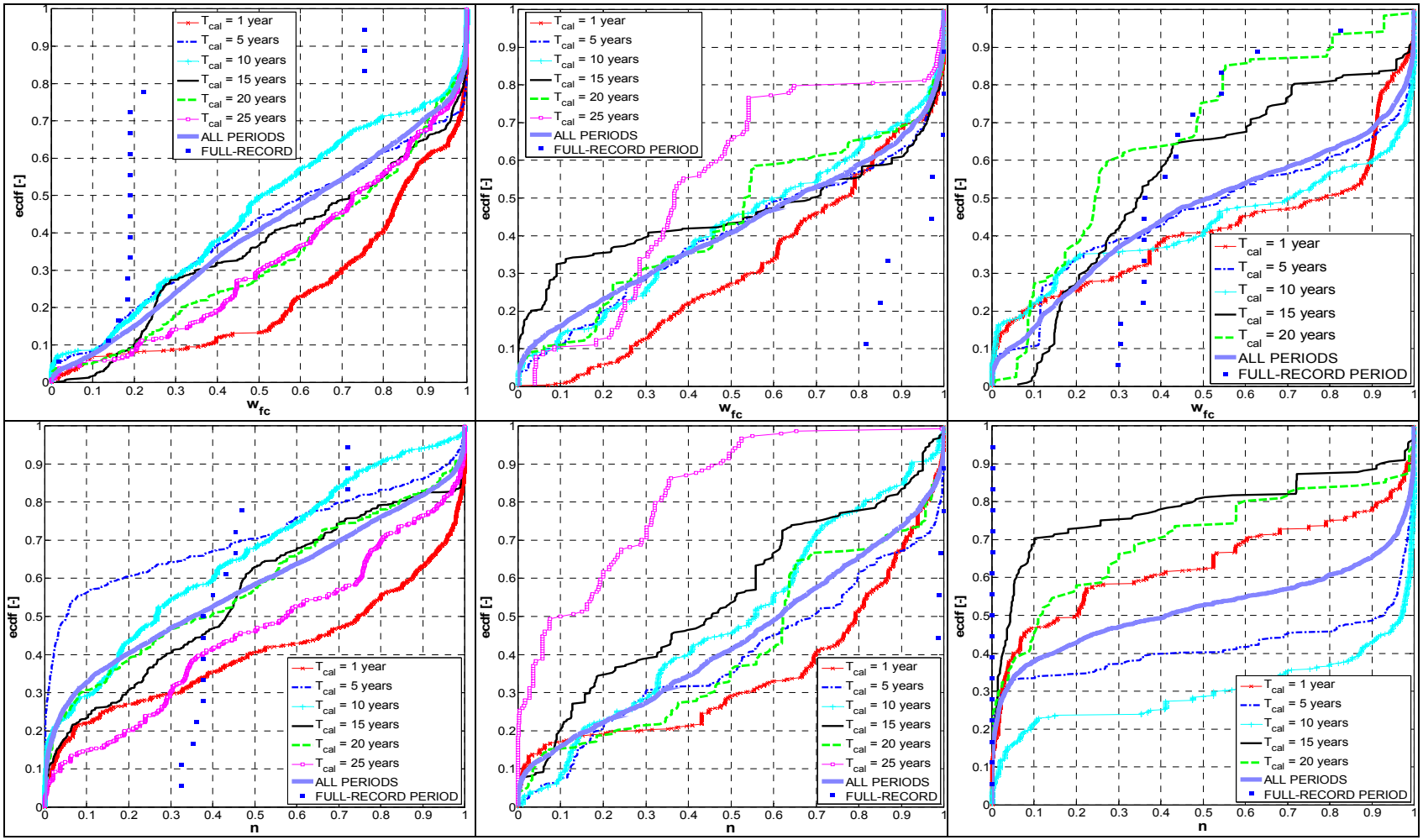
A. Figure 38. Empirical cumulative distribution functions of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



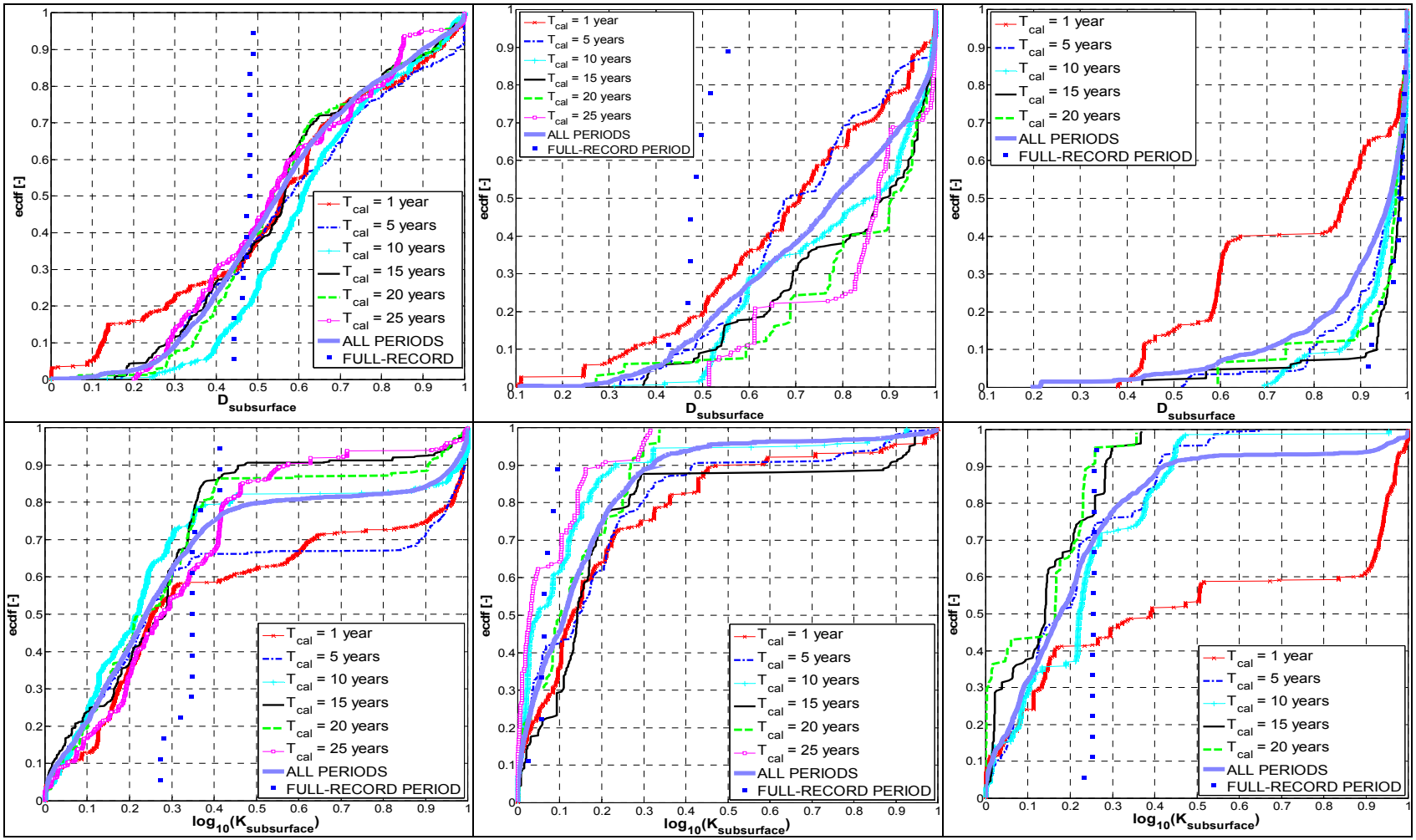
A. Figure 38 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



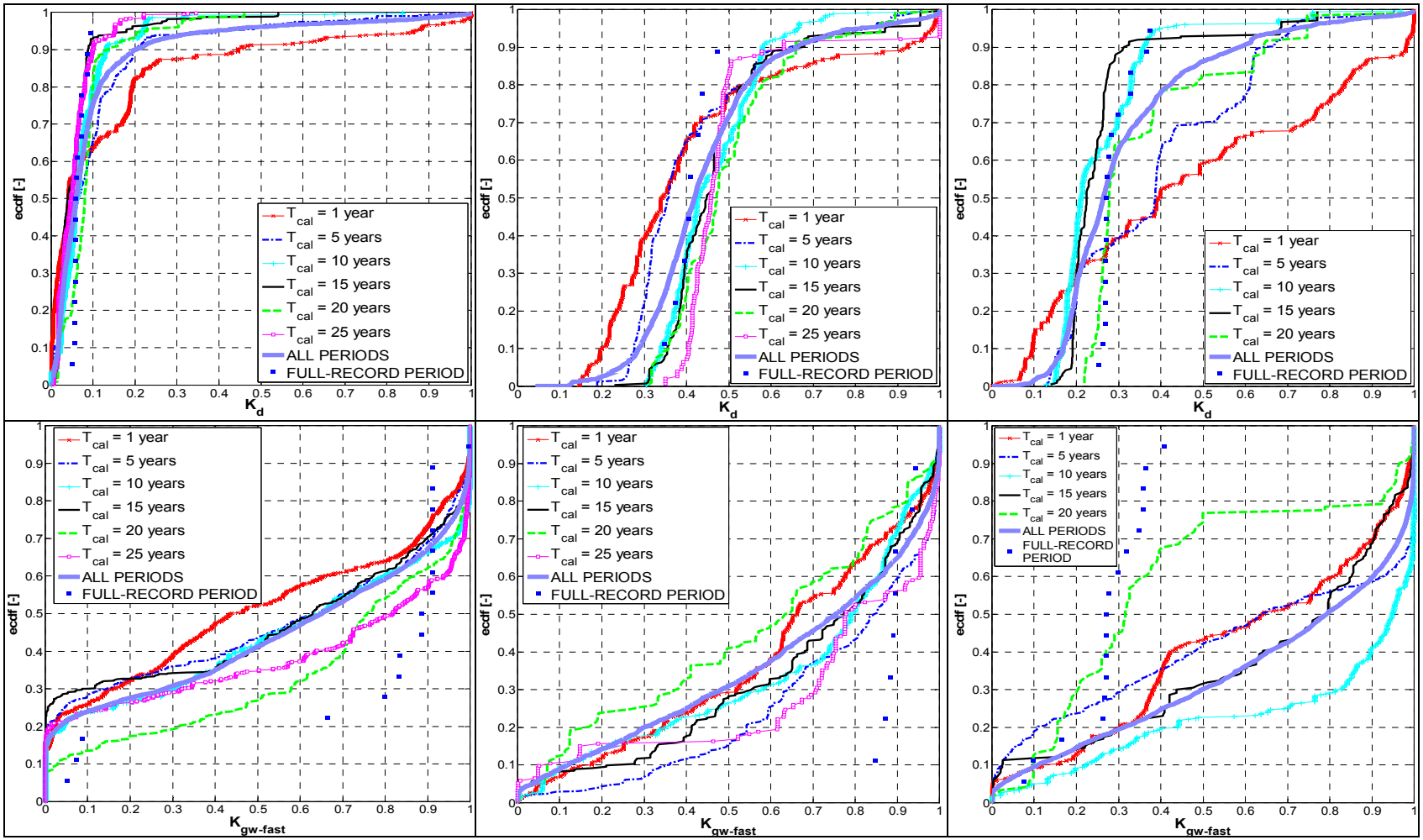
A. Figure 38 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



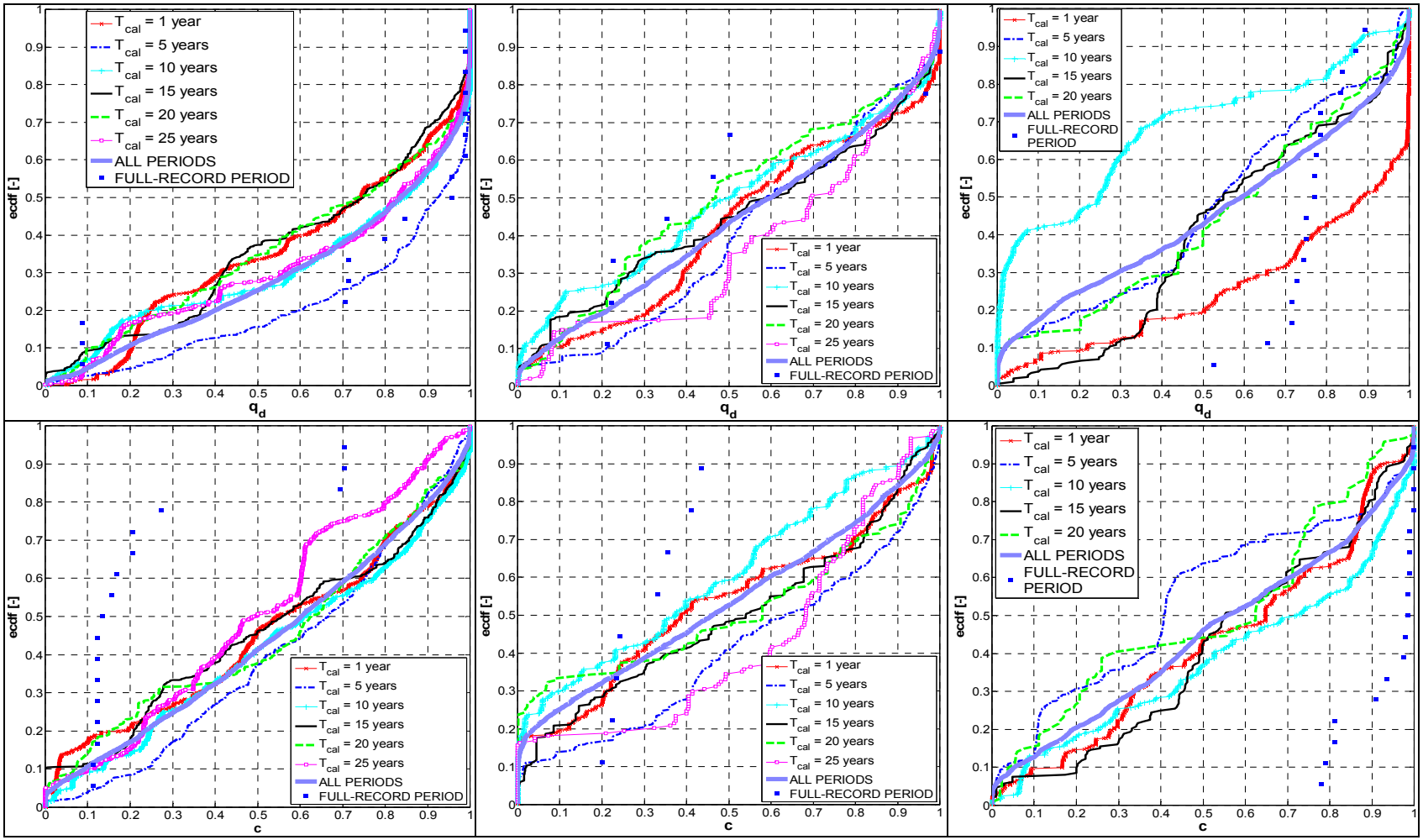
A. Figure 38 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



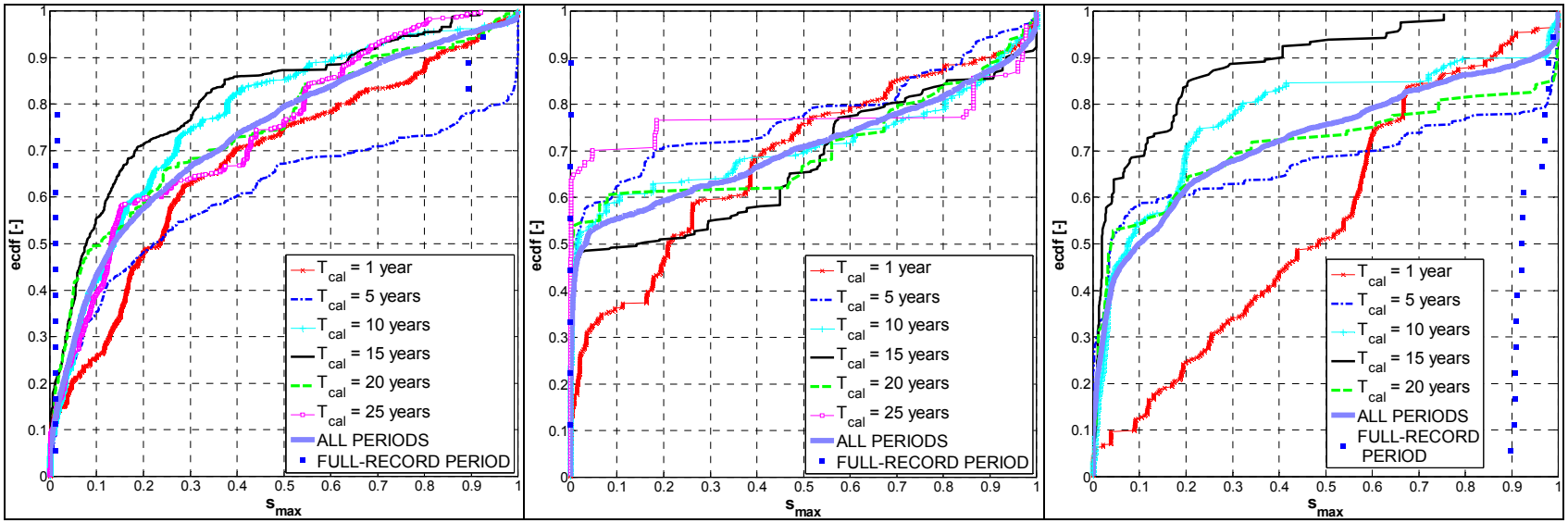
A. Figure 38 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 39. Empirical cumulative distribution functions of the Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



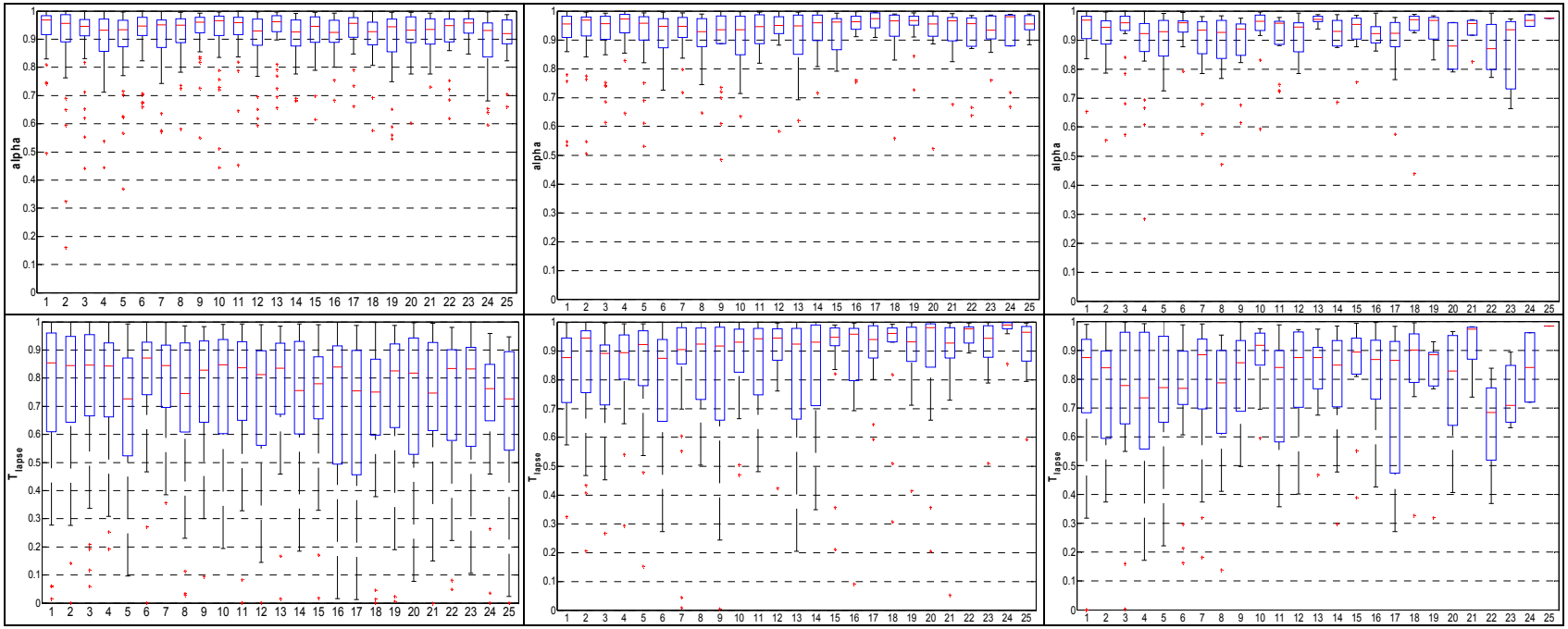
A. Figure 39 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



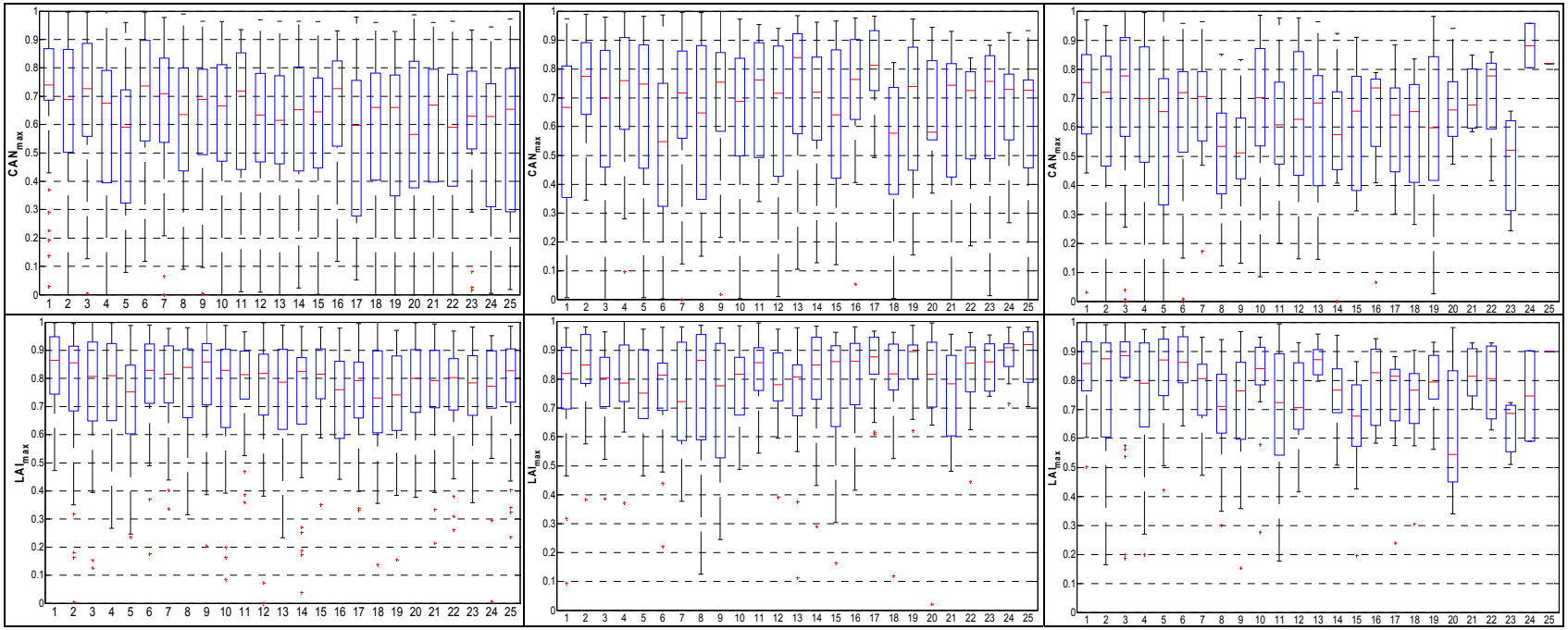
A. Figure 39 (continued). Empirical cumulative distribution functions of the Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

APPENDIX H. Parameter identifiability over calibration period of different lengths

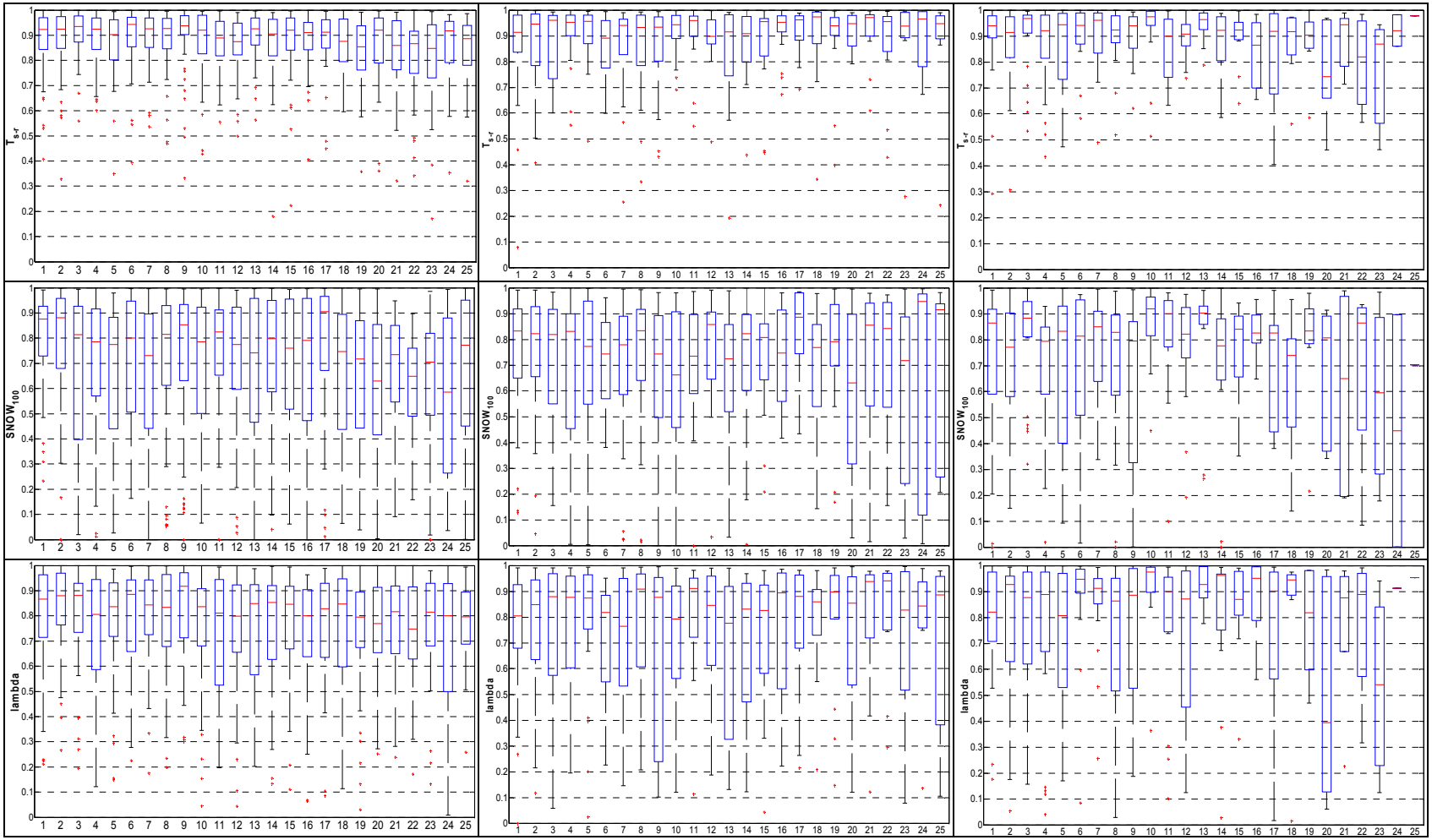
Box plots of the *IC* statistic are obtained over all calibration periods of given length (abscissa values).



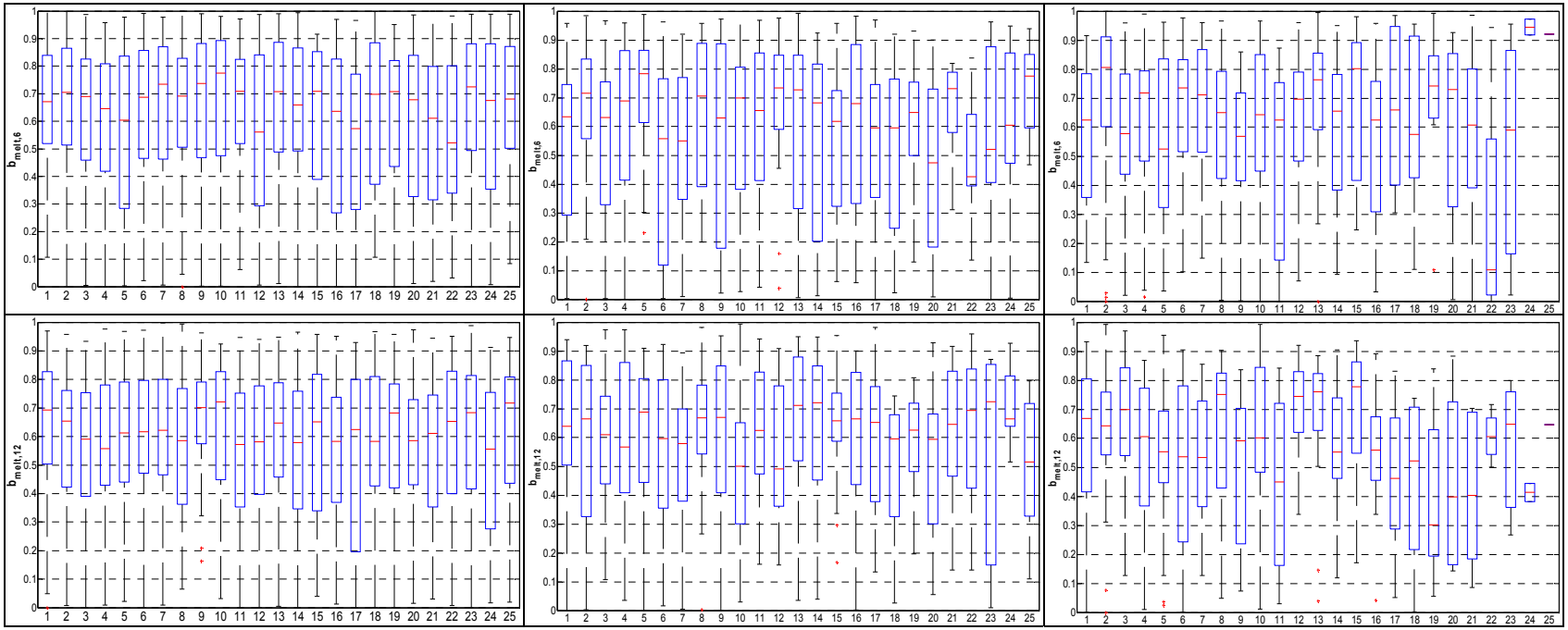
A. Figure 40. The *IC* statistic of the Pareto-optimal precipitation gradients and lapse rates: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



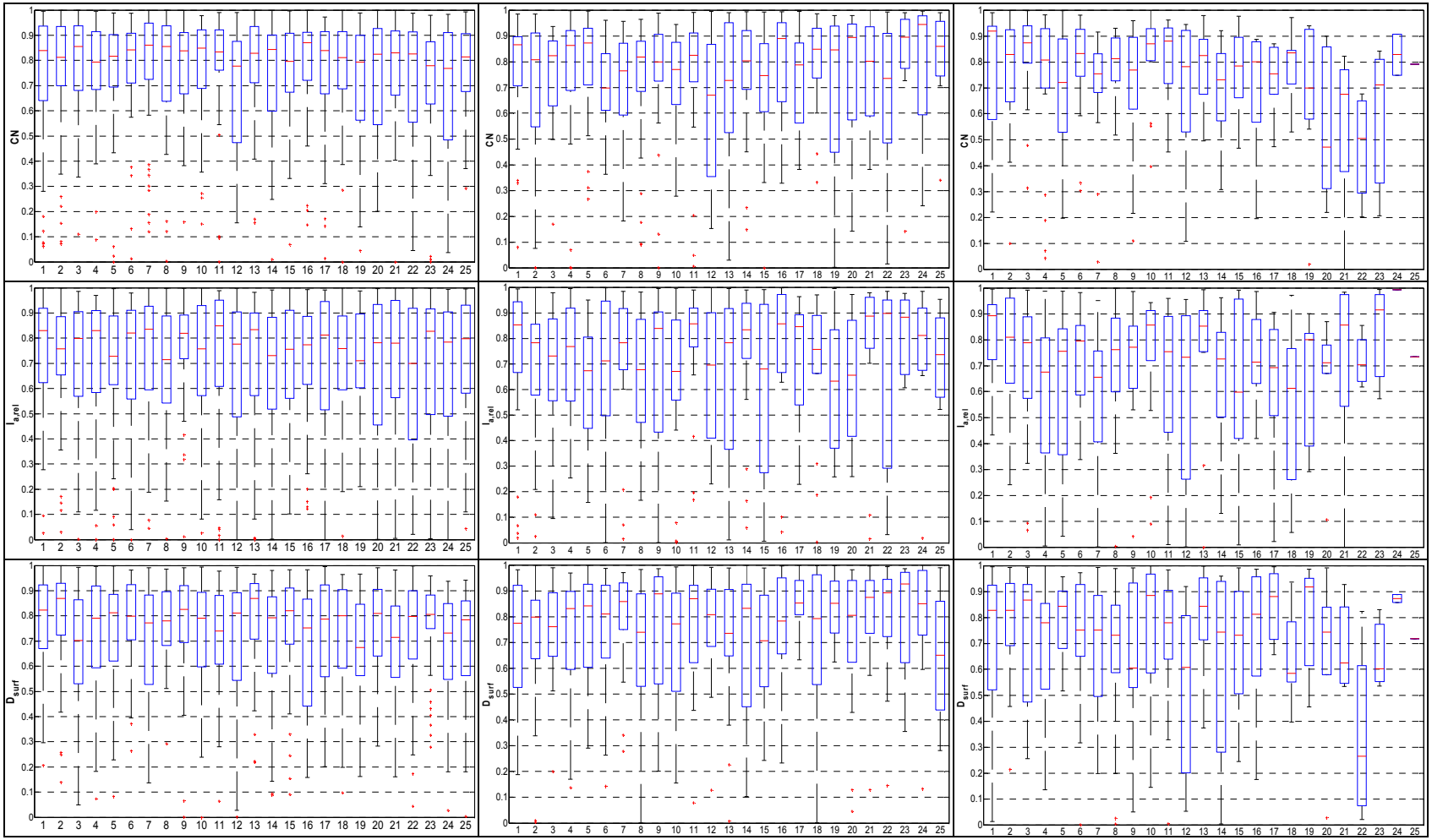
A. Figure 41. The *IC* statistic of the Pareto-optimal parameters of the interception routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



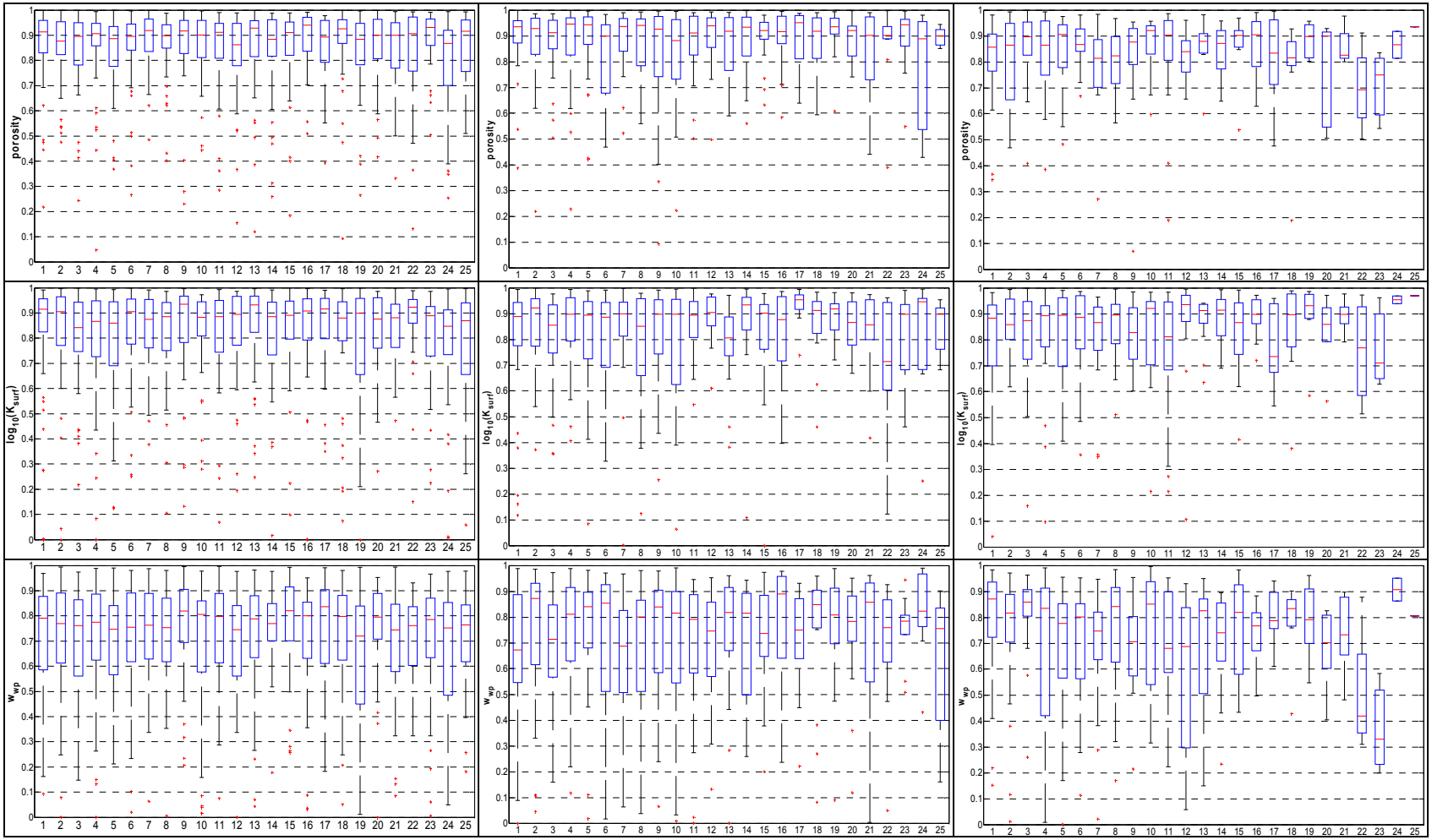
A. Figure 42. The *IC* statistic of the Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



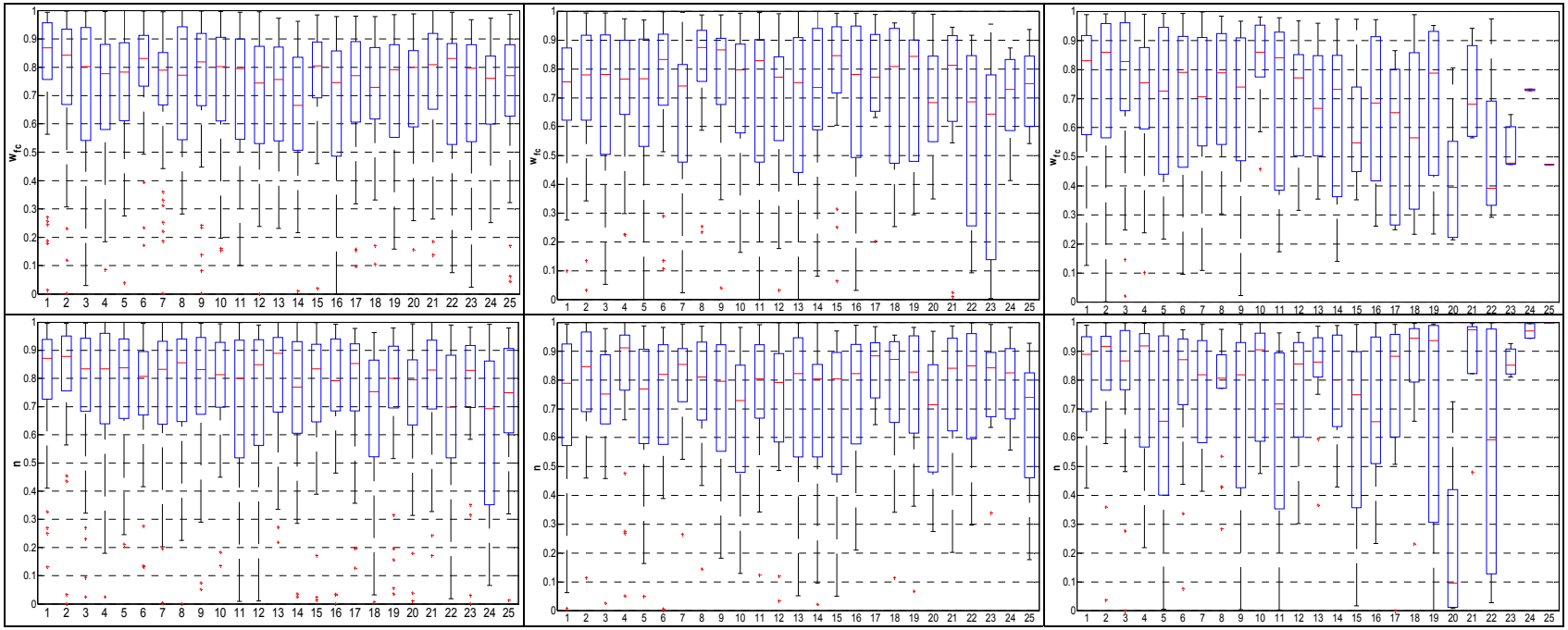
A. Figure 42 (continued). The *IC* statistic of the Pareto-optimal parameters of the snow routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



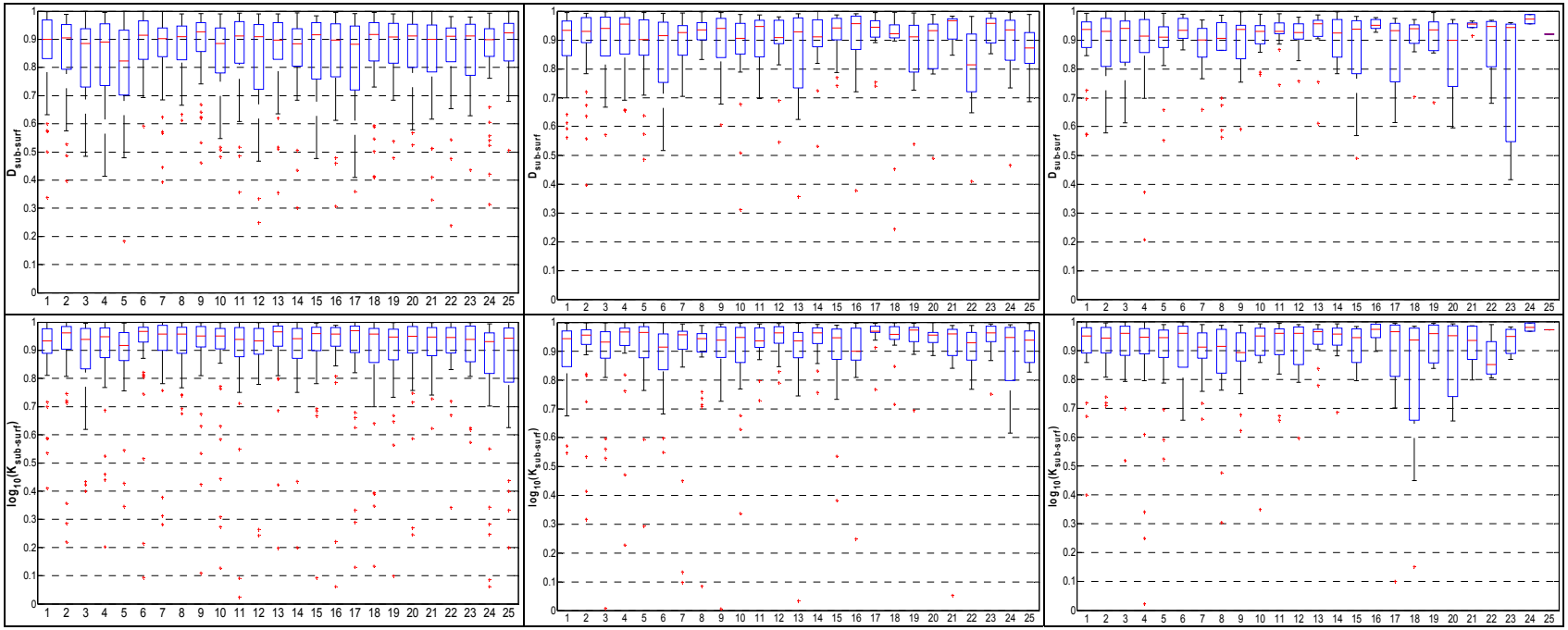
A. Figure 43. The *IC* statistic of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



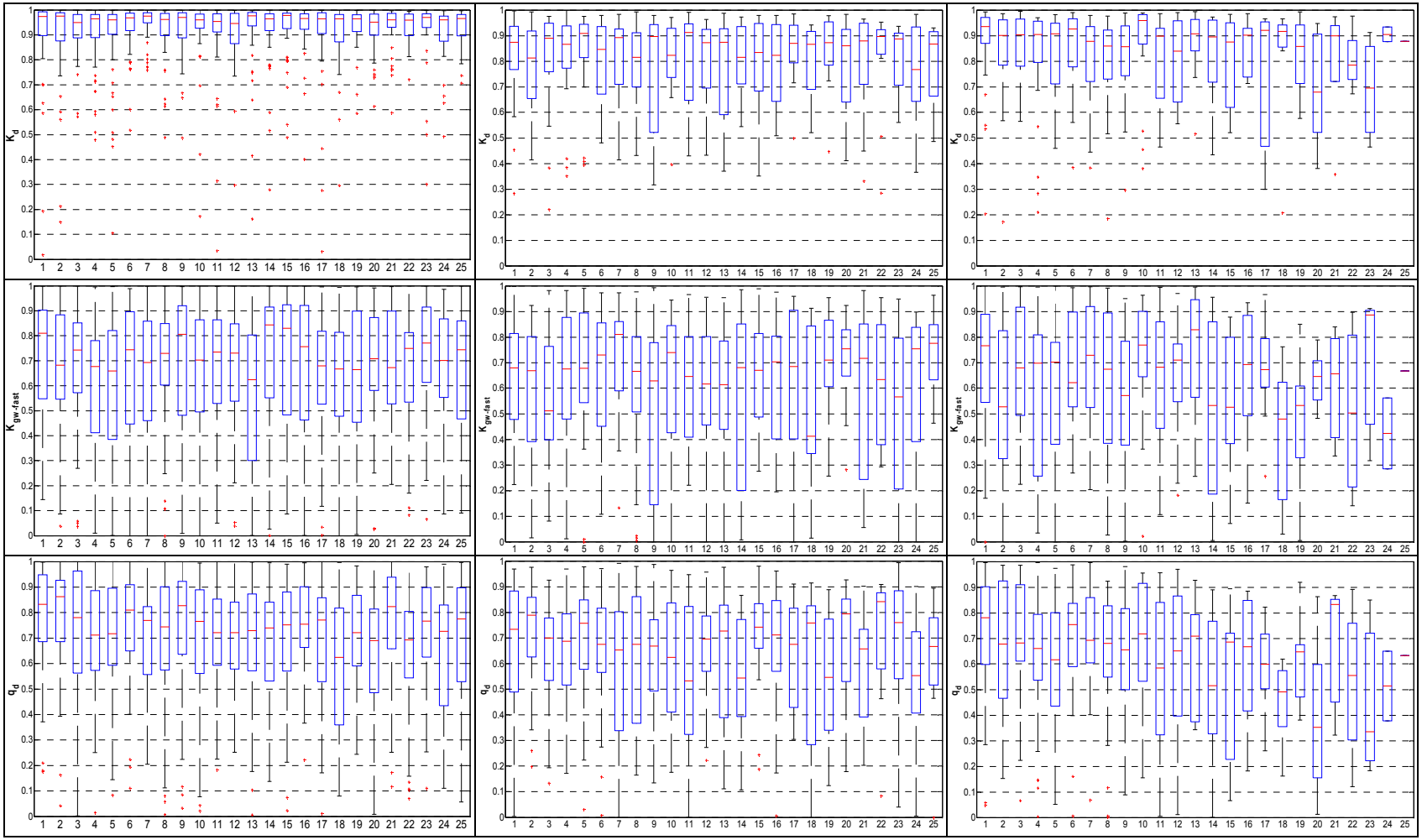
A. Figure 43 (continued). The *IC* statistic of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



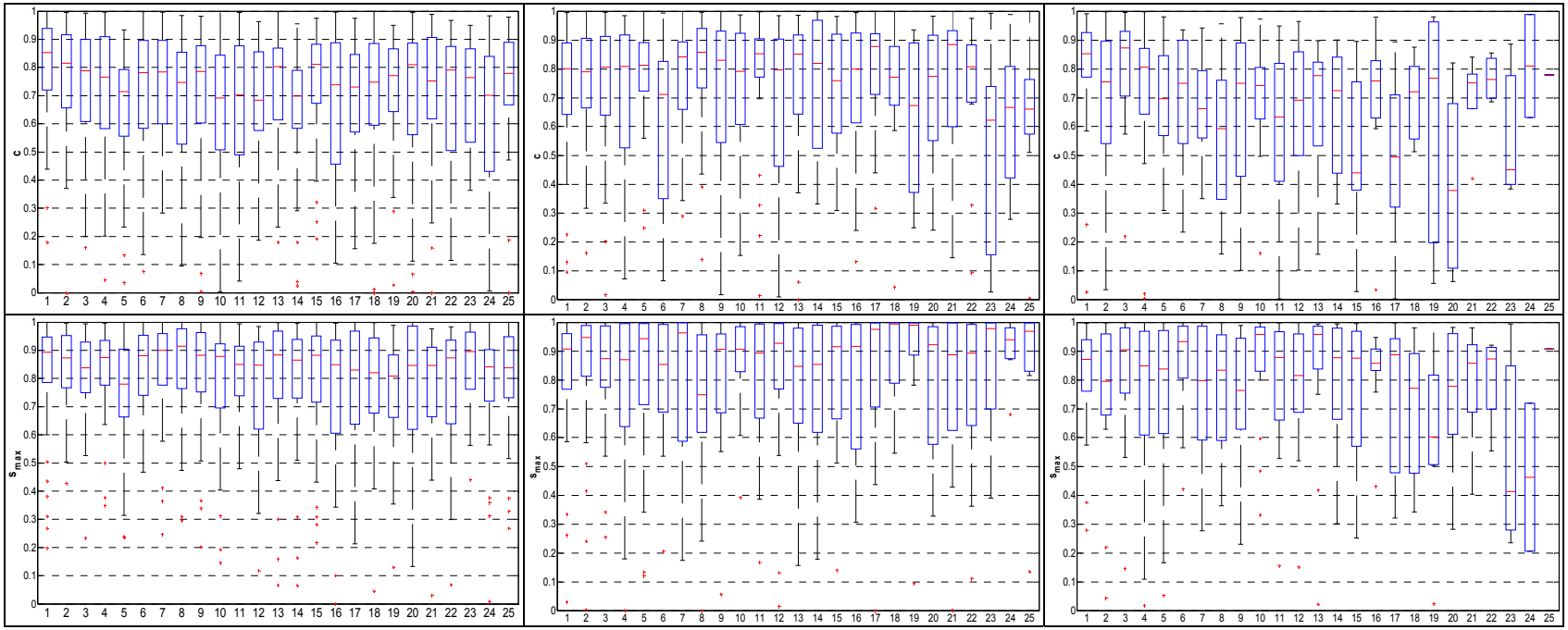
A. Figure 43 (continued). The *IC* statistic of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 43 (continued). The *IC* statistic of the Pareto-optimal parameters of the soil routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

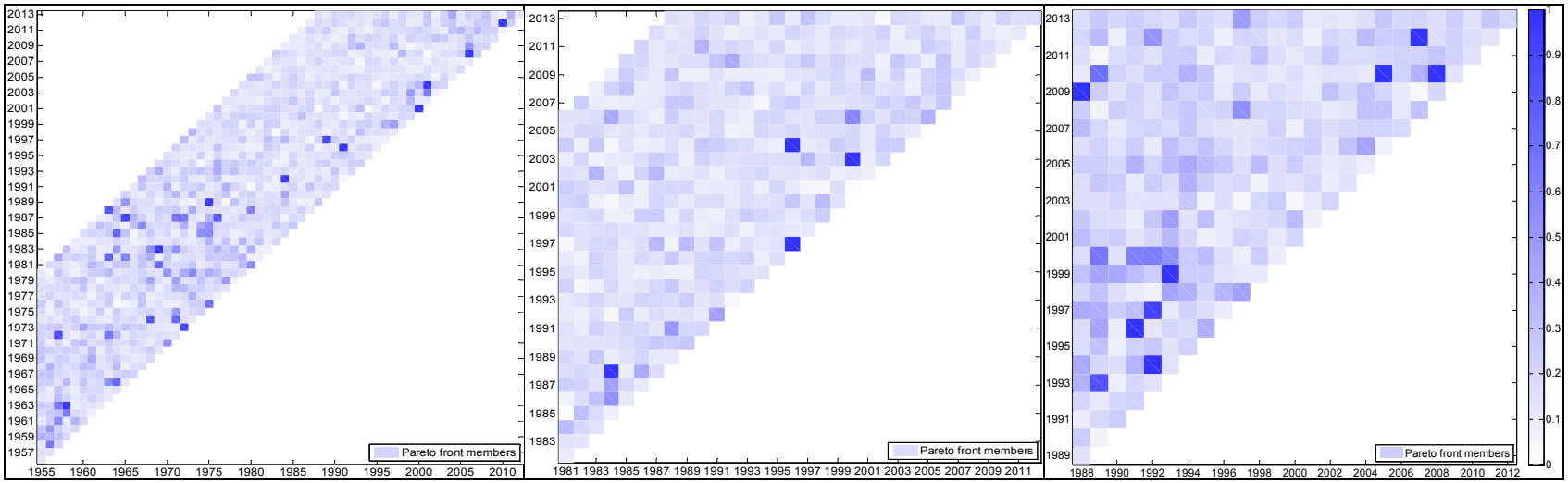


A. Figure 44. The *IC* statistic of the Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

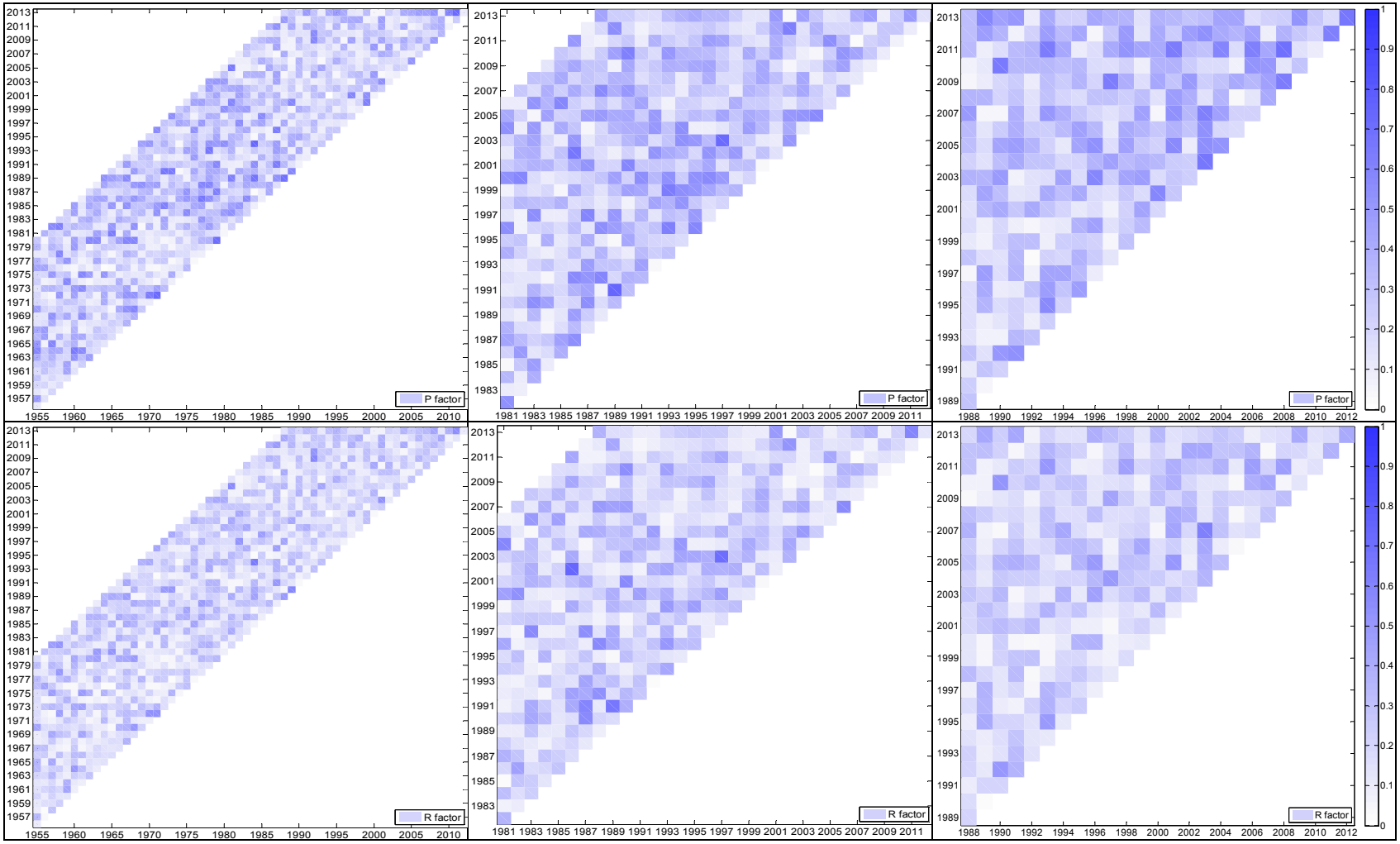


A. Figure 44 (continued). The *IC* statistic of the Pareto-optimal parameters of the response routine: semi-lumped BASIC version of the model, the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

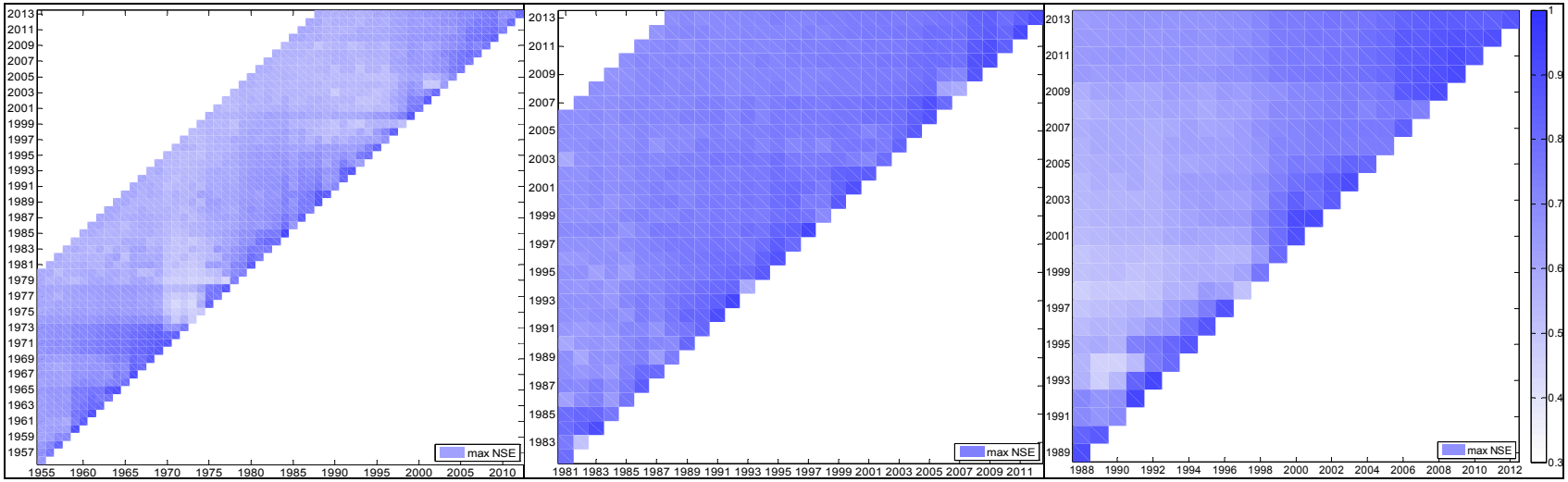
APPENDIX I. Model performance over different calibration periods: semi-lumped BASIC version of the model



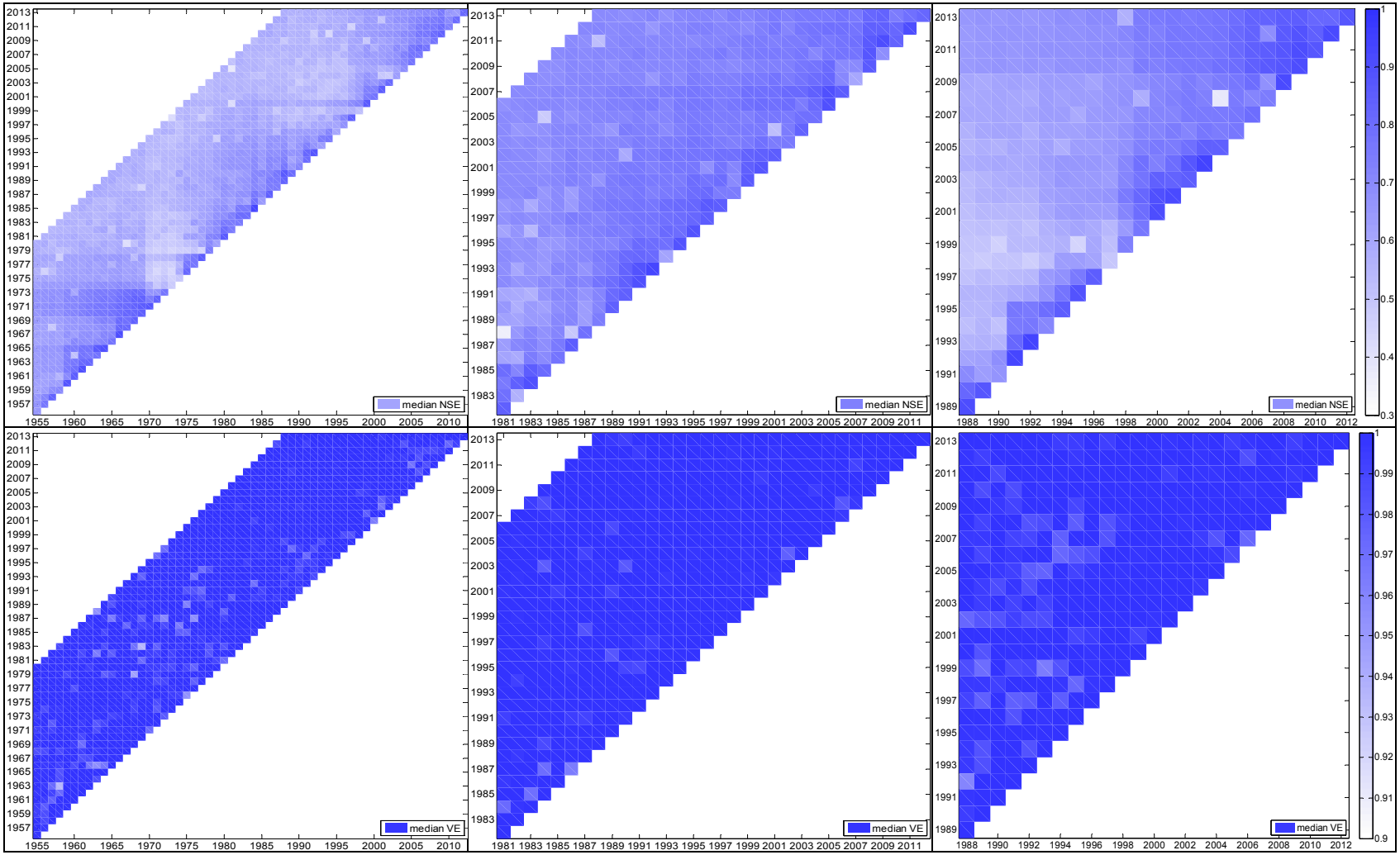
A. Figure 45. Relative number of the Pareto-optimal sets: the Kolubara (left panel), the Toplica (mid panel) and the Mlava River catchments.



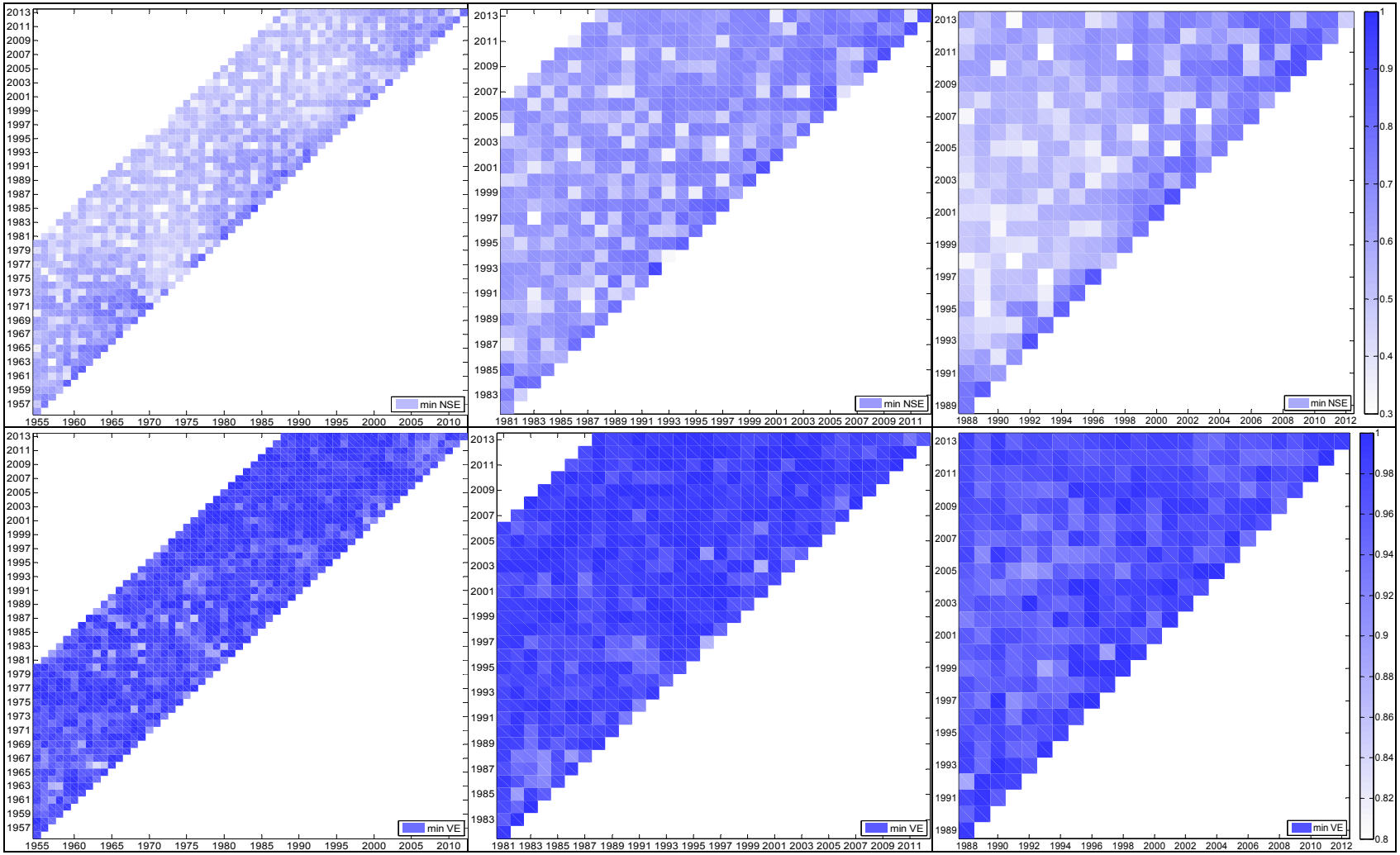
A. Figure 46. *p*-factor (top) and *r*-factor (bottom): the Kolubara (left), the Toplica (mid) and the Mlava River catchments (right panels).



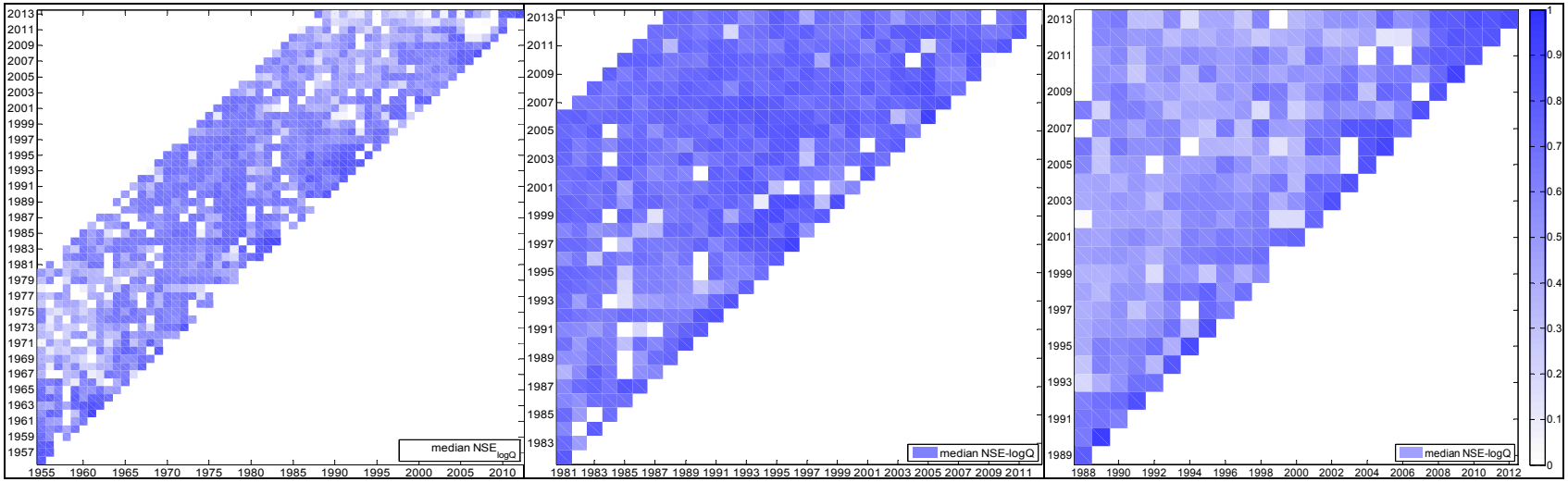
A. Figure 47. Maximum *NSE* values: the Kolubara (left panel), the Toplica (mid panel) and the Mlava River catchments (right panel).



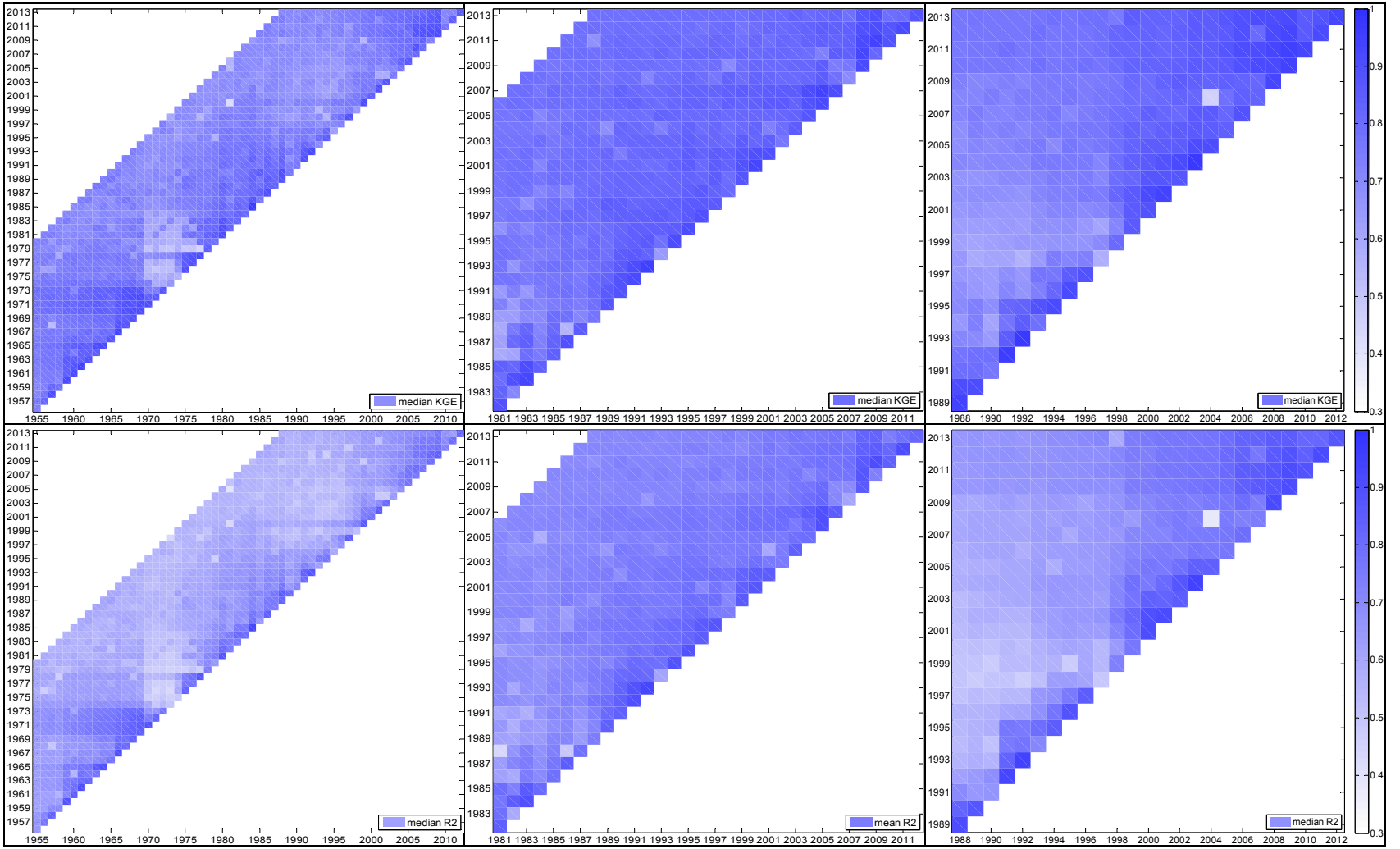
A. Figure 48. Median *NSE* (top) and *VE* values (bottom): the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).



A. Figure 49. Minimum *NSE* (top) and *VE* values (bottom): the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels)

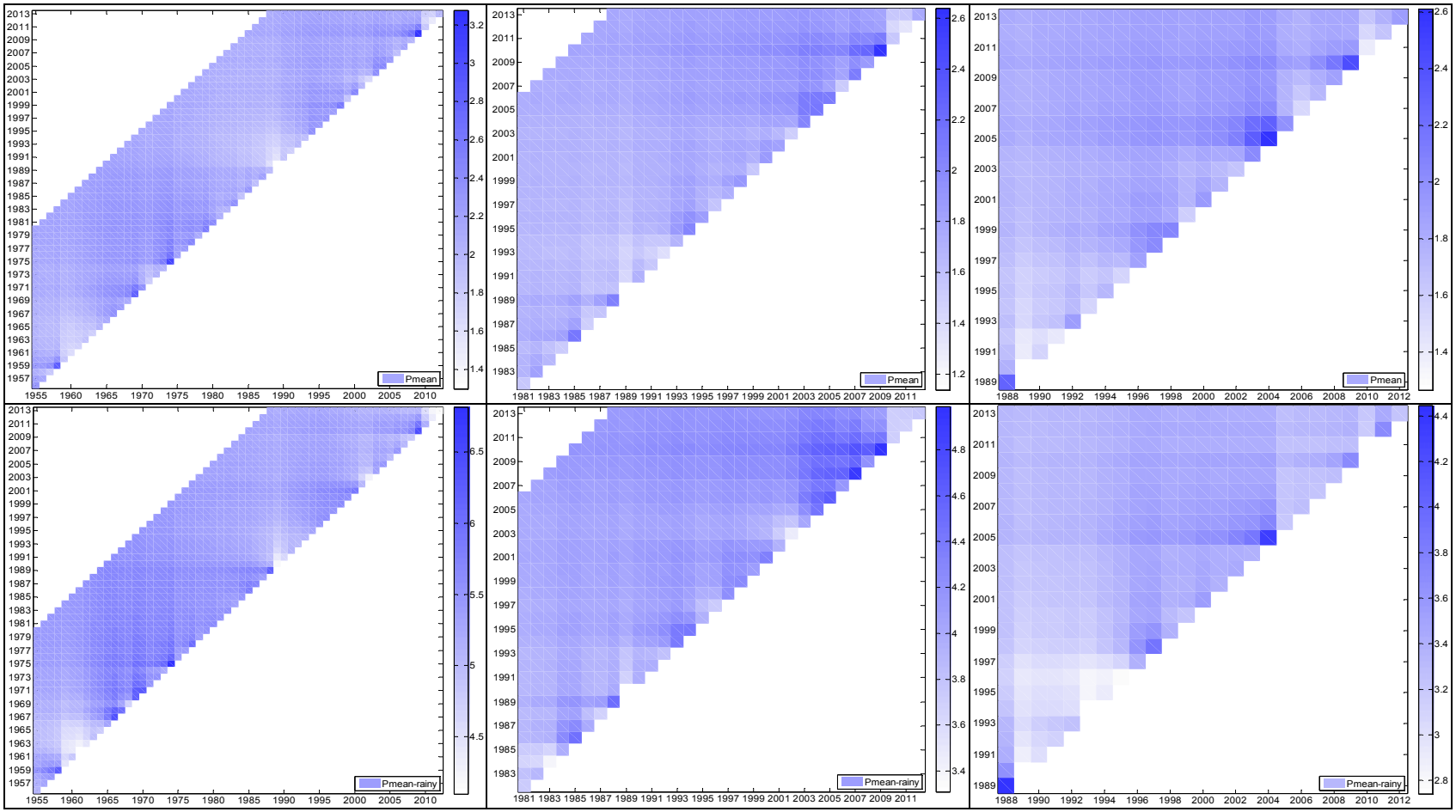


A. Figure 50. Median $NSE_{\log Q}$ values: the Kolubara (left), Toplica (mid) and Mlava River catchments (right panel).

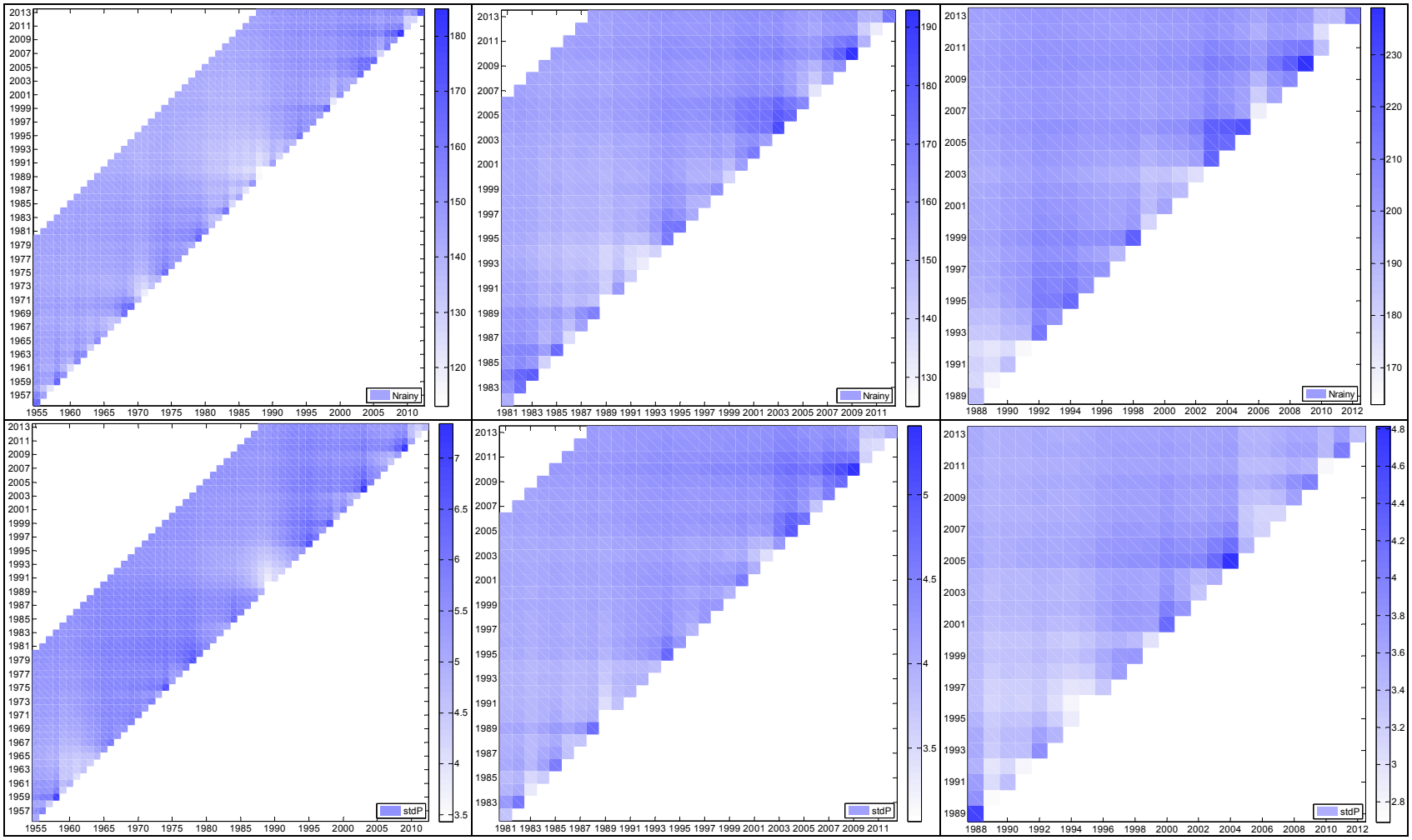


A. Figure 51. Median *KGE* (top) and R^2values (bottom): the Kolubara (left), Toplica (mid) and Mlava River catchments (right panels).

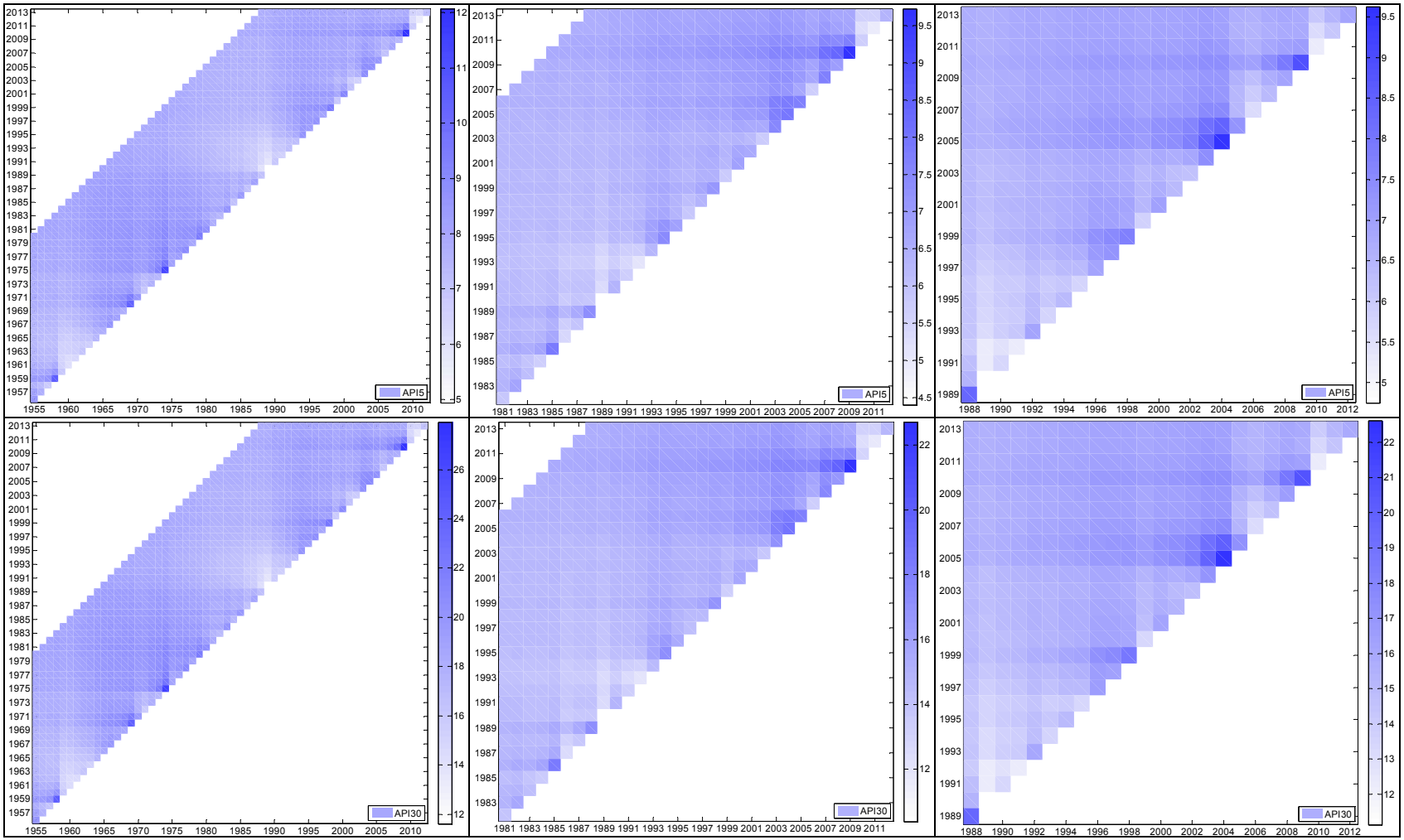
APPENDIX J. Temporal variability in hydro-meteorological characteristics in the catchments



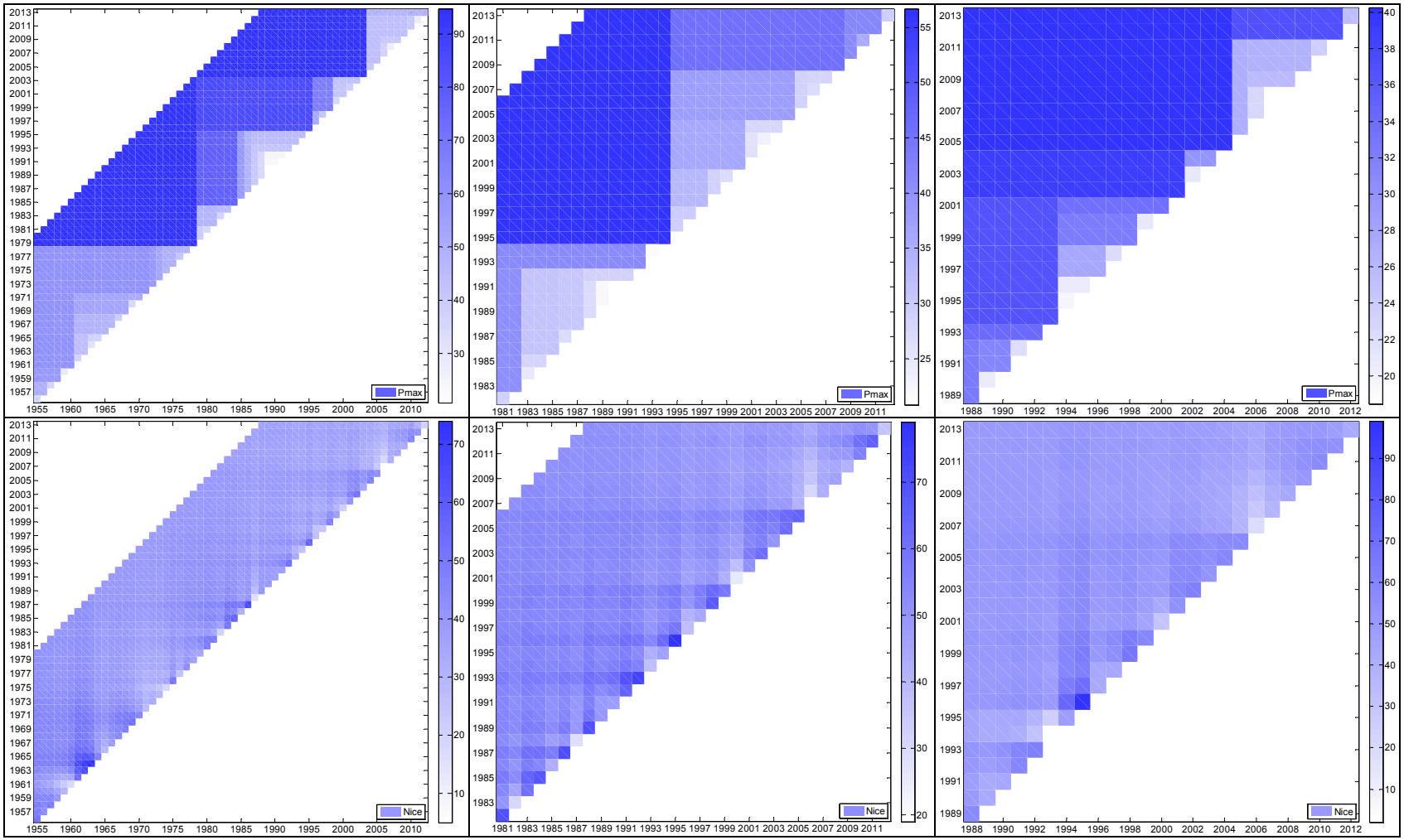
A. Figure 52. Mean daily precipitation depths and precipitation depths in wet days: the Kolubara, Toplica and Mlava River catchments.



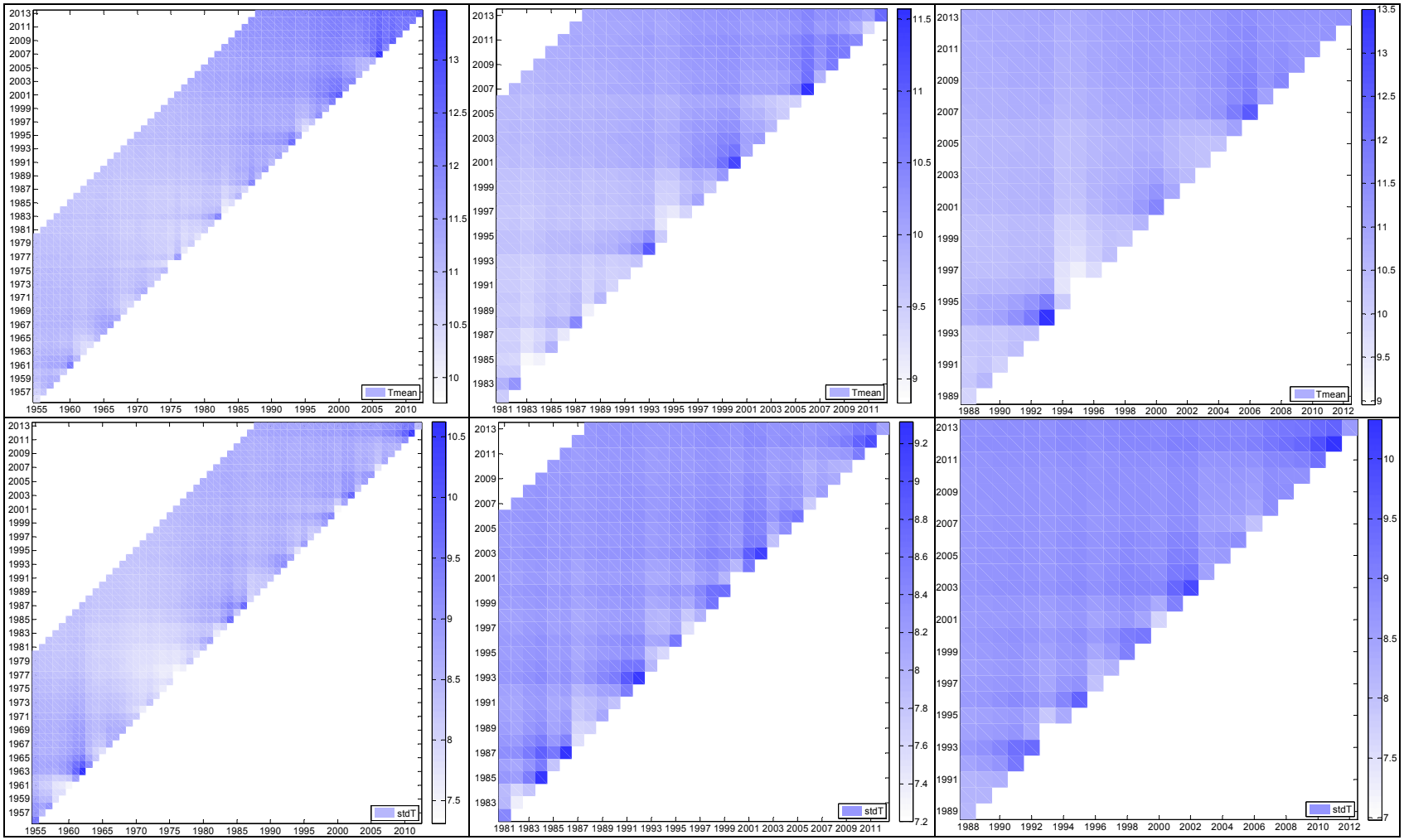
A. Figure 53. Mean annual number of rainy days (top panels) and standard deviation of precipitation (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



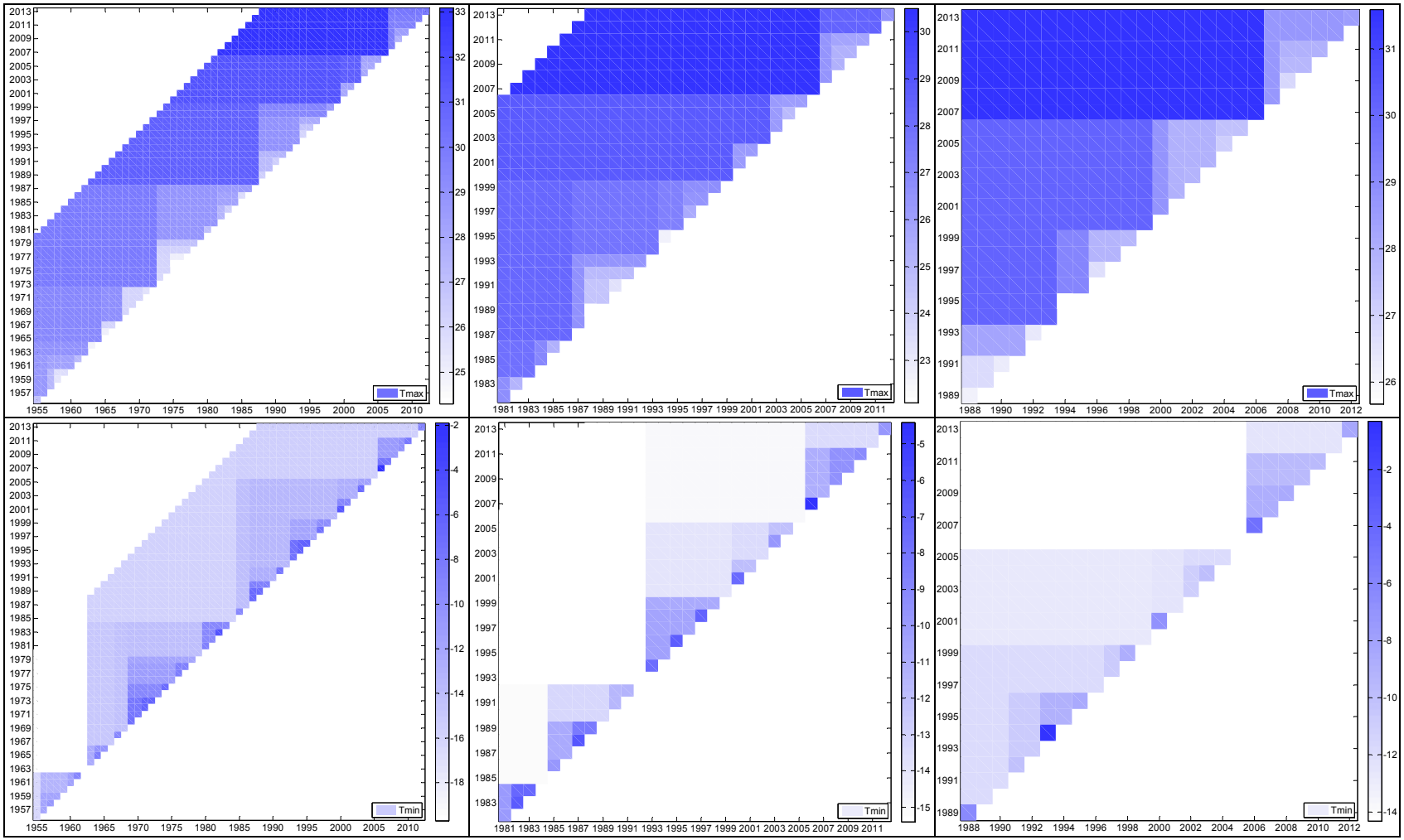
A. Figure 54. API 5 (top panels) and API 30 (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



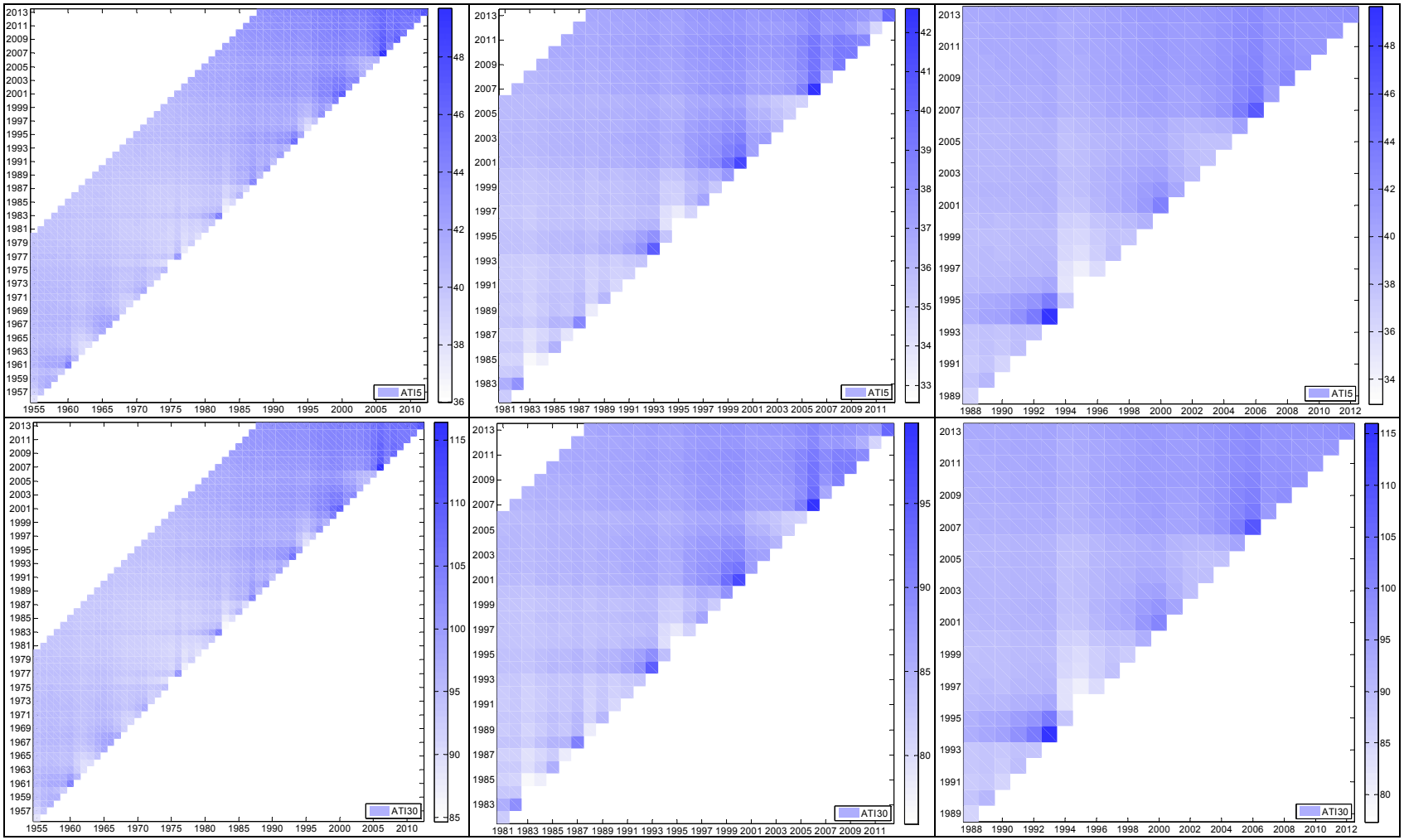
A. Figure 55. Maximum daily precipitation depths (top panels) and mean annual number of ice days (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



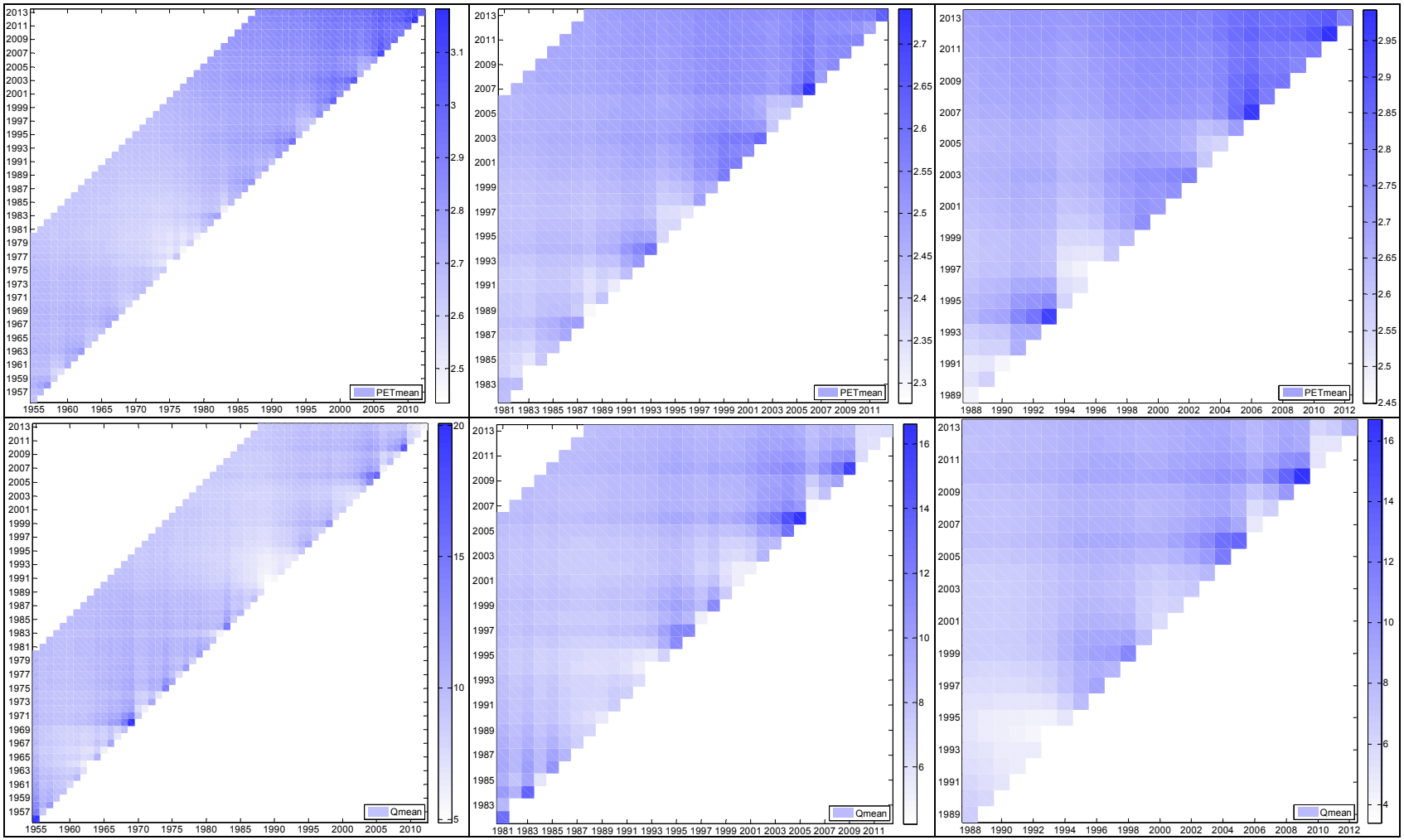
A. Figure 56. Mean daily temperatures (top panels) and standard deviation of mean daily temperatures (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



A. Figure 57. Maximum (top panels) and minimum mean daily temperatures (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



A. Figure 58. ATI 5 (top panels) and ATI 30 (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



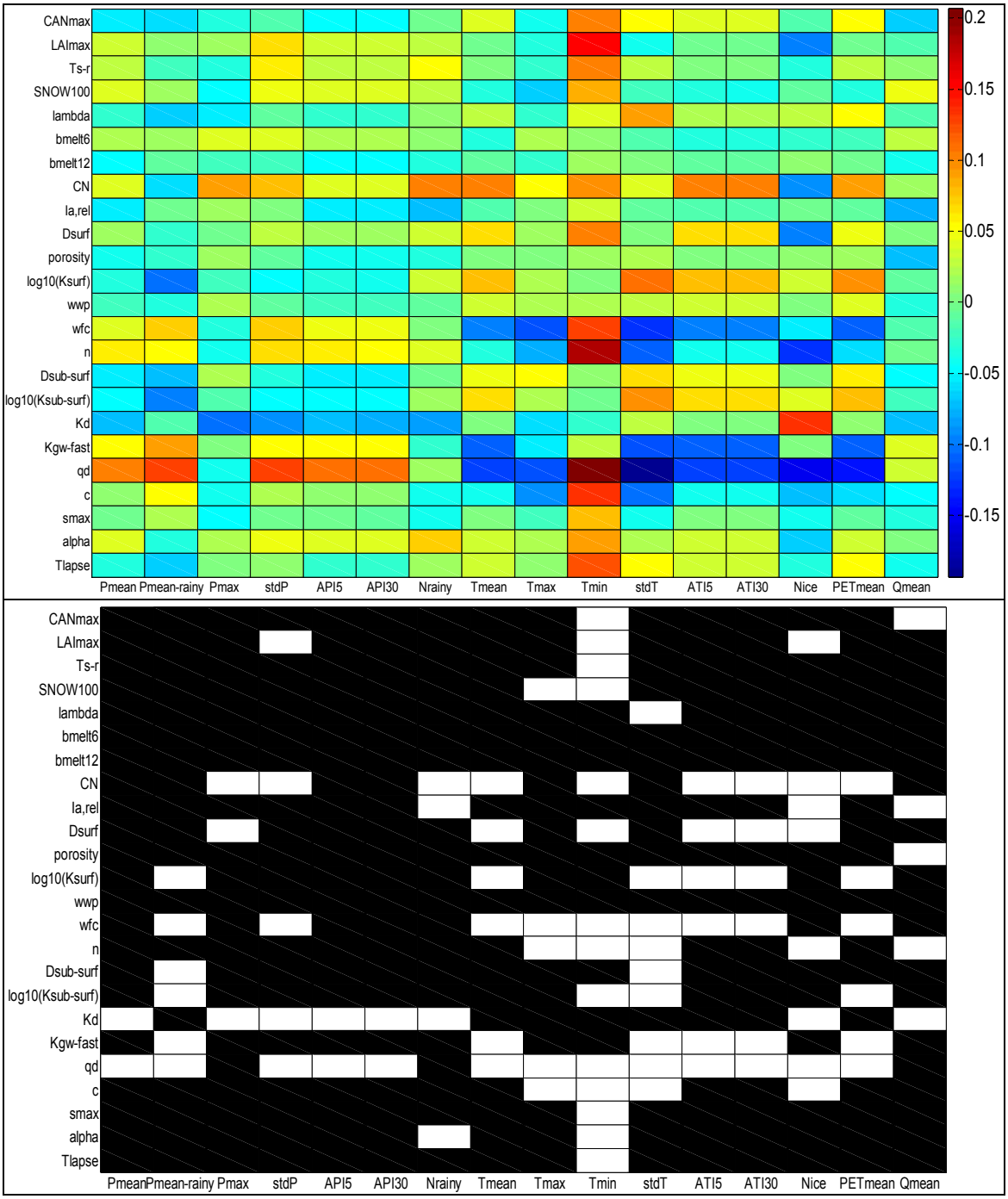
A. Figure 59. Mean daily *PET* rates (top panels) and mean daily flows (bottom panels): the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).

APPENDIX K. Relationship between the Pareto-optimal parameters and hydro-meteorological characteristics of the calibration period

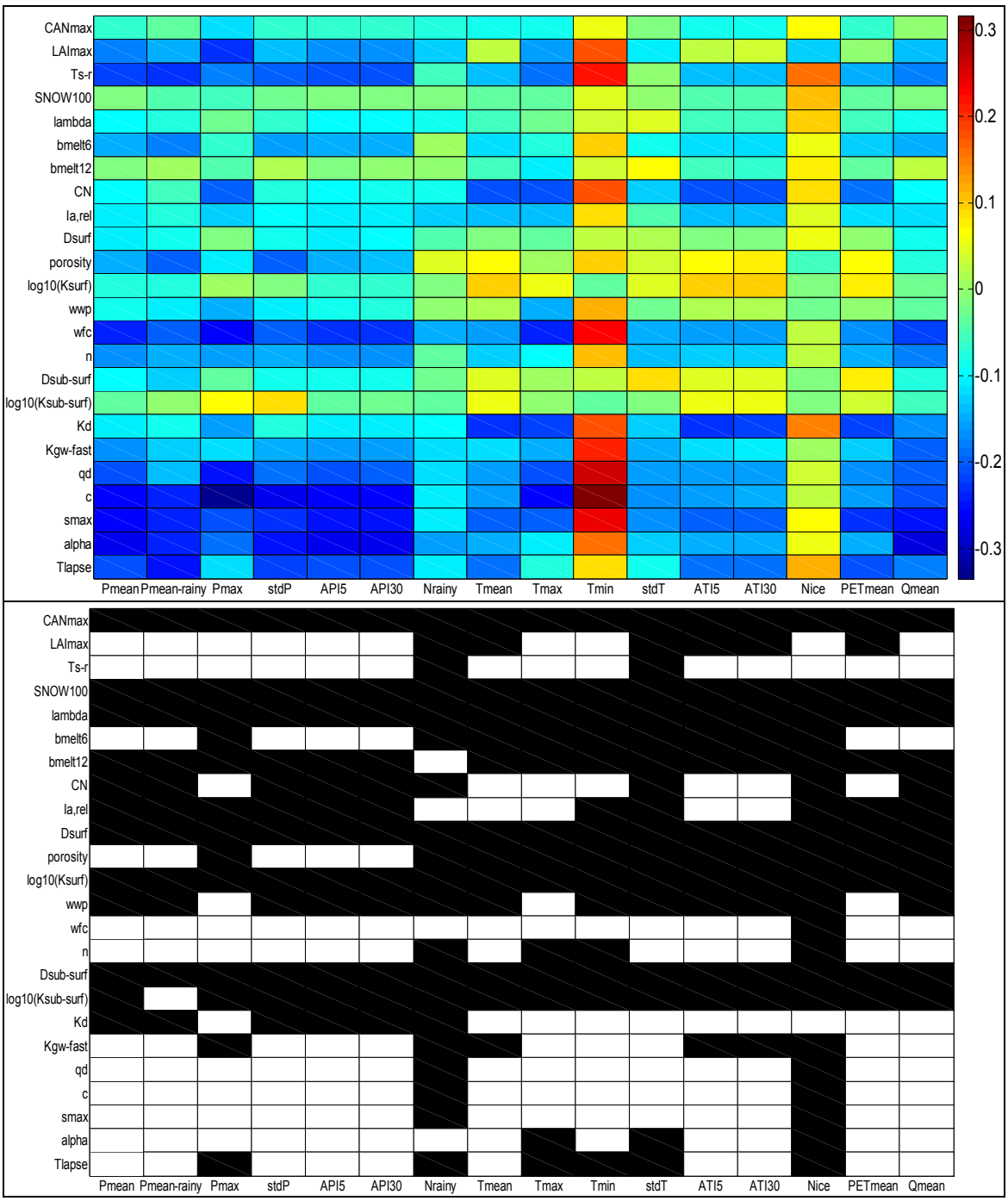
Relationships between median values of the Pareto-optimal parameters and hydro-meteorological indices are presented in A. Figure 60 through A. Figure 62 (Spearman correlation coefficients) and in A. Figure 63 through A. Figure 65 (variable importance in "tree bagging" metamodel).

Spearman correlation coefficients between the *IC* statistic and the hydro-meteorological indices are illustrated in A. Figure 66 through A. Figure 68. White fields in the bottom panels in these figures and in A. Figure 60 through A. Figure 62 denote statistically significant correlation at 95% significance level.

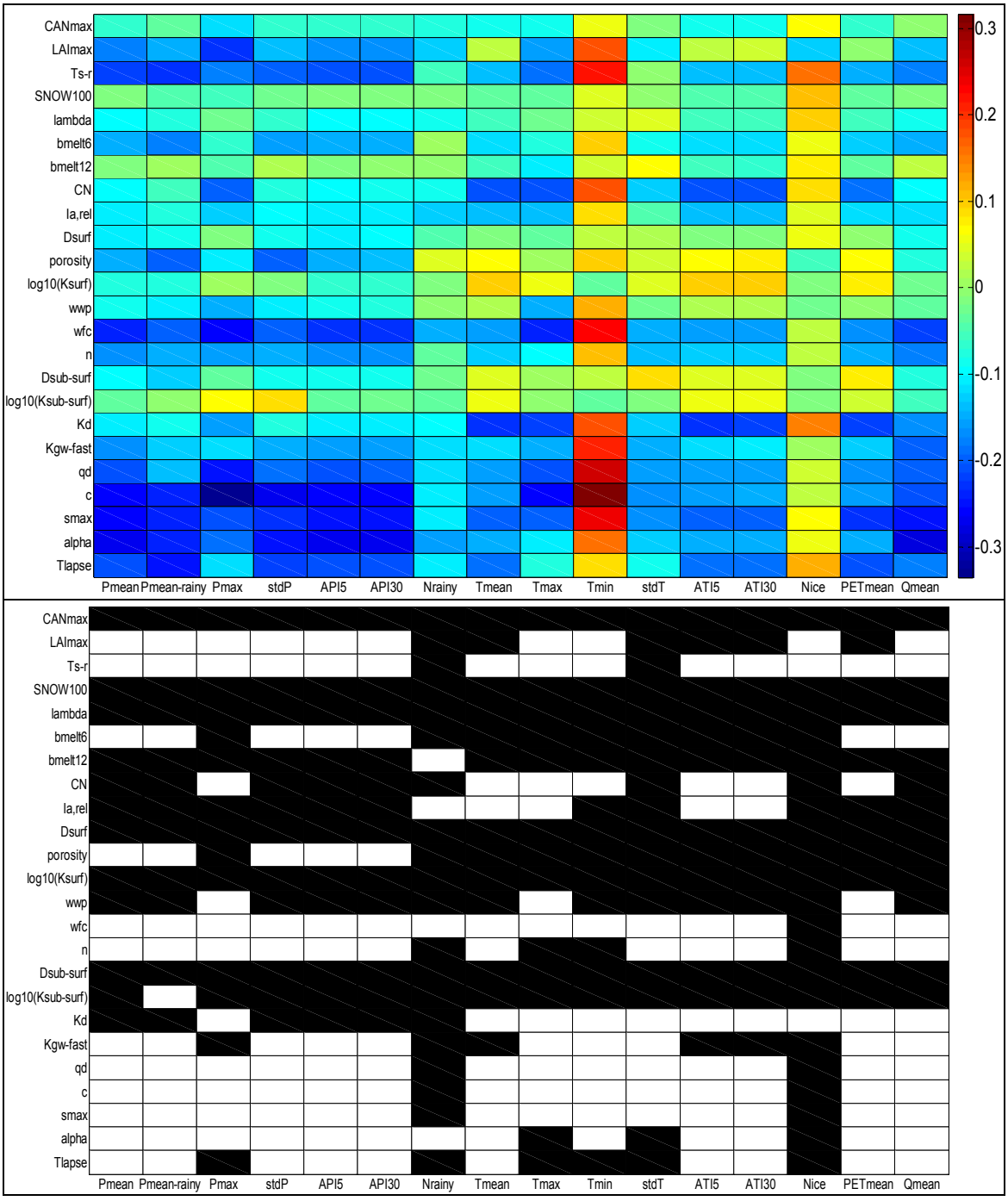
Correlation coefficients between median values of the objective functions and the hydro-meteorological indices are illustrated in A. Figure 69.



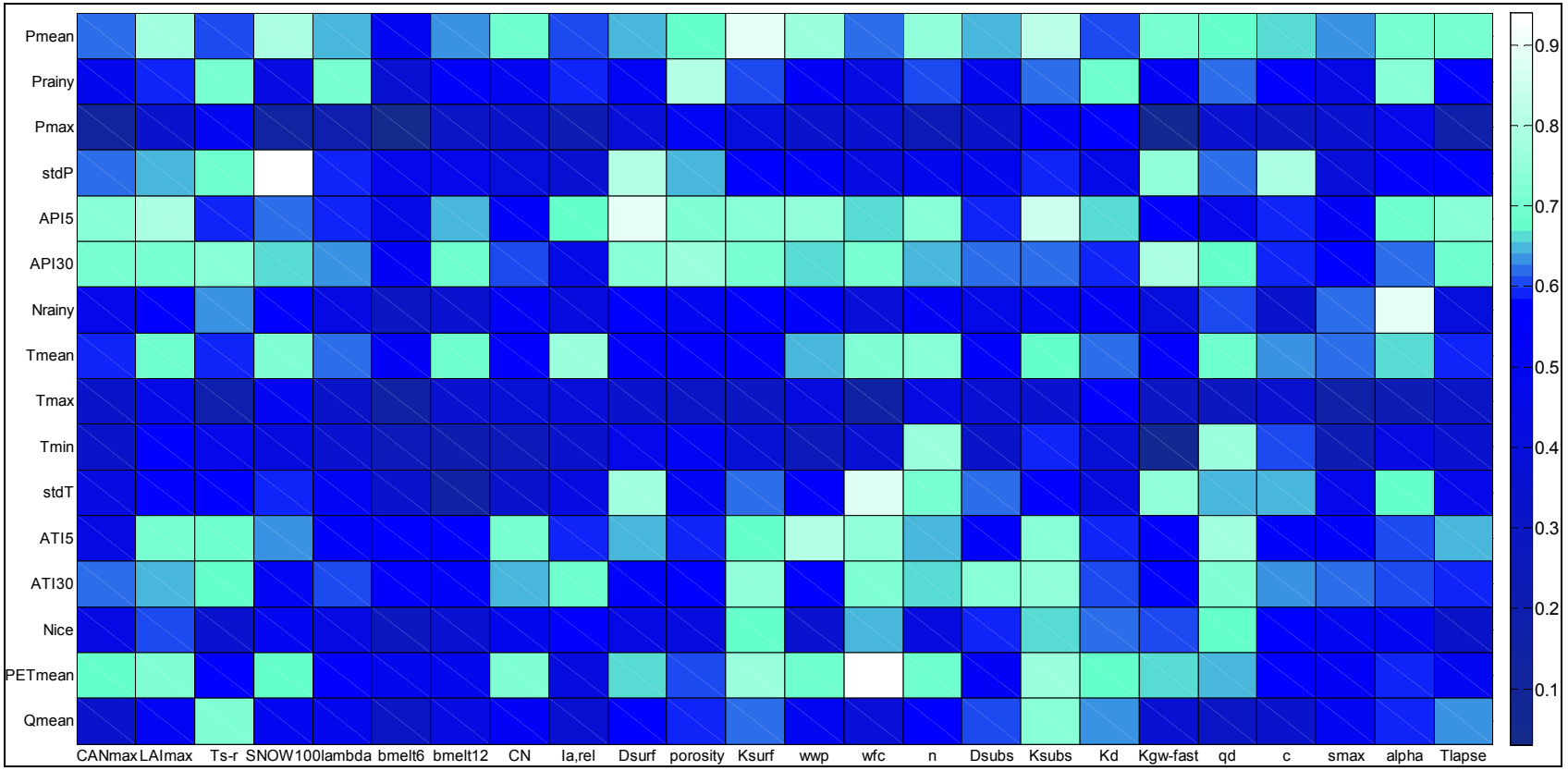
A. Figure 60. Spearman correlation coefficients between median values of the Pareto-optimal parameters and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Kolubara River catchment



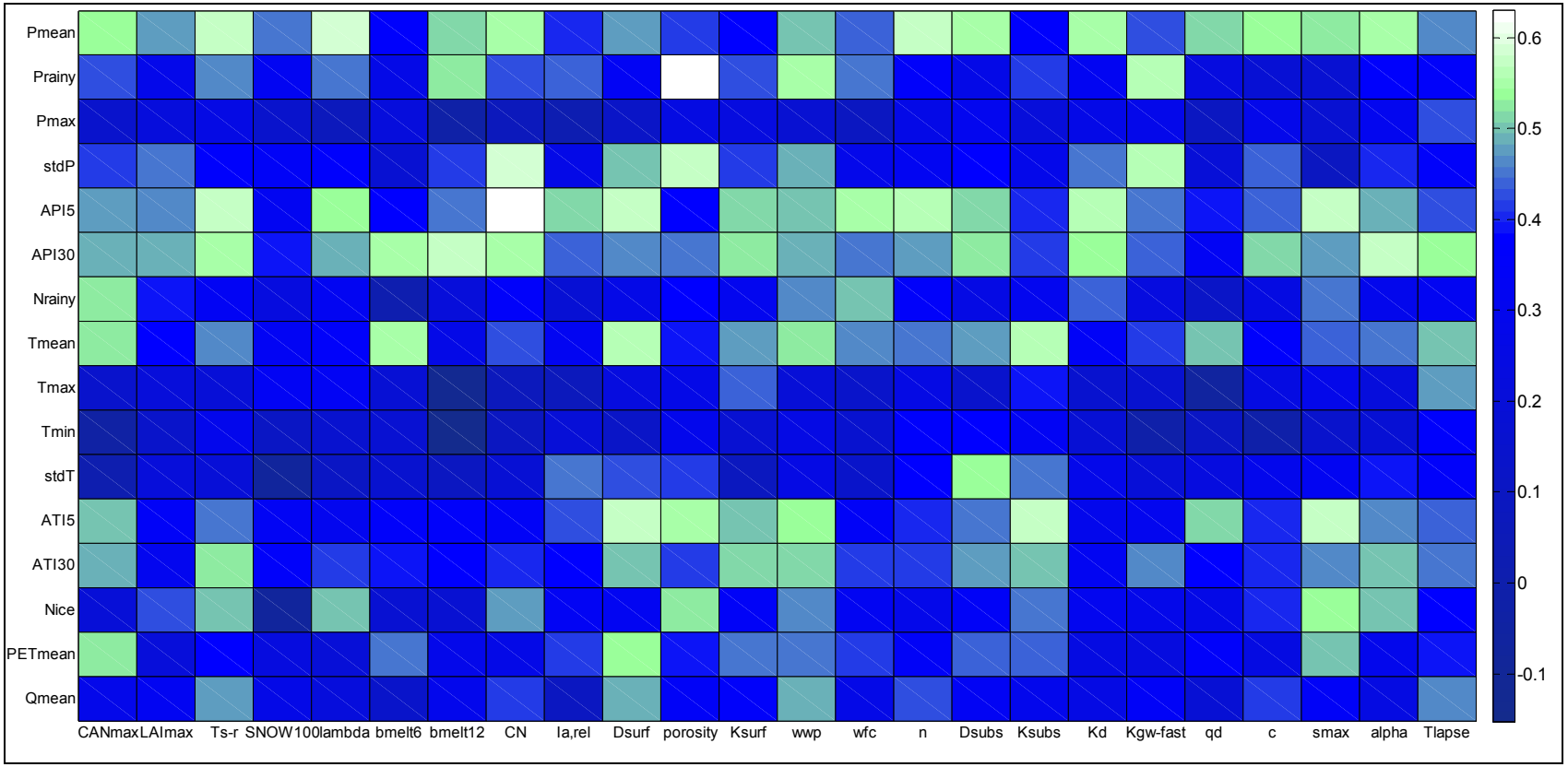
A. Figure 61. Spearman correlation coefficients between median values of the Pareto-optimal parameters and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Toplica River catchment.



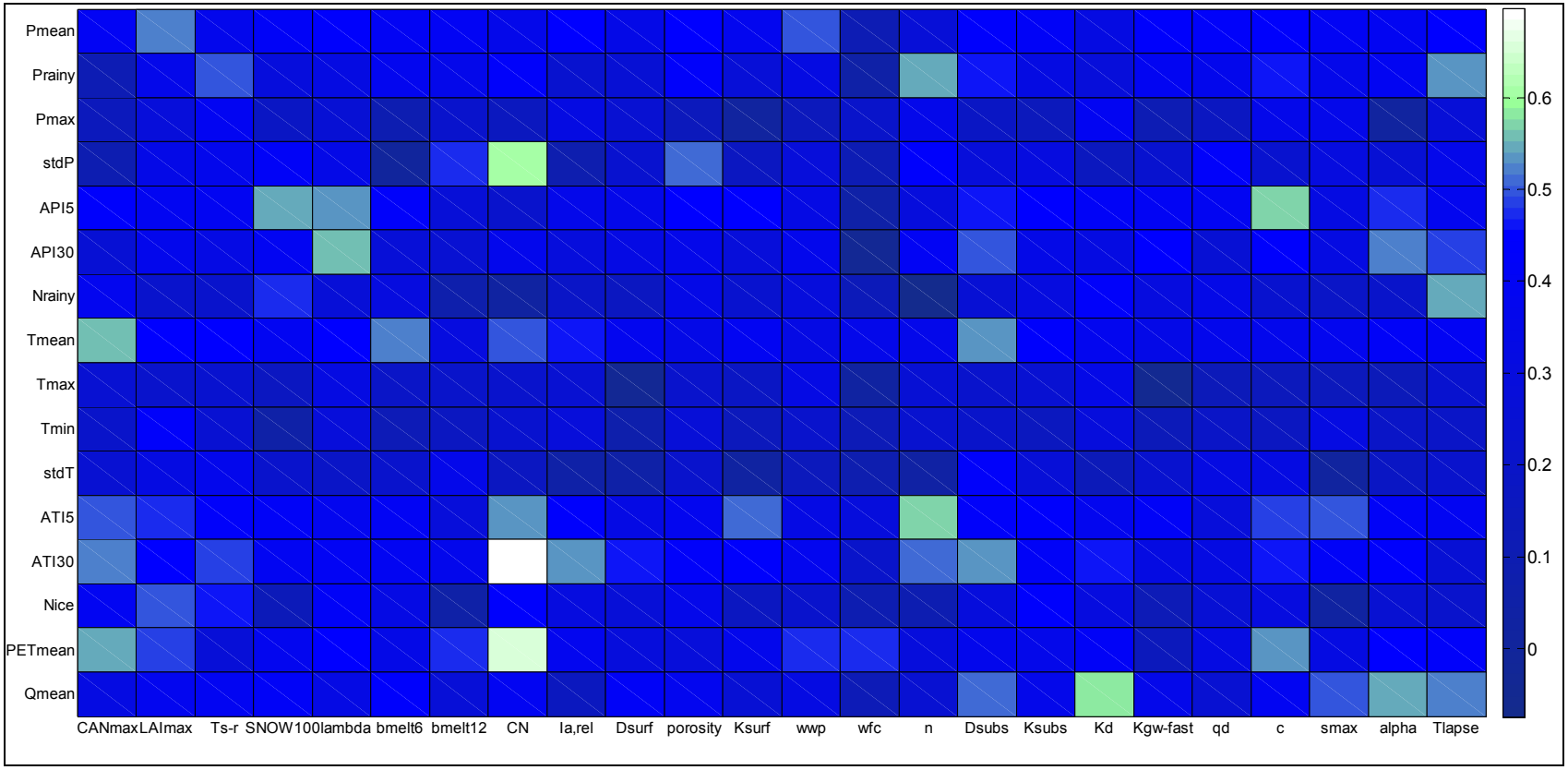
A. Figure 62. Spearman correlation coefficients between median values of the Pareto-optimal parameters and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Mlava River catchment.



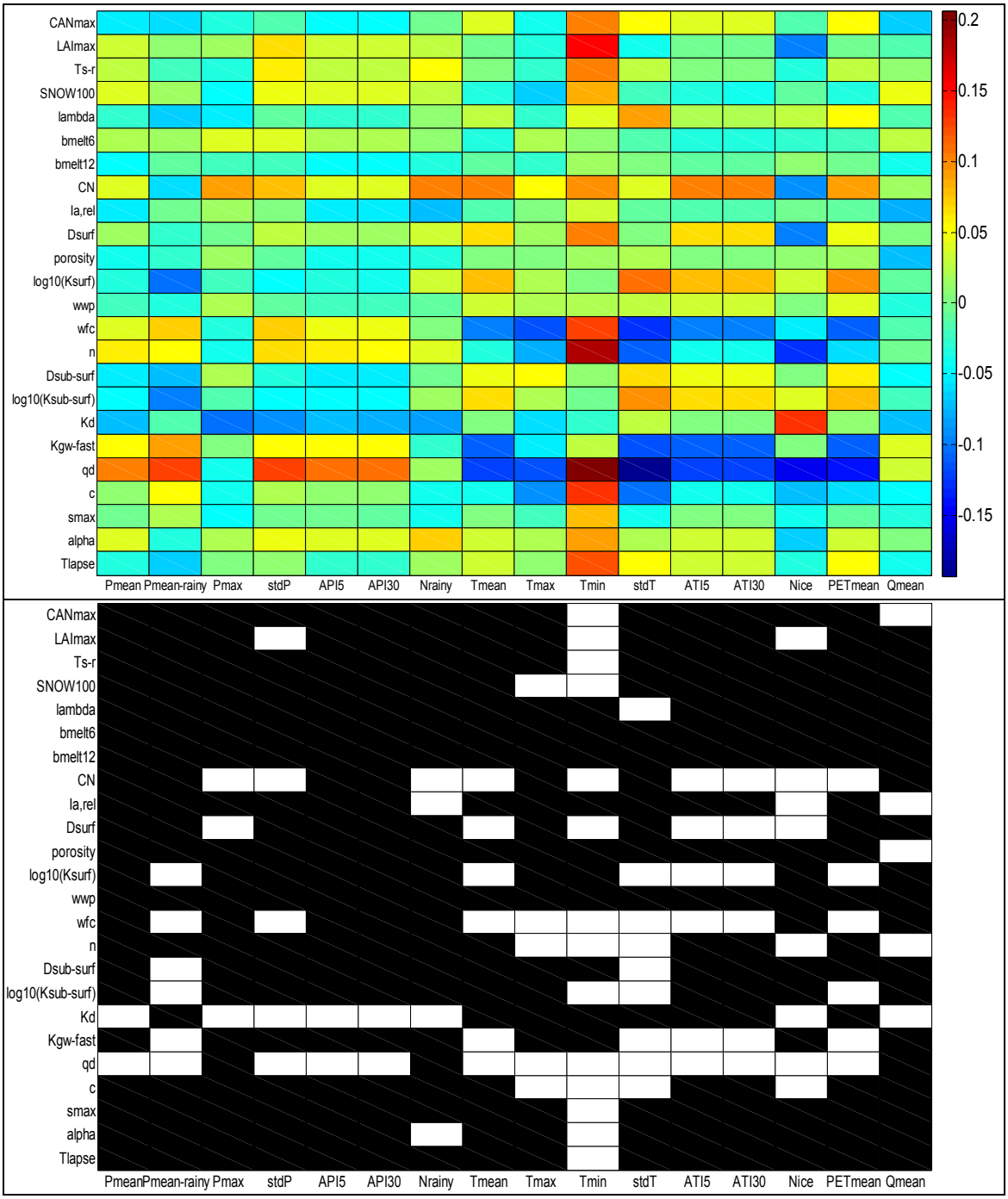
A. Figure 63. Variable importance in the "tree-bagging" metamodel of median values of the Pareto-optimal parameters: the Kolubara River catchment.



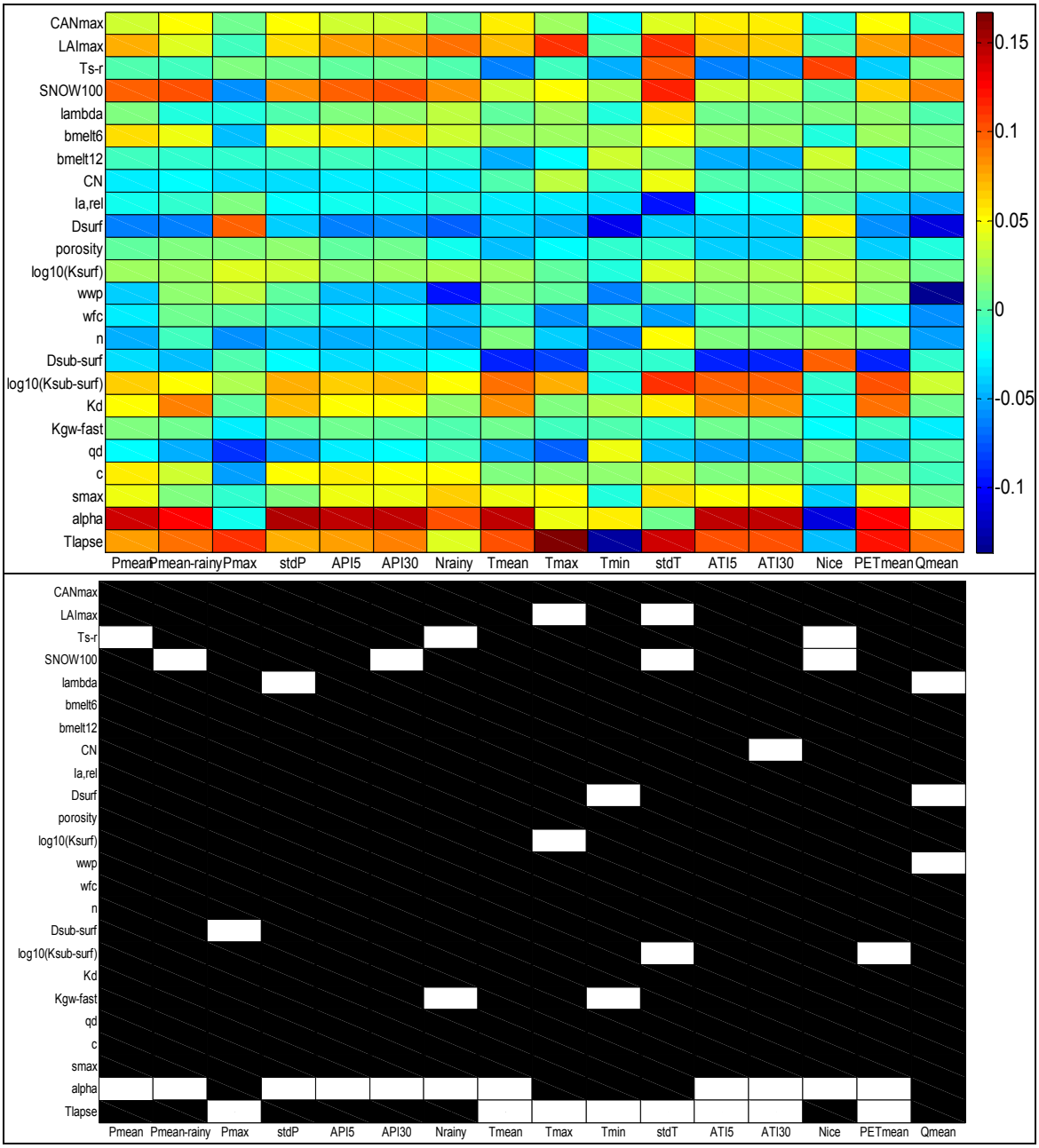
A. Figure 64. Variable importance in the "tree-bagging" metamodel of median values of the Pareto-optimal parameters: the Toplica River catchment.



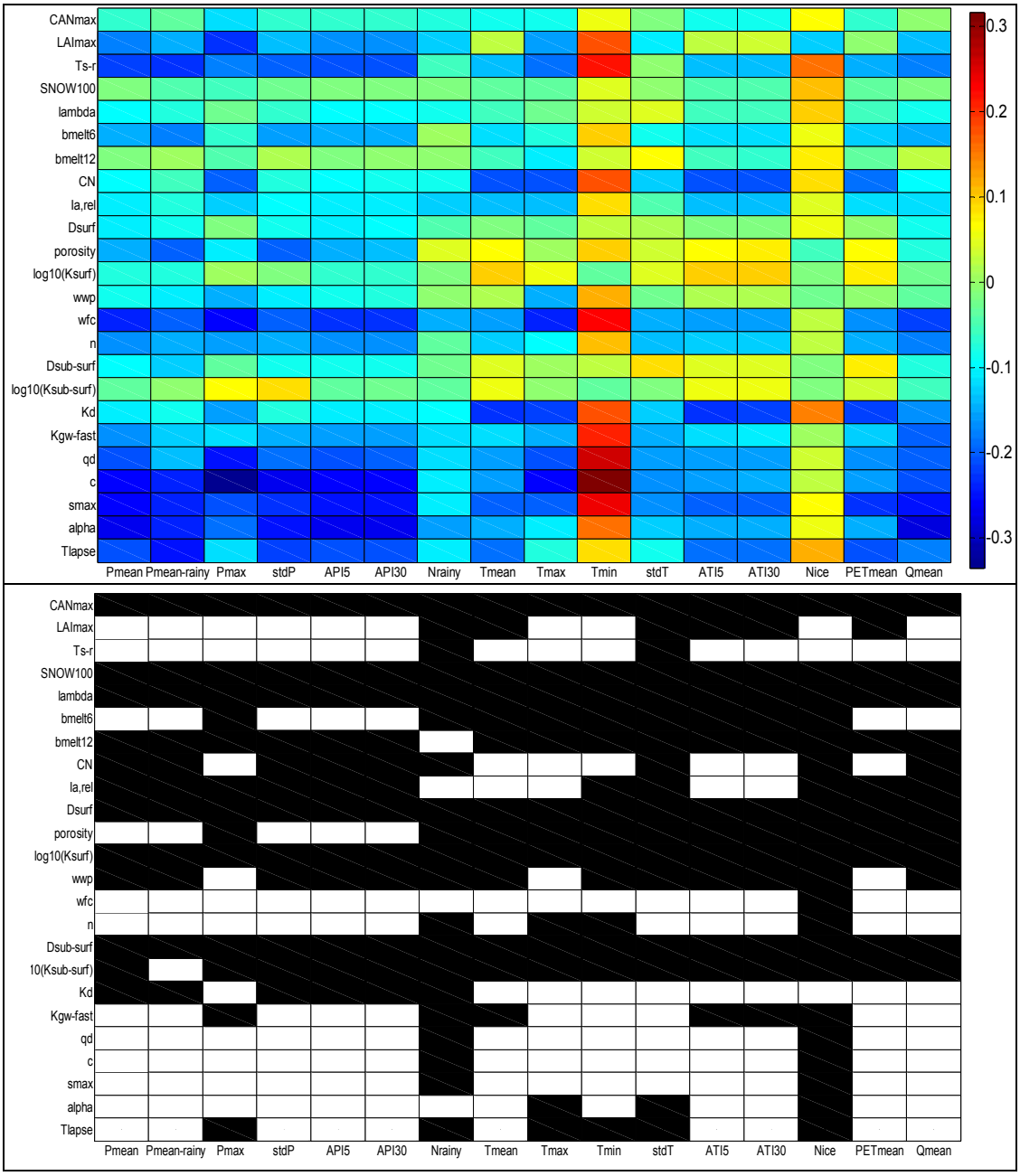
A. Figure 65. Variable importance in the "tree-bagging" metamodel of median values of the Pareto-optimal parameters: the Mlava River catchment.



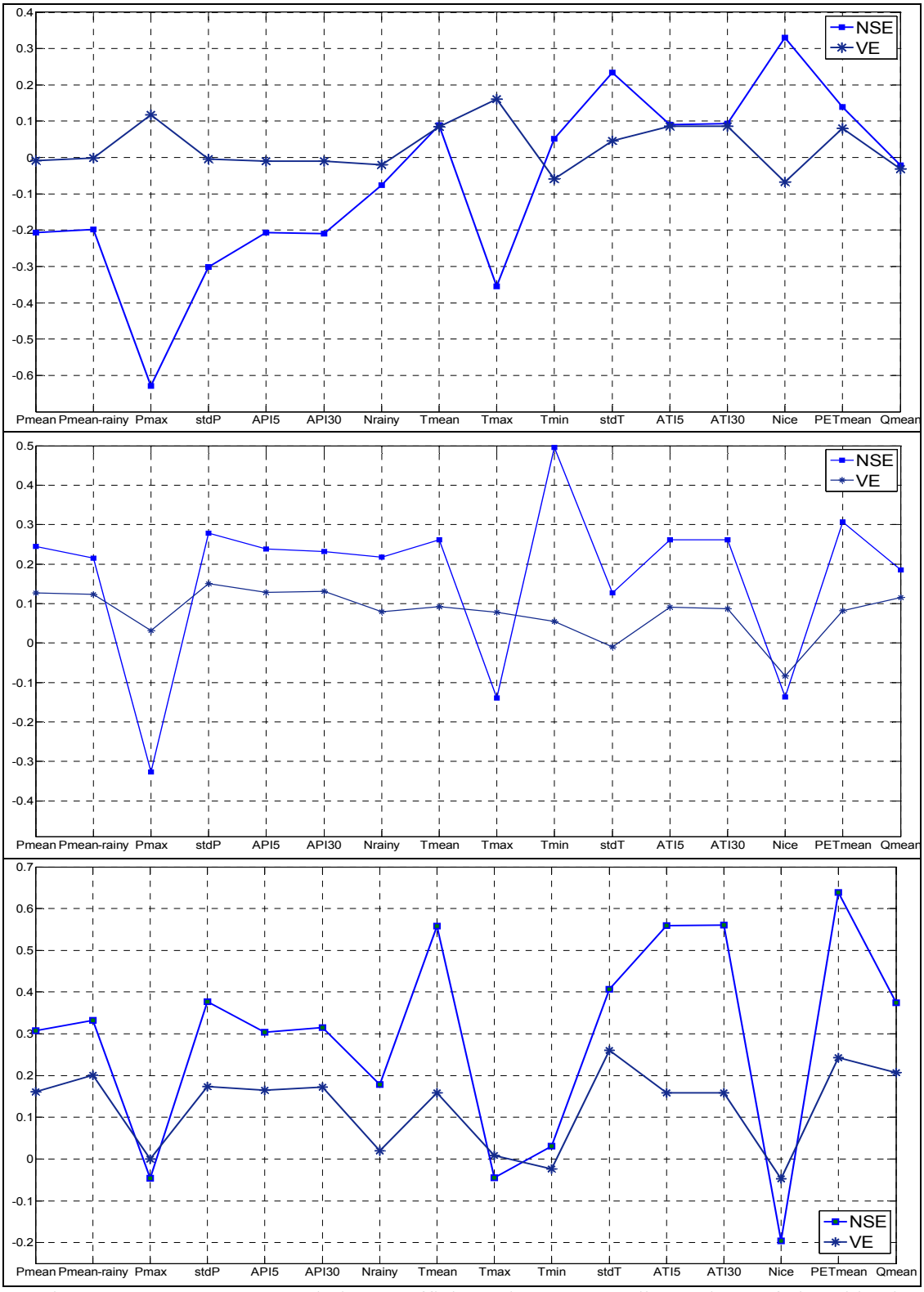
A. Figure 66. Spearman correlation coefficients between the *IC* statistic and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Kolubara River catchment.



A. Figure 67. Spearman correlation coefficients between the *IC* statistic and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Toplica River catchment.



A. Figure 68. Spearman correlation coefficients between the *IC* statistic and hydro-meteorological indices (top panel) and statistical significance at 95% significance level (bottom panel): the Mlava River catchment.



A. Figure 69. Spearman correlation coefficients between median values of the objective functions (Pareto-optimal ensemble) and hydro-meteorological indices: the Kolubara (top), Toplica (mid) and Mlava River catchments (bottom panel).

APPENDIX L. Model performance for different calibration strategies and for different model structures

A. Table. 8. Performance of the semi-lumped BASIC version of the model for different calibration strategies

Catchment	Objective functions	Mean N_{Pareto}(*)	Median *NSE*	Median *VE*	Median NSE_{logQ}	*p*-factor	*r*-factor	*p* / *r*
Kolubara	*NSE – VE*	24	0.65	0.996	0.44	0.268	0.218	1.2
	NSE - NSE_{logQ}	85	0.63	0.97	0.75	0.279	0.206	1.36
	KGE – VE	14	0.6	1	-0.25	0.242	0.237	1.02
	R^2 – *VE*	44	0.5	0.95	-4.2	0.301	0.308	0.98
	$RMSE_{HF} – RMSE_{LF}$	98	0.52	0.86	0.52	0.527	0.574	0.92
	HMLE – RMSE	39	0.61	0.94	0.55	0.07	0.045	1.5
	NSE - NSE_{logQ} – *VE*	92	0.63	0.98	0.73	0.365	0.267	1.36
Toplica	*NSE – VE*	15	0.75	0.998	0.62	0.291	0.225	1.29
	NSE - NSE_{logQ}	81	0.75	0.98	0.8	0.247	0.184	1.34
	KGE – VE	14	0.7	1	0.29	0.284	0.264	1.07
	R^2 – *VE*	69	0.53	0.92	-1.8	0.229	0.273	0.84
	$RMSE_{HF} – RMSE_{LF}$	99	0.67	0.91	0.25	0.402	0.416	0.96
	HMLE – RMSE	11	0.7	0.93	0.56	0.059	0.05	1.18
	NSE - NSE_{logQ} – *VE*	98	0.74	0.99	0.79	0.294	0.224	1.32
Mlava	*NSE – VE*	27	0.68	0.996	0.52	0.295	0.227	1.3
	NSE - NSE_{logQ}	87	0.6	0.96	0.74	0.274	0.199	0.77
	KGE – VE	13	0.58	1	-0.026	0.293	0.306	0.96
	R^2 – *VE*	43	0.46	0.95	-1.94	0.263	0.298	0.79
	$RMSE_{HF} – RMSE_{LF}$	100	0.57	0.88	0.32	0.474	0.544	0.87
	HMLE – RMSE	94	0.63	0.92	0.46	0.056	0.046	1.24
	NSE - NSE_{logQ} – *VE*	99	0.64	0.98	0.74	0.355	0.263	1.35

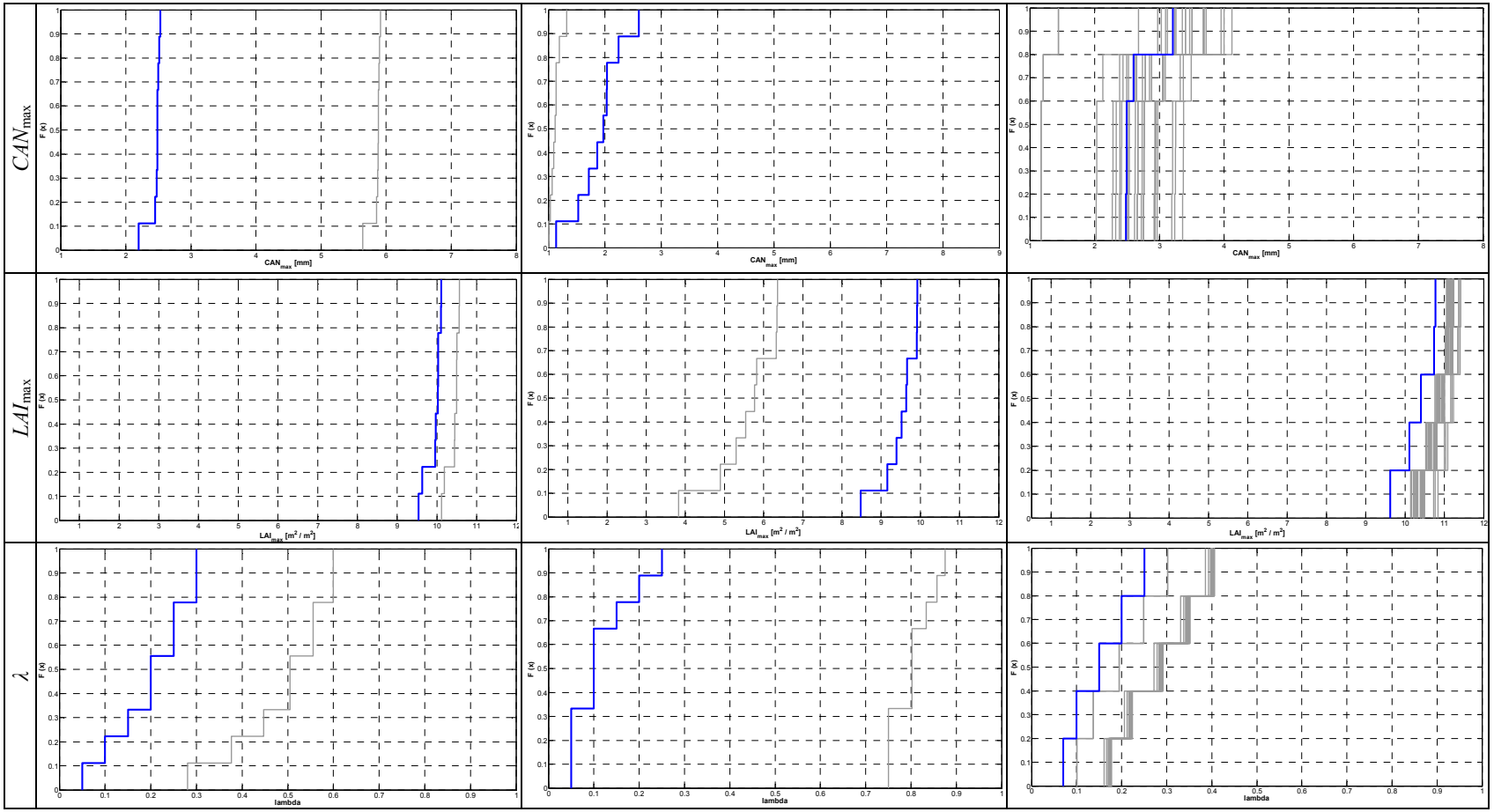
(*) *Values of the performance measures are calculated according to all 5-year long calibration periods.*

A. Table. 9. Performance of different versions of the 3DNet-Catch model

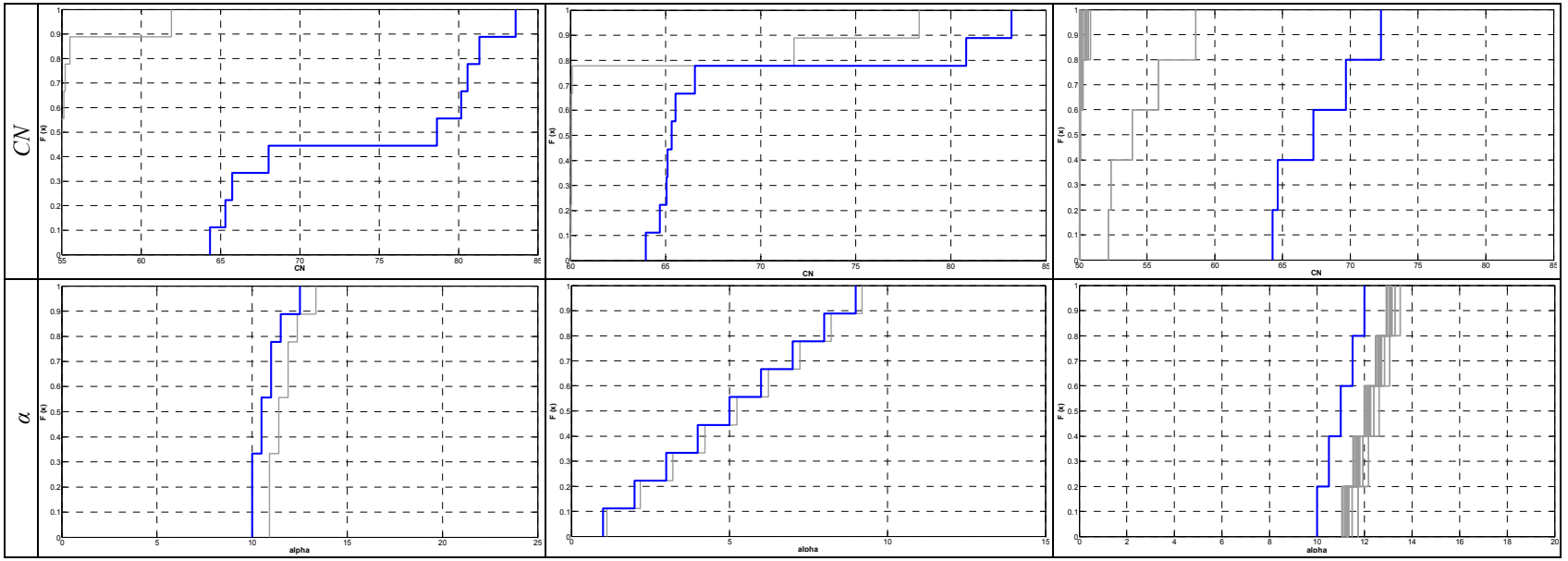
Catchment	Version of the model	N_{Pareto}	*NSE*(*)	*VE*	*p*-factor *r*-factor (**)
Kolubara	SIMPLE	26	0.66 0.65 0.56	1 0.996 0.97	0.26 0.2 (*1.31*)
	BASIC	24	0.66 0.65 0.57	1 0.996 0.96	0.27 0.22 (*1.23*)
	FULL	27	0.65 0.64 0.56	1 0.99 0.95	0.24 0.2 (*1.17*)
	BASIC_distributed	33	0.68 0.67 0.59	1 0.99 0.95	0.21 0.18 (*1.13*)
Toplica	SIMPLE	20	0.75 0.74 0.68	1 0.996 0.98	0.25 0.19 (*1.30*)
	BASIC	15	0.76 0.75 0.66	1 0.998 0.97	0.29 0.23 (*1.29*)
	FULL	22	0.73 0.72 0.60	1 0.997 0.97	0.3 0.27 (*1.10*)
	BASIC_distributed	21	0.73 0.72 0.62	1 0.999 0.98	0.28 0.23 (*1.21*)
Mlava	SIMPLE	34	0.67 0.60 0.59	1 0.99 0.96	0.212 0.17 (*1.27*)
	BASIC	27	0.69 0.68 0.58	1 0.999 0.96	0.3 0.23 (*1.3*)
	FULL	21	0.69 0.67 0.58	1 0.995 0.96	0.31 0.26 (*1.22*)
	BASIC_distributed	36	0.70 0.69 0.61	1 0.99 0.95	0.22 0.19 (*1.14*)

(*) *Values of the objective functions denote maximum, median and minimum of the Pareto ensembles, averaged over all 5-year long calibration periods.*

(**) *Values in parenthesis in the last column in A. Table. 9 denote ratios of p- factor to r-factor.*



A. Figure 70. Prior and posterior empirical cumulative distribution functions of the spatially distributed parameters in 2008-2013: the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).



A. Figure 70 (continued). Prior and posterior empirical cumulative distribution functions of the spatially distributed parameters in 2008-2013: the Kolubara (left panels), Toplica (mid panels) and Mlava River catchments (right panels).

БИОГРАФИЈА

Андријана Тодоровић је рођена 20. октобра 1983. године у Београду, где је завршила основну школу и VI Београдску Гимназију. Грађевински факултет Универзитета у Београду уписла је 2002. године. Дипломирала је 2008. године са просечном оценом 8,02. Дипломски рад „Картирање ризика од поплава у долини Топчидерске реке“ оцењен је највишом оценом и награђен од стране Привредне коморе Београда за најбољи Дипломски рад у 2008. години. Последипломске студије на Грађевинском факултету, одсек Грађевинарство, уписала је 2008. године. Од фебруара 2009. године запослена је на Грађевинском факултету као асистент-студент докторских студија на предметима на основним академским, мастер и специјалистичким студијама.

Ангажована је као истраживач на пројектима које финансира Министарство просвете, науке и технолошког развоја Републике Србије. Аутор је једног рада у међународном часопису са SCI листе, као и више радова у часописима од националног значаја и на научним скуповима.

Влада енглеским и служи се руским и немачким језиком.

Прилог 1.

Изјава о ауторству

Потписана Андријана Тодоровић

број индекса 907 / 08

Изјављујем

да је докторска дисертација под насловом

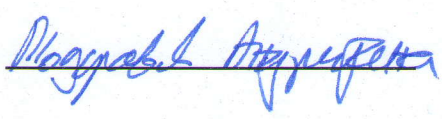
IMPACT OF CALIBRATION PERIOD ON PARAMETER ESTIMATES IN THE CONCEPTUAL HYDROLOGIC MODELS OF VARIOUS STRUCTURES

(Наслов на српском језику: УТИЦАЈ ПЕРИОДА КАЛИБРАЦИЈЕ НА ОЦЕНЕ ПАРАМЕТАРА КОНЦЕПТУАЛНИХ ХИДРОЛОШКИХ МОДЕЛА РАЗЛИЧИТИХ СТРУКТУРА)

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 1. 7. 2015.



Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Андријана Тодоровић

Број индекса 907 / 08

Студијски програм Грађевинарство

Наслов рада: IMPACT OF CALIBRATION PERIOD ON PARAMETER ESTIMATES IN THE CONCEPTUAL HYDROLOGIC MODELS OF VARIOUS STRUCTURES

Нслов на српском језику: УТИЦАЈ ПЕРИОДА КАЛИБРАЦИЈЕ НА ОЦЕНЕ ПАРАМЕТАРА КОНЦЕПТУАЛНИХ ХИДРОЛОШКИХ МОДЕЛА РАЗЛИЧИТИХ СТРУКТУРА

Ментори в. проф. Јасна Плавшић и в. проф. Милош Станић

Потписана: Андријана Тодоровић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 1.7.2015

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

IMPACT OF CALIBRATION PERIOD ON PARAMETER ESTIMATES IN THE CONCEPTUAL HYDROLOGIC MODELS OF VARIOUS STRUCTURES

(На српском језику: УТИЦАЈ ПЕРИОДА КАЛИБРАЦИЈЕ НА ОЦЕНЕ ПАРАМЕТАРА КОНЦЕПТУАЛНИХ ХИДРОЛОШКИХ МОДЕЛА РАЗЛИЧИТИХ СТРУКТУРА)

која је моје ауторско дело.

Дисертацију са свим прилозима предала сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучила.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 1. 7. 2015.