Teklu T. Hailegeorgis

Identification of spatially distributed Precipitation-Runoff response routines for hourly simulation in gauged and ungauged basins

Thesis for the degree of Philosophiae Doctor

Trondheim, January 2015

Norwegian University of Science and Technology
Faculty of Engineering Science and Technology
Department of Hydraulic and Environmental Engineering

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Engineering Science and Technology
Department of Hydraulic and Environmental Engineering

ISBN 978-82-326-0674-0 (printed ver.)
ISBN 978-82-326-0675-7 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2015:6

Printed by NTNU-trykk

Identification of spatially distributed Precipitation-Runoff response routines for hourly simulation in gauged and ungauged basins

A dissertation
submitted to
the Faculty of Engineering Science and Technology
at the
Norwegian University of Science and Technology
in fulfillment of the requirements for the degree of
Philosophiae Doctor (PhD)
by
Teklu T. Hailegeorgis

Research work supervised by:
Professor **Knut Alfredsen**
Norwegian University of Science and Technology, Trondheim (Norway)

Assessment Committee:
Professor **Jan Seibert**
University of Zürich, Zürich (Switzerland)

Dr. **Kolbjørn Engeland**
Norwegian Water Resources and Energy Directorate, Oslo (Norway)

Adjunct Associate Professor **Trond Rinde**
Norwegian University of Science and Technology, Trondheim (Norway)

Abstract

Due to sudden and considerable flow fluctuations during hydropeaking operation, forecasting of inflow to hydropower reservoirs at high temporal resolution (e.g. hourly) is required. Prediction of streamflow for both gauged and ungauged basins using Precipitation-Runoff (P-R) models is widely employed for operational purposes. However, there are various challenging factors and inherent uncertainties in the P-R modelling. Moreover, for the boreal Norwegian catchments there are research gaps for prediction of hourly streamflow related to identification of suitable parameterizations, model structures and regionalization methods for prediction in ungauged basins (PUB). Therefore, to address some of the research gaps, comprehensive calibration and post-calibration comparative evaluation of the performances of the different runoff response routines for both catchment and regional scale modelling are required. This is required to improve the runoff simulation based on observed input climate forcing (simulation mode) and hence for the improvement of hydrological forecast (forecast mode).

The objective of the first paper (P1) in this study was the identification of six different cases of explicitly resolved or probabilistic parameterizations of the spatial heterogeneity of a single state subsurface storage capacity. We conducted semi-distributed and distributed simulations based on the 'fill-and-spill' saturation excess, infiltration excess and subsurface drainage runoff mechanisms. Equivalent performances of simulation from the different cases indicate the unidentifiably of the parameterizations and hence a preference for a parsimonious simple distributed parameterization. Identification requires more representative input climate data than a mere calibration problem. In addition, calibration only to streamflow data cannot fully identify the parameterizations.

In light of the findings from the first paper, we conducted a study on identification of parametrical parsimonious and more complex configurations of the widely used conceptual Hydrologiska Byråns Vattenbalansavdelning (HBV) runoff response routine in the second paper (P2). Despite equivalent streamflow simulations for the tested HBV variants, a parametrical parsimonious HBV routine (HBV-Parsim) provided better parameter identifiability and more reliable baseflow simulation. In the other variants, considerable interactions between the soil moisture accounting and the response routine

parameters and compensation between the outflow from the upper reservoir and the baseflow from the lower reservoir affect the reliability of the simulation. Hence, evaluation of the reliability of internal simulations of baseflow and soil moisture by the widely used HBV routines against tailor-made analytical methods or observations is necessary.

Inspired by the preference to parsimony in P1 and P2 and the compensation between the fluxes from multiple storage reservoirs of the conceptual model in P2, our objective in P3 was geared towards the evaluation of the performance of a distributed version of a 'top-down' parsimonious single storage routine (hereafter named Kirchmod). We applied the principle of catchments as simple dynamical systems following Kirchner (2009) for a macroscale (3090 km^2) mountainous catchment of considerable runoff delay compared to the hourly simulation. In this case, we both set the response routine parameters by estimation from streamflow recession analysis and by calibration. We obtained simulated streamflow hydrographs and flow duration curves that are in good agreement with the observed and transferability of the optimal parameter sets to the interior subcatchments validated the model. However, the parameter calibration provides slightly better simulation of peak flows than estimation from the recession analysis. In addition, the various sources of uncertainty in parameter estimation needs thorough assessment. There is no marked influence of the runoff delay due to the correlation among the free parameters, which indicates problems of parameter non-identifiability even for the parsimonious routine.

Based on the findings from the catchment scale performances of the P-R response routines in P1, P2 and P3, we wished to address the issues of Prediction in Ungauged Basins (PUB) through multi-model identification of different regionalization methods on 26 catchments in a mid-Norway study region in P4. We found that the best performing regionalization methods for the catchments vary among the model structures and evaluation metrics. However, based on the regional performances, the regionalization methods based on the single-donor physical similarity and the multi-donor regional calibration corresponding to maximum regional weighted average (MRWA) performance measures (PM) performed better than the nearest neighbor and regional median parameters. The lack of data on the subsurface physical attributes and high- density hourly hydro-climatic gauging networks for the region can affect the performances of the regionalization methods, which needs scrutiny in future endeavors.

The fifth objective was the identification of distributed P-R response routines relevant to operational purposes based on a multi-basin (26 catchments) local and regional calibration in P5. The best performing model structure(s) vary among the catchments and the evaluation metrics and hence there is no unique model structure that performs best for all catchments in the region. However, the Kirchmod followed by the BGM perform better than the various configurations of the HBV routine for the majority of the catchments and in terms of the regional calibration (MRWA) for the PUB. Therefore, flexible models and a multi-basin modelling framework, which allow identification of models for hourly simulation among a pool of plausible options for several catchments in the region, is better than the common single catchment model for operational purposes.

In P1 to P5, we observed the challenges in identifying a unique regional P-R response routine due to the uniqueness of catchments runoff response and various sources of uncertainties. The last objective of the thesis was the development of data based statistical model and comparative evaluations against the best performing P-R response routine for hourly prediction in an ungauged and regulated basin for ecological applications in P6. A simple regional regression model based on the relationship among streamflow percentiles and catchment drainage areas, and regional transfer of streamflow information to the nearest neighbor catchment performed better than the MRWA based transfer of model parameters using the Kirchmod. We found the simple regional regression model to be useful to predict a natural time series of streamflow in a regulated river to derive ecologically relevant streamflow metrics, for assessing hydrological alterations due to regulation and hydropeaking and environmental flows.

Preface

I submit this thesis to the Norwegian University of Science and Technology (NTNU), the Faculty of Engineering Science and Technology (Trondheim) for fulfilment of the requirements for the degree of Philosophiae Doctor (PhD). I have conducted this doctoral study at the Department of Hydraulic and Environmental Engineering under supervision of Professor Knut Alfredsen.

The Center for Environmental Design of Renewable Energy (CEDREN), which obtained funding from the Research Council of Norway and some Norwegian hydropower companies, provided a financial support for this CEDREN-hydroPEAK-hydrology PhD project (Project number: 50043420).

In accordance with the guidelines of the Faculty of Engineering Science and Technology for paper-based thesis, the thesis is composed of an introduction to the research, summary of main findings, conclusions and six appended papers.

Acknowledgments

I have an interest for sustainable water resources development through a blend of engineering knowledge with socio-environmental and economic-financial dimensions. Back in 2005 while I was looking for a further study, I got an admission to the master's program in hydropower development (2005-2007) at NTNU (Department of Hydraulic and Environmental Engineering). While working on a master's thesis on distributed hydrological modelling, I became interested to conduct a PhD study in the field.

I obtained a funded PhD 'stipendiat' position in 2010 at the same institution under the CEDREN-hydroPEAK-hydrology. It was very appealing to me to contribute towards endeavors for tackling the challenges in hydrological predictions. The path was very long to reach to this destination, which gave me opportunities to learn many lessons and gain skills along the way. I would like to express my deepest thanks for all of you, who have helped me to accomplish this work. Professor Knut Alfredsen, you have always been available for me to supervise the overall PhD research. Your encouraging advices, inspirational thoughts and positive attitudes are highly appreciated. Yisak S. Abdella and Sjur Kolberg (from SINTEF Energy Research AS), I strongly acknowledge your cooperation and support on ENKI hydrological modelling framework and object oriented programming, and sharing part of the data used for the research. Thanks to Yisak also for providing parallel computing and squeezed region versions of the dynamic-link libraries (dlls) of the ENKI routines to improve the computational efficiency of the regional calibration.

In addition, I am grateful to Professors at the department (Knut, Ånund, Leif, Sveinn and others) for the coursework. I would like to convey my thanks to the entire administrative and technical staffs at the department (Geir, Ingjerd, Hege, Hilbjørg, Varshita, Brit and others) for your support. My sincere thanks go to previous and current PhD and Postdoc colleagues at the department, members and financers of the CEDREN and institutions who provided the data used for the research.

My wife Hiwot and our little son Milki, you deserve utmost appreciation for your understanding. Thanks to all my relatives and friends for your encouragement. All praise and glory be to the Almighty God "With God All Things Are Possible".

Teklu T. Hailegeorgis 06.10.2014 Trondheim

Contents

List of Figures

List of Tables

Author's declaration

I, the author of this thesis, performed the works in this thesis from planning to running of experiments and writing of the papers. The co-authors in paper 1 (P1), paper 3 (P3) and paper 4 (P4) contributed valuable comments to improve the manuscripts. For the P4, the second author, Yisak S. Abdella, provided parallel computing and squeezed region versions of the dlls to improve the computational efficiency of the regional calibration. The Professor supervised the whole PhD research.

Chapter 1 INTRODUCTION

Context and motivation

Anticipating increased energy production from unregulated and intermittent renewable energy sources in the European power market, the Norwegian hydropower industry envisages utilization of the large storage capacity of reservoirs for hydropeaking. Therefore, the need for sustainable use of water resources for hydropower production is becoming even more important due to increased anthropogenic pressure on the resources. To achieve this, prediction of inflow to storage reservoirs are important for optimal utilization and an environmental friendly management. Both single catchment and regional scale predictions based on continuous simulation of streamflow by precipitation-runoff (P-R) models on a daily temporal resolution have wide and still evolving applications within hydrology. However, the hydropeaking operation would require improved inflow forecasting for short-term (e.g. hourly) peaking operation, flood management and assessments of negative ecological impacts of the operation. In addition, using parameters calibrated for a coarser temporal resolution for prediction of hourly streamflow creates an additional source of uncertainty in the hourly prediction.

In real-time forecasting for operational purposes, the accuracy of forecast is highly affected by the uncertainties in the simulation relevant to input data, parameter calibration and model structure and uncertainties in the forecast for instance due to weather forecast and model states at the start of forecast. Therefore, improved modelling approaches that allow opportunities for better representation of the spatial variability of input climate forcing especially of the precipitation variability in snow dominated and high altitude regions and utilization of available data from grid-based measurements would be necessary. Current applications of spatially distributed (gridded) models for scenario assessment of the impacts of land use and climate change are common. However, there are growing interests towards utilization of distributed precipitation-runoff models for prediction purposes and inclusion to a real-time forecasting tool. Therefore, the potential utilities of distributed precipitation-runoff model algorithms for improved prediction of hourly runoff require thorough investigations. Identification of parameterization, model structure and modelling approaches that better simulate the dominant hydrological processes for hourly runoff responses based on observed input climate forcing (simulation mode) and hence a potential for better hydrological forecast (forecast mode) are required through comprehensive calibration and post-calibration comparative evaluations, which are the focus of this thesis.

Precipitation-runoff modelling entails different procedures for transformation of precipitation to runoff and hence several factors affect the performance. The major factors are the quality of climate forcing, parameterization of spatial heterogeneities and issues of scales, model structure and procedures related to model calibration and validation. The quality of the climate forcing and streamflow data used for model calibration is a crucial factor that affects calibration and simulation performance. Explicit or probabilistic parameterization of grid-to-grid heterogeneities and probabilistic parameterization of subgrid or subelement scale heterogeneities involve the issues of selection of spatial scale and parametrical parsimony versus complexity. Due to the uniqueness of catchments, the performance of the model structures that are different in terms of conceptualizations and modelling paradigms could be markedly different for different catchments. The model calibration involves calibration algorithms, objective functions, performance measures, parameter identifiability and uncertainty whereas the model validation involves temporal and spatial transferability of the calibrated parameter set to interior catchments or homogeneous catchments in the region.

However, prediction of streamflow based on at-site or local calibration of precipitation and streamflow relationships is possible only for gauged basins, where streamflow records of sufficient length is available. This can be a challenging problem in hydrology particularly when catchments contributing to the reservoir inflow are ungauged or poorly gauged. In addition, it is difficult to predict natural flow in many regulated rivers for assessment of flow alterations and impacts on the ecological integrity. Therefore, there need to be methods available for prediction of streamflow contributions for ungauged basins. In this regard, the issues of Prediction in Ungauged Basins (PUB) (Sivapalan, 2003) have become an active research area in hydrology especially during the decade of PUB (Hrachowitz *et al*., 2013). However, the focus has been mainly on prediction with a daily or coarser temporal resolution.

Therefore, prediction in ungauged basins for hourly temporal resolution through regional calibration or other regionalization techniques require a thorough assessment. In addition to the PUB, the regional calibration allows for input data augmentation by utilizing all the available data in the region and multi-basin based evaluation of the P-R response routines. However, regional scale prediction by utilizing the P-R response routines through transfer of model parameters is a challenging task due to the heterogeneities in catchments runoff response, the inherent uncertainties in input data and limitations in parameter calibration procedures. Therefore, comprehensive attempts for identification of more suitable P-R response routines based on both local and regional calibration and identification of regionalization methods for the PUB are very important. In addition, investigation of the potential utility of statistical models for hourly prediction in ungauged basins allows data based transfer of information, which avoids the challenges associated to calibration of the P-R routines.

Current state of research and research questions

There are several studies on precipitation-runoff models for forecasting purposes in different geographic zones. Some examples are the Probability Distributed model or PDM model (Moore *et al*., 2005; Moore, 2007) for United Kingdom, the HBV model e.g. Huttunen and Vehvilainen (2001) for Finland, Kobold and Brilly (2006) for Slovenia, Blöschl *et al*. (2008) for Austria, Olsson and Lindström (2008) for Sweden, Şorman et al. (2009) for Turkey, Renner et al. (2009) for Rhine basin, Engeland *et al*. (2010) for northern Norway and Engeland and Steinsland (2014) for south-western Norway, the TOPKAPI model (Bartholmes and Todini, 2005) for Italy and multi-models (Velàzquez et al., 2011) for France.

However, many of the previous studies conducted model calibration and forecast at daily time scales while the hydropeaking operation requires forecasts for short-term peaking operation (e.g. hourly). Several studies indicated the time scale dependencies of conceptual model parameters. Kavetski *et al*. (2011) investigated the time scale dependencies of information content of data, parameter calibration and identifiability, quick flow and hydrograph peak simulation. Bastola and Murphy (2013) illustrated considerable loss in performance by using parameters calibrated for a daily streamflow for simulation of hourly streamflow, which introduces additional sources of uncertainty to the hourly prediction. Therefore, even if there is probably lack of hourly records from high-density hydrometric networks compared to the daily resolution, there are clear advantages of hourly prediction based on model parameters calibrated utilizing the hourly observations. Regarding the spatial scale, most of the previous forecasting models were lumped conceptual models as noted in the distributed model intercomparison project or DMIP (Smith *et al*., 2004; Reed *et al*., 2004). Further, Smith *et al*. (2012) noted an expanded use of spatially distributed watershed models by the US National Weather Service (NWS). Therefore, there are several open research questions pertinent to the utility of spatially distributed P-R models for hourly prediction for operational purposes in the boreal climate regime and landscape features.

However, thorough diagnostic evaluation of the behavior of the P-R models in simulation mode is indispensable since the quality of real-time forecast is dependent on the process simulation (e.g. Bell and Moore, 1998; Refsgaard, 1997). The simulation performance of the P-R models is subject to uncertainties from various sources, which include climate forcing and streamflow observations used for the calibration, model structure, parameter calibration, spatial scales for parameterization of the spatial heterogeneities and temporal scales of observations. Moreover, the discrepancy between optimal rainfall-runoff models for engineering purposes and optimal rainfall-runoff models for scientific investigations of the overall basin behavior as noted by Wagener and McIntyre (2005) requires further research progress.

Continuous simulations of streamflow in different climate regimes for the PUB based on different regionalization methods to mention a few include regional calibration (e.g. Fernandez *et al*., 2000; Beldring *et al*., 2003; Engeland *et al*., 2006), parameter averaging (e.g. Kokkonen *et al*., 2003), nearest neighbor (e.g. Merz and Blöschl, 2004; Parajka *et al*., 2005; Oudin *et al*., 2008), physical similarity (e.g. McIntyre *et al*., 2005; Reichl *et al*., 2009) and regression (e.g. Bárdossy, 2007). However, the majority of the works used lumped conceptual models at daily or coarser temporal scales. In addition, regionalization for hourly prediction in boreal catchments is not common in literature. Moreover, the regional modelling offers additional advantages of multi-basin based comprehensive evaluation of predictions. Therefore, the main research questions lie in the identification of the distributed Precipitation-Runoff routines for better simulation of hourly runoff response based on both catchment and regional scale modelling towards a potential improvement in the forecast mode. We classified the work in this thesis in to six main objectives presented in six papers, where the first three papers are based on catchment scale modelling and the last three are at both catchment and regional scales.

Research papers

The lists of papers (P1 to P6), which form this thesis, are as listed below:

P1: Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed *(under review)*.

Teklu T. Hailegeorgis, Knut Alfredsen, Yisak S. Abdella and Sjur Kolberg

P2: Hailegeorgis, T. T. and Alfredsen, K. **Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments**.

Hydrology Research (IWA Publishing), article in press, 2014.

http://www.iwaponline.com/nh/up/default.htm

P3: Distributed hourly runoff computations in mountainous boreal catchments from 'catchments as simple dynamical systems' storage-discharge relationships *(under review).*

Teklu T. Hailegeorgis, Knut Alfredsen, Yisak S. Abdella and Sjur Kolberg

P4: Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments *(under review).*

Teklu T. Hailegeorgis, Yisak S. Abdella, Knut Alfredsen and Sjur Kolberg

P5: Multi-basin and regional calibration based identification of distributed Precipitation-Runoff models for prediction of hourly streamflow on 26 catchments in mid-Norway *(under review)*.

Teklu T. Hailegeorgis and Knut Alfredsen

P6: Regional statistical and Precipitation-Runoff modelling for ecological applications: prediction of hourly streamflow in regulated rivers and ungauged basins *(under review)*.

Teklu T. Hailegeorgis and Knut Alfredsen

Thesis objectives

The general objectives of the research are the development, calibration and evaluation of different distributed precipitation-runoff response routines for prediction of hourly streamflow. The focuses of the work are the identification of parameterizations of spatial heterogeneity, model configurations and structures, and regionalization methods. We also performed comparative evaluations of the regional calibration of the P-R response routines and the regional regression model for the PUB. Six sub-objectives were specified that together answer the main objective. Each of these is analyzed and presented in the six scientific papers P1 to P6. The specific objectives for each paper can be summarized as:

O1: Investigation of the performance of parameterizations of the spatial heterogeneity of subsurface storage capacity across different spatial scales for semi-distributed and distributed hourly simulation of streamflow;

O2: Investigation of the performance of parametrical parsimonious and more complex configurations of a distributed conceptual runoff response routine for hourly simulation of streamflow including parameter identifiability and uncertainty;

O3: Evaluation of the performance of a distributed 'top-down' response routine for hourly runoff simulation for a macroscale catchment, when parameters are both estimated from streamflow recession and calibrated based on observed streamflow;

O4: Comparative evaluation of the performance of different regionalization methods for the PUB based on multi-models, different performance measures and evaluation metrics;

O5: Identification of distributed P-R models for prediction of hourly streamflow relevant to operational purposes based on multi-basin and regional calibration; and

O6: Developing a regional regression model for prediction of natural hourly streamflow in regulated and ungauged rivers for ecological applications, and comparison with the regional calibration of the P-R model.

Thesis organization

The first chapter outlines the context, motivation and objectives of the thesis, current state of research and lists of the papers that forms the Thesis. Chapter 2 provides a brief description of the study region, input data and the general methods and tools used. Chapter 3 presents summary of results from the six papers (P1 to P6). Chapter 4 summarizes the conclusions based on the findings of the research and presents perspectives for future research for the study region. Appendix A contains the six research papers, which form the thesis, and Appendix B contains co-authors and publishers' declarations.

Chapter 2 STUDY REGION, DATA AND RESEARCH METHODOLOGY

Study region and data

The study region consists of 26 unregulated gauged catchments ranging in size from 39 to 3090 km^2 in boreal mid Norway (catchments no. 1 to 26 in Figure 2). We also used the Lundesokna catchment, where a regulated hydropeaking river exists, for the study in paper 6. The dominant land uses/land covers in the study area are mountains above timberline and forests, and the predominant soil or loose material is glacial tills. For the P1, P2 and P3, we implemented catchment scale modelling by utilizing data from 4 streamflow gauging stations inside the Gaula watershed and 12 nearby precipitation-gauging stations (Figure 1). The catchments are catchment nos. 3, 6, 8 and 14 in Figure 2. For the P4, P5 and P6, we implemented regional scale modelling on 26 catchments (Figure 2) using precipitation and temperature data from 44 and 54 hourly stations, wind speed data from 40 stations and global radiation and relative humidity data from 12 stations in the region. We used hourly records from 2008 to 2010 for calibration. We obtained the streamflow, hypsography and land use data from the Norwegian Water Resources and Energy Directorate (NVE), the climate data from the Norwegian Meteorological Institute, Statkraft, TrønderEnergi, Nord Trøndelag Elektrisitetsverk (NTE) and Bioforsk, data on loose material (soil) and bedrock geology from the Norwegian Geological Survey (http://www.ngu.no) and stream networks map from the Norwegian Mapping Authority.

Research methodology

For the P-R modelling, we performed the local and regional calibration using continuous streamflow series. We used the Nash-Sutcliff efficiency (NSE) or R^2 and for a log-transformed series (NSEln) or R^2ln performance measures (Nash and Sutcliffe, 1970) and various runoff signatures for comparative evaluations of the runoff response routines for hourly prediction. We also developed and evaluated a regional regression model for prediction of hourly streamflow in regulated ungauged basin. For all the P-R models, we used the ENKI modular hydrological modelling framework (Kolberg and Bruland, 2012) for model set-up, parameter calibration and runoff simulation. ENKI allows building a model from a library of routines for model calibration and simulation, making it very flexible for analyzing various combinations of routines and for testing new developments. New routines were coded and compiled as dynamic-link libraries (dlls) using the Microsoft Visual Studio and dynamically added to the ENKI framework. The required input static maps such as elevation and land use were prepared in ArcGIS (http://www.esri.com) and converted to Idrisi raster files using Geospatial Data Abstraction Library (GDAL) for use as input in ENKI. The input and output databases in ENKI are based on the NetCDF. ENKI allows import of TAB-delimited text file input time series and export of output time series for instance to MS-excel for further processing of the results. We interpolated the input climate forcing from gauging stations on a 1x1 km^2 grids using the Inverse Distance Weighted (IDW) interpolation routine in ENKI. We used the Matlab software for development and evaluation of the regional regression model (P6).

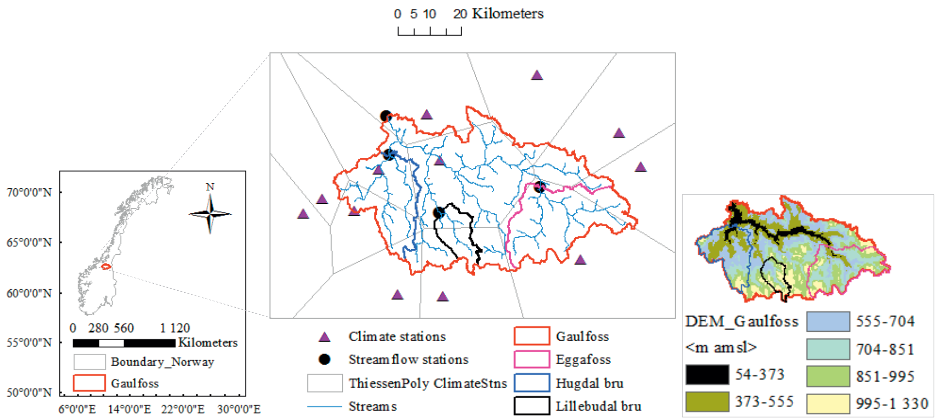


Figure 1. *Location map of the Gaula catchment, precipitation and streamflow gauging stations, and elevation map (DEM) used for the local calibration (P1, P2 and P3).*

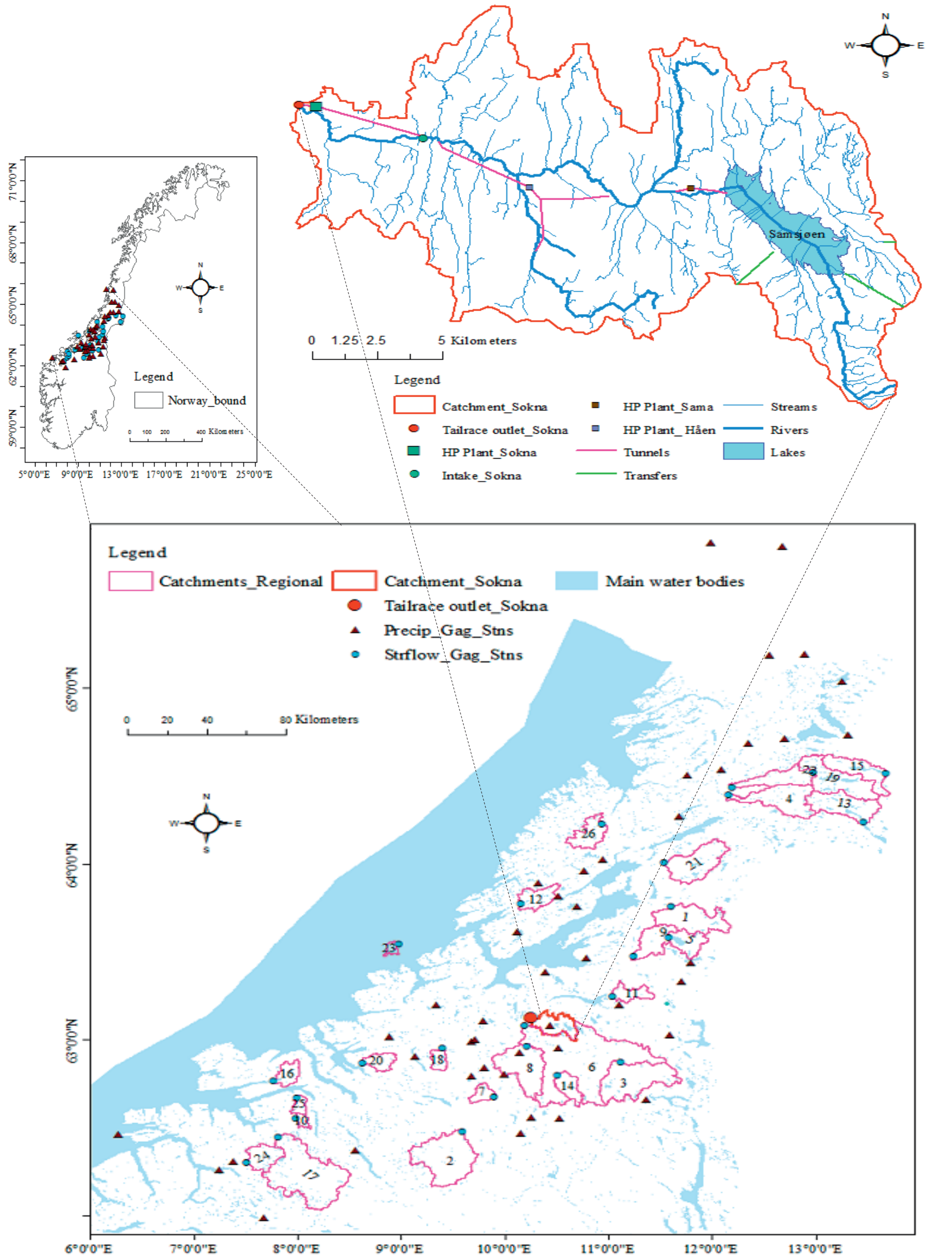


Figure 2. *Location map of the study region and catchments (nos. 1 to 26 and the Lundesokna (sokna) catchment, and precipitation and streamflow gauging stations used for P4, P5 and P6.*

Chapter 3 SUMMARY OF MAIN RESULTS

One of the main challenges in precipitation-runoff modelling is identification of suitable parameterization for instance discretization techniques and scales of representation of spatial heterogeneities either explicitly or by a probability distribution. We investigated the performance of six different cases of parameterizations based on the probability-distributed model (Moore, 1985) for representing the spatial heterogeneity of the subsurface storage capacity for semi-distributed (elements) and gridded (1x1 km^2) simulation of hourly streamflow. Table 1 shows summary of the key features of the six-parameterization cases.

Table 1. *Summary of the key features of the six parameterization cases.*

Cases	Scales of representation of spatial heterogeneity and computation or calibration				
	Heterogeneity by a probability distribution	S_{max}	Shape parameter 'b' of the distribution	Runoff generation (response)	STS runoff routing
1	Subcatchment	Catchment (calibrated)	Catchment (calibrated)	Element	Grid
2	Subelement	Element (from topographic gradient)	Element (from topographic gradient)	Element	Grid
3	-	Catchment (calibrated)	-	Element	Grid
1G	Subgrid	Catchment (calibrated)	Catchment (calibrated)	Grid	Grid
2G	Subgrid	Grid (from topographic gradient)	Grid (from topographic gradient)	Grid	Grid
3G	-	Catchment (calibrated)	-	Grid	Grid

S_{max}: maximum subsurface storage capacity and STS: source-to-sink routing.

We calibrated the routines for the Gaulfoss catchment and transferred the calibrated parameters to interior catchments of Eggafoss, Hugdal bru and Lillebudal bru (Figure 1) for 'proxy basin' based model validation. Calibration of the six-parameterization cases provided satisfactory but indistinguishable simulation of the hourly runoff hydrographs (Figure 3 and Figure 4) and to some extent reproduced the temporal variability of streamflow in terms of the flow duration curves (FDCs). Spatial transfer of the calibrated parameters reproduced hydrographs at the two interior catchments with relatively representative climate data (Eggafoss and Hugdal bru), which indicate the validation of the models.

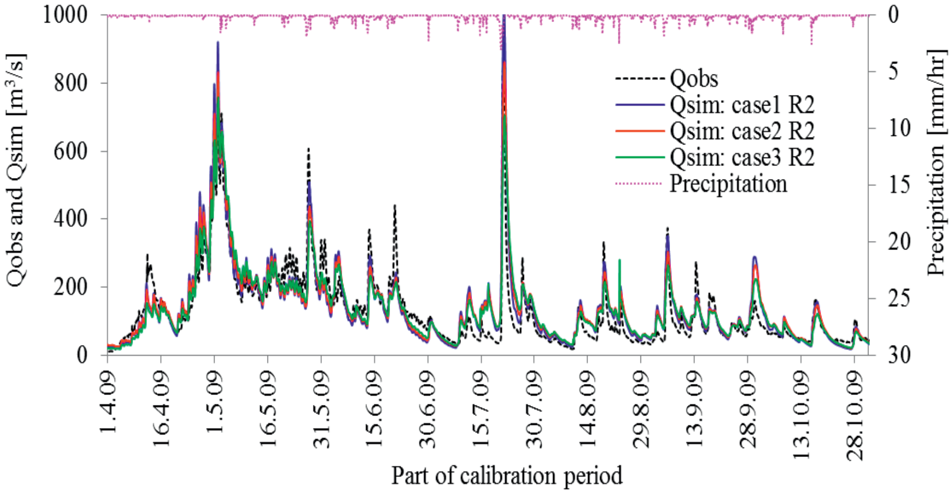


Figure 3. *Hourly observed and simulated streamflow hydrographs for part of the calibration period corresponding to the maximum Nash-Sutcliffe efficiency (R^2) for Gaulfoss for cases 1 to 3.*

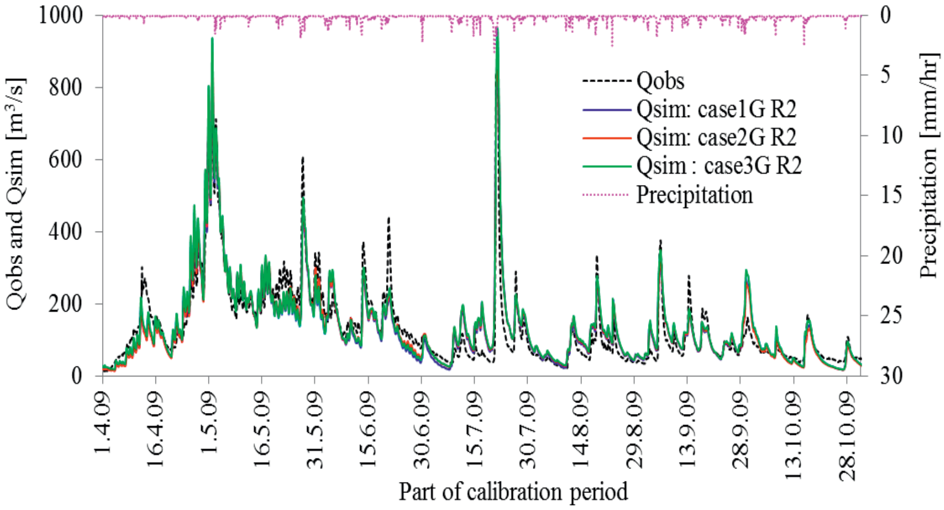
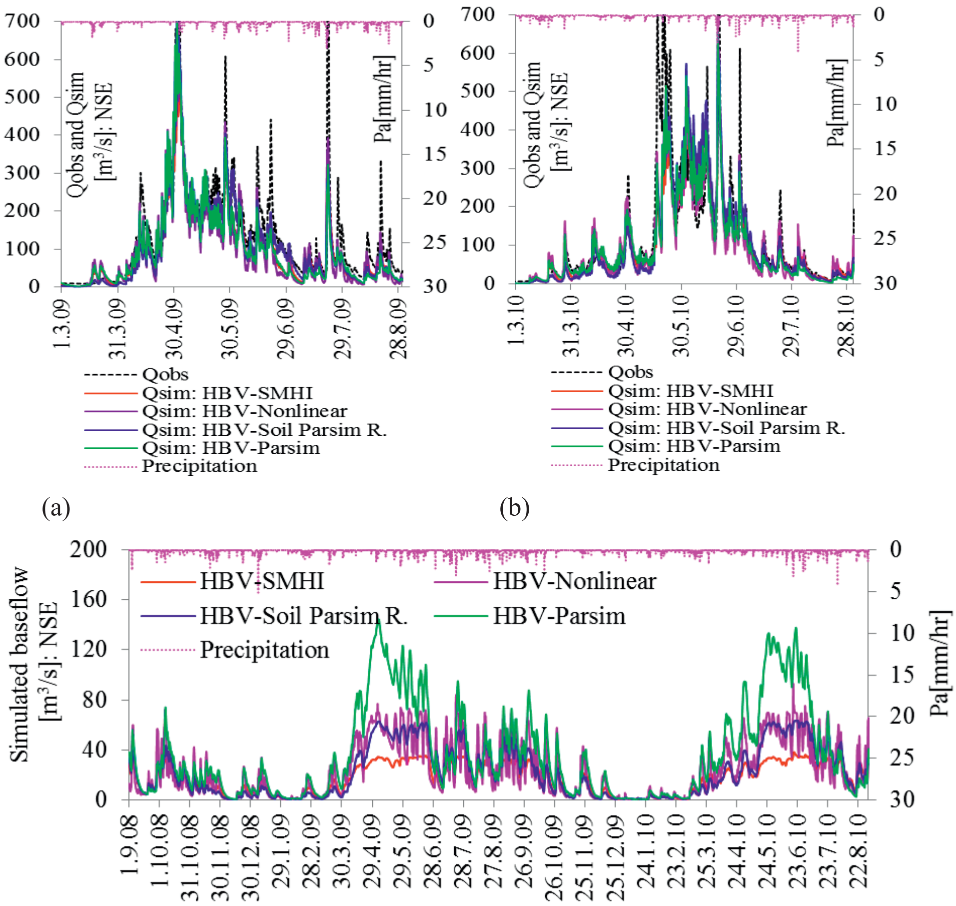


Figure 4. *Hourly observed and simulated streamflow hydrographs for part of the calibration period corresponding to maximum R^2 for Gaulfoss for cases 1G to 3G.*

Unidentifiability and equifinality of parameters related to overparameterization pose challenges to calibration and prediction by precipitation-runoff models. In light of the findings in P1 regarding the equivalent performances by the simple and more complex parameterizations, we wanted to evaluate the simulation performances of parametrical parsimonious versus more complex configurations of the HBV runoff response routines, which is lacking for hourly simulation using the widely used HBV model in the Nordic watersheds. The evaluated model configurations are the distributed (1x1 km^2) standard HBV from the Swedish Meteorological and Hydrological Institute (SMHI) (Bergström, 1976) or HBV-SMHI, a non-linear storage-discharge relationship HBV (HBV-Nonlinear), a standard soil-moisture accounting and linear reservoir, and parsimonious runoff response routine (HBV-Soil Parsim R) and a parsimonious and linear configurations of both the soil-moisture accounting and the runoff response routines (HBV-Parsim). We calibrated the routines using four streamflow gauging stations in the Gaula watershed (Figure 1) and conducted spatial, temporal and spatio-temporal validation of the routines through parameter transfer among the catchments.

The different configurations provided nearly indistinguishable simulation of streamflow hydrographs (Figure 5a&b). However, the HBV-Parsim routine provided reliable baseflow simulation (Figure 5c) complying with the high baseflow contribution to the streamflow for the study catchments as demonstrated by analytical baseflow separation techniques in this study and in previous studies (e.g. Beldring *et al*., 2000). This complies with the principle of parsimony (Jakeman, 2000) and the simplest model is more likely to be correct among models of equal performance (Forster, 2000). The reliable baseflow simulation by the HBV-Parsim was obtained due to the maximum soil moisture storage capacity or the field capacity (FC) assigned based on the land use classes. In the other configurations, we calibrated the FC as a free parameter, which resulted in considerable negative correlation between the FC and the percolation rate and hence a decrease in the baseflow. In addition, the compensation between the outflow from the upper reservoir and the baseflow from the lower reservoir resulted in equivalent prediction of the total streamflow, which masked the lack of realism in the internal simulation of the baseflow.



(c)

Figure 5. *(a) and (b) streamflow hydrographs for part of calibration period, and (c) baseflow hydrographs for the calibration period for Gaulfoss.*

The models and equations used for the P1 and P2 were mainly conceptual types. In addition, the basis of some of the physical equations used in P1 is scaling-up of the theories of small-scale processes or the 'bottom-up' modelling paradigm. The non-identifiability of parameterizations and model configurations, and problems of compensation between the runoff fluxes substantiates the need for parsimonious model structure and alternative modelling paradigm. In P3, we wanted to evaluate a parsimonious distributed runoff response routine based on a 'top-down' modelling paradigm. We evaluated the performance of a response routine based on functional relationships between catchment storage and discharge. The routine is named Kirchmod in the further discussion. We estimated the parameters from streamflow recession analysis following the method outlined by Kirchner (2009) and as an alternative we tested the case where the response routine parameters of Kirchmod are calibrated. We estimated and calibrated the parameters for a macroscale (3090 km^2) Gaulfoss catchment and validated through spatial transfer to interior catchments of Eggafoss, Hugdal bru and Lillebudal bru (Figure 1).

The routine provided simulated streamflow hydrographs and flow duration curves that are in good agreement with the observed (Figure 6 and Figure 7) while simulation from parameters estimated from recession analysis resulted in slight underestimation of the peak flows (Figure 7). In addition, there are various sources of uncertainties in parameter estimation from the recession analysis, which requires an assessment of parameter uncertainty and identifiability to evaluate the reliability of inferences and predictions. However, the results obtained from the study generally encouraged further evaluation of the 'top-down' routine for operational purpose based on larger number of catchments or the regional scale modelling.

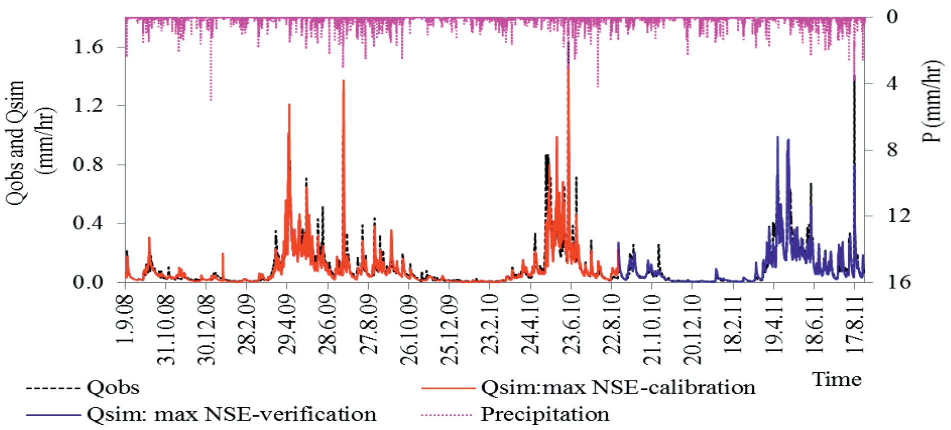


Figure 6. *Hydrographs for catchment no. 6 (Gaulfoss) from calibration corresponding to max NSE (response routine parameters calibrated).*

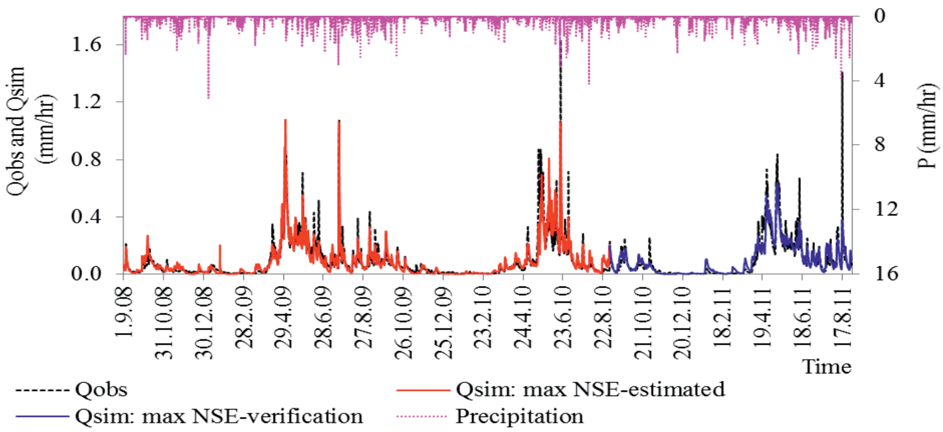


***Figure** 7. Hydrographs for catchment 6 (Gaulfoss) from calibration corresponding to max NSE (response routine parameters estimated from streamflow recession).*

In the P1, P2 and P3, we performed the parameter calibration based on catchment scale modelling and validation of the models based on parameter transfer to interior or neighboring catchments following the 'proxy-basin' test. However, regional scale modelling or regionalization studies for prediction in ungauged basins with an hourly resolution are important but are not common for the boreal catchments. Therefore, in P4 we evaluated a regional scale modelling approach based on calibration of 26 unregulated catchments in boreal mid-Norway with size ranging from 39 to 3090 km^2 using the Kirchmod, the HBV and the Basic Grid Model or BGM (Bell and Moore, 1998) based model structures. We evaluated four regionalization methods, which include the regional calibration in terms of the parameter vectors that yield the maximum regional weighted average (MRWA) performance measures, regional median of optimal parameters (RMedP), nearest neighbor (NN) and physical similarity.

For the study region and the set of hydro-climatological data, the best performing regionalization method tends to vary among the model structures, performance measures and evaluation metrics for the PM. Based on the regional median and mean of the raw values of the Nash-Sutcliffe efficiencies R^2 and R^2ln, and the regional median and mean of the losses or gains of the PM between the local calibration and the regionalization, the MRWA regionalization method performs slightly better than the other methods. However, more comprehensive comparisons based on the cumulative distribution functions (CDF) of the losses or gains in the performance measures indicated that the physical similarity in the combined physical attributes spaces (PSCOMB) performed slightly better for the R^2 (Figure 8) while the physical similarity in soil types (PSSOIL) performed slightly better for the R^2ln (Figure 9). Therefore, for the study region the physical similarity and the regional calibration based regionalization methods are more suitable than the NN and RMedP.

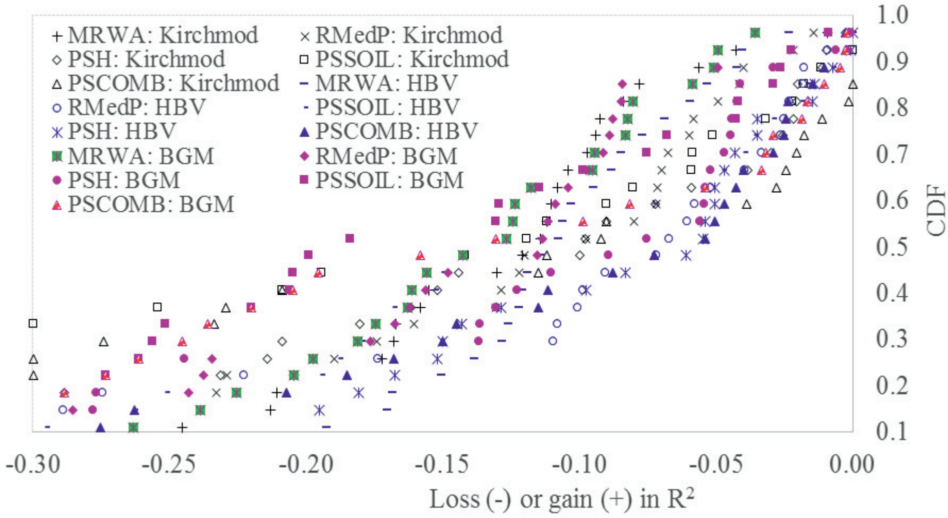


Figure 8. *Cumulative distribution functions (CDF) of the losses or gains in the performance measures from the local calibration due to the regionalization for R^2. Some portions of the larger losses in the PM are not displayed for clarity of the figures.*

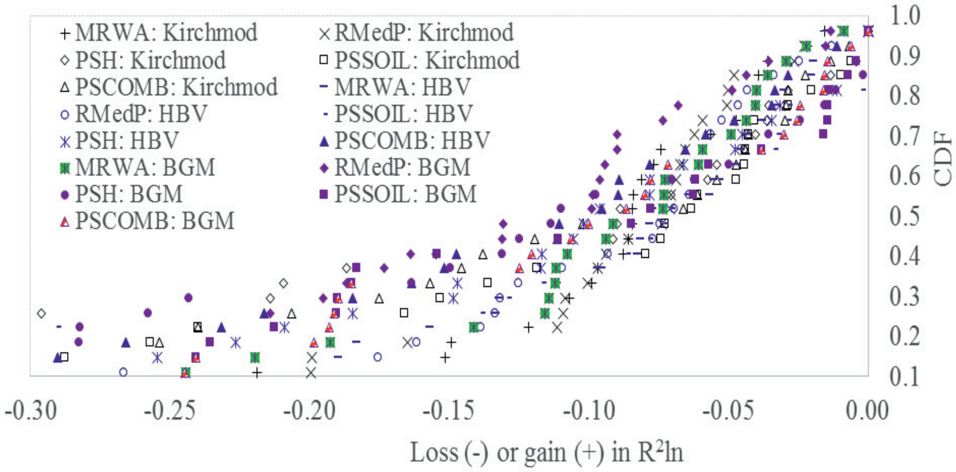


Figure 9. *Cumulative distribution functions (CDF) of the losses or gains in the performance measures from the local calibration due to the regionalization for R^2ln. Some portions of the larger losses in the PM are not displayed for clarity of the figures.*

The study in P4 was a multi-model based comparative evaluation of regionalization methods for prediction in ungauged basins. There are additional merits of multi-basin based local and regional calibrations in terms of identification of suitable P-R response routines for operational uses. However, such study is lacking for hourly prediction of streamflow in the boreal catchments. Therefore, in P5 we investigated the performances of five distributed P-R response routines on the 26 catchments in the region. The routines include the Kirchmod, the BGM and three variants of the HBV response routine (HBV1, HBV2 and HBV3). The evaluation and identification criteria include (a) parameter uncertainty and identifiability; (b) the Nash-Sutcliffe efficiencies NSE and NSEln from the local and regional calibration (MRWA) and (c) 'realism of simulation' for the local calibration based on different runoff signatures for some selected catchments.

We obtained narrow posterior parameter distributions for the Kirchmod compared to the other routines indicating that small number of free parameters provides a narrow range of parameter uncertainty. Considerable correlation among some pairs of calibrated parameters for all of the routines indicate problem of parameter interactions, which is the potential cause of non-identifiability of the performances of the runoff response routines. Based on the performance measures of the local and regional calibration, a good performing model structure vary among the catchments and the PM. Therefore, there is no single best performing model structure for the whole region due to uniqueness of catchments and the optimal parameter set cannot provide good calibration performance across full ranges of streamflow. Based on the box plots of the PM for both the local and regional calibration, the Kirchmod and BGM routines provide higher NSE values between the 25^{th} and 75^{th} percentiles (Figure 10). The Kirchmod, HBV2 and HBV3 provided slightly higher NSEln values between the 25^{th} and 75^{th} percentiles (Figure 11). However, the regional performance in terms of the regional mean and regional median PM of the Kirchmod and the BGM are higher than the HBV based routines, and hence they are more suitable for the regional calibration (MRWA) based PUB in the region. Simulation of multiple runoff signatures indicate that different model structures appeared to be relatively better in reproducing runoff ratio for different catchments, the Kirchmod and BGM routines reproduced relatively better FDCs than the other routines and there is systematic under prediction of high flows by all the routines.

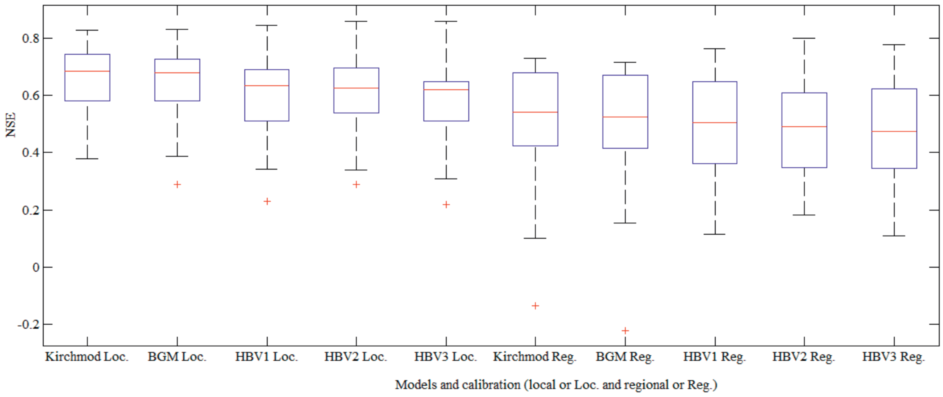


Figure 10. *Box plots of the NSE performance measures (PM) for the local and regional calibration.*

On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as "+" marks.

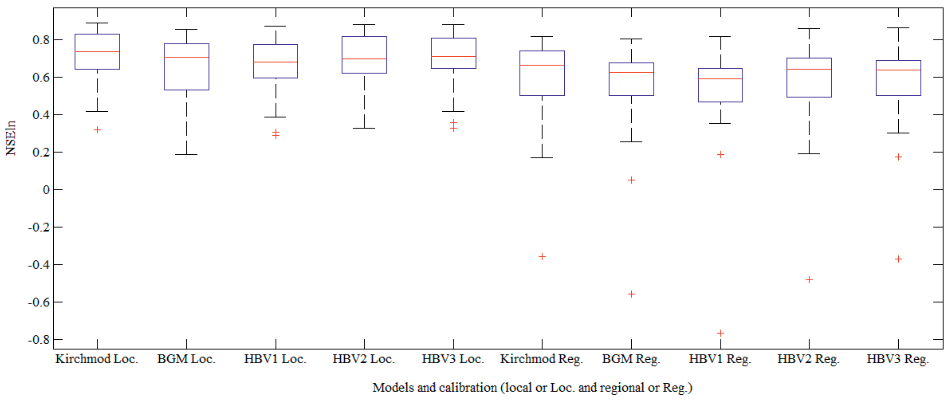


Figure 11. *Box plots of the NSEln performance measures (PM) for the local and regional calibration.*

The regional studies conducted in P4 and P5 are based on simulation of the dominant runoff response mechanisms using the P-R routines that involve the input climate forcing, model structure and procedures for model calibration. Further endeavors towards more improved prediction are indispensable to reduce the uncertainty in prediction and associated risks for instance in assessments of the impacts of regulation on ecological integrity. Despite the limitations of the black-box models, the regional regression model allows data mining based predictions and inferences without employing the simulation of the real-world hydrological processes. However, the use of such models for the PUB for hourly resolution is not common. In P6, we developed the regional regression model using the 26 catchments for prediction of flow duration curves (FDCs) and streamflow time series at regulated (ungauged) basin. We then compared the transfer of hourly observed streamflow information using the regional regression model together with the nearest neighbor (NN) regionalization method with the Kirchmod response routine together with MRWA regionalization method, which are selected based on the performances in P4 and P5.

The simple regional regression model outperformed the P-R response routine (Table 2) for instance the regression based prediction by transferring information from catchment no. 6 (Gaulfoss) to its interior catchment of catchment no. 3 (Eggafoss) indicated the NSE value of 0.89 versus the local P-R calibration (NSE = 0.81) and the MRWA (NSE = 0.68). The results for other catchments further illustrate the usefulness of the improved prediction by the regression model when applied to a regulated hydropeaking catchment inside the study region. The natural streamflow hydrographs and FDCs predicted by the regression model shows smoothly varying hydrographs at downstream of the tailrace of Lundesokna hydropeaking plant while the observed (regulated) flow exhibits continuous sudden fluctuations of streamflow magnitudes (Figure 12 and Figure 13). The within a year FDC for observed (regulated) flow under hydropeaking operation exhibits sharp transitions from high to medium flows and from medium to low flows (Figure 12). Alteration in the FDC and hydrographs due to hydropeaking operation indicate alterations in several streamflow characteristics, which affects the ecological integrity in regulated rivers. Therefore, improved prediction of natural time series of streamflow at regulated river is useful to derive ecologically relevant streamflow metrics for assessments of environmental flows and ecological responses. In addition, the model allows prediction at any ungauged or regulated catchments for assessment of contribution to environmental flow or reservoir inflow. The concept of an inflow controlled environmental flow regime (e.g. Alfredsen *et al.*, 2012) would also benefit from a more reliable prediction of continuous natural flow in regulated rivers, where environmental legislations on riverine systems are stringent.

Table 2. *NSE for regional transfer of information based on regional regression (2006-2011) and P-R model calibration (2008-2010) for prediction of hourly streamflow.*

Donor catchments (Catchment No.)	Recipient catchments (Catchment No.)										
Regression(reg.)	1	3	6	10	12	14	16	17	20	21	26
1	**0.95**	0.31	0.39	-1.10	0.50	-0.07	-1.21	-0.44	0.42	0.70	0.49
3	0.35	**0.96**	0.88	-0.40	-0.13	0.55	-2.29	0.48	0.33	0.02	-0.22
6	0.42	0.89	**0.95**	-0.39	-0.07	0.54	-2.02	0.42	0.38	0.12	-0.16
10	0.04	0.39	0.42	**0.43**	-0.21	0.23	-2.49	0.61	0.44	-0.21	-0.36
12	0.53	-0.35	-0.26	-1.96	**0.80**	-0.77	-0.68	-1.08	0.15	0.67	0.80
14	0.21	0.80	0.79	-0.30	-0.19	**0.70**	-2.26	0.47	0.34	-0.09	-0.30
16	-0.06	-0.69	-0.54	-2.29	0.26	-0.98	**0.45**	-1.22	-0.10	0.13	0.25
17	-0.17	0.49	0.46	-0.06	-0.37	0.22	-2.72	**0.97**	0.12	-0.39	-0.51
20	0.50	0.34	0.42	-0.44	0.22	0.11	-1.48	-0.14	**0.83**	0.24	0.10
21	0.72	-0.07	0.04	-1.68	0.61	-0.50	-0.98	-0.83	0.20	**0.90**	0.66
26	0.48	-0.37	-0.30	-2.07	0.73	-0.82	-0.71	-1.12	0.06	0.67	**0.88**
P-R model											
1	**0.74**	0.79	0.82	0.48	0.63	0.53	0.03	0.44	0.56	0.71	0.66
3	0.73	**0.81**	0.83	0.51	0.71	0.53	0.14	0.23	0.53	0.72	0.71
6	0.70	0.79	**0.83**	0.54	0.49	0.56	-0.59	0.15	0.65	0.67	0.54
10	-0.14	0.10	0.18	**0.58**	0.19	0.23	-0.23	0.16	0.31	0.24	0.06
12	0.68	0.78	0.70	0.50	**0.75**	0.52	0.12	0.41	0.51	0.70	0.71
14	0.62	0.66	0.74	0.47	0.27	**0.58**	-0.34	0.18	0.62	0.57	0.47
16	0.44	0.48	0.50	0.40	0.50	0.38	**0.67**	0.66	0.28	0.46	0.42
17	0.29	0.31	0.44	0.35	0.29	0.28	0.56	**0.77**	0.20	0.30	0.21
20	0.58	0.69	0.75	0.56	0.23	0.56	-1.16	-0.11	**0.67**	0.55	0.32
21	0.69	0.73	0.74	0.56	0.69	0.54	0.17	0.35	0.53	**0.71**	0.63
26	0.64	0.71	0.61	0.45	0.71	0.48	0.30	0.56	0.44	0.65	**0.72**
MRWA NSE	0.65	0.68	0.72	0.42	0.69	0.47	0.55	0.68	0.43	0.63	0.64

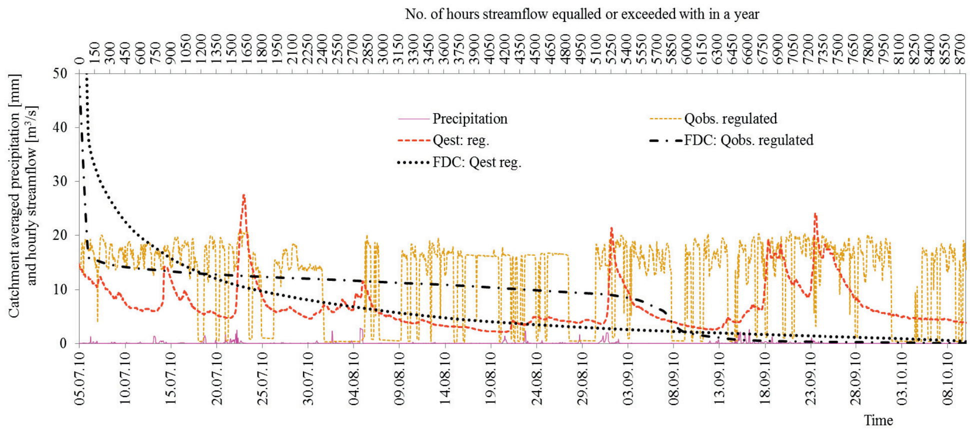


Figure 12. *Typical comparisons of summer and fall observed/regulated streamflow versus the predicted 'unimpaired' or natural streamflow for Lundesokna river (transferred from Gaulfoss by the regional regression).*
FDC: Qobs (observed streamflow 2008-2011) and FDC: Qpred (predicted 2006-2011).

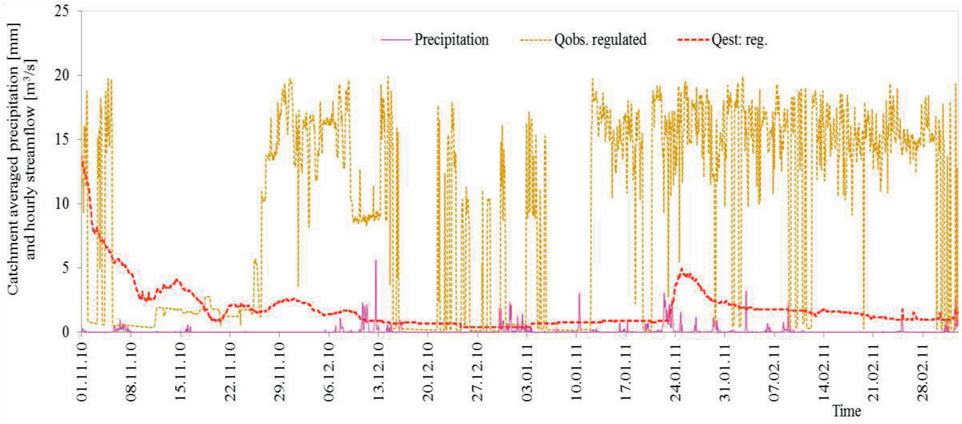


Figure 13. *Typical comparisons of winter observed/regulated streamflow versus the predicted 'unimpaired' or natural streamflow for Lundesokna river (transferred from Gaulfoss by the regional regression).*

Chapter 4 CONCLUSIONS AND PERSPECTIVES FOR FUTURE RESEARCH

Conclusions

The main objective of the thesis is development, calibration and comparative evaluation of distributed Precipitation-Runoff response routines for hourly simulation. The work includes the identification of different parameterizations, model structures, modelling paradigms and regionalization methods for hourly continuous streamflow simulation based on both catchment and regional scale modelling including the PUB. Development of a regional regression model and the comparisons of this with the P-R model was also an additional objective of the Thesis. We conducted the studies in P1, P2 and P3 based on catchment scale local calibration and the 'proxy basin' approach for model validation while the focus in P4, P5 and P6 was regional calibration for identification of the regionalization methods and the P-R response routines. The main conclusions drawn from the research findings are as follows:

The study in P1 indicated that identification of parameterizations of spatial heterogeneities based on model calibration only to precipitation and streamflow relationships is challenging and identification requires measurements from high density precipitation networks than that is required for a mere parameter calibration. However, the equivalent simulation performances of the simple semi-distributed and simple distributed (BGM) parameterizations to the subelement or subgrid scale parameterizations show the potential suitability of the parametrical parsimonious routine for operational prediction in the boreal catchments.

The study in P2 based on different configurations of the HBV response routine indicated that overparameterization is the main explanation for the non-identifiability of parameterizations. Parsimonious configuration provides equivalent streamflow simulation with improved parameter identifiability. In addition, parametrical parsimony allows for reliable internal simulation by minimizing the correlations between the soil moisture and runoff response parameters and compensation between the outflow from the upper reservoir and the baseflow from the lower reservoir. Therefore, predictions by the conceptual HBV runoff response routine variants require evaluation of the reliability of internal simulation of the variables other than that used for model calibration, for instance the baseflow against analytical estimation (e.g. tailor-made baseflow separation techniques).

The distributed version of a parsimonious response routine (Kirchmod) based on the ‘top-down’ modelling paradigm and functional relationships between catchment storage and discharge (P3) reproduces the streamflow signatures for a macroscale (3090 km^2) boreal mountainous catchment, which exhibit a considerable runoff delay compared to the hourly simulation. In addition, validation of parameter transfer to the interior catchments encourages regional scale evaluations of the routine. However, the correlation among the free parameters or the parameter non-identifiability is the main causes of the insensitivity of the calibration to the runoff routing. Moreover, we observed different sources of uncertainties and non-identifiability of the response routine parameters estimated from the streamflow recession. Therefore, the results indicate the need for assessments of uncertainty and non-identifiability even for parametrical parsimonious models.

The multi-donor regional calibration (MRWA) performs nearly equivalent to the single-donor regionalization method based on similarity of catchments in their physical attributes for transfer of parameters of the P-R response routines for the mid-Norway region (P4). In addition, both methods perform better than the nearest neighbor regionalization method. Comprehensive selection of performance measures and their evaluation metrics for a specific modelling objective (e.g. high flow, low flow and water balance) are required for reliable identification of suitable regionalization methods. Further regionalization attempts for the boreal region should focus on improving the density of both climate and streamflow gauging networks to obtain a more representative hourly input and to allow regional study on a confined region to reduce the heterogeneities among the pooled catchments. In addition, measurements and use of attributes related to soil hydraulic properties, which mainly influences the runoff response in the region, are important for further assessment of the physical similarity regionalization.

Multi-basin local and regional calibration based attempts for comprehensive identification of reliable P-R routines indicates that there is no single best performing model structure solution for the mid-Norway region due to ‘uniqueness’ of catchments and change in the optimal parameter vectors with the changes in the PM that are selected to suit specific calibration objectives (P5). In addition, narrow posterior parameter distributions (less parameter uncertainty), which are obtained for parametrical parsimonious routines, does not guarantee less parameter correlations and identifiability. Therefore, combined flexible models and multi-basin regional scale modelling framework provides an opportunity for comprehensive identification of reliable model structure, parameterizations and modelling paradigms among a pool of competing plausible options for improved prediction than the contemporary operational prediction using a fixed model at a catchment scale. Systematic under prediction of high flows by all the routines indicates unrepresentativeness of the precipitation input leading to less accurate estimation of spatially interpolated precipitation on the 1x1 km^2 grids, which requires upgrading of the hourly gauging networks in the region.

For the boreal study region, a data based simple regional regression model with the nearest neighbor method regional transfer of streamflow information performs better than parameter transfer by the regional calibration of the P-R response routine for prediction of hourly flow duration curves and streamflow (P6). The improved prediction for the hourly natural flow at regulated rivers is very useful for estimation of ecologically relevant streamflow metrics for studies related to ecological impacts of regulation and environmental flow assessment. Hydrological alterations in terms of the hourly hydrographs and flow duration curves due to the regulation (hydropeaking operation) compared to the predicted natural streamflow indicate potential alterations in several streamflow characteristics, which require further scrutiny relevant to the ecological integrity in regulated rivers.

Perspectives for future research

The thesis mainly focused on evaluating the performances of spatially distributed (1x1 km^2 grids) P-R response routines in the simulation mode while the study also demonstrated the potential merits of a data based regional regression model for prediction in regulated and ungauged basins. The main premise is that for an improved performance and reliability in the simulation mode, there is a likelihood of improved performance in the forecast mode. Even though the spatially distributed models allow better representation of the spatial variability of precipitation, availability of adequate precipitation data in terms of the density of the gauging networks and quality is mandatory for reliable calibration and identification of the P-R response routines. The low-density of available gauging networks and hence low resolution of the hourly climate forcing field compared to the extent of the modelled catchments or region is the major limitation of this thesis. The sparse streamflow gauging networks on the unregulated catchments for the regional scale modelling using the P-R response routines has the drawbacks of pooling catchments that are heterogeneous in their runoff response and affects the performance of parameter transfer based on the regionalization methods. In addition, calibration based on a single objective function and model evaluation based on some performance measure cannot yield a unique versatile optimal parameter vector to reproduce the whole portions of the hydrograph and other runoff signatures relevant for operational purposes. Therefore, future perspectives for improved prediction by the P-R response routines in the region should include:

- The limitations due to the lack of sufficient hourly measurements need to be addressed through upgrading of the existing climate and streamflow gauging networks to high-density networks of sufficient and quality-controlled hydro-climatic records for research purposes and data dissemination through coordinated involvement of the various stakeholders;
- Additional observations for instance soil moisture and ground water level in selected catchments, which help for multi-criteria calibration and validation of the response routines and the physics-based modelling;
- Study that clearly illustrate the merits of transition from lumped to spatially distributed models, and from daily to hourly and sub-hourly temporal resolution for operational forecasting in terms of simulation of the hydrographs (e.g. low flow, medium flow, peak flow and time to peak) and other runoff signatures;
- More comprehensive studies on flexible multi-model, multi-basin and multi-objective parameter calibration strategy for identification of models among a pool of plausible competing options;
- Study on methods how to handle the varying optimal (calibrated) parameter sets ,which correspond to the different objective functions or performance measures, in the forecast mode; and

- Study on identification of models based on their performance in the forecast mode.

References

Alfredsen K., Harby A., Linnansaari T. and Ugedal O. (2012). Development of an inflow controlled environmental flow regime for a Norwegian river. *River Res. Applic.* 28 (6), 731-739.

Bartholmes, J. and Todini, E. (2005). Coupling meteorological and hydrological models for flood forecasting. *Hydrol. Earth. Syst. Sc.*, 9(4), 333-346.

Bastola, S. & Murphy, C. (2013). Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. *Hydrology Research* 44 (3), 484-494.

Beldring, S., Gottschalk, L., Rodhe, A. & Tallaksen, L.M. (2000) Kinematic wave approximations to hillslope hydrological processes in tills. *Hydrol. Process.* 14, 727-745.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., and Vøkso, A. (2003). Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrol. Earth. Syst. Sc.*, 7(3), 304-316.

Bell, V.A., Moore, R.J. (1998). A grid-based distributed flood forecasting model for use with weather radar data: Part 1. Formulation. *Hydrol. Earth. Syst. Sc.* 2, (2-3), 265-281.

Bergström, S. (1976). Development and application of a conceptual runoff model for Scandinavian catchments, *SMHI Reports RHO*, No. 7, Norrköping.

Blöschl, G., Reszler, C. & Komma, J. (2008). A spatially distributed flash flood forecasting model. *Environmental Modelling & Software* 23, 464-478.

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrol. Earth Syst. Sci.*, 11, 703-710.

Engeland, K., Braud, I., Gottschalk, L., and Leblois, E. (2006). Multi-objective regional modelling. *J. Hydrol.*, 327, 339-351

Engeland, K. and Steinsland, I. (2013). Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water Resour. Res.* 50 (1), 182-197.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A. (2000). Regional calibration of a watershed model. *Hydrolog. Sci. J.*, 45, 689-707.

Foster, M. R. (2000). Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology* 44, 205-231.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)-a review. *Hydrolog. Sci. J.*, 58(6), 1198-1255, DOI:10.1080/02626667.2013.803183.

Huttunen, M.; Vehvilainen, B. (2001). The Finnish Watershed Simulation and Forecasting System (WSFS). *American Geophysical Union, Fall Meeting 2001, abstract* #H22I-09.

Jakeman, A. J. & Hornberger, G. M. (1993). How Much Complexity Is Warranted in a Rainfall-Runoff Model?. *Water Resour. Res.* 29 (8), 2637-2649.

Kavetski, D., Fenicia, F. & Clark, M. P. (2011). Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resour. Res.,* 47, W05501.

Kirchner, J.W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* 42, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res.* 45, W02429, doi:10.1029/2008WR006912.

Kolberg, S. A. & Bruland, O. (2012). ENKI-An Open Source environmental modelling platform. *Geophysical Research Abstracts* 14, EGU2012-13630, EGU General Assembly.

Kobold, M. & Brilly, M. (2006). The use of HBV model for flash flood forecasting. *Nat. Hazards Earth Syst. Sci*. 6, 407-417.

Kokkonen, T., Jakeman, A. J., Young, P. C., and Koivusalo, H. J. (2003). Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrol. Process.*, 17(11), 2219-2238.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res.*, 41, W12434, doi: 10.1029/2005WR004289.

Merz, R., and Bloschl, G. (2004). Regionalization of catchment model parameters. *J. Hydrol.*, 287, 95-123.

Moore, R.J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrolog. Sci. J.* 30 (2), 273-297.

Moore, R. J., Bell, V. A., Jones, D. A. (2005). Forecasting for flood warning. *C. R. Geoscience,* 337, 203-217.

Moore, R.J. (2007). The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.*11 (1), 483-499.

Nash, J. E., and Sutcliffe, J. V. (1970). flow forecasting through conceptual models, I. A discussion of principles. *J. Hydrol.* 10, 228-290.

Olsson, J. and Lindström, G. (2008). Evaluation and calibration of operational hydrological ensemble forecasts in Sweden. *J. Hydrol.*, 350, 14-24.

Oudin, L., Andre´assian, V., Perrin, C., Michel, C., and Le, M., N. (2008). Spatial proximity, physical similarity, regression and un-gaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resour. Res.*, 44, W03413, Doi: 10.1029/2007WR006240.

Parajka, J., Merz, R., and Blöschl, G. (2005). A comparison of regionalization methods for catchment model parameters. *Hydrol. Earth Syst. Sci.*, 9, 157-171.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. -J. and DMIP participants (2004). Overall distributed model intercomparison project results. *J. Hydrol.*, 298, 27-60.

Refsgaard, J.C. (1997). Validation and Intercomparison of different updating procedures for real-time forecasting. *Nord. Hydrol.* 28, 65-84.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S. (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resour. Res.*, 45(10), W10423.

Renner, M., Werner, M. G. F., Rademacher, S., and Sprokkereef, E. (2009). Verification of ensemble flow forecast for the River Rhine. *J. Hydrol.*, 376, 463-475.

Smith, M.B., Seo, D. -J., Koren, V. I., Reed, S., Zhang, Z., Duan, Q.-Y., Moreda, F., Cong, S. (2004). The distributed model intercomparison project (DMIP): motivation and experiment design. *J. Hydrol.* 298, 1-4, 4-26.

Smith, M.B., Koren, V., Zhang, Z., Zhang, Y., Reed, S.M., Cui, Zhengtao, Moreda, F., Cosgrove, B.A., Mizukami, N., Anderson, E.A., and DMIP 2 participants (2012). Results of the DMIP 2 Oklahoma experiments. *J. Hydrol.*, 418-419, 17-48.

Velàzquez, J. A., Anctil, F., Ramos, M. H. and Perrin, C. (2011). Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Adv. Geosci.*, 29, 33-42.

Wagener, T. & McIntyre N. (2005). Identification of rainfall–runoff models for operational applications. *Hydrolog. Sci. J.*, 50(5), 735-751.

Şorman, A. A., Şensoy, A., Tekeli, A. E., Şorman A. Ü. and Akyürek, Z. (2009). Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey. *Hydrol. Process.*, 23 (7), 1031-1040.

Appendices

Appendices

Appendix A

Appendix A: Research papers

Paper I

Paper 1 (P1)

Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed

Teklu T. Hailegeorgis, Knut Alfredsen, Yisak S. Abdella and Sjur Kolberg

Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed

Teklu T. Hailegeorgis*[1], Knut Alfredsen[1], Yisak S. Abdella[2] and Sjur Kolberg[2]

[1]Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim.

[2]SINTEF Energi AS, Sem Sælands vei 11, NO-7465 Trondheim.

*Corresponding author (e-mail: tekhi09@gmail.com)

Abstract

Identification of proper parameterizations of spatial heterogeneity is required for precipitation-runoff models. However, relevant studies with a specific aim at hourly runoff simulation in boreal mountainous catchments are not common.

We conducted calibration and evaluation of hourly runoff simulation in a boreal mountainous watershed based on six different parameterizations of the spatial heterogeneity of subsurface storage capacity for a semi-distributed (subcatchments hereafter called elements) and distributed (1x1 km^2 grid) setup. We evaluated representations of element-to-element, grid-to-grid, and probabilistic subcatchment or subbasin, subelement and subgrid heterogeneities.

The parameterization cases satisfactorily reproduced the streamflow hydrographs with the Nash-Sutcliffe efficiency (R^2) values up to 0.84/0.86 and for a log-transformed streamflow (R^2ln) up to 0.85/0.90 for the calibration and validation periods respectively. However, the more stringent test for predictive reliability in terms of quantile-quantile (Q-Q) plots indicated marked over and under predictions and the parameterizations slightly reproduced the flow duration curves. The simple and parametrical parsimonious parameterizations with no subelement or no subgrid heterogeneities provide equivalent simulation performance compared to the more complex cases. The results indicate that (i) identification of parameterizations require measurements from dense precipitation networks than a mere calibration of the precipitation-streamflow relationships, (ii) challenges in identification of parameterizations based on calibration to only the catchment integrated streamflow observations (iii) potential preference for the simple and parsimonious parameterizations for operational forecast due to their equivalent simulation performance to the more complex subgrid scale parameterizations and (iv) the

effects of non-identifiability of parameters (correlations and equifinality) needs assessment.

Keywords:

Parameterization; Spatial heterogeneity; Subsurface storage capacity; Semi-distributed and distributed; Calibration and evaluation; boreal mountainous watershed.

Introduction

Heterogeneities across spatial scales require either explicit resolving or proper parameterization procedures, which are prevailing challenges in catchment scale precipitation-runoff modelling. Previous studies such as Myrabø (1986; 1997), Gottschalk et al. (2001), Singh et al. (2002), Smith et al. (2004) and Bogaard et al. (2005) showed growing opportunities for distributed precipitation-runoff modelling, which allow for better representation of the spatial heterogeneity in climate forcing, catchment characteristics, runoff responses and state variables. Advances in measurement techniques of input variables such as precipitation from weather radar and remotely sensed snow accumulation data can resolve some of the spatial heterogeneities. There are also several efforts to improve model calibration algorithms for parameter identification in distributed hydrological models. However, a thorough diagnostic evaluation of the behavior of the prediction models is indispensable since the quality of real-time forecast is dependent on the process simulation (e.g. Bell and Moore, 1998; Refsgaard, 1997).

One of the main challenges related to predictions based on distributed precipitation-runoff models is the sensitivity of the results to the degree of the spatial resolution of inputs and the computational units used to address the spatial heterogeneity. The heterogeneities to model may include those of model parameters, climate forcing, land surface characteristics, storage capacity of soils and runoff delay (travel lag). Various discretization techniques in precipitation-runoff modelling for the representation of the spatial heterogeneities are available in literature. Selection of discretization of catchments into a number of units based on the various catchment characteristics that govern the hydrological processes namely hydrological response units (HRUs) (Leavesley and Stannard, 1990), topographic wetness index (Beven and Kirby, 1979), topographic drainage divide based subcatchments (Sivapalan and Viney, 1994) hereafter called elements, hillslopes (Goodrich, 1990) and grid squares (Abott et al., 1986) are useful depending on the objectives of the study and the availability of data. Parameterizations of internal heterogeneities within the catchments or within units (e.g. elements, hillslopes, HRUs or grids) by probability distribution functions (e.g. Moore, 1985) are common. Aggregation of inputs and state variables for instance based on simple averaging are also common in catchment modelling (see a review by Blöschl and Sivapalan, 1995).

There are challenges related to parameterizations and scales for the boreal mountainous region. Halldin et al. (1999) noted for northern (boreal) catchments with distinct topographic features that small-scale phenomena influence the exchange processes between the land surface and the atmosphere and the lateral distribution of water through subsurface and surface flows. The spatial observation scale of the input climate forcing is usually coarse (low resolution) from sparse hydro-meteorological networks compared to a fine (high) resolution discretization that may be required to represent the underlying heterogeneity explicitly or probabilistically. In addition, there are scale mismatches between the spatial heterogeneities of climate forcing and topographic controls (e.g. the fine scale topographic driven spatial heterogeneity is dominating the grid-to-grid variability of the low intensity precipitation).

Therefore, for a reliable prediction augmented by sensitivity analysis and hence insights in to the dominant hydrological processes, it is indispensable to investigate the effects of heterogeneities at different spatial scales (i.e. subcatchment or subbasin, subelement, subgrid, element-to-element and grid-to-grid) on the simulation of runoff responses. The subcatchment, subelement and subgrid scale runoff parameterization may also enhance our understanding of saturation excess runoff generation and it allows for validation of models against spatial observations.

Previous work reported in literature introduced different probability distribution function based models (PDM) to reduce the complexity of the 'fully' distributed precipitation-runoff models by parameterizing the spatial heterogeneity for instance the subbasin or subgrid scale heterogeneity of the subsurface storage and infiltration capacity by a probability distribution to model the dynamics of runoff contributing areas. Examples of these are the Hydrological Byråns Vattenballansavdelning (HBV) model (Bergström, 1976), the Xinanjiang model (Zhao, 1980;1992), the Probability distributed model or PDM (Moore and Clarke, 1981; Moore, 1985), the ARNO model (Todini, 1988; 1996), the variable infiltration capacity or VIC (Wood et al, 1992) and the Improved

Arno-IA model (Hagemann and Gates, 2003), Ducharne et al. (1998), Manfreda (2008) and Beven and Kirby (1979). Bell and Moore (1998), Cole and Moore (2009) further illustrated the performances of a grid based PDM variants based on both rain gauge and radar precipitation data.

The main tasks addressed in the present study are investigation of the performances of different parameterization approaches related to representation of the spatial heterogeneity of the subsurface storage capacity for boreal mountainous watershed. The approaches range from the explicit representation of element-to-element and grid-to-grid heterogeneities to probabilistic parameterization of subcatchment, subelement and subgrid heterogeneities. The main objective of the present study is to investigate performances of six different parameterizations for representing the spatial heterogeneity of the subsurface storage capacity for semi-distributed (elements) and gridded (1x1 km^2) cases for prediction of streamflow hydrographs and flow duration curves (FDC). We calibrated the routines for the Gaulfoss gauge in the Gaula catchment in mid-Norway and evaluated the calibrated parameters through spatial transfer to the internal catchments of Eggafoss, Hugdal bru and Lillebudal bru for model validation. To our knowledge, this study is the first attempt at evaluating the performance of different levels of parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in a boreal mountain watershed. For the study region, there are growing interest in streamflow prediction at fine temporal resolution for instance hourly for hydropeaking operation of reservoirs for production scheduling, flood forecasting and environmental flow assessment. Bastola and Murphy (2013) illustrated a marked loss in performance of the parameters calibrated for a daily time step when used for hourly simulation, which substantiates the need for hourly predictions based on parameters calibrated for hourly observations.

The study watershed and data

We used hourly streamflow data from four gauges located inside the 3600 km^2 Gaula watershed located in mid Norway (Gaulfoss, Hugdal bru, Eggafoss and Lillebudal bru) (Figure 1). The last three catchments are located inside the Gaulfoss catchment, but are not nested. For the elements based simulation, we topographically delineated 33 elements within Gaulfoss. Seven of these elements (1-7) are located inside Eggafoss, another seven elements (9-15) are located inside Hugdal bru, the smallest gauged catchment (Lillebudal bru) was discretized as element 8, and the elements 16-33 are parts of the Gaulfoss catchment outside of Eggafoss, Hugdalbru and Lillebudalbru. Generally, the discretized elements are mesoscale sizes, which are less than but comparable to the size of the smallest gauged catchment of Lillebudal bru. Figure 1a shows the locations of the study catchments, hydro-climatic stations, elevation, and different discretization schemes.

The main land use is mountainous terrain above the tree line, forests dominated by conifers and riparian areas (marshes/bogs) as shown in Figure 1b. Hypsometric curves (Strahler, 1952) indicate considerable variations in the elevations of the catchments (Figure 1c). The dominant loose material (soil) in the Gaula watershed is glacial till deposits underlain predominantly by metamorphic and igneous bedrock geology (http://www.ngu.no) (Table 1).

The characteristics of the watershed are humid temperate climate and snowmelt dependent high-flow regime (Figure 1e). The flow duration curves (Figure 1d) show marked contribution of the subsurface flow to the streamflow. Precipitation occurs mainly in the form of rainfall (April-October) and mainly snowfall (November-March). The climate input data are precipitation (P), temperature (T), wind speed (W_s), relative humidity (H_R) and global radiation (R_G) of hourly resolution, which matches to the simulation time step. The model was forced by a climate input distributed on a 1x1 km^2 grid scale based on the inverse distance weighed (IDW) spatial interpolator from the point measurement gauges. We used precipitation data from 12 gauging stations, three of which are located inside the Gaulfoss catchment. Table 1 provides a summary of some characteristics of the catchments and the hydro-climatic data.

Models and methods

Probability distributed parameterization of runoff response routines

The model structure used in the present study is based on the probability-distributed model or PDM (Moore, 1985). For the sake of generality we presented the parameterization by the general form of the double bounded distribution (DB-PDF) on [0, 1] or the Kumarswamy distribution (*KwD*) (Kumarswamy, 1980). Although the 2-shape parameter KwD can capture a wide range of variability depending on the values of its two shape parameters (see Fig 2b), it entails an inconvenient analytical solution. Therefore, a special form of the KwD obtained by setting the shape parameter 'a' to 1, which then forms the 1-shape parameter Pareto distribution, is used due to its analytical solution and computational feasibility. However, the performance of the PDM based models may depend on the parameterization approaches used to represent the spatial heterogeneities that we wanted to investigate. We provide detailed descriptions of six parameterization cases evaluated in the present study as below while Table 2 provides a summary. Table 3 gives the lists of calibrated model parameters and their uniform prior ranges.

Case 1: Subcatchment heterogeneity by a probability distribution, catchment scale S_{max} and calibrated shape parameter 'b'

This case is similar to the probability distribution functions based parameterizations in the PDM (Moore, 1985) and the VIC model (Wood et al., 1992). The PDM model is collection of subsurface reservoirs with different storage capacities (*c*) and maximum storage capacity (c_{max}). We take into account the pattern of subcatchment scale runoff by parameterizing the heterogeneity of the subsurface storage capacity in the catchment by a probability distribution. We computed the maximum storage capacity on the catchment scale (the catchment scale S_{max}) from the calibrated parameters c_{max} and c_{min} and the shape parameters according to eq. (A.3) or the analytical solution in Appendix A. This case does not represent the element-to-element heterogeneity of the S_{max}.

The effective precipitation (TOSTORAGE) is partitioned into saturation excess runoff or 'saturation from below' (Dunne and Black, 1970 a&b) and change in storage based on the probability distribution following the 'equal storage redistribution of interacting storage elements' concept of Moore (1985) as shown in Figure 2a. The subsurface storage is a conceptual 'bucket type' single reservoir of finite storage capacity (equal to S_{max}) depleted by the subsurface drainage and evaporation from the subsurface (soil).

The cumulative distribution function (CDF) and the probability density function (PDF) for the *KwD* for a random variable of c_n (Figure 2b) are:

$$
\begin{aligned}
&CDF: F(c_n) = 1 - \left\{1 - (c_n)^a\right\}^b;\ c_n[0,1] = \frac{c - c_{\min}}{c_{\max} - c_{\min}} \\
&PDF: f(c_n) = \frac{dF(c_n)}{dc_n} = \begin{cases} ab(c_n)^{a-1}\left\{1-(c_n)^a\right\}^{b-1} & ,a>0, b>0 \\ 0, \text{otherwise} \end{cases}
\end{aligned} \tag{1}
$$

By inverting the cumulative distribution function, the quantile function for the local storage capacity (*c*) is:

$$
c = c_{\min} + (c_{\max} - c_{\min})\left\{1 - \left[1 - F(c)\right]^{\frac{1}{b}}\right\}^{\frac{1}{a}}, \tag{2}
$$

where the c_n [0, 1] is the normalized local storage capacity, $c_{\min}$ is a parameter which represents the minimum (threshold) local storage capacity below which there is no saturation excess runoff generation (Hegemann and Gates, 2003) and it also represents the threshold storage below which there is no drainage and water held under soil tension (Moore and Bell, 2002; Moore, 2007). The $c_{\max}$ is the maximum local storage capacity and 'a' and 'b' are the shape parameters of the distribution. Appendix A contains further details of the *KwD* and analytical solutions for the Pareto distribution.

The direct runoff generated due to infiltration excess or R_{iex} [L] (Horton, 1933) and the actual infiltration to the soil or *TOSOIL* [L] are:

$$
\begin{aligned}
&R_{iex} = \max\left[0, (SNOWOUT - INFCAP)\right] \\
&TOSOIL = SNOWOUT - R_{iex}
\end{aligned}, \tag{3}
$$

where the INFCAP [L] is a free parameter and *SNOWOUT* [L] is outflow from the snow routine to the soil. The saturated excess direct runoff or *R* [L] is the amount of runoff in excess of the storage capacity. The change in storage with time is:

$$
\begin{aligned}
&\frac{dS}{dt} \approx \Delta S = S(t + \Delta t) - S(t) = TOSTORAGE - R; \\
&TOSTORAGE = TOSOIL - AET - D_{rv}
\end{aligned} \tag{4}
$$

We computed the actual evapotranspiration from the soil (*AET* [L]) as a linear function of potential evapotranspiration rate (*PET*) from the storage, the total storage (S_T) and the total maximum storage capacity ($S_{T\max}$):

$$
AET = PET \times \frac{S_T}{S_{T\max}} \tag{5}
$$

We used the following conceptual relationships between storage and drainage for the subcatchment based runoff response:

$$D_{rv} = k\left[S\left(t\right)\right]^{n}, \tag{6}$$

where k is in $mm^{1-n}h^{-1}$, S [L] is the storage in mm, D_{rv} [L] is the drainage volume per unit area computed before saturation excess runoff and n is a dimensionless exponent. We computed the drainage or D_r [L^3T^{-1}] from the D_{rv} [L]. From eqn. (4), we derived the following equation for simulation of saturation excess direct runoff over the interval t, t+Δt:

$$R\left(t\right) = \begin{cases} TOSTORAGE - \left(S_{\max} - S(t)\right) + S_{\max}\left\{\left[1 - \dfrac{S(t)}{S_{\max}}\right]^{\frac{1}{b+1}} - \dfrac{TOSTORAGE}{\left(b+1\right)S_{\max}}\right\}^{b+1} & ; S(t+\Delta t) < S_{\max} \\ TOSTORAGE - \left(S_{\max} - S(t)\right); S(t+\Delta t) \geq S_{\max} & \end{cases} \tag{7}$$

We computed the rate of total direct runoff (R_r [L^3T^{-1}]) as:

$$R_r = \frac{A_i}{\Delta t}\left\{R \times F\left(c^*\left(t\right)\right) + R_{iex}\right\}, \tag{8}$$

where the $F(c^*(t))$ is the fraction of the catchment saturated to generate the saturation excess runoff (see Appendix A) and A_i is the catchment area.

Case 2: Subelement heterogeneity by a probability distribution, element-to-element heterogeneities of the S_{max} and the shape parameter 'b'

In case 2, we investigated the case when the element-to-element heterogeneity of the maximum storage capacity or S_{max} and the shape parameter are modelled (i.e. the S_{max} and the shape parameter are computed for each element).

The role of topography in runoff response dynamics have been widely studied (e.g. Beven and Kirby, 1979; Wood et al., 1990; Blöschl and Sivapalan, 1995; Wood et al., 1990; Bell and Moore, 1998). In a study of Norwegian catchments, Beldring et al. (2003) noted a relationship between the maximum soil moisture storage and altitude with larger soil moisture storage for lowland areas than for mountains as the average thickness of surface deposits tends to decrease with altitude. Therefore, depending on the distribution of topographic and soil characteristics in the catchment, the maximum storage capacity (S_{max}) may vary throughout the catchment and hence the effects of the lumped representation of the maximum storage capacity on the runoff simulation needs further investigation.

Dumenil and Todini (1992) computed the shape parameter 'b' of the distribution based on the standard deviation of the subgrid elevation. Bell and Moore (1998), Hagemann and Gates (2003), Manfreda and Fiorentino (2008), Manfreda (2008) and Liu et al. (2012) also noted the topographic influence on ′b'. Figure 2b illustrates as the value of 'b' increases, the fraction of element with small storage capacity increases that increases the likelihood of more saturation excess runoff.

This case considers the influence of topography on the storage capacity and hence on the dynamics of runoff generation by directly utilizing the topographic information. It is useful to represent the spatial heterogeneity of hydrological variables based on readily available high-resolution spatial information such as topographic features derived from the Digital Elevation Model (DEM) to reduce the number of model parameters, to allow transfer of parameters to ungauged catchments and in parameterization for climate models (e.g. Ducharne et al., 1998).

The maximum storage capacity and shape parameter 'b' for each element are computed from a functional relationship between the equations for maximum storage capacity (S_{max}) from integration of eq. (A.3) or its analytical solution and based on the topographic gradient (eq. 9). This approach is similar to the linkage function in the grid-to-grid (G2G) model of Bell and Moore (1998). We related the parameter 'b' to the maximum storage capacity of the Pareto distribution:

$$S_{\max} = \left(\frac{MaxMFDslope_{\max} - MaxMFDslope_{avg}}{MaxMFDslope_{\max}} \right) \left(c_{\max} - c_{\min} \right); \; S_{\max} = \frac{c_{\max} - c_{\min}}{b+1} \tag{9}$$

Equating the above two equations for S_{max}, the following relationship for 'b' and the topographic gradient:

$$b = \frac{MaxMFDslope_{avg}}{MaxMFDslope_{\max} - MaxMFDslope_{avg}} \tag{10}$$

The above relations provide

$$S_{max} = c_{max} - c_{min} \{ if\ MaxMFDslope_{avg} = 0,\ b = 0$$
$$S_{max} = \frac{c_{max} - c_{min}}{2} \{ if\ MaxMFDslope_{avg} = 0.5 MaxMFDslope_{max},\ b = 1\ , \tag{11}$$
$$S_{max} = 0 \{ if\ MaxMFDslope_{avg} = MaxMFDslope_{max}\ ,\ b\ is\ undefined$$

where the $MaxMFDslope_{avg}$ represent the average of the gradients of the 1x1 km^2 grid cells within the element while $MaxMFDslope_{max}$ is a regional parameter representing the maximum of gradients for the 1x1 km^2 grid cells in the whole catchment.

The MaxMFDslope for the grid cell is the topographic gradient in the steepest flow direction among its eight neighbors in a 3x3 window computed from the DEM as MFDslope = (Elevation $_{\text{upstream cell}}$ - Elevation $_{\text{downstream cell}}$) / Flow travel length. Flow travel length = (grid cell size) for the cardinal flow direction and (grid cell size)*√2 for diagonal flow directions. For the element based simulation (cases 1 to 3), the case of $MaxMFDslope_{avg}$ = $MaxMFDslope_{max}$ is not an issue, but for the grid based simulation (cases 1G to 3G) a $storage_{min}$ calibrated parameter was introduced to avoid S_{max} and 'b' to become zero in flat areas. Besides allowing study on the sensitivity of runoff generation to spatial distribution of S_{max} and b, case 2 also reduces the number of free parameters.

Case 3: No subcatchment and subelement heterogeneity of the storage capacity and no element-to-element heterogeneity of S_{max}

In case 3, there is no parameterization of the spatial heterogeneity by a probability distribution. There is no element-to-element heterogeneity of the S_{max}, rather it is a free parameter. Therefore, case 3 is a simple semi-distributed model. We computed the direct runoff and update of the storage for the elements as follow:

$$R = \max\left[0, \left(S(t) + TOSTORAGE - S_{max}\right)\right]; S(t+\Delta t) = \max\left\{0, \left[S(t) + TOSTORAGE - R\right]\right\}$$
$$R_r = \left\{R + R_{iex}\right\} \times \frac{A_i}{\Delta t} \tag{12}$$

Cases 1G, 2G and 3G: Grid based runoff simulation

Case 1G is parameterization of the subgrid heterogeneity by a probability distribution, but we calibrated the parameters for the catchment scale similar to that of case 1. Case 2G is parameterization of the subgrid heterogeneity by the probability distribution and it accounts for the grid-to-grid heterogeneity of S_{max} based on the linkage function between

S_{max} and the topographic gradient. We derived the following equations for case 2G from the linkage function between topographic gradients and the S_{max}:

$$S_{max} = Storage_{min} + \left\{ \left(c_{max} - Storage_{min} \right) \times \left(\frac{MaxMFDslope_{max} - MaxMFDslope}{MaxMFDslope_{max}} \right) \right\} \tag{13}$$

$$b = \frac{\left(MaxMFDslope \times \left(c_{max} - Storage_{min} \right) \right) - \left(c_{min} \times MaxMFDslope_{max} \right)}{c_{max}\left(MaxMFDslope_{max} - MaxMFDslope \right) + \left(Storage_{min} \times MaxMFDslope \right)} \tag{14}$$

$$a = \frac{c_{max}\left(MaxMFDslope_{max} - MaxMFDslope \right) + \left(Storage_{min} \times MaxMFDslope \right)}{\left(c_{min} \times MaxMFDslope_{max} \right) + \left(MaxMFDslope \times \left(c_{max} - Storage_{min} \right) \right)} \tag{15}$$

Case 3G donot consider both the subgrid heterogeneity of storage capacity by a probability distribution and the grid-to-grid heterogeneity of the S_{max}. We set the S_{max} by calibration. Therefore, case 3G is a simple distributed model.

The main difference between the distributed simulations (cases 1G to 3G) and the semi-distributed simulations (cases 1 to 3) is in the equations used for simulation of the subsurface drainage. In a boreal landscape dominated by till soils, hydraulic conductivity decreases with depth, the groundwater table largely follows the topography and the catchment runoff depend on soil moisture conditions and the depth to groundwater (Lind and Lundin, 1990; Hinton et al., 1993; Myrabø, 1997; Beldring, 1999; 2002). Therefore, topography has a significant impact on runoff by controlling movement and storage of water through convergence and divergence of flow (Beldring, 2002). We computed the rate of subsurface drainage or flow from derived equation based on assumptions of Dupuit-Forchheimer to Darcy's law for saturated subsurface flow (Freeze and Cherry, 1979; Wigmosta and Lettenmaier, 1999) by assuming a power-law transmissivity decay with depth (Ambroise et al. 1996; Wigmosta and Lettenmaier, 1999):

$$D_r = \left\{ \frac{w}{n_T} \psi MaxMFDslope \left[S_{max} \right]^{1-n_T} \frac{\Delta t}{A_i} \right\} \left[S(t) \right]^{n_T}, \tag{16}$$

where ψ [LT^{-1}] is diffusivity or saturated hydraulic conductivity at the surface divided by porosity, w [L] is size of the grid cell, and n_T is the transmissivity decay exponent, D_r [L^{3}T^{-1}] is drainage and D_{rv} (L) is drainage volume per unit area computed from the D_r and the MaxMFDslope is as defined earlier. Eq. (16) accounts for the grid-to-grid heterogeneity of topographic gradient. Based on preliminary tests of parameter sensitivity, we used a hyperbolic (Duan and Miller, 1997) transmissivity decay profile (i.e. exponent $n_T = 2.0$) for the study catchments.

The evapotranspiration routine

We used the Priestley Taylor method (Priestley and Taylor. 1972) to estimate the potential evapotranspiration:

$$PET = \alpha \frac{\Delta}{\Delta + \gamma} R_n , \tag{17}$$

where α is Priestley Taylor constant, the Δ is the slope of saturation vapor pressure curve corresponding to an air temperature at 2m (T2m), Y is the psychometric constant (0.67 $hPaK^{-1}$), R_n is the net radiation, which is the sum of net shortwave radiation and the net longwave radiation. Following Teuling et al. (2010), we used α = 1.26 rather than calibrating to reduce the number of parameters. We computed the net short wave radiation from global radiation and land albedo, and the net long wave radiation based on Sicart et al. (2006). We used eq (5) for the computation of the actual evapotranspiration (*AET*) for all cases. The *AET* is set to zero when the surface is snow covered.

The snow routine

The snow accumulation and snowmelt processes exert significant influence on the hydrological cycle of the study area. The outflow melt water release from saturated snow (i.e. *SNOWOUT*) is computed by a snow routine based on the Gamma distributed snow depletion curve (SDC) (Kolberg and Gottschalk, 2006; 2010), which was implemented in ENKI hydrological modelling platform (Kolberg and Bruland, 2012). The free parameters in this routine are snow-rain threshold temperature parameter (TX) and snowmelt sensitivity to wind speed or the windscale (WS). We simulated the potential evapotranspiration, snow accumulation and snowmelt-runoff processes based on the 1x1 km^2 grid scale. For the semi-distributed (element) simulations (cases 1, 2 and 3), we aggregated the gridded (1x1 km^2) outflow from the snow routine (*SNOWOUT*) and potential evapotranspiration (*PET*) to the element scale based on simple averaging and provided as an input to the runoff response routines.

Runoff routing

We used the source-to-sink (STS) routing algorithm (Olivera, 1996; Olivera and Maidment, 1999) to route the generated runoff at each source to the sink (catchment outlet). In the STS, a flow path response function or $U_i(t)$ [T^{-1}] relates the instantaneous runoff generated at the source to the outlet response. The flow path response function is based on the first passage time distribution (Hayami, 1951; Nauman, 1981). Olivera (1996) showed the relationships among the total expected travel time from the source to the outlet (T_i), its corresponding variance or $Var(T_i)$, the flow dispersion coefficient (D_i) and Peclet number or Π_i[-] based on the statistical properties of mean and variance. The gridded (1x1 km^2) flow path response function or $U_i(t)$ [T^{-1}] is given by:

$$U_i(t) = \frac{1}{2t\sqrt{\pi\left(\frac{t}{T_i}\right)/\Pi_i}} \exp\left\{ -\frac{\left[1-\left(\frac{t}{T_i}\right)\right]^2}{4\left(\frac{t}{T_i}\right)/\Pi_i} \right\} \tag{18}$$

The flow path responses function for grid cell '*i*' represents the probability distribution of flow travel time (t) from the source (grid cell) to the sink (outlet) which has a mean value of T_i . The flow path Peclet number is a representative measure of the relative importance of advection with respect to dispersion, whereby the flow dispersion coefficient represents the effects of storage and spreading. Figure 3c shows typical response functions.

For the semi-distributed runoff simulations (cases 1, 2 and 3), we distributed the generated runoff at the element scale over the 1x1 km^2 grid cells within the elements. We coupled the generated runoff to the flow path response function to perform the flow routing on the grid scale, rather than aggregating to the element scale. The grid scale would enable us to account for the differences in the flow travel time and hence response functions among the grid cells especially for elongated elements. Beldring et al. (2003) noted that the permanent river network including streams and lakes lies within 1x1 km^2 of almost every point in the Norwegian landscape and all the lateral transfers of water at 1x1 km^2 grid cells take place within the river network.

We routed the sum of direct runoff and subsurface drainage generated at the source grid cell to the outlet. We performed the runoff routing by a convolution from the unit hydrograph model for a spatially distributed linear system subdivided into uniform non-overlapping sub-areas (Maidment et al., 1996; Olivera and Maidment, 1999):

$$Q(t) = \sum_{i=1}^{N} \left\{ R_{ri}(t) + D_{ri}(t) \right\} \otimes U_i(t) \tag{19}$$

Q [L^3T^{-1}] is routed simulated streamflow, N is the total number of grids in the watershed, and $\otimes$ is the convolution operator. The routing routine involves two free parameters namely the velocity of flow (V) and the dispersion coefficient of flow (D). We assumed the parameters to be 'time-invariant' and set as calibration parameters for the whole catchments (i.e. not spatially distributed).

Model calibration and evaluation

We used the Differential Evolution Adaptive Metropolis algorithm or DREAM (Vrugt et al, 2008; 2009) with residuals based log-likelihood objective function for the calibration. DREAM is an adaptive random walk Metropolis algorithm to enhance the applicability of the MCMC methods to complex, non-linear and high-dimensional problems such as calibration of watershed models (Vrugt et al, 2008; 2009).

For the hourly streamflow series, the serial correlation is high and hence the actual amount of information obtained from the data is much less. Therefore, we introduced the fraction of effectively independent observations from the total observations denoted as 'f'. We used the logarithmic likelihood function for simplicity and numerical stability (Vrugt et al., 2012). The residual based log-likelihood (L-L) is:

$$L-L\left(\delta / \sigma_{\varepsilon}^{2}, \sum_{t=1}^{n_i}\left(Qsim_t^{(\theta)} - Qobs_t^{(\theta)}\right)^2\right) = \left(\frac{-n_i}{2}\log(2\pi) - \frac{n_i}{2}\log\left(\sigma_{\varepsilon}^{2}\right) - \frac{\sum_{t=1}^{n_i}\left(Qsim_t^{(\theta)} - Qobs_t^{(\theta)}\right)^2}{2\sigma_{\varepsilon}^{2}}\right) \times f, \tag{20}$$

where $Qsim^{(\theta)}$ and $Qobs^{(\theta)}$ respectively are Box-Cox (Box and Cox, 1964) transformed observed and simulated streamflow time series (t), n_i is the length of non-missing records of streamflow for the catchment, δ represents model parameter, θ is the Box-Cox transformation parameter and σ_ε^2 is variance of error.

We carried out the transformation in order to obtain an approximately Normal distributed series with homoscedastic residuals. If $\theta = 0$, the streamflow is assumed to be lognormal distributed (i.e. high weightage to low flows). If $\theta = 1$, the streamflow series is assumed to be Gaussian (i.e. high weightage to high flow). A value of $\theta = 0.3$ is common in literature (e.g. Vrugt and Bouten 2002 and references therein). However, we computed the θ values from the observed streamflow data set using the 'fminsearch' algorithm in matlab, which calls for finding the θ value that maximizes a likelihood function (http://www.mathworks.com). We computed the fraction f based on a AutoRegressive or AR(1) model of error covariance (Zięba, 2010).

The maximum Nash-Sutcliffe efficiency or R^2 (Nash and Sutcliffe, 1970), which emphasizes high flows, and the maximum R^2ln for log-transformed series, which emphasizes low flows), were used for further comparisons and evaluations. We also evaluated the performances of the routines based on their predictive reliability (Kavetski and Fenicia, 2011) using quantile-quantile (Q-Q) plots. The plots were in the form of the empirical cumulative distribution functions (CDF) of the observed and simulated streamflow. The departures of the plots from the theoretical uniform distribution indicate the discrepancy between the predictive distribution and the observed data. In addition, we evaluated the routines based on their prediction performances of temporal variability of the streamflow or the flow duration curves. We used the 'split-sample' test and 'proxy basin' test (Klemeś, 1986) for temporal, spatial and spatio-temporal validation of the models against independent data to test the reliability of model prediction outside the calibration conditions (Seibert, 2003). We performed the spatial and spatio-temporal validation of the models through direct transfer of calibrated parameter vectors, which correspond to the maximum R^2 and maximum R^2ln, of the Gaulfoss catchment to the internal catchments of Eggafoss, Hugdalbru and Lillebudal bru both for the calibration and validation periods.

Results and discussion

Model calibration

Figure 3a-b, Figure 4 and Figure 5 respectively present the hydrographs, quantile-quantile (Q-Q) plots and flow duration curves of observed versus simulated streamflow for Gaulfoss. We presented the hydrographs only for the R^2 performance measure for a part of calibration period (for clear presentation of the Figures), the Q-Q plots for the R^2 for the calibration and validation results and the flow duration curves for both the R^2 and R^2ln for the results of calibration and validation periods. We presented the performance measures for calibration, temporal, spatial and spatio-temporal validation of the calibrated parameters in Table 4.

We obtained the goodness-of-fits of R^2/R^2ln respectively up to 0.84/0.86 for calibration and up to 0.85/0.90 for the temporal and spatial validation, which indicate satisfactory fits between the observed and simulated hydrographs for the six different parameterization cases for semi-distributed and distributed runoff simulation. Therefore, the simulations based on different parameterizations of the single state and single drainage outlet subsurface storage provided satisfactory runoff simulation in terms of the goodness-of-fit tests.

However, a more stringent test for reliability of prediction based on the Q-Q plots of the observed and simulated streamflow indicate that there is a considerable prediction uncertainty for all the parameterization cases as shown in Figure 4. Nearly symmetrical deviations from the perfect fit uniform distribution (diagonal line) show both under and over predictions. The results of the flow duration curves indicate better simulation of the temporal variability of the high flow compared to the low flow as shown in Figure 5. The performance of the calibration of the parameterizations in reproducing the hydrographs (based on R^2 and R^2ln performance measures) found to be better than the Q-Q plots and the FDC. Calibration of hydrological models for specific objectives of reproducing the flow-duration curves (e.g. Westerberg et al., 2011) and the Q-Q values may improve their respective performances.

It was not possible to consistently distinguish the best performing parameterization since the different parameterizations provided only marginally different performances for different seasons (snow melt versus rainfall) and ranges of streamflow (low, medium and high).

Model validation

Investigation of the performances of distributed models calibrated to the streamflow at basin outlet for hydrologic simulation at internal catchments was one of the science question tested by the Distributed Model Intercomparison Project, DMIP (Smith et al., 2004). Spatial transferability of calibrated parameters from the Gaulfoss catchment to the internal catchments of Eggafoss and Hugdal bru (Table 4) provided indistinguishable satisfactory performances for all the parameterization cases. However, parameter transfer to Lillebudal catchment resulted in poor performance especially for R^2ln (low flows). For

Lillebudal bru catchment, Hailegeorgis and Alfredsen (2014, article in press) found poor performance of parameter transfer from the Gaulfoss watershed for the HBV conceptual model especially for low flow simulation. Mountainous terrain dominates the Lillebudal bru catchment with a mean altitude above the altitude of all climate stations used for the calibration (Table 1). Effects of the dominantly mountainous terrain may cause significant temporal and spatial variability of precipitation fields. Moreover, there are no climate stations inside or close to the Lillebudal bru catchment and hence less representativeness in precipitation data may cause poor streamflow simulation. Performance in simulation of low flow from the baseflow is also rather poor for Lillebudal bru. Therefore, the quality of observed streamflow data needs scrutiny for the Lillebudal bru gauge.

Parametrical parsimony

The effects of correlations among the parameters during calibration may cause the poor identifiability of the parameterizations. One of the solutions to reduce parameter interactions is improving the parametrical parsimony of the routines. Parametrical parsimony involves reducing the number of free parameters for instance by fixing the insensitive parameters. The parameter c_{min} is introduced to represent the minimum (threshold) local storage capacity below which there is no saturation excess runoff generation (Hegemann and Gates, 2003) and also the threshold storage below which there is no drainage, water being held under soil tension (Moore, 2007). The calibrated values for this parameter were less than 7.5 mm against a uniform prior range of (0.0 -100 mm) and hence the c_{min} can be set to zero and excluded from the free parameters to improve the parsimony and to avoid its interaction with c_{max} and other parameters.

Also, the calibrated values for the exponent parameter of the conceptual subsurface drainage-storage relationship (n) of the parameterization cases 1 to 3 ranges from 1.5 to 3.0 against a uniform prior range of (0.20-5.0). Hence, there is a possibility to fix this parameter to some representative value within this range to improve the parsimony and to avoid its interaction with k and other parameters. For instance, Wittenberg and Sivapalan (1999) and Moore and Bell (2002) respectively assumed quadratic ($n = 2.0$) and cubic ($n = 3.0$) relationships between ground water storage and baseflow.

However, parametrical parsimony alone may not guarantee improvement in the identifiability and predictive uncertainty of the parameterizations since there are also other sources of uncertainty related to the input data and scale issues.

The effects of input data for model calibration

We conducted the semi-distributed and distributed runoff simulations for the boreal mountainous catchments based on precipitation data from 12 gauging stations spatially interpolated by inverse distance weighing (IDW) on 1x1 km^2 grids. However, Goodrich et al. (1995) reported an inadequacy of meteorological gauging networks in the higher altitudes. Also for high latitude mountainous regions, Moine et al. (2003) noted the complexity of hydrological modelling due to the complexity of local processes and the difficulty of estimating spatially distributed inputs such as rainfall and temperature due to sparse networks. Beldring et al. (2003) noted that the spatial interpolation procedure with correction for altitude differences is unable to describe all effects caused by the various precipitation formation mechanisms and wind directions in Norwegian catchments. Das et al. (2008) found that a distributed HBV model structure do not outperform the simpler model structures, which they attributed to the interpolated climate inputs that cannot reflect the true spatial variability. Wrede et al. (2013) compared a distributed HBV model complemented by subgrid scale parameterization for distinct land use classes to a less parameterized lumped HBV model for a Swedish lowland catchment. The authors found that the results were indistinguishable, which they attributed to the deficiency of the traditional model calibration against only the observed streamflow at the catchment outlet. Calibration based on climate data from high density gauging networks and spatial distributed observations, which are not available for the present study, may provide more insights out of the simulations. In addition, only the catchment integrated observed streamflow data is available for calibration for the study region.

Parameterization and scale issues

Both discretization and aggregation techniques in precipitation-runoff models are dependent on the scales and hence the results of simulation from parameterization across a range of scales may be sensitive to the spatial scales used (e.g. Wood et al., 1990; Becker and Braun, 1999; Koren et al., 1999; Haddeland et al., 2002; Merz et al., 2009). Beldring et al. (1999; 2000) suggested elements at scales of approximately 1x1 km^2 sufficient to parameterize the hydrological processes in till soils. Gottschalk et al. (2001) also identified a hillslope scale of 1-2 km^2 for the NOPEX region. In addition, Wood et al. (1988; 1990) identified a ‘Representative Elementary Area (REA)’ of subcatchments of about 1x1 km^2. Scale issues in hydrological modelling (Blöschl and Sivapalan, 1995) are one of the major challenges in parameterization of precipitation-runoff models.

Due to the sparse hydro-meteorological networks (i.e. only 12 precipitation stations distributed over the boreal mountainous study region), it is clear that the resolution of the forcing field is low. Even though the resolution of climate forcings is much lower than the resolution of the parameterization for the case 2G, the performances of the case 2G and case 2 are indistinguishable. There are scale mismatch between spatial heterogeneities of climate control and topographic control due to the prevailing terrain heterogeneity at a finer hillslope scale (e.g. 25mx25m). For the boreal watershed, the topographic driven influence on the spatial heterogeneity of soil moisture, subsurface storage and hence lateral movement of subsurface flow dominate the grid-to-grid variability of the low intensity precipitation. However, the only advantage of distributed (gridded) simulations (cases 1G to 3G) over the semi-distributed (cases 1 to 3) in the present study is found to be the simplicity in preparing gridded input maps for the distributed model than preparing topographically delineated elements for the semi-distributed model rather than marked improvement in the runoff simulation.

For the boreal catchments, topographic control heterogeneities at finer spatial scales dominates the runoff generation processes and hence parameterizations for the finer scale hillslope processes may be required (see Halldin et al., 1999). Therefore, the grid cell-to-grid cell routing in the hillslopes (e.g. 25mx25m grids) towards the stream networks by considering the hillslope topographic gradients within the 1x1 km^2 grid like in the distributed hydrology-soil-vegetation model or DHSVM (Wigmosta et al., 1994) may further allow more representativeness and utility of the terrain features.

Conclusions

We examined the performances of six different parameterizations of the subsurface storage capacity based on the probability distributed model for semi-distributed and distributed (1x1 km^2 grids) hourly runoff simulation in boreal mountainous watershed in mid Norway.

The study for the boreal watershed of variable relief showed that the subelement and subgrid scale parameterizations of the subsurface storage capacity donot provide better results for the hourly runoff simulation than the coarser parameterizations, which indicate:

i. Identification of parameterizations require measurements from dense precipitation networks than what is required for a mere calibration of precipitation-streamflow relationships;
ii. Challenges towards identification of parameterizations based on model calibration only to the catchment integrated streamflow observations;
iii. Equivalent simulation performance for the available data set showed a potential preference for the simple and parsimonious parameterizations e.g. case 3G (a simple distributed routine) in operational forecast mode related to model updating.
iv. The effects of input uncertainties related to precipitation and streamflow, and parameter non-identifiability on identification of the parameterizations require scrutiny.

Previous studies are lacking pertinent to comparisons of different parameterizations of the subsurface storage capacity for hourly runoff simulation in boreal catchments. Both the precipitation control e.g. density of the climate networks and topographic control driven heterogeneities at further finer spatial scales need thorough exploration. In addition, the present study donot consider the preferential flow that may be apparent in the glacial till soils.

Acknowledgements

The Center for Environmental Design of Renewable Energy (CEDREN) supported the research financially under HydroPEAK hydrology (Project number: 50043420). We obtained the climate data from the Norwegian Meteorological Institute, TrønderEnergi, Statkraft and bioforsk, and the streamflow data from the Norwegian Water Resources and Energy Directorate.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.,1986. 'An Introduction to the European Hydrological System - Systéme Hydrologique

Européen, "SHE", 2: structure of a physically-based, distributed modelling system', *J. Hydrol*. 87, 61-77.

Ambroise, B., Beven, K., Freer J., 1996. Towards a generalization of the TOPMODEL concepts: topographic indices of hydrological similarity. *Water Resour. Res*. 32(7), 2135–2145.

Bastola, S. & Murphy, C., 2013. Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. *Hydrology Research* 44(3), 484-494.

Becker, A., and P. Braun, 1999. Disaggregation, aggregation, and spatial scaling in hydrological modelling. *J. Hydrol*. 217, 239–252.

Beldring, S., Gottschalk, L., Seibert, J., Tallaksen, L.M., 1999. Distribution of soil moisture and groundwater levels at patch and catchment scales. *Agr. Forest Meteorol*. 98-99, 305-324.

Beldring, S., Gottschalk, L., Rodhe, A., Tallaksen, L.M., 2000. Kinematic wave approximations to hillslope hydrological processes in tills. *Hydrol. Process*. 14, 727-745.

Beldring, S., 2002. Runoff Generating Processes in Boreal Forest Environments with Glacial Tills. *Nord. Hydrol*. 33(5), 347-372.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., Vøkso, A., 2003. Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrol. Earth. Syst. Sc*. 7(3), 304-316.

Bell, V.A., Moore, R.J., 1998. A grid-based distributed flood forecasting model for use with weather radar data: Part 1. Formulation. *Hydrol. Earth. Syst. Sc*. 2(2-3), 265-281.

Bergström, S.,1976. Development and application of a conceptual runoff model for Scandinavian catchments, *SMHI Reports RHO*, No. 7, Norrköping.

Beven, K. J., Kirby, M. J.A, 1979. Physically-based variable contributing area model of basin hydrology. *Hydrolog. Sci. Bull*. 24(1), 43–69.

Beven, K. J., Freer, J., 2001. Equifinality, data assimilation, and un-certainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol*. 249, 11-29.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modeling: a review. *Hydrol. Process*. 9, 251-290.

Bogaard, T.A., Luxemburg, W.M.J., Wit, M. de., Douben, N., Savenije, H.H.G., 2005. Some hydrological challenges in understanding discharge generation processes in the Rhine and Meuse basins. *Phys. Chem. Earth, Pt. A/B/C* 30 (4-5), 262-266.

Box, G. E. P., Cox, D. R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society*, Series B *26*, 211-252.

Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. *J. Hydrol*. 356, 106-118.

Duan, J., Miller, N.L., 1997. A generalized power function for the subsurface transmissivity profile in TOPMODEL. *Water Resour. Res*. 33(11), 2559–2562.

Ducharne, A., Laval, K., Polcher, J., 1998. Sensitivity of the hydrological cycle to the parameterization of soil hydrology in GCM. *Clim. Dynam*.14, 307-327.

Dumenil, L. and Todini, E., 1992. A rainfall-runoff scheme for use in the Hambrurg climate model, in: O'Kane, J.P. (Ed), *Advances in Theoretical Hydrology, a Tribute to James Dooge*, Amsterdam, Netherlands: Elsevier,129-157.

Dunne, T., Black, R. D., 1970a. An experimental investigation of runoff production in permeable soils. *Water Resour. Res*. 6(2), 478–490.

Dunne, T., and Black, R. D., 1970b. Partial area contributions to storm runoff in a small New England watershed. *Water Resour. Res*. 6(5), 1296–311.

Freeze, R.A., Cherry, J.A., 1979. Groundwater, Prentice-Hall, Englewood Cliffs, N.J.

Goodrich, D. C., 1990. 'Geometric simplification of a distributed rainfall-runoff model over a large range of basin scales', *PhD Thesis*, Univ. Arizona, Tucson, 361 pp.

Goodrich, D.C., Faures, J.-M., Woolhiser, D.A., Lane, L.J., Sorooshian, S., 1995. Measurment and analysis of small-scale convective storm rainfall variability. *J. Hydrol*. 173, 283-308.

Gottschalk, L., Beldring, S., Engeland, K., Tallaksen, L.M., Sælthun, N. R., Kolberg, S.,Yury, M., 2001. Regional/macroscale hydrological modelling: a Scandinavian experience. *Hydrolog. Sci. J*. 46 (6), 963-982.

Haddeland, I., Matheussen, B.V., Lettenmaier, D.P., 2002. Influence of spatial resolution on simulated streamflow in a macroscale hydrological model. *Water Resour. Res*. 38 (7), 1124.

Hagemann, S., Gates, L.D., 2003. Improving a subgrid runoff parameterization scheme for climate models by the use of high-resolution data derived from satellite observations. *Clim. Dynam*. 21, 349-359.

Hailegeorgis, Teklu T., and Alfredsen, Knut, 2014. Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrology Research*, article in press.

Halldin, S., Gottschalk, L., Gryning, S.E., Jochum, A., Lundin, L.C., Van de Griend, A.A., 1999. Energy, water and carbon exchange in a boreal forest-NOPEX experiences. *Agric. For. Meteorol*. 98-99, 5-29.

Hayami, S., 1951. 'On the propagation of flood waves', *Disaster Prev. Res. Inst. Bull*.1, 1-16.

Hinton, M.J., Schiff, S.L., English, M.C., 1993. Physical properties governing groundwater flow in a glacial till catchment. *J. Hydrol*. 142*,* 229-249.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle. *Trans. Am. Geophys. Union* 14, 446-460.

Jansson, C. Espeby, B., and Jannson, P-E., 2004. Preferential flow in glacial till soil. *Nord. Hydrol*. 36 (1), 1-11.

Kavetski, D., Fenicia, F., 2011. Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resour. Res*. 47(11) DOI: 10.1029/2011WR010748.

Klemes, V, 1986. Operational testing of hydrological simulation models. *Hydrolog. Sci. J*. 31, 13-24.

Kolberg, S. A., Bruland, O., 2012. ENKI - An Open Source environmental modelling platform. *Geophysical Research Abstracts* 14, EGU2012-13630, EGU General Assembly.

Kolberg, S. A., Gottschalk, L., 2006. Updating of snow depletion curve with remote sensing data. *Hydrol. Process*. 20(11), 2363-2380.

Kolberg, S. A., Gottschalk, L., 2010. Interannual stability of grid cell snow depletion curves as estimated from MODIS images. *Water Resour. Res*., 46 (11), doi:10.1029/2008WR007617.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J.M., 1999, A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *Journal of Geophysical Research* 104 (D16), 19,569-19,585.

Kumaraswamy, P, 1980. A generalized probability density function for double-bounded random processes. *J. Hydrol*. 46 (1-2), 79–88.

Leavesley, G. H. and Stannard, L. G. 1990. 'Application of remotely sensed data in a distributed-parameter watershed model' in Kite, G. W. and Wankiewicz, A. (Eds), Proc. *Workshop on Applications of Remote Sensing in Hydrology*, Saskatoon, February 1990. pp. 47-64.

Lind, B.B., Lundin, L., 1990. Saturated hydraulic conductivity of Scandinavian tills. *Nord. Hydrol*. 2(1), 107- 118.

Liu, J., Chen, X., Wu, J., Zhang, X., Feng, D., Xu, C.-Y, 2012. Grid parameterization of a conceptual distributed hydrological model through integration of a sub-grid topographic index: necessity and practicability. *Hydrolog. Sci. J*. 57(2).

Manfreda, S., Fiorentino, M., 2008. A stochastic approach for the description of the water balance dynamics in a river basin. *Hydrol. Earth Syst. Sci*. 12, 1189–1200.

Manfreda, S., 2008. Runoff generation dynamics within a humid river basin. Nat. *Hazards Earth Syst. Sci*. 8, 1349–1357.

Maidment, D.R., Olivera, J.F., Calver, A., Eatherral, A., Fraczek, W., 1996. A unit hydrograph derived from a spatially distributed velocity field. *Hydrol. Process*.10 (6), 831–844.

Merz, R., Parajka, J, Blöschl, G., 2009. Scale effects in conceptual hydrological modelling. *Water Resour. Res*. 45, W09405, doi:10.1029/2009WR007872.

Moine, N., Hendrickx, F., Gailhard, J., 2013. Rainfall–runoff modelling as a tool for constraining the reanalysis of daily precipitation and temperature fields in mountainous regions Cold and Mountain Region Hydrological Systems Under Climate Change: Towards Improved Projections. *Proceedings of H02, IAHS-IAPSO-IASPEI Assembly*, Gothenburg, Sweden, July 2013 (IAHS Publ. 360, 2013).

Moore, R. J., Clarke, R. T., 1981. A distribution function approach to rainfall runoff modeling. *Water Resour. Res*. 17, 1367–1382.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrolog. Sci. J*. 30 (2), 273-297.

Moore, R.J., Bell, V.A., 2002. Incorporation of groundwater losses and well level data in rainfall-runoff models illustrated using the PDM. *Hydrol. Earth Syst. Sci*. 6 (1), 25-38.

Moore, R.J., 2007. The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci*.11 (1), 483-499.

Myrabo, S., 1986. Runoff studies in a small catchment. *Nord. Hydrol*. 17, 335–346.

Myrabo, S., 1997. Temporal and spatial scale of response area and groundwater variation in Till. *Hydrol. Process*.11,1861–1880.

Nash, J. E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models, I. A discussion of principles. *J. Hydrol*. 10, 228–290.

Nauman, E.B., 1981. Residence time distributions in systems governed by the dispersion equation. *Chem. Eng. Sci*. 36, 957-966.

Olivera, F., 1996. Spatially distributed modeling of storm runoff and nonpoint source pollution using geographic information systems, *PhD Thesis*, Department of Civil Engineering, University of Texas at Austin, USA.

Olivera, F., Maidment, D.R.,1999. Geographic information system (GIS)-based spatially distributed model for runoff routing. *Water Resour. Res*. 35 (4), 1155–1164.

Priestley, C.H.B.,Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* 100, 81-82.

Refsgaard, J.C., 1997. Validation and Intercomparison of different updating procedures for real-time forecasting. *Nord. Hydrol*. 28, 65-84.

Seibert, J., 2003. Reliability of model predictions outside calibration conditions. *Nordic Hydrology* 34(5), 477-492.

Sicart, J. E. Pomeroy, J. W., Essery, R. L. H. and Bewley, D., 2006. Incoming longwave radiation to melting snow: observations, sensitivity and estimation in northern environments. *Hydrol. Process*. 20, 3697-3708.

Singh, V.P., Woolhiser, D.A., 2002. Mathematical modeling of watershed hydrology. *J. Hydrol. Eng*. 7 (4), 270-292.

Sivapalan, M. and Viney, N. R., 1994. 'Application of a nested catchment model for predicting the effects of changes in forest cover'. In: Ohta,T. (Ed.) *Proc. Int. Symp. Forest Hydrology*, Univ. Tokyo, Tokyo. IUFRO, pp. 315-322.

Smith, M.B., Seo, D. -J., Koren, V. I., Reed, S., Zhang, Z., Duan, Q.-Y., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. *J. Hydrol*. 298, 1-4, 4-26.

Strahler, A.N., 1952. Hypsometric (area–altitude) analysis of erosional topography. *Geol Soc Amer Bull* 63, 1117–1141.

Teuling, J. Lehner, I., Kirchner, J. W. & Seneviratne, S. I., 2010. Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment. *Water Resour. Res*. 46, W10502, doi:10.1029/2009WR008777.

Todini, E., 1988. Rainfall-runoff modeling-past, present, and future. *J. Hydrol*. 100, 341–352.

Todini, E., 1996. The Arno rainfall-runoff model. *J. Hydrol*. 175, 339-382.

Vrugt, J.A., Bouten, W., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resour. Res*. 38 (12), doi:10.1029/2001WR001118.

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 Resour. Res*. 44, W00B09, doi:10.1029/2007WR006720.

Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M., Higdon D., 2009, Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Journal of Nonlinear Sciences and Numerical Simulation* 10(3), 273-290.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Schoups, G., 2012. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water resources*, article in press.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., Xu, C.-Y., 2011. Calibration of hydrological models using flow-duration curves. *Hydrol. Earth Syst. Sci*, 15, 2205-2227.

Wigmosta, M.S., Vail L.W., Lettenmaier, D.P., 1994. A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res*. 30 (6), 1665-1679.

Wigmosta, M.S., Lettenmaier, D.P., 1999. A comparison of simplified methods for routing topographically driven subsurface flow. *Water Resour. Res*. 35 (1), 255-264.

Wittenberg, H., Sivapalan, M., 1999. Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. *J. Hydrol*. 219, 20-33.

Wood, E.F., Sivapalan, M., Beven, K.J., B and, L., 1988. Effects of spatial variability and scale with implications to hydrologic modeling. *J. Hydrol.* 102, 29-47.

Wood, E.F., Sivapalan, M., Beven, K., 1990. Similarity and scale in catchment storm response. *Rev. Geophys*. 28 (1), 1-18.

Wood, E.F., Lettenmaier, D.P., Zartarian, V.G., 1992. A Land-Surface Hydrology Parameterization with subgrid variability for General Circulation Models. *J. Geophys. Res*. 97, 2717-2728.

Zhao, R.J., Zuang, Y., Fang, L.R., Lin, X.R., Zhang, Q.S., 1980. The Xinanjiang model, in Hydrological Forecasting (*Proc. Oxford Symp*. April 1980), 351-356, IAHS Publ.

Zhao, R.-J., 1992. The Xinanjiang model applied in China. *J. Hydrol*. 135, 371-381.

Zięba, A., 2010. Effective number of observations and unbiased estimators of variance for autocorrelated data an overview. *Metrol. Meas. Syst*. XVII (1), 3-16.

Appendix A: Further details on the PDM

The actual storage S [L] is the sum of the unsaturated (S_{US}) and saturated (S_S) portions (Figure 2a):

$$S_{US}(t+\Delta t) = \{1 - F(c^*(t+\Delta t))\}c^*(t+\Delta t)$$

$$S_S(t+\Delta t) = F(c^*(t+\Delta t))c^*(t+\Delta t) - \int_{c_{\min}}^{c^*(t+\Delta t)} F(c)dc \tag{A.1}$$

$$S(t+\Delta t) = c^*(t+\Delta t) - \int_{c_{\min}}^{c^*(t+\Delta t)} F(c)dc = \int_{c_{\min}}^{c^*(t+\Delta t)} \{1 - F(c)\}dc = \int_{c_{\min}}^{c^*(t+\Delta t)} \left(1 - c_n{}^a\right)^b dc$$

The total actual storage, S_T [L] for the grid cell is:

$$S_T(t+\Delta t) = c_{\min} + S(t+\Delta t), \tag{A.2}$$

where $F(c^*(t))$ = probability ($c \leq c^*(t)$) indicates the fraction of grid cell with local storage capacity less than or equal to $c^*(t)$ and is saturated to generate runoff at time t (Figure 2a and b). The c is the local storage capacity, c_n is the normalized storage capacity, $c_{\min}$ is the minimum local storage capacity, and 'a' and 'b' are the shape parameters. Based on the 'equal storage redistribution of interacting storage elements' assumption, $c^*(t)$ is the critical store capacity at which all stores have water content of c^*, irrespective of their capacity, unless this is less than c^* when they will be full at time t (Moore, 1985). The maximum possible storage at saturation ($S_{\max}$ [L]) and the total maximum possible storage at saturation ($S_{T\max}$ [L]) for the grid cell are:

$$S_{\max} = \int_{c_{\min}}^{c_{\max}} \{1 - F(c)\}dc = \int_{c_{\min}}^{c_{\max}} \left(1 - c_n{}^a\right)^b dc \text{ and } S_{T\max} = c_{\min} + S_{\max} \tag{A.3}$$

The analytical solutions for the Pareto distribution are as below:

$$F(c) = 1 - \left(1 - \frac{c - c_{\min}}{c_{\max} - c_{\min}}\right)^b = 1 - \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}}\right)^b$$

$$\frac{c - c_{\min}}{c_{\max} - c_{\min}} = c_n\,[0,1]$$

$$f(c) = \frac{dF(c)}{dc} = \left(\frac{b}{c_{\max} - c_{\min}}\right)\left(\frac{c_{\max} - c}{c_{\max} - c_{\min}}\right)^{b-1}$$

$$c = c_{\max} - \left\{(c_{\max} - c_{\min})[1 - F(c)]^{\frac{1}{b}}\right\}$$

$$S_{T\max} = c_{\min} + \int_{c_{\min}}^{c_{\max}} 1 - F(c)dc = c_{\min} + \int_{c_{\min}}^{c_{\max}} \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right)^b dc = \frac{bc_{\min} + c_{\max}}{b+1}$$

$$S_{\max} = S_{T\max} - c_{\min} = \frac{c_{\max} - c_{\min}}{b+1}$$

$$S_T(t) = c_{\min} + \int_{c_{\min}}^{c^*(t)} \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right)^b dc = S_{T\max} \left\{ 1 - \left(\frac{c_{\max} - c^*(t)}{c_{\max} - c_{\min}} \right)^{b+1} \right\}$$

$$S(t) = \int_{c_{\min}}^{c^*(t)} \left(\frac{c_{\max} - c}{c_{\max} - c_{\min}} \right)^b dc = S_{\max} \left\{ 1 - \left(\frac{c_{\max} - c^*(t)}{c_{\max} - c_{\min}} \right)^{b+1} \right\}$$

$$c^*(t) = c_{\max} - \left\{ (c_{\max} - c_{\min}) \left[1 - \frac{S(t)}{S_{\max}} \right]^{\frac{1}{b+1}} \right\}$$

$$R(t) = TOSTORAGE - (S_{\max} - S(t)) + S_{\max} \left\{ \left[1 - \frac{S(t)}{S_{\max}} \right]^{\frac{1}{b+1}} - \frac{TOSTORAGE}{(b+1)S_{\max}} \right\}^{b+1} ; S(t+\Delta t) < S_{\max}$$

$$R(t) = TOSTORAGE - (S_{\max} - S(t)); S(t+\Delta t) \geq S_{\max}$$

Tables

Table 1. Main characteristics of the study catchments and hydro-climatic data.

Catchments	Gaulfoss	Eggafoss	Hugdal bru	Lillebudal bru
Catchment area, km^2	3090	668	546	168
Major soil types (%)				
Till soils (thick layer)	26.9	39.9	23.2	14.5
Till soils (thin layer)	33.0	27.8	38.7	29.5
Peat and marsh (organic material)	8.6	6.8	8.1	2.9
Bare mountains	24.0	23.9	20.1	49.1
Major types of bed rock geology (%)				
Gabbro and amphibolite	2.4	6.8	0.0	0.0
Amphibolite and schist	4.4	21.2	0.0	0.0
Greenstone and amphibolite	11.0	22.4	11.8	0.0
Quartzite	9.7	0.0	15.7	1.2
Mica gneiss, schist, amphibolite and metasandstone	48.2	8.7	41.3	97.7
Phyllite and schist	17.8	31.1	26.9	0.0
Elevations at climate stations	127-885	127-885	127-885	127-885
Elevations at streamflow stations	45	135	330	515
Mean annual precipitation at stations [mm]	670-1105	670-1105	670-1105	670-1105
Catchment averaged (interpolated) annual precipitation [mm]	874.21	922.6	884.06	864.7

Table 2. Summary of the key features of the parameterization cases.

Cases	Scales of representation of spatial heterogeneity and computation or calibration				
	Heterogeneity by a probability distribution	S_{max}	Shape parameter 'b' of the distribution	Runoff generation (response)	STS runoff routing
1	Subcatchment	Catchment (calibrated)	Catchment (calibrated)	Element	Grid
2	Subelement	Element (from topographic gradient)	Element (from topographic gradient)	Element	Grid
3	-	Catchment (calibrated)	-	Element	Grid
1G	Subgrid	Catchment (calibrated)	Catchment (calibrated)	Grid	Grid
2G	Subgrid	Grid (from topographic gradient)	Grid (from topographic gradient)	Grid	Grid
3G	-	Catchment (calibrated)	-	Grid	Grid

S_{max}: the maximum subsurface storage capacity and STS: source-to-sink routing.

Table 3. Lists of the calibrated model parameters and their uniform priors.

No.	Parameters	Routines	Cases	Uniform priors	Description
1	c_{max} [mm]	PDM	1-2, 1G-2G	150-1000	Local max. storage
2	c_{min} [mm]	PDM	1-3, 1G-3G	0-100	Local min. storage
3	S_{max} [mm]	PDM	3, 3G	150-1000	Catchment scale max. storage
4	$Storage_{min}$ [mm]	PDM	2G	0.001-20	Min. storage capacity
5	b [-]	PDM	1, 1G	0.001-1	Shape parameter of distn.
6	k [$mm^{1-n}h^{-1}$]	PDM	1-3	10^{-7}-10^{-3}	Coefficient of Q-S r/shp
7	n [-]	PDM	1-3	0.2-5	Exponent of Q-S r/shp
8	ψ [m/s]	PDM	1G-3G	0.005-1.5	Diffusivity
9	Ic [mm/hr]	PDM	1-3, 1G-3G	0.1-40	Infiltration capacity
10	TX [^{0}C]	Snow	1-3, 1G-3G	-3-2	Snowfall-rainfall threshold temp.
11	WS [-]	Snow	1-3, 1G-3G	1-6	Snowmelt sensitivity to wind speed
12	V [m/s]	Routing	1-3, 1G-3G	1.9-2.6	Velocity of flow
13	D [m^2/s]	Routing	1-3, 1G-3G	200-1500	Flow dispersion coefficient

Table 4. Calibration, temporal and spatial validation of model parameters corresponding to maximum R^2 and R^2ln performance measures for Gaulfoss.

Calibrated catchment (calibration period)	Cases (calibration or validation period)	Parameter transferability in time and space (to internal catchments)							
		Gaulfoss		Eggafoss		Hugdal bru		Lillebudal bru	
		R^2	R^2ln	R^2	R^2ln	R^2	R^2ln	R^2	R^2ln
Gaulfoss (2008-2010)	1 (2008-2010)	**0.84**	**0.86**	0.75	0.78	-	-	0.58	-2.38
	1 (2010-2011)	0.78	0.90	0.64	0.85	0.80	0.87	-0.01	-3.79
	2 (2008-2010)	**0.82**	**0.81**	0.75	0.73	-	-	0.56	-6.11
	2 (2010-2011)	0.83	0.90	0.61	0.82	0.82	0.77	-0.01	-3.46
	3 (2008-2010)	**0.80**	**0.83**	0.74	0.77	-	-	0.30	0.34
	3 (2010-2011)	0.83	0.88	0.59	0.85	0.77	0.79	-0.03	-3.39
	1G (2008-2010)	**0.84**	**0.83**	0.73	0.76	-	-	0.57	-3.59
	1G (2010-2011)	0.85	0.87	0.62	0.82	0.81	0.78	-0.03	-4.03
	2G (2008-2010)	**0.84**	**0.85**	0.73	0.78	-	-	0.58	-3.85
	2G(2010-2011)	0.85	0.90	0.62	0.84	0.81	0.83	-0.03	-4.03
	3G (2008-2010)	**0.84**	**0.84**	0.73	0.75	-	-	0.58	-3.77
	3G (2010-2011)	0.84	0.89	0.64	0.84	0.81	0.79	-0.01	-3.86

*The bold values indicate results of the calibration period for Gaulfoss and the validation period is 2010-2011.

Figures

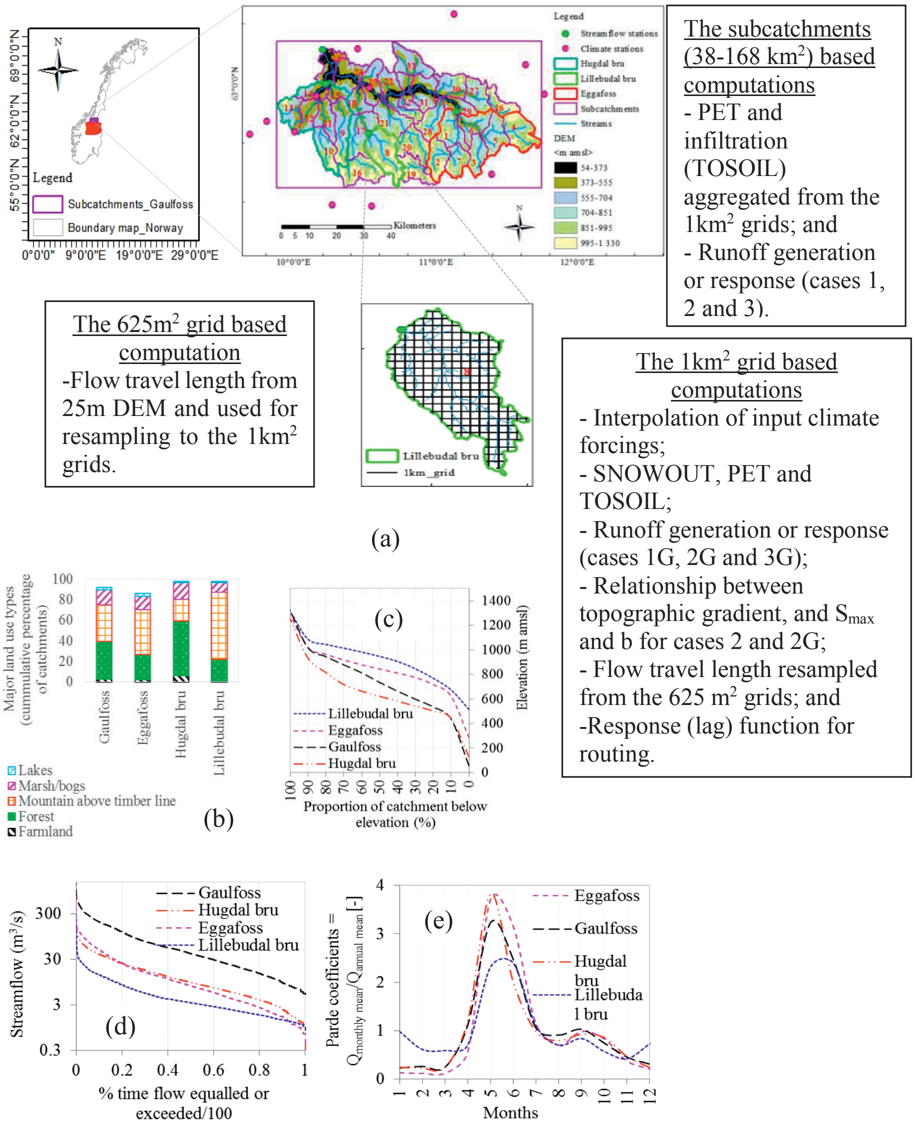
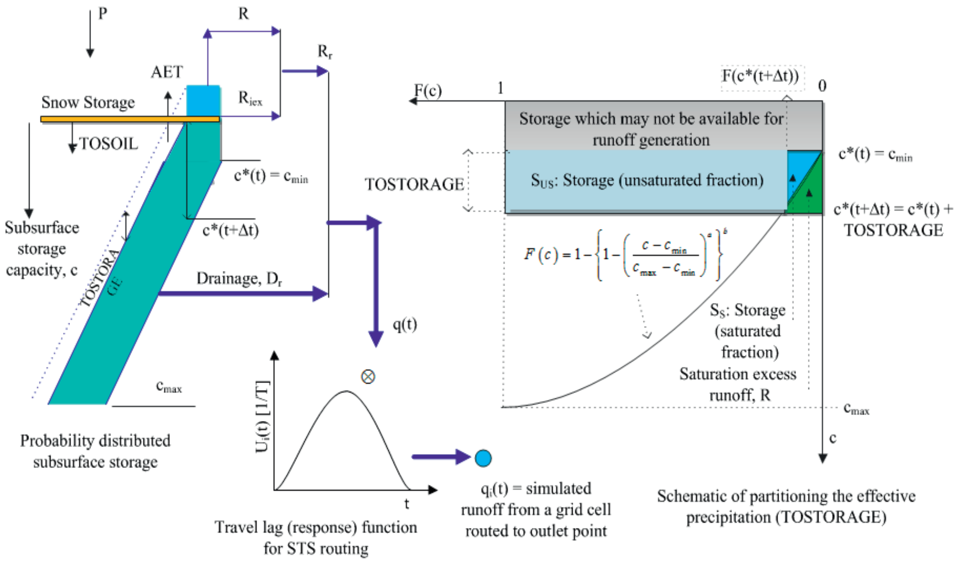
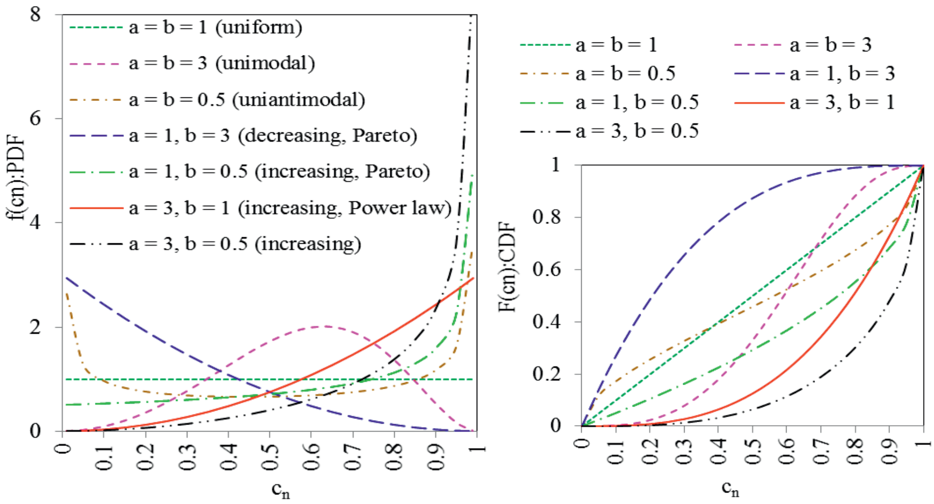


Figure 1. (a) Locations of the study catchments and gauging stations, elevation map, and different discretization schemes used for representation or simulation of the different variables as explained in the boxes, (b) Land use/cover cumulative percentages, (c) Hypsometric curves (d) Semi-log plot of daily streamflow duration curves, (e) Flow regimes or Pardé coefficients from daily streamflow data.

The numbers from 1-33 on Figure 1a (top right) are topographically discretized subcatchments called elements. The descriptions in the boxes in Figure 1a are the spatial scale used for the representation of the spatial heterogeneities and for the simulation of different variables.



(a)



(b)

Figure 2. (a) Probability distributed subsurface storage based model structure and schematic of partitioning effective precipitation (TOSTORAGE), (b) Probability density functions (PDFs-KwD) and cumulative distribution functions (CDFs-KwD).

At initial, $c^*(t) = c_{min}$ is considered.

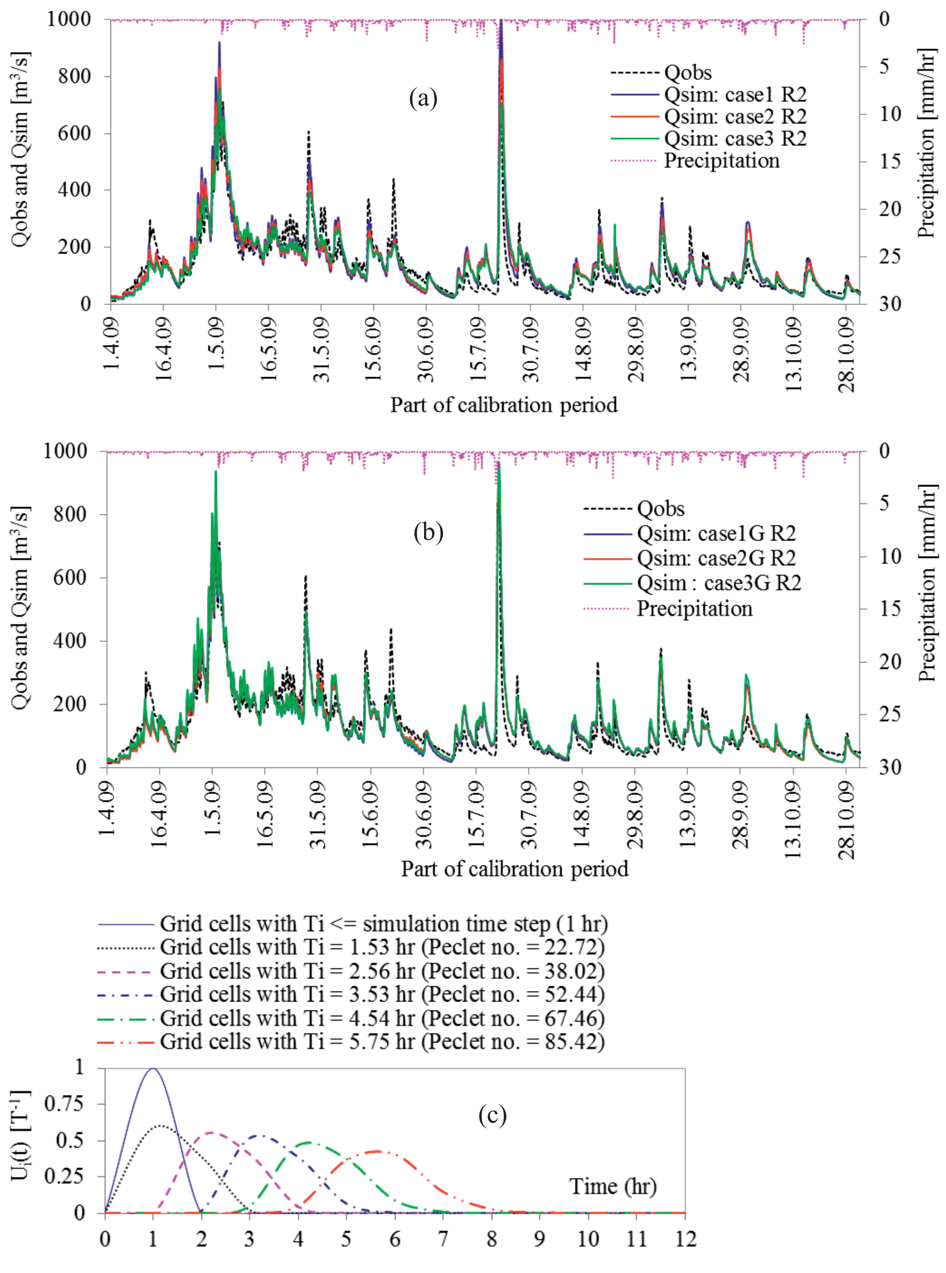


Figure 3. Hourly observed and simulated streamflow hydrographs for part of the calibration period corresponding to maximum R^2 for Gaulfoss: (a) Cases 1 to 3, (b) Cases 1G to 3G, and (c) Typical source-to-sink (STS) routing flow path response (lag) functions.

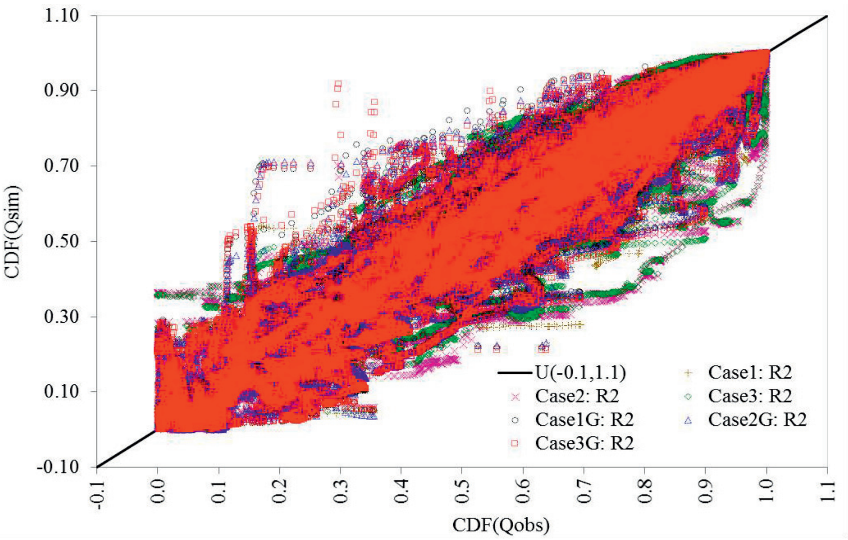


Figure 4. Quantile-quantile (Q-Q) plots for observed versus simulated flows for calibration and validation periods for Gaulfoss for R^2.

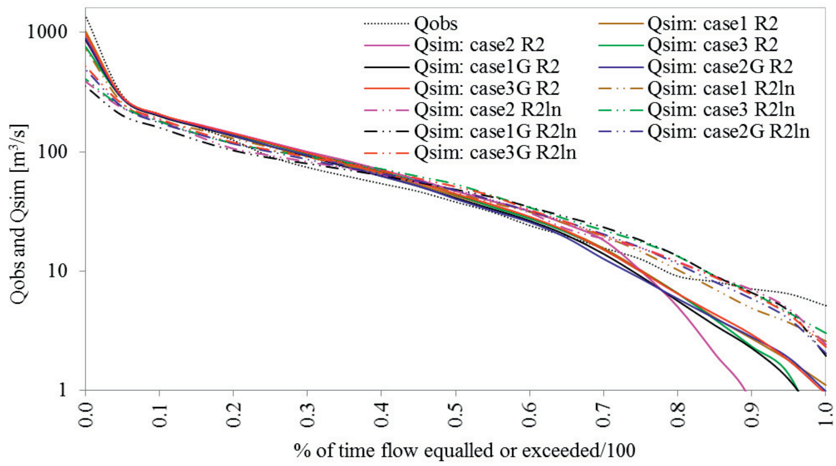


Figure 5. Flow duration curves for observed versus simulated flows for calibration and validation periods for R^2 and R^2ln performance measures for Gaulfoss.

Paper II

Paper 2 (P2)

Hailegeorgis, T. T. and Alfredsen, K. **Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments**

Hydrology Research (IWA Publishing), article in press, 2014.

http://www.iwaponline.com/nh/up/default.htm

The article is in press (uncorrected proof). Better quality figures for Figure 1 and Figure 3a&b are presented in the introduction to the thesis as Figure 1 and Figure 5a&b respectively.

Is not included due to copyright

Paper III

Paper 3 (P3)

Distributed hourly runoff computations in mountainous boreal catchments from 'catchments as simple dynamical systems' storage-discharge relationships

Teklu T. Hailegeorgis, Knut Alfredsen, Yisak S. Abdella and Sjur Kolberg

Distributed hourly runoff computations in mountainous boreal catchments from 'catchments as simple dynamical systems' storage-discharge relationships

Teklu T. Hailegeorgis[1], Knut Alfredsen[1], Yisak S. Abdella[2] and Sjur Kolberg[2]

[1]*Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology (NTNU), S.P. Andersens vei 5, NO-7491 Trondheim.*

[2]*SINTEF Energi AS, Sem Sælands vei 11, NO-7465 Trondheim.*

Correspondence to: Teklu T. Hailegeorgis (teklu.hailegeorgis@ntnu.no), Tel.: [+47] 73592411 and fax: [+47] 735 91298.

Abstract

We evaluated the 'top-down' runoff response routine based on the principles of 'catchments as simple dynamical systems' for distributed hourly runoff simulation for macroscale (3090 km^2) mountainous boreal catchment dominated by glacial till soils and considerable proportion of snowfall precipitation. The main research question lies in how the approach performs when we both estimate the parameters from streamflow recession and calibrated (for both runoff routed and unrouted cases) for a macroscale boreal catchment. The routine reproduced the streamflow hydrographs and duration curves with Nash-Sutcliffe efficiency (NSE) up to 0.83. Moreover, transferability of the parameters to internal mesoscale (168-653 km^2) subcatchments validated the 'top-down' paradigm and indicated an opportunity for prediction in ungauged subcatchments. However, both estimating the parameters from recession and neglecting runoff delay underestimate the peak flows slightly. In addition, the lower end of recession and the minimum length of recession segments included in the analysis affect the estimated parameters. Despite the parsimony, there are considerable parameter uncertainty and non-identifiability for both estimation and calibration. Therefore, detailed evaluation of reliability of the 'top-down' routine is necessary based on other observed variables not used for calibration. Study on combined utility of the 'top-down' and process based 'bottom-up' paradigms would be essential. However, the results encourage further comparative evaluations of the routine on large number of catchments.

Key words

Distributed hourly runoff simulation; Catchments as simple dynamical systems; Streamflow recession; Discharge sensitivity function; Parameter calibration and transferability; Source-to-sink routing

1 INTRODUCTION

Several studies indicated the uniqueness of watersheds in runoff response due to natural heterogeneities in catchment characteristics, climate forcing, dominant hydrological processes and process interactions. Uniqueness of watersheds has been studied in literature, e.g. Beven (2000) and McDonnell et al. (2007). To model catchment uniqueness, modelling of the dominant hydrological processes based on small-scale physics such as Darcy's and Richard's equations and upscaling of the theories to the catchment scale following the 'bottom-up' approach is common. However, Sivapalan et al. (2003) presented an overview of the 'top-down' and discussed the importance of the approach for parsimony and learning from observed data. The authors also suggested adopting the approach in a comparative mode for many catchments in different climatic and hydrologic conditions. McDonnell et al. (2007) suggested a combination of the 'bottom-up' and 'top-down' worldviews for descriptions of watershed function based on the conditions that constrain the watershed throughout its long-term evolution. Kirchner (2006) asserted the need for modelling based on analyses of contemporary hydrological observations to infer model structure, equations and parameters for improved hydrological modelling. The main motivation was to reduce relying on the traditional approach of parameter tuning to a calibration data set in overparameterized and poorly identified 'bottom-up' models.

Rainfall-runoff modelling based on storage-discharge relationships and streamflow recession analysis has long been conducted (e.g. Ambroise et al., 1996; Lamb and Beven, 1997; Wittenberg and Sivapalan, 1999 and Rees et al., 2004). Kirchner (2009) proposed a simplified 'top-down' approach for a lumped modelling study of headwater catchments in Mid Wales (United Kingdom) based on a functional relationship between catchment storage and discharge and parameter estimation from both streamflow recession analysis and direct calibration of the rainfall-runoff relationships. The author inferred a single storage model structure based on a variable known as the discharge sensitivity function, g(Q) = dQ/dS, where S is catchment storage and Q is discharge. If discharge is a function of storage, then the catchments antecedent moisture will be implicitly measured by stream discharge and the catchment response to a unit increase in storage will be directly quantified by the hydrologic sensitivity function (Kirchner, 2009). Parameter estimation from only streamflow observation would avoid the uncertainty of the quality of climate data on the parameter calibration, which may provide significant advantages in areas with a sparse climate network. Moreover, the computational demand of calibration, for instance calibration of a distributed model running at hourly time-step is high with well-known pitfalls. Additional expected advantages of such parsimonious precipitation-runoff models are reduction of problems related to overparameterization since one can estimate the parameters in the runoff response routine from the recession plots of observed streamflow at times when effects of evapotranspiration and precipitation are negligible.

The defining feature of the downward approach to hydrologic modelling is the attempt to predict overall catchment response and the catchment functioning based on an interpretation of the observed response at the catchment scale (Sivapalan et al., 2003). Therefore, an additional advantage of the approach proposed by Kirchner (2009) is the ability to make inference on the unobservable catchment storage from streamflow observations. We are particularly limited due to our inability to 'see' the subsurface of a catchment, in which much of the hydrologic response often remains hidden from our current measurement techniques (Wagener et al, 2007).

However, previous studies such as Myrabo (1997), Beldring (2002) and Jansson (2005) shows the dominance of glacial tills in the boreal catchments. Jansson et al. (2005) further reported preferential flow in glacial tills. Zehe and Sivapalan (2009) classified the preferential flow and connectivity of flow paths to the outlet as a runoff response threshold. The preferential flow violates one of the main assumptions of the runoff response routine evaluated in the present study, which is the hydraulic connectivity of unsaturated and saturated storages. Hence, the validity of the prevailing assumptions in the approach needs assessment across catchments in different climate regimes, landscape features and spatial scales (from micro to macro scales). Blöschl (2006) endorsed the need for hydrological synthesis across places and scales for improved knowledge based on comparative analysis. Kirchner (2009) also concluded with a statement on the importance of assessing the applicability of the approach to diverse hydrologic setting.

In addition, to our knowledge all the previous studies or applications of the approach were lumped. The main examples are Teuling et al. (2010) for streamflow simulation in Swiss prealpine catchment and Krier et al. (2012) for inferring basin-averaged effective precipitation rates for 24 small to mesoscale catchments in Luxembourg. More recently, Brauer et al. (2013) applied the predetermined power-law relation of Brutsaert and Nieber (1977) for a less humid lowland catchment (6.5 km^2) in the Netherlands. However, Kirchner (2009) noted that the storage-discharge relationship that characterizes the catchment's behavior might not describe any individual point on the landscape. Clark et al. (2009) illustrated that for a mountain research watershed in Georgia (USA) the recession relationships of (dQ/dt) and Q is approximately consistent with a linear reservoir at a hillslope scale and deviation from linearity that becomes progressively larger with increasing spatial scale. Spence et al. (2010) also illustrated a hysteretic relationship between storage and streamflow at a catchment scale for subarctic catchment with dominant shallow organic soils and numerous water bodies. Therefore, the performance of the distributed version of the approach needs further exploration.

Relevant to the catchment size and the effects of runoff delay, the potential applicability of the approach for runoff simulation in macroscale catchments is lacking in literature. Kirchner (2009) stated that the method must break down for catchments that are too large, but Krier et al. (2012) illustrated the validity of Kirchner's 'doing hydrology backward' approach for small to mesoscale catchments ($\leq$ 1092 km^2) which exhibit heterogeneous lithology. One can also speculate for large size watersheds that the spatial heterogeneity of precipitation and the effects of runoff routing in the hillslopes and river networks may influence the precipitation-runoff relationship more than the catchment storage and discharge relationships. Therefore, the effects of the runoff delay on calibration of parameters, the g(Q) and streamflow simulation require further study.

Overparameterization in precipitation-runoff models introduce more degrees of freedom than the data can adequately constrain, which leads to equifinality problem (Beven and Binley, 1992; Kirchner, 2006). Since precipitation-streamflow relationships can provide only limited information, uncertainty and identifiability problems of model parameters are not completely avoidable even in parsimonious models. However, the likelihood of reliability of prediction is highly influenced by the reliability of calibrated parameters. Therefore, the evaluation and assessment of uncertainties in estimation of parameters related to extraction of recession segments, interaction among the parameters and the model equation inferred from the recession analysis are required. In addition, the uncertainty and identifiability assessment of the calibrated parameters is necessary.

The objectives of the present study are:

(1) To evaluate the calibration and validation performances of a spatially distributed version of the 'top-down' approach of single catchment storage runoff response routine for hourly runoff simulation when parameters are both estimated from streamflow recession and calibrated to observed discharge;

(2) To assess the parameter uncertainty and identifiability for both parameter estimation and calibration; and

(3) To study the effects of parameter uncertainty and runoff delay on the g(Q).

2 THE STUDY REGION AND DATA

The study area is the Gaula watershed in mid Norway. We used streamflow data from Gaulfoss and its internal subcatchments Eggafoss, Hugdal bru and Lillebudal bru. The catchments exhibit boreal climate with seasonal snow. The dominant land covers are conifer forests, mountains above timberline and marsh land/bogs. Lillebudal has a mean catchment elevation of 915 masl and 65 % of the catchment area above the timberline, which is higher than the other catchments. There is no considerable proportion of farmland and lakes, and none of the catchments has glacier coverage. The dominant soil type is glacial tills. The underlain bedrocks are mainly metamorphic (mica gneiss, mica slate, phyllite, green stone, quartzite, mica schist and amphibolite), and metasandstone (Figure 1a). All climate input data are of hourly time resolution similar to the computational time step. The climate data used are precipitation (P) from 12 stations, temperature (T) from 11 stations, wind speed (W_s) from 9 stations, and relative humidity (H_R) and global radiation (R_G) from 3 stations. The spatial resolution for the climate measurement is at a very coarse scale compared to the 1x1 km^2 computational grid size. Spatial fields of the climate data are computed by inverse distance weighing (IDW). A catchment map and characteristics of the study catchments are given in Figure1a and Table 1.

3 METHODS AND MODELS

3.1 Kirchner's runoff response routine

Kirchner's method (Kirchner, 2009) was developed based on the main assumptions that:

(i) The streamflow Q depends solely on the amount of water stored in the catchment (S) and f is a strictly monotonically increasing function (i.e. invertible);

$$Q = f(S);\ S = f^{-1}(Q) \tag{1}$$

(ii) The discharge in the catchment is mainly controlled by release of water from the storage rather than 'bypassing' flow from direct precipitation; and

(iii) The unsaturated and saturated storages are hydraulically connected and the net groundwater flow across watershed boundary (G_{in}-G_{out}) is zero.

From conservation of mass equation:

$$\frac{dS}{dt} = I - AET - Q + \left(G_{in} - G_{out}\right) = \left(I - AET - Q\right) \tag{2}$$

The water balance based response routine is:

$$\frac{dQ}{dt} = \frac{dQ}{dS}\frac{dS}{dt} = \frac{dQ}{dS}\left(I - AET - Q\right) = g\left(Q\right)\left(I - AET - Q\right), \tag{3}$$

where the actual evapotranspiration (AET), infiltration (I) = rainfall + snow melt (SM) and discharge (Q) are given in mm/hr, storage (S) is in mm, t is a time variable and g(Q) is the sensitivity of discharge to change in storage. The reciprocal of the sensitivity function or 1/g(Q) is system 'response time' or 'memory' (Teuling et al, 2010) usually denoted as ԏ (Tau) and indicates how rapidly streamflow recedes.

Based on the lumped water balance model of Kirchner (2009), I, AET, Q and S in the above equations are lumped for the catchment scale. But, the differences in the usage of eq. (1-3) in the present study is that we simulated distributed runoff for each grid cell by considering spatially distributed climate inputs, fluxes and storage. The grid based computations in this study accounts for the spatial variability of climate forcing. Since the seasonal snow accumulation and snowmelt highly influence the water balance of the catchment, spatial variability of temperature based on elevation is also important.

3.1.1 Estimation of the regression parameters and g(Q) from streamflow recession

Kirchner (2009) inferred the model structure and parameters from observed streamflow during recession periods. The recession curve describes in an integrated manner how different factors in a catchment influence the generation of streamflow in dry weather periods (Tallaksen, 1995). Recession plots provide information on how the rate of streamflow recession (-dQ/dt) varies with discharge (Q) when effects of evapotranspiration and precipitation or infiltration are assumed to be negligible and hence eq. (3) for dQ/dt can be reduced to eq. (4). The main advantages of recession analysis are that rainfall can be assumed to be zero, or at least small (so difficulties with any errors in catchment rainfall estimation are avoided), and that the hydrograph represents an aggregate measure of catchment behavior (Sivapalan, et al., 2003).

We followed the refined recession analysis (extraction and binning) procedures by Kirchner (2009). We used hourly data from 1995–2011 for Gaulfoss, Eggafoss and Hugdal bru and 2004–2011 for Lillebudal bru. We extracted only night-time recessions to filter out periods with significant effect of evapotranspiration but we did not exclude periods with precipitation due to the lack of long hourly series of precipitation data. From the recession plots of observed streamflow (Figure 2), we inferred a second order polynomial fit between ln(-dQ/dt) and ln(Q):

$$\frac{dQ}{dt} \approx g(Q)(-Q)\Big|_{I<<Q,AET<<Q} \tag{4}$$

As already stated, the 'top-down' paradigm provides an opportunity to infer model equation and structure from the streamflow observations. Rearranging for g(Q) in eq. (4) with log transformation for numerical stability following Kirchner (2009), the following polynomial regression based storage-discharge relationship was fitted from the recession analysis:

$$\ln\left(g(Q)\right) \approx \ln\left(\frac{dQ}{dS}\right) \approx \ln\left(\frac{-dQ/dt}{Q}\Big|_{P<<Q,AET<<Q}\right) \approx \beta_0 + \beta_1 \ln(Q) + \beta_2 \left(\ln(Q)\right)^2, \tag{5}$$

where β_0, β_1 and β_2 are parameters of the polynomial regression model. The rate of flow recession (-dQ/dt) is computed as differences in discharge between two successive hours and the discharge Q is computed as average discharge over the two hours following Brutsaert and Nieber (1977) and Kirchner (2009).

However, the validity of the results depends on the adequacy of the fitted regression model. Hence, we tested the adequacy of the selected polynomial regression model. We diagnosed the multicollinearity, significance of the regression model and parameters, key features of residuals and parameter uncertainty based on the following regression model and ordinary least square or minimizing the standard error of estimates (SSE):

$$\ln\left(g\left(Q\right)\right) = \beta_0 + \beta_1 \ln\left(Q\right) + \beta_2 \left(\ln\left(Q\right)\right)^2 + \varepsilon$$
$$\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)^T = \underset{\beta_0, \beta_1, \beta_2}{\arg\min} \sum_{i=1}^{n_b} \left\{ \ln\left(g\left(Q_i\right)\right) - \left(\beta_0 + \beta_1 \ln\left(Q_i\right) + \beta_2 \left(\ln\left(Q_i\right)\right)^2\right)\right\}^2, \tag{6}$$

where $\mathcal{E}$ is an error term, Q_i represents bin-averaged discharges, n_b is the number of bin-averaged discharges.

Though a polynomial regression with only two parameters obtained by setting the quadratic term $\beta_2 = 0$ may reduce problems related to correlation among the parameters, the regression model may not remain significant due to the lack of fit because of the missing quadratic term. We performed significance test for the regression parameters and the regression model by the *t*-test and *F*-test. We diagnosed the residuals for the normality assumption of the linear regression model by the *Z*-score test, which is the inverse of the standard normal distribution corresponding to a probability (p_r) from ranking of the residuals $p_r = (i\text{-}0.5)/N$, where i is the ranks of the residuals in ascending order and N is the number of samples. We carried out residuals diagnosis for homoscedasticity, correlation, systematic lack of fit and outliers from plots of the estimates of the response variable ln(g(Q)) versus the residuals, which should be random plots around an expected value of zero.

We estimated the Individual Confidence Levels (ICL) for the parameters from the *t*-test. To assess identifiability of regression parameters through their joint confidence regions, we wanted to compute the Joint Confidence Region (JCR), which simultaneously bounds the joint parameters, to assess the effects of parameters correlation or interaction based on elliptical confidence regions (Bard, 1974). Elliptical joint confidence regions are better predictors of regression model uncertainty because they capture the parameter correlation (Rooney and Biegler, 2001). From the multivariate normal distribution of regression parameters given in Appendix A, the sum of squares function in the exponent term of eq. (A1) is an equation of a hyper-ellipse centred at the parameter estimates. The joint confidence region is all the points in the ellipsoid region and computed from the *F*-distribution as:

$$\frac{\left(\underline{\theta}-\hat{\underline{\theta}}\right)^{T}\underline{C}^{-1}\left(\underline{\theta}-\hat{\underline{\theta}}\right)}{p'S^{2}}\leq f_{p',n-p,\alpha}, \qquad (7)$$

where $\underline{\theta}$ is the subset of $\underline{\beta}$, $\underline{C}$ is the part of the $(X^TX)^{-1}$ matrix, which is corresponding to the parameters for which the joint confidence region is to be constructed, p' < p is the dimension of the parameters for which the joint confidence region is to be constructed. In the present study p' = 2 since the joint confidence regions of two parameters at a time are computed. The α is significance level, S^2 is estimated error variance = SSE/N-p and p denotes the number of parameters. Equation (7) is exact for the linear regression model (Rooney and Biegler, 2001; Vurgin et al, 2007).

For estimation of g(Q) from the recession analysis, we estimated the parameters based on bin-averaged discharges extracted from streamflow recessions. Therefore, we also employed an alternative parameter calibration. However, we calibrated the model parameters as effective parameters for the catchment. Hydrological models are often lumped spatially with calibrated effective parameters, which are assumed to take into account all of the local scale heterogeneity of land surface characteristics, meteorological variables and hydrological processes and fluxes for large areas (Gottschalk et al., 2001; Beldring et al., 2003). In addition, Pokhrel et al. (2008) and Pokhrel and Gupta (2011) illustrated the limitations of making inferences on the spatial properties of a distributed model when only information about catchment output stream response is available.

3.1.2 Model parameters and g(Q) from direct calibration

Direct calibrations based on optimization algorithms involve comparisons of simulated streamflow with the observed streamflow at the outlet. In this case, we calibrated the response routine parameters in eq. (5) based on precipitation-streamflow relationships to compute the g(Q) and simulate the streamflow. We used the Differential Evolution Adaptive Metropolis algorithm or DREAM (Vrugt et al., 2009) with residual based log-likelihood objective function implemented in ENKI hydrological modelling framework (Kolberg and Bruland, 2012):

$$\log\left\{l\left(\delta/\sigma_{\varepsilon}^{2},\sum_{i=1}^{n}\left(Qsim_{i}^{(\theta)}-Qobs_{i}^{(\theta)}\right)^{2}\right)\right\}=\left(\frac{-n}{2}\log(2\pi)-\frac{n}{2}\log\left(\sigma_{\varepsilon}^{2}\right)-\frac{\sum_{i=1}^{n}\left(Qsim_{i}^{(\theta)}-Qobs_{i}^{(\theta)}\right)^{2}}{2\sigma_{\varepsilon}^{2}}\right)\times f, \quad (8)$$

where $Qsim^{(\theta)}$ and $Qobs^{(\theta)}$ respectively are Box-Cox (Box and Cox, 1964) transformed simulated and observed streamflow series of length n, δ denotes model parameter, θ is the Box-Cox transformation parameter and σ_{ε}^2 is variance of error. We computed the θ from observed streamflow records based on the 'fminsearch' algorithm in matlab, which finds the θ value that maximizes a log-likelihood function (http://www.mathworks). The *f* is a fraction of effectively independent observations estimated from the autoregressive or AR (1) model of error covariance (Zięba, 2010).

We assessed the uncertainty and identifiability of the calibrated parameters from the last 50 % of marginal posterior parameters. The calibrated parameters in the response routine are β_0, β_1, β_2 and EvR, where EvR represents a discharge at which AET equals 0.95*PET (Figure 1(b)). Runoff is computed by integrating eq. (3) in time using an adaptive Bogacki-Shampine (Bogacki and Shampine, 1989) numerical method, which is implemented in ENKI (Kolberg and Bruland, 2012). We used the observed discharge [mm/h] before the start of simulation period as an initial state for all the grid cells to infer initial storages. We started the simulation in September and provided 'burn-in' period before the calibration period to reduce the effects of initial snow state.

Despite the log-likelihood objective function for the DREAM calibration, we used the parameter set yielding maximum NSE (max NSE) performance measure for further analyses since it is a suitable scale for comparison of streamflow hydrographs. We evaluated the simulation based on the two common runoff signatures of hydrographs and flow duration curves. Hydrographs are catchment-integrated signatures explaining how catchments respond to climate forcing and its own states. Flow duration curves express the temporal variability of flow in terms of the percentage of time a flow of a certain magnitude is available within a year. We also tested the spatial transferability (validation) of both the estimated and calibrated parameters of the Gaulfoss catchment (3090 km^2) to internal subcatchments of Eggafoss (653 km^2), Hugdal bru (549 km^2) and Lillebudal bru (168 km^2).

3.2 Snow routine

The snow processes are dominant in the study area during winter and spring seasons. The snow routine is based on the gamma distributed snow depletion curve or SDC (Kolberg and Gottschalk, 2006), which computes the outflow melt water release from saturated snow (Q_s). The free parameters in the routine are rainfall-snowfall threshold temperature (TX) and snowmelt sensitivity to wind speed (WS).

3.2 Evapotranspiration routine

We computed the potential evapotranspiration (PET) by the Priestley Taylor method:

$$PET = \alpha \frac{\Delta}{\Delta + \gamma}\left(R_n\right), \tag{9}$$

where α is the Priestley Taylor constant = 1.26 (Priestley and Taylor. 1972), Δ is the slope of saturation vapor pressure curve at air temperature at 2m (T2m), ϒ is the psychometric constant (0.67 $hPaK^{-1}$), R_n is net radiation. The AET is computed from the PET and streamflow, which is used as a proxy to indicate the soil-moisture state according to the equation given in Figure 1(b). There is no free parameter in this routine.

3.3 Runoff routing routine

We linked the response function based source-to-sink (STS) routing (Naden, 1992; Olivera, 1996) to the runoff response routine to account for the effects of runoff delay both in hillslopes and river networks. Travel time lag influences the hydrologic behavior of large basins. The runoff response at the outlet for runoff signal at each grid cell *i* is given by the response function ($U_i(t)$ $[T^{-1}]$):

$$U_i(t) = \frac{1}{2t\sqrt{\pi\left(\frac{t}{T_i}\right)/\Pi_i}} \exp\left\{-\frac{\left[1-\left(\frac{t}{T_i}\right)\right]^2}{4\left(\frac{t}{T_i}\right)/\Pi_i}\right\}, \tag{10}$$

where Π_i [-] = $\Sigma(l_iV_i/D_i)$ is the flow path Peclet number, T_i is the expected flow travel time to the outlet for grid cell *i*, l_i is flow travel length in grid cell *i*, D_i (flow dispersion coefficient) and V_i (velocity of flow) are effective calibrated parameters representing all the grid cells. We performed the runoff routing by a convolution following Maidment et al. (1996):

$$Q_{sim}(t) = \left\{ \frac{\sum_i A_i Q_{gi} \otimes U_i(t)}{A} \right\}, \qquad (11)$$

where Q_{sim} [L/T] is catchment averaged routed simulated flow at the time step t, A_i [L^2] is area of grid cell, A [L^2] is catchment area, Q_{gi} [L/T] is average runoff (over time step) generated at each grid cell *i* and $\otimes$ is the convolution operator.

We presented the model structure and equations in Figure 1(b). We conducted calibration of parameters with and without including the runoff routing.

4 RESULTS

4.1 Recession plots and estimated g(Q)

Kirchner (2009) discussed the importance of measuring g(Q) across nested networks to understand how storage-discharge relationships vary across the landscape. For the four catchments in the present study, flow recession rates and recession plots or recession relationships fitted to bin-averaged discharges with their corresponding polynomial regression equations are given in Figure 2(a) and (b). The rate of streamflow recession (-dQ/dt) ranges from 0.0000-0.031 mm/hr^2 for Gaulfoss, 0.000055-0.0359 mm/hr^2 for Eggafoss, 0.0000-0.0369 mm/hr^2 for Hugdal bru and 0.00022-0.259 mm/hr^2 for Lillebudal bru. The 'response time' ʊ ranges 16-237 h for Gaulfoss, 19-145 h for Eggafoss, 22-126 h for Hugdal bru and 15-131 h for Lillebudal bru. The corresponding bin-averaged discharges (Q) range from 0.0032-0.698 mm/h for Gaulfoss, 0.0020-0.744 mm/h for Eggafoss, 0.00298-0.948 mm/h for Hugdal bru and 0.0189-0.942 mm/h for Lillebudal bru. Response time (τ) is dependent on catchment size or the effects of runoff delay. The larger Gaulfoss catchment exhibits slow response time while the smaller Hugdal bru and Lillebudal bru catchments exhibit fast response time. Catchments with slow recession rate are typically groundwater dominated, while impermeable catchments with little storage show faster recession rates (Staudinger et al., 2011).

4.2 Hydrographs and flow duration curves

We presented the simulated versus the observed streamflow hydrographs of Gaulfoss for calibration and validation periods in Figure 3a-c. Figure 3a corresponds to the regression parameters estimated from recession (runoff routed), Figure 3b corresponds to parameters calibrated (runoff routed) while Figure 3c corresponds to parameters calibrated (runoff unrouted). Parameter estimation from recession and calibration resulted in NSE up to 0.77 and 0.82 respectively and the model reproduced the hydrographs for the different seasons for Gaulfoss (Table 5). In addition, the spatial transferred parameters to internal subcatchments (Eggafoss, Hugdal bru and Lillebudal bru) based on the traditional 'split sample' and 'proxy ungauged basin' tests (Klemeś, 1986) provide NSE up to 0.81 and 0.83 for parameter estimation and calibration respectively (Table 5). This also shows that both the estimated g(Q) from recession segments of 17 years and from calibration based on continuous streamflow observations of two years provide representative parameters to capture seasonal variations of streamflow hydrographs. However, both parameter estimation from recession analysis and neglecting the runoff routing slightly underestimates the peak flows compared to simulation based on direct calibration and runoff-routed. Figure (5a) displays the plots of the observed versus the simulated flow duration curves from combined calibration and validation periods. The model reproduced the temporal variability of streamflow in terms of the flow duration curve.

4.3 Parameter uncertainty and identifiability

Table 2 shows the lists of calibrated parameters along with their ranges of uniform prior whereas Table 3 gives the values of the parameters estimated and calibrated both for runoff routed and unrouted cases corresponding to the max NSE performance measure for Gaulfoss. Figure 4(a) and 4(b) respectively show typical results from diagnostics of the polynomial regression and uncertainty bounds of regression parameters for parameter estimation from streamflow recession. Figure 4(c) presents the uncertainty of the calibrated response routine parameters (runoff routed) in terms of histograms of the marginal posterior distributions from the DREAM calibration. The parameters β_0 and β_1 exhibit wider posterior distributions (large uncertainty) compared to the β_2 and EvR.

Diagnostics of the fitted second order polynomial regression of the recession analysis revealed the adequacy of the model. The parameters and the model are significant, and normality and randomness of the residuals comply with the key assumptions in the regression model. However, the residual plots indicating the systematic lack of fit indicated that the regression model appeared to be insignificant for some of the catchments when there is no quadratic term.

The rectangular region created by individual 95% confidence limits based on the *t*-test indicates wide uncertainty ranges. We performed tests on whether it is necessary to consider the joint confidence regions to account for the correlation among the regression parameters for Gaulfoss. It was observed that the majority of ellipsoid joint confidence regions lie inside the rectangular individual confidence limits only for β_0 and β_1 (Figure 4(b)) which indicates that considering joint confidence region is not necessary for the two parameters. However, the elliptical joint confidences regions involving the quadratic parameter β_2 (not shown here) are far off their corresponding rectangular individual confidence limits, which indicate the poor identifiability due to correlation between the parameters and suggest the removal of β_2 from the regression model.

However, significance test based on residuals analyses (Figure 4(a)) revealed that the regression model is not adequate for some cases without the quadratic term as discussed earlier. Despite these two contradicting results, we preferred to use the second order polynomial regression with the three parameters (constant, linear and quadratic) of eq. (5) in the present study. Even though the optimal regression parameters obtained from the recession analysis and direct calibration are different (Table 3), their uncertainty bounds somehow overlap, which can be observed from comparing the confidence limits or regions (Figure 4b) versus histograms of posterior parameters for Gaulfoss (Figure 4c).

Table 4 contains the results of parameter correlation in terms of the linear correlation matrix and ranges of posterior parameters for Gaulfoss. The large linear correlations among the response routine parameters for the direct calibration indicate poor identifiability of parameters. We presented the effects of parameter estimation and calibration and the runoff delay on the g(Q) along with ensemble mean of g(Q), which is computed from posterior parameter sets from the calibration, in Figure 5(b) to address the effects of parameter uncertainty on the g(Q).

4.4 Parameters transferability or model validation

The transferability of model parameters from Gaulfoss (3090 km^2) to its three internal subcatchments namely Eggafoss (653 km^2), Hugdal bru (549 km^2) and Lillebudal bru (168 km^2) indicated the validity of the 'top-down' modelling paradigm. However, the performance of parameter transfer to Lillebudal is lower than that of the others. In addition, the hypsometric curves (Figure 1a) and mean catchments elevation (Table 1) indicate that the major portion of Lillebudal bru is located at higher elevation than the climate stations, which indicates the unrepresentativeness of the climate input for the catchment. The landuse map (Figure 1a) and Table 1 show higher proportion of mountains above timberline (bare rock) for the Lillebudal bru. Contingent on availability of dense streamflow and climate gauging networks, it would be possible to test how the parameter transferability works down to the 1x1 km^2 computational grid.

5 DISCUSSION

Motivated by Kirchner (2009), we evaluated the applicability of the principle of 'catchments as simple dynamical systems' for macroscale to mesoscale (3090-168 km^2) mountainous boreal catchments in mid Norway for distributed hourly runoff simulation.

Model calibration and validation

Both parameter estimation from recession analysis and neglecting the runoff routing slightly underestimates the peak flows compared to simulation based on direct calibration (runoff-routed case). This result do not comply with that of Brauer et al. (2013), who found that for a less-humid catchment in the Netherlands parameter calibration from direct storage-discharge fitting of the power-law relationship led to a strong underestimation of the response of runoff to rainfall while recession analysis lead to an overestimation. The differences may also arise from the difference between the polynomial relationship derived from the recession analysis and/or the power-law relationship, the differences in calibration algorithms used and the runoff routing.

Transferability of model parameters indicated spatial validation of the model and an opportunity for regionalization of the parameters towards prediction in ungauged rivers in the catchment. If g(Q) can be estimated from some combination of catchment characteristics, then it may help in solving the problem of hydrologic prediction in ungauged basins (Kirchner, 2009). However, previous attempt by Krakauer and Temimi (2011) to identify first order controls of recession time scale indicated that the used predictor variables were significant at only high flow or low flow rates. Moreover, observations on the geological characteristics of the catchments, which influences the recession behaviors, are not easily available and there are limitations associated with the data mining or spatial analysis methods. Further work in this regard on large number of catchments would be interesting. Nevertheless, the temporal and spatial transferability of the estimated and calibrated model parameters are promising for further evaluation of the routine on large number of catchments.

The results from the present study indicate that the principle of 'catchments as simple dynamical systems' in which the streamflow is assumed to be mainly controlled by the release of water from the storage allows us to simulate the runoff responses in the study region. Though the occurrence of preferential flow in glacial till soil was reported for instance by Jannson et al. (2005), the present study shows streamflow simulation from the method proposed by Kirchner (2009) that is based on the main assumption of hydraulic connectivity of storages and flow pathways provided validated results. Graham et al. (2010) and Graham and McDonnell (2010) reported the existence of connected preferential flow paths at the soil-bedrock interface in their study on hillslope threshold runoff responses to rainfall. However, both estimation of parameters from streamflow recession and calibrated effective parameters could represent effective characteristics of the heterogeneous catchment system (see Beven et al., 2000; Wagener and Wheater, 2006).

Effects of parameter uncertainty on the g(Q)

Parameter uncertainty affects the observed discharge sensitivity to storage or g(Q) and hence streamflow simulation. We found considerable differences between the results of the g(Q) that is computed from estimated and calibrated parameters for recession segments of low flows for Gaulfoss catchment (Figure 5(b)). From recession based inference, the expected behavior of g(Q) as an increasing function of Q or decreasing 'response time' with increasing discharge are observed only above certain limits of Q. This problem is attributed to the recession segments at the lower end of recession. Kirchner (2009) discussed the significant scatter at the lower end of recession particularly for $Q < 0.1$ mm/h and attributed it to the effects of measurement noises. As we can observe from the lower ranges of recession rates in (Figure 2(a)), there are equal recession rates over the ranges of bin-averaged discharges. The ln (-dQ/dt) versus ln(g(Q)) plots in Figure 2(b) also shows higher observed g(Q) for the lower ends of recession plots, which may not be related to fluctuations in catchment storage rather potentially related to resolution of loggers and errors in measurements of low winter flows. These figures suggest cutting of the lower end of recession below ln(-dQ/dt) < -

8.0 or nearly below ln(Q) < -3.20 for Gaulfoss and similarly for the other catchments. However, the values of estimated parameters are sensitive to the cut limits of the lower end of recession that we obtained markedly different recession parameters from different cut limits of the lower end recession.

Moreover, the estimated parameter sets based on different lower cut limits provided equivalently good performances of runoff simulation, which obviously indicate a major source of uncertainty. Therefore, in the present study we kept the lower ends of recession segments while estimating the parameters. However, we limited the upper end of recession to ln(Q) = 0 to remove outliers above this limit, which are most probably attributable to the errors in streamflow measurements during high flow recessions. A further study is required to address uncertainties due to the lower end of recession and other sources in a comprehensive manner, which was not the main objective in the present study. Stoelzle et al. (2012) compared different recession extraction and fitting procedures and found significant differences in the results. Differences between the g(Q) from recession and direct calibration may also arise since the g(Q) from recession analysis was obtained from parameter estimation based on nighttime hourly recessions for 17 years streamflow data while g(Q) from calibration was obtained from calibration based on 2 years hourly continuous records. Continuous records in the case of calibration allow inclusion of all ranges of streamflow (low flow to high flow) which have different degrees of sensitivity to the catchment storage.

We also found that incorporating the shorter recession segments in the analysis provided a nearly constant τ and g(Q). Since streamflow fluctuations causing recessions only for short periods are not related to catchment storage, we set minimum length of recession segments to be included in the analysis to exclude the shorter discharge fluctuations. Selection of recession segments longer than 9 to 15 h provided nearly similar patterns of g(Q) for Gaulfoss and Eggafoss and hence we extracted recession segments $\geq$ 9h for the two catchments while we extracted recession segments $\geq$ 4h for Hugdal bru and Lillebudal bru. In addition, the quickly draining storages are more prone to evaporation (see Staudinger et al., 2011).

The differences in the recession versus calibration and the routed versus unrouted are observed both in the slope and intercept parameters. Therefore, it is not easy to distinguish the effects of the procedures in the recession analysis from the effects of the runoff delay related to the river networks or catchment size. However, the model calibration allows quantification of uncertainty in the g(Q) from the posterior parameters. The ensemble mean of g(Q) from the last 1000 posterior parameters sets are lower than the g(Q) corresponding to the optimal parameter sets for both the runoff routed and unrouted cases while the difference is more exaggerated for runoff routed case (Fig. 5b).

Effects of catchment size and runoff delay

We obtained a maximum runoff delay or travel time lag between headwater and outlet of 14.81h for Gaulfoss based on parameter calibration including the source-to sink routing algorithm. When we consider the runoff routing explicitly during the calibration, the routing parameters accounts for runoff delay in the hillslopes and channel networks and hence it is expected that the response routine regression parameters do not compensate for the runoff delay. Therefore, we obtained lower 'response time' ($\tau = 1/g(Q)$) when routing is included in the model compared to the case when runoff is unrouted (Figure 5(b)). Despite the significant runoff travel time lag in the Gaulfoss catchment compared to the hourly runoff simulation, we found slight underestimation of peak flows due to the effects of neglecting runoff delay or runoff routing during the calibration (Fig. 3c). In addition, there is no considerable difference in the maximum performance measures (max NSE). This shows that interaction or compensation among model parameters during calibration partially conceals the sensitivity of the outlet hydrographs to the effect of runoff delay, which is the unidentifiability problem. However, generally importance of runoff routing decreases with the catchment size and almost negligible for the smallest modelled catchment of Lillebudal bru (Table 5).

Effects of data quality

The climate stations are available only inside the Gaulfoss and Hugdal bru catchments and hence more representative climate input is expected for these catchments than for Eggafoss and Lillebudal bru. Therefore, for study catchments with few precipitation stations the performance of precipitation interpolator affects the inferences made on the calibrated parameters. Moreover, sparse gauging networks may not capture localized precipitation events. Slightly better transferability of parameters to internal subcatchments for the parameter sets based on recession analysis over parameter sets based on direct calibration (Table 5) may be attributable to the fact that estimation from recession is dependent only on streamflow while the representativeness of climate input affects the calibration.

6 CONCLUSION

A parsimonious ‘top-down’ modelling is also prone to the parameter uncertainty and non-identifiability problems, which are the main challenges in the ‘bottom-up’ paradigm. The correlation among the parameters during calibration and hence their non-identifiability masks the sensitivity of streamflow simulation from catchment storage-discharge relationships to the runoff delay even for a macroscale catchment. Detailed evaluation of the reliability of runoff simulation and inferences made from the ‘top-down’ approach is required for any catchment size for instance based on variables other than the catchment-integrated streamflow used for model calibration, which requires observations other than streamflow. Study on combined utility of the ‘top-down’ and process-based ‘bottom-up’ approaches by utilizing the strengths of each would be essential.

Calibration based on data from high-density climate stations would improve the uncertainty and poor identifiability problems for improved predictions and inferences on the hydrological behavior of catchments. Evaluation of possible effects of precipitation during recessions is required rather than assuming negligible precipitation during recessions. However, the results obtained from the present study encourage further comparative evaluation of the routine for prediction purposes based on large number of catchments.

Acknowledgements

This work was a part of Centre for Environmental Design of Renewable Energy’s, (CEDREN’s) hydroPEAK projects under hydrology sub project (Project number: 50043420). CEDREN substantially funded the project. We wish to express our thanks to the Norwegian Meteorological Institute, Statkraft, TrønderEnergi and bioforsk for the climate data and the Norwegian Water Resources and Energy Directorate for the streamflow data used in the present study.

REFRENCES

Ambroise, B., Beven, K. & Freer, J. (1996) Towards a generalization of the TOPMODEL concepts: topographic indices of hydrological similarity. *Water Resour. Res.* 32(7), 2135–2145.

Bard, Y. (1974) *Nonlinear parameter estimation*. Academic Press, New York, NY.

Beldring, S. (2002) Runoff Generating Processes in Boreal Forest Environments with Glacial Tills. *Nord. Hydrol*. 33 (5), 347-372.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R. & Vøkso, A. (2003) Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrol. Earth. Syst. Sc*. 7(3), 304-316.

Beven, K. J. & Binley, A. M. (1992) The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process*. 6, 279-298.

Beven, K. J. (2000) Uniqueness of place and process representations in hydrological modelling. *Hydrol. Earth Syst. Sci*. 4(2), 203-213.

Beven, K.J., Freer, J., Hankin, B. & Schulz, K. (2000) The use of generalized likelihood measures for uncertainty estimation in higher-order models of environmental systems. In: *Nonlinear and Nonstationary Signal Processing* (Fitzgerald, Smith, R.C., Walden, A.T., & Young, P.C., eds), Cambridge University Press, UK.

Blasone, R. S. (2007) Parameter Estimation and Uncertainty Assessment in Hydrological Modelling. *Ph.D. Thesis*, Institute of Environment & Resources Technical University of Denmark

Blöschl, G. (2006) Hydrologic synthesis: Across processes, places, and scales. *Water Resour. Res*. 42, W03S02, doi:10.1029/2005WR004319.

Bogacki, P. & Shampine, L. F. (1989) "A 3(2) pair of Runge–Kutta formulas". *Applied Mathematics Letters* **2** (4), 321-325, doi:10.1016/0893-9659(89)90079-7, ISSN 0893-9659.

Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal* Statistical Society, Series B 26, 211-252.

Brauer, C. C., Teuling, A. J.,Torfs, P. J. J. F. & Uijlenhoet, R. (2013) Investigating Storage-Discharge Relations in a Lowland Catchment Using Hydrograph Fitting, Recession Analysis and Soil Moisture Data. *Water Resour. Res*. 49, 4257–4264, doi:10.1002/wrcr.20320.

Brutsaert, W., Nieber, J. L., 1977. Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resour. Res*. 13, 637–643.

Brutsaert, W. (2008) Long-term groundwater storage trends estimated from streamflow records: Climatic perspective. *Water Resour. Res*. 44, W02409, doi:10.1029/2007WR006518.

Bárdossy, A. (2007) Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth. Syst. Sc*. 11, 703-710.

Clark, M.P., Rupp, D.E., Woods, R.A., Tromp-van Meerveld, H.J, Peters, N.E. & Freer, J.E. (2009) Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrol. Process. 23*, 311-319.

Gottschalk, L., Beldring, S., Engeland, K., Tallaksen, L., Sælthun, N.R. & Kolberg, S. (2001) Regional/macroscale hydrological modelling: a Scandinavian experience. *Hydrolog. Sci. J*. 46 (6), 963-982.

Graham, C., B., Woods, R. A., McDonnell, J. J. (2010) Hillslope threshold response to rainfall: (1) A field based forensic approach. *J. Hydrol*. 393, 65–76.

Graham, C. B. & McDonnell, J. J. (2010) Hillslope threshold response to rainfall: (2) Development and use of a macroscale model. *J. Hydrol*. 393, 77-93.

Götzinger, J. & Bárdossy, A. (2007) Comparison of four regionalization methods for a distributed hydrological model. *J. Hydrol*. 333, 374-384.

Jansson, C. Espeby, B. & Jannson, P-E. (2005) Preferential flow in glacial till soil. *Nord. Hydrol*. 36 (1), 1-11.

Kirchner, J. (2003) A double paradox in catchment hydrology and geo-chemistry. *Hydrol. Processes* 17, 871 – 874.

Kirchner, J.W. (2006) Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res*. 42, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W. (2009) Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res*. 45, W02429, doi:10.1029/2008WR006912.

Klemeś, V. (1986) Operational testing of hydrological simulation models. *Hydrolog. Sci. J.* 31, 13-24.

Kolberg, S. A. & Bruland, O. (2012) ENKI - An Open Source environmental modelling platform. *Geophys. Res. Abstracts* 14, EGU2012-13630, EGU General Assembly.

Kolberg, S. A. & Gottschalk, L. (2006) Updating of snow depletion curve with remote sensing data. *Hydrol. Process*. 20(11), 2363–2380.

Krakauer, N.Y. & Temimi, M. (2011) Stream recession curves and storage variability in small watersheds. *Hydrol. Earth Syst. Sci*. 15, 2377–2389.

Krier, R. Matgen, P. Goergen, K., Pfister, L., Hoffmann, L., Kirchner, J. W., Uhlenbrook,

S. & Savenije, H. H. G. (2012) Inferring catchment precipitation by doing hydrology backward: A test in 24 small and mesoscale catchments in Luxembourg. *Water Resour. Res*. 48, W10525, doi: 10.1029/2011WR010657.

Lamb, R, & Beven, K. J. (1997) Using interactive recession curve analysis to specify a general catchment storage model. *Hydrol. Earth Syst. Sci*. 1,101-113.

Maidment, D.R., Olivera, J.F., Calver, A., Eatherral, A. & Fraczek, W. (1996) A unit hydrograph derived from a spatially distributed velocity field. *Hydrol. Process*.10 (6), 831–844.

McDonnell, J. J., Sivapalan, M., Vaché, K. Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J. & Weiler, M. (2007) Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. *Water Resour. Res.* 43, W07301, doi:10.1029/2006WR005467.

Myrabo, S. (1997) Temporal and spatial scale of response area and groundwater variation in Till. *Hydrol. Process*.11, 1861–1880.

Naden, P. S. (1992) Spatial variability in flood estimation for large catchments: The exploitation of channel network structure. *J. Hydrol. Sci*. 37, 53–71.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models, I. A discussion of principles. *J. Hydrol*. 10, 228–290.

Olivera, F. (1996) Spatially distributed modeling of storm runoff and nonpoint source pollution using geographic information systems, *PhD Thesis*, Department of Civil Engineering, University of Texas at Austin, USA.

Oudin, L., Kay, A., Andréassian, V. & Perrin, C. (2010) Are seemingly physically similar catchments truly hydrologically similar?. *Water Resour. Res*. 46, W11558, doi: 10.1029/2009WR008887.

Parajka, J., Merz, R. & Bloschl, G. (2007) Regional calibration of catchment models: Potential for ungauged catchments. *Hydrol. Earth Syst. Sci*. 9, 157-171.

Pokhrel, P., Gupta, H. V. & Wagener, T. (2008) A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resour. Res*. 44, W12419, doi:10.1029/2007WR006615.

Pokhrel, P. & Gupta, H. V. (2010) On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resour. Res*. 46, W01505, doi:10.1029/2009WR008066.

Pokhrel, P. & Gupta, H. V. (2011) On the ability to infer spatial catchment variability using streamflow hydrographs. *Water Resour. Res*. 47, W08534, doi:10.1029/2010WR009873.

Priestley, C.H.B. & Taylor, R.J. (1972) On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev*. 100, 81-82.

Rees, H.G.; Holmes, M.G.R.; Young, A.R. & Kansakar, S.R. (2004) Recession-based hydrological models for estimating low flows in ungauged catchments in the Himalayas. *Hydrol. Earth Syst. Sci*. 8(5), 891-902.

Rooney, W. C. & Biegler. L.T. (1999) Incorporating joint confidence regions into design under uncertainty. *Comput. Chem. Eng*. 23, 1563-1575.

Rooney, W. C. & Biegler, L. T. (2001) Design for model parameter uncertainty using nonlinear confidence regions. *Proc. Syst. Eng*. 47, 1794-1804.

Shrestha, S., Bastolab, S. Babelc, M.S., Dulalb, K.N., Magomeb, J. Hapuarachchid, H.A.P., Kazamaa, F., Ishidairab, H. & Takeuchid, K. (2007) The assessment of spatial and temporal transferability of a physically based distributed hydrological model parameters in different physiographic regions of Nepal. *J. Hydrol*. 347, 153-172.

Sivapalan, M. Blöschl, G., Zhang, L. & Vertessy, R. (2003) Downward approach to hydrological prediction. *Hydrol. Process*. 17, 2101 – 2111.

Spence, C., Guan, X.J., Phillips, R., Hedstrom, N., Granger, R. & Reid, B. (2010) Storage dynamics and streamflow in a catchment with a variable contributing area. *Hydrol. Process*. 24, 2209-2221.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P. & Tallaksen, L. M. (2011) Comparison of hydrological model structures based on recession and low flow simulations. *Hydrol. Earth Syst. Sci*. 15, 3447–3459.

Stedinger, J. R., Vogel, R. M., Lee, S. U. & Batchelder, R. (2008) Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res.* 44, W00B06, doi:10.1029/2008WR006822.

Stoelzle, M., Stahl, K. & Weiler, M. (2012) Are streamflow recession characteristics really characteristic?. *Hydrol. Earth Syst. Sci. Discuss*. 9, 10563-10593.

Tallaksen, L. M. (1995) A review of baseflow recession analysis. *J. Hydrol*. 165, 349-370.

Teuling, J. Lehner, I., Kirchner, J. W. & Seneviratne, S. I. (2010) Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment. *Water Resour. Res*. 46, W10502, doi:10.1029/2009WR008777.

Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M. & Higdon D. (2009) Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *Journal of Nonlinear Sciences and Numerical Simulation* 10(3), 273-290.

Vugrin, K. W., Swiler, L. P., Roberts, R. M., Stucky-Mack, N. J., & Sullivan, S. P. (2007) Confidence region estimation techniques for nonlinear regression in groundwater flow: Three case studies. *Water Resour. Res.* 43, W03423, doi:10.1029/2005WR004804.

Wagener, T., Sivapalan, M., Troch, P. A. & Woods, R. A. (2007) Catchment classification and hydrologic similarity. *Geogr. Compass* 1, doi:10.1111/j.1749-8198.2007.00039.x.

Wagener, T. & Wheater, H. S. (2006) Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *J. Hydrol*. 320, 132-154.

Wittenberg, H. & Sivapalan, M. (1999) Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation. *J. Hydrol*. 219, 20-33.

Zehe, E. & Sivapalan, M. (2009) Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications. *Hydrol. Earth Syst. Sci*. 13,1273- 1292.

Zięba, A. (2010) Effective number of observations and unbiased estimators of variance for autocorrelated data an overview. *Metrol. Meas. Syst*. XVII (1), 3-16.

Appendix

The multivariate normal probability density function for the regression parameters can be written as:

$$f\left(\hat{\underline{\beta}}\right)=\frac{1}{\left(2\pi\right)^{\frac{p}{2}}\left[\left|\left(\underline{X}^{T}\underline{X}\right)^{-1}\sigma^{2}\right|\right]^{\frac{1}{2}}}\exp\left\{-\frac{1}{2\sigma^{2}}\left(\underline{\beta}-\hat{\underline{\beta}}\right)^{T}\left(\underline{X}^{T}\underline{X}\right)\left(\underline{\beta}-\hat{\underline{\beta}}\right)\right\}, \tag{A1}$$

where $\underline{X}$ is a matrix of exploratory variables, $\underline{\beta}$ is vector of parameters, p is the total number of parameters, $|(\underline{X}^T\underline{X})^{-1}\sigma^2|$ is determinant of the covariance matrix of the parameter estimates, σ^2 is variance, the underline represents a vector or a matrix, T denotes transpose and the hat notation represents an estimate.

Tables

Table 1. Some major characteristics of the study catchments.

Description [units]	Gaulfoss	Hugdal bru	Eggafoss	Lillebudal bru
Lat./long. of flow stns [^{0}]	63.12/10.25	63.01/10.19	62.93/11.08	62.83/10.48
Catchment area, A [km^2]	3090	549	653	168
Elev. 25 m DEM [m.a.m.s.l.]	54-1330	130-1257	285-1286	516-1304
Mean elev. catchment [m.a.m.s.l.]	730.6	651.27	832.33	915.20
Elev. of climate stations	127-885	127-885	127-885	127-885
Catch. averaged annual precip.[mm]	874.21	863.53	884.06	864.7
Lake percentage [%]	2.05	1.0	2.84	1.14
Forest percentage [%]	36.72	53.59	24.55	21.7
Bare rock/mountain above TL* [%]	35.8	20.69	43.96	65.33
Marsh/Bog [%]	14.53	16.70	12.57	8.98
Farm land [%]	2.66	5.99	2.11	0.52

* TL denotes timberline.

Table 2. Lists of free parameters and their uniform priors.

No.	Calibrated parameters	Description	Unit	Uniform prior range
	Snow			
1	TX	Threshold temperature	^{0}C	[-3,2]
2	WS	Snow melt sensitivity to wind speed	-	[1,6]
	Response			
3	EvR	Discharge at which AET equals 0.95*PET	mm/h	[0.1,Q_{max}]
4	β_0	Regression parameter (constant term)	-	[-4.2,-1]
5	β_1	Regression parameter (linear term)	-	[0.2,1.5]
6	β_2	Regression parameter (quadratic term)	-	[-0.5,0,5]
	Routing			
7	V	Velocity of flow	m/s	[1.9,2.6]
8	D	Dispersion coefficient of flow	m^2/s	[200,1500]

Q_{max} is maximum streamflow for the calibration period.

Table 3. Calibrated and estimated parameters (corresponding to max NSE) for Gaulfoss. * Parameters other than regression are taken from calibrated (routed). Bold fonts are recession/regression parameters.

Cases	TX	WS	EvR	β_2	β_1	β_0	V	D
1	*Max NSE: regression parameters estimated from recession**							
	-1.482	5.708	0.273	**0.130**	**0.920**	**-2.850**	2.198	774.215
2	*Max NSE: calibrated (routed)*							
	-1.482	5.708	0.273	**-0.040**	**0.684**	**-2.477**	2.198	774.215
	Max NSE: calibrated (unrouted)							
3	-0.999	5.895	0.670	**-0.127**	**0.286**	**-3.234**		

Table 4. Parameters correlation matrix (r), and maximum and minimum values of marginal posteriors parameters for the calibration (runoff-routed) for Gaulfoss.

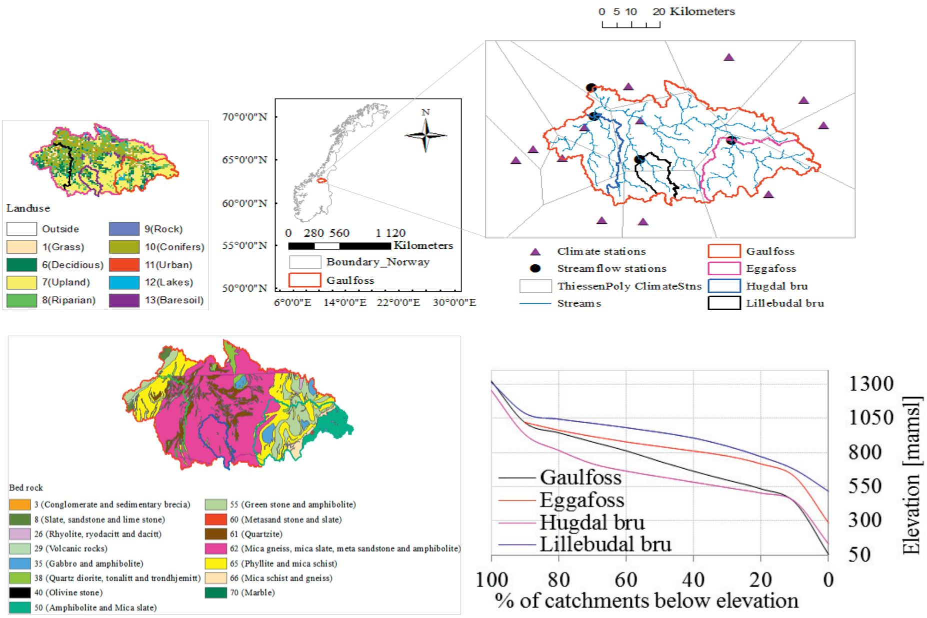
Parameters	TX	WS	EvR	β_1	β_0	β_2	V	D
TX	1.00	0.44	0.27	0.03	0.12	-0.07	0.06	0.05
WS		1.00	0.23	0.25	0.27	0.23	-0.28	0.12
EvR			1.00	0.51	**0.64**	0.40	-0.17	-0.37
β_1				1.00	**0.91**	**0.97**	-0.28	-0.23
β_0					1.00	**0.80**	-0.36	-0.23
β_2						1.00	-0.23	-0.20
V							1.00	-0.08
D								1.00
Max	-0.73	6.00	0.34	1.49	-1.55	0.10	2.59	1466.24
Min	-1.41	2.71	0.13	0.27	-3.63	-0.13	1.91	201.11

Bold fonts are $|r| > 0.6$.

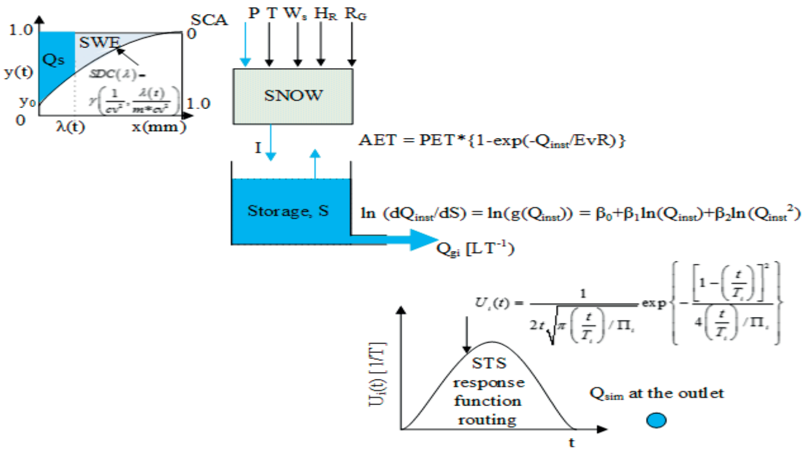
Table 5. Results for Max NSE from parameter estimation from recession and calibration for Gaulfoss (01.09.2008-01.09.2010), and temporal and spatial transferability/verification of model parameters for internal subcatchments.

Cases	Max NSE from calibration/estimation for Gaulfoss	Temporal verification (01.09.2010-01.09.2011)	Spatial validation or transferability ('proxy ungauged' internal catchments		
			Eggafoss	Hugdal bru	Lillebudal bru
1	*Max NSE: regression parameters estimated from recession*				
	0.77	0.81	0.70	0.80	0.55
2	*Max NSE: calibrated (routed)*				
	0.82	0.83	0.68	0.75	0.52
3	*Max NSE: calibrated (unrouted)*				
	0.80	0.82	0.65	0.79	0.51

Figures

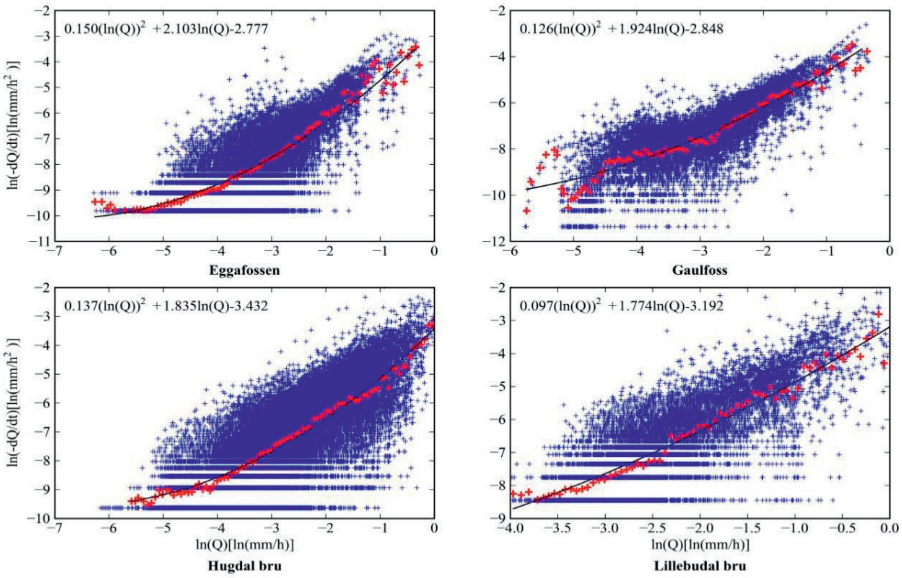


(a)

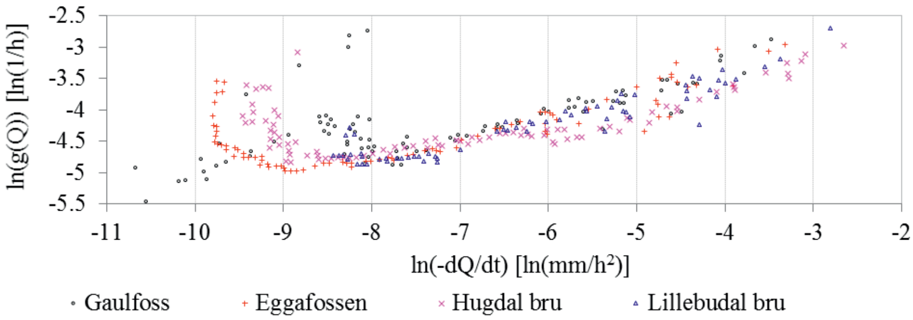


(b)

Figure 1. (a) Maps of landuse, locations, bedrock geology and hysometric curves for the study catchments and (b) model structure (grid cell) based on Kirchner's runoff response routine. *The snow routine is based on Kolberg et al. (2006). The* Q_{inst} *is instantaneous runoff,* Q_{gi} *is average runoff (over the time step) generated at the grid cell and* Q_{sim} *is the* Q_{gi} *routed to the outlet.*



(a)



(b)

Figure 2. (a) Flow recession rates and fitted recession plots, and (b) ln (-dQ/dt) versus ln(g(Q)) plots showing effects of lower end of recession.

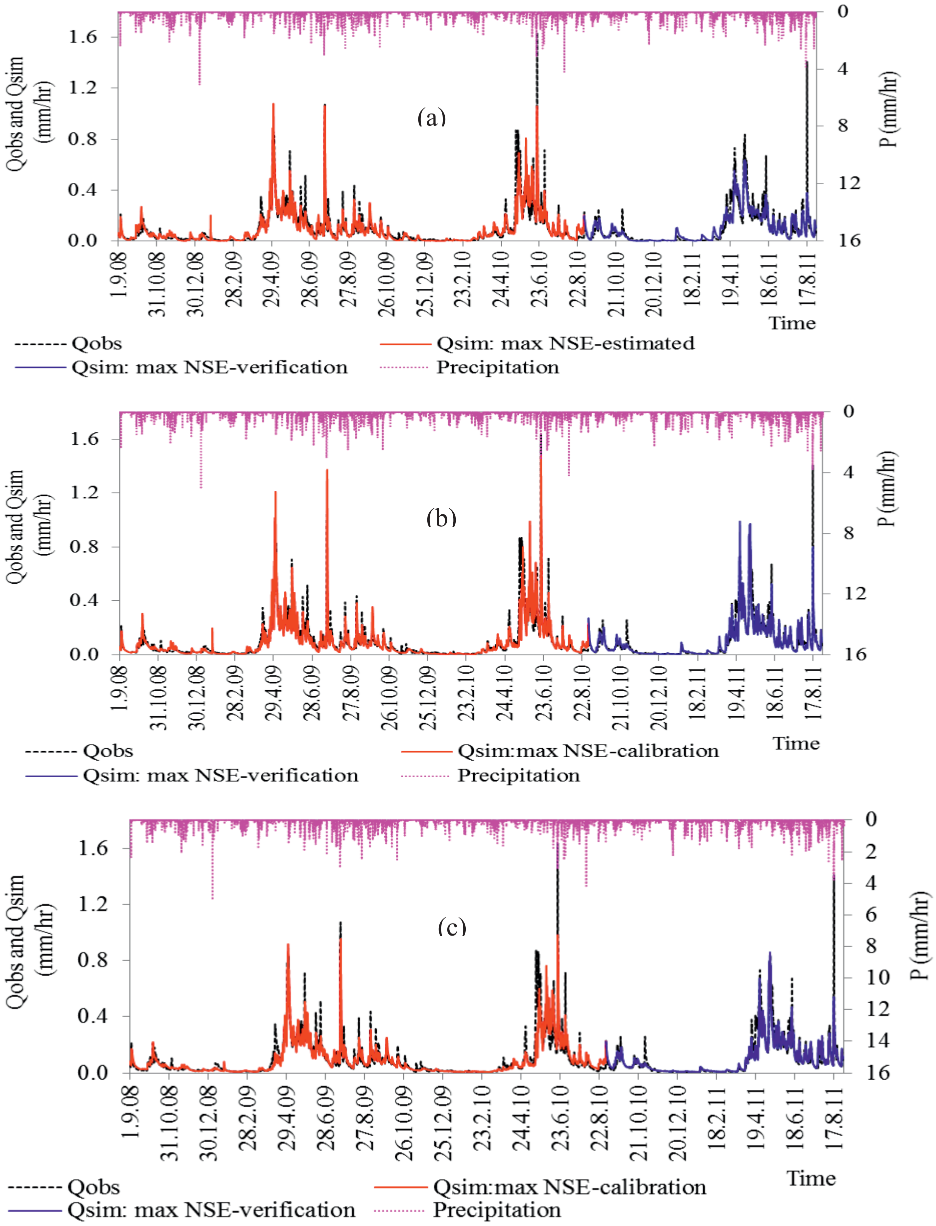
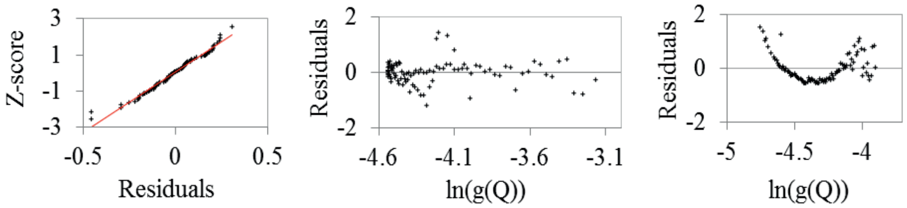
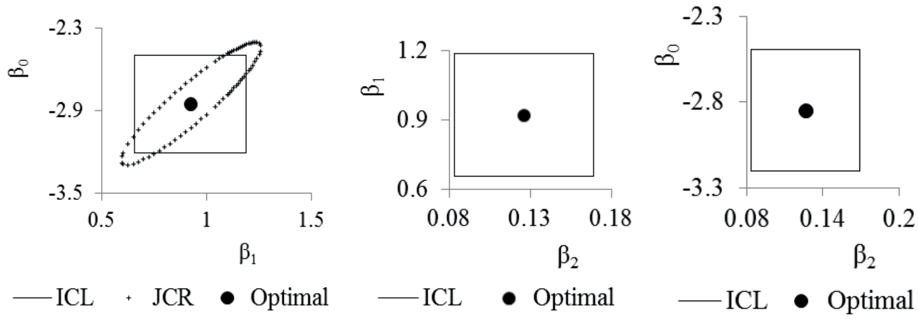


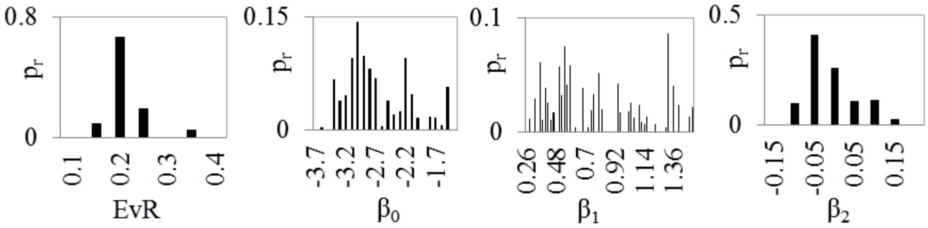
Figure 3. Hydrographs for Gaulfoss from DREAM log-likelihood performance measure calibration corresponding to max NSE (a) regression parameters estimated from recession (runoff routed), (b) calibrated (runoff routed) and (c) calibrated (runoff unrouted).



(a)

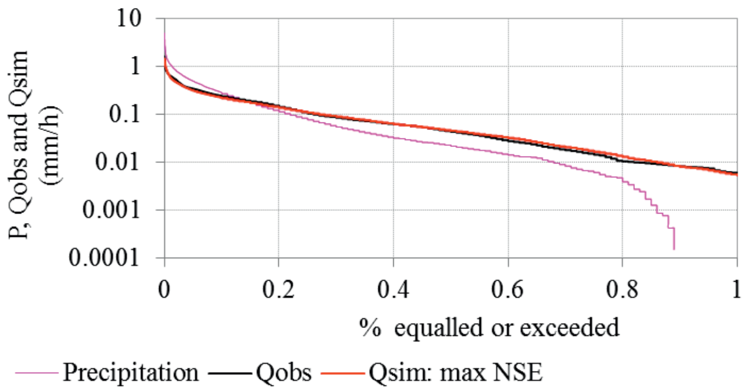


(b)

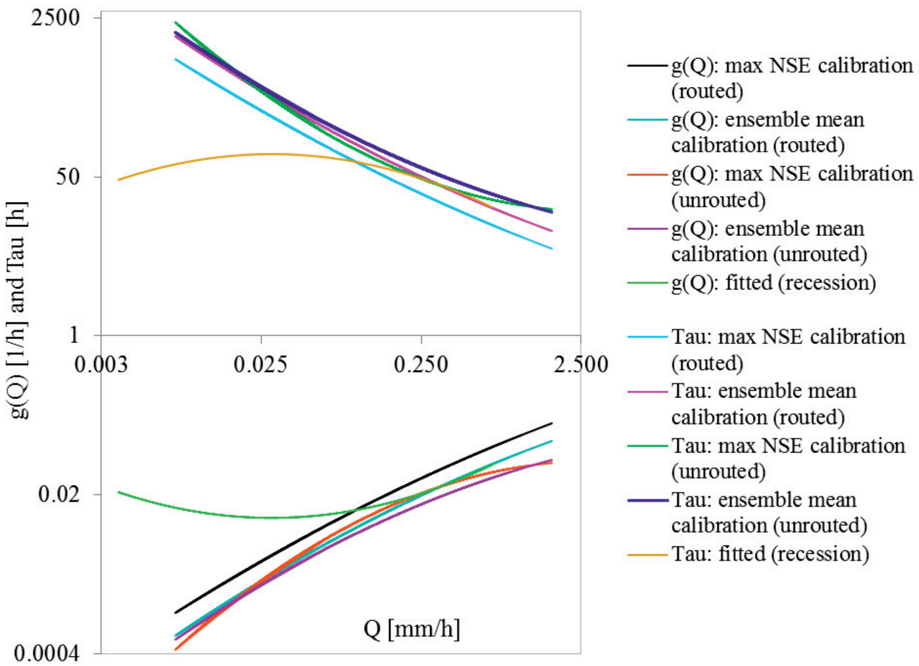


(c)

Figure 4. (a) Typical results from diagnostic of the regression for recession analysis: normality test for Hugdal bru (left), residuals plot for Gaulfoss (middle) and systematic lack of fit due to missing quadratic term for Eggafoss (right), (b) 95% Individual Confidence Limits (ICL) and Joint Confidence Regions (JCR) for regression parameters for Gaulfoss, and (c) Response routine parameters uncertainty in terms of histograms of marginal posteriors from calibration (runoff-routed) for Gaulfoss.



(a)



(b)

Figure 5. (a) Precipitation and streamflow duration curves and (b) Observed discharge sensitivity function (g(Q)) and 'response time' (Tau) for Gaulfoss.

The g(Q) and Tau are computed based on eq. 5 from fitted and calibrated regression parameters.

Paper 4 (P4)

Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments

Teklu T. Hailegeorgis, Yisak S. Abdella, Knut Alfredsen and Sjur Kolberg

Paper IV

Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments

Teklu T. Hailegeorgis*[1], Yisak S. Abdella[2], Knut Alfredsen[1] and Sjur Kolberg[2]

[1]Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim.

[2]SINTEF Energi AS, Sem Sælands vei 11, NO-7465 Trondheim.

*Corresponding author: Teklu T. Hailegeorgis (tekhi09@gmail.com and teklu.hailegeorgis@ntnu.no), Tel.: [+47] 73592411 and fax: [+47] 735 91298.

Abstract

Regionalization for continuous hourly streamflow simulation is not common in literature especially for boreal catchments. In this paper we performed calibration of 26 catchments (39-3090 km^2) in boreal mid-Norway using three different model structures: a first-order nonlinear system model (hereafter Kirchmod), the HBV model and the Basic-Grid-Model (BGM). We evaluated four regionalization methods: regional calibration or parameter vector yielding maximum regional weighted average (MRWA) performance measures (PM), regional median of optimal parameters (RMedP), nearest neighbor (NN) and physical similarity. The physical similarity attributes include hypsometric curves (PSH), land use (PSLU), drainage density (PSDD), catchment area (PSCA), terrain slope (PSSL), bedrock geology (PSBRG), soil type (PSSOIL) and combinations of all (PSCOMB).

Based on the regional median and mean of the raw values and regional median and mean of losses or gains of the PM (from local calibration due to the regionalization) evaluation metrics, a multi-donor MRWA method performed slightly better than the other methods. For the single-donor methods, the physical similarity could explain hydrological homogeneity better than the NN. The more comprehensive evaluation metrics of the cumulative distribution functions (CDF) of the losses or gains in the PM indicated that the PSCOMB and PSSOIL are more suitable for the R^2 and R^2ln respectively and hence the single-donor physical similarity performed slightly better than the multi-donor MRWA. However, due to the marginal differences between the multi-donor regional calibration (MRWA) and the single-donor physical similarities (PSCOMB and PSSOIL),

both methods can be considered suitable for the study region. The study indicated the importance of comprehensive evaluations of the regionalization methods by considering the objectives of prediction (high flow, low flow and water balance) for selection of the PM and various evaluation metrics for the PM.

Introduction

Continuous streamflow simulation for prediction in ungauged basins or PUB (Sivapalan et al., 2003) is a fundamental and challenging task in hydrology. Precipitation-runoff models are widely used to transfer hydrological information (through model parameters) for continuous simulation in ungauged basins. The approach also has the potential to provide generalized or regionalized hydrological understanding and inferences. In the present study, regionalization method is defined as any method for transfer of hydrological information from gauged to ungauged basins (Blöschl and Sivapalan, 1995; Oudin et al., 2010).

However, there are different challenging factors pertinent to identification of suitable regionalization method(s) namely the selection of proper donor catchment(s), identification of proper model structure and performance measures. Wagener and Wheater (2006) noted the uncertainties pertinent to estimation of continuous streamflow time-series in ungauged basins.

Regionalization attempts based on different methods have been conducted in literature. The first category of methods is based on regional calibration by utilizing data from multi-sites in the region or a multi-donor approach. Fernandez et al. (2000) implemented a regression model based regional calibration. Beldring et al. (2003) conducted a multi-criteria calibration of a distributed HBV model for 141 catchments in Norway. Engeland et al. (2006) conducted a multi-objective model calibration for a physically-based distributed Ecomag model (Motovilov et al., 1999) based on a multi-site streamflow observations. Parajka et al. (2007) proposed an iterative regional calibration (IRC) by defining a combined objective function that linearly combines local and regional information. Donnelly et al. (2009) demonstrated comparable performance of a spatially homogeneous parameter set to the local calibrated parameter sets. Vaze et al. (2013) showed nearly equivalent performance of the regional calibration and the transfer of local calibrated parameter values from gauged catchments using geographical proximity.

The second category of multi-donor regionalization methods are based on pooling of model parameters (e.g. transferring of averages of optimal parameters for the catchments in the region and methods based on averaging of streamflow) and also ensemble simulations. Kokkonen et al. (2003), Oudin et al. (2008) and Kim and Kaluarachchi (2008) computed regional parameters as averages of local calibrated parameters. Goswami et al. (2007) followed regional averaging of discharge from multi-donor catchments for model calibration at pseudo-ungauged catchments. McIntyre et al. (2005) employed an output averaging method where a regional computed streamflow was derived from simulations obtained from acceptable parameter sets for a number of different gauged catchments. Recently, Cibin et al. (2014) illustrated ensemble simulations for the PUB based on transferring derived probability distributions of parameters rather than transferring a single optimal parameter vector or averaged or interpolated parameter values.

The third category of regionalization method is the geographic distance based nearest neighbor (NN). The NN method is a spatial proximity approach based on the assumption that hydrological properties may vary smoothly in space and hence geographic proximity between the donor and the recipient catchments could explain hydrological similarity. This method is based on a single-donor catchment. Some of the previous regionalization attempts based on the NN are Merz and Blöschl (2004), Parajka et al. (2005), Oudin et al. (2008) and Samuel et al. (2011).

The fourth category of regionalization method is based on physical similarity. The method of physical similarity is based on the assumption that similarity in some physical attributes, which govern the runoff response, could explain the hydrological similarity. This method is also based on a single-donor catchment. Kokkonen et al. (2003), McIntyre et al. (2005), Parajka et al. (2005), Oudin et al. (2008) and Reichl et al. (2009) and Zhang and Chiew (2009) conducted regionalization based on the physical similarity.

The fifth category of regionalization methods common in literature are regression based which are also a multi-donor approach. However, Bárdossy (2007) illustrated that regionalization should not focus on relating catchment properties to individual parameter values but on relating them to compatible parameter sets or vectors. Parameter equifinality (Beven, 2006) or interactions among parameters during calibration may not retain the expected relationships between model parameters and catchment attributes. In addition, several studies among others Fernandez et al (2000), Kokkonen et al. (2003), McIntyre et al. (2005), Wagener and Wheater (2006), Oudin et al. (2008), Kim and Kaluarachchi (2008), Bastola et al. (2008) and Parajka et al. (2013) noted various limitations in the regression based regionalization methods.

Comparative evaluations of the performances of different regionalization methods are commonly performed in literature in order to identify suitable method(s) for a region (e.g. Parajka et al., 2005; Oudin et al., 2008; Zhang and Chiew, 2009). However, endeavors for identification of suitable regionalization method(s) are affected by several factors such as selection of the model structure and performance measures, selection of the physical attributes for similarity measures, climatic and landscape characteristics of the study region, quality of input data and density of hydro-climatic gauging networks that require comprehensive study.

In addition, the previous comparisons of regionalization methods were mainly based on simulations for low spatial resolution (i.e. lumped or semi-distributed) and also low temporal resolution (e.g. daily runoff simulation). However, distributed or gridded spatial resolution allows better representation of the spatial heterogeneity for prediction in ungauged basins. Moreover, evaluation of performances of the regionalization methods based on hourly temporal resolution are also important for management of water resources e.g. inflow prognosis for hydropeaking operation of hydropower reservoirs, flood prediction and monitoring of environmental flows. Also, better insights on the regionalization endeavors would be expected from the fine resolution inputs. Where sub-daily data exist, it would appear to be wise to use the extra information they contain, leading to more accurate calibrated model parameters (Littlewood et al., 2013).

Regionalization for continuous simulation of streamflow using precipitation-runoff models are affected by the model structure and model calibration (Wagener and Wheater, 2006; Oudin et al., 2008; Oudin et al., 2010; Kim and Kaluarachchi, 2008). Engeland and Gottschalk (2002) noted that structural errors in the model are more important than parameter uncertainties. Oudin et al. (2010) while testing hypothesis of inferring hydrological homogeneity based on spatial proximity of catchments pointed out that the physical meaning of calibrated model parameters suffers from problems in model identification, model structural errors, and difficulties in finding an appropriate calibration strategy.

However, previous attempts for continuous streamflow simulation by rainfall-runoff models for the PUB were mainly based on conceptual model structures such as the HBV model (e.g. Siebert, 1999; Merz and Blöschl, 2004; Parajka et al., 2005; Götzinger and Bárdossy, 2007; Samuel et al., 2011), Probability Distributed Model based on parameterization of sub-basin heterogeneity (e.g. McIntyre et al., 2005; Zhang and Chiew, 2009). Rare applications of physics-based distributed model for regionalization include the Mike 11 NAM (Reed et al., 2004; Makungo et al., 2010). Hrachowitz et al. (2013) and Croke and McIntyre (2013) discussed the utility of the 'top-down' modelling approach for the PUB.

Parajka et al. (2007) suggested varying the model structure between catchments to improve model efficiency for local calibration. Wagener and Wheater (2006) also noted that progress in parameter estimation procedures and parsimonious modelling still has to be fully incorporated in regionalization approaches. Lee et al. (2005) while selecting conceptual models for regionalization of catchments in UK noted that the study provided no evidence of relationships between the catchment structures and the model structures. Therefore, comprehensive evaluation of the regionalization methods based on different modelling paradigms and model structures is indispensable.

Furthermore, Wagener and McIntyre (2005), Wagener and Wheater (2006) and references therein demonstrated the incapability of current model structures to simulate both high flow and low flow behaviors of catchments simultaneously with a single parameter set. The results of calibration and evaluation are also affected by the choice of objective functions (Gupta et al., 1998; Madsen, 2003; Muleta, 2012) for model calibration and any performance measures (PM) used for model evaluations. Hence, evaluations based on various performance measures may not provide similar regionalization solutions since the performance measures give different weightages to the high flows and low flows. In addition, some catchments that are similar in their rainfall and snow melt dependent high flow regime may not be necessarily similar in their catchment storage related low flow regime or vice versa due to differences in their precipitation patterns and subsurface characteristics. Dependency of catchment similarity on the flow conditions was illustrated by Patil and Stieglitz (2011) for flow duration curves for catchments in the United States. Hence, dependency of the regionalization on the performance measures is also an additional challenge in the regionalization endeavors, which needs to be studied. Further references on regionalization works can be found from reviews by He et al. (2011), Razavi and Coulibaly (2013) and Hrachowitz et al. (2013), and the synthesis by Parajka et al. (2013).

Despite of several attempts of regionalization for prediction in ungauged basins, there are still challenges in the transfer of hydrological information through rainfall-runoff model parameters from gauged to ungauged catchments within a certain region (see a recent review by Hrachowitz et al., 2013). No universally best performing regionalization method, model structure or evaluation criteria could be suggested due to the peculiarities of catchments in different climate regimes and landscapes (see a recent synthesis by Parajka et al., 2013). Therefore, the major challenges in precipitation-runoff model calibration based regional transfer of model parameters pose research questions pertaining to (a) are the performances of the regionalization methods consistent among the model type and the selected performance measures (PM)?; (b) Is the best regionalization method consistent among different evaluation metrics of the PM (i.e. based on raw values of performance measures and loss or gain in the performance measures from the local calibration due to the regionalization)? and (c) which regionalization method performs best for the specific region of interest?. The main objectives of the present study are to conduct comparative evaluations of the performances of four types of regionalization methods based on the above research questions. To our knowledge, the present study is the first attempt for distributed hourly runoff simulation in the study region and globally regionalization for distributed hourly simulation is not common in literature. Moreover, there are growing interests for hourly prediction in ungauged basins pertinent to hydropower operation (e.g. hydropeaking, floods, environmental flow assessments and prediction of natural flow series or the flow regime in regulated rivers).

The Study Region and Data

The study region consists of 26 unregulated gauged catchments ranging in size from 39 to 3090 km^2 in boreal mid Norway (Table 1 and Figure 1). There are five subcatchments nested in the main catchments. Catchments 3, 8 and 14 are subcatchments of catchment 6, catchment 5 is subcatchment of catchment 9, and catchment 19 is subcatchment of catchment 22.

Streamflow and climate records of hourly time resolution from 2008 to 2010 are used for model calibration. The climate forcing are precipitation (P), temperature (T), wind speed (W_s), relative humidity (H_R) and global radiation (R_G). Locations of precipitation and streamflow gauging networks are given in Figure 1. Lists of the catchments and streamflow stations, and some characteristics of the catchments are given in Table 1.

Precipitation occurs in the form of snowfall during winter and rainfall dominates during summer, spring and autumn. Therefore, runoff dynamics of the catchments is influenced by both rainfall-runoff and snowmelt-runoff processes. High flow regimes for most of the study catchments are related to snow-melt events (nival regime). Also, some catchments exhibit precipitation on snow melt events (pluvial and nival combined) and precipitation events (pluvial) high flow regime.

The catchments exhibit wide range variations of elevation and terrain slope. The dominant land uses/land covers in the study area are mountains above timberline and forests (Figure 2). There are also significant proportions of marshes/bogs and lakes for some of the catchments. Five of the study catchments have glacier coverage (maximum 5.25 % for catchment 10). Predominant soil or loose material is glacial tills (Figure 3). The dominant bedrock types for the study catchments are metamorphic and igneous rocks. We obtained the hypsometric curves and land use date from http://www.nve.no and the soil and bedrock geology data from http://www.ngu.no. We used the stream network from the 1:50000 maps produced by the Norwegian Mapping Authority (http://www.statkart.no) to estimate the drainage density [km/km^2] (Horton, 1932; Dingman 1978) as the total length of stream networks divided by the catchment area (Table 1).

Models and Methods

Three different distributed (1x1 km^2 grid) runoff response routines namely the 'top-down' water balance model proposed by Kirchner (2009) (hereafter Kirchmod), the Hydrologiska Byråns Vattenballansavdelning model (Bergström, 1976) (hereafter HBV) and the Basic-Grid-Model (Bell and Moore, 1998) (hereafter BGM) were used. The numbers of free parameters for the Kirchmod, HBV and BGM are 6, 10 and 7 respectively (Table 2). Brief descriptions of the models are given here.

Kirchner's runoff response routine

Kirchner (2009) inferred the model structure, equations and parameters from binned analysis of streamflow recession (i.e. the ‘top-down’ modelling paradigm). The method is based on a nonlinear single storage-discharge relationship. Kirchner (2009) applied the model for headwater catchments in Mid Wales (United Kingdom) by estimating the discharge sensitivity function g(Q) = dQ/dS, where S is catchment averaged storage and Q is discharge. Further applications of the methodology for individual catchments or at-site applications include Teuling et al. (2010), Krier et al. (2012), Ajami et al. (2011), Birkel et al. (2011) and Brauer et al. (2013). Hailegeorgis and Alfredsen (2014a) conducted regional calibration of the routine.

The main assumption in the Kirchner’s method is the streamflow Q depends solely on the amount of water stored in the catchment S. The water balance response routine is given as:

$$\frac{dQ}{dt} = \frac{dQ}{dS}\frac{dS}{dt} = \frac{dQ}{dS}\left(I - AET - Q\right) = g\left(Q\right)\left(I - AET - Q\right) \tag{1}$$

The following linear regression relationship was considered between the g(Q) and Q:

$$\ln\left(g\left(Q\right)\right) \approx \beta_0 + \beta_1 \ln\left(Q\right), \tag{2}$$

where the actual evapotranspiration (AET), infiltration (I) = rainfall + snow melt (SM) and discharge (Q) are given in mm/hr and catchment storage (S) is in mm, t is a time variable, β_0 and β_1 are parameters. We computed the AET from the potential evapotranspiration (PET) and discharge, which is used as a proxy to indicate the soil-moisture state according to:

$$AET = PET\left\{1 - \exp\left(-\frac{Q_{inst}}{EvR}\right)\right\}, \tag{3}$$

where Q_{inst} is an instantaneous discharge in mm/hr and EvR is a parameter which denotes a discharge at which AET equals 0.95*PET.

The HBV Runoff Response Routine

The distributed version of the HBV runoff response routine used in the present study contains two conceptual storage reservoirs. The relationship between the single outlet upper storage reservoir and outflow is non-linear while the relationship between the single outlet lower storage reservoir and baseflow is linear. The upper and lower reservoirs conceptually represent quick flow (Q_{UZ}) and baseflow (Q_{LZ}) respectively:

$$Q_{UZ} = k_1 \times (UZ)^{n_u} \; ; \; Q_{LZ} = k_0 \times (LZ), \tag{4}$$

where n_u is exponent of non-linearity for the upper zone (UZ), LZ denotes the lower zone, k_1 and k_0 are recession coefficients (parameters). Percolation from the upper to the lower reservoir is controlled by a parameter PERC. The soil moisture accounting routine is based on a non-linear function partitioning the infiltration from rainfall and snowmelt (I) into change in soil moisture storage (Δ_{SM}) and recharge (R) to upper reservoir (Bergström, 1976).

Basic Grid Model (BGM) Runoff Response Routine

The BGM is a simple distributed model based on Bell and Moore (1998). The runoff generation mechanisms are the Hortonian or infiltration excess runoff, R_{iex} [L] (Horton, 1933) and the 'fill and spill' type saturation excess runoff, R [L] (Dunne and Black, 1970 a&b) and a subsurface flow or drainage based on a non-linear catchment storage-discharge relationship:

$$R_{iex} = \max\left[0, (SNOWOUT - I_c)\right] \; ; \; TOSOIL = SNOWOUT - R_{iex} \tag{5}$$

$$R = \max\left[0, \left(S(t) + TOSTORAGE - S_{\max}\right)\right]; S(t+\Delta t) = \max\left\{0, \left[S(t) + TOSTORAGE - R\right]\right\} \tag{6}$$

$$TOSTORAGE = TOSOIL - AET - D_{rv} \; ; \; AET = PET \times \frac{S}{S_{\max}} \; ; \; D_{rv} = k\left[S(t)\right]^n , \tag{7}$$

where I_c [L/T] is an infiltration capacity parameter, SNOWOUT [L] is rainfall and snowmelt outflow from the snow routine, TOSOIL [L] is infiltration in to the soil, TOSTORAGE is net input to the subsurface storage (S[L]), PET [L] and AET [L] are as defined earlier, D_{rv}[L] is subsurface flow or drainage per unit area, and k [L^{1-n}/T], n[-] and maximum subsurface storage capacity or S_{max}[L] are the free parameters.

Snow Routine

The influences of snow processes are dominant in the study area during winter and spring seasons. The snow routine simulates the outflow melt water release from saturated snow (Q_s) based on the gamma distributed snow depletion curve or SDC (Kolberg and Gottschalk, 2006), which was implemented in ENKI hydrological modelling platform (Kolberg and Bruland, 2012). We used the same snow routine with all of the three runoff response routines.

Potential Evapotranspiration Routine

We used the PriestleyTaylor method (Priestley and Taylor, 1972) for the calculation of potential evapotranspiration for all the routines:

$$PET = \quad \alpha \frac{\Delta}{\Delta + \gamma}\left(R_n\right), \tag{8}$$

where Δ is the slope of saturation vapor pressure curve at air temperature at 2m (T2m), Y is the psychometric constant (0.67 $hPaK^{-1}$), R_n is net radiation and α is the Priestley Taylor constant. Following Teuling et al. (2010), we used an alpha value of 1.26 rather than setting it by calibration in order to reduce the number of free parameters.

Runoff Routing

A simple translation based on 1-hr isochrones was implemented for all the routines. The hillslope runoff response of each 1x1 km^2 grid cell was translated to the catchment outlet based on travel time lags. Routed simulated streamflow at the outlet was computed from the sum of contributions from each grid cell:

$$Qsim_t = \sum_{i=1}^{N} qsim_{t-T_i}^{i} \;, \mathrm{T_i} = \frac{L_i}{V}, \tag{9}$$

where *t* and *i* represent time and grid cells, N is the number of grid cells in the catchments, Qsim [LT^{-3}] is streamflow at the outlet, qsim [LT^{-3}] is runoff generated at each grid cell, T_i [T] is flow travel time lag to the outlet for each grid, L_i [L] is flow travel path length computed from 25m Digital Elevation model (DEM) and V [LT^{-1}] is 'effective' velocity of flow set by calibration.

Model Calibration and Diagnostics

We used the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2009) with residuals based log-likelihood (L-L) objective function for model calibration, which was implemented in ENKI hydrological modelling platform (Kolberg and Bruland, 2012). The log-likelihood (L-L) function for the regional calibration is:

$$L-L\left(\delta / \sigma_i^2, \sum_{i=1}^{N_C}\sum_{t=1}^{n_i}\left(Qsim_{t,i}^{(\theta)} - Qobs_{t,i}^{(\theta)}\right)^2\right) = \left\{\sum_{i=1}^{N_C}\left(\frac{-n_i}{2}\log\left(2\pi\right) - \frac{n_i}{2}\log\left(\sigma_i^2\right) - \frac{\sum_{t=1}^{n_i}\left(Qsim_{t,i}^{(\theta)} - Qobs_{t,i}^{(\theta)}\right)^2}{2\sigma_i^2}\right)\right\} \times f, \tag{10}$$

where δ denotes model parameter, σ_i^2 and n_i respectively are error variance and the length of non-missing records of streamflow for catchment i, N_C is the total numbers of modelled catchments in the region, $Qsim^{(\theta)}$ and $Qobs^{(\theta)}$ respectively are Box-Cox (Box and Cox, 1964) transformed observed and simulated streamflow time series, θ is the Box-Cox transformation parameter and f represents a fraction of effectively independent observations of the hourly time series, which can be estimated from the autoregressive (AR1) model of error covariance (Zięba, 2010).

The objective of Box-Cox transformation was to obtain an approximately Normal distributed series with homoscedastic residuals (i.e. variance of residuals is independent of streamflow). If θ = 0, the streamflow is assumed to be lognormal distributed and low flows get high weightages similar to the Nash-Sutcliffe efficiency for log-transferred series or R^2ln. If θ = 1 (i.e. no transformation), the streamflow series is assumed to be Gaussian and high flows get high weightages similar to the Nash-Sutcliffe efficiency or R^2 (Nash and Sutcliffe, 1970). Values of θ between 0.25 and 0.30 are common in literature (e.g. Vrugt et al., 2002 and references therein; Willems, 2009) but it can also be estimated by optimization. However, for the sake of consistency among the catchments, we used θ = 0.3 and f = 0.001.

We performed the performance evaluations of the local calibration and the regionalization methods based on the R^2 and R^2ln:

$$R_i^2 = 1 - \frac{\sum_{t=1}^{n_i}\left(Q_{obs_{t,i}} - Q_{sim_{t,i}}\right)^2}{\sum_{t=1}^{n_i}\left(Q_{obs_{t,i}} - Q_{obsavg.i}\right)^2} , \tag{11}$$

where $Q_{obsavg.i}$ is the mean value of observed streamflow time series for catchment *i* with length of non-missing streamflow record n_i. The evaluation metrics for the performance measures (PM) include their regional median and mean of the raw values, regional median and mean of their losses or gains from the local calibration due to the regionalization, cumulative distribution functions (CDF) of their raw values and CDF of their losses or gains from the local calibration due to the regionalization.

Parameter vectors which yield maximum values of R^2 and R^2ln for each catchment from the DREAM regional calibration are identified as optimal parameter vectors for each catchment and defined as local calibration (local calib.). These optimal parameter vectors for each catchment (local calibration) are transferred among the catchments based on the different regionalization methods. Regional calibration by the DREAM algorithm can be regarded as an 'importance sampling' strategy for each catchment, where we sample according to an ‘importance surface’ reflecting where we believe the optimum is likely to be. DREAM seeks and converges to the posterior distribution and DREAM’s regional posterior distribution is an importance surface for each catchment. The regional calibration approach is an acceptable calibration strategy for each single catchment without considerable loss in performance from the DREAM at-site calibration, which utilizes only streamflow data from each individual catchment.

Regionalization Methods

The performances of the following four regionalization methods were explored in order to identify suitable regionalization or parameter transfer method(s) in the region. Regionalization based on the regression method is not conducted in the present study due to its limitations discussed in the introduction.

Regionalization method 1: Regional calibration i.e. parameter vectors yielding maximum regional weighted average performance measures or the MRWA

This method explores the performance of the regional calibration based on a parameter vector which provide the maximum regional weighted average (MRWA) performance measures (R^2 and R^2ln) among those computed using eq. (12) for each parameter vector accepted by the DREAM calibration algorithm. It involves pooling of performance measures from multiple donor catchments to identify the parameter sets which provide the MRWA performance measures for the region. In this method a homogeneous parameter vector is derived for the region for each performance measure.

$$R_{RWA}{}^2 = \frac{1}{N_C}\sum_{i=1}^{N_C}\left(\frac{n_i}{N_{TS}}\right)R_i^2 \qquad (12)$$

N_{TS} is the length of calibration time stamps (including the missing records), N_C and n_i are as defined earlier and RWA represents regional weighted averaged, where the weights are the term in the parenthesis assigned for each catchment based on the length of their non-missing streamflow records in the calibration period. The regional calibration regionalization method used in the present study is similar to previous works on the regional calibration based regionalization in terms of utilizing the streamflow data from all the catchments in the region. However, the regional parameter vectors corresponding to the maximum regional weighted average (MRWA) performance measures are used in the present study.

Regionalization method 2: Regional median parameters (RMedP)

This method evaluates the performance of homogeneous parameter vector derived for a region from the optimal parameter vectors for each catchment or the local calibration. Performances of the regional median parameters (RMedP) were evaluated:

$$RMedP_j = Median\left\{P_j^1, P_j^2, ..., P_j^{N_C}\right\}, \tag{13}$$

where j is subscript for the free parameters (j = 1 to N_p, N_p is the total number of free parameters). This method allows pooling of parameters from multiple donor catchments (i.e. multi-donor median) and deriving unique parameter sets for the whole region for each performance measure. The only difference between this method and the parameter averaging method by Kokkonen et al. (2003), Oudin et al. (2008) and Kim and Kaluarachchi (2008) is that the median rather than the mean values of parameters were computed. The limitation of this method is it transfers the regional median of each parameter rather than a set or vector of optimal parameters and hence does not keep the correlation structure of the calibrated model parameters.

Regionalization method 3: Nearest neighbor (NN)

This method tests whether the nearest neighbor (NN) or the spatial proximity explains hydrological homogeneity or not. The method is based on transfer of the local calibration optimal parameter vector from the nearest neighbor single-donor catchment. The Euclidean distance in the geographic co-ordinates spaces is used to identify the nearest neighbor for the catchments.

Regionalization method 4: Physical similarity (PS)

The physical similarity based regionalization method tests the assumption that similarity of catchments in physical attributes could explain their homogeneity in runoff responses or the hydrological/functional similarity. This method is based on transfer of the local calibration optimal parameter vectors from a single donor catchment to a similar (in physical attributes spaces) recipient catchment. The method is highly affected by identification of the physical attributes governing the runoff response, the pre-processing of the selected attributes and the types of distance metrics, which entail some elements of subjectivity. Comprehensive work is required for identification of the physical, climate and runoff attributes (e.g. Sawicz et al., 2011; Viglione et al., 2013).

However, the attributes related to the physical and climate characteristics of catchments are more useful for prediction in ungauged basins. In the present study, we evaluated eight different cases of physical similarity based on seven attributes namely hypsometric curves (PSH), land use (PSLU), drainage density (PSDD), catchment area (PSCA), cumulative distribution functions of terrain slope (PSSL), bedrock geology (PSBRG), soil types (PSSOIL) and combinations of all attributes (PSCOMB). Despite of potential interactions among the attributes, combinations of large numbers of attributes are expected to provide more information on runoff responses of the catchments. The PSLU, PSBRG and PSSOIL are based on the percentages of catchment areas covered by the types or classes of the respective attributes.

Dominant topographic influences on the runoff response for boreal catchments were reported in Halldin et al. (1999) and Beldring et al. (2003). The hypsometric curves express how the area of the catchment is distributed according to elevation and it is expected to provide more information than using only the mean and median values of altitudes. Elevation variations can affect the precipitation pattern, snowmelt and the land cover. Catchments with steeper slopes are expected to have flashy response than catchments with gentle slopes. Large drainage density signifies dominant quick flow in stream channels. The land use mainly controls the water balance through evapotranspiration and snow processes. The soil types are used as a proxy for soil characteristics (e.g. infiltration capacity and soil depth) and bedrock types are used as a proxy for bed rock hydraulic properties mainly that influence the subsurface storage and release of water. The scale of the catchment (catchment area) mainly controls the runoff delay.

The Euclidean distance similarity metric between the two catchments ($Dist_{j,h}$) calculated in the Min Max normalized [0,1] attributes space for the sake of simplicity or scaling of the combined attributes is used:

$$\mathrm{Dist}_{\mathrm{j,h}} = \left\{ \sum_{i=1}^{m} \left(Nx_i^j - Nx_i^h \right)^2 \right\}^{\frac{1}{2}},$$
$$Nx_i^j = \frac{x_i^j - \min(x_i)}{\max(x_i) - \min(x_i)}, \; Nx_i^h = \frac{x_i^h - \min(x_i)}{\max(x_i) - \min(x_i)} \qquad (14)$$

where i is the index for the attributes 'x', j and h are indexes for the two catchments to be compared, m is the total numbers of attributes and N stands for normalized. The min and max are the minimum and the maximum values for the whole catchments of the attribute i. We assigned equal weightages to each attributes. All the regionalization methods are evaluated when catchments are recipients or the model parameters are donated to them from multi-donor catchments or a best single-donor catchment in the region.

Results

Performance measures (R^2 and R^2ln) for each catchment corresponding to the local calibration, parameter transfer to recipient catchments based on arbitrary best single-donor catchment(s) and parameter transfer based on best performing regionalization method(s) for the individual catchments are given in Table 3, Table 4 and Table 5 respectively. The best regionalization method(s) vary among the catchments, model structures and performance measures but the results generally showed that the physical similarity based regionalization methods performed best for the majority of the catchments. We identified the best single-donor catchment for each recipient catchment from arbitrary transfer of the optimal parameter vector of the local calibration from the 25 potential donor catchments without employing any regionalization method. It provides the maximum possible performance measure (PM) that can be obtained from the single-donor based transfer of optimal parameters among the catchments.

Tables 3 to 5 also indicate that for some catchments two or more regionalization methods performed equally good when more than one donor catchments have the top similarity rank (rank no. 1) to the recipient catchment in terms of two or more physical attributes spaces or when different best-donor catchments in different physical attributes spaces have the same optimal parameter set from the local calibration. In some cases for instance, for catchment 16-HBV-R^2, catchment 17-HBV-R^2ln and catchment 5-BGM-R^2, the multi-donor based RMedP regionalization method provided higher PM than the maximum possible performance that can be obtained by the best arbitrary single-donor catchment. In addition, the RmedP provided higher PM than the local calibration for some of the catchments (catchment 16-HBV-R^2 and catchment 5-BGM-R^2) in Table 4 and Table 5 respectively. These unlikely cases of higher PM from the median parameter vector than the local calibrated (optimal) parameter vectors from the DREAM algorithm does not guarantee better performance of the RMedP, which does not keep the correlation structure of the parameters.

Differences in performance among the catchments and models can be observed from the results in Tables 3 to 5. For instance, the local calibration and regionalization methods resulted in low performance (for instance $R^2 < 0.6$) for six catchments (2, 10, 11, 15, 22 and 25) for the Kirchmod and the BGM and for eleven catchments (4, 8, 10, 11, 14, 16, 20, 22, 23, 24 and 25) for the HBV. The R^2ln < 0.6 were observed for five catchments (8, 11, 14, 22 and 25) for the Kirchmod and the HBV and for eight catchments (8, 11, 14, 15, 20, 22, 23 and 25) for the BGM.

The performance measures obtained from the best regionalization method(s) of the individual catchments and their regional median and mean are comparable to the maximum possible PM and their regional median and mean values based on arbitrary transfer of local calibration parameters (Tables 3 to 5). However, the main objective of the regionalization is to derive the best performing regionalization method(s) for the whole region or sub-regions which generally involves compromises in the performances for the individual catchments. Therefore, further comparisons of the regionalization methods based on the regional performance in terms of the regional median and mean performance measures are presented in Table 6. Based on the regional median and mean R^2 and R^2ln for the three models, the MRWA regionalization method performed slightly better than the others.

However, the regional median and mean performance measures are affected by the results of poorly performing catchments. Therefore, improved comparisons which reduces the effects of poor simulation for some catchments for instance due to poor or unrepresentative input data are required. To this end, losses in the regional performances (i.e. losses in the regional median and mean performance measures) from the local calibration due to the regionalization are presented in Table 7. Based on Table 7, the MRWA also provided relatively better results in terms of the losses in the regional median and mean performance measures (i.e. lower losses). However, the PSCOMB provided better results in terms of the losses in the regional median and mean R^2 for the HBV model. The RMedP, PSH and PSOIL regionalization methods also provided equivalent performance for some of the cases (Tables 6 and 7). Therefore, based on their relative merits of regional performances, only the MRWA, RMedP, PSH, PSOIL and PSCOMB regionalization methods are considered for further comparative evaluations based on graphical plots.

Further evaluations based on the cumulative distributions functions (CDF) of performance measures rather than the regional median and mean statistics are presented in Figure 4(a) for R^2 only for Kirchmod and BGM and Figure 4(b) for R^2ln only for Kirchmod and HBV for the sake of clear presentation. However, a consistently best performing regionalization method is nearly indistinguishable. But, the PSCOMB gave better performances for relatively large proportions (i.e. 38 % to 46 % of the catchments) while the PSSOIL performed slightly better for the HBV model for the R^2ln performance measure, which gives higher weightages to low flows. A more improved evaluation based on the CDF of the loss or gain in performance measures are presented in Figure 5(a) and Figure 5(b). By looking at the higher portions of the CDF of Figure 5(a) and Figure 5(b), the PSCOMB and the PSSOIL appeared to perform better for R^2 and R^2ln respectively. The performance losses or gains for each catchment indicated that transfer of model parameters resulted in high losses in R^2 for catchments 2, 15, 16 and 17 and high losses in R^2ln for catchments 14 and 22 though the performance losses vary among the models. Catchments 2 and 17 are dominated by high relief. Hailegeorgis and Alfredsen (2014a) found catchment 22 to be an outlier in the region for regional regression model between flow percentiles and catchment areas. Hailegeorgis and Alfredsen (2014b, article in press) for the Gaulfoss watershed (catchment 6) found poor performances of parameter transfer to its internal sub-catchment of Lillebudal bru (catchment 14) especially for low flow simulation, which may be attributable to less representativeness of climate data or low quality streamflow data for catchment 14.

Discussion

The results of the present study revealed that selection of proper regionalization methods are dependent on the model structure used, the selected performance measures and evaluation metrics for the performance measures. There was also no consistent trend to explain the variations but generalization of the results to infer the best regional solution would be possible through comparative evaluations. The results reflect the limitations of the contemporary regionalization endeavors using precipitation-runoff modelling due to their high dependence on model structure, input data and procedures related to model calibration, which require comprehensive comparative evaluations for identification of suitable regionalization methods.

Evaluation of the regionalization methods, the PM and their evaluation metrics

Evaluations of regional performance based on the regional median and mean of R^2 and R^2ln (Table 6) and losses in the regional median and mean of R^2 and R^2ln from the regional calibration due to the regionalization (Table 7) revealed that the MRWA regionalization method performed slightly better than the other methods. These results generally indicated good performance of the regional calibration (MRWA) for prediction in ungauged basins in the region. The regional median and mean based evaluation in the present study is expected to favor the MRWA due to the facts that the regional calibration algorithm utilizes the streamflow data from all the catchments in the region and the regional calibration or the MRWA parameter vector is corresponding to the maximum regional weighted average of the performance measures (eq. 12). Moreover, removing the poorly simulated catchments from the analyses would be expected to benefit more the multi-donor based regionalization methods (MRWA and RMedP) than the single-donor regionalization methods. However, a study on the effects of the number of donor catchments (e.g. Oudin et al., 2008) on the multi-donor regionalization methods was not the objectives of the present study. Oudin et al. (2008) based on a regionalization of 913 catchments excluded poorly performing catchments (i.e. $R^2 < 0.70$) from the donor set. However, owing to the small number of catchments in the present study we did not exclude the poorly performing catchments from the analyses.

Alternatively we performed comparisons based on the regional losses or gains in the performance measures rather than the regional mean and median of the raw performance measures to remove the effects of poor simulation for some of the catchments due to the unequal representativeness and quality of the input data among the catchments. Though seemingly useful to draw general regional conclusions, the discriminatory power of the regional statistics (e.g. the regional median and mean) of the raw PM and their corresponding regional losses to identify differences among the regionalization methods appeared to be less. From the more comprehensive comparisons based on the CDF of the losses or gains in the performance measures (Figure 5(a) and Figure 5(b)), looking at the higher frequency portions revealed that physical similarity in the combined physical attributes spaces (PSCOMB) regionalization method performed slightly better for the R^2 while the physical similarity in soil types (PSSOIL) performed slightly better for the R^2ln. The R^2 gives higher weightage to high flow (which is mainly contributed from rainfall and snowmelt events) while the R^2ln gives higher weightage to the low flow (which is mainly contributed by the subsurface flow and affected by the soil hydraulic properties). Beldring et al. (1999; 2000) based on studies on boreal glacial tills dominated catchments noted significant contribution of the subsurface flow from the subsurface storage to the runoff hydrographs. The better performances of the soil types attribute would substantiate the need for data acquisition on soil hydraulic properties for further attempts of regionalization based on the physical similarity in the region.

The better performance of transfer of the local calibration parameters in the region based on the physical similarity generally indicated the physical control of runoff processes. Low performances of the nearest neighborhood (NN) regionalization method in terms of the regional mean and regional median (Table 6 and 7) indicated the lack of smooth spatial variations of dominant hydrological processes in the study region. However, the results also indicated the potential effects of geographical proximity on the performances of regionalization or the interaction between the geographic proximity and physical similarity which substantiates the need for further study how to integrally utilize the two methods for prediction in ungauged basins. For instance, catchments 15 and 24 are geographically farthest north and south respectively of the region and catchments 13, 19, 22, 10 and 16 are least close to the rest of the catchments in their geographical proximity. Catchments 23, 10, 2, 24, 14, 17, 25, 15 are far from the rest of catchments in terms of Euclidian distances in their combined physical attributes or the soil attribute spaces. The majority of the catchments, which are far in their geographical and/or physical attributes (i.e. combined or soil attributes) Euclidian distance space, are among those catchments which exhibited $R^2 < 0.6$ and/or $R^2\text{ln} < 0.6$ (Tables 3-5) and with high performance losses in R^2 and/in R^2ln. Therefore, evaluation of the representativeness of climate stations and the quality of streamflow data need to be scrutinized in regionalization endeavors.

The findings of the present study compared to the previous studies

The findings of the present study do not support the results of previous studies that reported better performances of the nearest neighbor (spatial proximity) than the physical similarity. Parajka et al. (2005) compared the CDF, median and scatter of R^2 and found that spatial proximity based on Kriging performed best followed by the physical similarity for Austrian catchments for daily runoff simulation by a semi-distributed conceptual HBV model. Oudin (2008) demonstrated the suitability of spatial proximity followed by physical similarity based on the median R^2 for daily lumped runoff simulation using GR4J and TOPMO models for non-snow catchments of dense stream networks in France. Zhang and Chiew (2009) based on comparisons in terms of the mean, median and percentiles of R^2 found that the spatial proximity approach performed better than a physical similarity approach for daily runoff simulation by lumped conceptual Xinanjiang and SimHyd rainfall-runoff models for Australian catchments. Parajka et al. (2013) from synthesis of previous regionalization studies and comparisons based on median R^2 noted that the spatial proximity and similarity methods performed best in humid catchments than simple averaging of parameters and parameter regression.

The main reasons for the differences among the findings of the present study and the previous works may be attributed to the differences in the hydrological behavior of the boreal snow-dominated catchments. Also, the present study is based on simulation of hourly runoff response compared to the daily or monthly simulation in the previous studies. The effects of the choices of performance measures and their evaluation metrics are also the causes of the differences as observed even with in the present study. Moreover, the results of the physical similarity is affected by the selection of the physical attributes and the similarity distance metrics. However, there are considerable similarities in the types of attributes used in the previous studies and the present study except that the hypsometric curves rather than the mean elevation and the CDF of the slope rather than the mean slope are used in the present study.

The density of hydro-climatic gauging networks of the present study can also have considerable influences on the performances of the nearest neighbor (NN) donor catchments based regionalization. Parajka et al (2005) and Oudin et al. (2008) noted the better performance of the nearest neighbor (NN) based regionalization contingent on the availability of dense stream-gauging networks and nested modelled catchments. Therefore, the sparse hydro-climatic gauging networks in the present study might have contributed to the less performance of the nearest neighbor method. Only two of the modelled catchments have precipitation gauging stations inside their boundary. The effects of the quality (less representativeness) of precipitation data may also result in pronounced effects on the high flow simulation that is highly influenced by rainfall and snowmelt events. Several studies (e.g. Croke and McIntyre, 2013;Yadav et al., 2007; Oudin et al., 2008; Samuel et al., 2011; Kim and Kaluarachchi, 2008) noted the importance of considering the representativeness of input climate data for prediction in ungauged basins. However, Samuel et al. (2011) demonstrated that coupling of spatial proximity and physical similarity methods provided better performance than model-averaging and regression methods even for less dense stream-gauging network in Ontario (Canada). Dense hourly measurement networks inside the unregulated catchments would generally benefit the target hourly runoff simulation in the study region.

The differences in the performance between the two performance measures used in the present study supports the previous studies by Wagener and McIntyre (2005) and Wagener and Wheater (2006), which demonstrated the incapability of the current model structures to simulate both high flow and low flow behaviors of catchments simultaneously with a single parameter set. Dependency of catchment similarity on the flow conditions was also demonstrated by Patil and Stieglitz (2011). The dependency of regionalization on the performance measures substantiates the need for selection of the PM depending on the modelling objectives (e.g. high flow, low flow and water balance simulation). Catchments which are similar in their rainfall and snow melt related high flow regime may not be necessarily similar in their low flow behavior or vice versa due to the likely regional similarities or differences in precipitation patterns and subsurface characteristics. Also, comprehensive evaluation of the PM based on different metrics such as regional median and mean and also CDF of the raw PM values and their losses or gains from the local calibration due to the regionalization found to be important for more improved identification of the regionalization methods.

Conclusions

We performed comparative evaluations of regionalization methods for transfer of local calibration parameters for hourly runoff simulation in boreal mid-Norway based on regional calibration of 26 catchments using three distributed runoff response routines, and two performance measures and various evaluation metrics.

For the study region and available set of hydro-climatological data, the results indicate that the best performing regionalization method(s) vary among the model structures, and performance measures and their evaluation metrics. Therefore, multiple models, and various PM and their evaluation metrics are important for comprehensive identification of suitable regionalization method(s) and hence for reduction of uncertainty in the PUB. For the mid Norway study region, the marginal differences between the regional calibration (MRWA) and the physical similarity indicate the suitability of both methods. The physical similarity based regionalization method is a more suitable regionalization method than the nearest neighbor, but the effects of hydro-climatic network density on the nearest neighbor requires further study.

The present study was the first attempt for regionalization of hourly runoff simulation in boreal mid Norway and regionalization for hourly runoff prediction are not common in literature. Therefore, the findings from the present study provide comprehensive information relevant to distributed continuous hourly simulation in ungauged basins to the general scientific community. Learning the differences or the similarities among catchments in different climate regimes and landscapes would also be important for the advancement of the PUB. Further regionalization study at hourly temporal resolution for a boreal region should focus on input data (both climate and streamflow data) from high-density and large pool of gauging networks and additional physical attributes related to soil hydraulic properties. Comparative regionalization studies for distributed hourly runoff simulation among catchments in different climate regimes and landscape features would provide further insights.

Acknowledgments

The Center for Environmental Design of Renewable Energy or CEDREN (http://www.cedren.no) funded the project. We obtained the climate data from the Norwegian Meteorological Institute (http://met.no/Klima), Statkraft, TrønderEnergi, Nord Trøndelag Elektrisitetsverk (NTE) and bioforsk (http://www.bioforsk.no). We

found the streamflow, hypsography and land use data from the Norwegian Water Resources and Energy Directorate (http://www.nve.no). We found the loose material (soil) and bedrock geology data from the Norwegian Geological Survey (http://www.ngu.no) and the stream networks from a 1:50000 map from the Norwegian Mapping Authority (http://www.statkart.no).

References

Ajami, H., Troch, A., Maddock,T., Meixner, T., and Eastoe, C. (2011). "Quantifying mountain block recharge by means of catchment-scale storage-discharge relationships." *Water Resour. Res*., 47, W04504, doi: 10.1029/2010WR009598.

Bastola, S., Ishidaira, H., and Takeuchi, K. (2008). "Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe." *J. Hydrol*., 357(3-4), 188-206.

Bardossy, A. (2007). "Calibration of hydrological model parameters for ungauged catchments." *Hydrol. Earth Syst. Sci*., 11, 703-710.

Beldring, S., Gottschalk, L., Seibert, J., and Tallaksen, L.M. (1999). "Distribution of soil moisture and groundwater levels at patch and catchment scales." *Agr. Forest Meteorol*. 98-99, 305-324.

Beldring, S., Gottschalk, L., Rodhe, A., Tallaksen, L.M. (2000). "Kinematic wave approximations to hillslope hydrological processes in tills." *Hydrol. Process*., 14, 727-745.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., and Vøkso, A. (2003). "Estimation of parameters in a distributed precipitation-runoff model for Norway." *Hydrol. Earth. Syst. Sc*., 7(3), 304-316.

Bell, V. A., and Moore, R. J. (1998). "A grid-based distributed flood forecasting model for use with weather radar data: Part 1. Formulation." *Hydrol. Earth. Syst. Sc.,* 2 (2-3), 265-281.

Bergström, S. (1976). "Development and application of a conceptual runoff model for Scandinavian catchments." *SMHI Reports RHO*, No. 7, Norrköping.

Beven, K. J., and Westerberg, I. (2011). "On red herrings and real herrings: disinformation and information in hydrological inference." *Hydrolog. Process*., 25, 1676-1680, DOI: 10.1002/hyp.7963.

Birkel, C., Soulsby, C., and Tetzlaff, D. (2011). “Modelling catchment-scale water storage dynamics: reconciling dynamic storage with tracer inferred passive storage.” *Hydrolog. Process.*, 25, 3924-3936.

Blöschl, G., and Sivapalan, M. (1995). “Scale issues in hydrological modeling: a review.” *Hydrolog. Process.,* 9, 251-290.

Box, G. E. P., and Cox, D. R. (1964). “An analysis of transformations.” *Journal of the Royal Statistical Society*, Series B 26, 211-252.

Brauer, C. C., Teuling, A. J.,Torfs, P. J. J. F., and Uijlenhoet, R. (2013) “Investigating Storage-Discharge Relations in a Lowland Catchment Using Hydrograph Fitting, Recession Analysis and Soil Moisture Data.” *Water Resour. Res*., 49(7), 4257–4264.

Cibin, R., Athira, P., Sudheer, K. P., and Chaubey, I. (2014). “Application of distributed hydrological models for predictions in ungauged basins: a method to quantify predictive uncertainty.” *Hydrolog. Process.,* 28, 2033-2045.

Croke B., and McIntyre, N. (2013). “Data-based perceptions on Predictions in Ungauged Basins.” *Hydrology Research*, 44, 399-400.

Dingman, S., L. (1978). “Drainage Density and Streamflow: A Closer Look.” *Water Resour. Res*., 14 (6), 1183-1187.

Donnelly, C., Dahne, J., Lindström, G., Rosberg, J., Strömqvist, J., Pers, C., Yang W., and Arheimer, B. (2009). “An evaluation of multi-basin hydrological modelling for predictions in ungauged basins. In: K. Yilmaz, et al., eds. New approaches to hydrological prediction in data sparse regions.” Wallingford: IAHS Press, *IAHS Publ*. 333,112–120.

Dunne, T., Black, R. D. (1970a). “An experimental investigation of runoff production in permeable soils.” *Water Resour. Res*. 6(2), 478-490.

Dunne, T., and Black, R. D. (1970b). “Partial area contributions to storm runoff in a small New England watershed.” *Water Resour. Res*. 6(5), 1296-1311.

Engeland K., and Gottschalk L. (2002). “Bayesian estimation of parameter in a regional hydrologic model.” *Hydrol. Earth System Sci*., 6(5), 883-898.

Engeland, K., Braud, I., Gottschalk, L., and Leblois, E. (2006). “Multi-objective regional modelling.” *J. Hydrol*., 327, 339-351

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A. (2000). “Regional calibration of a watershed model.” *Hydrolog. Sci. J.*, 45, 689-707.

Goswami, M., O'Connor, K. M., and Bhattarai, K. P. (2007). "Development of regionalization procedures using a multi-model approach for flow simulation in an ungauged catchment." *J. Hydrol.*, 333(2-4), 517–531.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1998). "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information." *Water Resour. Res.*, 34(4), 751-763.

Götzinger, J., and Bárdossy, A. (2007). "Comparison of four regionalization methods for a distributed hydrological model." *J. Hydrol*., 333, 374-384.

Hailegeorgis, Teklu T., and Alfredsen, Knut (2014a). "Regional statistical and precipitation-runoff modelling for ecological applications: prediction of hourly streamflow in regulated rivers and ungauged basins." Ecohydraulics 2014, *Proceedings of the 10th International Symposium on Ecohydraulics (ISE)*, Trodheim, Norway, 23-27 June.

Hailegeorgis, Teklu T., and Alfredsen, Knut (2014b). "Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments." *Hydrology Research*, article in press.

Halldin, S., Gottschalk, L., Gryning, S.E., Jochum, A., Lundin, L.C., and Van de Griend, A.A. (1999). "Energy, water and carbon exchange in a boreal forest-NOPEX experiences." *Agric. For. Meteorol*. 98-99, 5-29.

He, Y., Bárdossy, A., and Zehe, E. (2011). "A review of regionalisation for continuous streamflow simulation." Hydrol. Earth Syst. Sci., 15(11), 3539-3553.

Horton, R. E. (1932). "Drainage-basin characteristics", *Eos Trans*. AGU 3, 350-361.

Horton, R.E. (1933). "The role of infiltration in the hydrologic cycle." *Trans. Am. Geophys. Union* 14, 446-460.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C. (2013). "A decade of Predictions in Ungauged Basins (PUB)-a review." *Hydrolog. Sci. J*, 58(6), 1198-1255, DOI:10.1080/02626667.2013.803183.

Kim, U., and Kaluarachchi, J. J. (2008). “Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia.” *J. Hydrol.,* 362(1-2), 39-56.

Kirchner, J.W. (2009). “Catchments as simple dynamical systems: catchment characterization, rainfall–runoff modeling, and doing hydrology backward.” *Water Resour. Res*., 45, W02429.

Kolberg, S. A., and Bruland, O. (2012). “ENKI-An Open Source environmental modelling platform.” *Geophys. Res. Abstracts* 14, EGU2012-13630, EGU General Assembly.

Kolberg, S. A., and Gottschalk, L. (2006) “Updating of snow depletion curve with remote sensing data.” *Hydrol. Process*. 20(11), 2363-2380.

Kokkonen, T., Jakeman, A. J., Young, P. C., and Koivusalo, H. J. (2003). “Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina.” *Hydrol. Process*., 17(11), 2219-2238.

Krier, R. Matgen, P. Goergen, K., Pfister, L., Hoffmann, L., Kirchner, J. W., Uhlenbrook, S., and Savenije, H. H. G. (2012). “Inferring catchment precipitation by doing hydrology backward: A test in 24 small and mesoscale catchments in Luxembourg.” *Water Resour. Res*. 48, W10525, doi: 10.1029/2011WR010657.

Lee, H., McIntyre, N. Wheater, H., and Young, A. (2005). “Selection of conceptual models for regionalization of the rainfall-runoff relationship.” *J. Hydrol*., 312, 125-147.

Li, H., Zhang, Y., Chiew, F. H. S., and Shiguo, X. (2009). “Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index.” *J. Hydrol*., 370(1-4), 155-162.

Littlewood, I. G., Barry, F., and Croke, W. (2013). “Effects of data time-step on the accuracy of calibrated rainfall–streamflow model parameters: practical aspects of uncertainty reduction” *Hydrology Research,* 44 (3), 430-440.

Madsen, H. (2003). “Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives.” *Adv. Wat. Res*., 26, 205-216.

Makungo, R., Odiyo, J. O., Ndiritu, J. G., and Mwaka, B. (2010). "Rainfall–runoff modelling approach for ungauged catchments: A case study of Nzhelele River sub-quaternary catchment." *Phys. Chem. Earth*, 35(13-14), 596-607.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). "Ensemble predictions of runoff in ungauged catchments." *Water Resour. Res*., 41, W12434, doi: 10.1029/2005WR004289.

Merz, R., and Bloschl, G. (2004). "Regionalization of catchment model parameters," *J. Hydrol*., 287, 95-123.

Motovilov, Y.G. Gottschalk, L., Engeland, K., and Rodhe, A. (1999). "Validation of a distributed hydrological model against spatial observations." *Agr. Forest Meteorol*. 98-99, 257-277.

Muleta, M. K. (2012). "Model Performance Sensitivity to Objective Function during Automated Calibrations." *J. Hydrol. Eng*. 17 (6), 756-767.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models, I. A discussion of principles." *J. Hydrol*. 10, 228-290.

Oudin, L., Andre´assian, V., Perrin, C., Michel, C., and Le, M., N. (2008). "Spatial proximity, physical similarity, regression and un-gaged catchments: A comparison of regionalization approaches based on 913 French catchments." *Water Resour. Res*., 44, W03413, Doi: 10.1029/2007WR006240.

Oudin, L., Kay, A., Andre´assian, V., and Perrin, C. (2010). "Are seemingly physically similar catchments truly hydrologically similar?." *Water Resour. Res*., 46, W11558, doi: 10.1029/2009WR008887.

Parajka, J., Merz, R., and Blöschl, G. (2005). "A comparison of regionalization methods for catchment model parameters." *Hydrol. Earth Syst. Sci.*, 9, 157-171.

Parajka, J., Blöschl, G., and Merz, R. (2007). "Regional calibration of catchment models: potential for ungauged catchments." *Water Resour. Res.*, 43, W06406.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G. (2013). "Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies." *Hydrol. Earth Syst. Sci*., 17, 1783-1795.

Patil, S., and Stieglitz, M. (2011). "Hydrologic similarity among catchments under variable flow conditions." *Hydrol. Earth Syst. Sci.*, 15, 989-997.

Priestley, C.H.B., and Taylor, R.J. (1972). “On the assessment of surface heat flux and evaporation using large-scale parameters.” *Mon. Weather Rev.,* 100, 81-82.

Razavi, T. and Coulibaly, P. (2012). “Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods” *J. Hydrol. Eng.,* 17, 756-767.

Reed, S., Koren, V., Smith M., Zhang, Z., Moreda, F., Se, Dong-Jun, and DMIP Participants (2004). “Overall distributed model intercomparison project results.” *J. Hydrol*., 298, 27-60.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S. (2009). “Optimization of a similarity measure for estimating ungauged streamflow.” *Water Resour. Res*., 45(10), W10423.

Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2011). “Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods.” *J. Hydrol. Eng*., 16(5), 447-459.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). “Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA.” *Hydrol. Earth Syst. Sci*, 15, 2895-2911.

Seibert, J. (1999). “Regionalization of parameters for a conceptual rainfall runoff model.” Agric. For. Meteorol., 98-99, 279-293.

Sivapalan, M, Takeuchi, K., Franks, S.W., Gupta, V.K., Harambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., mendiondo, E.,M., O’connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe, E. (2003). “IAHS decade on predictions in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences.” *Hydrol. Sci. J*., 48(6), 857-880.

Teuling, J., Lehner, I., Kirchner, J. W., and Seneviratne, S. I. (2010). “Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment.” *Water Resour. Res*. 46, W10502, doi:10.1029/2009WR008777.

Vaze, J., Zhang, Y., Chiew, F.H.S., Wang, B., and Teng, J. (2013). “Regional calibration against multiple data sources to predict streamflow.” Climate and Land Surface Changes in Hydrology *Proceedings of H01, IAHS-IAPSO-IASPEI Assembly*, Gothenburg, Sweden, July (IAHS Publ. 359, 2013), 165-170.

Viglione, A., Parajka, J., Rogger, M. Salinas, J. L., Laaha, G., Sivapalan, M., and Bloschl,G. (2013). "Comparative assessment of predictions in ungauged basins-Part 3: Runoff signatures in Austria." *Hydrol. Earth Syst. Sci. Discuss*., 10, 449-485.

Vrugt, J.A., Gupta, H.V., Bouten, W., and Sorooshian, S. (2003). "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." *Water Resour. Res*. 39 (8), 1201, doi: 10.1029/2002WR001642.

Vrugt, J.A., Bouten, W., Gupta, H.V., and Sorooshian, S. (2002). "Toward improved identifiability of hydrologic model parameters: The information content of experimental data". *Water Resour. Res.* 38 (12), doi:10.1029/2001WR001118.

Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M., and Higdon D. (2009). "Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling." *Journal of Nonlinear Sciences and Numerical Simulation,* 10(3), 273-290.

Wagener, T., and Wheater, H. S. (2006). "Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty." *J. Hydrol*., 320(1-2), 132-154.

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

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

Zhang, Y. and Chiew, F. H. S. (2009). "Relative merits of different methods for runoff predictions in ungauged catchments." *Water Resour. Res*., 45, W07412, doi: 10.1029/2008WR007504.

Zięba, A. (2010). "Effective number of observations and unbiased estimators of variance for auto correlated data-an overview." *Metrol. Meas. Syst*. XVII (1), 3-16.

Tables

Table 1. Characteristics of the study catchments.

Catchment No.	Catchment name	Station No. (NVE)	Station altitude masl	Catchment area, km^2	Drainage density km/km^2	Mean hourly flow, m^3/s	Catchment averaged P, mm
1	Dillfoss	127.13	25	480	2.1	16.6	1332.94
2	Driva v/Risefoss	109.9	550	745	0.8	14.5	1029.06
3	Eggafoss	122.11	330	668	2.0	16.8	1041.80
4	Embrethølen	139.26	160	495	1.8	20.1	1152.83
5	Feren	124.13	401	220	2.0	9.3	1354.49
6	Gaulfoss	122.9	45	3090	1.9	83.5	981.87
7	Gisnås	121.29	580	95	1.2	2.4	928.76
8	Hugdal bru	122.17	135	546	1.6	12.5	913.79
9	Høggås bru	124.2	97	495	2.1	20.6	1361.58
10	Isa v/Morstøl bru	103.2	103	44	1.0	3.1	1217.40
11	Kjeldstad i Garb.	123.31	173	145	1.8	5.7	1441.30
12	Krinsvatn	133.7	87	207	1.2	13.1	1860.95
13	Lenglingen	308.1	354	450	1.7	14.7	1173.91
14	Lillebudal bru	122.14	515	168	3.1	4.9	930.07
15	Murusjø	307.5	311	346	1.8	8.5	1192.55
16	Osenelv v/Øren	105.1	12	138	1.3	6.4	1337.01
17	Rauma v/Horgheim	103.4	60	1100	0.9	36.6	1186.46
18	Rinna	112.8	460	91	1.2	3.6	1344.66
19	Skjellbreivatn	139.25	354	546	1.8	19.2	1126.52
20	Søya v/Melhus	111.9	40	138	1.9	8.4	1676.71
21	Støafoss	128.5	80	477	1.9	14.8	1268.16
22	Trangen	139.35	137	852	1.8	35.1	1115.87
23	Valen	117.4	10	39	1.2	1.2	1385.99
24	Valldøla v/Alstad	100.1	265	226	0.9	11.0	1246.80
25	Vistdal	104.23	50	67	2.2	3.9	1258.18
26	Øyungen	138.1	103	239	1.7	12.1	1353.40

We estimated catchment averaged precipitation (P) and temperature (T) by the inverse distance weighing (IDW) interpolation.

Table 2. Lists of calibrated parameters and their uniform prior ranges for the models.

Parameters	Description	Kirchmod	HBV	BGM	Units	Uniform prior ranges
Snow						
TX	Threshold temperature	x	x	x	^{0}C	[-3,2]
WS	Snow melt sensitivity to wind speed	x	x	x	-	[1,6]
Soil and evapotranspiration						
FC	Maximum soil moisture (SM) or field capacity		x		mm	[50,600]
LP	(SM/FC) value at and above which AET= PET		x		-	[0.5,0.99]
β	Shape parameter of the curve partioning the infiltration in to UZ recharge and change in soil moisture		x		-	[0.5,5]
Runoff response						
EvR	Discharge at which AET equals 0.95*PET	x			mm h^{-1}	[0.1,6.15]
β_0	Regression parameter 1	x			-	[-8,0]
β_1	Regression parameter 2	x			-	[-1,1]
k_1	Recession coefficient of the upper reservoir		x		d^{-1}	[0.001,1.5]
k_0	Recession coefficient of the lower reservoir		x		d^{-1}	[0.0005,0.5]
n_u	Non-linearity exponent of the upper reservoir		x		-	[0.2,5]
PERC	Percolation rate from the upper to lower reservoir		x		mm d^{-1}	[0,6]
S_{max}	Maximum storage capacity			x	mm	[150,1000]
I_c	Infiltration capacity of the soil			x	mm h^{-1}	[0.1,40]
k	Coefficient of storage-drainage relationship			x	$mm^{1-n}h^{-1}$	[10^{-7}-10^{-3}]
n	Exponent of storage-drainage relationship			x	-	[0.2,5.0]
Routing						
V	Velocity of flow	x	x	x	m s^{-1}	[0.25,3.5]

Table 3. The R^2 and R^2ln performance measures for the local calibration, best arbitrary single-donor and best regionalization method(s) for the Kirchmod.

Catchment No.	Local calib.		Best arbitrary single-donor		Best regionalization method(s) for the individual catchment	
	R^2	R^2ln	R^2	R^2ln	R^2	R^2ln
1	0.74	0.84	0.73	0.83	0.73	0.83
2	0.49	0.69	0.28	0.66	0.37	0.66
3	0.81	0.87	0.75	0.87	0.79	0.87
4	0.71	0.77	0.70	0.72	0.70	0.72
5	0.70	0.73	0.67	0.70	0.64	0.70
6	0.83	0.89	0.83	0.88	0.83	0.87
7	0.64	0.68	0.81	0.68	0.64	0.68
8	0.60	0.54	0.58	0.54	0.58	0.53
9	0.73	0.79	0.72	0.78	0.71	0.78
10	0.58	0.77	0.58	0.74	0.58	0.72
11	0.38	0.42	0.34	0.42	0.38	0.41
12	0.75	0.82	0.71	0.82	0.71	0.82
13	0.81	0.84	0.73	0.84	0.75	0.84
14	0.58	0.52	0.57	0.28	0.57	-0.13
15	0.43	0.60	0.22	0.58	0.37	0.58
16	0.67	0.82	0.55	0.77	0.64	0.81
17	0.78	0.86	0.68	0.82	0.76	0.82
18	0.63	0.75	0.61	0.75	0.63	0.75
19	0.81	0.83	0.75	0.79	0.75	0.79
20	0.67	0.65	0.58	0.62	0.65	0.62
21	0.72	0.70	0.71	0.79	0.72	0.69
22	0.51	0.32	0.51	0.30	0.51	0.22
23	0.61	0.64	0.44	0.49	0.44	0.49
24	0.74	0.73	0.74	0.73	0.74	0.73
25	0.51	0.54	0.54	0.54	0.51	0.54
26	0.72	0.83	0.71	0.81	0.71	0.79
Regional median	0.69	0.74	0.68	0.73	0.65	0.72
Regional mean	0.66	0.71	0.62	0.68	0.63	0.66

Table 4. The R^2 and R^2ln performance measures for the local calibration, best arbitrary single-donor and best regionalization method(s) for the HBV model.

Catchment No.	Local calib.		Best arbitrary single-donor		Best regionalization method(s) for the individual catchment	
	R^2	R^2ln	R^2	R^2ln	R^2	R^2ln
1	0.68	0.82	0.68	0.82	0.68	0.82
2	0.70	0.75	0.42	0.72	0.42	0.71
3	0.72	0.84	0.72	0.83	0.72	0.83
4	0.56	0.70	0.52	0.70	0.52	0.70
5	0.64	0.63	0.59	0.61	0.59	0.60
6	0.79	0.85	0.76	0.85	0.76	0.85
7	0.64	0.66	0.64	0.64	0.64	0.63
8	0.54	0.53	0.54	0.50	0.52	0.50
9	0.66	0.75	0.64	0.75	0.63	0.75
10	0.42	0.69	0.41	0.68	0.41	0.68
11	0.34	0.44	0.32	0.41	0.32	0.41
12	0.65	0.81	0.64	0.81	0.64	0.81
13	0.76	0.82	0.71	0.82	0.71	0.82
14	0.48	0.33	0.47	-0.13	0.47	-0.13
15	0.74	0.73	0.54	0.66	0.59	0.20
16	0.47	0.62	0.61	0.61	0.69*	0.54
17	0.86	0.88	0.81	0.85	0.80	0.86
18	0.61	0.72	0.61	0.69	0.61	0.69
19	0.85	0.86	0.76	0.86	0.76	0.86
20	0.52	0.64	0.52	0.63	0.52	0.61
21	0.62	0.68	0.62	0.68	0.62	0.68
22	0.55	0.40	0.53	0.40	0.53	0.32
23	0.56	0.62	0.48	0.62	0.48	0.62
24	0.58	0.67	0.51	0.66	0.51	0.64
25	0.29	0.44	0.27	0.35	0.28	0.34
26	0.63	0.87	0.63	0.87	0.63	0.87
Regional median	0.63	0.70	0.60	0.68	0.60	0.68
Regional mean	0.61	0.68	0.57	0.65	0.58	0.62

*PM of the multi-donor RmedP regionalization method is higher than the local calibration PM.

Table 5. The R^2 and R^2ln performance measures for the local calibration, arbitrary best donor and best regionalization method(s) for the BGM model.

Catchment No.	Local calib.		Best arbitrary single-donor		Best regionalization method(s) for the individual catchment	
	R^2	R^2ln	R^2	R^2ln	R^2	R^2ln
1	0.72	0.78	0.69	0.78	0.69	0.78
2	0.49	0.68	0.30	0.68	0.49	0.58
3	0.81	0.85	0.81	0.85	0.81	0.85
4	0.70	0.76	0.70	0.74	0.65	0.74
5	0.63	0.62	0.58	0.62	0.72*	0.62
6	0.83	0.86	0.81	0.82	0.81	0.84
7	0.68	0.64	0.67	0.64	0.67	0.64
8	0.58	0.51	0.56	0.50	0.56	0.50
9	0.73	0.75	0.72	0.73	0.72	0.72
10	0.59	0.74	0.59	0.73	0.55	0.72
11	0.39	0.40	0.36	0.38	0.38	0.38
12	0.76	0.79	0.71	0.79	0.71	0.79
13	0.80	0.78	0.76	0.78	0.76	0.78
14	0.58	0.19	0.58	0.07	0.57	0.07
15	0.29	0.50	0.02	0.46	0.02	0.46
16	0.65	0.79	0.62	0.77	0.54	0.77
17	0.79	0.81	0.76	0.79	0.70	0.73
18	0.68	0.74	0.68	0.74	0.68	0.74
19	0.72	0.72	0.69	0.72	0.69	0.72
20	0.68	0.59	0.67	0.59	0.67	0.58
21	0.72	0.68	0.70	0.68	0.70	0.68
22	0.54	0.33	0.46	0.33	0.50	0.33
23	0.58	0.53	0.56	0.43	0.43	0.43
24	0.75	0.69	0.62	0.67	0.60	0.67
25	0.55	0.53	0.54	0.50	0.54	0.50
26	0.72	0.75	0.71	0.75	0.71	0.75
Regional median	0.68	0.71	0.67	0.70	0.67	0.70
Regional mean	0.65	0.65	0.61	0.64	0.61	0.63

*PM of the multi-donor RmedP regionalization method is higher than the local calibration PM.

Table 6. Regional median and mean performance measures (PM) for the local calibration (local calib.) and the regionalization methods.

	R^2			R^2 ln		
Description	Kirchmod	HBV	BGM	Kirchmod	HBV	BGM
Regional median performances						
Local calib.	0.69	0.63	0.68	0.74	0.70	0.71
MRWA	**0.54**	0.49	**0.53**	0.67	**0.64**	**0.63**
RMedP	**0.54**	**0.50**	0.52	**0.69**	0.63	0.59
NN	0.51	0.47	0.50	0.55	0.59	0.53
PSH	0.51	**0.50**	**0.53**	0.53	0.51	0.48
PSLU	0.45	0.44	0.48	0.62	0.60	0.57
PSDD	0.51	0.44	0.45	0.61	0.56	0.54
PSCA	0.43	0.39	0.52	0.64	0.59	0.59
PSSL	0.49	0.48	0.51	0.61	0.55	0.54
PSBRG	0.52	**0.50**	0.46	0.55	0.55	0.49
PSSOIL	0.53	**0.50**	**0.53**	0.67	0.63	0.56
PSCOMB	**0.54**	**0.50**	0.52	0.60	0.60	0.55
Regional mean performances						
Local calib.	0.66	0.61	0.65	0.71	0.68	0.65
MRWA	**0.51**	0.48	**0.50**	**0.59**	**0.56**	**0.53**
RMedP	0.49	0.48	0.47	0.58	0.54	0.43
NN	0.35	0.45	0.39	0.52	0.54	0.48
PSH	0.45	0.47	0.47	0.51	0.52	0.48
PSLU	0.40	0.42	0.41	0.48	0.42	0.40
PSDD	0.35	0.40	0.27	0.51	0.46	0.45
PSCA	0.30	0.36	0.33	0.54	0.49	0.50
PSSL	0.35	0.40	0.33	0.51	0.47	0.47
PSBRG	0.41	0.43	0.39	0.51	0.48	0.48
PSSOIL	0.46	0.48	0.44	0.58	**0.56**	0.51
PSCOMB	0.48	**0.50**	0.47	0.56	0.51	**0.53**

The best performances are given in bold.

Table 7. Losses in the regional mean and median PM from the local calibration due to the regionalization.

Regionalization methods	Models	Loss in regional mean R^2	Loss in regional median R^2	Loss in regional mean R^2ln	Loss in regional median R^2ln
MRWA	Kirchmod	**-0.154**	-0.143	**-0.120**	-0.073
MRWA	HBV	-0.128	-0.135	**-0.119**	**-0.056**
MRWA	BGM	**-0.157**	-0.155	**-0.122**	**-0.078**
RMedP	Kirchmod	-0.175	-0.151	-0.133	**-0.049**
RMedP	HBV	-0.132	-0.125	-0.141	-0.068
RMedP	BGM	-0.187	-0.166	-0.219	-0.116
NN	Kirchmod	-0.309	-0.177	-0.186	-0.185
NN	HBV	-0.159	-0.157	-0.145	-0.108
NN	BGM	-0.260	-0.184	-0.170	-0.174
PSH	Kirchmod	-0.202	-0.126	-0.189	-0.141
PSH	HBV	-0.137	-0.107	-0.156	-0.096
PSH	BGM	-0.182	**-0.125**	-0.163	-0.146
PSLU	Kirchmod	-0.256	-0.200	-0.232	-0.103
PSLU	HBV	-0.192	-0.117	-0.264	-0.092
PSLU	BGM	-0.241	-0.181	-0.252	-0.091
PSDD	Kirchmod	-0.308	-0.173	-0.196	-0.108
PSDD	HBV	-0.213	-0.145	-0.221	-0.135
PSDD	BGM	-0.380	-0.207	-0.202	-0.125
PSCA	Kirchmod	-0.354	-0.322	-0.171	-0.113
PSCA	HBV	-0.245	-0.147	-0.197	-0.129
PSCA	BGM	-0.319	-0.252	-0.153	-0.140
PSSL	Kirchmod	-0.313	-0.128	-0.200	-0.100
PSSL	HBV	-0.207	-0.077	-0.208	-0.164
PSSL	BGM	-0.325	-0.143	-0.185	-0.168
PSBRG	Kirchmod	-0.253	-0.127	-0.200	-0.164
PSBRG	HBV	-0.178	-0.101	-0.206	-0.144
PSBRG	BGM	-0.267	-0.240	-0.177	-0.149
PSSOIL	Kirchmod	-0.201	-0.130	-0.130	-0.068
PSSOIL	HBV	-0.134	-0.129	-0.127	-0.071
PSSOIL	BGM	-0.213	-0.191	-0.140	-0.081
PSCOMB	Kirchmod	-0.176	**-0.102**	-0.152	-0.084
PSCOMB	HBV	**-0.105**	**-0.063**	-0.176	-0.103
PSCOMB	BGM	-0.181	-0.144	-0.126	-0.093

Smaller losses for the models and regionalization methods are shown in bold and colored respectively.

Figures

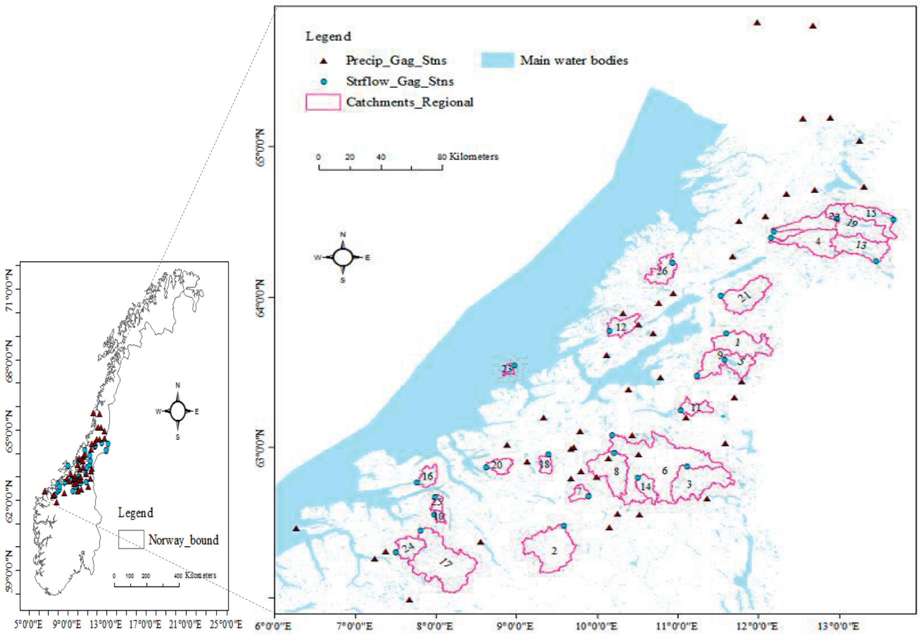


Figure 1. Locations of the study region and the hydro-climatic stations.

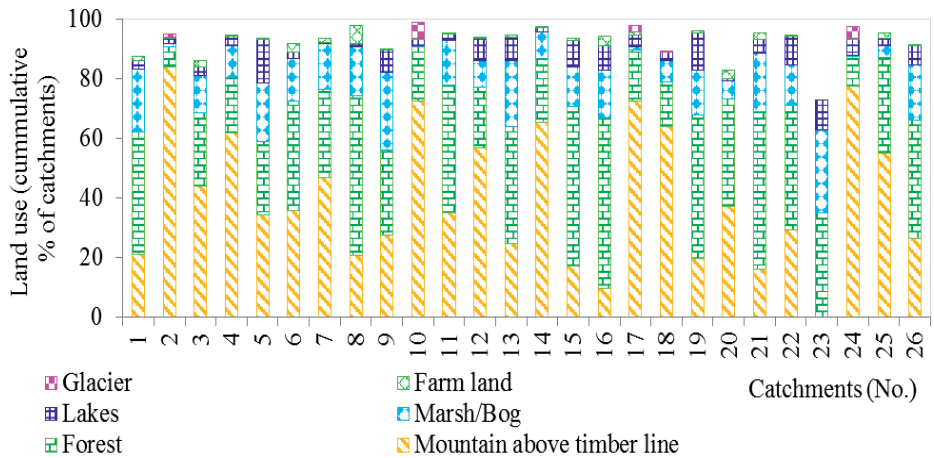


Figure 2. Major land use/cover types or classes (given in cumulative percentages of the catchments).

The sum less than 100 indicate some areas are unclassified by the source institute.

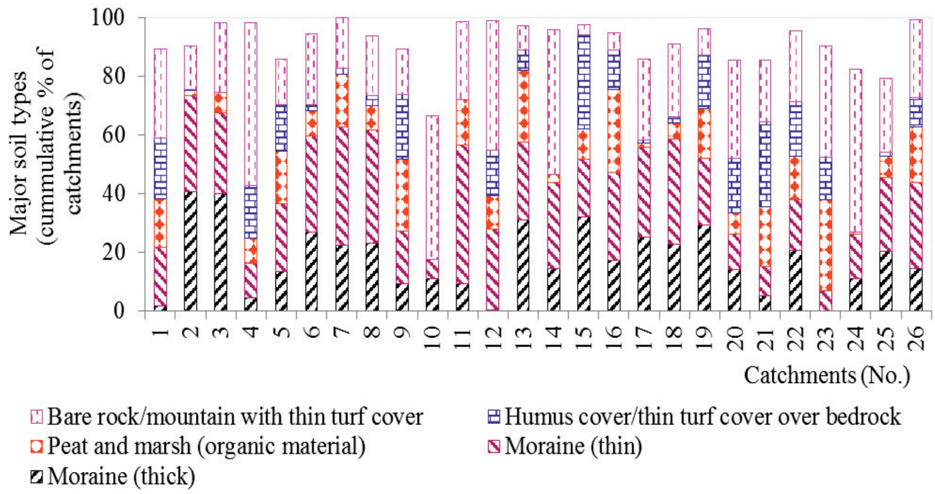


Figure 3. Major soil types (given in cumulative percentages of the catchments).

The sum less than 100 indicate some areas are unclassified by the source institute.

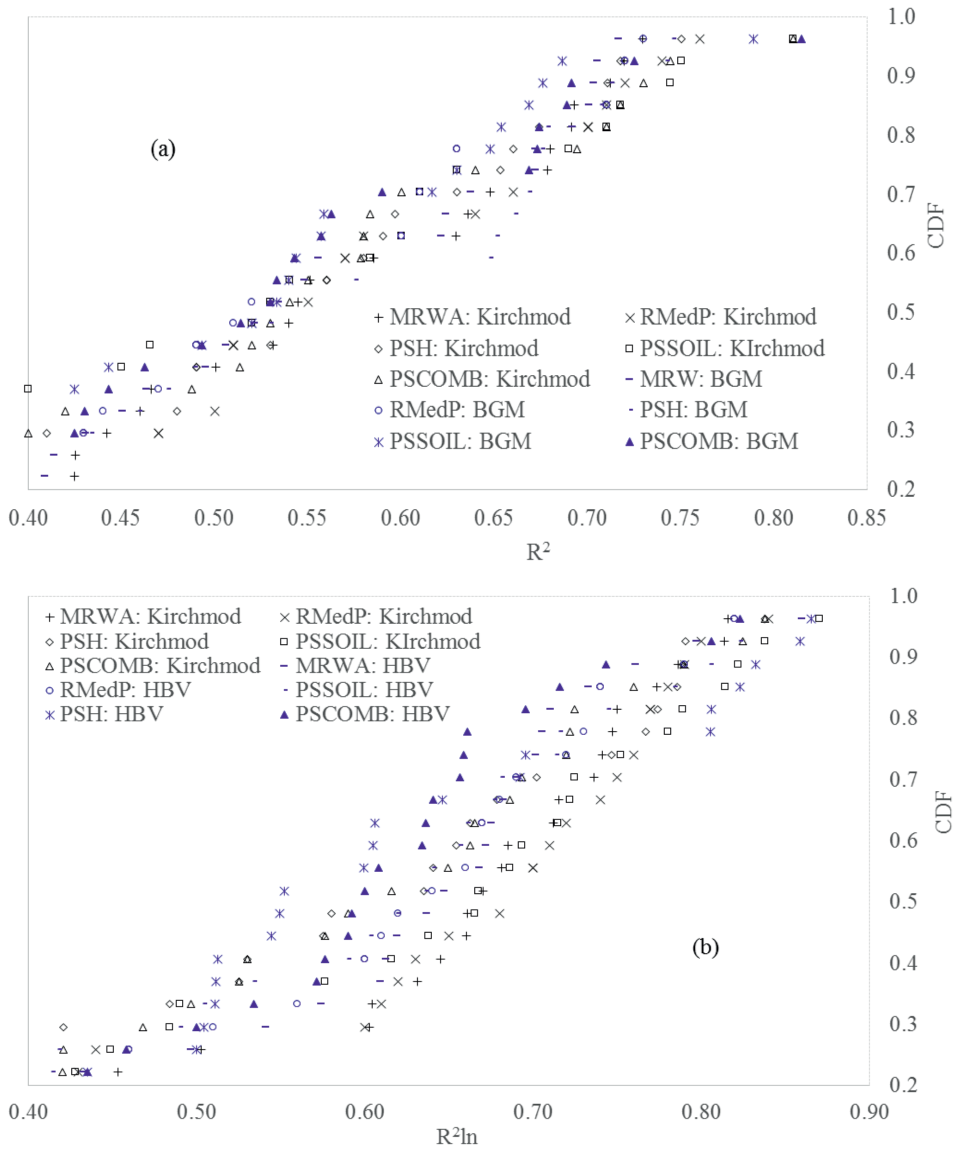


Figure 4. Cumulative distribution functions (CDF) of the performance measures for the local calibration and the regionalization methods (a) for R^2 for Kirchmod and BGM and (b) for R^2ln for Kirchmod and HBV.

Lower portions of the PM are not displayed for clarity of the figures.

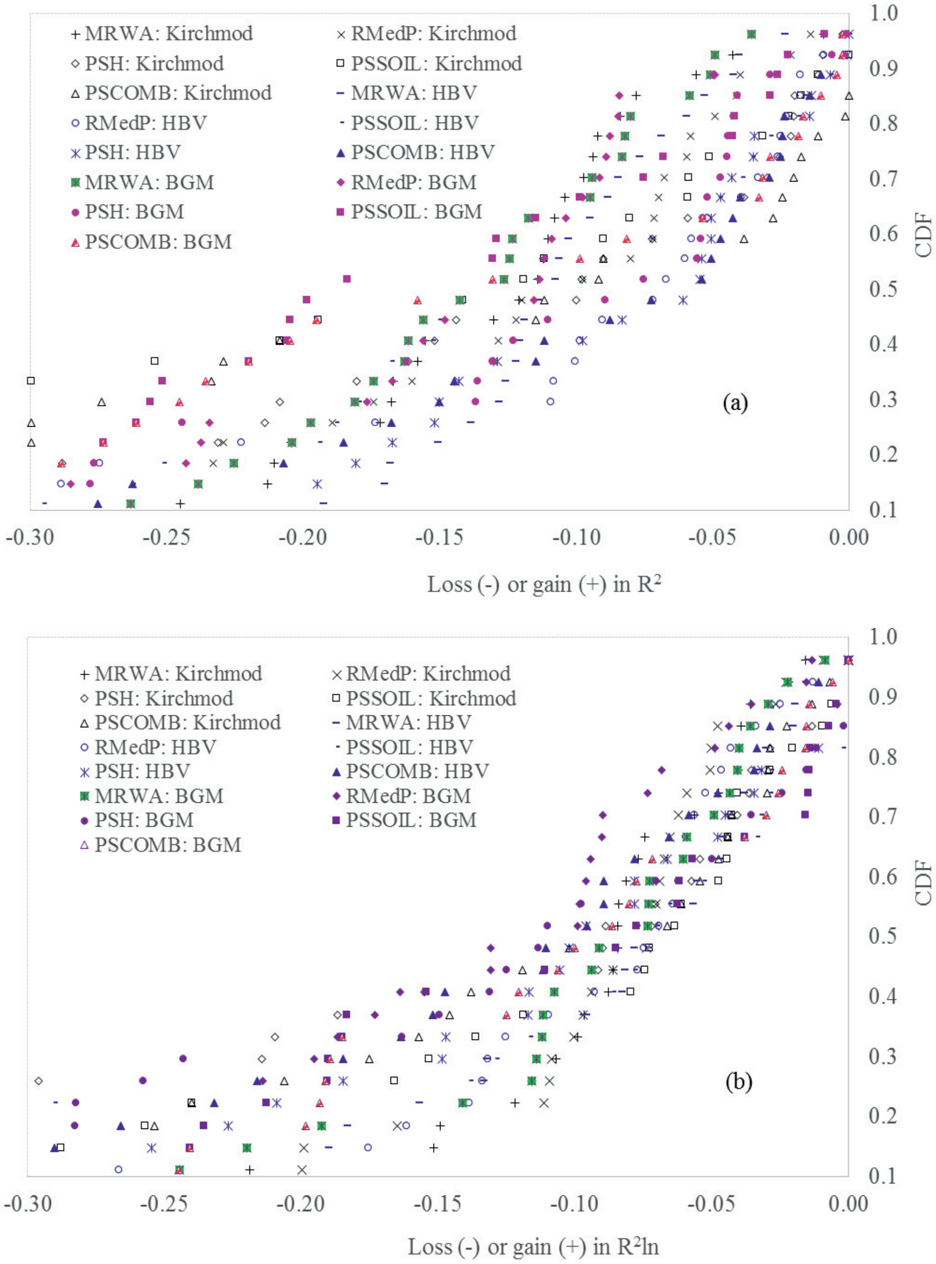


Figure 5. Cumulative distribution functions (CDF) of the losses or gains in the performance measures from the local calibration due to the regionalization (a) R^2 and (b) R^2ln.

Some portions of larger losses in the PM are not displayed for clarity of the figures.

Paper 5 (P5)

Multi-basin and regional calibration based identification of distributed Precipitation-Runoff models for prediction of hourly streamflow on 26 catchments in mid-Norway

Teklu T. Hailegeorgis and Knut Alfredsen

Paper V

Is not included due to copyright

Paper 6 (P6)

Regional statistical and Precipitation-Runoff modelling for ecological applications: prediction of hourly streamflow in regulated Rivers and ungauged basins

Teklu T. Hailegeorgis and Knut Alfredsen

We presented some portion of this study on the poster session of the 10th International Symposium on Ecohydraulics 2014, Trondheim, 23-27 June 2014.

Paper VI

REGIONAL STATISTICAL AND PRECIPITATION-RUNOFF MODELLING FOR ECOLOGICAL APPLICATIONS: PREDICTION OF HOURLY STREAMFLOW IN REGULATED RIVERS AND UNGAUGED BASINS

Teklu T. Hailegeorgis

Norwegian University of Science and Technology (NTNU), Department of Hydraulic and Environmental Engineering, S.P. Andersens vei 5, N-7491, Trondheim, Norway.

Knut Alfredsen

Norwegian University of Science and Technology (NTNU), Department of Hydraulic and Environmental Engineering, S.P. Andersens vei 5, N-7491, Trondheim, Norway.

Abstract

River regulation for hydropower may create significant alterations of natural river flow characteristics or regime that have profound ecological, geomorphologic and hydraulic repercussions. Pre-regulation or 'baseline' natural flow regime can be obtained from pre-regulation observed streamflow if available, which frequently is not the case. Moreover, humanities have regulated rivers for generations and hence pre-regulation natural flow characteristics may not represent contemporary post-regulation natural flow characteristics mainly due to land use and climate change. In addition, it is impossible to observe post-regulation natural flow directly and hence local or at-site calibration of the P-R model is not possible for regulated river reaches. Furthermore, previous ecological studies mainly focused on coarse temporal resolutions such as annual, monthly and daily streamflow while contemporary operation rules or practices in regulated rivers for instance hydropeaking require examination of relevant ecological indicators at high temporal resolution. To obtain information on temporal flow pattern alterations compared to the natural (i.e. pretending no regulation) to study the impacts of regulations, there need to be a methodology for predicting a time series of natural streamflow that excludes the effects of regulation. In addition, prediction in ungauged basins (PUB) where ecological data are available and for environmental flow assessment are required.

Therefore, in the present study regional prediction through transfer of information

from gauged donor catchment(s) in the region based on both regional regression model and regional calibration of a 'top-down' Precipitation-Runoff (P-R) model were conducted for prediction of hourly time series of 'unimpaired' or natural flow and flow duration curves (FDCs) downstream of a regulation. We evaluated the models and the regionalization methods based on the 'proxy ungauged catchment' technique. For the study region of boreal Norway, the results indicated that prediction of hourly streamflow from simple regression model relating streamflow percentiles and catchment drainage area outperformed prediction by regional calibration of the P-R model. Therefore, the approach has significant advantage for derivation of ecologically relevant indices to study ecological impacts of river regulation and hydropeaking operation, and for the PUB. Regionalization method based on spatial proximity explained homogeneity of catchments in streamflow characteristics in the case of the regression model while the regional calibration in terms of transferring of model parameters corresponding to maximum regional weighted average (MRWA) Nash-Sutcliffe efficiency (NSE) performance measure performed better in the case of P-R model.

Plots of hydrographs and FDC of observed regulated flow versus predicted 'unimpaired' or natural flow indicated significant hydrological alterations due to the regulation. The within a year FDC for observed regulated hydropeaking flow exhibits sharp bend transitions from high to medium flows and from medium to low flows. High flow (e.g. $Q > 22$ m^3/s) occurs only about 1 % of the time (< 100 hours duration), low flow (e.g. $Q < 5.0$ m^3/s) occurs for more than 68 % of the time (6000 hours) while the middle portion of the FDC which is 1% to 60 % of the time (100-5250 hours) is characterized by a nearly constant streamflow. The observed regulated hydrograph also shows continuous sudden fluctuations of streamflow while the predicted natural flow hydrographs and FDC exhibit smoothly varying patterns, which are typical characteristics of a natural flow. Alteration in FDCs and hydrographs would also indicate potential alteration in other streamflow characteristics, which are relevant for assessment of ecological integrity. The predicted natural time series is useful to derive any ecologically relevant streamflow metrics (ERSFM). Comparison of the indices derived from the predicted (natural) versus the actual flow under regulation would help to characterize flow related changes and devise improved mitigation and management in regulated rivers. The predicted natural flow provides useful information related to the concept of natural flow regime for environmental flow (e-flow). Moreover, the methods are also applicable for prediction at any ungauged sites.

1 INTRODUCTION

Study on environmental flow and flow regime [e.g. 1, 2, 6, 26, 32, 36, 37], alterations of natural flow regime due to regulation or hydropeaking [e.g. 11, 25] and impacts of climate change [e.g. 8, 35, 39] are indispensable to investigate the effects of hydrological alterations on ecological integrity. In addition, there are also interests to predict contributions from ungauged streams downstream of regulations to the environmental flow (e-flow). In addition, prediction of streamflow may be required at ungauged sites where ecological data is available in order to study the impacts of hydrological alteration on the riverine ecology. It is not possible to measure post-regulation natural flow for regulated rivers due to disturbance of the natural flow, and prediction in ungauged basins for water resources development planning and management is one of the challenging tasks in hydrology. Moreover, the effects of operation practices in hydropeaking rivers require prediction for hourly or finer temporal scale. Therefore, methodologies for prediction of time series of natural flow regime that excludes the effects of regulation (i.e. 'unimpaired' or natural flow) and in ungauged basins are required as a decision support for ecological friendly management of water resources.

Continuous streamflow prediction in ungauged basins through regional Precipitation-Runoff (P-R) modelling is one of a highly researched area [e.g. see review and comparison papers 9, 10, 24, 28] especially from 2003 to 2012 which was a Prediction in Ungauged Basins (PUB) decade for the International Association of Hydrological Sciences (IAHS) [33]. Regional transfer of information from gauged site to ungauged site for the PUB are performed for various objectives such as derivation of ecologically relevant streamflow metrics (ERSFM) for environmental flow assessment, to study the ecological and physical processes in riverine ecosystem and for water resources planning and management. Current scientific understanding of hydrologic controls on riverine ecosystems and experience gained from individual river studies support development of environmental flow standards at the regional scale [26]. Several regionalization attempts for prediction of flow characteristics (e.g. regime, hydrograph, seasonality, frequency, extremes such as flood and drought) at ungauged basins are available in literature on both hydrological and ecological sciences but the task remain challenging. A number of approaches for prediction of ERSFM at ungauged sites are available in literature. In a more recent time, the two main modelling approaches for prediction of ERSFM are statistical regression [5, 15, 31, 34] and the P-R models [e.g. 13, 20, 26].

However, these ecological applications are on only coarse temporal resolutions such as annual, monthly and daily streamflow while contemporary operation practices in regulated rivers such as hydropeaking require high-resolution prediction, which allows close examination of relevant ecological indicators from high-resolution hydrographs. The lists of suggested ERSFM and extraction software available in literature [e.g. 15, 18, 23, 25, 29] are based on daily or coarser time series which may not be representative for hydropeaking flow, which is variable at high resolution (e.g. hourly).

The main objectives and scope of this study are: (i) developing statistical regression models for prediction of streamflow 'signatures' such as flow duration curves (FDCs) and streamflow time series from relationship between streamflow and watershed characteristics; (ii) evaluating regionalization methods for regional transfer of streamflow information and Precipitation-Runoff model parameters; (iii) demonstration through comparative evaluation of the methods for prediction of flow duration curves and time series of natural streamflow by transferring of regional information to a hydropeaking river.

To our knowledge, thorough study on continuous prediction of 'unimpaired' or natural hourly streamflow for regulated rivers, from which ecologically relevant hourly ERSFM can be derived, are lacking. The concept of inflow controlled environmental flow regime can also be well evaluated if a continuous time series of natural flow can be predicted rather than assessing based on a pre-regulation natural flow regime. This work is supposedly an initial attempt of regional prediction of hourly streamflow for ecological applications in boreal Norway where regulation of rivers is common and hydropeaking operation is increasing while environmental legislations are stringent and aquatic ecosystems (e.g. salmonid fish) are abundant. However, the methods and procedures are applicable for other climate regimes too. Identification of ERSFM, which are relevant for hourly streamflow, was not an objective and scope of the present study.

2 THE STUDY REGION

The study region is boreal mid Norway (Figure 1) and the first portion of Table 1 presents the lists of catchments along their drainage areas. We used hourly streamflow data (2006-2011) from 26 stations in unregulated catchments (40 to 3090 km^2). Four of the catchments (3, 6, 8 and 14) are located inside the Gaula watershed. Hourly climate forcing include precipitation (P) from 44 stations, temperature (T) from 54 stations, wind speed (Ws) from 40 stations, relative humidity (HR) and global radiation (RG) from 12 stations, which are spatially interpolated on 1x1 km^2 grids for calibration of the gridded P-R model. The high flow regime for the study catchments are from snowmelt events in most cases, but some of the catchments exhibit precipitation on snowmelt or summer precipitation events. The dominant land use/land cover types in the study area are bare rock mountainous above timberline and forests. Predominant soil formation is glacial tills. We applied the proposed methods in the present study to predict the 'unimpaired' or natural streamflow series for the regulated Lundesokna catchment at tailrace outlet of the Sokna hydropower plant (total catchment area 243.4 km^2). Lundesokna river is a tributary of the Gaula river and flows from Samsjøen reservoir (487-473 masl, 9.8 km^2) to Gaula. Gaula one of the best salmon rivers in Norway. Sokna hydropower plant (commissioned in 1964) is a hydropeaking plant and has the following salient features: installed capacity of 30 MW, a gross head of 185 meters, intake regulation height of 9 meters and total catchment area at intake of 217 km^2. For a regulated observed time series, a time series of discharge data from the Sokna power plant is available while contribution of the local catchment between Sokna intake and tailrace outlet (area = 26.4 km^2) to the streamflow was predicted by the P-R model.

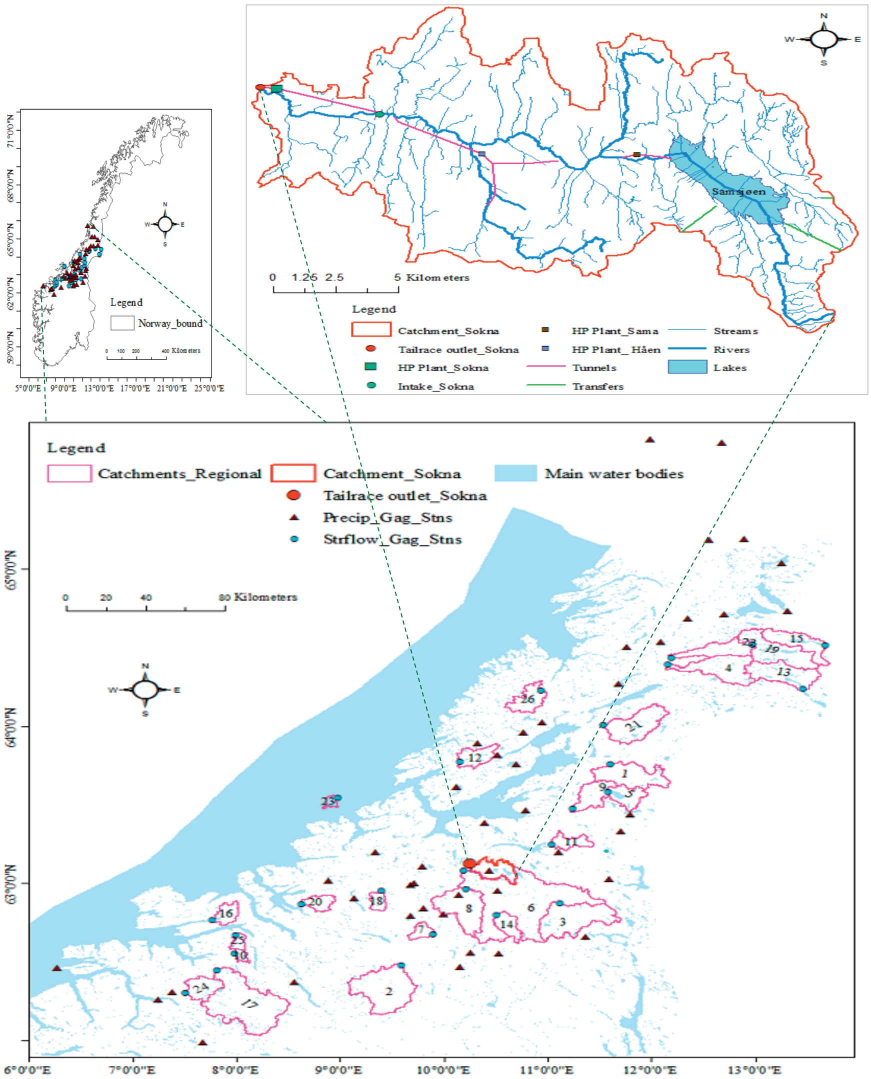


Figure 1. Maps indicating locations of regional catchments and the Sokna catchment-hydropower systems.

3 METHODS AND MODELS

There are several inherent uncertainties associated to both of regional regression and P-R modelling. The accuracy of streamflow characteristic predictions is important because of the potential consequences a poor prediction can have on estimates of ecological health [20]. Some of the main problems associated to P-R models are predictive uncertainty due to uncertainty in inputs, parameter calibration and model structure, and regionalization methods. Similarly, the data driven regression approach is associated with several assumptions such as randomness, normality and homoscedascity of residuals and non-collinearity of the independent variables. In addition, dependence between regression model parameters and subjectivity in selection and pre-processing of the independent variables are prevailing challenges. For instance, [15] conducted regional regression analyses based on 16 potential independent variables (watershed characteristics) to predict 19 presupposed ecologically relevant streamflow characteristics for Tennessee and Cumberland river basins (USA). The P-R models allow prediction of continuous time series of streamflow, from which one can derive any ecological indices of interest, while the regression approach focuses on deriving separate relationships among various dependent variables (i.e. each ERSFM) and selected catchment attributes. However, selection of a small number of independent variables would reduce the number of regression parameters and hence the uncertainties related to dependence among parameters and collinearity among independent variables. Therefore, the modelling approach in the present study geared towards parametrical parsimony, simplicity and consistency.

3.1 Statistical (regression) model

Some applications of regression or other statistical models for direct prediction of ERSFM from watershed characteristics include [7, 15, 19, 20, 22]. [7] compared regional regression based on 24 potential catchment characteristics as independent variables versus the HBV model for prediction of low flow index for daily streamflow from 51 catchments in Southern Norway. [19, 22] demonstrated regressing different catchment attributes versus discharge relationships for prediction of low flow hydrograph in ungauged basins in Australia.

Identification of independent variables and choice of dependent variables are important for the regression model. In the study on global environmental assessment methodologies, [36] stated that flow duration curves and other single flow indices comprise the second largest subgroup of hydrological approach for environmental flow. [13] used ratio of 25 % to 75 % exceedance flows. [39] used frequency of high flows

(Q_{95} and Q_{99}) during winter and summer, mean annual and mean summer flows, and frequency of summer low flows (Q_{10} and Q_{20}). [40] used 10 % and 20 %, and 30% respectively of average daily flow for baseflow in dry and wet seasons and the 25th percentile flow as a minimum high pulse discharge. [1] defined flow regimes based on flow percentiles (low < 25%, high > 75% and normal 25% to 75%) to represent dry, wet and normal years to develop an inflow controlled environmental flow regime. The natural flow regime paradigm [25, 30] focuses on a full range of intra-and inter annual variability of streamflow characteristics pertinent to magnitude, frequency, duration, timing and rate of change to study comparative relationships between natural and altered hydrological conditions and riverine ecology. Therefore, due to various utility of streamflow characteristics the focus of the present study was to evaluate methods to derive the two main runoff 'signatures' namely flow duration curves and time series of streamflow or hydrographs for prediction in regulated rivers and ungauged basins from which further ERSFM can be extracted, rather than identifying and extracting each ERSFM.

Flow duration curves

We fitted separate linear regressions between each streamflow percentiles of 0 % to 100 % at 1% intervals (response or dependent variables) with the independent variable (drainage areas of catchments). The flow percentiles rather than the various ERSFM exhibit similar relationships to watershed characteristics and make the regression more consistent. A simple linear regression model with assumptions of normal and homoscedastic residuals is:

$$\underline{Y} = \underline{X}\underline{\beta}^* + \underline{\varepsilon} \text{ or } Y_i = \beta_0^* + \beta_1^* x_{1i} + \varepsilon \quad (1)$$

$$\underline{\varepsilon} \approx N\left(\underline{0}, \underline{I}\sigma^2\right) \text{ and } \underline{Y} \approx N\left(X\underline{\beta}^*, \underline{I}\sigma^2\right) \quad (2)$$

We estimated the set of parameters by minimizing the standard error of estimates or the ordinary least-square technique and their lower and upper confidence levels (UCL and LCL) from the *t*-statistics:

$$\hat{\underline{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(\underline{X}'\underline{X}\right)^{-1} \underline{X}'\underline{Y}\ ;\ \left[\text{UCL, LCL}\right] = \hat{\beta}_i \pm t_{\alpha/2,n-p}\sqrt{S^2\left(\underline{X}'\underline{X}\right)_{ii}^{-1}}, \quad (3)$$

where $\underline{Y}$ is nx1 column vector of response (dependent) variable, $\underline{X}$ is nxp matrix of independent variable, $\underline{\varepsilon}$ is nx1 column vector of the error term that indicates the deviation of the estimate from the true value, $\underline{I}$ is a nxn identity matrix, σ^2 is a variance, * represents the 'true' values, the underline represents the vector or matrix notation, n is the number of observations (data points), p is the number of model parameters, N represents the assumed Normal distribution, $(X'X)ii^{-1}$ is the main diagonal element corresponding to i^{th} row and i^{th} column of a $(X'X)ii^{-1}$ matrix of size pxp, ′ represents a transpose and α is the significance level.

We performed diagnostic analyses of residuals to verify the adequacy of the model. We performed the significance tests for model parameters by *F*-test. We computed the percentage of variability in the data explained by the regression model (R^2). We tested randomness of the residuals and outliers by plots of residuals versus the dependent variable. We also verified the normality of the residuals by probability plots and presented the model prediction error in terms of 95 % confidence intervals (CI). Once, we develop regional regression equations between streamflow percentiles and catchment areas, we can estimate streamflow percentiles and hence flow duration curves for ungauged or regulated basins in the region. Flow duration or the percentage of time flow equaled or exceeded F (%) = 100-percentiles.

Time series of streamflow (hydrographs)

Streamflow percentiles or flow duration alone cannot provide sufficient information for ecological studies and hence prediction of complete time series of streamflow hydrographs for regulated or ungauged catchments is required. In the present study, a simple method was proposed to derive streamflow time series for ungauged catchments from flow percentiles of both gauged and ungauged catchments (which are derived from the regional regression) and observed streamflow data for the gauged catchments. The main premise or assumption in the method is that for catchments that are homogeneous in terms of their streamflow percentiles, the streamflow time series at a similar time *t* exhibit the same percentile for the homogenous donor (gauged) and the recipient (ungauged) catchments. We used a simple lookup function in Microsoft excel:

$$Q_t^{ungauged} = lookup\left(Per_t^{Qgauged}, Per^{0:1:100}, Q_{Per}^{reg:ungauged}\right), \qquad (4)$$

where Q_t is time series of streamflow, Per_t are percentiles for time series of streamflow, $Per^{0:1:100}$ are percentiles from 0 % to 100 % at 1% intervals and Q_{Per} is streamflow corresponding to percentile Per.

3.2 Precipitation-Runoff (P-R) model

3.2.1 *Runoff response routine*

Based on plausible assumptions, [14] proposed a 'top-down' modelling approach with possibility of inferring model structure, equations and parameters from observed streamflow during recession. The parsimonious water balance based P-R response routine is as below:

$$\frac{dS}{dt} = I - AET - Q = \left(I - AET - Q\right) \tag{5}$$

$$\frac{dQ}{dt} = \frac{dQ}{dS}\frac{dS}{dt} = \frac{dQ}{dS}\left(I - AET - Q\right) = g\left(Q\right)\left(I - AET - Q\right) \approx g\left(Q\right)\left(-Q\right)\Big|_{I<<Q,\, AET<<Q}, \tag{6}$$

where actual evapotranspiration (AET), infiltration (I) = rainfall + snow melt (SM) and discharge (Q) are given in mm/hr, ground water storage (S) is in mm depth of water, t is a time variable. The g(Q) is discharge sensitivity function [14], which is the sensitivity of discharge to changes in storage. The response routine used in the present study is based on 1x1 km^2 grid cells or the hillslope scale rather than lumped for the whole catchment.

$$\ln\left(g\left(Q\right)\right) \approx \ln\left(\frac{dQ}{dS}\right) \approx \ln\left(\frac{-dQ/dt}{Q}\Big|_{P<<Q,\, AET<<Q}\right) \approx \alpha_0 + \alpha_1 \ln\left(Q\right), \tag{7}$$

where α_0 and α_1 are model parameters set by calibration. Runoff computation follows integrating in time the storage-discharge relationship:

$$S\left(Q\right) = \int dS = \int \frac{1}{g\left(Q\right)} dQ \tag{8}$$

3.2.2 *Other routines*

We computed the potential evapotranspiration (PET) based on Priestly Taylor method [27] and actual evapotranspiration from the PET, discharge and evapotranspiration ratio parameter (EvR). We simulated the snow accumulation and snowmelt outflow (SNOWOUT) based on gamma distributed snow depletion curve or SDC [17]. We implemented a simple travel time zone isochrones routing (pure translation) to translate the hillslope runoff response of each 1x1 km^2 grid cell to the catchment outlet based on travel time lags.

3.2.3 *Calibration of the P-R model*

We performed regional calibration of the P-R model based on streamflow data from the 26 gauged catchments. We used the Differential Evolution Adaptive Metropolis algorithm or DREAM [38] with residual based log-likelihood (L-L) objective function implemented in ENKI hydrological modelling platform [16]:

$$L-L\left(\delta/\sigma_i^2,\sum_{i=1}^{N_C}\sum_{t=1}^{n_i}\left(Qsim_{t,i}^{(\theta)}-Qobs_{t,i}^{(\theta)}\right)^2\right)=\left\{\sum_{i=1}^{N_C}\left(\frac{-n_i}{2}\log(2\pi)-\frac{n_i}{2}\log\left(\sigma_i^2\right)-\frac{\sum_{t=1}^{n_i}\left(Qsim_{t,i}^{(\theta)}-Qobs_{t,i}^{(\theta)}\right)^2}{2\sigma_i^2}\right)\right\}\times f\,, \quad (9)$$

where $Qsim^{(\theta)}$ and $Qobs^{(\theta)}$ respectively are Box-Cox [3] transformed observed and simulated streamflow time series of length n, N_C is the total numbers of the catchments, δ denotes model parameter, θ is the Box-Cox transformation parameter, *f* is fraction of effectively independent observations and σ_ε^2 is variance of error. Nash-Sutcliffe efficiency, NSE [21] performance measure was used for evaluation of both the P-R and regression models. The six free parameters are threshold temperature (TX), wind speed sensitivity of snow (WS), the EvR, velocity of flow for runoff routing (V), and α_1 and α_0, which are response routine parameters. For the P-R model, we evaluated the regionalization method or parameter transferability based on the spatial proximity, arithmetic and weighted averaged parameters, and parameters corresponding to maximum regional arithmetic and maximum regional weighted average (MRWA) performance measure (NSE). We assigned the weights based on the length of the non-missing records.

4 RESULTS

Results of parameter estimation for the regional regression model and regional calibration of the P-R model (i.e. parameters corresponding to maximum regional weighted average or MRWA NSE) are given in the second portion of Table 1. We presented the regional regression results at 5 % percentiles intervals. The diagnostics of residuals identified the Trangen catchment as an outlier and hence excluded from the regression analysis. Table 2 shows the NSE results for 11 catchments, which has no missing streamflow data for P-R model calibration period. The bold values along the diagonals indicate the performances of regression model and local calibration of the P-R model for a particular catchment. Figure 2 shows predictions for Eggafoss watershed by transferring regional regression information from the Gaulfoss watershed, regional transfer of P-R model parameters corresponding to MRWA NSE and results for the local calibration. In this case, we evaluated the transferability of regional information for predictions in internal subcatchments based on spatial proximity. In Figure 3, we presented the transferability of regional regression information to a neighboring catchment based on spatial proximity and regional transfer of P-R model parameters corresponding to MRWA NSE. Figure 4 presents observed hydrographs and FDC under hydropeaking versus the predicted 'unimpaired' or natural streamflow and FDCs for Lundesokna river. Figure 5 and Figure 6 respectively provide typical prediction by the regional regression versus the observed regulated hydrographs for summer and fall, and winter seasons.

5 SUMMARY AND CONCLUSIONS

The study geared towards evaluation of parametrically parsimonious, simple and more consistent approaches for prediction of hourly streamflow for regulated or ungauged basins, which is useful for deriving streamflow characteristics of ecological relevance. We proposed a simple (two parameters) linear regression model with catchment area as independent variable to predict streamflow percentiles and hence duration curves. For the regression model, we predicted for the ungauged sites based on regional transfer of predicted streamflow percentiles and observed streamflow from the nearest gauged site. For the P-R model, we conducted regional calibration of a parsimonious 'top-down' model and regional transfer of model parameters corresponding to the MRWA NSE performance measure performed slightly better than the other methods. Table 2 provides the results from comparative evaluation of the models through spatial transfer of information.

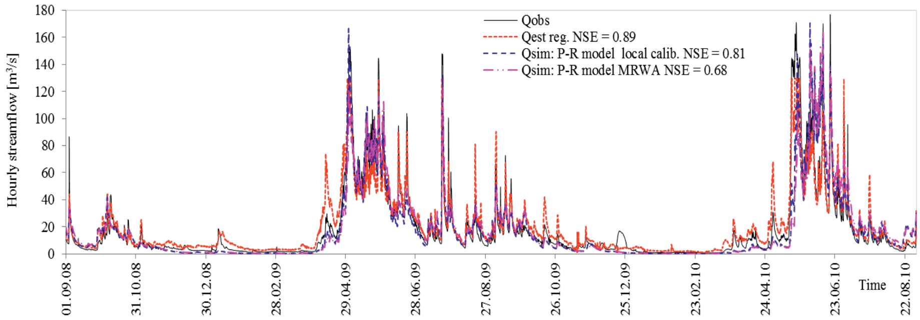


Figure 2. Transfer of regional information to Eggafoss from Gaulfoss to its internal subcatchment (regression) and based on regional P-R model calibration (2008-2010).

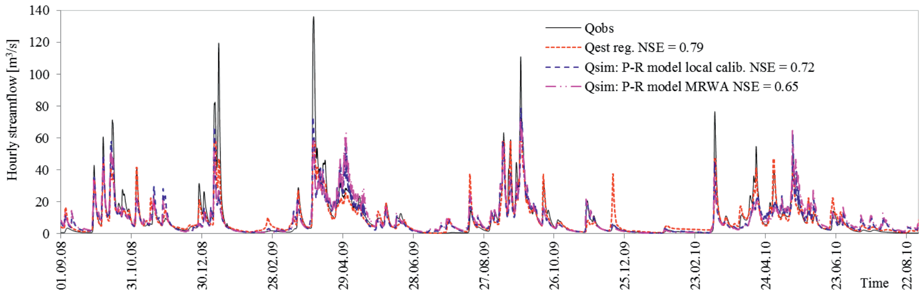


Figure 3. Transfer of regional information to Øyungen from a nearby catchment Krinsvatn (regression) and based on P-R model calibration (2008-2010).

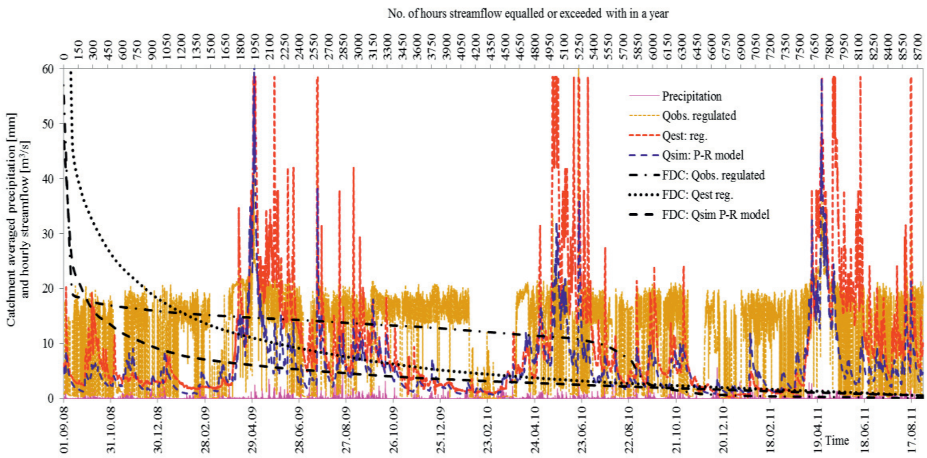


Figure 4. Transfer of regional information from Gaulfoss catchment to a nearby regulated Lundesokna river (regression) and based on parameter corresponding to MRWA NSE (P-R model calibration).

Table 1. The study catchments, regional regression and calibrated MRWA NSE parameters for the P-R model. Regression (2006-2011) and P-R model calibration (2008-2010). The bold fonts are model parameters.

Catchments and streamflow stations				Parameters for regional regression and P-R model							
Catch. No.	Catchment name	NVE Station No.	Drainage area, km^2	Flow percentiles, P	β_0LCL	$\boldsymbol{\beta_0}$	β_0UCL	β_1LCL	$\boldsymbol{\beta_1}$	β_1UCL	R^2
1	Dillfoss	127.13	480	0	-0.24	-0.02	0.20	0.001	0.002	0.002	0.84
2	Driva v/Risefoss	109.9	745	5	0.10	0.47	0.84	0.002	0.002	0.003	0.80
3	Eggafoss	122.11	668	10	0.28	0.75	1.22	0.002	0.003	0.004	0.80
4	Embrethølen	139.26	495	15	0.32	0.86	1.41	0.003	0.004	0.004	0.83
5	Feren	124.13	220	20	0.44	1.06	1.67	0.004	0.005	0.005	0.86
6	Gaulfoss	122.9	3090	25	0.56	1.23	1.90	0.005	0.006	0.006	0.88
7	Gisnås	121.29	95	30	0.68	1.39	2.09	0.006	0.007	0.008	0.90
8	Hugdal bru	122.17	546	35	0.82	1.58	2.35	0.007	0.008	0.009	0.91
9	Høggås bru	124.2	495	40	0.93	1.75	2.58	0.008	0.009	0.010	0.93
10	Isa v/Morstøl bru	103.2	44	45	1.10	2.02	2.95	0.009	0.011	0.012	0.93
11	Kjeldstad i Garb.	123.31	145	50	1.25	2.29	3.34	0.011	0.013	0.014	0.94
12	Krinsvatn	133.7	207	55	1.39	2.56	3.74	0.013	0.015	0.017	0.94
13	Lenglingen	308.1	450	60	1.71	3.14	4.57	0.015	0.017	0.019	0.94
14	Lillebudal bru	122.14	168	65	2.03	3.79	5.55	0.018	0.020	0.023	0.93
15	Murusjø	307.5	346	70	2.30	4.54	6.79	0.021	0.024	0.027	0.92
16	Osenelv v/Øren	105.1	138	75	2.44	5.12	7.79	0.026	0.030	0.033	0.93
17	Rauma v/Horgheim	103.4	1100	80	2.10	5.28	8.46	0.035	0.039	0.043	0.94
18	Rinna	112.8	91	85	2.29	5.90	9.52	0.046	0.051	0.056	0.95
19	Skjellbreivatn	139.25	546	90	3.98	8.24	12.49	0.059	0.065	0.071	0.96
20	Søya v/Melhus	111.9	138	95	5.27	11.6	17.92	0.085	0.094	0.102	0.96
21	Støafoss	128.5	477	100	19.48	99.0	178.5	0.321	0.428	0.534	0.75
22	Trangen	139.35	852		**TX, °C**	**WS**	**EvR**	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_0}$	**V, m/s**	
23	Valen	117.4	39		-0.73	3.82	0.85	0.71	-3.52	2.93	
24	Valldøla v/Alstad	100.1	226								
25	Vistdal	104.23	67								
26	Øyungen	138.1	239								

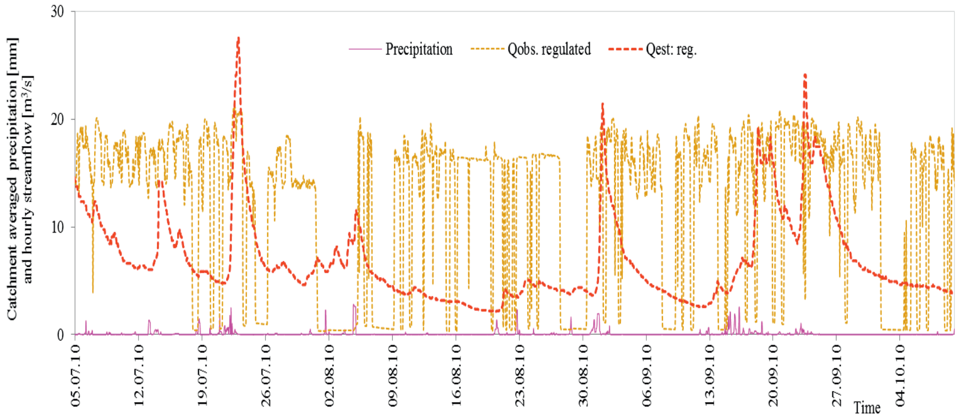


Figure 5. Typical comparisons of summer and fall observed regulated streamflow versus the predicted ‘unimpaired’ or natural streamflow for Lundesokna river (transferred from Gaulfoss through regional regression).

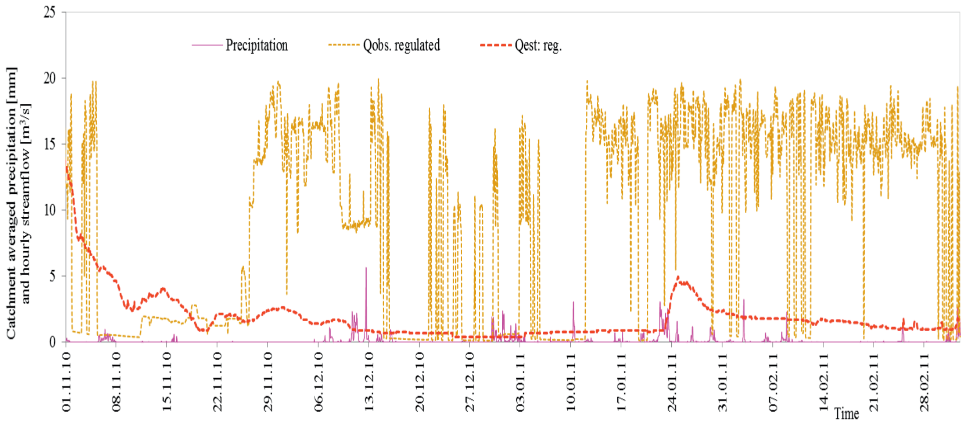


Figure 6. Typical comparisons of winter observed regulated streamflow versus the predicted ‘unimpaired’ or natural streamflow for Lundesokna river (transferred from Gaulfoss through regional regression).

FDC: Qobs (2008-2011) and FDC: Qpred (2006-2011).

Table 2. NSE for regional transfer of information based on regional regression (2006-2011) and P-R model calibration (2008-2010) for prediction of hourly streamflow.

Donor catchments (Catchment No.)	Recipient catchments (Catchment No.)										
Regression(reg.)	1	3	6	10	12	14	16	17	20	21	26
1	**0.95**	0.31	0.39	-1.10	0.50	-0.07	-1.21	-0.44	0.42	0.70	0.49
3	0.35	**0.96**	0.88	-0.40	-0.13	0.55	-2.29	0.48	0.33	0.02	-0.22
6	0.42	0.89	**0.95**	-0.39	-0.07	0.54	-2.02	0.42	0.38	0.12	-0.16
10	0.04	0.39	0.42	**0.43**	-0.21	0.23	-2.49	0.61	0.44	-0.21	-0.36
12	0.53	-0.35	-0.26	-1.96	**0.80**	-0.77	-0.68	-1.08	0.15	0.67	0.80
14	0.21	0.80	0.79	-0.30	-0.19	**0.70**	-2.26	0.47	0.34	-0.09	-0.30
16	-0.06	-0.69	-0.54	-2.29	0.26	-0.98	**0.45**	-1.22	-0.10	0.13	0.25
17	-0.17	0.49	0.46	-0.06	-0.37	0.22	-2.72	**0.97**	0.12	-0.39	-0.51
20	0.50	0.34	0.42	-0.44	0.22	0.11	-1.48	-0.14	**0.83**	0.24	0.10
21	0.72	-0.07	0.04	-1.68	0.61	-0.50	-0.98	-0.83	0.20	**0.90**	0.66
26	0.48	-0.37	-0.30	-2.07	0.73	-0.82	-0.71	-1.12	0.06	0.67	**0.88**
P-R model											
1	**0.74**	0.79	0.82	0.48	0.63	0.53	0.03	0.44	0.56	0.71	0.66
3	0.73	**0.81**	0.83	0.51	0.71	0.53	0.14	0.23	0.53	0.72	0.71
6	0.70	0.79	**0.83**	0.54	0.49	0.56	-0.59	0.15	0.65	0.67	0.54
10	-0.14	0.10	0.18	**0.58**	0.19	0.23	-0.23	0.16	0.31	0.24	0.06
12	0.68	0.78	0.70	0.50	**0.75**	0.52	0.12	0.41	0.51	0.70	0.71
14	0.62	0.66	0.74	0.47	0.27	**0.58**	-0.34	0.18	0.62	0.57	0.47
16	0.44	0.48	0.50	0.40	0.50	0.38	**0.67**	0.66	0.28	0.46	0.42
17	0.29	0.31	0.44	0.35	0.29	0.28	0.56	**0.77**	0.20	0.30	0.21
20	0.58	0.69	0.75	0.56	0.23	0.56	-1.16	-0.11	**0.67**	0.55	0.32
21	0.69	0.73	0.74	0.56	0.69	0.54	0.17	0.35	0.53	**0.71**	0.63
26	0.64	0.71	0.61	0.45	0.71	0.48	0.30	0.56	0.44	0.65	**0.72**
MRWA NSE	0.65	0.68	0.72	0.42	0.69	0.47	0.55	0.68	0.43	0.63	0.64

Comparisons of the NSE along the diagonals (i.e. bold values) in Table 2 for the regression (reg.) versus the P-R model indicate that for most of the catchments the regional regression model outperformed the local calibration of the P-R model. This shows that the prediction errors for those particular catchments of the regression model are less than the errors in prediction by the P-R model. The 'proxy ungauged basin' parameter transfer indicated the regression model + spatial proximity resulted in regional performances better than the regional calibration (MRWA) NSE of the P-R model parameters. The regression based prediction transfers characteristics of the observed streamflow from gauged to ungauged catchments while the P-R model transfers the model parameters affected by rigorous calibration procedures and various sources of uncertainties. Regression based prediction by transferring from Gaulfoss to an internal subcatchment of Eggafoss (Figure 2) indicated NSE value of 0.89 versus the local P-R calibration (NSE = 0.81) and the MRWA (NSE = 0.68). In addition, prediction for Øyungen based on transfer of information on streamflow from the neighboring Krinsvatn catchment (Figure 3) indicated NSE value of 0.79 versus the local P-R calibration (NSE = 0.72) and the MRWA (NSE = 0.65). The NSE values in Table 2 further explained the spatial proximity based transferability of regional regression based information among pairs of catchments such as between catchments 1 and 21, 3 and 6, 10 and 17, 12 and 26, 14 and 3, 14 and 6, 21 and 1, and 21 and 26. However, some catchments such as catchments 10 and 16 exhibit poor NSE and hence poor spatial transferability due to large prediction error associated to them, but these catchments are not outliers. However, if a large number of catchments are available for the analyses certain additional criteria can be set to exclude less performing catchments.

Results of application of the proposed tools for prediction (2008-2011) in the regulated Lundesokna river at downstream of the tailrace discharge of Sokna hydropower plant are given in Figure 4. The observed streamflow (Qobs. regulated in Figure 4) for Lundesokna river at downstream of the tailrace is highly influenced by the hydropeaking plant discharge. We used the developed tools to predict or simulate the FDC and time series of hourly 'unimpaired' streamflow and then transferred the regional regression information from the Gaulfoss catchment to the Lundesokna river based on spatial proximity or nearest neighbor. The predicted hydrographs show that the catchment runoff generation well responds to the catchment-averaged precipitation events. Even though there are similarities between the general patterns of hydrographs predicted from the regression-based approach (Qest.) and the P-R model (Qsim) and their corresponding FDCs, the results show significant differences in predicted streamflow magnitudes for specific time and duration. However, as demonstrated by transfer of information among the catchments, the regression model + spatial proximity outperformed the regional calibration (MRWA) NSE of the P-R model. [7, 20] respectively also found that regional regression model outperformed the P-R models for Southern Norway and Kentucky (USA).

Figure 4, Figure 5 and Figure 6 show significant hydrological alterations due to regulation and hydropeaking for Lundesokna. The within a year FDC for observed regulated hydropeaking flow exhibits sharp bend transitions from high to medium flows and from medium to low flows. High flow (e.g. $Q > 22$ m^3/s) occurs only about 1 % of the time (< 100 hours duration), low flow (e.g. $Q < 5.0$ m^3/s) occurs for more than 68 %

of the time (6000 hours) while the middle portion of the FDC which is 1% to 60 % of the time (100-5250 hours) is characterized by a nearly constant streamflow. The observed regulated hydrograph also shows continuous sudden fluctuations of streamflow magnitudes while the predicted streamflow hydrographs and FDCs exhibit smoothly varying patterns, which are typical characteristics of an ‘unimpaired’ or natural flow. Alteration in the FDCs and hydrographs would also indicate alterations in several streamflow characteristics, which affect the ecological integrity in regulated rivers. Figure 5 and Figure 6 for summer and fall, and winter seasons respectively clearly indicate typical differences between the regulated or hydropeaking flow and the predicted natural flow. The ecological impacts of such hydrological alterations require further study, which was not an objective and scope of the present study.

In conclusion, the study illustrated that a simple data based regression model from relationships among streamflow percentiles and catchment drainage areas for prediction of streamflow percentiles and FDCs, and regional transfer of observed streamflow time series outperformed the regional calibration of the P-R model for prediction of hourly streamflow at regulated or ungauged site in the boreal study region. The simple approaches for derivation of FDCs and hydrographs and hence for estimation of ERSFM for studies related to ecological impacts of river regulation provide significant contribution for operational environment and research purposes. It would relieve people working with management issues from relying on scarce or short data series and it would contribute to endeavors for the PUB, which is one of the important but challenging tasks in hydrology. The followed methodologies are also applicable in other climate regimes. We expect improved results for a large set of catchments.

ACKNOWLEDGMENTS

The Center for Environmental Design of Renewable Energy or CEDREN (http://www.cedren.no) provided the funding for the research. We also would like to acknowledge the Norwegian Water Resources and Energy Directorate (NVE), the Norwegian Meteorological Institute, Statkraft, TrønderEnergi, bioforsk, and Nord Trøndelag energi for the hydro-meteorological data used in the present study.

REFERENCES

[1] Alfredsen K., Harby A., Linnansaari T. and Ugedal O., Development of an inflow controlled environmental flow regime for a Norwegian river, *River Res. Applic*., Vol. 28, No. 6, (2012), pp 731-739.

[2] Arthington A.H, Bunn S.E., Poff N.L., and Naiman R.J., The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Applications*, Vol. 16, No. 4, (2006), pp 1311-1318.

[3] Box G. E. P. and Cox, D. R., An analysis of transformations, *Journal of the Royal Statistical Society*, Series B Vol. 26, (1964), pp 211-252.

[4] Carlilse D.M., Falcone J., Wolock D.M., Meador M.R., and Norris R.H., Predicting the natural flow regime: models for assessing hydrological alteration in streams, *River Res. Applic*., Vol. 26, No. 2, (2009) pp 118-136.

[5] Carlisle D.M., Wolock D.M. and Meador M.R., Alteration of streamflow magnitudes and potential ecological consequences: a multiregional assessment, *Frontiers in Ecology and the Environment*, Vol. 9, (2011), pp 264-270.

[6] Costa R., Martínez-Capel F., Muñoz-Mas R. Alcaraz-Hernández J. and Garófano-Gómez V., Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar Nase, Parachondrostoma arrigonis (River Cabriel, Spain), *River Res. Applic*.,Vol. 28, No. 6, (2012), pp 740–752.

[7] Engeland K. and Hisdal H., A Comparison of Low Flow Estimates in Ungauged Catchments Using Regional Regression and the HBV-Model, *Water Resour Manage*, Vol. 23, (2009), pp 2567–2586.

[8] Gibson C.A., Meyer J.L, Poff N.L, Hay L.E., Georgakakos A., Flow regime alterations under changing climate in two river basins: implications for freshwater ecosystems, *River Res. Applic*., Vol. 21, No. 8, (2005), pp 849-864

[9] He Y., Bárdossy A. and Zehe E., A review of regionalisation for continuous streamflow simulation, *Hydrol. Earth Syst. Sci*., Vol. 15, No. 11, (2011), pp 3539-3553.

[10] Hrachowitz M., Savenije H.H.G., Blöschl G., McDonnell J.J., Sivapalan M., Pomeroy J.W., Arheimer B., Blume T., Clark M.P., Ehret U., Fenicia F., Freer J.E., Gelfan A., Gupta H.V., Hughes D.A., Hut R.W., Montanari A., Pande S., Tetzlaff D., Troch P.A., Uhlenbrook S., Wagener T., Winsemius H.C., Woods R.A., Zehe E., Cudennec C., A decade of Predictions in Ungauged Basins (PUB)-a review, *Hydrolog. Sci. J*., Vol. 58, No.6, (2013), pp 1198-1255.

[11] Jones N.E., The dual nature of hydropeaking rivers: Is ecopeaking possible?. *River Res. Applic*., (2013) Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/rra.2653.

[12] Kennard M.J., Pusey B.J., Olden J.D., Mackay S.J., Stein J.L. and Marsh N., Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*, Vol. 55, No.1, (2010), pp 171-193.

[13] Kennen J.G., Kauffman L.J., Ayers M.A. and Wolock D.M., Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites, *Ecological Modelling*, Vol. 211, (2008) pp 57-76.

[14] Kirchner J.W., Catchments as simple dynamical systems: catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res*., 45, No. 2, (2009), DOI: 10.1029/2008WR006912.

[15] Knight R.R., Gain W.S., and Wolfe W.J., Modelling ecological flow regime: An example from the Tennessee and Cumberland River Basins, *Ecohydrology*, Vol. 5, No. 5, (2011), pp 613-627.

[16] Kolberg S. A. and Bruland, O., ENKI - An Open Source environmental modelling platform. *Geophysical Research Abstracts* 14, (2012), EGU2012-13630, EGU General Assembly.

[17] Kolberg S. A. and Gottschalk L., Updating of snow depletion curve with remote sensing data, *Hydrol. Process*., Vol. 20, No. 11, (2006), pp 2363–2380.

[18] Mathews R. and Richter B., Application of the Indicators of Hydrologic Alteration software in environmental flow-setting, *Journal of the American Water Resources Association*, Vol. 43, No. 6, (2007) 1400-1413.

[19] Moliere D., Lowry J., Staben G. and Humphrey C. (2006), Flow characteristics of streams in the tropical rivers region, *In Proceedings of the 30th Hydrology and Water Resources Symposium*: Past, present and future, Sandy Bay, Tas., (2006), pp 43-48.

[20] Murphy J.C., Knight R.R., Wolfe W.J., and Gain W.S., Predicting ecological flow regime at ungaged sites: A comparison of methods, *River Res. Applic.*, Vol. 29, No. 5, (2012), pp 660-669.

[21] Nash J. E., Sutcliffe J. V., River flow forecasting through conceptual models, I. A discussion of principles, *J. Hydrol*., Vol. (10), (1970), pp 228–290.

[22] Nathan R.J. and McMahon T.A., Identification of homogeneous regions for the

purposes of regionalization, *J. Hydrol*., Vol. 121, No. 1-4, (1990), pp 217-238.

[23] Olden J.D. and Poff N.L., Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, *River Res. Applic*.,19, No. 2, (2003), pp 101-121.

[24] Parajka J., Viglione A. Rogger M., Salinas J. L., Sivapalan M. and Blöschl G., Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci*., Vol. 17, (2013), pp 1783-1795.

[25] Poff N.L., Allan J.D., Bain M.B., Karr J.R., Prestegaard K.L., Richter B.D., Sparks R.E. and Stromberg J.C., The natural flow regime, *BioScience*, Vol. 47, No. 11, (1997), pp 769-784.

[26] Poff N.L., Richter B.D., Arthington A.H, Bun S.E, Naiman R.J., Kendy E., Acreman M., Apse C., Bledsoe B.P., Freeman M.C., Henriksen J., Jacobson R.B., Kennen J.G., Merritt D.M., O'keeffe J.H., Olden J.D., Rogers K.,Tharme R.E. and Warner A., The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards, *Freshwater Biology*, Vol. 55, No. 1, (2010), pp 147-170.

[27] Priestley C.H.B. and Taylor R.J., On the assessment of surface heat flux and evaporation using large-scale parameters, *Mon. Weather Rev*. Vol. 100, (1972), pp 81-82.

[28] Razavi T. and Coulibaly P., Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods, *J. Hydrol. Eng*., Vol. 18, (2013), pp 958-975.

[29] Richter B.D., Baumgartner J.V., Powell, J., and Braun, D.P. 1996, "A method for assessing hydrologic alteration within ecosystems", *Conservation Biology*, Vol. 10, No. 4, pp 1163-1174.

[30] Richter B.D., Baumgartner J.V., Wigington R. and Braun D.P., How much water does a river need? *Freshwater Biology*, Vol. 37, (1997), pp 231-249.

[31] Sanborn S.C. and Bledsoe B.P., Predicting streamflow regime metrics for ungaged streams in Colarado, Washington, and Oregon, *Journal of Hydrology*, Vol. 325, (2006), pp 241-261.

[32] Schofield N. and Burt A., Issues in environmental water allocation-an Australian perspective, *Water Science and Technology*, Vol. 48, No. 7, (2003), pp 83-88.

[33] Sivapalan M., Takeuchi K, Franks S.W., Gupta V.K., Karambiri H., Lakshmi V., Liang X., McDonnel J.J., Mendiondo E.M., O'connell P.E., Oki T., Pomeroy J.W.,

Schertzer D., Uhlenbrook S. and Zehe E., IAHS Decade on Predictions in Ungauged Basins, PUB 2003-2012: shaping an exciting future for the hydrological sciences, *Hydrolog. Sci. J.*, Vol. 48, No. 6, (2003), pp 857-880.

[34] Sickle V. J., Stoddard J. L., Paulsen S. P. and Olsen A. R., Using Relative Risk to Compare the Effects of Aquatic Stressors at a Regional Scale, *Environmental Management*, Vol. 38, No.6, (2006), pp 1020-1030.

[35] Shrestha R.R., Peters D.L. and Schnorbus M. A., Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators, *Hydrol. Process*, (2013) Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/hyp.9997.

[36] Tharme R.E., A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers, *River Res. Applic*., 19, No. 5-6, (2003), 397-441.

[37] Vezza P., P., Rosso P. M. and Comoglio C., Defining minimum environmental flows at regional scale: Application of mesoscale habitat models and catchments, *River Res. Applic*., Vol. 28, No. 6, (2012), pp 717-730.

[38] Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M. and Higdon D., Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Journal of Nonlinear Sciences and Numerical Simulation*, Vol. 10, No. 3, (2009), pp 273-290.

[39] Wenger S.J., Luce C.H., Hamlet A.F., Isaak D.J. and Neville H.M., Macroscale hydrologic modeling of ecologically relevant flow metrics, *Water Res. Research*, Vol. 46, No. 9, (2010), doi:10.1029/2009WR008839.

[40] Yin X.A., Yang Z.F. and Petts G.E., Optimizing environmental flows below dams, *River Res. Applic*., Vol. 28, No. 6, (2012), pp 703-716.

Appendix B: Co-authors and publishers' declarations

Appendix B

NTNU

**Encl. to app-
lication for
assessment of
PhD thesis**

STATEMENT FROM CO-AUTHOR

(cf. section 10.1 in the PhD regulations)

Teklu Tesfaye Hailegeorgis
....applies to have the following thesis assessed:
Name of candidate

Identification of spatially distributed Precipitation-Runoff response routines for hourly simulation in gauged and ungauged basins

....
title

*) The statement is to describe the work process and the sharing of work and approve that the article may be used in the thesis.

*)
Statement from co-author on article:
I, Sjur Kolberg, hereby declare that I am aware that the work in the three articles entitled:
1. Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed,
2. Distributed hourly runoff computations in mountainous boreal catchments from 'catchments as simple dynamical systems' storage-discharge relationships and
3. Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments, of which I am a co-author, will form part of the PhD thesis by the candidate who made major contributions in the design and running of experiments, and writing of the papers.

Trondheim 23/9 2014
....
Place, date

Sjur Kolberg
....
Signature co-author

*)
Statement from co-author on article:
I, Yisak S. Abdella, hereby declare that I am aware that the work in the three articles entitled:
1. Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed,
2. Distributed hourly runoff computations in mountainous boreal catchments from 'catchments as simple dynamical systems' storage-discharge relationships and
3. Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments, of which I am a co-author, will form part of the PhD thesis by the candidate who made major contributions in the design and running of experiments, and writing of the papers.

TRONDHEIM 23.09.2014
Place, date

Yisak Abdella
Signature co-author

IWA PUBLISHING
the international
water association

Alliance House
12 Caxton Street
London SW1H 0QS
United Kingdom
Tel: +44 (0)20 7654 5500
Fax: +44 (0)20 7654 5555
Email:
publications@iwap.co.uk
www.iwapublishing.com

Teklu T. Hailegeorgis,
Department of Hydraulic and Environmental Engineering
NTNU
NO-7491
Trondheim
Norway

23 September 2014

Dear Teklu

Permissions request relating to material to be published in *Hydrology Research*

In response to your request for copyright clearance to use the following paper, due for publication in *Hydrology Research*, in your PhD thesis:

> Teklu T. Hailegeorgis and Knut Alfredsen 2014 Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrology Research (*in press) doi 10.2166/nh.2014.051

We are very happy to grant you permission to reproduce the material specified above at no charge, provided that:

- the material to be used has appeared in our publication without credit or acknowledgement to another source;
- suitable acknowledgement to the source is given in accordance with standard editorial practice, e.g.,

  "Reprinted from *Hydrology Research*, volume *x*, issue number *y*, pages *zz-zzz*, with permission from the copyright holders, IWA Publishing."

- reproduction of this material is confined to the purpose for which this permission is given.

I trust this permission will be satisfactory; if any point needs clarification or you have any further queries, please do not hesitate to contact us again.

Yours sincerely

Michelle Herbert
Journals Editorial Co-ordinator
IWA Publishing