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[3]; and Sjur Kolberg[4]

Abstract: Regionalization for prediction in ungauged basins at hourly resolution is important for water resources management (e.g., floods and hydropeaking). In the research reported in this paper, calibration of 26 catchments (39–3,090 km^2) in mid-Norway was performed using hourly records and three spatially distributed (1×1 km^2) precipitation–runoff models, as follows: (1) first-order nonlinear system model, (2) Hydrologiska Byråns Vattenballansavdelning (HBV) model, and (3) basic grid model. Four regionalization methods for each model [(1) parameter set yielding maximum regional weighted average performance measures (PMs), (2) regional median of optimal parameters, (3) nearest neighbor (NN), and (4) physical similarity (PS)] were evaluated and compared with three benchmarks. Parameter transfer from best regional donor and from an ideal best arbitrary single-donor, and local calibration (LC) were as benchmarks. The PS attributes include hypsometric curves, land use, drainage density, catchment area, terrain slope, bedrock geology, soil type, and combination of all. Comprehensive evaluation of single-donors and multidonors, simple benchmarks, and more advanced regionalization methods using multimodels, two PMs, and their statistical evaluation indicate that the identification of regionalization methods is dependent on the models, the PM, and their statistical evaluation. In general, the hypsometric curves, land use, and best regional donor methods performed better for the Nash–Sutcliffe efficiency based on boxplots and regional median values of both the Nash–Sutcliffe efficiency and relative deterioration or improvement of the Nash–Sutcliffe efficiency from the LC due to the regionalization. The methods also performed better for the individual catchments. The terrain slope, regional median of optimal parameters, maximum regional weighted average, and best regional donor methods performed better for the natural-logarithm-transformed streamflow (i.e., regarding the Nash–Sutcliffe efficiency) based on the same evaluation criteria. Similar performance to the more advanced regionalization methods of transfer of homogeneous parameter sets across the whole region from the best regional donor for both Nash–Sutcliffe efficiency and natural-logarithm-transformed streamflow (i.e., regarding the Nash–Sutcliffe efficiency) indicate the potential of the simple regionalization approach for predicting runoff response in the region.
DOI: 10.1061/(ASCE)HE.1943-5584.0001218.

Introduction

Continuous time precipitation–runoff modeling has been used to represent the hydrological processes and understand the basin response. The modeling task entails parameter calibration procedures for gauged basins. However, model identification is not always straightforward due to uncertainties in input data (e.g., climate forcing and streamflow), model structure, and potential nonidentifiability of some model. Moreover, there are further challenges for prediction in ungauged basins (PUBs; Sivapalan et al. 2003) through transfer of information from the calibrated gauged basins to ungauged sites. The uncertainty in precipitation measurments due to the inability of the existing gauges to properly capture the spatial variability of precipitation is a major source of data uncertainty that affects parameter calibration and model prediction. Application of spatially distributed hydrological modeling has been encouraged due to the availability of spatial data and their potential to simulate streamflow at interior catchments. Distributed model efficiency seems to depend on rainfall and model spatial resolution (Pechlivanidis et al. 2011). When the model can capture the spatial information content of precipitation (e.g., McIntyre and Al-Qurashi 2009), effective parameters that are calibrated for a catchment based on spatially distributed inputs, and computations of fluxes and states may have the potential for better performance when transferred to interior locations within the catchment over a calibration based on the lumped counterpart. Hence, spatially distributed modeling and parameter regionalization are important for the PUBs. Regionalization methods have previously been used to transfer knowledge from gauged to ungauged basins (Blöschl and Sivapalan 1995; Oudin et al. 2010), which require evaluation and identification of the methods.

[1]Researcher, Dept. of Hydraulic and Environmental Engineering, Norwegian Univ. of Science and Technology (NTNU), NO-7491 Trondheim, Norway (corresponding author). E-mail: tekhi09@gmail.com

[2]Researcher, Dept. of Energy Systems (Water Resources), SINTEF Energi AS, Sem Sælands vei 11, NO-7465 Trondheim, Norway.

[3]Professor, Dept. of Hydraulic and Environmental Engineering, Norwegian Univ. of Science and Technology (NTNU), NO-7491 Trondheim, Norway.

[4]Researcher, Dept. of Energy Systems (Water Resources), SINTEF Energi AS, Sem Sælands vei 11, NO-7465 Trondheim, Norway.

Note. This manuscript was submitted on July 11, 2014; approved on February 25, 2015; published online on April 13, 2015. Discussion period open until September 13, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699/04015028(20)/$25.00.

Regionalization Methods

Several methods for parameter regionalization have been reported. Parajka et al. (2013) categorized parameter regionalization methods into five groups. The first method is based on regional calibration by utilizing data from multisites in the region. Fernandez et al. (2000) implemented a regression-model-based regional calibration and concluded that improved regional relationships between watershed model parameters and basin characteristics did not result in improvements in the ability to model streamflow at ungauged sites. Beldring et al. (2003) conducted simultaneous calibration of distributed Hydrologiska Byråns Vattenballansavdelning (HBV) model for 141 catchments in Norway by assigning similar parameter values for spatial units with identical landscape classification and concluded that the method provides satisfactory calibration and validation results. Engeland et al. (2006) conducted a multiobjective calibration of a physically based distributed model using multisites streamflow observations to obtain regional parameter sets and concluded that the estimated parameter and streamflow uncertainty depends on the applied method, the chosen objective functions, and the data used. Parajka et al. (2007) proposed an iterative regional calibration (IRC) in which the model parameters of catchments are calibrated simultaneously by defining a combined objective function that linearly combines the local and regional information and concluded that the regional calibration method reduced the uncertainty of most parameters as compared to local calibration (LC). Donnelly et al. (2009) illustrated comparable performance of a simultaneously calibrated spatially homogeneous parameter sets for a multibasin hydrological predictions for the environment (HYPE) hydrological model with locally calibrated parameters. Vaze et al. (2013) showed similar performance of the regional calibration where one parameter set is used to model an entire subregion or region and transfer of local calibrated parameter sets from gauged to ungauged catchments using geographical proximity. Hence, study is required to evaluate the performance for PUB of homogeneous parameter set derived for a region from the regional calibration.

The second method is also multidonor regionalization method based on transferring of averages of optimized parameters for the catchments in the region or so-called parameter averaging (e.g., Kokkonen et al. 2003; Oudin et al. 2008; Kim and Kaluarachchi 2008), using averages of streamflow that are simulated from parameter sets transferred from the donor catchments or so-called output averaging (e.g., Oudin et al. 2008), and ensemble simulations. Using the median statistics rather than the mean may be necessary as the median value is less affected than the mean by poor performance of some donors. McIntyre et al. (2005) illustrated ensemble predictions for the PUBs and concluded that it is important to integrate the results of a wider range of model types. Cibin et al. (2014) transferred probability distributions of parameters rather than transferring a single optimal parameter vector or averaged or interpolated parameter values and obtained that the observed streamflow in the so-called proxy-ungauged basin lies well within the estimated confidence interval of predicted streamflow.

The third regionalization method is based on geographic distance, i.e., nearest neighbor (NN). The NN method is based on the assumption that hydrological properties vary smoothly in space and hence spatial proximity between the donor and the recipient catchments can explain hydrological similarity. Hence, the density of hydrometric gauging stations may affect the performance of the method as the heterogeneity in runoff response may increase as the distance between the donor and recipient catchments increases. Among others, Oudin et al. (2008), Zhang et al. (2009), and Samuel et al. (2011) used distances between catchment centroids; other distances (Gottschalk et al. 2011) can also be used. This method is based on either a single-donor nearest neighbor catchment (e.g., Merz and Blöschl 2004; Parajka et al. 2005) or multidonor nearby catchments (e.g., Parajka et al. 2005; Oudin et al. 2008; Samuel et al. 2011; Arsenault and Brissette 2014).

The fourth regionalization method is physical similarity (PS; e.g., Kokkonen et al. 2003; McIntyre et al. 2005; Parajka et al. 2005; Oudin et al. 2008; Reichl et al. 2009; Zhang and Chiew 2009; Samuel et al. 2011; Arsenault and Brissette 2014), which is based on the assumption that similarity in some physical attributes that govern the runoff response could explain the hydrological similarity. This method is also based on either a single-donor catchment or multidonor catchments. The method requires identification of the physical and climate attributes (e.g., Sawicz et al. 2011; Viglione et al. 2013), which influences the runoff response in the region of the research reported in this paper.

The fifth regionalization method is regression-based, which uses data from many catchments to develop regression relationships for instance among model parameters and catchment characteristics. Bárdossy (2007) concluded that regionalization should not focus on relating catchment properties to individual parameter values but on relating them to compatible parameter sets. 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 illustrated various limitations in the regression-based regionalization methods (e.g., Fernandez et al. 2000; Lamb and Kay 2004; Wagener and Wheater 2006; Bastola et al. 2008; Bulygina et al. 2009; Pechlivanidis et al. 2010).

Factors That Influence Regionalization Performance

There are different challenging factors pertinent to the identification of suitable regionalization methods for instance selection of donor catchments, identification of models, and performance measures (PMs). Wagener and Wheater (2006) noted uncertainties pertinent to estimation of continuous streamflow time-series in ungauged basins. Evaluations of the performance of different regionalization methods are performed in literature in order to identify suitable methods (e.g., Parajka et al. 2005; Oudin et al. 2008; Zhang and Chiew 2009). However, comprehensive comparative investigations related to PUB are required for a geographic region of interest.

In general, previous comparisons of regionalization methods were mainly conducted at low spatial and temporal resolution (i.e., lumped-semidistributed and/or daily time step). Distributed models are expected to provide more opportunity for prediction at ungauged locations. In addition, predictions at 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. Littlewood and Croke (2013) indicated the importance of a fine temporal resolution also for extraction of the information content of the data for accuracy of calibrated parameters, which would allow more reliable parameter regionalization.

Uncertainties related to model structure, parameter calibration, and input data affect the performance of the regionalization of precipitation–runoff models for continuous simulation of streamflow (Wagener and Wheater 2006; Oudin et al. 2008, 2010; Kim and Kaluarachchi 2008; Gupta et al. 2008). Engeland and Gottschalk (2002) noted that structural errors in the model are more important than parameter uncertainties. Oudin et al. (2010) noted 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. Several previous 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 or quality of input climate data for PUBs.

Previous attempts for continuous streamflow simulation by rainfal–runoff models for PUBs were mainly based on conceptual rainfall–runoff modeling, i.e., the HBV model (e.g., Siebert 1999; Merz and Blöschl 2004; Parajka et al. 2005; Götzinger and Bárdossy 2007; Samuel et al. 2011), the probability distributed model (PDM; Moore 1985), and its variants (e.g., McIntyre et al. 2005; Zhang and Chiew 2009; Pechlivanidis et al. 2010). Regionalization based on data based so-called top–down rainfall–runoff modeling paradigm, for instance based on model equations and parameters that can be inferred from catchment storage–discharge relationships as illustrated by Kirchner (2009), is not common. Croke and McIntyre (2013) and Hrachowitz et al. (2013) noted the importance of such parsimonious approach for PUBs.

Selection of proper model structure is important in regionalization study. Parajka et al. (2007) concluded that it would be worth improving the model efficiency for the local calibration case by varying the model structure between catchments depending on regional runoff processes. Lee et al. (2005) while selecting conceptual models for regionalization of catchments in int United Kingdom concluded that the study provided no evidence of relationships between catchment types and model structures. Therefore, comprehensive evaluation of the regionalization methods based on different modeling paradigms and several model structures is indispensable.

Furthermore, Lee et al. (2005) and Wagener and Wheater (2006) illustrated the incapability of regionalized models to simulate both high-flow and low-flow behaviors of catchments simultaneously. Model identification or performance is dependent on the objective functions used (Gupta et al. 1998; Madsen 2003; Muleta 2012). Patil and Stieglitz (2011) also illustrated based on flow duration curves for catchments in the United States that similarity among catchments is not preserved at all flow conditions. Some catchments that are similar in their rainfall-dependent and snowmelt-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. However, Oudin et al. (2006) illustrated good efficiency for both low and high flows through model combination. Hence, dependency of the regionalization on the performance measures is also an additional challenge in the regionalization endeavors, which needs consideration. Further references on regionalization works can be found from review papers 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 transferring hydrological information through rainfall–runoff model parameters from gauged to ungauged catchments within a certain region [a recent review was provided 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 [a recent synthesis was provided by Parajka et al. (2013)].

Scientific Questions and Objectives of the Paper

The research reported in this paper will attempt to answer the following questions: (1) whether the performance of the regionalization methods are consistent among model structures, performance measures, and statistical approaches used for evaluation of the PM; and (2) which regionalization method performs best for the specific region of study. The research reported in this paper is the first attempt for distributed hourly runoff simulation in the region of the research reported in this paper, and it would contribute scientifically to the growing interest for hourly prediction in ungauged basins pertinent to hydropower operation (e.g., hydropeaking, floods, environmental flow assessments, and prediction of natural flow).

The paper is organized as described next. The next two sections provide brief explanation of the region of the research reported in this paper and data, and models and methods used. Then, the results of the regionalization study are presented and then discussions are presented. The last section provides conclusions.

Investigated Region and Data

The study catchments were selected from large set of catchments in the mid-Norway region. However, due to the extensive regulation, only 26 unregulated catchments in the region that are gauged by the Norwegian Water and Energy Directorate (NVE) and range in drainage area from 39 to 3,090 km^2 were considered. Streamflow and climate records of hourly time resolution from September 2007 to September 2010 for parameter calibration were used. The relatively short period is due to a lack of long, good-quality hourly climate data. The model run was started in September and the first year was used for model warm-up to reduce the effects of initial states. The hourly climate forcing used in the research reported in this paper are precipitation (P) in millimeters, temperature (T) in degrees Celsius, wind speed (W_s) in meters per second, relative humidity (H_R) in percentage, and global radiation (R_G) in Watts per square meter. The climate data from point measurements are spatially interpolated on 1×1 km^2 using the inverse distance weighing (IDW) method. Lists of the catchments and streamflow stations, and some characteristics of the catchments are given in Table 1. The climate gauging stations are mostly outside the study catchments and are not equally representative for the catchments. Locations of catchments, precipitation, and streamflow gauging stations are given in Fig. 1.

Precipitation occurs mainly in the form of snowfall during winter [December, January, and February (DJF)] and as rainfall during summer, spring, and autumn. Hence, runoff dynamics is influenced by both rainfall–runoff and snowmelt–runoff processes. High flow regimes for most of the study catchments are related to snowmelt events (nival regime). In addition, some catchments exhibit rainfall on snowmelt events (pluvial and nival combined) and rainfall events (pluvial) dependent high-flow regime.

Table 2 presents lists of seven physical attributes that were used for physical similarity-based regionalization method [(1) hypsometric curves (termed PSH), (2) land use (termed PSL), (3) drainage density (termed PSD), (4) catchment area (termed PSA), (5) cumulative distribution function (CDF) of terrain slope (termed PSS), (6) bedrock geology (termed PSR), and (7) soil types (termed PSSO)]. The dominant land use/land cover in the study area is bare rocks in mountains above timberline and forests. There is also significant proportion of marshes/bogs and lake areas for some of the catchments. Five of the study catchments have glacier coverage. Predominant soil or loose material is glacial tills. The dominant bedrock types for the study catchments are metamorphic and igneous rocks. Hypsometric curves and land-use data were obtained from the website of Norwegian water and Energy Directorate (http://www.nve.no) and the soil and bedrock geology data were obtained from the website of Norwegian Geological Survey or NGU (http://www.ngu.no). Stream networks from the 1:50,000 maps produced by the Norwegian Mapping Authority and obtained from the website http://www.statkart.no were used.

J. Hydrol. Eng., 2015, 20(11): 04015028

Table 1. Some Characteristics of the Investigated Catchments

Catchment number[a]	Catchment name	Station number[b]	Station altitude (m)[c]	Catchment area (km^2)	P_{ca} (mm/year)[d]	T_{maxca} (°C)[d]	T_{minca} (°C)[d]
1	Dillfoss	127.13	25	480	1,333	27	−27
2	Driva v/Risefoss	109.9	550	745	1,029	23	−32
3	Eggafoss	122.11	330	668	1,042	26	−31
4	Embrethølen	139.26	160	495	1,153	26	−33
5	Feren	124.13	401	220	1,354	26	−28
6	Gaulfoss	122.9	45	3,090	982	27	−31
7	Gisnås	121.29	580	95	929	26	−30
8	Hugdal bru	122.17	135	546	914	28	−31
9	Høggås bru	124.2	97	495	1,362	26	−27
10	Isa v/Morstøl bru	103.2	103	44	1,217	24	−25
11	Kjeldstad i Garb	123.31	173	145	1,441	27	−31
12	Krinsvatn	133.7	87	207	1,861	29	−25
13	Lenglingen	308.1	354	450	1,174	25	−32
14	Lillebudal bru	122.14	515	168	930	26	−32
15	Murusjø	307.5	311	346	1,193	26	−31
16	Osenelv v/Øren	105.1	12	138	1,337	28	−19
17	Rauma v/Horgheim	103.4	60	1,100	1,186	23	−27
18	Rinna	112.8	460	91	1,345	25	−28
19	Skjellbreivatn	139.25	354	546	1,127	26	−32
20	Søya v/Melhus	111.9	40	138	1,677	27	−28
21	Støafoss	128.5	80	477	1,268	27	−27
22	Trangen	139.35	137	852	1,116	26	−34
23	Valen	117.4	10	39	1,386	29	−23
24	Valldøla v/Alstad	100.1	265	226	1,247	24	−24
25	Vistdal	104.23	50	67	1,258	25	−23
26	Øyungen	138.1	103	239	1,353	27	−26
	Median values	—	136	293	1,232	26	−28

[a]Catchments 3, 8, and 14 are subcatchments of Catchment 6; Catchment 5 is a subcatchment of Catchment 9; and Catchment 19 is a subcatchment of Catchment 22.
[b]As per the Norwegian Water and Energy Directorate.
[c]Meters above sea level.
[d]Catchment averaged precipitation (P_{ca}) and catchment averaged maximum and minimum temperature (T_{maxca} and T_{minca}) are spatially interpolated based on IDW and catchment averaged values of mean annual precipitation, and maximum and minimum temperature of the hourly observations for the calibration period of 2008–2010.

Models and Methods

Three distributed (1×1 km^2 grid) precipitation–runoff models [(1) so-called top–down water balance model (termed Kirchmod; Kirchner 2009), (2) Hydrologiska Byråns Vattenballansavdelning model (Lindström et al. 1997), and the basic grid model (BGM; Bell and Moore 1998) were used. Lists of calibrated parameters along with their prior ranges are given in Table 3. Model structures of the runoff response routines are presented in Figs. 2(a–c). Brief descriptions of the models are given in the subsequent subsections.

Runoff Response Routine

Kirchner (2009) inferred model equations and parameters from analysis of streamflow recession (i.e., the so-called top–down modeling paradigm). The method is based on a catchment storage–discharge relationship. The main assumption in the Kirchner (2009) method is that the streamflow depends solely on the amount of water stored in the catchment. The water balance response routine is given in terms of the actual evapotranspiration (AET) in millimeters and infiltration in to the soil (I) in millimeters, the latter of which is the rainfall plus the snowmelt, as

$$\frac{dQ}{dt} = \frac{dQ}{dS}\frac{dS}{dt} = \frac{dQ}{dS}(I - \mathrm{AET} - Q) = g(Q)(I - \mathrm{AET} - Q) \quad (1)$$

where $dQ/dS = g(Q)$, the discharge sensitivity function (Kirchner 2009); t = time; S = catchment averaged storage (mm); and Q = discharge (mm). The runoff simulation was based on the subsequent integral equation [Eq. (2a)] and the subsequent regression relationship [Eq. 2b] between the $g(Q)$ and Q was considered

$$S(Q) = \int \frac{1}{g(Q)} dQ \quad (2a)$$

$$\ln[g(Q)] \approx \beta_0 + \beta_1 \ln(Q) \quad (2b)$$

where β_0 and β_1 are calibrated model parameters. A Runge–Kutta fourth-order method was used to integrate the equation over the time step. The AET were computed from the potential evapotranspiration (PET) and discharge according to

$$\mathrm{AET} = \mathrm{PET}\left[1 - \exp\left(-\frac{Q}{EvR}\right)\right](1 - \mathrm{SCA}) \quad (3)$$

where EvR is a parameter which denotes a discharge at which AET equals $0.95 \times$ PET; and the snow-covered area (SCA) = fraction of grid cell that is snow covered, i.e., AET is set to zero for snow-covered areas. The Q is an instantaneous simulated discharge solved by the numerical solver while the average Q over the computational time step of 1 h (millimeters per hour) is used for calibration against a hourly averaged observed discharge. The observed hourly averaged discharge one time step before the start of model run was used as an initial discharge for the numerical solver.

J. Hydrol. Eng., 2015, 20(11): 04015028

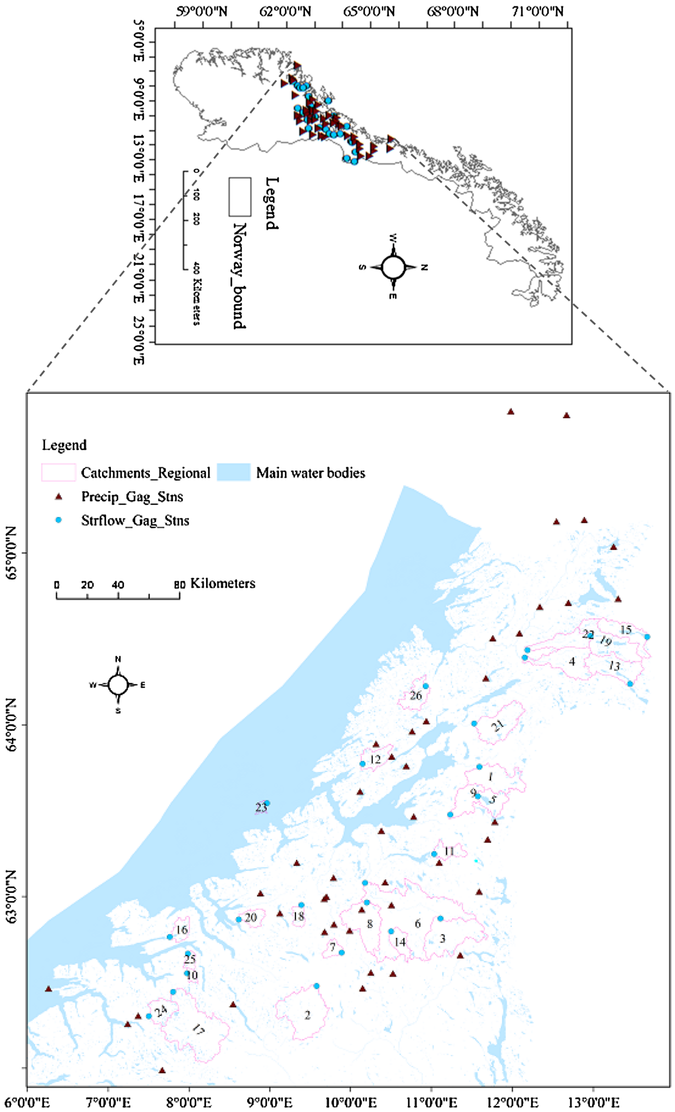

Fig. 1. Locations of the region of the research reported in this paper, catchments, and hydroclimatic stations

Hydrologiska Byråns Vattenballansavdelning Model

The HBV runoff response routine used in the research reported in this paper contains two conceptual storage reservoirs. The relationship between the single-outlet upper storage reservoir and outflow is nonlinear while the relationship between the single-outlet lower storage reservoir and baseflow is linear. The outflow from the upper and lower reservoirs [in terms of the upper zone (UZ) and lower zone (LZ)] conceptually represent the quick flow from overland flow and so-called superficial drainage (Q_{UZ}), and baseflow from ground water storage (Q_{LZ}), respectively

$$Q_{UZ} = k_1 \times (UZ)^{n_u} \tag{4a}$$

$$Q_{LZ} = k_0 \times LZ \tag{4b}$$

where n_u is an exponent of nonlinearity for the upper zone; UZ and LZ are both in millimeters; and k_1 and k_0 are recession coefficient parameters. Percolation (PERC) is a parameter that modulates percolation from the upper to the lower reservoir. The soil moisture accounting routine is based on Bergström (1976). A nonlinear function partitions the infiltration into change in soil moisture storage (Δ_{SM}) and recharge to the upper zone (R_{uz}). If $SM > (LP \times FC)$, AET = PET but if $SM < (LP \times FC)$, AET = $(\text{PET} \times SM)/(LP \times FC)$, in terms of soil moisture (SM), field capacity (FC), and the so-called limit for potential evaporation (LP).

Basic Grid Model

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 so-called fill-and-spill-type saturation excess runoff, R (L; Dunne and Black 1970a, b) and a subsurface flow or drainage based on a nonlinear catchment storage–discharge relationship

$$R_{\text{iex}} = \max[0, (\text{SNOWOUT} - I_c)] \tag{5a}$$

$$I = \text{SNOWOUT} - R_{\text{iex}} \tag{5b}$$

$$R = \max\{0, [S(t) + \text{TOSTORAGE} - S_{\max}]\} \tag{6a}$$

$$S(t + \Delta t) = \max\{0, [S(t) + \text{TOSTORAGE} - R]\} \tag{6b}$$

$$\text{TOSTORAGE} = I - \text{AET} - D_{rv} \tag{7a}$$

$$\text{AET} = \text{PET} \times \frac{S}{S_{\max}} \tag{7b}$$

$$D_{rv} = k[S(t)]^n \tag{7c}$$

where I_c (L/T) is an infiltration capacity parameter; SNOWOUT (L) = rainfall and snowmelt outflow from snow routine; TOSTORAGE = net input to the subsurface storage variable [S (L)]; $D_{rv}(L)$ = subsurface flow or drainage per unit area variable; and k (L^{1-n}/T), n (dimensionless), and maximum subsurface storage capacity [$S_{\max}(L)$] are calibrated parameters. The optimal values of the parameters were determined by the calibration algorithm.

Snow Accumulation and Melting Routine

The influences of snow processes are dominant in the area of the research reported in this paper during winter and spring seasons. The snow routine simulates snow accumulation and the outflow melt water release from saturated snow (Q_s) based on the gamma distributed snow depletion curve (SDC; Kolberg and Gottschalk 2006), which was implemented in the ENKI hydrological modeling platform (Kolberg and Bruland 2012). The calibrated parameters in this routine are TX and WS; i.e., snow–rain threshold temperature parameter and snowmelt sensitivity to wind speed or windscale, respectively. The same snow routines were used with all three runoff response routines.

Potential Evapotranspiration Routine

The Priestley–Taylor method (Priestley and Taylor 1972) was used for the calculation of potential evapotranspiration (millimeters per hour) for all routines

$$\text{PET} = \alpha \frac{\Delta}{\Delta + \gamma}(R_n)\left(\frac{\Delta t}{L_v}\right) \tag{8}$$

where α = Priestley Taylor constant; Δ = slope of saturation vapor pressure curve at air temperature at 2 m (kPa/°C); γ = psychrometric constant (0.066 kPa/°C); R_n (W/m^2) = net radiation = net shortwave radiation (SR_n) + net longwave

Table 2. Descriptions and Ranges of the Physical Attributes for the Investigated Catchments

Attributes[a]	Description[b]	Unit[c]	Minimum	Maximum
PSH	Hypsography, 0%	m	7	582
PSH	Hypsography, 20%	m	56	1,141
PSH	Hypsography, 40%	m	71	1,290
PSH	Hypsography, 60%	m	81	1,400
PSH	Hypsography, 80%	m	97	1,529
PSH	Hypsography, 100%	m	299	2,283
PSL	Farm land	%	0	6
PSL	Forest	%	7	57
PSL	Mountains above timberline and bare rock	%	0	84
PSL	Glacier	%	0	5
PSL	Marshes, i.e., bog	%	0	28
PSL	Lakes	%	1	15
PSD	Drainage density	$\mathrm{km/km^2}$	1	3
PSA	Catchment area	$\mathrm{km^2}$	39	3,090
PSS	Cummulative distribution function of slope, 0.20	degree	1	12
PSS	Cummulative distribution function of slope, 0.40	degree	3	22
PSS	Cummulative distribution function of slope, 0.60	degree	5	30
PSS	Cummulative distribution function of slope, 0.80	degree	8	38
PSS	Cummulative distribution function of slope, 1.00	degree	44	86
PSR	Sandstone and slate	%	0	35
PSR	Granite and granodiorite	%	0	61
PSR	Diorite and monzodiorite	%	0	82
PSR	Rhyolite, rhyodacite, and dacite	%	0	56
PSR	Greenstone and amphibolites	%	0	31
PSR	Meta sandstone and slate	%	0	30
PSR	Mica gneiss, schist, amphibolite, and metasandstone	%	0	98
PSR	Phyllite and schist	%	0	32
PSR	Dioritic to granitic gneiss and migmatite	%	0	100
PSSO	Thick moraine	%	0	41
PSSO	Thin moraine	%	7	47
PSSO	Peat and marsh, i.e., organic material	%	0	31
PSSO	Humus/thin turf cover	%	0	32
PSSO	Bare rock/mountains with thin turf	%	4	56

[a]PSA refers to physical similarity in catchment area, PSD refers to physical similarity in drainage density, PSH refers to physical similarity in hypsometric curves, PSL refers to physical similarity in land use, PSR refers to physical similarity in bedrock geology, PSS refers to physical similarity in cumulative distribution function of terrain slope, PSSO refers to physical similarity in soil types, and PSC refers to physical similarity in all attributes.
[b]Data points of the hypsography and cumulative distribution function of terrain slope used for similarity distance calculation are more than what is presented in Table 2.
[c]Meters refers to meters above sea level.

radiation (LR_n); L_v ($\mathrm{kJ/m^3}$) = volumetric latent heat of vaporization; and Δt = simulation time step (s). The SR_n was computed from the global radiation (R_G) and land albedo, and the LR_n was computed based on Sicart et al. (2006). In accordance with Teuling et al. (2010), $\alpha = 1.26$ was used rather than setting it by calibration in order to reduce the number of calibrated parameters.

Runoff Routing

A simple translation based on 1-h isochrones was implemented for all models. The hillslope runoff response of each 1×1 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

$$Q_{\mathrm{sim}_t} = \sum_{i=1}^{N} q_{\mathrm{sim}^i_{t-T_i}} \tag{9a}$$

$$T_i = \frac{L_i}{V} \tag{9b}$$

where t and i represent time and grid cells, respectively; N = number of grid cells in the catchment; Q_{sim} ($\mathrm{LT^{-3}}$) = streamflow at the outlet; q_{sim} ($\mathrm{LT^{-3}}$) = runoff generated at each grid cell; $T_i(T)$ = flow travel time lag to the outlet for each grid (T); L_i (L) = flow travel path length computed from 25-m digital elevation model (DEM); and V ($\mathrm{LT^{-1}}$) = velocity of flow, which was set by calibration.

Model Identification

The Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al. 2009) was used for model calibration based on the log–likelihood (L–L) objective function, which was implemented in the ENKI hydrological modeling platform (Kolberg and Bruland 2012). A regional calibration by the DREAM algorithm can be regarded as a so-called importance sampling strategy for each catchment, where sampling is according to a so-called importance surface reflecting where the optimum is likely to be. The DREAM seeks and converges to the posterior distribution and the DREAM 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.

J. Hydrol. Eng., 2015, 20(11): 04015028

Table 3. Lists of Calibrated Parameters and Their Prior Ranges for the Models

Parameters	Description	Kirchmod	HBV[a]	BGM[b]	Units	Prior ranges minimum, maximum
Snow						
TX	Threshold temperature	Yes	Yes	Yes	°C	−3, 2
WS	Snowmelt sensitivity to wind speed	Yes	Yes	Yes	—	1, 6
Soil and evapotranspiration						
FC	Maximum soil moisture or field capacity	No	Yes	No	mm	50, 600
LP	Value at and above which AET = PET	No	Yes	No	—	0.5, 0.99
β	Shape parameter of the curve partioning the infiltration in to UZ recharge and change in soil moisture	No	Yes	No	—	0.5, 5
Runoff response						
EvR	Discharge at which AET equals 0.95 × PET	Yes	No	No	$\mathrm{mm\,h^{-1}}$	0.1, 6.15
β_0	Regression Parameter 1	Yes	No	No	—	−8, 0
β_1	Regression Parameter 2	Yes	No	No	—	−1, 1
k_1	Recession coefficient of the upper reservoir	No	Yes	No	$\mathrm{day^{-1}}$	0.001, 1.5
k_0	Recession coefficient of the lower reservoir	No	Yes	No	$\mathrm{day^{-1}}$	0.0005, 0.5
n_u	Nonlinearity exponent of the upper reservoir	No	Yes	No	—	0.2, 5
$PERC$	Percolation rate from the upper to lower reservoir	No	Yes	No	$\mathrm{mm\,day^{-1}}$	0, 6
$S_{\max}$	Maximum storage capacity	No	No	Yes	mm	150, 1,000
I_c	Infiltration capacity of the soil	No	No	Yes	$\mathrm{mm\,h^{-1}}$	0.1, 40
k	Coefficient of storage–drainage relationship	No	No	Yes	$\mathrm{mm^{1-n}\,h^{-1}}$	10^{-7}, 10^{-3}
n	Exponent of storage–drainage relationship	No	No	Yes	—	0.2, 5.0
Routing						
V	Velocity of flow	Yes	Yes	Yes	$\mathrm{m\,s^{-1}}$	0.25, 3.5
Total numbers of calibrated parameters		6	10	7	—	—

[a]HBV = Hydrologiska Byråns Vattenballansavdelning.
[b]BGM = basic grid model.

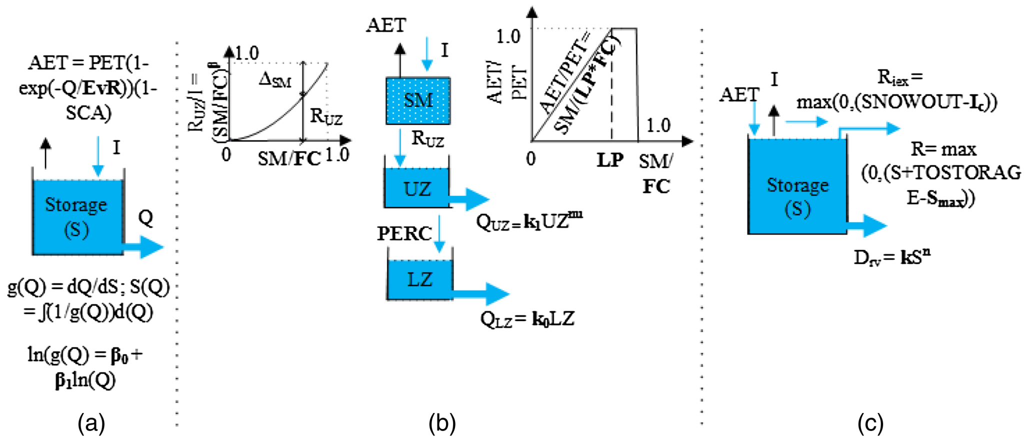

Fig. 2. Model structures (response routines): (a) Kirchmod model; (b) HBV model; (c) BGM model. Bold-font terms are calibrated parameters

The log–likelihood objective function used for the regional calibration is

$$L - L\left\{\delta/\sigma_i^2, \sum_{i=1}^{N_C}\sum_{t=1}^{n_i}[Q^{(\theta)}_{\mathrm{sim}_{t,i}} - Q^{(\theta)}_{\mathrm{obs}_{t,i}}]^2\right\}$$
$$= \left(\sum_{i=1}^{N_C}\left\{\frac{-n_i}{2}\log(2\pi) - \frac{n_i}{2}\log(\sigma_i^2) - \frac{\sum_{t=1}^{n_i}[Q^{(\theta)}_{\mathrm{sim}_{t,i}} - Q^{(\theta)}_{\mathrm{obs}_{t,i}}]^2}{2\sigma_i^2}\right\}\right) \times f \quad (10)$$

where δ = model parameter; σ_i^2 and n_i = error variance and the length of nonmissing records, respectively, of streamflow for catchment i; $N_C = 26$ = total number of modeled catchments in the region; $Q^{(\theta)}_{\mathrm{sim}}$ and $Q^{(\theta)}_{\mathrm{obs}}$ are Box–Cox (Box and Cox 1964) transformed observed and simulated streamflow time series, respectively, to approximate normal distributed and homoscedastic series; θ is the Box–Cox transformation parameter; and f = fraction of effectively independent observations. Values of θ between 0.25 and 0.30 are common in literature (e.g., Vrugt et al. 2002; Willems 2009). For the sake of consistency, $\theta = 0.3$ and $f = 0.001$ were used for all catchments.

Evaluation of the local calibration and the regionalization methods were carried out based on two PM {(1) Nash–Sutcliffe efficiency (Nash and Sutcliffe 1970), i.e., [−∞, 1.0], and (2) NSEln, i.e., the NSE equivalent for natural-logarithm-transformed series i.e., [−∞, 1.0]}

$$\mathrm{NSE}_i = 1 - \frac{\sum_{t=1}^{n_i}(Q_{\mathrm{obs}_{t,i}} - Q_{\mathrm{sim}_{t,i}})^2}{\sum_{t=1}^{n_i}(Q_{\mathrm{obs}_{t,i}} - \bar{Q}_{\mathrm{obs},i})} \quad (11)$$

For the NSEln calculation, time steps with zero values of either observed or simulated streamflow were skipped. Adding an arbitrary selected value is highly influential, and setting it closer to zero makes it even worse due to the behavior of $\ln(x)$ as x approaches zero. In the region and catchment sizes studied in the research reported in this paper, zero discharge is not an issue even in midwinter. It is not a sign of catchment behavior that the model should be required to mimic. However, rare zero values of streamflow in the research reported in this paper could also be avoided by increasing the accuracy of the streamflow database to three decimals. The NSE gives higher weightage to high flow while the NSEln gives higher weightage to low flow. Better statistic for drought performance was not the objective of this paper as low flow performance in Norway are linked to legal minimum flows in regulated rivers, not to drought applications. The evaluation metrics for the PM include their regional median values, regional median of deterioration or improvement of PM from the local calibration due to the regionalization, box plots of PM values, and their deterioration or improvement from the local calibration due to the regionalization. Optimized (locally calibrated) parameter sets, which yield maximum values of NSE and NSEln for each catchment, were defined as local calibration. These locally identified complete sets of parameters were transferred from each catchment to the remaining ones based on the different regionalization methods.

Regionalization Methods

In total seven parameter transfer experiments were investigated in which three are used as benchmark, i.e., reference model performance, while four regionalization methods are further assessed against benchmarking performance.

Benchmarks

Three basic parameter transfers were considered as benchmarks to compare the performance of the more advanced regionalization method(s). A best arbitrary single-donor (BASD), which is an ideal case of parameter transfer, was used as a benchmark. The BASD for each recipient catchment was identified from an arbitrary transfer of the optimal parameter sets from the $N_c - 1$ potential donors without employing any regionalization method. It provides the maximum possible PM that can be obtained from the single-donor transfer of parameters. Each donor parameter set on all catchments was also tested to identify the parameter sets that provides the highest regional median PM and named it best regional donor (BRD) as a second benchmark. The performance of the LC was also used as a benchmark to evaluate the deterioration or improvement or the so-called spatial loss or gain of the PM from the LC due to the regionalization.

Regional Calibration

This method explores the performance of the regional calibration based on parameter sets that provide the maximum regional weighted average (MRWA) PM among those computed using Eq. (6) for both NSE and NSEln for each parameter set accepted by the DREAM algorithm. It involves identification of parameter sets that provide the MRWA PM for the region. In this method, a homogeneous parameter set is derived for the region corresponding to each PM in terms of the regional weighted average (RWA) NSE

$$\mathrm{NSE}_{\mathrm{RWA}} = \frac{1}{N_C}\sum_{i=1}^{N_C}\left(\frac{n_i}{N_{TS}}\right)\mathrm{NSE}_i \qquad (12)$$

where N_{TS} = length of timestamp for the calibration period (including the missing records); and the weights are the term in the parenthesis assigned for each catchment based on the length of their nonmissing streamflow records during the calibration period. The regional calibration regionalization method used in the research reported in this paper is similar to previous works on the regional calibration in terms of utilizing the streamflow data from multiple catchments in the region. However, the regional parameter sets corresponding to the maximum regional weighted average PM obtained from the total modeled catchments are used in the research reported in this paper.

Regional Median Parameters

This method evaluates the performance of regional median parameter (RMedP) set derived for a region from the optimized parameter set for each catchment

$$\mathrm{RMedP}_j = \mathrm{Median}(P_j^1, P_j^2, \ldots, P_j^{N_C}) \qquad (13)$$

where the subscript j denotes the free parameters ($j = 1$–N_p, where N_p = total number of calibrated parameters). This method allows pooling of parameters for each PM from multiple donor catchments (i.e., multidonor median of parameter set) and then transferring homogeneous parameter sets for the whole region for prediction. The only difference between this method and the so-called parameter averaging pooling option 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 used. However, a limitation of transferring either median or mean parameters is that the method transfers the regional median or mean of each parameter rather than a set of optimal parameters and hence does not keep the correlation structure of the calibrated model parameters.

Nearest Neighbor

In general, this method assumes that the spatial proximity could explain hydrological homogeneity. The optimized parameters were transferred from the nearest neighbor single-donor catchment to the recipient catchment(s). The Euclidean distance in the geographic coordinate spaces of the streamflow gauging stations was used to identify the nearest neighbor for the catchments. This distance was chosen since streamflow at the catchment outlets, which integrates the effects of catchment area in terms of the spatial variability of runoff dynamics and the effects of runoff delay, was used for calibration. In addition, the so-called top–down modeling paradigm of Kirchmod was based on the use of observed streamflow at the gauging stations.

Physical Similarity

The method assumes that similarity of catchments in physical attributes could explain their homogeneity in runoff response. In the research reported in this paper, the optimized parameter sets from a single donor catchment were transferred to a recipient catchment that is most similar in the physical attributes. The method is subject to the selected physical attributes governing the runoff response (e.g., Sawicz et al. 2011; Viglione et al. 2013), which require availability of reliable data and subjective judgment in selection of attributes. In the research reported in this paper, selection of the attributes was based on their relevance in influencing the runoff response of the study area, availability of representative data, and findings from previous studies, e.g., the dominant topographic influences on the runoff response for boreal catchments as reported

J. Hydrol. Eng., 2015, 20(11): 04015028

in Halldin et al. (1999) and Beldring et al. (2003). Eight different cases of physical similarity were evaluated, as follows: (1) similarity in hypsometric curves, PSH; (2) land use, PSL; (3) drainage density, PSD; (4) catchment area, PSA; (5) cumulative distribution function of terrain slope, PSS; (6) bedrock geology, PSR; (7) soil types, PSSO; and (8) combinations of all attributes, PSC.

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 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 bedrock hydraulic properties, which mainly 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 was calculated between the catchments ($\mathrm{Dist}_{j,h}$) after the [0,1] normalization of the classes in each attribute for the sake of simplicity or scaling of the values

$$\mathrm{Dist}_{j,h} = \sum_{i=1}^{a} \left\{ \left[\sum_{k=1}^{c} (Nx_{i,k}^{j} - Nx_{i,k}^{h})^2 \right]^{1/2} \right\} \tag{14a}$$

$$Nx_{i,k}^{j} = \frac{x_k^j - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{14b}$$

$$Nx_{i,k}^{h} = \frac{x_k^h - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{14c}$$

where i and k are indices for the attribute x and classes, respectively; j and h are indices for the two catchments to be compared; a = total number of the attributes considered; c = total numbers of classes in the attribute; N refers to normalized; and min and max are the minimum and the maximum values, respectively, for the whole catchments of the classes in the attribute i. Equal weightages were assigned to each class in the attributes and to each attribute for the combined attributes. For the sake of explanation, for the land-use attribute in Table 2, there are six classes [(1) farmland, (2) forest, (3) mountains, (4) glacier, (5) marshs (bog), and (6) lakes] and for the PSC there are seven attributes.

The NN and PS methods were used based on a single-donor and rank-similarity approach. For each recipient catchment, the $N_C - 1$ potential donor catchments were ranked based on their geographic proximity and physical similarity distances (i.e., eight cases), respectively, for the NN and PS. A rank value of 1 refers to the nearest or most similar catchment, i.e., best single donor, from which the complete set of local calibrated parameters is transferred to the recipients. Hence, the regionalization methods were evaluated without employing any sophisticated clustering methods to form homogeneous subregions for parameter transfer within subregions. The small number of unregulated gauged catchments in the region (i.e., only 26) compared to the wide extent of the region of the research reported in this paper also suggests analyses based on the total $N_c - 1$ potential donors. In addition, it keeps consistency in the numbers of potential donors among the regionalization methods.

Results

Regional Performance

Box plots of the values of the PM (both NSE and NSEln) for the benchmarks LC and BRD, and regionalization methods for the Kirchmod, HBV, and BGM models are given in Figs. 3(a–c), respectively. For the Kirchmod, regional median NSE and NSEln are higher for the PSH and RMedP regionalization methods respectively compared to the other methods. For the HBV model, regional median NSE and NSEln are higher for the benchmark BRD, and both MRWA and BRD, respectively, compared to the other methods. For the BGM model, regional median of both NSE and NSEln are higher for the benchmark BRD compared to the other more advanced methods. The PM of the Kirchmod model is slightly higher for RMedP for NSE, and for both RMedP and BRD for NSEln compared to the other methods. The PM of HBV is higher for PSH and BRD for NSE and NSEln, respectively, compared to the other methods. Similarly, the PM of BGM is higher for the multidonor based on RMedP and MRWA for NSE and NSEln, respectively.

Since regionalization focuses on identifying the best regional solution, it would generally involve compromises in the PM of the LC for the individual catchments. Comparisons of the regionalization methods based on PM are affected by the results of poorly performing catchments. Therefore, it was aimed to perform comparisons of the regional performance of the regionalization methods based on a relative measure that has the potential to reduce the effects of poor LC for some catchments, for instance, due to poor or unrepresentative input data. To this end, box plots of relative deterioration or improvement in the PM from the local calibration due to the regionalization, which is calculated as (regionalization PM−local calibration PM)/local calibration PM are presented in Figs. 4(a–c) for the Kirchmod, HBV, and BGM models, respectively. The results indicate that the PSC and PSS methods provided low regional median relative deterioration for NSE and NSEln, respectively, for both Kirchmod and HBV models. The PSH and MRWA methods provided low regional median relative deterioration for the NSE and NSEln, respectively, for the BGM model. In terms of NSEln, the BRD provided similar performance to the MRWA for the HBV model and to the PSS for the BGM model.

Performance for the Individual Catchments

Even though the main objective of the regionalization procedures in the research reported in this paper was to identify the best-performing regionalization methods for the whole region, evaluation of performance for each catchment could provide additional clues, for instance, why some catchments performed badly. Values of PM for each catchment corresponding to the benchmarks (LC, BASD, and BRD) and for a hypothetical best regionalization methods performance (BRMP) from both single-donor and multidonors for each catchment are given in Tables 4 and 5 for the NSE and NSEln, respectively. The best regional donor catchments vary with the model and PM; Catchment 9 is the best regional donor for both NSE and NSEln for the Kirchmod; Catchments 22 and 21 are the best regional donors for NSE and NSEln, respectively, for the HBV model; and Catchments 1 and 9 respectively are the best regional donors for NSE and NSEln for the BGM model. These catchments performed better, as best regional donors, than Catchments 3 and 6 that have climate stations inside their boundary. However, more reliable parameter calibration based on data from high-density climate stations has the potential to further improve the performance of the BRD. The PM obtained from the hypothetical BRMP

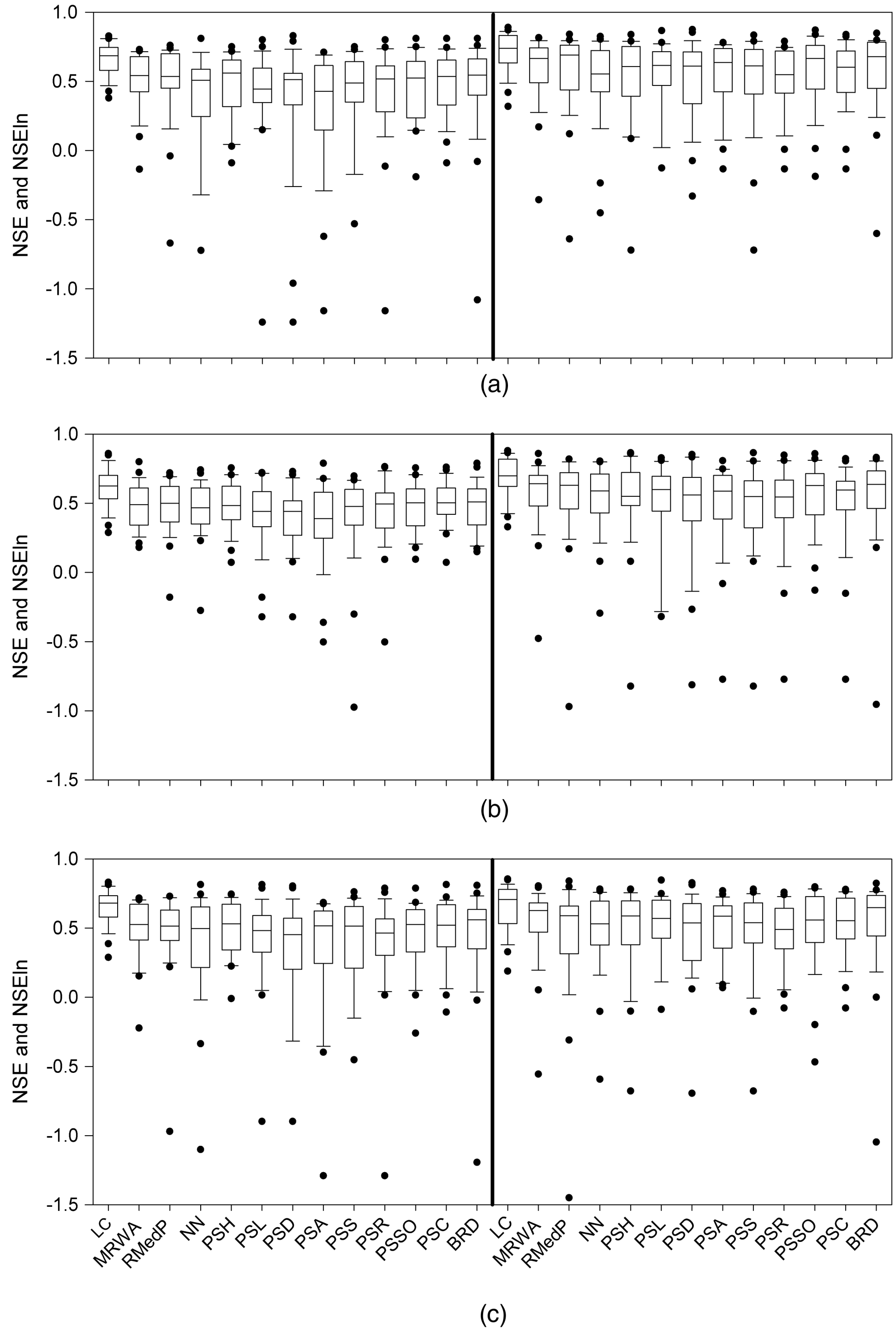

Fig. 3. Box plots of the performance measures for the benchmarks (LC and BRD) and regionalization methods for the following: (a) Kirchmod model; (b) HBV model; (c) BGM model; first 13 plots (from LC to BRD) correspond to the NSE and the last 13 plots (LC to BRD) correspond to the NSEln; on each box, 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 plus signs but the *y*-axis is truncated (some outliers are not shown)

(i.e., from both single-donors and multidonors) of the individual catchments and their regional median are nearly equivalent to the maximum possible performance measures and their regional median values of the ideal BASD (i.e., from single-donor; Tables 4 and 5). The regional median NSEln of BASD and BRMP for Kirchmod are 0.73 and 0.72, respectively. The regional median NSE of both BASD and BRMP for the Kirchmod, HBV, and BGM models are 0.64, 0.60, and 0.67 respectively. The regional median NSEln of both BASD and BRMP for HBV and BGM models are 0.68 and 0.70, respectively.

For the analyses based on the performance of regionalization for individual catchments, two or more regionalization methods performed equally, and the best regionalization methods varies among the catchments, model structures, and performance measures (Table 6). There are two cases for equal performance of different regionalization methods to happen, as follows: (1) when more than

J. Hydrol. Eng., 2015, 20(11): 04015028

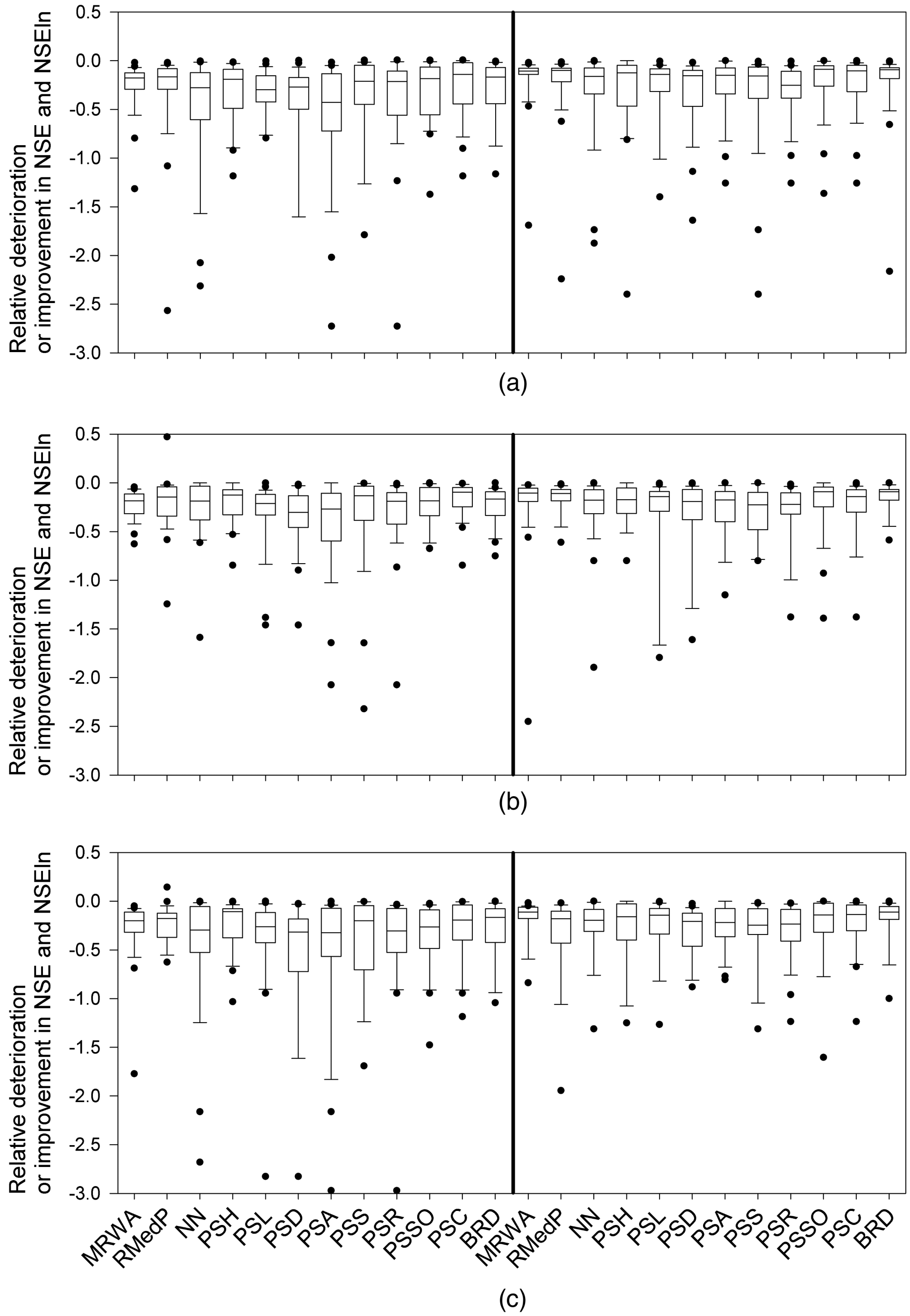

Fig. 4. Box plots of relative deterioration or improvement of performance measures from the LC due to the regionalization methods and the benchmark BRD for the following: (a) Kirchmod model; (b) HBV model; (c) BGM model; first 12 plots (MRWA to BRD) correspond to the NSE and the last 12 plots (MRWA to BRD) correspond to the NSEln; *y*-axis is truncated (some outliers are not shown)

one single-donor catchments have the top similarity rank (rank value of 1) to the recipient catchment in more than one physical attribute, and (2) when more than one best performing single-donor catchments in terms of different physical attributes have the same optimized parameter set. However, in terms of the regional median performance for the NSE, the PSC is the most frequent best regionalization method for the Kirchmod and BGM, and both PSC and PSH are the most frequent best regionalization methods for the HBV model. For the NSEln, the PSSO and PSC, respectively, are the most frequent best regionalization methods for both Kirchmod and HBV model, and BGM model. Overall, the PSC regionalization method provided the most frequent best performance compared to PSSO and PSH.

For each catchment, there are differences in the PM among the catchments and models (Tables 4 and 5). For instance, all benchmarks and regionalization methods resulted in low performance (NSE < 0.6) at six catchments [(1) Catchment 2, (2) Catchment 10, (3) Catchment 11, (4) Catchment 15, (5) Catchment 22, and (6) Catchment 25] for the Kirchmod and BGM models, and at ten catchments [(1) Catchment 4, (2) Catchment 8, (3) Catchment 10,

Table 4. Nash-Sutcliffe Efficiency Performance Measures for the Benchmarks and the Best Regionalization Method or Methods for Individual Catchments

	Kirchmod[a]				HBV[b]				BGM[c]			
Catchment number	LC	BASD	BRD	BRMP	LC	BASD	BRD	BRMP	LC	BASD	BRD	BRMP
1	0.74	0.73	0.73	0.73	0.68	0.68	0.59	0.68	0.72	0.69	**0.72**	0.69
2	0.49	0.28	−0.08	0.28	0.70	0.42	0.17	0.42	0.49	0.49	−0.02	0.49
3	0.81	0.75	0.76	0.75	0.72	0.72	0.61	0.72	0.81	0.81	0.75	0.81
4	0.71	0.70	0.66	0.66	0.56	0.52	0.51	0.52	0.70	0.70	0.63	0.65
5	0.70	0.67	0.40	0.67	0.64	0.59	0.58	0.59	0.63	0.58	0.30	0.72[d]
6	0.83	0.83	0.81	0.83	0.79	0.76	0.66	0.76	0.83	0.81	0.81	0.81
7	0.64	0.61	0.59	0.61	0.64	0.64	0.50	0.64	0.68	0.67	0.60	0.67
8	0.60	0.58	0.50	0.58	0.54	0.54	0.38	0.52	0.58	0.56	0.54	0.56
9	0.73	0.72	**0.73**	0.72	0.66	0.64	0.60	0.63	0.73	0.72	0.72	0.72
10	0.58	0.58	0.52	0.58	0.42	0.41	0.35	0.41	0.59	0.59	0.52	0.55
11	0.38	0.34	0.25	0.34	0.34	0.32	0.15	0.32	0.39	0.39	0.28	0.38
12	0.75	0.71	0.64	0.71	0.65	0.64	0.61	0.64	0.76	0.71	0.63	0.71
13	0.81	0.75	0.67	0.75	0.76	0.71	0.66	0.71	0.80	0.76	0.59	0.76
14	0.58	0.57	0.55	0.57	0.48	0.47	0.39	0.47	0.58	0.58	0.55	0.57
15	0.43	0.22	−1.08	0.22	0.74	0.59	0.29	0.59	0.29	0.02	−1.19	0.02
16	0.67	0.55	0.17	0.55	0.47	0.61	0.30	0.69[d]	0.65	0.62	0.07	0.54
17	0.78	0.68	0.40	0.68	0.86	0.81	0.79	0.80	0.79	0.76	0.47	0.70
18	0.63	0.61	0.56	0.61	0.61	0.61	0.38	0.61	0.68	0.68	0.57	0.68
19	0.81	0.75	0.42	0.75	0.85	0.76	0.76	0.76	0.72	0.69	0.36	0.69
20	0.67	0.58	0.55	0.58	0.52	0.52	0.35	0.52	0.68	0.67	0.58	0.67
21	0.72	0.71	0.71	0.71	0.62	0.62	0.59	0.62	0.72	0.70	0.68	0.69
22	0.51	0.51	0.41	0.51	0.55	0.53	**0.55**	0.53	0.54	0.54	0.37	0.50
23	0.61	0.44	0.15	0.44	0.56	0.48	0.33	0.48	0.58	0.56	0.06	0.43
24	0.74	0.74	0.54	0.74	0.58	0.51	0.49	0.51	0.75	0.62	0.59	0.60
25	0.55	0.54	0.42	0.54	0.29	0.28	0.20	0.28	0.55	0.54	0.44	0.54
26	0.72	0.71	0.66	0.71	0.63	0.63	0.53	0.63	0.72	0.71	0.64	0.71
Regional median	0.69	0.64	0.55	0.64	0.63	0.60	0.51	0.60	0.68	0.67	0.56	0.67

Note: Bold-font formatting indicates the best regional donor catchments for the models and their corresponding NSE.
[a]BASD = best arbitrary single-donor; BRD = best regional donor; BRMP = best regionization methods performance; and LC = local calibration.
[b]HBV = Hydrologiska Byråns Vattenballansavdelning.
[c]BGM = basic grid model.
[d]Multidonor RMedP as a BRMP resulted in NSE > BASD. Also, multidonor RMedP as a BRMP resulted in NSE > LC.

(4) Catchment 11, (5) Catchment 14, (6) Catchment 20, (7) Catchment 22, (8) Catchment 23, (9) Catchment 24, and (10) Catchment 25] for the HBV model. Poor NSEln values (<0.6) were observed at five catchments [(1) Catchment 8, (2) Catchment 11, (3) Catchment 14, (5) Catchment 22, and (5) Catchment 25] for the Kirchmod and the HBV models, and at eight catchments [(1) Catchment 8, (2) Catchment 11, (3) Catchment 14, (4) Catchment 15, (5) Catchment 20, (6) Catchment 22, (7) Catchment 23, and (8) Catchment 25] for the BGM model. This shows that in terms of performance for the individual catchments, the HBV and BGM models performed poorly in terms of NSE and NSEln, respectively, for a large number of catchments. The relative deterioration of performance from the local calibration due to the regionalization is high for Catchments 15, 16, and 2 for the NSE and for Catchment 14 for the NSEln.

Therefore, it is worth investigating why do models perform better or worst for some catchments. The quality of observations and potentially less representativeness of the precipitation and other climate data in terms of the density and altitude of gauging stations, and hence less accurate estimation of spatially interpolated climate forcing on the 1×1 km^2 computational grids, is one of the main factors for poor PM for these catchments. The altitudes of the climate gauging stations used in the research reported in this paper range from 15 to 885 meters above sea level (MASL). However, hypsometric curves for the high altitude catchments indicate that about only 6, 46, 48, 38, 17, and 24% of catchments lie below 885 MASL, respectively, for Catchments 2, 7, 10, 14, 17, and 24. Therefore, in addition to influencing the runoff generation, hypsography affects the representativeness of climate records, as the low-lying gauging stations may not capture the precipitation in the mountainous regions for Catchments 2, 10, 14, and 24. The effect of catchment size on the performance is also an additional factor i.e., Catchments 23, 10, 25, 16, 20, and 11 are <150 km^2 (Table 1). The correlations analysis among performance measures, catchment area, and streamflow characteristics at the end of this results section indicate decrease in performance with flow magnitudes (i.e., catchment area). Hailegeorgis and Alfredsen (2014a) and Hailegeorgis et al. (2015) for Catchment 6 (the Gaulfoss watershed) found poor performance of parameter transfer to its internal subcatchment of Catchment 14 (Lillebudal bru) especially for low flow simulation, which may be attributable to less representativeness of climate data for Catchment 14.

Selection of Physical Attributes

Selection of proper catchment attributes for the physical similarity is necessary. The selected seven physical attributes were used separately and all combined together. However, several subsamples of attributes are possible owing to multitudes of possible combinations of the seven attributes. However, single-donor and transfer of model parameters based physical similarity regionalization that is performed in the research reported in this paper resulted in PM that are nearly equivalent to the maximum possible performance that can be obtained from arbitrary transfer of optimized parameters (BASD). For instance, comparisons of the best performing physical similarity attribute(s) with the BASD in terms of regional median

J. Hydrol. Eng., 2015, 20(11): 04015028

Table 5. Nash–Sutcliffe Efficiency for Natural-Logarithm-Transformed Series (NSEln) Performance Measures for the Benchmarks and the Best Regionalization Method or Methods for Individual Catchments

	Kirchmod[a]				HBV[b]				BGM[c]			
Catchment number	LC	BASD	BRD	BRMP	LC	BASD	BRD	BRMP	LC	BASD	BRD	BRMP
1	0.84	0.83	0.79	0.79	0.82	0.82	0.74	0.81	0.78	0.78	0.76	0.77
2	0.69	0.66	0.63	0.66	0.75	0.72	0.62	0.71	0.68	0.68	0.49	0.58
3	0.87	0.87	0.80	0.87	0.84	0.83	0.80	0.83	0.85	0.85	0.78	0.85
4	0.77	0.72	0.70	0.72	0.70	0.70	0.65	0.70	0.76	0.74	0.67	0.74
5	0.73	0.70	0.67	0.70	0.63	0.61	0.58	0.60	0.62	0.62	0.54	0.62
6	0.89	0.88	0.85	0.87	0.85	0.85	0.83	0.85	0.86	0.82	0.82	0.84[d]
7	0.68	0.68	0.61	0.68	0.66	0.64	0.61	0.63	0.64	0.64	0.57	0.64
8	0.54	0.54	0.45	0.53	0.53	0.50	0.46	0.50	0.51	0.50	0.45	0.50
9	0.79	0.78	**0.79**	0.78	0.75	0.75	0.75	0.75	0.75	0.73	**0.75**	0.72
10	0.77	0.74	0.70	0.72	0.69	0.68	0.67	0.68	0.74	0.73	0.66	0.72
11	0.42	0.42	0.33	0.41	0.44	0.41	0.38	0.41	0.40	0.38	0.33	0.38
12	0.82	0.82	0.78	0.82	0.81	0.81	0.69	0.81	0.79	0.79	0.75	0.79
13	0.84	0.84	0.79	0.84	0.82	0.82	0.73	0.82	0.78	0.78	0.74	0.78
14	0.52	0.28	−0.60	−0.13	0.33	−0.13	−0.95	−0.13	0.19	0.07	−1.05	0.07
15	0.60	0.58	0.45	0.58	0.73	0.66	0.45	0.66	0.50	0.46	0.33	0.46
16	0.82	0.81	0.72	0.81	0.62	0.61	0.56	0.54	0.79	0.77	0.70	0.77
17	0.86	0.82	0.79	0.82	0.88	0.85	0.82	0.85	0.81	0.79	0.74	0.73
18	0.75	0.75	0.64	0.75	0.72	0.69	0.66	0.69	0.74	0.74	0.64	0.74
19	0.83	0.79	0.76	0.79	0.86	0.86	0.74	0.86	0.72	0.72	0.67	0.72
20	0.65	0.62	0.60	0.62	0.64	0.63	0.60	0.61	0.59	0.59	0.54	0.58
21	0.70	0.79	0.69	0.69	0.68	0.68	**0.68**	0.68	0.68	0.68	0.67	0.68
22	0.32	0.30	0.11	0.22	0.40	0.40	0.26	0.32	0.33	0.33	0.00	0.33
23	0.64	0.49	0.42	0.49	0.62	0.62	0.46	0.62	0.53	0.43	0.42	0.43
24	0.73	0.73	0.62	0.73	0.67	0.66	0.61	0.64	0.69	0.67	0.57	0.67
25	0.54	0.54	0.30	0.54	0.44	0.35	0.18	0.34	0.53	0.50	0.26	0.50
26	0.83	0.81	0.75	0.79	0.87	0.87	0.71	0.87	0.75	0.75	0.74	0.75
Regional median	0.74	0.73	0.68	0.72	0.70	0.68	0.64	0.68	0.71	0.70	0.65	0.70

Note: Bold-font formatting indicates the best regional donor catchments for the models and their corresponding Nash–Sutcliffe efficiency for natural-logarithm-transformed series (NSEln).
[a]BASD = best arbitrary single-donor; BRD = best regional donor; BRMP = best regionization methods performance; and LC = local calibration.
[b]HBV = Hydrologiska Byråns Vattenballansavdelning.
[c]BGM = basic grid model.
[d]The multidonor RMedP as a BRMP resulted in NSEln > BASD for the model.

NSE indicate 0.56 (PSH) versus 0.64 (BASD), 0.50 (PSR, PSSO, and PSC) versus 0.60 (BASD) and 0.53 (PSH) versus 0.67 (BASD) for Kirchmod, HBV, and BGM models, respectively. Similarly for the NSEln are 0.67 (PSSO) versus 0.73 (BASD), 0.63 (PSSO) versus 0.68 (BASD), and 0.59 (PSH and PSA) versus 0.68 (BASD) for Kirchmod, HBV, and BGM models, respectively. Hence, improvements that can be obtained for the regional median PM from other combinations of the physical attributes are ≤ 0.10. Such regional performance gain that may be obtained from further combinations of physical similarity attributes is not feasible compared to the large numbers of possible combinations of the physical attributes. Furthermore, the regional performance of the physical similarity based on each attribute is nearly similar for most of the cases [Figs. 3(a) and 4(c)], which would not guide dropping out of less-informative attributes. However, combination of the whole attributes (PSC) appeared to be the most frequent best-performing method for the individual catchments and exhibits less regional median relative deterioration in performance compared to the individual attributes.

Single-Donor and Multidonor Methods

Comparisons of the performance of single-donor versus multidonor regionalization methods are an important aspect of analysis. For Catchment 16–HBV–NSE, Catchment 5–BGM–NSE, and Catchment 6–HBV–NSEln, the multidonor based RMedP regionalization method provided higher PM than the maximum possible performance that can be obtained by the BASD (Tables 4 and 5). In addition, the RMedP provided higher performance than the LC for Catchment 16–HBV–NSE and Catchment 5–BGM–NSE (Table 4). These specific cases show the higher performance of the RMedP, which does not keep the correlation structure of the parameters, than the LC parameter set and the BASD. The performance of the multidonors over the single-donor methods for Catchments 6 and 16, and 5, respectively, for the HBV and BGM models may arise because of questions related to high sensitivity of runoff simulation to some parameters.

The increase in the performance of the multidonor method compared to the ideal BASD and the LC may indicate the importance of selection of proper donors for each target catchment or groups of catchments. This can be performed through identification of subregions, which are homogeneous in runoff response. In this case, selection of proper donor catchments is more important than identifying optimal number of donors as the latter may vary among the subregions and target ungauged catchments. The simplest approach is to exclude catchments with poor LC performance and hence potentially less-accurate calibrated parameters from the pool of donors. However, large numbers of catchments are required to form several homogeneous subregions and hence reduce the heterogeneity among the pooled catchments. Owing to the small number of catchments in the research reported in this paper over a relatively wide geographic extent, the research reported in this paper focused on evaluation of the performance of the total potential donors rather than attempting to identify the optimal number of donors.

Table 6. Best-Performing Regionalization Methods for Individual Catchments

	Nash–Sutcliffe efficiency for natural-logarithm-transformed series					
Catchment number	Kirchmod	HBV	BGM	Kirchmod	HBV	BGM
1	PSS/PSR/PSC	PSA/PSSO	PSS/PSR/PSC	PSS/PSR/PSC	PSH/PSS/PSR/PSC	RMedP/PSA/PSSO
2	NN	PSH/PSC	RMedP	PSH/PSC	MRWA	PSH/PSC
3	PSL/PSS	PSL	PSL	PSL	PSL/PSD	PSL
4	PSA	RMedP	PSA	PSR/PSSO/PSC	PSA	PSR/PSSO/PSC
5	PSH	PSH	RMedP	RMedP	PSA	PSA
6	PSD	PSH/PSSO	NN/PSC	PSD/PSSO	PSD	RMedP
7	PSA	PSH/PSA/PSS	PSL/PSC	PSH/PSA	NN/PSS	PSH/PSA
8	PSH/PSC	NN/PSR/PSSO	NN/PSR/PSSO	PSH/PSC	MRWA/PSH/PSC	MRWA/PSH/PSC
9	RMedP	RMedP	PSS/PSC	RMedP	PSA	PSS/PSC
10	NN/PSS/PSR/PSSO/PSC	PSS/PSR/PSSO/PSC	NN	PSS/PSR/PSSO/PSC	MRWA	PSSL/PSR/PSSO/PSC
11	PSC	PSC	PSS/PSSO	PSC	PSS/PSSO/PSC	MRWA/PSC
12	NN/PSH/PSS	NN/PSH/PSS/PSSO	RMedP/NN/PSH/PSS	PSSO	NN/PSH/PSS/PSSO	PSSO
13	MRWA	PSH/PSSO/PSC	PSR	PSH/PSSO/PSC	PSH/PSSO/PSC	PSH/PSSO/PSC
14	NN	NN/PSD	NN/PSD	PSA/PSR/PSC	PSSO	PSA/PSR/PSC
15	PSL/PSR/PSSO/PSC	PSL/PSR/PSSO/PSC	PSL/PSR/PSSO/PSC	PSR/PSSO/PSC	NN/PSL/PSA/PSR/PSSO/PSC	NN/PSL/PSR/PSSO/PSC
16	MRWA/PSL	RMedP	PSSO	PSSO	MRWA	PSSO
17	MRWA	MRWA	MRWA	MRWA	MRWA	MRWA
18	PSSO	NN/PSH/PSA	NN/PSH/PSA	NN/PSH/PSA	NN/PSH/PSA	NN/PSH/PSA
19	PSH/PSSO	PSR/PSC	PSR/PSC	PSH/PSSO	PSH/PSSO	PSH/PSSO
20	NN/PSS/PSC	NN/PSS/PSC	NN/PSS/PSC	PSL	MRWA	PSL
21	NN/PSA/PSSO	NN/PSH/PSA	PSSO	NN/PSA/PSSO	PSSO	PSH
22	PSR/PSC	MRWA/PSR/PSC	MRWA	PSA	PSSO	PSA
23	PSD	PSH	RMedP	PSL/PSSO	NN/PSSO	PSC
24	PSSO/PSC	PSA	PSA	PSS	MRWA/PSA/PSSO/PSC	PSS
25	PSS/PSC	PSS/PSC	PSC	NN/PSA/PSR	PSR	PSR
26	PSH/PSS	NN	PSH/PSS	PSH	PSH/PSS	NN

Note: BGM = basic grid model; HBV = Hydrologiska Byråns Vattenballansavdelning; MRWA = maximum regional weighted average; NN = nearest neighbour; PSA = physical similarity in catchment area; PSC = physical similarity in all attributes; PSD = physical similarity in drainage density; PSH = physical similarity in hypsometric curves; PSL = physical similarity in land use; PSR = physical similarity in bedrock geology; PSS = physical similarity in cumulative distribution functions of terrain slope; PSSO = physical similarity in soil types; RMedP = regional median of optimal parameters.

Performance of the Multimodels

The Kirchmod provided the highest regional median NSE for the LC and the regionalization methods MRWA, RMedP, NN, PSH, PSD, PSR, PSSO, and PSC, and highest regional median NSEln for the LC and all regionalization methods except the NN (Tables 4 and 5). However, the LC PM of the Kirchmod for some individual catchments is poor, for instance Catchments 2 and 15 for NSE whereas the HBV model resulted in good LC performance for these catchments (Table 4). Generally, the results indicate that the Kirchmod outperforms the two other models for both LC and regionalization methods.

Optimized Parameter Values

Information gained on calibrated parameter values is interesting since evaluation of the regionalization methods were performed based on the transferability of model parameters. The main question lies in whether the calibrated common parameters are similar among the models. Since the three models have the same snow and routing routines, it was checked whether the common parameters (TX, WS, and V) converged to similar parameter values for each catchment or to their regional statistics for the three models. Plots of optimized values of these parameters for the models for each catchment are given in Fig. 5. The results indicate that calibration of the three models converged to different parameter values for the majority of catchments. This indicates important information on implications of the correlation and nonidentifiability of parameters on parameter transferability for regionalization. However, the TX is in accordance with similar trends for the three models. The RMedP values of the TX, WS, and V are nearly the same, respectively, for HBV and BGM, Kirchmod and BGM, and Kirchmod and BGM.

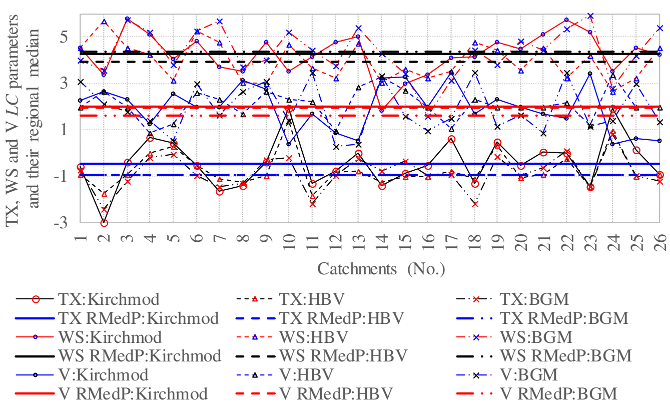

Fig. 5. Plots of common calibrated parameters for the three models

Correlation Analysis

The relationships among streamflow characteristics, catchment physical attributes, model parameters and performance measures were further analyzed to investigate whether the results of the regionalization would allow process understanding and identification of proper physical attributes. The streamflow characteristics analyzed in this context refer to the hourly streamflow at different percentages of time flow equaled or exceeded or the flow duration curves (FDC) for the calibration period.

Assessment of relationships between catchment attributes and optimized parameters is useful for spatial transfer of parameters for prediction in ungauged basins. Linear correlation coefficients (r) among some attributes, which are used for the physical similarity based regionalization, and the optimized runoff response routine parameters are given in Tables 7–9, respectively, for the Kirchmod, HBV, and BGM models. Considerable correlations (e.g., $r \geq 0.60$) are obtained only between median terrain slope and some of optimized parameters (i.e., drainage coefficients for the HBV and BGM models, S_{max} for BGM model, and EvR for Kirchmod).

Assessment of relationship between catchment attributes and the streamflow characteristics is useful for spatial transfer of streamflow characteristics for prediction in ungauged basins. The linear correlation coefficients (r) only between catchment area, among the physical attributes used in the research reported in this paper, and the streamflow characteristics were considerable (e.g., $r \geq 0.85$) as given in Table 10, which indicate important relationships for transfer of streamflow characteristics for prediction in ungauged basins in the boreal region.

Knowledge of the relationship between optimized parameters and streamflow characteristics may provide useful guidance for preliminary selection of a pool of gauged catchments for the regionalization. However, maximum positive and negative correlation values of only 0.36 and −0.36, respectively, were observed (data not shown).

Relationship between PM of both LC and regionalization methods and the streamflow characteristics can also provide useful guidance for preliminary selection of donor catchments and suitable regionalization methods. However, positive correlations with a maximum value of 0.51 were obtained for the majority of the cases (data not shown). The positive correlations i.e., increase or decrease

Table 7. Linear Correlation Coefficients between Some Physical Attributes and Calibrated Parameters for the Kirchmod Model

Physical attributes	Nash–Sutcliffe efficiency for natural logarithm transformed series					
	EvR	β_1	β_0	EvR	β_1	β_0
Hypsography, median elevation (m)[a]	−0.10	−0.23	−0.13	0.25	0.22	−0.06
Land use, forest (%)	−0.02	0.26	0.02	−0.31	−0.34	−0.20
Land use, mountains (%)	0.11	−0.26	0.04	0.43	0.29	0.04
Land use, lakes (%)	0.06	0.04	−0.44	−0.25	−0.41	−0.30
Drainage density (km/km^2)	0.28	0.38	0.44	−0.03	−0.21	−0.24
Catchment area (km^2)	−0.12	−0.26	−0.12	−0.29	0.19	0.23
Slope, median (degree)	0.24	0.01	0.35	**0.90**	0.09	−0.15
Bedrock, dioritic to granitic gneiss and migmatite (%)	0.15	−0.02	0.19	**0.63**	−0.09	−0.16
Soil, glacial tills (%)	−0.48	−0.17	−0.25	−0.32	−0.15	−0.11

Note: Correlation coefficient r values >0.6 or <-0.6 are given in bold.
[a]Meters above sea level.

J. Hydrol. Eng., 2015, 20(11): 04015028

Table 8. Linear Correlation Coefficients between Some Physical Attributes and Calibrated Parameters for the HBV Model

Physical attributes	Nash–Sutcliffe efficiency for natural logarithm transformed series							
	k_1	k_0	PERC[a]	n_u	k_1	k_0	PERC	n_u
Hypsography, median elevation (m)[b]	0.46	0.36	**0.61**	−0.14	0.32	0.01	0.35	−0.31
Land use, forest (%)	−0.35	−0.35	−0.41	0.21	−0.19	−0.09	−0.23	0.14
Land use, mountains (%)	0.49	0.42	0.58	−0.17	0.33	−0.03	0.31	−0.20
Land use, lakes (%)	−0.48	−0.18	−0.14	−0.05	−0.28	0.03	−0.20	−0.01
Drainage density (km/km^2)	−0.31	−0.36	−0.44	0.51	−0.27	−0.14	−0.27	0.22
Catchment area (km^2)	0.17	0.11	0.00	−0.08	−0.18	−0.17	−0.32	−0.03
Slope, median (degree)	**0.68**	**0.74**	0.52	−0.21	**0.65**	−0.09	0.43	−0.43
Bedrock, dioritic to granitic gneiss and migmatite (%)	0.53	0.43	0.27	−0.33	0.44	−0.08	0.32	−0.34
Soil, glacial tills (%)	0.06	−0.24	0.06	0.00	−0.01	0.15	0.08	−0.01

Note: Correlation coefficient r values >0.6 or < −0.6 are given in bold.
[a]PERC = percolation, i.e., from the upper to lower reservoir.
[b]Meters above seal level.

Table 9. Linear Correlation Coefficients between Some Physical Attributes and Calibrated Parameters for the BGM Model

Physical attributes	Nash–Sutcliffe efficiency for natural logarithm transformed series							
	$S_{\max}$	k	n	I_c	$S_{\max}$	k	n	I_c
Hypsography, median elevation (m)[a]	0.03	0.01	0.07	−0.22	0.28	0.32	−0.43	−0.28
Land use, forest (%)	0.02	−0.15	−0.01	0.35	−0.39	−0.20	0.19	0.22
Land use, mountains (%)	0.08	0.29	0.00	−0.30	0.42	0.27	−0.33	−0.26
Land use, lakes (%)	0.03	−0.13	−0.47	0.22	0.06	0.03	−0.11	0.30
Drainage density (km/km^2)	−0.20	0.09	0.08	0.04	0.05	−0.13	0.19	0.15
Catchment area (km^2)	−0.10	−0.12	−0.06	0.29	−0.26	0.09	0.18	−0.14
Slope, median (degree)	0.43	**0.60**	−0.12	−0.24	**0.70**	0.06	−0.23	0.14
Bedrock, dioritic to granitic gneiss and migmatite (%)	0.53	0.52	−0.14	−0.05	0.45	−0.05	−0.13	0.25
Soil, glacial tills (%)	−0.13	−0.41	0.27	0.29	−0.28	0.21	−0.31	−0.24

Note: Correlation coefficient r values >0.6 or < −0.6 are given in bold.
[a]Meters above seal level.

Table 10. Linear Correlation Coefficients among Streamflow and Catchment Area

Streamflow (m^3/s) at FDC (%)	Catchment area (km^2)
0	0.94
10	0.98
20	0.97
30	0.94
40	0.94
50	0.92
60	0.88
70	0.88
80	0.89
90	0.86
100	0.94

in the PM of the models for the LC and regionalization methods with the streamflow magnitudes and hence the catchment area were observed.

Discussion

Regional Performance

The results of the research reported in this paper justify that selection of proper regionalization methods are dependent on the model structure used, the selected PM and their statistical evaluation (e.g., Parajka et al. 2005; Lee et al. 2005; Oudin et al. 2008). In general, the results [Figs. 3(a) and 4(c)] for the three models indicate that the PSH and BRD performed better for the NSE, and the RMedP, MRWA, and BRD performed better for the NSEln. However, in general, the regional median relative loss of deterioration or improvement from the local calibration of Figs. 4(a–c) indicate that the PSC and PSH methods performed better for the NSE, and PSS, MRWA, and BRD methods performed better for the NSEln. There was also no consistent trend to explain the variations. However, generalization of the results to infer the best regional solution is required for application to PUBs. The results reflect the limitations of the contemporary regionalization endeavors using precipitation–runoff modeling, which require comprehensive comparative evaluations for identification of suitable regionalization methods.

The better or similar performance of the benchmark BRD compared to the more advanced regionalization methods based on both regional median values of the PM [Figs. 3(a–c)] and regional median of relative deterioration and improvement of the PM from the local calibration [Figs. 4(a–c)] indicate an opportunity for regionally homogeneous parameter set from single best regional donor regionalization solution for the boreal mid-Norway. This result complies with Haddeland et al. (2002) who found that snow-dominated catchments are less sensitive to the aggregation of model parameters than are rainfall-dominated catchments. However, the poor performance of the single-donors mainly of the NN method do not reflect this, which indicate for the research reported in this paper that identification of proper regional donor was better than identification of proper donors for each target catchment or groups of catchment.

Arsenault and Brissette (2014) obtained for Quebec (Canada) that the physical similarity approach performs better than the spatial proximity, which complies with the research reported in this paper. The findings of the research reported in this paper do not support the results of previous studies that reported better performance of the nearest neighbor than the physical similarity (e.g., Merz and Blöschl 2004; Parajka et al. 2005; Oudin et al. 2008; Zhang and Chiew 2009; Parajka et al. 2013). Low performance of the NN regionalization method in terms of the regional median [Figs. 3(a) and 4(c)] indicated the lack of smooth spatial variations of dominant hydrological processes in the region of the research reported in this paper. The main reasons for the differences among the findings of the research reported in this paper and the previous works may be attributed to the differences in the hydrological behavior of the boreal catchments. For instance, Beldring et al. (1999, 2000) noted significant contribution of the subsurface flow from the subsurface storage, which is highly influenced by the spatial variability of terrain characteristics, to the runoff hydrographs of boreal glacial tills dominated catchments. In addition, the research reported in this paper is based on simulation of hourly runoff response in which a simple translation based routing accounts for the runoff delay compared to the daily or monthly simulation in the previous studies. Moreover, the results of the physical similarity are 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 and the research reported in this paper except that the hypsometric curves rather than the mean elevation and the cumulative distribution functions of the slope are used in the research reported in this paper rather than the mean or median slope.

In addition, runoff responses in boreal region are also influenced by the spatial variability of both rainfall, and snow accumulation and melt processes. The effects of the less representativeness of precipitation data may result in pronounced effects on the high-flow simulation that is influenced by both rainfall and snowmelt events. The density of hydroclimatic gauging networks can also have considerable influences on the performance of the nearest neighbor donor catchments based regionalization (Parajka et al. 2005; Oudin et al. 2008). Therefore, the low density of hydroclimatic gauging networks in the research reported in this paper might have contributed to the lesser performance of the nearest neighbor method. Only two of the modeled catchments [(1) Catchment 3, and (2) Catchment 6] have precipitation gauging stations inside their boundary. Dense hourly measurement networks inside the catchments of the research reported in this paper would generally benefit the runoff simulation in the region of the research reported in this paper. Evaluation of the representativeness of climate stations both in terms of density and location altitude of the gauging stations need to be scrutinized in regionalization endeavors. With respect to the length of time series for calibration, Merz et al. (2009) suggested that a calibration period of 5 years of daily data captures most of the temporal hydrological variability. In the research reported in this paper, it is expected that the 2-year period hourly data was used for the calibration, due to the lack of long hourly climate data, would provide robust calibration. However, if there were no limitations of hourly climate data, calibration based on long time series would be important. Coupling of spatial proximity and physical similarity methods may provide improved performance for the less dense stream-gauging network (Samuel et al. 2011).

The results also indicated the potential effects of geographical proximity on the performance of regionalization or the interaction between the geographic proximity and physical similarity, which suggests integrated utilization of 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, and 15 are far from the rest of catchments in terms of their combined physical attributes. The majority of these catchments are among those catchments which exhibited NSE < 0.6 and/or NSEln < 0.6 (Tables 4 and 5) for both the local calibration and regionalization methods and with high relative deterioration in the PM.

Physical Similarity Attributes

The better performance of transfer of the locally calibrated parameters in the region based on the physical similarity methods generally indicated the physical control of runoff processes. The better performance of the physical similarity method would substantiate the need for data acquisition on some influential attributes (e.g., soil hydraulic properties) for further attempts of regionalization based on physical similarity in the region. The mean annual precipitation climate attribute (e.g., Parajka et al. 2005; Kim and Kaluarachchi 2008; Sawicz et al. 2011) is also a potentially relevant attribute. In addition, other climate attributes such as mean annual potential evapotranspiration (e.g., Kim and Kaluarachchi 2008), aridity index (e.g., Zhang and Chiew 2009; Oudin et al. 2008), percentage of snow in total precipitation, and mean annual temperature may also be important. However, if representative climate data were readily available for the catchments, use of physioclimatic similarity for instance by including mean annual and seasonal precipitation attributes would be important.

Single-Donors and Multidonors

The best performance of the multidonor MRWA and RMedP methods for the NSEln [Figs. 3(a-c)] indicate better performance of homogeneous parameter set for the region for low-flow than high-flow simulation. The differences in the performance between the two PM even for the hourly simulation in the research reported in this paper supports the previous studies by Lee et al. (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. The dependency of regionalization on the performance measures substantiates the need for selection of the PM depending on the modeling objectives (e.g., high-flow, low-flow, and water-balance simulation). Excluding the poorly calibrated catchments from the donors set (e.g., Oudin et al. 2008) or transferring average or median of parameter values of only the immediate upstream and downstream neighbors of each catchment (e.g., Merz and Blöschl 2004) would be expected to benefit more the multidonor-based regionalization methods (MRWA and RMedP) than the single-donor regionalization methods. Due to the small number of catchments in the research reported in this paper, the poorly performing catchments were not excluded from donors.

Performance of the Multimodels

The highest regional median performance is achieved by the Kirchmod [Figs. 3(a) and 4(c)] probably due to its so-called top–down modeling paradigm and parsimony. The principle of model parsimony (Jakeman and Hornberger 1993) endorses parsimonious models as long as they perform similar to the more complex models. Complexity alone cannot guarantee good and reliable performance (Perrin et al. 2001). Hailegeorgis and Alfredsen (2014a) obtained similar performance of different simple to complex HBV model configurations based on calibration, and so-called split-sample and proxy basin (Klemeś 1986) validation tests. Hrachowitz et al. (2014) illustrated that expert-knowledge of model constraints i.e., process and parameter, and model improvement through diagnostic evaluation based on several runoff signatures can lead to more consistent models despite the number of parameters. The poor performance of Kirchmod model for some catchments suggest the use of multimodels than a single model, which results in ensemble predictions to represent model parameter uncertainty (e.g., McIntyre et al. 2005).

Parameter Identifiability

The implications of the differences in the calibrated values of the parameters that are common for the three models (Fig. 5) for the regionalization endeavors is that there is interactions among the parameters within the runoff response, and snow and runoff routing routines (Wagener and Wheater 2006). Another aspect is the possible impact of the number of parameters on the regionalization performance. For instance, the HBV model is more complex in terms of numbers of parameters and storage states but provided the poorest regional performance. Results from Uhlenbrook et al. (1999), Merz and Blöschl (2004) and Hailegeorgis and Alfredsen (2014a) indicated that some of the HBV parameters are difficult to identify due to their lesser sensitivity or their uncertainty. To improve parameter identifiability, it may be necessary to fix the less-sensitive parameters at predefined values (e.g., their median values or the RMedP) and only regionalize the optimized runoff response routines parameters.

Correlation Analysis

The large positive linear correlation (>0.6) between the median terrain slope and EvR, k_1, and k_0 (Tables 7 and 8) might be reflected in the low regional median relative deterioration of the NSEln by the PSS method for both the Kirchmod and the HBV models. However, the large positive correlation between the bedrock type and EvR, hypsography and PERC, and median slope, and both k and $S_{\max}$ did not result in high performance of similarity in the respective attributes for the respective models. This may be related to parameter nonidentifiability problems and the need for relating catchment attributes to parameter set rather than to individual parameter values as suggested by Bárdossy (2007).

The large positive correlation between catchment area and streamflow corresponding to different flow durations (Table 10) indicate an opportunity for prediction of flow duration curves and time series in ungauged basins using statistical approach (e.g., Yadav et al. 2007; Hailegeorgis and Alfredsen 2014b). However, PM based on regionalization of precipitation–runoff models using the physical similarity method did not outperform the other regionalization methods except the fact that there is no outlier catchment for the PSA method for NSEln for the BGM model [Fig. 3(c)]. This may be attributed to the fact that no defined correlation between streamflow characteristics, which is correlated to catchment scale, and model parameters was found probably due to parameter nonidentifiability problems. Merz et al. (2009) illustrated the lesser scale dependence of the HBV model parameters. The low PM for small size catchments, which can be inferred from the positive correlations between streamflow characteristics and the PM, indicate the scale dependance of the PM in the region of the research reported in this paper; this may be attributed to exaggerated effects of the low-resolution climate forcing field for the smaller catchments. The result is in agreement with Merz et al. (2009).

Conclusions

Four regionalization methods [which include (1) regional calibration (i.e., MRWA), (2) regional median parameters, (3) nearst neighbor, and (4) physical similarity] were evaluated to transfer locally calibrated parameters of three distributed (1×1 km^2 grid) hourly precipitation–runoff models in 26 catchments in mid-Norway. The physical similarity method incorporates eight different cases based on seven attributes. The performances were evaluated using the NSE and NSEln PM and their statistical evaluation approaches, box plots and regional median of PM values, and relative deterioration or improvement of PM from the local calibration due to the regionalization.

For the region of the research reported in this paper and available set of hydroclimatological data, identification of the regionalization methods depends on the model structures, and PM and their statistical evaluation approach. In general, the single-donor physical similarity (i.e., PSH and PSC) methods and the simple benchmark (i.e., BRD) performed better for the NSE based on boxplots and associated regional median values of both the NSE values and relative deterioration or improvement of the NSE from the local calibration, and regionalization performance for the individual catchments. The single-donor physical similarity, multidonors (i.e., RMedP and MRWA) methods, and BRD performed better for the NSEln based on the same evaluation criteria. Similar performance of the benchmark BRD, which is based on transfer of regionally homogeneous parameter sets, for both NSE and NSEln compared to the more advanced regionalization methods signify the merits of the method as a simple regionalization solution for the region and the need for high-density climate gauging networks for more reliable calibration of parameters for the potential donors. Comparisons of the multimodels indicated that the Kirchmod outperformed the other models, which indicates the relative merits of the so-called top–down and parsimonious modeling. Nearly equivalent performance of the single-and multidonor methods, the simple benchmark and more advanced regionalization methods, and the effects of model structure and modeling paradigm indicate that comprehensive identification of suitable regionalization methods is necessary for more reliable prediction in ungauged basins. The research reported in this paper also indicated the importance of considering the objectives of prediction (e.g., high flow or low flow) for selection of the PM and their evaluation metrics.

The research reported in this paper was the first attempt for regionalization of hourly runoff simulation in the region. Further regionalization study at hourly temporal resolution for the region of the research reported in this paper should focus on representative (i.e., high-density gauging networks and longer records of climate and streamflow input) and physioclimatic attributes including precipitation and soil hydraulic attributes. The findings from the research reported in this paper provide important information relevant to distributed continuous hourly runoff simulation in ungauged basins. Regionalization studies for distributed hourly runoff simulation for catchments in other climate regimes and landscape features would also provide further insights.

Acknowledgments

The Center for Environmental Design of Renewable Energy (CEDREN; http://www.cedren.no) funded the research reported in this paper). The climate data were obtained from the Norwegian

Meteorological Institute, Nord Trøndelag Elektrisitetsverk (NTE), and Bioforsk. The streamflow, hypsography, and land-use data were found from the Norwegian Water Resources and Energy Directorate. The loose material (soil) and bedrock geology data were obtained from the Norwegian Geological Survey. Stream networks of a 1:50,000 map from the Norwegian Mapping Authority were used. The anonymous reviewers are also thanked for their constructive comments, which helped to improve the paper.

References

Arsenault, R., and Brissette, F. P. (2014). "Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches." *Water Resour. Res.*, 50(7), 6135–6153.

Bárdossy, A. (2007). "Calibration of hydrological model parameters for ungauged catchments." *Hydrol. Earth Syst. Sci.*, 11(2), 703–710.

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

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. Sci.*, 7(3), 304–316.

Beldring, S., Gottschalk, L., Rodhe, A., and Tallaksen, L. M. (2000). "Kinematic wave approximations to hillslope hydrological processes in tills." *Hydrol. Process.*, 14(4), 727–745.

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.

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. Sci.*, 2(2–3), 265–281.

Bergström, S. (1976). "Development and application of a conceptual runoff model for Scandinavian catchments." *Rep. RHO 7*, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden.

Beven, K. (2006). "A manifesto for the equifinality thesis." *J. Hydrol.*, 320(1–2), 18–36.

Blöschl, G., and Sivapalan, M. (1995). "Scale issues in hydrological modeling: A review." *Hydrol. Process.*, 9(3–4), 251–290.

Box, G. E. P., and Cox, D. R. (1964). "An analysis of transformations." *J. Roy. Stat. Soc.*, 26(2), 211–252.

Bulygina, N., McIntyre, N., and Wheater, H. (2009). "Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis." *Hydrol. Earth Syst. Sci.*, 13(6), 893–904.

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." *Hydrol. Process.*, 28(4), 2033–2045.

Croke, B., and McIntyre, N. (2013). "Data-based perceptions on predictions in ungauged basins." *Hydrol. Res.*, 44(3), 399–400.

Donnelly, C., et al. (2009). "An evaluation of multi-basin hydrological modelling for predictions in ungauged basins." *New approaches to hydrological prediction in data-sparse regions*, K. Yilmaz, et al., eds., IAHS, Wallingford, U.K., 112–120.

Dunne, T., and 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., Braud, I., Gottschalk, L., and Leblois, E. (2006). "Multi-objective regional modelling." *J. Hydrol.*, 327(3–4), 339–351.

Engeland, K., and Gottschalk, L. (2002). "Bayesian estimation of parameter in a regional hydrologic model." *Hydrol. Earth Syst. Sci.*, 6(5), 883–898.

Fernandez, W., Vogel, R. M., and Sankarasubramanian, A. (2000). "Regional calibration of a watershed model." *Hydrol. Sci. J.*, 45(5), 689–707.

Gottschalk, L., Leblois, E., and Skøien, J. O. (2011). "Distance measures for hydrological data having a support." *J. Hydrol.*, 402(3–4), 415–421.

Götzinger, J., and Bárdossy, A. (2007). "Comparison of four regionalization methods for a distributed hydrological model." *J. Hydrol.*, 333(2–4), 374–384.

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.

Gupta, H. V., Wagener, T., and Liu, Y. Q. (2008). "Reconciling theory with observations: Elements of a diagnostic approach to model evaluation." *Hydrol. Process.*, 22(18), 3802–3813.

Haddeland, I. B., Matheussen, V., and Lettenmaier, D. P. (2002). "Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model." *Water Resour. Res.*, 38(7), 1124–1134.

Hailegeorgis, T., and Alfredsen, K. (2014a). "Comparative evaluation of performance of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments." *Hydrol. Res.*, in press.

Hailegeorgis, T., and Alfredsen, K. (2014b). "Regional statistical and precipitation-runoff modelling for ecological applications: Prediction of hourly streamflow in regulated rivers and ungauged basins." *Proc., 10th Int. Symp. on Ecohydraulics (ISE)*, SINTEF, Trondheim, Norway.

Hailegeorgis, T., Alfredsen, K., Abdella, Y., and Kolberg, S. (2015). "Evaluation of different parameterizations of the spatial heterogeneity of subsurface storage capacity for hourly runoff simulation in boreal mountainous watershed." *J. Hydrol.*, 522, 522–533.

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 regionalization for continuous streamflow simulation." *Hydrol. Earth Syst. Sci.*, 15(11), 3539–3553.

Horton, R. E. (1933). "The role of infiltration in the hydrologic cycle." *Trans. Am. Geophys. Union*, 14(1), 446–460.

Hrachowitz, M., et al. (2013). "A decade of predictions in ungauged basins (PUB)—A review." *Hydrol. Sci. J.*, 58(6), 1198–1255.

Hrachowitz, M., et al. (2014). "Process consistency in models: The importance of system signatures, expert knowledge, and process complexity." *Water Resour. Res.*, 50(9), 7445–7469.

Jakeman, A. J., and Hornberger, G. M. (1993). "How much complexity is warranted in a rainfall-runoff model?" *Water Resour. Res.*, 29(8), 2637–2649.

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(2), W02429.

Klemeś, V. (1986). "Operational testing of hydrological simulation models." *Hydrol. Sci. J.*, 31(1), 13–24.

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.

Kolberg, S. A., and Bruland, O. (2012). "ENKI—An open source environmental modelling platform." *Geophysical Research Abstracts 14, European Geophysical Union (EGU) General Assembly, EGU2012-13630*, Copernicus, Göttingen, Germany.

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

Lamb, R., and Kay, A. L. (2004). "Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain." *Water Resour. Res.*, 40(7), in press.

Lee, H., McIntyre, N., Wheater, H., and Young, A. (2005). "Selection of conceptual models for regionalization of the rainfall-runoff relationship." *J. Hydrol.*, 312(1–4), 125–147.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S. (1997). "Development and test of the distributed HBV-96 hydrological model." *J. Hydrol.*, 201(1–4), 272–288.

Littlewood, I. G., and Croke, B. F. W. (2013). "Effects of data time-step on the accuracy of calibrated rainfall–streamflow model parameters: Practical aspects of uncertainty reduction." *Hydrol. Res.*, 44(3), 430–440.

Madsen, H. (2003). "Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives." *Adv. Water Res.*, 26(2), 205–216.

McIntyre, N., and Al-Qurashi, A. (2009). "Performance of ten rainfall-runoff models applied to an arid catchment in Oman." *Environ. Modell. Softw.*, 24(6), 726–738.

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). "Ensemble predictions of runoff in ungauged catchments." *Water Resour. Res.*, 41(12), W12434.

Merz, R., and Blöschl, G. (2004). "Regionalization of catchment model parameters." *J. Hydrol.*, 287(1–4), 95–123.

Merz, R., Parajka, J., and Blöschl, G. (2009). "Scale effects in conceptual hydrological modeling." *Water Resour. Res.*, 45(9), W09405.

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

Muleta, M. K. (2012). "Model performance sensitivity to objective function during automated calibrations." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000497, 756–767.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models: I. A discussion of principles." *J. Hydrol.*, 10(3), 282–290.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., and Michel, C. (2006). "Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations." *Water Resour. Res.*, 42(7), W07410.

Oudin, L., Andréassian, 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(3), W03413.

Oudin, L., Kay, A., Andréassian, V., and Perrin, C. (2010). "Are seemingly physically similar catchments truly hydrologically similar?" *Water Resour. Res.*, 46(11), W11558.

Parajka, J., Blöschl, G., and Merz, R. (2007). "Regional calibration of catchment models: Potential for ungauged catchments." *Water Resour. Res.*, 43(6), W06406.

Parajka, J., Merz, R., and Blöschl, G. (2005). "A comparison of regionalization methods for catchment model parameters." *Hydrol. Earth Syst. Sci.*, 9(3), 157–171.

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(5), 1783–1795.

Patil, S., and Stieglitz, M. (2011). "Hydrologic similarity among catchments under variable flow conditions." *Hydrol. Earth Syst. Sci.*, 15(3), 989–997.

Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., and Wheater, H. S. (2011). "Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications." *Global NEST J.*, 13(3), 193–214.

Pechlivanidis, I. G., McIntyre, N. R., and Wheater, H. S. (2010). "Calibration of the semi-distributed PDM rainfall-runoff model in the upper Lee Catchment, UK." *J. Hydrol.*, 386(1–4), 198–209.

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

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(2), 81–92.

Razavi, T., and Coulibaly, P. (2013). "Streamflow prediction in ungauged basins: Review of regionalization methods." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0000690, 958–975.

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.*, 10.1061/(ASCE)HE.1943-5584.0000338, 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(9), 2895–2911.

Seibert, J. (1999). "Regionalisation of parameters for a conceptual rainfall-runoff model." *Agr. Forest Meteorol.*, 98–99, 279–293.

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(17), 3697–3708.

Sivapalan, M, et al. (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(10), W10502.

Uhlenbrook, S., Seibert, J., Leibundgut, C., and Rodhe, A. (1999). "Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure." *Hydrol. Sci. J.*, 44(5), 779–797.

Vaze, J., Zhang, Y., Chiew, F. H. S., Wang, B., and Teng, J. (2013). "Regional calibration against multiple data sources to predict streamflow." *Proc., H01, The Int. Association of Hydrological Sciences (IAHS)–The Int. Association for the Physical Sciences of the Oceans (IAPSO)–The Int. Association of Seismology and Physics of the Earth's Interior (IASPEI) Assembly*, IAHS, Wallingford, U.K., 165–170.

Viglione, A., et al. (2013). "Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria." *Hydrol. Earth Syst. Sci. Discuss.*, 10(1), 449–485.

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), 1312–1325.

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." *J. Nonlinear Sci. Numer. Simul.*, 10(3), 271–288.

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." *Environ. Modell. Softw.*, 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." *Adv. Water Resour.*, 30(8), 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(7), W07412.