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]

[1]Researcher, Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim (corresponding author). e-mail: TEKHI09@GMAIL.COM, Tel.: [+47] 73592411 and Fax: [+47] 735 91298.

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

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

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

Abstract

Regionalization for prediction in ungauged basins at hourly resolution is important for water resources management (e.g. floods and hydropeaking). In this paper, calibration of 26 catchments (39-3090 km^2) in mid-Norway was performed using hourly records and three spatially distributed (1x1 km^2) precipitation-runoff models: a first-order nonlinear system model (hereafter Kirchmod), the HBV model and the Basic-Grid-Model (BGM). Four regionalization methods for each model namely parameter set yielding maximum regional weighted average (*MRWA*) performance measures (PM), regional median of optimal parameters (*RMedP*), nearest neighbor (*NN*) and physical similarity (*PS*) were evaluated and compared with three benchmarks. Parameter transfer from best regional donor (*BRD*) and from an ideal best arbitrary single-donor (*BASD*), and local calibration (*LC*) were as benchmarks. The physical similarity attributes include hypsometric curves (*PSH*), land use (*PSL*), drainage density (*PSD*), catchment area (*PSA*), terrain slope (*PSS*), bedrock geology (*PSR*), soil type (*PSSO*) and combination of all (*PSC*).

Comprehensive evaluation of single-and multi-donors, simple benchmarks and more advanced regionalization methods using multi-models, two PM 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 *PSH, PSC* and *BRD* methods performed better for the Nash-Sutcliffe efficiency (*NSE)* based on boxplots and regional median values of both the *NSE* and relative deterioration or improvement of the *NSE* from the local calibration due to the regionalization. The methods also performed better for the individual catchments. The *PSS*, *RMedP*, *MRWA* and *BRD* methods performed better for the log-transformed streamflow

(*NSEln)* 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 *BRD* for both *NSE* and *NSEln* indicate the potential of the simple regionalization approach for predicting runoff response in the region.

Introduction

Continuous time precipitation-runoff modelling has been used to represent the hydrological processes and understand the basin response. The modelling 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 non-identifiability of some model. Moreover, there are further challenges for prediction in ungauged basins (PUB; 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 modelling 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 modelling and parameter regionalization are important for the PUB. 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.

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 multi-sites 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 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 multi-objective calibration of a physically based distributed model using multi-sites 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. Donnelly et al. (2009) illustrated comparable performance of a simultaneously calibrated spatially homogeneous parameter sets for a multi-basin 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 sub-region 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 multi-donor regionalization method based on transferring of averages of optimized parameters for the catchments in the region or “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 “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 PUB 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 ‘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 (see 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 multi-donor 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 (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 multi-donor 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 study region.

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 and 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. 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-semi-distributed 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 (see Wagener and Wheater, 2006; Oudin et al., 2008; Oudin et al., 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 PUB.

Previous attempts for continuous streamflow simulation by rainfall-runoff models for PUB were mainly based on conceptual rainfall-runoff modelling, 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 'top-down' rainfall-runoff modelling 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 PUB.

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 UK 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 modelling 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 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 (see a recent review by Hrachowitz et al., 2013). No universally best-performing regionalization method, model structure or evaluation criteria could be suggested due to the peculiarities of catchments in different climate regimes and landscapes (see a recent synthesis by Parajka et al., 2013).

Scientific questions and objectives of the paper

The present study will attempt to answer the following questions: (a) are the performance of the regionalization methods consistent among model structures, performance measures (PM) and statistical approaches used for evaluation of the PM? and (b) which regionalization method performs best for the specific region of study?. The present study is the first attempt for distributed hourly runoff simulation in the study region, 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 follows. The next two sections provide brief explanation of the study region and data, and models and methods used. Then, the results of the regionalization study are presented followed by discussions. The last section provides conclusions.

The Study 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 3090 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 study are precipitation (P) in mm, temperature (T) in ^{o}C, wind speed (W_s) in m/s, relative humidity (H_R) in percentage and global radiation (R_G) in W/m^2. The climate data from point measurements are spatially interpolated on 1x1 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 Figure 1.

Precipitation occurs mainly in the form of snowfall during winter (*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 namely the hypsometric curves, land use, drainage density, catchment area, cumulative distribution functions of terrain slope, bedrock geology and soil types. 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 http://www.nve.no and the soil and bedrock geology data was obtained from the Norwegian Geological Survey (NGU) (http://www.ngu.no). Stream networks from the 1:50000 maps produced by the Norwegian Mapping Authority (http://www.statkart.no) were used.

Models and Methods

Three distributed (1x1 km^2 grid) precipitation-runoff models namely the 'top-down' water balance model (hereafter Kirchmod; Kirchner, 2009), the Hydrologiska Byråns Vattenballansavdelning model (hereafter HBV; Lindström et al., 1997) and the Basic-Grid-Model (hereafter 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 Figure 2a-c. Brief descriptions of the models are given here.

Kirchner's runoff response routine

Kirchner (2009) inferred model equations and parameters from analysis of streamflow recession (i.e. the 'top-down' modelling paradigm). The method is based on a catchment storage-discharge relationship. The main assumption in the Kirchner's method is that the streamflow depends solely on the amount of water stored in the catchment. The water balance response routine is given as:

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

The $dQ/dS = g(Q)$ or discharge sensitivity function (Kirchner, 2009), where S is catchment averaged storage and Q is discharge. The runoff simulation was based on the following integral equation and the following regression relationship between the $g(Q)$ and Q was considered:

$$S\left(Q\right) = \int \frac{1}{g\left(Q\right)} dQ \; ; \; \ln\left(g\left(Q\right)\right) \approx \beta_0 + \beta_1 \ln\left(Q\right) , \qquad (2)$$

where the actual evapotranspiration (AET), infiltration in to the soil (I) = rainfall + snow melt (SM), Q and S are all in mm, t is a time variable, β_0 and β_1 are calibrated model parameters. A Runge Kutta 4th 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:

$$AET = PET\left\{1 - \exp\left(-\frac{Q}{EvR}\right)\right\}\left(1 - SCA\right) , \qquad (3)$$

where the *EvR* is a parameter which denotes a discharge at which *AET* equals 0.95**PET* and *SCA* is the 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 1h (mm/hr) is used for calibration against a hourly averaged observed discharge. The observed hourly averaged discharge one timestep before the start of model run was used as an initial discharge for the numerical solver.

The HBV model

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

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

where n_u is exponent of non-linearity for the upper zone, *UZ* (mm) and *LZ* (mm) respectively denote the upper and lower zones, k_1 and k_0 are recession coefficient parameters. Percolation from the upper to the lower reservoir is controlled by the parameter (*PERC*). The soil moisture accounting routine is based on Bergström (1976). A non-linear function partitions the infiltration into change in soil moisture storage (Δ_{SM}) and recharge to the upper zone (R_{uz}). If *SM* > (*LP*FC*), *AET* = *PET* but if *SM* < (*LP*FC*), *AET* = (*PET*SM*)/(*LP*FC*), where *SM* is soil moisture, *FC* is the field capacity and *LP* is 'the limit for potential evaporation'.

Basic Grid Model (BGM)

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

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

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

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

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

Snow Accumulation and Melting routine

The influences of snow processes are dominant in the study area 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 or SDC (Kolberg and Gottschalk, 2006), which was implemented in the ENKI hydrological modelling platform (Kolberg and Bruland, 2012). The calibrated parameters in this routine are snow-rain threshold temperature parameter (*TX*) and snowmelt sensitivity to wind speed or windscale (*WS*). The same snow routine were used with all three runoff response routines.

Potential Evapotranspiration Routine

The PriestleyTaylor method (Priestley and Taylor, 1972) was used for the calculation of potential evapotranspiration, PET (mm/h) for all routines:

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

where α is the Priestley Taylor constant, Δ is the slope of saturation vapor pressure curve at air temperature at 2m (kPa/^{o}C), γ is the psychrometric constant (0.066 kPa/^{o}C), R_n (W/m^2) is net radiation = net shortwave radiation (SR_n) + net longwave radiation (LR_n), L_v (kJ/m^3) is volumetric latent heat of vaporization and Δt (s) is the simulation time step in seconds. 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). Following Teuling et al. (2010), α = 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-hr isochrones was implemented for all models. The hillslope runoff response of each 1x1 km^2 grid cell was translated to the catchment outlet based on travel time lags. Routed simulated streamflow at the outlet was computed from the sum of contributions from each grid cell:

$$Qsim_t = \sum_{i=1}^{N} qsim^i_{t-T_i} \quad , T_{\mathrm{i}} = \frac{L_i}{V}, \tag{9}$$

where t and i represent time and grid cells respectively, N is the number of grid cells in the catchment, Q_{sim} [LT^{-3}] is the streamflow at the outlet, q_{sim} [LT^{-3}] is the runoff generated at each grid cell, T_i [T] is the flow travel time lag to the outlet for each grid, L_i [L] is the flow travel path length computed from 25m Digital Elevation Model (DEM) and V [LT^{-1}] is 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 modelling platform (Kolberg and Bruland, 2012). A regional calibration by the DREAM algorithm can be regarded as an 'importance sampling' strategy for each catchment, where sampling is according to an ‘importance surface’ reflecting where the optimum is likely to be. DREAM seeks and converges to the posterior distribution and DREAM’s regional posterior distribution is an importance surface for each catchment. The regional calibration approach is an acceptable calibration strategy for each single catchment without considerable loss in performance from the DREAM at-site calibration, which utilizes only streamflow data from each individual catchment.

The log-likelihood (*L-L*) objective function used for the regional calibration is:

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

where δ denotes model parameter, σ_i^2 and n_i respectively are error variance and the length of non-missing records of streamflow for catchment i, $N_C = 26$ is the total numbers of modelled catchments in the region, $Q_{sim}^{(\theta)}$ and $Q_{obs}^{(\theta)}$ respectively are Box-Cox (Box and Cox, 1964) transformed observed and simulated streamflow time series to approximate normal distributed and homoscedastic series, θ is the Box-Cox transformation parameter and f represents a 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, i.e. Nash-Sutcliffe efficiency or *NSE* (Nash and Sutcliffe, 1970) [$-\infty$, 1.0] and its equivalent for log-transformed series (*NSEln*) [$-\infty$, 1.0]:

$$NSE_i = 1 - \frac{\sum_{t=1}^{n_i} \left(Q_{obs_{t,i}} - Q_{sim_{t,i}} \right)^2}{\sum_{t=1}^{n_i} \left(Q_{obs_{t,i}} - \bar{Q}_{obs,i} \right)^2} \tag{11}$$

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 behaviour of ln(x) as x approaches to zero. In our region and catchment sizes, zero discharge is not an issue even in mid-winter. It is not a sign of catchment behaviour that the model should be required to mimic. However, rare zero values of streamflow in the present study 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 our objective 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 (*LC*). 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 we investigate 7 parameter transfer experiments in which 3 are used as benchmark, i.e. reference model performance, while 4 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. The authors also tested each donor parameter set on all catchments 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 '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. (12) 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:

$$NSE_{RWA} = \frac{1}{N_C} \sum_{i=1}^{N_C} \left(\frac{n_i}{N_{TS}} \right) NSE_i \ , \tag{12}$$

where N_{TS} is the length of timestamp for the calibration period (including the missing records) and *RWA* stands for regional weighted averaged, where the weights are the term in the parenthesis assigned for each catchment based on the length of their non-missing streamflow records during the calibration period. Note that the regional calibration regionalization method used in the present study 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 (*MRWA*) PM obtained from the total modelled catchments are used in the present study.

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:

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

where j is subscript for the free parameters (j = 1 to N_p, where N_p is the total number of calibrated parameters). This method allows pooling of parameters for each PM from multiple donor catchments (i.e. multi-donor median of parameter set) and then transferring homogeneous parameter sets for the whole region for prediction. The only difference between this method and the "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 (NN)

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 co-ordinates spaces of the streamflow gauging stations was used to identify the nearest neighbor for the catchments. This distance was 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 'top-down' modelling paradigm of Kirchmod was based on the use of observed streamflow at the gauging stations.

Physical similarity (PS)

The method assumes that similarity of catchments in physical attributes could explain their homogeneity in runoff response. In the present study, 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 present study, 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 in Halldin et al. (1999) and Beldring et al. (2003). Eight different cases of physical similarity were evaluated: similarity in hypsometric curves (*PSH*), land use (*PSL*), drainage density (*PSD*), catchment area (*PSA*), cumulative distribution functions of terrain slope (*PSS*), bedrock geology (*PSR*), soil types (*PSSO*) and 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 ($Dist_{j,h}$) after the [0,1] normalization of the classes in each attribute for the sake of simplicity or scaling of the values:

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

$$Nx_{i,k}^{j} = \frac{x_k^j - \min(x_i)}{\max(x_i) - \min(x_i)}, \quad Nx_{i,k}^{h} = \frac{x_k^h - \min(x_i)}{\max(x_i) - \min(x_i)}$$

where *i and k* respectively are the indexes for the attribute *x* and classes, *j* and *h* respectively are indexes for the two catchments to be compared, *a* is the total numbers of the attributes considered, *c* is the total numbers of classes in the attribute and *N* stands for normalized. The min and max are the minimum and the maximum values 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 landuse attribute in Table 2, there are six classes i.e. farmland, forest, mountains, glacier, marshs (bog) and 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*. Rank no. 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 sub-regions for parameter transfer within sub-regions. The small number of unregulated gauged catchments in the region (i.e. only 26) compared to the wide extent of the study region 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 Figure 3a-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 are slightly higher for *RMedP* for *NSE,* and for both *RMedP* and *BRD* for *NSEln* compared to the other methods. The PM of HBV are higher for *PSH* and *BRD* for *NSE* and *NSEln* respectively compared to the other methods. Similarly, the PM of BGM are higher for the multi-donor 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, the authors now aim 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 Figure 4a-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 present study 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 multi-donors for each catchment are given in Table 4-5 for the *NSE* and *NSEln* respectively. The best regional donor (*BRD*) 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* (i.e. from both single-and multi-donors) 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) (see Table 4-5). 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 (see Table 6). There are two cases for equal performance of different regionalization methods to happen. The first case is when more than one single-donor catchments have the top similarity rank (rank no. 1) to the recipient catchment in more than one physical attribute. The second case is 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 (see Table 4-5). For instance, all benchmarks and regionalization methods resulted in low performance (*NSE* < 0.6) at six catchments (2, 10, 11, 15, 22 and 25) for the Kirchmod and BGM models, and at ten catchments (4, 8, 10, 11, 14, 20, 22, 23, 24 and 25) for the HBV model. Poor *NSEln* values (< 0.6) were observed at five catchments (8, 11, 14, 22 and 25) for the Kirchmod and the HBV models, and at eight catchments (8, 11, 14, 15, 20, 22, 23 and 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 1x1 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 present study range from 15 to 885 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, 11 are < 150 km^2 (see 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 (2014b, article in press) and Hailegeorgis et al. (2015) for catchment 6 (the Gaulfoss watershed) found poor performance of parameter transfer to its internal sub-catchment 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 sub-samples 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 present study 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 *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 are nearly similar for most of the cases (see Figure 3a-4c), 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-and multi-donor methods

Comparisons of the performance of single-donor versus multi-donor regionalization methods are an important aspect of analysis. For catchment 16-HBV-*NSE*, catchment 5-BGM-*NSE* and catchment 6-HBV-*NSEln*, the multi-donor based *RMedP* regionalization method provided higher PM than the maximum possible performance that can be obtained by the *BASD* (see Table 4-5). In addition, the *RMedP* provided higher performance than the *LC* for catchment 16-HBV-*NSE* and catchment 5-BGM-*NSE* (see 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 multi-donors over the single-donor methods for catchments 6 and 16, and 5 respectively for the HBV and BGM models may arise questions related to high sensitivity of runoff simulation to some parameters.

The increase in the performance of the multi-donor method compared to the ideal *BASD* and the *LC* may indicate the importance of selection of proper donors for each targer catchment or groups of catchments. This can be performed through identification of sub-regions, 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 sub-regions 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 sub-regions and hence reduce the heterogeneity among the pooled catchments. Owing to the small number of catchments in the present study over a relatively wide geographic extent, the present study focused on evaluation of the performance of the total potential donors rather than attempting to identify the optimal number of donors.

Performance of the multi-models

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* (Table 4-5). However, the *LC* PM of the Kirchmod for some individual catchments are 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, we 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 Figure 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 non-identifiability of parameters on parameter transferability for regionalization. However, the *TX* follows 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.

Correlation analysis

We further analyzed the relationships among streamflow characteristics, catchment physical attributes, model parameters and performance measures to investigate whether the results of the regionalization would allow process understanding and identification of proper physical attributes. The streamflow characteristics analyzed in here 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 Table 7, Table 8 and Table 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 present study, and the streamflow characteristics were found to be 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 on 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 (Table not reported here).

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 (Table not reported here). The positive correlations i.e. increase or decrease 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 present study 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 (Figure 3a-4c) 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 Figure 4a-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 PUB. The results reflect the limitations of the contemporary regionalization endeavors using precipitation-runoff modelling, 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 (Figure 3a-c) and regional median of relative deterioration and improvement of the PM from the local calibration (Figures 4a-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 this particular study that identification of proper regional donor found to be 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 this study. The findings of the present study 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 (Figure 3a-4c) indicated the lack of smooth spatial variations of dominant hydrological processes in the study region. The main reasons for the differences among the findings of the present study and the previous works may be attributed to the differences in the hydrological behavior of the boreal 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 present study 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 present study except that the hypsometric curves rather than the mean elevation and the cumulative distribution functions (CDF) of the slope are used in the present study rather than the mean or median slope.

In addition, runoff responses in boreal region is 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 hydro-climatic gauging networks can also have considerable influences on the performance of the nearest neighbor (*NN*) donor catchments based regionalization (see Parajka et al., 2005; Oudin et al., 2008). Therefore, the low density of hydro-climatic gauging networks in the present study might have contributed to the less performance of the nearest neighbor method. Only two of the modelled catchments (3 and 6) have precipitation gauging stations inside their boundary. Dense hourly measurement networks inside the study catchments would generally benefit the runoff simulation in the study region. 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 daily data captures most of the temporal hydrological variability. In the present study, the authors expect that the two-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 (see 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, 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 (Table 4-5) 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 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 physio-climatic similarity for instance by including mean annual and seasonal precipitation attributes would be important.

Single-and multi-donors

The best performance of the multi-donor *MRWA* and *RMedP* methods for the *NSEln* (Figure 3a-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 present study 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 modelling 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 multi-donor based regionalization methods (*MRWA* and *RMedP*) than the single-donor regionalization methods. Note that due to the small number of catchments in the present study, the poorly performing catchments were not excluded from donors.

Performance of the multi-models

The highest regional median performance is achieved by the Kirchmod model (Figure 3a-4c) probably due to its 'top-down' modelling 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 (2014b) obtained similar performance of different simple to complex HBV model configurations based on calibration, and '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 multi-models 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 (Figure 5) for the regionalization endeavors is that there is interactions among the parameters within the runoff response, and snow and runoff routing routines (see also 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 (2014b) 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 pre-defined 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 (Table 7 and Table 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 non-identifiability 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 2014a). 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 (Figure 3c). 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 non-identifiability problems. Merz et al. (2009) illustrated the less 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 study region; 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 the regional calibration (*MRWA*), regional median parameters (*RMedP*), nearst neighbor (*NN*) and physical similarity (*PS*) were evaluated to transfer locally calibrated parameters of three distributed (1x1 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 performance 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 study region and available set of hydro-climatological 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 (*PSH* and *PSC*) methods and the simple benchmark (*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 (*PSS),* multi-donors (*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 multi-models indicated that the Kirchmod outperformed the other models, which indicates the relative merits of the ‘top-down’ and parsimonious modelling. Nearly equivalent performance of the single-and multi-donor methods, the simple benchmark and more advanced regionalization methods, and the effects of model structure and modelling paradigm indicate that comprehensive identification of suitable regionalization methods is necessary for more reliable prediction in ungauged basins. The study 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 present study was the first attempt for regionalization of hourly runoff simulation in the region. Further regionalization study at hourly temporal resolution for the study region should focus on representative (i.e. high-density gauging networks and longer records of climate and streamflow input) and physio-climatic attributes including precipitation and soil hydraulic attributes. The findings from the present study 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 or CEDREN (http://www.cedren.no) funded the project. 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 1:50000 map from the Norwegian Mapping Authority were used. The authors are very thankful to the anonymous reviewers 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, 6135–6153.

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.

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

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

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

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

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

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

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

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

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

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, 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.” *Hydrolog. Process.*, 28, 2033–2045.

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

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

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

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

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

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

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

Gottschalk, L., Leblois, E., and Skøien, J.O. 2011. “Distance measures for hydrological data having a support.” *J. Hydrol*., 402, 415–421.

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.

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

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, doi:10.1029/2001WR000854.

Hailegeorgis, T., and Alfredsen, K. (2014a). “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),* Ecohydraulics 2014, Trodheim, Norway, 23-27 June.

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

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, 446–460.

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

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H.H.G., Gascuel-Odoux, C. (2014). “Process consistency in models: The importance of system signatures, expert knowledge, and process complexity.” *Water Resour. Res*. 50, 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, W02429.

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

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

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

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

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), doi: 10.1029/2003WR002428.

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

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, 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” *Hydrology Research,* 44 (3), 430–440.

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

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

McIntyre N. and Al-Qurashi A., (2009). “Performance of ten rainfall-runoff models applied to an arid catchment in Oman.” *Environmental Modelling & Software*, 24(6), 726-738.

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

Merz, R., Parajka, J., and Blöschl, G. (2009). “Scale effects in conceptual hydrological modeling.” *Water Resour. Res.,* 45, W09405, doi:10.1029/2009WR007872.

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

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

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

Oudin, L., 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, W07410, doi:10.1029/2005WR004636.

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, W03413, Doi: 10.1029/2007WR006240.

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

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

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

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

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

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, 198–209.

Pechlivanidis, I.G, Jackson, B.M., McIntyre, N.R., 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." *Global NEST Journal*, 13(3), 193-214.

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, 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, 81–82.

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

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

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

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

Seibert, J. (1999). “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, 3697–3708.

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

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

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.” *Hydrolog. 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.” Climate and Land Surface Changes in Hydrology *Proceedings of H01, IAHS-IAPSO-IASPEI Assembly*, Gothenburg, Sweden, July (IAHS Publ. 359, 2013), 165–170.

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

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

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

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

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

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

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