Direct statistical downscaling of monthly streamflow from atmospheric variables in catchments with differing contributions from snowmelt

Shaochun Huang | Deborah Lawrence | Nilsen Irene Brox | Hong Li

Hydrology Department, Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

Abstract

The climate–hydrological modelling chain is the most widely used approach for assessing likely future changes in streamflow but is also associated with large uncertainties. We present an alternative approach for modelling monthly streamflow directly from large-scale atmospheric variables using statistical downscaling to streamflow for 42 catchments with flow generation regimes ranging from largely snowmelt-driven to entirely rainfall-driven. It aims to find the most appropriate combination of predictors and methods, which can best include information pertaining to snow storage and provide robust streamflow simulations under changing climate conditions. The effects of selection of probable predictors (five groups of potential variables from reanalysis data, three moving windows, calibration periods of two different lengths) and methods (eight combinations of pre-processing methods and statistical models) were considered. Results indicate that the following selection of probable predictors gave a good and robust model performance: a combination of 13 independent climate variables and their lagged information, a 3-month moving window and a calibration period of 31 years. Three combinations of methods performed best: (a) principal component analysis (PCA) and the relevance vector machine (RVM) statistical model using gridded data, (b) PCA and RVM using interpolated data and (c) the Pearson correlation analysis and RVM using interpolated data. Monthly streamflow was well reproduced for most catchments, with a median NSE (Nash and Sutcliffe efficiency) of 0.82 and PBIAS (percentage bias) of 0 in the calibration period and a median NSE over 0.60 and PBIAS less than −6 in the validation period. The best performance of the methods was associated with snow-dominated catchments in inland regions. We also established that lagged information on atmospheric variables gave better results than snow variables from reanalysis data. The poorest results were associated with rainfall-dominated catchments, and this was
strongly correlated with the inter- versus the intra-annual variability in monthly streamflow.

**KEYWORDS**
empirical downscaling, machine learning, monthly streamflow, multi-linear regression, Norway, predictor selection

## 1 | INTRODUCTION

Rapid climate change urgently requires credible climate and hydrological projections for developing adaptation strategies. This is especially true for the northern high latitudes, that are expected to warm more rapidly than mid-latitudes under climate change (IPCC, 2014). Consequently, significant changes in the hydrological cycle are likely to occur in northern Europe, including increases in mean and extreme precipitation, changes in soil moisture and runoff, and shifts in snow regimes (Raisanen, 2016).

To date, most hydrological impact studies of climate change involve a three-step approach (Olsson et al., 2016): the so-called climate–hydrological modelling chain. The first step is to generate climate projections using global climate models (GCMs). The GCM outputs are either dynamically downscaled by regional climate models (RCMs) or statistically downscaled for use in regional scale analyses. The second step is to simulate streamflow or hydrological processes using hydrological models and time series derived from the downscaled climate projections. The final step is to analyse the projected changes in the hydrological cycle using the simulations, and these analyses are then applied in developing adaptation strategies. This three-step approach has been, for example, applied in the Norwegian government report “Climate in Norway 2100” (Hanssen-Bauer et al., 2017).

Several weaknesses and large uncertainties exist within such a complex climate-hydrological modelling chain. First, the RCMs are computationally expensive, so it is not possible to apply them to all GCM outputs. For example, within the World Climate Research Program Coordinated Regional Downscaling Experiment (CORDEX) initiative for Europe (www.euro-cordex.net), several RCMs could only provide downsampling of one GCM per RCM. Second, post-processing methods, which statistically downscale the coarse-resolution GCM/RCM outputs into finer resolution climate information and correct biases that exist in climate model outputs (Mueller and Seneviratne, 2014), are often criticized for having too simple statistical methods and implicit assumptions (Ehret et al., 2012; Hewitson et al., 2014; Maraun, 2016; Maraun et al., 2017). Last, but not least, the choice of RCMs, statistical post-processing methods, and hydrological models, both their structures and their parameterizations, contributes to substantial uncertainties in impact projections (Hagemann et al., 2011; Lawrence and Haddeland, 2011; Dobler et al., 2012; Chen et al., 2013; Meresa and Romanowicz, 2017).

To avoid these problems, one alternative is to apply a direct statistical downscaling of GCM outputs to hydrological variables, such as streamflow, assuming that the statistical relationship is also valid under changing climatic conditions in the future. The direct downscaling method generally applies synoptic-scale atmospheric variables that are better resolved by GCMs than the local or regional-scale temperature and precipitation series (Cannon and Whitfield, 2002). By directly linking GCM outputs and hydrological variables, uncertainty introduced by different RCMs, post-processing methods and hydrological models can, in principle, be avoided. In addition, such a method is simple and computationally efficient, so it can be easily applied to a full ensemble of GCM outputs to account for the uncertainty contributed by the GCMs (Das and Nanduri, 2018). However, the direct downscaling method has also been criticized for its oversimplification of the hydrological cycle and its lack of water stores and transfers within the soils and groundwater of a catchment (Xu, 1999). Additionally, some early studies indicated a poor performance of the direct downscaling methods (Wilby et al., 1999) and discouraged the further application of this approach. Furthermore, the direct downscaling method generally cannot take into account other important factors affecting streamflow variability, such as variations in land use and soil cover over time.

Nevertheless, recent studies applying direct downscaling methods using advanced statistical models show reasonable results for monthly streamflow (Noori et al., 2011; Sachindra et al., 2013; Tofiq and Guven, 2014; Okkan and Inan, 2015; Sachindra et al., 2015) and seasonal flows (Landman et al., 2001; Ghosh and Mujumdar, 2008; Das and Nanduri, 2018) across the world. Some studies have even succeeded in simulating streamflow with a higher temporal resolution, such as daily or 5-day streamflow (Cannon and Whitfield, 2002; Tisseuil et al., 2010). Joshi et al. (2016) found that the direct downscaling using Sparse Bayesian Learning outperformed the climate-
hydrological modelling approaches with respect to low-flow indices and Sidibe et al. (2020) found that this method was more efficient than hydrological models for poor-data regions.

Most of the abovementioned studies focus on one catchment, where streamflow is mainly generated from rainfall. Only Cannon and Whitfield (2002) and Tisseuil et al. (2010) provided a comprehensive analysis of various downscaling methods for 21 and 51 catchments in different flow regimes, respectively. The number of potential variables, which were selected directly from the reanalysis data based on literature review or experiences, varied substantially in these studies, ranging from 1 to 26. The time steps of the variables were generally consistent with the time step of predictands, but some studies (e.g., Cannon and Whitfield 2002) also applied information on antecedent conditions to describe lagged effects, for example, snow accumulation, soil and groundwater retention. The gridded climate variables can be used either directly as probable predictors or interpolated relative to the catchment considered, and they were usually standardized or normalized prior to the downscaling process. Principal component analysis (PCA) and correlation analysis were the most used pre-processing methods for reducing the number of probable predictors. Various machine learning techniques have been tested in recent studies, and they were often compared with linear regression models. There were no clear guidelines for selecting calibration and validation periods. In general, the training data accounted for 70–80% of the time series and the rest was used as the testing subset (Tisseuil et al., 2010; Joshi et al., 2013; Tofiq and Guven, 2014; Okkan and Inan, 2015). Das and Nanduri (2018), however, used 90% of the data as the training subset in order to maximize the number of data available for training the nonlinear models. Detailed information about these studies is listed in Appendix A.

There has been very little previous work using direct downscaling methods to predict hydrological variables in Norway, and no studies have shown that such methods can be applied to the full range of seasonal flow regimes found in Norway, ranging from snow-dominated inland regions to rainfall-dominated regions along the coast. The only available previous study was reported by the Norwegian Meteorological Institute (Engen-Skaugen et al., 2007) and considered precipitation and temperature as separate univariate predictors of monthly streamflow in several catchments. They found that the best statistical relationships between monthly rainfall and monthly streamflow were during the winter months in catchments with predominantly rainfall-driven flow regimes. Overall, poor results were reported for most catchments, particularly in cases where seasonal flow regimes are strongly controlled by snowmelt processes. This is not surprising, as one expects a one-to-one relationship between monthly streamflow and precipitation or temperature to be most evident when rainfall rather than snowmelt drives streamflow. We, therefore, reconsider here the general downscaling approach using more advanced statistical methods, a wider range of atmospheric variables as potential variables, and a larger selection of catchments than Engen-Skaugen et al. (2007).

Unlike many previous studies that have only applied the direct downscaling method to one catchment, we selected up to 42 unregulated catchments distributed across the country, having flow regimes ranging from rainfall-dominated to snow-dominated (Section 2). We selected the methods from the previous direct downscaling studies with most promising results (Section 3). In addition, we tested different techniques for the selection of predictors, and these were rarely reported in other studies. As a result, we were able to present an extensive statistical downscaling analysis relating large-scale climate information to catchment-scale streamflow in Norway (Section 4). We discuss and compare our results with other direct downscaling studies in Section 5. Finally, we summarize the findings and explain the further application of the method for projecting climate impacts on monthly runoff in the gauged catchments in Norway (Section 6). The objectives of this study can be formulated as (a) to apply different predictors and downscaling methods to 42 Norwegian catchments representing a range of flow regimes, and (b) to find the most appropriate combination of predictors, pre-processing methods and statistical models for further use in the assessment of climate change impact on streamflow in Norway.

2 | STUDY AREA

Since statistical relationships derived between GCM outputs and streamflow cannot take into account flow regulation, we selected 42 catchments without significant regulation representing different regions in Norway (Figure 1a). All the daily discharge data for these catchments have been quality checked by the Norwegian Water Resources and Energy Directorate (NVE) for their suitability for analysing monthly streamflow, and daily flow records are continuously available for the period 1960–2010 for the 42 catchments. In addition, 35 of the 42 catchments have longer time series starting in 1950 or earlier. The size of the catchments varies from 7 to 14,156 km², and 83% of them are smaller than 1,000 km².

The distribution of the catchments represents various climate and hydrological regimes, geographic conditions and landscape types in Norway. In general, these
catchments are located in three climate regimes (the oceanic climate, continental climate and polar tundra climate) according to the Köppen-Geiger climate classification scheme (Peel et al., 2007). The mean annual temperature during 1971–2000 is highest (up to 7°C) along the coast of southern Norway, while it is as low as −4°C in the high mountains and northern Norway (Hanssen-Bauer et al., 2017). The annual precipitation also shows a distinct spatial distribution, ranging from 300 mm in south-eastern Norway and inland areas in northern Norway to more than 3,500 mm in the west. Consequently, the catchments in the north and in the mountainous regions receive large amounts of snow, with a snow ratio (snowfall/total precipitation) larger than 0.5 (Figure 1b), while the catchments along the coast mainly receive rainfall with a snow ratio less than 0.25.

The spatial patterns of rainfall and snow accumulation exert a strong control on the distribution of annual and seasonal streamflow in Norway. Although the average annual runoff is estimated to be 1,100 mm from 1971 to 2000 for the whole of Norway, this varies from under 400 mm in several catchments in the inland regions of eastern Norway and northernmost Norway (Finnmark) to over 3,500 mm in areas of western Norway (Hanssen-Bauer et al., 2017). The seasonal distribution of streamflow varies considerably from catchment to catchment, depending on the location, the median elevation and distribution of elevation in the catchment. However, it is generally found that catchments in or proximal to coastal regions have the highest streamflow in autumn and winter months during periods of higher rainfall. Catchments in inland regions and throughout Finnmark, on the other hand, have high streamflow due to snowmelt in late spring and summer months and very low streamflow in winter months.

The differences in primary runoff source, that is, rainfall versus snowmelt, leads also to pronounced differences in inter-annual versus intra-annual variability in monthly streamflow. A streamflow variability index can be calculated as the ratio given by the average inter-annual variance in monthly streamflow (i.e., year-to-year variation...
for individual months) divided by the average intra-annual variance in monthly streamflow (i.e., month-to-month variation in a given year). Figure 1b shows that catchments with snow-dominated flow regimes tend to have low values of the variability index, whereas catchments with rainfall-dominated flow regimes have high values. The highest values are found along the coast in western and mid- to northern Norway. Detailed information on the 42 catchments is listed in Appendix B.

3 | METHOD AND DATA

3.1 | Data

As in previous direct downscaling studies, we selected the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data (Kalnay et al., 1996) from 1949 to 2010 to test the downscaling models. The NCEP/NCAR reanalysis data were aggregated into a monthly time step, at 2.5° resolution covering the study domain of 2.5–32.5° E longitude and 55–72.5° N latitude (Figure 1). Monthly time series of streamflow were obtained from NVE. The effect of the calibration period length, that is, the subset used as training data, was investigated in this study. This has not been considered in past studies, but can be a crucial issue when considering use of other reanalysis products with a shorter time series length, for example, ERA-Interim reanalysis (Dee et al., 2011) starting in 1979. The monthly time series of streamflow and the NCEP/NCAR data, which cover the 41-year period of 1950–1990 and the 31-year period of 1960–1990, were used for model calibration. The remaining data, covering the 20-year period from 1991 to 2010 were used for the model validation. In other words, we used 67% and 60% of the data as the training subset and 33% and 40% of data for validation, respectively.

The procedure for selecting appropriate variables for downscaling streamflow involves identifying potential variables, testing for correlations and developing approaches that account for snow storage. This produces five groups of potential variables: group 1, 22 variables; group 2, 13 independent variables; group 3, 13 independent variables, plus lagged information; group 4, 22 variables plus snow; and group 5, 13 independent variables plus snow. We explain each group of potential variables below.

First, up to 22 climate variables were identified as potential variables based on the previous studies (Appendix A). They are 2 m air temperature (T_2m), air temperature (T) at 500, 850, 1,000 hPa, surface air temperature (T_surface), surface skin temperature (T_skin), geopotential height (hgt) at 500, 850, 1,000 hPa, surface pressure, relative humidity (rhum) at 500, 700, 850, 1,000 hPa, 2 m specific humidity (shum_2m), specific humidity (shum) at 500, 850, 1,000 hPa as well as the average of specific humidity at all hPa levels (shum_all). Precipitation was not included as a probable variable, because the main objective of this study is to test direct downscaling methods using atmospheric variables that are better resolved by GCMs and not necessarily those used by hydrological models as input.

Since some of the variables can be highly correlated, we computed the Pearson correlation coefficient (Pearson, 1895) for each pair of them (Figure 2). The correlation matrix indicates high positive correlations between the temperature variables and some of the specific humidity and pressure variables. Based on this matrix, we selected 13 relatively independent variables with correlation coefficients less than 0.7 as group 2 of potential variables. This group includes T at 500, 1,000 hPa, hgt at 500, 850, 1,000 hPa, surface pressure, rhum at 500, 700, 850, 1,000 hPa and shum at 500, 850, 1,000 hPa.

Because streamflow in the summer half-year is determined not only by the summer climatic conditions but also by the amount of snow storage in the previous winter in many catchments in Norway, it is important to consider the antecedent conditions contributing to snow accumulation. We applied the same lagged information suggested by Cannon and Whitfield (2002) in this study, that is, the average of the variables for the previous two 3-month periods as model inputs. The lagged information was only estimated for the 13 variables because it leads to an excessively large number of variables if used with the 22 variables (i.e., 22 × 3 = 66 variables).

The fourth and fifth groups of variables include snow water equivalent (SWE), which is available from the NCEP/NCAR reanalysis data. We assumed that the difference in the SWE between two adjacent months could also be an indicator of snowmelt in summer and can possibly be used as a probable variable to replace the lagged information. In order to test this assumption, we added the snow information to the first two groups of potential variables.

3.2 | Modelling structure

The major novelty of this study is the application of various combinations of probable predictors, pre-processing methods and statistical models. Investigation of these combinations allows us to find the most appropriate downscaling method for Norway, where seasonal flow regimes cover a wide range of combinations of snow versus rainfall as the dominant streamflow generation process. The
combinations of probable predictors involve the five groups stated in Section 3.1, sampled using three alternative moving windows (1-, 3-, 5-month moving averages) and two calibration periods. We consider eight combinations of pre-processing and statistical models: two types of predictor data (gridded vs. interpolated data), two techniques for reducing the number of probable predictors (Pearson correlation analysis and PCA) and two statistical models (multi-linear regression and relevance vector machine (RVM)).

In order to explain the methods and data in detail, we use one small catchment (catchment ID 24.9—Tingvatn, shown in red in Figure 3) in the south of Norway as an example. First, we selected six grid points from the reanalysis data adjacent to this catchment (the blue dots in Figure 3). For large catchments with a drainage area larger than 10,000 km², 10 grid points were selected. The number of grid points were determined based on a pre-analysis for all catchments, and the six and 10 grid points gave relatively better downscaling results than other numbers.

In previous downscaling studies, the gridded data were used directly, for example, in Sachindra et al. (2013), or were interpolated to the catchment, for example, in Tisseuil et al. (2010). These two choices of data sources lead to substantial differences in the number of probable predictors. For the current example, the number of gridded probable predictors is six times the number of interpolated ones (Figure 3). If we select group 1 of potential variables, there will be in total $6 \times 22$ probable predictors from the gridded data, while only 22 from the interpolated data. The advantage of using the interpolated data is the shorter computational time, but the correlation between the predictors and streamflow can be weaker than that between the climate data from the nearest grid point to the catchment. Both the gridded and interpolated climate data as well as the streamflow data were standardized by subtracting the mean and dividing by the standard deviation in the calibration period. The standardization of the probable predictors in the validation period was performed with the mean and standard deviation corresponding to the calibration period of the data set.

Sachindra et al. (2013) simulated monthly streamflow using 12 statistical models, each derived separately from the historical climate and streamflow data for the calendar month of interest. The previous studies as well as our preliminary test showed that use of 12 models, one for
each month, gave better monthly streamflow than use of one model that predicts streamflow for all months simultaneously. Such monthly statistical models may be valid for the historical period for which they were developed, when changes in climate are not substantial. However, climate projections show that there is a likely shift in seasonal climatic conditions in Norway by the end of this century under the high emissions RCP scenario (RCP 8.5) (Hanssen-Bauer et al., 2017). Thus, the statistical downscaling models derived from historical data should be capable of accommodating such a shift if they are to be used for climate impact assessments in Norway.

In order to improve the model robustness for use with future projections, we therefore applied 1-, 3- and 5-month moving windows to the probable predictors and streamflow to derive the monthly statistical models (Figure 3). The 1-month moving window performs exactly the same as the method presented in Sachindra et al. (2013), i.e., 12 statistical models are derived from the climate variables and streamflow in the individual months separately. The 3- and 5-moving window methods include 3- and 5-month data to derive the models for the centred month. For example, the statistical model for February is derived from the data in January, February and March using the 3-month moving window. Such a moving window method was also applied to correct the bias of RCM outputs for Norway to improve the robustness of transfer functions (Wong et al., 2016).

As discussed above, there is a large number of probable predictors, especially when using data from grid points. To reduce the dimensionality of the data, two techniques for selecting the best explanatory predictors (Pearson correlation analysis and PCA) were applied (Figure 3). For the entire study period (1950–2010), the Pearson correlation coefficients between each probable predictor and streamflow were calculated. Ideally, the probable predictors that showed the best, statistically significant (95% confidence level, $p = .05$) correlations with the streamflow should be considered as potential predictors, which are used for calibrating statistical models for each calendar month. However, for some catchments in winter months, dozens of predictors would be selected based on this criterion, especially using the gridded data. In summer months, no potential predictors were found for some catchments, mainly due to lack of snow storage information in the probable predictors. In order to solve these problems, we selected a maximum of 15 probable predictors with the highest correlation coefficients as the potential predictors. This maximum number of predictors was determined based on a preliminary downscaling run for all catchments. For the PCA, we selected the first eight principal components (PCs) as the potential predictors because they can explain more than 95% of the variances in almost all our cases.

The potential predictors were then used in two statistical models (Figure 3): the multi-linear regression model and the machine learning technique RVM (https://cran.r-project.org/web/packages/kernlab/kernlab.pdf). We applied the train() function from the caret package in R (http://topepo.github.io/caret/index.html) to fit linear regression with backward selection, which starts with all predictors in the model, iteratively removes the least contributive predictors until all predictors are statistically significant (Kassambara, 2018). The RVM method was chosen because it was applied in several previous direct downscaling studies.
and always showed good model performance (Ghosh and Mujumdar, 2008; Okkan and Inan, 2015; Das and Nanduri, 2018). Both models used 10-fold cross-validation to estimate the mean squared error and to select the best model based on the smallest error.

The simulated monthly streamflow in the calibration and validation periods were evaluated using the non-dimensional Nash and Sutcliffe (1970) efficiency criterion (NSE) and the percentage bias (PBIAS) in streamflow. According to the model performance ratings suggested by Moriasi et al. (2015), NSE ≥ 0.7 and PBIAS ≤ 10 are considered to be “good” simulation results, while NSE ≥ 0.50 and PBIAS ≤ 15 are considered to be “satisfactory” for monthly streamflow simulations.

4 | RESULTS

4.1 | Effects of selection of different probable predictors

In this section, we analysed the effects of the potential variables, moving windows and calibration periods, which can influence the selection of probable predictors, on the simulated streamflow in terms of the NSE and the PBIAS values (Figures 4 and 5). Each boxplot in the figures consists of the criteria generated by eight combinations of pre-processing and statistical models for 35 catchments (calibration period 1950–1990, Figure 4) and 42 catchments (calibration period 1960–1990, Figure 5), respectively.

FIGURE 4  Comparison of Nash and Sutcliffe efficiency (NSE) and percentage bias (PBIAS) in streamflow using five groups of potential variables (22 variables, 13 variables, 13 variables plus lagged information, 22 variables plus snow information and 13 variables plus snow information) and three moving windows (1-, 3- and 5-month moving windows) in the calibration (1950–1990) and validation (1991–2010) periods. Each boxplot consists of the results generated by eight combinations of pre-processing methods and statistical models for 35 catchments.
When the 41-year time series were used for calibration (red boxes in Figure 4), the statistical models developed using data from a single month (1-month moving window) gave better results in terms of both the NSE and the PBIAS than did the models based on 3- and 5-month moving windows. The median NSE for all catchments varied from 0.81 to 0.84, from 0.71 to 0.73 and from 0.66 to 0.67 during the calibration period using the 1-, 3- and 5-month moving windows, respectively. The PBIAS also showed a similar pattern during the calibration period, with a median value of 0 for the 1-month moving window and from $-2$ to $-1$ for the 3- and 5-month moving windows.

During the validation period (blue boxes), the model performance did not differ as substantially as it did in the calibration period between different moving windows. The median NSE varied from 0.57 to 0.68, from 0.57 to 0.61, from 0.54 to 0.56 using the 1-, 3- and 5-month moving windows, respectively. The absolute value of the median PBIAS was between 4 and 8 using all three moving windows. The loss of model performance between the calibration and the validation periods was largest when the 1-month moving window was used. Such results indicated that although the models derived from the single monthly data gave the best model performance during the calibration period, they were not as stable in the validation period as were the models that used the broader windows and thus sampled a wider range of conditions. Considering model performance versus transferability in different periods as well as the possible seasonal shift under a warming climate in Norway, the 3-month moving window seemed to be an appropriate compromise method for projecting future streamflow using climate change scenarios.

Compared with the effects of moving windows, the five groups of potential variables did not influence the model performance substantially. The largest difference among the medians of NSE from the different groups was

**Figure 5** As in Figure 4, but for the calibration period 1960–1990 and 42 catchments
0.11 using the 1-month moving window during the validation period. Despite the minor differences, the best model performance was found using group 3, consisting of 13 variables plus lagged information, and there was almost no difference between the model performance using the 22 and 13 variables. The additional snow information only improved the streamflow simulations marginally in the calibration period as compared with the groups of 22 and 13 variables, but it leaded to the poorest model performance in the validation period when the 1-month moving window was used instead of a broader window.

A similar pattern of results for the various combinations of moving windows and potential variables were found when only 31-year time series were used for calibration (Figure 5). The model performance during the 31-year calibration period was generally better than in the 41-year period. The median NSE varies from 0.83 to 0.87, from 0.73 to 0.75, from 0.67 to 0.69 using the 1-, 3-, and 5-month moving windows, respectively. Although the calibration periods cover different lengths in Figures 4 and 5, the model performances in the validation period are similar, with the median NSE ranging from 0.58 to 0.66, from 0.59 to 0.61, from 0.54 to 0.56 using different moving windows. The PBIAS values are also comparable for the two calibration period lengths considered. This result suggests that a 31-year calibration period may be sufficient for deriving robust statistical downscaling models for Norwegian catchments.

### 4.2 The best downscaling methods for Norwegian catchments

Based on the best model performance shown in Section 4.1, we analysed the eight combinations of pre-processing and statistical models using the 13 variables plus lagged information, a 3-month moving window and a calibration period from 1960 to 1990 (Figure 6). In the calibration period, the two combinations of methods using PCA and RVM for both gridded and interpolated data gave the best model performance in terms of both median NSE (0.85) and PBIAS (0). The second best combination consists of the interpolation, Pearson correlation analysis and RVM with NSE of 0.82 and PBIAS of 0. The use of multi-linear regression model showed poorer performance than the ones using RVM and the methods using PCA and multi-linear regression models gave the poorest results in terms of NSE.

In the validation period, the model performance is degraded by 0.07–0.23 in terms of the NSE values and by up to 6 percentage points in terms of the PBIAS. The combination of interpolation, correlation analysis and RVM showed the best results in terms of median NSE (0.68) while the two combinations using PCA and RVM performed best in terms of median PBIAS (0). Taking the performance in both calibration and validation periods into account, we suggest that all three abovementioned combinations are appropriate to statistically downscale monthly streamflow in Norwegian catchments.
For the sake of simplicity, we focused on the combination of gridded data, PCA and RVM in the following results. Figure 7 shows the spatial distribution of NSE values. In the calibration period, all the catchments had NSE values larger than 0.7, with the exception of a few catchments along the western coast. In the validation period, the spatial variability in the NSE values was more pronounced than in the calibration period. The downscaling models did not satisfactorily simulate the streamflow in several catchments along the western and southern coastlines, and the poorest NSE values (NSE < 0.5) were mainly found near the coast in western, mid- and northern Norway. In contrast, the statistical downscaling models were robust in most inland areas with good model performance for both calibration and validation periods.

This distinct spatial pattern was also seen in the relationship between the snow ratio of the catchments and the NSE values for the validation period (Figure 8a), having an $r^2$ value of 0.49. The snow-dominated catchments tended to have higher NSE values than the rainfall-dominated catchments. In contrast, the rainfall-dominated catchments tended to have lower NSE values due to higher inter-annual streamflow variability in the month-to-month pattern of flow, as defined in Section 2. Figure 8b shows a high correlation between the calculated variability index and the overall performance of the downscaling approach. A similar spatial pattern of model performance as well as the relationships can be found using other combinations of methods (results not shown).

We also compared the observed and simulated monthly streamflow for one rainfall-dominated catchment (catchment ID 97.1) along the western coast and one snow-dominated catchment (catchment ID 2.614) to consider how the model performance differs in the two catchments, given their very different seasonal flow regimes (Figure 9). The statistical downscaling models can give good simulation results in terms of NSE values...
and the PBIAS for both catchments in the calibration period. It seems, however, more difficult to reproduce the complex variability of monthly streamflow which varies highly both between months in a given year and between years (catchment ID 97.1) than monthly streamflow in a catchment in which a single, snowmelt-driven peak dominates the annual pattern. The poorer simulation results for the rainfall-dominated catchment are associated with an underestimation of winter low flows and an overestimation of the autumn peaks. These results indicate a high sensitivity to the input data and a relatively poor temporal transferability of the downscaling model for a rainfall-dominated catchment.

Finally, it is interesting to know the most important climate variables that influence streamflow generation in Norway. Thus, we analysed how often the 13 potential variables and their lagged information were selected as the real predictors by the method of interpolation, Pearson correlation analysis and linear models. The linear models were chosen due to their simplicity in interpreting the predictors and this combination of methods gave a relatively good model performance for the validation period. On average, there are three real predictors used by each linear model. Figure 10 shows that the most frequently used climate variables are the specific humidity at 1,000 hPa and temperature at 1,000 hPa and its lagged information. The various geopotential heights form the secondary important group of climate variables and the relative humidity variables play the least important role in predicting streamflow in Norway.

5 | DISCUSSION

This study has investigated the effects of five groups of potential variables for developing downscaling models, which are rarely reported in other studies but are important for simulations in snow-dominated catchments. Our results confirm that lagged information is necessary and useful for obtaining good models in snow-dominated catchments, as suggested by Cannon and Whitfield (2002). The use of lagged information also outperforms the use of snow storage information available from the reanalysis...
data (see Figures 4 and 5 between the variable groups “13 Var + lag” and “13 Var + snow”). This can be partly explained by the fact that the large-scale atmospheric variables are better resolved than are the local land-surface process variables, such as snow storage, by the coarse resolution climate models.

Recent studies published during the last decade have used dozens of potential variables to simulate the streamflow. Our results show, however, that by using only 13 independent variables one obtains practically the same simulation results as those generated by the full set of 22 variables. The advantages of using a small number of variables are the shorter computational time to calibrate the models and the avoidance of overfitting the statistical models; thus, a preliminary selection of potential variables, in which relationships between variables are identified, is recommended prior to the pre-processing of the climate data.

The test of moving windows is a novelty relative to previous studies of the direct downscaling of streamflow. Before we investigated the use of moving windows, we constructed one statistical model using the entire time series to simulate the relationship between atmospheric variables and streamflow in all months. The model performance (not shown) was poor in terms of the NSE values and was, therefore, rejected for further use in developing climate impact projections. We then applied the method suggested by Sachindra et al. (2013) in which 12 statistical models are developed using the time series corresponding to each month separately, that is, the use of a 1-month moving window. This method improved the simulated monthly streamflow substantially in the calibration period, but the model performance is reduced dramatically in the validation period. Because our ultimate goal is to project streamflow under a changing climate, this method is most likely not robust enough for non-stationary conditions. As a compromise solution, we retained the concept of 12 statistical models, but tested the use of 3-month and 5-month moving windows to select a broader seasonal sample of data. This solution improves the transferability of the statistical models and provides good model performance in both the calibration and validation periods. Hence, we would also recommend testing the use of different moving windows in similar downscaling studies.

Regarding pre-processing methods, interpolation of variables between grid points appears to be advantageous because it reduces the number of probable predictors. However, we should again highlight that most of the catchments considered in this study are much smaller than the grid cells of the reanalysis data. For larger catchments, different results may be obtained if the gridded climate data is interpolated to the hydrological stations (such as Joshi et al. (2013)) or to the centroids of the catchments (such as Huang et al. (2013)) because the distance between the outlet and centroid of the catchment is long. Thus, use of original gridded data may be preferred for large catchments in other regions.

PCA is a common method for reducing the dimension of input data and has been used in several downscaling studies. Our results showed the best model performance using the combination of PCA and RVM and the poorest performance using PCA and linear models. Hence, it...
would be good to test the combination of PCA with different statistical models. However, an advantage of non-PC models is that the predictors used retain their physical meaning and their roles can therefore be interpreted in the final results. This is often not true for sets of variables after they are reduced into their corresponding PCs. In addition, our results confirm the conclusion obtained by previous studies (Cannon and Whitfield, 2002; Tisseuil et al., 2010; Joshi et al., 2013; Sachindra et al., 2013), that the nonlinear models generally give better simulation results than do linear models.

The downscaling models have a poorer performance for rainfall-dominated catchments along the coast of western, mid-, and northern Norway, but give good simulation results for most snow-dominated catchments in inland areas. This result can be somewhat surprising, as it is often assumed that it is more difficult to apply direct downscaling to snow-dominant catchments, due to the weaker correspondence between the dominant atmospheric conditions in a given season and streamflow. It is also in contrast with early work on direct downscaling of runoff in Norway using precipitation and temperature as individual univariate predictors (Engen-Skaugen et al., 2007).

In the present study, however, we have included the use of lagged information and a range of possible atmospheric variables, and this has led to a superior performance in the snow-dominated catchments, particularly relative to catchments in which rainfall is entirely dominant in streamflow generation. The weaker performance of the models for the rainfall-driven catchments may be partly due to the poor representation of coarse gridded data for the transitional areas between the ocean and the land surface. However, it also reflects the higher variability in the runoff time series in rainfall-dominated catchments; and, indeed, it is in these catchments that calibration of hydrological models using observed data is most difficult (e.g., Lawrence et al. (2009)). In addition, it may well be that atmospheric variables with a higher spatial resolution than the NCEP/NCAR data used here are necessary for modelling monthly streamflow in the smaller, rainfall-dominated catchments along the western coast of Norway. To test the effect of resolution, we applied the ERA-Interim data at the resolution of 0.75° for the same catchments using the best combination of methods obtained in this study. The median NSE of the simulated monthly discharge for the 42 gauges is 0.83 and 0.61 for the calibration period 1981–2005 and validation period 2006–2015. These results are comparable with the ones using NCEP/NCAR data, and thus the best methods and results found in this study should be robust. Finally, it may be that the inclusion of precipitation as a predictor variable would improve the model performance in these catchments for the higher resolution data.

6 | CONCLUSION

This paper presents a methodological study to systematically analyse various combinations of predictors and methods for downscaling of monthly streamflow from atmospheric variables. We tested different predictors (five alternative groups of potential variables, three possible choices of moving windows, two calibration periods of different lengths), and methods (eight combinations of pre-processing methods and statistical models) for developing monthly streamflow for 42 Norwegian catchments. Out of the tested data, we found that group 3, consisting of 13 independent variables plus their lagged information, represents the most suitable set of potential variables for simulating monthly streamflow, especially for snow-dominated catchments. The use of a 3-month moving window to derive the statistical models for the centred month gave good simulation results and improved the stability of model performance in different periods. The validation results based on a 31-year calibration period were comparable with results using a 41-year calibration period, indicating that a 31-year time series was sufficient for training the statistical models for the range of catchments considered here. Out of the tested methods, we found that the best combinations of preprocessing and statistical models were (a) PCA combined with RVM using gridded data, (b) PCA and RVM using interpolated data, and (c) the Pearson correlation analysis and RVM using interpolated data. The temperature and specific humidity variables were the most frequently selected variables for simulating streamflow generation in Norway for linear models.

Based on the best combinations of predictors and methods, the simulated monthly streamflow for the 42 catchments generally show a good agreement with the observations, with a median NSE over 0.82 and PBIAS of 0 in the calibration period and a median NSE over 0.60 and PBIAS less than −6 in the validation period. The downscaling models had a poorer performance for some rainfall-dominated catchments along the western coast of western, mid- and northern Norway coastline, but gave good simulation results for most snow-dominated catchments in the inland areas. Thus, we conclude that the direct statistical downscaling approach is capable of reproducing the recent observed changes in monthly streamflow in most of the Norwegian catchments we have considered.

Based on the conclusion, we should be able to apply the direct downscaling models for the same gauged catchments using input data derived from climate models, rather than reanalysis data. Similarly to other climate impact studies using hydrological models, we will firstly train models for each catchment using ERA-Interim data for the transitional areas between the ocean and the land surface. However, it also reflects the higher variability in the runoff time series in rainfall-dominated catchments, particularly relative to catchments in which rainfall is entirely dominant in streamflow generation. The weaker performance of the models for the rainfall-driven catchments may be partly due to the poor representation of coarse gridded data for the transitional areas between the ocean and the land surface. However, it also reflects the higher variability in the runoff time series in rainfall-dominated catchments; and, indeed, it is in these catchments that calibration of hydrological models using observed data is most difficult (e.g., Lawrence et al. (2009)). In addition, it may well be that atmospheric variables with a higher spatial resolution than the NCEP/NCAR data used here are necessary for modelling monthly streamflow in the smaller, rainfall-dominated catchments along the western coast of Norway. To test the effect of resolution, we applied the ERA-Interim data at the resolution of 0.75° for the same catchments using the best combination of methods obtained in this study. The median NSE of the simulated monthly discharge for the 42 gauges is 0.83 and 0.61 for the calibration period 1981–2005 and validation period 2006–2015. These results are comparable with the ones using NCEP/NCAR data, and thus the best methods and results found in this study should be robust. Finally, it may be that the inclusion of precipitation as a predictor variable would improve the model performance in these catchments for the higher resolution data.
reanalysis data that are upscaled (i.e., aggregated) to match the GCM resolution. Secondly, we will run the calibrated downscaling models using GCM projections that have been bias corrected relative to the upscaled ERA-Interim reanalysis data using an anomaly approach for multi variables (e.g., wind, temperature, humidity and pressure). An example of this bias correction method can be found in Pontoppidan et al. (2018). Thirdly, changes in monthly runoff can be estimated by comparing the simulation results in the scenario and reference periods. The ultimate goal of our work is to compare the projected changes using the direct downscaling methods with those simulated using hydrological models and the downscaled bias-corrected climate data to assess the similarities and differences in the outcomes from these two distinct approaches.

ACKNOWLEDGEMENTS
The authors would like to thank Norwegian Research Council for funding the R3 project ( Relevant, Reliable and Robust local-scale climate projections for Norway, grant 255397) and Abdelkader Mezghani, Rasmus E. Benestad and Helene Birkelund Erlandsen from the Norwegian Meteorological Institute for discussions of the statistical downscaling methods.

ORCID
Shaochun Huang  
https://orcid.org/0000-0001-7426-5181

Nilsen Irene Brox  
https://orcid.org/0000-0003-2828-8854

REFERENCES
Cannon, A.J. and Whitfield, P.H. (2002) Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. Journal of Hydrology, 259 (1–4), 136–151. https://doi.org/10.1016/S0022-1694(01)00581-9.

Chen, J., Brissette, F.P., Chaumont, D. and Braun, M. (2013) Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two north American river basins. Journal of Hydrology, 479, 200–214. https://doi.org/10.1016/j.jhydrol.2012.11.062.

Das, J. and Nanduri, U.V. (2018) Assessment and evaluation of potential climate change impact on monsoon flows using machine learning technique over Wainganga River basin, India. Hydrological Sciences Journal, 63(7), 1020–1046. https://doi.org/10.1080/02626667.2018.1469757.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Drago, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N. and Vitart, F. (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597. https://doi.org/10.1002/qj.828.

Dobler, C., Hagemann, S., Wilby, R.L. and Stotter, J. (2012) Quantifying different sources of uncertainty in hydrological projections in an Alpine watershed. Hydrology and Earth System Sciences, 16(11), 4343–4360. https://doi.org/10.5194/hess-16-4343-2012.

Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P. and Ludwig, F. (2017) Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level. Climatic Change, 143(1–2), 13–26. https://doi.org/10.1007/s10584-017-1971-7.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K. and Liebert, J. (2012) Should we apply bias correction to global and regional climate model data? Hydrology and Earth System Sciences, 16(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012.

Engen-Skaugen, T., Benestad, R. and Roald, L.A. (2007) Empirically Downscaling of Runoff in Norway: Is it Feasible? MET Report 15/2017. Oslo, Norway: Norwegian Meteorological Institute.

Ghosh, S. and Mujumdar, P.P. (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in Water Resources, 31(1), 132–146. https://doi.org/10.1016/j.advwatres.2007.07.005.

Hagemann, S., Chen, C., Haerter, J.O., Heinke, J., Gerten, D. and Piani, C. (2011) Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. Journal of Hydrometeorology, 12(4), 556–578. https://doi.org/10.1175/2011jhm1336.1.

Hanssen-Bauer, I., Førland, E.J., Haddeland, I., Hisdal, H., Lawrence, D, Mayer, S., Nesje, A., Nilsen, J.E.O., Sandven, S., Sandø, A.B., Sorteberg, A., Ådlandsvik, B. (2017). Climate in Norway 2100 – a knowledge base for climate adaptation. NCICS report 1/2017, The Norwegian Centre for Climate Services.

Hewitson, B.C., Daron, J., Crane, R.G., Zermoglio, M.F. and Jack, C. (2014) Interrogating empirical-statistical downscaling. Climate Change, 122(4), 539–554. https://doi.org/10.1007/s10584-013-1021-z.

Huang, S.C., Krysanova, V. and Hattermann, F.F. (2013) Projection of low flow conditions in Germany under climate change by combining three RCMs and a regional hydrological model. Acta Geophysica, 61(1), 151–193.

IPCC. (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kottlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.F., Teichmann, C., Valentini, R., Vautard, R., Weber, B. and Yio, P. (2014) EURO-CORDEX: new high-resolution climate change projections for European impact
Wilby, R.L., Hay, L.E. and Leavesley, G.H. (1999) A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *Journal of Hydrology*, 225(1–2), 67–91.

Wong, W.K., Haddeland, I., Lawrence, D., Beldring, S. (2016). Gridded 1 x 1 km climate and hydrological projections for Norway. NVE Report 59/2016, Norwegian Water Resources and Energy Directorate.

Xu, C.Y. (1999) From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography*, 23(2), 229–249.

---

**How to cite this article:** Huang S, Lawrence D, Irene Brox N, Li H. Direct statistical downscaling of monthly streamflow from atmospheric variables in catchments with differing contributions from snowmelt. *Int J Climatol*. 2021;41 (Suppl. 1): E2757–E2777. [https://doi.org/10.1002/joc.6878](https://doi.org/10.1002/joc.6878)
APPENDIX A: List of studies published after 2000 which consider direct downscaling of climate variables to streamflow

| Literature            | Region                                    | Hydrological regime | Predictors                                                                 | Preprocessing                                                                 | Model/method                                                                 | Predictand                        |
|-----------------------|-------------------------------------------|---------------------|----------------------------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------|
| Landman et al. (2001) | Vaal and upper Tugela river catchments in South Africa | Pluvial             | 850, 700, 500 and 200 hPa geopotential heights 700 hPa relative humidity 850–500 hPa thickness 500–200 hPa thickness | Empirical orthogonal function (EOF) | Canonical correlation analysis (CCA) | December–February seasonal streamflow |
| Cannon and Whitfield (2002) | British Columbia, Canada | Pluvial, nival and glacial | Sea-level pressure (+previous two 3-month periods) 500 hPa geopotential height (+previous two 3-month periods) 850 hPa specific humidity (+previous two 3-month periods) 1,000–500 hPa thickness fields (+previous two 3-month periods) | Standardization, stepwise variable selection | The multi-layer perceptron (MLP) neural network + bootstrap aggregation, linear regression models | 5-day discharge |
| Ghosh and Mujumdar (2008) | Mahanadi river basin in India | Pluvial             | 2 m surface air temperature Mean sea level pressure 500 hPa geopotential height Surface specific humidity | Standardization, principal component analysis (PCA), fuzzy clustering | Support vector machine (SVM) and relevance vector machine (RVM) | Monsoon season discharge |
| Tisseuil et al., 2010 | Southwest France | Pluvial, nival | 1,000, 850 and 500 hPa daily air temperature Mean daily convective precipitation rate at surface Mean daily clear sky downward longwave flux at surface Mean daily clear sky downward solar flux at surface Mean daily clear sky upward solar flux at surface Mean daily downward longwave radiation flux at surface Mean daily downward solar radiation flux at surface Mean daily upward longwave radiation flux at surface Mean daily precipitation rate at surface Mean daily precipitation rate at surface Mean daily surface pressure 1,000, 850 and 500 hPa mean daily relative humidity Mean daily upward solar radiation flux at surface 1,000, 850 and 500 hPa mean daily specific humidity | Interpolation, standardization, PCA, hierarchical ascending cluster analysis | Generalized linear models (GLM), generalized additive models (GAM), aggregated boosted trees (ABT) and artificial neural networks (ANN) | Fortnightly, daily, daily integrated |
| Literature | Region | Hydrological regime | Predictors | Preprocessing | Model/method | Predictand |
|------------|--------|---------------------|------------|---------------|--------------|------------|
| Noori et al. (2011) | Sofichay River in Iran | Pluvial, nival | Mean daily SST/land skin temperature Mean daily sea level pressure Mean daily total cloud cover Mean daily latent heat net flux at surface Mean daily sensible heat net flux at surface Monthly rainfall \((t, t-1, t-2)\) Monthly sun radiation \((t, t-1, t-2)\) Monthly temperature (min, max, average) \((t, t-1, t-2)\) | Normalization, PCA, gamma test and forward selection | Support vector machine (SVM) | Monthly flow |
| Joshi et al. (2013, 2016) | Moisie, Romaine and Ouelle rivers in eastern Canada | Nival | Mean temperature at 2 m (yearly, seasonly and monthly) Specific humidity at 1,000, 850 and 500 hPa Relative humidity at 1,000, 850 and 500 hPa Zonal wind component at 1,000, 850 and 500 hPa Meridional wind component at 1,000, 850 and 500 hPa Vertical wind component at 1,000, 850 and 500 hPa Geopotential height at 1,000, 850 and 500 hPa Heating degree days | Interpolation, normalization, canonical correlation analysis, backward stepwise predictor selection | Relevance vector machine (RVM) and multiple linear regression | Low flow indices |
| (Continues) | (Continues) | (Continues) | (Continues) | (Continues) | (Continues) | (Continues) |

APPENDIX A (Continued)
| Literature                  | Region                             | Hydrological regime | Predictors                                                                 | Preprocessing                                      | Model/method                   | Predictand                      |
|-----------------------------|------------------------------------|---------------------|----------------------------------------------------------------------------|-----------------------------------------------------|---------------------------------|---------------------------------|
| Tofiq and Guven (2014)      | The Diyala River in Iraq            | Pluvial             | Mean sea level pressure \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa wind speed \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa zonal velocity \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa Meridional velocity \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa Vorticity \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa wind direction \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa divergence \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 850 and 500 hPa Geopotential \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | Accumulated precipitation \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | 1,000, 850, 500 hPa specific humidity \((t, t-1, t-2, t-3, t-4, t-5, t-6)\)  
                           | Screen air temperature \((2 \text{ m})\) \((t, t-1, t-2, t-3, t-4, t-5, t-6)\) | The natural logarithm of the inflow data | Linear genetic programming | Peak monthly discharge |
| Okkan and Inan (2015)       | The Kemer watershed in Turkey       | Pluvial             | Surface, 200, 500 and 850 hPa mean air temperature  
                           | 200, 500 and 850 hPa geo-potential height  
                           | 500 and 850 hPa relative humidity  
                           | Surface large scale precipitation  
                           | Surface Sea level pressure | PCA | Feedforward neural network (FFNN), least squares support vector machine (LSSVM), multi-linear regression (MLR) and relevance vector machine (RVM) | Monthly flow |
| Sachindra et al. (2015)     | The Fyans Creek catchment in South Australia | Pluvial             | The same as Sachindra et al. (2013)  
                           | Downscaled precipitation \((t, t-1, t-2)\)  
                           | Downscaled evaporation \((t, t-1, t-2)\) | Standardization, Pearson correlation, stepwise predictor selection | Multi-linear regression (MLR) | Monthly flow |
| Das and Nanduri (2018)      | The Wainganga River basin, India    | Pluvial             | Mean sea-level pressure  
                           | Specific humidity  
                           | 500 hpa geopotential height  
                           | 2-m surface air temperature | Standardization, PCA, fuzzy c-means (FCM) clustering | Support vector machine (SVM) and relevance vector machine (RVM) | Monsoon flows |
| Sidibe et al. (2020)        | West and Central Africa             | Pluvial             | Sea surface temperature       | EOF, PCA | Linear step-wise regression model | Monthly flow |
APPENDIX B: List of the 42 selected catchments, the first year of data used in this study, and catchment characteristics

| Catchment ID | Gauging station        | Discharge data starting from | Area (km²) | Snow/total precipitation | Streamflow variability index |
|--------------|------------------------|-----------------------------|------------|--------------------------|------------------------------|
| 2.11         | Narsjø                 | 1950                        | 119        | 0.41                     | 0.2                          |
| 2.142        | Knappom                | 1950                        | 1,648      | 0.31                     | 0.49                         |
| 2.32         | Atnasjø                | 1950                        | 463        | 0.41                     | 0.22                         |
| 2.614        | Rosten                 | 1950                        | 1828       | 0.5                      | 0.17                         |
| 12.7         | Etna                   | 1950                        | 570        | 0.42                     | 0.21                         |
| 15.79        | Orsjoren               | 1960                        | 1,177      | 0.56                     | 0.19                         |
| 16.66        | Grosettjern            | 1950                        | 6          | 0.46                     | 0.23                         |
| 20.2         | Austenå                | 1950                        | 276        | 0.39                     | 0.52                         |
| 22.16        | Myglevatn ndf,         | 1960                        | 182        | 0.29                     | 0.76                         |
| 24.9         | Tingvatn (Lytne)       | 1950                        | 272        | 0.32                     | 0.74                         |
| 28.7         | Haugland               | 1950                        | 140        | 0.07                     | 0.6                          |
| 41.1         | Stordalsvatn           | 1950                        | 129        | 0.31                     | 0.84                         |
| 48.1         | Sandvenvatn            | 1950                        | 468        | 0.56                     | 0.28                         |
| 48.5         | Reinsnosvatn           | 1950                        | 120        | 0.61                     | 0.24                         |
| 50.1         | Hølen                  | 1950                        | 232        | 0.65                     | 0.23                         |
| 55.4         | Røykenes               | 1950                        | 50         | 0.15                     | 0.88                         |
| 62.5         | Bulken (Vangsvatnet)   | 1950                        | 1,094      | 0.46                     | 0.41                         |
| 72.5         | Brekke bru             | 1950                        | 267        | 0.59                     | 0.22                         |
| 82.4         | Nautsundvatn           | 1950                        | 219        | 0.23                     | 0.9                          |
| 83.2         | Viksvatn (Hestadfjorden) | 1950                     | 507        | 0.47                     | 0.4                          |
| 97.1         | Fetvatn (Fitjavatnet)  | 1950                        | 89         | 0.42                     | 0.66                         |
| 103.4        | Horgeheim              | 1950                        | 1,099      | 0.66                     | 0.17                         |
| 105.1        | Osenelv v/Øren         | 1950                        | 138        | 0.27                     | 0.79                         |
| 109.9        | Driva v/Risefoss       | 1960                        | 744        | 0.55                     | 0.16                         |
| 122.11       | Eggafoss               | 1950                        | 653        | 0.48                     | 0.18                         |
| 123.31       | Kjelstad               | 1960                        | 142        | 0.4                      | 0.3                          |
| 124.2        | Høggås Bru             | 1950                        | 495        | 0.36                     | 0.35                         |
| 133.7        | Krinsvatn (Kringsvatnet) | 1950                     | 207        | 0.29                     | 0.73                         |
| 138.1        | Øyungen                | 1950                        | 239        | 0.31                     | 0.74                         |
| 152.4        | Fustvatn               | 1950                        | 526        | 0.43                     | 0.52                         |
| 157.3        | Vassvatn               | 1950                        | 17         | 0.33                     | 0.68                         |
| 162.3        | Skarsvatn              | 1950                        | 146        | 0.46                     | 0.33                         |
| 163.5        | Junkerdalselv         | 1950                        | 419        | 0.59                     | 0.17                         |
| 165.6        | Strandå                | 1950                        | 24         | 0.28                     | 0.9                          |
| 191.2        | Øvrevatn               | 1950                        | 526        | 0.53                     | 0.22                         |
| 200.4        | Skogsfjordvatn         | 1960                        | 135        | 0.44                     | 0.44                         |
| 223.2        | Lombola                | 1960                        | 878        | 0.51                     | 0.23                         |
| 234.18       | Polmak nye             | 1950                        | 14,157     | 0.42                     | 0.28                         |
| 247.3        | Karpelva               | 1950                        | 138        | 0.45                     | 0.32                         |
| 311.46       | Engeren                | 1950                        | 395        | 0.37                     | 0.25                         |
| 311.6        | Nybergsund             | 1960                        | 4,420      | 0.38                     | 0.25                         |