# Predicting low flows in ungauged catchments

**KOLBJØRN ENGELAND, HEGE HISDAL & STEIN BELDRING**
*NVE, Postboks 5091 Majorstua, N-0301 Oslo, Norway*
koe@nve.no

**Abstract** The aim of this study was to compare the regression method and a regional precipitation–runoff model to estimate low flow indices at ungauged sites. Southwestern Norway was chosen as the study region. As the first step in the regression method, two homogeneous sub-regions were established according to the dominant low flow season, winter or summer. Then individual regression equations were established for each sub-region between the low flow index "common low flow" and catchment characteristics using a step-wise procedure. A gridded version of the HBV model was applied as a regional precipitation–runoff model. The model was calibrated to a subset of the catchments and validated on independent catchments. The calibration criterion was selected to fit the low flow part of the streamflow record. Comparison of the predicted low flow index by the regression method and the HBV model showed that the regression method gave the best estimates.

**Key words** low flow indices; prediction in ungauged basins; regionalization; HBV model

## INTRODUCTION

Estimation of low flow indices at ungauged sites is needed for many decisions in water resources management. In Norway an increasing demand for low flow data, especially for small catchments, is related to the increasing request to build small hydropower plants. Also, related to other water management issues such as river pollution and ecological aspects, irrigation, reservoir design and management, drinking water supply and fish farming, there is a need to estimate low flow indices.

Due to the Norwegian Water Resources Act the construction of small hydropower plants requires estimation of the low flow index "common low flow" (clf) in small ungauged catchments. This index is a starting point to set the residual flow when a licence is needed, and is often used as the residual flow if a licence is not needed. Clf is defined as follows (preferably based on at least 15–20 years of data): (a) remove the 15 smallest values every year in a daily streamflow record, (b) calculate the annual minimum series, (c) rank the values in the annual minimum series, and (d) remove 1/3 of the smallest values. The smallest value left is the clf.

A standard procedure for obtaining the best possible estimation of low flow indices in ungauged catchments is needed. In this paper, two methods are studied: multiple regression and precipitation–runoff modelling.

Regression techniques aim to establish regression equations between low flow indices and catchment characteristics. An overview is given in Demuth & Young (2004). Norwegian studies were recently presented in Væringstad & Hisdal (2005). In heterogeneous areas it is necessary to establish individual regression equations for sub-regions that are homogeneous with regard to the low flow generating processes (e.g. Laaha & Blöschl, 2006). In Norway, there is a major difference between catchments with dominant summer and winter low flow (Væringstad & Hisdal, 2005).

A regional precipitation–runoff model can produce streamflow series from which the desired low flow indices can be calculated. This method requires good procedures to transfer model parameters from gauged to ungauged catchments and for interpolation of the meteorological input variables (temperature, precipitation, etc.). The gain is that you can calculate any low flow index from one model. The loss is the increased model complexity that might increase the uncertainties in the estimation.

The aim of this study is to evaluate and compare regression and precipitation–runoff modelling methods for estimation of clf in ungauged catchments in Norway. This was achieved by estimating regression equations between clf and catchments characteristics for 56 catchments in southern Norway. The performance of the regression equations was evaluated by a cross-validation procedure. A gridded version of the HBV-model was calibrated to a subset of the 56 catchments using objective criterions that give good fit to low flows. The calibration results were validated on an independent set of catchments. The two methods were compared using the root mean square error, explained variance ($R^2$) and bias for the predicted low flow.

This paper starts with a presentation of the streamflow and geographical data. Then the regression method is described and regional regression equations are derived. It is followed by a presentation of the HBV model and how the two methods were compared before the results are presented and some conclusions are drawn.

## DATA

The study region is the southwestern part of Norway (Fig. 1). In the inland and the mountainous areas the low flow period is in the winter due to precipitation being stored as snow, whereas in the coastal lowlands the low flow period is in the summer due to increased evapotranspiration and slightly lower rainfall. The vegetation cover is mainly coniferous and deciduous forests in the lowlands, and grass and bushes in the mountains. The area is covered by several lakes and mires that are very important for the hydrological response. The soils are mainly thin till deposits on bedrock. Fluvial deposits are mostly found in the valley bottoms.

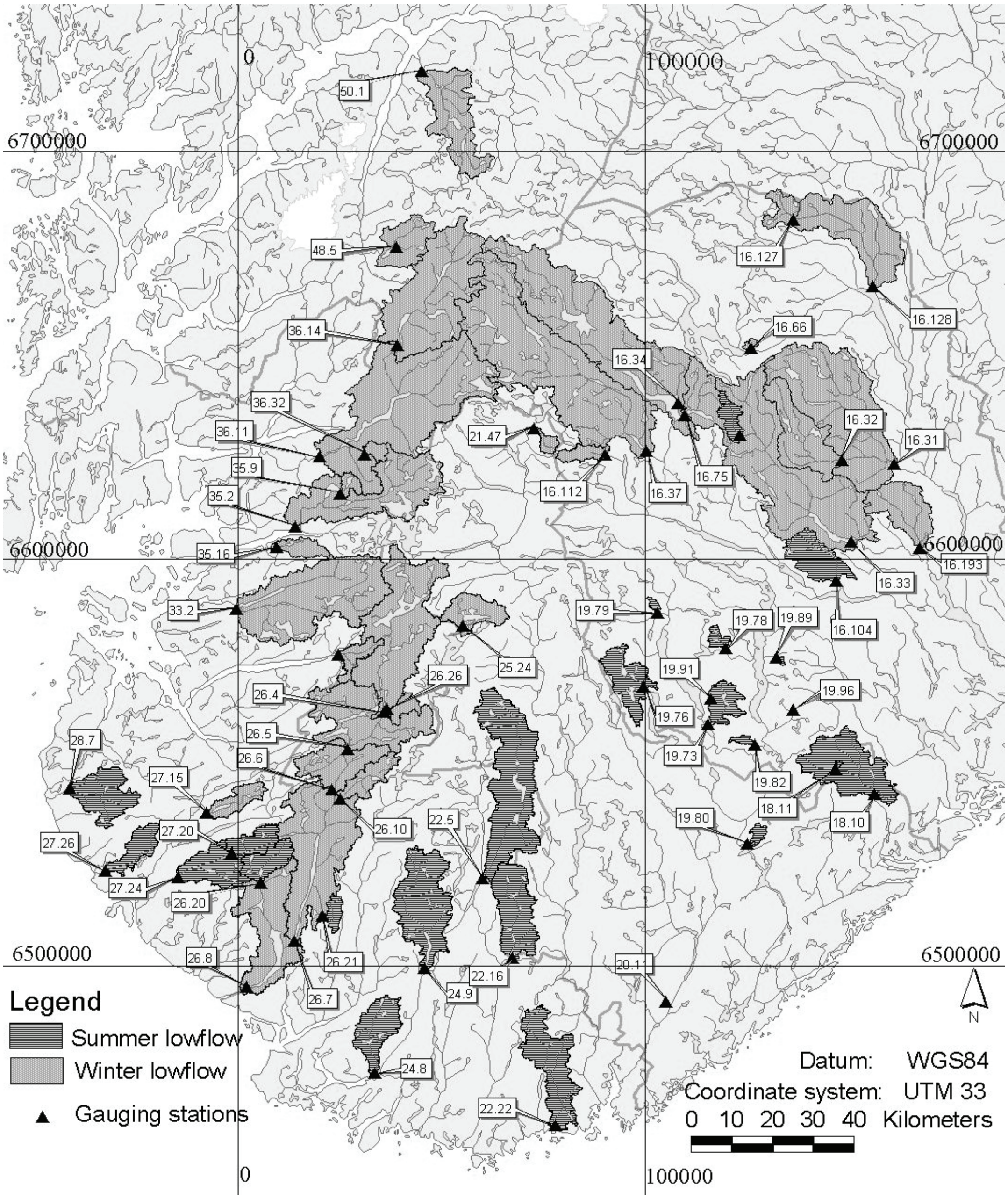


**Fig. 1** Catchments and corresponding streamflow stations used in this study.

**Table 1** The catchment characteristics included in the regression analysis.

| Symb. | Description | Symb. | Description |
|---|---|---|---|
| $A$ | Catchment area (km$^2$) | $C_G$ | Catchment gradient (m km$^{-1}$) |
| $Q_M$ | Mean annual runoff (l s$^{-1}$ km$^{-2}$) obtained from the runoff map of Norway | $G_{1085}$ | River gradient excluding the 10% lowest parts and the 15% highest parts 1085 (m km$^{-1}$) |
| $R_L$ | Length of main river (km) from the outlet to the most distant river string. | $C_L$ | Catchment length (km) from outlet to the most distant point at the water divide |
| $R_G$ | River gradient (m km$^{-1}$) | $M_{\%}$ | Mountainous areas (%) |
| $F_{\%}$ | Forested areas (%) | $L_{\%}$ | Lake percentage (%) |
| $B_{\%}$ | Bogs (%) | $L_{eff}$ | Effective lake percentage (%) |
| $C_W$ | Catchment width (km) | $T_A$ | Average annual temperature ($^o$C) |
| $A_{\%}$ | Agricultural areas (%) | $T_S$ | Average summer temperature ($^o$C) |
| $H_{max}$ | Maximum elevation (m a.s.l.) | $T_W$ | Average winter temperature ($^o$C) |
| $H_{min}$ | Minimum elevation (m a.s.l.) | $P_A$ | Annual precipitation (mm) |
| $D_H$ | Elevation gradient (m) | $P_S$ | Summer precipitation (mm) |
| $U_{\%}$ | Urbanised areas (%) | $P_W$ | Winter precipitation (mm) |

Daily streamflow data was obtained from 56 stations (Fig. 1). The stations were selected according to the record length (a minimum of 20 years) and the quality of the low flow measure ments. The dominant low flow season is indicated in Fig. 1. The summer was defined as May–October and winter as November–April. The average flows for the three winter months and the three summer months with the lowest streamflow were used to find the dominant low flow period.

Table 1 lists the physiographic and climatic catchments descriptors. All the land cover percentages are based on the N50 maps (scale 1:50 000). The gradients are based on a digital elevation model with resolution of 100 × 100 m. A digital river network was used to calculate the river gradients. The average precipitation: $P_A$, $P_S$ and $P_W$, and temperature: $T_A$, $T_S$ and $T_W$, were provided by the Norwegian meteorological institute. They were given as average values for the period 1961–1990 on a regular grid with a 1 × 1 km resolution and aggregated to catchment averages.

## REGIONAL REGRESSION ANALYSIS

The regional regression analysis was performed in two steps: (a) divide the data into groups that can be regarded as homogeneous with respect to their low flow behaviour, (b) establish regression equations for each sub-region.

### Classification

The catchments with observations were divided into two groups according to their dominating low flow period, as previously described. In order to decide whether an ungauged catchment has low flow during winter or summer, the average July temperature in combination with the ratio between summer and winter precipitation was found to be the best indicator. If the July temperature is higher than 10.4°C and the winter precipitation is more than 0.7 times the summer precipitation, the catchment has summer low flow (Fig. 2). It should be noted that the latter criterion is based only on one station (16.193 Hørte) and is therefore not very robust.

### Regression models

In the second step multiple linear regression was used to obtain relationships between clf and catchment characteristics for the winter and summer regions separately. In total, 24 catchment characteristics (Table 1) were potential candidates for the regression equation. A stepwise procedure based on the Akaikes information criterion was used to select the most important characteristics to explain the clf. In order to evaluate the predictive capability of the model, a cross validation test was carried out. Each observation was successively left out in the estimation of the regression parameters. The clf was then predicted at the independent site. The explained variance $R^2_{CV}$ for the predictions was then calculated.

Different transformations of the dependent variable, clf, were tested in order to obtain homoscedasdic residuals and the log-transformation was found to be the best alternative. Several

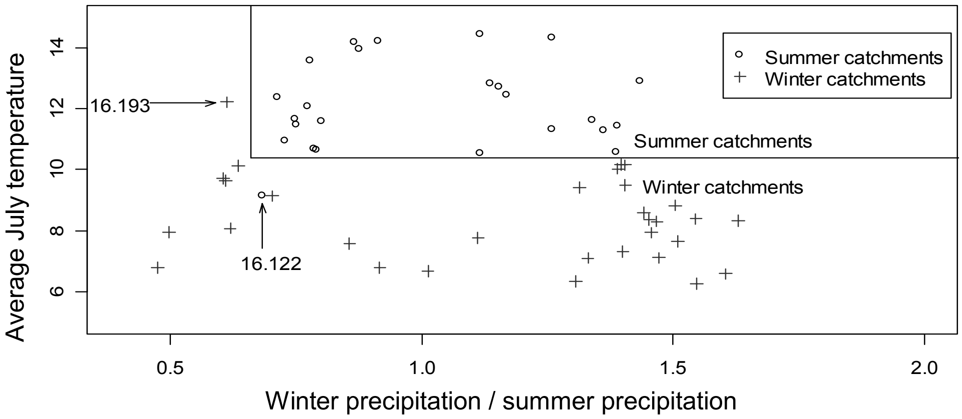


**Fig. 2** Classification of summer and winter catchments. The circles and crosses indicate the summer and winter catchments respectively according to the initial classification, whereas the lines indicate the limits according to the classification based on climatic conditions.

alternative models were evaluated. The best model fit was obtained when the independent variables were either log-transformed or kept un-transformed. Equations (1) and (2) show the estimated regression coefficients for the winter and summer region respectively. $R^2_{CV}$ for the winter region was 0.816/0.711 (log-transformed clf/un-transformed clf), and for the summer region 0.820/0.757.

$$\ln(Q_c) = -3.3325 + 0.0102C_L + 0.03298\ln(L_{eff}) + 0.026485Q_M + 1.601\ln(T_S) - 0.215\ln(F_{\%} + 0.1) - 0.0173M_{\%} \quad (1)$$

$$\ln(Q_c) = -4.734 + 1.301\ln(Q_M) - 0.448\ln(B_{\%} + 0.1) + 0.102L_{\%} + 0.0130C_L \quad (2)$$

Figure 3 shows some diagnostics. There are two plots for the summer catchments and two plots for the winter catchments. All four show predicted clf based on equations (1) and (2) *versus* observed clf. For a good model fit the points should be close to the 1:1 line. In the first plots clf is untransformed, whereas the second plots show the results as log-transformed clf.

The bias of the residuals is centred on zero, but the model underestimates the highest values. A better fit is obtained for summer catchments than for winter catchments since the low flow data in the winter are more uncertain than the summer low flow. During winter the instrument might freeze, there are often measurement errors due to ice at the gauging station, or the low flow might actually be an estimated flow based on an ice reduction procedure.

The average runoff, the lake percentage and the catchment length is included for both summer and winter catchments. The clf increases with increasing values for all these three variables. Bogs have the opposite effect of lakes. Increasing bog percentage gives decreasing clf. Bogs seem to act as swamps in the landscape. This variable is included for the summer region. For the winter region also temperature, forest areas and mountainous areas are selected. The clf increases with increasing temperatures since less precipitation is stored as snow.

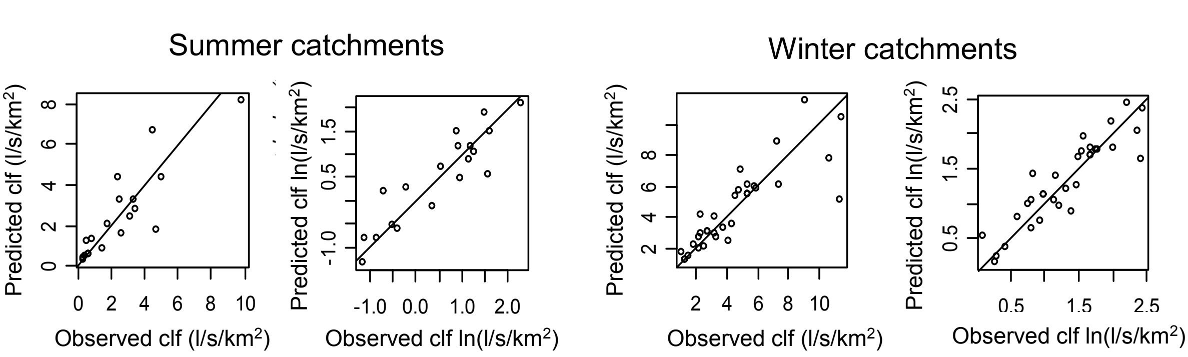


**Fig. 3** Cross-validation of regression model. The left plot of each pair of plots show the results as untransformed clf, whereas the right plot shows the results with clf log-transformed.

## THE HBV MODEL

A gridded version of the Norwegian HBV model (Beldring *et al.*, 2003) was used. The HBV model operates on daily time steps and calculates the water balance for grid cells of 1 × 1 km. For each grid cell the percentage of lake was decided in addition to the proportion of the two dominant (out of five) land-use classes. Some of the model parameters were common for the whole region whereas others were determined for each land-use class. The same process parameterizations were applied to all grid-cells.

Precipitation and temperature observations were interpolated to each grid cell using an inverse distance weighting routine with elevation correction. The precipitation gradients were calibrated according to the procedure described in Beldring *et al.* (2003).

### Calculation of common low flow

We chose to let the HBV model give the clf for each grid cell. The clf for the catchment is the average of the clfs in the grid cells within the catchment. This procedure might be difficult to use for large catchments where all parts of the catchment do not simultaneously contribute to the low flow events, e.g. some parts of the catchment have summer low flow and other parts winter low flow. Other alternatives exist, but would be complicated if applied for operational purposes.

### Calibration and validation

To evaluate the predictability of clf in ungauged catchments using the HBV model, a split sample test was applied. Daily streamflow observations from 36 stations were used for calibration. A calibration was performed using the average Nash-Sutcliffe coefficient for log-transformed streamflow to give weights on low streamflow values. In order to compare the prediction of clf using the regression method and the HBV model in a proper way, a split sample test was also performed for the regression method using the same 36 catchments to estimate the regression coefficients. The explained variance $R^2$ and bias for the clf estimated by both methods was calculated both for the calibration and the validation sets.

## RESULTS AND DISCUSSION

Figure 4 shows the observed and HBV-estimated clf for the calibration catchments and the validation catchments. The figure also shows the results for the regression method.

The regression method in general gives better prediction of clf in ungauged catchments than the HBV model. The regression method is especially superior to the HBV model for the smallest clf-values. The predictive power for clf-values smaller than 2-3 L $s^{-1}$ $km^{-2}$ for the HBV-model is rather limited.

The use of log-transformation to obtain heteroscedastic residuals indicates that the error is relative for the regression model. This means that the absolute error is small for small predicted values of clf and larger for larger predicted values of clf. For the HBV model, however, the absolute error is independent of the magnitude of the clf. This means that the precision of the lowest clf predictions is rather low.

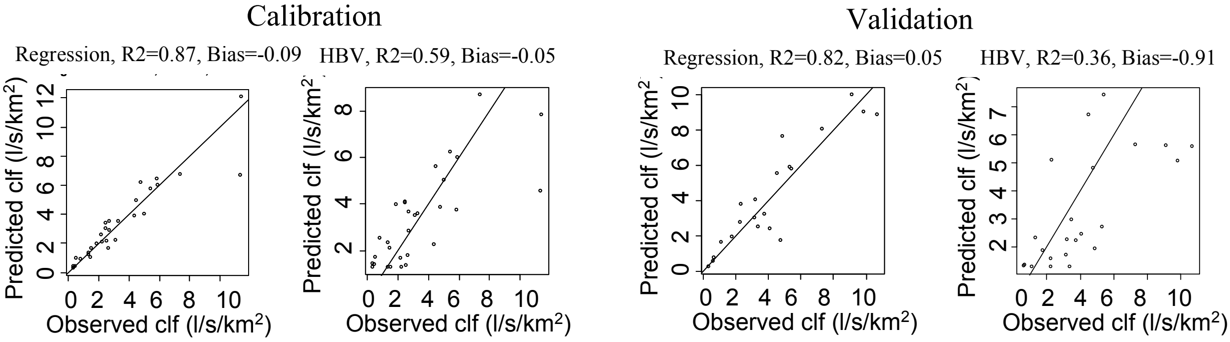


**Fig. 4** The observed and simulated clf for the calibration and validation catchments.

The use of the HBV model to calculate clf demands high performance for the simulated recession and low flow periods. Since the estimates are to be used in ungauged catchments, it is necessary to use a regional parameter set where the parameters depend on landscape characteristics. The regression method indicates that, in addition to climatic descriptors, lakes and bogs are important landscape characteristics that control the low flow. Better results might be obtained by using improved interpolation of precipitation, explicit representation of lake and bog elements and introduction of soil and land-use classes important to recession and low flow.

The regression results are limited to the selected region, and the quality of the estimates might decrease if the equations are applied in an extrapolation mode for the catchment characteristics. The regression procedure does not account for strongly correlated clf along the river network, so application of streamflow data further up- or downstream can give better results than the regression equations.

The regression method provides the best prediction of low flow indices, but separate regression equations are needed for each index and region. Also, a routine to automatically estimate the catchment characteristics of ungauged basins need to be at hand. The HBV model allows calculation of any low flow index based on the simulated time series and would be able to provide a static map of clf and other indices.

## CONCLUSIONS

A multiple regression procedure to estimate common low flow in ungauged catchments has been compared to the application of the HBV model. The following conclusions can been drawn: (a) the regression method gives better predictions of clf than the HBV model; (b) for the regression method, the best results are obtained when the clf is log-transformed, and the independent variables should be either log-transformed or kept un-transformed; and (c) important catchment characteristics are average runoff, lakes, bogs, catchment length and temperature.

For the development of a low flow map for Norway, several challenges are identified of which the most important are the need for quality control of low flow data, the development of regression equations for several low flow indices for the whole of Norway and the development of automatic procedures to extract the catchment characteristics required.

**Acknowledgements** The study presented is funded by the "Small hydropower" project at the Norwegian Water Resources and Energy Directorate.

## REFERENCES

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

Demuth, S. & Young, A. R. (2004) Regionalisation procedures. In: *Hydrological Drought – Processes and Estimation Methods for Streamflow and Groundwater* (ed. by L. M. Tallaksen & H. A. J. van Lanen), 307–343. Developments in Water Sciences 48. Elsevier Science Publisher, The Netherlands.

Laaha, G. & Blöschl, G. (2006) A comparison of low flow regionalisation methods—catchment grouping. *J. Hydrol.* (in press).

Væringstad, T. & Hisdal, H. (2005) Estimation of common low flow in ungauged catchments (in Norwegian), Miljøbasert vannføring rapport 06-2005. Norwegian Water Resources and Energy Directorate, Oslo, Norway.