

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds

J. E. Shortridge¹, S. D. Guikema², and B. F. Zaitchik³

¹Department of Geography and Environmental Engineering, Johns Hopkins University, Baltimore, USA

²Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, USA

³Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, USA

Received: 25 September 2015 – Accepted: 4 October 2015 – Published: 28 October 2015

Correspondence to: J. E. Shortridge (jshortridge@jhu.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

In the past decade, certain methods for empirical rainfall–runoff modeling have seen extensive development and been proposed as a useful complement to physical hydrologic models, particularly in basins where data to support process-based models is limited. However, the majority of research has focused on a small number of methods, such as artificial neural networks, despite the development of multiple other approaches for non-parametric regression in recent years. Furthermore, this work has generally evaluated model performance based on predictive accuracy alone, while not considering broader objectives such as model interpretability and uncertainty that are important if such methods are to be used for planning and management decisions. In this paper, we use multiple regression and machine-learning approaches to simulate monthly streamflow in five highly-seasonal rivers in the highlands of Ethiopia and compare their performance in terms of predictive accuracy, error structure and bias, model interpretability, and uncertainty when faced with extreme climate conditions. While the relative predictive performance of models differed across basins, data-driven approaches were able to achieve reduced errors when compared to physical models developed for the region. Methods such as random forests and generalized additive models may have advantages in terms of visualization and interpretation of model structure, which can be useful in providing insights into physical watershed function. However, the uncertainty associated with model predictions under climate change should be carefully evaluated, since certain models (especially generalized additive models and multivariate adaptive regression splines) became highly variable when faced with high temperatures.

1 Introduction

Hydrologists and water managers have made use of observed relationships between rainfall and runoff to predict streamflow ever since the creation of the rational method in the 19th century (Beven, 2011). However, the development of increasingly so-

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

context, it has been argued that ANNs are better suited for problems aimed at prediction without any need for model interpretation, rather than those where understanding the process generating predictions and the role of input variables is important (Hastie et al., 2009). Given the importance that this interpretation plays in understanding the contexts in which a hydrologic model is appropriate and reliable, the strong opinions surrounding the use of ANNs for water resource management are perhaps not surprising.

While a number of comparison studies exist that apply multiple empirical models to a given problem, finding generalizable insights from these studies is hindered because of the limited number of models and datasets evaluated. Perhaps the most comprehensive comparison to date is that of Elshorbagy et al. (2010a, b), who compared six methods for data-driven modeling of daily discharge in the Ourthe River in Belgium. This work found that linear models were able to perform comparably to much more complex methods when the data content of the models were limited, or when system input–output behavior was close to linear. However, other studies have demonstrated the value of using more complex approaches when modeling more complex rainfall–runoff behavior (e.g., Abrahart and See, 2007; Asefa et al., 2006). The differing results obtained across these studies indicate that no single method is likely to be suitable for all basins, timescales, or applications. However, it is important to recognize that predictive accuracy alone is not necessarily sufficient justification for applying a model to a given problem. Models should not only be accurate, but also be fit-for-purpose (Beven, 2011; Van Griensven et al., 2012). For instance, accurate representation of low return period flows is more important in a flood forecasting model than one aimed at predicting average amounts of water available for withdrawal and human consumption. Similarly, the ability to provide insights into physical watershed function may be more important in basins where land-use change could alter the hydrologic regime, compared to a basin that is heavily urbanized and expected to remain so. This is particularly true for data-driven models; as pointed out by Solomatine and Ostfield (2008), there is a need for additional research on making models more understandable and

useful for water managers. More comprehensive consideration of model strengths and limitations should be standard practice in model development and selection, rather than simply evaluating global error metrics.

In this work, we compare six methods for empirical streamflow prediction (linear models, generalized additive models, multivariate adaptive regression splines, random forests, M5 model trees and ANNs) in their ability to predict monthly streamflow in five rivers in the Lake Tana basin in Ethiopia. This study region was selected to provide a counterpart to previous comparative studies that have largely focused on rivers in temperate regions. Furthermore, physically-based hydrological process models are a challenge in these basins due to data limitations – soil and vegetation parameters are poorly characterized and high frequency, spatially-distributed precipitation estimates are highly uncertain – and complex hydrodynamic processes, including lake backwater effects, that are neglected by most watershed models. There are, however, relatively long time series of streamflow available, and estimates of historical precipitation and temperature are available at a monthly timescale. These data, combined with information on relevant landscape change, can be leveraged to create reasonably accurate empirical models. Models are compared not only in terms of their predictive accuracy, but also in terms of model error structure and the implications that this structure may have for water resource applications. Additionally, we evaluate the methods by which model structure and predictor variable influence can be evaluated to gain insights into physical system function for each model type. Finally, we assess the suitability of using different model types for climate change impact assessment by comparing model uncertainty in projections made for increasingly extreme climate conditions. The overall objective of this research is not to identify a single “best” model, but rather to highlight some of the strengths and limitations of different approaches, as well as demonstrate important issues that should be kept in mind for model comparisons in the future.

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



2 Methods

2.1 Study area

Lake Tana is located at an elevation of approximately 1800 m in the highlands of north-west Ethiopia (Fig. 1). The catchment draining to the lake encompasses approximately 12 000 km², and the four main tributaries providing water to the lake are the Gilgel Abbay (including its tributary, the Koga River), Ribb, Gumara, and Megech Rivers. Collectively, these rivers account for 93 % of the inflow to the lake (Alemayehu et al., 2010). Ninety percent of rainfall in the basin occurs during the wet season from May until October, and there is significant interannual variability in precipitation with annual rainfall levels ranging from below 1000 mm to over 1800 mm (Achenef et al., 2013). Population growth and expansion of agricultural and pastoral land use in the region has resulted in substantial deforestation and land degradation, with agricultural, pastoral and settled land cover comprising over 70 % of the basin's surface area (Rientjes et al., 2011; Garede and Minale, 2014; Gebrehiwot et al., 2010). There is some evidence that this has impacted the hydrology of the rivers draining into the lake (Gebrehiwot et al., 2010). A summary of basin characteristics for the evaluation period of 1960–2004 is presented in Table 1.

Approximately 2.6 million people live in the basin, and are largely settled in rural areas and reliant on rainfed subsistence agriculture. This makes the region quite vulnerable to climate variability and change, and a number of water resources infrastructure projects are planned to better manage this vulnerability and support economic development (Alemayehu, 2010). To better understand the potential implications of this development, extensive effort has been put towards developing rainfall–runoff models for the Lake Tana basin, as well as other areas of the Ethiopian highlands with similar characteristics (van Griensven et al., 2012). Many of these studies rely on Soil and Water Assessment Tool (SWAT) models, although there are some that use water balance approaches (Van Griensven et al., 2012). While these models have in some cases demonstrated reasonably high accuracy, previous evaluations were largely based on

**Empirical streamflow
simulation for water
resource
management**

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Nash–Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) which can be a flawed performance metric in highly seasonal watersheds (Schaeffli and Gupta, 2007; Legates and McCabe, 1999). More importantly, the limited data available for physical parameterization of these models required a heavy reliance on model calibration, which sometimes resulted in parameterization schemes that are inconsistent with physical understanding of the region’s hydrology (Steenhuis et al., 2009; van Griensven et al., 2012). Furthermore, a number of studies relied on empirical relationships such as curve numbers and the Hargreaves equation that were developed for temperate regions (e.g., Mekonnen et al., 2009; Setegne et al., 2009). While these limitations are likely to introduce considerable uncertainty into model projections, particularly in situations where climatic or environmental conditions differ from those experienced in the calibration period, few studies include any sort of uncertainty analysis in model predictions. Empirical models could provide a useful complement to physical models developed for the region by providing insights into physical system function and allowing for more comprehensive uncertainty analysis.

2.2 Model development

Models were developed using monthly streamflow, climate, and land cover data for the period from 1961 to 2004. In each of the five major rivers in the basin, we developed empirical models that estimated monthly streamflow as a function of climate conditions and agricultural land cover in each basin. Monthly streamflow data was taken from historic stream gauge records for each basin, as reported in feasibility studies developed for proposed irrigation projects (Alemayehu, 2010). Historic data for monthly daily average temperature, monthly total precipitation, and monthly wet days in each river basin were derived from the University of East Anglia Climate Research Unit (CRU) TS3.10 gridded meteorological fields (Harris et al., 2014), which are based on meteorological station observations. Historic estimates of rainfall intensity were also calculated by dividing monthly total precipitation by CRU TS3.10 records of the number of wet days in that month. However, this data was found to be highly correlated with monthly precipi-

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



better match a normal distribution, a requirement for the use of Gaussian linear models and generalized additive models.

In the second formulation, streamflow and climate anomalies were used as the response and predictor variables to better account for the highly seasonal nature of streamflow and precipitation in the region. Streamflow anomalies were calculated for each observation by subtracting the long-term average streamflow for that month (m) from the observed value and dividing this number by the long-term standard deviation of that month's streamflow as in Eq. (2). This procedure was repeated for precipitation and temperature, and these values were then used to fit models of the form described in Eq. (3).

$$Q_{b,t}^{AN} = \frac{Q_{b,t} - \bar{Q}_{b,m}}{SD(Q_{b,m})} \quad (2)$$

$$Q_{b,t}^{AN} = (P_{b,t}^{AN}, P_{b,t-1}^{AN}, P_{b,t-2}^{AN}, T_{b,t}^{AN}, T_{b,t-1}^{AN}, T_{b,t-2}^{AN}, AgLC_{b,t}) + \varepsilon_{b,t} \quad (3)$$

Seven different types of models were compared using each formulation in each basin:

1. A Gaussian linear regression model (GLM) using the basic stats package in the R statistical computing software (R Development Core Team, 2014).
2. Gaussian generalized additive model (GAM): GAMs are a semi-parametric regression approach where the response variable is estimated as the sum of smoothing functions applied over predictor variables. These functions allow the model to capture non-linear relationships between the predictor and response variables without a priori assumptions about the form (e.g., quadratic, logarithmic) of these functions, and are fit using penalized likelihood maximization to prevent model overfitting (Hastie and Tibshirani, 1990). GAMs were fit using the mgcv package in R (Wood, 2011).
3. Multivariate adaptive regression splines (MARS): MARS are a non-parametric regression approach where the response variable is estimated as the sum of basis

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



functions fit to recursively partitioned segments of the data (Friedman, 1991). MARS models were fit using the earth package in R (Milborrow, 2015).

4. Artificial neural network (ANN): ANNs are a non-parametric regression approach represented by a network of nodes and links that connects predictor variables to the response variable. Each link in the network represents a function that maps the input nodes into the output node (Ripley, 1996). ANN models were fit using the nnet package in R (Venables and Ripley, 2013).
5. Random forest (RF): random forests are a rule-based, non-parametric regression approach where the model prediction is created by averaging the predicted value from multiple regression trees which are trained on separate bootstrapped resamples of the data. Each tree is fit using a small, randomly selected subset of predictor variables, resulting in reduced correlation between trees (Breiman, 2001). Random forest models were fit using the randomForest package in R (Liaw and Wiener, 2002).
6. M5 model: M5 models are a rule-based, non-parametric regression approach that fits a linear regression model to each terminal node of a regression tree (Quinlan, 1992). M5 models were fit using the Cubist package in R (Kuhn et al., 2014).
7. Climatology model: a climatology model that simply predicted each month's streamflow as equivalent to the long-term average streamflow for that month was included for comparison purposes.

2.3 Model evaluation

When using non-parametric regression approaches, it is important to avoid overfitting a model to a given dataset because this can result in large errors in out-of-sample predictions (Hastie et al., 2009). To avoid model overfit, the caret package in R (Kuhn, 2015) was used to determine model parameters for the MARS, ANN, RF and M5 models. This package uses resampling to evaluate the effect that model parameters have

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



on the model's predictive performance and chooses the set of parameters that minimizes out-of-sample error (Kuhn, 2015). In this evaluation, 25 bootstrap resamples of the training dataset were generated for each parameter value to be assessed. A model was fit using each bootstrap sample and used to predict the remaining observations, and the parameter values that minimized average RMSE across all resamples. Details on the specific parameters evaluated for each model are presented in Table 1. While the development of more complex structures are possible for some models, this process can result in over-parameterization and poor model performance (Gaume and Gosset, 2003; Han et al., 2007). Additionally, the use of a standardized parameterization procedure allows for a more even comparison between different model types.

The predictive ability of each model was assessed using 50 random holdout cross-validation samples. In each sample, a random selection of years were selected, and observations from these years were removed ("held-out") from the dataset. The size of the held-out sample ranged from 1 to 9 years. Each model was then fit to the remaining portion of the data, using the caret package described above to determine model parameters for the MARS, ANN, RF and M5 models. These models were then used to predict streamflow for the held-out portion of the data, and both the mean absolute error (MAE) and NSE were calculated after transforming model predictions after back to the original streamflow units. While NSE values are acknowledged to be a flawed performance metric in highly seasonal watersheds (Schaefli and Gupta, 2007; Legates and McCabe, 1999), this metric was included to provide a rough comparison of how empirical model performance compared to the performance of physical models developed for the region. Mean MAE and NSE were calculated for each model across the 50 cross-validation samples and used to choose the model with the highest predictive accuracy in each basin. This cross-validation procedure provides a mechanism for evaluating how well a model will generalize to an unseen set of data while avoiding some of the problems that can arise from the use of a single calibration and validation dataset (Elshorbagy et al., 2010a; Han et al., 2007).

**Empirical streamflow
simulation for water
resource
management**

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

As a rough point of comparison for the statistical models developed in this research, we also evaluated discharge estimates derived from a process-based hydrological model. The model used in this application is the Noah Land Surface Model version 3.2 (Noah LSM; Ek et al., 2003; Chen et al., 1996). Noah LSM was implemented for offline simulations of the Lake Tana basin at a gridded spatial resolution of 5 km for the period 1979–2010 using a time step of 30 min. Meteorological forcing was drawn from the Princeton 50 year reanalysis dataset (Sheffield et al., 2006), downscaled to account for Ethiopia’s steep terrain using MicroMet elevation correction equations (Liston and Elder, 2006). The Princeton reanalysis was selected because it provides relatively high resolution meteorological fields, including all variables required to run a water and energy balance LSM like Noah, for the period 1948–present. While higher resolution and possibly higher quality datasets are available for recent years, this longer dataset was utilized to compare the process-based model to statistical models developed for a long historical period. Soil parameters for the Noah simulation were drawn from the FAO global soil database, land use was defined according to the United States Geological Survey (USGS) global 1 km land cover product, and vegetation fraction was derived from MODerate Imaging Spectroradiometer (MODIS) imagery. Land cover was treated as a static parameter over the full length of the simulation, as spatially complete estimates of historical land use were not available at the required resolution and specificity.

The highest performing model in each basin based on MAE was retained for more detailed evaluation of model error structure, covariate influence, and uncertainty in climate change sensitivity analysis. To generate a complete time-series of out-of-sample model predictions for error analysis, the holdout cross validation procedure was repeated for the highest performing standard-formulation and anomaly-formulation models for each basin, but this time holding out a single year of observations in each iteration. The predictions from this cross validation were used to evaluate the how model error structure might impact model predictions used for water resource applications. The influence of different predictor variables on model predictions was also assessed for the highest performing model in each basin after being fit to the complete dataset.

Each predictor variable was assessed using metrics for covariate importance and influence that are unique to that model type, demonstrating how models could be used to gain physical insights about data-scarce regions and the mechanisms for generating these insights for each type of model. Partial dependence plots (Hastie et al., 2009) were also generated for each covariate for the highest performing model in each basin to provide insights about how covariate influence compared across different basins and model types.

Finally, two evaluations were conducted to assess uncertainty in model projections of streamflow under increasingly extreme climate conditions to better understand the implications of using different model formulations for climate change impact studies. Model projections of streamflow in different climate conditions are likely to be accompanied by considerable uncertainty, particularly when climate conditions exceed those experienced historically. To assess this uncertainty, the best performing model in each basin was used to generate streamflow predictions for (1) changes in temperature from 0 to 5 °C, (2) changes in precipitation from -30 to +30 %, (3) an increase in temperature to 5 °C combined with a decrease in precipitation to -30 %, and (4) an increase in temperature to 5 °C combined with an increase in precipitation to +30 %. For each of the four assessments, the models generated predictions for the 45 year historic climate record adjusted for a given degree of climate change using the delta-change method (Gleick, 1986), while holding agricultural land cover constant at 60 %. Model predictions for the altered climate record were then used to calculate the average annual streamflow in each river. These should not be interpreted as a prediction of actual climate change impacts, but rather a measurement of the sensitivity of streamflow in the basin to different climate conditions. This process was repeated 100 times for models fit on random bootstrap resamples of the historic dataset to generate uncertainty bounds surrounding model predictions and evaluate how the uncertainty in these predictions increased as climate conditions became more extreme.

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

3 Results

3.1 Model accuracy and error structure

5 Table 2 shows the out-of-sample cross validation errors for each model assessed in each basin. The random forest model had the lowest mean absolute error for the standard-formulation model in four of the five basins, with the M5 model performing
10 best for the Koga basin. These models outperformed the Noah LSM simulations in all basins assessed. The Noah LSM errors are for a single period of analysis and thus don't present an exact corollary to the cross validation performed for the empirical models. Nevertheless, the significant increases in errors associated with the Noah
15 LSM model demonstrates the difficulty associated with the use of process-based models in the region, particularly when relying on global datasets that may be unreliable at the spatial and temporal resolutions required for physical modeling. Physical models developed for monthly streamflow prediction in other basins within the Ethiopian highlands have reported NSE values ranging from 0.53 to 0.92 (van Griensven et al.,
20 2012), compared to values ranging from 0.71 to 0.87 for the random forest models developed here. If this measure alone was used for model evaluation, these empirical models would generally be classified as having good performance based on the guidelines suggested by Moraisi et al. (2007). However, the climatology model outperforms the best standard formulation models in all basins except Megech, indicating that in the majority of basins the errors from the fitted empirical models are higher than those that result from simply using the long-term monthly average for each month's prediction. This demonstrates how high NSE values can be quite easy to obtain in seasonal basins.

25 Evaluation of anomaly model errors indicates that the models using this formulation achieve better predictive accuracy than those using the standard formulation, and are able to outperform the climatology model based on both NSE and MAE in all basins. However, the highest performing models in each basin varies more when the anomaly formulation is used, with the GLM, GAM, random forest, and M5 models all minimizing

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



MAE in different basins. In all basins except Koga, the highest performing model significantly outperformed the climatology model based on paired Wilcoxon rank-sum tests (Bonferroni-corrected p value < 0.01).

Further exploration of model residuals indicates another important advantage of using the anomaly model formulation. In the standard model formulation, model residuals appear to be non-random. Example autocorrelation plots are shown for the Gilgel Abbay and Ribb Rivers in Fig. 2, and demonstrate that a positive autocorrelation exists at the 12 month time lag. For brevity, only plots for two rivers are shown, although this autocorrelation existed in the standard-formulation models for all basins except Megech (Table 3). This autocorrelation occurs because the standard-formulation models consistently underestimate wet-season streamflow while overestimating dry-season flows, as is apparent in hydrographs of observed and predicted streamflow (Fig. 3). Because wet-season flows contribute such a large portion of the total annual flow volume, this results in regular underestimation of aggregate values such as mean annual flow (Table 3). This autocorrelation is reduced in the anomaly-formulation models, meaning that they are better able to capture the peak flow volumes experienced in the wet season and do not underestimate mean annual flow to the same degree that the standard formulation models do.

3.2 Model structure and covariate influence

Evaluating the relationship between predictor covariates and streamflow response can lend insight into the physical processes underlying runoff generation in each basin. There are two components of this relationship that can be evaluated: how much each covariate contributes to model accuracy (covariate importance), and the direction and nature of the relationship between covariate values and model response (covariate influence). In many machine-learning models, complete description of the all of the mathematical relationships within the model (for instance, through description of each tree comprising a random forest model) is infeasible, requiring the use of other mechanisms for understanding covariate importance and influence. However, because each

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	⏩
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	

model type is structured in a different way, these mechanisms differ. This section first describes the mechanisms available for obtaining insights about covariate influence in each of the highest performing models. To provide a mechanism for comparing results across different basins, each basin model is then assessed using the general approach of partial dependence plots.

In the Gilgel Abbay and Koga basins, the highest performing model was a simple linear regression model. These models can be evaluated by reviewing model coefficients and associated p values, as shown in Table 4. In a standard linear regression, model coefficients can be interpreted as the mean change in the response variable that results from a unit change in that covariate when all others are held constant. These coefficients are for streamflow anomalies rather than raw values, making their immediate interpretation less intuitive. For instance, in the Gilgel Abbay model an increase of one standard deviation in precipitation results in an increase of 0.22 standard deviations in flow. The associated p value for each coefficient evaluates a null hypothesis that the true coefficient value is equal to zero given the other covariates in the model, and thus has no influence on the response variable.

Evaluating model structure based on regression coefficients is appealing due to their simplicity and familiarity. However, it is important to keep in mind that the above interpretations rely on specific assumptions regarding model error distributions. Examination of fitted model residuals from both basins indicate that errors are autocorrelated in the Koga basin and not normally distributed due to the presence of outliers in both basins. Non-normality and autocorrelation both impact the t statistics and f statistics used to test for the significance of model coefficients, and thus the p values for these models are likely biased (Montgomery et al., 2012).

Interpretation of variable influence in GAMs is based on the estimated degrees of freedom (EDF) a covariate's smoothing function $s(X_i)$ uses within a model (Hastie and Tibushini, 1986). An EDF value of one or below indicates a linear function relating the response variable to that covariate, while values greater than one represent a non-linear smoothing function. An EDF value of zero indicates that the covariate smoothing



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



function is penalized to zero (meaning it has no influence on model predictions). In the model for the Megech River, the terms for lagged temperature at one and two months, as well as precipitation lagged at two months were all smoothed to zero. Of the remaining covariates, lagged precipitation has a linear impacts on model response, while precipitation, temperature and land cover have non-linear impacts. Smoothing functions can be plotted to gain more insight about these relationships (Fig. 4). The functions for precipitation anomaly, lagged (one month) precipitation anomaly, and agricultural land cover show a positive relationships with streamflow, while the function for temperature anomaly predicts low streamflow at both high and low anomalies.

ρ values test the null hypothesis that a covariate's smoothing function is equal to zero, but rest on the assumption that model residuals are homoscedastic and independent (Wood, 2012). Similar to the linear models, residuals in the Megech GAM model appear to be both autocorrelated and heteroscedastic, meaning that a formal statistical interpretation of this value may be inappropriate and that confidence bounds around smoothing functions might be misleading.

The M5 cubist model fit for the Gumara basin is an ensemble of 100 small M5 regression trees. In each tree, the model splits observations based on logical rules related to one or more covariates and fits a linear regression model to each set of observations. The final model prediction is the average across all of the individual trees. Using this sort of ensemble approach can reduce model variance and improve accuracy if the individual trees are unbiased, uncorrelated predictors (Breiman, 1996). This can be useful in avoiding models that are overfit to the data, but can reduce model interpretability since direct visualization of model structure becomes impractical as the number of trees increases. However, the frequency with which individual covariates are used as splitting points within trees and as regression coefficients can provide some insights about covariate importance (Table 4; note that because multiple covariates can be used for rules and linear models, these don't necessarily add to 100 %). Model rules were largely based on land cover, with some rules based on precipitation. These two covariates were also used most frequently in linear regressions at model nodes,

followed by temperature (current and 1 month lag) and 1 month lagged precipitation. Notably, climate data from 2 months lagged was not used at all. While this can be useful in identifying which covariates have the largest impact on model predictions, it does not provide any information regarding the nature or direction of that influence.

Similarly, the random forest model developed for the Ribb basin is an ensemble of regression trees in which the final model prediction is the average of the predictions from each individual tree. However, random forests use standard regression trees that do not incorporate linear regression models at terminal nodes. Variable importance within the final model is measured by recording the increase in out-of-sample MSE that results when a covariate is randomly permuted for each tree in the ensemble. This increase in error is then averaged across all trees in the ensemble. In our model, the largest increases in error resulted from permutation of land cover and temperature, followed by 2 month lagged temperature and precipitation. Covariate influence can be evaluated through the use of partial dependence plots, which measure the change in model predictions that result from changing the value of one parameter while leaving all other covariates constant (Hastie et al., 2009). Partial dependence plots indicate that model predictions of streamflow are higher when the percent of agricultural land cover is greater than approximately 75 %, when temperatures anomalies are low, and when precipitation anomalies are high. However, it appears that the plot for lagged temperature might be sensitive to outliers at high temperature anomalies as evidenced by the large increase that occurs above an anomaly of +2, in a region where very few data points are present.

Many of the measures used to evaluate covariate importance and influence are model specific, making inter-basin and inter-model comparisons difficult. However, the partial dependence plots used in the randomForest R package can be developed for any model and provide a mechanism for comparing the influence that covariates have in the different models and basins (Shortridge et al., 2015). Partial dependence plots were generated for each basin's best performing model and results are shown for climatic variables in Fig. 6. As expected, models generally respond positively to increases

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



in precipitation and negatively to increases in temperature, with the greatest influence in the current month and decreasing influence at one and two months prior. The influence of the current month's precipitation is linear in three of the five basins; while this is constrained to be the case in the Gilgel Abbay and Koga basins due to the use of a linear model, the linear response in Gumara is not required from the M5 model structure. Interestingly, both Megech and Ribb demonstrate a linear response to negative precipitation anomalies, but little response to positive anomalies. Streamflow response to temperature is strongest in the Gumara basin; interestingly, this is the basin with the smallest response to precipitation.

The partial dependence plots for the percentage of the basin classified as agricultural land cover indicates a positive relationship between agricultural land cover and streamflow in all basins except for the Gilgel Abbay (Fig. 7). This would be expected if deforestation had contributed to a decrease in evapotranspiration in the contributing watersheds. The exact nature of this response differs across the different rivers, with the relatively minor responses in Koga and Ribb, and much stronger responses in the Gumara and Megech basins. However, this plot also demonstrates some of the limitations associated with different model structures. The plot for Gumara is highly erratic, indicating that the M5 model might be overfit to the training dataset, despite the use of model averaging to reduce model variance. Additionally, the GAM used in the Megech basin was only trained on agricultural land cover values up to 77 %; while this model may be accurately representing the impact of land cover changes within this range, extrapolating this relationship to higher values leads to predictions that may not be physically realistic.

3.3 Climate change sensitivity and uncertainty assessment

Figure 8 shows the results of the climate change sensitivity analysis for total flow from all five tributaries, with dashed lines representing 95 % confidence intervals obtained through 100 bootstrapped resamples of the data set. As would be expected, increasing temperature independently of precipitation results in decreasing total flows while

increasing precipitation results in higher flows. However, the uncertainty surrounding temperature sensitivity increases at higher changes in temperature, while the uncertainty surrounding precipitation sensitivity remains relatively constant, even at extreme changes in annual precipitation. The bottom panels of the figure show the sensitivity of total inflows to concurrent changes in temperature and precipitation. Unsurprisingly, decreasing precipitation combined with higher temperatures results in greater decreases in total flow than when temperature and precipitation are varied independently. However, even if temperature increases are combined with higher precipitation, total flows decline in the majority of bootstrap resamples.

The uncertainty surrounding temperature sensitivity is a key limitation to using data-driven approaches for climate impact assessment. To better understand which models and basins are contributing to this uncertainty, Fig. 9 shows how the coefficient of variation (the standard deviation of predictions from all bootstrap samples divided by the mean of these predictions) varies as a function of temperature change in each basin. From this figure, it is apparent that the Megech model is by far the largest contributor to model uncertainty; however, it is not clear whether this contribution is due to model structure (the GAM model used for the Megech River) or characteristics associated with the basin itself. To investigate how different model structures contributed to this uncertainty, the bootstrap resampling procedure was used to assess uncertainty in streamflow predictions in the Gumara River from all model types. This basin was chosen because all six models were able to outperform the climatology model, and thus could be considered good choices for model selection based on predictive accuracy alone. The results indicate that the increase in uncertainty is highest, and increases non-linearly, in the GLM, GAM, and MARS models. Uncertainty increases more slowly in the ANN and M5 models, and no noticeable increase in uncertainty is apparent in the random forest model.

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

4 Discussion

The objective of this study was not to identify the “best” approach for empirical rainfall–runoff modeling, as this is likely to be highly specific to the basin and problem to which a model is applied. However, we hope that the comparison conducted here can highlight some of the strengths and limitations of different approaches, as well as demonstrate some important issues that should be kept in mind for model comparisons in the future. One important finding was the limitation with using NSE as an error metric. Our results confirm previous studies that found that even uninformative models able to capture basic seasonality are able to achieve high NSE values (Legates and McCabe, 1999; Schaefli and Gupta, 2007), and provide further evidence indicating that high NSE values should be considered a necessary but not sufficient requirement for model usage in planning situations. In particular, understanding error structure can be valuable in evaluating whether model biases might undermine the model’s suitability for management activities. In our example, the autocorrelation present in the standard-formulation models meant that these models were consistently underestimating wet-season flows, resulting in low estimates of the total annual flow in the rivers. Since multiple reservoirs are planned for construction on these rivers to support irrigation activities, this bias could lead to poor estimates of how much water is available for agricultural use in the short term (i.e., seasonal forecasting) and long-term (due to climate change). Interestingly, difficulties in accurately capturing high flows has been observed in physical hydrologic models for Ethiopia (e.g., Setegne et al., 2011; Mekonnen et al., 2009) and more generally (e.g., Wilby, 2005). The implications of this limitation should be carefully evaluated before using models for water resource planning or (more importantly) flood risk evaluation.

Depending on the model type used, different mechanisms are available to evaluate covariate importance and influence within the model. This evaluation can be useful in both confirming that the model is replicating physically realistic relationships between input and output variables, and in characterizing watershed behavior in a manner that

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



could shed light on underlying physical processes. The easy manner in which covariate relationships within the GAM and random forest models can be visualized using a single command within their respective R packages is a strong advantage to these approaches compared to methods such as M5 model trees and artificial neural networks. Of course, partial dependence plots can be developed for any model type (as was done in this research), but code must be written by the user and thus requires a higher degree of effort than is necessary for in-package functions. A downside to most machine-learning models is that they do not support the statistical formalism in assessing variable importance that is possible when linear models and GAMs are used. However, this formalism often rests on assumptions regarding model residuals that are unlikely to be met in many hydrologic models (Sorooshian and Dracup, 1980).

Within the Lake Tana basin, evaluation of covariate influence indicates that each basin's model is performing in a physically realistic manner, with a runoff increasing with higher precipitation levels and decreasing with higher temperatures. The influence of precipitation and temperature is greatest in the current month, and progressively declines to a very small influence after two months. This suggests that long-term (multi-month) storage does not significantly contribute to variability in flow volumes. One interesting finding is the non-linear relationship between concurrent month precipitation and runoff that exists in the Megech and Ribb basins, which suggests that above a certain point increasing rainfall does not result in a commiserate increase in streamflow. Other studies have noted the dampening effect that wetlands and floodplains have had on river flows in the region (Dessie et al., 2014; Gebrehiwot et al., 2010); this phenomenon could explain the non-linear relationship identified in this work. The clearly negative relationship between temperature and runoff demonstrates the degree to which upstream evapotranspiration impacts streamflow and suggests that evapotranspiration is largely energy-limited, rather than water-limited. Increasing agricultural land-use appears to be associated with higher runoff in all rivers except for Gilgel Abbay (where no clear relationship between land cover and runoff was observed), and suggests that agricultural expansion at the expense of forest cover has reduced the evaporative component of

the water balance in these basins. Finally, the relative performance of different model formulations themselves can also be informative. For instance, the improved performance of the anomaly-formulation models indicates that the relationship between precipitation and runoff varies throughout the year and could point towards differences in runoff-generating mechanisms in the wet and dry seasons that have been observed in other case studies (Wilby, 2005).

One limitation with data-driven approaches for streamflow prediction is that the relationships they model can only generate reliable predictions for conditions that are comparable to those experienced historically. Using these models to generate predictions for conditions that exceed historic variability is likely to introduce considerable uncertainty into their projections. Our results indicate that uncertainty in projections of streamflow under changing precipitation is relatively constant, whereas uncertainty increases markedly in projections of streamflow under increasing temperature. This result is not surprising when one considers the basin's climate, which is characterized by highly variable rainfall but fairly consistent temperatures (Table 5). A temperature increase of 3 °C equates to almost two standard deviations beyond historic variability, whereas a change in precipitation of 30 % is well within the range of conditions experienced historically. One would expect that in other climates (for example, temperate watersheds with only minor changes in rainfall throughout the year), this relationship could be reversed. Despite the uncertainty that exists in projections of streamflow under changing temperature, total annual flow appears to be quite sensitive to increasing temperatures. In fact, the decreases in streamflow due to increasing temperature appears likely to be more than enough to counteract any increases in streamflow resulting from higher precipitation that is projected for the region in some global circulation models (GCMs). This is consistent with the work of Setegne et al. (2011), who used projections from multiple GCMs as input for a SWAT model developed for the region and found that streamflow decreased in the majority of emissions scenarios and models, even when precipitation increased. Unfortunately, this suggests that any hopes for a

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



“windfall” of additional water to support agriculture and hydropower in the region under climate change may be unfounded.

Repeating the climate change sensitivity experiment with multiple models fit to the Gumara watershed indicated that the MARS, GAM, and linear models all result in the largest increase in uncertainty at high temperatures. This indicates that when models are fit to slightly different bootstrap resamples of the historic dataset, the projected changes in streamflow at high temperature changes can be highly erratic. This is likely due to the fact that extrapolating the relationships that are observed between historic temperature and streamflow to higher temperatures can lead to very large changes in streamflow. Fitting the models to bootstrap resamples of the data results in minor changes to these relationships that can result in widely varying projections when the models are used to predict streamflow at higher temperatures, particularly when these relationships are nonlinear (as in the GAM). At the other end of the spectrum, the random forest model exhibits almost no increase in uncertainty at high temperatures, meaning that projections of streamflow at high temperatures are consistent across the bootstrap resamples. This is likely the result of the random forest model structure. The predicted value for each of a regression tree’s terminal nodes is the average of all observations that meet the conditions described for that node. Thus, the model will not predict values beyond those experienced historically, even if covariate values exceed those contained within the historic dataset. Thus, this model is likely to underestimate the change in streamflow that results from increasing temperatures.

5 Conclusions

In this work, we compared multiple methods for data-driven rainfall–runoff modeling in their ability to simulate streamflow in five highly-seasonal watersheds in the Ethiopian highlands. Despite the popularity of ANNs in research on streamflow prediction to date, ANNs were not found to be the most accurate model in any of the five basins evaluated. Other methods, in particular GAMs and random forests, are able to capture non-linear

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



relationships effectively and lend themselves to simpler visualization of model structure and covariate influence, making it easier to gain insights on physical watershed functions and confirm that the model is operating in a physically realistic manner. However, it is important to carefully evaluate model structure and residuals, as these can contribute to biased estimates of water availability and uncertainty in estimating sensitivity to potential future changes in climate. In particular, autocorrelation in model residuals can result in underestimation of aggregate metrics such as annual flow volumes, even in models with high NSE performance. Uncertainty in GAM projections was found to rapidly increase at high temperatures, whereas random forest projections may be underestimating the impact of high temperatures on river flows. Thorough consideration of this uncertainty and bias is important any time that models are used for water planning and management, but especially crucial when using such models to generate insights about future streamflow levels. By considering multiple model formulations and carefully assessing their predictive accuracy, error structure and uncertainties, these methods can provide an empirical assessment of watershed behavior and generate useful insights for water management and planning. This makes them a valuable complement to physical models, particularly in data-scarce regions with little data available for model parameterization, and warrants additional research into their development and application.

Acknowledgements. We would like to gratefully acknowledge the Ethiopian Ministry of Water and Energy, the Tana Sub Basin Organization, and the International Water Management Institute for making available the data used to perform this analysis. All data for this paper is properly cited and referred to in the reference list. The source code for the models developed in this study is available from the authors upon request. Empirical modeling work was supported by a National Defense Science and Engineering Graduate Fellowship and by National Science Foundation Grant 1069213 (IGERT). Noah LSM simulations presented here were performed under NASA Applied Sciences Program grant NNX09AT61G. This research was conducted while Guikema was affiliated with the Department of Geography and Environmental Engineering at Johns Hopkins University. This support is gratefully acknowledged. Any opinions, find-

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

ings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sources.

References

- Abraham, R. J. and See, L. M.: Neural network modelling of non-linear hydrological relationships, *Hydrol. Earth Syst. Sci.*, 11, 1563–1579, doi:10.5194/hess-11-1563-2007, 2007.
- Acheneb, H., Tilahun, A., and Molla, B.: Tana Sub Basin Initial Scenarios and Indicators Development Report, Tana Sub Basin Organization, Bahir Dar, Ethiopia, 2013.
- Alemayehu, T., McCartney, M., and Kebede, S.: The water resource implications of planned development in the Lake Tana catchment, Ethiopia, *Ecohydrol. Hydrobiol.*, 10, 211–221, doi:10.2478/v10104-011-0023-6, 2010.
- Asefa, T., Kemblowski, M., McKee, M., and Khalil, A.: Multi-time scale stream flow predictions: the support vector machines approach, *J. Hydrol.*, 318, 7–16, doi:10.1016/j.jhydrol.2005.06.001, 2006.
- Beven, K. J.: *Rainfall–Runoff Modelling: The Primer*, John Wiley and Sons, New York, 2011.
- Breiman, L.: Bagging predictors, *Mach Learn*, 24, 123–140, doi:10.1007/BF00058655, 1996.
- Breiman, L.: Random forests, *Mach Learn*, 45, 5–32, 2001.
- Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M., and Betts, A.: Modeling of land surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, 101, 7251–7268, doi:10.1029/95JD02165, 1996.
- Dessie, M., Verhoest, N. E. C., Admasu, T., Pauwels, V. R. N., Poesen, J., Adgo, E., Deckers, J., and Nyssen, J.: Effects of the floodplain on river discharge into Lake Tana (Ethiopia), *J. Hydrol.*, 519, 699–710, doi:10.1016/j.jhydrol.2014.08.007, 2014.
- Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarp-ley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, 108, 8851, doi:10.1029/2002JD003296, 2003.
- Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology – Part 1: Concepts and methodology, *Hydrol. Earth Syst. Sci.*, 14, 1931–1941, doi:10.5194/hess-14-1931-2010, 2010a.

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Elshorbagy, A., Corzo, G., Srinivasulu, S., and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology – Part 2: Application, *Hydrol. Earth Syst. Sci.*, 14, 1943–1961, doi:10.5194/hess-14-1943-2010, 2010b.
- Friedman, J. H.: Multivariate adaptive regression splines, *Ann. Stat.*, 1–67, 1991.
- 5 Galelli, S. and Castelletti, A.: Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling, *Hydrol. Earth Syst. Sci.*, 17, 2669–2684, doi:10.5194/hess-17-2669-2013, 2013.
- Garede, N. M. and Minale, A. S.: Land use/cover dynamics in ribb watershed, North Western Ethiopia, *J. Nat. Sci. Res.*, 4, 9–16, 2014.
- 10 Gaume, E. and Gosset, R.: Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology?, *Hydrol. Earth Syst. Sci.*, 7, 693–706, doi:10.5194/hess-7-693-2003, 2003.
- Gebrehiwot, S. G., Taye, A., and Bishop, K.: Forest cover and stream flow in a headwater of the Blue Nile: complementing observational data analysis with community perception, *Ambio*, 39, 284–294, doi:10.1007/s13280-010-0047-y, 2010.
- 15 Han, D., Kwong, T., and Li, S.: Uncertainties in real-time flood forecasting with neural networks, *Hydrol. Process.*, 21, 223–228, doi:10.1002/hyp.6184, 2007.
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34, 623–642, doi:10.1002/joc.3711, 2014.
- 20 Hastie, T. and Tibshirani, R.: Generalized additive models, *Stat. Sci.*, 1, 297–310, 1986.
- Hastie, T. and Tibshirani, R.: *Generalized Additive Models*, Chapman, Hall, London, 1990.
- Hastie, T., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd Edn., Springer, New York, 2009.
- 25 Iorgulescu, I. and Beven, K. J.: Nonparametric direct mapping of rainfall–runoff relationships: an alternative approach to data analysis and modeling?, *Water Resour. Res.*, 40, W08403, doi:10.1029/2004WR003094, 2004.
- Jain, A., Sudheer, K. P., and Srinivasulu, S.: Identification of physical processes inherent in artificial neural network rainfall runoff models, *Hydrol. Process.*, 18, 571–581, doi:10.1002/hyp.5502, 2004.
- 30 Kuhn, M.: caret: Classification and regression training, available at: <http://CRAN.R-project.org/package=caret> (last access: 09/06/2015), 2015.

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

- Kuhn, M., Weston, S., Keefer, C., and Coulter, N.: Cubist: rule- and instance-based regression modeling, available at: <http://CRAN.R-project.org/package=Cubist> (last access: 09/06/2015), 2014.
- Legates, D. R. and McCabe Jr, G. J.: Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, 35, 233–241, 1999.
- Liaw, A. and Wiener, M.: Classification and regression by randomForest, *R News*, 2, 18–22, 2002.
- Lin, J.-Y., Cheng, C.-T., and Chau, K.-W.: Using support vector machines for long-term discharge prediction, *Hydrolog. Sci. J.*, 51, 599–612, doi:10.1623/hysj.51.4.599, 2006.
- Liston, G. E. and Elder, K.: A meteorological distribution system for high-resolution terrestrial modeling (MicroMet), *J. Hydrometeorol.*, 7, 217–234, doi:10.1175/JHM486.1, 2006.
- Mekonnen, M. A., Wörman, A., Dargahi, B., and Gebeyehu, A.: Hydrological modelling of Ethiopian catchments using limited data, *Hydrol. Process.*, 23, 3401–3408, doi:10.1002/hyp.7470, 2009.
- Milborrow, S.: earth: Multivariate Adaptive Regression Splines, available at: <http://CRAN.R-project.org/package=earth> (last access: 09/06/2015), 2015.
- Montgomery, D. C., Peck, E. A., and Vining, G. G.: *Introduction to Linear Regression Analysis*, John Wiley and Sons, Hoboken, NJ, 2012.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *T. Asabe*, 50, 885–900, 2007.
- Quinlan, J. R.: *Learning with Continuous Classes*, in: *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, Singapore, 1992.
- R Development Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org> (last access: 09/06/2015), 2014.
- Rientjes, T. H. M., Haile, A. T., Kebede, E., Mannaerts, C. M. M., Habib, E., and Steenhuis, T. S.: Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin – Ethiopia, *Hydrol. Earth Syst. Sci.*, 15, 1979–1989, doi:10.5194/hess-15-1979-2011, 2011.
- Ripley, B. D.: *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- Schaeffli, B. and Gupta, H. V.: Do Nash values have value?, *Hydrol. Process.*, 21, 2075–2080, doi:10.1002/hyp.6825, 2007.

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



See, L., Solomatine, D., Abrahart, R., and Toth, E.: Hydroinformatics: computational intelligence and technological developments in water science applications – Editorial, *Hydrolog. Sci. J.*, 52, 391–396, doi:10.1623/hysj.52.3.391, 2007.

Setegn, S. G., Srinivasan, R., Melesse, A. M., and Dargahi, B.: SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia, *Hydrol. Process.*, doi:10.1002/hyp.7457, 2009.

Setegn, S. G., Rayner, D., Melesse, A. M., Dargahi, B., and Srinivasan, R.: Impact of climate change on the hydroclimatology of Lake Tana Basin, Ethiopia, *Water Resour. Res.*, 47, W04511, doi:10.1029/2010WR009248, 2011.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, *J. Climate*, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Shortridge, J. E., Falconi, S. M., Zaitchik, B. F., and Guikema, S. D.: Climate, agriculture, and hunger: statistical prediction of undernourishment using nonlinear regression and data-mining techniques, *J. Appl. Stat.*, 42, 2367–2390, doi:10.1080/02664763.2015.1032216, 2015.

Solomatine, D. P. and Ostfeld, A.: Data-driven modelling: some past experiences and new approaches, *J. Hydroinf.*, 10, 3–22, doi:10.2166/hydro.2008.015, 2008.

Sorooshian, S. and Dracup, J. A.: Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: correlated and heteroscedastic error cases, *Water Resour. Res.*, 16, 430–442, doi:10.1029/WR016i002p00430, 1980.

Steenhuis, T. S., Collick, A. S., Easton, Z. M., Leggesse, E. S., Bayabil, H. K., White, E. D., Awulachew, S. B., Adgo, E., and Ahmed, A. A.: Predicting discharge and sediment for the Abay (Blue Nile) with a simple model, *Hydrol. Process.*, 23, 3728–3737, doi:10.1002/hyp.7513, 2009.

Sudheer, K. P. and Jain, A.: Explaining the internal behaviour of artificial neural network river flow models, *Hydrol. Process.*, 18, 833–844, doi:10.1002/hyp.5517, 2004.

van Griensven, A., Ndomba, P., Yalaw, S., and Kilonzo, F.: Critical review of SWAT applications in the upper Nile basin countries, *Hydrol. Earth Syst. Sci.*, 16, 3371–3381, doi:10.5194/hess-16-3371-2012, 2012.

Venables, W. N. and Ripley, B. D.: *Modern Applied Statistics with S-PLUS*, Springer Science and Business Media, New York, 2013.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, *Hydrol. Process.*, 19, 3201–3219, doi:10.1002/hyp.5819, 2005.

Wood, S. N.: *Generalized Additive Models: an Introduction with R*, CRC Press, Boca Raton, FL, 2006.

- 5 Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *J. R. Stat. Soc. B*, 73, 3–36, doi:10.1111/j.1467-9868.2010.00749.x, 2011.

Wood, S. N.: On p-values for smooth components of an extended generalized additive model, *Biometrika*, 100, 221–228, doi:10.1093/biomet/ass048, 2012.

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)**Table 1.** Study basin characteristics over the evaluation period of 1961 to 2004.

Basin	Drainage area above gauge (km ²)	Average annual streamflow at gauge (MCM)	Standard deviation of annual streamflow (MCM)	Average temp (°C)	Average monthly rainfall [mm]	
					May–Oct	Nov–Apr
Gilgel Abbay	2664	1883	217	15.7	206	39.3
Gumara	385	236	71	17.7	186	29
Koga	200	114	31	15.7	206	39.3
Megech	424	172	66	20.6	234	41.4
Ribb	677	210	83	18.2	263	45.8

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)**Table 2.** Model parameters evaluated through cross validation.

Model type	R package	Parameters evaluated
GLM	stats	NA
GAM	mgcv	$k = 3$ (for each predictor)
MARS	earth	degree = {1, 2, 3} nprune = {5, 10, 15, 20, 25}
ANN	nnet	size = {1, 2, 4, 8, 20} decay = {0.0, 0.1, 0.5, 1.0, 2.0}
RF	randomForest	mtry = {2, 3, 4, 5, 6, 7}
M5	Cubist	committees = {10, 50, 100} neighbors = {0, 5, 9}

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 3. Cross validation errors for each assessed model.

Standard Formulation		GLM	GAM	MARS	RF	M5	ANN	Climatology	Noah LSM
MAE	Gilgel Abbay	30.78	18.54	16.75	14.89	15.11	17.22	10.42	28.11
	Gumara	4.29	3.41	3.28	2.67	2.96	3.15	2.57	3.95
	Koga	1.50	1.30	1.38	1.20	1.17	1.23	1.06	1.97
	Megech	4.45	2.64	2.83	2.37	2.53	3.04	2.54	4.09
	Ribb	4.69	2.98	3.50	2.97	3.27	3.17	2.81	7.01
NSE	Gilgel Abbay	-0.02	0.81	0.83	0.87	0.86	0.84	0.95	0.59
	Gumara	0.04	0.51	0.61	0.80	0.66	0.70	0.81	0.48
	Koga	0.45	0.71	0.65	0.76	0.77	0.76	0.83	0.25
	Megech	-1.85	0.63	0.46	0.73	0.65	0.52	0.71	0.41
	Ribb	-1.14	0.71	0.39	0.71	0.31	0.67	0.73	-0.75
Anomaly Formulation		GLM	GAM	MARS	RF	M5	ANN	Climatology	Noah LSM
MAE	Gilgel Abbay	9.73	9.82	10.10	10.12	9.94	9.79	10.42	28.11
	Gumara	2.22	2.25	2.43	2.23	2.16	2.22	2.57	3.95
	Koga	1.03	1.06	1.08	1.09	1.05	1.05	1.06	1.97
	Megech	2.49	2.48	2.63	2.66	2.69	2.50	2.54	4.09
	Ribb	2.79	2.76	2.84	2.70	2.78	2.77	2.81	7.01
NSE	Gilgel Abbay	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.59
	Gumara	0.85	0.85	0.82	0.85	0.86	0.86	0.81	0.48
	Koga	0.83	0.82	0.81	0.81	0.82	0.82	0.83	0.25
	Megech	0.73	0.72	0.65	0.66	0.61	0.72	0.71	0.41
	Ribb	0.73	0.75	0.72	0.75	0.73	0.74	0.73	-0.75

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Table 4. Residual autocorrelation factors at a 12 month lag for the standard formulation and anomaly formulation models, and resulting mean annual observed and predicted flow.

	Autocorrelation factors		Mean annual flow (MCM)		
	Standard	Anomaly	Observed	Standard	Anomaly
Gilgel	0.33	0.11	22 925	20 703	22 958
Gumara	0.29	0.07	2870	2392	2734
Koga	0.04	0.10	1383	1333	1386
Megech	0.05	0.04	2035	1637	2028
Ribb	0.21	−0.01	2575	1969	2615

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Table 5. Covariate importance measurements from each basin’s model.

Model type	Linear model				Generalized additive model		M5 model tree		Random forest	
Measure of influence	Linear regression coefficients and associated p values				Estimated degrees of freedom (EDF) and associated p values		Covariate usage in tree rules and model coefficients		Increase in MSE when covariate is randomly permuted	
Basin	Gilgel Abbay		Koga		Megech		Gumara		Ribb	
Covariate	Coefficient estimate	p value	Coefficient estimate	p value	EDF	p value	Tree rules	Model coefficients	Percent increase in MSE	
Prec	0.22	< 0.01	0.24	< 0.01	1.346	< 0.01	5%	58%	7.71%	
Prec (lag 1)	0.10	0.03	0.16	< 0.01	0.624	0.08	0%	19%	2.79%	
Prec (lag 2)	0.01	0.74	0.05	0.26	0	0.29	0%	0%	1.10%	
Temp	-0.09	0.08	-0.07	0.17	1.023	0.07	0%	47%	12.74%	
Temp (lag 1)	-0.04	0.49	-0.06	0.22	0	0.32	0%	46%	4.97%	
Temp (lag 2)	-0.01	0.81	-0.09	0.08	0	0.56	0%	0%	8.16%	
Agr. LC	0.00	0.33	0.02	0.01	1.986	< 0.01	86%	73%	15.21%	

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Table 6. Mean and standard deviation values for temperature, wet-season rainfall, and dry-season rainfall in each basin.

	Temperature (°C)		Wet season rainfall (mm month ⁻¹)		Dry season rainfall (mm month ⁻¹)	
	Mean	SD	Mean	SD	Mean	SD
Gilgel Abbay	15.7	1.54	206	145	39.3	56.5
Gumara	17.7	1.55	186	137	29.0	43.6
Koga	15.7	1.54	206	145	39.3	56.5
Megech	20.6	1.75	234	118	41.4	60.9
Ribb	18.2	1.61	263	115	45.8	57.0

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

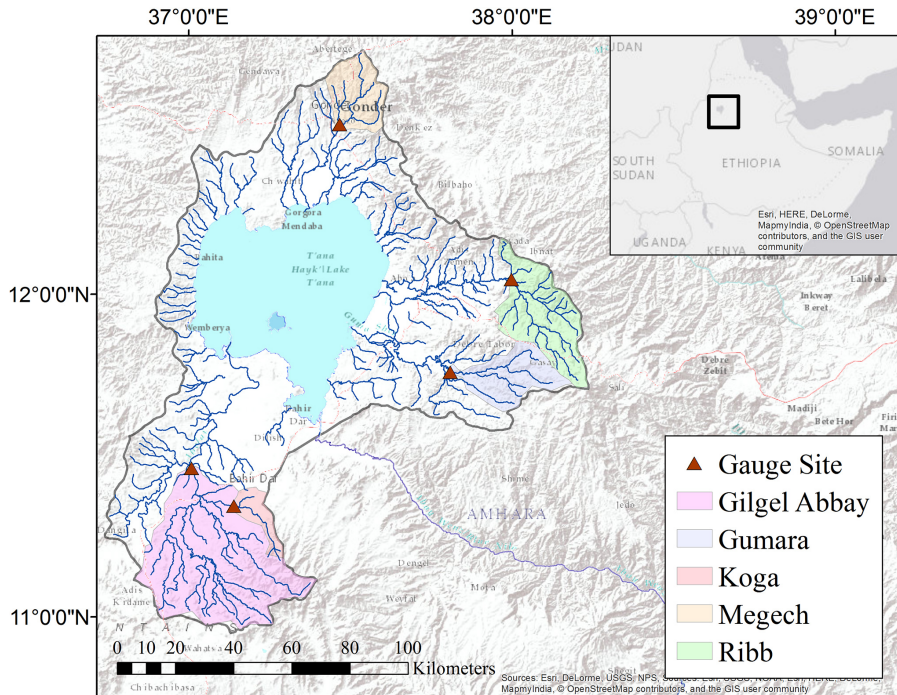


Figure 1. Map of Lake Tana and surrounding rivers.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

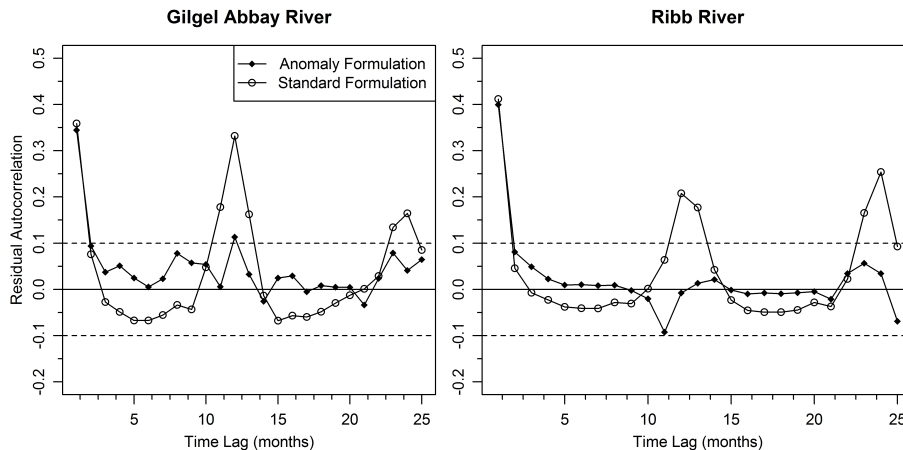


Figure 2. Autocorrelation in model residuals for the Gilgel Abbay and Ribb Rivers.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

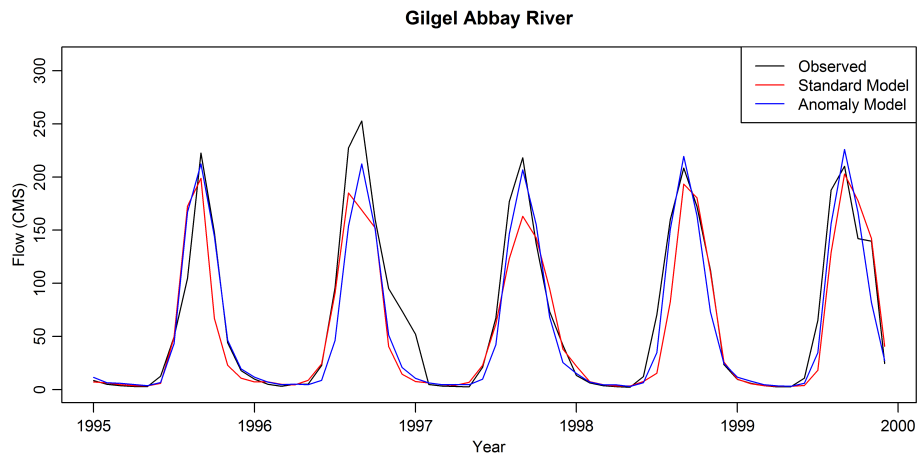


Figure 3. Example observed and predicted flows for Gilgel Abbay River from 1995 to 2000.

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

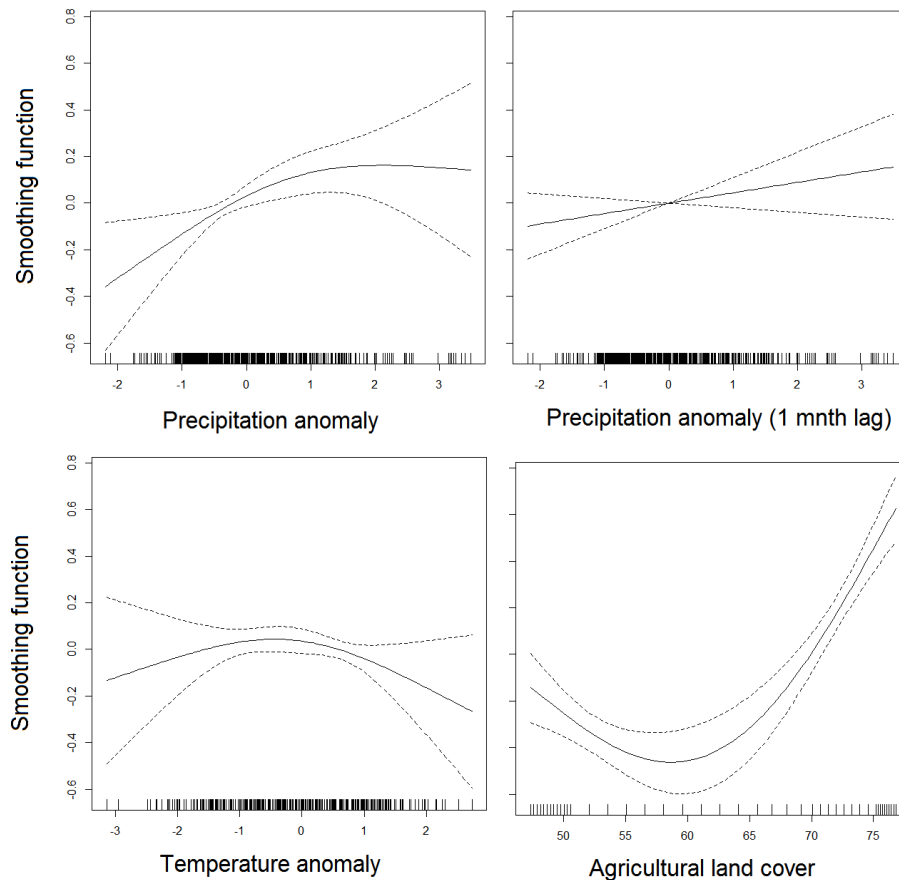


Figure 4. Plots of the smoothing functions used in the Megech River GAM. Hash marks along the x axis indicate observation values of each covariate.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

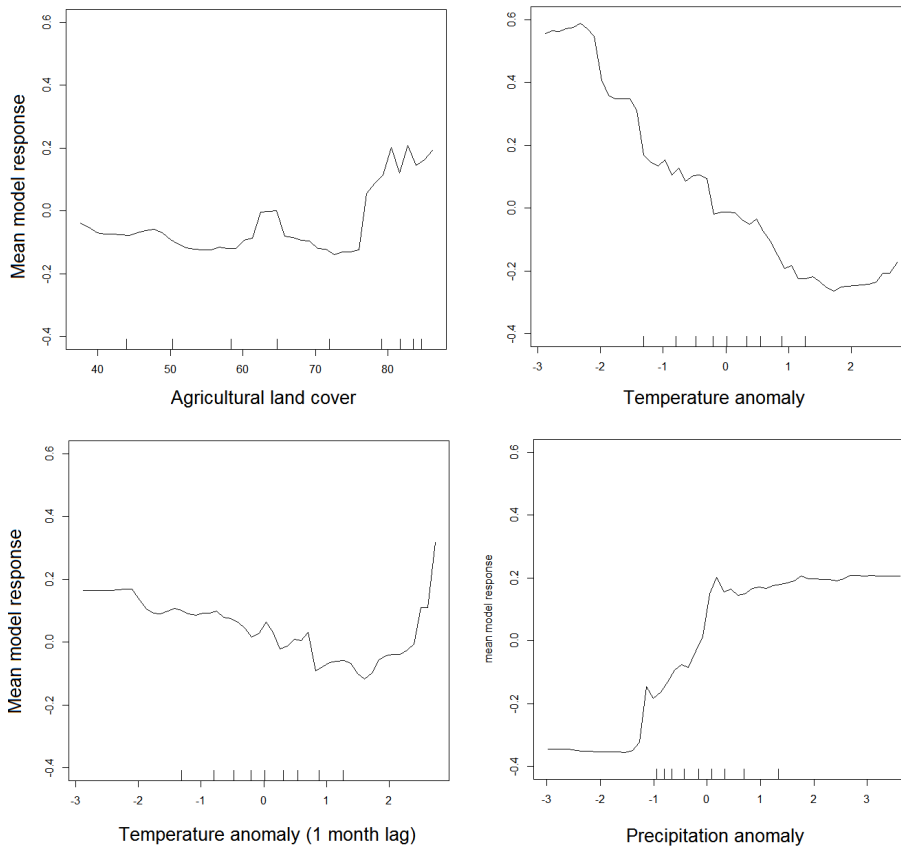


Figure 5. Partial dependence plots for the Ribb River random forest model. Hash marks along the x axis show covariate sample decile values.

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[◀](#) | [▶](#)

[◀](#) | [▶](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

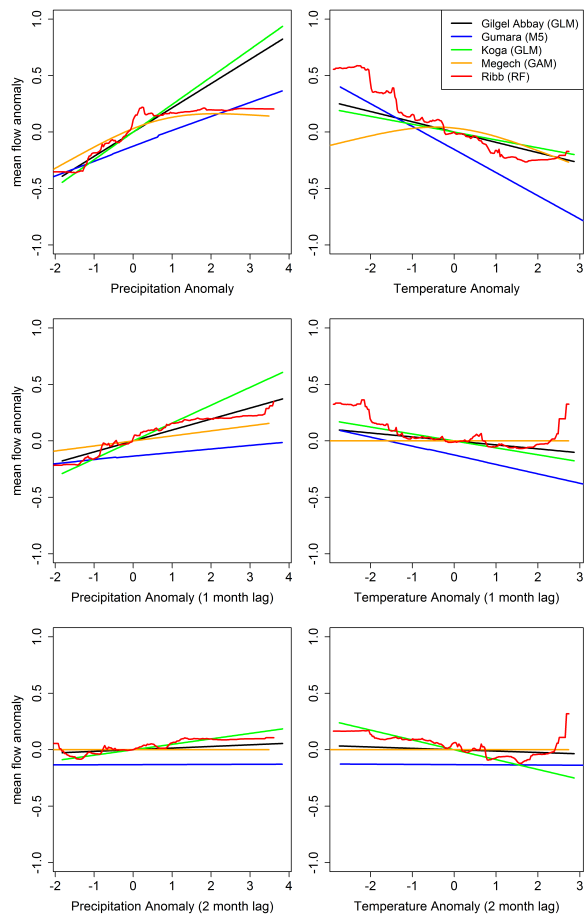


Figure 6. Partial dependence plots for climate covariates in the highest performing model in each basin. Model type is indicated in parentheses.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)

[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

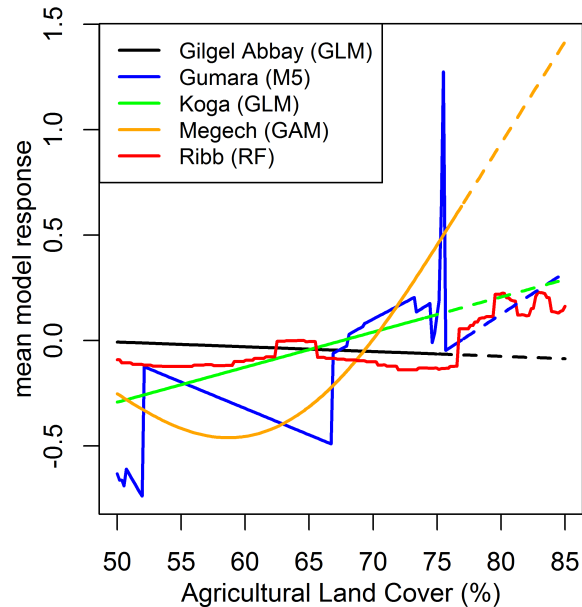


Figure 7. Partial dependence plot for agricultural land cover in the highest performing model in each basin. Model type is listed in parentheses for each basin. Dashed lines indicate values that exceed historic levels of agricultural land cover experienced in that basin.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

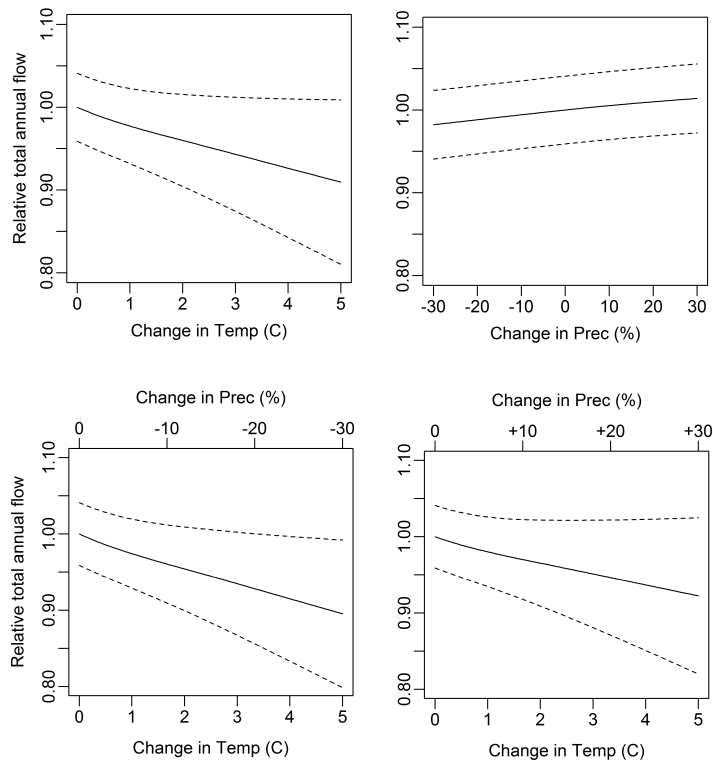


Figure 8. Projected changes in total streamflow (relative to current long-term average) under changing climate conditions. The top two panels show the sensitivity to changes in temperature and precipitation when they are varied independently. The bottom panel shows sensitivity to changing temperature in conjunction with decreasing (left panel) and increasing (right panel) precipitation. Dashed lines represent 95 % confidence bounds from bootstrap resampling.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

HESSD

12, 11083–11127, 2015

Empirical streamflow simulation for water resource management

J. E. Shortridge et al.

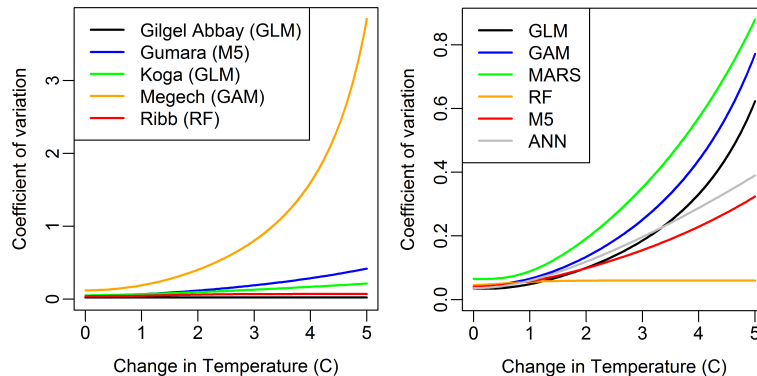


Figure 9. Changes in the coefficient of variation across bootstrap resamples from the highest performing model in each basin (left panel) and multiple models all applied to the Gumara basin (right panel).