

Monthly Streamflow Forecasting across Contrasting Hydro-Climatic Regimes of France Using Support Vector Regression, Ensemble Trees and Neural Networks

Amin Barani¹✉

¹ CEO, Sistan and Baluchestan Regional Water Company, Zahedan, Iran. (✉ Aminbarani2008@gmail.com)

Received: 05 May 2026

Accepted: 07 June 2026

Available online: 25 June 2025

Abstract: Reliable monthly streamflow forecasts support reservoir operation, water allocation and drought preparedness, yet the skill of data-driven models is rarely compared across hydrologically contrasting catchments under an identical protocol. This study evaluated multiple linear regression (MLR), support vector regression (SVR), random forest (RF), gradient boosting machine (GBM) and a multilayer perceptron (MLP) for monthly streamflow prediction in six French catchments (Bruche, Meuse, Loing, Odet, Ubaye and Esteron) that span oceanic, lowland karstic, groundwater-dominated, mid-mountain, Alpine nival and Mediterranean regimes. Daily SAFRAN-based precipitation, temperature and potential evapotranspiration and observed streamflow from the CAMELS-FR sample (1999–2018) were aggregated to 240 monthly values per catchment. Three input scenarios were built from autocorrelation and cross-correlation analyses: antecedent flows only (S1), antecedent flows plus antecedent climate and a seasonal signal (S2, a true one-month-ahead forecast), and S2 plus same-month climate (S3, a perfect-forcing upper bound). Models were tuned by five-fold forward-chaining cross-validation on 1999–2012 and tested on 2013–2018 against persistence and climatology benchmarks. In the forecast setting (S2) the best models reached a Nash–Sutcliffe efficiency (NSE) of 0.25–0.87 and beat climatology in every catchment, but gains over MLR were small. Adding same-month climate (S3) raised the best NSE to 0.63–0.94 and reduced RMSE by 21–65%, the largest gains occurring in the rain-fed Esteron, Odet and Bruche catchments and the smallest in the snowmelt-driven Ubaye. The MLP ranked first in S3 (mean NSE = 0.85; best in five of six catchments), followed by SVR (0.80), whereas RF and GBM overfitted and could not extrapolate beyond the largest training flows. Permutation importance showed that one-month lagged flow and same-month precipitation dominated, while seasonal terms and 12-month lagged flow were informative only in the nival Ubaye. Forecast skill at the monthly scale is therefore governed more by catchment memory and by the quality of meteorological forecasts than by the choice of algorithm.

Keywords: Monthly streamflow forecasting, Support vector regression, Multilayer perceptron, Gradient boosting, CAMELS-FR, Hydrological regime

1. Introduction

Monthly streamflow forecasts are a basic input to water resources planning. Reservoir release rules, irrigation quotas, hydropower scheduling and drought management plans are all set weeks to months ahead, and errors in the expected inflow translate directly into water shortages or unnecessary spills. Physically based and conceptual rainfall–runoff models remain the reference tools for this task, but they require careful calibration, continuous forcing data and, for forecasting, a meteorological forecast chain. Data-driven models offer a lighter alternative that learns the input–output relationship directly from observed records, and over the past two decades they have become a standard component of the hydrological forecasting toolbox (ASCE Task Committee, 2000; Maier and Dandy, 2000; Yaseen et al., 2015).

Artificial neural networks (ANNs) were the first machine learning (ML) models to be widely used for streamflow prediction (Kisi, 2007; Piri et al., 2009). Support vector regression (SVR), which follows the structural risk minimisation principle (Vapnik, 1995; Drucker et al., 1997), later gave more accurate long-term discharge predictions than autoregressive moving average (ARMA) models and ANNs (Lin et al., 2006). Wang et al. (2009) compared ARMA models, ANNs, adaptive neuro-fuzzy inference systems (ANFIS),

genetic programming and support vector machines for monthly discharge and found that the best model depended on the evaluation criterion. Tree ensembles such as random forest (RF; Breiman, 2001) and gradient boosting machines (GBM; Friedman, 2001) have since gained popularity because they handle non-linear interactions with little tuning (Tyralis et al., 2019). Deep learning, in particular the long short-term memory network, has pushed daily rainfall–runoff simulation to new levels of accuracy in large-sample settings (Kratzert et al., 2018; Feng et al., 2020; Ni et al., 2020; Nearing et al., 2021).

Despite this progress, three issues limit what can be concluded from much of the monthly forecasting literature. First, most studies examine one or two catchments, so it is unclear whether the ranking of algorithms is a property of the models or of the particular hydrological regime (Papacharalampous et al., 2019). Second, the input set often mixes antecedent information with same-month meteorological observations. Such models are useful for simulation, but in operational forecasting the same-month precipitation is itself unknown and must come from a weather or seasonal forecast (Akbarian et al., 2023). Reporting the two settings separately shows how much of the apparent skill depends on the forcing rather than on the learning algorithm. Third, ML results are seldom compared with

naive benchmarks such as persistence and monthly climatology, even though the Nash–Sutcliffe efficiency (NSE) is itself a comparison with the mean flow and can look favourable for strongly seasonal rivers (Knoben et al., 2019). Recent multi-model comparisons for monthly flow (Ayana et al., 2023) and daily flow (Rasouli et al., 2012) also show that performance depends strongly on the predictors that are supplied.

Large-sample hydrology datasets make it possible to address these issues with consistent data. The CAMELS family of datasets provides quality-controlled forcing, streamflow and catchment attributes for many basins (Addor et al., 2017), and the recent CAMELS-FR dataset extends this approach to France, one of the most hydrologically diverse countries in Europe (Delaigue et al., 2025). Within a few hundred kilometres, French rivers range from rain-fed oceanic streams in Brittany and chalk- and limestone-fed rivers in the Paris Basin to snowmelt-dominated Alpine rivers and flashy Mediterranean systems.

The objective of this study was to compare five data-driven models, namely multiple linear regression (MLR), SVR, RF, GBM and a multilayer perceptron (MLP), for monthly streamflow prediction in six French catchments that represent contrasting hydro-climatic regimes. Specifically, the study aimed to (i) characterise the monthly regimes, memory and recent trends of the catchments; (ii) identify input combinations from autocorrelation and cross-correlation analyses; (iii) quantify skill in a genuine one-month-ahead forecast setting and in a perfect-forcing setting, each against persistence and climatology benchmarks; and (iv) interpret the models by permutation importance and seasonal error analysis. All data are openly available, so the full workflow can be reproduced.

2. Materials and Methods

2.1. Study Area and Data

Six catchments were selected from the CAMELS-FR sample distributed with the airGRdatasets package (Delaigue et al., 2024, 2025). They were chosen to cover the main hydrological regimes of metropolitan France and to have almost complete streamflow records (Figure 1; Table 1). The Bruche at Russ drains the eastern slopes of the Vosges massif and has a mixed rain–snow regime. The Meuse at Saint-Mihiel is a large lowland river on the limestone plateaus of Lorraine. The Loing at Épisay drains the sedimentary Paris Basin, where the chalk and

limestone aquifers sustain baseflow and part of the recharge leaves the topographic catchment (Le Moine et al., 2007). The Odet at Ergué-Gabéric is an oceanic, rain-fed river in Brittany. The Ubaye at Lauzet-Ubaye is a high Alpine catchment (median elevation 2128 m) with a snowmelt peak in May–June, and the Esteron at Broc is a Mediterranean pre-Alpine river with autumn floods and summer low flows.



Figure 1. Location of the six study catchment outlets in France and their hydrological regimes

For each catchment, CAMELS-FR provides daily catchment-average total precipitation (P) and air temperature (T) from the SAFRAN reanalysis of Météo-France (8 km grid; Quintana-Seguí et al., 2008; Vidal et al., 2010), potential evapotranspiration (PET) computed with the temperature-based formula of Oudin et al. (2005), and observed daily streamflow (Q, mm d⁻¹) from the French national hydrometric archive. The common period was 1 January 1999 to 31 December 2018. Daily values were aggregated to monthly totals of P, PET and Q (mm month⁻¹) and to monthly mean T (°C). Monthly Q was computed as the mean of the valid daily values multiplied by the number of days in the month, and months with fewer than 20 valid daily flows were treated as missing. This rule removed one month for the Ubaye (November 2009) and four months for the Esteron (September–October 2004 and June–July 2014). All other series were complete (240 months).

Table 1. Main characteristics of the study catchments (1999–2018)

Catchment	Station (code)	Lon (°E)	Lat (°N)	Area (km ²)	Elevation, median (range) (m)	P (mm yr ⁻¹)	PET (mm yr ⁻¹)	Q (mm yr ⁻¹)	T (°C)	Q/P	PET/P	Regime
Bruche	La Bruche à Russ [Wisches] (A273011002)	7.28	48.51	224	603 (264–1102)	1244	620	769	8.6	0.62	0.50	Mid-mountain rain-snow (Vosges)
Meuse	La Meuse à Saint-Mihiel (B222001001)	5.53	48.87	2543	351 (199–509)	954	665	366	9.9	0.38	0.70	Lowland pluvial, karstic (Lorraine)
Loing	Le Loing à Épisy (F439000101)	2.79	48.34	3917	137 (48–380)	754	712	166	11.2	0.22	0.94	Groundwater-dominated sedimentary (Paris Basin)
Odet	L'Odet à Ergué-Gabéric - Treodet (J421191001)	-4.06	48.01	203	140 (13–301)	1297	675	722	11.1	0.56	0.52	Oceanic pluvial (Brittany)
Ubaye	L'Ubaye au Lauzet-Ubaye [Roche-Rousse] - DREAL PACA (X045401001)	6.40	44.45	943	2128 (798–3306)	998	436	649	3.5	0.65	0.44	Alpine nival (Southern Alps)
Esteron	L'Esteron au Broc [La Clave] (Y643401001)	7.16	43.85	442	900 (137–1875)	1072	727	449	10.6	0.42	0.68	Mediterranean pre-Alpine (Provence)

P = precipitation; PET = potential evapotranspiration (Oudin et al., 2005); Q = streamflow depth; T = mean air temperature; Q/P = runoff coefficient; PET/P = aridity index. Elevations are catchment hypsometric values from CAMELS-FR.

2.2. Statistical Characterisation and Trend Analysis

Descriptive statistics (minimum, maximum, mean, standard deviation, coefficient of variation, skewness and excess kurtosis) were computed for monthly Q and P, and normality was tested with the Shapiro–Wilk test. Monotonic trends in annual P, PET, T and Q were tested with the Mann–Kendall test (Mann, 1945; Kendall, 1975), using the variance correction of Hamed and Rao (1998) to account for serial correlation. Trend magnitudes were estimated with the Sen (1968) slope. For Q, the annual value was taken as twelve times the mean of the available monthly values, so that the few missing months did not bias the annual totals. The Mann–Kendall statistic is:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (1)$$

where $x_{\{i\}}$ and $x_{\{j\}}$ are the annual values in years i and j and $n = 20$ is the record length. The standardised statistic Z is obtained from S and its corrected variance, and $|Z| > 1.96$ indicates a significant trend at the 5% level. The Sen slope is the median of all pairwise slopes $(x_{\{j\}} - x_{\{i\}})/(j - i)$.

2.3. Input Selection and Forecasting Scenarios

Candidate predictors were identified from the autocorrelation (ACF) and partial autocorrelation (PACF) functions of log-transformed monthly flow and from the cross-correlation between flow and precipitation, all computed on the training period only (Bowden et al., 2005). Three input scenarios were then defined (Table 2). Scenario S1 uses only antecedent flows, i.e. the lags with significant PACF (1–3 months) plus the 12-month seasonal lag. Scenario S2 adds antecedent precipitation, temperature and PET together with a sine–cosine encoding of the calendar month. Because every S2 predictor is known at the end of month $t-1$, S2 is a true one-month-ahead forecast. Scenario S3 adds same-month P, T and PET. It represents simulation with observed forcing, or forecasting with a perfect meteorological forecast, and therefore gives an upper bound on what better meteorological forecasts could deliver. The month of the year m was encoded as:

$$s_m = \sin\left(\frac{2\pi m}{12}\right), \quad c_m = \cos\left(\frac{2\pi m}{12}\right) \quad (2)$$

Table 2. Input combinations used for the monthly streamflow models

Scenario	Inputs	No.	Setting
S1	Q(t-1), Q(t-2), Q(t-3), Q(t-12)	4	Flow memory only
S2	S1 + P(t-1), T(t-1), PET(t-1), sin(m), cos(m)	9	One-month-ahead forecast
S3	S2 + P(t), T(t), PET(t)	12	Perfect forcing (simulation)

Flow inputs and the target were transformed with $\ln(1 + Q)$ to reduce skewness, and all inputs were standardised to zero mean and unit variance using training-period statistics. Predictions were back-transformed to mm month⁻¹ and negative values were set to zero. The first 12 months (1999) were used only to build lagged inputs. Samples with a missing target or missing lagged flow were removed, which left 156 training months (2000–2012) and 72 test months (2013–2018) for four catchments, 151/72 for the Ubaye and 149/65 for the Esteron.

2.4. Machine Learning Models

Multiple linear regression. MLR was used as the linear reference. It is fitted by ordinary least squares, and its predicted log-flow is a weighted sum of the standardised inputs.

Support vector regression. SVR fits a function that deviates from the targets by at most ε while remaining as flat as possible (Vapnik, 1995; Drucker et al., 1997). With a radial basis function (RBF) kernel the prediction is:

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2) + b \quad (3)$$

where α_i and α_i^* are Lagrange multipliers bounded by the regularisation constant C , γ is the kernel width and b is the bias. The values of C , γ and ε were tuned (Table 3).

Random forest. RF averages B regression trees, each grown on a bootstrap sample with a random subset of predictors considered at each split (Breiman, 2001):

$$\hat{y}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x}) \quad (4)$$

Gradient boosting machine. GBM builds an additive model in which each new shallow tree $h_{\{k\}}$ is fitted to the negative gradient of the squared loss, i.e. to the residuals of the current ensemble, and is added with a learning rate ν (Friedman, 2001). Stochastic subsampling of 80% of the training data was applied at each iteration:

$$F_k(\mathbf{x}) = F_{k-1}(\mathbf{x}) + \nu h_k(\mathbf{x}), \quad k = 1, \dots, K \quad (5)$$

Multilayer perceptron. The MLP is a feed-forward network with one or two hidden layers (Rumelhart et al., 1986). For one hidden layer with H neurons, the output is:

$$\hat{y} = \sum_{j=1}^H w_j \varphi\left(\sum_{i=1}^p v_{ji} x_i + b_j\right) + b_0 \quad (6)$$

where φ is the activation function (tanh or ReLU), $v_{\{ji\}}$ and $w_{\{j\}}$ are weights and b are biases. The weights were estimated with the quasi-Newton L-BFGS algorithm, and an L2 penalty α controlled overfitting.

Hyperparameters were selected by an exhaustive grid search (Table 3) with five-fold forward-chaining cross-validation on the training period, so that each validation fold always followed its training data in time (Bergmeir and Benítez, 2012). The root mean square error of log-flow was the selection criterion. RF used 500 trees. The final models were refitted on the full training period and applied once to the test period. All computations were made in Python with scikit-learn 1.8 (Pedregosa et al., 2011) and a fixed random seed.

Table 3. Hyperparameter search spaces

Model	Hyperparameters and candidate values
MLR	None (ordinary least squares)
SVR	Kernel = RBF; $C \in \{0.1, 1, 10, 100\}$; $\gamma \in \{0.01, 0.05, 0.1, 0.5\}$; $\epsilon \in \{0.01, 0.05, 0.1\}$
RF	Trees = 500; max depth $\in \{4, 8, \text{none}\}$; min samples leaf $\in \{1, 3, 5\}$; max features $\in \{0.33, 0.66, 1.0\}$
GBM	Trees $\in \{100, 300, 600\}$; learning rate $\in \{0.01, 0.05, 0.1\}$; max depth $\in \{2, 3, 4\}$; subsample = 0.8
MLP	Hidden layers $\in \{(4), (8), (16), (8, 4)\}$; $\alpha \in \{0.001, 0.01, 0.1, 1\}$; activation $\in \{\text{tanh}, \text{ReLU}\}$; solver = L-BFGS

2.5. Benchmarks and Evaluation Criteria

Two naive benchmarks were evaluated on the same test months: persistence, in which the forecast equals the previous month's flow, and climatology, in which the forecast equals the 1999–2012 mean flow for the same calendar month. Model performance was measured with the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), NSE (Nash and Sutcliffe, 1970), Kling–Gupta efficiency (KGE;

Gupta et al., 2009) and percent bias (PBIAS; Moriasi et al., 2007):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - S_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - S_i| \quad (8)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - S_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (9)$$

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (10)$$

$$PBIAS = 100 \frac{\sum_{i=1}^n (S_i - O_i)}{\sum_{i=1}^n O_i} \quad (11)$$

where $O_{\{i\}}$ and $S_{\{i\}}$ are observed and predicted monthly flows, r is the Pearson correlation between them, $\alpha = \sigma_S/\sigma_O$ is the ratio of their standard deviations and $\beta = \mu_S/\mu_O$ is the ratio of their means. Models were ranked by RMSE within each catchment and scenario, and the ranks were averaged over the catchments. The significance of differences in squared errors was tested with the Diebold–Mariano test (Diebold and Mariano, 1995) with the small-sample correction of Harvey et al. (1997). Input relevance was assessed by permutation importance (Breiman, 2001), i.e. the increase in test RMSE when one input is randomly shuffled (30 repetitions). It was averaged over the RF and GBM models of scenario S3 and normalised to percentages for each catchment.

3. Results

3.1. Hydro-Climatic Characteristics

The six catchments differ strongly in water balance (Table 1). Mean annual precipitation ranged from 754 mm in the Loing to 1297 mm in the Odet, and runoff ranged from 166 mm to 769 mm. The runoff coefficient was lowest in the Loing (0.22), where the aridity index approaches unity ($PET/P = 0.94$), and highest in the Ubaye (0.65) and the Bruche (0.62). The monthly regimes (Figure 2) show the expected contrasts. The Bruche, Meuse, Loing and Odet have winter high flows and late-summer low flows that follow the seasonal cycle of PET. The Ubaye has a single snowmelt peak in May (149 mm month⁻¹ on average) despite fairly uniform precipitation. The Esteron shows a precipitation maximum in November and a strong autumn–winter flow response. In the Loing, the monthly flow varies only between about 5 and 24 mm month⁻¹, reflecting the damping effect of groundwater storage.

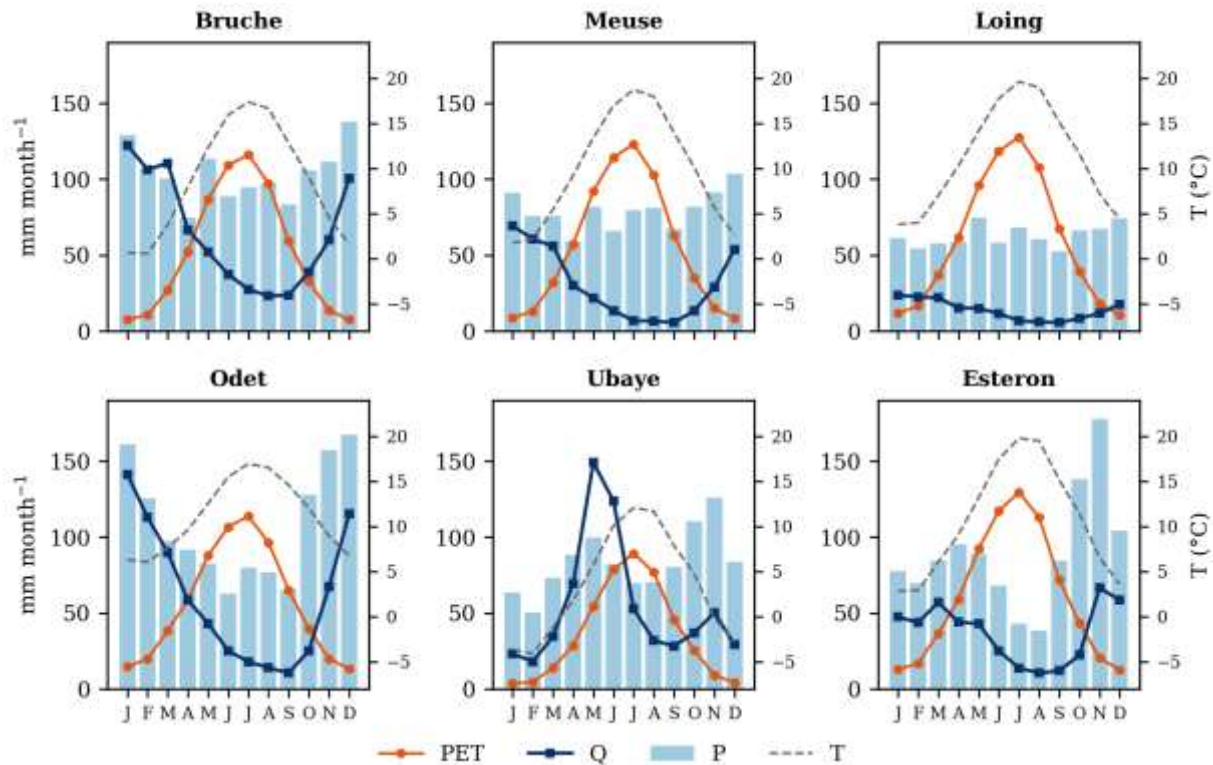


Figure 2. Mean monthly precipitation (P, bars), potential evapotranspiration (PET), streamflow (Q) and air temperature (T, right axis) of the study catchments, 1999–2018

Monthly flows were positively skewed in all catchments (skewness 1.43–1.92; Table 4), with coefficients of variation between 0.76 (Loing) and 1.04 (Meuse). The Shapiro–Wilk test rejected normality for all raw series ($p < 0.001$) and, although the logarithmic transformation reduced the skewness, the transformed series also remained non-normal ($p \leq 0.001$). This supports the use of transformation together with non-parametric learners. The time series (Figure 3) show that several of the largest monthly flows occurred in the test period. These include January 2018 in the Bruche (286 mm) and the Meuse (175 mm), and June 2016 in the Loing (67 mm), which exceeded the corresponding training maxima.

Table 4. Descriptive statistics of monthly streamflow (mm month⁻¹), 1999–2018

Catchment	N	Min	Max	Mean	SD	CV	Skewness	Kurtosis
Bruche	240	10.7	285.7	64.1	51.5	0.80	1.43	1.91
Meuse	240	1.6	174.7	30.5	31.7	1.04	1.47	2.13
Loing	240	2.7	66.9	13.9	10.5	0.76	1.67	3.52
Odet	240	4.2	296.3	60.2	56.9	0.95	1.59	2.71
Ubaye	239	8.6	241.2	54.1	46.5	0.86	1.80	2.87
Esteron	236	4.9	213.1	37.4	37.0	0.99	1.92	3.83

SD = standard deviation; CV = coefficient of variation; Kurt = excess kurtosis.

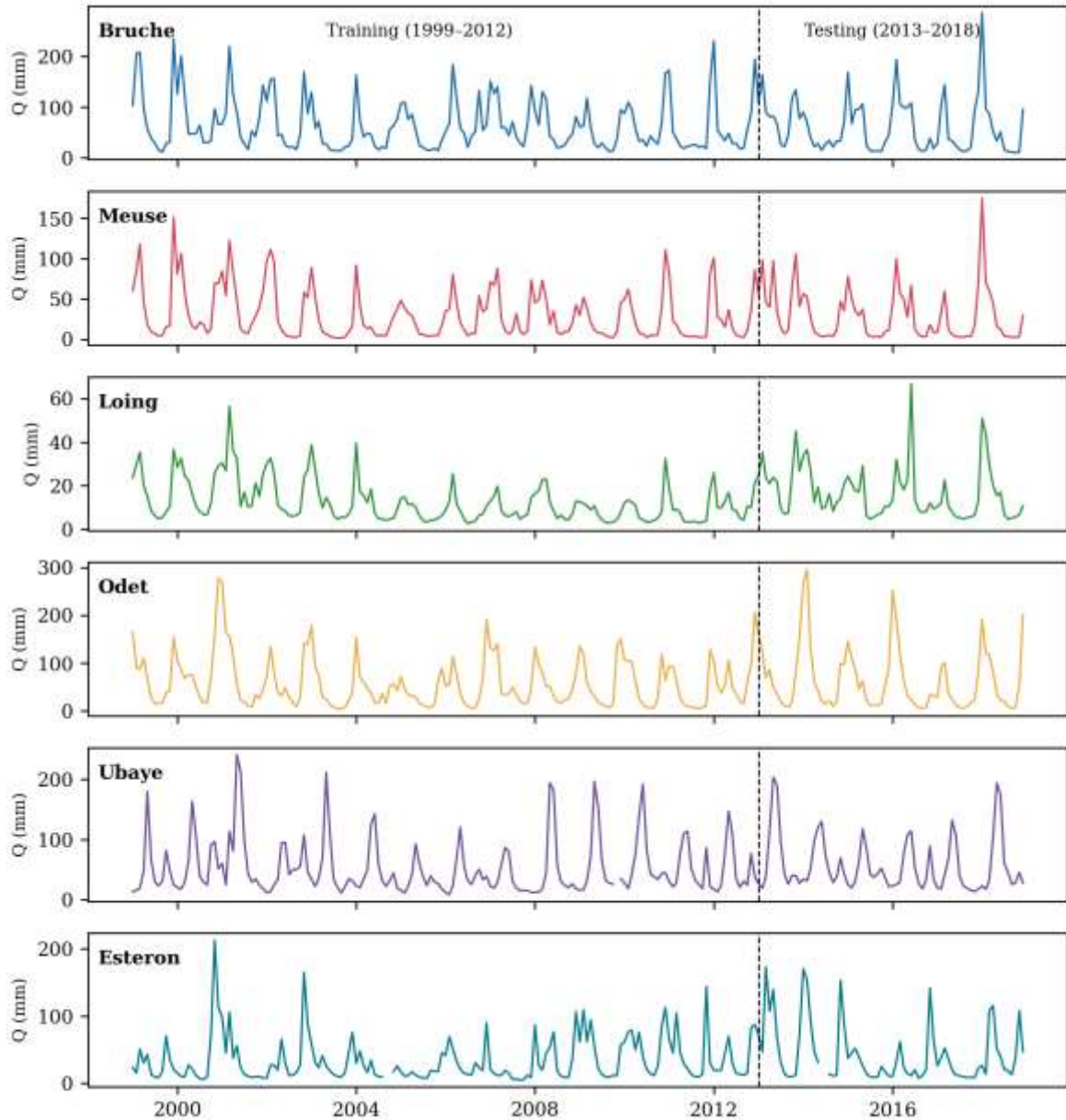


Figure 3. Monthly streamflow series of the six catchments; the dashed line separates the training (1999–2012) and testing (2013–2018) periods

The modified Mann–Kendall test (Table 5) showed significant decreases in annual precipitation in the Bruche (-19.6 mm yr^{-1}) and the Meuse (-14.0 mm yr^{-1}), and significant warming in the two southern catchments, the Ubaye ($+0.073 \text{ }^{\circ}\text{C yr}^{-1}$) and the Esteron ($+0.061 \text{ }^{\circ}\text{C yr}^{-1}$), where PET also increased significantly. None of the annual streamflow series had a significant trend over 1999–2018. The 20-year record is short, so these results describe recent variability rather than long-term change, but they show that the test period is not a simple continuation of the training period.

Table 5. Sen slopes (per year) and modified Mann–Kendall Z statistics (in brackets) of annual series, 1999–2018

Catchment	P (mm)	PET (mm)	T ($^{\circ}\text{C}$)	Q (mm)
Bruche	-19.6^* (-2.11)	$+1.4$ ($+1.91$)	$+0.031$ ($+1.27$)	-9.9 (-1.20)
Meuse	-14.0^* (-2.04)	$+1.2$ ($+1.11$)	$+0.024$ ($+1.13$)	-7.3 (-1.07)
Loing	-3.3 (-0.68)	$+0.6$ ($+0.62$)	$+0.016$ ($+0.49$)	-1.0 (-0.31)
Odet	-8.0 (-1.01)	$+0.0$ ($+0.00$)	$+0.002$ ($+0.00$)	-2.1 (-0.03)
Ubaye	-2.0 (-0.16)	$+2.6^*$ ($+3.03$)	$+0.073^*$ ($+3.21$)	$+1.7$ ($+0.23$)
Esteron	$+8.6$ ($+0.49$)	$+2.3^*$ ($+2.91$)	$+0.061^*$ ($+3.35$)	$+8.9$ ($+1.01$)

significant at the 5% level ($|Z| > 1.96$), Hamed and Rao (1998) variance correction

3.2. Memory Structure and Input Selection

The lag-1 autocorrelation of log-flow ranged from 0.63 (Esteron) to 0.78 (Loing) and was well above the

95% confidence limit (± 0.15) in every catchment (Figure 4). The PACF was significant at lag 1 everywhere and significantly negative at lags 2–3 in most catchments, which motivated the use of three flow lags. The 12-month ACF was high in the Loing (0.70), the Meuse (0.65) and the Ubaye (0.65) but weak in the Esteron (0.28), consistent with a regular annual cycle in the first three and irregular flood timing in the

Mediterranean catchment. The cross-correlation between flow and precipitation was highest at lag 0 in all catchments except the Odet, where the correlation with the previous month's precipitation was slightly higher ($r = 0.63$ vs. 0.60). In the Ubaye the lag-3 cross-correlation was negative ($r = -0.17$), reflecting the delay between winter snowfall and spring melt.

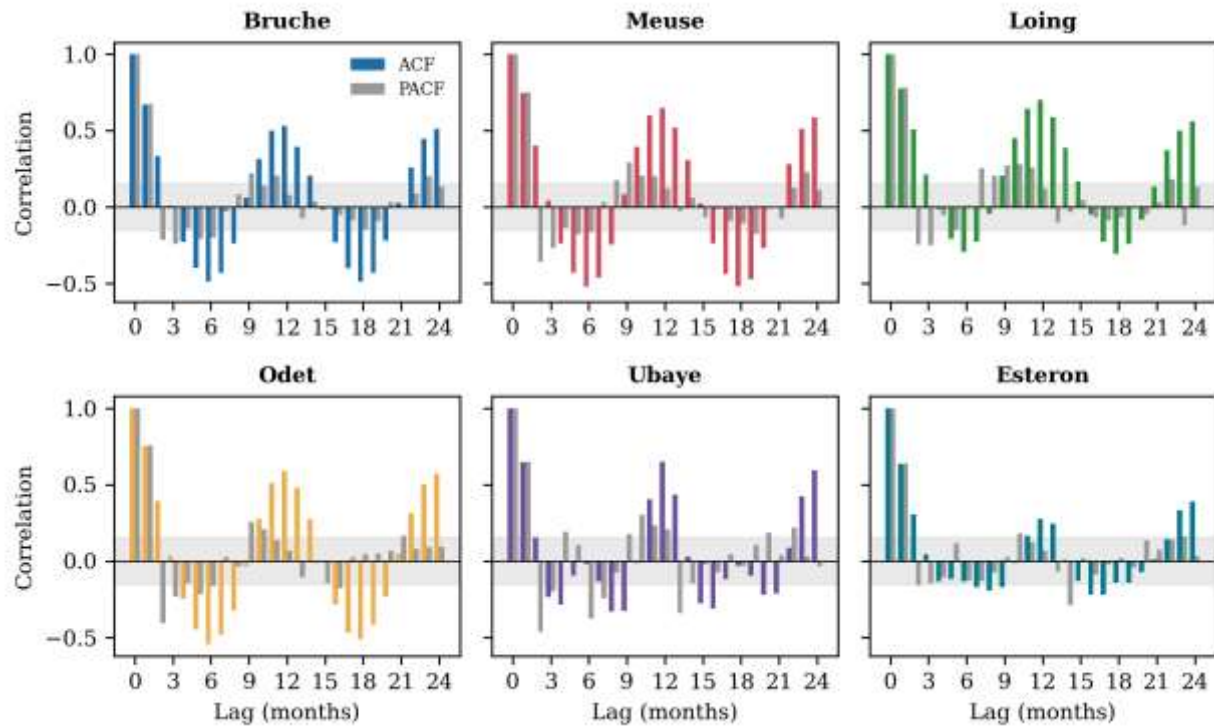


Figure 4. Autocorrelation (ACF) and partial autocorrelation (PACF) functions of log-transformed monthly streamflow in the training period; the shaded band is the 95% confidence interval

3.3. Forecasting Performance

Figure 5 summarises the test-period NSE of all models and benchmarks. Persistence was the weakest benchmark ($\text{NSE} \approx 0.00\text{--}0.43$) because monthly flows

change quickly relative to their memory. Climatology was much stronger in the regular-regime catchments and reached $\text{NSE} = 0.77$ in the Ubaye and 0.57 in the Odet, but only 0.11–0.12 in the Loing and the Esteron.

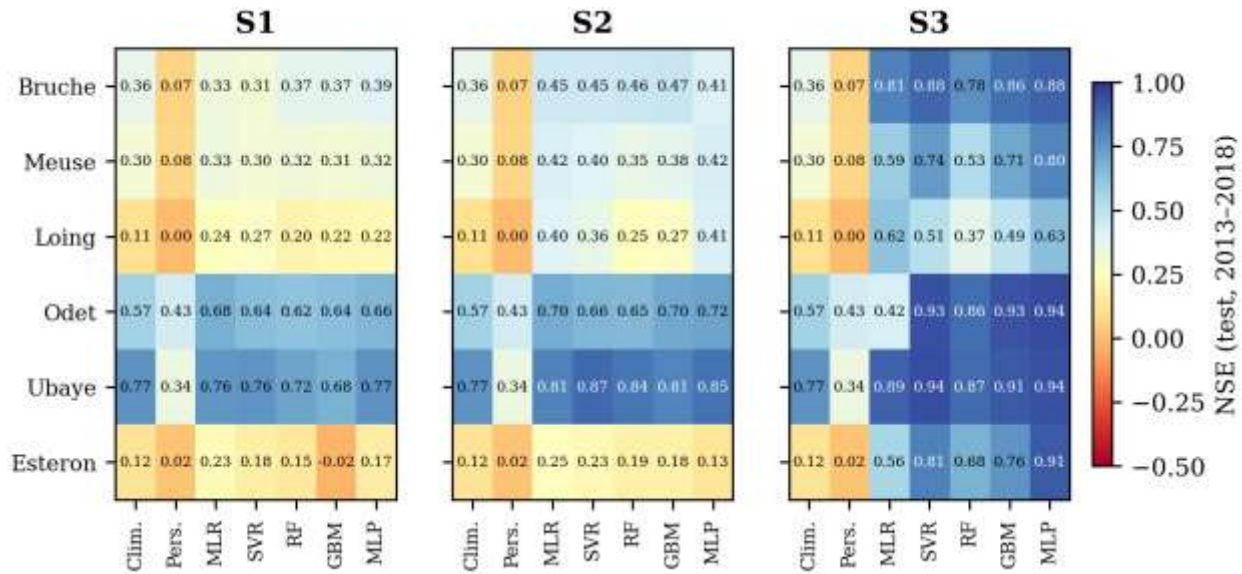


Figure 5. Test-period (2013–2018) Nash–Sutcliffe efficiency of the benchmarks (climatology, persistence) and the five models under input scenarios S1–S3

With flow memory alone (S1), the best models reached $NSE = 0.23$ – 0.77 . These values were only marginally above climatology in the Ubaye (0.77 vs. 0.77) and, in the Bruche, MLR (0.33) and SVR (0.31) were below the climatology benchmark (0.36). Adding antecedent climate and the seasonal signal (S2) improved all six catchments. The best S2 models achieved $NSE = 0.47$ (Bruche, GBM), 0.42 (Meuse, MLP), 0.41 (Loing, MLP), 0.72 (Odet, MLP), 0.87 (Ubaye, SVR) and 0.25 (Esteron, MLR), and all of these exceeded climatology. However, the five algorithms were close to each other in S2. The mean NSE over the six catchments was 0.50 for MLR, 0.50 for SVR, 0.49 for MLP, 0.47 for GBM and 0.46 for RF (Table 7). In other words, when the future forcing is unknown, the non-linear models add little to a linear model built on the same predictors.

Supplying same-month climate (S3) changed the picture. The best NSE rose to 0.63 – 0.94 and the best KGE to 0.73 – 0.96 (Table 6). Compared with the best S2 model, RMSE fell by 65% in the Esteron, 55% in the Odet, 54% in the Bruche, 41% in the Meuse, 29% in the Ubaye and 21% in the Loing. The ranking of the algorithms also became clearer. The MLP was the best model in five catchments and had the best mean rank (1.17) and the highest mean NSE (0.85). SVR was second (mean rank 2.00 ; mean NSE 0.80) and was the best model in the Ubaye ($NSE = 0.94$, $KGE = 0.96$). GBM ranked third (0.78), while MLR (0.65) and RF (0.68) ranked last. The MLR result in the Odet ($NSE = 0.42$ in S3 vs. 0.70 in S2) shows that adding correlated same-month inputs can harm a linear model in log space when the test period contains flows outside the training range.

Table 6. Test-period (2013–2018) performance of the models under the forecast (S2) and perfect-forcing (S3) scenarios

Catchment	Model	S2 NSE	S2 KGE	S2 RMSE	S2 MAE	S2 PBIAS (%)	S3 R ²	S3 NSE	S3 KGE	S3 RMSE	S3 MAE	S3 PBIAS (%)
Bruche	MLR	0.45	0.54	38.4	26.1	-10.9	0.84	0.81	0.87	22.6	12.3	-7.3
	SVR	0.45	0.48	38.4	25.1	-15.6	0.90	0.88	0.87	18.0	10.9	-10.8
	RF	0.46	0.48	38.1	25.5	-12.7	0.86	0.78	0.65	24.5	15.8	-13.0
	GBM	0.47	0.46	37.8	25.4	-14.0	0.91	0.86	0.74	19.6	12.4	-11.6
	MLP	0.41	0.53	39.9	26.4	-8.6	0.90	0.88	0.85	17.6	10.9	-10.2
Meuse	MLR	0.42	0.57	24.9	14.9	-15.3	0.76	0.59	0.69	21.0	9.5	-3.2
	SVR	0.40	0.44	25.2	14.4	-26.2	0.81	0.74	0.81	16.5	7.6	-7.4
	RF	0.35	0.41	26.3	14.2	-21.2	0.58	0.53	0.52	22.4	12.0	-20.3
	GBM	0.38	0.40	25.8	13.8	-21.4	0.76	0.71	0.65	17.5	9.0	-17.4
	MLP	0.42	0.50	24.8	14.5	-15.0	0.82	0.80	0.87	14.5	7.6	-9.0
Loing	MLR	0.40	0.47	9.4	4.9	-15.0	0.64	0.62	0.67	7.4	3.8	-10.2
	SVR	0.36	0.43	9.6	5.0	-18.6	0.60	0.51	0.76	8.4	4.4	-2.5
	RF	0.25	0.33	10.5	5.7	-19.7	0.47	0.37	0.42	9.6	4.8	-20.3
	GBM	0.27	0.37	10.3	5.4	-19.1	0.55	0.49	0.53	8.6	4.2	-16.7
	MLP	0.41	0.48	9.3	4.9	-15.2	0.65	0.63	0.73	7.3	3.6	-8.9
Odet	MLR	0.70	0.73	36.6	21.2	-11.4	0.80	0.42	0.47	50.8	18.9	6.8
	SVR	0.66	0.56	38.7	20.8	-20.8	0.94	0.93	0.92	17.4	10.8	-6.8
	RF	0.65	0.55	39.1	20.4	-19.1	0.94	0.86	0.69	24.9	14.2	-15.2
	GBM	0.70	0.66	36.3	20.5	-14.5	0.94	0.93	0.86	17.7	10.3	-8.6
	MLP	0.72	0.70	35.2	19.5	-13.7	0.95	0.94	0.89	15.8	9.8	-8.8
Ubaye	MLR	0.81	0.89	20.0	12.5	-4.6	0.90	0.89	0.94	15.0	10.1	-0.9
	SVR	0.87	0.92	16.1	10.5	-5.9	0.94	0.94	0.96	11.5	7.9	-1.8
	RF	0.84	0.86	18.2	13.2	-6.8	0.88	0.87	0.84	16.4	10.7	-7.2
	GBM	0.81	0.79	20.0	13.2	-8.7	0.91	0.91	0.94	13.9	9.9	-0.6

	MLP	0.85	0.90	17.6	10.7	-0.6	0.94	0.94	0.95	11.5	7.6	0.4
Esteron	MLR	0.25	0.37	38.5	22.3	-20.7	0.76	0.56	0.65	29.6	12.7	6.5
	SVR	0.23	0.25	39.1	21.4	-32.8	0.85	0.81	0.86	19.1	9.8	0.4
	RF	0.19	0.19	40.1	23.0	-32.4	0.79	0.68	0.56	25.2	14.8	-19.5
	GBM	0.18	0.19	40.1	22.9	-31.8	0.78	0.76	0.76	21.6	12.8	-11.6
	MLP	0.13	0.25	41.5	27.1	-24.9	0.91	0.91	0.90	13.6	8.9	-2.4

RMSE and MAE in mm month⁻¹. Bold rows mark the best S3 model (highest NSE) in each catchment.

Table 7. Mean test NSE and mean RMSE rank of the models over the six catchments

Model	NSE S1	NSE S2	NSE S3	Rank S1	Rank S2	Rank S3
MLR	0.429	0.503	0.647	2.00	2.67	4.00
SVR	0.410	0.495	0.803	3.00	2.67	2.00
RF	0.398	0.456	0.680	3.83	3.83	4.67
GBM	0.365	0.467	0.775	4.00	3.33	3.17
MLP	0.421	0.489	0.851	2.17	2.50	1.17
Persistence	0.156	0.156	0.156	–	–	–
Climatology	0.373	0.373	0.373	–	–	–

Rank 1 = lowest RMSE among the five models in a catchment.

Figure 6 compares the observed test hydrographs with the best S2 and S3 models. The S2 forecasts follow

the seasonal cycle but systematically lag or damp the peaks, because a flood caused by heavy rain in month t cannot be anticipated from information available at $t-1$. This is clearest for the January 2018 flood in the Bruche and the Meuse and for the autumn floods of the Esteron. The S3 models reproduce the timing and most of the magnitude of these events. The Ubaye is the exception: its S2 and S3 hydrographs are almost the same, because the May–June melt peak is largely predetermined by the snowpack and the calendar.

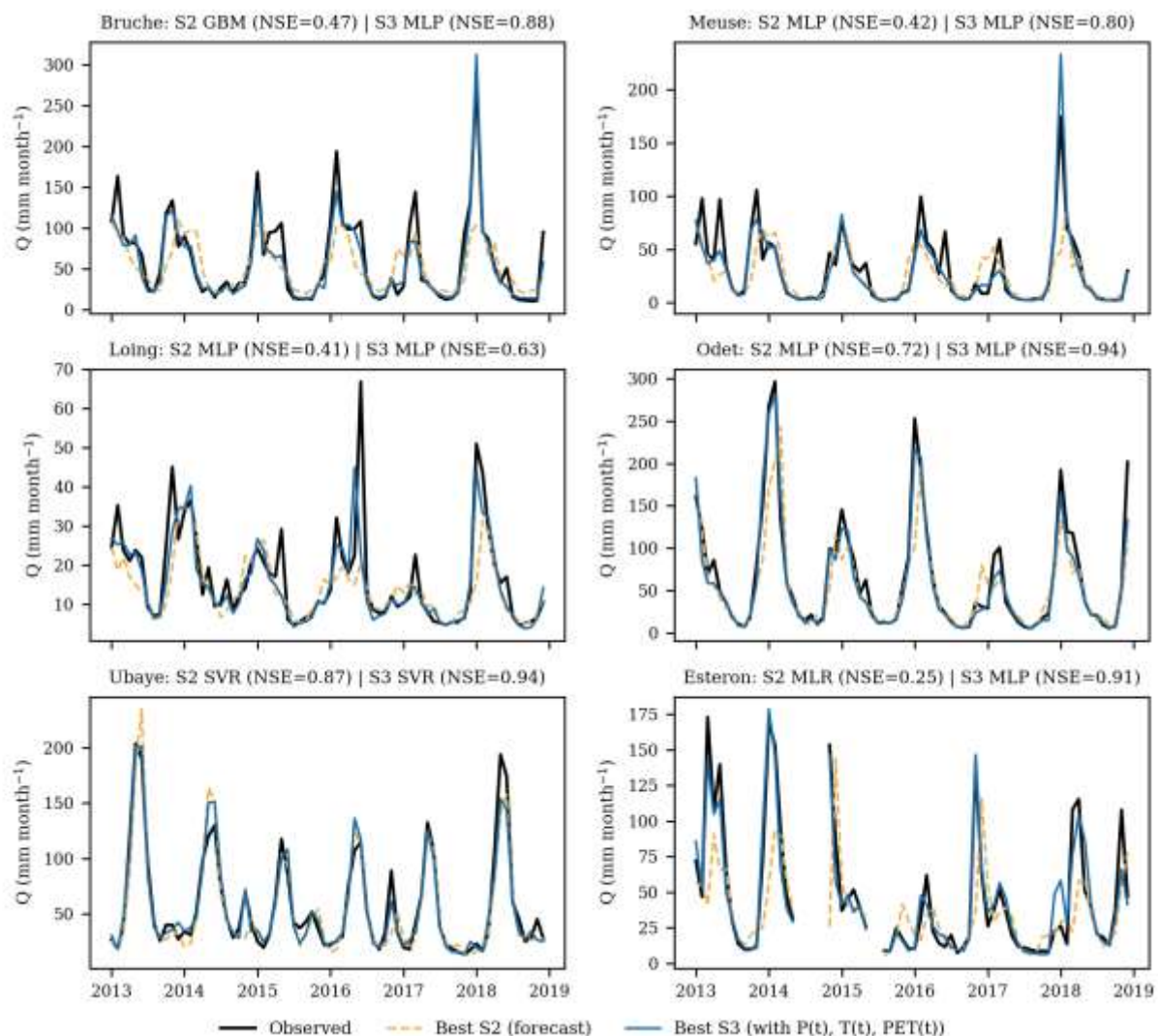


Figure 6. Observed and predicted monthly streamflow in the test period for the best forecast (S2) and perfect-forcing (S3) models

The scatter plots of the best S3 models (Figure 7) show slopes of 0.67–0.95, i.e. a tendency to underestimate high flows. This was strongest in the Loing, where the June 2016 flood (67 mm) was

predicted as 23 mm. PBIAS was negative for almost all models and scenarios (Table 6), ranging from -1.8% (Ubaye, SVR) to -10.2% (Bruche, MLP) for the best S3 models. Part of this bias comes from back-transforming

log-space predictions, which estimates the conditional median rather than the mean (Duan, 1983). Part comes from the wetter test period in the Loing, the Odet and the

Esteron, whose mean test-period flows were 34%, 11% and 28% higher than in training.

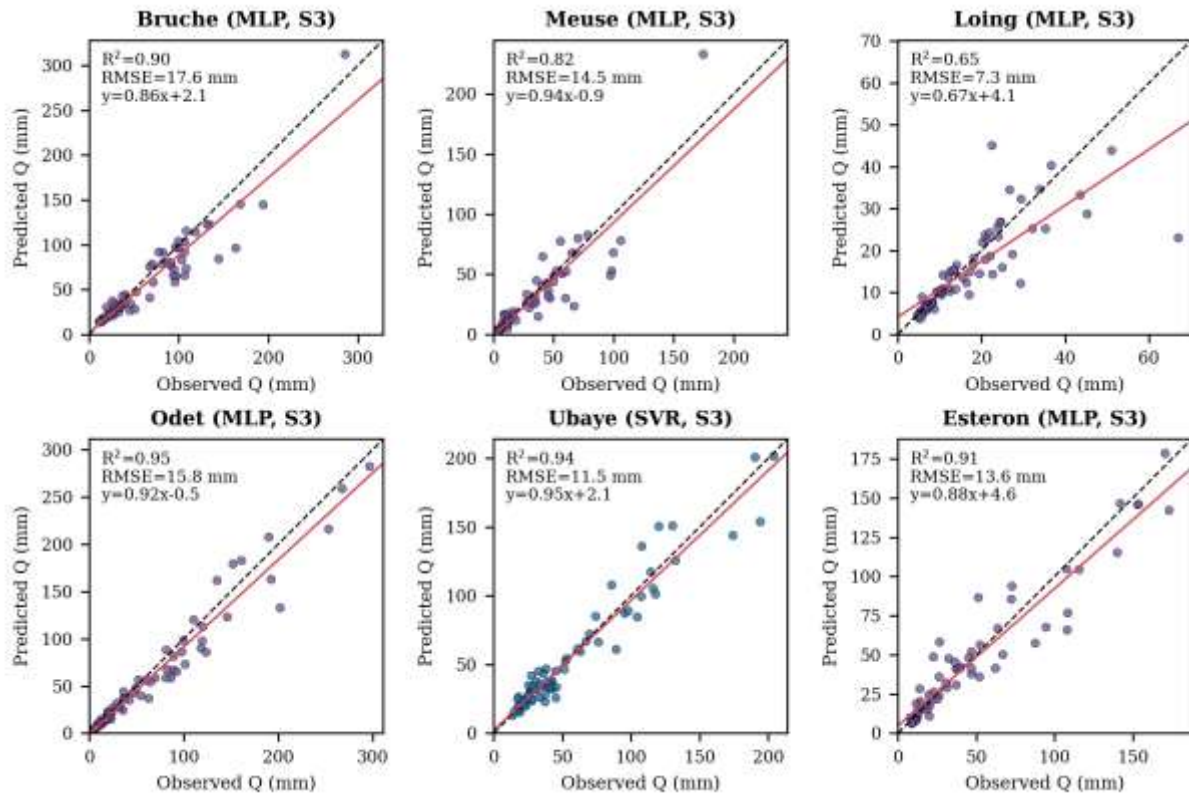


Figure 7. Observed versus predicted monthly streamflow of the best S3 model in each catchment (dashed: 1:1 line; red: least-squares fit)

3.4. Generalisation and Statistical Significance

Train–test differences in NSE (S3) revealed clear overfitting by the tree ensembles. The mean gap was 0.28 for RF and 0.21 for GBM, compared with 0.09 for MLP and 0.13 for SVR. The largest gaps occurred in the Loing (RF: training NSE 0.95, test 0.37). Because a tree cannot predict values beyond the range of its training targets, RF and GBM were additionally penalised in the catchments whose test period contained record monthly flows. In S2 and S3, the selected MLPs all used the strongest regularisation in the grid ($\alpha = 1$), and the selected SVRs all used the smallest kernel width ($\gamma = 0.01$). These choices indicate that smooth response surfaces generalise best with about 150 training months.

The Diebold–Mariano test (Table 8) showed that S3 errors were significantly smaller than S2 errors ($p < 0.05$) in five catchments. The exception was the Loing ($p = 0.094$). In contrast, the best S3 model was significantly better than MLR only in the Ubaye ($p = 0.022$). In the other catchments the lower RMSE of the MLP was not statistically distinguishable from MLR over 65–72 test months, because a few large events dominate the squared-error differences.

Table 8. Diebold–Mariano tests on squared errors in the test period

Catchment	Best S3	DM (vs. MLR S3)	p	Best S2	DM (S3 vs. S2)	p
Bruche	MLP	-1.00	0.320	GBM	-2.42	0.018
Meuse	MLP	-1.10	0.277	MLP	-2.20	0.031
Loing	MLP	-0.40	0.693	MLP	-1.70	0.094
Odet	MLP	-1.24	0.218	MLP	-2.60	0.011
Ubaye	SVR	-2.34	0.022	SVR	-2.26	0.027
Esteron	MLP	-1.15	0.255	MLR	-2.90	0.005

Negative DM values indicate smaller squared errors for the first model; Harvey et al. (1997) small-sample correction.

3.5. Input Importance and Seasonal Error Structure

Permutation importance (Figure 8) confirmed that two inputs carry most of the information. The one-month lagged flow accounted for 20–46% of the importance and was most influential in the groundwater-dominated Loing, and same-month precipitation accounted for 18–51% and was most influential in the flashy Esteron. Lagged PET was also important in the Meuse, Odet and Bruche (27%, 22% and 17%), where it acts as an indicator of the seasonal state of soil moisture. The Ubaye differed from the other catchments: the cosine of the month (18%) and the 12-month lagged flow (15%) were nearly as important as precipitation, which reflects the regular timing of the melt season. Additional flow lags ($t-2$, $t-3$) contributed less than 5% everywhere.

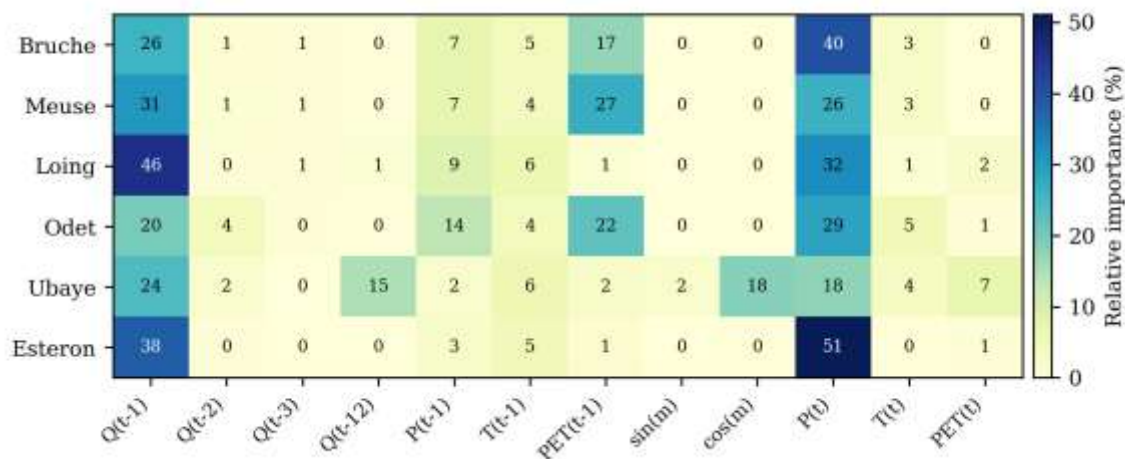


Figure 8. Normalised permutation importance of the inputs (mean of RF and GBM, scenario S3)

Seasonal error analysis (Figure 9) showed that forecast errors were largest in winter. The median normalised RMSE of the best S2 models peaked in January (1.17) and was lowest in July–September (0.08–0.13). Same-month forcing reduced the January error to 0.30, so the added value of meteorological information is concentrated in the wet season, when flows are most variable and most relevant for flood and storage management.

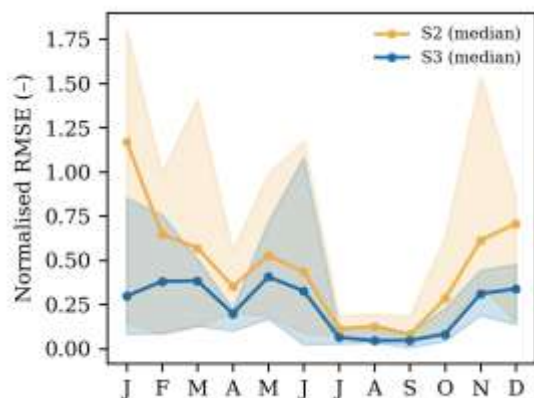


Figure 9. Median (line) and range (shading) over catchments of the RMSE normalised by mean flow, by calendar month, for the best S2 and S3 models

4. Discussion

The central result of this study is that monthly forecast skill depends more on what is known about the coming month than on the learning algorithm. In the true forecast setting (S2), the differences between MLR, SVR, RF, GBM and MLP were small and inconsistent across catchments. When same-month climate was supplied (S3), NSE increased by 0.06–0.66 and the differences between algorithms became clear. This agrees with Akbarian et al. (2023), who found that the skill of ML models driven by ECMWF forecasts over Iranian basins fell with lead time, and with Rasouli et al. (2012), who found that local observations and forecast weather variables were the most useful predictors at

short lead times. For operational use, the S3 results should be read as an upper bound. A real system would replace the observed same-month forcing with seasonal or extended-range forecasts, and its skill would lie between the S2 and S3 values reported here.

Catchment memory explains much of the between-catchment variability. The snowmelt-dominated Ubaye had the highest skill in all scenarios and the smallest benefit from same-month forcing, because winter snow storage determines spring runoff several months in advance. The Loing had the strongest flow autocorrelation but the lowest S3 skill. In this groundwater-dominated catchment the monthly signal is small and smooth, a few summer storm events (e.g., June 2016) dominate the errors, and inter-catchment groundwater exchanges weaken the link between catchment precipitation and outlet flow (Le Moine et al., 2007). The rain-fed Esteron and Odet benefited most from same-month precipitation, which is consistent with their fast response and weak 12-month autocorrelation.

Among the algorithms, the regularised MLP and the RBF-SVR generalised best, whereas RF and GBM overfitted despite tuning. This differs from the generally strong performance of random forests reported in water resources applications (Tyrallis et al., 2019), but it follows from two features of the present design. First, the training samples were small (about 150 months), so the ensembles could nearly memorise the training set. Second, the chronological split placed several record flows in the test period, and tree-based models cannot extrapolate beyond the training target range. Smooth function approximators such as the MLP and SVR are better suited to this situation, which is common in non-stationary hydrological records. The finding that MLR was statistically indistinguishable from the best model in five of six catchments supports reporting a linear benchmark and a significance test. It is also consistent with Papacharalampous et al. (2019), who found that simple stochastic methods were often as accurate as ML methods for hydrological time series.

The study has limitations. The selected SVR and MLP hyperparameters lay at the edge of the search grid ($\gamma = 0.01$, $\alpha = 1$), so a wider grid might improve these models slightly. The log transformation introduced a small negative bias, which a smearing correction (Duan, 1983) or direct optimisation of KGE could reduce. Only six catchments and one-month lead time were considered, and deep sequence models such as LSTM (Kratzert et al., 2018), which need longer records, were not included. Future work should couple the S2 framework with ensemble seasonal forecasts, extend it to the full CAMELS-FR sample to relate skill to catchment attributes, and quantify predictive uncertainty.

5. Conclusions

Five data-driven models were compared for monthly streamflow prediction in six French catchments with contrasting hydro-climatic regimes, using open CAMELS-FR data for 1999–2018, a chronological test period and naive benchmarks. The main conclusions are:

(1) In a genuine one-month-ahead forecast (S2), the best models achieved NSE = 0.25–0.87 and KGE up to 0.92 and outperformed climatology and persistence in all catchments. However, the non-linear models improved little on multiple linear regression.

(2) Supplying same-month precipitation, temperature and PET (S3) increased NSE to 0.63–0.94 and reduced RMSE by 21–65%, and the improvement over S2 was statistically significant in five catchments. The value of meteorological forecasts is therefore largest in rain-fed catchments and in winter, and smallest in snowmelt-dominated catchments.

(3) The regularised MLP was the most accurate model overall (mean NSE = 0.85 in S3; best in five catchments), followed by SVR (0.80), which was best in the Alpine Ubaye. RF and GBM overfitted the short training records and underestimated floods that exceeded the training range.

(4) One-month lagged flow and same-month precipitation were the dominant predictors. Seasonal encoding and 12-month lagged flow were useful only in the nival regime, and flow lags beyond one month contributed little.

These results suggest that operational monthly forecasting systems should give priority to good initial-state information and to reliable precipitation forecasts, and should always report a linear model and climatology as benchmarks when evaluating machine learning models.

Data Availability Statement

The data used in this study are openly available in the CAMELS-FR dataset (Delaigue et al., 2024; <https://doi.org/10.57745/WH7FJR>) and in the airGRdatasets R package (<https://CRAN.R-project.org/package=airGRdatasets>), distributed under the CC BY 4.0 licence. The processing and modelling

scripts are available from the corresponding author on request.

Conflict of Interest

The authors declare no conflict of interest.

References

- Addor, N., Newman, A.J., Mizukami, N., & Clark, M.P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21, 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Akbarian, M., Saghaian, B., & Golian, S. (2023). Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran. *Journal of Hydrology*, 620, 129480. <https://doi.org/10.1016/j.jhydrol.2023.129480>
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124–137.
- Ayana, Ö., Kanbak, D., Kaya Keleş, M., & Turhan, E. (2023). Monthly streamflow prediction and performance comparison of machine learning and deep learning methods. *Acta Geophysica*, 71, 2905–2922. <https://doi.org/10.1007/s11600-023-01023-6>
- Bergmeir, C., & Benítez, J.M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213. <https://doi.org/10.1016/j.ins.2011.12.028>
- Bowden, G.J., Dandy, G.C., & Maier, H.R. (2005). Input determination for neural network models in water resources applications. Part 1—background and methodology. *Journal of Hydrology*, 301, 75–92.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Delaigue, O., Guimarães, G.M., Brigode, P., Génot, B., Perrin, C., & Andréassian, V. (2024). CAMELS-FR dataset, V1. *Recherche Data Gouv*. <https://doi.org/10.57745/WH7FJR>
- Delaigue, O., Guimarães, G.M., Brigode, P., Génot, B., Perrin, C., Soubeyroux, J.-M., Janet, B., Addor, N., & Andréassian, V. (2025). CAMELS-FR dataset: a large-sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking. *Earth System Science Data*, 17, 1461–1479. <https://doi.org/10.5194/essd-17-1461-2025>
- Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155–161.
- Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383), 605–610.
- Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, 56,

- e2019WR026793.
<https://doi.org/10.1029/2019WR026793>
- Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Gupta, H.V., Kling, H., Yilmaz, K.K., & Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Hamed, K.H., & Rao, A.R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204, 182–196.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.
- Kendall, M.G. (1975). *Rank Correlation Methods*. Charles Griffin, London.
- Kisi, O. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, 12(5), 532–539. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:5\(532\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532))
- Knoben, W.J.M., Freer, J.E., & Woods, R.A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23, 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22, 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Le Moine, N., Andréassian, V., Perrin, C., & Michel, C. (2007). How can rainfall–runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. *Water Resources Research*, 43, W06428. <https://doi.org/10.1029/2006WR005608>
- Lin, J.Y., Cheng, C.T., & Chau, K.W. (2006). Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, 51(4), 599–612. <https://doi.org/10.1623/hysj.51.4.599>
- Maier, H.R., & Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15, 101–124.
- Mann, H.B. (1945). Nonparametric tests against trend. *Econometrica*, 13, 245–259.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., & Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.
- Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., & Gupta, H.V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57, e2020WR028091. <https://doi.org/10.1029/2020WR028091>
- Ni, L., Wang, D., Singh, V.P., Wu, J., Wang, Y., Tao, Y., & Zhang, J. (2020). Streamflow and rainfall forecasting by two long short-term memory-based models. *Journal of Hydrology*, 583, 124296. <https://doi.org/10.1016/j.jhydrol.2019.124296>
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, 303, 290–306.
- Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2019). Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochastic Environmental Research and Risk Assessment*, 33, 481–514. <https://doi.org/10.1007/s00477-018-1638-6>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D., & Remesan, R. (2009). Daily pan evaporation modeling in a hot and dry climate. *Journal of Hydrologic Engineering*, 14(8), 803–811. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000056](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000056)
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchistéguy, L., & Morel, S. (2008). Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47, 92–107.
- Rasouli, K., Hsieh, W.W., & Cannon, A.J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 414–415, 284–293. <https://doi.org/10.1016/j.jhydrol.2011.10.039>
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American Statistical Association*, 63, 1379–1389.
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 11(5), 910. <https://doi.org/10.3390/w11050910>
- Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.
- Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., & Soubeyrou, J.-M. (2010). A 50-year high-resolution

- atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, 30, 1627–1644. <https://doi.org/10.1002/joc.2003>
- Wang, W.C., Chau, K.W., Cheng, C.T., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374, 294–306. <https://doi.org/10.1016/j.jhydrol.2009.06.019>
- Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., & Sayl, K.N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, 530, 829–844. <https://doi.org/10.1016/j.jhydrol.2015.10.038>