Benchmarking seasonal forecasting skill using river flow persistence in Irish catchments

Dáire Foran Quinn, Conor Murphy, Robert L. Wilby, Tom Matthews, Ciaran Broderick, Saeed Golian, Seán Donegan & Shaun Harrigan

View supplementary material

Published online: 16 Mar 2021.

Submit your article to this journal

Article views: 847

View related articles

View Crossmark data

Citing articles: 2 View citing articles

IAHS AISH

Taylor & Francis
Taylor & Francis Group

Benchmarking seasonal forecasting skill using river flow persistence in Irish catchments

Dáire Foran Quinn[a], Conor Murphy[a], Robert L. Wilby[b], Tom Matthews[b], Ciaran Broderick[c], Saeed Golian[a], Seán Donegan [iD][a] and Shaun Harrigan[d]

[a]Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Co., Kildare, Ireland; [b]Department of Geography and Environment, Loughborough University, Loughborough, UK; [c]Flood Forecasting Division, Met Éireann, Dublin, Ireland; [d]Forecast Department, European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK

ABSTRACT

This study assesses the seasonal forecast skill of river flow persistence in 46 catchments representing a range of hydrogeological conditions across Ireland. Skill is evaluated against a climatology benchmark forecast and by examining correlations between predicted and observed flow anomalies. Forecasts perform best when initialized in drier summer months, 87% of which show greater skill relative to the benchmark at a 1-month horizon. Such skill declines as forecast horizon increases due to the longer time a catchment has to "forget" initial anomalous flow conditions and/or to be impacted by "new" events. Skill is related to physical catchment descriptors such as the baseflow index (correlation $\rho = 0.86$) and is greatest in permeable high-storage catchments. The distinct seasonal and spatial variations in persistence skill allow us to pinpoint when and where this method can provide a useful benchmark in the future development of more complex seasonal hydrological forecasting approaches in Ireland.

1 Introduction

Seasonal hydrological forecasting (SHF) can play an important role in the operation and management of water resources by enhancing preparedness and informing decision-making (Wilby 2001, Wedgbrow et al. 2002, Anghileri et al. 2016, Tang et al. 2016, Viel et al. 2016, Prudhomme et al. 2017, Dixon and Wilby 2019). For instance, skilful predictions of future streamflow weeks to months in advance can help reservoir managers balance flood-control safety (Amnatsan et al. 2018) with water security during drought conditions (Watts et al. 2012). Such forecasts can also improve hydropower productivity (Hamlet et al. 2002), agriculture (Mushtaq et al. 2012), tourism (Fundel et al. 2013) and inland water transport (Meißner et al. 2017). This is because foresight enables planning and management of anomalous hydrological conditions in advance of their adverse effects, which is more cost effective than reactive management (Pappenberger et al. 2015b).

Although SHF is growing in both feasibility and importance globally, a universally best-performing method does not exist (Yuan et al. 2015). Therefore, when establishing seasonal forecasting capability, it makes sense to begin with basic approaches to river flow prediction as the various techniques show skill in different hydro-climatic contexts. The simplest possible model is sometimes referred to as a zero-order forecast (Dixon and Wilby 2016) and is usually based on climatology (i.e. long-term mean flow) or persistence. SHF based on river flow persistence is straightforward to implement because the method uses the most recently observed flow anomaly as the forecasted anomaly.

Assessing the persistence of standardized flow anomalies, rather than absolute flow values, captures the distinct seasonal cycle of river flows. Factors to be considered in the generation of persistence forecasts include the duration of the predictor period (i.e. the time span over which the "most recent" flow anomaly is calculated) and the forecast horizon (i.e. the time interval for which this forecast is made). In an analysis of persistence forecast skill for the UK, Svensson (2016) found statistically significant correlations between hindcasts and observations for 78% of station–month combinations using a 1-month predictor period and a 1-month forecast horizon. However, skill depended greatly on the "memory" inherent to each catchment. The memory time scale essentially reflects the amount of storage and can be defined as the length of time needed to "forget" an imposed anomaly (Ghannam et al. 2016). Hence, hydrological memory of antecedent conditions generally provides a baseline source of predictive skill for slowly changing flows in catchments with a relatively large groundwater storage capacity (Svensson et al. 2015).

The temporal and spatial distribution of persistence skill can be compared with other methods that do not depend on information about future atmospheric conditions as the source of predictability but instead rely solely on initial hydrological conditions, as with ensemble streamflow prediction (ESP) (e.g. Day 1985, Harrigan et al. 2018). Such methods are particularly important in Northern Europe due to the limited long-term predictability of precipitation, especially in the summer (Weisheimer and Palmer 2014). For instance, Arnal et al.

(2018) observe that in Europe, an SHF system driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 dynamical seasonal climate forecasts surpasses the ESP method at predicting river flow anomalies for the first-month forecast horizon only, with ESP on average more skilful at longer forecast horizons.

Operational SHF systems already exist in some countries. For example, the Hydrological Outlook has been delivering streamflow and groundwater forecasts for UK regions on a monthly basis since June 2013 (Prudhomme *et al.* 2017). Although the European Flood Awareness System (EFAS) provides a seasonal flow forecast up to 7 months in advance across 74 European regions, one of which is the island of Ireland (Arnal *et al.* 2018), no operational service currently exists in the Republic of Ireland to deliver such forecasts at catchment scales. Given the heterogeneous geological and hydrological conditions across the country, as well as the variable climate, streamflow conditions can vary significantly between catchments (Webster *et al.* 2014). Following the summer drought of 2018, water managers in Ireland expressed interest in the potential application of SHF. However, to the authors' knowledge, no previous research has assessed the potential of SHF, or the value of different forecast methods for Ireland.

Using Ireland as a test case, we develop a simple statistical approach to SHF based on flow persistence that can be used to benchmark more sophisticated procedures. In order to justify the additional overhead associated with using more complex techniques, their skill should exceed the baseline set by the zero-order forecast (Barnston *et al.* 1994). Having a benchmark besides streamflow climatology is important when evaluating SHF because reference methods that are easy to outperform can create overconfidence in such systems (Pappenberger *et al.* 2015a).

In addition, we look at the potential for the persistence forecasts themselves to be used in an operational context in Ireland, as they are in the UK. A thorough investigation of flow persistence contributes to a deeper understanding of the predictability of Irish river flows. Hence, we (1) evaluate the monthly to seasonal hydrological forecast skill of flow persistence for Irish catchments; (2) investigate where and when persistence offers skill beyond climatology, using forecast horizons from 1 week to 3 months; and (3) identify which catchment characteristics determine when and where persistence skill is greatest.

2 Data and methods

2.1 River flow data and catchment characteristics

Forty-six river flow records were selected from the national hydrometric register (see Supplementary material, Table S1) following the broad criteria used by Murphy *et al.* (2013b) when creating the Irish Reference Network. These standards require consistent hydrometric data quality; at least 25 years of record; and a "near-natural" flow regime, identified in catchments which are minimally impacted by human activity (e.g. zero or stable water abstractions). These catchments are, therefore, relatively free from factors that could confound the interpretation of hydrological forecasting skill. The catchment sample is representative of the 215 gauged catchments of the Office of Public Works (OPW) Flood Studies Update (FSU)

(Mills *et al.* 2014). Our set also provides good spatial coverage across Ireland's diverse hydrological, hydrogeological and climatic conditions (Broderick *et al.* 2016, 2019).

For the purpose of analysing the spatial variations in forecast performance, stations were categorized into geographic regions broadly corresponding to the third level of the European Union Nomenclature of Territorial Units for Statistics (NUTS) (Fig. 1). As only one catchment in the sample lies in the NUTS III Dublin region, this region was combined with the NUTS III Mid-East region to form the East region. Daily mean river flow series (m^3/s) for each gauging station were obtained from the OPW and the Environmental Protection Agency. The longest record used in this network begins in 1954, but other series typically extend from the 1970s onwards. After removing missing data, the maximum and minimum record lengths in the station network are 63 and 25 years, respectively, with 43 years being the mean.

A wide range of physical attributes were specified for each catchment by the OPW as part of the FSU (Mills *et al.* 2014). These physical descriptors characterize catchment morphology, soil and climate (see Supplementary material, Table S2). The baseflow index (BFI) measures the proportion of catchment streamflow that is derived from baseflow or saturated groundwater storage as opposed to direct runoff (Gustard *et al.* 1992). By quantifying the contribution of stored sources to total runoff, the BFI is indicative of the inertia in streamflow variation and is, therefore, similar to the Richards-Baker flashiness index (Baker *et al.* 2004). The FSU developed the soil baseflow index (BFI_{soil}) to enable estimation of BFI at ungauged sites. Mills *et al.* (2014) modelled catchment BFI_{soil} using descriptors of soil drainage, subsoil permeability, aquifer productivity and climate. BFI_{soil} is essentially an index of catchment permeability and is strongly correlated with the gauged BFI values of catchments in our sample (Spearman's rank correlation, $\rho = 0.97$).

2.2 Persistence forecast initialization

The river flow anomaly observed over a specific "predictor period" (e.g. the month of January) is used as the predicted anomaly for the immediately following "forecast horizon" period (e.g. the month of February). Variations in persistence forecast skill were examined with respect to the duration of these predictor and forecast horizon periods. In each case, the same six durations were tested: 1, 2 and 3 weeks; and 1, 2 and 3 months. For consistency, a "week" is defined as a quarter-monthly period and a "month" then represents any four consecutive "weeks." As such, all predictor periods end on one of the year's 48 so-defined weeks, and every year, a persistence forecast is initialized on the last day of each predictor period (assuming, for future operational purposes, that all relevant flow data are available by the end of the last day). As an example, Table 1 illustrates a range of possible predictor periods that could be used for a persistence forecast initialized at the end of the first week of August for a particular catchment.

Each of the six predictor periods was combined with each of the six forecast horizon periods, meaning that 36 predictor–forecast horizon period combinations were tested for each catchment. Using a combination of a 1-week predictor period and a 1-month
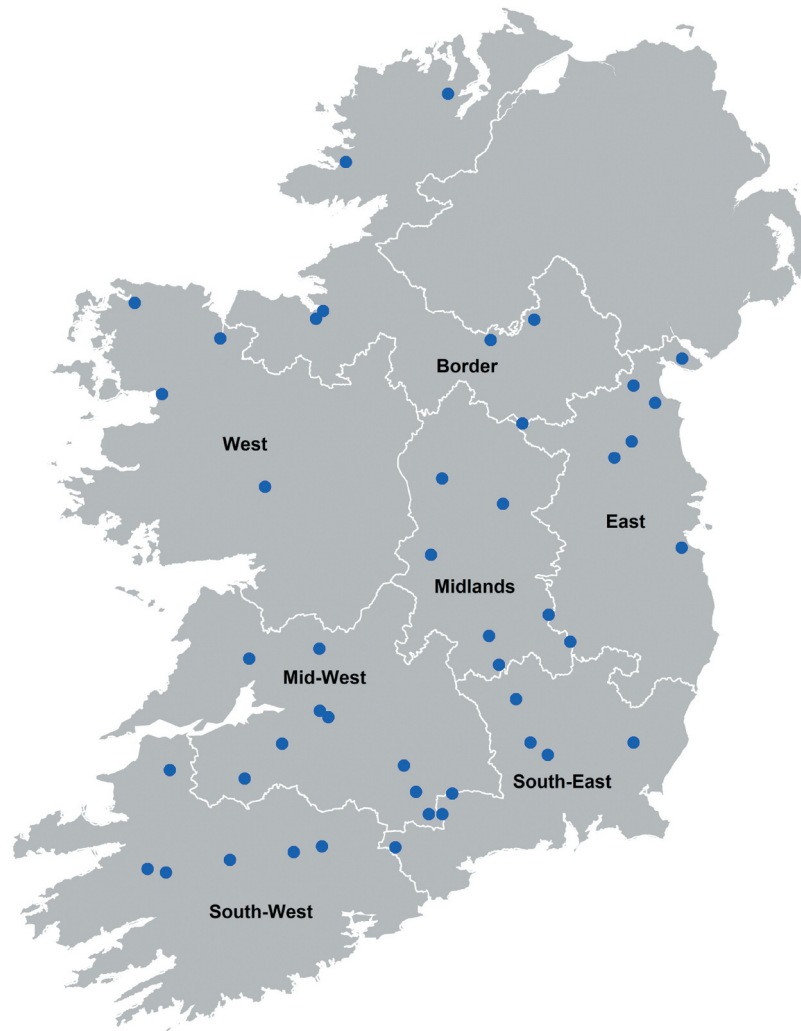
Figure 1. Regions in the Republic of Ireland, based on the Nomenclature of Territorial Units for Statistics (NUTS) Level III classification, and location of the 46 stations (blue dots).

Table 1. Sample predictor periods used for the persistence forecast initialized at the end of the first week of August.

Duration of predictor period	Weeks that form the predictor period
1 week	First week of August
1 month	Last three weeks of July and first week of August
3 months	Last three weeks of May, all weeks in June and July, and first week of August

forecast horizon, for example, the river flow anomaly over the final week of October would be used to predict the anomaly for the entire month of November. For any given predictor–forecast horizon combination, the predictor-based flow anomalies are therefore used as the hindcast series; the observations in the forecast period are used for model evaluation.

2.3 Calculation of the standardized river flow anomalies

Hindcast and observed series of standardized flow anomalies were generated using daily river flow series at each station. These data were used to calculate the mean flow experienced over the predictor period or forecast horizon being tested. For example, a hindcast series of mean weekly flow values was created

for the 1-week predictor period, while an observation series of mean monthly flow values was created for the 1-month forecast horizon period. If more than 20% of daily observations in a given period were missing, then this block was not used in the analysis. Following Svensson (2016), mean flows were log-transformed to reduce the influence of extreme flows. For each station, i, log-transformed mean flow values, $x_{i,p}$, for a given period, p, of each year were used to calculate the long-term climatological mean flow, $\bar{x}_{i,p}$, (Equation (1)) and standard deviation, $s_{i,p}$ (Equation (2)) as follows:

$$\bar{x}_{i,p} = \frac{\sum x_{i,p}}{n_{i,p}} \tag{1}$$

$$s_{i,p} = \sqrt{\frac{\sum \left(x_{i,p} - \bar{x}_{i,p}\right)^2}{n_{i,p} - 1}} \tag{2}$$

In any given year, t, the period's log-transformed mean flow value, $x_{i,p,t}$, was then converted into a standardized flow anomaly, $z_{i,p,t}$, using the following equation:

$$z_{i,p,t} = \frac{x_{i,p,t} - \bar{x}_{i,p}}{s_{i,p}} \qquad (3)$$

In this way, the distribution of standardized flow anomalies for a given predictor or forecast period at a given station has a mean of 0 and a standard deviation of 1. This standardization approach takes the seasonal cycle of flows into account and enables the comparison of different hydrological regimes. Predicted anomalies can be converted back into predicted flows by reversing the standardization of Equation (3).

2.4 Evaluation of the forecast performance

The performance of persistence forecasts produced from each predictor–forecast combination was evaluated using Pearson's correlation coefficient (r). This quantifies the linear association between various hindcast and observed series. As outlined in the following sections, the relative predictive performance of the persistence forecasts was then evaluated (1) against the performance of the streamflow climatology benchmark forecasts, and (2) using the persistence forecast skill metrics applied by Svensson (2016) for the UK.

2.4.1 Performance against the benchmark

To assess the usefulness of river flow persistence as a potential benchmark in the development of SHF in Ireland, persistence forecasts were evaluated against a river flow climatology benchmark. To measure the forecast accuracy of both approaches, the mean squared error (MSE) between hindcast and observed standardized flow anomalies was calculated for each station and predictor–forecast horizon period combination. Although the persistence method produces a different MSE in each case, climatology produces an MSE of 1. This is because climatology uses the long-term mean as the predicted flow anomaly, hence the values of a standardized distribution have an average squared difference of 1 from the mean. The mean squared error skill score (MSESS) was then used as the deterministic verification metric to assess the improvement (or lack thereof) of the persistence method over climatology in each case. Adapted from the generic skill score equation of Murphy (1988), the MSESS was calculated as follows:

$$\text{MSESS} = -\text{MSE}_\text{p} + 1 \qquad (4)$$

where MSE_p represents the MSE of the persistence method. MSESS values range from $-\infty$ (least skilful) to 1 (perfectly skilful), with any positive value representing skill relative to the benchmark.

2.4.2 Comparison with persistence forecasts for the UK

In order to evaluate the relative performance of persistence forecasts in Ireland, they were compared with the method of Svensson (2016) for measuring usability of persistence forecasts across 93 UK stations using a 1-month predictor period and a 1- to 3-month forecast horizon. This involved evaluating the significance of the correlations between each month's series of hindcast anomalies and observed anomalies at the 95% confidence interval using a Pearson's one-tailed correlation test. The significant correlation threshold varied between $r \geq 0.34$ for the shortest record length and $r \geq 0.21$ for the longest record length; but, in line with Svensson (2016), only station–month combinations with significant positive correlations exceeding 0.23 were regarded as potentially "usable" forecasts.

Moderate yet significant autocorrelation ($p < 0.05$) was detected at lag = 1 in observation series for a greater number of stations than would be expected by chance (>5%) in February (33%), June, July, August and October (8–10%). In these cases, the observed standardized anomalies are significantly correlated with the standardized anomalies experienced in the same month the year prior. As such, we note that the critical r values for rejecting the null hypothesis inferred from the t-test are likely to be biased in these cases (Santer et al. 2000). However, given the aim of comparing with the UK, where such biases were not accounted for, we do not adjust for this autocorrelation.

2.5 Assessing the influence of catchment characteristics

The relationship between each physical catchment descriptor and the predictive skill of river flow persistence was evaluated using Spearman's rank correlation (ρ) to help explain why flow anomalies tend to persist more in some catchments than others. An all-subsets regression approach was then used in conjunction with cross-validation to find the best multivariate regression models at predicting catchment annual and seasonal average persistence skill. Despite many of the physical catchment descriptors being significantly cross-correlated, all were considered potential predictors. This is because the relationship between the physical descriptors is not always linear and can vary depending on other factors. The "leaps" package in R (Lumley and Miller 2009) was employed to perform an exhaustive search for the best subsets of predictor variables using a branch-and-bound algorithm. Models were generated using these predictor subsets and evaluated based on the significance of each predictor ($p > 0.05$) and the adjusted coefficient of determination (R^2). The latter identifies the proportion of variability in forecast skill across the sampled catchments that is explained by the statistical model, after adjusting for the number of predictors. Potential multicollinearity between the explanatory variables was assessed using the variance inflation factors and tolerance values.

A leave-one-out cross-validation (LOOCV) was then performed to formally assess the predictive capability of the models. This involved successively leaving out one station from the training dataset and estimating models based on the remaining 45 stations, keeping the model structure the same. The predictive accuracy of each model was indicated by the root mean square error (RMSE) and the mean absolute error (MAE) between the predicted and observed skill values. To illustrate the regression for identifying catchments where persistence-based forecasts are likely to be successful, the best-performing models were employed to predict the persistence skill at an additional 169 stations within the entire FSU gauged catchment set.

2.6 Categorical forecast verification

To assess the potential of persistence forecasts in an operational context, the standardized river flow anomalies were categorized as high, medium or low flows and a forecast verification was carried out for each set of forecasts using contingency tables. These show the hindcast distribution along rows and the

observed distribution in columns. The 70–99% time exceedance range of the flow duration curve is widely used to classify low flows (Smakhtin 2001). As the flow duration curve varies by season, we define a standardized anomaly as a low flow if it falls below the 30th-percentile standardized anomaly threshold of the period. Meanwhile, the high-flow category consists of standardized anomalies above a 70th-percentile threshold, and the medium flow category contains the standardized anomalies between these two percentile thresholds.

3 Results

3.1 Influences on the performance of persistence forecasts

3.1.1 Predictor–forecast horizon period combination

Across all catchments and predictor periods, persistence skill declines with increasing forecast horizon. Figure 2(a) shows the decay in network-wide correlations between hindcast and observed anomalies for the 1-week forecast ($r = 0.69$) through to the 3-month ahead forecast ($r = 0.31$) when holding the predictor period at 1 week. The upper (95th percentile) and lower (5th percentile) bands show considerable variation in forecast performance at each horizon across the station network. For example, the 1-month forecast horizon has mean $r = 0.52$ with 95th percentile range spanning $r = 0.18$ to $r = 0.81$.

Similarly, across all catchments and forecast horizons, annual average persistence skill declines as the duration of the predictor period increases. Figure 2(b) shows a gradual decrease in network-wide correlations between hindcast and observed anomalies for the 1-week predictor period ($r = 0.52$) through to the 3-month predictor period ($r = 0.24$), when holding the forecast horizon at 1 month. Again, there is significant variation about the mean depending on catchment. There are also certain times of the year when this decline in performance is less marked, particularly for forecasts initialized in early January and in July, where the 1-week and 1-month predictor periods produce forecasts of similar accuracy. Overall, the persistence forecasts perform best using the 1-week predictor period and 1-week forecast horizon. However, as we focus on monthly to seasonal hydrological forecasting, this 1-week predictor period was combined with 1- to 3-month forecast horizons.

3.1.2 Forecast initialization month

Performance of the flow persistence forecasts varies throughout the year. Figure 3 shows the network-wide mean correlations between hindcasts and observations for each initialization week using the 1-week predictor period and 1-month forecast horizon. A distinct seasonal pattern can be identified in forecast performance, with summer months (JJA) having the highest seasonal mean correlation coefficient ($r = 0.66$) and winter months (DJF) having the lowest ($r = 0.44$). Spring (MAM) and autumn (SON) have similar mean correlation values of $r = 0.49$ and 0.48, respectively. This seasonality of forecast performance becomes more pronounced as longer predictor periods and/or forecast horizons are used. Figure 3 also shows negative correlation between forecast performance and both average daily precipitation ($r = −0.51$) and river flow ($r = −0.79$). Therefore, despite week-to-week variability, a general pattern of forecast improvement is evident

as conditions become drier. One notable deviation from this trend is seen for forecasts initialized in mid-winter, with the correlation between hindcasts and observations in the last week of December rising to $r = 0.58$, despite being on average one of the wettest times of the year.

3.1.3 Physical catchment characteristics

Across all predictor and forecast horizon periods, the physical catchment descriptors that had the strongest association with catchment persistence skill were BFI_{soil} and standard-period (1961–1990) annual average rainfall (SAAR). Figure 4 illustrates these network-wide relationships using a 1-week predictor period and 1-month forecast horizon. A strong positive correlation is seen between median persistence skill and BFI_{soil} ($\rho = 0.86$), indicating that persistence skill is greater for catchments with higher storage capacities. This is consistent with a negative relationship between persistence and the flashiness index of a catchment ($\rho = −0.68$). Additionally, moderate positive correlations were found between persistence skill and the physical descriptors related to the size of the catchment, such as area ($\rho = 0.39$) and main-channel length ($\rho = 0.36$). Weak insignificant positive correlations were found with the flood attenuation indicators (FAI, $\rho = 0.27$; FARL, $\rho = 0.14$) and standard-period average annual potential evapotranspiration (SAAPE, $\rho = 0.13$).

Meanwhile, skill is moderately negatively correlated with SAAR ($\rho = −0.66$) and other physical descriptors linked to the wetness of a catchment, including forest cover ($\rho = −0.57$), peat bog cover ($\rho = −0.56$) and the proportion of time soils can expect to be typically quite wet (FLATWET, $\rho = −0.36$). Catchment mean elevation and slope are also negatively correlated with skill, indicating that flows are less likely to persist in more upland catchments with steep gradients. This correlation is stronger when using the S1085 gradient measure ($\rho = −0.59$), which calculates main-channel slope excluding the bottom 10% and top 15% of its length, than when using the Taylor-Schwartz measure (TAYSLO, $\rho = −0.46$). The latter divides the main-channel route into a series of individual slopes, each 500 m in length, and finds the average gradient of these slopes.

Generally, similar relationships emerge between these physical catchment descriptors and persistence skill using longer forecast horizons, but the correlation between BFI_{soil} and skill weakens as the forecast horizon increases ($\rho = 0.77$ at the 3-month horizon). Multicollinearity was an issue when using some combinations of physical catchment descriptors in the regression models. For example, predictor subsets that included both BFI_{soil} and descriptors related to catchment size showed high cross-correlations, as surface runoff dominates the hydrology of smaller headwater catchments, whereas larger catchments incorporate flatter spaces with greater storage potential.

The best-performing multiple linear regression model used BFI_{soil}, SAAR, SAAPE, S1085 and TAYSLO as predictors of annual median catchment persistence skill for the 1-week predictor period and 1-month forecast horizon (Table 2), yielding an adjusted R^2 of 0.90. Overall, BFI_{soil} is the most important predictor, explaining 78% of the variation in skill across the sample. This rises to 87% by including SAAR, 89% by adding SAAPE and 90% when incorporating both slope measures. Under cross-validation, the R^2 value declined slightly to 0.89
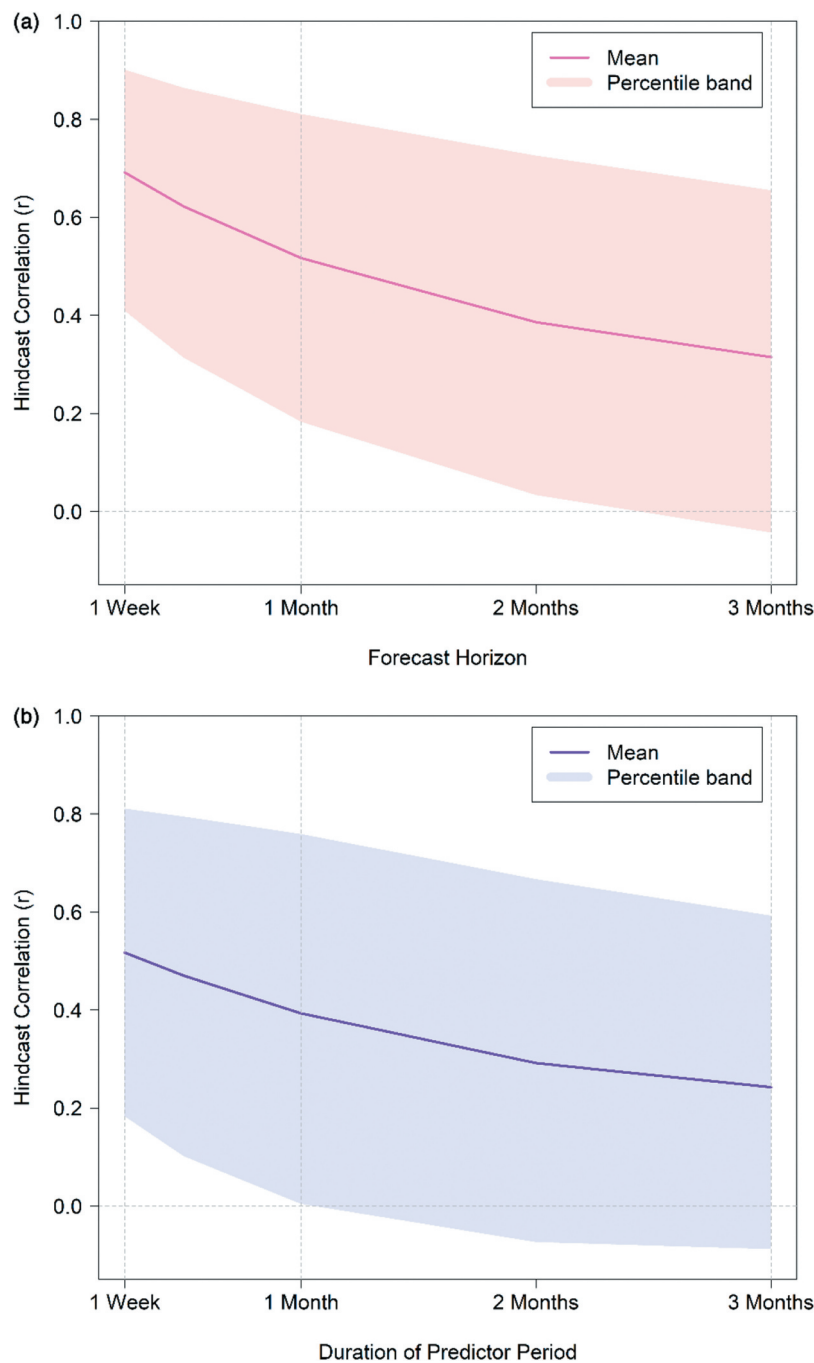
Figure 2. Network-wide persistence forecast performance, measured by the correlation (*r*) between hindcasts and observations, plotted against: (a) forecast horizon, using a 1-week predictor period, and (b) duration of the predictor period, using a 1-month forecast horizon. The spread of persistence skill across the network is indicated by the upper (95th percentile) and lower (5th percentile) bands.

and the average error made by the model in predicting catchment skill (MSESS) was found to be relatively low (MAE = 0.056). As the RMSE squares errors before they are averaged, the relatively higher RMSE of 0.072 indicates variation in the magnitude of errors. Using the principle of parsimony, a three-predictor model excluding the slope measures yields a comparably high adjusted R^2 value (0.89) and, under cross-validation, only a slightly higher MAE (0.061), and so may be more widely applicable. Satisfactorily low variance inflation factors (< 3) and high tolerance levels (> 0.4) indicate that multicollinearity was not an issue in either the three- or five-predictor models.

Both main-channel slope measures only contribute significantly (p < 0.05) to the model when applied in tandem. Catchments that rank among the steepest in the sample using the S1085 measure, but rank notably lower based on the TAYSLO measure, tend to be smaller and more responsive. Meanwhile, the catchments that also have a relatively high gradient using TAYSLO do not tend to be as small. This is because TAYSLO is sensitive not only to overall catchment gradient but also to the range of individual slopes between reaches. For example, the Kings at Annamult (444 km^2) and the Graney at Scariff (280 km^2) have comparable gradients of 3.6 m/km and 3.9 m/km, respectively,
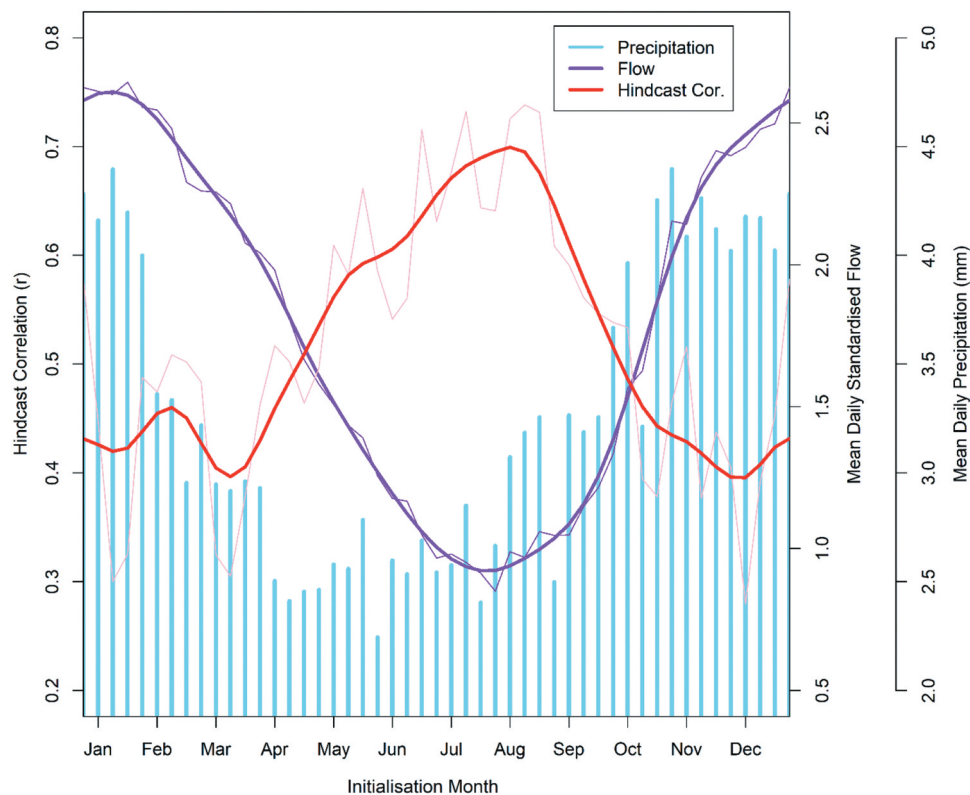
Figure 3. Network-wide weekly mean daily precipitation (mm) and weekly mean standardized flow compared with mean persistence forecast performance, measured by the correlation (*r*) between hindcasts and observations using a 1-week predictor period and 1-month forecast horizon. Smoothing splines (smoothing parameter = 0.4) highlight the general patterns of river flow and skill throughout the year.

using S1085, but the gradients are 1.05 m/km and 0.26 m/km, respectively, with TAYSLO. Thus, despite the covariance between slope descriptors ($\rho = 0.83$), they are more useful in estimating catchment skill when used together in regression models.

The best regression model for predicting skill at the 3-month horizon utilizes a different predictor subset (Table 2), which yields both a lower adjusted R^2 ($R^2 = 0.82$; cross-validated $R^2 = 0.79$) and a higher error (MAE = 0.069, RMSE = 0.096). The optimal predictor-subset and model performance also varies by season (Table 3). When BFI_{soil} is used as the sole predictor of 1-month persistence skill, a notably higher R^2 value is produced for winter (0.76) and autumn (0.71) skill than for summer (0.60) or spring (0.54) skill. It is on the rising limb of the annual hydrograph, as the proportional split between baseflow and quick-response runoff increases, that the estimated baseflow contribution becomes an even more important determinant of whether flow anomalies will persist or not. The decline in regression model performance in predicting annual and seasonal skill at the 3-month forecast horizon (see Supplementary material, Table S3) is explained by the weakening correlation between BFI_{soil} and flow persistence over longer durations.

3.2 Relative skill of the persistence forecasts

Results presented in this section, unless otherwise specified, refer to persistence forecasts based on the 1-week predictor period. Similar analysis with lower levels of usability of persistence forecasts based on the 1-month predictor period is summarized in the Supplementary material (Figs S1–S3). The following qualitative descriptors are used to categorize MSESS values as high (0.5–1), moderate (0.25–0.5), low (0–0.25) and no skill (≤ 0). Figure 5 uses four example hindcast time series to illustrate variations in skill across these different categories.

3.2.1 Performance against the benchmark

The majority (58%) of persistence forecasts produced across the catchment sample perform better than the streamflow climatology benchmark at the 1-month horizon (Fig. 6). This proportion is higher for forecasts initialized during the summer months (87%), compared with spring and autumn (both 53%). Winter is the least skilful season, with only 41% of simulations outperforming climatology. The most skilful predictor month is August, with a high median MSESS in its first 2 weeks and moderate median MSESS in the final 2 weeks. March has the lowest average skill score, with a significant proportion of forecasts surpassing the benchmark only when the final week is used as the predictor period. Figure 6 shows the decay in skill when using longer forecast horizons with only 33% of persistence forecasts outperforming the benchmark for the 2-month horizon and 23% for the 3-month horizon – mainly confined to the summer months in both cases. The general seasonal pattern of forecast skill remains similar at each horizon. The level of variation across the catchment sample in each initialization week decreases when sub-samples of catchments with similar physical characteristics, such as storage capacity and annual
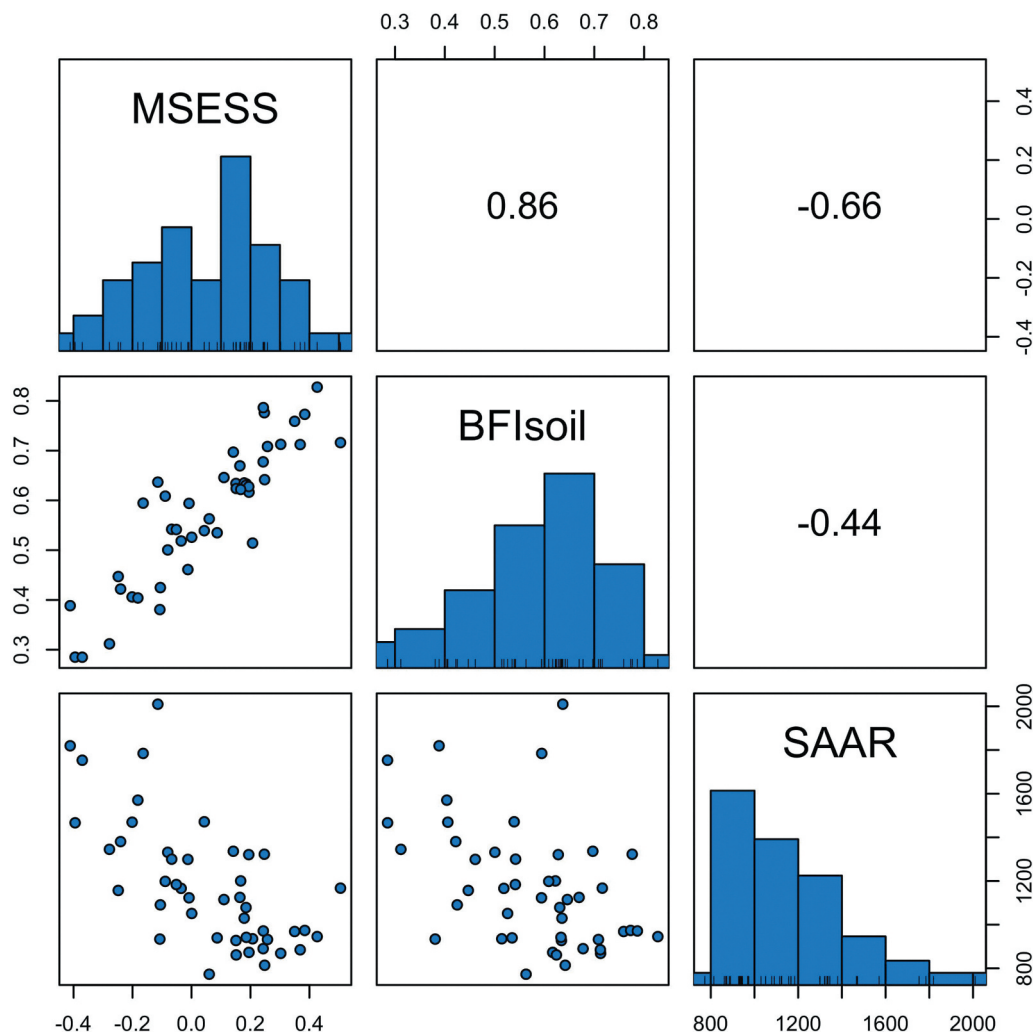
Figure 4. Spearman's correlation (ρ) between the median persistence skill (MSESS), baseflow index (BFIsoil), and standard-period annual average rainfall (SAAR) of each catchment in the sample. The median skill is for the 1-week predictor period and 1-month forecast horizon. Histograms show the distribution of each variable.

average rainfall, are considered alone. For example, using a sample conditioned on a BFI_{soil} threshold of 0.6 (which includes approximately 52% of the stations), almost 95% of forecasts are skilful in summer (see Supplementary Fig. S4).

3.2.2 Comparison with the UK

Figure 7 compares the proportion of "usable" forecasts in the Irish station network with the proportion of "usable" forecasts reported by Svensson (2016) for the UK using the same criteria of usability (i.e. the correlation between hindcast and observed flow anomalies exceeds 0.23 with $p \leq 0.05$). The usability rates of forecasts initialized at the end of each calendar month using a 1-month predictor period are presented to enable direct comparison with Svensson (2016) for the UK.

A similar pattern can be identified in the seasonality of persistence forecast performance in both countries at each forecast horizon. Looking at the 1-month forecast horizon (Fig. 7(a)), it can be seen that the forecast usability rate peaks during the summer in Ireland (93%) and the UK (94%) before declining in autumn, albeit more markedly in Ireland (where the rate drops to 55%, compared with 75% in the UK). A significant increase in this ratio is noted in both countries

for December, rising to 89% in Ireland and 86% in the UK; but a steep decline is observed in the following months, with the lowest forecast usability rate being found in March for both Ireland (35%) and the UK (52%). For the whole year, a relatively high proportion of persistence forecasts are "usable" in both Ireland (69%) and the UK (78%) at this 1-month horizon. However, when the forecast horizon is increased to 3 months (Fig. 7(b)) there is a greater reduction in the overall usability ratio for Ireland (where it falls to 46%) than in the UK (66%). Although Irish forecast usability levels are comparable to those of the UK during summer and in January, persistence skill almost disappears in Ireland during February (2%), March (10%) and November (6%).

3.3 Spatial distribution of skill

A broad pattern of persistence skill can be identified across Ireland, influenced by the spatial variation in catchment storage and wetness characteristics. Stations with the best-performing persistence forecasts are mainly found in the Midlands, East and South-East regions – collectively, these have median $r = 0.59$ between hindcasts and observations.

Table 2. The best-performing multiple linear regression equations in predicting annual median catchment persistence skill (MSESS) using a 1-week predictor period.

Forecast horizon	Regression equation
1 month	MSESS = 1.232*BFI$_{soil}$ − 0.00023*SAAR + 0.0018*SAAPE − 0.0069*S1085 + 0.0385*TAYSLO
3 months	MSESS = 1.077*BFI$_{soil}$ − 1.058*FLATWET + 0.512*FARL + 2.219*ALLUV −0.358*FOREST

Table 3. The best-performing multiple linear regression models in predicting annual and seasonal median catchment persistence skill (MSESS) using the 1-week predictor period and 1-month forecast horizon, alongside their adjusted R^2 values. Their respective root mean square error (RMSE) and mean absolute error (MAE) values, measured during cross-validation, are included.

Time of year	Predictors	Adjusted R^2	Cross-validated R^2	MAE	RMSE
Overall	BFI$_{soil}$, SAAR, SAAPE, S1085, TAYSLO	0.90	0.89	0.056	0.072
Summer (JJA)	BFI$_{soil}$, SAAR, S1085, TAYSLO, FAI, SAAPE, FOREST	0.88	0.86	0.074	0.094
Autumn (SON)	BFI$_{soil}$, FARL, FOREST, ALLUV	0.87	0.85	0.085	0.105
Winter (DJF)	BFI$_{soil}$, SAAPE	0.85	0.84	0.067	0.081
Spring (MAM)	BFI$_{soil}$, SAAR	0.71	0.68	0.102	0.124

Conversely, significantly lower median r = 0.41 is found between hindcast and observed anomalies in the catchments of the Border, West and South-West regions. Figure 8 shows the variation in forecast skill (MSESS) across the year in all 46 stations, grouped by region. The most skilful region (the Midlands) is found at the base of the plot, with regions becoming, on average, progressively less skilful moving up the graph. This "DNA of persistence" skill conveys the heterogeneity within these broad regions (e.g. station 6030, River Big at Ballygoly, is notably less skilful than most other catchments in the East region). The regional differentiation of persistence skill outside summer becomes less apparent as the forecast horizon increases. As skill is highest when initialized in the summer months at the 1-month horizon, Fig. 9 maps these median summer MSESS values. The median BFI$_{soil}$ (Fig. 9(a)) and SAAR (Fig. 9(b)) of the studied catchments in each region reveal the influence of these catchment characteristics on the spatial distribution of skill.

The regression models that provided the most accurate predictions of the sample catchments' median seasonal MSESS were used to infer the likely average persistence skill of the larger FSU 215-catchment set (Fig. 10). This provides an overview of the expected performance of the flow persistence method outside the training set, allowing us to explore potential utility of the technique as an operational forecasting tool at the national scale. Consistent with the spatial pattern of persistence skill shown by the sample, the Midlands, East and South-East are the only regions where most stations outperform the climatology benchmark across all seasons. When the forecast horizon is increased to 3 months, a negative median seasonal skill is predicted for virtually all stations outside summer (see Supplementary material, Fig. S5). However, the regression models are less accurate at predicting catchment skill at this horizon (see Supplementary material, Table S3).

3.4 Precision and verification of the forecasts

Contingency tables reveal the performance of categorical forecasts by cross-referencing hindcasts with observations (Fig. 11). A perfect forecast would have all entries along the diagonal, as the forecasted flow anomaly category would always be the same

as the observed category. For comparison, these tables also show the percentages that would be expected in each bin if the hindcasts were randomly distributed (i.e. using a forecasting method with no skill). Compared to the random distribution, the proportion of each flow category that is correctly (incorrectly) forecasted is found to increase (decrease), most notably at the 1-month horizon. For example, the 1-month forecast contingency table shows that 2.7% of total flows were observed to be low but inaccurately forecasted to be high. As low flow observations represent 30% of the total flows, this means that 9% of all low flows were predicted in the wrong extreme. When more extreme flow categories are used (e.g. the 10th and 90th percentile flow thresholds), the percentage of observed flows forecasted as the opposite extreme drops significantly in all cases.

The performance of the persistence forecasts is generally better for low flows. For example, across the entire catchment sample, high flows are 30% more likely than low flows to be predicted in the wrong extreme. However, this difference is seldom observed in catchments with a low storage capacity. High flows are only 6% more likely than low flows to be predicted as the wrong extreme when looking at the 10 lowest BFI$_{soil}$ stations, compared with a 62% greater likelihood for the 10 highest BFI$_{soil}$ stations.

4 Discussion

4.1 When is river flow persistence skilful?

The highest overall flow persistence skill was found when the predictor and forecast horizon period each had a duration of 1 week, with skill generally declining as their respective durations increased. This is because persistence skill hinges on the strength of catchment "memory." The longer the forecast horizon, the more time a catchment has to "forget" the anomalous river flow conditions of the predictor period, due to an increased chance of weather perturbing the status quo. This is consistent with Svensson (2016), who found that the proportion of usable persistence forecasts declines from 78% to 66% when moving from a 1-month to a 3-month forecast horizon in the UK. Similarly, Meißner et al. (2017) noted that
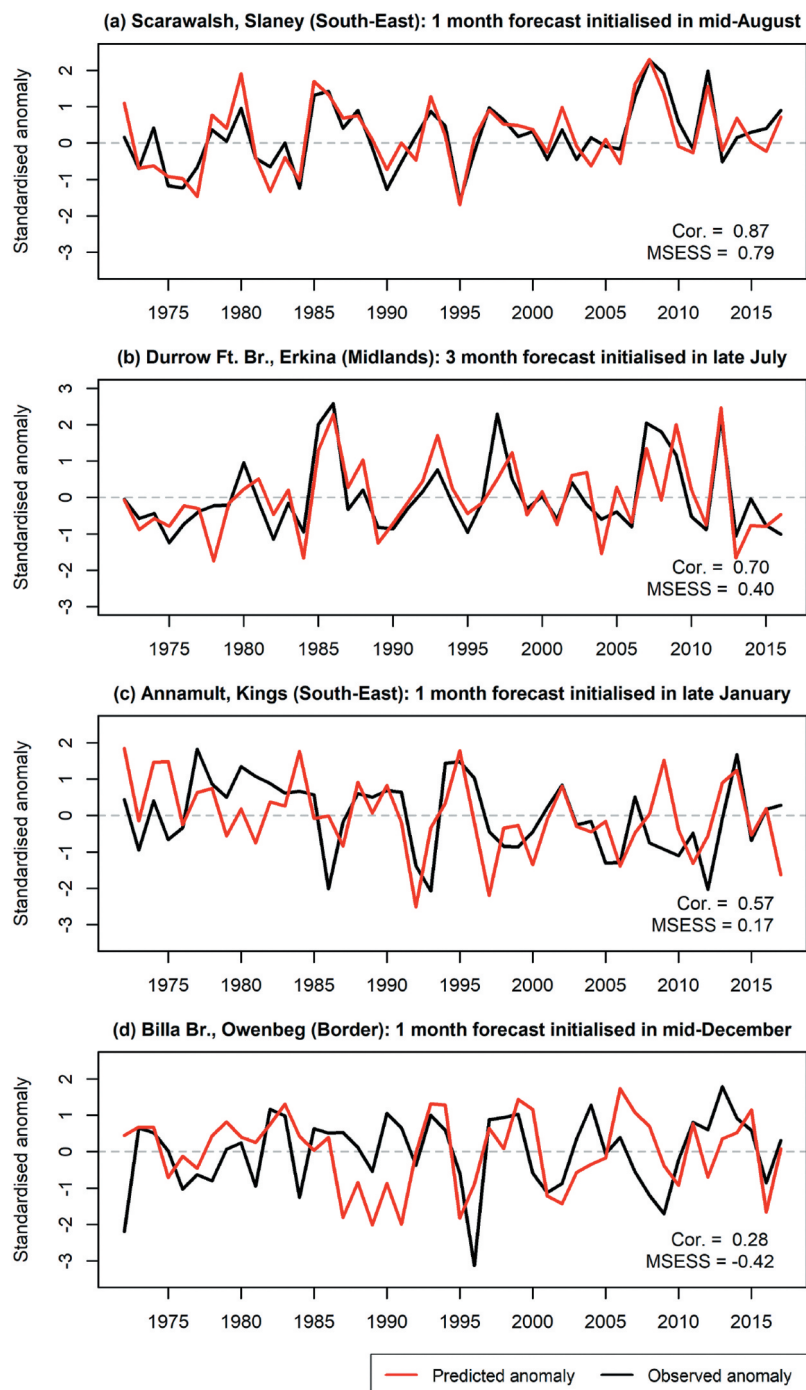
Figure 5. Hindcast time series comparing the observed (black) and predicted river flow anomaly using the persistence method (red) at sample forecast horizons for four example stations at four example initialization times. The examples show skill scores (MSESS) that are: (a) high, (b) moderate, (c) low and (d) negative.

catchment hydrological memory only acted as a sufficient source of predictability for flows up to a lead time of 1 month in major German rivers.

Shorter predictor periods mainly produce more skilful persistence forecasts because a more recent average flow anomaly is utilized. Owens *et al.* (2003) also found flow persistence skill to decline with longer predictor periods across Australia. The more comparable performance of both the 1-week and 1-month predictor periods in July can be attributed to the gradual recession of river flow at this time,

when flows are less variable across the month as they are predominantly maintained by slowly released groundwater. Meanwhile, in a number of mountainous catchments along the Atlantic margins, the 1-month predictor period even outperforms the 1-week predictor period in early January. At this time of year, river flow observed over a single week is more random, especially in such low-storage catchments. Aggregating over a longer period therefore increases the likelihood of identifying a flow anomaly that will persist over the forecast window.
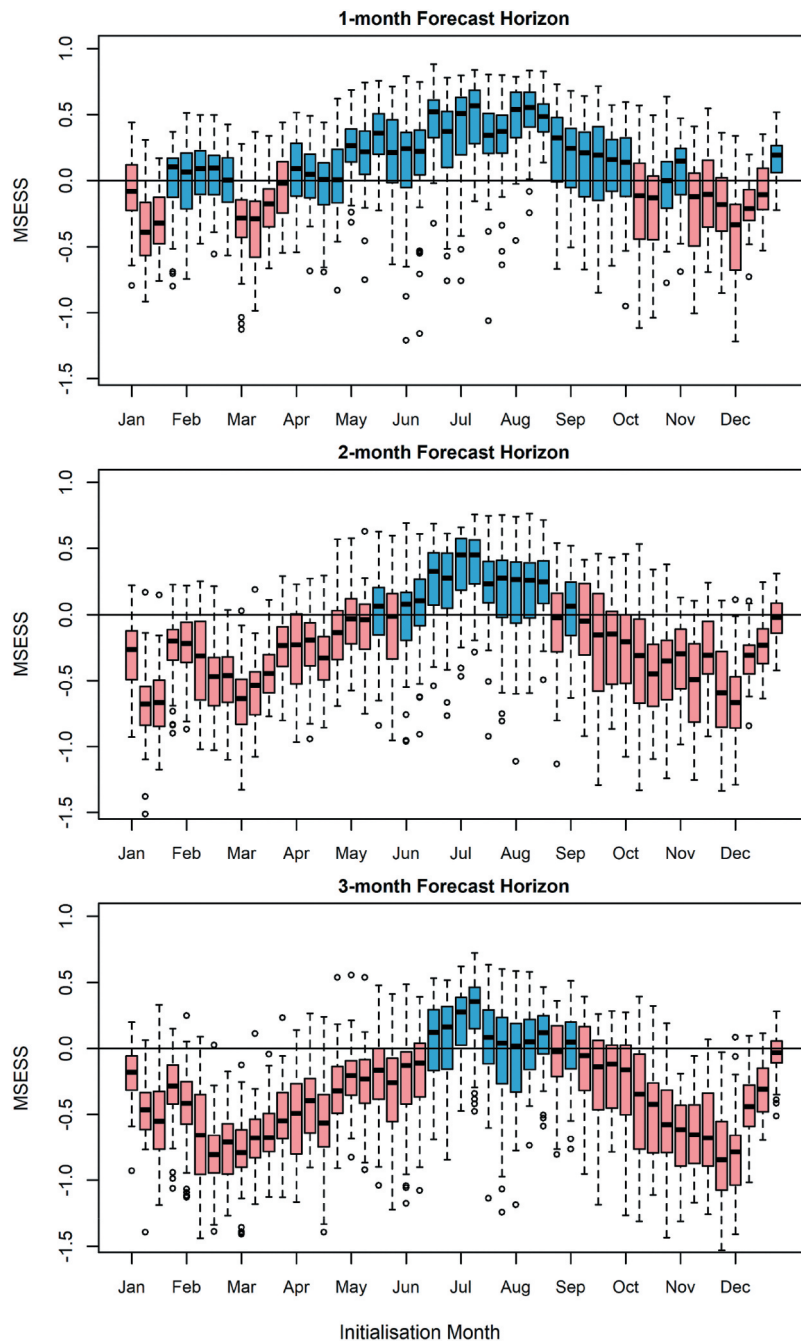
Figure 6. Network-wide skill scores (MSESS) at each of the 48 forecast initializations at the 1-, 2-, and 3-month horizons, in each case using the 1-week predictor period. Forecasts above the zero line are more skilful than the benchmark. Blue (red) boxes indicate a majority (minority) of forecasts outperform the benchmark.

River flow persistence skill is also strongly conditional on the time of year at which the forecast is initialized, with highest average skill in summer. Greater forecast skill has also been reported in drier months using other approaches that do not utilize information about future atmospheric conditions as a source of predictability; for example, in the UK (Svensson 2016, Harrigan *et al.* 2018), Denmark (Lucatero *et al.* 2018) and China (Yang *et al.* 2014). The important role of river "memory" in providing flow predictability in the south-east of the Amazon Basin was also partly attributed to the existence of a marked dry season in this area of the catchment, during which time

discharge becomes dominated by baseflow (Paiva *et al.* 2012).

Seasonality of skill can be explained by negative correlations with rainfall and the positive correlation with temperature. In the warmer summer months, when precipitation is relatively low and evapotranspiration is relatively high, small and medium rainfall amounts may evaporate before contributing to river discharge. In wetter seasons, more frequent and intense rainfall events disrupt the persistence of streamflow, particularly in low-storage catchments, essentially causing the catchment to "forget" prior conditions. A decrease in evaporation also contributes to the sharp deterioration in skill in late autumn, as excess
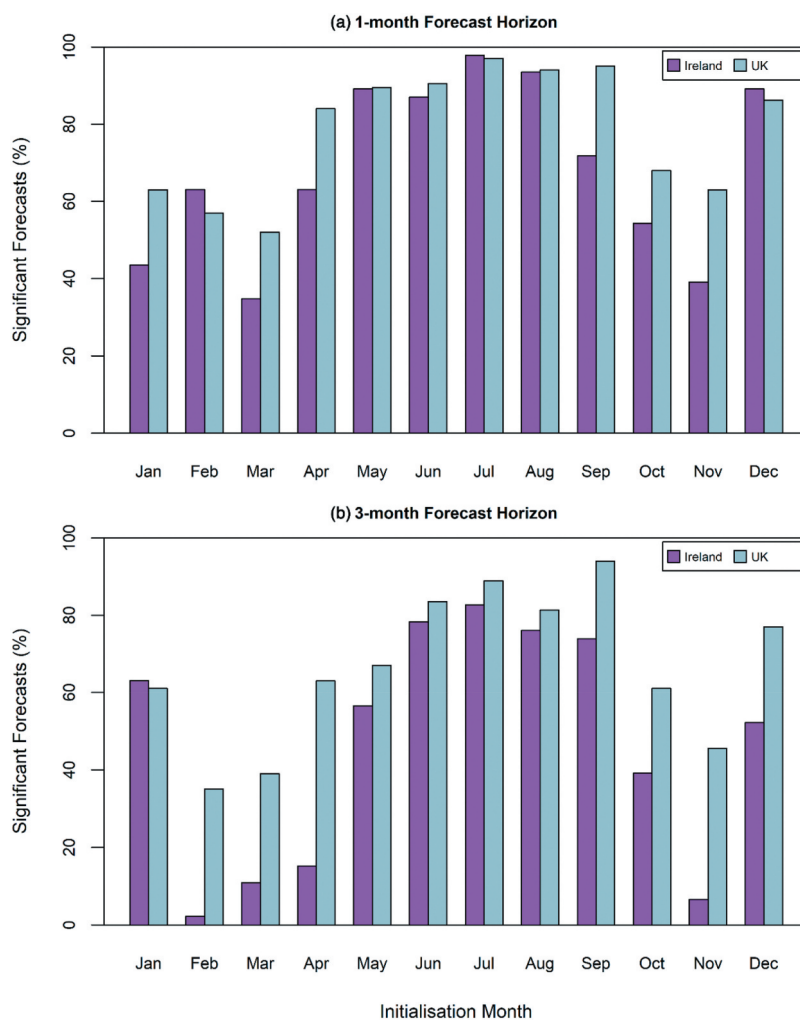
Figure 7. Percentage of "usable" forecasts across the Irish network, compared with the percentage of "usable" forecasts found by Svensson (2016) for a UK-based network. In both cases a 1-month predictor period is combined with a 1- and 3-month forecast horizon.

rainfall starts to fill up both manmade and natural reservoirs, creating unpredictable amounts of runoff in the river. In north Queensland (Australia), Owens *et al.* (2003) found that streamflow persistence skill deteriorated in late spring due to the onset of the storm season; in South Australia, high skill in spring was connected to a decrease in rainfall. Meanwhile, in the south-east US, Li *et al.* (2009) found that the influence of a catchment's initial hydrological conditions was longer lasting in the warm seasons, when the soil moisture status is low, due to high evaporation demand.

Although winter (December to February) is, on average, both the wettest and the least skilful season, March and November are the initialization months with the lowest persistence forecast skill. This becomes particularly apparent at the 3-month forecast horizon and partly stems from the fact that these months fall in periods of transition between wetter and drier times of the year. The persistence of initial river flow conditions is negatively impacted by high variability in rainfall across the forecast window (Harrigan *et al.* 2018). Svensson (2016) also found a low level of flow persistence skill for these months in the UK, which has similar precipitation climatology to Ireland. Meanwhile, a relatively weak but

significant ($p < 0.05$) month-to-month autocorrelation is found in precipitation levels from December to January ($r = 0.25$), reaching up to $r = 0.36$ in some catchments along the Atlantic margins. The increase in the average skill of flow persistence forecasts initialized in late December is consistent with the fact that persistence in precipitation is strongest in this month. Despite mid-winter being one of the wettest times of the year, flow anomalies have a higher chance of persisting when rainfall variability over the forecast window decreases.

4.2 Where is river flow persistence skilful?

The spatial distribution of persistence skill reflects variations in catchment permeability and wetness (Fig. 9). The strong positive correlation between catchment skill and BFI_{soil} ($\rho = 0.86$ at the 1-month horizon) indicates that flows tend to persist in rivers with more permeable underlying lithologies and soils. The high storage capacities of these catchments mean that flow regimes are dominated by slowly released groundwater (Sear *et al.* 1999, Chiverton *et al.* 2015). As BFI_{soil} can be estimated
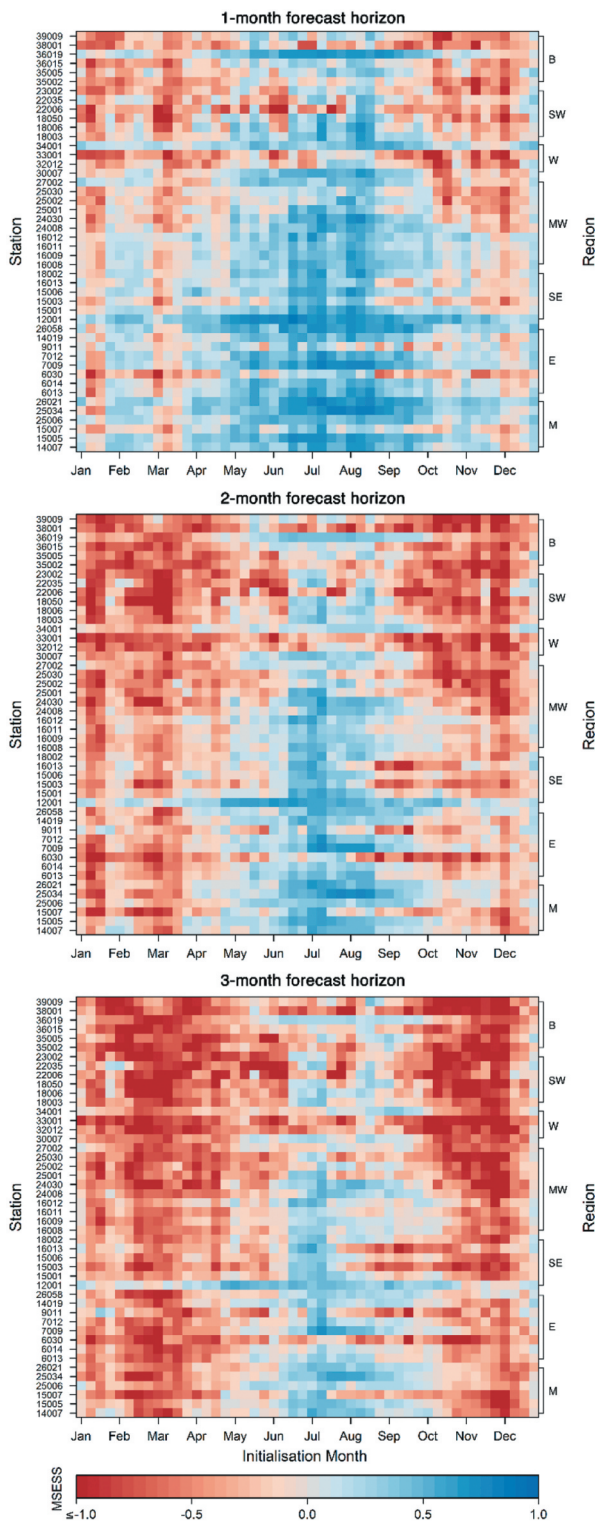
Figure 8. Variation in forecast skill based on a 1-week predictor period and 1- to 3-month forecast horizons across all 48 forecast initializations for each station, grouped by their respective regions: B (Border), SW (South-West), W (West), MW (Mid-West), SE (South-East), E (East) and M (Midlands).

for ungauged catchments, the regression models presented in this paper could be used to inform strategic investment in hydrometric networks by identifying where gauges could be most successfully used for persistence-based forecasts.

Long-term average rainfall (SAAR) and potential evapotranspiration (SAAPE) are also useful indicators of average

catchment persistence skill due to their role in predicting catchment effective storage capacity. The negative correlation between skill and SAAR ($\rho = -0.66$), for example, may reflect the greater likelihood that high-SAAR catchments have to be saturated, due to the greater frequency and/or intensity of rainfall events in them. However, it is difficult to separate the short-term influence of these rainfall events from the long-term influence that precipitation climate has on catchment hydrogeology, including the formation and hydrological behaviour of soils. Wetter catchments in Ireland, for example, tend to have poorly drained soils and less permeable subsoils, further reducing their water storage capacity. As such, catchment BFI_{soil} and SAAR are inextricably linked to each other, as they are to other significant catchment skill predictors such as the slope measures (S1085 and TAYSLO). The least skilful catchments, for example, are mainly found along the western seaboard where mountainous terrain both orographically enhances rainfall (Broderick *et al.* 2016) and contributes to low catchment storage capacity (Chiverton *et al.* 2015, Nkiaka *et al.* 2017). Nonetheless, each descriptor plays an important role in estimating persistence skill. For example, the wettest station in the network, the Laune at Laune Bridge, has higher-than-average $BFI_{soil} = 0.64$; yet it is the only station with $BFI_{soil} > 0.6$ that does not outperform the climatology benchmark for the majority of forecasts, reflecting perhaps the influence of high annual average rainfall (2010 mm).

Overall, the highest persistence skill is found in the generally drier and more permeable Midlands and eastern catchments. Stations in the region with highest median forecast skill, the Midlands, have the highest median BFI_{soil} (0.71) and one of the lowest median SAAR amounts (939 mm). This is largely a lowland karst region, and the underlying Carboniferous limestone geology is associated with regionally and locally important aquifers. When combined with a relatively high proportion of soils that are well drained, this explains the relatively strong memory of catchments. The South East, however, has a more extensive cover of well-drained soils and is predicted by the regression models to be the most skilful region when the entire 215-catchment set is considered.

Meanwhile, stations in the least skilful region, the Border, have the lowest median BFI_{soil} (0.42) and one of the highest median SAAR amounts (1289 mm). This region is poorly drained and has a large proportion of bedrock aquifers characterized as unproductive (except for local zones). The underlying lithology of the Border region is, however, quite varied, so when a wider range of catchments outside the training set is considered, median skill is predicted to rise. The poorly drained South-West region, then, has the lowest average persistence skill. Due to the low storage capacity of these catchments, river flow responds more rapidly to precipitation, and thus the influence of the initial river flow generally only persists for brief periods.

In comparison to Ireland, the higher persistence skill found by Svensson (2016) in the UK, particularly during "transition" seasons at longer forecast horizons, is likely influenced by both the greater average size of catchments across the UK sample and the larger underground storage in the major aquifers of south-east England. For example, Svensson (2016) found correlations between persistence hindcasts and observations that exceeded $r = 0.9$ for several
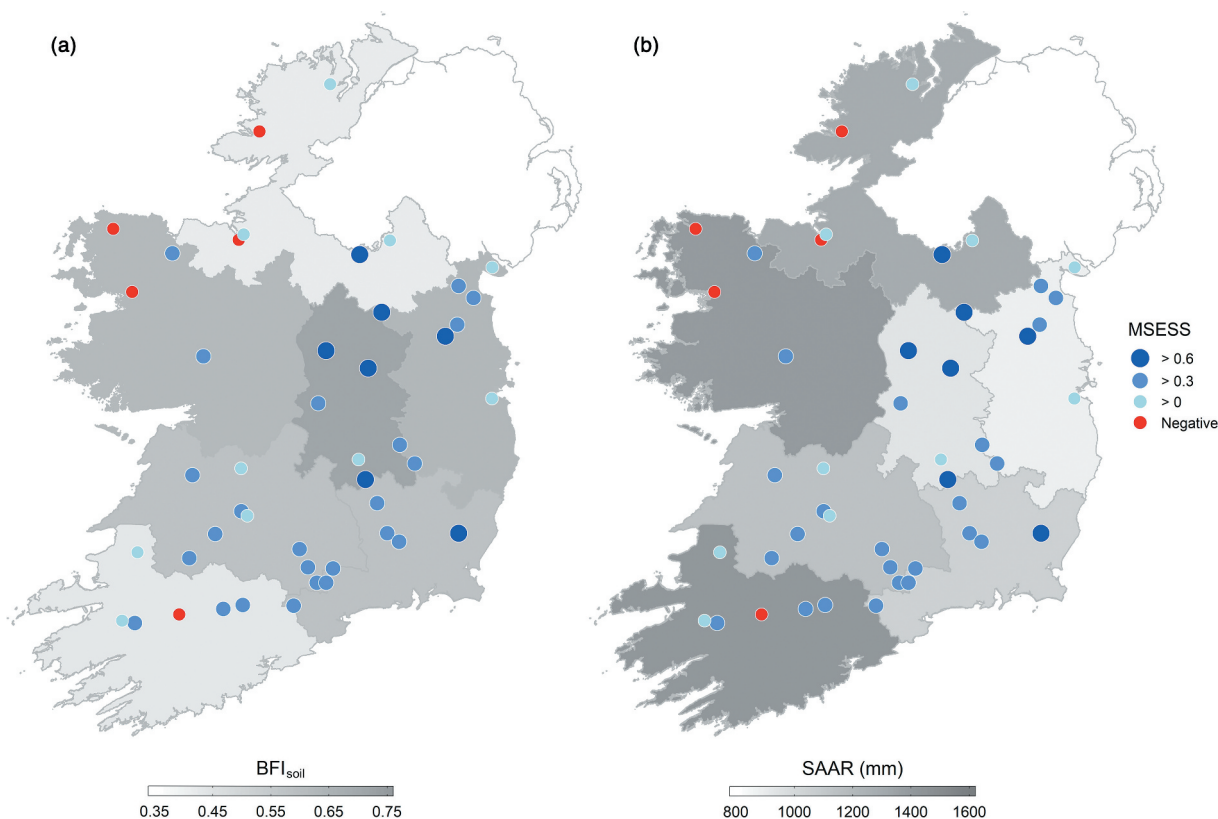
Figure 9. Median (a) baseflow index (BFIsoil) values and (b) standard-period annual average rainfall (SAAR) values for stations in each region, compared with the median summer persistence skill score (MSESS) of each individual station for the 1-week predictor period and 1-month forecast horizon.

catchments on the highly permeable English Chalk outcrops, in some places reaching $r = 0.98$. No such correlations were observed in any Irish catchments using the same predictor–forecast combinations (the highest being $r = 0.89$ in July for a Midlands catchment). Similar to Ireland, however, Svensson (2016) identified a distinct spatial pattern in skill across the UK. The greatest skill was found in the more permeable south-eastern catchments, whereas lower skill was found in more responsive north-western catchments characterized by steep gradients and impermeable bedrock.

4.3 Usefulness of the persistence forecasts

The river flow anomaly persistence approach has the potential to be used as an easy-to-implement benchmark in the evaluation of more complex forecasting techniques in Ireland. Using the standardized flow anomaly observed over the last week of each month to create a 1-month forecast, the persistence method outperforms the river flow climatology benchmark in 63% of the catchment sample, on average, across all months. This includes over 70% of catchments from May to September and in December, and 50–69% of catchments in the remaining months, except March and November. However, at the 3-month forecast horizon, persistence only provides a tougher-to-beat benchmark in the majority of stations during the summer initialization months. Therefore, the usefulness of persistence as a benchmark for skill beyond the 1-month forecast horizon is limited.

Nonetheless, having multiple potential benchmark methods decreases the risk that a new forecasting system will only be perceived as skilful compared to a benchmark that is easy to beat in a given hydrological context.

By pinpointing exactly when and where persistence is currently the toughest reference forecast to beat, we guide future development of such medium- to long-term hydrological forecasting systems in Ireland, including more rigorous benchmarking of skill at the catchment scale. Moreover, as the skill of persistence in any given catchment is dependent on the strength of local hydrogeological memory, this study highlights where in Ireland more sophisticated methods based on meteorological predictability might add value. For example, winter flows in the more responsive north-western and south-western catchments have strong positive correlations with the North Atlantic Oscillation (NAO) index (Murphy et al. 2013a), which is highly predictable on decadal time scales (Smith et al. 2020). The skill of NAO-conditioned ESP methods should therefore be assessed in these regions. There is also potential to incorporate river flow persistence into hybrid approaches that leverage the skill derived from different approaches to SHF.

Persistence forecasts may also have some potential in a practical setting in Ireland, particularly for the Midlands and eastern stations where there is medium to high skill. In the UK, the 1- to 3-month flow outlooks based on hydrological persistence indicate the forecast confidence level for each station based on their respective correlations between hindcasts and observations (Prudhomme et al. 2017). Based on our comparison with UK hindcast correlations (Svensson 2016),
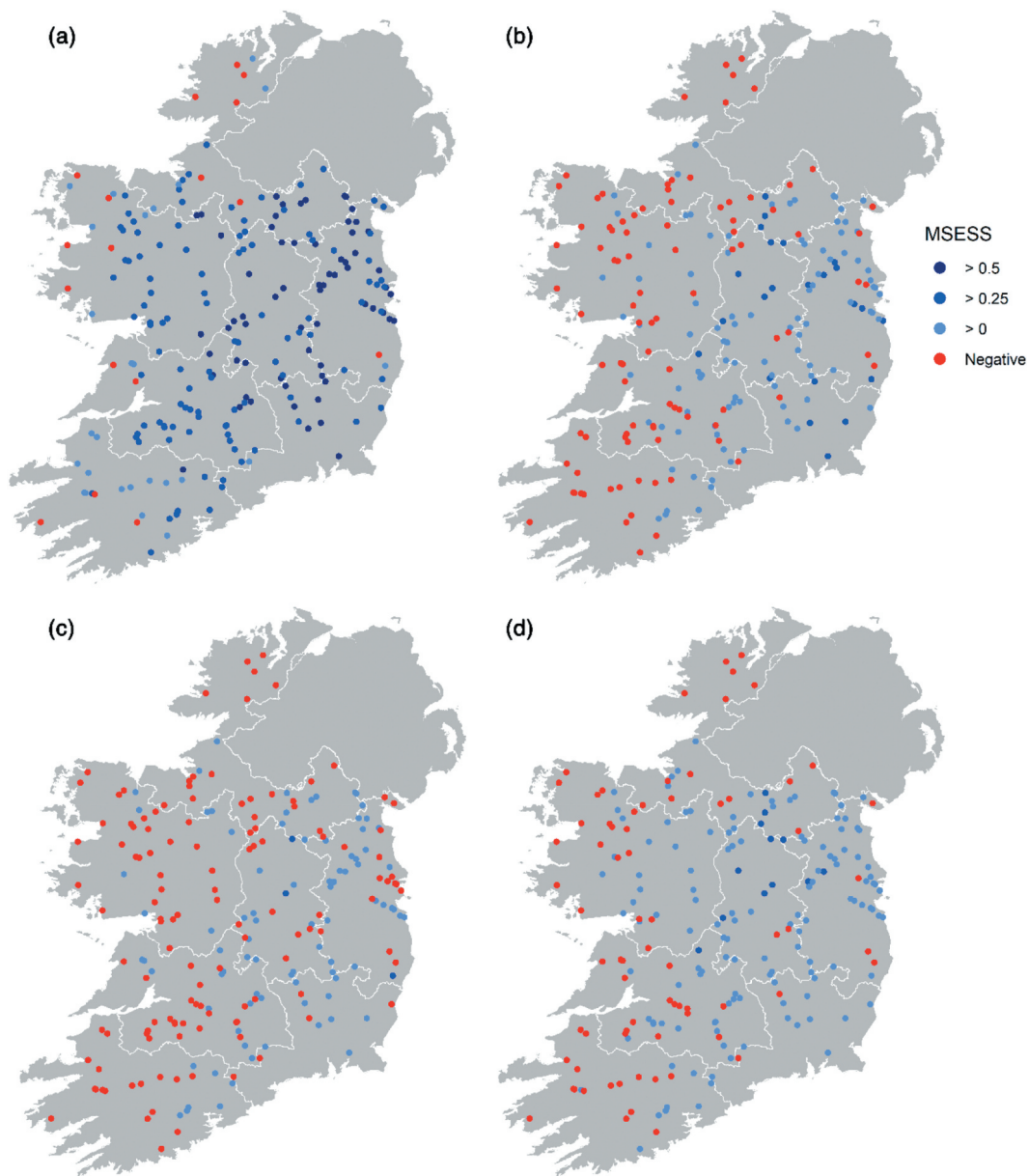
Figure 10. Median predicted seasonal persistence skill scores (MSESS) for FSU catchments in: (a) summer, (b) autumn, (c) winter and (d) spring based on a 1-week predictor period and 1-month forecast horizon.

1-month forecasts with comparable confidence levels can potentially be produced for many Irish catchments. The prospect for 3-month persistence-based outlooks is, however, more limited in the Irish context (Fig. 7).

Perhaps the greatest practical use of the persistence forecasts is in the prediction of sustained low-flow anomalies during summer months. Accurate low-flow forecasts would enable the management of adverse impacts, including reduced power production and crop yields, impaired stream navigability and restricted supplies of water to households and businesses. The superior performance of persistence in predicting low flows, in comparison to high flows, can be attributed to the slower evolution of low flows. Rainfall deficits generally take longer to evolve into anomalously low flows than heavy rainfall events take to be translated into anomalously high flows. As the influence of initial conditions is lost quickly in the case of peak flow forecasts, their accuracy is generally more dependent on skilful

meteorological forecasts (Fundel *et al.* 2013). However, no significant difference was noted in accurately forecasting either high or low flow extremes in the runoff-dominated catchments. This likely stems from their flashier regimes, with rainfall anomalies propagating into streamflow anomalies at a much faster rate (Barker *et al.* 2016). Therefore, in more responsive catchments, even accurate low-flow forecasts are more reliant on rainfall amounts being skilfully predicted; again, highlighting the need to explore the potential of incorporating meteorological forecasts and climate indices into SHF models.

Another limitation currently applicable to the operationalization of persistence forecasts in Ireland is the lag between the observation of flows and the availability of quality-controlled data, as relying on raw flow data can produce spurious forecasts. It should, however, be noted that a similar level and seasonality of forecast performance was found by applying the anomaly persistence model to water level (stage height) data (see
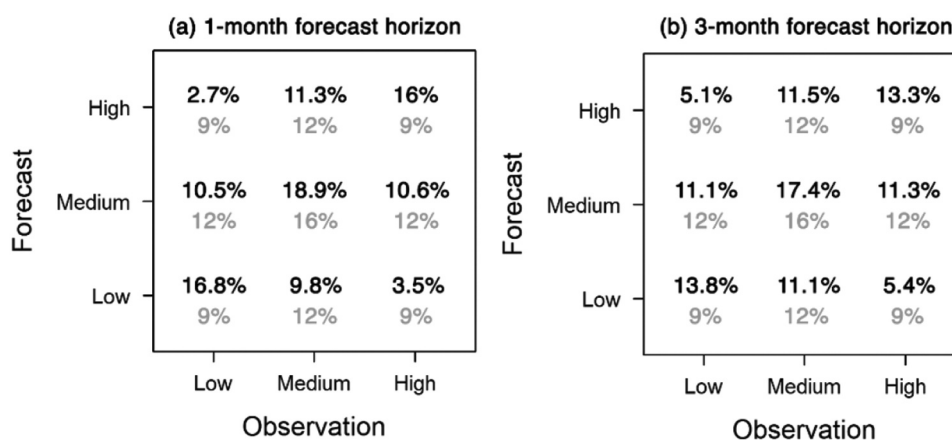
Figure 11. Categorical persistence forecast verification using: (a) 1-month forecasts and (b) 3-month forecasts. The grey percentages represent the values expected from a random distribution.

Supplementary material, Fig. S6), which in Ireland are available from the OPW in real time (albeit not quality controlled). This highlights the transferability of the method to other regions beyond Ireland where only water level data may be available.

5 Conclusions

This paper shows that river flow persistence offers a source of skill for monthly to seasonal hydrological forecasting in Irish catchments. This skill is conditional on the duration of the predictor and forecast horizon periods, the forecast initialization month and catchment characteristics. Using the most skilful predictor period (1 week), most persistence forecasts outperform the river flow climatology benchmark in February and from April to September at the 1-month horizon. However, this narrows to the summer months when using 2- and 3-month forecast horizons. This reduction in skill results from the longer time a catchment has to "forget" initial anomalous river flow conditions and/or to be impacted by "new" anomalies. Large rainfall events, for example, tend to disrupt the persistence of flows, and greater forecast skill is thus found in relatively dry, warmer months when smaller rainfall amounts often evaporate before contributing to river discharge. Predictability relies on catchment memory, and hence skill is strongly positively correlated with BFI_{soil} ($\rho = 0.86$). As such, persistence forecast skill is greatest in lowland regions that are characterized by permeable lithologies, well-drained soils and lower annual average rainfall totals. By using such physical catchment descriptors and/or similar hydrological metrics (such as the flashiness index), it may be possible to anticipate the level of streamflow persistence skill for catchments beyond Ireland.

The seasonal and spatial distribution of persistence skill shows when and where initial hydrological conditions provide useful seasonal flow forecasting skill (as in the slowly responding catchments predominantly located in the Midlands, east and south-east of the island). It also underscores the scope for development of dynamical hydrological forecasting approaches in the wetter, poorly drained catchments underlain by impermeable lithologies, found mainly in the Border, West and South-West regions. By highlighting exactly when and where flow persistence provides higher predictive skill than the reference flow climatology forecast, we show the potential value of using persistence methods in

benchmarking the skill of 1-month hydrological forecasts and, for the summer initialization months, the skill of 2- to 3-month forecasts. Establishing the skill of such simple benchmarks is an essential step towards the development of more sophisticated river flow forecasting methods at the catchment scale.

ORCID

Seán Donegan http://orcid.org/0000-0002-5957-2201

References

Amnatsan, S., Yoshikawa, S., and Kanae, S., 2018. Improved forecasting of extreme monthly reservoir inflow using an analogue-based forecasting method: a case study of the Sirikit Dam in Thailand. *Water*, 10 (11), 1614. doi:10.3390/w10111614

Anghileri, D., *et al.*, 2016. Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, 52 (6), 4209–4225. doi:10.1002/2015WR017864

Arnal, L., *et al.*, 2018. Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, 22 (4), 2057–2072. doi:10.5194/hess-22-2057-2018

Baker, D.B., *et al.*, 2004. A new flashiness index: characteristics and applications to Midwestern rivers and streams. *Journal of the American Water Resources Association*, 40 (2), 503–522. doi:10.1111/j.1752-1688.2004.tb01046.x

Barker, L.J., *et al.*, 2016. From meteorological to hydrological drought using standardised indicators. *Hydrology and Earth System Sciences*, 20 (6), 2483–2505. doi:10.5194/hess-20-2483-2016

Barnston, A.G., *et al.*, 1994. Long-lead seasonal forecasts - where do we stand? *Bulletin of the American Meteorological Society*, 75, 2097–2113. doi:10.1175/1520-0477(1994)075%3C2097:LLSFDW%3E2.0.CO;2

Broderick, C., *et al.*, 2016. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resources Research*, 52 (10), 8343–8373. doi:10.1002/2016WR018850

Broderick, C., *et al.*, 2019. Using a scenario-neutral framework to avoid potential maladaptation to future flood risk. *Water Resources Research*, 55 (2), 1079–1104. doi:10.1029/2018WR023623

Chiverton, A., *et al.*, 2015. Which catchment characteristics control the temporal dependence structure of daily river flows? *Hydrological Processes*, 29 (6), 1353–1369. doi:10.1002/hyp.10252

Day, G.N., 1985. Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, 111 (2), 642–654. doi:10.1061/(ASCE)0733-9496(1985)111:2(157)

Dixon, S.G. and Wilby, R.L., 2016. Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan. *Hydrological Sciences Journal*, 61 (1), 107–122. doi:10.1080/02626667.2015.1006227

Dixon, S.G. and Wilby, R.L., 2019. A seasonal forecasting procedure for reservoir inflows in Central Asia. *Rivers Research and Applications*, 35 (8), 1141–1154. doi:10.1002/rra.3506

Fundel, F., Jörg-Hess, S., and Zappa, M., 2013. Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. *Hydrology and Earth System Sciences*, 17, 395–407. doi:10.5194/hess-17-395-2013

Ghannam, K., *et al.*, 2016. Persistence and memory timescales in root-zone soil moisture dynamics. *Water Resources Research*, 52 (2), 1427–1445. doi:10.1002/2015WR017983

Gustard, A., Bullock, A., and Dixon, J.M., 1992. *Low flow estimation in the United Kingdom*. Wallingford: Institute of Hydrology.

Hamlet, A.F., Huppert, D., and Lettenmaier, D.P., 2002. Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management*, 128 (2), 91–101. doi:10.1061/(ASCE)0733-9496(2002)128:2(91)

Harrigan, S., *et al.*, 2018. Benchmarking ensemble streamflow prediction skill in the UK. *Hydrology and Earth System Sciences*, 22, 2023–2039. doi:10.5194/hess-22-2023-2018

Li, H., *et al.*, 2009. The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research*, 114 (D4). doi:10.1029/2008JD010969

Lucatero, D., *et al.*, 2018. Seasonal streamflow forecasts in the Ahlergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical consistency. *Hydrology and Earth System Sciences*, 22, 3601–3617. doi:10.5194/hess-22-3601-2018

Lumley, T. and Miller, A., 2009. Leaps: regression subset selection. R package version 2.9. Available from: http://CRAN.R-project.org/package=leaps [Accessed 10 September 2020].

Meißner, D., Klein, B., and Ionita, M., 2017. Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe. *Hydrology and Earth System Sciences*, 21, 6401–6423. doi:10.5194/hess-21-6401-2017

Mills, P., Nicholson, O., and Reed, D., 2014. *Physical catchment descriptors. Volume IV, Flood studies update technical research report*. Dublin: Office of Public Works. Available from: https://opw.hydronet.com/data/files/Technical%20Research%20Report%20-%20Volume%20IV%20-%20Physical%20Catchment%20Descriptors.pdf

Murphy, *et al.*, 2013b. *HydroDetect: the identification and assessment of climate change indicators for an Irish reference network of river flow stations*. Wexford: Environmental Protection Agency.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116, 2417–2424. doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2

Murphy, C., *et al.*, 2013a. Climate-driven trends in mean and high flows from a network of reference stations in Ireland. *Hydrological Sciences Journal*, 58 (4), 755–772. doi:10.1080/02626667.2013.782407

Mushtaq, S., *et al.*, 2012. The economic value of improved agrometeorological information to irrigators amid climate variability. *International Journal of Climatology*, 32 (4), 567–581. doi:10.1002/joc.2015

Nkiaka, E., Nawaz, N.R., and Lovett, J.C., 2017. Using standardized indicators to analyse dry/wet conditions and their application for monitoring drought/floods: a study in the Logone catchment, Lake Chad Basin. *Hydrological Sciences Journal*, 62 (16), 2720–2736. doi:10.1080/02626667.2017.1409427

Owens, D.T., Clarkson, N.M., and Clewett, J.F., 2003. Persistence of Australian streamflow and its application to seasonal forecasts. *In*: D.A. Post, ed. *MODSIM 2003 International Congress on Modelling and Simulation, Townsville, Australia, July 2003*. Canberra: CSIRO Land and Water.

Paiva, R.C.D., *et al.*, 2012. On the sources of hydrological prediction uncertainty in the Amazon. *Hydrology and Earth System Sciences*, 16, 3127–3137. doi:10.5194/hess-16-3127-2012

Pappenberger, F., *et al.*, 2015a. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology*, 522, 697–713. doi:10.1016/j.jhydrol.2015.01.024

Pappenberger, F., *et al.*, 2015b. The monetary benefit of early flood warnings in Europe. *Environmental Science & Policy*, 51, 278–291. doi:10.1016/j.envsci.2015.04.016

Prudhomme, C., *et al.*, 2017. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal timescales. *Hydrological Sciences Journal*, 62 (16), 2753–2768. doi:10.1080/02626667.2017.1395032

Santer, B.D., *et al.*, 2000. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research: Atmospheres*, 105 (D6), 7337–7356. doi:10.1029/1999JD901105

Sear, D.A., Armitage, P.D., and Dawson, F.H., 1999. Groundwater dominated rivers. *Hydrological Processes*, 13 (3), 255–276. doi:10.1002/(SICI)1099-1085(19990228)13:3%3C255::AID-HYP737%3E3.0.CO;2-Y

Smakhtin, V.U., 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240, 147–186. doi:10.1016/S0022-1694(00)00340-1

Smith, D.M., *et al.*, 2020. North Atlantic climate far more predictable than models imply. *Nature*, 583, 796–800. doi:10.1038/s41586-020-2525-0

Svensson, C., *et al.*, 2015. Long-range forecasts of UK winter hydrology. *Environmental Research Letters*, 10 (6), 064006. doi:10.1088/1748-9326/10/6/064006

Svensson, C., 2016. Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. *Hydrological Sciences Journal*, 61 (1), 19–35. doi:10.1080/02626667.2014.992788

Tang, Q., *et al.*, 2016. Hydrological monitoring and seasonal forecasting: progress and perspectives. *Journal of Geographical Sciences*, 26 (7), 904–920. doi:10.1007/s11442-016-1306-z

Viel, C., *et al.*, 2016. How seasonal forecast could help a decision maker: an example of climate service for water resource management. *Advances in Science & Research*, 13, 51–55. doi:10.5194/asr-13-51-2016

Watts, G., *et al.*, 2012. Testing the resilience of water supply systems to long droughts. *Journal of Hydrology*, 414-415, 255–267. doi:10.1016/j.jhydrol.2011.10.038

Webster, K.E., *et al.*, 2014. *Environmental flow assessment for Irish Rivers*. Wexford: Environmental Protection Agency.

Wedgbrow, C.S., *et al.*, 2002. Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales. *International Journal of Climatology*, 22 (2), 219–236. doi:10.1002/joc.735

Weisheimer, A. and Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, 11 (96), 20131162. doi:10.1098/rsif.2013.1162

Wilby, R.L., 2001. Seasonal forecasting of river flows in the British Isles using preceding North Atlantic pressure patterns. *Water and Environment Journal*, 15 (1), 56–63. doi:10.1111/j.1747-6593.2001.tb00305.x

Yang, L., *et al.*, 2014. Attribution of hydrologic forecast uncertainty within scalable forecast windows. *Hydrology and Earth System Sciences*, 18, 775–786. doi:10.5194/hess-18-775-2014

Yuan, X., Wood, E.F., and Ma, Z., 2015. A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *Wiley Interdisciplinary Reviews: Water*, 2 (5), 523–536. doi:10.1002/wat2.1088