Received: 18 December 2018 | Accepted: 3 March 2019

DOI: 10.1002/hyp.13445

RESEARCH ARTICLE

WILEY

A classification approach to reconstruct local daily drying dynamics at headwater streams

Aurélien Beaufort[1] | Julie Carreau[2] | Eric Sauquet[1]

[1] UR RiverLy centre de Lyon-Villeurbanne, Irstea, Villeurbanne, France

[2] Université Montpellier, CNRS, IRD, Maison des Sciences de l'Eau, HydroSciences Montpellier, Montpellier, France

Correspondence
Aurélien Beaufort, UR RiverLy, centre de Lyon-Villeurbanne, Irstea, 5 rue de la Doua CS 20244, 69625 Villeurbanne, France.
Email: aurelien.beaufort@irstea.fr

Abstract

Headwater streams (HSs) are generally naturally prone to flow intermittence. These intermittent rivers and ephemeral streams have recently seen a marked increase in interest, especially to assess the impact of drying on aquatic ecosystems. The two objectives of this work are (a) to identify the main drivers of flow intermittence dynamics in HS and (b) to reconstruct local daily drying dynamics. Discrete flow states—"flowing" versus "drying"—are modelled as functions of covariates that include information on climate, hydrology, groundwater, and basin descriptors. Three classifiers to estimate flow states using covariates are tested on four contrasted regions in France: (a) a linear classifier with regularization (LASSO for least absolute shrinkage and selection operator) and two non-linear non-parametric classifiers, (b) a one-hidden-layer feedforward artificial neural network (ANN) classifier, and (c) a random forest (RF) classifier. The three classifiers are compared with a benchmark classifier (BC) that simply estimates dominant flow state for each month based on observations (without using covariates). The performance assessment over the period 2012–2016 carried out by cross-validation shows that the three classifiers for flow state based on covariates outperformed the BC. This demonstrates the predictive power of the covariates. ANN is the classifier that globally achieves the best performance to predict the daily drying dynamics whereas both RF and LASSO tend to underestimate the proportion of drying states. The covariates are ranked in terms of relevance for each classifier. The monthly proportion of drying states provided by the discrete observation network has a major importance for the three classifiers ANN, LASSO, and RF. This may reflect the proclivity of a site to flow intermittence. ANN gives higher importance to climatic and hydrological covariates and its non-linearity allows a greater degree of freedom.

KEYWORDS

artificial neural network, drying prediction, flow state, intermittent rivers, least absolute shrinkage and selection operator, random forest

1 | INTRODUCTION

Headwater streams (HSs) are generally defined as the uppermost streams in a watershed and represent a large part of hydrographical networks (Leopold, Wolman, & Miller, 1964; Nadeau & Rains, 2007). HS can be fed by groundwater, precipitation, and run-off from small drainage areas. They contribute to the good functioning of rivers (sediment flux, inputs of particulate organic matter and nutrients), provide primordial ecosystem services (biogeochemical cycling, sources of aquatic organisms, aquatic habitat, thermal refuge; Finn, Bonada, Mùrria, & Hughes, 2011; Larned, Datry, Arscott, & Tockner, 2010; Meyer et al., 2007), and constitute reference areas to be preserved (Lowe & Likens, 2005).

Due to their upstream position in the network, their size, and their high reactivity to natural or human disturbances, HS are generally naturally prone to flow intermittence (Datry, Larned, & Tockner, 2014; Fritz et al., 2013). Intermittent rivers and ephemeral streams (IRES) are defined by periodic flow cessation and may experience partial or complete dry up at some location in time and space (Datry, Fritz, & Leigh, 2016; Larned et al., 2010; Leigh et al., 2016). They range from ephemeral streams that flow a few days after rainfall to intermittent rivers that recede to isolated pools (Datry et al., 2018). IRES have seen a marked increase in interest stimulated by the challenges of water management facing the global change context (Acuña et al., 2014; Datry et al., 2016) and by the need to improve existing knowledge on aquatic ecosystems in IRES (Larned et al., 2010; Leigh & Datry, 2017; Sarremejane et al., 2017; Stubbington, England, Wood, & Sefton, 2017).

Recent studies underline the importance to better characterize the hydrological functioning of IRES (Acuña, Hunter, & Ruhí, 2017; Boulton, 2014; Leigh & Datry, 2017). To study the impact of flow intermittence on the composition and persistence of aquatic species, freshwater biologists use quantitative metrics (e.g., to characterize the drying duration, magnitude, frequency, timing, and rate of change) determined from continuous flow series (Bunn, Thoms, Hamilton, & Capon, 2006; Datry et al., 2014; Kennard et al., 2010; Vadher, Millett, Stubbington, & Wood, 2018). However, the understanding of the spatio-temporal variability of IRES and their localization within the river network is challenging. In addition, considering only continuous gauging stations could lead to an underestimation of the proportion of IRES and of their regional extent (De Girolamo, Lo Porto, Pappagallo, & Gallart, 2015; Eng, Wolock, & Dettinger, 2016; Snelder et al., 2013).

Citizen science creates opportunities to overcome the lack of hydrological data and may contribute to densify the flow-state observation network (Buytaert et al., 2014; Datry, Pella, Leigh, Bonada, & Hugueny, 2016; Turner & Richter, 2011; van Meerveld, Vis, & Seibert, 2017). In France, new sources of observational data are available thanks to the Observatoire National des Etiages Network (ONDE; https: //onde.eaufrance.fr; Nowak & Durozoi, 2012). This unique network in Europe by its coverage, the number of monitored sites and the regularity of the observations, provides frequent discrete field observations (at least five inspections per year) of flow intermittence on more than 3,300 sites throughout France that are located mostly in headwater areas.

However, discrete observations of intermittence with irregular and at most weekly frequency cannot provide information on the persistence of dry conditions at daily temporal resolution. Thus, continuous time series of flow states are needed. Beaufort, Lamouroux, Pella, Datry, and Sauquet (2018) succeeded to relate ONDE observations to continuous hydrological and groundwater level data for predicting the daily probability of drying at the regional scale and obtained robust predictions over France. However, as predictions are aggregated over large areas, this approach does not allow to differentiate the temporal variability of "drying" state for neighbouring streams. Spatial variability of flow intermittence may be high and the understanding of local drying dynamic is crucial. Hence, the main objective of this work is to extend this previous study. Specifically, we aim at (a) identifying the main drivers of the flow intermittence dynamics in HS and (b) reconstructing the daily drying dynamics of HS at the local scale. To achieve these objectives, discrete flow states—"flowing" versus "drying"—from the ONDE observations are modelled as functions of covariates having continuous time series that include information on climate, hydrology, groundwater level, and basin descriptors.

The paper is organized in six parts. Section 2 describes the general modelling framework developed to predict flow states. Section 3 introduces the study area and the data, and Section 4 presents the performance assessment protocol. Results are presented in Section 5 and discussed in Section 6 before drawing general conclusions in Section 7.

2 | STATISTICAL FRAMEWORK FOR MODELLING DAILY DRYING DYNAMICS

Drying dynamics can be reconstructed from a classifier that relates flow states to covariates. More specifically, the classifier is calibrated in order to, for each day, estimate the probability of the drying state given a set of covariates, described in Sections 3.2 to 3.5 and summarized in Table 1, which are meant to introduce information on climate, groundwater level, hydrology, and basin descriptors. The flow state predicted at each ONDE site at a given day relies on the information provided by local and regional covariates observed at various dates by other ONDE sites in a same region. Due to the limitation of the observation period, the predictions are restricted to the period between the 1st May and the 30th September of each year (see Section 3.2).

Four classifiers are considered: (a) a benchmark classifier (BC), which does not use any covariates but is entirely based on the historical proportions of observed flow state; (b) the so-called LASSO (least absolute shrinkage and selection operator) classifier, which is a linear classifier; (c) an artificial neural network (ANN) classifier, which is potentially non-linear but encompasses a linear classifier as a special case; (d) a random forest (RF) classifier, which can also be non-linear but with a different strategy than ANN. Each classifier

TABLE 1 List of covariates

Type	Covariate	Description	Frequency	Spatial aggregation
Climate	R0	Rainfall accumulation over the day of observation (dayObs)	Day	Aggregating values for catchment site
	R1	Rainfall accumulation over the day before dayObs	Day	Aggregating values for catchment site
	R10	Rainfall accumulation over the 10 days before dayObs	Day	Aggregating values for catchment site
	R20	Rainfall accumulation over the 20 days before dayObs	Day	Aggregating values for catchment site
	R30	Rainfall accumulation over the 30 days before dayObs	Day	Aggregating values for catchment site
	T0	Air temperature over the day of observation (dayObs)	Day	Aggregating values for catchment site
	T1	Air temperature over the day before dayObs	Day	Aggregating values for catchment site
	T10	Air temperature average over the 10 days before dayObs	Day	Aggregating values for catchment site
	T20	Air temperature average over the 20 days before dayObs	Day	Aggregating values for catchment site
	T30	Air temperature average over the 30 days before dayObs	Day	Aggregating values for catchment site
	PET0	Evapotranspiration over the day of observation (dayObs)	Day	Aggregating values for catchment site
	PET1	Evapotranspiration over the day before dayObs	Day	Aggregating values for catchment site
	PET10	Evapotranspiration accumulation over the 10 days before dayObs	Day	Aggregating values for catchment site
	PET20	Evapotranspiration accumulation over the 20 days before dayObs	Day	Aggregating values for catchment site
	PET30	Evapotranspiration accumulation over the 30 days before dayObs	Day	Aggregating values for catchment site
	AI	Aridity index	Annual	Aggregating values for catchment site
	WR	Winter rainfall accumulation (December to March)	Annual	Aggregating values for catchment site
Hydrology	FQ0	Mean non-exceedance frequency of discharge at dayObs	Day	Based on HER-HR combination
	FQ5	Mean non-exceedance frequency of discharge average over the 5 days before dayObs	Day	Based on HER-HR combination
	FQ10	Mean non-exceedance frequency of discharge average over the 10 days before dayObs	Day	Based on HER-HR combination
Groundwater level	FGw0	Mean non-exceedance frequency of groundwater level at dayObs	Day	Based on HER
	FGw5	Mean non-exceedance frequency of groundwater level average over the 5 days before dayObs	Day	Based on HER
	FGw10	Mean non-exceedance frequency of groundwater level average over the 10 days before dayObs	Day	Based on HER
ONDE sites characteristics	Alti	Mean altitude of the catchment (m)	Constant	Based on catchment
	Area	Drainage area (km^2).	Constant	Based on catchment
	Slope	Mean hill slope of the catchment ($m.km^{-1}$).	Constant	Based on catchment
	MPD	Monthly proportion of days with dry states observed between 2012 and 2016 (%)	Month	Based on ONDE data

Note. B-Catchment: Based on catchment; B-HER: Based on HER; B-HER-HR: Based on HER-HR; B-ONDE: Based on ONDE data; HER: hydro-ecoregion; ONDE: Observatoire National des Etiages.

relies on a function f to estimate the flow state at day "d" at each location (o_x),

$$FlowState(d, o_x) = f(g_1(d, o_x), ..., g_n(d, o_x)), \quad (1)$$

where g_1 to g_n are the covariates depending on time including hydro-climatic covariates and "x" represents the location of the site.

Calibrating classifiers independently at each site is not possible because there are too few ONDE observations (32 observations per sites on average between 2012 and 2016). The suggested approach as well as the related assumptions for calibrating f is commonly adopted in regionalization in hydrology (e.g., the index flood method for flood frequency analysis; Dalrymple, 1960). By transferring information from different sites located in the same hydrologically homogeneous region, the sample size is increased and more robust estimates of the parameters are obtained. This implies that all the ONDE sites share the same classifier with the same set of parameters and that the drivers of flow intermittence are the same. Nevertheless, drying dynamics differ from site to site due to local factors. Thus, a calibration at the regional scale is considered to derive the model f_R valid for all ONDE sites o_x located in the same region R,

$$FlowState(d, o_x) = f_R(g_1(d, o_x), ..., g_n(d, o_x), e_1(o_x), ..., e_m(o_x)), \quad (2)$$

where e_1 to e_m are covariates that characterize the location of the site o_x.

A performance analysis, focusing on four contrasted regions in France, is carried out over the 6-year period 2012–2017 to assess their ability to simulate the daily drying dynamics at ONDE sites and to compare the accuracy of classifiers. In a second step, the influence of covariates in each classifier was examined and main environmental drivers of flow intermittence are identified.

2.1 | Benchmark classifier

BC is a simple classifier without any covariates that predicts, at a given site, the flow state that is the most frequently observed historically for the month considered. When there is a tie, drying is predicted. Cross-validation is implemented as follows. To predict the daily flow state of an ONDE site "*o*" at the month "*M*" of the year "*Y*," we select all the flow states observed at this site *o* at months *M* during the reference period of observation excluding the year *Y* (2012–2016\Y).

2.2 | LASSO classifier

The LASSO (Tibshirani, 1996) classifier estimates the drying state probability as a linear function of the covariates transformed with a sigmoid function to constrain the range to [0,1]. LASSO includes a regularization mechanism that may lead to a sparse model in which the coefficients of less relevant covariates are driven to zero (Bishop, 2006). The amount of regularization is determined through cross-validation (*R* package "elasticnet"; Zou & Hastie, 2018). To classify into either flowing or drying states, an optimal threshold is set after a second cross-validation procedure using the amount of regularization determined previously and leading to the best F-score (see Section 4.4; Equation 9). The relevance of each covariate is inspected directly through the magnitude of the associated coefficient estimated by the LASSO classifier. The LASSO method was recently considered in a hydrological application with other linear and non-linear regression techniques to predict synthetic design hydrographs for ungauged catchments (Brunner et al., 2018).

2.3 | ANN classifier

ANNs—feed-forward neural networks with one hidden layer—estimate the drying state probability as a potentially non-linear function of covariates. This is a non-parametric approach that combines the contribution of the neurons in the hidden layer to build an approximation. The number of neurons is related to the number of coefficients and hence the complexity of the classifier. As for LASSO, a sigmoid function is applied to constrain the range of the ANN output to [0,1] (see the implementation in the *R* package "nnet"; Venables & Ripley, 2002). We include a direct connection between inputs and outputs so that the case with zero hidden units corresponds to a linear relationship. Weight decay regularization, also known as ridge regression, is considered to control overfitting by decreasing less relevant coefficients. Both the number of hidden units and the amount of weight decay are selected with a first cross-validation procedure (Bishop, 2006). As for the LASSO classifier, an optimal threshold is set with a second cross-validation procedure using the number of hidden units and the amount of weight decay determined previously and leading to the best F-score. The LASSO classifier can be thought of as a particular case of the ANN with no hidden units although the regularization mechanisms are different.

To quantify the relevance of the different covariates, the connection weight approach (Olden & Jackson, 2002; Olden, Joy, & Death, 2004) is employed,

$$W_V = \sum_{h=1}^{nhu} A_{V,h} B_h, \quad (3)$$

where W_V (−) is the relevance of covariate V, $A_{V,h}$ (−) are the ANN coefficients connecting hidden unit h to covariate V, B_h (−) are the ANN coefficients connecting hidden unit *h* to the output, and nhu is the number of hidden units. ANNs have been widely used as black box tools for modelling the rainfall-runoff transform (see (ASCE, 2000a, 2000b) for a review and (Abdollahi, Raeisi, Khalilianpour, Ahmadi, & Kisi, 2017) for a recent application to an intermittent river).

2.4 | RF classifier

An RF combines decision trees obtained by resampling the calibration set (Breiman, 2001). Each tree is a structure made of binary nodes associated to binary rules of the type $V \leq s$ versus $V > s$ where V is one of the covariates and s is a bound. When reaching a terminal node, a majority vote is taken amongst the observations belonging to the node. A single decision tree tends to yield nonrobust estimation (very dependent on the selected calibration set) and the process of combining the trees in a forest circumvents this issue. We use the implementation in the *R* package "randomForest" (Liaw & Wiener, 2002). The covariates relevance is given directly by the randomForest package, which determines how much the mean square errors in prediction increases when that covariate is randomly permuted within the tree. RF models have been recently used to predict the spatial distribution of intermittent and perennial rivers at the basin scale (González-Ferreras & Barquín, 2017) and at the national scale (Snelder et al., 2013).

3 | STUDY AREA AND DATA AVAILABLE

3.1 | Study area

France is located in a temperate zone characterized by a variety of climates due to the influences of the Atlantic Ocean, the Mediterranean Sea, and mountain areas. The study area is composed of four hydro-ecoregions (HERs) located in continental France (Figure 1). The HERs correspond to a typology based on geology, topography, and climate and accounting for stream size. They have been developed for river

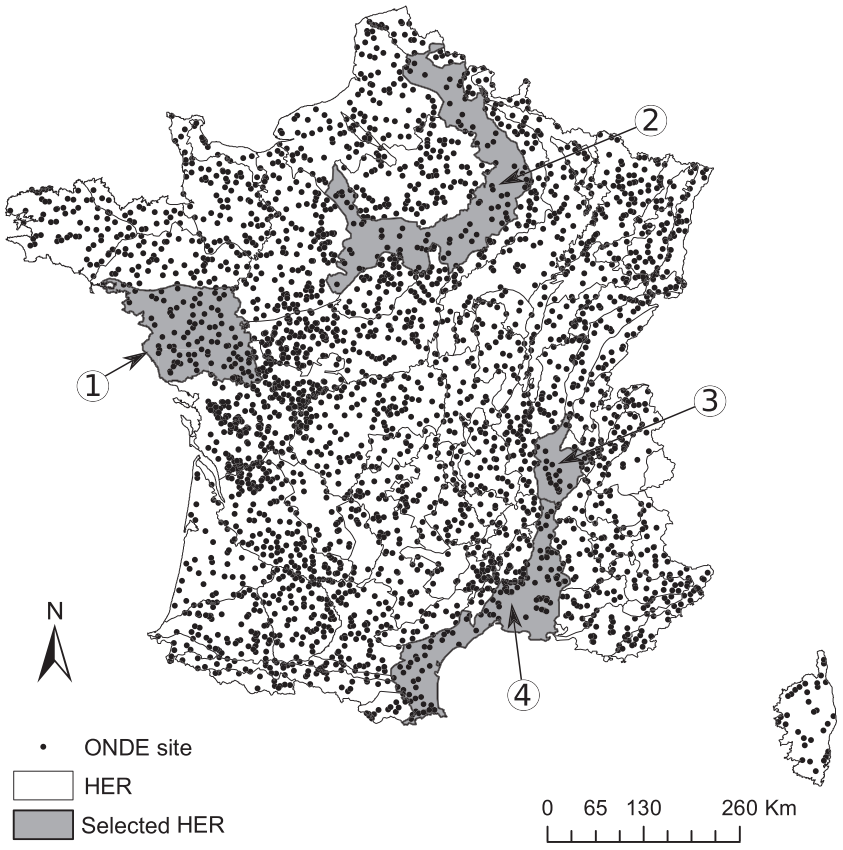

FIGURE 1 Location of the 3,300 Observatoire National des Etiages (ONDE) sites and partition of France into hydro-ecoregions (HERs)

management in accordance with the European Water Framework Directive (Wasson, Chandesris, Pella, & Blanc, 2002). The HERs were not specifically developed to discriminate river flow regimes. However, they have demonstrated their relevance as homogeneous regions in regionalization application (e.g., Cipriani, Toilliez & Sauquet, 2012; Sauquet & Catalogne, 2011) and are chosen as the regions over which f_R is calibrated (Equation 2).

Four HERs are selected amongst the partition of France into 114 HERs based on the presence of several gauging stations monitoring IRES (see Section 3.3). They have contrasted characteristics (Table 2) and are representative of most HERs in France except for mountainous regions. HER1 is distinguished by its hard, impermeable and noncarbonated primary rocks, a landform of hills, and an oceanic climate. HER2 is a lowland region with an altitude of less than 200 m.

TABLE 2 Characteristics of each selected HER to assess and compare the classifiers performance calculated between 2012 and 2016

HER	1	2	3	4
ONDE site	98	93	25	111
Intermittent ONDE site	79	42	14	79
Gauging station	26	13	4	22
Intermittent gauging station	11	2	1	9
Number of piezometers	15	119	36	130
Number of observations (ONDE)	3,906	2,223	785	3,381
Number of "drying" states (ONDE)	710 (18.2%)	213 (9.6%)	113 (14.4%)	892 (26.4%)
Mean altitude (m)	67	105	324	132
Area (km^2)	17,800	27,300	5,000	17,600
Precipitation (mm)	895	765	1,137	676
PET (mm)	715	616	711	1,051
AI (–)	1.25	1.24	1.60	0.64
Air temperature (°C)	12.2	11.0	11.5	14.5

Note. HER: hydro-ecoregion; ONDE: Observatoire National des Etiages.

The subsoil is mainly composed of carbonated sedimentary rocks. However, the upper layers of these rocks have varying surface characteristics of permeability inducing differences in the density of the drainage network. HER3 is an entity of accentuated relief, with heterogeneous basements dominated by massive limestones and carbonate rocks, and a humid continental mountain climate. HER4, with its plains and hills, is characterized by a Mediterranean climate with an extended summer drought in comparison with other regions. The geology is very heterogeneous, varying from the alluvial plain to the granite massifs and massive limestone hills.

3.2 | ONDE dataset: A discrete national flow-state observations network

The ONDE network was set up in 2012 by the French Biodiversity Agency (AFB, formerly ONEMA). Its aim is to constitute a perennial network recording summer low-flow levels that can be used to anticipate and manage water crisis during severe drought events (Nowak & Durozoi, 2012).

The ONDE network remains stable over time and distributed throughout France with 3,300 sites regularly inspected (Figure 1). ONDE sites are located on HS with a Strahler order strictly less than five and balanced across HER to take the representativeness of the hydrological contexts into account (Nowak & Durozoi, 2012). There are two types of monitoring: an ongoing monitoring and an additional monitoring. The ongoing monitoring provides a reliable baseline of knowledge over time with at least one observation per month for each ONDE site (around the 25th) between May and September. The additional monitoring has a frequency determined by local stakeholders (maximum weekly observations) in case of severe low-flow events for both drought preparedness and drought mitigation. The average number of ongoing and additional inspections at each ONDE site between 2012 and 2017 reaches 5 and 1.5 per year, respectively, over all ONDE sites.

One of the flow states is assigned at each observation amongst "visible flow," "no visible flow," and "dried out." In this work, we pool flow states into two classes: flowing when there is visible flow across the channel (visible flow) and drying when the channel is entirely devoid of surface water (dried out) or when there is still water in the river bed but without visible flow (disconnected pools, lentic systems; no visible flow). Beaufort et al. (2018) showed the complementary qualities of the ONDE network, more homogeneously distributed in France and the already existing French river flow monitoring network HYDRO (http://www.hydro.eaufrance.fr).

One covariate, derived from ONDE observations, is calculated to reconstruct the daily drying dynamics at each ONDE site: the monthly proportion of observations with drying states (%) observed between 2012 and 2016 (MPD). The ONDE sites were projected on the river network RHT (Theoretical Hydrographic Network; Pella, Lejot, Lamouroux, & Snelder, 2012) and the drainage area (Area), the altitude (Alti), and the stream slope (Slope) were estimated and used as covariates to characterize ONDE sites.

3.3 | Explanatory discharge dataset

ONDE sites are ungauged but gauging stations located in the same HER could potentially inform about the hydrological state of rivers at a regional scale. Time series of daily discharge were extracted from the French River discharge monitoring network (HYDRO, http://www.hydro.eaufrance.fr/). This network is composed of 1,600 gauging stations distributed across France for which flow data are available between 2011 and 2017. According to the hydrometric services in charge of the selected gauging stations, high quality of measurements is ensured and observed discharges were either not altered or only slightly altered by human actions.

Hydrological variables derived from gauging stations located in the same HER as the ONDE site, are also considered as covariates. To ensure homogeneity, only gauging stations with the same river flow regime as the ONDE site were kept. The river flow regime (HR) is taken from the classification for France suggested by Sauquet, Gottschalk, and Krasovskaia (2008) and each HER is partitioned into subregions defined by the river flow regimes. For the four selected HERs, a total number of three, six, one, and eight HER-HR combinations are outlined in HER1, HER2, HER3, and HER4, respectively. The HER1 and HER2 are mainly composed of pluvial river flow regimes and the HER3 and HER4 are composed of a mix of pluvial river flow regimes, transition river flow regimes, and Mediterranean river flow regimes. Due to scale effect, discharge time series from different gauging stations cannot be combined directly and a post-processing is required. The first step consists in selecting all gauging stations located in a given HER-HR combination. The total number of gauging stations per HER-HR combination is at least five. In a second step, the flow duration curve is determined for each selected gauging station and three covariates are computed: the average nonexceedance frequency of the observed discharge at gauging stations (a) at the dayObs (FQ0), (b) over the 5 days before dayObs (FQ5), and (c) over the 10 days before dayObs (FQ10). The preprocessing methodology is illustrated in the appendix (Figure A1).

Each stream where a HYDRO gauging station is located is defined as IRES or perennial streams. The identification of intermittent gauging stations was carried out with the aim of (a) selecting the four HERs where model performance is assessed (see section 3.1) and (b) validating against continuous time series of flow states (see section 4.3). Several definitions of IRES can be found in the literature (Eng et al., 2016; Huxter & van Meerveld, 2012; Reynolds, Shafroth, & LeRoy Poff, 2015). In this study, we consider stations as intermittent when five consecutive days with discharge less than 1 l/s is observed during the observation period.

3.4 | Explanatory groundwater level dataset

Daily groundwater levels are provided by the ADES database (http://www.ades.eaufrance.fr) at sites involved in groundwater/surface water exchanges (Brugeron, Allier, & Klinka, 2012). This dataset is composed by 750 piezometers with daily groundwater level data

available from 2011 to 2017 with less than 5% of missing data (continuous or not). The level of alteration of groundwater levels by water withdrawal is unknown because no information is available at this scale.

Groundwater level data are used as covariates. A post-processing similar to the one applied to daily discharge is applied to groundwater levels (Figure A1) except for the first step, which consists in selecting all piezometers located in a same HER instead of HER-HR combination. In a second step, in each HER, three covariates are computed: the average nonexceedance frequency of the observed groundwater level (a) at the dayObs (FGw0), (b) over the 5 days before dayObs (FGw5), and (c) over the 10 days before dayObs (FGw10).

3.5 | Explanatory meteorological dataset

Daily meteorological covariates are taken from the SAFRAN dataset (Quintana-Seguí et al., 2008; Vidal, Martin, Franchistéguy, Baillon, & Soubeyroux, 2010) that has an 8-km resolution grid and is available from August 1, 1958 to July 31, 2017. The SAFRAN dataset provides precipitation and air temperature. The daily reference evapotranspiration is computed with the Penman–Monteith formulation (Allen, Pereira, Raes, & Smith, 1998). Daily catchment-scale data are computed for each gauging station and ONDE site using weighted mean (for temperature) or sum (for precipitation and evapotranspiration) of each contributive cell of the 8-km grid to the catchment surface (Sauquet & Catalogne, 2011).

The daily catchment-scale air temperature (T), rainfall (R), and potential evapotranspiration (PET) data are used as covariates in order to reconstruct the daily drying dynamics at ONDE sites. For each covariate, we consider daily values observed on the current day (R0, PET0, T0) and on the previous day (R1, PET1, T1) along with values accumulated or averaged over 10, 20, or 30 days preceding the current day (Rp, PETp, Tp) with p = 10, 20, or 30.

We derive two additional annual climate descriptors, namely the aridity index (AI) and the winter rainfall accumulation. AI is given by the ratio between the mean annual precipitation and the mean annual PET. The catchment is considered as "hyper-arid" if AI < 0.03, "arid" if $0.03 \leq AI \leq 0.2$, "semi-arid" if $0.2 \leq AI \leq 0.5$, "dry subhumid" if $0.5 \leq AI \leq 0.65$, and "humid" if AI > 0.65. Winter rainfall is determined each year as the rainfall accumulation between December of the precedent year and March of the current year (4 months).

4 | PERFORMANCE ASSESSMENT AND COMPARISON

The performance of the classifiers is evaluated on the four selected HERs (Figure 1). The calibration and validation methods are described in the next sections.

4.1 | Cross-validation over 2012–2016

A cross-validation procedure is carried out for each classifier in each HER. The calibration set is constituted by selecting randomly 80% of the observations. The test set consists of the remaining 20%. Once the classifiers are trained on the calibration set, the evaluation criteria are calculated (see Section 4.4) based on the prediction on the test set. This step is repeated 20 times in order to evaluate the uncertainty associated to the selection of the calibration set.

4.2 | Extrapolation ability over 2017

In order to assess the extrapolation ability of the classifiers, they were calibrated over the period 2012–2016 and their performance is evaluated over the first 3 months of 2017 (due to SAFRAN data availability), namely May, June, and July. In each HER, all the covariates associated with the ONDE observations between 2012 and 2016 are used to calibrate classifiers. Then, the covariates of the year 2017 are used as input to predict the daily flow states at the ONDE sites. For BC, benchmark values are computed excluding the year 2017.

4.3 | Spatio-temporal extrapolation ability

In order to evaluate their spatio-temporal extrapolation ability over the period 2012–2016, the classifiers calibrated on the ONDE sites are applied to 65 gauging stations—38 and 27 located on perennial and intermittent streams, respectively. Based on the continuous observations from the gauging stations, an independent test set of daily flow states is constituted and serves to evaluate the classifiers at the daily time step between May and September. The classifiers remain calibrated against observations available at the ONDE sites located in the same HER and used the covariates (Table 1) computed for 65 gauged basins to predict the flow state. Gauging stations covariates used for prediction are calculated as for the ONDE sites except for the monthly proportion of observations with drying states (MPD). MPD is calculated by taking into account the flow observed at the gauging station on the 25th of each month corresponding approximately to the ONDE observation dates. A day in the discharge time series is classified in a drying state if the observed daily flow is less than 1 $l.s^{-1}$. This way, potential false-positive detection of zero flows associated with measurement resolution constraints and uncertainty in discharge observations are accounted for. For BC, to predict the flow state of a gauging station i during the month M of the year Y, we select the most frequent flow states observed on the 25th at this station i on months M during the reference period excluding the year Y (2012–2016\Y). The flow state predicted is assigned to the 15 days preceding and following the 25th of month M and year Y considered in order to obtain continuous daily flow states comparable to predictions of the other classifiers.

4.4 | Evaluation criteria

Several validation criteria are calculated to compare the performance of the classifiers. First, criteria based on a 2 × 2 contingency table, see Table 3, are used to evaluate the ability of classifiers to accurately predict flow states for a stream at a given day.

Derived from the contingency table, five criteria (see Equations (5) to (9)), are calculated to assess classifiers performance: the probability of detection (POD; best value is 100%), the false alarm ratio (FAR; best value is 0%), the precision (best value is 1), the recall (best value is 1), and the F-score (best value is 1).

$$\mathrm{POD} = \frac{a}{a+c} \times 100. \tag{5}$$

$$\mathrm{FAR} = \frac{b}{a+b} \times 100. \tag{6}$$

$$\mathit{Precision} = \frac{a}{a+b}. \tag{7}$$

$$\mathit{Recall} = \frac{a}{a+c}. \tag{8}$$

$$\mathrm{F\text{–}score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{9}$$

In addition, the proportions of observed and predicted days with a drying state—named P_{obs} and P_{pred}, respectively—at gauging station i are compared to measure the spatio-temporal ability (Section 4.3),

$$P(i) = \frac{(N_{dr})_i}{(N_{fl} + N_{dr})_i} \times 100, \tag{10}$$

with N_{dr}: the number of observed or predicted daily drying states between the beginning of May to the end of September; N_{fl}: the number of observed or predicted daily flowing states between the beginning of May to the end of September. N_{dr} and N_{fl} can be calculated annually or over the period between 2012 and 2016.

TABLE 3 Contingency table

		Observations	
		Drying	Flowing
Predictions	Drying	a	b
	Flowing	c	d

Note. a (b) represents the number of correctly (incorrectly) predicted "drying" states; c (d) represents the number of incorrectly (correctly) predicted "flowing" states.

The bias and the root mean square error are calculated for a performance assessment at the HER scale

$$\mathrm{Bias(HER)} = \frac{1}{G}\sum_{i=1}^{G}\left(P_{pred}(i) - P_{obs}(i)\right), \tag{11}$$

$$\mathrm{RMSE(HER)} = \sqrt{\frac{1}{G}\sum_{i=1}^{G}\left(P_{pred}(i) - P_{obs}(i)\right)^2}, \tag{12}$$

where i is a gauging station located inside a given selected HER and G is the total number of gauging stations located in each HER.

4.5 | Evaluation of covariate contribution

In this evaluation, the covariates are grouped according to their type defined in Table 1 except for MPD because this covariate is directly related to local observations. The methods for estimating the contribution of covariates to the prediction are different for each classifier (see Sections 2.2 to 2.4) and the magnitude of the contribution of each covariate cannot be directly compared across classifiers. Cumulative or averaged percentages representing the relative contribution of the covariates in each classifier are used to rank the groups of covariates.

5 | RESULTS

5.1 | Classifier assessment over four contrasted HER

5.1.1 | Cross-validation results between May and September over the period 2012–2016

The three classifiers ANN, RF and LASSO outperforms BC in all HERs. This can be seen from the POD, FAR, and F-scores in Figure 2. The LASSO classifier, although linear, performed only slightly worse than the ANN classifier, its non-linear counterpart (see the POD and FAR scores in Figure 2). Amongst the two non-linear classifiers, RF achieved the overall best performance in the four HERs (see the three scores in Figure 2).

The performance obtained by the two non-linear classifiers is very close and the best POD is obtained by ANN in all selected HERs (Figure 2a). However, RF minimizes the FAR and obtains a better F-score than ANN on average over the four HERs.

5.1.2 | Performance of classifiers in extrapolation over 2017

The results of the classifier predictions in extrapolation in 2017 show that BC obtains the best POD, which is greater than 60% in the four HERs (Figure 3a). However, the FARs are very high and exceed 50% so BC tends to strongly overestimate drying states in 2017 especially in HER1 and HER3 (Figure 3b).

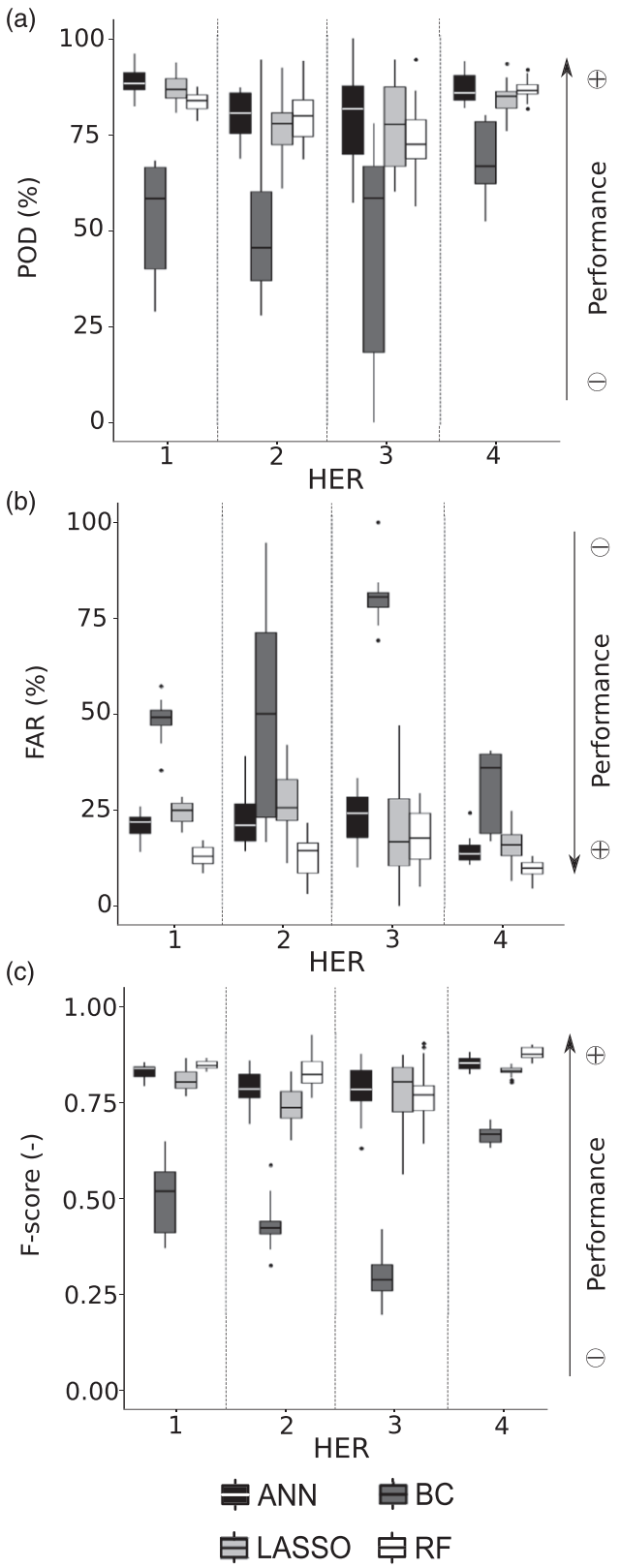

FIGURE 2 Evaluation criteria calculated with the cross-validation process over the period 2012–2016 with the four classifiers: (a) probability of detection (POD); (b) false alarm ratio (FAR); (c) F-score

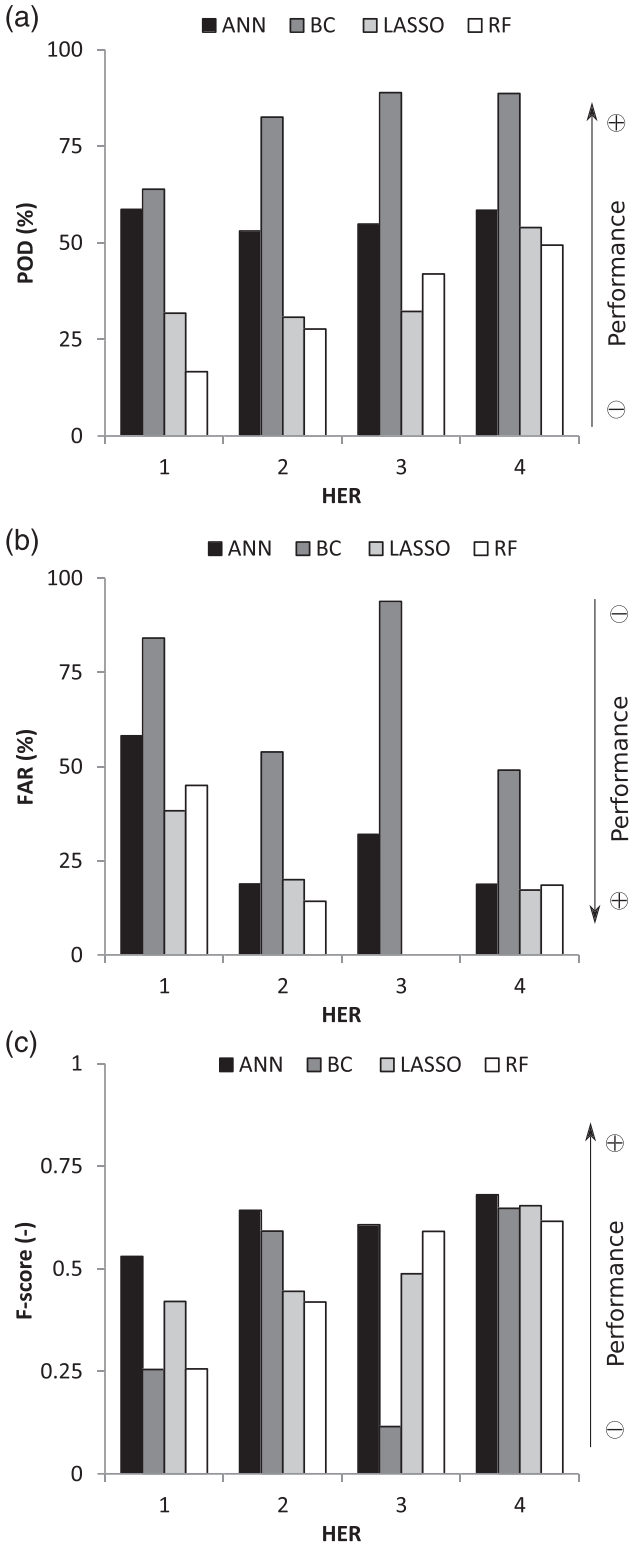

FIGURE 3 Evaluation criteria calculated in extrapolation over 2017 with the four classifiers: (a) probability of detection (POD); (b) false alarm ratio (FAR); (c) F-score

ANN obtains a POD higher than 50% and a FAR significantly lower than BC whatever the HER. The ANN classifier obtains the best F-score for each selected HER (F-score > 0.5).

The performance of the LASSO and RF are rather moderate. They obtain very low POD lower than 30% on the HER1, HER2, and HER3 but also a very low FAR reaching zero for the HER3. We deduce that these two classifiers tend to underestimate the number of drying states extrapolated in 2017. Their F-score is lower than BC in the HER2 but remains very close to the F-score of ANN in the HER4, which corresponds to the HER where the most drying state is observed.

5.1.3 | Performance of the spatio-temporal extrapolation over the period 2012–2016

For the three classifiers ANN, LASSO, and RF, P_{pred} are close to P_{obs} especially for gauging stations whose proportions of drying states are greater than 20% (Figure 4). For stations where the proportion of drying states is lower than 20%, the accuracy of the classifiers is more contrasted.

Overall, ANN achieves the best performance with an average root mean square error (RMSE) of 3.3% over the four HERs (Table 4). The RMSE is similar and close to 3% for each HER. ANN tends to slightly overestimate the proportion of observed drying states illustrated by positive biases and a FAR close to 40% on the four HERs (Table 4). RF underestimates the proportions of drying states especially in the HER1, HER3, and HER4 where the biases are below −1.5%. The predictions of LASSO are more contrasted with an overestimation of the drying state proportions in the HER1, an underestimation of the drying proportions in the HER3 and biases close to zero in the HER2 and HER4.

For the 27 gauging stations with at least one observed drying state between 2012 and 2016, the ANN classifier can detect drying states on 23 gauging stations but failed to detect drying states on four of them. These four gauging stations are located in the HER1 and showed a proportion of drying states less than 5%. In comparison, RF and LASSO failed to detect drying states for seven and nine gauging stations, respectively.

On the other hand, ANN tends to predict very short drying states, persisting less than 2 days, on perennial stations leading to a prediction of drying states of less than 1% (Figure 4a). These incorrectly predicted drying states correspond to periods of severe low flows, that is, when most of the daily flows stay below the 90th quantile of the flow duration curve. Thus, although not strictly speaking in a drying state, they are consistent with the ANN classifier predictions.

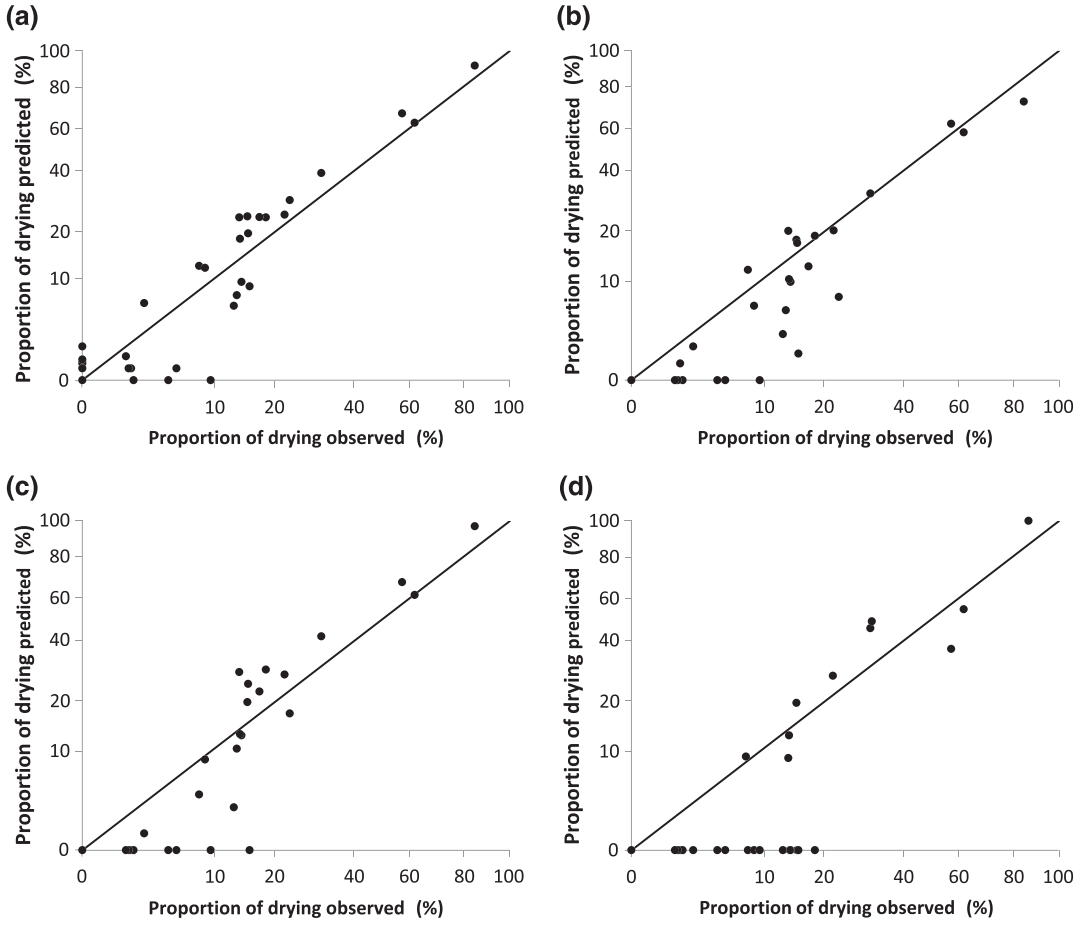

FIGURE 4 Predicted and observed proportions of "drying" states over the period 2012–2016 at gauging stations located in the four selected hydro-ecoregions. Drying states predictions are obtained respectively with (a) artificial neural network, (b) random forest, (c) least absolute shrinkage and selection operator, and (d) benchmark classifier. Each dot is a gauging station and the y-axis displays a square root scale

TABLE 4 Biases, RMSE, POD, FAR, and F-score calculated at the gauging stations located in the four HERs over the period 2012–2016. The bolded values correspond to the region HER where the best performance is achieved by model for each criterion

	CRITERIA	HER1	HER2	HER3	HER4
ANN	Bias	**0.3**	1.2	1.7	0.6
	RMSE	3.6	**3**	3.2	3.2
	POD	80	**96**	72	79
	FAR	39	41	44	**36**
	F-score	0.71	**0.73**	0.63	0.69
RF	Bias	−1.5	**0.5**	−4.3	−1.8
	RMSE	3.9	**1.4**	8.6	4.8
	POD	**73**	60	38	63
	FAR	36	**32**	48	42
	F-score	**0.62**	0.58	0.44	0.61
LASSO	Bias	1	**0.1**	−1.6	−0.2
	RMSE	5.9	**1.9**	3.1	4.7
	POD	58	**92**	18	59
	FAR	32	**31**	35	35
	F-score	0.60	**0.79**	0.28	0.62
BC	Bias	**−1.6**	−2.7	4.2	−1.8
	RMSE	**4.3**	7.2	8.4	6.3
	POD	**73**	60	38	63
	FAR	36	**32**	48	42
	F-score	**0.62**	0.58	0.44	0.61

Note. ANN: artificial neural network; BC: benchmark classifier; FAR: false alarm ratio; HER: hydro-ecoregion; LASSO: least absolute shrinkage and selection operator; POD: probability of detection; RF: random forest; RMSE: root mean square error.

The nonexceedance frequency of discharge observed at four gauging stations, representing each selected HER and having a proportion of drying states less than 20%, is compared with the flow states predicted by the four classifiers (Figure 5).

During the warmest year in 2012, ANN, RF, and LASSO give consistent results and suggest dry periods relatively close to observations for three of the four HERs (HER1, HER2, and HER3). Conversely, all the classifiers predict drying states in HER4 whereas no event is observed but these predicted drying states remain concomitant with a period of severe low flow beginning in June (most of the daily flows stay below the 80th quantile of the flow duration curve).

On the other hand, during wet years (in 2013 for HER2 and HER3 and in 2014 for HER1 and HER2; Figure 6), ANN, RF, and LASSO tend to overestimate drying states but the periods with drying states are very short.

Globally, BC failed to detect drying states in HER1 and falsely detect drying states in HER2 and HER4. Drying states incorrectly detected by BC are not always concomitant with a period with low flows and BC is unable to reproduce the hydrological dynamics of gauging stations.

RF and LASSO can perform better than ANN in some years, especially in the HER1 and HER3. However, their accuracy is less regular and in some instances, they largely underestimate the duration of drying states during some years, for example, in 2015 and 2016 for HER3 with RF and in 2014 and 2015 for HER4 with LASSO (Figure 5).

5.2 | Identification of covariate contributions to drying state predictions

The analysis of the contributions of the covariates is summarized in Table 5. Results are displayed by type of covariate. The pre-eminence of the type of covariate differs depending on the classifier.

For the ANN classifier, climatic covariates are identified as the main drivers in cumulative value for the prediction of daily drying states. The percentage of the contributions of climatic covariates reaches as much as 67.5% in the HER3. The second most relevant covariates for predicting drying states are related to river flow for HER1 and HER3 or to groundwater level for HER2 and HER4. The MPD is the third most significant covariate. However, it is worthy to note that there is a balance between the relative weight of MPD and the climatic covariates: MPD has a high value whereas climatic covariates are found less important, and conversely. Despite a large number of climatic covariates (17; Table 1), ANN uses all of them but assigns them less weight than MPD.

For the LASSO classifier, the climatic covariates and the MPD seem to be the most relevant in cumulative values for predicting drying states. In HER1, hydrology is the third most important driver and small predictive power is given to covariates describing groundwater whereas in HER4, the results bring forward groundwater instead of river hydrology. In HER2 and HER3, groundwater and hydrology are less differentiated with covariates importance above 10%. The pre-eminence of hydrology against groundwater level covariates is not linked to their respective sites number located in each HER (Table 2).

In the RF classifier, the most important covariate, in cumulated or averaged value, is MPD that obtains a percentage higher than 40% for the four HERs. In HER2, HER, and HER4, climate is the second main driver followed by hydrology and groundwater. The importance of hydrological covariates is, however, more important in terms of averaged values than climate covariates in all HERs.

The ONDE site characteristics, constant in time at a given site, have a minor importance for predicting daily drying state for the three classifiers.

6 | DISCUSSION

6.1 | Global comparison of the classifiers

This first conclusion is that the covariates bring valuable information in drying state predictions as can be deduced from the fact that the three classifiers, ANN, LASSO, and RF, perform better than BC which does not use any covariates. The cross-validation procedure between 2012 and 2016 shows similar performance between the classifiers and suggests that the non-linearity taken into account by both ANN and RF is one possible cause of improved POD and FAR scores in comparison with LASSO (Figure 2). The performance of classifiers assessed in extrapolation over the period May–July 2017 (F-score ~ 0.5) is lower than their performance obtained in cross-validation

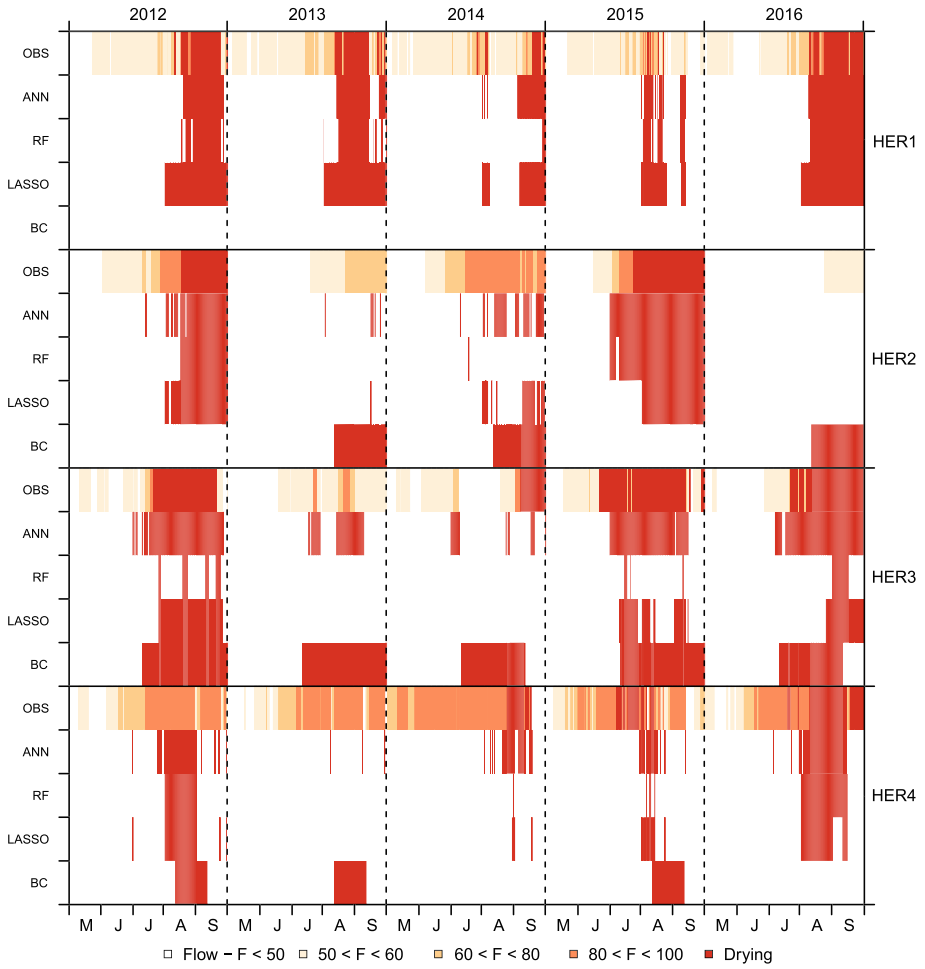

FIGURE 5 Observed and predicted flow states for four gauging stations located in the four hydro-ecoregions (HERs). Observed daily flows are categorized in four classes based on their associated quantile of the flow duration curve (F)

(F-score > 0.75). It can be partly explained by the particularity of the year 2017 (Figure 6) when an early drought beginning in May is observed over a large part of France. Between 2012 and 2016, the south-east of France is the only region that was affected by droughts and dryings in spring whereas dryings started from July in the other regions. This unusual situation was never encountered in the calibration dataset and the estimates of flow states are less accurate except in HER4 where the F-score is higher than 0.6 for the four classifiers. Nevertheless, there is a clear gain in performance with the ANN classifier in comparison with RF and LASSO. This added value is even more important when classifiers are tested against continuous time series of daily discharges. The ANN classifier seems to better predict the year-to-year variability (dry year vs. wet year) at a given station. The conclusions of this comparison are drawn only on analyses based on the 2012–2017 observation period (Figure 6). The ONDE network is still functional and we could expect that the additional flow states observations in contrasted years would contribute to improve the calibration of classifiers and this could lead to improved performances in extrapolation ability.

The main difficulty encountered by the three classifiers comes from the complexity of predicting rare events. In cross-validation over the period 2012–2016 and in extrapolation in year 2017, the best F-scores amongst the three classifiers are obtained in HER4 where the drying state frequency is the most important (Table 2). ANN is the classifier that achieves the best performance for weakly intermittent HS. Classifiers tend to underestimate drying states with an observed frequency smaller than 20% (Figure 4). It may occur that no drying state is observed due to the rarity of zero-flow events and the value of MPD is zero whatever the month of the year. The high importance of this covariate in classifiers leads to an underestimation of predicted drying states or to consider the HS as perennial. This underlines the importance of taking into account a calibration dataset composed of years with contrasted climate thus allowing a better representation of the extreme events. Future observations of the ONDE network will make it possible to sample more contrasted situations and to better identify sites impacted by flow intermittence leading to an updated ranking of the three classifiers.

6.2 | Drivers of flow intermittence

MPD is considered as very important for the three classifiers. This covariate reflects the level of flow intermittence of each ONDE site

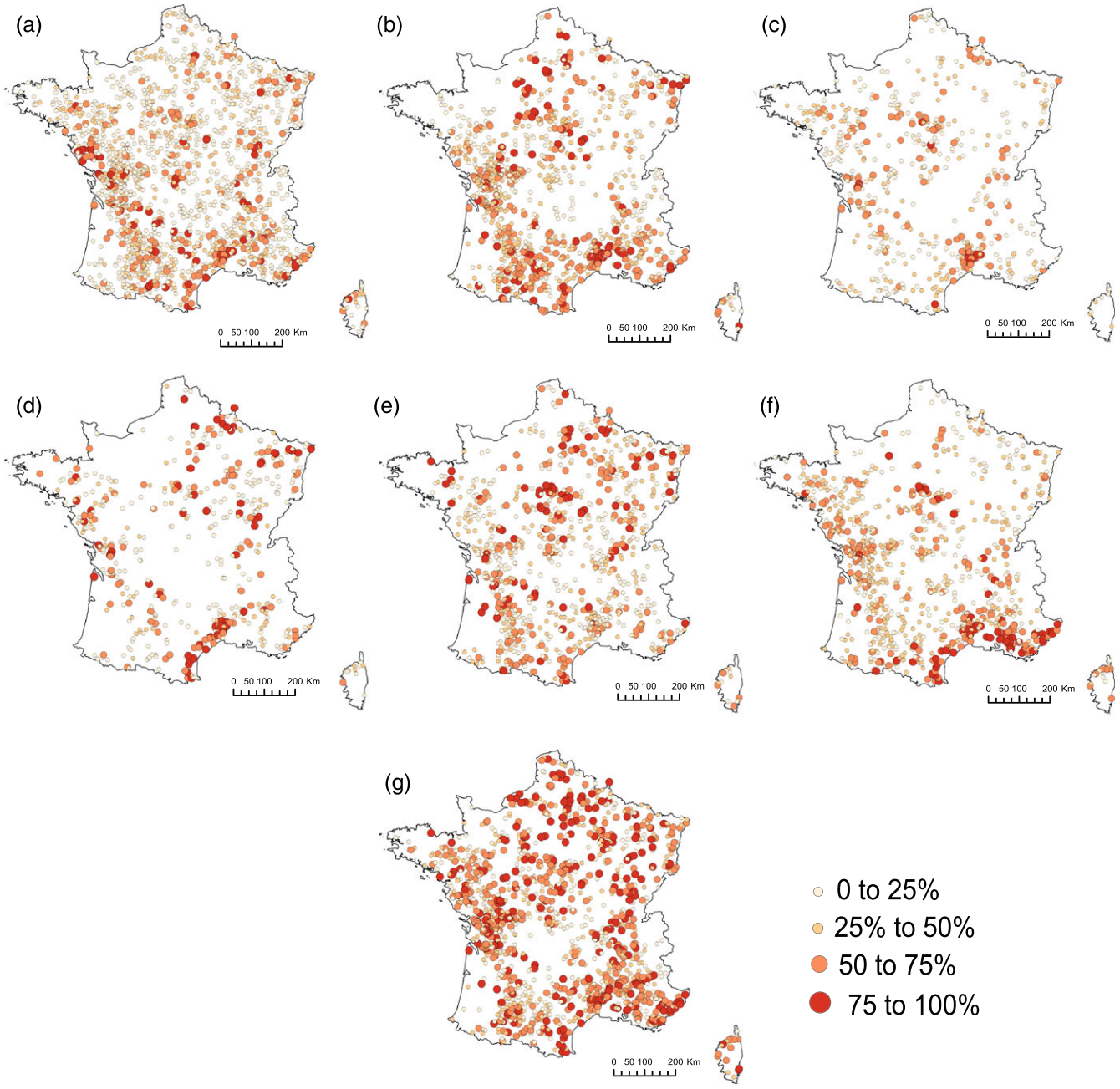

FIGURE 6 Distribution of the percentages of drying observed at Observatoire National des Etiages sites for the years (a) 2012–2017, (b) 2012, (c) 2013, (d) 2014, (e) 2015, (f) 2016, and (g) 2017

and the classifiers aim at reproducing the variability around this average value.

ONDE site characteristics are not identified as drivers by the three classifiers. This is surprising because several studies have shown that the catchment area (Area) is a very important explanatory variable in hydrology and specifically for classifying streams as temporary or perennial (González-Ferreras & Barquín, 2017; Snelder et al., 2013). The catchment altitude (Alti) and streams slope (Slope) were also identified as relevant variable to detect intermittent streams (D'Ambrosio, De Girolamo, Barca, Ielpo, & Rulli, 2017; Snelder et al., 2013). One possible reason is that the HERs are homogeneous. We may expect also that other covariates related, for example, to exchanges between open channels and groundwater systems operating at small scale, riverbed permeability, and aquifer structure, would be relevant to capture the variability between sites. Datry et al. (2016) showed that river network could be very fragmented at the local scale (< 100 m) and two nearby sites can experience a very different drying dynamics albeit sharing a similar climate. Despite our effort to select unaltered ONDE sites, unexpected water withdrawal or release may govern the drying dynamics.

Classifiers use climatic, groundwater, and hydrological information to capture and predict the daily variability of drying state. Both RF and

TABLE 5 Covariates importance (%) determined for each classifier and for each selected HER aggregated as a function of their type defined in Table 1. The bolded values correspond to the highest covariates importance (cumulative or average) for each region HER

	HER	Climate	Hydrology	Groundwater level	ONDE sites characteristics	MPD
ANN	1	**38.5** (2.3)	22.4 (7.5)	13.9 (4.6)	6.5 (2.2)	18.7 **(18.7)**
	2	**66.4** (3.9)	5.5 (1.8)	14.0 (4.7)	6.7 (2.2)	7.4 **(7.4)**
	3	**67.5** (4.2)	10.8 (3.6)	6.1 (2.0)	5.7 (1.9)	9.9 **(9.9)**
	4	**49.8** (3.1)	8.7 (2.9)	25.6 (8.5)	1.6 (0.5)	14.3 **(14.3)**
LASSO	1	**37.6** (9.4)	24.7 (8.2)	0.1 (0.0)	0.1 (0.0)	37.5 **(37.5)**
	2	22.4 (2.3)	11.1 (3.7)	18.2 (6.1)	1.8 (0.6)	**46.4 (46.4)**
	3	15.5 (4.2)	16.8 (5.6)	22.2 (7.4)	0.1 (0.0)	**45.5 (45.5)**
	4	**46.2** (2.6)	1.8 (0.6)	12.7 (4.2)	0.0 (0.0)	39.3 **(39.3)**
RF	1	18.0 (1.4)	20.2 (6.7)	4.1 (1.4)	6.4 (2.1)	**51.3 (51.3)**
	2	37.9 (1.4)	7.7 (2.6)	6.8 (2.3)	6.2 (2.1)	**41.3 (41.3)**
	3	22.4 (1.3)	15.3 (5.1)	11.8 (3.9)	7.5 (2.5)	**42.9 (42.9)**
	4	27.5 (1.8)	11.2 (3.7)	6.6 (2.2)	5.9 (2.0)	**48.8 (48.8)**

Note. ANN: artificial neural network; HER: hydro-ecoregion; LASSO: least absolute shrinkage and selection operator; ONDE: Observatoire National des Etiages; RF: random forest. The first value is the cumulative importance of each covariate in a given type and the value in brackets represents the average importance of covariates.

LASSO seem to put too much weight on MPD that leads them to underestimate drying states when a new situation is encountered in extrapolation as in 2017. ANN gives more importance to climatic covariates in comparison with RF and LASSO. Some studies underlined the importance of meteorological features in drying variability. Abdollahi et al. (2017) have shown the importance to combine precipitation and flow patterns for predicting daily mean streamflow of an IRES. De Girolamo, Barca, Pappagallo, and Lo Porto (2017) identified that errors in meteorological inputs are responsible of the limited performance of the model in predicting stream flow and hydrological indicators in an IRES. The regional climate patterns (Snelder et al., 2013) and the minimum monthly precipitation values in August (González-Ferreras & Barquín, 2017) were relevant variables to both detect and map IRES at the regional scale. Furthermore, the non-linear structure of ANN might enable it to better exploit the day-to-day variability of climate, groundwater, and hydrological data in order to reproduce drying dynamics. Both RF and LASSO may not be flexible enough to take benefit from covariates other than MPD. ANN makes it possible to predict drying states on sites where no drying state was observed when the drought conditions become severe and seems to be more suitable in extrapolation. LASSO and RF seem to require a longer observation period of the flow states provided by the ONDE observations in order to more accurately estimate the MPD and thus hope to improve the drying predictions in extrapolation.

7 | CONCLUSION

The main scientific contribution concerns the exploitation of ONDE observations, discontinuous in time and space, to improve our knowledge about the headwater catchment functioning. Despite the low frequency of the observations, they provide essential information to be combined with statistical methods adapted to discontinuous data in time for reconstructing the dynamics of the local drying states at a daily resolution.

The main conclusion of this study is that statistical classifiers predicting at-site drying states from covariates are better than using the dominant observed flow states. In this context, taking into account of the non-linearity in classification model is of importance to ensure the best predictive performance. The non-linearity of the ANN classifier gives it a greater degree of freedom and, according to our results, made it possible to improve flow state predictions, especially in extrapolation. Moreover, our results underline the importance to consider calibration datasets that span a full range of expected hydrological conditions leading to a better representation of extreme events. Another major conclusion concerns the importance of the predictive information provided by the monthly proportion of observations with drying states (MPD). This covariate reflects the level of flow intermittence of each ONDE site and the classifiers aim at reproducing the variability around this average value. Meteorological variables as covariates appear as important drivers of the flow intermittence although the site characteristics have unexpectedly low relevance.

One of the perspectives to this work would be to explore how this non-linearity is used to predict daily drying dynamics. Studying the statistical decision boundary of ANN especially for weakly intermittent stations constitute a first step toward this goal. Another perspective concerns the determination of additional covariates more locally defined that could be tested to analyse their added value in local drying predictions. All these approaches will have to be studied in contrasted climatic and environmental situations in order to accurately assess the performance of each classifier.

Thanks to this first application, the next step will be to extend this approach in all HERs in France. It is therefore conceivable to use the results of our models, that is, the reconstructed drying dynamics, in the context of ecological studies that focus on the distribution and persistence of aquatic communities in response to flow alterations. In addition, this work could be relevant for watershed management. The knowledge of the duration and the variability of the drying phase is very important for assessing the ecological status of IRES (Mazor, Stein, Ode, & Schiff, 2014; Prat et al., 2014). It could also help to

identify which metrics are the most relevant for detecting human impacts (Datry, Bonada, & Boulton, 2017). Many IRES occur in the headwater of perennial systems and their conservation is widely recognized, especially for the supply of good quality water (Lowe & Likens, 2005). The reconstruction of local drying dynamics could guide stakeholders to improve the ecological restoration and protection of IRES.

ACKNOWLEDGMENTS

The authors wish to thank the two reviewers for their valuable comments, suggestions, and positive feedback on the manuscript. The research project was partly funded by the French Agency for Biodiversity (AFB, formerly ONEMA). This study is based upon works from COST Action CA15113 (SMIRES, Science and Management of Intermittent Rivers and Ephemeral Streams, www.smires.eu), supported by COST (European Cooperation in Science and Technology).

ORCID

Aurélien Beaufort https://orcid.org/0000-0002-6523-8269
Julie Carreau https://orcid.org/0000-0002-0935-9138
Eric Sauquet https://orcid.org/0000-0001-9539-7730

REFERENCES

Abdollahi, S., Raeisi, J., Khalilianpour, M., Ahmadi, F., & Kisi, O. (2017). Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resources Management*, *31*(15), 4855–4874. https://doi.org/10.1007/s11269-017-1782-7

Acuña, V., Datry, T., Marshall, J., Barceló, D., Dahm, C. N., Ginebreda, A., … Palmer, M. A. (2014). Why should we care about temporary waterways? *Science*, *343*(6175), 1080–1081. https://doi.org/10.1126/science.1246666

Acuña, V., Hunter, M., & Ruhí, A. (2017). Managing temporary streams and rivers as unique rather than second-class ecosystems. *Biological Conservation*, *211*, 12–19. https://doi.org/10.1016/j.biocon.2016.12.025

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998) Crop evapotranspiration-Guidelines for computing crop water requirements. *FAO Irrigation and drainage paper*, (56). Fao, Rome, 300(9), D05109.

ASCE. (2000a). Artificial neural networks in hydrology – I: Preliminary concepts. *Journal of Hydrologic Engineering*, *5*(2), 115–123.

ASCE. (2000b). Artificial neural networks in hydrology – II: Hydrologic applications. *Journal of Hydrologic Engineering*, *5*(2), 124–137.

Beaufort, A., Lamouroux, N., Pella, H., Datry, T., & Sauquet, E. (2018). Extrapolating regional probability of drying of headwater streams using discrete observations and gauging networks. *Hydrology and Earth System Sciences*, *22*(5), 3033–3051. https://doi.org/10.5194/hess-22-3033-2018

Bishop, C. M. (2006). *Pattern recognition and machine learning* (p. 229). New York: Springer-Verlag.

Boulton, A. J. (2014). Conservation of ephemeral streams and their ecosystem services: What are we missing?: Editorial. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *24*(6), 733–738. https://doi.org/10.1002/aqc.2537

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brugeron, A., Allier, D., & Klinka, T. (2012). Approche exploratoire des liens entre référentiels hydrogéologique et hydrographique: Premières identifications des piézomètres potentiellement représentatifs d'une relation nappe/rivière et contribution à leur valorisation, *Rapport final BRGM/RP-61047-FR*, 241 pp.

Brunner, M. I., Furrer, R., Sikorska, A. E., Viviroli, D., Seibert, J., & Favre, A.-C. (2018). Synthetic design hydrographs for ungauged catchments: A comparison of regionalization methods. *Stochastic Environmental Research and Risk Assessment*, *32*(7), 1993–2023. https://doi.org/10.1007/s00477-018-1523-3

Bunn, S. E., Thoms, M. C., Hamilton, S. K., & Capon, S. J. (2006). Flow variability in dryland rivers: Boom, bust and the bits in between. *River Research and Applications*, *22*(2), 179–186. https://doi.org/10.1002/rra.904

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., … Zhumanova, M. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, *2*. https://doi.org/10.3389/feart.2014.00026

Cipriani, T., Toilliez, T., & Sauquet, E. (2012). Estimating 10 year return period peak flows and flood durations at ungauged locations in France (in French). *La Houille Blanche*, *4-5*, 5–13. https://doi.org/10.1051/lhb/2012024

Dalrymple, T. (1960). Flood frequency analysis, *U.S. Geol. Surv. Water Supply Pap*., 1543-A.

D'Ambrosio, E., De Girolamo, A. M., Barca, E., Ielpo, P., & Rulli, M. C. (2017). Characterising the hydrological regime of an ungauged temporary river system: A case study. *Environmental Science and Pollution Research*, *24*(16), 13950–13966. https://doi.org/10.1007/s11356-016-7169-0

Datry, T., Bonada, N., & Boulton, A. J. (2017). Chapter 1 - General introduction. In T. Datry, N. Bonada, & A. Boulton (Eds.), *Intermittent rivers and ephemeral streams* (pp. 1–20). Academic Press. https://doi.org/10.1016/B978-0-12-803835-2.00001-2

Datry, T., Boulton, A. J., Bonada, N., Fritz, K., Leigh, C., Sauquet, E., … Dahm, C. N. (2018). Flow intermittence and ecosystem services in rivers of the Anthropocene. *Journal of Applied Ecology*, *55*(1), 353–364. https://doi.org/10.1111/1365-2664.12941

Datry, T., Fritz, K., & Leigh, C. (2016). Challenges, developments and perspectives in intermittent river ecology. *Freshwater Biology*, *61*(8), 1171–1180. https://doi.org/10.1111/fwb.12789

Datry, T., Larned, S. T., & Tockner, K. (2014). Intermittent rivers: A challenge for freshwater ecology. *Bioscience*, *64*(3), 229–235. https://doi.org/10.1093/biosci/bit027

Datry, T., Pella, H., Leigh, C., Bonada, N., & Hugueny, B. (2016). A landscape approach to advance intermittent river ecology. *Freshwater Biology*, *61*(8), 1200–1213. https://doi.org/10.1111/fwb.12645

De Girolamo, A. M., Barca, E., Pappagallo, G., & Lo Porto, A. (2017). Simulating ecologically relevant hydrological indicators in a temporary river system. *Agricultural Water Management*, *180*, 194–204. https://doi.org/10.1016/j.agwat.2016.05.034

De Girolamo, A. M., Lo Porto, A., Pappagallo, G., & Gallart, F. (2015). Assessing flow regime alterations in a temporary river—the River Celone case study. *Journal of Hydrology and Hydromechanics*, *63*(3), 263–272. https://doi.org/10.1515/johh-2015-0027

Eng, K., Wolock, D. M., & Dettinger, M. D. (2016). Sensitivity of intermittent streams to climate variations in the USA: Sensitivity of intermittent streams. *River Research and Applications*, *32*(5), 885–895. https://doi.org/10.1002/rra.2939

Finn, D. S., Bonada, N., Mùrria, C., & Hughes, J. M. (2011). Small but mighty: Headwaters are vital to stream network biodiversity at two levels of organization. *Journal of the North American Benthological Society*, *30*(4), 963–980. https://doi.org/10.1899/11-012.1

Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington, P. J., Leibowitz, S. G., … Nadeau, T.-L. (2013). Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *JAWRA Journal of the American Water Resources Association*, *49*(4), 867–882. https://doi.org/10.1111/jawr.12040. https://doi.org/10.1111/jawr.12040

González-Ferreras, A. M., & Barquín, J. (2017). Mapping the temporary and perennial character of whole river networks: Mapping flow permanence in river network. *Water Resources Research*, *53*, 6709–6724. https://doi.org/10.1002/2017WR020390

Huxter, E. H. H., & (Ilja) van Meerveld, H. J. (2012). Intermittent and perennial streamflow regime characteristics in the Okanagan. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *37*(4), 391–414. https://doi.org/10.4296/cwrj2012-910

Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L., & Marsh, N. (2010). Classification of natural flow regimes in Australia to support environmental flow management: Classification of natural flow regimes in Australia. *Freshwater Biology*, *55*(1), 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x

Larned, S. T., Datry, T., Arscott, D. B., & Tockner, K. (2010). Emerging concepts in temporary-river ecology. *Freshwater Biology*, *55*(4), 717–738. https://doi.org/10.1111/j.1365-2427.2009.02322.x

Leigh, C., Boulton, A. J., Courtwright, J. L., Fritz, K., May, C. L., Walker, R. H., & Datry, T. (2016). Ecological research and management of intermittent rivers: An historical review and future directions. *Freshwater Biology*, *61*(8), 1181–1199. https://doi.org/10.1111/fwb.12646

Leigh, C., & Datry, T. (2017). Drying as a primary hydrological determinant of biodiversity in river systems: A broad-scale analysis. *Ecography*, *40*(4), 487–499. https://doi.org/10.1111/ecog.02230

Leopold, L. B., Wolman, M. G., & Miller, J. P. (1964). *Fluvial processes in geomorphology*. New York, USA: Dover Publications.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Lowe, W. H., & Likens, G. E. (2005). Moving headwater streams to the head of the class. *Bioscience*, *55*(3), 196–197. https://doi.org/10.1641/0006-3568(2005)055[0196:MHSTTH]2.0.CO;2

Mazor, R. D., Stein, E. D., Ode, P. R., & Schiff, K. (2014). Integrating intermittent streams into watershed assessments: Applicability of an index of biotic integrity. *Freshwater Science*, *33*(2), 459–474. https://doi.org/10.1086/675683

van Meerveld, H. J. I., Vis, M. J. P., & Seibert, J. (2017). Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences*, *21*(9), 4895–4905. https://doi.org/10.5194/hess-21-4895-2017

Meyer, J. L., Strayer, D. L., Wallace, J. B., Eggert, S. L., Helfman, G. S., & Leonard, N. E. (2007). The contribution of headwater streams to biodiversity in river networks1: The contribution of headwater streams to biodiversity in river networks. *JAWRA Journal of the American Water Resources Association*, *43*(1), 86–103. https://doi.org/10.1111/j.1752-1688.2007.00008.x

Nadeau, T.-L., & Rains, M. C. (2007). Hydrological connectivity between headwater streams and downstream waters: how science can inform policy1: Hydrological connectivity between headwater streams and downstream waters: How science can inform policy. *JAWRA Journal of the American Water Resources Association*, *43*(1), 118–133. https://doi.org/10.1111/j.1752-1688.2007.00010.x

Nowak, C., & Durozoi, B. (2012). Observatoire national des Etiages, Note technique, ONEMA.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1–2), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Olden, J. D., Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, *178*(3–4), 389–397. https://doi.org/10.1016/j.ecolmodel.2004.03.013

Pella, H., Lejot, J., Lamouroux, N., & Snelder, T. (2012). Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. *Géomorphologie: Relief, Processus, Environnement*, *18*(3), 317–336. https://doi.org/10.4000/geomorphologie.9933

Prat, N., Gallart, F., Von Schiller, D., Polesello, S., García-Roger, E. M., Latron, J., … Froebrich, J. (2014). The mirage toolbox: An integrated assessment tool for temporary streams: Mirage toolbox. *River Research and Applications*, *30*(10), 1318–1334. https://doi.org/10.1002/rra.2757. https://doi.org/10.1002/rra.2757

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., … Morel, S. (2008). Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, *47*(1), 92–107. https://doi.org/10.1175/2007JAMC1636.1

Reynolds, L. V., Shafroth, P. B., & LeRoy Poff, N. (2015). Modeled intermittency risk for small streams in the Upper Colorado River Basin under climate change. *Journal of Hydrology*, *523*, 768–780. https://doi.org/10.1016/j.jhydrol.2015.02.025

Sarremejane, R., Cañedo-Argüelles, M., Prat, N., Mykrä, H., Muotka, T., & Bonada, N. (2017). Do metacommunities vary through time? Intermittent rivers as model systems. *Journal of Biogeography*, *44*(12), 2752–2763. https://doi.org/10.1111/jbi.13077

Sauquet, E., & Catalogne, C. (2011). Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France. *Hydrology and Earth System Sciences*, *15*(8), 2421–2435. https://doi.org/10.5194/hess-15-2421-2011

Sauquet, E., Gottschalk, L., & Krasovskaia, I. (2008). Estimating mean monthly runoff at ungauged locations: An application to France. *Hydrology Research*, *39*(5–6), 403. https://doi.org/10.2166/nh.2008.331–423.

Snelder, T. H., Datry, T., Lamouroux, N., Larned, S. T., Sauquet, E., Pella, H., & Catalogne, C. (2013). Regionalization of patterns of flow intermittence from gauging station records. *Hydrology and Earth System Sciences*, *17*(7), 2685–2699. https://doi.org/10.5194/hess-17-2685-2013

Stubbington, R., England, J., Wood, P. J., & Sefton, C. E. M. (2017). Temporary streams in temperate zones: Recognizing, monitoring and restoring transitional aquatic-terrestrial ecosystems: Temporary streams in temperate zones. *Wiley Interdisciplinary Reviews: Water*, *4*(4), e1223. https://doi.org/10.1002/wat2.1223

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Turner, D. S., & Richter, H. E. (2011). Wet/dry mapping: Using citizen scientists to monitor the extent of perennial surface flow in dryland regions. *Environmental Management*, *47*(3), 497–505. https://doi.org/10.1007/s00267-010-9607-y

Vadher, A. N., Millett, J., Stubbington, R., & Wood, P. J. (2018). Drying duration and stream characteristics influence macroinvertebrate survivorship within the sediments of a temporary channel and exposed gravel bars of a connected perennial stream. *Hydrobiologia*, *814*, 121–132. https://doi.org/10.1007/s10750-018-3544-9

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth ed.). New York: Springer. 0-387-95457-0. https://doi.org/10.1007/978-0-387-21706-2

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., & Soubeyroux, J.-M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, *30*(11), 1627–1644. https://doi.org/10.1002/joc.2003

Wasson, J.-G., Chandesris, A., Pella, H., & Blanc, L. (2002). Typology and reference conditions for surface water bodies in France: The hydro-ecoregion approach. *TemaNord*, *566*, 37–41.

Zou, H., & Hastie, T. (2018). Elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA. R package version 1.1. Available at: https://CRAN.R-project.org/package=elasticnet

How to cite this article: Beaufort A, Carreau J, Sauquet E. A classification approach to reconstruct local daily drying dynamics at headwater streams. *Hydrological Processes*. 2019;33:1896–1912. https://doi.org/10.1002/hyp.13445

APPENDIX A

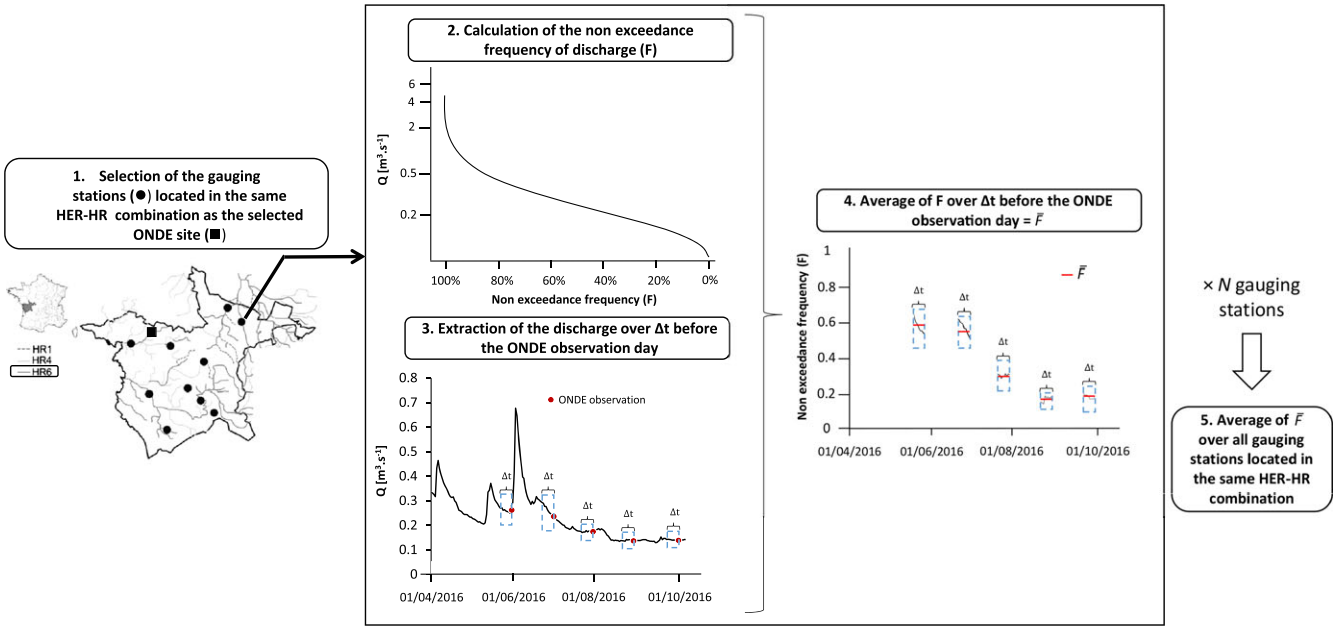

FIGURE A1 Procedure to calculate the hydrological covariates FQΔt at a given Observatoire National des Etiages site located in a subregion hydro-ecoregion (HER)–HR. The value of Δt is 0, 5, and 10 days to calculate FQ0, FQ5, and FQ10, respectively