Extrapolating regional probability of drying of headwater streams using discrete observations and gauging networks

Aurélien Beaufort, Nicolas Lamouroux, Hervé Pella, Thibault Datry, and Eric Sauquet
Irstea, UR RiverLy, centre de Lyon-Villeurbanne, 5 rue de la Doua CS 20244, 69625 Villeurbanne, France

Correspondence: Aurélien Beaufort (aurelien.beaufort@irstea.fr)

Received: 9 November 2017 – Discussion started: 15 November 2017
Revised: 23 March 2018 – Accepted: 29 April 2018 – Published: 24 May 2018

Abstract. Headwater streams represent a substantial proportion of river systems and many of them have intermittent flows due to their upstream position in the network. 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 objective of this paper is to quantify how discrete (in space and time) field observations of flow intermittence help to extrapolate over time the daily probability of drying (defined at the regional scale). Two empirical models based on linear or logistic regressions have been developed to predict the daily probability of intermittence at the regional scale across France. Explanatory variables were derived from available daily discharge and groundwater-level data of a dense gauging/piezometer network, and models were calibrated using discrete series of field observations of flow intermittence. The robustness of the models was tested using an independent, dense regional dataset of intermittence observations and observations of the year 2017 excluded from the calibration. The resulting models were used to extrapolate the daily regional probability of drying in France: (i) over the period 2011–2017 to identify the regions most affected by flow intermittence; (ii) over the period 1989–2017, using a reduced input dataset, to analyse temporal variability of intermittence at the national level. The two empirical regression models performed equally well between 2011 and 2017. The accuracy of predictions depended on the number of continuous gauging/piezometer stations and intermittence observations available to calibrate the regressions. Regions with the highest performance were located in sedimentary plains, where the monitoring network was dense and where the regional probability of drying was the highest. Conversely, the worst performances were obtained in mountainous regions. Finally, temporal projections (1989–2016) suggested the highest probabilities of intermittence (>35%) in 1989–1991, 2003 and 2005. A high density of intermittence observations improved the information provided by gauging stations and piezometers to extrapolate the temporal variability of intermittent rivers and ephemeral streams.

1 Introduction

Headwater streams represent a substantial proportion of river systems (Leopold et al., 1964; Nadeau and Rains, 2007; Benstead and Leigh, 2012). From an ecological point of view, headwater catchments are at the interface between terrestrial and aquatic ecosystems and they often harbour a unique biodiversity with a very high spatial turnover (Meyer et al., 2007; Clarke et al., 2008; Finn et al., 2011). Their contribution to the functioning of hydrographic networks is essential: sediment flows, inputs of particulate organic matter and nutrients, refugia/colonisation, sources of aquatic organisms (Meyer et al., 2007; Finn et al., 2011).

Headwater streams are generally naturally prone to flow intermittence, i.e. streams which stop flowing or dry up at some point in time and space, mainly due to their upstream position in the network and their high reactivity to natural or human disturbances (Benda et al., 2005; Datry et al., 2014b). These waterways which cease flow and/or dry are referred to as intermittent rivers and ephemeral streams (IRES). The geographic extent of IRES is poorly documented due to mapping limitations (digital elevation models, satellite images, aerial photos) and because of their size and their location (Leopold et al., 1994; Nadeau and Rains, 2007; Benstead and Leigh, 2012; Fritz et al., 2013). However, the proportion of
IRES in hydrological networks can be very large: for example, they represent 60% of the length of rivers in the United States (Nadeau and Rains, 2007) and are considered to represent probably more than 50% of the global hydrological network (Larned et al., 2010; Datry et al., 2014b). Considering only gauging stations with continuous records may lead to severe underestimation of their regional extent (Snelder et al., 2013; De Girolamo et al., 2015; Eng et al., 2016).

Recently, IRES have seen a marked increase in interest stimulated by the challenges of water management facing the global change context (water scarcity issues, climate change impact, etc.) (Acuña et al., 2014; Datry et al., 2016b). Studies have characterised the hydrological functioning of IRES (Gallart et al., 2012; Costigan et al., 2016; Sarremejane et al., 2017) to assess the effects of flow intermittence on aquatic ecosystems (Larned et al., 2010; Datry et al., 2016b; Leigh et al., 2016; Leigh and Datry, 2017). IRES have been altered due to human actions (abstraction, hill dams, low-water support, pollution, etc.) despite their high and unique biodiversity (Datry et al., 2014a; Garcia et al., 2017a). In addition, some perennial streams are becoming intermittent due to global change, water abstraction or river damming (Skoulikidis et al., 2009), and the extent of IRES may increase in the future (Döll and Schmied, 2012; Jaeger et al., 2014; Pumo et al., 2016; Garcia et al., 2017b; De Girolamo et al., 2017a).

A better hydrological understanding of IRES is now essential and improved management requires knowledge of both the spatial extent and arrangement of IRES within the river network (Boulton, 2014; Acuña et al., 2017). Efforts have been made to estimate the spatial distribution of IRES at the catchment scale (Skoulikidis et al., 2011; Datry et al., 2016a), at the regional scale (Gómez et al., 2005) and at the national scale (Snelder et al., 2013). In France, Snelder et al. (2013) suggested a classification of IRES regimes and spatialised their distribution. Based on an analysis of the continuous gauging network, they showed that the proportion of IRES accounted for 20 to 39% of the hydrographic network. The accuracy of the obtained map is highly dependent on the density of the flow monitoring network. The installation of additional gauging stations is expensive and headwater systems may be difficult to monitor due to active geomorphology processes or to difficult access.

As a promising tool to advance the mapping of IRES, citizen science creates opportunities to overcome the lack of hydrological data, contributes to densifying the flow-state observation network (Turner and Richter, 2011; Buytaert et al., 2014; Datry et al., 2016b) and could be used for hydrological model calibration (van Meerveld et al., 2017). In France, Datry et al. (2016a) used such data to describe the spatio-temporal dynamics of aquatic and terrestrial habitats within five river catchments located in the western part of France. They showed that processes resulting in flow intermittence were complex at a fine scale and could vary substantially among nearby catchments. However, these data were only available in a few catchments, limiting any attempt to map large-scale patterns of flow intermittence in river networks. Since this first attempt, new sources of observational data have become available in France thanks to the ONDE network (Observatoire National des Etiages, https://onde.eaufrance.fr, last access: 22 May 2018). This unique network in Europe provides frequent discrete field observations (five inspections per year) of the flow intermittence across more than 3300 sites throughout France and located mostly in headwater areas.

However, discrete observations of intermittence do not provide any information on the persistence of dry conditions between two consecutive dates of observation. The rewetting–drying events could have significant impacts on communities whose survival is conditioned by the duration/frequency of drying. The duration of drying is of importance for ecologists, as one key driver for the composition and persistence of aquatic species (Vardakas et al., 2017; Kelso and Entrekin, 2018; Vadher et al., 2018). Temporal extrapolations of river flow regimes are thus necessary to summarise the different facets of flow intermittence at various timescales, from daily to inter-annual.

The main objective of this paper is to use discrete (in space and time) field observations of flow intermittence to extrapolate over time the daily probability of drying (averaged at the regional scale). We first carried out a quantitative analysis of the ONDE network data in order to characterise the information that they contribute in comparison with the data resulting from the conventional hydrological monitoring. Then, we developed two empirical models based on linear or logistic regressions to convert discontinuous series of flow intermittence observation from ONDE into continuous daily probability of drying, defined at the regional scale across France. Explanatory variables were derived from available continuous daily discharge and groundwater-level data of a dense gauging/piezometer network, and models were calibrated using the ONDE discrete observations. The robustness of the models was tested using (1) an independent, dense regional dataset of intermittence observations and (2) observations of the year 2017 excluded from the calibration. Finally, the resulting models were used to extrapolate the regional probability of drying in France: (i) over the period 2012–2017 to identify the regions most affected by flow intermittence; (ii) over the period 1989–2017, using a reduced input dataset, to analyse temporal variability of flow intermittence at the national level.

2 Material and methods

2.1 Study area

The study area is continental France and Corsica (550,000 km²). France is located in a temperate zone characterised by a variety of climates due to the influences
of the Atlantic Ocean, the Mediterranean Sea and mountain areas.

We defined regions as combinations of “level-2 Hydro-
EcoRegions” (HER2) and classes of hydrological regimes
(HR). Hydro-EcoRegion (HER) corresponds to a typology
developed for river management in accordance with the Eu-
ropean Water Framework Directive. The Hydro-EcoRegion
classification includes 22 “level-1 Hydro-EcoRegions”
(HER1) based on geology, topography and climate, and con-
sidered the primary determinants of the functioning of wa-
ter ecosystems (Wasson et al., 2002). HER2 correspond
to a finer classification accounting for stream size. HER2
have a mean drainage area of 5000 km$^2$ (between 100 and
27 000 km$^2$). The hydrological regime classes (HR) were
identified by reference to the work carried out by Sauquet
et al. (2008) where it was possible to distinguish rainfall-
fed regimes, transition and snowmelt-fed river flow regimes.
Overall, we used 280 regions (that is, HER2–HR combina-
tions) with a mean drainage area of 1400 km$^2$ (between 4
and 20 000 km$^2$).

2.2 ONDE dataset discrete national flow-state
observations

The ONDE network was set up in 2012 by the French Bio-
diversity Agency (AFB, formerly ONEMA) with the aim of
constituting a perennial network recording summer low-flow
levels and used to anticipate and manage water crisis during
severe drought events (Nowak and Durozoi, 2012).

There are 3300 ONDE sites distributed throughout France
(Fig. 1). ONDE sites are located on headwater streams with
a Strahler order of strictly less than 5 and balanced across
HER2 regions to take into account the representativeness of
the hydrological contexts (Nowak and Durozoi, 2012). The
ONDE network is stable over time. Observations have been
made monthly (around the 25th) by trained AFB staff, be-
tween April and September, every year since 2012. One of
the statuses is assigned at each observation among “visible
flow”, “no visible flow” and “dried out”. Here, we consider
two intermittency statuses: “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”).
The proportion of drying sites determined on the basis of
the ONDE network for each HER2–HR combination is con-
sidered a good estimate of the daily regional probability of
drying (RPoDONDE) of streams with a Strahler order of less
than 5. Observed values of RPoDONDE are calculated as follows:

$$\text{RPoDONDE}(d) = \frac{N_{\text{drying}}^{\text{HER2} \rightarrow \text{HR}}}{N_{\text{flowing}} + N_{\text{drying}}^{\text{HER2} \rightarrow \text{HR}}},$$

where $d$ denotes the observation date of the ONDE network,
$N_{\text{drying}}$ and $N_{\text{flowing}}$ are the number of drying and of flowing

![Figure 1. Location of the 3300 ONDE sites and partition into HER2.](image)

status observed at ONDE sites located in the same HER2–
HR combination at the observation date $d$, respectively.

Figure 2 illustrates the complementary nature of the
ONDE network to the already existing HYDRO (http://www.
hydro.eaufrance.fr, last access: 22 May 2018) French river
flow monitoring network. The ONDE sites and a set of 1600
gauging stations available in the HYDRO database have been
projected on the RHT (Theoretical Hydrographic Network;
Pella et al., 2012) river network and the drainage area and
the elevation have been estimated. A large part of ONDE
sites are located on small headwater streams, with 70% of
the sites with a drainage area of less than 50 km$^2$, while most
of the gauging stations record flows of catchments of medium
size (between 100 and 500 km$^2$). Only four stations display
a drainage area of more than 1000 km$^2$. The distributions
of elevation of the two databases look similar. The ONDE sites
are mostly located on rivers with an elevation below 200 m
(75% of sites). The ONDE sites are sparse at high elevations
(95 sites located above 1000 m). This bias is likely due to
access difficulties in mountainous areas.

2.3 POC dataset: a denser regional dataset used for
independent validation

A spatially denser citizen science dataset of flow-
state observations in western France (Poitou-Charente re-
region) (http://atlas.observatoire-environnement.org, last ac-
cess: 22 May 2018) has been used as a validation dataset to
test the robustness of our models calibrated with the ONDE
dataset. The POC monitoring (2011–2013) covered more
than 4000 km of river length across 20 catchments. Each river
was entirely surveyed every 1st and 15th of each month be-
between June and October, resulting in eight observations per year. Four intermittency statuses were available in the POC dataset (Datry et al., 2016a), but to allow comparisons with the ONDE network, we pooled the two “Flowing” and “Low Flow” POC statuses into a single “Flowing” status and the two “No flow” and “Dry” statuses into the “Drying” status. This dataset is available as maps with flow states assigned to the inspected streams. Values of RPoD at each POC observation date are calculated in the same way as RPoD\textsubscript{ONDE}. Thus RPoD\textsubscript{POC} is given by the ratio between the number of drying statuses and the total number of observations at each inspected stream located in the same HER2–HR.

2.4 Explanatory discharge dataset

Two discharge datasets (continuous daily time series) were used as explanatory variables of discrete intermittence observations, with the objective of extrapolating the intermittence frequency over time. The two datasets included time series of daily discharge extracted from the French river flow monitoring network (HYDRO database, http://www.hydro.eaufrance.fr/, last access: 22 May 2018): (i) the 2011–2017 dataset with full records available between 1 January 2011 and 31 June 2017; (ii) the 1989–2017 dataset concerning a reduced number of gauging stations and providing daily discharges between 1 January 1989 and 31 June 2017. According to the hydrometric services in charge of the selected gauging stations, high quality of measurements was ensured and observed discharges were not or only slightly altered by human actions.

The 2011–2017 dataset was composed of 1600 gauging stations distributed across France. Each stream where a HYDRO gauging station is located has been defined as IRES or perennial. Several definitions of IRES can be found in the literature (Huxter and van Meerveld, 2012; Eng et al., 2016; Reynolds et al., 2015). In this study, we considered stations to be intermittent when five consecutive days with discharge less than 1 L s\(^{-1}\) have been observed during the period of record.

The 1989–2017 dataset consisted of 630 gauging stations selected with less than 5% of missing data (continuous or not) during the period 1989–2017. This dataset was thereafter used to estimate the regional probability of drying before the creation of the ONDE network.

2.5 Explanatory groundwater-level dataset

Because groundwater resources influence stream intermittence, we used available time series of the daily groundwater level available in the ADES database (http://www.ades.eaufrance.fr, last access: 22 May 2018) at sites identified as involved in groundwater–surface water exchanges (Brugeron et al., 2012). Similarly to the discharge data, two sets of groundwater-level data with records available over the two periods 2011–2017 and 1989–2017 have been selected. The level of alteration of groundwater levels by water withdrawal is unknown because no information is available at this scale.

The 2011–2017 dataset was composed of 750 piezometers with daily groundwater-level data with less than 5% of missing data (continuous or not). The selection of the 1989–2017 dataset was not easy because few groundwater-level measurements were available in the database before 2000. For example, only five piezometers met the tolerance limit on missing values considered for the 1989–2017 discharge dataset. In order to extend the dataset and because groundwater levels were less variable than stream discharges, the proportion of permitted gaps was fixed to 20% between 1989 and 2017. This led us to select 150 piezometers. Thereafter, when the missing data period was less than 10 days, groundwater levels were reconstructed by linear interpolation in order to reduce the proportion of missing values to less than 5% for the 150 piezometers selected.
2.6 Statistical modelling of regional probability of drying

The parametric modelling strategy was based on five main steps (Fig. 3). The first step consisted in selecting all ONDE sites, gauging stations and piezometers located in the same HER2–HR combination. When the total number of gauging stations and piezometers was less than five for a HER2–HR combination, we merged the HER2–HR combination with a neighbouring one located in the same HER1. This was done for 20 of the 280 regions. The second step consisted in calculating the RPoD_{ONDE} for each observation date (5 year⁻¹) and for all selected ONDE sites. In a third step, a flow duration curve was determined for each selected HYDRO gauging station. The average non-exceedance frequency of the observed discharge at gauging stations was averaged for the date of observation (d) at ONDE sites and the 5 days preceding the observation. The lag of 6 days accounted for the fact that ONDE survey dates in a region could differ by 5 days, and accounted for the inertia of physical processes (e.g. storage capacity). The same operation was carried out with selected piezometers. Finally the hydrological conditions are described by the average (across stations) \( F \) of the non-exceedance frequencies of discharge (\( F_q \)) and groundwater levels (\( F_{gw} \)) with respect to the relative proportions of gauging stations and piezometers:

\[
F(d) = \frac{\sum_{i=1}^{N_q} F_q_i + \sum_{j=1}^{N_{gw}} F_{gw_j}}{(N_q + N_{gw})},
\]

where \( F_q_i \) denotes the average non-exceedance frequency of discharge at the gauging station \( i \) calculated between \( d \) and \( d-5 \); \( F_{gw_j} \) the average non-exceedance frequency of groundwater levels at the piezometer \( j \) calculated between \( d \) and \( d-5 \); \( N_q \) the number of gauging stations selected in a HER2–HR combination; and \( N_{gw} \) the number of selected piezometers selected in the HER2–HR combination. The fourth step consisted in estimating the RPoD_{ONDE} as a function of \( F \). Two types of regression were fitted for each HER2–HR combination across France.
A truncated logarithmic linear regression (LLR), with two parameters $\alpha_1$ and $\beta_1$:

$$\text{RPoD}_{\text{LLR}}(d) = \begin{cases} \min(1; \alpha_1 \times \ln(F(d)) + \beta_1) & \text{when } F < F_0 \\ 0 & \text{when } F \geq F_0. \end{cases}$$

$F_0$ was fixed as the value of non-exceedance frequencies of discharge and groundwater levels at which no more drying status was observed across the ONDE network ($\text{RPoD}_{\text{ONDE}} = 0$).

A logistic regression (LR) with two parameters $\alpha_2$ and $\beta_2$:

$$\text{Logit}(\text{RPoD}_{\text{LR}}(d)) = \ln \left( \frac{\text{RPoD}_{\text{LR}}(d)}{1 - \text{RPoD}_{\text{LR}}(d)} \right) = \alpha_2 \times F(d) + \beta_2.$$  \hfill (4)

LR is a multivariate analysis method well known for its relevance in binary classification issues (Lee, 2005). The RPoD$_{\text{LR}}$ was then calculated as follows in Eq. (5):

$$\text{RPoD}_{\text{LR}}(d) = \frac{\exp(\alpha_2 F(d) + \beta_2)}{1 + \exp(\alpha_2 F(d) + \beta_2)}.$$  \hfill (5)

Models were calibrated against observations available during the same period, 2012–2016, leaving out the year 2017 for an independent validation test. However, for the continuous temporal extrapolations (one over 2011–2017, the other over 1989–2017), two models were built with different piezometers and gauging stations selected as explanatory variables (see Sects. 2.4 and 2.5). Thus there are two sets of regression parameters specific to each dataset for both the LLR and LR models, leading to different prediction of RPoD.

Finally, in a fifth step, a daily regional probability of drying (RPoD) could be predicted for each HER2–HR combination with both models following analytical formulas (Eqs. 3 and 5).

### 2.7 Model robustness: validation using independent datasets

We used (1) the POC independent data and (2) the 2017 ONDE year to test the robustness of the LLR and LR model to predict the intermittence frequency (1) in space and (2) over time. Note that when predicting on the POC datasets, a new model was calibrated using only ONDE sites located out of POC streams.

For both datasets (POC and ONDE, 2017), the relative performance of the LLR and LR models was compared in multiple ways using both the 2011–2017 and 1989–2017 datasets. The performance of each model was evaluated by the Nash–Sutcliffe efficiency criterion (NSE) (Nash and Sutcliffe, 1970):

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N} (\text{RPoD}_{\text{ONDE}}(i) - \text{RPoD}_{\text{pri}}(i))^2}{\sum_{i=1}^{N} (\text{RPoD}_{\text{ONDE}}(i) - \text{RPoD}_{\text{ONDE}}^0)^2},$$

where RPoD$_{\text{ONDE}}(i)$ is the average proportion of drying statuses over the ONDE sites located in the HER2–HR combination at the $i$th observation date, RPoD$_{\text{pri}}(i)$ is the predicted regional probability of drying at the $i$th observation date, RPoD$_{\text{ONDE}}^0$ is the mean of RPoD$_{\text{ONDE}}(i)$ over the period and $N$ is the total number of observations in the ONDE network for each HER2–HR combination.

### 2.8 Model prediction

Both models have been calibrated over the period 2012–2016 and were then applied in a fifth step to predict the daily RPoD in France (Fig. 3). The RPoD was firstly predicted over the period 2012–2016 in order to identify the most affected regions by flow intermittence using the 2011–2017 datasets. The second application concerned the extrapolation of RPoD in France over a longer period using the 1989–2017 dataset to analyse the temporal variability of flow intermittence at the national level. It should be noted that model predictions only concern streams with a Strahler order of lower than 5 due to the ONDE site location.

### 3 Results

#### 3.1 Quantitative analysis

##### 3.1.1 Inter-annual intermittence according to the raw discrete ONDE network

A total of 1127 ONDE sites have recorded at least one drying event during the period 2012–2016 representing 35 % of the 3300 ONDE sites. From the ONDE database the probability of drying at the country scale was computed as the total number of drying statuses over France divided by the total number of ONDE observations available during statuses the same year (Fig. 4a). Between 2012 and 2016, the most critical year is 2012, with 15 % of drying statuses followed by 2015 (14 %) and 2014 (14 %) (Fig. 4a). The years 2013 and 2014 are less affected with only 6 % of drying statuses observed (Fig. 4a).

Drying events mainly occur between July and September, but the evolution of the month’s proportion of drying can differ between years (Fig. 4b). In more detail, water levels in 2012 decrease in August when the proportion of drying is 27 %, and the situation lasts until the end of September with 25 % of drying (Fig. 4b). In 2013, the proportion of drying is
lower than in 2012, but follows the same pattern with an increase at the end of July (3 %), reaching 9 % in August and in September. In 2014, the first peak of drying (5 %) is reached early in June. Then, the proportion of drying decreases in July (3 %) and increases slightly in August (4 %), reaching 7 % in September. In 2015, the critical period occurs at the end of July with 19 % of drying statuses, and the proportion of drying decreases slightly at the end of August (17 %) until it reaches 9 % in September. Finally, in 2016, the situation gradually deteriorates every month, reaching 20 % of drying statuses in August, and 28 % in September.

Between 2012 and 2016, a proportion of drying higher than 50 % is recorded on 93 ONDE sites and their spatial distribution is very patchy at the scale of France (black and dark grey dots, Fig. 5a). There are only 158 ONDE sites with at least one drying event every year and a variability of drying locations can be observed across years. The south-east of France is heavily affected by rivers drying, where the proportion of drying can exceed 75 % annually (black dots, Fig. 5b–f). The north-western part of France is less affected, although many ONDE sites show a proportion of drying observed above 50 % in 2014 and 2016 (Fig. 5d and f). Northeastern France is rather affected in 2012, 2014 and 2015, where several ONDE sites have more than 75 % of drying statuses (Fig. 5b, d and e). South-western France is particularly affected in 2012 and 2015 (Fig. 5b and e).

3.1.2 Comparison of flow intermittence between the raw ONDE and HYDRO datasets

The HYDRO dataset includes 90 gauging stations located on streams considered IRES, which represents only 5.6 % of the 1600 gauging stations against 35 % for ONDE sites. At the national scale, the number of IRES seems underrepresented in the south-western, central, and north-eastern parts of France and Corsica in comparison with sites experiencing drying in the ONDE network (Fig. 6).

The number of gauging stations with at least one drying event (discharge < 1 L s⁻¹) observed between May and September varies between 79 in 2012 and 47 in 2014 (Table 1). The lowest numbers of gauging stations with drying events are observed in the years 2013 and 2014, while the highest numbers are related to the years 2012, 2015 and 2016. This finding is consistent with the analysis of the ONDE network (Fig. 5a, d). The frequency of drying, corresponding to the ratio between the number of dry days and the total number of days between 1 May and 30 September (153 days), in contrast, is quite constant over the years (~30 %). The number of gauging stations with drying events over more than 50 % of the time varies little between wet years (14 in 2013) and dry years (21 in 2015), unlike ONDE observations, suggesting a significant temporal variability in the frequency of drying between dry and wet years (Fig. 5).

### Table 1. Annual statistics on flow intermittence calculated on HYDRO gauging stations between 1 May and 30 September.

<table>
<thead>
<tr>
<th>Year</th>
<th>Stations with at least one drying event</th>
<th>Stations with drying &gt; 50 %</th>
<th>Frequency of discharge &lt; 1 L s⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>79</td>
<td>19</td>
<td>32.7</td>
</tr>
<tr>
<td>2013</td>
<td>47</td>
<td>14</td>
<td>37.9</td>
</tr>
<tr>
<td>2014</td>
<td>54</td>
<td>15</td>
<td>32.9</td>
</tr>
<tr>
<td>2015</td>
<td>76</td>
<td>21</td>
<td>31.1</td>
</tr>
<tr>
<td>2016</td>
<td>71</td>
<td>19</td>
<td>28.6</td>
</tr>
</tbody>
</table>

3.2 Validation of the predicted regional probability of drying

3.2.1 Regression results

The LLR and LR models, calibrated over the period 2012–2016, perform well with the 2011–2017 dataset, with a mean NSE of 0.8 with the LR model against 0.7 with the LLR
Figure 5. Distribution of the percentages of drying observed at ONDE sites for the years (a) 2012–2016, (b) 2012, (c) 2013, (d) 2014, (e) 2015 and (f) 2016.

Figure 6. Map of ONDE sites and HYDRO gauging stations with at least one drying.

model (Fig. 7a and b). With the LR model, 50% of the HER2–HR combinations obtain a NSE greater than 0.8, representing a coverage of 65% of the French territory, while 33% of HER2–HR combinations display a NSE higher than 0.8 (50% of coverage of France) with the LLR model. Regions with the highest performances are located in sedimentary plains, in the south-east of France and in the Pyrenees Mountains. Conversely, the worst performances are obtained in the mountainous regions of the Alps as well as in the Massif Central. In these regions the size of the HER2 is rather small and the number of ONDE sites, gauging stations and piezometers per HER2–HR combination are certainly too few to derive reliable relations. Despite pooling, estimating RPoD remains impossible for nine HER2–HR combinations (4.5% of coverage of France) because the number of ONDE sites, gauging stations and piezometer sites is insufficient (less than five) to perform the regression analysis.

The performance level is lower when the 1989–2017 dataset is used in models: the mean NSE with the LR and LLR models is 0.7 and 0.6, respectively (Fig. 7c and d).
The LR and LLR models lead to a similar performance range. However, the LR model outperforms the LLR model in terms of number of HER2–HR combinations, with NSE greater than 0.8 (Fig. 7c and d). The performance is sensitive to the dataset. As expected, the best results are obtained with the denser network. A decrease in NSE by more than 0.2 is identified for 5% of the French territory when the 1989–2017 dataset is used (black areas; Fig. 7e and f). The regions
with the most degraded values of NSE are small HER2–HR combinations located in eastern France (Fig. 7e and f).

The decrease in performance is mainly due to the difference in the number of gauging stations and piezometers between the two datasets (Fig. 8). The most degraded NSEs correspond to HER2–HR combinations where the number of gauging stations and piezometers considered in regressions is the most reduced, i.e. with a loss higher than 50% of the stations (black and dark grey dots; Fig. 8a and b). However, the decrease in performance remains low (the difference in NSE is below 0.1 for 75 and 64% of HER2–HR combinations with the LLR and LR models, respectively).

3.2.2 Comparison to the POC database

The observed proportion of drying RPODPOC is rather well simulated by both LLR and LR models with the 2011–2017 explanatory dataset (NSE > 0.7 except for the year 2011; Fig. 9). In addition, the models are able to capture small fluctuations of RPODPOC during the summer period. The best results during the year 2011 are obtained with the LLR model (black curve; Fig. 9) and the LR model overestimates RPODPOC by 3% (dashed grey curve; Fig. 9). In 2012, the decline in water levels is more gradual than in 2011 and a marked peak is reached in September with 40% of RPODPOC (Fig. 9). This pattern is well reproduced by both models with a good fit to all observation points (Fig. 9). The year 2013 is less affected by drying occurrence and the maximum RPODPOC does not exceed 20% (Fig. 9). Curves of both models fit to observations well until the end of August. Note that the LR model is slightly closer to the observations around the peak in September compared to the LLR model. However, the LR model overestimates the RPODPOC at the end of September and in October.

When the 1989–2017 dataset is used for explanatory variables, the simulations of RPD are weakly degraded with both models (Fig. 9d, e, f). However, the simulated pattern is similar to the observed one. The LLR model outperforms the LR model during the 3 years of validation with the 1989–2017 dataset (black curve; Fig. 9d, e, f).

3.2.3 Temporal patterns assessment of models between 2012 and 2017

During the calibration period, the LLR and LR models tend to better simulate the RPD during dry years 2012 and 2016 (NSE = 0.8 with the LLR and LR models; Table 2) than during wet years (e.g. 2014 with NSE < 0.7). The NSEs are lower during the months of May and June when few drying events are observed, while NSEs are much better during the driest months of August and September.

During the validation year of 2017, both models obtain a similar performance over the year independent of datasets (NSE = 0.7).

Monthly NSEs in 2017 follow the same trend as monthly NSEs of the calibration period, with lower NSEs in May (NSEs < 0.4) and June (NSEs = 0.5) and higher NSEs in July, August and September (NSEs = 0.6) with both models independent of datasets. Figure 10 shows the dispersion between predicted RPD and drying statuses observed at ONDE sites in the scatter plot during the validation year 2017 (Fig. 10a and b) in comparison with the year 2012, which obtains the better NSE during calibration period (Fig. 10c and d). The NSEs obtained in 2017 are 0.72 with the LLR model and 0.68 with the LR model against 0.83 and 0.81 in 2012, respectively. The performance is slightly lower in 2017 but remains acceptable with NSEs close to 0.7, and both models seem able to predict RPD from the calibration period.

3.3 Application of regional models

3.3.1 Modelling of intermittencies severity between 2012 and 2016

Both models have been applied using the 2011–2017 dataset. Figure 11 displays the maximum number of consecutive days (DRPOD > 20%) with RPD higher than 20% simulated by both the LLR and LR models. The most affected regions are located in the south-east of France and in the sedimentary plains, which is consistent with the spatial pattern obtained from the ONDE observations (Fig. 5). The most impacted year followed the same hierarchy: the year 2012 is the most critical year, with 30% of France displaying DRPoD > 20% higher than 60 days followed by the year 2015 (20% of France with DRPoD > 20% > 60 days) and 2016 (15% of France with DRPoD > 20% > 60 days) (Fig. 11). The years 2013 and 2014 are weakly affected, with 5% and 6% of France with DRPoD > 20% higher than 60 days, respectively.

The LR model tends to simulate shorter periods of drying, particularly in HER2–HR combinations located in southeastern France in 2013 and 2014 (Fig. 11). However, there is an overall agreement between RPD simulated by both models in terms of spatial and temporal extent of dry streams.

3.3.2 Reconstitution of historical regional probability of drying

The trend temporal patterns of RPD predicted by the two models, considering the 1989–2017 dataset, look similar between 1989 and 2016, and the simulated RPD fit well to RPDONDE (Fig. 12).

The proportion of drying is highly variable over the total simulation period, with alternating dry (1989 to 1991, 2003 to 2006, 2009 to 2012) and wet (1994 to 1995, 2000 to 2002, 2013 to 2014) phases. In spite of inter-annual variability, peaks of RPD occur regularly between August and September, whether in dry years or wet years. This finding is consistent with the preeminence of rainfall-fed river flow regimes with low flows in summer, in France.
4 Discussion

4.1 ONDE network complementarity with conventional flow monitoring network

The analysis of the ONDE observations shows that the proportion of rivers undergoing drying is significantly higher (35%) than that observed with the conventional monitoring (HYDRO database, 8%). This proportion, although related to a short period of records 2012 and 2016, is consistent with the percentage of 39% of river segments classified as intermittent by Snelder et al. (2013). This analysis confirms the under-representation of IRES in the French HYDRO database, and probably others in other countries (flows are often uncontrolled in IRES). Without gauging stations located on headwaters, Snelder et al. (2013) were unable to predict IRES in eastern France (see Fig. 9, p. 2694). The high density of ONDE sites makes it possible to improve the detection of drying events and lead to better understand the spatial distribution of IRES located at the upstream extent of the hydrographic network. The ONDE network encompasses various hydrological conditions, which provides a more accurate assessment of inter-annual variability, differentiating between dry years (2012, 2015 and 2016) and wet years (2013, 2014) with clearly few drying occurrences.

The validation of the LR and LLR models against the spatially dense POC database also demonstrates the spatial representativeness of the ONDE network. Thanks to the qualitative information provided and to models such as the statistical models developed here, it is now possible to capture drying events at the regional scale.

The ONDE sites are located on small headwater streams which can be very reactive to external disturbances (rainfall deficit, change in air temperature, increase in water withdrawals, etc.) and by nature are more likely to be IRES. The gauging stations available in the HYDRO database are located on larger streams and their hydrologic response to changes in external factors (environmental or human) is slower and drying occurred with greater inertia under temperate climate. Their uneven distribution across France does not allow us to accurately characterise the inter-annual variability of drying development. Overall, the ONDE network provides very complementary information to conventional flow monitoring, leading to a better understanding of the processes of drying in upstream catchments.

4.2 Dependency on spatial gauging network density

The performance obtained with the LR and LLR models is slightly better with the 2011–2017 dataset (mean NSE = 0.75) than those obtained with the 1989–2017 dataset (mean NSE > 0.65), whose network is less dense. HER2–HR combinations are the most degraded where the number of monitoring stations is the most decreased between the two
Figure 9. Comparison between observed proportion of drying RPoD<sub>POC</sub> and RPoD predicted by the LLR and LR models with the 2011–2017 dataset in (a) 2011, (b) 2012 and (c) 2013 and with the 1989–2017 dataset in (d) 2011, (e) 2012 and (f) 2013.

Table 2. NSE criteria obtained between 2012 and 2017 with the LLR and LR models calibrated over the period 2012–2016.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LLR model</td>
<td>May</td>
<td>0.2</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>0.6</td>
<td>0.3</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>0.7</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>0.8</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>May–Sep</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>LR model</td>
<td>May</td>
<td>0.2</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Jun</td>
<td>0.6</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Sep</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>May–Sep</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>
datasets. The accuracy of the predictions is dependent on the number of gauging stations, ONDE sites and piezometers available to calibrate the regressions. The highest NSEs are obtained in western sedimentary plains and south-eastern France, where a significant number of streams have dryings regardless of years (Fig. 5). The dominant river flow regime in these regions is mainly influenced by precipitation and the lowest water levels are reached in August and September,
Figure 12. RPoD simulated between 1989 and 2016 with the 1989–2017 dataset with (a) the LR model and (b) the LLR model. The grey area represents the RPoD between the 90th percentile and the 10th percentile simulated on the HER2–HR combination, the black curve represents the average RPoD simulated by the HER2–HR combination and white dots represent the mean RPoD\textsubscript{ONDE} for each observation date. Dates mentioned correspond to the day of the maximum average RPoD simulated by the HER2–HR combination (black curve) of each year.

which corresponds to the monitoring period of the ONDE database. They benefit from a dense monitoring network (gauging stations, ONDE sites, piezometers), which allows a better representation of the hydrological functioning of streams located within the same HER2. Conversely, performance was poor in mountainous areas such as in the Alps or the Massif Central (NSE < 0.4) where river flow regimes are diversified combining rainfall and snowmelt influences. By construction, the area of HER2–HR combination in mountains is reduced, which leads to a limited number of monitoring stations, certainly not sufficient to fit the models. Moreover, the observation period for ONDE sites was limited between May and September and dryings can be missed, particularly for streams influenced by snow or ice melting with potential drying periods in winter. In regions potentially concerned with drying events from the May–September period, the actual ONDE monitoring strategy needs to be adapted to provide reliable temporal observations and extrapolations of drying frequencies.

We have chosen to average the non-exceedance frequencies of flows and groundwater levels in order to increase the monitoring network. If models had been calibrated using only gauging stations, performance will have been globally similar, or slightly better, in some HER2–HR combinations (Fig. 13). Therefore, we could not validate the real gain of using groundwater-level data in addition to discharge data. This is certainly due to the dominant proportion of the gauging stations compared to the piezometers. Indeed, in the 2011–2017 dataset, the proportion of gauging stations is greater than 75 % for more than 70 % of HER2–HR combinations, whereas the proportion of piezometers exceeds 70 % in only 5 % of HER2–HR combinations. Groundwater-level data thus have a small weight in regressions for this dataset. However, in the 1989–2017 dataset, the proportion
of piezometers is greater than 70% in more than 30% of HER2–HR combinations. The presence of piezometers increases the density of the monitoring network in HER2–HR combinations with few available gauging stations. Thanks to groundwater-level data, RPoD can be predicted on more HER2–HR combinations.

4.3 Interest in reconstructing the dynamic regional probability of drying

Spatio-temporal simulation of the probability of drying is crucial for advancing our understanding of IRES ecology and management. Some aquatic species can persist in a dry reach for a few days, weeks or months, while some are highly sensitive to desiccation (Datry, 2012; Storey and Quinn, 2013; Stubbington and Datry, 2013). Estimation of the total duration of days with drying at the reach scale is therefore needed to understand biological patterns in river networks (Kelso and Entreklin, 2018). To our knowledge, no study has proposed to reconstruct daily flow-state time series of headwater streams at the country scale such as France (> 500,000 km²) using discrete observations in time and space. In the literature, studies at national scale remain focused on the detection and the mapping of IRES because these rivers are historically poorly investigated and their proportion in existing hydrographic networks remains inaccurate or misunderstood (Nadeau and Rains, 2007; Snelder et al., 2013). Recently, several studies proposed alternative methodologies in order to estimate metrics in ungauged IRES (Gallart et al., 2016) or to predict daily streamflow in river basins experiencing flow intermittence (De Girolamo et al., 2017b), but remain applicable at the local scale.

This study provides a first regional approach to use discrete data obtained from regular observations. The average non-exceedance frequency is a global hydrological statistic that only captures the hydrological conditions at the regional scale in modelling the RPoD. For rainfall-driven river flow regimes, the effect of rainfall events on flow intermittence at the HER2–HR scale is probably indirectly reflected by the daily discharge and groundwater levels used to calculate the average non-exceedance frequency. However, when more observation data are available, it is likely that including more detailed descriptors of rainfall events and local geology could improve our approach. In France, based on the 2011–2017 dataset, both models suggest the highest values of RPoD along the Mediterranean coast (D_{RPoD} > 20% > 100 days each year). Rivers in this region are subject to a predominantly pluvial regime (Class 7; Sauquet et al., 2008), i.e. hot and dry summers followed by intense rainfall events in autumn, leading to high flows in November (Skoulikidis et al., 2017b). The catchments in this region are small and particularly reactive to environmental changes, making them highly sensitive to flow intermittence. Rivers located in the sedimentary plain in western France are also very impacted by flow intermittence. The regime is also influenced by precipitation and for the basins subject to intense agriculture significant water abstractions during summer in this region reduce water availability in rivers and in aquifers which are no longer able to support the low water levels, leading to increased flow intermittence. Regarding alteration issues in our datasets, we do not have access to the exact location and the volumes of water withdrawal for irrigation purposes. However, due to their upstream location, water availability is expected to be low, which may limit potential withdrawals and as a consequence flow alteration at ONDE sites. The alteration of groundwater levels is unknown because no information is available. However, in sedimentary plains where agricultural crops dominate the landscape, we are not sure that no human action affects low flows. It is important to note that the responses of biological communities to artificial flow intermittence are still poorly understood compared to natural IRES (Datry et al., 2014b; Skoulikidis et al., 2017a).

Figure 13. Comparison of NSE obtained with regression including only the discharge variable as a function of NSE obtained with discharge and groundwater-level variables in the 2011–2017 dataset with (a) the LLR model and (b) the LR model.
4.4 Validity of historical regional probability of drying during severe low-flow period

The second application aimed at reconstructing historical RPoD over the period 1989–2016. Both models suggest the highest values of mean RPoD (> 35 %) in 1989–1991, 2003 and 2005. During these dry years, predicted values of RPoD result from extrapolation but are consistent with published studies (Mérillon and Chaperon, 1990; Moreau, 2004). For example, Mérillon (1992) estimated that for the whole of France, 11,000 km of rivers were dried at the end of the summers of 1989 and 1990. Caillouet et al. (2017) found that the low-flow event observed in 1989–1990 was particularly severe in terms of duration and affected 95 % of France. Snelder et al. (2013) showed from 628 gauging stations that the years 1989–1991, 2003 and 2005 witnessed particularly high values of duration and frequency of drying events. They found that regions with the highest probability of drying were located along the Mediterranean and Atlantic coasts, which is consistent with ONDE observations and with our results.

Both models suggest the same sequence of dry and wet years. However, the application of the LLR model leads to less contrasted RPoD than the LR model (Fig. 12). To illustrate these differences, the RPoD has been simulated by both models with an extreme $F$ of 1 % (Fig. 14). The RPoD_{LLR} is significantly higher and exceeds 80 % in 30 % of the study area against only 5 % of the area with the RPoD_{LR}. On the other hand, models simulate low RPoD in HER2–HR combinations where the RPoD_{ONDE} is very low between 2012 and 2016, even when $F$ was 1 %, because this situation never occurred during the calibration period (Fig. 14). The logistic function of the LR model takes an S-shape which induced a decrease in the slope of the curve toward extreme values observed during the calibration period (2012–2016). The truncated logarithmic function of the LLR model is not bounded and RPoD can reach 100 % during extreme low-flow events by extrapolation. Since the ONDE network monitoring period does not include a period with drought as severe as in the 1990s, it is not currently possible to assess the relative performance of the two models. Refining extrapolated values requires additional information on headwater collected during more severe droughts than those observed during the last 5 years and then gives support to the pursuit of the ONDE network.

5 Conclusion

This paper investigates the spatial and temporal dynamics of the regional probability of drying (RPoD) of headwater streams by taking benefit from qualitative and discontinuous data provided by the ONDE network. Two models based on linear or logistic regressions have been developed and succeeded in reconstructing the temporal dynamics of RPoD. They are based on a strong relationship between the non-exceedance frequencies of discharges and groundwater levels as a function of the proportion of drying statuses observed at ONDE sites per HER2–HR combination. LLR and LR models show similar performance and perform well between 2011 and 2017. The accuracy of predictions is dependent on the number of gauging stations, ONDE sites and piezometers available to calibrate the regressions. Regions with the highest performance are located in the sedimentary plains, where the monitoring network is dense and where the RPoD is the highest. Conversely, the worst performances are obtained in the mountainous regions. Finally, both models have been used to reconstruct historical RPoD between 1989 and 2016 and suggest the highest values of mean RPoD (> 35 %) in 1989–1991, 2003 and 2005. This is consistent with other published studies, but the high density of ONDE sites makes it possible to improve the detection of drying events and lead to better capturing of the spatial distribution of IRES located at the upstream extent of the hydrographic network. Moreover, the duration of drying is of importance for ecologists.
and the prediction of a daily RPoD provides one key driver for the composition and persistence of aquatic species.

From a methodological point of view, our method relating discrete drying observation obtained by citizen science networks to continuous daily gauging data seems robust across the highly diverse (climate and topography) regions of France, and provides good predictions in an independent region excluded from the calibration process (PoC). These two results suggest a potential application of our approach in other countries. Citizen science creates opportunities to overcome the lack of hydrological data, contributes to densifying the flow-state observation network (Turner and Richter, 2011; Buytaert et al., 2014) and remains less expensive than the installation of additional gauging stations to survey flow intermittence. The next step will be to use this regional approach to simulate the RPoD in future periods by taking into account effects of climate change through predicted discharge and groundwater-level data. This would allow quantification of the evolution of the probability of drying between the current period and the different climate projections provided by the latest IPCC Report (IPCC 2014) and would assist decision makers in defining protocols for restoring flows with appropriate measures to preserve aquatic ecosystems (Woelfle-Erskine, 2017).

Secondly, further work is needed to develop an approach capable of reconstructing the drying dynamics locally by differentiating each stream. Our approach remains spatially valid for estimating RPoDs at the scale of HER2–HR combinations, but does not allow characterisation of the variability of drying occurrence between nearby streams within these regions. From a methodological point of view, statistical tools such as neural networks (Breiman, 2001) have shown good ability to assess both the occurrence and extent of perennial and temporary segments (González-Ferreras and Barquín, 2017) and could be investigated as an alternative method to reconstruct locally the temporal variability of drying.

Data availability. Data used in this study are freely available on the Eaufrance data portal (http://www.eaufrance.fr; AFB, 2018). Time series of daily regional probability of drying are available upon request from the corresponding author.

Author contributions. AB, NL and ES developed the main ideas and designed the experiments. AB implemented the algorithms and analysed the results. AB prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors wish to thank Anne van Loon and Catherine Sefton 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).

Edited by: Laurent Pfister
Reviewed by: Catherine Sefton and Anne Van Loon

References


