Universität Zürich[UZH]

Geographisches Institut

# Evaluating the application of extended NWP-driven streamflow forecasts as socio-economic drought warning tool in Swiss catchments

ESS 511 Master's Thesis


**Author:** Luca Panigada, 20-707-824

**Supervised by:** Dr. Massimiliano Zappa (massimiliano.zappa@wsl.ch), Dr. Konrad Bogner (konrad.bogner@wsl.ch)

**Faculty representative:** Prof. Dr. Jan Seibert


24.04.2026

# Abstract

Prolonged low flow periods pose significant challenges to water resources management in Switzerland, affecting both economic activities and ecological stability. Climate projections indicate that Swiss rivers will undergo substantial changes in discharge, with an expected decrease during summer months, coinciding with the peak in water demand. In this context, the ability to anticipate periods of water scarcity is increasingly crucial for an efficient water supply management and for the prevention and mitigation of conflicts in water allocation. This thesis, therefore, seeks to evaluate the possibility of providing early warnings for approaching water scarcity periods and to offer information on water availability and potential withdrawals from rivers over the coming weeks. To this end, extended streamflow ensemble forecasts driven by Numerical Weather Prediction (NWP), with a lead time of up to 46 days, were combined with local water demand estimates. Specifically, a multi-model ensemble configuration composed of 11 different hydrological models was used. A total of 20 Swiss meso-scale catchments, highly influenced by anthropogenic withdrawals, were analyzed over a period spanning from 2019 to 2025. For each catchment, an optimal probabilistic decision threshold was identified by maximizing the $F_1$ score. This threshold was subsequently used to assess the practical issuance of drought warnings. The results showed the feasibility of issuing drought warnings for almost all analyzed catchments. Forecast ability in anticipating low flow conditions exhibited variability across catchments, and years, but showed greater reliability for lead times of up to almost two weeks prior to event onset. Furthermore, the analysis highlighted the potential to provide valuable insights on water availability and possible abstractions over a forecast horizon of 7–14 days.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Low flows in Switzerland

Switzerland is often referred to as the "Water Tower of Europe" because with its rivers, lakes, and glaciers, it holds about 6% of Europe's freshwater resources, while its land area accounts for just 0.4% of the continent (FOPH & FOEN, 2010; Wechsler et al., 2025). Despite this designation, Swiss water bodies are not exempt from periods of water scarcity, with lakes and rivers that can experience prolonged periods of low water levels and reduced flow.

Rivers low flow occurrence is the result of the interaction between climate drivers and catchment storage processes (Kohn et al., 2019). In Switzerland, low flows are often associated with prolonged periods of little or no precipitation due to persistent high-pressure systems, especially in summer, but they can also occur during intense cold periods, when snow prevents immediate water recharge into the rivers (Kohn et al., 2019). The high topographical variability of Switzerland also plays an important role in shaping river flow regimes and the seasonality of low flows. Catchments with high elevations, such as those in the Alps, usually experience a minimum flow during the winter months, while for the catchments on the Swiss Plateau, the low flow season generally occurs during the summer and autumn months (Blanc & Schädler, 2014).

In recent decades Swiss rivers have experienced prolonged periods of low flow. Notable examples include the hydrological droughts of 2003, 2011, 2015, 2018 and 2022 (Floriancic et al., 2020). Extreme droughts, such as that of the summer of 2003 represent an emblematic example in illustrating the issues that intense and long-lasting periods of low flow can cause. Following this event, the Federal Office for the Environment (FOEN) conducted a detailed analysis of the effects of this drought on Swiss water bodies, examining the impacts observed on both humans and on the environment (BUWAL et al., 2004). One of the sectors hardest hit by the 2003 water scarcity was agriculture, with losses estimated at approximately 350 million CHF by the Federal Office for Agriculture (FOAG). From July to October, due to low river flows, many cantons were forced to restrict or prohibit water withdrawals for irrigation, leading in some cases to tensions between farmers and authorities (BUWAL et al., 2004). In several cantons, small hydropower plants had to suspend their operations to comply with residual flow requirements, while drinking water supply issues were limited to small and isolated waterworks not connected to the main network (BUWAL et al., 2004). Aquatic fauna was also severely affected, with approximately 245 km of rivers experiencing partial or total drying, resulting in the death of tens of thousands of fish (BUWAL et al., 2004).

A comparable event also occurred in the summer of 2018. Starting in July, many small streams across Switzerland recorded very low flows, and in August several rivers experienced discharge values even lower than those observed in 2003 (BAFU et al. (Hrsg.), 2019). In 13 cantons, restrictions on water withdrawals from rivers for irrigation were issued, and in the cantons of Aargau and Thurgau, 25 small hydropower plants were forced to suspend operations for several weeks (BAFU et al. (Hrsg.), 2019). In August, the army was mobilized to transport water to approximately 91 alpine pastures where water for livestock was no longer available, with helicopters deployed to supply even the most remote areas (BAFU et al. (Hrsg.), 2019). Fish kills were recorded in several rivers due to both low water levels and elevated water temperatures, which reduce dissolved oxygen and create additional stress, particularly for cold water species such as trout (BAFU et al. (Hrsg.), 2019).

As reported by BAFU (2022), the most recent drought in the summer of 2022 once again presented the challenges and issues observed in 2003 and 2018. This sequence of events clearly demonstrates that periods of low water flow in Switzerland are not solely a hydrological deficit, but can affect a wide variety of activities and sectors, with significant repercussions for both humans and the environment.

## 1.2 Future expected changes

In the future, extreme droughts similar to those experienced in 2003 and 2018 are expected to become more frequent and severe. According to the climate scenario without climate change mitigation (RCP8.5), by the end of the century, in addition to rising temperatures, Switzerland is projected to experience a shift in the distribution of precipitation throughout the year, with a trend toward wetter winters and drier summers (CH2018, 2018; Riahi et al., 2011). These changes in precipitation and temperature will also have an important impact on river flow regimes. Higher winter precipitation combined with the rising of snowfall limit will lead to an average increase in river flows during winter months up to 30%, especially in snow-dominated catchments, while in summer, due to reduced snowmelt, lower rainfall, and increased evaporation, average flows are expected to decrease by up to 40%, with rivers in the Swiss plateau, Jura, and Pre-Alps regions experiencing the most significant reductions (Federal Office for the Environment (FOEN), 2021).

Beyond its impact on river flow regimes, climate change is projected to significantly alter future water demand across multiple sectors. Under the climate scenario without climate change mitigation, by the end of the century, water demand for irrigation in Switzerland could increase by up to 40% (HAFL, 2023). Parallel to agricultural needs, the rising snowfall limit will likely necessitate expanded artificial snowmaking to maintain the viability of the ski tourism sector. For instance, Vorkauf et al. (2024) estimated that by 2100, water demand for snow production at the Andermatt-Sedrun-Disentis ski resort could increase by 79% during an average winter.

Furthermore, future population growth could also lead to an increase in the demand for drinking water. Although approximately 80% of Switzerland's drinking water is sourced from groundwater and springs (SVGW, 2020), and therefore is not directly withdrawn from rivers, the extraction of groundwater can place additional pressure on the river discharge, especially during periods of drought. Specifically in catchments without glaciers or persistent snowpack, river flow during drought consists almost entirely of baseflow (Kaule & Gilfedder, 2021). Groundwater pumping can therefore reduce the baseflow recharge to the river, thus further decreasing river discharge (Liu et al., 2020; Navas et al., 2025).

The summer period is therefore expected to become even more critical in the future for water resources management, as the projected decline in water availability coincides with the period of highest water demand (Brunner et al., 2019a). Sectors such as agriculture and industry, which draw directly from surface waters for their activities, may face increasing challenges in meeting their water requirements, especially as competition for water resources intensifies during prolonged drought (Lanz et al., 2021).

## 1.3 Streamflow forecasts

River discharge is strongly dependent on meteorological variables such as precipitation and temperature. Nowadays, weather forecasts provide reliable information up to a time horizon of 1–2 weeks (Zhang et al., 2019). Beyond this period, forecasting skill decreases significantly due to the high sensitivity to initial conditions and the chaotic nature of the atmosphere (Lorenz, 1969). To account for this uncertainty, Ensemble Prediction Systems (EPS) are employed, in which initial conditions are perturbed to simulate multiple scenarios (Molteni et al., 1996). This approach provides probabilistic information, extending the forecasting horizon into the sub-seasonal range (Domeisen et al., 2022).

The integration of ensemble Numerical Weather Prediction (NWP) for streamflow forecasting holds great potential for enhancing water resources management (Bogner et al., 2018; Monhart et al., 2019). Such hydrometeorological forecasting systems use variables derived from weather predictions, primarily temperature and precipitation, as inputs for hydrological models to simulate and predict future river discharge (Smiatek et al., 2012). As with weather forecasts, streamflow predictions extending into the sub-seasonal range, rely on ensembles to account for uncertainty and provide probabilistic information on the future state of streamflow (Troin et al., 2021).

Several studies conducted on Swiss rivers highlight the benefits of using NWP-driven hydrological models for streamflow forecasting. Monhart et al. (2019) evaluated the performance of streamflow predictions across three small-to-medium-sized alpine catchments in Switzerland. They compared the traditional Ensemble Streamflow Prediction (ESP), which relies on known initial conditions forced by historical climate sequences, with forecasts generated by coupling NWP with hydrological models. Their results demonstrated that the NWP-based approach outperforms the ESP in all three catchments and across most seasons, particularly for short lead times of up to approximately five days. On the other hand, Bogner et al. (2018) assessed the application of NWP-based extended-range streamflow forecasts across Switzerland to predict variables relevant for agricultural management and hydropower production, including soil moisture and total runoff. Their study demonstrated that the spatial and temporal aggregation of hydrologically relevant variables predicted through NWP-based forecasts can increase forecast skill over climatology for lead times up to 32 days, emphasizing how this forecasting approach can lead to significant benefits for water resource management in both the agricultural sector and hydroelectric power production. Still analyzing NWP-based streamflow forecasting for Swiss rivers, Bogner et al. (2022) showed that the pre- and post-processing of meteorological and streamflow data, combined with the aggregation of the weekly discharge forecast from continuous to terciles classes, can enhance the hydrological forecast skill up to 4 weeks.

## 1.4 Research motivation

As previously highlighted, extended periods of low flow can have a significant impact on both society and the environment. Future projections indicate that Swiss rivers are likely to face an increase in the frequency and severity of drought events. Consequently, the ability to anticipate low flow periods can play a fundamental role both in safeguarding aquatic ecosystems and managing water resources, helping to prevent shortages and mitigate conflicts over water use. The skill of NWP-driven streamflow forecasts in providing information on river discharge several weeks in advance can, therefore, potentially be used to anticipate the onset of low flow periods. Furthermore, integrating water demand data with these forecasts could provide crucial insights into potential water availability for withdrawals in the coming weeks, allowing for early warnings of approaching water scarcity periods.

The goal of this thesis is therefore to evaluate the application of extended NWP-driven streamflow forecasts in Swiss catchments for predicting water scarcity periods. By integrating water demand estimates with streamflow forecasts, this study aims to assess the possibility of providing information on river water availability in the coming weeks, relative to the needs of both society and the environment, and to understand how drought warnings can be provided.

The research questions for this thesis are:

- RQ1: How can water demand data be integrated into streamflow forecasts to provide indications on water abstraction possibilities for the coming weeks?

- RQ2: How can streamflow forecasts be used to issue warnings regarding drought periods?

- RQ3: How far in advance can reliable information regarding approaching water scarcity periods in the studied catchments be provided?

# Chapter 2

# Methods and data

To address the research questions, the thesis was structured following the workflow shown in Figure 2.1. In the first phase, the study catchments were selected, and their respective water demand was estimated. Subsequently, a forecast verification was conducted to assess the skill of streamflow predictions during drought periods, enabling the identification of the most appropriate forecast ensemble configuration. Based on this configuration, a probabilistic decision threshold was tuned for each catchment, which was used to trigger drought warnings. Finally, water demand was integrated into the streamflow forecasts, and drought warnings were tested across two different streamflow scenarios, allowing for the extraction of information regarding water availability. All data processing and analysis were conducted within the RStudio environment.

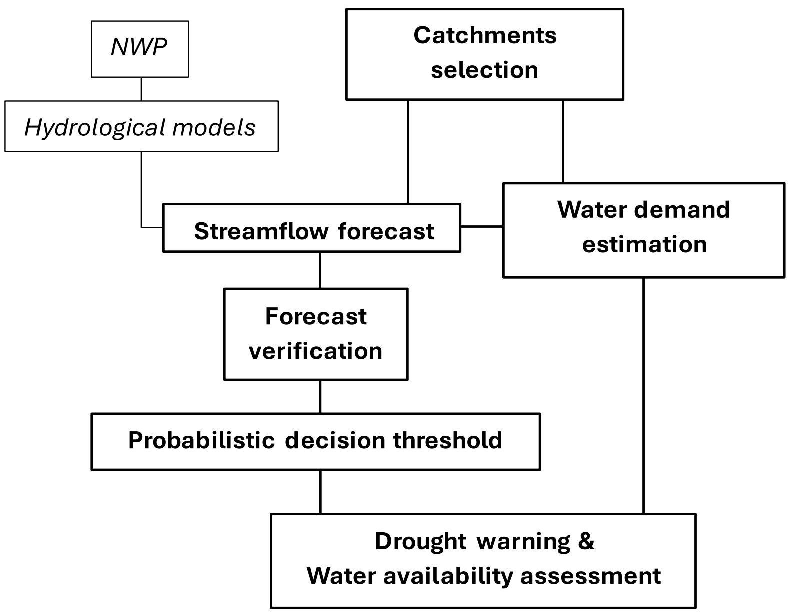


Figure 2.1: *Overview of the research workflow.*

## 2.1 Study area

For this analysis, 20 Swiss meso-scale catchments part of the National Surface Water Quality Monitoring Programme (NAWA), were selected. The initial pool of 87 catchments for which streamflow forecasts were available was screened using the HydCHeck tool, which enables the evaluation of anthropogenic influences on Swiss rivers (Steeb et al., 2024). The selection aimed to completely exclude rivers regulated by artificial reservoirs and hydroelectric power plants, focusing instead on those subject to significant pressures from sectors such as agriculture, industry, and drinking water supply.

As shown in Figure 2.2, the selected catchments are mostly distributed across the Jura, Swiss Plateau, and Pre-Alps regions.

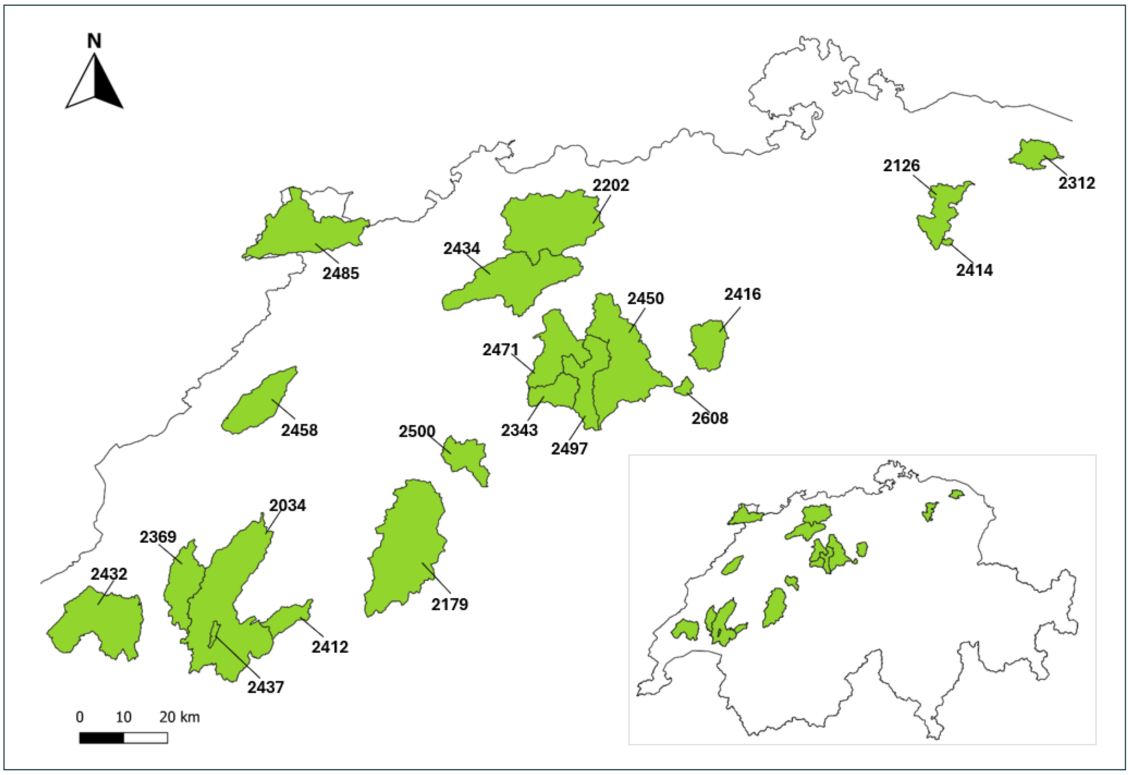


Figure 2.2: *Map with the selected Swiss catchments and their identification code.*

The catchment areas are quite varied, with three particularly small catchments in the range of 3–10 km$^2$, while the others range from 43 to 416 km$^2$, as summarized in Table 2.1. The mean elevation varies from 450 m a.s.l. for the Aach catchment (ID: 2312), located in the Lake Constance region, to 1071 m a.s.l. for the Sense (ID: 2179), a river that for many kilometers marks the boundary between the cantons of Fribourg and Bern.

Table 2.1: *Main information of the selected catchments, including BAFU ID, name, catchment area, and mean elevation.*

| ID | Catchment/River name | Area (km²) | Mean elevation (m a.s.l.) |
|---|---|---|---|
| 2034 | Broye | 416 | 715 |
| 2126 | Murg | 80 | 652 |
| 2179 | Sense | 351 | 1071 |
| 2202 | Ergolz | 261 | 588 |
| 2312 | Aach | 47 | 467 |
| 2343 | Langeten | 60 | 760 |
| 2369 | Mentue | 105 | 675 |
| 2412 | Sionge | 43 | 865 |
| 2414 | Rietholzbach | 3 | 794 |
| 2416 | Aabach | 73 | 581 |
| 2432 | Venoge | 227 | 686 |
| 2434 | Dünnern | 234 | 711 |
| 2437 | Parimbot | 7 | 716 |
| 2450 | Wigger | 366 | 656 |
| 2458 | Seyon | 112 | 978 |
| 2471 | Murg | 183 | 653 |
| 2485 | Allaine | 212 | 562 |
| 2497 | Luthern | 105 | 749 |
| 2500 | Worble | 67 | 666 |
| 2608 | Sellenbodenbach | 10 | 608 |

Figure 2.3 shows the monthly mean specific discharge for the period 1991–2020 for each of the rivers. Climatological data clearly show that most rivers present a pluvial regime, with peak flows occurring in the winter and spring months and minimum values between summer and early fall (Blanc & Schädler, 2014). The Sense (ID: 2179) and, to a lesser extent, the Sellenbodenbach (ID: 2608) are the only rivers exhibiting the influence of snowmelt, which results in a discharge peak during spring and early summer.

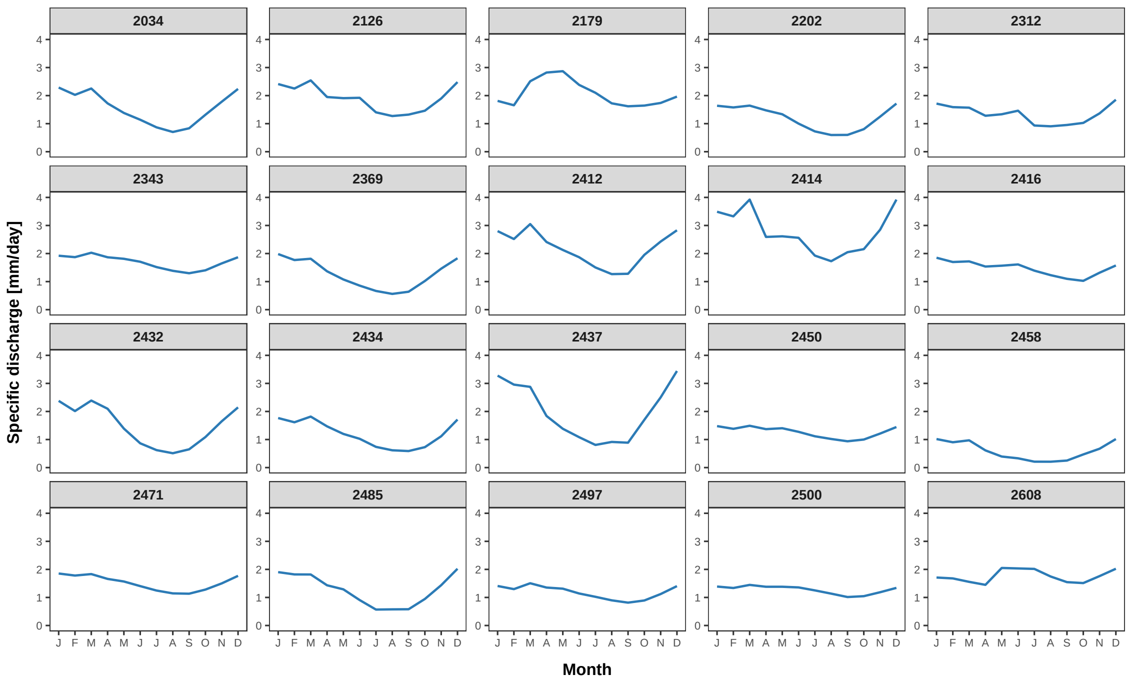


Figure 2.3: *Monthly specific discharge for the period 1991-2020.*

## 2.2 Water demand estimation

The water demand for each catchment was estimated following the methodology and data used by Brunner et al. (2019b) and Steeb et al. (2024), with additional modifications to try to account for seasonal variations in water use. The analysis considered water demand for drinking water, industry, agriculture, artificial snow production, and the ecological water requirements.

### 2.2.1 Drinking water

Drinking water demand was estimated for households and tourism sector, considering the usage rate of 142 liters per person per day (Brunner et al., 2019b; Freiburghaus, 2015). To account for seasonal variations, drinking water use was adjusted according to Eq. (2.1) by Wada et al. (2011), where domestic water use is calculated as a function of monthly temperature as follows:

$$W_{\mathrm{Dom}_m} = \frac{W_{\mathrm{Dom}_a}}{12}\left[\left(\frac{T - T_{\mathrm{avg}}}{T_{\mathrm{max}} - T_{\mathrm{min}}}\right) R_{\mathrm{Dom}} + 1.0\right], \tag{2.1}$$

with $W_{Dom_a}$ being the annual domestic water withdrawal, $T$ and $T_{avg}$ the average temperature of the target month and annual average temperature, $T_{max}$ and $T_{min}$ the average temperatures of the warmest and coldest months, $R_{Dom}$ the relative difference in domestic water demand between the warmest and the coldest month, and $W_{Dom_m}$ the target month resulting water use.

The adjustment was applied separately to each catchment, using the monthly average temperature of the period 1991-2020, calculated from the temperature time series of the CAMELS-CH dataset (Höge et al., 2023). An $R_{Dom}$ of 0.25 was calculated from the monthly supplied drinking water for the region of Bellinzona in Ticino (Azienda Multiservizi Bellinzona (AMB), 2024), and assumed to be representative for all catchments. The adjusted water use per person of each month $W_{Dom_m}$, was then multiplied by the number of inhabitants within the catchment (Federal Statistical Office (FSO), 2022c), and the expected number of tourists in that month (Federal Statistical Office (FSO), 2022b, 2024a), to determine the total monthly drinking water demand.

### 2.2.2 Industry

Following Brunner et al. (2019b), industrial water demand was calculated for the secondary and tertiary sectors. The number of employees per sector (Federal Statistical Office (FSO), 2021), was multiplied by the consumption rate of 148 $m^3$ per employee per year for the secondary sector and 85 $m^3$ per employee per year for the tertiary sector (Freiburghaus, 2009). Industrial water demand was assumed to remain constant throughout the year (Brunner et al., 2019b).

### 2.2.3 Agriculture

Agricultural water demand was estimated for livestock and irrigation. Livestock water demand was calculated by multiplying the number of livestock in each catchment (Federal Statistical Office (FSO), 2022a), by a use rate of 110 l $day^{-1}$ per unit (Brunner et al., 2019b; Freiburghaus, 2009). Livestock water consumption was assumed to remain constant throughout the year (Brunner et al., 2019b).

For the estimation of irrigation water demand, a rate of 377 $m^3$ ha $yr^{-1}$ was adopted, following Steeb et al. (2024). They derived this rate by dividing the 144 million $m^3$ of water estimated by Weber and Schild (2007) as the irrigation requirement for Switzerland in an average dry year by the 382'200 ha of Swiss agricultural land requiring irrigation, resulting from the "Irrigation requirement" raster map by Fuhrer (2010). To find the irrigation water need for each catchment, the rate was multiplied by the catchment's potential area requiring irrigation.

In Switzerland, the irrigation season generally extends from April to September, with a peak around July (Eisenring et al., 2021). To reproduce the seasonal variability of irrigation activity, the total estimated water demand was redistributed daily between April and September following a Gaussian distribution centered in July, while preserving the calculated volume, so that it resembles the trend simulated by Eisenring et al. (2021).

### 2.2.4 Artificial snow

Water demand for artificial snow production, was based on the results directly obtained by Steeb et al. (2024), where they calculated the water volume for artificial snow by multiplying the rate of 1120 $m^3$ ha $yr^{-1}$ by the area of ski slopes below 1950 m a.s.l. present in the catchments. Water demand for snowmaking was assumed to be abstracted from the rivers during the melt season from April to September (Federal Office for the Environment (FOEN), 2021).

### 2.2.5 Ecological water requirement

Water abstraction from rivers in Switzerland is regulated under the Federal Act on the Protection of Waters (Gewässerschutzgesetz, GSchG). By law, the minimum flow required to safeguard the functions of the aquatic ecosystem is determined based on the $Q_{347}$ flow, which represents the flow reached or exceeded on average during 347 days per year, calculated over a 10 year period for rivers free from influences such as withdrawals or water inputs (GSchG). The $Q_{347}$ flow value was therefore used as the ecological flow threshold to be maintained, as done by Brunner et al. (2019b), and was calculated for each catchment as the 95% quantile of daily observed discharges for the period 2010-2019 available from the CAMELS-CH dataset.

### 2.2.6 Total water demand

The total estimated water demand for each catchment was calculated by summing the demands across all categories, excluding the ecological water requirement. Subsequently, the monthly and yearly demands were disaggregated to a daily resolution and normalized by the catchment area to obtain the specific water demand in mm $day^{-1}$.

## 2.3 Streamflow forecast system

The streamflow forecasts used in this thesis are the result of the forecasting system *Hydrologisches Multimodell* ("Hydrologisches Multimodell", n.d.), a product that is part of the SDSC-ETH-WSL MACH-Flow project ("MACH-Flow", n.d.).

The system uses sub-seasonal 51 members ensemble forecasts of precipitation and temperature from the ECMWF Integrated Forecasting System (IFS), with a horizontal resolution of 32 km and a lead time up to 46 days provided by MeteoSwiss. These forecasts are bias-corrected and downscaled (Monhart et al., 2018), and then used as input for hydrological models to predict streamflow over the following 46 days.

The streamflow forecasting system consists of 11 lumped conceptual hydrological models combined with three different snow modules, including CemaNeige, the TUW snow module, and snow. For each model, five different transformations were applied to the streamflow before the calibration to focus on different flow characteristics, thereby ensuring a balanced representation of low, intermediate, and high flows. The models were calibrated using the Kling-Gupta Efficiency (KGE) as the objective function. The calibration and evaluation processes were performed using observed meteorological and discharge data from 1991 to 2023. The first two years were used for warm-up, the period up to 2010 was used for calibration, and the evaluation was carried out from 2011 to 2023 (*). In total, there are thus 55 hydrological members, or 55 different ensemble configurations, each comprising 51 discharge members. Table 2.2 summarizes the 11 hydrological models, as well as the transformations applied to the discharge.

Table 2.2: *Hydrological models and streamflow transformations.*

| | |
|---|---|
| **Models** | CemaNeigeGR4J, CemaNeigeGR5J, CemaNeigeGR6J, CemaNeige_awbm, CemaNeige_bucket, CemaNeige_cmd, CemaNeige_cwi, CemaNeige_sacramento, TUW, TUWsnow_topmodel, snow |
| **Transformations** | power 0.2, power -0.5, log, Box-Cox, none |

(*Details concerning the structure and functioning of the forecasting system were provided confidentially by Dr. Massimiliano Zappa)

## 2.4 Drought definition and forecast

### 2.4.1 Definition

The World Meteorological Organization (WMO) defines socio-economic and ecological drought as the situation in which water needs of the population, economy and environment exceed the available supply (Federal Office for the Environment (FOEN), 2026). Based on this definition, events of water scarcity or drought within the studied catchments were defined. Adapted to the studied case, the available water supply is represented by the natural discharge ($Q_{\mathrm{nat}}$), the water needs of population and economy is the water demand ($W_{\mathrm{dem}}$) and the water need of the environment is represented by the $Q_{347}$ threshold. A socio-economic drought event can therefore be expressed as follows:

$$Q_{\mathrm{nat}} - W_{dem} < Q_{347}. \tag{2.2}$$

Streamflow measuring stations provide for each studied catchment the observed discharge. Since the analysis includes rivers that are potentially subject to significant water withdrawal, the observed discharge should already reflect the subtraction of water demand from the natural discharge. Subtracting water demand from the observed discharge would overestimate observed drought events and intensity, distorting what actually happened, which is represented by the observation itself. Terrier et al. (2021) conducted a review of methods for the naturalization of discharge observation time series, highlighting that the most common method is to adjust the observations by adding or removing the volume of water disturbance. Based on this, for the present analysis, the observed discharge ($Q_{\mathrm{obs}}$) is considered as:

$$Q_{\mathrm{obs}} = Q_{\mathrm{nat}} - W_{\mathrm{dem}}, \tag{2.3}$$

and the definition of water scarcity event of equation (2.2) is therefore modified as follows:

$$Q_{\mathrm{obs}} < Q_{347}, \tag{2.4}$$

assuming that this situation already captures and contains the water demand abstraction influence on streamflow.

### 2.4.2 Drought forecast

Given that the streamflow forecast is provided in the form of an ensemble, the empirical probability of drought occurrence can be calculated as follows:

$$f_d = P(Q_{\mathrm{for},d} < Q_{347}) = \frac{n_{\mathrm{for},d}}{N_{\mathrm{tot},d}}, \tag{2.5}$$

where $f_d$ is the forecasted drought probability on day $d$, $Q_{\mathrm{for},d}$ is the forecasted discharge on day $d$, while $n_{\mathrm{for},d}$ and $N_{\mathrm{tot}}$ are the number of members below the $Q_{347}$ threshold and the total number of members.

In their research, Koch et al. (2018) emphasize that, for rivers heavily influenced by human activities, the naturalization of observations is necessary for the calibration and validation of hydrological models. Since the hydrological models used in this thesis for the streamflow forecast were calibrated using observed discharge, which as assumed before already reflects the remaining flow after an eventual water abstraction, the simulated or forecasted discharge was also assumed to already account for water demand removal. This ensures consistency with the observed measurements and avoids overestimating the predicted events.

## 2.5 Drought forecast verification

### 2.5.1 Brier score

The Brier Score (BS) was used to evaluate the accuracy of the forecast probabilities regarding the occurrence of water scarcity events. This score measures the mean squared error of probabilistic forecasts for a dichotomous event (Brier, 1950), and is defined as follows:

$$BS = \frac{1}{N} \sum_{d=1}^{N} (f_d - o_d)^2, \tag{2.6}$$

where $N$ is the total number of observations, $f_d$ is the forecast probability for day $d$, and $o_d$ is the observed binary outcome for day $d$ defined as:

$$o_d = \begin{cases} 1 & \text{if } Q_{\text{obs}} < Q_{347}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

#### 2.5.1.1 Brier score decomposition

To gain better understanding on the forecast error, the Brier score decomposition according Murphy (1973) was performed. Murphy (1973) suggested that Brier score can be divided in three components, reliability, resolution and uncertainty as follows:

$$BS = \underbrace{\sum_{k=1}^{K} \frac{n_k}{N} (f_k - o_k)^2}_{\text{Reliability}} - \underbrace{\sum_{k=1}^{K} \frac{n_k}{N} (o_k - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}(1 - \bar{o})}_{\text{Uncertainty}}, \tag{2.8}$$

where $n_k$ denotes the number of forecasts in bin $k$, $f_k$ the mean forecast probability in bin $k$, $o_k$ the observed event frequency in bin $k$, $\bar{o}$ the overall event frequency, and $N$ the total number of observations. The reliability term indicates how the forecast probabilities correspond to observed relative frequency of the event. The resolution term tells the capability of the model to distinguish between situations, while the uncertainty term is independent of the forecast and reflects the inherent uncertainty of the event being predicted (S. J. Mason, 2004). For a clearer assessment of reliability, a reliability diagram was constructed, relating the forecasted probabilities to the observed event frequencies (Hsu & Murphy, 1986).

#### 2.5.1.2 Brier skill score

The Brier Skill Score (BSS) was calculated to observe the level of improvement of the forecast compared to the climatological reference, and is defined by Toth et al.

(2003) and Wilks (1995) as:

$$BSS = 1 - \frac{BS_{\text{for}}}{BS_{\text{clim}}}, \tag{2.9}$$

where $BS_{\text{for}}$ is the Bries score calculated from the forecast, while $BS_{\text{clim}}$ is the Brier score according to the climatology (S. J. Mason, 2004).

### 2.5.2 Relative Operating Characteristic

To test the ability of the forecast to distinguish between event and non-event the relative operating characteristic (ROC) was used (I. Mason, 1982; Wilks, 2011). The ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR) obtained by iterating through all possible thresholds in the range of predicted probabilities (Richardson et al., 2020). For a better understanding of the ROC curve the area under the ROC curve (AUC) was computed and plotted (Richardson et al., 2020). True positive rate (TPR) and the false positive rate (FPR) are defined as follows:

$$\text{TPR} = \frac{TP}{TP + FN} \tag{2.10}$$

$$\text{FPR} = \frac{FP}{FP + TN}, \tag{2.11}$$

where $TP$ and $TN$ denote true positives and true negatives, respectively, and $FP$ and $FN$ denote false positives and false negatives.

### 2.5.3 Ensemble configurations comparison

Since discharge forecasts are provided by 11 different models, each with 5 different transformations, in order to assess whether one ensemble configuration may perform better in predicting drought, forecast verification was carried out for three different ensemble configurations, differing in size. The first configuration considered a single hydrological member ensemble, i.e., a single model transformation combination (1 $\times$ 1 $\times$ 51 discharge members). The second considered a single model ensemble, i.e., a single hydrological model including all five transformations (1 $\times$ 5 $\times$ 51 discharge members) while for the third a multi-model ensemble was considered, including all models and all transformations (11 $\times$ 5 $\times$ 51 discharge members).

To identify the most representative single hydrological ensemble member and single model ensemble across all study areas, the Brier Score was calculated separately for each catchment and ensemble during the summer and autumn months. For each catchment, ensembles were ranked based on their Brier Score performance. Finally, the ensemble with the best average rank across all catchments was selected.

## 2.6 Drought warning

To test the application of the forecasting system as warning tool for water scarcity events, a warning threshold based on probabilities was established for each catchment. The threshold was calculated over a training period and evaluated on a separate testing period. The estimated water demand was incorporated into the discharge forecasts to assess its impact on low flow events and to determine whether the system can provide information regarding the feasibility of water abstraction from the river in the coming weeks.

### 2.6.1 Warning threshold determination

Warning systems based on probability of event occurrence utilize a probability threshold as a decision rule to turn the probability signal into binary event/non-event outcomes, thus determining whether to issue a warning or not (Ambühl, 2010). The choice of the appropriate warning threshold is fundamental to ensure maximum precision and reliability of the warning system (Ambühl, 2010), and usually must take into account the costs of missed events and false alarms. The ROC curve is often used to derive the optimal threshold, for example by selecting the threshold that provides the best trade-off between true positive rate and false positive rate (Fawcett, 2006; Zou et al., 2016). In case of rare events, where the non-event class represents the majority of the observations, the ROC curve can provide misleadingly optimistic view of classifier performance, due to its insensitivity to class imbalance in the false positive rate (Imani et al., 2026; Zou et al., 2016), making it an inadequate method for estimating the optimal threshold.

When dealing with imbalanced datasets, a more appropriate approach for the analysis of classifier performance is the use of the precision-recall curve (Boyd et al., 2013; Imani et al., 2026), where precision and recall are defined as:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{2.12}$$

$$\text{Recall} = TPR = \frac{TP}{TP + FN}. \tag{2.13}$$

The combination of these two measures allows for a better evaluation of the model's ability to identify the rare class, as true negatives, which are the majority class, are not considered (Miao & Zhu, 2022). A metric commonly used in machine learning to evaluate classifier performance on imbalanced datasets is the $F_1$ score (Lipton et al., 2014; Sujon et al., 2025; Yuan et al., 2024; Zou et al., 2016), which combines precision and recall using the harmonic mean, giving them equal weight (Obi, 2023), and is defined as follows:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{2.14}$$

The $F_1$ score can therefore be used as objective metric to find the optimal probability threshold for an imbalanced dataset, testing for each probability threshold from 0 to 1 which one maximizes the score (Bouttier & Marchal, 2024; Hancock et al., 2022; Miao & Zhu, 2022). The optimal threshold based on the $F_1$ score can be determined as follows:

$$P_{\text{opt}} = \arg\max_{t \in [0,1]} \left[ 2 \cdot \frac{Precision(t)\, Recall(t)}{Precision(t) + Recall(t)} \right]. \tag{2.15}$$

Where $t$ are the tested thresholds from 0 to 1, and $P_{\text{opt}}$ is the optimal probability threshold.

Since forecasts are issued twice a week, a complete set of lead times is not available for every target day. To simplify data management and analysis, a single probability value per target day was computed by averaging the probabilities of all forecasts with a lead time falling within the same week (Table 2.3), thereby simulating a lagged ensemble in which forecasts issued on different dates for the same target day are combined (Vitart & Takaya, 2021). This procedure results in seven probability values for each target day, corresponding to the seven weekly lead time, thereby ensuring a continuous time series for each forecast horizon.

Table 2.3: *Weekly lead time aggregation for the same target day.*

| **Lead time (week(s))** | **Lead time (days)** |
|---|---|
| 1 | 1–7 |
| 2 | 8–14 |
| 3 | 15–21 |
| 4 | 22–28 |
| 5 | 29–35 |
| 6 | 36–42 |
| 7 | 43–46 |

For each catchment, the optimal threshold $P_{opt}$ was determined by selecting the threshold $t$ (tested at 0.01 step increments), that maximizes the $F_1$ score (2.15). The threshold was calculated based solely on the probabilities predicted for lead time week 1, on the assumption that shorter lead times better reflect reality and therefore yield a more representative threshold. To verify the robustness of the threshold and issued probabilities, the threshold was determined on a training period (2019–2022) and subsequently evaluated on an independent test dataset (2023–2025), across all weekly lead times to verify how far in time drought signal can be identified.

A drought warning is therefore issued when the probability $f_d$ of the forecasted discharge being below the physical $Q_{347}$ threshold is equal or higher than $P_{opt}$, as follow:

$$\text{Drought warning} = \begin{cases} 1 & \text{if } f_d \geq P_{\text{opt}} \\ 0 & \text{otherwise} \end{cases}. \tag{2.16}$$

### 2.6.2 Integration of water demand

To understand the influence of water abstraction from the river on low flows and to assess whether the total water demand in the catchments can be met, the estimated water demand was added to the observed discharge $Q_{\text{obs}}$ and the forecasted discharge $Q_{\text{for}}$ in order to reproduce a natural flow scenario for both observations and forecasts. It is therefore assumed that the total water demand of each catchment is directly withdrawn from the river. The two scenarios of natural observed and forecasted discharge are described as follows:

$$\text{Observed natural scenario:} \quad Q_{\text{nat},d} = Q_{\text{obs},d} + W_{\text{dem},d}, \tag{2.17}$$

$$\text{Forecasted natural scenario:} \quad Q_{\text{for,nat},d} = Q_{\text{for},d} + W_{\text{dem},d}, \tag{2.18}$$

As regards the forecasted natural scenario, the water demand was added to each ensemble member, and the new probabilities $f_{\text{nat},d}$ of being below the $Q_{347}$ threshold were calculated as follows:

$$f_{\text{nat},d} = p(Q_{\text{for},d} + W_{\text{dem},d} < Q_{347}) = \frac{n_{\text{for,nat},d}}{N_{\text{tot}}}, \tag{2.19}$$

The forcasted probability $f_{\text{nat},d}$ was then compared with the operational warning threshold $P_{\text{opt}}$ to verify the drought warnings issued under the natural flow scenario.

$$\text{Natural scenario drought warning} = \begin{cases} 1 & \text{if } f_{nat,d} \geq P_{\text{opt}} \\ 0 & \text{otherwise} \end{cases}. \tag{2.20}$$

### 2.6.3 Drought warnings from a physical perspective

With the aim of better understanding and visualizing how the warning system operates, the warning issuance rule:

$$P(Q_{\text{for},d} < Q_{347}) \geq P_{\text{opt}}, \tag{2.21}$$

can be transformed from the probabilistic space to the physical space of the ensemble. Bouttier and Marchal (2024), in their study used the $F_2$ score to determine the optimal probabilistic threshold for classifying events of heavy precipitation exceedance. With the aim of making forecast communication easier for an end user, instead of producing probability exceedance maps of precipitation, they summarized the ensemble prediction by creating maps showing the rainfall intensity quantile associated with the probability $P_{\text{opt}}$. Based on this approach, the drought warning issuance rule can be elaborated as follows:

$$\begin{aligned} P\big(Q_{\text{for},d} < Q_{347}\big) &\geq P_{\text{opt}} \\ \Rightarrow P\big(Q_{\text{for},d} \leq Q_{347}\big) &\geq P_{\text{opt}} \\ \Rightarrow F_{Q_{\text{for},d}}(Q_{347}) &\geq P_{\text{opt}} \\ \Rightarrow Q_{347} &\geq F^{-1}_{Q_{\text{for},d}}(P_{\text{opt}}) \\ \Rightarrow Q_{347} &\geq Q_{P_{\text{opt}},d}, \end{aligned} \tag{2.22}$$

where $F_{Q_{\text{for},d}}(.)$ is the cumulative distribution function of the forcasted discharge $Q_{\text{for},d}$ variable, while $F^{-1}_{Q_{\text{for},d}}(.)$ is the quantile function coverting probability $P_{\text{opt}}$ into the respective discharge threshold $Q_{P_{\text{opt}},d}$. On each day $d$, the discharge $Q_{P_{\text{opt}},d}$ is extracted from the forecast ensemble, consisting of 2805 members, by identifying it as the value of the member with rank $r = \lceil P_{\text{opt}} \times 2805 \rceil$ within the empirical cumulative distribution, and given the discrete ensemble distribution, it is identified as the first member whose empirical cumulative probability equals or exceeds the

$P_{opt}$ threshold. $Q_{P_{\text{opt}},d}$ therefore represents the ensemble member having value such that there is a probability $P_{\text{opt}}$ of non-exceedance, and can be interpreted as the decision-making discharge, extracted daily from the ensemble, that when compared with the $Q_{347}$ threshold, determines whether the drought warning is activated or not.

For each day, the forecasted distance between $Q_{P_{\text{opt}},d}$ and the $Q_{347}$ threshold, was calculated. These values were then averaged across all available lead times within the same weekly horizon (see Table 2.3).

$$\textit{Forecasted Risk-Based Distance}_d = Q_{P_{\text{opt}},d} - Q_{347} \tag{2.23}$$

# Chapter 3

# Results

## 3.1 Water demand visualization

The catchments under study exhibit varying magnitudes of water demand when normalized by their respective surface areas (Figure 3.1), with catchment with ID 2500 showing the highest overall demand. In line with the selection criteria, all studied catchments show water demand from multiple sectors. Irrigation, livestock, and drinking water account for the largest share of demand in most catchments. In contrast, the water volumes associated with tourism and artificial snow production are minimal, with the latter present in only two catchments. As sought, a clear seasonality emerges, with water demand peaking during the summer months due to the intensification of irrigation activities and higher drinking water use.

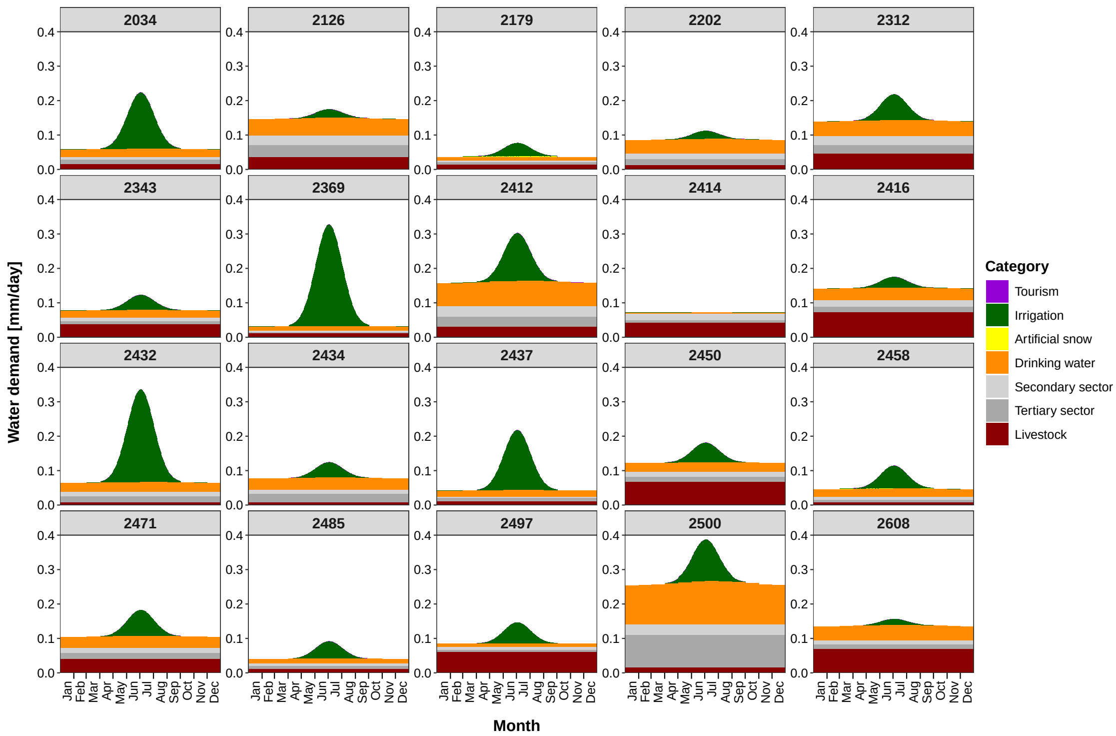


Figure 3.1: *Daily specific water demand by catchment throughout the year.*

## 3.2 Forecast accuracy

Three different ensemble configurations were evaluated to assess their impact on the accuracy of drought forecasts. Among all the studied catchments the best single hydrological member was found to be the log-transformed CemaNeige GR6J, which achieved the lowest mean rank (16 out of 55), while the best model ensemble corresponds to the CemaNeige GR6J with an average rank of 4.7 out of 11. In addition the so-called multimodel ensemble was included in the comparison. Table 3.1 summarizes the characteristics of the ensembles used for the comparison.

Table 3.1: *Summary of ensemble configurations.*

| **Ensemble** | **Model** | **Transformations** | **Hydrological members** | **Total members** |
|---|---|---|---|---|
| Single hydrological member | CemaNeige GR6J | log | 1 | 51 |
| Single model | CemaNeige GR6J | none, log, 0.2, -0.5, Box-cox | 5 | 255 |
| Multimodel | All 11 models | none, log, 0.2, -0.5, Box-cox | 55 | 2805 |

### 3.2.1 Brier Skill Score

Figure 3.2 compares the median Brier Skill Score (BSS) obtained by the three ensemble configurations across all catchments for the summer and autumn seasons of the 2019–2025 period, plotted as a function of the lead time. The Brier Skill Score indicates that forecast skill decreases with increasing lead time, particularly after 14-21 days, when the median skill converges to that of climatology and even drops below zero in autumn for all ensemble configurations. Within this forecast horizon the multimodel ensemble configuration yielded the highest overall median skill in both seasons. For some individual catchments, however, the forecasts exhibit negative skill even at short lead times, as shown by the shaded area, representing the range of BSS across all catchments. Specifically, at a one-day lead time, the multimodel ensemble resulted in negative BSS for 4 catchments in summer and 5 in autumn. In comparison, these numbers increase for the other configurations, with the single hydrological member that exhibited negative skill in 6 catchments in summer and 8 in autumn, while the single model ensemble reached 7 and 8, respectively. For some of these catchments, the BSS was found to be negative in both seasons.

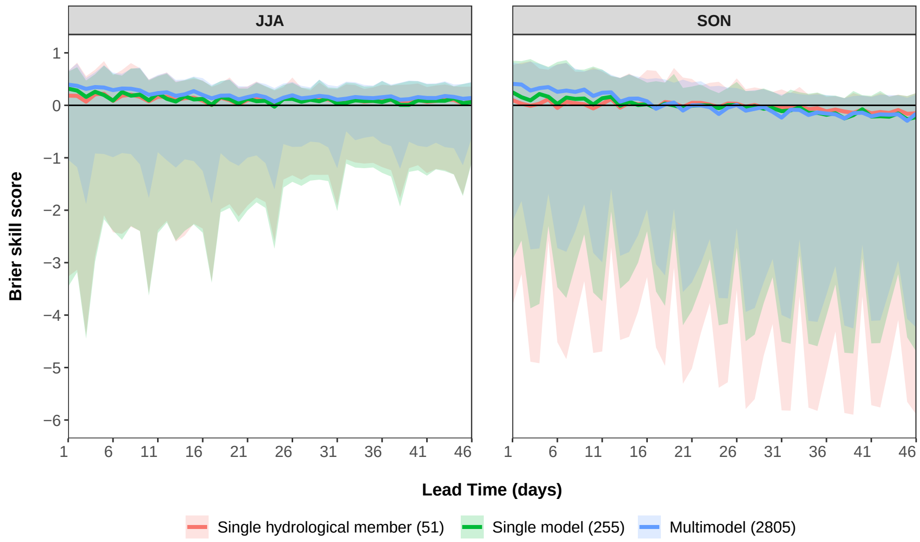


Figure 3.2: *Comparison of Brier Skill Score (BSS) across ensemble configurations for summer and autumn of the period 2019-2025. The plots illustrate the evolution of forecast skill as a function of lead time. Solid lines represent the median BSS calculated across all catchments for each ensemble configuration, while the shaded areas indicate the full range of BSS values encountered across the catchments.*

The Brier Skill Score for winter and spring was not calculated, as the climatology Brier Score for all catchments was effectively zero due to the near absence of events during this seasons. The Brier score of the forecasts for all three ensemble configurations and all catchments in winter and spring was found to range from 0 to 0.04 across all lead times, thus indicating a highly accurate probabilistic forecast.

### 3.2.2 Reliability, resolution and uncertainty

The Brier Score was decomposed into reliability, resolution, and uncertainty terms. Figure 3.3 illustrates the reliability values for the three configurations with daily lead times aggregated into weekly blocks. As can be seen, the multimodel ensemble exhibits lower values, indicating superior reliability compared to the other two ensembles across all lead times in summer, and especially during the first two weeks of lead time in autumn.

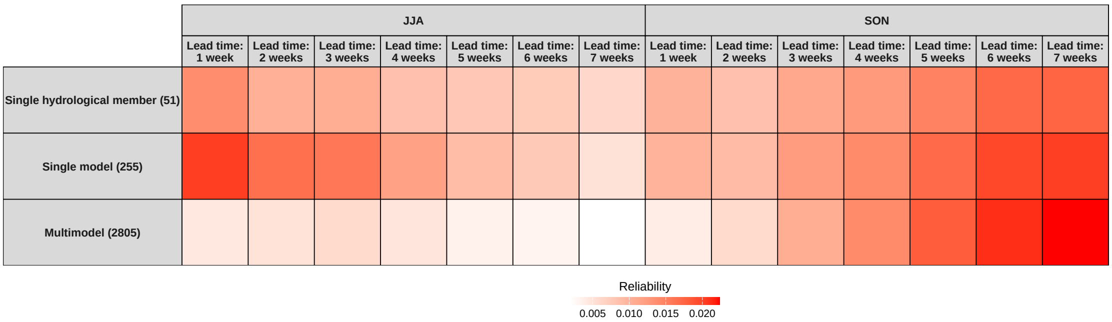


Figure 3.3: *Reliability term heat map per ensemble configuration and lead time. To assess the reliability of the ensemble configurations, probabilities and observation pairs from all 20 catchments were aggregated into common bins. Daily lead times are grouped into weekly blocks.*

The reliability diagram in Figure 3.4 shows that for forecasts up to a lead time of two weeks, all ensembles provide better calibrated probabilities. However, beyond this horizon, they gradually lose their calibration, especially for the higher probability classes. Particularly for forecasts with a lead time of up to one week, the probabilities issued by the multimodel configuration are the closest to the diagonal, indicating superior calibration compared to the other ensembles. In summer, probabilities up to 50% issued with lead times of 5–7 weeks appear better calibrated than those issued with shorter lead times, as they lie closer to the diagonal. This better alignment may explain the apparent improvement in reliability at longer lead times observed in Figure 3.3 during summer.

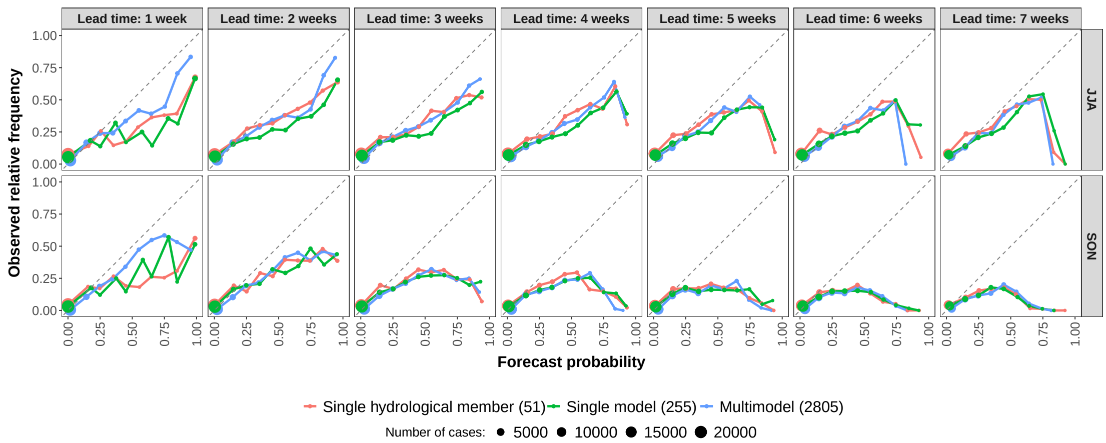


Figure 3.4: *Reliability diagram for the three ensemble configurations across different lead times. Each data point represents the aggregated performance across all 20 catchments, with predictions grouped into common probability bins. The diagonal dashed line indicates perfect calibration, while the point size shows the number of cases within each bin.*

The resolution term was computed individually for each catchment, and the median value for each ensemble configuration is presented in Figure 3.5. In this case, both the multimodel and single model ensemble configurations exhibit a similar pattern, with higher and therefore better resolution at shorter lead times compared to the single hydrological member ensemble. In general can be observed that for all configurations the resolution appears better in summer compared to autumn.


Figure 3.5: *Heat map of the resolution term for each ensemble configuration and lead time, for summer and autumn. Values represent the median resolution calculated across all 20 catchments.*

The uncertainty term of the Brier Score decomposition, while independent of model performance, shows that summer is the most challenging season for drought prediction, with a mean uncertainty value across catchments of 0.14. In autumn, the average uncertainty was 0.05, while in winter and spring it was found to be near zero.

### 3.2.3 ROC-AUC

Figure 3.6 shows the area under the ROC curve (AUC) values obtained from the three ensemble configurations as a function of the lead time for both summer and autumn seasons. As can be seen, the multimodel ensemble configuration provides a higher capability in distinguishing between classes of drought event and non-event, especially for a lead time up to 4 weeks in both seasons, with AUC values above 0.8. Beyond this time horizon, the median AUC values across all catchments for all ensemble configurations tend to converge around a value of 0.75. The shaded area indicates that the use of the multimodel ensemble maintained AUC values above 0.5 across all catchments and in both seasons, a threshold corresponding to a random classifier. It is important to acknowledge that such high AUC values may be the result of the large proportion of non-event occurrences in the dataset even though the analysis is restricted to the seasons most prone to drought. However, this does not diminish the higher ability of the multimodel ensemble in assign higher probabilities to drought events compared to non-events.

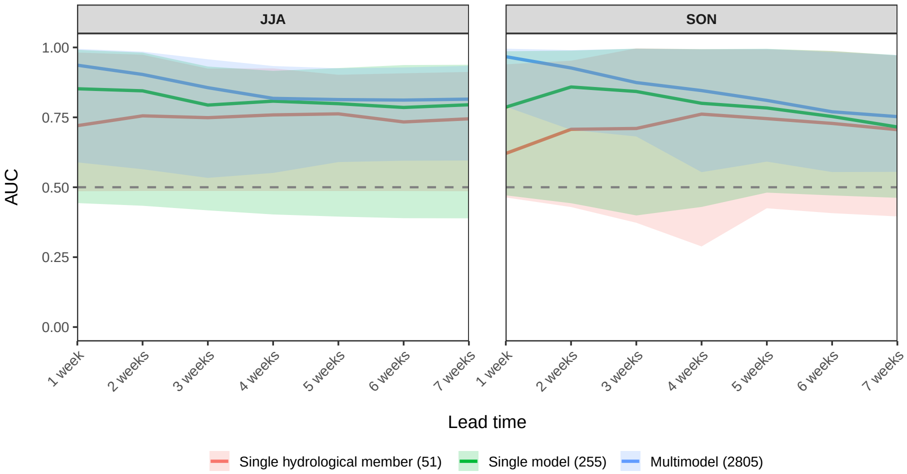


Figure 3.6: *Area under the ROC curve (AUC) across all catchments for the three ensemble configurations and different lead times for the period 2019-2025. The line represents the median across all catchments, while the shaded areas show the range of AUC values.*

## 3.3 Drought warning system

Based on the results obtained from the verification of the accuracy of drought event forecasts, the multimodel ensemble configuration was applied to all catchments, in order to test the issuance of drought warnings.

### 3.3.1 Threshold tuning

The optimal probability $P_{\text{opt}}$, was tuned over the training period 2019-2022 based on drought probabilities issued from the forecast with a lead time of one week. To verify its robustness and consistency, the threshold was then applied to the test period 2023–2025.

Table 3.2 reports, for each catchment, the $P_{\text{opt}}$ thresholds derived from the training period, with their associated performance metrics. Additionally, it presents the results obtained by applying these thresholds to forecasts with a one week lead time issued during the test period 2023–2025. As can be seen from the table, the $P_{\text{opt}}$ threshold varies significantly across catchments. At the extremes, there are catchment with ID 2471 showing a very low threshold of just 0.05, and catchment 2500, where the threshold that maximized the $F_1$ score during the training period was 0.95. For 14 out of the 20 catchments, the optimal threshold was found to be below 0.5.

In the training period, comparing the observed drought days (Obs.) with the correctly predicted ones, i.e., the true positives (TP), can be seen that for most catchments, many of the observed events were successfully classified as drought events and therefore predicted. The recall metric (Rec.) quantifies exactly the proportion of observed drought events that were correctly forecasted. Catchments 2500 and 2608 show the lowest recall values, with only 35% and 42% of drought days correctly identified, respectively, resulting in a high number of missed events. All other catchments show a recall above 50%, with catchment 2432 having 98% of drought days

Table 3.2: *Performance metrics for the training and test periods for forecasts issued with a lead time of one week. $P_{opt}$ denotes the optimal probability, (Obs.) indicates days of observed drought, (TP) and (FP) represent the number of true positives and false positives, (Prec.) is the precision metric, (Rec.) is the recall metric, and $F_1$ refers to the $F_1$ score.*

| | **Training period 2019-2022** | | | | | | | **Test period 2023-2025** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Catchment (ID)* | $P_{opt}$ | *Obs.* | *TP* | *FP* | *Prec.* | *Rec.* | $F_1$ | *Obs.* | *TP* | *FP* | *Prec.* | *Rec.* | *F1* |
| 2034 | 0.40 | 96 | 68 | 27 | 0.72 | 0.71 | 0.71 | 35 | 22 | 29 | 0.43 | 0.63 | 0.51 |
| 2126 | 0.49 | 34 | 25 | 20 | 0.56 | 0.74 | 0.63 | 0 | 0 | 24 | 0.00 | 0.00 | 0.00 |
| 2179 | 0.78 | 53 | 47 | 19 | 0.71 | 0.89 | 0.79 | 25 | 10 | 1 | 0.91 | 0.40 | 0.56 |
| 2202 | 0.27 | 82 | 62 | 45 | 0.58 | 0.76 | 0.66 | 32 | 25 | 13 | 0.66 | 0.78 | 0.71 |
| 2312 | 0.68 | 67 | 39 | 40 | 0.49 | 0.58 | 0.53 | 50 | 25 | 17 | 0.60 | 0.50 | 0.54 |
| 2343 | 0.14 | 159 | 122 | 57 | 0.68 | 0.77 | 0.72 | 84 | 71 | 38 | 0.65 | 0.85 | 0.74 |
| 2369 | 0.13 | 23 | 17 | 107 | 0.14 | 0.74 | 0.23 | 57 | 0 | 70 | 0.00 | 0.00 | 0.00 |
| 2412 | 0.70 | 97 | 50 | 22 | 0.69 | 0.52 | 0.59 | 44 | 19 | 7 | 0.73 | 0.43 | 0.54 |
| 2414 | 0.27 | 81 | 59 | 28 | 0.68 | 0.73 | 0.70 | 52 | 38 | 43 | 0.47 | 0.73 | 0.57 |
| 2416 | 0.41 | 47 | 42 | 12 | 0.78 | 0.89 | 0.83 | 26 | 18 | 39 | 0.32 | 0.69 | 0.43 |
| 2432 | 0.60 | 52 | 51 | 20 | 0.72 | 0.98 | 0.83 | 72 | 46 | 4 | 0.92 | 0.64 | 0.75 |
| 2434 | 0.39 | 36 | 28 | 18 | 0.61 | 0.78 | 0.68 | 18 | 17 | 3 | 0.85 | 0.94 | 0.89 |
| 2437 | 0.80 | 66 | 51 | 10 | 0.84 | 0.77 | 0.80 | 32 | 18 | 0 | 1.00 | 0.56 | 0.72 |
| 2450 | 0.16 | 101 | 67 | 35 | 0.66 | 0.66 | 0.66 | 68 | 52 | 22 | 0.70 | 0.76 | 0.73 |
| 2458 | 0.10 | 63 | 45 | 41 | 0.52 | 0.82 | 0.64 | 38 | 30 | 35 | 0.46 | 0.83 | 0.59 |
| 2471 | 0.05 | 148 | 119 | 107 | 0.53 | 0.80 | 0.64 | 73 | 60 | 80 | 0.43 | 0.82 | 0.56 |
| 2485 | 0.26 | 103 | 89 | 45 | 0.66 | 0.86 | 0.75 | 78 | 65 | 18 | 0.78 | 0.83 | 0.81 |
| 2497 | 0.18 | 69 | 50 | 52 | 0.49 | 0.72 | 0.58 | 41 | 32 | 46 | 0.41 | 0.78 | 0.54 |
| 2500 | 0.95 | 40 | 14 | 7 | 0.67 | 0.35 | 0.46 | 1 | 1 | 16 | 0.06 | 1.00 | 0.11 |
| 2608 | 0.19 | 60 | 23 | 14 | 0.62 | 0.42 | 0.50 | 47 | 21 | 16 | 0.57 | 0.50 | 0.53 |

correctly predicted. Precision (Prec.), on the other hand, indicates the proportion of forecasted events that were correctly identified as droughts, thereby accounting for false positives (FP) and therefore drought false alarms. Catchment 2369 exhibits the lowest precision across all catchments. In fact, while 17 out of 23 observed drought days were correctly predicted, an additional 107 false alarms were issued over the four years of the training period, resulting in a precision of only 14%. For all the other catchments, precision is significantly higher, ranging from 48% to 84%, indicating a much greater reliability of the issued warnings.

The $F_1$ score can reach a maximum value of 1 when both precision and recall are equal to 1. Since it represents the harmonic mean of precision and recall, the $F_1$ score is significantly affected if either of these metrics is very low. During the training period, catchment 2369 results in the lowest $F_1$ score, with a value of 0.23, mainly penalized by its low precision, followed by catchment 2500, whose score of 0.46 is instead strongly penalized by the low recall. For all the other catchments, the $F_1$ score during the training period is equal to or greater than 0.5, with catchments 2416 and 2432 achieving a score of 0.83, indicating greater capability in correctly forecasting events and in reducing false alarms, as reflected by their high recall and precision values.

The threshold $P_{\text{opt}}$ was then applied to the test period 2023–2025, again considering only forecasts within a lead time of one week, in order to compare the metrics with those of the training period. From the table, it can be observed that, in general, the number of observed drought days during the test period is much lower across all catchments compared to the training period, which spans one additional year. Comparing the recall between the training and test periods, it can be seen that for 9 catchments the recall improved in the test period, for 4 catchments it remained very similar, and for 7 catchments it decreased. Precision, on the other hand, decreased significantly for 6 catchments, remained similar for 6 catchments, and improved for 8 catchments. Among all catchments, the $F_1$ score was found to be very similar between the training and test periods for 12 of them. Catchment 2434 showed the greatest improvement, with the score increasing from 0.68 to 0.89, while 5 other catchments experienced a pronounced decrease. In catchment 2126, no drought days were observed between the 2023-2025 period, making it impossible to compute recall, precision, and $F_1$ score. However, 24 false alarms were issued over the course of these three years. Similarly, for catchment 2369, which had the lowest $F_1$ score during the training period, the score in the test period is 0. This occurred because none of the 70 issued warnings corresponded to the 57 observed drought days, resulting all in false alarms. Catchment 2500 also exhibits a very low $F_1$ score of 0.11. Although the only observed drought day event was correctly predicted, it is penalized by 16 false positives days.

### 3.3.2 Drought identification

The optimal threshold was applied to forecasts with lead times from two to seven weeks to evaluate the lead time at which a reliable drought signal first emerges. The year 2023 proved to be the most interesting within the test period in terms of both the frequency and intensity of observed droughts compared to 2024 and 2025, therefore, it was analyzed in greater detail.

Figure 3.7 shows, for all catchments, the observed and forecasted drought days in 2023 for both the measured streamflow, therefore already influenced by water abstraction, and for the natural streamflow scenario, reconstructed by adding water demand volumes back to the measured flow. By observing the figure, it can be seen that in 2023, 18 out of the 20 catchments experienced days in which the observed streamflow fell below the $Q_{347}$ threshold (indicated in gray), particularly between late June and late October. Catchments identified by IDs 2343, 2471, and 2485 experienced drought periods lasting nearly nine consecutive weeks. Catchment 2414, on the other hand, experienced two clearly distinct drought periods, one between June and August, and another between October and November. In other catchments, drought periods were shorter and scattered over time, or entirely absent, as observed in catchments 2126 and 2369. Looking at the observed natural streamflow scenario (indicated in black), six catchments would still have experienced periods with streamflow below the $Q_{347}$ threshold, indicating a high severity of flow deficit. Conversely, for 12 other catchments, the absence of water demand abstraction would have completely prevented drought days in 2023.

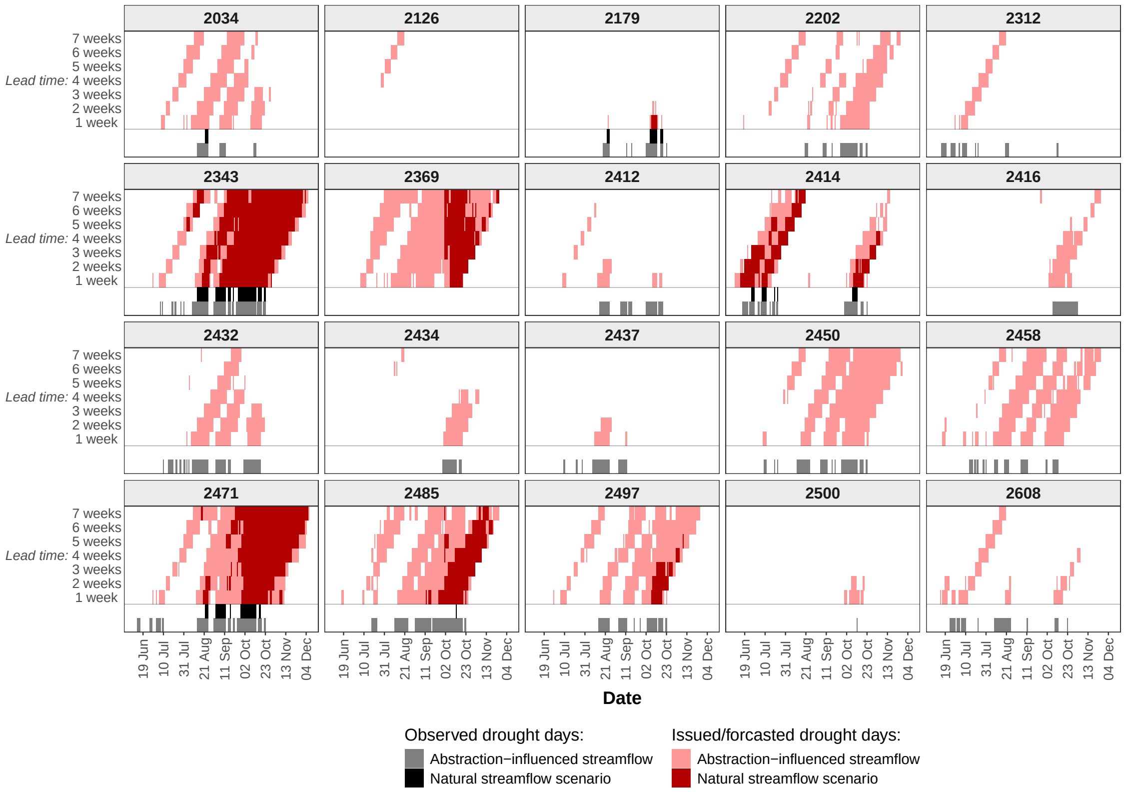


Figure 3.7: *Observed and forecasted/issued drought days in 2023 across the catchments. Observed drought days are shown in gray for measured streamflow and in black for the observed natural streamflow scenario. Forcasted and issued drought days are indicated in light red for the abstraction-influenced streamflow and in dark red for the natural streamflow scenario. Forecasts for the same target period are shown for lead times ranging from seven weeks to one week.*

By analyzing the warnings, it can be observed that in most catchments where drought days were recorded, warnings were also issued. It can also be observed that often, drought warnings were issued for all the weekly lead times. However, for longer lead times these warnings often exhibit a delay in both the onset and the end of the drought compared to observations, thus generating false alarms in the latter case. This delay gradually decreases as the lead time shortens, until the lead times of two and one week, where the forecasted drought days align more closely with the observations in most cases. Especially regarding the onset of drought periods, it can be observed that in almost none of the catchments were warnings accurately forecasted and therefore issued in time, with respect to the observations, beyond a lead time of three weeks. For catchment 2343, the signal indicating the onset of the first prolonged low flow period around 21 August aligns better with the observations already at a lead time of three weeks, as does for the low flow period observed in catchment 2432 between 2 and 23 October. For catchment 2416 as well, warnings show better agreement with the observations only starting from a lead time of two weeks. It can also be noted that very short periods of only a few consecutive days in which streamflow fell below the $Q_{347}$ threshold are not always captured by the forecast, as observed in catchments 2179, 2312, and 2437. Catchment 2369 is the only one in which no days of drought were observed, even though the forecast predicted a significant drought period with warnings issued for all lead times, resulting false alarms.

For 5 of the 6 catchments where the observed natural streamflow scenario recorded days of drought, the forecast also issued drought warnings, with catchments 2343, 2414, and 2471 showing the greatest consistency with the observations. For the catchment 2497, the natural streamflow forecast scenario detected possible warning days even though none were observed. Conversely, for catchment 2034, where the observed natural streamflow scenario showed only a few drought days, the forecast failed to issue any warnings, unlike the similar situation in catchment 2485, where the signal was correctly captured. In catchment 2369 the forecast for the natural streamflow scenario is clearly incorrect. For most of the other catchments, no warnings were issued under the natural streamflow scenario, which is consistent with the observations, thus correctly identifying that, in the absence of water abstractions no drought days would have occurred.

Table 3.3 reports the number of warnings issued in the forecasts for each catchment for each of the seven lead time weeks, specifically for the abstraction-influenced streamflow case. From the table, it can be seen that generally for shorter lead times (1–3 weeks), the number of warnings is closer to the observed days of drought, compared to longer lead times (4-7). Table 3.4 further confirms that for longer lead times, many of the issued warnings are false alarms, primarily caused by the apparent temporal lag relative to the observations. Table 3.5 shows, instead, the warnings for the natural streamflow scenario. It can be noted that by avoiding water abstractions in 12 catchments, the reconstructed natural flow resulted in zero observed drought days and that for 11 of these catchments, the forecast correctly issued no warnings.

Table 3.3: *Issued and observed drought days in 2023, for the abstraction-influenced streamflow.*

| Catchment (ID) | Issued/forecasted drought days per lead time week | | | | | | | Observed drought days |
|---|---|---|---|---|---|---|---|---|
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| 2034 | 31 | 37 | 31 | 40 | 52 | 49 | 51 | 22 |
| 2126 | 7 | 7 | 6 | 3 | 0 | 0 | 0 | 0 |
| 2179 | 0 | 0 | 0 | 0 | 0 | 3 | 11 | 25 |
| 2202 | 31 | 29 | 35 | 49 | 36 | 36 | 38 | 32 |
| 2312 | 7 | 7 | 7 | 10 | 10 | 10 | 9 | 26 |
| 2343 | 108 | 105 | 98 | 97 | 95 | 86 | 84 | 72 |
| 2369 | 113 | 100 | 101 | 94 | 87 | 79 | 70 | 0 |
| 2412 | 0 | 2 | 4 | 4 | 4 | 9 | 25 | 39 |
| 2414 | 25 | 38 | 43 | 52 | 53 | 58 | 58 | 44 |
| 2416 | 9 | 4 | 6 | 19 | 21 | 19 | 22 | 26 |
| 2432 | 12 | 15 | 17 | 28 | 41 | 54 | 50 | 64 |
| 2434 | 3 | 3 | 0 | 13 | 21 | 23 | 20 | 18 |
| 2437 | 0 | 0 | 0 | 0 | 0 | 11 | 18 | 32 |
| 2450 | 78 | 82 | 71 | 67 | 62 | 50 | 53 | 52 |
| 2458 | 58 | 59 | 60 | 64 | 61 | 53 | 65 | 36 |
| 2471 | 117 | 109 | 108 | 102 | 93 | 90 | 94 | 70 |
| 2485 | 76 | 83 | 87 | 80 | 73 | 74 | 83 | 70 |
| 2497 | 61 | 70 | 66 | 65 | 60 | 51 | 55 | 41 |
| 2500 | 0 | 0 | 0 | 0 | 0 | 8 | 17 | 1 |
| 2608 | 7 | 10 | 10 | 14 | 11 | 24 | 37 | 40 |

Table 3.4: *Total false positives/alarms per weekly lead time for the year 2023.*

| Catchment (ID) | False positives per lead time week | | | | | | |
|---|---|---|---|---|---|---|---|
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| 2034 | 23 | 28 | 24 | 33 | 44 | 32 | 29 |
| 2126 | 7 | 7 | 6 | 3 | 0 | 0 | 0 |
| 2179 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2202 | 26 | 29 | 33 | 33 | 26 | 18 | 13 |
| 2312 | 6 | 7 | 7 | 10 | 8 | 8 | 2 |
| 2343 | 53 | 54 | 49 | 46 | 36 | 31 | 21 |
| 2369 | 113 | 100 | 101 | 94 | 87 | 79 | 70 |
| 2412 | 0 | 2 | 4 | 4 | 4 | 2 | 6 |
| 2414 | 25 | 32 | 35 | 36 | 28 | 21 | 23 |
| 2416 | 9 | 4 | 4 | 5 | 0 | 1 | 4 |
| 2432 | 11 | 10 | 4 | 16 | 18 | 18 | 4 |
| 2434 | 3 | 3 | 0 | 11 | 13 | 9 | 3 |
| 2437 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| 2450 | 49 | 53 | 41 | 36 | 33 | 21 | 13 |
| 2458 | 51 | 43 | 46 | 51 | 47 | 39 | 35 |
| 2471 | 64 | 66 | 64 | 56 | 44 | 39 | 35 |
| 2485 | 32 | 41 | 43 | 32 | 22 | 19 | 18 |
| 2497 | 45 | 52 | 48 | 39 | 38 | 23 | 23 |
| 2500 | 0 | 0 | 0 | 0 | 0 | 8 | 16 |
| 2608 | 0 | 2 | 9 | 14 | 10 | 20 | 16 |

Table 3.5: *Issued and observed drought days in 2023, for the natural stramflow scenario.*

| Catchment (ID) | Issued/Forecasted drought days per lead time week | | | | | | | Observed drought days |
|---|---|---|---|---|---|---|---|---|
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| 2034 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 2126 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2179 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 14 |
| 2202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2312 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2343 | 83 | 87 | 80 | 68 | 69 | 61 | 59 | 52 |
| 2369 | 26 | 25 | 35 | 35 | 30 | 19 | 13 | 0 |
| 2412 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2414 | 9 | 10 | 16 | 17 | 31 | 38 | 33 | 17 |
| 2416 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2432 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2434 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2437 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2450 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2458 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2471 | 78 | 72 | 64 | 56 | 47 | 40 | 38 | 34 |
| 2485 | 9 | 18 | 27 | 29 | 28 | 25 | 31 | 1 |
| 2497 | 0 | 0 | 1 | 4 | 14 | 17 | 11 | 0 |
| 2500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2608 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The year 2024 recorded higher streamflows, with 15 rivers experiencing no drought days and for which no warnings were issued. Forecasted warnings were issued for only five rivers, and in most cases, these signals disappeared for forecasts with a lead time of one week. Similarly to 2023, forecasts for the catchment 2369 proved to be the most problematic in 2024. In fact, despite the 56 observed drought days, only seven warning days were issued, none of which occurred for forecast with lead time of one week. In contrast, 2025 turned out to be a slightly drier year than 2024, though not as dry as 2023. Some rivers experienced only a few days of drought, while others showed no observed drought days. From Figure 3.8 it can be observed that the issued warnings for 2025 are more uncertain for some rivers, with false alarms occurring in catchments 2126, 2416, 2471, and 2497.

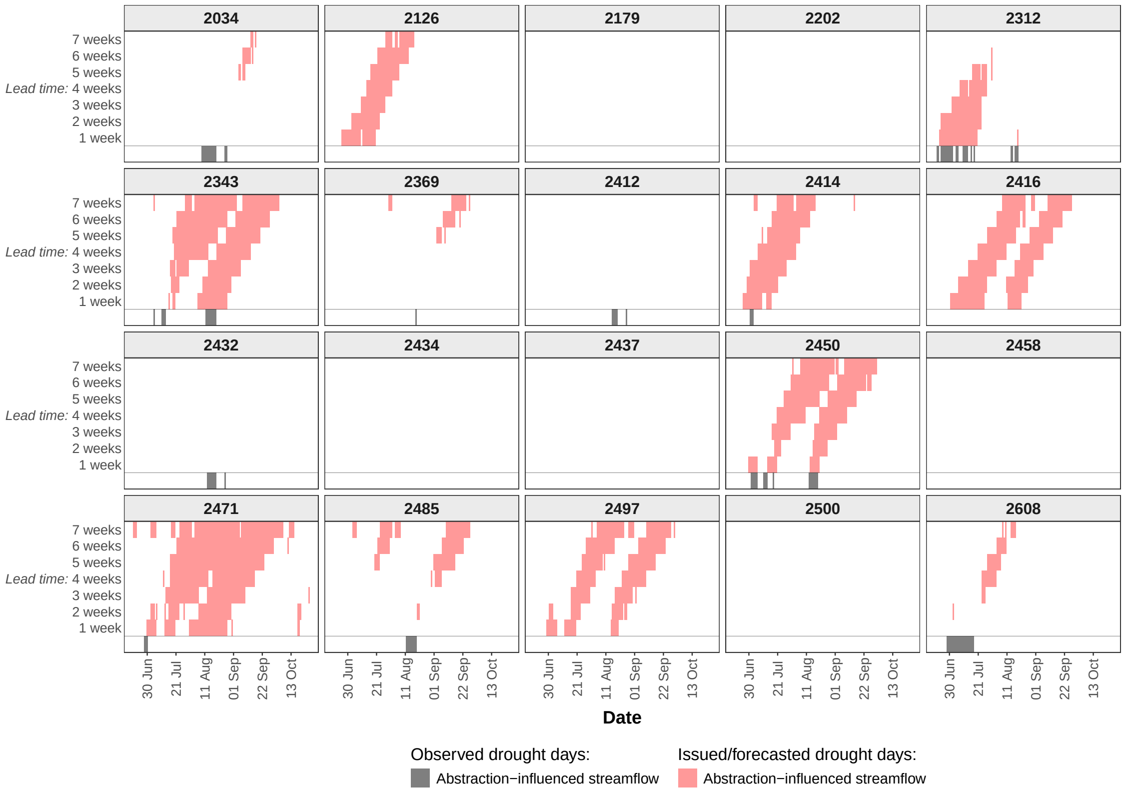


Figure 3.8: *Observed and forecasted/issued drought days in 2025 across the catchments.*

The year 2022 stands out as one of the years most severely affected by drought between 2019 and 2025. Since this year falls within the training period, its analysis is intended solely for visualization and comparison, particularly regarding the forecast with lead time of one week, which was actually used for threshold determination. In Figure 3.9 can be noted that, unlike in 2023, the observed drought periods in 2022 across the catchments are much more compact and less interrupted by weeks with higher streamflow. It can also be noted that for some catchments, such as the 2343 and 2471, the delay in the forecast onset of drought for lead times of 7 to 3 weeks is smaller compared to 2023, however, the delay regarding the end of the drought period remains, although it is less pronounced for some catchments, as can be clearly observed in catchment 2416.

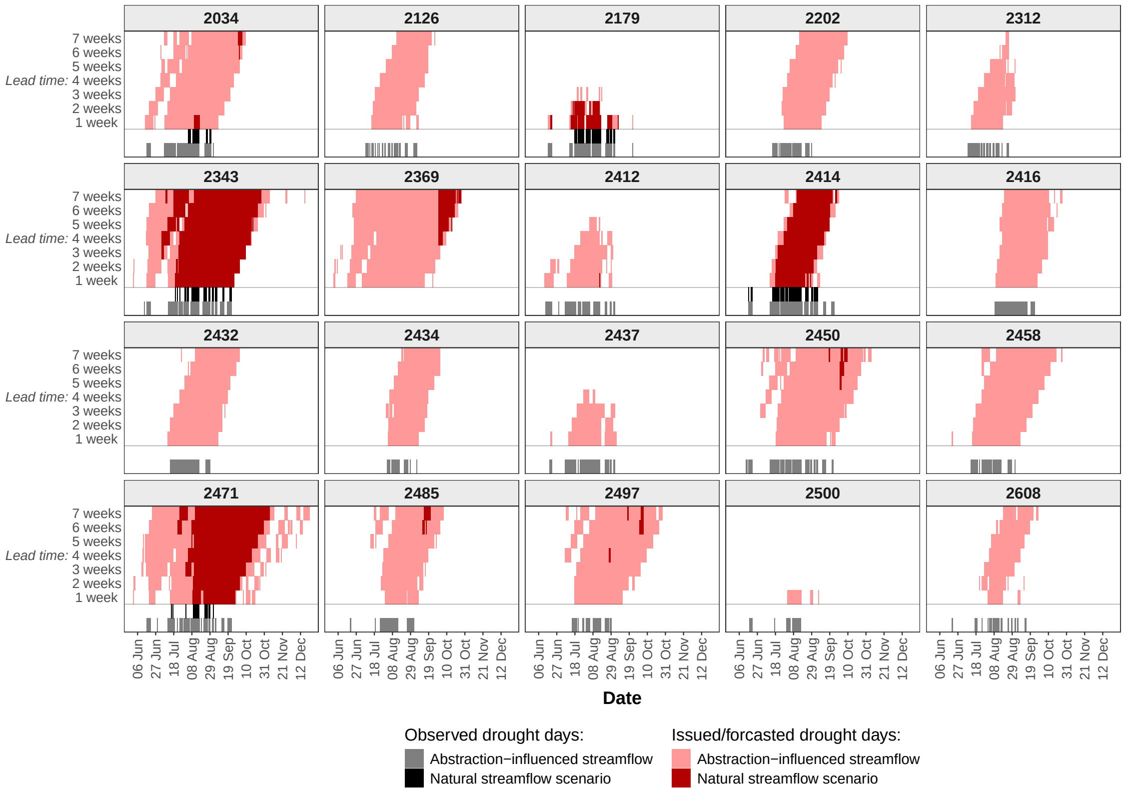


Figure 3.9: *Observed and forecasted/issued drought days in 2022 across the catchments.*

### 3.3.3 Drought warning in the physical ensemble space

Another way to interpret the issuance of the drought warning is to examine, within the ensemble distribution, the position of the member that leads to the optimal non-exceedance probability $P_{\text{opt}}$ and compare the discharge value of this member ($Q_{P_{opt}}$) with the $Q_{347}$ threshold. In this way, instead of providing a binary output (warning or no warning) as shown in the figures in the previous chapter, for each day can be observed the distance of the decision-making discharge $Q_{P_{opt}}$ from the $Q_{347}$ threshold, allowing to visualize how close the warning was to being triggered or how far below the physical threshold the respective member lies.

Figure 3.10 shows the drought period of the Broye River (ID: 2034) in 2023. In the figure the positive blue values, indicate the observed water surplus relative to the ecological flow threshold and can be interpreted as the additional amount of water that could have been abstracted from the river, while negative values in red, indicate observed days of drought and how much the measured discharge fell below the $Q_{347}$ threshold, thus representing the deficit. As a reference for the deficit, the corresponding daily water demand is also shown in green. In this way, it is possible to observe how much of the water demand would have needed to be unmet in order to respect the ecological $Q_{347}$ threshold.

The black line, on the other hand, indicates the forecasted distance between the decision-making discharge value $Q_{P_{opt}}$ and the $Q_{347}$ threshold. Positive values of the line indicate that no warning was issued and can be interpreted as the safety margin, representing the surplus discharge before the risk of drought reaches the $P_{opt}$ threshold, whereas negative values indicate days when a warning was issued, quantifying the extent to which the discharge falls below the $Q_{347}$ threshold according to the chosen $P_{\text{opt}}$ risk level. For the same target date, the forecasted distance is shown for different weekly lead times.

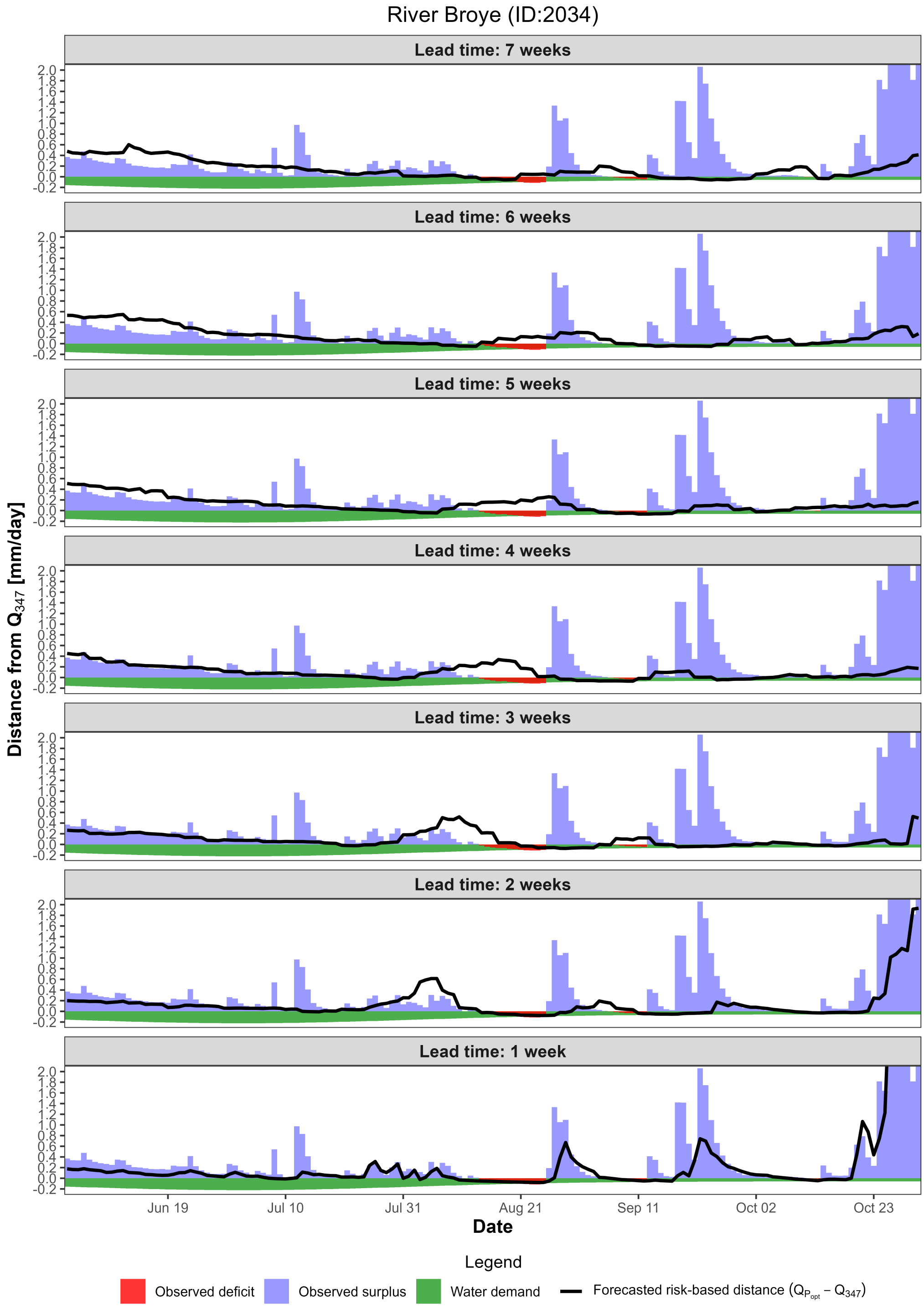


Figure 3.10: *Surplus and deficit of the Broye River (ID: 2034) during the 2023 drought period. Blue and red bars represent the observed water surplus and deficit compared to the $Q_{347}$ threshold, while the green area shows the water demand. The black line denotes the forecasted risk-based distance between the decision-making discharge $Q_{P_{opt}}$ and the $Q_{347}$ threshold. Forecasted distances are displayed for the same target date across different lead times.*

By analyzing the observed surplus and deficit in the Broye River in 2023, it can be observed that until the middle of August the streamflow remained above the threshold. From the middle of August to the end of the month, an initial period of deficit occurred, followed by a marked peak in surplus. In September, a new decrease in discharge occurred, with some days of deficit, followed again by a peak in surplus between the end of September and the beginning of October. In October, the discharge decreased again, but without reaching a deficit, although it remained very close to the threshold, before increasing steadily from the middle of October onward. It thus becomes evident that the Broye River's discharge was highly dynamic during the summer of 2023, with periods of peak flow and resulting surplus followed by periods of deficit.

Examining the predicted distance of the decision-making discharge from the $Q_{347}$ threshold, it can be seen that as the lead time decreases, the line progressively adjusts to the timing of deficit and surplus periods, which are always delayed for longer lead times compared to the observation. It is also observed that for shorter lead times, not only is the timing more accurate, but the predicted distance is also closer to the observed values, both for periods of surplus and deficit. For short lead times, the forecasted distance is indeed much more responsive and sharp, whereas for the lead times longer than three weeks, the line becomes much flatter. These kind of behavior regarding the forecasted distances was observed for most of the rivers in 2023.

#### 3.3.3.1 Examples of severe drought

In Chapter 3.3.2, it was observed that in 2023, in some catchments, even under the natural streamflow scenario, the river would still have experienced days of drought. Two of these catchments were those of the Langeten River (ID: 2343) and the Rietholzbach River (ID: 2414). From Figure 3.11 can be seen that in both catchments, the measured deficit intensity exceeds the water demand on most days, indicating that even if the total water demand had not been extracted from the river, the discharge would still have fallen below the $Q_{347}$ threshold. The forecasted distance, especially regarding deficits, is particularly accurate for both rivers in terms of timing and intensity, already for a lead time of 2 weeks. For the Langeten River, the forecasted $Q_{P_{opt}}$ discharge level relative to the $Q_{347}$ threshold is slightly lower than the observed values, especially from the middle of October. Thus, for these two rivers, the forecast, in addition to correctly indicating the days of drought, would also have correctly signaled that no water abstraction would have been possible during deficit periods, since even under natural flow conditions, the $Q_{347}$ threshold could not have been maintained. It is also interesting to note that, despite covering the same period of interest, the two rivers experienced completely different conditions. The Langeten River had an almost continuous deficit period from early August to the end of October, likely interrupted only by sporadic summer storms, whereas the Rietholzbach River already experienced deficit periods in June and August, which were interrupted in August and later in September by high discharge peaks, followed by a new deficit period in October. Despite these differences, the ensemble was able, at least visually, to adapt well to each individual river.

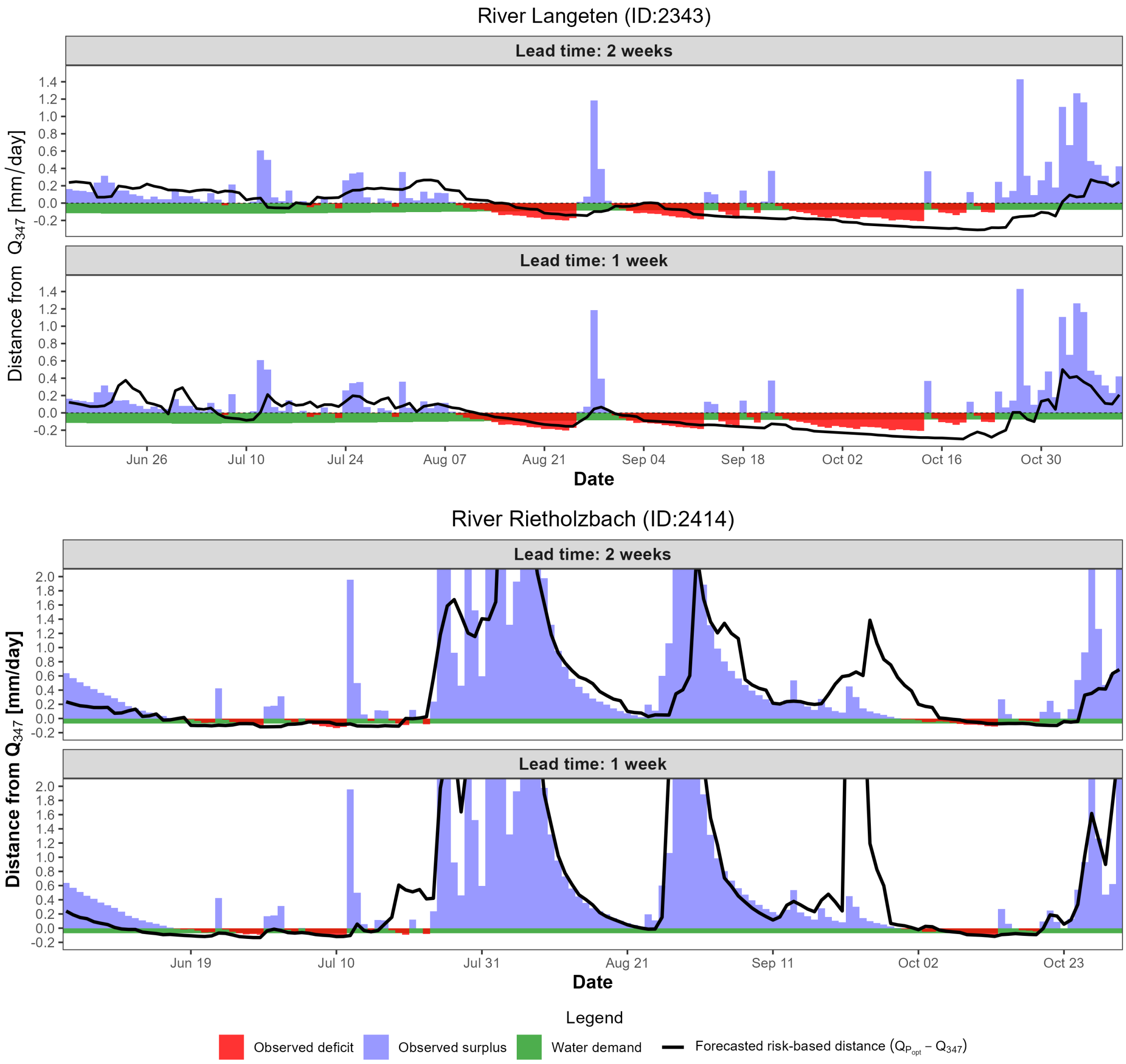


Figure 3.11: *Observed and forcasted distance from the $Q_{347}$ threshold for the Langeten and Rietholzbach rivers in the 2023 drought period. Note that the upper limit of the y-axis has been truncated to provide better visual focus on the deficit and water demand.*

### 3.3.3.2 Examples of avoidable drought

In 2023, the Wigger and Venoge rivers also experienced periods of deficit. However, as can be seen from Figure 3.12, the observed deficit severity on most days does not exceed the water demand, unlike what was previously observed for the Langeten and Rietholzbach rivers. This means that to comply with the $Q_{347}$ threshold, it would not have been necessary to waive the entire water demand volume, but only the amount required to maintain the ecological flow. The forecasted distance from the threshold provides a fairly accurate picture of this situation for both rivers, especially with a lead time of 1 week.

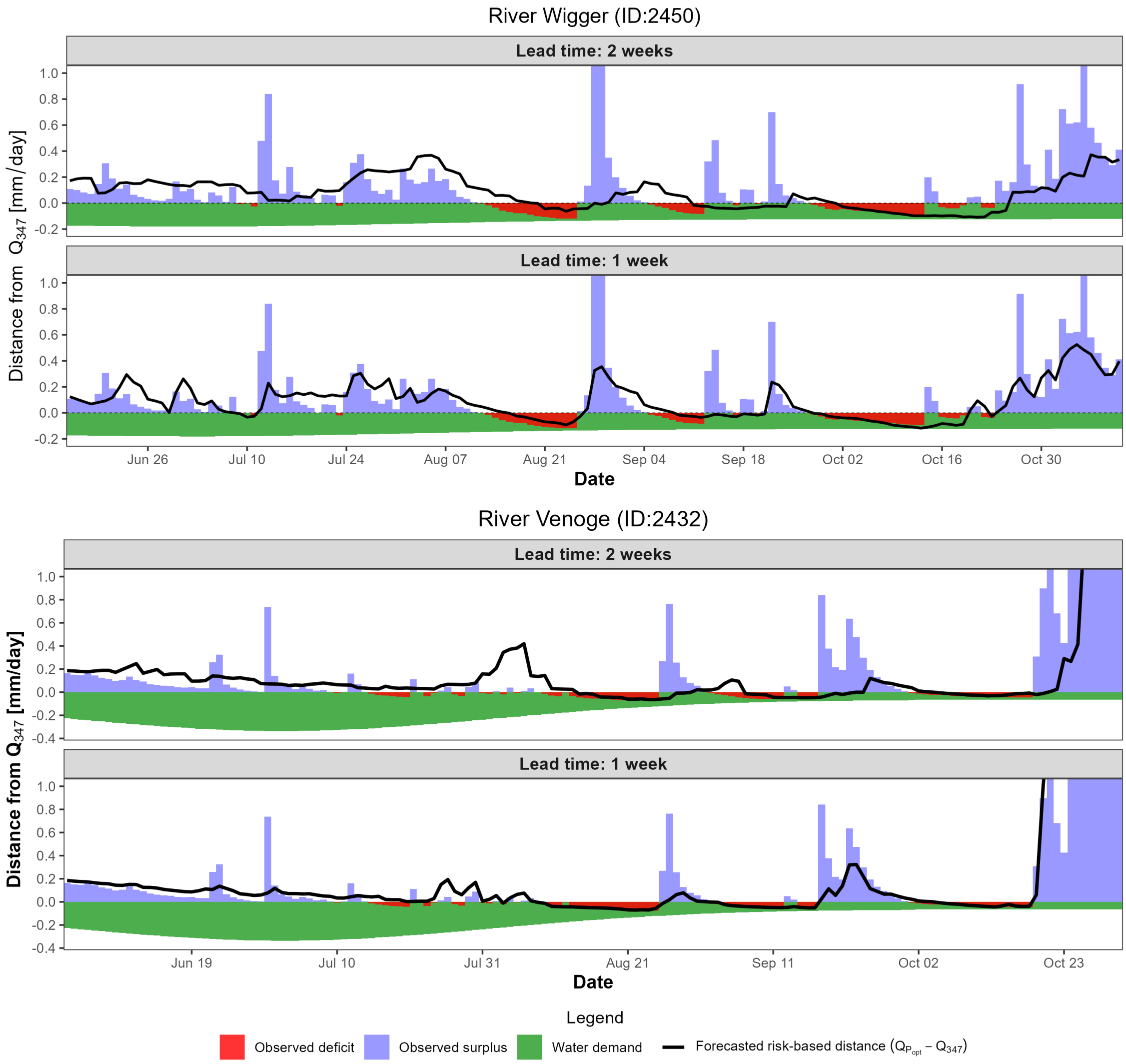


Figure 3.12: *Observed and forcasted distance from the $Q_{347}$ threshold for the Wigger and Venoge rivers in the 2023 drought period.*

### 3.3.3.3 False drought warnings

For some rivers, false drought warnings have been issued. In Figure 3.13, it can be observed that although the Worble River experienced only one day of drought in 2023, the forecasted distance remains below the $Q_{347}$ threshold for a much longer period. However, it can be noted that this false alarm period corresponds to an observed discharge very close to $Q_{347}$ threshold, where the surplus is almost negligible. Thus, in this case, although the alarm was incorrect, it was issued during a period of near-drought when the probabilistic drought signal was indeed very strong. For the Luthern River, however, it can be seen that the signaled periods of low flow approximately correspond to the observed ones, but especially in October, the forecasted distance overestimates the severity of the low flow compared to the observed data. The warning would therefore suggest that no water could be abstracted from the river to meet the $Q_{347}$ threshold requirement, whereas the observations show that reducing only part of the demand would be sufficient.

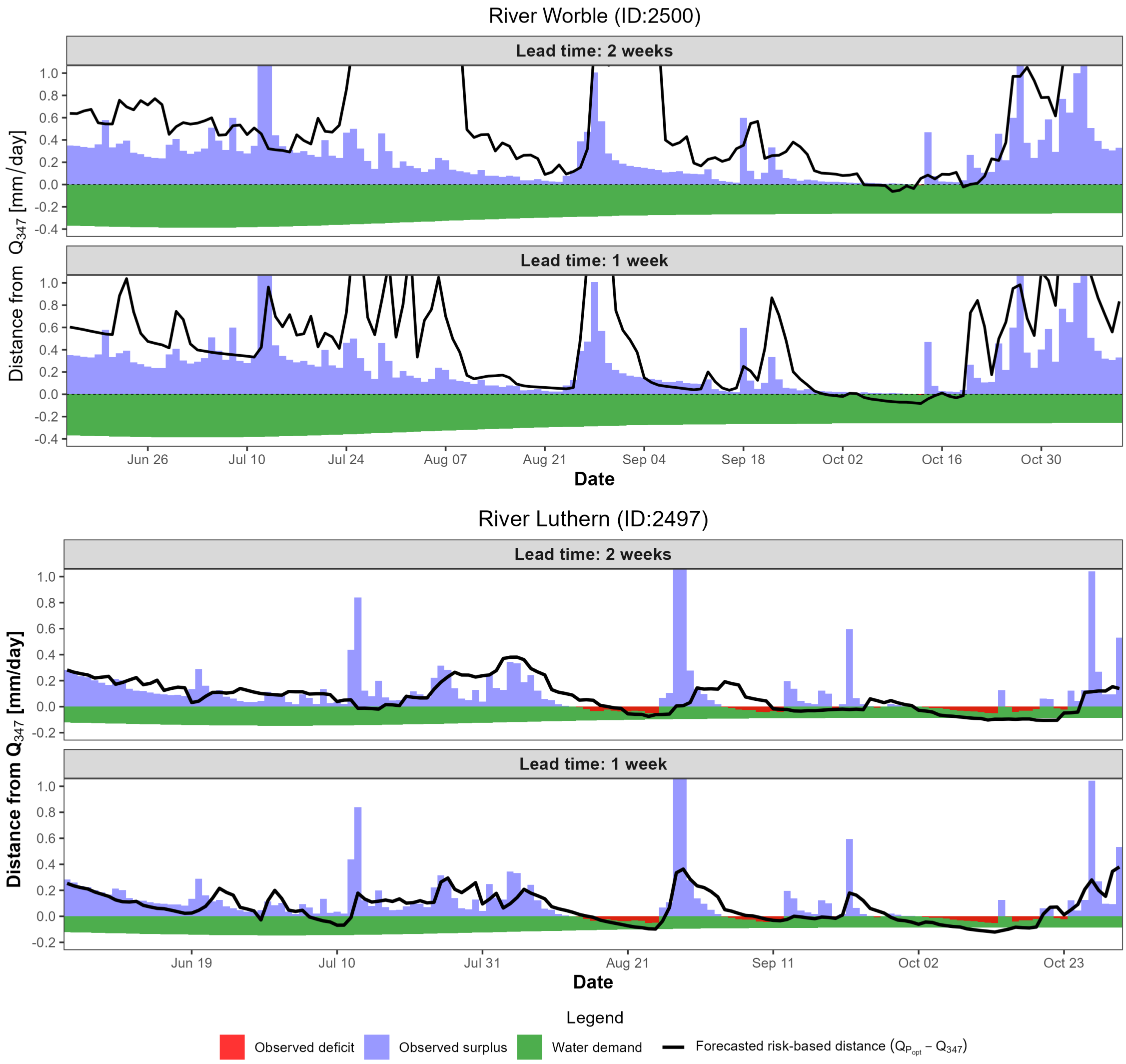


Figure 3.13: *Observed and forcasted distance from the $Q_{347}$ threshold for the Worble and Luthern rivers in the 2023 drought period.*

As noted in Chapter 3.3.2, the Mentue River catchment (ID: 2369) had the most false alarms in 2023, with no actual drought days observed. From Figure 3.14 it can be observed that even with lead times of just 1-2 weeks, the decision-making discharge lies too low resulting in a prolonged periods of deficit between August and October, while the observed discharge remained above the $Q_{347}$ threshold. Unlike other catchments, it can be noted that the forecast is less responsive and sharp, even for lead times of just a few weeks. The predicted drought days corresponded to periods when the observed discharge tended to decrease but still remained above the threshold. For the Murg River, where in 2023 there was also a period in which the discharge was very close to the $Q_{347}$ threshold but did not fall below it, the forecast closely followed the observations without generating false alarms.

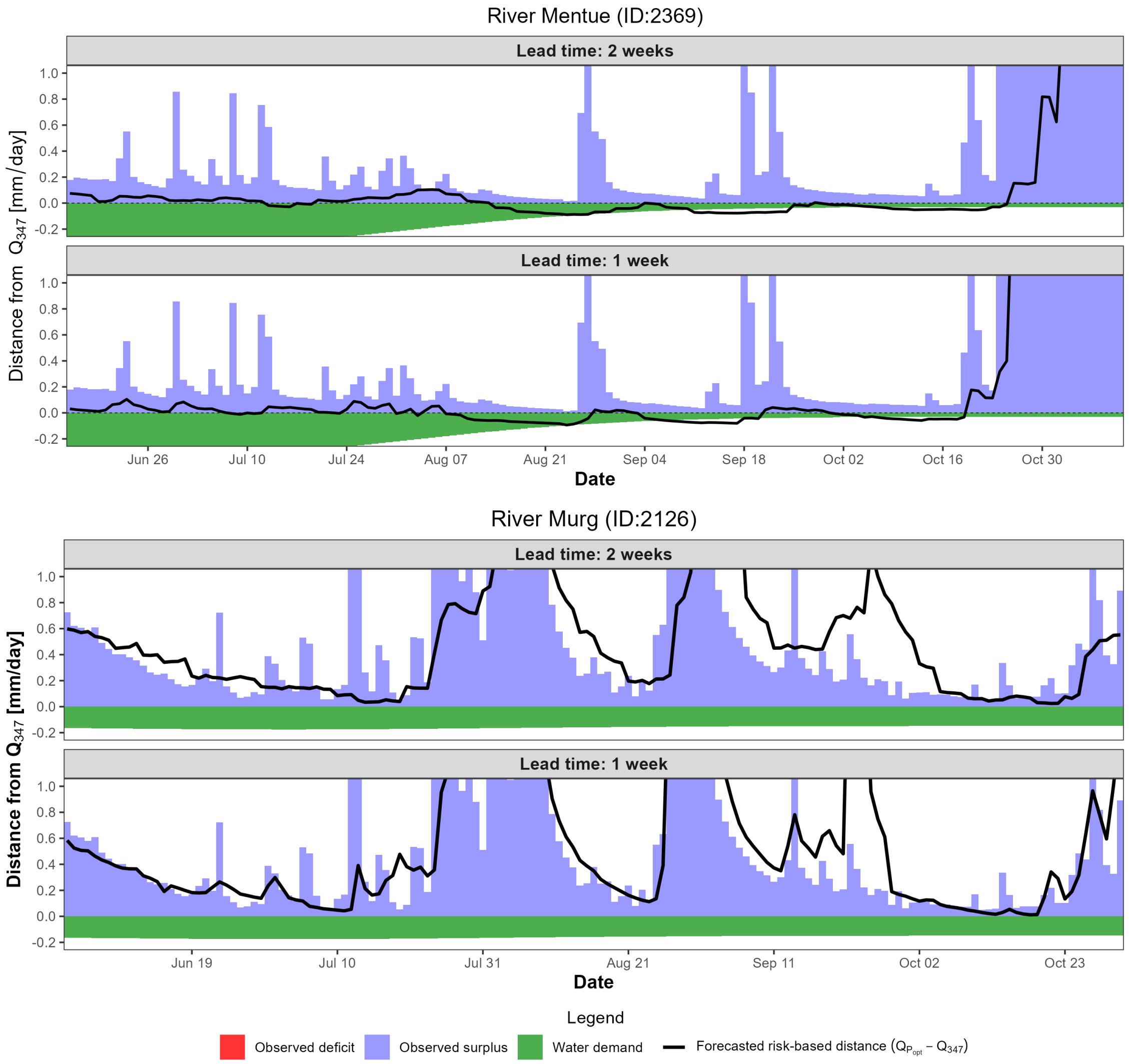


Figure 3.14: *Observed and forcasted distance from the $Q_{347}$ threshold for the Mentue and Murg rivers in the 2023 drought period.*

#### 3.3.3.4 Total deficit amount of 2023

To assess how effectively the $Q_{P_{opt}}$ captures the overall severity of the drought signal, the total annual deficit was calculated for each river in 2023, both for the observations and for the forecasted distances from the $Q_{347}$ threshold, by summing the daily values. As shown in Table 3.6, for some catchments, such as the 2202, 2412, 2432, and 2450, the annual sum of the predicted distances is consistent with the observed values. In catchments 2343 and 2471, where the measured deficit was the highest in 2023, the forecasted cumulative distance also indicates strongly negative values. In catchments 2126 and 2312, where no deficit was observed, the cumulative forecasted distance is equal to or close to zero. The risk-based $Q_{P_{opt}}$ threshold, therefore shows its potential as a good proxy for assessing drought severity up to two weeks of lead time.

Table 3.6: *Observed and forecasted cumulative deficit in 2023. The forecasted cumulative deficit is calculated as the sum of the predicted $Q_{P_{opt}}$ and $Q_{347}$ distances per lead time of 1 and 2 weeks.*

| Catchment (ID) | Forecasted cumulative deficit [mm] per weekly lead time | | Observed cumulative deficit [mm] |
|---|---|---|---|
| | 2 | 1 | |
| 2034 | −1.72 | −1.78 | −1.17 |
| 2126 | 0.00 | 0.00 | 0.00 |
| 2179 | −0.04 | −0.59 | −1.19 |
| 2202 | −0.83 | −0.83 | −0.89 |
| 2312 | −0.23 | −0.21 | 0.00 |
| 2343 | −13.15 | −12.67 | −8.67 |
| 2369 | −3.63 | −3.10 | 0.00 |
| 2412 | −0.54 | −1.38 | −1.59 |
| 2414 | −4.05 | −4.18 | −2.56 |
| 2416 | −0.62 | −0.76 | −1.27 |
| 2432 | −1.45 | −1.94 | −2.37 |
| 2434 | −0.62 | −0.65 | −0.27 |
| 2437 | −0.12 | −0.29 | −0.93 |
| 2450 | −2.66 | −2.46 | −2.82 |
| 2458 | −0.43 | −0.48 | −0.16 |
| 2471 | −10.71 | −10.87 | −7.97 |
| 2485 | −3.69 | −4.04 | −1.77 |
| 2497 | −2.73 | −2.78 | −1.10 |
| 2500 | −0.20 | −0.70 | −0.01 |
| 2608 | −0.35 | −0.68 | −2.16 |

# Chapter 4

# Discussion

## 4.1 Forecast performance

From the verification of drought events forecast accuracy, the ensemble composed of all models and all transformations was found to perform better than the other configurations, ensuring, especially for lead times up to 14-21 days, the highest Brier Skill Score for the majority of catchments. This configuration also demonstrated a strong ability to assign higher probabilities during drought events, effectively distinguishing events from non-events, as shown by the resolution term and AUC values.

The use of multiple hydrological models for streamflow simulation and forecasting is widely documented in the literature. A key motivation for adopting a multi-model approach is its ability to better account for forecast uncertainty associated with structural deficiencies and errors in individual hydrological models, which can be mitigated by exploiting model diversity (Troin et al., 2021).

The three analyzed ensemble configurations differ not only in the number of models and transformations used, but also significantly in their total number of members, 51, 255, and 2805. The large number of members in the multimodel configuration may have further contributed to better managing forecast uncertainty, especially during the first few days of forecast. In the configuration with only 51 members, composed of a single model and a single transformation, the individual members tend to produce very similar values during the first days of forecast, resulting in a reduced spread and thus providing little useful probabilistic information. In contrast, the two larger configurations generate a greater spread, more effectively capturing forecast uncertainty from the very first days of prediction (Buizza & Palmer, 1998).

Decomposing the forecast accuracy analysis by season proved essential for a proper interpretation of performance. Since the studied event is highly seasonal and rare, the annual Brier Score is heavily influenced by the majority of days when the event does not occur, during which the forecast correctly assigns a probability close to zero. This results in low Brier Scores that can mask the true performance of the forecasts during event periods. The ROC and AUC depend entirely on the presence of both event and non-event classes. By excluding forecasts and observations from winter and spring, which mostly correspond to non-events, the strong sensitivity of the ROC curve to the over representation of the negative class is reduced. This provides, in the case of a rare positive event such as drought, a more meaningful assessment of the model's ability to correctly classify the classes.

## 4.2 Drought warning issuance

Probabilistic forecasts of drought events have shown, across all catchments, a good ability to distinguish between drought and non-drought days. Based on this strength of the forecasts, the idea of classifying predictions into drought events or non-events was then developed. At the core of the classification between drought and non-drought events lies what has been defined as the optimal probabilistic threshold $P_{opt}$, which determines whether a warning is issued based on the daily probability of drought occurrence. For most catchments, the $P_{opt}$ threshold identified during the training period through $F_1$ score maximization, proved to be representative also during the test period, resulting in very similar $F_1$ scores across the two periods. For some catchments, however, the threshold in the test period often led to false alarms or missed events.

The split of forecasts into training and test periods led, for some catchments, to an imbalance in the number and type of events between the two datasets. Prolonged drought periods were found to exhibit a stronger signal and higher probabilities compared to short sequences of low flow days. Consequently, the range of probabilities between the training and test datasets also differed substantially in some cases. One of the most representative examples was catchment with ID 2369. As illustrated in Figure 4.1, during the training period (2019–2022) for this catchment, only two short drought periods were recorded, in which the predicted probability for the lead time of one week, remained below 40%. In 2022, the predicted drought probability was very high, but no events were observed. The $P_{opt}$ threshold was therefore optimized over a period characterized by very different probabilistic responses to events. In the test period, it can be observed that in 2023, although no events were recorded, the forecast produced a very high probability peak. With the threshold calibrated on the low probability response of the events in the training period, false alarms were inevitably issued in 2023. Moreover, for this river, streamflow predictions remain relatively inaccurate with respect to observations, featuring a false drought signal in 2022 and an almost absent predictive signal for the 2024 drought. In contrast, for the catchment with ID 2202, the probabilities during event days are much more consistent between the training and test periods, making the threshold representative for separating events from non-events also in the latter.

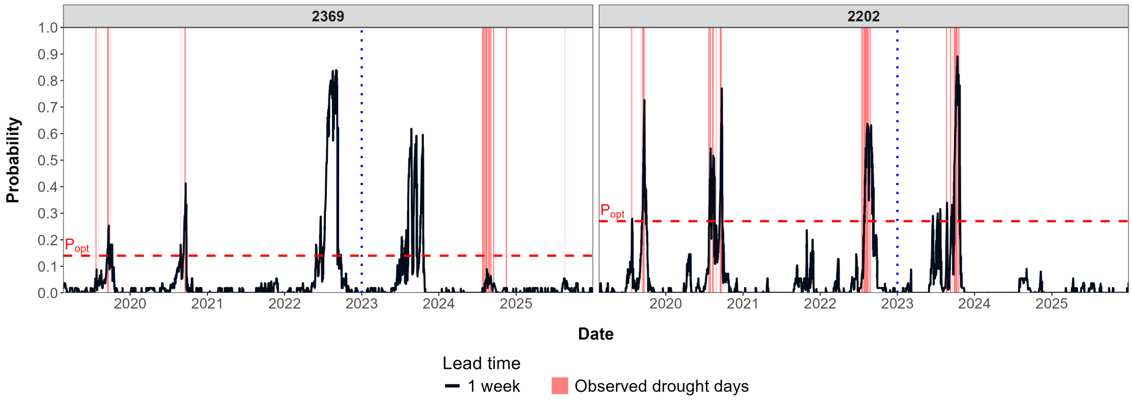


Figure 4.1: *Issued probabilities for a lead time of one week in training and test periods for catchments 2369 and 2202.*

A key limitation of this approach lies therefore in the division of the full dataset into training and test periods. Catchments where a strong decrease in $F_1$ score from training to test periods was observed, are typically those where both observed events and their associated probabilistic signals differ substantially between the two periods, resulting in a threshold calibrated on the training period that is not representative for event prediction and classification in the test period. Due to how it has been defined, the drought event constitutes a rare and extreme phenomenon, with some years in which no events are recorded. As a result, splitting the dataset in a way that is balanced in terms of events becomes even more challenging. Furthermore, in the absence of events, the $F_1$ score can not be computed, making it impossible to estimate the $P_{opt}$ threshold.

Despite these limitations, tuning the probabilistic threshold allows for the identification of an individual warning threshold for each catchment. The reliability diagrams indicated that the multimodel ensemble configuration yielded better calibrated probabilities in relation to the observed event frequencies, compared to the other two configurations. From Figure 4.2 it can be observed that, in reality, the calibration of probabilities is not consistent across different catchments. While in catchment 2412 the predicted event probabilities reach up to 100% , as can be seen for the drought periods in 2022 and 2023, probabilities in catchment 2450 reaches a maximum of around 70% and in catchment 2348 do not exceed 50%. While a threshold of around 75% would identify and predict most of the events for catchment 2412, such probability levels are not even reached for the other two catchments. Consequently, no warning would be issued, despite the presence of a probability peak during the observed events. For catchments 2450 and 2348, it can therefore be inferred that the discharge ensemble during low flow events is systematically too high, leading to underestimated drought probability. Therefore, in these cases, the issued probabilities can not be interpreted as reliable risk information, as they do not reflect the actual expected frequency of occurrence.

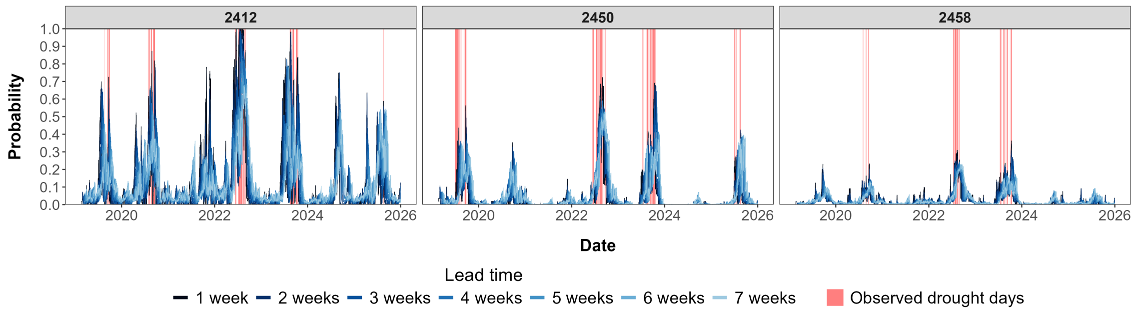


Figure 4.2: *Variability in forecasted drought probabilities across different catchments.*

The advantage of finding the warning threshold by optimizing the $F_1$ score is that it focuses exclusively on the separation between event and non-event based on probability values, regardless of their absolute calibration. The optimal threshold is therefore identified within the specific range of drought probabilities issued for each catchment, prioritizing the effective discrimination between occurrences. Bouttier and Marchal (2024) define this approach as implicit calibration, emphasizing that the probabilities and the $P_{opt}$ threshold should not be interpreted as true probabilities of event occurrence, unless they are properly calibrated. Silva Filho et al. (2023) emphasize that when a classifier is intended for tasks where performance is measured through contingency matrix, such as the $F_1$ score, there is no specific requirement for probability calibration, provided that the classification performance remains sufficiently high. While calibrating the probabilities would improve their interpretability, it would not resolve the fundamental issue of ensemble streamflow overestimation during drought conditions.

The threshold $P_{opt}$ thus represents, within the ensemble probability space, the value that maximizes the classification performance between event and non-event according to the $F_1$ metric. For the majority of the rivers, the forecast successfully classified and issued drought warnings correctly, particularly for lead times of 1 and 2 weeks. Warnings were also generated for lead times of up to 7 weeks, but with higher uncertainty regarding the exact onset and ending of the drought period. In some catchments, $P_{opt}$ was shown to identify the ensemble position $Q_{P_{opt}}$ that not only provides the best event discrimination but also captures the severity of observed deficits, at least for the first two weeks of forecasting, thus potentially representing an accurate connection point between the ensemble system and the observed discharge.

### 4.2.1 False alarms and missed events

An interesting aspect emerged when analyzing the observed streamflow conditions under which false alarms were issued. Figure 4.3 illustrates the frequency of false alarms across discharge percentile classes, pooling all 20 rivers together. It can be observed that most false alarms occur in the percentile classes immediately above the one defining drought conditions (0–5%), with nearly 40-50% of false alarms issued within the 5-10% and 10-15% discharge percentile classes, for the forecast with shorter lead times. Therefore, while these false alarms are statistically recorded as errors, most of them are actually near-drought events. This suggests that relying on a single threshold that classify only in event and non event classes can be penalizing for the forecast. It can also be observed that with increasing forecast lead time, the frequency of false positives recorded in the high flow percentile classes increases, especially in the test period. This is consistent with what was observed in the warning issued in 2023, where it was noted that beyond a two week lead time, warnings are delayed and therefore issued under non critical flow conditions.

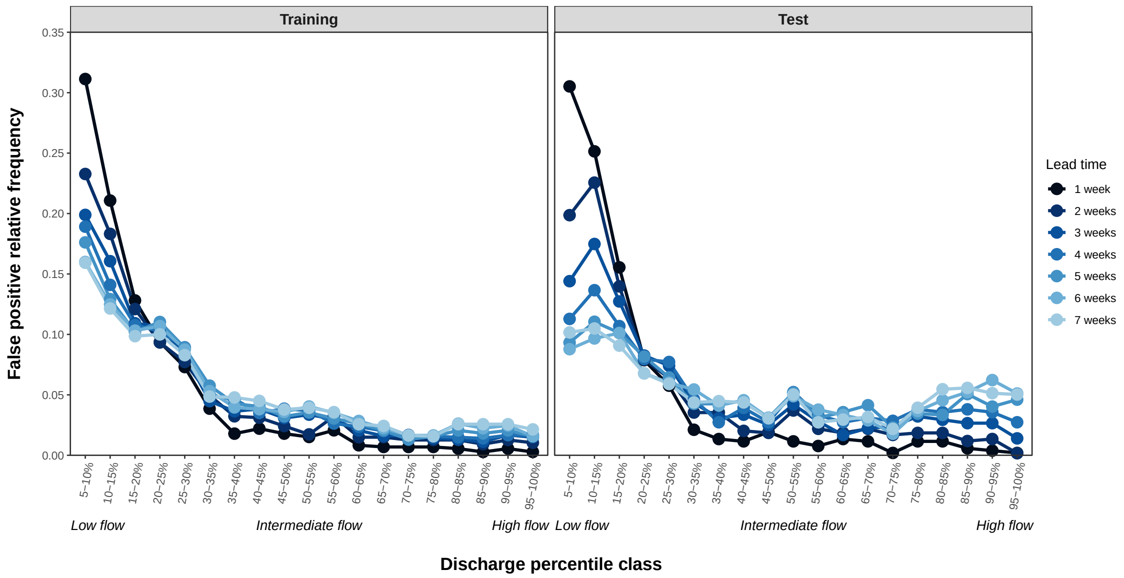


Figure 4.3: *False positive relative frequency across discharge percentile classes, for training and test periods, pooling all catchments.*

Regarding missed events, Figure 4.4 shows that the highest frequency of false negatives occurs for isolated days or very short periods of drought. For longer drought durations, the frequency of false negatives is lower, approaching zero for forecasts with shorter lead times, especially in the training period. This information is particularly relevant, as it implies that the forecast is highly reliable regarding the signal of prolonged drought period, which are the most important ones to foresee in the context of water resources management.

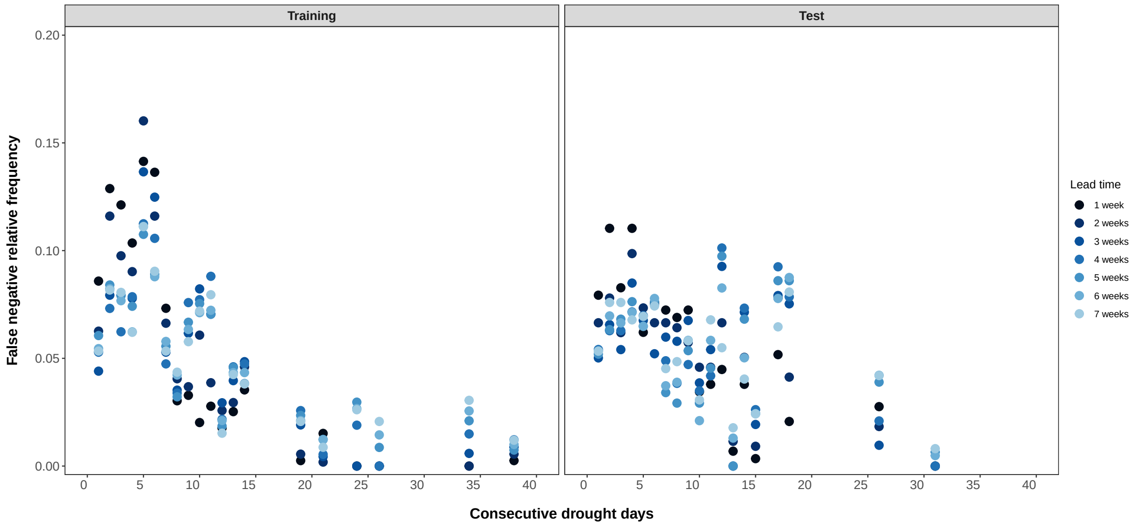


Figure 4.4: *False negative frequency (missed events) as a function of drought duration for the training and test periods, pooling all catchments.*

## 4.3 Drought forecast "delay"

The forecasts proved capable of provide a signal of potential drought conditions up to a lead time of seven weeks, as evidenced by the issuance of drought warnings in many catchments already at this time horizon. Despite this, in all catchments was observed that for lead times beyond two weeks, the timing of the forecasted drought periods becomes progressively delayed compared to the observation, consequently missing the onset of the events and prolonging the drought duration. To better understand this dynamic, the daily forecasted probabilities for all available lead times were visualized by plotting the target date on the x-axis and the lead times on the y-axis. Specifically, the 2023 drought period was examined in greater detail for catchment with ID 2034. In the upper panel of Figure 4.5 the raw plot is presented, where each diagonal of squares represents a single forecast run initialized on the same date and containing forecasts for the subsequent 46 consecutive days. Since forecasts are issued only twice a week, they do not form a continuous grid, resulting in visible missing data. Therefore, to enhance the plot readability, the colors have been interpolated to provide a continuous probability value for every day, as shown in the bottom panel of Figure 4.5.

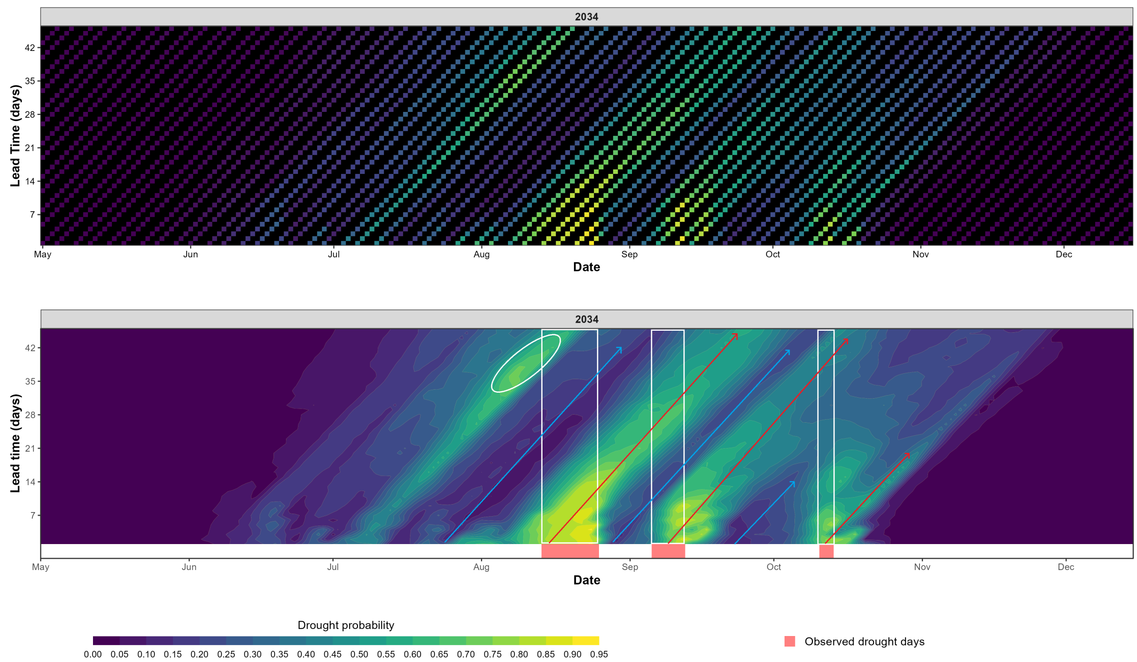


Figure 4.5: *Evolution of forecasted drought probabilities by lead time for the Broye catchment (ID: 2034) during the 2023 drought period. The x-axis represents the target date, while the y-axis shows the lead time. The upper panel shows the raw forecast output, where each diagonal band of squares represents a single forecast run initialized on the same date and covering the subsequent 46 days. The lower panel shows an interpolated version of the same data, providing a continuous probability field for each day.*

From these figures, interesting insights can be drawn. Similar to what has been observed for the drought warnings in some catchments, in the first two weeks of forecast the probabilistic drought signal remains better aligned with observations, thereafter, it begins to deviate progressively to the right. Looking at the first drought period in the middle of August, it can be seen that forecasts issued in early July, therefore 46 days in advance, had predicted a stronger drought signal specifically for the first decade of August (white circle). Subsequently, however, a period of higher flow because of precipitation occurred around the middle of July, and the forecasts issued during this time no longer indicated a drought signal for the middle of August (blue arrow). Only as August approaches, that is, two weeks before the event, do the issued forecasts begin to correctly predict the possibility of the beginning of the low flow event. The individual forecast runs issued, especially at the start of this drought period, are not yet seeing the brief but sharp increase in discharge towards the end of August and the beginning of September, likely due to an intense rainfall event. As a result, they tend to extend the low discharge signal over the subsequent 46 days of the same forecast run (red arrow), with the drought probability gradually decreasing but remaining relatively high. This dynamic is also observable in subsequent drought periods, which are interspersed with discharge peaks and for all catchments, especially in the forecasts of 2023 as can be seen in Figure 4.6.

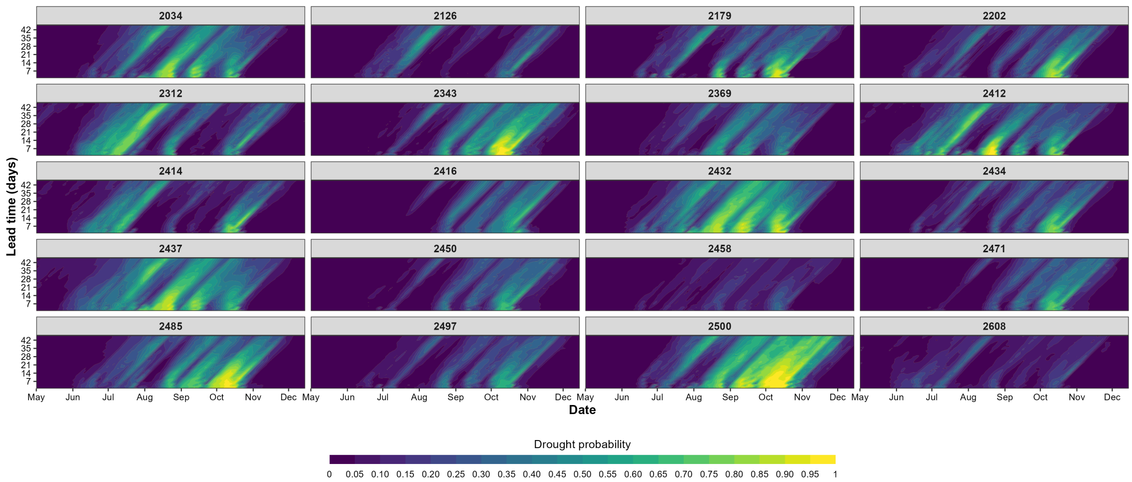


Figure 4.6: *Forecasted probabilities for all catchments across different lead times during the 2023 drought period.*

The 2022 drought period was sustained by more persistent low flow conditions, with fewer storm-driven discharge interruptions compared to 2023. For catchment 2034, it can be seen in Figure 4.7 that in 2022 the observed discharge remained below the ecological threshold for almost two consecutive months, between July and August. Examining the probabilities it can be observed that the signal is more compact and with higher probabilities even for longer lead times compared to 2023. Between late June and early July, there was a period of higher discharge, and in this case as well can be observed that the probabilities issued by the forecast runs of that period tend to strongly reduce the drought signal.

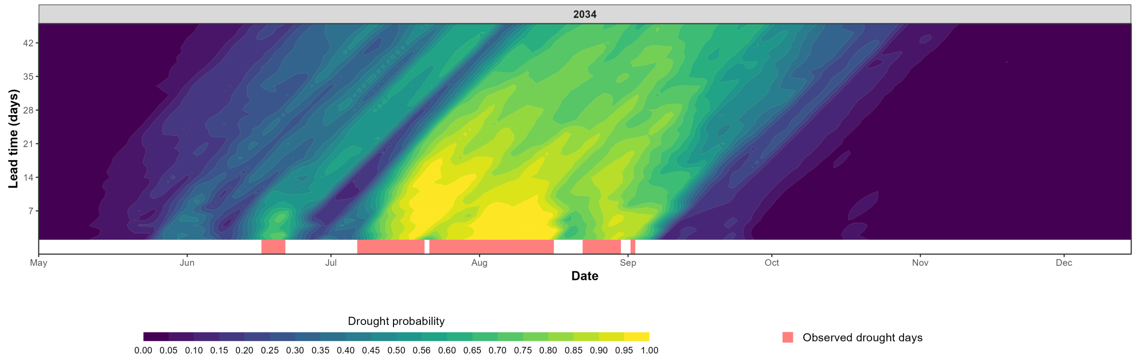


Figure 4.7: *Forecasted probabilities for the Broye catchment (ID: 2034) across different lead times during the 2022 drought period.*

It appears, therefore, that beyond two weeks, the forecasts are often influenced by initial conditions and are unable to change signal as rapidly as they do for shorter lead times, consequently, they tend to extend the same discharge conditions throughout the 46 day series, which appears to be late compared to the new observed conditions. The prolonged drought signal overlaps in some cases with subsequent periods of drought and in this case it is difficult to understand how to interpret it since from the warning point of view it would result in a correct alarm. As highlighted, however, in the white circle in Figure 3.11, there are also situations where single forecast runs starting with lower drought probability signal actually see the correct increase of drought probability at longer lead time, showing that there are periods when the forecast is more certain. The signal indicating the onset of drought is therefore consistently strong and stable across consecutive forecast runs, especially for lead times shorter than two weeks, where the forecast is also more capable of capturing the arrival of periods of increased flow.

This lack of responsiveness may stem from the large ensemble spread after week two, also considering the fact that a 2805 members was used. Furthermore, since the forecasts are issued only twice a week, does not help with consistency between different forecast runs, especially in summer when the atmosphere is highly dynamic and unpredictable. Therefore, it would be interesting to understand whether the fact that, beyond a certain time horizon, the ensemble fail to capture precisely the onset of a low flow period or a high flow event is due to the actual absence of this information in weather forecasts, especially regarding precipitation, or whether the issue lies in the structure and discharge generation process of hydrological models and how they convert meteorological input into runoff.

This rightward shift of the probabilistic signal therefore explains the pattern observed in drought warnings and their apparent delay observed at longer lead times. From the point of view of signal classification, the warnings are therefore correct but they are not with respect to the observations. Furthermore, by aggregating the daily lead times for each target day into weekly averages, this shift is further amplified and the near vertical signal observed in the first two weeks is hidden and almost no longer visible.

## 4.4 Water demand integration

Initial attempts to treat streamflow forecasts directly as total available water, from which water demand was subtracted, showed a marked overestimation of predicted drought durations compared to observed days. To be consistent with the observations, which reflect the reality of what occurred, the forecasted streamflow has been considered as already affected by the withdrawals. In doing so, it was possible to observe that, with regard to the duration of the events, the warning periods align more with the observations.

The warnings issued directly from the streamflow forecasts therefore already indicate that, with water abstraction from the river, the forecasted streamflow will fall below the ecological discharge $Q_{347}$, which must be respected by law, implying that, to meet this requirement, part or potentially all of the requested water demand in the catchment may not be abstracted from the river. The water demand was therefore added to the streamflow forecasts to investigate its potential influence on drought days, thereby reconstructing what has been defined as the natural streamflow scenario. In this way, it was possible to simulate what would have happened if no water had been abstracted and to compare the warning days with the abstraction-influenced scenario directly provided by the streamflow forecasts.

In particular, in 2023, the forecasts based on the natural scenario did not issue any drought warnings for most rivers, unlike the streamflow forecasts influenced by water abstraction. In this situation, therefore, the forecast indicated that water abstraction must be reduced or potentially entirely suspended in order to meet the residual flow requirement, meaning that not all water demand could be fulfilled. For some rivers, however, even the natural streamflow scenario issued drought warnings, indicating that despite the absence of water abstraction, the flow was still expected to fall below the threshold, implying that in such cases the full water demand could not be satisfied.

Information on potential water abstraction can therefore be inferred from the difference between the warnings issued under the natural streamflow scenario and those derived from abstraction-influenced streamflow forecasts. However, in this way, the natural streamflow scenario depends entirely on the estimate of total water demand, which, being an estimate, does not necessarily represent the actual amount of water abstracted from the river.

Visualizing the distance between the decision-making discharge $Q_{P_{opt}}$, and the $Q_{347}$ threshold, has proven to be highly informative regarding water abstraction feasibility during drought periods. By identifying the optimal probabilistic threshold ($P_{opt}$) that best distinguishes between drought events and non-events, the distance between the decision-making discharge $Q_{P_{opt}}$, and the physical $Q_{347}$ can be evaluated daily. This approach translates probabilistic uncertainty captured by the ensemble into a tangible physical margin, representing either the available water for extraction or the deficit that must be avoided to comply with ecological flow regulations, making the information more intuitive and directly comparable with legal requirements as well as the water demand. Moreover, it provides a better insight into why warnings are issued or not.

For catchments where the $P_{opt}$ threshold proved effective during the training period, this ensemble perspective demonstrated the ability to discriminate between different hydrological conditions and withdrawal scenarios, particularly within the first two weeks of the forecast. One of the most interesting aspects is certainly the possibility of visualizing those situations in which it is not necessary to sacrifice the entirety of the water demand withdrawal to restore the river flow above the required threshold. From a resource management perspective, and especially for water users, this is a crucial piece of information, as it makes it possible to foresee that a total restriction of water abstractions from the river may not be necessary. This information cannot be directly extracted through the binary classification of the event, as the total water demand is accounted as a single value, thereby masking the possibility of more flexible management strategies. For a lead time of about two weeks, the ensemble therefore reveals its potential in extracting insights into drought severity.

In some rivers, however, it was noted that the distance between $Q_{opt}$ and $Q_{347}$ does not align well with observations, accounting for the issuance of false alarms or missed events. A poor calibration of $P_{opt}$ can lead to a decision-making $Q_{Popt}$ value that is either too high or too low, thus providing misleading information regarding water availability or the severity of the deficit. Consequently, in such cases, the threshold fails to serve as a reliable indicator for assessing water abstraction potential.

## 4.5 Limitations

A major limitation of this study stems from the restricted dataset available for analysis. With only six years of forecasts available and focusing on extreme and rare seasonal events, the number of observable cases was limited relative to the total period. Indeed, only 2022 and 2023 featured low flow conditions persisting for several consecutive weeks across most of the studied rivers. In several years, no drought events were recorded across any of the rivers, further narrowing the sample of analyzable events. This aspect also had repercussions on the data analysis. Both the Brier Score and the ROC curve are heavily influenced by the fact that the positive class, or events, is very small compared to that of non-events. This often yields overly optimistic results regarding the accuracy of forecast and event classification. Furthermore, splitting the total dataset into training and testing periods further exacerbated the scarcity of events. This resulted for some catchments, in a non-representative $P_{opt}$ thresholds, which can make the forecast performance appear poor during the test period, even when the underlying limitation stems solely from the event classification process rather than the accuracy of the streamflow prediction itself. A larger dataset would be necessary, above all, to make a study of this kind more robust and to increase the representativeness of the results. It has thus been highlighted that studying rare events represents a significant challenge, both in terms of data availability and the correct interpretation and management of the resulting analysis.

Another limitation arises from the estimation of water demand and its integration into the forecast. In Switzerland today, precise data regarding water requirements are lacking, as regional and seasonal data are not collected at a national level (Federal Office for the Environment (FOEN), 2025). Agricultural water demand is undoubtedly the most dependent on river discharge, as evidenced by the damages and conflicts that emerged during the droughts of 2003 and 2018. In this thesis, a rate of 377 $m^3$ ha $yr^{-1}$ was applied, derived from an estimated water requirement for Switzerland of 144 million $m^3$ distributed across the $382,200$ hectares of Swiss agricultural land requiring irrigation. The census of agricultural holdings conducted by the Federal Statistical Office (FSO) estimated that Switzerland's water demand for irrigation in 2023 was approximately 37.4 million $m^3$, with an irrigated agricultural area of 34'200 hectares (Federal Statistical Office (FSO), 2024b). Similar values were also found in the modeling of irrigation water demand for Switzerland conducted by Baumgartner et al. (2025), which estimated a water requirement of 41 million $m^3$ for 2022 and approximately 31 million $m^3$ for 2023. A significant discrepancy is therefore observed in both the irrigated area and the water volume compared to the values used in this thesis. These differences would lead to substantially different results and interpretations regarding the potential for water abstraction from the rivers.

In Switzerland, water abstraction from both groundwater and surface water is subject to concessions issued by municipalities or cantons, however, the actual volumes withdrawn are rarely documented (Lanz et al., 2021). According to 2006 estimates, 73% of the water used for industry and commerce in Switzerland is obtained through self-supply, of which 77% is sourced from surface waters (Freiburghaus, 2009). Both agriculture and industry therefore exploit surface water in large quantities for their activities. Consequently, knowing the specific volume of water abstracted from rivers is of primary importance for providing accurate forecasts of water availability in the coming weeks. One possible approach to determining the extent of river abstraction could be to observe discharge variations at a higher temporal resolution than a daily scale. Observing hourly discharge rates could provide insights into flow fluctuations throughout the day, which could serve as a more precise proxy for estimating the actual water volume abstracted from each river.

The probabilistic forecasts, based on the selected ensemble configuration, showed a limit beyond a time horizon of two weeks in identifying the onset of water scarcity periods, especially when these periods are interspersed with discharge peaks. Monitoring the daily probability of low flow beyond this lead time may be too restrictive from a forecasting signal perspective. A weekly trend might provide a more indicative information regarding water low flow signal. Despite these limitations, a gradual change in probability was often observed both within the same 46 days forecast run and across consecutive runs, even for longer lead times. This trend can represent valuable information regarding a potential change in streamflow conditions. Having set a single threshold calibrated specifically on the probabilities issued during the first week, this gradual change in probabilities over time is lost, as the evolution of the forecast signal is masked by the binary classification. Implementing a specific threshold for each weekly lead time could maybe better adapt to their varying probability ranges thereby allowing even the weak signals that anticipate drought periods at longer lead times to be captured. Converting the continuous probability signal into a binary event/non-event signal can therefore lead to a significant loss of information regarding temporal variation of the low flow probability signal.

## 4.6 Further considerations

A key strength demonstrated by the forecasting system is its ability to adapt to each individual river. It was observed that during the same period, discharge conditions can vary significantly between catchments, particularly in summer when precipitation events can be highly localized. Consequently, some rivers may experience low flow periods while others simultaneously exhibit discharge peaks. The forecasting flexibility allows for site specific resource management, which is essential given the diverse water demand observed across different catchments. As highlighted by Brunner et al. (2019b), in a country like Switzerland, where each catchment possesses unique topographical and meteorological characteristics, water scarcity assessment should be conducted at the local level, as regional aggregation can lead to either over- or underestimating the actual water deficit. However, the necessity of treating each catchment individually also presents a significant challenge, particularly regarding data analysis and management. In this thesis, identifying the appropriate level of information to prioritize proved challenging in several instances. This was particularly evident when interpreting accuracy and verification metrics, such as the Brier Score and ROC-AUC, and when deciding which catchment-specific details were most critical to report in order to provide a balanced view of both the system's strengths and its limitations.

# Chapter 5

# Conclusion

Prolonged periods of low flow pose a significant challenge for water resource management. In Switzerland, such conditions have been shown to affect numerous activities, generating both economic and environmental impacts. In the context of established climate change, the hydrological regime of Swiss rivers is expected to undergo major variations, with the projected reduction in summer discharge coinciding with the period of peak water demand. Being able to anticipate periods of water scarcity is therefore a key element for water resource management, as it would enable more effective resource allocation and help mitigate the risk of conflicts over water use.

This thesis therefore aimed to evaluate the application of extended streamflow forecasts driven by Numerical Weather Prediction (NWP) for predicting periods of water scarcity, with the objective of understanding what information can be provided regarding water availability and its abstraction from watercourses in the coming weeks. To this end, 20 Swiss catchments potentially heavily influenced by surface water abstraction were considered. For each catchment, water demand was estimated by accounting for its seasonal variability and subsequently integrated with streamflow forecasts to gain insight into availability conditions and potential abstractions.

In accordance with the World Meteorological Organization (WMO) definition of socio-economic drought, water scarcity events were defined as periods when the river discharge falls below the ecological $Q_{347}$ threshold, which under Swiss law serves as the basis for determining the mandatory residual flow that must be guaranteed. Since the hydrological models were calibrated on observed discharge, the streamflow forecasts were considered to be already influenced by water abstraction. Consequently, the water demand was added back to reproduce a naturalized flow scenario, enabling a comparison between this reconstructed scenario and the original influenced forecasts.

An ensemble configuration composed of 11 lumped conceptual models, each calibrated with five different transformations, resulting in a total of 2'805 streamflow members and a lead time of 46 days, was therefore used to estimate, on a daily basis, the probability of discharge falling below the $Q_{347}$ threshold. With this configuration, the probabilistic drought forecasting system demonstrated, on average for the summer and autumn periods between 2019 and 2025, a skill generally superior to climatology for lead times up to 2–3 weeks, as indicated by the Brier Skill Score, while also showing a high discriminative capacity between drought days and non-events. This capacity to assign higher probabilities to drought event days than to non-event was leveraged to test the potential of issuing probability based drought warnings. For each river, an optimal probabilistic decision threshold was determined by maximizing the $F_1$ score, thereby balancing the system's ability to correctly identify drought events with the need to limit false alarms. The threshold was calibrated over the 2019–2022 training period and subsequently tested on the 2023–2025 period. A more in-depth analysis was conducted for the year 2023, which proved to be the most severe in terms of observed drought within the test period.

The results showed that it is possible to issue drought warnings for almost all catchments, and that these are particularly reliable for forecasts issued with lead times of up to two weeks. Forecast ability in anticipating drought periods proved to be variable, particularly across low flow periods in different rivers and years, however, the probabilistic signal indicating the onset of drought conditions appears to be more consistent across consecutive forecast runs for lead times shorter than two weeks relative to the beginning of the event. Beyond this time horizon, the forecasts show limited ability to capture sudden changes in discharge, such as those associated with abrupt rainfall events generating peak flows or with post-peak discharge recession, often exhibiting sensitivity to the initial conditions. For lead times shorter than two weeks, the potential to extract information from the ensemble regarding drought severity was also identified, thus allowing the provision of different scenarios concerning water availability in the coming days and the possibility of water abstraction.

In conclusion, this thesis highlighted the challenges related to data availability and management when analyzing rare events such as severe low flow events. Furthermore, it emphasized that a deep understanding of forecast system dynamics is essential for providing reliable information on approaching water scarcity periods and, consequently, on abstraction potential. While probabilistic streamflow forecasts have been shown to represent a valuable tool for early drought warnings and proactive water resource management, further analysis and, more importantly, additional data are required to refine and strengthen the robustness of these forecasts and the associated decision-making processes.

# Bibliography


Ambühl, J. (2010). *Customer oriented warning systems*. Veröffentlichungen der MeteoSchweiz, 84, 86 pp.

Azienda Multiservizi Bellinzona (AMB). (2024). *Rendiconto d'esercizio 2024*. Retrieved October 7, 2025, from https://www.amb.ch/chi-siamo/dati-su-amb/

BAFU. (2022). *Trockenheit im sommer 2022: Befragung der kantonalen gewässerschutz- und fischereifachstellen zu auswirkungen und massnahmen, bericht vom 24. februar 2023*. Bundesamt für Umwelt. Bern.

BAFU et al. (Hrsg.) (2019). *Hitze und trockenheit im sommer 2018: Auswirkungen auf mensch und umwelt* (No. 1909). Bern, Bundesamt für Umwelt, Umwelt-Zustand Nr. 1909: 91 S.

Baumgartner, T., Heinz, M., Turek, M., Erismann, G., Kohling, M., Aasen, H., & Holzkämper, A. (2025). *SwissIrrigationInfo: Modellbasierte Schätzungen des Wasserverbrauchs für Bewässerung in der Schweiz*. Agroscope. https://doi.org/10.34776/AS212G

Blanc, P., & Schädler, B. (2014). *Water in switzerland – an overview*. Bern, Swiss Hydrological Commission, 28 pages.

Bogner, K., Chang, A. Y.-Y., Bernhard, L., Zappa, M., Monhart, S., & Spirig, C. (2022). Tercile Forecasts for Extending the Horizon of Skillful Hydrological Predictions. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-21-0020.1

Bogner, K., Liechti, K., Bernhard, L., Monhart, S., & Zappa, M. (2018). Skill of Hydrological Extended Range Forecasts for Water Resources Management in Switzerland. *Water Resources Management*, *32*(3), 969–984. https://doi.org/10.1007/s11269-017-1849-5

Bouttier, F., & Marchal, H. (2024). Probabilistic short-range forecasts of high-precipitation events: Optimal decision thresholds and predictability limits. *Natural Hazards and Earth System Sciences*, *24*(8), 2793–2816. https://doi.org/10.5194/nhess-24-2793-2024

Boyd, K., Eng, K. H., & Page, C. D. (2013). Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals. In H. Blockeel, K. Kersting, S. Nijssen, & F. Železný (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 451–466). Springer. https://doi.org/10.1007/978-3-642-40994-3_29

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2

Brunner, M. I., Björnsen Gurung, A., Zappa, M., Zekollari, H., Farinotti, D., & Stähli, M. (2019a). Present and future water scarcity in Switzerland: Potential for alleviation through reservoirs and lakes. *Science of The Total Environment*, *666*, 1033–1047. https://doi.org/10.1016/j.scitotenv.2019.02.169

Brunner, M. I., Zappa, M., & Stähli, M. (2019b). Scale matters: Effects of temporal and spatial data resolution on water scarcity assessments. *Advances in Water Resources*, *123*, 134–144. https://doi.org/10.1016/j.advwatres.2018.11.013

Buizza, R., & Palmer, T. N. (1998). Impact of Ensemble Size on Ensemble Prediction. *Monthly Weather Review*, *126*(9), 2503–2518. https://doi.org/10.1175/1520-0493(1998)126⟨2503:IOESOE⟩2.0.CO;2

BUWAL, BWG, & MeteoSchweiz. (2004). *Auswirkungen des hitzesommers 2003 auf die gewässer* (No. 369) (Schriftenreihe Umwelt 174 pp.). Bundesamt für Umwelt, Wald und Landschaft. Bern.

CH2018. (2018). *Ch2018 – climate scenarios for switzerland: Technical report*. National Centre for Climate Services (NCCS). Zurich.

Domeisen, D. I. V., White, C. J., Afargan-Gerstman, H., Muñoz, Á. G., Janiga, M. A., Vitart, F., Wulff, C. O., Antoine, S., Ardilouze, C., Batté, L., Bloomfield, H. C., Brayshaw, D. J., Camargo, S. J., Charlton-Pérez, A., Collins, D., Cowan, T., Chaves, M. d. M., Ferranti, L., Gómez, R., . . . Tian, D. (2022). Advances in the Subseasonal Prediction of Extreme Events: Relevant Case Studies across the Globe. *Bulletin of the American Meteorological Society*, *103*(6), E1473–E1501. https://doi.org/10.1175/BAMS-D-20-0221.1

Eisenring, S., Holzkämper, A., & Calanca, P. (2021). Berechnung der bewässerungsbedürfnisse unter aktuellen und zukünftigen bedingungen in der schweiz. *Agroscope Science*, *107*, 55. https://doi.org/https://doi.org/10.34776/as107g

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Federal Office for the Environment (FOEN). (2021). *Effects of climate change on swiss water bodies: Hydrology, water ecology and water management* (No. 2101). Federal Office for the Environment. Bern, Switzerland.

Federal Office for the Environment (FOEN). (2025, September 25). *Water use*. Retrieved April 13, 2026, from https://www.bafu.admin.ch/en/water-use

Federal Office for the Environment (FOEN). (2026). *Drought: Monitoring, forecasting and warning*. Retrieved 2026, from https://www.bafu.admin.ch/en/drought

Federal Statistical Office (FSO). (2021). *Arbeitsstätten und beschäftigte nach gemeinde und wirtschaftssektor*. STAT-TAB – interaktive Tabellen (BFS). https://www.bfs.admin.ch/asset/de/26665325

Federal Statistical Office (FSO). (2022a). *Nutztiere und landwirtschaftliche nutzflächen im jahr 2022*. https://www.bfs.admin.ch/news/de/2023-0407

Federal Statistical Office (FSO). (2022b). *Schweizer tourismusstatistik 2022*. https://www.bfs.admin.ch/bfs/de/home/statistiken/tourismus/beherbergung.assetdetail.30405413.html

Federal Statistical Office (FSO). (2022c). *Statistik der bevölkerung und der haushalte (statpop)*. https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/erhebungen/statpop.assetdetail.22406441.html

Federal Statistical Office (FSO). (2024a). Hotel sector: Arrivals and overnight stays of open establishments by year, month, canton and visitors' country of residence. https://www.pxweb.bfs.admin.ch/pxweb/en/px-x-1003020000_102/px-x-1003020000_102/px-x-1003020000_102.px

Federal Statistical Office (FSO). (2024b, November 22). *Landwirtschaftliche betriebszählung 2023 – zusatzerhebung herbst 2023: Schweizer landwirtschaft setzt 2023 über 37 millionen kubikmeter wasser für die bewässerung ein* [Accessed: 2026-04-13]. https://www.bfs.admin.ch/bfs/en/home.assetdetail.33026282.html

Floriancic, M. G., Berghuijs, W. R., Jonas, T., Kirchner, J. W., & Molnar, P. (2020). Effects of climate anomalies on warm-season low flows in Switzerland. *Hydrology and Earth System Sciences*, *24*(11), 5423–5438. https://doi.org/10.5194/hess-24-5423-2020

FOPH & FOEN. (2010). *Reporting for switzerland under the protocol on water and health* (Public health, Department of Food Security). Bern, Switzerland.

Freiburghaus, M. (2015). Wasserverbrauch. *Aqua & Gas*, 72–79.

Freiburghaus, M. (2009). Wasserbedarf der schweizer wirtschaft. *GWA (Zürich)*, *89*(12), 1001–1009.

Fuhrer, J. (2010). *Abschätzung des bewässerungsbedarfs in der schweizer landwirtschaft*. Agroscope Reckenholz-Tänikon Research Station ART, Switzerland.

HAFL. (2023). *Datengrundlage und künftige datenerfassung zur landwirtschaftlichen bewässerung in der schweiz: Projekt swiss irrigation info – schlussbericht modul 1*. Hochschule für Agrar-, Forst- und Lebensmittelwissenschaften (HAFL). Zollikofen, Schweiz.

Hancock, J., Johnson, J. M., & Khoshgoftaar, T. M. (2022). A Comparative Approach to Threshold Optimization for Classifying Imbalanced Data. *2022 IEEE 8th International Conference on Collaboration and Internet Computing (CIC)*, 135–142. https://doi.org/10.1109/CIC56439.2022.00028

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., & Fenicia, F. (2023). CAMELS-CH: Hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland. *Earth System Science Data*, *15*(12), 5755–5784. https://doi.org/10.5194/essd-15-5755-2023

Hsu, W., & Murphy, A. H. (1986). The attributes diagram a geometrical framework for assessing the quality of probability forecasts. *International Journal of Forecasting*, *2*(3), 285–293.

*Hydrologisches multimodell.* (n.d.). drought.ch. Retrieved October 9, 2025, from https://www.drought.ch/de/prognosen/hydrologisches-multimodell/

Imani, M., Joudaki, M., Bagheri, A., & Arabnia, H. R. (2026). Why ROC-AUC Is Misleading for Highly Imbalanced Data: In-Depth Evaluation of MCC, F2-Score, H-Measure, and AUC-Based Metrics Across Diverse Classifiers. *Technologies*, *14*(1), 54. https://doi.org/10.3390/technologies14010054

Kaule, R., & Gilfedder, B. S. (2021). Groundwater Dominates Water Fluxes in a Headwater Catchment During Drought. *Frontiers in Water*, *3*. https://doi.org/10.3389/frwa.2021.706932

Koch, H., Liersch, S., de Azevedo, J. R. G., Silva, A. L. C., & Hattermann, F. F. (2018). Assessment of observed and simulated low flow indices for a highly managed river basin. *Hydrology Research*, *49*(6), 1831–1846. https://doi.org/10.2166/nh.2018.168

Kohn, I., Stahl, K., & Stoelzle, M. (2019). *Low flow events - a review in the context of climate change in switzerland*. Commissioned by the Federal Office for the Environment (FOEN). Bern, Switzerland. https://doi.org/10.6094/UNIFR/150448

Lanz, K., et al. (2021). *Auswirkungen des klimawandels auf die wasserwirtschaft der schweiz* (Beiträge zur Hydrologie der Schweiz, Nr. 43). Bern.

Lipton, Z. C., Elkan, C., & Naryanaswamy, B. (2014). Optimal Thresholding of Classifiers to Maximize F1 Measure. In T. Calders, F. Esposito, E. Hüllermeier, & R. Meo (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 225–239). Springer. https://doi.org/10.1007/978-3-662-44851-9_15

Liu, W., Park, S., Bailey, R. T., Molina-Navarro, E., Andersen, H. E., Thodsen, H., Nielsen, A., Jeppesen, E., Jensen, J. S., Jensen, J. B., & Trolle, D. (2020). Quantifying the streamflow response to groundwater abstractions for irrigation or drinking water at catchment scale using SWAT and SWAT–MODFLOW. *Environmental Sciences Europe*, *32*(1), 113. https://doi.org/10.1186/s12302-020-00395-6

Lorenz, E. N. (1969). Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, 636–646. https://doi.org/https://doi.org/10.1175/1520-0469(1969)26⟨636:APARBN⟩2.0.CO;2

*Mach-flow*. (n.d.). Swiss Data Science Center. Retrieved October 9, 2025, from https://www.datascience.ch/projects/mach-flow

Mason, I. (1982). A model for assessment of weather forecasts. *Australian Meteorological Magazine*, *30*(4), 291–303.

Mason, S. J. (2004). On Using "Climatology" as a Reference Strategy in the Brier and Ranked Probability Skill Scores. *Monthly Weather Review*, *132*(7), 1891–1895. https://doi.org/10.1175/1520-0493(2004)132⟨1891:OUCAAR⟩2.0.CO;2

Miao, J., & Zhu, W. (2022). Precision–recall curve (PRC) classification trees. *Evolutionary Intelligence*, *15*(3), 1545–1569. https://doi.org/10.1007/s12065-021-00565-2

Molteni, F., Buizza, R., Palmer, T. N., & Petroliagis, T. (1996). The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, *122*(529), 73–119. https://doi.org/10.1002/qj.49712252905

Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C., & Liniger, M. A. (2018). Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Observations. *Journal of Geophysical Research: Atmospheres*, *123*(15), 7999–8016. https://doi.org/10.1029/2017JD027923

Monhart, S., Zappa, M., Spirig, C., Schär, C., & Bogner, K. (2019). Subseasonal hydrometeorological ensemble predictions in small- and medium-sized mountainous catchments: Benefits of the NWP approach. *Hydrology and Earth System Sciences*, *23*(1), 493–513. https://doi.org/10.5194/hess-23-493-2019

Murphy, A. H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology and Climatology*, *12*(4), 595–600.

Navas, R., Gelós, M., & Bailey, R. (2025). Quantifying Baseflow Changes Due to Irrigation Expansion Using SWAT+gwflow. *Water*, *17*(11), 1680. https://doi.org/10.3390/w17111680

Obi, J. C. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, *8*(1), 308–314. https://doi.org/10.30574/wjaets.2023.8.1.0054

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., & Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1), 33. https://doi.org/10.1007/s10584-011-0149-y

Richardson, D., Fowler, H. J., Kilsby, C. G., Neal, R., & Dankers, R. (2020). Improving sub-seasonal forecast skill of meteorological drought: A weather pattern approach. *Natural Hazards and Earth System Sciences*, *20*(1), 107–124. https://doi.org/10.5194/nhess-20-107-2020

Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M., & Flach, P. (2023). Classifier calibration: A survey on how to assess and improve predicted class probabilities. *Machine Learning*, *112*(9), 3211–3260. https://doi.org/10.1007/s10994-023-06336-7

Smiatek, G., Kunstmann, H., & Werhahn, J. (2012). Implementation and performance analysis of a high resolution coupled numerical weather and river runoff prediction model system for an Alpine catchment. *Environmental Modelling & Software*, *38*, 231–243. https://doi.org/10.1016/j.envsoft.2012.06.001

Steeb, N., Lustenberger, F., & Zappa, M. (2024). Beurteilung der Beeinflussung des Abflusses an NAWA-Messstellen. Detailbericht des BAFU-Projekts HydCHeck. Eidgenössische forschungsanstalt für wald, schnee und landschaft (WSL). https://doi.org/10.55419/wsl:37799

Sujon, K. M., Hassan, R., Choi, K., & Samad, M. A. (2025). Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. *Journal of Big Data*, *12*(1), 268. https://doi.org/10.1186/s40537-025-01313-4

SVGW. (2020). *Statistische erhebungen der wasserversorgungen in der schweiz, betriebsjahr 2016* (Information W 15 001). SVGW – Schweizerischer Verein des Gas- und Wasserfaches. Zürich.

Terrier, M., Perrin, C., de Lavenne, A., Andréassian, V., Lerat, J., & Vaze, J. (2021). Streamflow naturalization methods: A review. *Hydrological Sciences Journal*, *66*(1), 12–36. https://doi.org/10.1080/02626667.2020.1839080

Toth, Z., Talagrand, O., Candille, G., & Zhu, Y. (2003). Probability and ensemble forecasts. *Forecast verification: A practitioner's guide in atmospheric science*, *137*, 163.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J.-L. (2021). Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research*, *57*(7), 48. https://doi.org/10.1029/2020WR028392

Vitart, F., & Takaya, Y. (2021). Lagged ensembles in sub-seasonal predictions. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3227–3242. https://doi.org/10.1002/qj.4125

Vorkauf, M., Steiger, R., Abegg, B., & Hiltbrunner, E. (2024). Snowmaking in a warmer climate: An in-depth analysis of future water demands for the ski resort Andermatt-Sedrun-Disentis (Switzerland) in the twenty-first century. *International Journal of Biometeorology*, *68*(3), 565–579. https://doi.org/10.1007/s00484-022-02394-z

Wada, Y., Beek, L. P. H. v., Viviroli, D., Dürr, H. H., Weingartner, R., & Bierkens, M. F. P. (2011). Global monthly water stress: 2. Water demand and severity of water stress [Publisher: John Wiley & Sons, Ltd]. *Water Resources Research*, *47*(7). https://doi.org/10.1029/2010WR009792

Weber, M., & Schild, A. (2007). Stand der bewässerung in der schweiz. *Bericht zur Umfrage 2006*, 17.

Wechsler, T., Björnsen, A., Wartmann, F., Rohr, C., & Schädler, B. (2025). Das wasserschloss europas im wandel. eine zeitungsartikelanalyse zur verwendung der metapher im kontext der klimaerwärmung. *Aqua & Gas*, *105*(12), 64–69.

Wilks, D. S. (1995). *Statistical methods in the atmospheric sciences* [pp. 467]. Academic Press.

Wilks, D. S. (2011). Forecast verification. In *International geophysics* (pp. 301–394, Vol. 100). Elsevier.

Yuan, Y., Wang, D., Zhang, L., Wu, Q., & Guo, H. (2024). Data-driven urban waterlogging risk management approach considering efficiency-equity trade-offs and risk mitigation capability evaluation. *Journal of Hydrology*, *634*, 131004. https://doi.org/10.1016/j.jhydrol.2024.131004

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., & Emanuel, K. (2019). What Is the Predictability Limit of Midlatitude Weather? [Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences]. *Journal of the Atmospheric Sciences*, *76*(4), 1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding the Best Classification Threshold in Imbalanced Classification. *Big Data Research*, *5*, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001

# Appendix A

# Appendix

## A.1 Brier Score

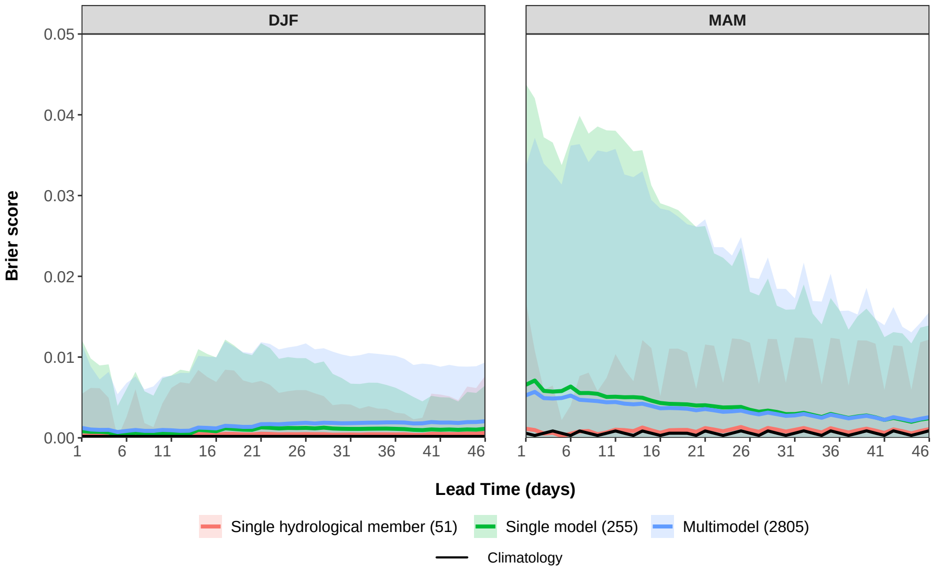


Figure A.1: *Brier score comparison between ensemble configurations for winter and spring seasons per lead time. The line shows the median Brier skill score for each configuration aggregating all catchments, while the shaded area shows the range of Brier scores across catchments.*

## A.2 Threshold tuning

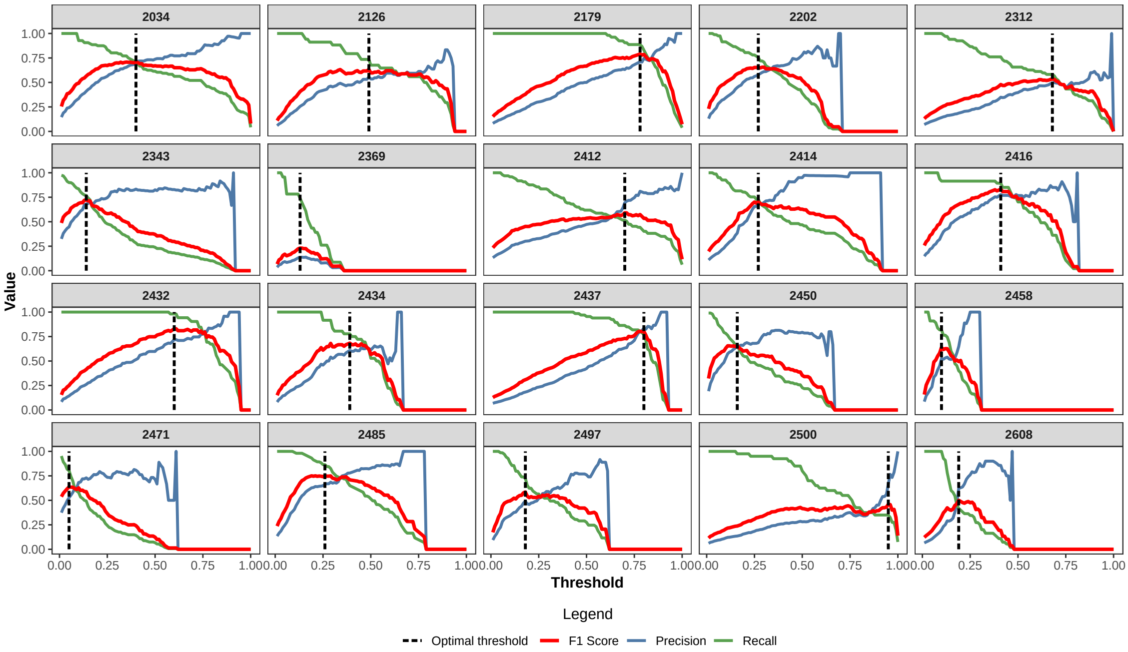


Figure A.2: *Threshold tuning process for all the catchments. The figure shows the $F_1$ score, precision, and recall values obtained by testing probability thresholds from 0 to 1 with a step of 0.01. The maximum $F_1$ score corresponds to the optimal probability threshold position.*

## A.3 Drought warnings

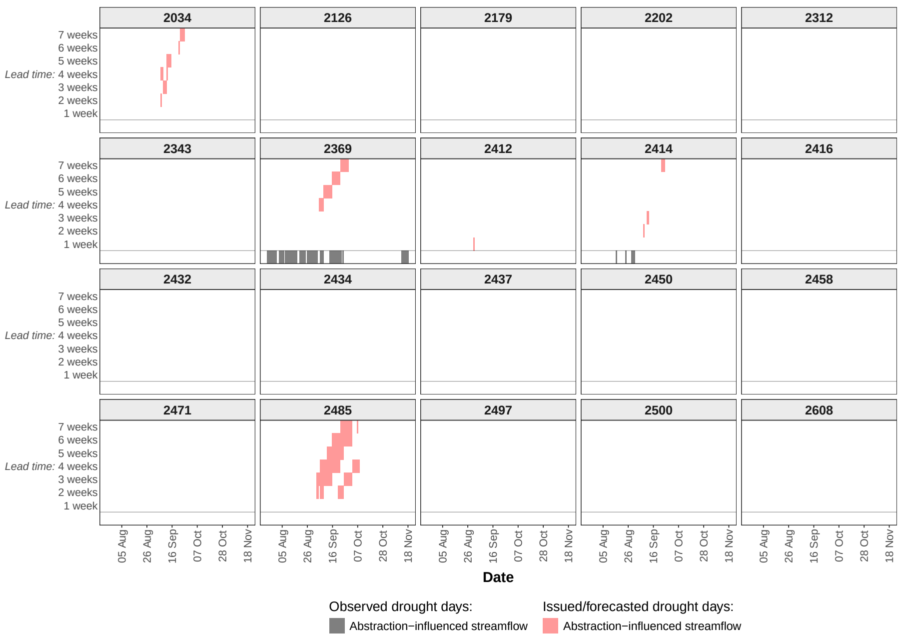


Figure A.3: *Observed and forecasted drought days in 2024 across the catchments.*

Table A.1: *Issued and observed drought days in 2024, including water demand abstraction.*

| Catchment (ID) | Issued/forecasted drought days per lead time week | | | | | | | Observed drought days |
|---|---|---|---|---|---|---|---|---|
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| 2034 | 4 | 1 | 4 | 3 | 3 | 1 | 0 | 0 |
| 2126 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2312 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2343 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2369 | 7 | 7 | 7 | 4 | 0 | 0 | 0 | 56 |
| 2412 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 2414 | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 5 |
| 2416 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2432 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2434 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2437 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2450 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2458 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2471 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2485 | 11 | 17 | 14 | 23 | 20 | 10 | 0 | 0 |
| 2497 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2608 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table A.2: *Issued and observed drought days in 2025, including water demand abstraction.*

| Catchment (ID) | Issued/forecasted drought days per lead time week | | | | | | | Observed drought days |
|---|---|---|---|---|---|---|---|---|
| | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| 2034 | 3 | 7 | 4 | 0 | 0 | 0 | 0 | 13 |
| 2126 | 18 | 23 | 21 | 19 | 18 | 21 | 24 | 0 |
| 2179 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2202 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2312 | 0 | 1 | 11 | 19 | 22 | 31 | 33 | 24 |
| 2343 | 64 | 62 | 58 | 50 | 37 | 27 | 25 | 12 |
| 2369 | 15 | 10 | 5 | 0 | 0 | 0 | 0 | 1 |
| 2412 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 2414 | 30 | 28 | 25 | 28 | 27 | 29 | 23 | 3 |
| 2416 | 38 | 36 | 38 | 38 | 35 | 37 | 35 | 0 |
| 2432 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| 2434 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2437 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2450 | 52 | 52 | 47 | 41 | 31 | 16 | 21 | 16 |
| 2458 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2471 | 87 | 72 | 69 | 60 | 53 | 41 | 46 | 3 |
| 2485 | 34 | 25 | 20 | 6 | 0 | 2 | 0 | 8 |
| 2497 | 44 | 41 | 36 | 32 | 28 | 20 | 23 | 0 |
| 2500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2608 | 6 | 7 | 10 | 11 | 3 | 1 | 0 | 2 |

## A.4 Discharge distance from $Q_{347}$

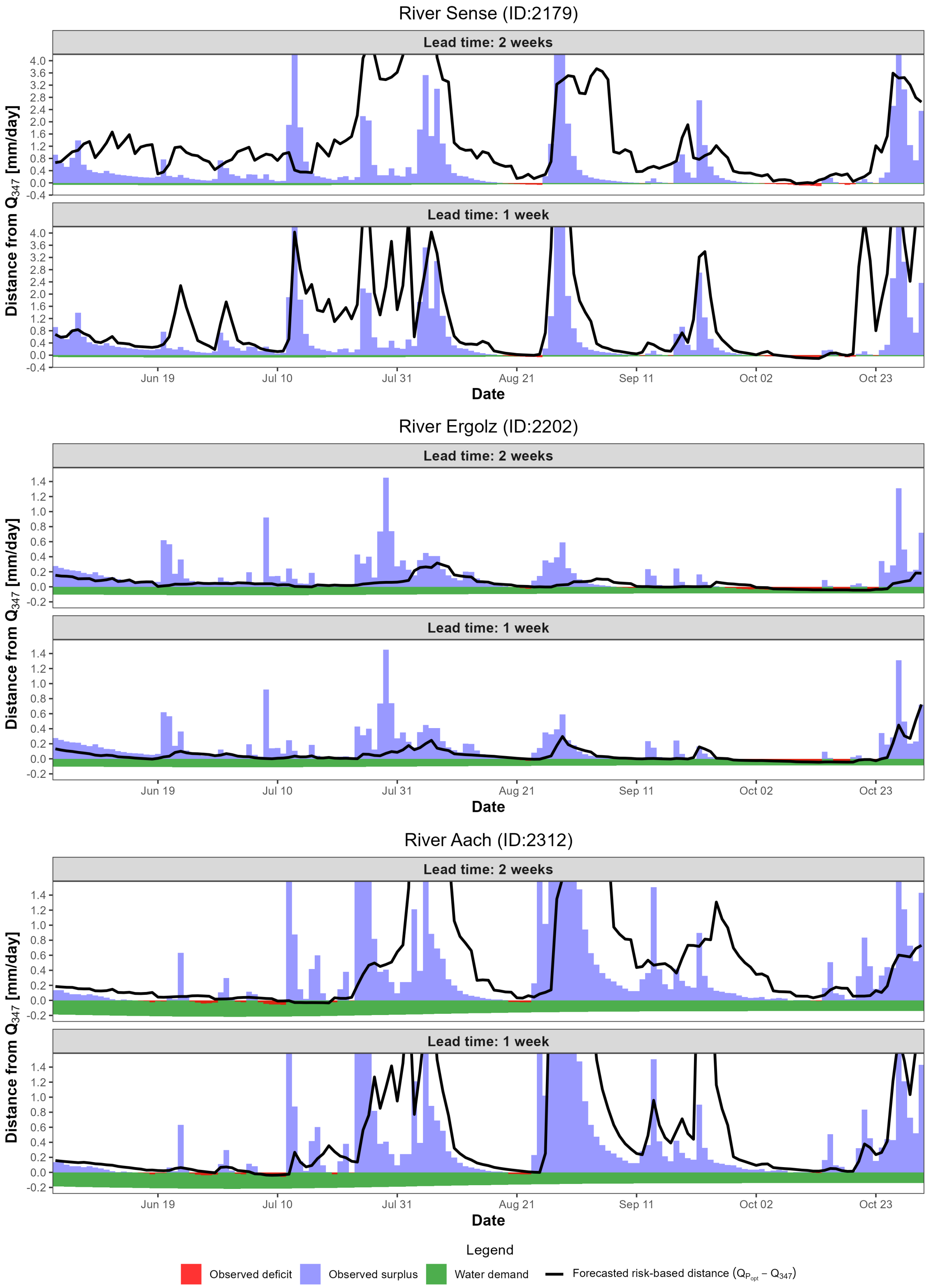


Figure A.4: *Observed and forecasted distance from the $Q_{347}$ threshold for the Sense, Ergolz, and Aach rivers in the 2023 drought period.*

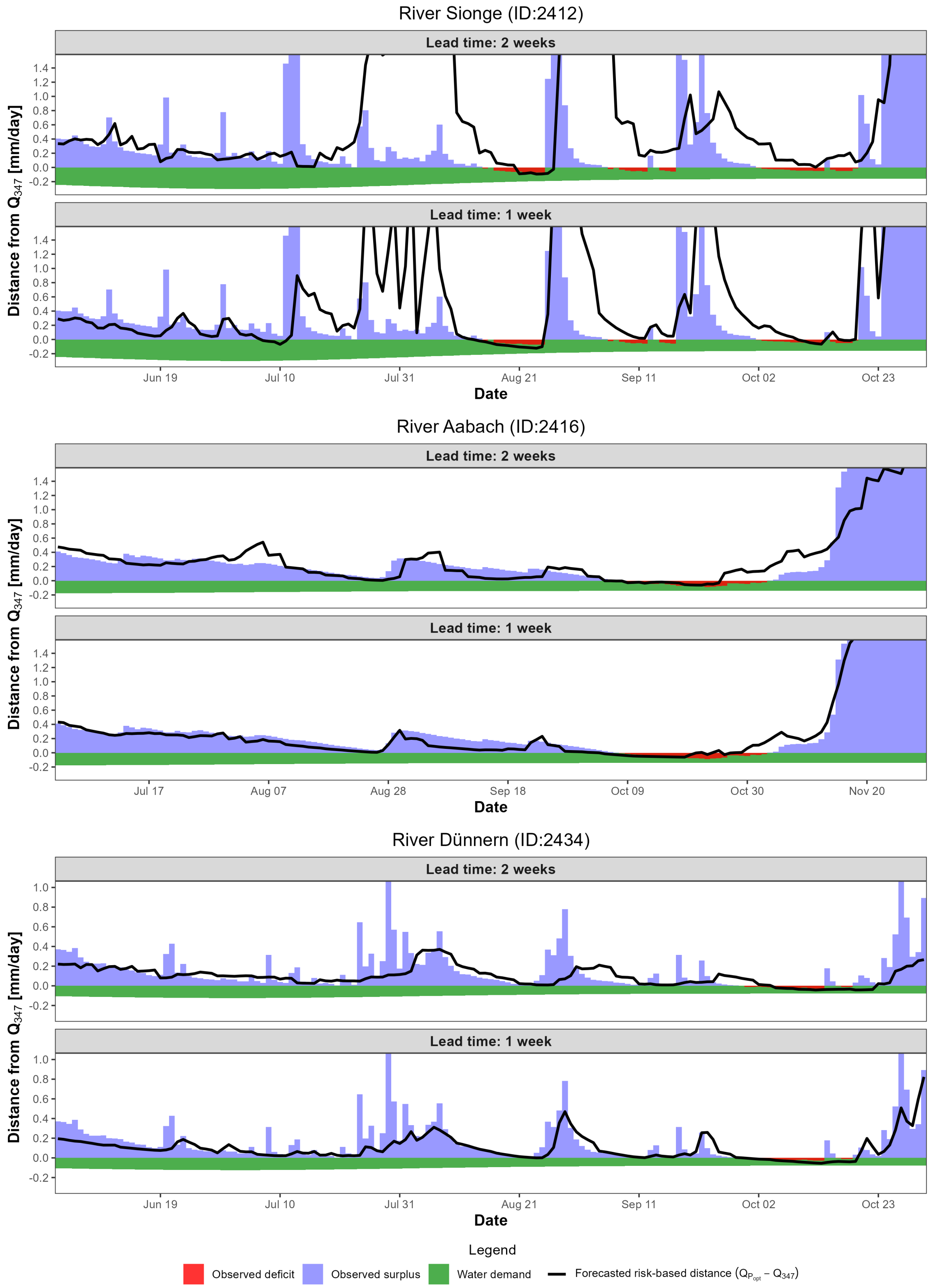


Figure A.5: *Observed and forecasted distance from the $Q_{347}$ threshold for the Sionge, Aabach, and Dünnern rivers in the 2023 drought period.*

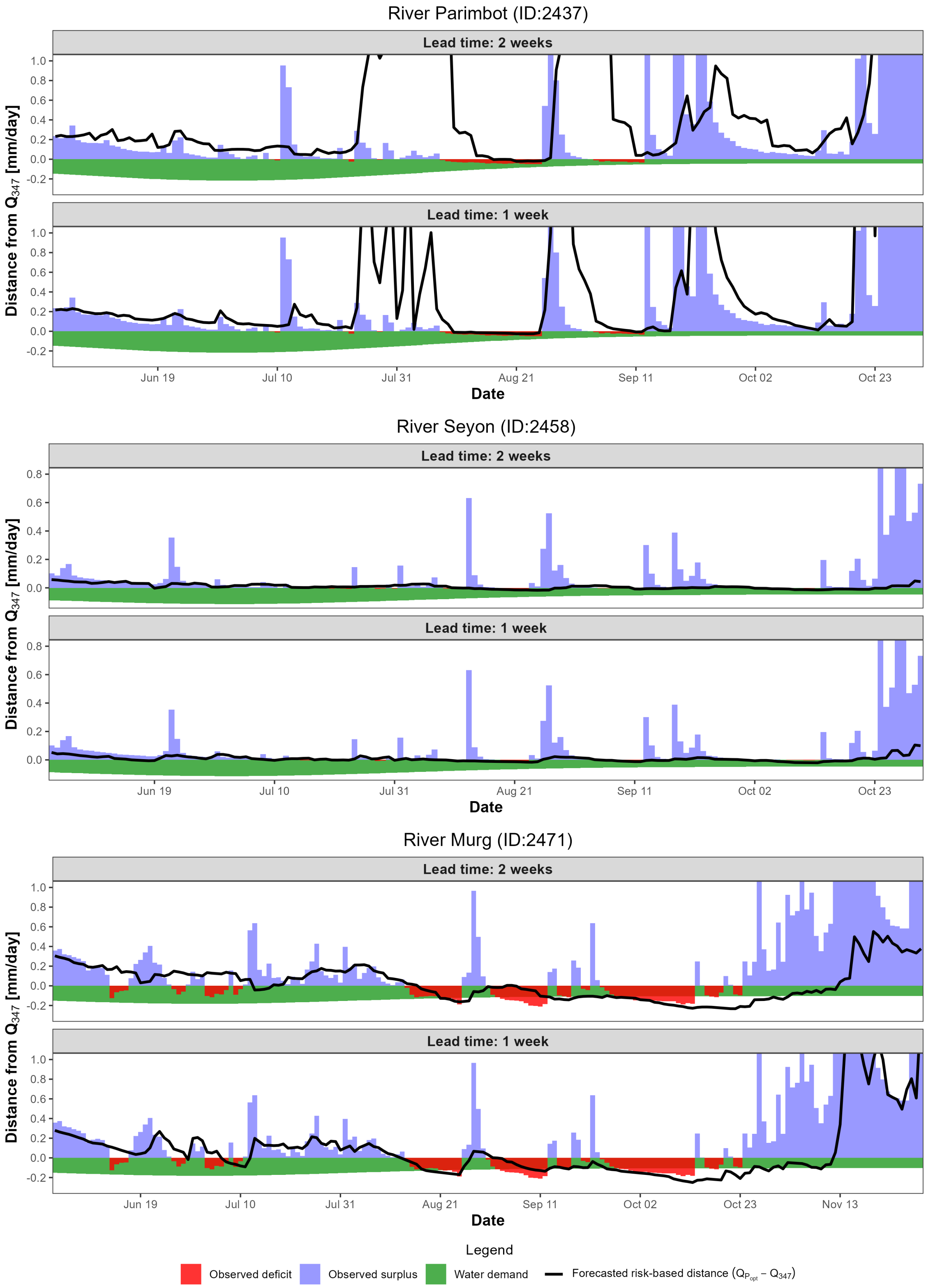


Figure A.6: *Observed and forecasted distance from the $Q_{347}$ threshold for the Parimbot, Seyon, and Murg rivers in the 2023 drought period.*

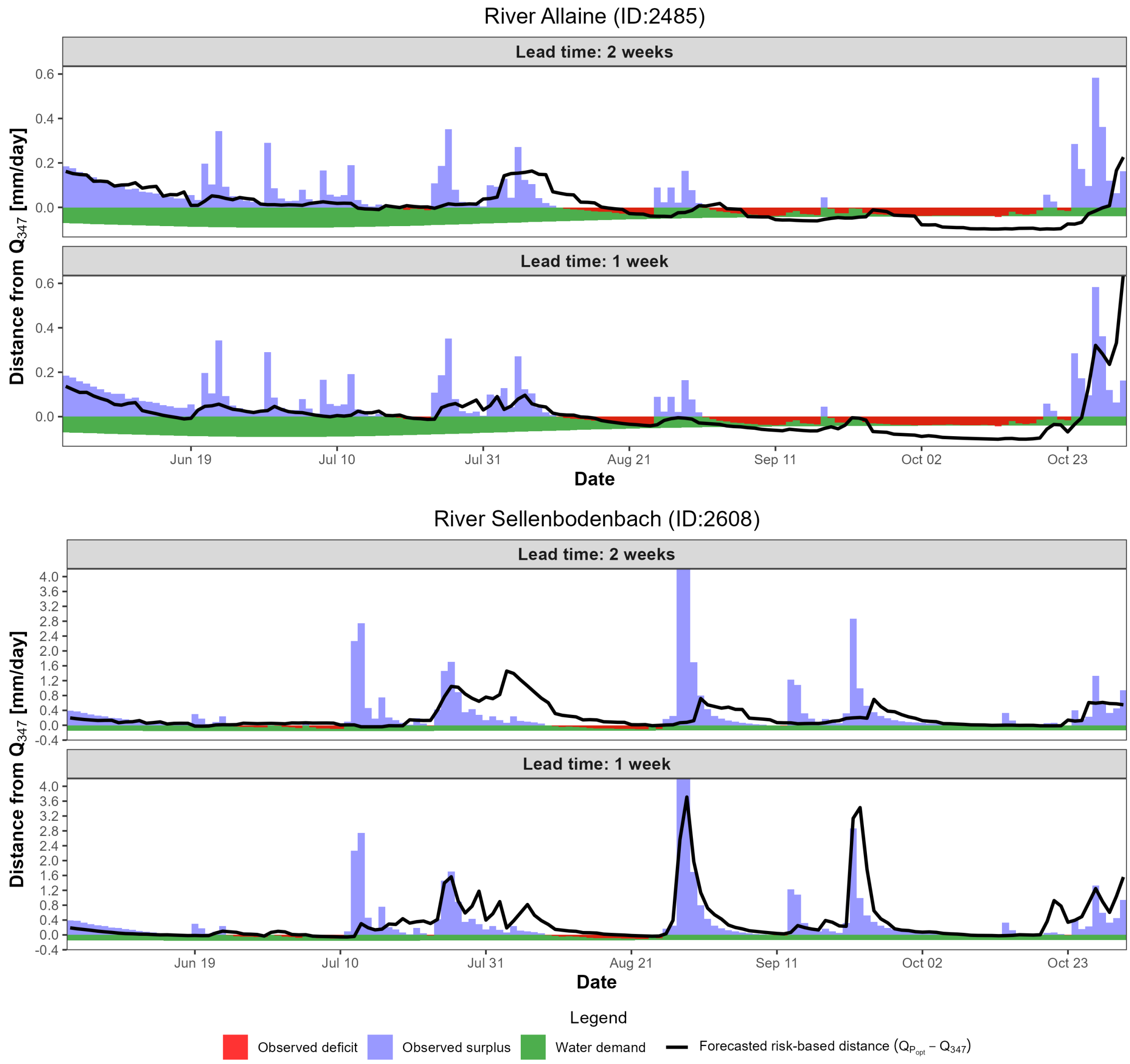


Figure A.7: *Observed and forcasted distance from the $Q_{347}$ threshold for the Allaine and Sellenbodenbach rivers in the 2023 drought period.*

## A.5 Drought probability

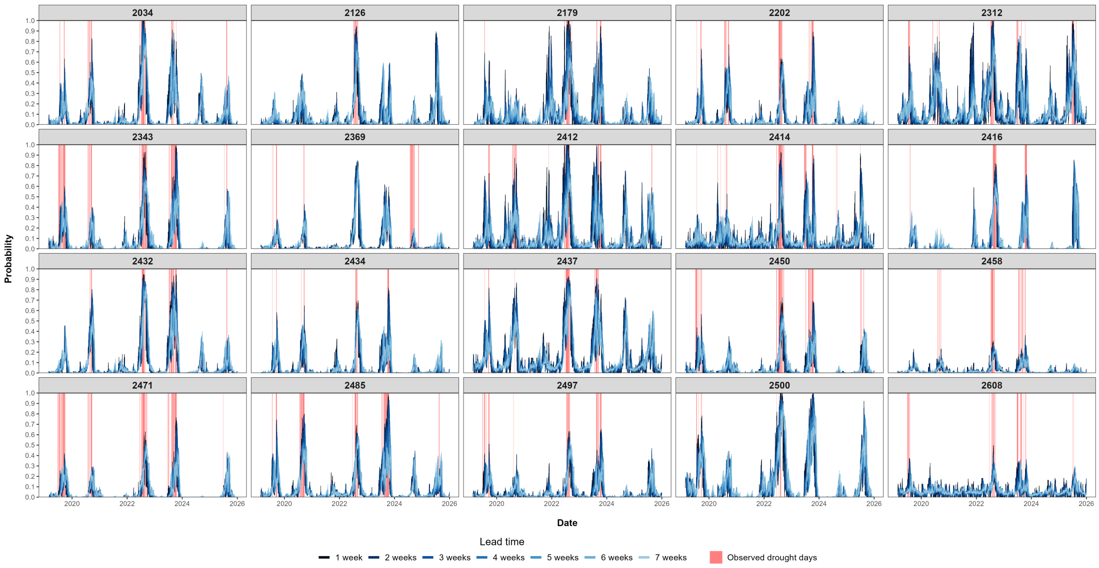


Figure A.8: *Forecasted probabilities for all catchments over the 2019–2025 period, categorized by weekly lead times.*

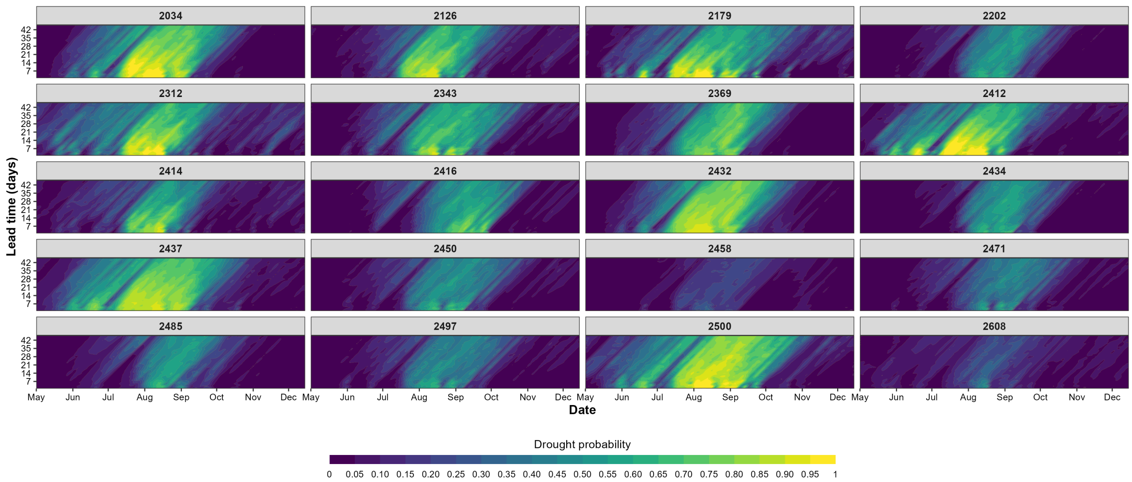


Figure A.9: *Forecasted probabilities for all catchments across different lead times during the 2022 drought period. The x-axis represents the target date, while the y-axis shows the lead time.*

# Acknowledgments

I would like to express my sincere gratitude to Dr. Massimiliano Zappa for giving me the opportunity to carry out this thesis and for providing the essential data for my research. I am also deeply grateful to Dr. Konrad Bogner for his help and the insightful suggestions provided throughout the project. Special thanks to Prof. Dr. Jan Seibert for his availability and for serving as the faculty member representing the University of Zurich. I would also like to thank the entire Mountain Hydrology and Mass Movements research unit at WSL for the time spent together and the supportive environment. Finally, I wish to thank my family for their constant support during my studies.

# Personal declaration

Personal declaration: I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Artificial Intelligence Statement (AI)

AI was used in this thesis:

- as support for writing R code, debugging errors, and assisting in the creation of figures (ChatGPT).
- for translation from Italian to English and for checking vocabulary and grammar (ChatGPT, DeepL)

All content has been personally reviewed and verified. The intellectual contribution remains my own.

Luca Panigada | Zurich, 24.04.2026