hydrology

IMPACT FACTOR 3.1

CITESCORE 4.9

Article

Deep Learning Prediction of Streamflow in Portugal

Rafael Francisco and José Pedro Matos

Special Issue
Editorial Board Members' Collection Series: The Flood Estimation and Forecasting Chain: Meteorological–Hydrological–Hydraulic Forecasts and Predictive Uncertainty towards Operational Decisions

Edited by
Dr. Andrea Petroselli and Prof. Dr. Pingping Luo

MDPI

https://doi.org/10.3390/hydrology11120217 hydrology

MDPI

Article

Deep Learning Prediction of Streamflow in Portugal

Rafael Francisco [1,*] and José Pedro Matos [2]

[1] Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049–001 Lisboa, Portugal
[2] Civil Engineering Research and Innovation for Sustainability (CERIS), Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049–001 Lisboa, Portugal; jose.matos@tecnico.ulisboa.pt
* Correspondence: rafael.francisco@tecnico.ulisboa.pt

Abstract: The transformative potential of deep learning models is felt in many research fields, including hydrology and water resources. This study investigates the effectiveness of the Temporal Fusion Transformer (TFT), a deep neural network architecture for predicting daily streamflow in Portugal, and benchmarks it against the popular Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrological model. Additionally, it evaluates the performance of TFTs through selected forecasting examples. Information is provided about key input variables, including precipitation, temperature, and geomorphological characteristics. The study involved extensive hyperparameter tuning, with over 600 simulations conducted to fine–tune performances and ensure reliable predictions across diverse hydrological conditions. The results showed that TFTs outperformed the HBV model, successfully predicting streamflow in several catchments of distinct characteristics throughout the country. TFTs not only provide trustworthy predictions with associated probabilities of occurrence but also offer considerable advantages over classical forecasting frameworks, i.e., the ability to model complex temporal dependencies and interactions across different inputs or weight features based on their relevance to the target variable. Multiple practical applications can rely on streamflow predictions made with TFT models, such as flood risk management, water resources allocation, and support climate change adaptation measures.

Keywords: streamflow; hydrological prediction; hydrological model; deep learning; temporal fusion transformer; probabilistic prediction; forecasting

check for updates

Citation: Francisco, R.; Matos, J.P. Deep Learning Prediction of Streamflow in Portugal. *Hydrology* **2024**, *11*, 217. https://doi.org/10.3390/hydrology11120217

Academic Editors: Andrea Petroselli and Pingping Luo

Received: 27 November 2024
Revised: 13 December 2024
Accepted: 15 December 2024
Published: 19 December 2024

1. Introduction

Throughout history, rivers have played a significant role in human civilization, being vital for several key activities (e.g., agriculture, commerce, and culture). The United Nations Department of Economic and Social Affairs (UN DESA) and the International River Foundation underline this importance, stating that "Rivers are the lifeblood of the land, people, and economies they support. [. . .] if a river stops flowing, life stops working" [1]. With the global population surpassing 8 billion [2], the exploration of riverine resources has reached unprecedented levels. This, combined with the growing frequency of extreme weather events—such as floods and droughts—has placed significant strain on water management systems worldwide [3], also leading to massive economic investments in reconstruction and adaptation strategies capable of diminishing their effects.

Lee et al. [3] highlighted that over 5300 water—related disasters—WRDs (i.e., floods, storms, landslides, and droughts)—occurred between 2001 and 2018. In this period, over 3.4 billion people were affected, 300,000 died, and over 1.7 trillion USD were estimated in damages worldwide. Floods and droughts account for nearly 60% of WRDs. In 2020, the World Resources Institute claimed that "by 2030, [. . .] 132 million people and $535 billion in urban property will be impacted annually due to riverine flooding" [4]. The heart of the matter is that grave challenges are associated with streamflow's high temporal and spatial variability. Droughts impair energy production, disrupt ecosystems, and trigger famine, disease outbreaks, and economic recessions. Conversely, when water abounds, floods can cause unfortunate damage.

Water–related challenges are only expected to worsen, with pressure scaling from the supply and demand sides. In fact, Alfieri et al. [5] projected that the upsurge in atmospheric temperature will increase the flood risk on a global scale: "At 4 °C global warming, countries representing more than 70% of the global population [. . .] will face increases in flood risk in excess of 500%".

In Portugal, the unfortunate events of December 2022 and January 2023 constitute additional examples of the current challenges. Total damages were estimated at €293 million, also resulting in the death of one person [6].

Accurately predicting streamflow is essential for sound water resources management. Succeeding in doing so may have profound implications for saving human lives, bolstering the economy, and protecting the environment. It may be easier said than done. In reality, the task is inherently complex due to the dynamic and heterogeneous nature of hydrological processes (something the classical hydrological models struggle with), meteorological data imprecision, and streamflow measurement difficulties. Due to these reasons, decision–makers must also develop the capability to act based on imperfect information.

Indeed, this study explores the application of a deep learning model, the Temporal Fusion Transformer (TFT) [7] to predict daily streamflow in Portugal with the respective uncertainty in a set of different experiments. We also compare how TFT predictions fare against a "classical" hydrological model, that is, "Hydrologiska Byråns Vattenbalansavdelning" (HBV), resorting to the implementation of the RS Minerve software [8]. The results presented herein can contribute toward a new path for hydrological prediction (HP), where such applications are still rare (e.g., [9]).

The article is organized as follows: in Section 2, theoretical foundations and related studies are specified. Afterwards, in Section 3, the methodology is presented, providing insight into the hydrological and meteorological data that are used, the explored predictive models, and the metrics chosen for performance assessment. Section 4 is dedicated to the description of the case study—unregulated catchments in Portugal. Results and discussion are presented in Section 5. Finally, conclusions are drawn in Section 6.

2. Theoretical Framework and Related Studies

State–of–the–art studies have strengthened the scientific community's belief in the exacerbation of all the referred impacts. For example, countries like India [10] have been shown to have a high susceptibility to floods, while South Africa, Botswana, Zimbabwe, and Mozambique are particularly affected by flood and drought events [11]. Infamous flood events have also recently been recorded in Europe [12] and North America [13]. In May 2024, the late floods in Rio Grande do Sul (Brazil) almost totally cut off this city with 1.3 million inhabitants, causing a total of 95 confirmed deaths, leaving at least 130 missing, and leaving 80% of the population without access to drinkable water [14]. Clearly, water–related disasters are a global issue.

Since the 1850s, hydrological models have continuously been explored and improved [15], in particular, to predict hydraulic variables (e.g., streamflow) using measured inputs (e.g., precipitation and temperature or potential evapotranspiration). HP is essential for issuing alerts and evacuations before floods and constitutes a vital tool for managing reservoirs in tasks of flood lamination, hydropower production, irrigation, water supply, and others. Incorporating uncertainty in HP is a technically demanding task, but it is crucial to guarantee a reasonable analysis of the specific problems faced by decision–makers, especially when predictions are most challenging.

On the one hand, deterministic predictions, which offer a single–point estimate, usually fail to account for risk and uncertainty, limiting their utility for decision–making in critical situations (Figure 1a). On the other hand, probabilistic predictions, which provide a range of possible outcomes with different probabilities, offer a more robust framework for flood prediction and water resource management (Figure 1b).

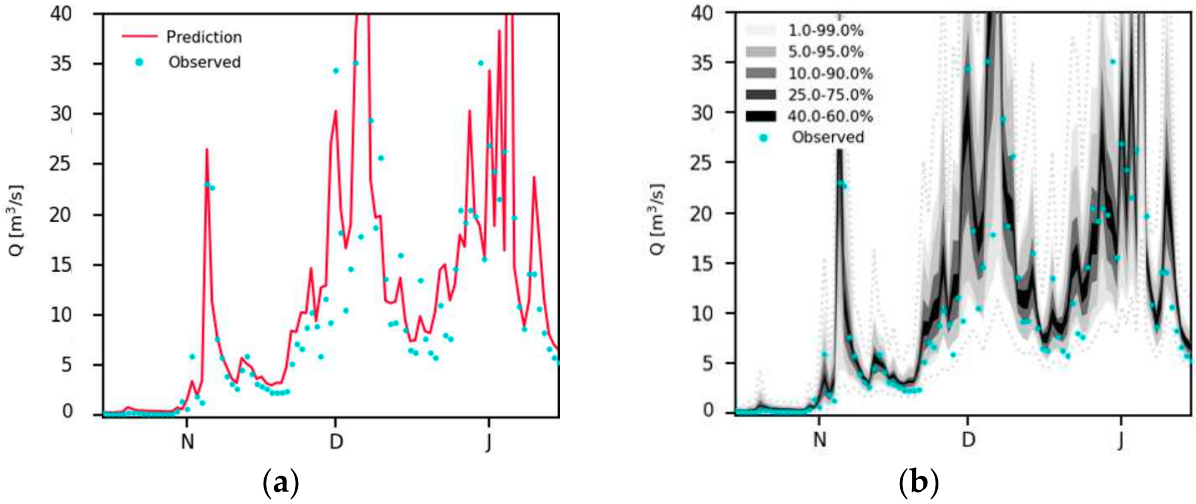


Figure 1. Examples of (**a**) deterministic and (**b**) probabilistic predictions for streamflow (Q). The months of the year are indicated along the horizontal axis. Bands indicate the percentage of observations expected to fall within their bounds.

Since its appearance in the 1950s [16], artificial intelligence (AI) and its sub–field of machine learning (ML) strove to enhance human productivity by automating complex tasks, achieving remarkable performance in different fields such as speech recognition [17], healthcare and prevention of diseases [18], autonomous vehicles [19], the energy sector [20], civil engineering [21–24], and many others.

Deep learning (DL) is a subset of ML that uses artificial neural networks (ANNs) to emulate behaviors similar to those of a human neural system [25]. In recent years, advancements in DL enabled the ongoing explosion in terms of large–language model (LLM) capabilities, including their application in hydrology (e.g., [26]). Noteworthy advancements in ANNs [27] include sophisticated architectures such as recurrent neural networks (RNNs) or Long Short–Term Memory (LSTM) networks [28–30]. One of the most recent innovations in this area is the "transformer" architecture [31], which has, alongside LSTMs, a central component of the TFT model [7] that is used in this study.

3. Methodology

3.1. Overview

Defining a methodology that balances data quality and availability, reasonable training and validation periods, appropriate model complexity, and coherent performance evaluation is essential in ML applications. Accordingly, and including the HBV model, the three main stages of this study are (Figure 2):

(1) Data collection, validation, and feature engineering: collecting and validating data, such as precipitation, temperature, and streamflow records, and performing feature engineering allows the creation of meaningful input variables to feed the model.

(2) Model development: whose main goal is to find a robust TFT model that can accurately replicate observed streamflow. This iterative process involves adjusting hyperparameters (external configuration variables used to manage model training such as dropout rate) such as learning rate, batch size, and dropout rate within the TFT, as well as defining model training and validation (with an emphasis on preventing overfitting). The deterministic models, the HBV, were also calibrated in this phase.

(3) Performance assessment: after the prediction task is complete, graphical methods (e.g., time series plots) and performance metrics particularly well–suited for hydrological applications are employed to assess the accuracy of the models. Eventually, if the obtained accuracy is not desired, the process returns to the model development phase.

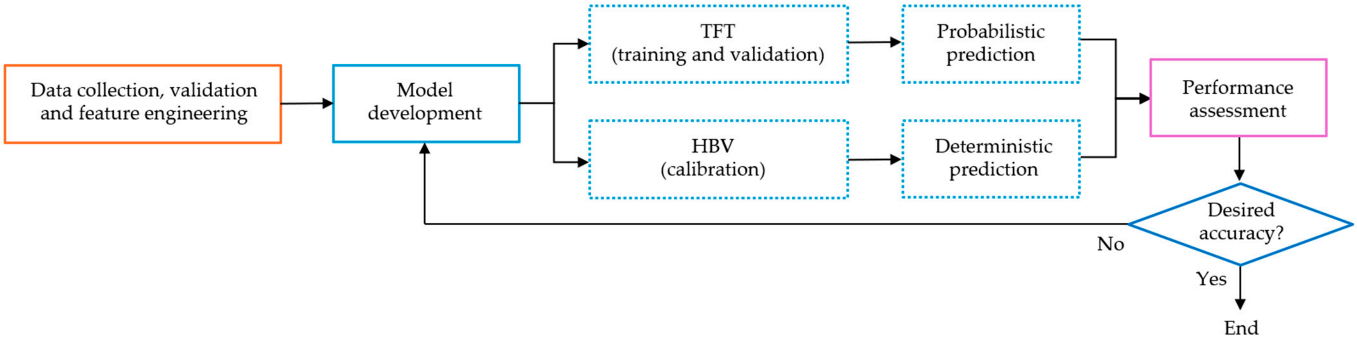


Figure 2. Work chain for model development.

Four representative catchments chosen from a broad set of unregulated catchments in Portugal were selected to dive into the capabilities of the TFT. In a first step, the model was applied to each catchment separately, generating individual predictions from meteorological and geomorphological data. In a second step, the full set of catchments was used to train and validate the model, producing new predictions. The main goal was to assess the model's capacity to incorporate hydrological data from all over the country to improve upon the initial results obtained with catchment–specific models. Finally, in a forecasting setup, observed streamflow values were also presented as input data. Two meteorological data sources were used: ERA5–Land (in the prediction tasks) and Global Forecasting System (GFS) (used in forecasting). More details on the different experiments that were undertaken can be found in Section 4.3.

3.2. Data Collection, Validation, and Feature Engineering

3.2.1. Hydrological Data

Hydrological data were retrieved from the Portuguese National Water Resources Information System (SNIRH) [32] as a series of average daily streamflow collected at monitored catchments. Data collection covered the period from 1980 to the present. A qualitative data analysis where all the selected time series were individually inspected was undertaken. Special attention was given to identifying and removing anomalous values inconsistent with plausible streamflow patterns. The main preoccupation was not to "pollute" the model with questionable values. Accordingly, in case of doubt, it was preferred to remove those values from the data set. Several criteria were adopted for exclusion, namely:

(i) Abrupt variations in streamflow on consecutive days, inconsistent with the patterns of the remainder of each series.
(ii) Consecutive years of streamflow records with zero values, suggesting faulty records in a considerable number of years.
(iii) Offsets in streamflow records, indicating potential calibration issues or other problems with the data.

Finally, data exclusions targeted whole periods. In other words, when problems were identified, whole hydrological years (and often larger periods) were removed from the series.

3.2.2. Meteorological Data

For model development purposes, precipitation and temperature were collected from the ERA5–Land reanalysis dataset [33], produced by the European Centre for Medium–Range Weather Forecasts (ECMWF) with global coverage (overland) in a $0.1 \times 0.1^{\circ}$ grid. Hourly data were downloaded from 1 January 1980 to 31 December 2022 for the whole Iberian Peninsula. The downloaded data were first aggregated for each catchment and then resampled to the daily scale after conversion to local time (UTC to Lisbon time, including daylight saving time shifts). Temperatures were resampled to the daily mean and precipitation to the daily sum. These data (and additional variables—see Section 3.2) were used to train and validate the models for streamflow prediction. No quality control measures were applied to the precipitation and temperature data, as previous studies have already demonstrated the relevance and reliability of ERA5–Land over the Iberian Peninsula and Portugal (e.g., [34,35]).

Additionally, the potential evapotranspiration (ET_0 in mm/day) was used as an input variable to the HBV model. ET_0 was calculated using the Hargreaves approach [36], following Equations (1) and (2).

$$ET_0 = 0.0135\left(\overline{T}_d + 17.78\right)R_s \tag{1}$$

$$R_s = R_a K_T \sqrt{T_{dmax} - T_{dmin}} \tag{2}$$

where $\overline{T}_d$ is the daily mean temperature (°C), R_s the incident solar radiation (mm/day), R_a the extraterrestrial solar radiation (mm/day) (calculated based on latitude and day of year), K_T an empirical constant, which depends on the region conditions (assumed equal to 1), T_{dmax} the maximum daily temperature from ERA5-Land (°C), and T_{dmin} the minimum daily temperature also from ERA5–Land (°C).

Forecasting attempts resorted to meteorological data from the Global Forecasting System (GFS) from the United States National Centers for Environmental Prediction (NCEP), which is a publicly available global weather forecasting model [37]. GFS forecasts with a three–hourly time discretization and a spatial grid of 0.25 × 0.25° were downloaded. These corresponded to the 00:00 UTC production cycle (additional runs are undertaken at 06:00, 12:00, and 18:00). Temperature and precipitation were pre–processed for the available area in the study from 1 January 2022 onwards. After spatial aggregation, they were resampled for 1h: temperature linearly and precipitation in 3 equal values for each 3–hourly block. Similarly to what was performed for ERA5–Land, the time zone was updated to Lisbon time, and the data aggregated was updated to daily.

Training and validation of the TFT model were performed with data from ERA5–Land, a decision due to both the short period of time during which GFS data are available and the comparatively greater homogeneity of the former. To make the GFS data more suitable to the trained TFT models and mitigate potential change of support problems, corrections were applied. Precipitation (P_{GFS}) and temperature forecasts (T_{GFS}) were thus transformed by applying Equations (3) and (4), respectively:

$$P^*_{GFS} = P_{GFS} \times \frac{\overline{P}_{ERA5}}{\overline{P}_{GFS}} \tag{3}$$

$$T^*_{GFS} = \left(T_{GFS} - \overline{T}_{GFS}\right) \times \frac{s_{ERA5}}{s_{GFS}} + \overline{T}_{ERA5} \tag{4}$$

where P^*_{GFS} is the transformed value of each precipitation forecast from GFS. Additionally, $\overline{P}_{ERA5}$ represents the average of the precipitation records from ERA5–Land, and $\overline{P}_{GFS}$ represents the average of the precipitation forecasts from GFS. T^*_{GFS}, $\overline{T}_{GFS}$, and $\overline{T}_{ERA5}$ represent the same variables as mentioned but for temperature. s_{ERA5} symbolizes the standard deviation of temperature records from ERA5–Land and s_{GFS} the standard deviation of temperature forecasts from GFS.

3.3. Model Development

3.3.1. Hydrologiska Byråns Vattenbalansavdelning

The HBV model is a conceptual hydrological model designed to simulate runoff in river catchments, developed by the Swedish Meteorological and Hydrological Institute (SMHI) [38]. Although initially tailored for application in Scandinavia, the HBV model quickly gained widespread adoption for hydrological modeling across the globe [39]. The model uses a lumped–parameter approach, representing the catchment as a single and uniform unit rather than accounting for spatial variations within the area. The RS–

Minerve software (v. 2.9.1.0) was used to carry out the simulations for this model, primarily developed by the Swiss Alpine Environment Research Center (CREALP) in collaboration with HydroCosmos SA [8]. The SCE–UA optimization algorithm, commonly used to calibrate hydrological models, was employed to find optimal parameters for each one of the considered catchments.

3.3.2. Temporal Fusion Transformer

The TFT is a class of deep neural networks designed to handle complex and multivariate data. It is a promising choice for prediction tasks based on observed trends [7]. It builds upon the "transformer" architecture by introducing additional features that improve performance, especially for time series prediction. Standard "transformers" rely on self–attention mechanisms, which can capture relationships between different time steps in sequential data [31]. Additionally to self–attention, the TFT resorts to gating mechanisms that help control the information flow, sample–dependent variable selection for focusing on the most relevant inputs, and static covariate encoders that handle time–invariant features, allowing the model to capture better short– and long–term dependencies in time series data [7]. Among such gating mechanisms are LSTMs, notorious for their increasing use in hydrological applications.

Modeling was carried out in Python 3.9.18 (December 2023) and relied heavily on the PyTorch library and its forecasting package (Pytorch Forecasting [40]). PyTorch Forecasting facilitates the use of state–of–the–art time series forecasting with neural networks, including building, training, and evaluating the performance of TFTs. Eclipse (with the PyDev extension) was used as the integrated development environment (IDE). A set of initial and simple trials with synthetic series (e.g., linear, sinusoidal, sinusoidal with heteroscedastic noise, sinusoidal with heteroscedastic noise and an exogenous variable, and autocorrelation) was conducted to understand the capabilities of TFTs and how different inputs can affect the predictions. In fact, some interesting conclusions led to the consideration of additional inputs as further explained in Section 4.2.

With respect to the iterative nature of the hyperparameter tuning, the process of their definition started with a trial–and–error approach in order to understand how models of different complexities (e.g., number of hidden neurons, attention heads, and LSTM layers) fared at predicting discharges. Initial values for hyperparameters (e.g., learning rate, dropout rate, network size) were selected based on the prior literature and domain knowledge. Subsequent adjustments were made based on performance metrics. This process was repeated until a choice of parameters was found that reliably yielded good validation and test metrics. More information on this is provided in Section 5.1.

For each experiment, the selection of training, validation, and test periods was defined as a fraction of the available data. More precisely:

Training period: 60% of the available data.
Validation period: 20% of the available data.
Test period: 20% of the available data.

The selection of the available data was random and accounted for the hydrological year in Portugal (from 1 October of a given year to 30 September of the following year). During preliminary testing and model development, numerous TFTs were trained on three machines with CUDA–capable Graphics Processing Units (GPUs).

3.4. Performance Assessment

3.4.1. Deterministic Predictions

The Nash–Sutcliffe efficiency (NSE) [41] is often used in hydrological applications to assess and quantify the quality of deterministic predictions. It is scale–invariant and valuable for evaluating the performance of predictive models. Its optimal value is 1 (unitless) and increasing deviations between the observed and simulated values diminish the value of the NSE. It can be calculated through Equation (5):

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(\hat{Q}_i - Q_i\right)^2}{\sum_{i=1}^{N}\left(\hat{Q}_i - \overline{Q}\right)^2} \tag{5}$$

where i is the time step, N is the total number of time steps considered, $\hat{Q}_i$ is the computed streamflow at the ith time step, Q_i is the observed streamflow at the ith time step, and $\overline{Q}$ is the mean observed streamflow.

3.4.2. Probabilistic Predictions

The Continuous Ranked Probability Score ($CRPS$) [42] measures how well a prediction matches the observed outputs—Equation (6)—and is particularly useful to quantify the performance of probabilistic or ensemble predictions. A value equal to 0 indicates an accurate prediction, while greater values suggest more significant differences between the probabilistic distribution of predictions and observations. Its units are the same as the variable that is being evaluated. CRPS has the particularly interesting feature of allowing the direct comparison of probabilistic and deterministic predictions, as in the case of the latter, it converges to the mean absolute error (MAE).

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left[\hat{F}_i(x) - \mathbb{H}(Q_i - x)\right]^2 dx \tag{6}$$

In the above equation and additionally to the description of the previous variables, $\hat{F}_i$ is the value of the cumulative distribution function of predicted streamflow at the ith time step, and $\mathbb{H}$ is the Heaviside step function.

A predictive distribution can be called reliable if the observations fall within the proposed distribution with matching probability. Predictive quantile–quantile (QQ) plots are valuable for assessing reliability (α)—Equation (7). They are based on the quantile distribution of the observed and predicted variables [43]. They consist of the complementary area between the obtained QQ plot for the prediction and a perfect diagonal (Figure 3). Therefore, values equal to 1 are desired (reliability is unitless). Deviations from this straight diagonal reflect different types of shortcomings in predictions [44]. Extreme deviations correspond to values closer to 0.

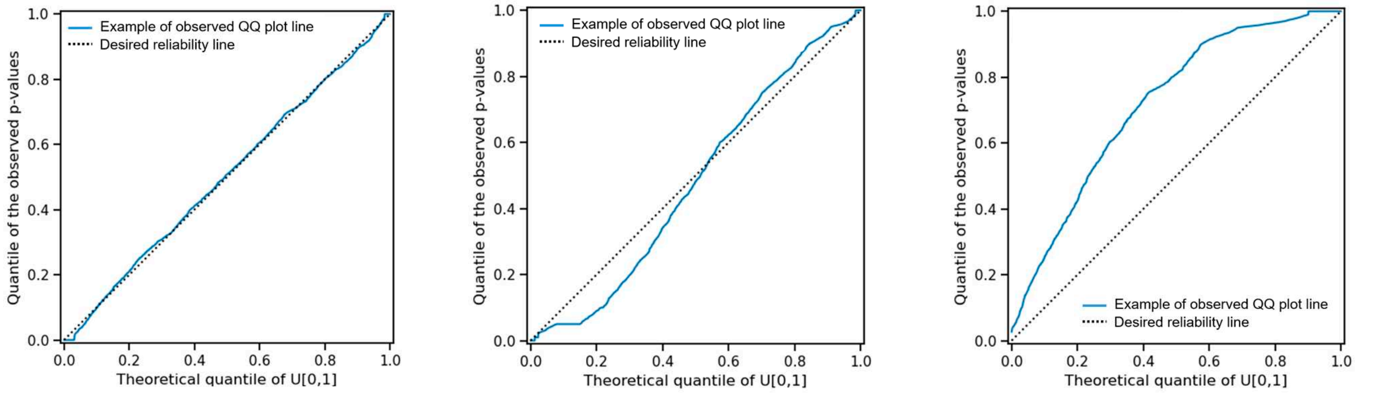
$$\alpha = 1 - \frac{2}{N}\sum_{i=1}^{N}\left|p_i - p_i^{th}\right| \tag{7}$$

Here, p_i is the p–value of the ith ranked prediction, and p_i^{th} is the theoretical p–value of the ith ranked prediction.

The relative resolution evaluates the predictive distribution's precision (or sharpness)—Equation (8) [44]. A prediction is said to be reliable when the predicted values are inserted in the uncertainty bands. The lower the variation between observed and predicted values, the greater the resolution. Figure 4 depicts an example of two different resolutions.

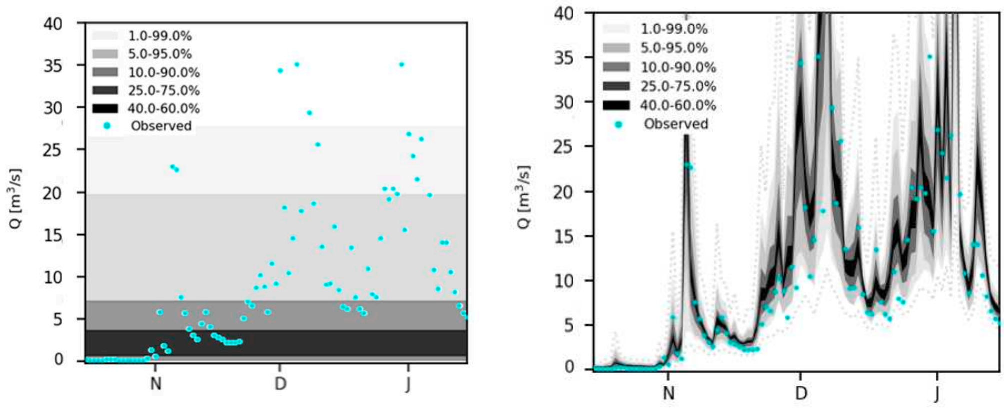
$$\pi^{rel} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{Q}_i}{S_{\hat{Q}_i}} \tag{8}$$

In the above equation, $S_{\hat{Q}_i}$ is the standard deviation of the predicted streamflow at the ith time step.



(**a**) (**b**) (**c**)

Figure 3. Examples of predictive QQ plots, more precisely of (**a**) an accurate prediction, (**b**) an over–prediction for lower theoretical *p*–values and under–prediction for higher theoretical *p*–values, and (**c**) an under–prediction.



(**a**) (**b**)

Figure 4. Example of a probabilistic prediction with different resolutions: (**a**) low resolution and (**b**) high resolution. The months of the year are indicated along the horizontal axis. Bands indicate the percentage of observations expected to fall within their bounds.

4. Case Study

4.1. Geographical and Hydrometeorological Context

Portugal is located on the Iberian Peninsula, in southwestern Europe. The country, which extends to the archipelagos of Madeira and Azores, in the North Atlantic, covers a total area of approximately 92,000 km^2 [45]. It is roughly bounded by 37° N and 42° N latitudes and 7° W and 9.5° W longitudes (WGS84 datum). Throughout this study, only Mainland Portugal is referred to, not including the Islands.

Figure 5 portrays the mean daily temperature (in °C) registered in Mainland Portugal from 1980 to 2022 (derived from hourly ERA5–Land data) and the annual average precipitation (in mm/year) for the same period, also derived from ERA5–Land. This visualization provides a better understanding of the meteorological conditions under which the models were applied, noticing that the north comprises wetter regions, while the south tends to be drier.

Figure 6 depicts the digital elevation model (DEM) [46] of Mainland Portugal and its main rivers (i.e., Douro, Tagus, and Guadiana) and hydrographic regions [47,48]. The northern region of the country is mainly characterized by rugged terrain.

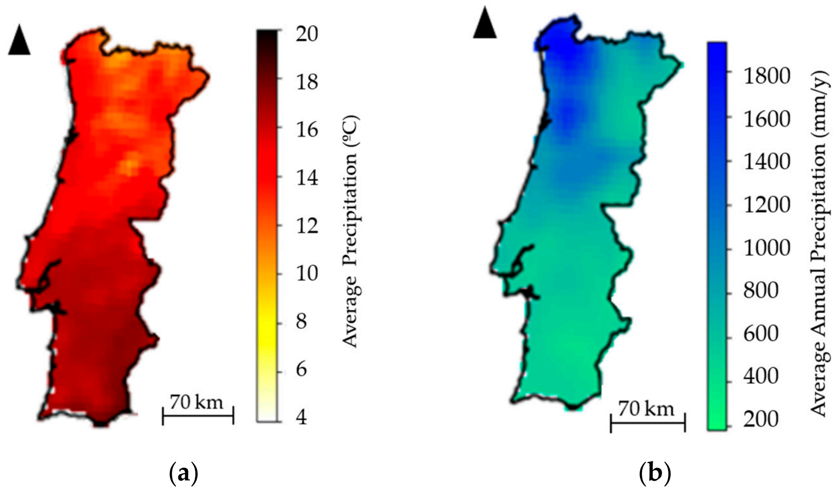


Figure 5. Variation of meteorological data in Mainland Portugal between 1980 and 2022, based on data retrieved from ERA5–Land, namely, (**a**) the daily average temperature and (**b**) the annual average precipitation.

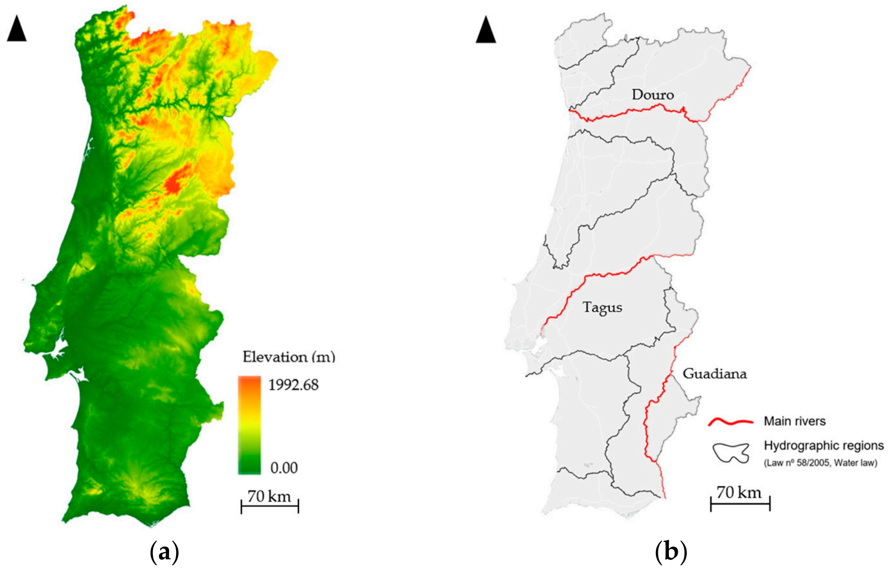


Figure 6. Geomorphological information about Mainland Portugal, specifically, (**a**) the digital elevation model and (**b**) the main hydrographic regions and rivers.

4.2. Additional Data and Feature Engineering

The available streamflow data correspond to 216 catchments monitored by a different hydrometric station in Mainland Portugal. In order to validate the generated information, ArcGIS Pro (v. 3.1 developed by Esri Inc., Redlands, CA, USA) [49] was used. The area calculated for each catchment was compared with the respective official value defined within SNIRH. All the catchments differing by more than 10% were judged to have problems and were therefore eliminated, which resulted in a revised total of 191. This threshold is meant to rule out problems associated with the derivation of each catchment from the DEM. The reasoning was that excluding some catchments is not critical to the process or the results, whereas including ill–derived catchments could lead to errors. Following this, all the catchments with a large part of their area in Spain were also removed, as well as those located downstream of dams and reservoirs, based on the extensive information provided by The Global Dam Tracker [50]. At the end of the process, a subset of 91 unregulated catchments remained.

Working with unregulated catchments is important to guarantee that no streamflow records are heavily influenced by man–made decisions that may be difficult to predict or reproduce. After a qualitative data analysis, anomalous values inconsistent with usual streamflow patterns were identified and removed, as referred to in Section 3.2.1. In total, 17 stations were judged problematic or not having enough data for training and/or validation (3 years at least of observed values) and were removed from the set, equaling the final 74 monitored catchments with useful data to "feed" the model. Figure 7a displays the distribution of the final catchments along the country, as well as dams/reservoirs and other eliminated streamflow gages. This process was complex and crucial to guarantee that the training and validation data accurately mirror the reality of these phenomena.

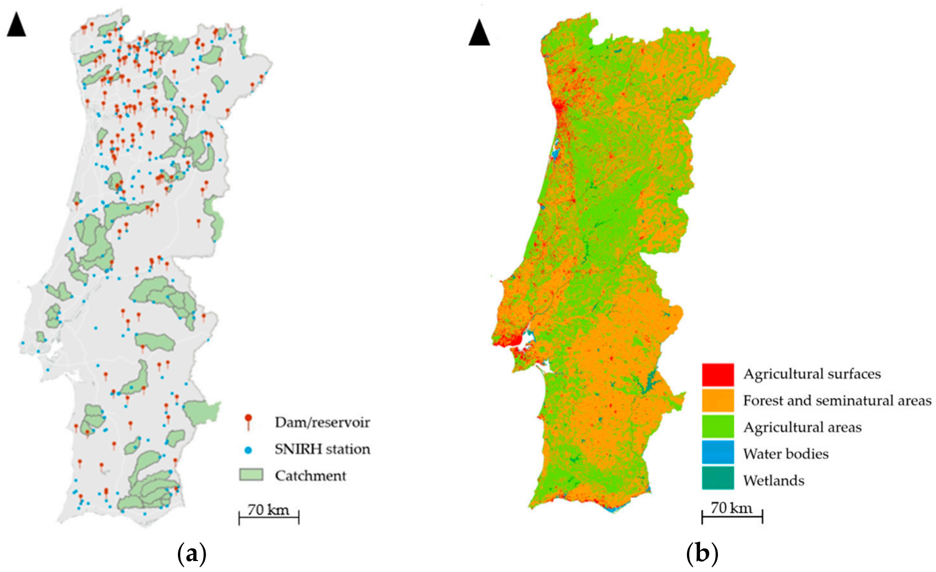


Figure 7. Representation of (**a**) the 74 catchments in the study, dams, and SNIRH stations, and (**b**) the main classes of land cover in Mainland Portugal.

Five variables were calculated and used as static inputs associated with each catchment:

- Drainage area (in km^2).
- Centroid's coordinates (in WGS84 datum).
- Gravelius compactness index (*GC*). It evaluates the resemblance of the shape of the catchment to a perfect circle (where *GC* = 1.0). *GC* increases as the shape distances itself from a circle. Equation (9) reveals how to obtain it:

$$GC = \frac{P_w}{2\sqrt{\pi A}} \quad (9)$$

  where P_w is the smoothed perimeter of the catchment.
- Mean elevation (in m.a.s.l.) was based on the information provided by the GLO–30 DEM, from Copernicus, with a 30 × 30 m spatial resolution [46].
- Land use (as an integer code) was introduced based on the information provided by the Copernicus CORINE data set [51], choosing the dominating land use class for each catchment. Figure 7b portrays the main classes of land use, proving the dominance of forest and seminatural areas (in orange) and agricultural areas (in green) in the selected catchments.

Following initial trials, it was decided to define the target variable to be predicted as not just streamflow, but specific streamflow—Equation (10). This decision was taken to facilitate the comparison between data from different catchments of different sizes. Additionally, this normalization avoids that the bigger catchments have an unwarranted larger weight on the training owing to larger streamflow, enabling the prediction for catchments with very different scales.

$$q = \frac{Q}{A} \times 10^3 \tag{10}$$

Here, q is the specific streamflow (m^3/s/km^2 $\times$ 10^3), Q is the observed streamflow (m^3/s), and A the catchment's drainage area (km^2).

4.3. Experiment Definition

This section aims to define concisely what experiments were performed. Each experiment was run using a trained TFT with a certain combination of hyperparameters, and a satisfactory achieved result. They were designed to evaluate the potential of TFT models to enhance hydrological predictions as follows:

(1) Experiment P1: This experiment is divided into four different "micro tests", where the model is trained for a single catchment. This setup simulates localized prediction tasks where models are tailored for individual catchments. The results reveal insights into the feasibility of using TFTs in situations where hydrological models are usually employed.

(2) Experiment P2: This experiment aims to use the information of all the 74 selected catchments for model training and validation simultaneously. The main idea is to assess whether the TFT can effectively gain general knowledge about hydrological processes in Portugal to improve upon predictions "specialized" for specific locations, in this case, the predictions from P1. This approach can be particularly relevant for managing interconnected catchments, where understanding regional patterns can improve predictions and lead to better resource allocation strategies.

(3) Experiment P3: Building on P2, this experiment includes past observed streamflow as training data, showcasing a near–operational application of TFTs. Although the meteorological data are from ERA5–Land rather than forecasts (the reason why we denominated P3 as "pseudo" forecasting), the experiment simulates real–world forecasting scenarios and allows the direct comparison with the other experiments.

(4) Comparison with the HBV model: To contrast with the performance of TFTs, an additional prediction was made for the control catchments with a calibrated HBV model. It demonstrates the practical utility of TFTs in possibly replacing traditional hydrological models, widely used in catchment management.

(5) Forecasting experiment using GFS meteorological data as input, instead of data from ERA5–Land. The used model is the same as the one from P3.

The four selected stations to test the central hypothesis of this research—that the TFT model can outperform streamflow predictions from the HBV model—are strategically located across the country, ensuring a diverse representation of regional hydrological conditions (Figure 8):

(i) 05K/01H lies in a small catchment in a relatively wet region and is isolated from other catchments.

(ii) 16G/01H is in a relatively large catchment in a moderately wet region and is involved with other catchments.

(iii) 24H/03H is moderately sized, lies in a relatively dry zone, and is isolated from other catchments.

(iv) 28L/02H is moderately sized, located in a relatively dry zone, and is involved with other catchments.

Table 1 shares the most significant information for each, such as their identification (ID), latitude (Lat.), and longitude (Long.) of its centroid, drainage area (A), average streamflow ($\overline{Q}$) average temperature ($\overline{T}$), and the average annual precipitation ($\overline{P}$). Table 1 depicts the geographical position and relative dimensions of the four catchments.

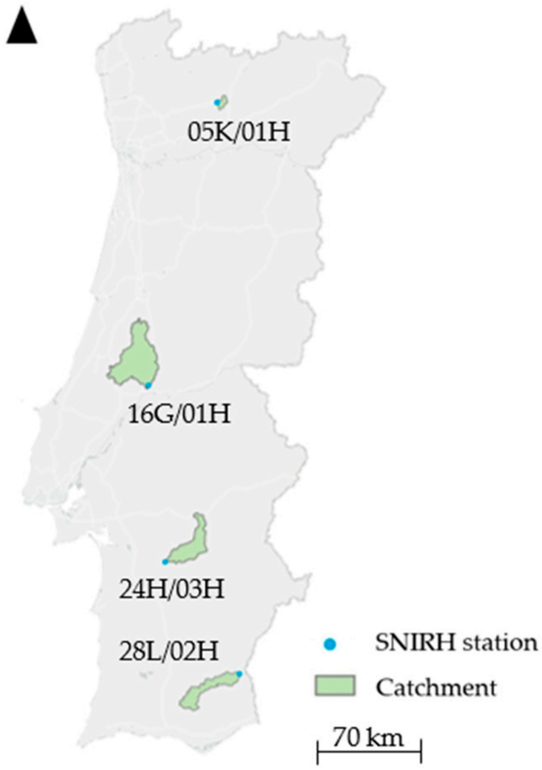


Figure 8. Representation of the four catchments used to compare the performance of each experiment and the HBV model. The comparison with Figure 7a can be beneficial to understand the relative position of these four comparative to the other catchments.

Table 1. Significant information of the considered stations.

Station	SNIRH ID	Lat. (°)	Long. (°)	A (km^2)	$\overline{Q}$ (m^3/s)	$\overline{P}$ (mm/y)	$\overline{T}$ (°C)
Santa Marta do Alvão	05K/01H	41.498	−7.754	48.76	1.4	1058.1	11.8
Fábrica da Matrena	16G/01H	39.532	−8.379	1047.15	7.3	797.2	15.0
Torrão do Alentejo	24H/03H	38.299	−8.229	468.35	1.6	539.0	16.8
Vascão	28L/02H	37.520	−7.579	409.89	1.7	460.2	16.8

5. Results and Discussion

5.1. Fine–Tuning of TFT Hyperparameters

Each of the experiments (P1, P2, and P3) involved different model configurations and datasets. P1, the simplest experiment, used data from one station to train each TFT, while subsequent models, from P2 to P3, incorporated data from multiple locations. This increased the model complexity and respective training times, with P1 taking an average of 40 min, P2 almost 5 h, and P3 over 6 h.

As outlined in Section 3.3.2, this study entailed a detailed exploration of hyperparameter combinations to optimize the performance of the TFT model, comprising more than 600 simulations and over 400 h of computational processing. Key parameters were systematically tested to understand their impact on the model's accuracy. A trial–and–error approach was employed, highlighting the iterative nature of the tuning process. Through numerous simulations, an understanding of which configurations resulted in better or poorer performance emerged. This process was crucial for ensuring reliable predictions across diverse hydrological conditions. The main conclusions obtained from the analysis were:

(1) For the simplest experiment (P1), the number of nodes in hidden layers (hidden_size parameter) equal to 45 had the best performance. As the complexity of the relations increased in the following experiments (P2 and P3), a value of 60 for this parameter produced more accurate predictions. Values above the referred tended to cause model overfitting, that is, the model replicated well the training data set but at the cost of producing poorer predictions in the test data set. Lower values did not allow the model to reproduce the patterns between the inputs and the streamflow, generally leading to poorer training and performance metrics.
(2) A minimum dropout rate (partial removal of information during part of training) is important to prevent overfitting. A high value is not desirable, at the risk of removing information essential to the model learning process. Fractions between 0.0 and 0.4 were tested. Values near 0.15 led to the best model performance. Results were less sensitive to dropout than to hidden layer size.
(3) Batch size is important. Lower batch sizes showed the ability to reduce the uncertainty of predictions, especially in higher values of quantiles. Finally, values in the range of 64 to 256 were used. Larger batches can lead to faster training but hinder model convergence.
(4) The learning rate plays a crucial role in model development. High learning rates can lead to overshooting of optimal weight configurations, while low learning rates may slow down the training process and cause the training procedure to stall at local minima. A set of values between 0.0001 and 0.01 was tested. In addition, 0.0001 showed adequacy to P1 and 0.0002 to P2 and P3.

Table 2 aggregates the chosen hyperparameters for each experiment, from P1 to P3 (for more information about the meaning of each hyperparameter, please consult PyTorch Forecasting and PyTorch Lightning documentations [52,53]), as well as additional information about the experiments with TFT models. The hyperparameters defined for each experiment correspond to the best combination identified for each case.

Table 2. Adopted hyperparameters and additional information about the experiments with the TFT models.

Parameter	P1	P2	P3
hidden_size	45	60	60
lstm_layers	3	3	3
attention_head_size	4	4	4
batch_size	64	64	64
limit_train_batch	1024	2048	1024
limit_val_batch	1024	1024	1024
learning_rate	0.0001	0.0002	0.0002
patience	24	24	24
dropout	0.15	0.15	0.15
Total number of parameters	6.10×10^4	5.61×10^5	5.77×10^5
Model size (MB)	0.24	2.25	2.31
Time to train (average)	40 min	4 h 50 min	6 h 20 min

5.2. Comparative Performance of Deep Learning Models in Prediction

The heterogeneous hydrological conditions in Portugal, from wetter regions in the north to drier areas in the south, present a significant challenge to the streamflow prediction model while also providing an opportunity to assess its robustness.

Figure 9 aggregates an example of the obtained predictions for one of the stations (i.e., 05K/01H) for the HBV model and the experiments from P1 to P3. Figure 9a illustrates the resampled daily temperature and precipitation data for the catchment monitored by this station in the period of the prediction.

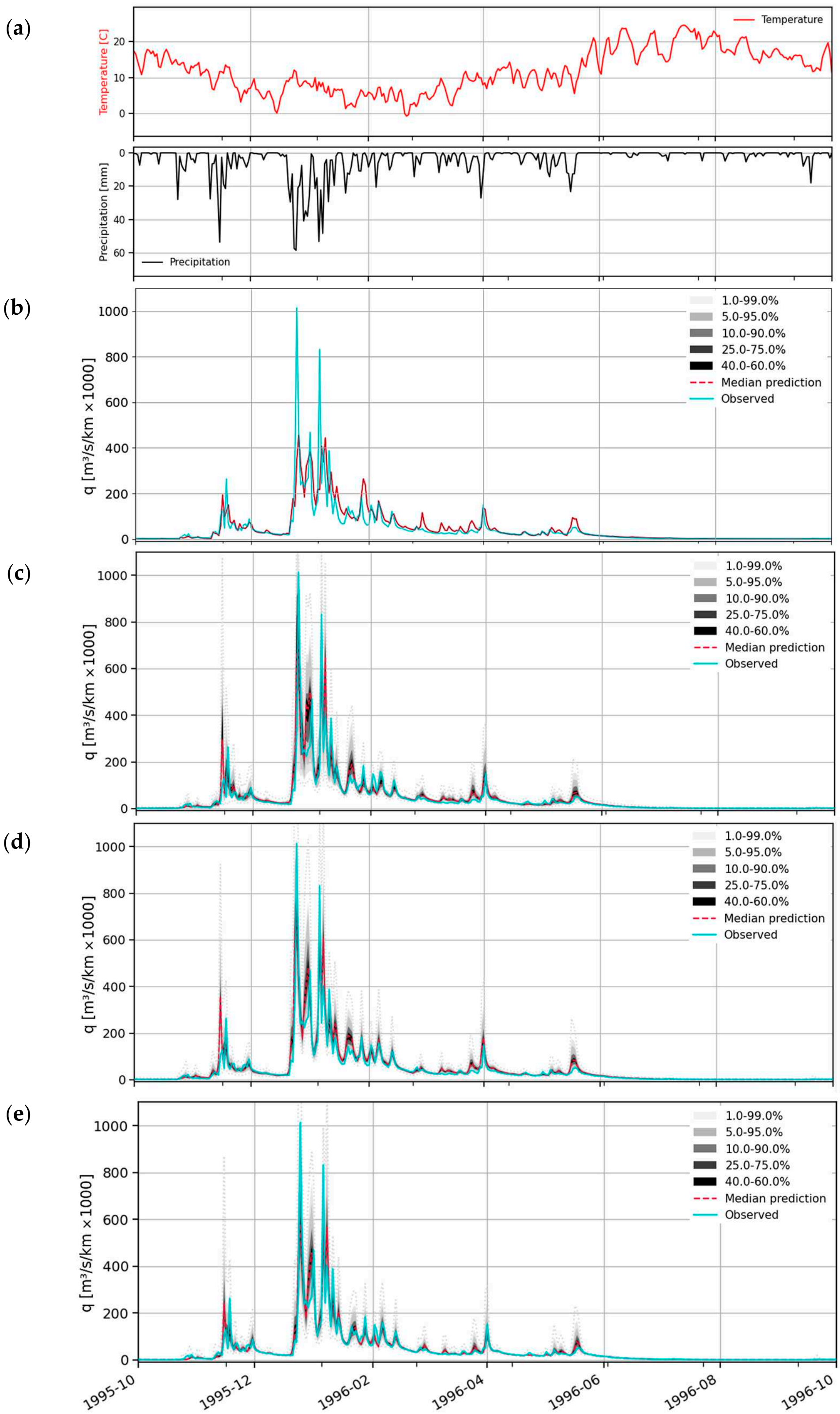


Figure 9. Set of predictions of streamflow at the hydrometric station 05K/01H with the correspondent (**a**) meteorological data (ERA5–Land); (**b**) deterministic prediction from the HBV model; and probabilistic predictions with the TFT for the experiments: (**c**) P1 (single catchment), (**d**) P2 (joint catchments), and (**e**) P3 ("pseudo" forecasting with joint catchments). The observed streamflow and its median prediction are represented in the plots, alongside the respective uncertainty bands.

The calibrated HBV model (classical approach), shown in Figure 9b, provides the deterministic prediction of streamflow obtained directly from it. As mentioned above, it does not account for uncertainty or any other possible outcomes. It also fails to capture two significant peaks in the beginning of 1996. This can be a major limitation in flood analysis, as missing such peaks could lead to misleading flood risk assessments. In Table 3, the obtained performance metrics can be consulted.

Table 3. Performance metrics obtained for each experiment. The color bars are proportional to the values in each metric.

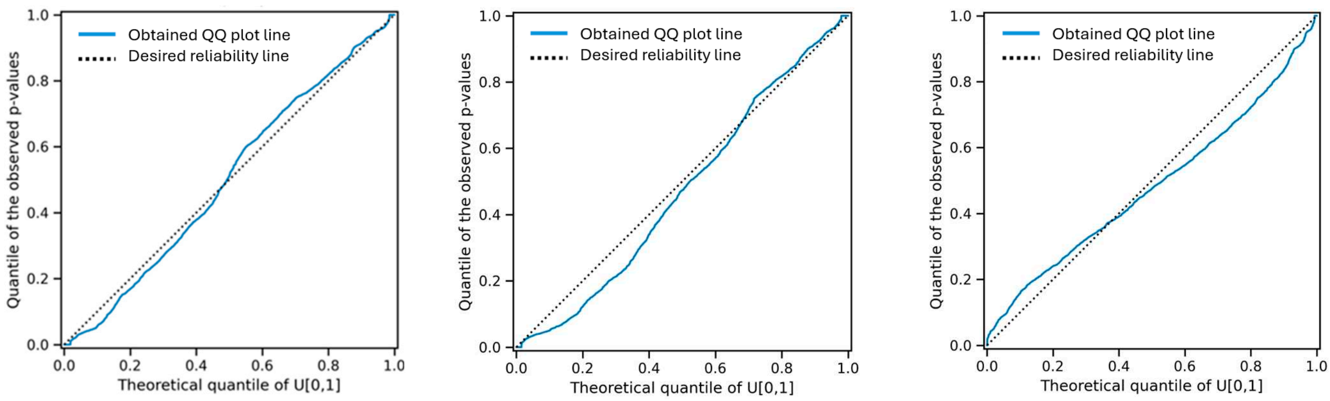
Station	Metric	HBV Model	P1 (Single Catchment)	P2 (Joint Catchments)	P3 ("Pseudo" Forecasting)
05K/01H	*NSE*	0.62	0.73	0.77	0.87
	CRPS	9.41	5.59	4.95	2.94
	α		0.95	0.92	0.92
	π^{rel}		2.79	3.43	8.35
16G/01H	*NSE*	0.71	0.65	0.80	0.81
	CRPS	2.79	2.04	1.70	0.95
	α		0.95	0.96	0.83
	π^{rel}		2.66	2.57	7.26
24H/03H	*NSE*	0.30	0.38	0.71	0.77
	CRPS	1.48	0.79	1.23	1.12
	α		0.89	0.77	0.92
	π^{rel}		0.68	0.68	1.68
28L/02H	*NSE*	0.48	0.45	0.58	0.67
	CRPS	3.22	2.11	1.73	1.08
	α		0.95	0.82	0.83
	π^{rel}		0.53	0.63	1.48

For P1 (TFT model trained with a single catchment), the obtained results are promising, as seen in Figure 9c. The model was capable of "learning" a great part of the inherent hydrological processes and still providing information about uncertainty, in particular the two peaks the HBV model struggled to replicate. The performance metrics shown in Table 3 emphasize the potential of the probabilistic prediction. For example, in stations 05K/01H and 24H/03H (the ones that monitor the isolated catchments), all metrics in P1 outperformed the HBV results. However, for stations 16G/01H and 28L/02H, the NSE in P1 showed poorer results compared to the HBV model, that is, 0.71 in HBV vs. 0.65 in P1 for the first and 0.48 in HBV vs. 0.45 in P1 for the second, possibly due to model training issues or the influence of local factors not fully captured by the TFT. Still, the CRPS is lower in P1 (which is desirable), and this experiment additionally offers insights into reliability and resolution, which HBV fails to do. Globally, this experiment suggests that TFTs can perform the same tasks as the HBV model with more accurate outcomes.

In P2 (the TFT model trained with joint catchments), the expanded dataset introduced a higher level of complexity into the model, which required significantly more time to train and validate. The model performance was expected to improve due to the increased exposure to hydrological information, which occurred. Figure 9d briefly confirms this by showing a prediction with narrower uncertainty bands and, therefore, increased accuracy in comparison with P1. Looking at Table 3, in fact, one can see resolution increased from 2.19 to 3.43. In the remainder of the stations, this improvement was not as strong, and resolution values are relatively similar. NSE improved in all stations, surpassing also all the values obtained by the HBV model. Reliability generally decreased slightly from P1 to P2, although remaining generally satisfactory. The CRPS globally improved from P1 to P2. For the relatively isolated catchments (05K/01H and 24H/03H), P2 has strongly increased the model performance. Experiment P2 showcases the capability of TFT to learn information about hydrological processes in such a way that can improve local predictions.

Experiment P3 (TFT model used for "pseudo" forecasting) introduced knowledge about past streamflow into the model at each time step. Once again, Figure 9e depicts an increase in the accuracy of the prediction with yet narrower uncertainty bands. Table 3 corroborates this by the abrupt increase in the resolution in comparison with prior experiments. Likewise, NSE and CRPS improved from the values obtained for HBV, P1, and P2. Reliability remained satisfactory.

Figure 10 presents the QQ plots that resulted from the three experiments (P1, P2, and P3) for the same catchment (05K/01H). Deviations in the predictions, as well as overestimations and underestimations of the predictive uncertainty, could be identified, to a greater or lesser extent, in all the experiments. This said, the calculated performance metrics (Table 3) for the four catchments suggest that the deviations are not major, especially when the results obtained using the "classical" HBV model are used to put them into perspective.



(**a**) (**b**) (**c**)

Figure 10. QQ plots for the streamflow prediction in the catchment gauged by the station 05K/01H in the experiments (**a**) P1 (**b**) P2 and (**c**) P3. All the obtained QQ plots are relatively close to the desired 1–1 line, evidencing good reliability.

5.3. Adaptation of Deep Learning Models to Forecasting Tasks

The forecasts obtained with GFS meteorology (item no. 5 in Section 4.3) are exemplified in Figure 11. Although the temperature forecasts from GFS are relatively accurate in this period, the precipitation presents some inaccuracies during particularly intense raining events. Data limitations (meteorology and streamflow observations) dictated that forecasts were only produced for stations 16G/01H and 24H/03H, between 1 January 2022 and 1 January 2024. The obtained results, covering this period, were benchmarked against persistence. In this context, persistence refers to a simple method where the forecast assumes that the value of the variable in the future will be equal to that of the last observation.

Looking at Table 4, it is possible to conclude that, for the period of 2 years, the TFT model had better results on the performance metrics when compared to the benchmark. The QQ plots for both stations are available in Figure 12, showing increased deviations when compared to Figure 10. This can be explained by the shorter period of analysis and the fact that GFS forecasts are expected to reproduce precipitation and temperature less accurately than ERA5–Land.

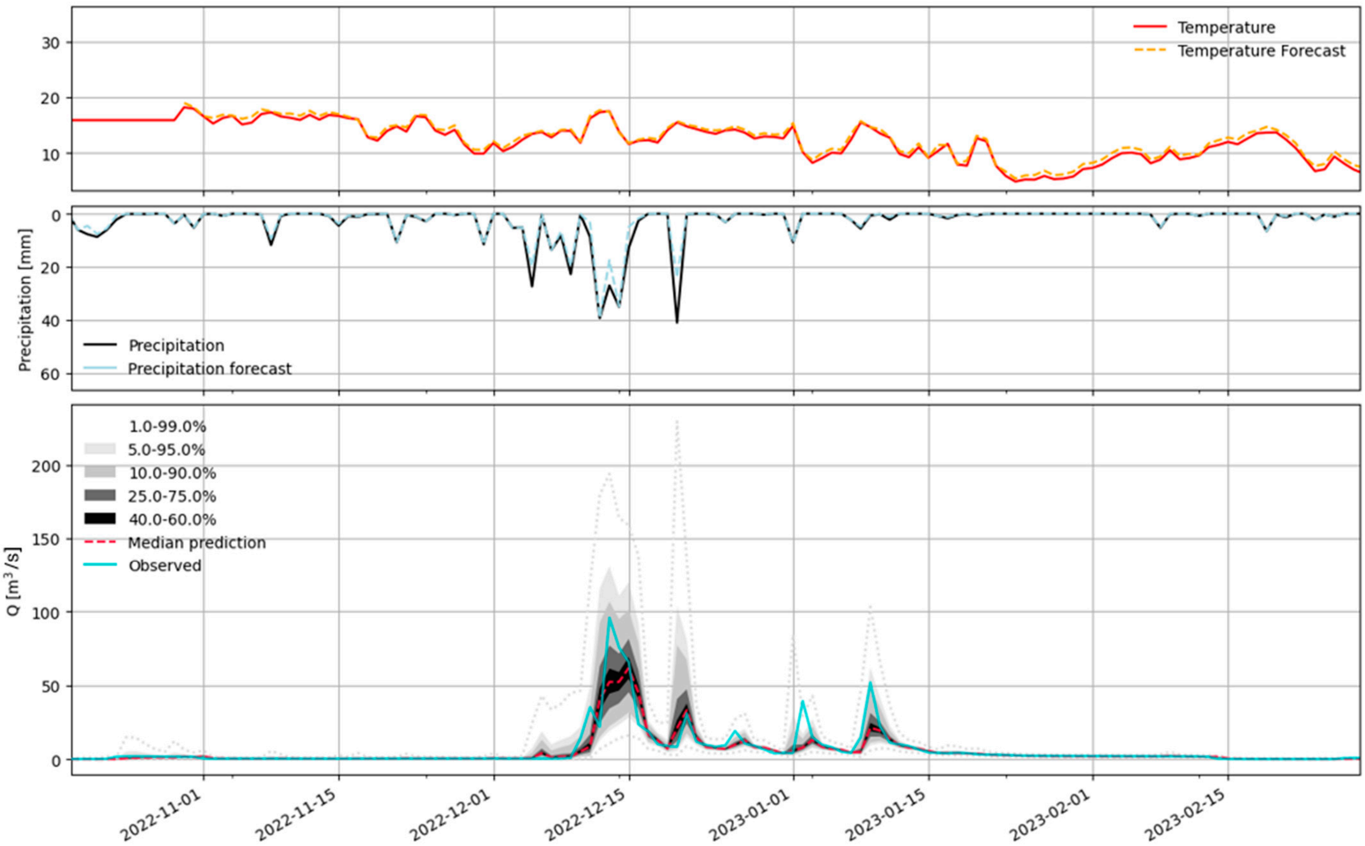
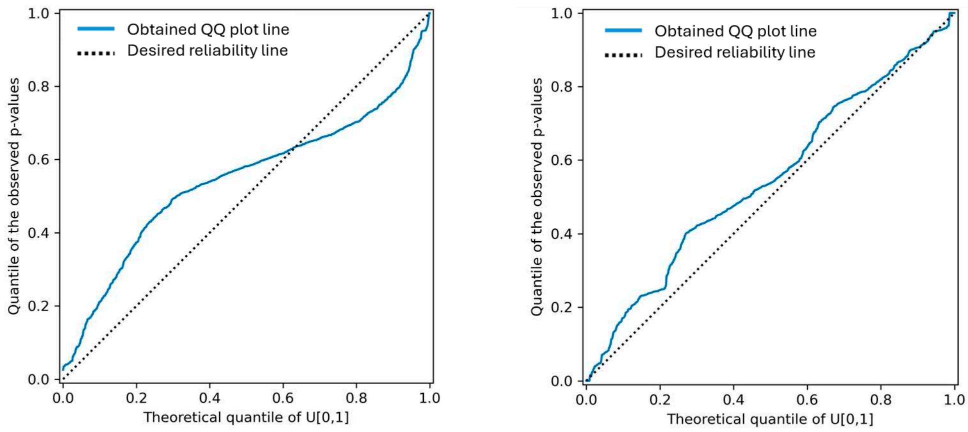


Figure 11. Forecast of streamflow at the hydrometric station 24H/03H with the correspondent meteorological forecasts from GFS in comparison with the values from ERA5–Land. The wet season (autumn and winter) of 2022/23 is depicted. The observed streamflow and its median prediction are represented in the plots, alongside the respective uncertainty bands.

Table 4. Performance metrics obtained for the forecasts. The color bars are proportional to the values in each metric.

Station	Metric	Persistence	Forecast
16G/01H	*NSE*	0.80	0.88
	CRPS	0.91	0.50
	α		0.80
	π^{rel}		6.98
24H/03H	*NSE*	0.50	0.76
	CRPS	0.82	0.47
	α		0.90
	π^{rel}		1.37



(**a**) (**b**)

Figure 12. QQ plots for the streamflow forecasts in the catchment gauged by the stations: (**a**) 16G/01H and (**b**) 24H/03H.

It can be concluded that the TFT models are highly promising for streamflow prediction applications, especially if compared with classical approaches such as the HBV model. TFT models can easily be used in an operational setting where forecasts are generated continuously as new input data (e.g., updated meteorological forecasts or streamflow measurements) become available. The showed plots, performance metrics, and following reasons support this rationale:

(i) Flexibility in the types of inserted inputs and model adaptability in producing predictions: The TFT model can incorporate different types of input data (categorical or real, constant or variable, and known or unknown), including past streamflow, meteorological data, and other geomorphological variables, allowing it to adapt to diverse hydrological contexts more effectively than the HBV model (and most likely other similar "classical" hydrological models).

(ii) Possibility of transposing knowledge from hydrological processes between catchments: Due to their nature, when trained with data from many catchments, TFTs can learn general features of the rainfall–runoff transformation processes that allow it to improve local predictions.

(iii) Probabilistic nature of TFT in comparison with deterministic nature of HBV: The HBV model, as a classical hydrological model, is not probabilistic, requiring either a post–processor to estimate uncertainty or a set of model parameters that "generates" an ensemble of model results capable of reproducing that uncertainty. In contrast, the TFT is inherently probabilistic, providing a distribution of possible outcomes along with associated uncertainties. Naturally, knowledge about predictive uncertainty allows for more informed decision–making in situations of risk management (e.g., in flood forecasting).

(iv) Simplicity of the TFT modeling chain: In practice, the modeling chain necessary for a TFT to produce forecasts is much simpler than the one required by classical hydrological models. Indeed, the latter cannot directly use past streamflow as an additional input and require that internal state variables (e.g., water storage in the soil layers and other reservoirs) are continuously updated so that predictions match the latest observations—a process that can be very demanding.

To balance the positive results, it is only fair to highlight some limitations. The following stand out:

(1) Data availability and quality: The performance of the TFTs highly depends on the quantity and quality of data, which sometimes can be difficult to guarantee. Even if long –time series exist, they may not be enough to ensure adequate performances, especially when observations are inaccurate.

(2) Complexity of hyperparameter definition: Finding adequate combinations of model hyperparameters can be very time–consuming and absorb significant computational resources. Optimizing batch size, dropout rate, learning rate, network architecture, and other factors requires a fine understanding of their hidden relationships.
(3) Some environmental and physical factors are not accounted for. Although the trained TFT models considered temperature, precipitation, and other relevant variables associated with the catchment, they did not include all factors that may affect hydrological responses. Notably, time was not accounted for at the level of catchment characterization (e.g., changes in land use over time).
(4) Overfitting: In smaller datasets (e.g., P1), the risk of overfitting is more pronounced, mainly when dropout rates are not carefully selected. Generically, when an excessively complex model is used, it tends to achieve a quick reduction of the training error but will soon stabilize and worsen in terms of validation (and test) groups. In such cases, the models can be generally improved by reducing the number of parameters, i.e., the complexity of the network.

These results show that TFTs can have operational applications in streamflow prediction. Namely, they have the potential to be beneficial in the following diverse activities:

(i) Predicting flood events in the scope of early warning systems and emergency response and acknowledging the different uncertainties associated with streamflow.
(ii) Optimizing reservoir operations and supporting decisions on water storage, release, and allocation across different sectors (e.g., agriculture, industry, and energy).
(iii) Optimizing hydropower production by adjusting power generation schedules to comply with energy demand while ensuring sustainable use.
(iv) Guiding the implementation of water–saving measures during low–flow periods or droughts.
(v) Managing the health of ecosystems and protecting wildlife highly sensitive to fluctuations in streamflow.

6. Conclusions

The presented results proved the potential of the TFT model in the prediction of daily streamflow, especially in contrast with the HBV model. Focusing on individual catchments (experiment P1), the model was able to generate reliable probabilistic predictions and thus provide information useful to decision–makers. In objective performance metrics, as well as capacity to generalize and inbuilt capacity to generate probabilistic predictions, it outperformed the "classical" HBV model. It also corresponded to a common site–based data–driven model application, fulfilling the objectives of the initially defined question. Very importantly, the P2 experiment showed that, trained with regional datasets, TFTs can learn about generic hydrological behaviors. Lastly, with P3, after providing the model streamflow awareness, the TFT performance improved, and the uncertainty decreased. Still, it is relevant to notice that this model has some limitations too (e.g., high dependence on data availability and quality, complexity of hyperparameter definition, and computational requirements).

The TFTs can be easily adapted to online forecasting of streamflow. Adding past streamflow to model inputs clearly improves performance, and even with a simple correction, the model fed with data not directly used in training (GFS replaced ERA5–Land in the forecasting experiment) could clearly outperform the benchmark.

The results can be applied to improve water management practices and operational forecasting, contributing to predicting future flood and drought events with uncertainty associated, optimizing reservoir operations and hydropower production, implanting water– and wildlife–saving measures, and mitigating the effects of climate change.

The potential improvement of the performance of the TFT model is enormous. Indeed, numerous additional features can be tested to increase the accuracy of predictions/forecasts as follows:

(i) Application of additional measures to mitigate model overfitting, including testing other regularization methods and conducting a more in–depth analysis of different combinations of model hyperparameters.
(ii) Attribution to greater weights to the data from under–represented catchments in the study area.
(iii) Specification of how the data are distributed within the catchment of different input variables (instead of one average value representing the entire catchment).
(iv) Extension to other variables besides streamflow (e.g., sediment transport in rivers, river stage, and hydropower scheduling).

Author Contributions: Both authors contributed in equal measure to this study. All authors have read and agreed to the published version of the manuscript.

Funding: The Foundation for Science and Technology's support through funding UIDB/04625/2020 from the research unit CERIS.

Data Availability Statement: The obtained predictions will be made available online along with instructions on how to read them prior to publication.

Acknowledgments: We are appreciative of the work required by the Portuguese Environment Agency to collect and share all the hydrometric information through SNIRH. This study would not have been possible without the ERA5–Land and GFS datasets, whose open–access publication is priceless. Finally, we acknowledge the central role that TFT code, part of the freely available PyTorch forecasting package, had in this contribution.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. United Nations Department of Economic and Social Affairs; International River Foundation. Thriving and Resilient Rivers for Future Generations—Addressing the Global Water Challenge. Sustainable Development. Available online: https://sdgs.un.org/partnerships/thriving-and-resilient-rivers-future-generations-addressing-global-water-challenges#description (accessed on 25 October 2024).
2. Ritchie, H.; Rodés-Guirao, L.; Mathieu, E.; Gerber, M.; Ortiz-Ospina, E.; Hasell, J.; Roser, M. Population Growth. OurWorldInData.Org. Available online: https://ourworldindata.org/population-growth (accessed on 6 May 2024).
3. Lee, J.; Perera, D.; Glickman, T.; Taing, L. Water-related disasters and their health impacts: A global review. *Prog. Disaster Sci.* **2020**, *8*, 100123. [CrossRef]
4. Kuzma, S.; Luo, T. *The Number of People Affected by Floods Will Double Between 2010 and 2030*; Word Resources Institute: Washington, DC, USA, 2020. Available online: https://www.wri.org/insights/number-people-affected-floods-will-double-between-2010-and-2030 (accessed on 9 September 2024).
5. Alfieri, L.; Bisselink, B.; Dottori, F.; Naumann, G.; de Roo, A.; Salamon, P.; Wyser, K.; Feyen, L. Global projections of river flood risk in a warmer world. *Earth's Future* **2017**, *5*, 171–182. [CrossRef]
6. Camilo, D. Cheias Fizeram 293 Milhões de Euros de Prejuízos. Governo Anuncia 185 Milhões em Apoios. Renascença. 12 January 2023. Available online: https://rr.sapo.pt/noticia/economia/2023/01/12/cheias-fizeram-293-milhoes-de-euros-de-prejuizos-governo-anuncia-185-milhoes-em-apoios/315654/ (accessed on 5 May 2024).
7. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
8. Centre de Recherche sur L'environnement Alpin (CREALP); HydroCosmos SA. RS Minerve v. 2.9.1.0 (Software). 2021. Available online: https://www.crealp.ch/en/rs-minerve-software/ (accessed on 7 April 2024).
9. Koya, S.R.; Roy, T. Temporal Fusion Transformers for Streamflow Prediction: Value of Combining Attention with Recurrence. *J. Hydrol.* **2024**, *637*, 131301. [CrossRef]
10. Yaseen, Z.M. Flood hazards and susceptibility detection for Ganga river, Bihar state, India: Employment of remote sensing and statistical approaches. *Results Eng.* **2024**, *21*, 101665. [CrossRef]
11. Franchi, F.; Mustafa, S.; Ariztegui, D.; Chirindja, F.J.; Di Capua, A.; Hussey, S.; Loizeau, J.-L.; Maselli, V.; Matanó, A.; Olabode, O.; et al. Prolonged drought periods over the last four decades increase flood intensity in southern Africa. *Sci. Total Environ.* **2024**, *924*, 171489. [CrossRef] [PubMed]
12. Lehmkuhl, F.; Schüttrumpf, H.; Schwarzbauer, J.; Brüll, C.; Dietze, M.; Letmathe, P.; Völker, C.; Hollert, H. Assessment of the 2021 summer flood in Central Europe. *Environ. Sci. Eur. (Ed.)* **2022**, *34*, 107. [CrossRef]
13. Ionno, A.; Arsenault, R.; Troin, M.; Martel, J.-L.; Brissette, F. Impacts of climate change on flood volumes over North American catchments. *J. Hydrol.* **2024**, *630*, 130688. [CrossRef]

14. Buschschlüter, V. Brazil Floods: Hundreds of Rio Grande do Sul Towns Under Water. *BBC News*, 7 May 2024. Available online: https://www.bbc.com/news/world-latin-america-68968987 (accessed on 17 May 2024).
15. Liu, J.; Yuan, X.; Zeng, J.; Jiao, Y.; Li, Y.; Zhong, L.; Yao, L. Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 265–278. [CrossRef]
16. Dick, S. Artificial Intelligence. *HDSR* **2019**, *1*. [CrossRef]
17. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]
18. Hunter, B.; Hindocha, S.; Lee, R.W. The Role of Artificial Intelligence in Early Cancer Diagnosis. *Cancers* **2022**, *14*, 1524. [CrossRef]
19. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Robot. Syst.* **2019**, *37*, 362–386. [CrossRef]
20. Şerban, A.C.; Lytras, M.D. Artificial Intelligence for Smart Renewable Energy Sector in Europe—Smart Energy Infrastructures for Next Generation Smart Cities. *IEEE Access* **2020**, *8*, 77364–77377. [CrossRef]
21. Chang, F.-J.; Chang, L.-C.; Chen, J.-F. Artificial Intelligence Techniques in Hydrology and Water Resources Management. *Water* **2023**, *15*, 1846. [CrossRef]
22. Geetha, G.K.; Sim, S.-H. Fast identification of concrete cracks using 1D deep learning and explainable artificial intelligence-based analysis. *Autom. Constr.* **2022**, *143*, 104572. [CrossRef]
23. Kim, Y.I.; Park, W.H.; Shin, Y.; Park, J.-W.; Engel, B.; Yun, Y.-J.; Jang, W.S. Applications of Machine Learning and Remote Sensing in Soil and Water Conservation. *Hydrology* **2024**, *11*, 183. [CrossRef]
24. Terlumun, S.; Onyia, M.E.; Okafor, F.O. Predicting the compressive strength of self-compacting concrete using artificial intelligence techniques: A review. *AI Civ. Eng.* **2024**, *3*, 14. [CrossRef]
25. Chollet, F. *Deep Learning with Python*, 2nd ed.; Manning Publications: Shelter Island, NY, USA, 2021; pp. 1–13; ISBN 9781617294433.
26. Kadiyala, L.A.; Mermer, O.; Samuel, D.J.; Sermet, Y.; Demir, I. The Implementation of Multimodal Large Language Models for Hydrological Applications: A Comparative Study of GPT-4 Vision, Gemini, LLaVa, and Multimodal-GPT. *Hydrology* **2024**, *11*, 148. [CrossRef]
27. Mohammadi, B. Modeling Various Drought Time Scales via a Merged Artificial Neural Network with a Firefly Algorithm. *Hydrology* **2023**, *10*, 58. [CrossRef]
28. Mehedi, M.A.A.; Khosravi, M.; Yazdan, M.M.S.; Shabanian, H. Exploring Temporal Dynamics of River Discharge Using Univariate Long Short-Term Memory (LSTM) Recurrent Neural Network at East Branch of Delaware River. *Hydrology* **2022**, *9*, 202. [CrossRef]
29. Dabire, N.; Ezin, E.C.; Firmin, A.M. Forecasting Lake Nokoué Water Levels Using Long Short-Term Memory Network. *Hydrology* **2024**, *11*, 161. [CrossRef]
30. Xiang, Z.; Yan, J.; Demir, I. A Rainfall-Runoff Model with LSTM-Based Sequence-to-Sequence Learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [CrossRef]
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on 17 March 2024).
32. Sistema Nacional de Recursos Hídricos (Agência Portuguesa do Ambiente). Daily Streamflow Records [Data Set]. 2024. Available online: https://snirh.apambiente.pt/ (accessed on 21 March 2024).
33. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
34. Almeida, M.; Coelho, P. A first assessment of ERA5 and ERA5-Land reanalysis air temperature in Portugal. *Int. J. Clim.* **2023**, *43*, 6643–6663. [CrossRef]
35. Barbosa, S.; Scotto, M.G. Extreme heat events in the Iberia Peninsula from extreme value mixture modeling of ERA5-Land air temperature. *Weather. Clim. Extrem.* **2022**, *36*, 100448. [CrossRef]
36. Hargreaves, G.H.; Samani, Z.A. Estimating Potential Evapotranspiration. *J. Irrig. Drain. Div.* **1982**, *108*, 225–230. [CrossRef]
37. U.S. National Centers for Environmental Prediction (NCEP). Global Forecasting System Documentation. Available online: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php (accessed on 1 October 2024).
38. Bergström, S. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. *SMHI Rep. RHO 7* **1976**, *134*, 1–8. Available online: https://urn.kb.se/resolve?urn=urn:nbn:se:smhi:diva-5738 (accessed on 7 August 2024).
39. Lindström, G.; Johansson, B.; Persson, M.; Gardelin, M.; Bergström, S. Development and Test of the Distributed HBV-96 Hydrological Model. *J. Hydrol.* **1997**, *201*, 272–288. [CrossRef]
40. Oreshkin, B.N.; Carpov, D.; Smelyanskiy, M. *PyTorch Forecasting: A Package for Time Series Forecasting with Deep Learning*; GitHub Repository: Cambridge, MA, USA, 2020. Available online: https://github.com/jdb78/pytorch-forecasting (accessed on 21 October 2024).
41. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

42. Hersbach, H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *WAF* **2000**, *15*, 559–570. [CrossRef]
43. Coccia, G.; Todini, E. Recent Developments in Predictive Uncertainty Assessment Based on the Model Conditional Processor Approach. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3253–3274. [CrossRef]
44. Renard, B.; Kavetski, D.; Kuczera, G.; Thyer, M.; Franks, S.W. Understanding Predictive Uncertainty in Hydrologic Modeling: The Challenge of Identifying Input and Structural Errors. *Water Resour. Res.* **2010**, *46*, 1–16. [CrossRef]
45. Portuguese Ministry of Foreign Affairs. About Portugal—Diplomatic Portal. Available online: https://portaldiplomatico.mne.gov.pt/en/about-portugal (accessed on 7 May 2024).
46. European Space Agency (ESA). *Copernicus DEM GLO-30 (Raster)*; European Space Agency: Paris, France, 2019. [CrossRef]
47. OpenStreetMap Contributors. OpenStreetMap Portugal continental e Madeira: Water Bodies. OpenStreetMap. Available online: https://www.openstreetmap.org (accessed on 14 November 2024).
48. Assembleia da República. Republic Diary No. 249/2005. *Diário República Ser. I* **2005**, *58*, 7280–7310. Available online: https://diariodarepublica.pt/dr/detalhe/lei/58-2005-469068 (accessed on 12 October 2024).
49. Esri Inc. *ArcGIS Pro (v. 3.1)*; Esri Inc.: Redlands, CA, USA, 2024; Available online: https://www.esri.com/en-us/arcgis/products/arcgis-pro (accessed on 10 March 2024).
50. Zhang, A.T.; Gu, V.X. Global Dam Tracker: A Database of More than 35,000 Dams with Location, Catchment, and Attribute Information. *Sci. Data* **2023**, *10*, 111. [CrossRef]
51. European Environment Agency (EEA). *Copernicus CORINE Land Cover (Vector)*; European Environment Agency: Copenhagen, Denmark, 2018. [CrossRef]
52. Beitner, J. PyTorch Forecasting Documentation: TemporalFusionTransformer. 2020. Available online: https://pytorch-forecasting.readthedocs.io/en/stable/api/pytorch_forecasting.models.temporal_fusion_transformer.TemporalFusionTransformer.html#pytorch_forecasting.models.temporal_fusion_transformer.TemporalFusionTransformer (accessed on 18 June 2024).
53. Falcon, W.; PyTorch Lightning Team. *PyTorch Lightning Documentation 2.2.5: Trainer*; San Francisco, CA, USA, 2023. Available online: https://lightning.ai/docs/pytorch/stable/common/trainer.html# (accessed on 18 June 2024).