# DEVELOPMENT OF DROUGHT EARLY WARNING SYSTEM FOR THE ALPS

*Muhammad Usman Liaqat[1], Mohammad Hussein Alasawedah[1], Albrecht Weerts[23], Alice Crespi[4], Suriyah Dhinakaran[4,5], Alexander Jacob[1], and Mariapina Castelli[1]*

[1]Institute for Earth Observations, Eurac Research, Bolzano, Italy, [2]Inland Water Systems, Operational Water Management, Deltares, The Netherlands, [3]Hydrology and Quantitative Water Management group, Department of Environmental Sciences, Wageningen University, The Netherlands, [4]Center for Climate Change and Transformation, Eurac Research, Bolzano, Italy ,[5]University of Münster, Münster, Germany

## ABSTRACT

Physically based distributed hydrological models have proven to be effective for simulating hydrological processes. Nevertheless, their applicability is limited by their intricate nature. Recently, deep learning-based models have gained significant attention for their ability to simulate high accuracy of hydrological fluxes. However, these models do not adhere to physical laws and remain difficult to comprehend. What if at the same time a data driven model that can achieve reliable accuracy, could also be clear about physical processes and provide interpretable outputs such as groundwater recharge, evapotranspiration, base flow and soil moisture? In this study, we propose a pipeline for optimizing hydrological model parameters by deep learning, exploiting the potential of earth observation data to train the model. The calibrated model will be used to formulate sub-seasonal drought predictions. The proposed framework will open new possibilities to exploit big data in hydrological modeling.

***Index Terms—*** Drought prediction, surrogate model, Deep Learning, Compute vision

## 1. INTRODUCTION

Human-caused climate change is affecting weather and climate extremes worldwide, with changes in their intensity and frequency [1]. In turn, frequency of natural disasters stemming from such events increased by a factor of five over the past fifty years [2]. Approximately 74% of these disasters were water-related, including droughts and floods [3].

The Alps are regarded as the water towers of Europe, containing glacier-rich mountains, snowfields, and underground aquifers. These water resources play a significant role in sustaining major European rivers such as the Adige, Danube, Rhine, Po, and Rhone [4]. Although the Alpine region has historically been protected from prolonged droughts by its temperate climate and high elevation, drought risks have increased in recent years [5]. This situation primarily stems from shifting climate patterns, which have resulted in a decline in water resources [6]. It is also anticipated that competing water demands will rise in the future, increasing the region's susceptibility to drought [7].

Studies have been carried out to examine current and future hydrological regimes in Alpine catchments [8]. Hydrological simulations can benefit from the use of physically based models. Fully distributed hydrological models are generally regarded as a reliable tool for hydrological modeling [9]. Nevertheless, the efficiency of physically based distributed hydrological models is greatly influenced by unobservable and underdetermined parameters that vary with time and space. Generally, calibration is employed to optimize the unobserved parameters at each location, in an attempt to reduce the difference between the model's outputs and observed measurements [10]. Parameter calibration has been widely used in earth science for many years. Several textbook chapters and research articles specifically address calibration techniques in earth science [11]. Most ecosystem dynamics [12] and rainfall runoff [13] models contain unknown parameters which vary in temporal and spatial scales and need to be calibrated. Additionally, these parameters are subject to change with changes in spatial and temporal resolution as well as other model parameters, model versions, and input data. This requires constant adjustment of previously calibrated parameters by thousands of model simulations, which is a repetitive, cumbersome, and time-consuming task [14]. To minimize these issues, there is a need for a hydrological model that is both user-friendly and highly efficient.

From 2017 the use of machine learning for hydrological applications has increased rapidly. [16]. Despite this progress, deep learning (DL) is rarely used by physical hydrologists. This is primarily due to the fact that the educational background necessary for implementing DL techniques is fundamentally different from that required in a traditional hydrology program [15]. Nevertheless, given the current growth rate, DL may eventually become an essential component of hydrology [16].

Data-driven hydrological models that utilize DL can offer systematic ways to optimize model parameters using big data. Unlike physically based hydrologic models, they do not rely on the specific meanings of variables and can effectively learn hydrologic processes from observed meteorological and hydrological data establishing rapid connections between meteorological predictors and hydrological responses, bypassing the need for detailed descriptions of physical processes [17]. As one of these methods, Long Short-Term Memory networks (LSTM) [18] have been widely studied for their ability to retain long-term relationships in time series data, to handle sequences with variable time lags efficiently, to model time patterns and non-linear relationships, and to make accurate predictions based on historical data. In recent research, LSTM exhibited their suitability for predicting different hydrological components such as rainfall runoff [19], soil moisture [20], and streamflow [21]. However, data-driven models can only be used to predict variables for which sufficient data is available. For other important simulated variables such as evapotranspiration, groundwater recharge, or root carbon storage they still rely on manually calibrated process-based models.

Modern DL networks, along with their efficient training techniques like backpropagation and gradient descent, are particularly advantageous in utilizing the vast amount of information present in large datasets. It is conceivable to employ DL to tackle parameter calibration challenges at a larger scale. However, despite numerous suggestions to integrate DL methods with physically based approaches, there are few frameworks [22] known to us that fully exploit modern DL for the parameter calibration problem.

In this study, we will develop a differentiable parameter learning (dPL) framework based on LSTM networks in two phases. In the first phase, deep learning model will be trained to reproduce the process-based model by keeping the physics of the model. In the second phase, the parameters of the hydrological (LSTM surrogate) model will be optimized based on historical observations and river basin physical attributes. The calibrated model will be finally fed by seasonal forecasts to generate medium term predictions of soil moisture and discharge. Based on these predictions, hydrological drought indices will be calculated to formulate drought predictions.

## 2. MATERIAL AND METHODS

### 2.1 Description of Study Area

The surrogate model will be developed over several Alpine catchments, but the Adige catchment (Fig.1) is used for the first tests and set up of the modeling framework. The Adige catchment is approximately 12,100 km$^2$. Its discharge exhibits the common characteristics of Alpine catchments, with peaks typically occurring during the melting season between June and September [23]. The mean Exploratory Data Analysis (EDA) curve (2000-2018) for Adige streamflow at Bronzolo is shown in Fig.2. The primary tributaries of the Adige catchment are situated in the Alpine region of the basin, making them highly dependent on snow dynamics. Climate change has already had significant impacts on water resources management in this area, particularly in terms of hydropower production and winter tourism [24]. The distribution of precipitation within the catchment is not uniform, with values ranging from 500 mm/yr in Val Venosta (located in the north-western part of the catchment) to 1600 mm/yr in the southern part of the basin, as reported by [25]. Temperature also varies greatly within the catchment due to its high elevation gradient. Average monthly temperatures range from 14 °C in July to -4 °C in January and December, as stated by [25].

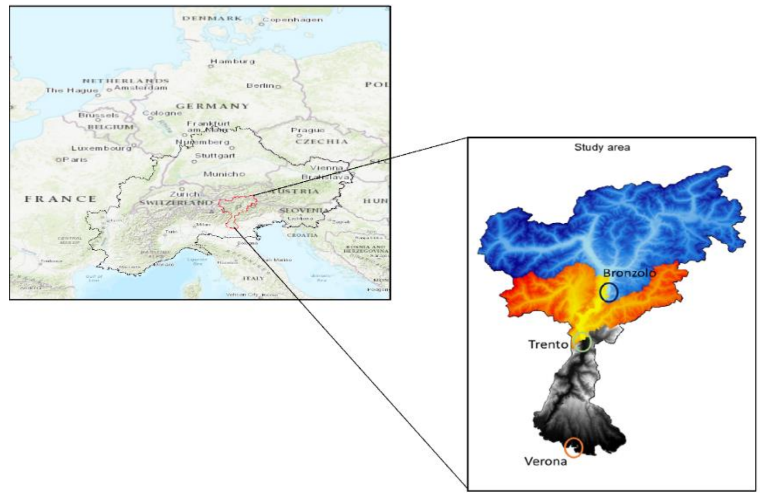


Fig.1. The Adige catchment and the gauging locations (the names of the gauging points are reported in the inset map).

### 2.2 Dynamic and Static Data Processing

In this study, both dynamic and static types of data are used as input to run hydrological as well as its surrogate model (Tab.1). The dynamic input data include meteorological variables, such as temperature, precipitation and potential evapotranspiration, and vegetation-related information, such as the leaf area index. The surrogate model is trained for 2000-2015 and tested for 2016-2020. Meteorological inputs from 2000 to 2020 derive from daily ERA5 reanalysis. Both the dynamic and static data are reprojected and resampled to the resolution of the hydrological model grid (~ 1 km). Once the model is set, it will be run with meteorological forcing inputs derived from a downscaled version of ECMWF SEAS5 seasonal forecasts in place of ERA5 reanalysis. The downscaling of forecast fields will be performed through machine learning schemes optimized for each specific variable.

### 2.3 Hydrological Model

This study uses the spatially distributed hydrologic model Wflow_sbm [26, 27] to estimate hydrological fluxes including snow accumulation and melt, interception, evapotranspiration, soil moisture, streamflow, actual evapotranspiration, surface water, and groundwater recharge. WFLOW uses kinematic wave routing for the routing of surface, channel, and lateral subsurface flows. At a given time interval, Wflow_sbm can simulate all hydrological fluxes at any grid cell using gridded topography, soil, land use, and meteorological data. The Wflow_sbm configuration allows to maximize the use of high-resolution spatial data from Earth Observation. Models can be set up for river basins around the globe using open data at various spatial resolutions. The table below shows all the possible geospatial, static data sources that can be used to run Wflow_sbm. This model is setup to represent a 0.008333-degree gird cell size, which corresponds to approximately one kilometer. However, it is possible to downgrade or upgrade model resolution according to user demand. In this study, a combination of multiple efficiency criteria such as Nash-Sutcliffe efficiency (NSE), Percent bias (PBIAS), Root-Mean-Square Error (RMSE), Coefficient of Determination ($R^2$), and Kling Gupta efficiency (KGE) are used to assess the model performance.

Table.1 List of dynamic and static data used in this study.

| Data | Data Source | Scale | Description |
|---|---|---|---|
| MERIT Hydro | Global hydrography datasets | Global | This dataset encompasses crucial information such as flow direction, flow accumulation, hydrologically adjusted elevations, and river channel width. |
| HydroLAKES | Hydrosheds | Global | shoreline polygons of all global lakes |
| Reservoir and Dam Database | GRanD | Global | Its purpose is to consolidate various dam and reservoir datasets into a single, accurate, and geographically explicit database that can be relied upon by the scientific community. |
| Corine Land Cover | Land cover | Europe | land cover in 44 different categories. |
| Soil Grids | ISRIC | Global | soil properties across the world. |
| GLIMS | NASA Earth Data | Global | GLIMS (Global Land Ice Measurements from Space) is an initiative designed to monitor the world's glaciers |
| Randolph Glacier Inventory | Glacier outlines | Global | Randolph Glacier Inventory contains a global overview of the glacier outlines. |
| LAI & $ET_{act}$ | MODIS | 500 m | MODIS 8-day composite LAI product MYD15A2H.006 |
| Dynamic Data | ERA5 | Daily | Temperature, Precipitation and Potential Evapotranspiration. |
| Discharge Data | Adige River | Daily | Mean Daily Discharge at Bronzolo, Trento and Verona |

## 2.4 Surrogate Model

A surrogate model for differentiable hydrological models, like Wflow_sbm, offers computational efficiency, improves convergence while reducing noise, making it suitable for distributed computing. In this study, surrogate model based on dPL framework is developed in two phases as shown in Fig.2. The first phase consists of training the LSTM to reproduce the performance of the process-based model Wflow_sbm by minimizing loss function (RMSE) while allowing for gradient tracking. This step is needed to support a differentiable workflow and to save computational time. The LSTM surrogate model is trained using dynamic forcings and static attributes, using soil moisture and evapotranspiration simulated by Wflow_sbm as a target of emulations. The data is preprocessed by removing outliers using Inter-Quartile Range (IQR) approach. Next, the data is normalized using the min-max normalization technique. Finally, the preprocessed data is split into a 70% training set and a 30% testing set to evaluate the model's performance. This process ensures that the LSTM model receives clean and standardized data for effective training and testing. Once the loss function is minimized, the parameters of the surrogate hydrological model are estimated based on historical observations and static inputs by minimizing a loss function between model output and observations, i.e., soil moisture as it is one of primary component for drought predictions. Once model is trained and parameters are calibrated, drought indices are estimated from streamflow and soil moisture to forecast drought conditions from 2 weeks up to 2 months.

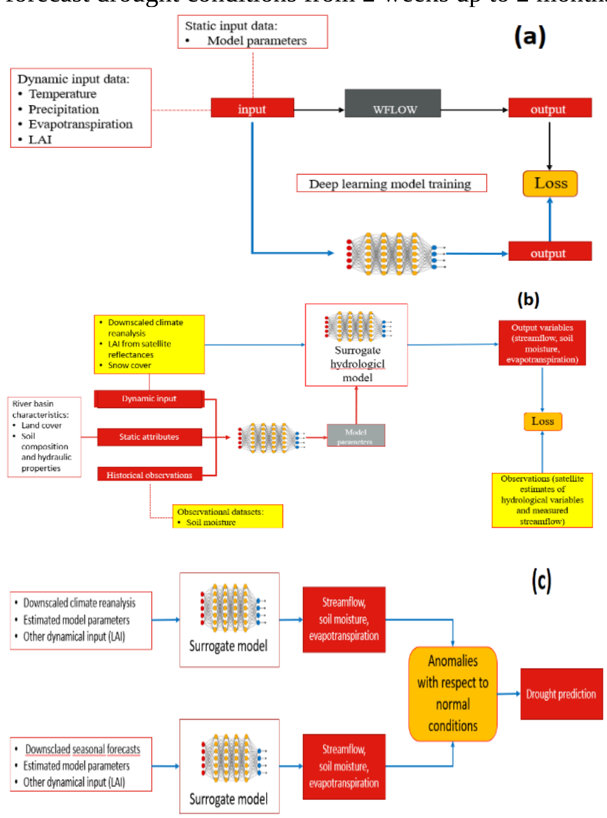

Fig.2. (a) A deep learning model is trained to emulate a physical based hydrological model (Wflow_sbm), (b) work plan of differential parameter learning (dPL) for parameter estimation using trained model in first phase, past observations, reanalysis inputs and static attributes (c) the trained model is finally run using downscaled dynamic forcing inputs derived from seasonal forecasts of ECMWF to predict drought conditions.

## 3. CONCLUSION AND OUTLOOK

This study proposes a framework to improve the prediction capability of deep learning-based hydrological models by incorporating a process-based hydrological model and satellite data for drought predictions. The model performance in learning the behavior of the hydrological model Wflow will be evaluated against Wflow_sbm simulations, and the prediction capability will be investigated using observed data.

## REFERENCES


1. IPCC, 2023: Summary for Policymakers. In: Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, 36 pages. (in press).
2. World Meteorological Association, 2021. WMO atlas of mortality and economic losses from weather, climate and water extremes (1970–2019). Technical Report.
3. Water, U.N., 2020. Water and climate change. The United Nations World Water Development Report.
4. Mastrotheodoros, T., et al., More green and less blue water in the Alps during warmer summers. Nature Climate Change, 2020. 10(2): p. 155-161.
5. Agency, E.E.C., R.; Thyssen, N.; Kristensen, P, Water Resources across Europe: Confronting Water Scarcity and Drought;Publications Office: Luxembourg
6. Diamantini, E., et al., Driver detection of water quality trends in three large European river basins. Science of the Total Environment, 2018. 612: p. 49-62.
7. Anghileri, D., et al., A comparative assessment of the impact of climate change and energy policies on Alpine hydropower. Water Resources Research, 2018. 54(11): p. 9144-9161.
8. Zander, M.J., et al., Future changes in flash flood frequency and magnitude over the European Alps. Hydrology and Earth System Sciences Discussions, 2022: p. 1-21.
9. Chen, Y., Li, J. and Xu, H., 2016. Improving flood forecasting capability of physically based distributed hydrological models by parameter optimization. Hydrology and Earth System Sciences, 20(1), pp.375-392.
10. Yang, S., et al., A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. Journal of Hydrology, 2020. 590: p. 125206.
11. Price, K., R.M. Storn, and J.A. Lampinen, Differential evolution: a practical approach to global optimization. 2006: Springer Science & Business Media.
12. Luo, Y. and E.A. Schuur, Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. Global change biology, 2020. 26(3): p. 1109-1117.
13. Paniconi, C. and M. Putti, Physically based modeling in catchment hydrology at 50: Survey and outlook. Water Resources Research, 2015. 51(9): p. 7090-7129.
14. Yang, Y., et al., In quest of calibration density and consistency in hydrologic modeling: distributed parameter calibration against streamflow characteristics. Water Resources Research, 2019. 55(9): p. 7784-7803.
15. Shen, C., A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 2018. 54(11): p. 8558-8593.
16. Shen, C., Deep learning: A next-generation big-data approach for hydrology. Eos, 2018. 99(1).
17. Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B. and Kisi, O., 2019. Daily streamflow prediction using optimally pruned extreme learning machine. Journal of Hydrology, 577, p.123981.
18. Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. 9(8): p. 1735-1780.
19. Hunt, K.M., et al., Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. Hydrology and Earth System Sciences, 2022. 26(21): p. 5449-5472.
20. Fang, K. and C. Shen, Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel. Journal of Hydrometeorology, 2020. 21(3): p. 399-413.
21. Song, Y.H., E.-S. Chung, and S. Shahid, Differences in extremes and uncertainties in future runoff simulations using SWAT and LSTM for SSP scenarios. Science of The Total Environment, 2022. 838: p. 156162.
22. Tsai, W.-P., et al., From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. Nature communications, 2021. 12(1): p. 5988.
23. Chiogna, G., et al., Effects of the 2017 drought on isotopic and geochemical gradients in the Adige catchment, Italy. Science of the Total Environment, 2018. 645: p. 924-936.
24. Majone, B., Villa, F., Deidda, R. and Bellin, A., 2016. Impact of climate change and water use policies on hydropower potential in the south-eastern Alpine region. Science of the Total Environment, 543, pp.965-980.
25. Laiti, L., et al., Testing the hydrological coherence of high-resolution gridded precipitation and temperature data sets. Water Resources Research, 2018. 54(3): p. 1999-2016.
26. Imhoff, R.O., Van Verseveld, W.J., Van Osnabrugge, B. and Weerts, A.H., 2020. Scaling point-scale (pedo) transfer functions to seamless large-domain parameter estimates for high-resolution distributed hydrologic modeling: An example for the Rhine River. Water Resources Research, 56(4), p. e2019WR026807.
27. van Verseveld, W. J., Weerts, A. H., Visser, M., Buitink, J., Imhoff, R. O., Boisgontier, H., Bouaziz, L., Eilander, D., Hegnauer, M., ten Velden, C., and Russell, B.: Wflow_sbm v0.6.1, a spatially distributed hydrologic model: from global data to local applications, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-182, in review, 2022.