Implementation and industrialization of a deep-learning model for flood wave prediction based on grid weather forecast for hourly hydroelectric plant optimization: case study on three alpine basins

Veronica Brizzi [1], Andreas Bordonetti [2] , Michele Comperini [3] , Giulia Baccarin [4]

[1] *MIPU Energy Data s.r.l. SB, Via della Repubblica 42, Dueville (VI), ITALY*
[2] *Alperia Greenpower s.r.l., Via Claudia Augusta 161, 39100 Bolzano (BZ), ITALY*
[3] *Alperia Greenpower s.r.l., Via Claudia Augusta 161, 39100 Bolzano (BZ), ITALY*
[4] *MIPU Predictive Hub s.r.l. SB, Via della Repubblica 42, Dueville (VI), ITALY*

Abstract

In the present study a data-driven approach is used to simulate the behavior of 3 alpine basins for hydroelectric energy production. A deep feedforward neural network is used to predict 3 different scenarios of flood wave, simulated starting from the weather forecast on a specific area. The three models present low error in the simulation, their prediction is used to optimize the management of the hydroelectric plant bottoming the basin. With respect to a traditional approach, the data-drive method enables a higher precision, a real-time prediction and relies only on weather forecast and historical flow.

Keywords

Sustainability, Flood wave management, Flow prediction, Hydroelectric plant management, Weather forecast, Machine learning, Deep Learning, Industrial Artificial Intelligence, Cloudburst

1. Introduction

In the hydroelectric energy production field, which covers more than 34% of the renewable energy production in Italy with 48.786 GWh produced in 2018 [1], and almost 40% in Europe [2], the effect of climate change is tangible. On the alpine chain the last years have been characterized by droughts and cloudbursts with an increased intensity in terms of precipitations. In this scenario, the flood wave prediction is becoming more and more challenging, as the soil changes properties with an increased speed and the behavior of alpine basins becomes more difficult to predict. This situation is being registered on the Alps, as well as on other mountain chains. [3].

The traditional approach to predict flood wave, based on empirical observations from reference tables or explicit description based on analytics, is failing to guarantee a high precision in prediction, and the international researches are focusing on different approaches to predict high intensity events as they become more and more common.

Probabilistic approaches [4] or empirical methods based on the statistical observation of flow [5] have been explored during the last years, while machine learning and deep learning has proven to be an efficient method to predict short and medium term flow for river basins. [6,7] with R2 higher than 90%. A specific study from Innsbruck University focused on an alpine basin, confirming the possibility to model the flood

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy
EMAIL: veronica.brizzi@mipu.eu (A. 1);
andreas.bordonetti@alperia.eu (A. 2);
michele.comperini@alperia.eu (A. 3);
giulia.baccarin@mipu.eu (A. 4)
ORCID: 0009-0001-2129-7099 (A. 1)

wave in the basin starting from historical rain registrations [8]. By now, none of the deep learning solutions have been industrialized to provide near real time predictions based on weather forecast, nor used to manage consequently a hydroelectric plant in order to maximize the power production and minimize risk an environmental impact of the flood. The present study starts from existing researches to implement a deep learning model flexible enough to be used in a running solution and efficient enough to enable the hydroelectric plant management based on its predictions.

2. Method

The study focused on 3 different hydrographic basins located in Bolzano area, Italy:

- Fortezza
- Gioveretto
- Rio Pusteria

Analyzing the flow curve of each specific basin and the data structure of the weather forecast, the analysis focused on an hourly frequency, with a forecast of 27 hours for the execution.

The solution aims at predicting three different flow scenarios in the worst conditions of rain precipitation, in order to manage the volume of water arriving due to intense precipitations and decrease the stress on the hydroelectric plant and basin.

Models implemented use deep learning algorithms to simulate the behavior of the basin based on the weather forecast of the following hours on the whole surface of the hydrographic basin. Weather forecast are pre-processed in order to represent three precipitation scenarios.

Since the study focusses on predicting critical scenarios during cloudbursts, historical flow data are selected to be representative of flood waves.

2.1. Available data

For each basin, the committee provided historical data of flood waves, real-time data of the flow as measured from the SCADA as well as the shape file containing the area distribution of the hydrographic basins. Flow data are collected with an hourly frequency. Registered precipitations are provided as well.

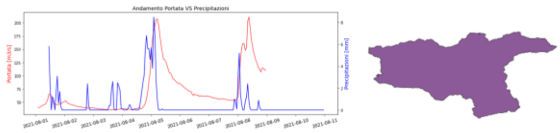

Figure 1. Example of historical flood wave (blue) with registered precipitation (red) and shape file provided, Fortezza basin.

The weather forecast is provided in grib format for the historical data, and netCDF format for real time data.

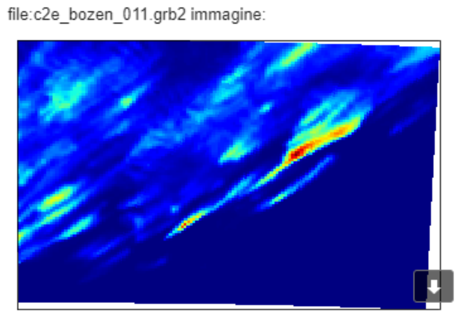

Figure 2. A single scenario of 21 for a single hour of 120 from Cosmo2 model, precipitation intensity

Two different weather forecast model have been explored during the study. Grib2 format files containing Cosmo1 model results are used for training and validation as they contain a higher amount of information, while netCDF format files containing Cosmo2 model results are used during the execution as are easier to preprocess. Both contain ensemble of equally-probable precipitation scenarios per hour, updated each 6 hours.

Table 1. Data structure for weather forecast models used

Weather model	Used for	N. scenarios	Hours of forecast	Grid [km]
Cosmo2	Training + Validation	21	120	2x2
Cosmo1	Execution	11	36	1x1

2.2. Data preprocessing for the training

The training is done on N-1 historical flood waves available for each basin, based on the best weather forecast from Cosmo2 models - the remaining flood wave is kept for the validation. Comparing the registered precipitation and the weather forecast of the same day, it emerged that the third quartile of the distribution for each scenario better represents the cloudburst and can be used for model implementation.

Weather forecast data in Cosmo2 models are preprocessed and merged in order to be transformed from grid-structure to time-series structure, following three main steps:

1. For each forecast the first 6 hours are selected and merged with the following forecast.
2. From the entire data structure of merged forecast are extracted 5 latitude-longitude points representing the basin area
3. For each hour and each point, the 21 values of precipitation are used to extract the third quartile of the distribution, obtaining 5 different time series of probable precipitation per point

The training dataset is structured for each flood wave by shifting both the historical flood data and the precipitation time-series of 23 hours, and including the cumulative sum of precipitation volume for the last 24 hours.

Each flood wave is treated individually in order to shift on the right axis without overlapping different periods, and then merged base on the basin.

2.3. Model implementation

For each basin a different model has been trained in order to maximize representativity and simulation performance, for a total of 3 different models.

The training period changes from basin to basin depending on the date and time of the specific floods.

Basin	Training period	N. of floods in the period
Fortezza	2021-08-01 02:00:00 - 2021-08-06 22:00:00	1
Gioveretto	2020-08-28 22:00:00 - 2020-10-06 00:00:00	3
Rio Pusteria	2020-08-27 11:00:00 - 2020-09-02 00:00:00	1

The algorithm and hyperparameters tuning has been done using a gridsearch approach, exploring different regression algorithms. The algorithm presenting the highest performance indicator (R2, MSE, MAPE) results in a feedforward neural network with a forced recursion in the execution loop. This algorithm enables a fast prediction during real-time use, as well as easy maintenance and management in its life cycle.

The FFNN has the following structure:

- 2 layers, 150 neurons on the first hidden layer and 185 on the second hidden layer
- Stochastic gradient descent solver
- hyperbolic tangent as activation function

The algorithm is trained based on 66% of observation extracted in a random way from the training dataset with the third quartile of each map point and the historical flood waves. The remaining 33% is used to test the model performances and avoid overfitting.

On the historical dataset, without the execution loop which guarantees recursion and is better detailed in the following paragraph, the test metrics are aligned with the expectation.

Table 2. Test metrics for each basin, using historical dataset and no recursion loop

Basin	R2 test	MSE test	MAPE test
Fortezza	99%	21.1	1.2%
Gioveretto	97%	22.2	0.9%
Rio Pusteria	99%	23.4	3.2%

3. Results

The model implemented for each basin is validated on a flood wave the model has never seen in order to test its capability of predicting new flood waves with a different behavior and evolution with respect to those used during the training.

3.1. Model validation

Differently form the training, each forecast is taken as a single forecast and not merged with the following. The point mapped during the training are used in validation and execution consistently. For each hour and point, the 21 values of precipitation are used to extract the third quartile of the distribution as well as the median and the whisker, representing respectively the most probable precipitation scenario, the best-case precipitation scenario and the worst-case precipitation scenario.

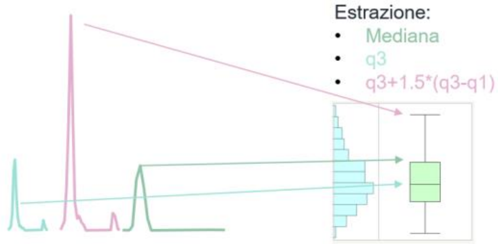

Figure 3. Extraction of probable, best-case and worst-case scenario from weather forecast

The test is done simulating the execution loop on the validation dataset.

The execution loop is structured in order to guarantee the recursion of the algorithm in a hard-coded way, externally from the algorithm structure. This is useful for a better control of the prediction.

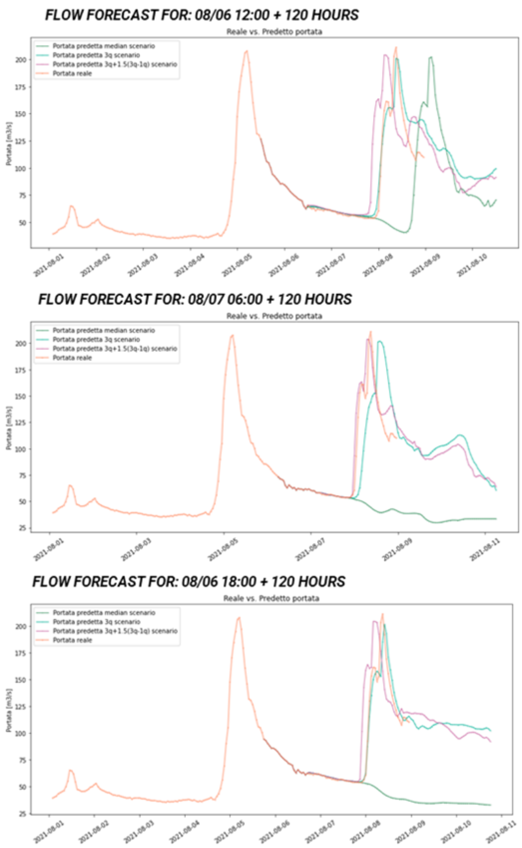

Figure 4. Validation for Fortezza basin in 3 different date and time and on the remaining flood wave.

The recursion loop runs for each hour of the forecast period (in the validation case, 120 times), and in each loop it copies the flow prediction of hour *t-1* in the input flow for hour *t*.

The prediction uses real flow values and predicted data with a different share in each additional prediction. Prediction starting at time *t* and predicting time *t+1* will use as input real flow rate and rain forecast from time *t-23* to time *t*. Prediction starting at time *t* and predicting time *t+10* will use as input rain forecast from time *t-13* to time *t+9*, real flow rate from time *t-13* to time *t*, and flow prediction done in previous iterations from time *t+1* to time *t+9*.

On the remaining flood wave three different validation are computed for each basin following the execution loop, simulating the prediction at three different date and time.

3.2. Model operation

With respect to the training and validation, during the execution Cosmo1 model is used to extract rain forecast. Data are preprocessed as seen during the validation, and the input dataset in structured accordingly. All models are retrained on the new weather forecast model to increase representativity of the prediction on the new data format.

The recursion loop runs only 30 times, which is the period of guaranteed availability of weather prediction (36 hours of forecast – 6 hours of forecast update periodicity). The results are shown for the following 27 hours, accordingly to the operative procedure of the study.

The model operates on cloud and in a proprietary platform. The architecture includes two different connectors to collect real-time flow data from the SCADA and weather forecast provided in a specific directory each 6 hours. The model runs hourly and for each hour prediction it uses the last weather forecast (updated 1 to 5 hours before) and the last flow value (updated hourly).

The flow prediction is sent to a hydroelectric plant simulator which tests different plant

Figure 5. Model operation in real-time, example of prediction

conduction opening and closing of bulkheads in order to maximize the exploitation of the flood wave minimizing risks and damages for the basin and the plant.

4. Conclusions

The proposed data-driven approach which simulates flood waves of a specific basin starting from rain forecast synthetized in 3 scenarios and using deep learning algorithms represents a pivot in the traditional method of flood management and flow prediction. The models for each basin are able to simulate the behavior of the flow during high-intensity precipitations with a high precision and an error on the peak of the flood wave as low as needed to manage the operations.

The main achievements of the study are:

- the proven possibility to model a generic basin starting solely from weather forecast and historical flow and flood waves
- the increased prediction performance in real-time which is able to absorb the intrinsic imprecision of weather forecast, thanks to the data preprocessing, the selection of a deep feedforward neural network as core algorithm, and the use of historical forecast during the training.

The study opens to new additional researches in the field, such as the use of convolutional neural networks to avoid time-series extraction from the weather forecast grid, the simulation of base flow in addiction to peak flow during flood waves, or the use of optimization algorithms to increase the precision of the hydroelectric plant management.

5. Acknowledgements

The intuition for the study and related funding has been provided by Alperia Greenpower after a study on the impact and frequency of recent flood waves in the basins under their control. The implementation work on the predictive model for flood wave has been completed by MIPU Energy Data industrial AI team, and industrialized on Rebecca platform.

Weather forecast based on Cosmo1 and Cosmo2 models is provided by MeteoSwiss.

This Word template was created by Aleksandr Ometov, TAU, Finland.

6. References

[1] Rapporto Statistico 2018 – Fonti Rinnovabili, GSE

[2] Renewable power generation costs in 2021, IRENA

[3] Environmental and economic impact of cludbursts-triggered debris flows and flash floods in Uttarakhans Himalaya: a case study, Vishwambhar Prasad Sati, Saurav Kumar Flood forecasting using medium-range probabilistic predictions

[4] DESIGN HYDROGRAPH ESTIMATION IN SMALL AND UNGAUGED BASINS. The EBA4SUB Software and Framework, Rodolfo Piscopia, Andrea Petroselli, Salvatore Grimaldi

[5] A deep learning approach for hydrological time-series prediction: A case study of Gilgit river basin - Dostdar Hussain, Tahir Hussain, Aftab Ahmed Khan, Syed Ali Asad Naqvi & Akhtar Jamil Earth Science Informatics volume 13, pages915–927 (2020)

[6] Daily reservoir inflow forecasting using weather forecast downscaling and rainfall-runoff modeling: Application to Urmia Lake basin, Iran Amirreza Meydani, Amirhossein Dehghanipour, Gerrit Schoups Massoud Tajrishy

[7] Sonderbetrieb für Bodenschutz, Wildbach- und Lawinenverbauung Autonome Provinz Bozen Südtirol Erstellung eines Niederschlagabflussmodells zur Bewertung eines möglichen Hochwasserschutzes durch Retention am Beispiel der Stauanlage Welsberg