Predication of missed monthly rainfall time series by using python-based ANNs: A case study of Porvoonjoki watershed, Finland.

1. Introduction

Artificial Neural networks(ANNs) application in hydro-climatological studies has been one of the main and growing approaches in modeling, prediction, and processing of long-term time-series such as rainfall-runoff processes across the hydrological units of different size such as Basin, watershed, and catchment areas (Salas et al.2000; Karamouz et al. 2008; Mahmoud Zemzami 2016). In this task, we will evaluate the ANNs performance in the forecasting of Monthly precipitation time series-which will be supposed to be missed or ungagged quantity of interest – through a training, validation, and testing scenario of ANNs (Nkuna, T.R., Odiyo and J. O 2011; Canchala-Nastar et al. 2019). Extracted time-series could be used by other python based Hydrological models which dependent on time-series data (D.J. Lampert, M. Wu 2015).

2. Problem Formulation(Definition)

Rainfall data as the main input of hydrological models may be collected in different temporal accuracies containing missed or ungagged values during some observation periods (Mohamed-Aymen Ben Aissia et al. 2017). So, in this task, we will try to formulate simple ANNs to evaluate their performance in the prediction of missed precipitation data on the monthly scale by using collected data from official hydro-climatological stations (Mislan et al. 2015). Also, we have selected the Porvoonjoki watershed -situated in the south region of Finland -and its three nearby meteorological stations as the study area.

3. Method

3.1. Study area and data collection

Meteorological data collected for this task belongs to the Porvoonjoki watershed (Fig 1) which is located in the south area of Finland (Veijalainen et al. 2010).

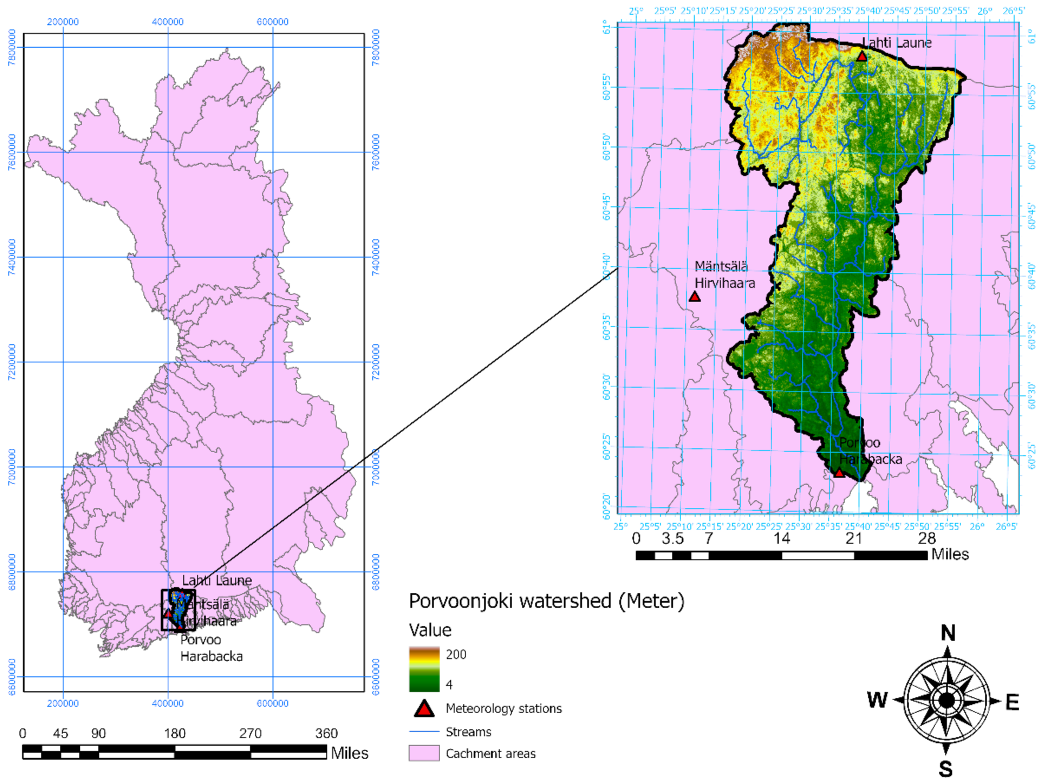

Fig 1: Location of Porvoonjoki watershed

Precipitation data used in this task has been collected from the database of three meteorological stations of Finish Meteorological Institute (FMI) including daily observed rainfall in mm. The location of stations was selected according to data availability and distance from the Porvoonjok watershed (Table 1).

Observation station	Station ID	Latitude	Longitude	Elevation	Usage type in this task
Lahti Laune	101152	60.96211	25.63090	85	Data points
Porvoo Harabacka	101028	60.39172	25.60730	22	
Mäntsälä Hirvihaara	103794	60.62773	25.19342	83	Quantity of interest (lables)

Table 1: observation Station used in this task.

3.2. Artificial neural networks

The machine-learning algorithm used in this task is an ANNs consists of two hidden layers with three neurons and an output layer with one neuron (Fig 2).

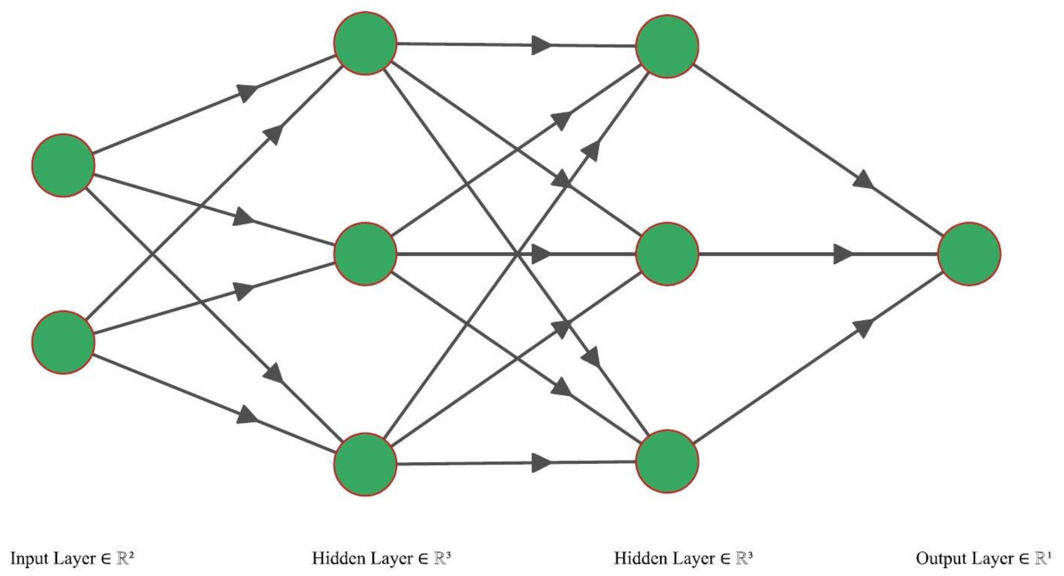

Fig2: the architecture of neural network used in the task.

3.3. Preprocessing and python

Original precipitation data gotten from the database of FMI was containing data gaps and unformal to handle and to import into the Python. So, some part of the Pre-processing has been accomplished by Arc GIS and excel, and some others are left to accomplish through python programming. Libraries that were being used to write the python codes were pandas, NumPy, Matplotlib, TensorFlow, seaborn, and Scikit-learn.

3.4. Training, validation, and testing

In this task daily rainfall data during 2001-2020 has been utilized to train, validate, and test the machine learning algorithm. After data preprocessing this data has been resampled to monthly data by Pandas in python and finally, we have 240 months (data point) to implement in the project. this total data point has been divided into three parts for each step of learning (training, validation, and testing) by using the Scikit-learn library and finally, we have 150 data points for training,50 data point for validation, and 40 points for the testing step.

4. *Results*

4.1. Training

In this task training of selected ANNs has been accomplished by the TensorFlow and the Keras library for the first 150 months (training data set) from 1/31/2001 up to 6/30/2013. obtained results for 500 epochs are as the flowing Fig 3 and Table 2:

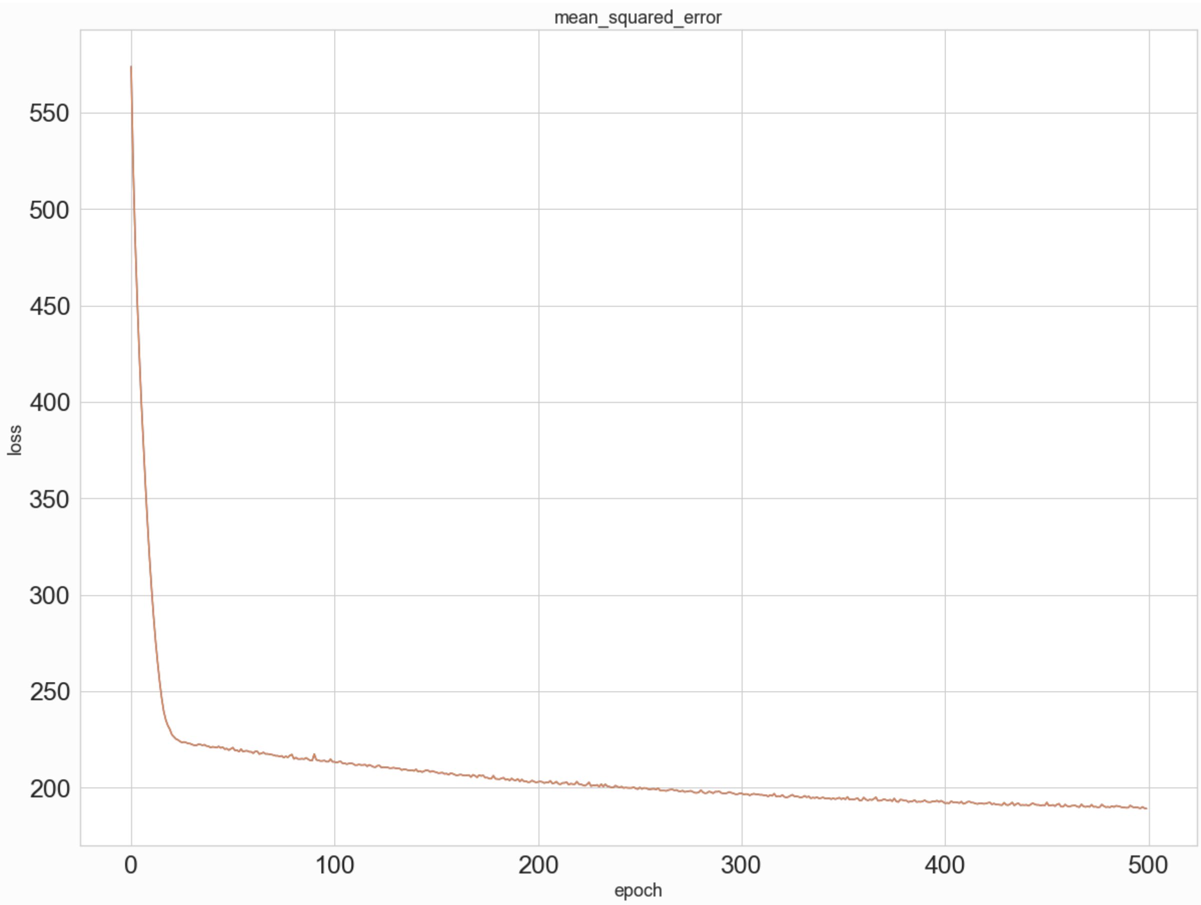

Fig 3: optimization results for trained ANNs.

```
Layer (type)                 Output Shape              Param #
=================================================================
hidden                             (3)                       9
_________________________________________________________________
hidden                             (3)                       12
_________________________________________________________________
output                             (1)                       4
=================================================================
Total params: 25
Trainable params: 25
Non-trainable params: 0
```

Table 2: obtained results for trained ANNs.

4.2. Validation

during the validation stage, monthly rainfall values from 7/31/2013 up to 8/31/2017(50 data points) have been inserted into the algorithm as input data. Obtained results are depicted by (Fig 4).

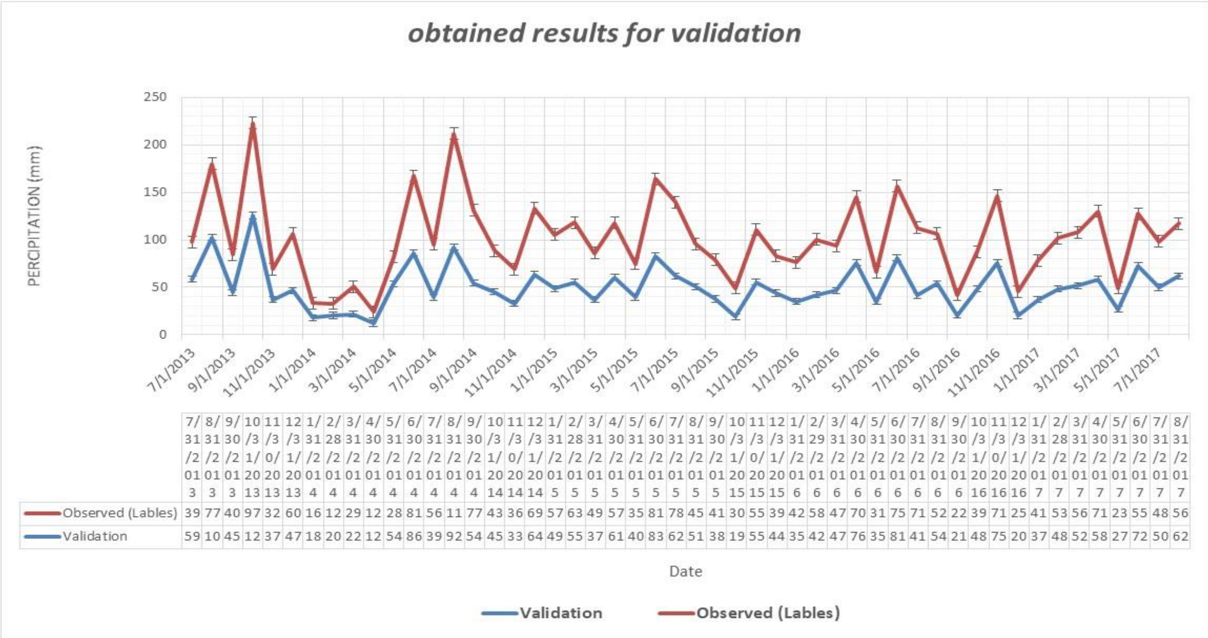

Fig 4: Obtained results for validation.

4.3. Test

during the testing stage, monthly rainfall values from 9/30/2017 up to 12/31/2020 (40 data points) have been inserted into the algorithm as input data. Obtained results are depicted in (Fig 5).

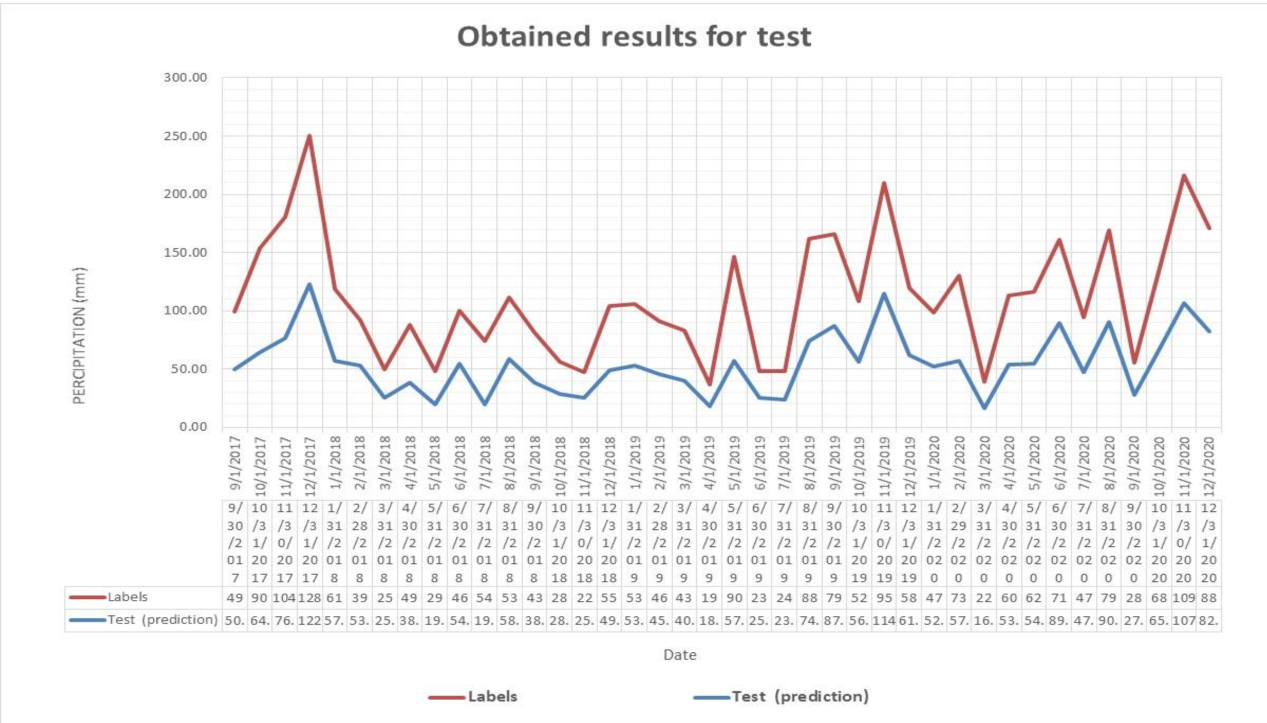

Fig 5: Obtained results for the test.

5. Conclusion

The task had aimed to evaluate and examine the application of ANNs to predict the missed time series of a watershed area located in the south region of Finland(Coşkun Hamzaçebi 2008). Fig 6 is depicting the total time series for 240 months during the study period. According to obtained results, the model can approximately predict the target values (labels) and obtained results can be improved by selecting more complex ANNs and other scenarios(Zeynep Idil Erzurum Cicek et al.2021; Hansika Hewamalage et al.2021).

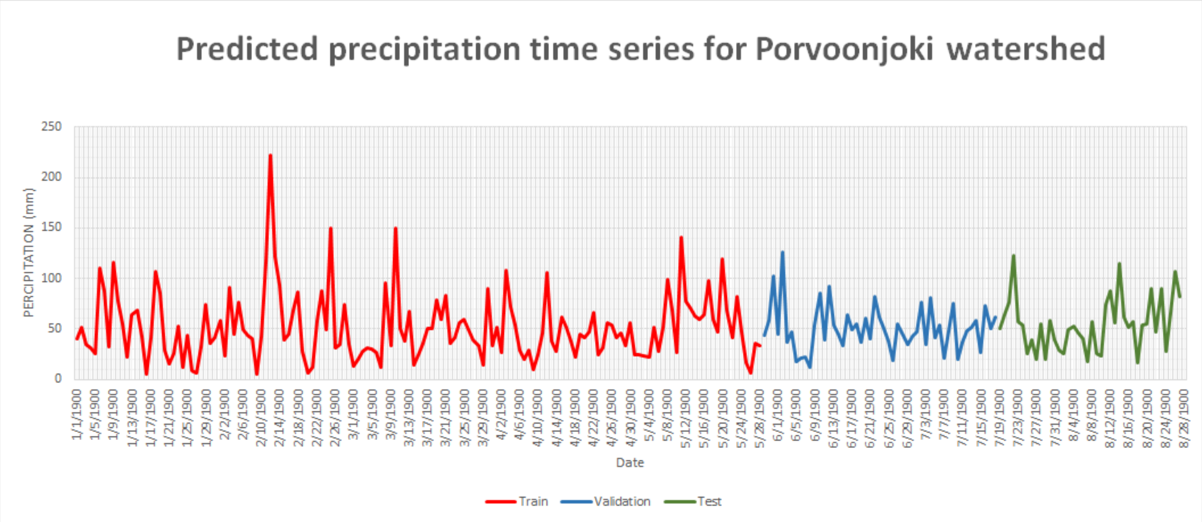

Fig 6: predicted precipitation time series of Porvoonjoki watershed

References

1. Salas, J.D., Markus, M., and Tokar, A.S., 2000. Streamflow forecasting based on artificial neural networks. Artificial neural networks in hydrology. *Water Science and Technology Library, Springer*, 36, 23–51.
2. Mahmoud Zemzami. (2016) Flow forecasts in a non-perennial river of an arid basin using neural networks. Journal of Applied Water Engineering and Research 4:2, pages 92-101.
3. Karamouz, M., Razavi, S., and Araghinejad, S., 2008. Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study. *Hydrological Processes*, 22, 229–241. doi:10.1002/(ISSN)1099-1085.
4. Nkuna, T.R., Odiyo, J.O. Filling of missing rainfall data in Luvuvhu River Catchment using artificial neural networks, Physics and Chemistry of the Earth, Volume 36, Issue 14-15, 2011.
5. Canchala-Nastar et al.Estimation of missing data of monthly rainfall in southwestern Colombia using artificial neural networks, Data in Brief, Volume 26, October 2019.
6. D.J. Lampert, M. Wu, Development of an open-source software package for watershed modeling with the Hydrological Simulation Program in Fortran, Environmental Modelling & Software, Volume 68, 2015, Pages 166-174, ISSN 1364-8152.
7. Mohamed-Aymen Ben Aissia, Fateh Chebana, Taha B.M.J. Ouarda, Multivariate missing data in hydrology – Review and applications, Advances in Water Resources, Volume 110, 2017, Pages 299-309, ISSN 0309-1708.
8. Mislan, Haviluddin, Sigit Hardwinarto, Sumaryono, Marlon Aipassa, Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan - Indonesia, Procedia Computer Science, Volume 59, 2015, Pages 142-151, ISSN 1877-0509.
9. Noora Veijalainen, Eliisa Lotsari, Petteri Alho, Bertel Vehviläinen, Jukka Käyhkö, National scale assessment of climate change impacts on flooding in Finland, Journal of Hydrology, Volume 391, Issues 3–4, 2010, Pages 333-350, ISSN 0022-1694.
10. Coşkun Hamzaçebi, Improving artificial neural networks' performance in seasonal time series forecasting, Information Sciences, Volume 178, Issue 23, 2008, Pages 4550-4559, ISSN 0020-0255.
11. Zeynep Idil Erzurum Cicek, Zehra Kamisli Ozturk, Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting, Applied Soft Computing, Volume 102, 2021, 107091, ISSN 1568-4946.
12. Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, Recurrent Neural Networks for Time Series Forecasting: Current status and future directions, International Journal of Forecasting, Volume 37, Issue 1,2021, Pages 388-427, ISSN 0169-207.