

Mixing the RiverLab and Deep Learning for augmenting Critical Zone monitoring

ANTOINE DOLANT¹, PAUL FLOURY¹, ARNAUD BLANCHOUIN², ROMANE NESPOULET³ AND LAURE CORDIER⁴

¹Extralab

²INRAE - U.R. HYCAR

³INRAe Antony - UR HYCAR

⁴IPGP

Presenting Author: dolant@extralab-system.com

The CZO of Orgeval is one of the most instrumented observatory. It is representative of actual issues, such as the environmental impact of intensive agriculture. Data is being recorded every two days since 1969 (50 years), and more recently (2015) every 30 minutes for major elements and every minute for physico-chemical parameters with the installation of a RiverLab. That makes more than 2.2 million records of probes data and more than 30,000 records of major elements in 6 years. Besides, Deep Learning are state-of-the-art AI algorithms omnipresent in almost all industries nowadays; they can detect and highlight trends and patterns by *learning* on data, and reproduce them for different applications (e.g. Tesla autopilot, Pl@ntnet's plant recognition, music generation, insurance scoring). Amongst the variety of different neural network combinations, Recurrent Neural Networks (RNNs) are rather fit to learn temporal-dependant data because of their cyclic neuronal structure, particularly the Long Short-Term Memory model (LSTM). LSTM is widely used in research applications that involve time series forecasting, such as weather forecasting [1], wind power forecasting [2], rainfall-runoff modelling [3] etc... The coupling of the Deep Learning's predicting power and the CZO's high frequency and historic data shows a strong potential for reproducing the behaviours of the measured elements through time. We discuss here the extent and limits of DL applied to geochemical problematics. We review the predictions through different time spans (10min, 1 day, 1 month, 1 year) into the future or in the past, and for different input parameters (size of the trainset, size of the input values, etc...). For instance, to predict the next 6 hours of Conductivity with a controlled error (MAE: 4.0 $\mu\text{S}/\text{cm}$, RMSE: 5.52 $\mu\text{S}/\text{cm}$, MAPE: 0.49%, R² Score: 0.99), the best parameters are a lookback of 18 hours, and an input dataset of hourly 31,000 values. The models are improved with a grid search of the best parameters, and evaluated with statistical metrics (MAE, RMSE, MAPE, sMAPE, R² Score...) as well as a visual valuation of the signal.

[1] <https://ieeexplore.ieee.org/abstract/document/7415154>

[2]

<https://www.sciencedirect.com/science/article/abs/pii/S03062619>

[3] <https://hess.copernicus.org/articles/22/6005/2018/>