Application of deep learning algorithms for the forecasting of electrical conductivity under different saline hydrogeochemical environments

V. Pisinaras[a], A. Nikolaidou[b], T. Semertzidis[b], P. Daras[b], C. Güler[c], M.A. Kurt[d]

[a] Soil and Water Resources Institute, Hellenic Agricultural Organization, Greece
[b] Center for Research and Technology - HELLAS, Information Technologies Institute, Greece
[c] Mersin University, Çiftlikköy Campus, Engineering Faculty, Department of Geological Engineering, Turkey
[d] Mersin University, Çiftlikköy Campus, Engineering Faculty, Department of Environmental Engineering, Turkey

INTRODUCTION

Groundwater constitutes the **primary, if not the only source of freshwater** in semi-arid coastal areas, which are experiencing rapid population and economic development. Therefore, **efficient groundwater management in coastal aquifers** is of **paramount importance** for socio-economic stability, but at the same time is challenging, since **groundwater salinization (GS) phenomenon in coastal aquifers is a complex process** driven by a wide variety of natural and anthropogenic stressors. The GS involves different mechanisms, including **seawater intrusion due to over-exploitation** of groundwater resources, **existence of trapped seawater** or connate saltwater, **agricultural return flow** and **dissolution of evaporites**, which **might co-exist**, thus increasing furthermore the complexity of simulation of GS and subsequently groundwater management. **Groundwater electrical conductivity** (EC) constitutes a physicochemical variable that can **efficiently represent GS** and it is widely applied in **operational groundwater management**, since it constitutes **a measure of water suitability** related to water uses (agricultural, domestic, etc.). The last years, artificial Intelligence (AI) has been widely used in modeling of non-linear hydrological systems. Based on the above, **our study aims to simulate and forecast EC in two aquifer systems located in Mediterranean Region, which experience complex GS processes based on 3 deep learning algorithms**.

STUDY AREAS

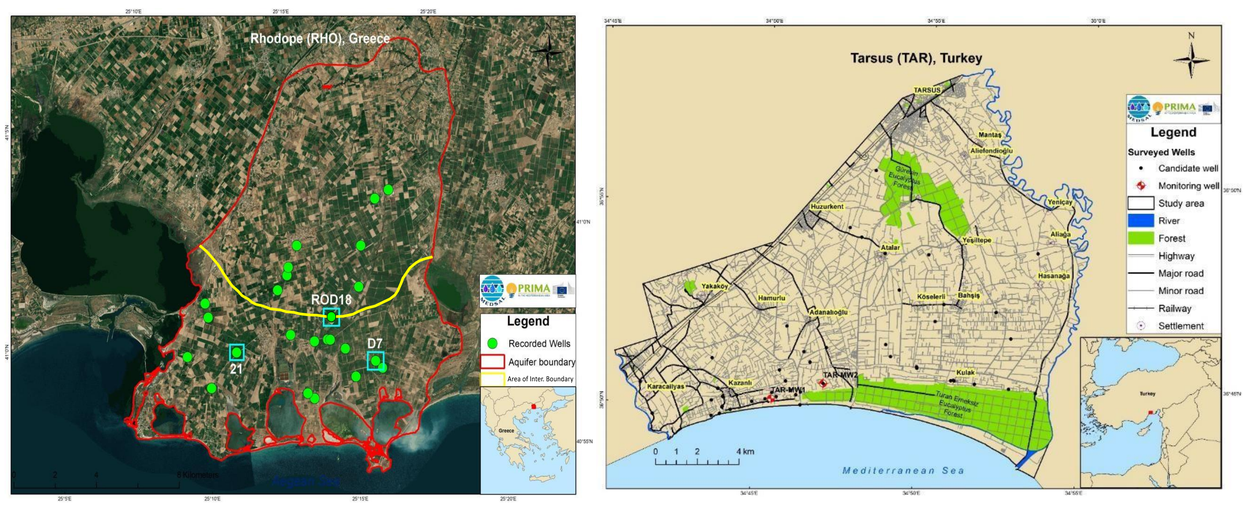

Fig. 1. Location maps of the 2 study areas. The GW monitoring wells are also presented

RHODOPE aquifer, GREECE

- Area = **120 km²**
- Porous aquifer: mainly **Miocene alluvial deposits**, thickness up to **200 m**
- **Agricultural activities** are dominating followed by **touristic activities** along the coast
- **More than 600 GW wells**
- Groundwater salinization mainly due to **seawater intrusion**, **geogenic factors** and **trapped saline lenses**
- **3** Electrical Conductivity, Temperature, Depth **sensors** installed in **operating GW wells**

TARSUS aquifer, TURKEY

- Area = **240 km²**
- Porous aquifer: **Quaternary-Recent fluvio-deltaic** coastal aquifer, thickness up to 500 m
- **Up to 2,000 GW wells**
- Increasing agricultural and industrial activities
- Groundwater salinization due to by **seawater intrusion**, **water-rock interaction** and **dissolution of evaporitic series**
- **2** Electrical Conductivity, Temperature, Depth **sensors**, pH, DO, REDOX sensors in **non-operating GW wells**

METHODOLOGY

Daily data of Precipitation (**P**), Air temperature (**T_a**) from a local meterological station and Electrical Conductivity (**EC**), Water temperature **T_w** and Groundwater level (**GWL**) from the installed sensors were used in order to **train AI models** that **forecast EC** based on the following algorthms:

- ❑ **Multilayer Perceptron (MLP)**
- ❑ **Bidirectional LSTM Network (BiLSTM)**
- ❑ **Convolutional Neural Network (CNN)**

One year of data was used for **models' training** in which **Gaussian noise** algorithms were applied in order to increase the training dataset. **One year of data** was used for **models' performance assessment** using standard metrics such as correlation coefficient (**R^2**), Nash-Suttclife Efficiency (**NSE**), Root Mean Squared Error (**RMSE**), Mean Absolute Error (**MAE**), **BIAS** and relative BIAS (**rBIAS**)

The following combinations of variables were used as input for EC forecasting, considering the last 14 days as input time window:

Table. 1. Combinations of input variables used for models' training

Input Setup ID	Parameters included	Input Setup ID	Parameters included
E0	P	E8	EC, P
E1	T_a	E9	EC, T_a
E2	T_w	E10	EC, GWL
E3	GWL	E11	EC, T_w
E4	EC	E12	EC, P, T_a
E5	P, T_a	E13	EC, P, T_a, GWL
E6	P, T_a, GWL	E14	EC, P, T_a, GWL, T_w
E7	P, T_a, GWLP, T_w		

RESULTS

RHODOPE

The results of selected metrics average values for AI models developed for **RHODOPE** aquifer considering all sensors are presented as heatmaps in Fig. 2, while selected forecasting time series that consider the mean performance of 10 different model initializations are presented in Fig. 3

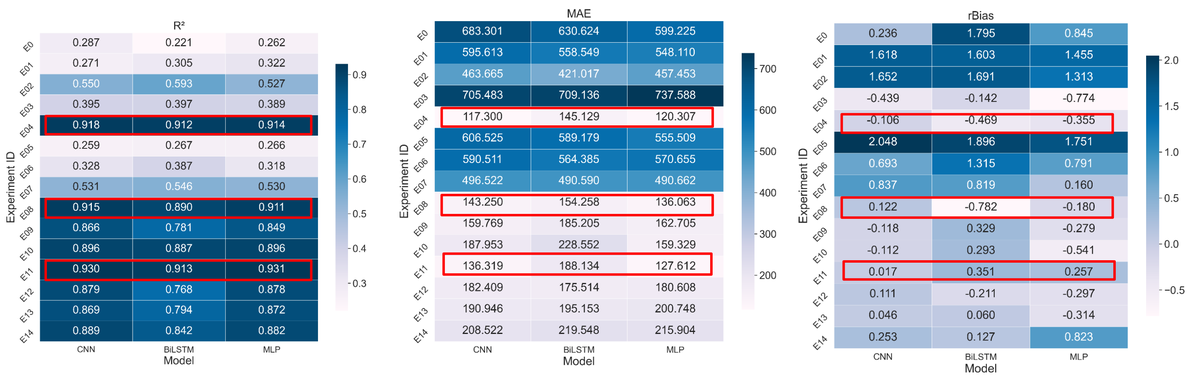

Fig. 2. Heatmaps of selected metrics for the performance assessment of AI models application in RHODOPE

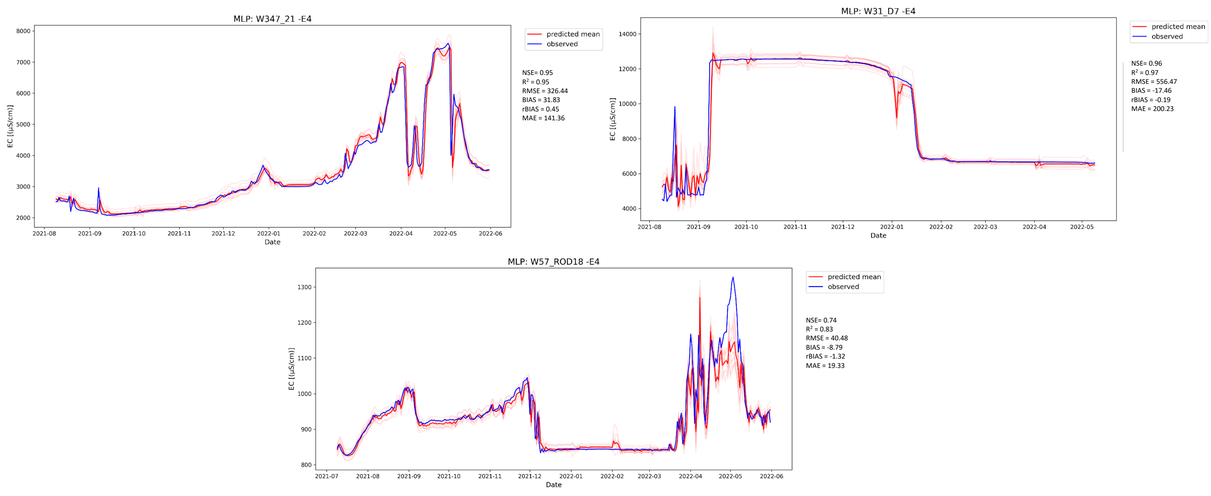

Fig. 3. Times series of EC forecasting for the 3 sensors in RHODOPE using MLP algorithms and "E4" input variables setup

TARSUS

The corresponding results for **TARSUS** aquifer are presented in Figs. 4 and 5

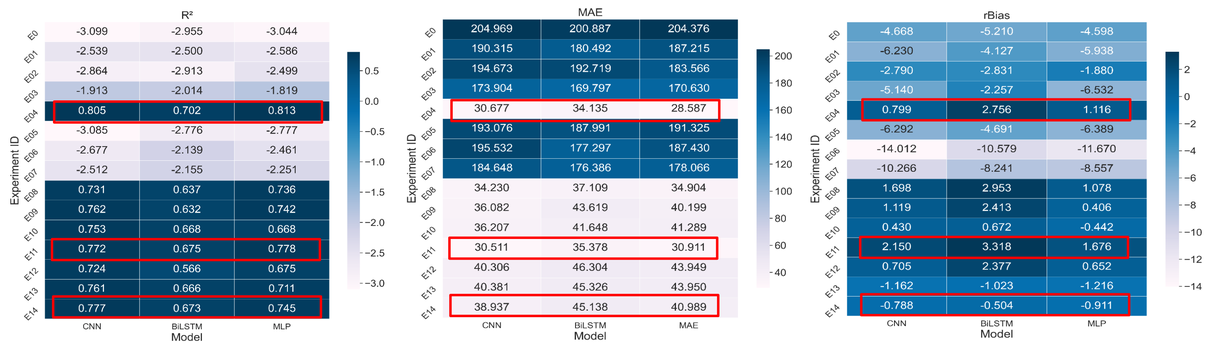

Fig. 4. Heatmaps of selected metrics for the performance assessment of AI models application in TARSUS

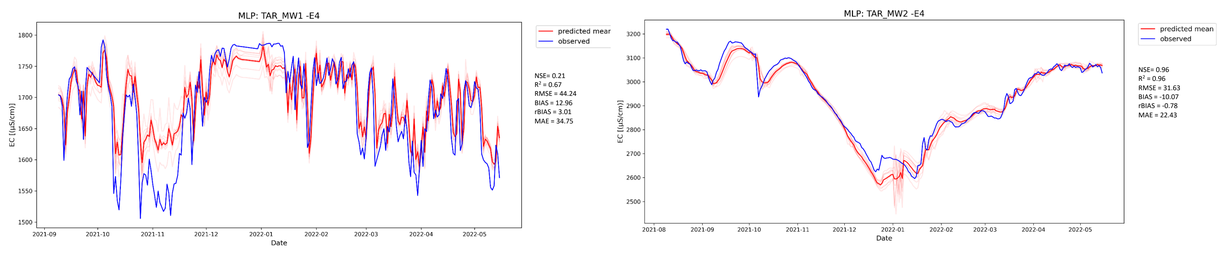

Fig. 5. Times series of EC forecasting for the 3 sensors in TARSUS using MLP algorithms and "E4" input variables setup

MAJOR OUTCOMES

- ❑ **Models' performance** was **satisfactory only when including the target variable (EC)** in input variables. Meteorological variables cannot efficiently simulate EC variation, either in operating (RHODOPE) or non-operating (TARSUS) GW wells.
- ❑ Using as **input variables** past **EC** and **EC-Tw** data proved to be the **best input variables combinations for both areas**.
- ❑ **Very good model performance** was achieved with **EC and P for RHODOPE** and **EC, P, T_a, GWL, T_w for TARSUS** as input variables. The **inclusion of hydrological and meteorological variables** into a groundwater time-series simulation model can **improve the physical significance of the models**.
- ❑ **CNN and MLP performed slightly better than BiLSTM**, but in general, the performance of the models was similar.
- ❑ **Sufficient model performance** was achieved when training the models with **1 year of daily data** which was **further augmented by applying Gaussian noise**.
- ❑ **More data is needed in order to assess overfitting** and **improve models training** in order to capture different EC variation patterns driven by differences in climate and water resources management conditions.

ACKNOWLEDGEMENTS

This research was supported by the MEDSAL Project © (www.medsal.net), which is part of the PRIMA Programme, supported by the European Union's Horizon 2020 Research and Innovation Programme and funded by the Greek national funding agency - GSRT (grant number 2018-7) and TÜBİTAK (Scientiêc and Technical Research Council of Turkey, No: 118Y366).