PREDICTING WATER TABLE DEPTH

Developing an LSTM based Prediction Model for Agricultural Areas

(Implementation: github.com/Akshayy99/Water_Table_Depth_Prediction_Pytorch)

AKSHAY KHARBANDA
20171037

TABLE OF CONTENTS

PROBLEM

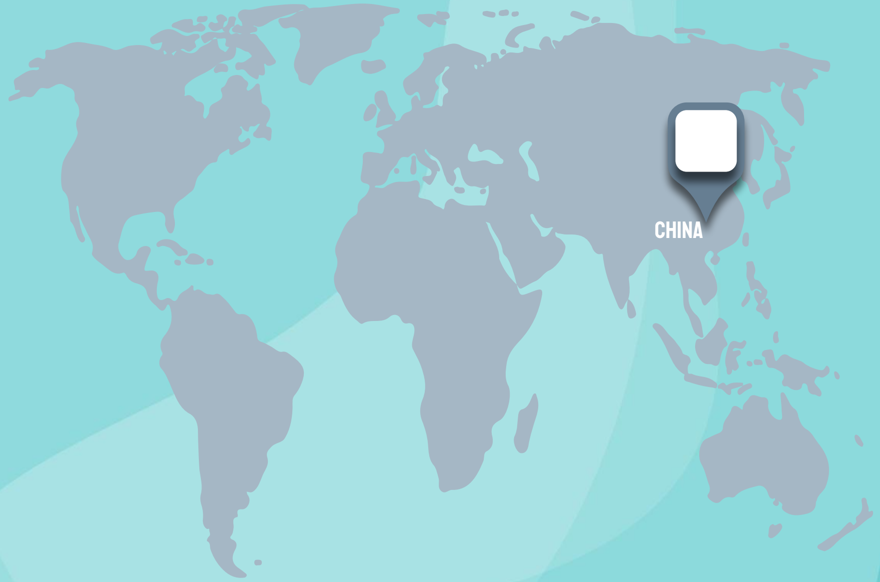
- Groundwater resources are vulnerable to **overexploitation**, **climate change**, and **biochemical pollution**.
- As a result, many areas over the world face groundwater shortages.
- The effective management of groundwater resources is necessary to:
  - develop optimal water resource allocation strategies
  - adjust crop patterns
  - develop optimal irrigation schedules

01

OBJECTIVE

OBJECTIVE

“Develop an effective and accurate method for predicting water table depth that can be used to help engineers and decision-makers manage groundwater resources and make management decisions.”

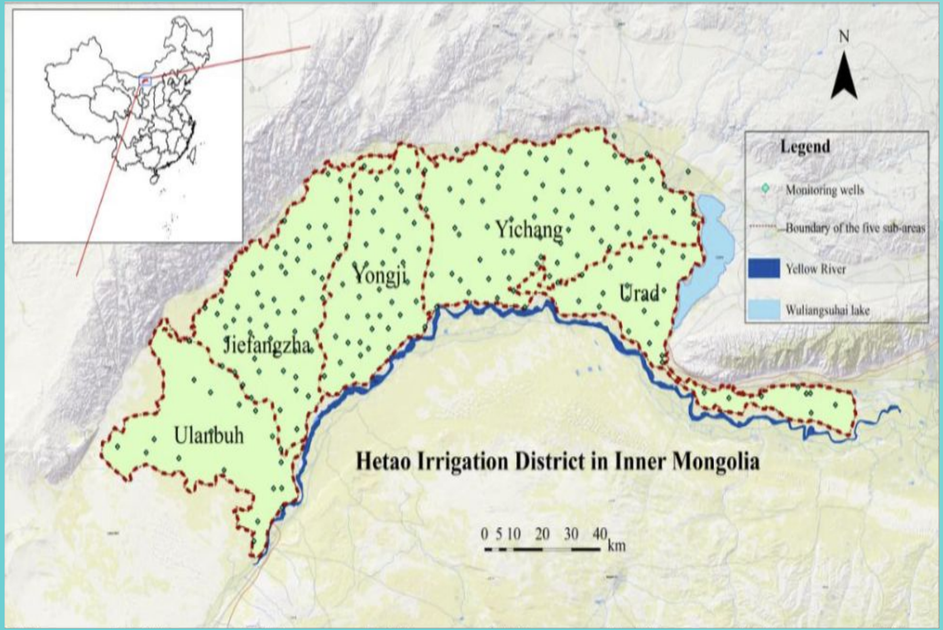


02

STUDY AREA AND DATA

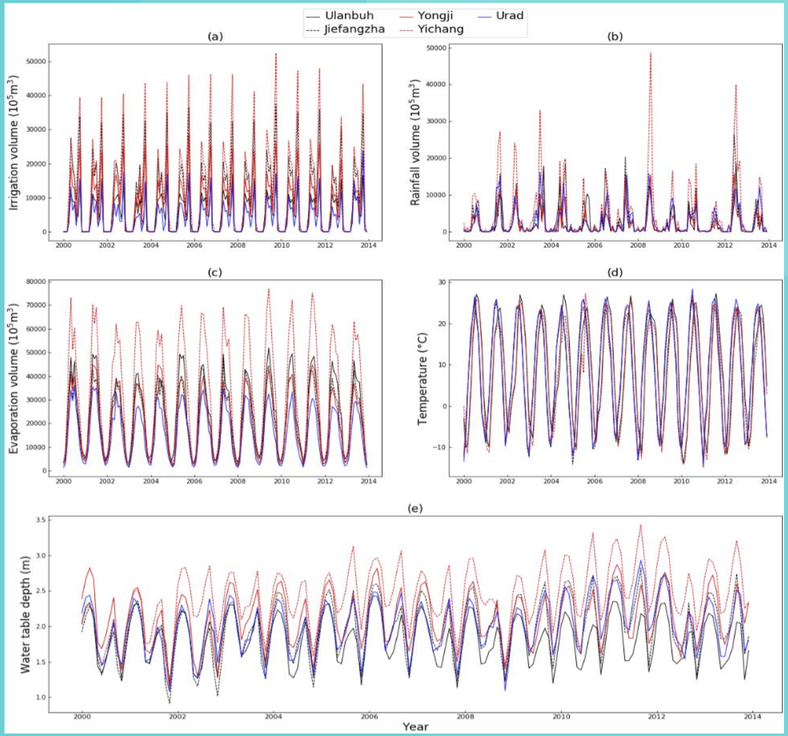
STUDY AREA (HETAO IRRIGATION DISTRICT, CHINA)

- Located in the arid area of the **Yellow River watershed**.
- District has been divided into five sub-areas from west to east as follows: Ulanbuh, Jiefangzha, Yongji, Yichang and Urad.
- Hetao Irrigation District covers 11,073 km^2; the central part of the district is about 180 km long and 60 km wide.



DATA AND STATISTICAL ANALYSIS

- I used 14 years (2000–2013) of time series data from the five sub-areas in Hetao Irrigation District.
- Six variables were chosen for this study: water diversion, precipitation, evaporation volume, temperature, water table depth, and time.
- Training set: first 12 years' time series data.
- Validation set: last two years of data.
- All six of the variables were standardized to ensure they remain on the same scale (Data pre-processing).



Statistic description of water table depth of five sub-areas.

AREAS	Average	Max (m)	Min (m)	Standard Deviation	Skewness
ULANBUH	1.817	2.430	1.133	0.331	-0.167
JIEFANGZHA	1.970	2.821	0.914	0.422	-0.140
YONGJI	2.111	2.870	1.129	0.379	-0.108
YICHANG	2.421	3.430	1.202	0.429	-0.222
URAD	2.002	2.932	1.080	0.375	-0.103

03

METHODOLOGY

LSTM WITH DROPOUT

1. Like RNNs, LSTMs also have chain-like modules.
2. Each repeating module of LSTMs contains a **memory block**. The memory block contains four parts: a CEC (the Constant Error Carousel) cell in addition to three special multiplicative units called gates.
3. LSTMs learn long-term dependencies more easily than RNNs because information can easily flow along the cells unchanged.
4. DNNs with large numbers of parameters can easily be **overfitting**, dropout provides an effective regularization method to solve this problem.
5. At each training iteration, when the neural network is updating a certain layer where the dropout is applied, it randomly does not update, or "dropout" some neurons in this layer.

OUR PROPOSED FRAMEWORK

1. Predicting water table depth is a **time series problem** since the current water table depth has changed in a way that is dependent on previous. This time series prediction was cast as a regression problem.
2. The input data were first put into the LSTM layer. The input gate of the LSTM layer will recompose input data and decide which input data is important (similar to PCA).
3. A fully connected layer is set atop the LSTM layer in order to improve the model's learning ability (since the fitting ability of just the LSTM layer might be insufficient).
4. Moreover, dropout is set on the LSTM layer to prevent overfitting.

MODEL EVALUATION CRITERIA

In this study, the Root Mean Square Error (RMSE), and coefficient of determination (R2) were used to evaluate the model’s accuracy.

1. R2 measures the degree of how well the outcomes are replicated by the model, ranging between [-∞, 1] where for optimal model prediction an R2 score close to 1 is preferred.
2. RMSE measures the prediction precision which creates a positive value by squaring the errors. RMSE scores range between[0, ∞], and the model prediction is ideal if RMSE is 0.

04

RESULTS

DEPTH PREDICTION RESULTS

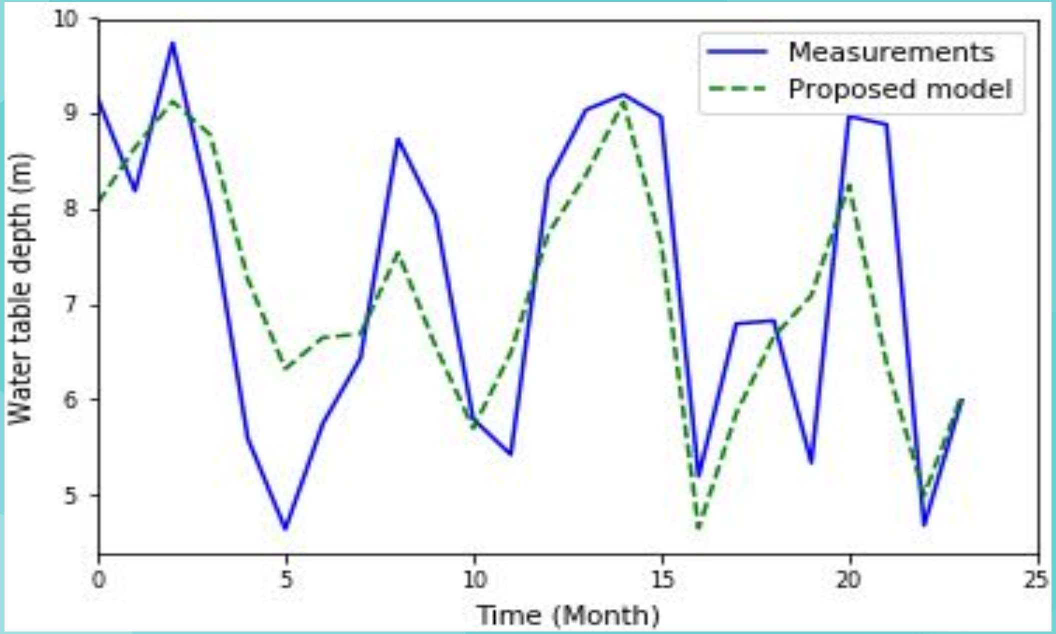
- The various hyperparameters were tweaked and tested out, and from the results, the optimal values for the hyperparameters (R2 = 0.82) was found out to be the following:
    - No. of Neurons: **40**
    - Learning Rate: **10^{-4}**
    - Dropout: **0.5**
    - Iterations: **20000**
- A higher learning rate (10^{-3}) may cause the optimization task to miss the optimal point (R2 was only 0.62), while a lower learning rate (10^{-5}) may help to avoid overshooting but the model would take longer to converge.

RESULTS (CONTINUED..)

- Too many training iterations may not ensure more optimal results. In addition, it is common to set a dropout probability to 0.5 in the deep learning field.
- Too many neurons are computationally expensive and often cause overfitting, while having an insufficient number of neurons (10 in this study) may decrease the network's learning ability.

RESULTS (CONTINUED..)

The Pytorch implementation of this LSTM shows that this model is able to predict the water table depths quite accurately, evident from the below output of the implementation.



05

CONCLUSION

CONCLUSION

1. The proposed model provides a promising method for predicting water table depth, as evidenced by satisfactory performance on water table depth prediction in five sub-areas and “Hetao”.
2. The architecture of the proposed model is reasonable. The LSTM layer helps to maintain previous information and contributes to learning the time series data. The dropout method helps to prevent overfitting during the training process.
3. The model can serve as an alternative model to predict water table depth in places with complex hydrogeological characteristics and hydrogeological data are difficult to obtain.

THANKS!