Predicting groundwater levels in the southeastern United States with the use of Long Short-Term Memory (LSTM) neural network

Mauricio Osorio Gonzalez, Georgios Boumis (Team D2)

Department of Civil, Construction, and Environmental Engineering

University of Alabama, Tuscaloosa, AL, USA

Abstract

For many households in the United States, groundwater is the only source of potable water. Additionally, humans are not the only species that rely on it, with entire ecosystems being dependent on groundwater. To achieve sustainable management of groundwater resources, good knowledge of present and future groundwater levels is required. In this study, a Long Short-Term Memory (LSTM) model was developed by making use of three climate variables in order to forecast groundwater levels in Alabama, USA. Results suggest that the proposed model is able to forecast groundwater levels up to 28 days ahead with ample skill, providing a reliable alternative to traditional physical models that require vast amounts of data.

1 Introduction

Groundwater is the primary source of drinking and industrial water supplies for many areas throughout the world. In the United States (US), past estimates assess that groundwater provides approximately 40% of the public water, while more than 40 million people, mainly in rural areas, rely on domestic wells for their drinking water needs [1]. Particularly in Alabama (AL), 27 of 36 south Alabama counties receive all of their public water supplies from groundwater sources [2]. Knowledge of groundwater levels (GWL) in aquifers is an imperative issue for smart and sustainable management of groundwater resources. Continuous records of GWL time series often embed crucial information about groundwater dynamics [3]. As a result, it is necessary for water managers and engineers to have access to accurate predictions of GWL in order to make more informed decisions.

Traditional GWL modeling has relied on the development of conceptual or physically-based models. However, such models often suffer from extensive and computationally expensive parameter calibration, as well as uncertainty associated with the representation of the physical processes involved. Most of the times, especially for small groundwater users/managers, it is not feasible to develop and maintain physics-based groundwater models, that are useful to inform management decisions. The advent of artificial intelligence (AI) and particularly deep learning (DL) algorithms, has provided an alternative approach to predict GWL in a much simpler and economic way than traditional modeling, which most of the times provides comparable or even more adequate results.

Rajaee et al. [4] provide a comprehensive overview of studies (67 articles) concerning the forecasting of GWL with AI, listing methodologies, the temporal resolution of the GWL forecasts, study areas, and the input variables (predictors) used . Figure 1 suggests that the prediction of GWL with the use of AI remains a greatly understudied problem. Indicatively, for the US there had been only 6 published studies up until 2019. To our knowledge, there are no studies that tackle GWL forecasting by making use of LSTM in the southeastern US.

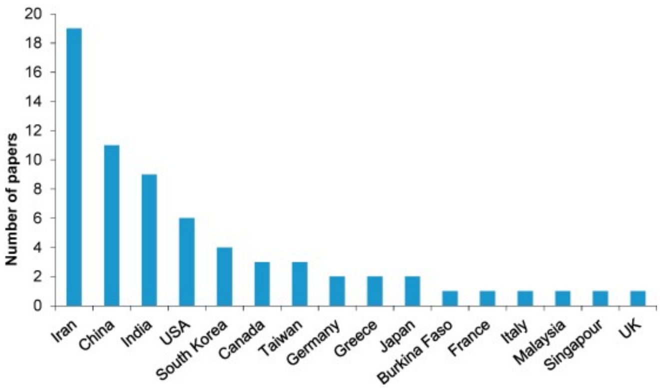

Figure 1: Number of published papers (up to 2019) regarding the use of AI for GWL prediction. Source: [4].

Here, we propose an LSTM model for short to long-term GWL forecasting for a well near Montgomery, AL, and compare its performance with a baseline model. Additionally, we explore the relevance of the selected climate variables as predictors.

2 Site of Interest & Data

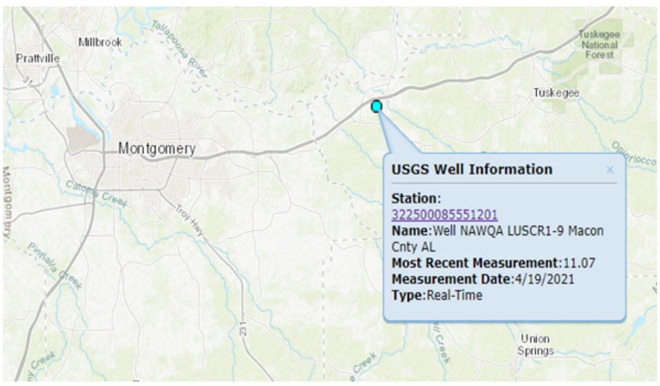

Figure 2: Well location

Daily GWL records from the US Geological Survey (USGS) well "NAWQA LUSCR1-9 Macon Cnty AL" were retrieved. This well was selected because it has one of the longest continuous records in Alabama that is part of the USGS Climate Network. This network aims to monitor the effects of climate variability (e.g. droughts) on groundwater levels. For this reason, the wells of this network meet the following criteria: they are minimally affected by pumping or artificial recharge, located in unconfined aquifers or near-surface confined aquifers that respond to climatic fluctuations. The location of the well, as seen in Figure 2, is approximately 15 miles east of Montgomery, AL, in Macon County, near the interstate 85. Table 1 presents relevant information concerning this well. It is screened in the Eutaw aquifer, which is a confined aquifer, that is part of Southeastern Coastal Plain aquifer system. At the location of the well, the water table is near the surface, therefore, GWL are subject to climate variations.

Table 1: Well summary

ID	322500085551201
Longitude	32.4169
Latitude	-85.9199
County	Macon, AL
Use	Monitoring
Depth	32
Aquifer	Eutaw (confined)

The period selected for this study comprises the water years (WY) 2007 to 2013. The GWL for well "NAWQA LUSCR1-9 Macon Cnty AL" is shown in Figure 3.

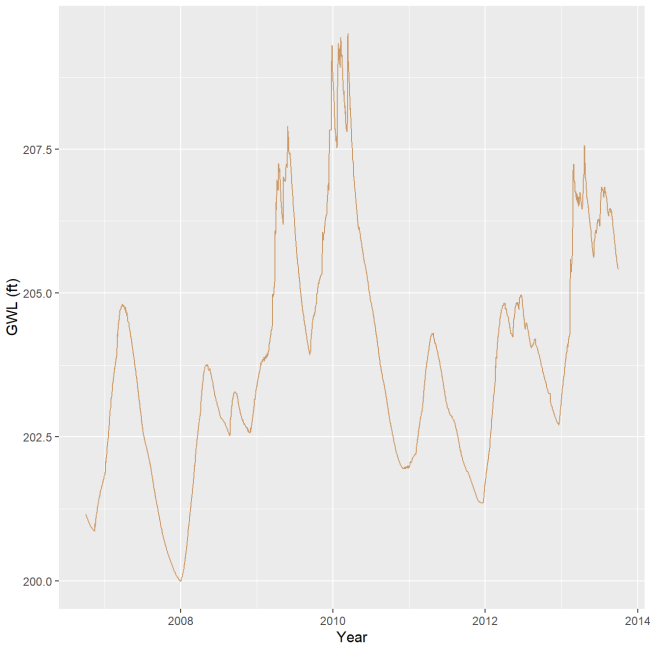

Figure 3: GWL at well "NAWQA LUSCR1-9 Macon Cnty AL"

Relevant climatic data variables were retrieved to use as predictors. Daily values of precipitation, air temperature and minimum and maximum vapor pressure were taken from the PRISM (Parameter-elevation Relationships on Independent Slopes Model) data set. Precipitation and air temperature were introduced directly into the LSTM model. Minimum and maximum vapor pressure were considered to estimate relative humidity, using the Magnus equation. Then, relative humidity was used as an additional predictor. PRISM is an interpolation method used to develop data sets that reflect spatial climate patterns in the United States [5]. Its inputs are climate observations from a wide range of monitoring networks [6]. For reference, Figure 4 shows the precipitation time series between WY 2007 and 2013.

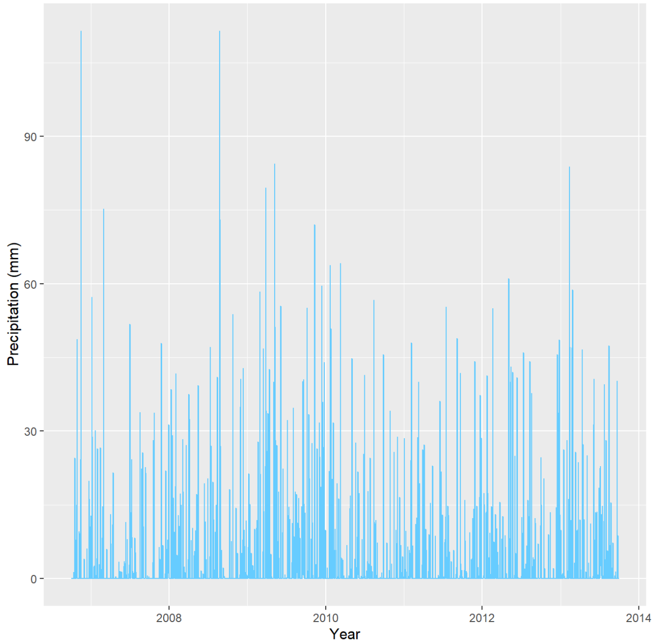

Figure 4: Precipitation in the vicinity of well "NAWQA LUSCR1-9 Macon Cnty AL"

3 Methodology

3.1 Long Short-Term Memory (LSTM)

An LSTM neural network expands the architecture of a simple RNN which is a sequence-based model capable of storing and relating information in a time series. LSTM neural network enables advanced time series prediction by adding a way to preserve information across multiple time steps, thus making it highly suitable for GWL forecasting. Unlike an RNN which cannot store long sequences due to the vanishing gradient problem, an LSTM overcomes this problem by adding more states (a forget state and a cell state) in the fundamental memory blocks of an RNN. A detailed schematization of the LSTM architecture is shown in Figure 5, while for more information regarding the building blocks of the LSTM neural network the reader is referred to [7].

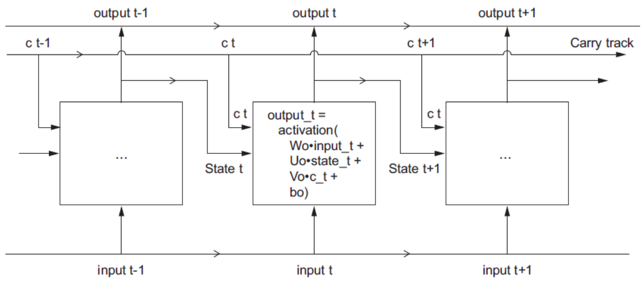

Figure 5: Memory blocks of a Long Short-Term Memory (LSTM) neural network.

3.2 Data pre-processing

The input data i.e., the four variables (GWL, precipitation, temperature and relative humidity), varied widely and therefore a standardization was necessary to ensure a smooth learning ability of the model. Prior to standardization, the 7 years of data were divided into training (5 years), development (1 year) and testing (1 year) sets. The scaler was fit considering only the training set and was subsequently used to scale the remaining data as well. The standardization formula that was utilized is described below:

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{1}$$

where z_i represents the scaled daily value and x_i the original daily value. This kind of transformation normalizes values i.e., constraints values inside the [0,1] interval.

3.3 Proposed model framework

After an extensive trial and error approach, the final proposed model is as follows:

1. The input training data are first put into an LSTM layer which consists of 30 neurons and no (recurrent) dropout.
2. Subsequently, the data are passed into a second LSTM layer with 10 neurons and, again, no dropout.
3. Finally, a simple output dense layer of N-unit dimension yields the predictions. N refers to the forecast lead time.

Dropout was deemed unnecessary for this regression problem since the amount of training data did not favor over-fitting. Adding dropout layers resulted in worse performance and thus it was discarded in both LSTM layers. Further parameterization of the model included an Adam optimizer with an initial learning rate of 0.01 and a callback for reducing the learning rate by a factor of 0.2, imposed on the validation loss function with a patience of 2. Additionally, a second callback was utilized for early stopping of the training process monitored by the validation loss function with a patience of 4. The loss function used is defined below:

$$LOSS = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{2}$$

where f_i is the predicted value at time i, and y_i is the measured value at time i. Finally, the optimal window size for the LSTM i.e., a "look-back" hyper-parameter that refers to how many time steps back the input data should go, was found to be 6 days.

3.4 Model evaluation criteria

The performance of our proposed model architecture was evaluated on the testing set by using the following metrics: a) Mean Squared Error (MSE), b) Mean Absolute Percentage Error (MAPE), and c) Mean Absolute Error (MAE). The expressions of these metrics are given below:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{3}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}|(y_i - f_i)/y_i| \times 100 \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i| \tag{5}$$

where f_i is the forecast of the model's ith observation, y_i is the target (real measurement), n denotes the number of instances in the testing set and μ is the mean of the instances.

The model's metrics for 1-day ahead predictions, were compared with the respective ones from the persistence forecast, a forecast that implies that the future condition will be the same as the present condition. The persistence forecast is widely used in hydro-meteorology to assess the degree of skill of forecasts prepared by other methods, especially for short-term predictions, and in practice it is one of the hardest to beat. For n-day ahead forecasts, where n>2, our model rivaled a benchmark model that projects the slope of the GWL between the previous day measurement and the n-days-back measurement. A schematization of the two baseline models is shown in Figure 6.

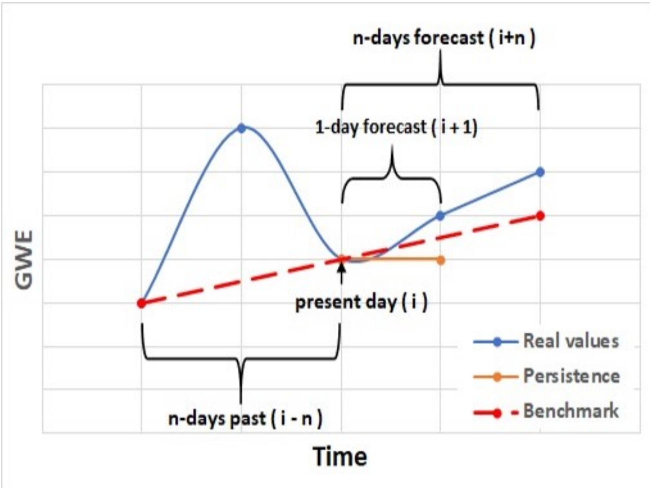

Figure 6: Schema of both persistence and benchmark models for comparison of 1-day and n-day ahead forecasts respectively.

4 Results

Figure 7 illustrates 14-day ahead forecasts created by the proposed LSTM structure over the span of 1 year (testing set). For clarity and reference, the respective observed values are also shown, along with the corresponding benchmark predictions. It is apparent that there is a quite good agreement between LSTM's forecasts and ground truth. Our proposed model was able to adequately capture the overall trend of the GWL including both upticks and recessions. However, it is evident that the model exhibited a slight timing offset from the measurements. The benchmark model tends to overestimate the magnitude of upticks and deep recessions.

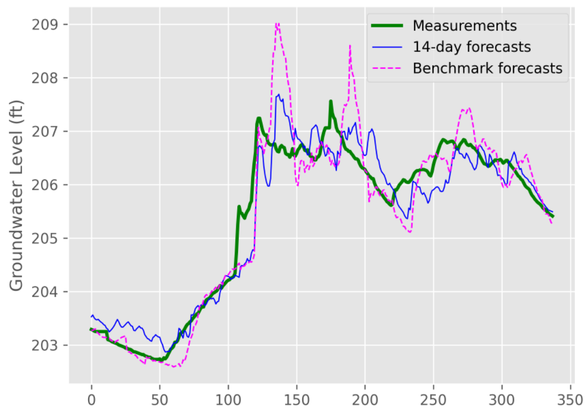

Figure 7: 14-day GWL forecasts produced by the LSTM and benchmark model. Actual GWL measurements are also presented for comparison.

The superiority of our LSTM structure over the benchmark model for long-term forecasting i.e., 14 days, is also reflected in Table 2; LSTM displayed lower error values in all three metrics examined. Characteristically, the benchmark model exhibits a 66%-increased MSE with respect to the LSTM.

Table 2: 14-day ahead forecast scores.

	MSE	MAPE	MAE
Proposed model	0.21	0.17	0.36
Benchmark	0.35	0.19	0.39

To assess the effect of forecast lead time on our model's predictive accuracy, various lead times where tested spanning from 1 day to 28 days, in increments of seven days. Figure 8 shows the MAE of forecasts with different horizons from both the LSTM and benchmark models.

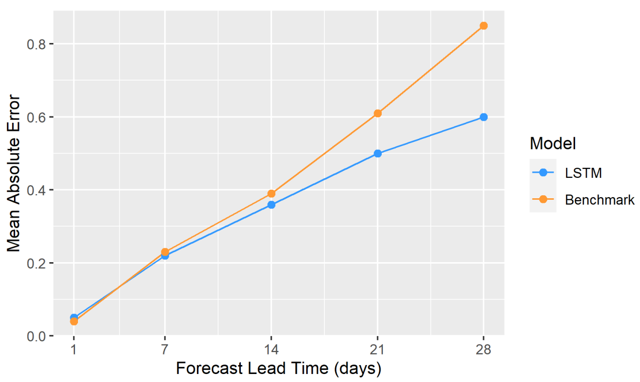

Figure 8: MAE of forecasts produced by the LSTM and benchmark model for different forecast lead times.

As expected, increased lead time resulted in reduced forecasting skill. In the short-term (1 day), LSTM demonstrated a slightly higher MAE than the persistence forecast though a conjecture is that with finer hyper-parameter tuning LSTM should be able to showcase improved performance. For mid-term forecasts i.e., up to 7 days, LSTM did not perform significantly better than the benchmark. However, for long-term GWL forecasting (more than 14 days), the proposed model provided added value over the benchmark model, with a fair rate of skill decline. Overall, it appears that the proposed model is capable of providing reliable prolonged forecasts; an increase of lead time by 14 days (14 to 28 days) results roughly to a 50% increase in the MAE.

To assess the importance of each predictor in the model, a forward step-wise selection was performed. The best 1-variable and 2-variable models were selected based on a satisfactory linear fit (adjusted R^2). The best 1-variable model included precipitation, whereas the most optimal 2-variable model comprised precipitation and relative humidity. In Figure 9 we can observe that our proposed model is able to outperform the benchmark by utilizing only one predictor i.e., precipitation (apart from GWL itself). However, adding subsequent variables results in increased performance.

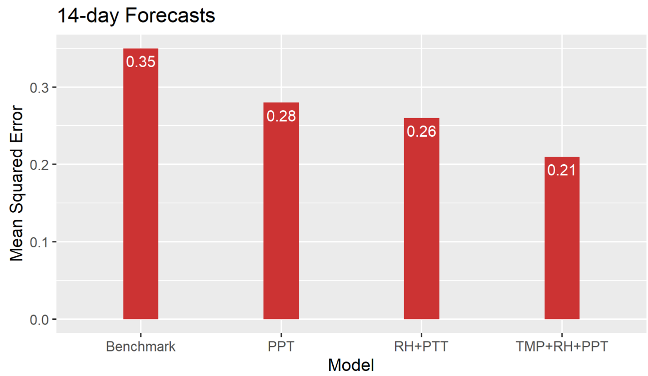

Figure 9: MSE of 14-day ahead forecasts produced by the benchmark model and LSTM with the use of different input data.

5 Conclusion & Discussion

Our analysis suggests that our proposed LSTM architecture has the potential to be an alternative to traditional GW models which require extensive sources of data. Results indicate that an LSTM model can be trained with few readily available climate variables to produce GWL forecasts from 1 to 28 days with a reasonable error for applications within the context of groundwater management (MSE $<$ 0.6 ft for 28-day forecast).

Precipitation, mean temperature, and relative humidity proved to add value to the LSTM model in contrast to just using past GWL for forecasting. Other climate/hydrological and soil properties could be included in the model to test if they would further improve model accuracy. However, additional variables are more expensive, require more time, or have greater uncertainty. Using only precipitation, mean temperature and relative humidity can be advantageous in the sense that with a weather station, which reports these values in near real-time, an automated forecasting system can be developed based on our proposed model. This has the potential to assist groundwater dependent ecosystem monitoring and management where groundwater fluctuations are influenced only by climate fluctuations.

Despite the adequacy of our proposed model architecture for the particular area studied here, future work should consider implementation of our LSTM neural network in a variety of wells with different hydrogeologic conditions so as to assess its translatability. The main limitation of the proposed model herein is that it yields adequate results for wells or portions of aquifers where human intervention does not affect GWL fluctuations. Variables like pumping could be introduced to the model in order to predict GWL in aquifers with anthropogenic impacts. This would allow practitioners to have a new tool for sustainable groundwater management, and would provide researchers with a valuable baseline model for future studies.

References

1. Alley, W. M., Reilly, T. E. & Franke, O. L. *Sustainability of ground-water resources* (US Department of the Interior, US Geological Survey, 1999).
2. *Water Information* `https://www.gsa.state.al.us/gsa/groundwater/waterinfo`.
3. Butler Jr, J., Stotler, R., Whittemore, D. & Reboulet, E. Interpretation of water level changes in the High Plains aquifer in western Kansas. *Groundwater* **51,** 180–190 (2013).
4. Rajaee, T., Ebrahimi, H. & Nourani, V. A review of the artificial intelligence methods in groundwater level modeling. *Journal of hydrology* **572,** 336–351 (2019).
5. Daly, C. *et al.* Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society* **28,** 2031–2064 (2008).
6. *PRISM Climate Group, Oregon State University* `http://prism.oregonstate.edu`. Created: 2004-02-04.
7. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9,** 1735–1780 (1997).