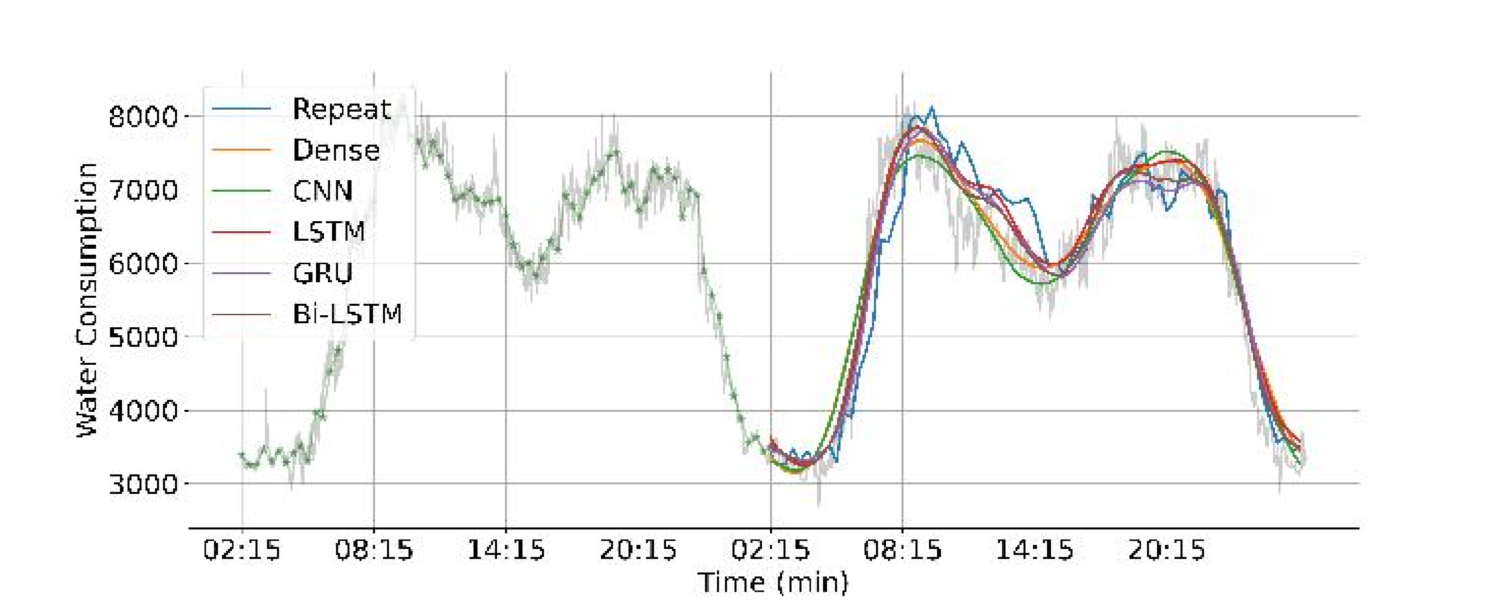
Real-time Water Consumption Prediction Based on Deep Learning

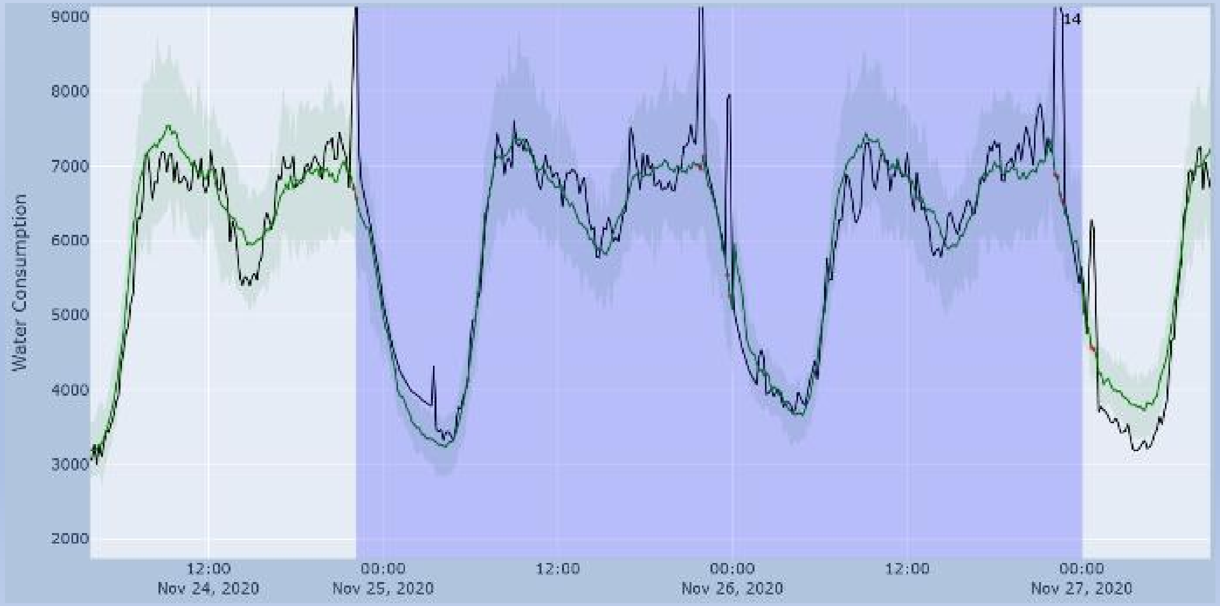
---Value and Probabilistic Prediction

Cao Jinpu
Civil and Environmental Engineering Department
jinpu@stanford.edu



Value prediction

Probabilistic prediction



Abstract—

- Urban water supply service is one of the key functions of urban infrastructure. Accurate prediction of water consumption in the future is helpful to detect the abnormalities of water supply systems including pipe bursts in real-time, and effectively improve the economy and stability of the water supply system.

- Based on two water consumption dataset, the paper finds that Recurrent Neural Networks model represented by GRU model outperforms other artificial neural networks models when conducting the value prediction of water consumption. The root mean square error of GRU model is only 80% of the basic model's error.

- Besides, the paper develops a probabilistic prediction model of water consumption based on Deep Autoregression (DeepAR) model whose mean absolute percentage error on test set is only 6%. It can be used in the pipe bursts alarming.

- ***Keywords—water consumption, seq to seq, RNN, probabilistic prediction, DeepAR***

LITERATURE REVIEW

- Markov modified autoregressive moving average (ARIMA) model (Hongyan)
- artificial neural networks (Altunkaynak)
- deep convolutional neural network DeepCNNs (Haytham)
- Shuffled Complex Evolution Metropolis algorithm (Cutore)
- …

- Need?

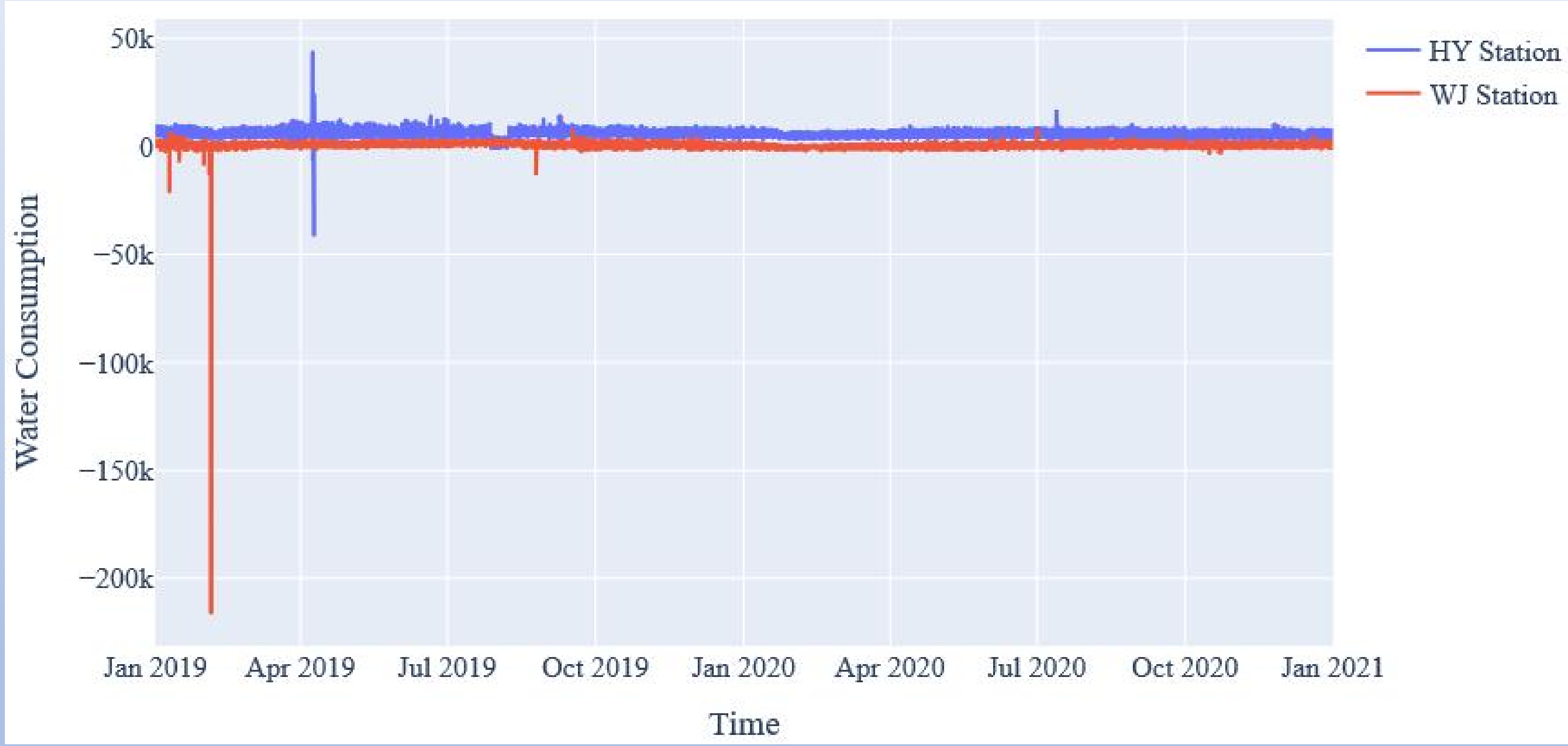
    - **More comparisons between different deep learning models**

    - **Probabilistic prediction model based on deep learning**

DATASET – RAW DATA

- Two water stations
  Hanyang Water Station (HY)
  Wujin Water station (WJ)

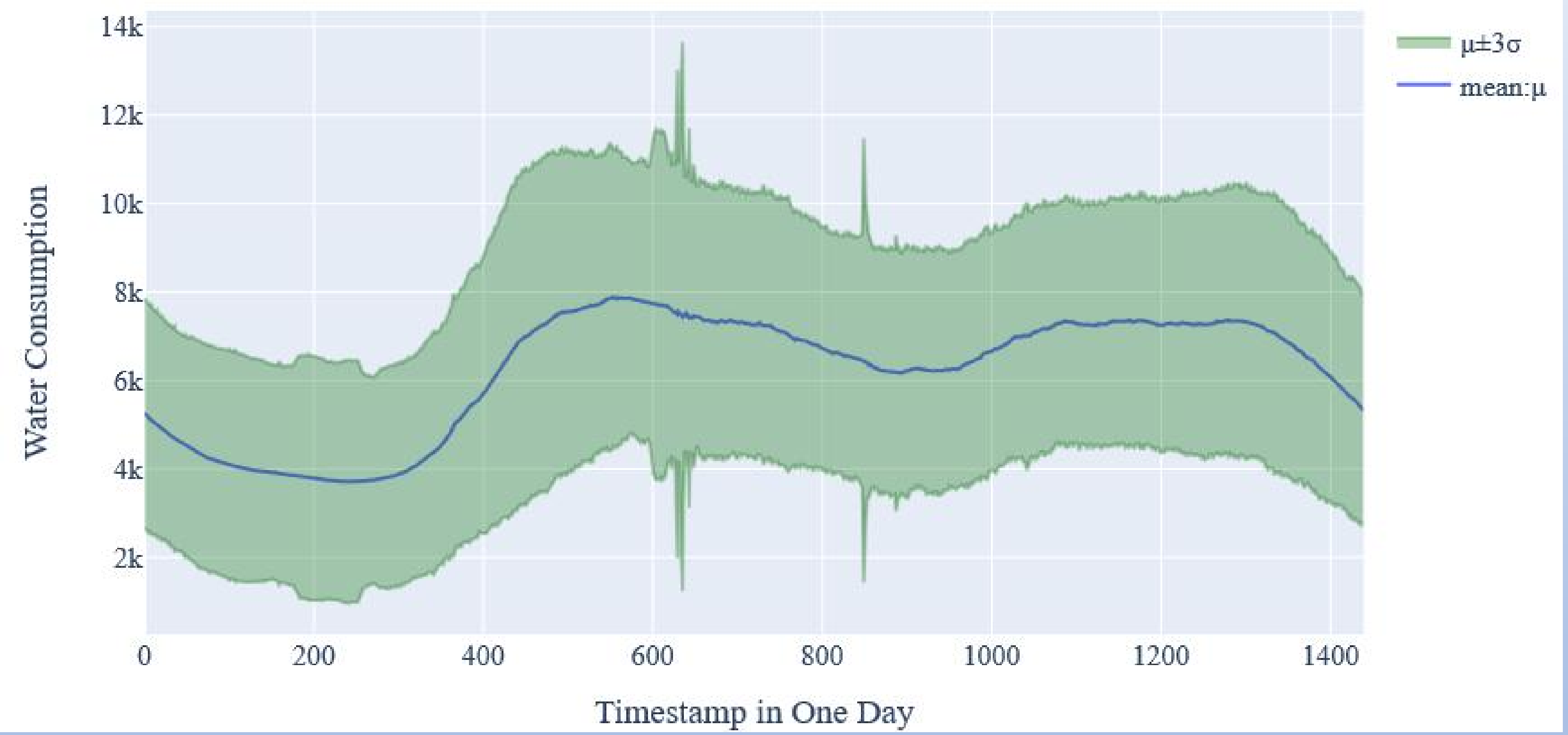
- Two years
  $(365 + 366) \times 24 \times 60 =$
  1,052,640 records



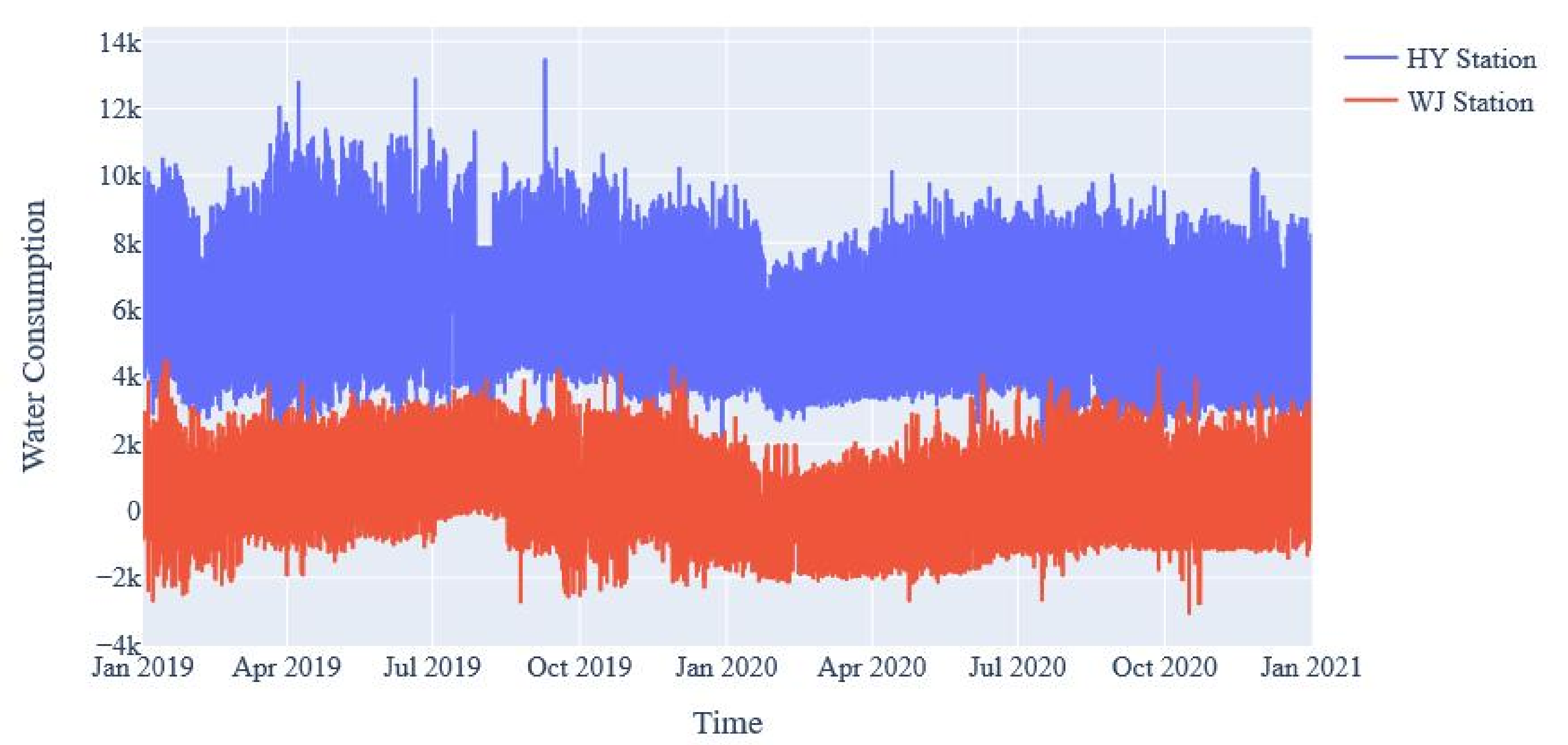
DATASET – PREPROCESS

- To clean the outliners, the water consumption is regrouped by day. The mean and standard are computed.

- Abnormal values are detected using the **3σcriterion** (Fig.2.) and mean values are used instead.

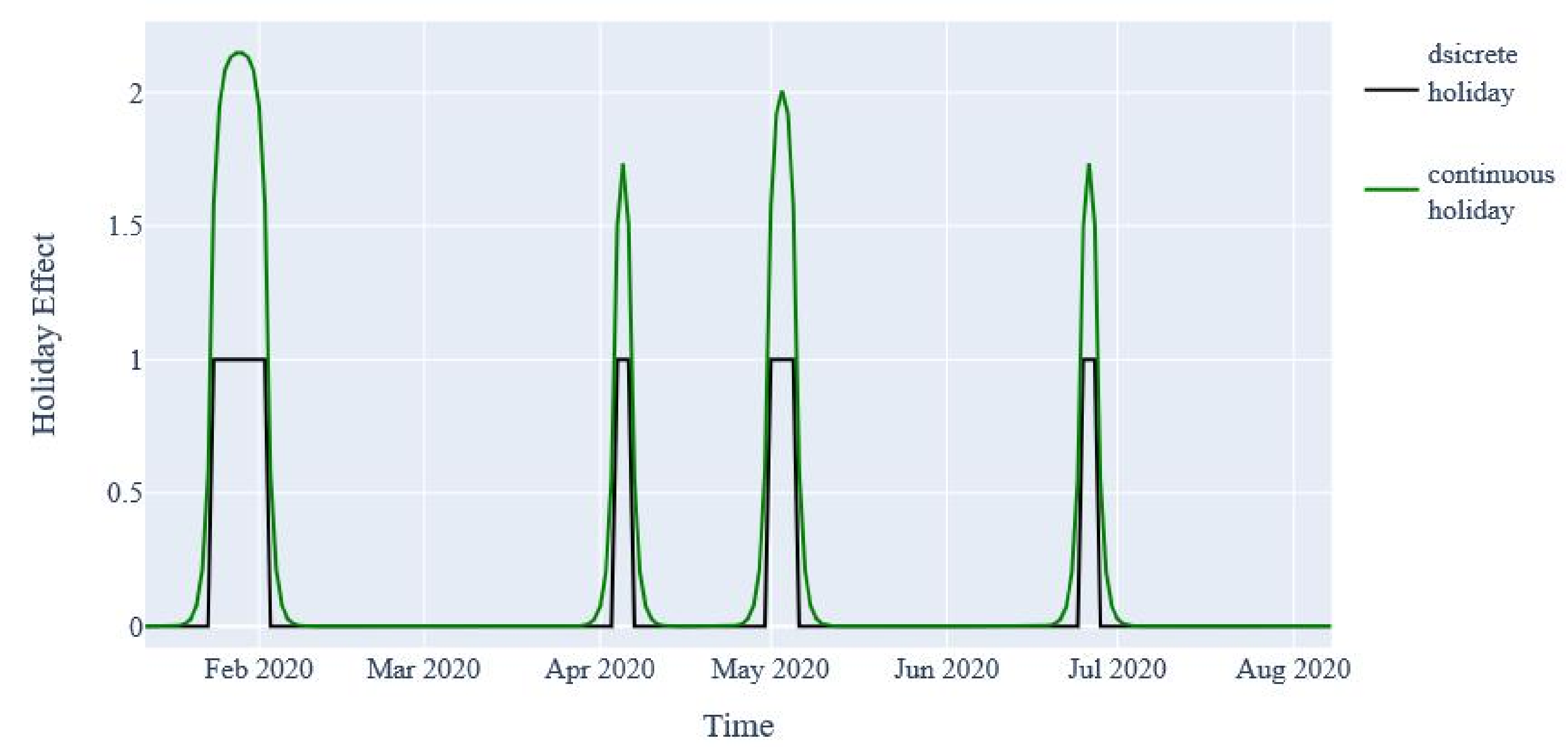


DATASET – CLEAN DATA



DATASET – FEATURES (HOLIDAY)

- Generally, the water consumption will increase before and after the festival, rather than only on the festival day.
- **Discrete to continuous**



Water Consumption Prediction (value) – input/output

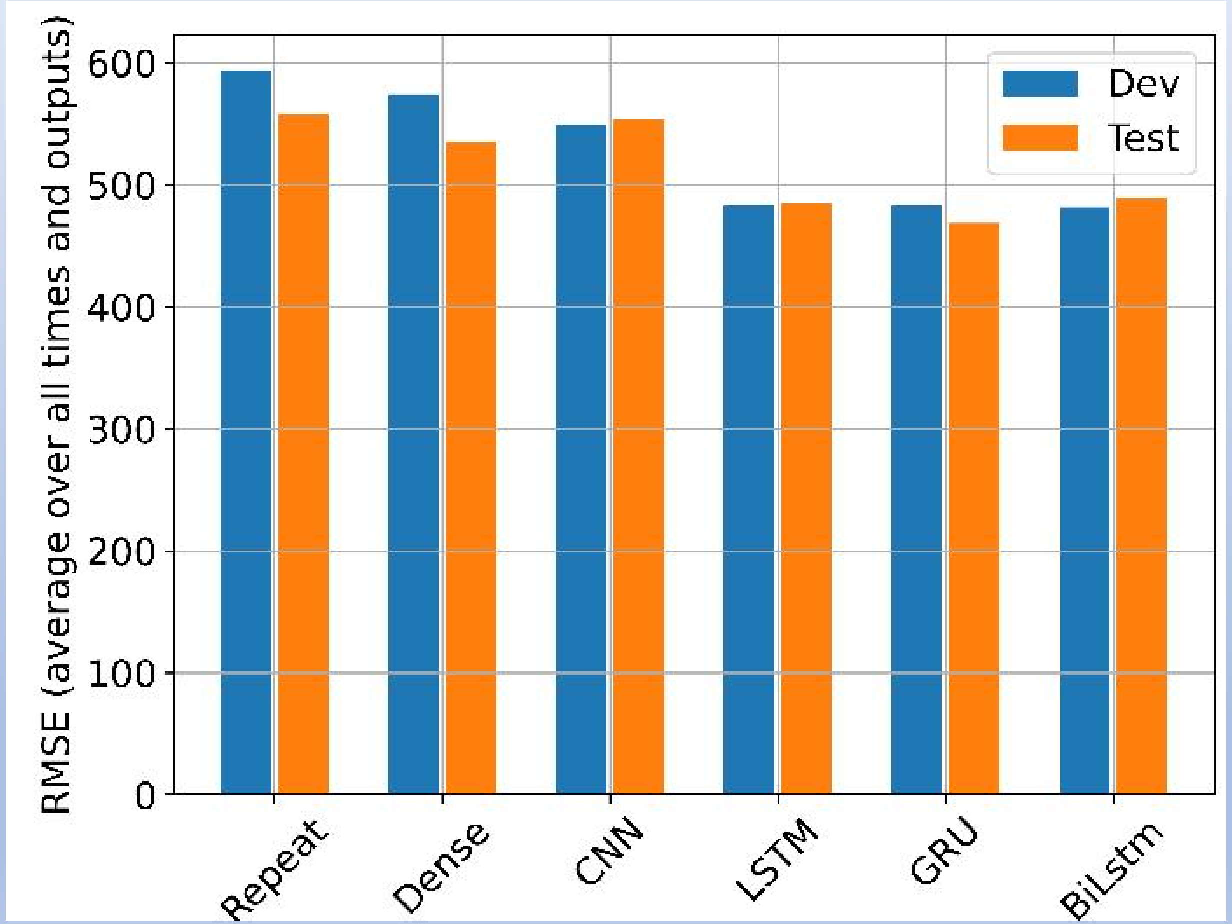
- Resample:20 minutes

- 1 day to 1 day
  - 1440/20 +1=73

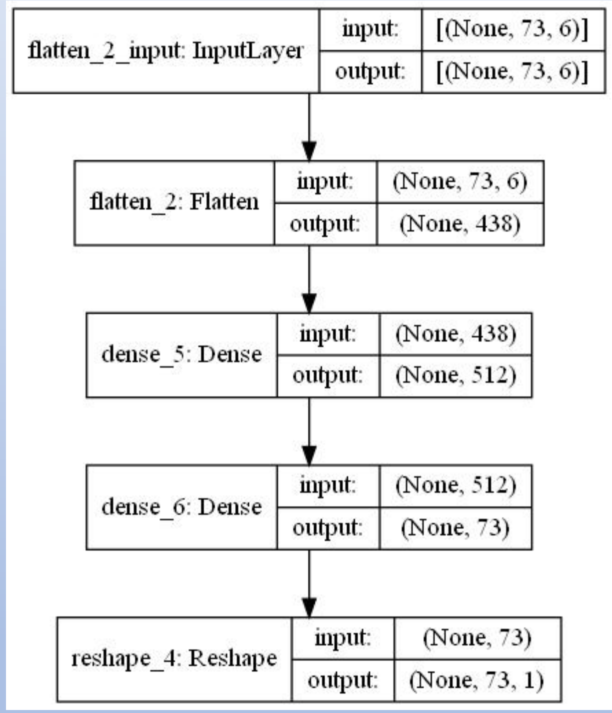
- Inputs: (None, 73, 6)
  - Daily time index,
  - Water consumption,
  - max temperature, min temperature
  - Weekend, holiday

- Outputs: (None, 73, 1)
  - Water consumption

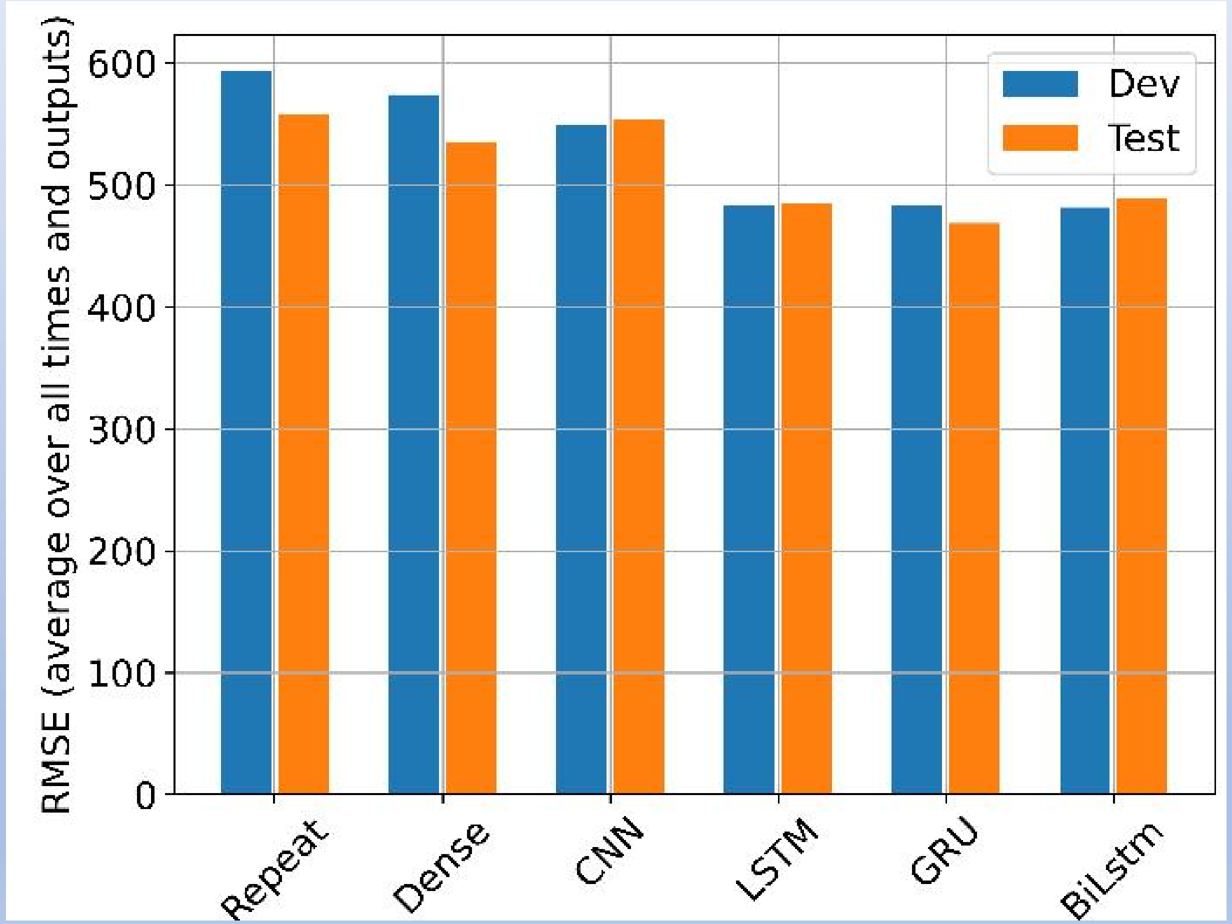
Water Consumption Prediction (value) – Models



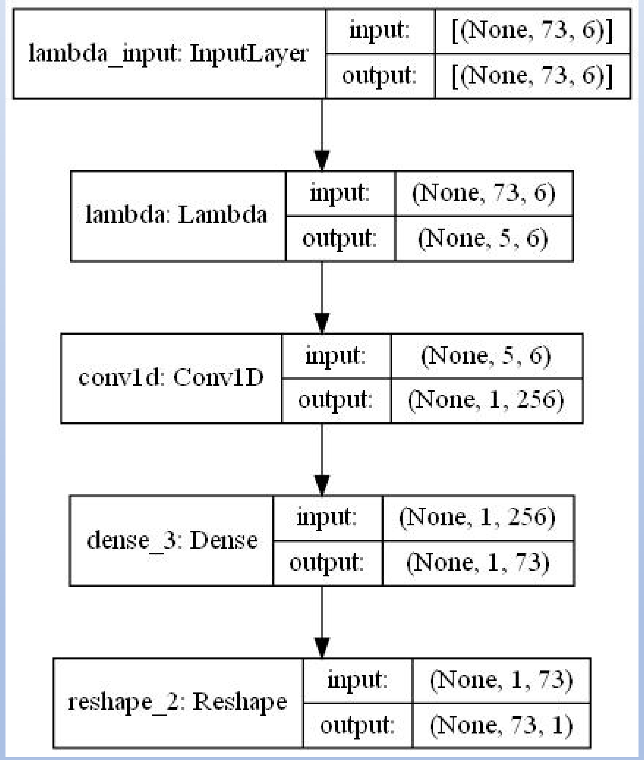
- **Dense (relu)**



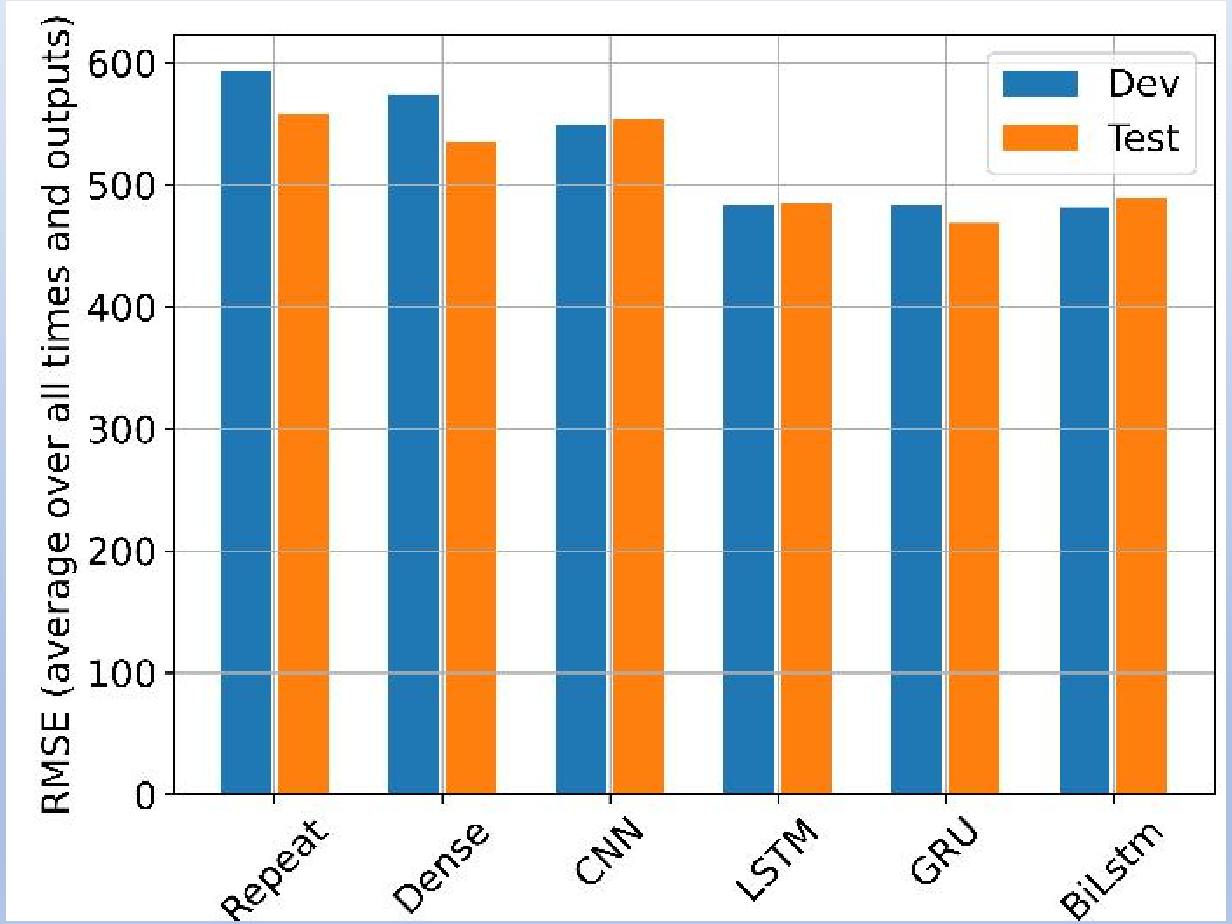
Water Consumption Prediction (value) – Models



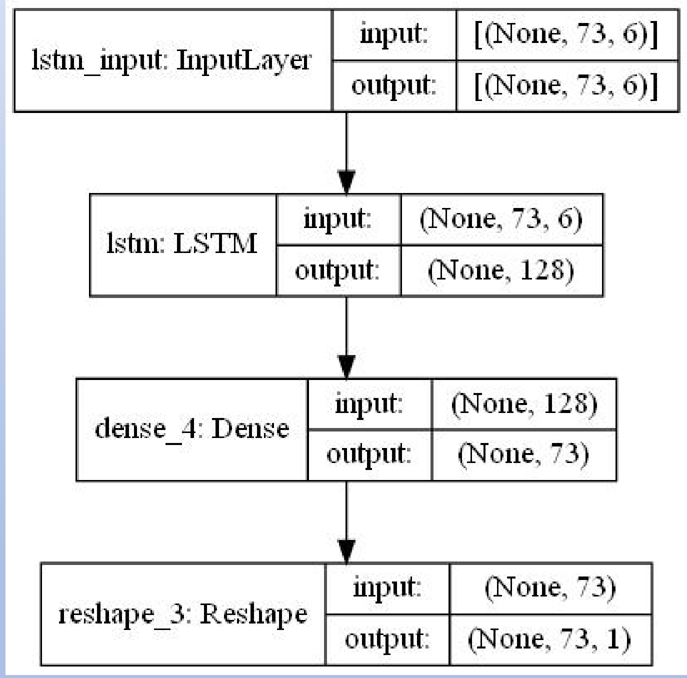
- **CNN**



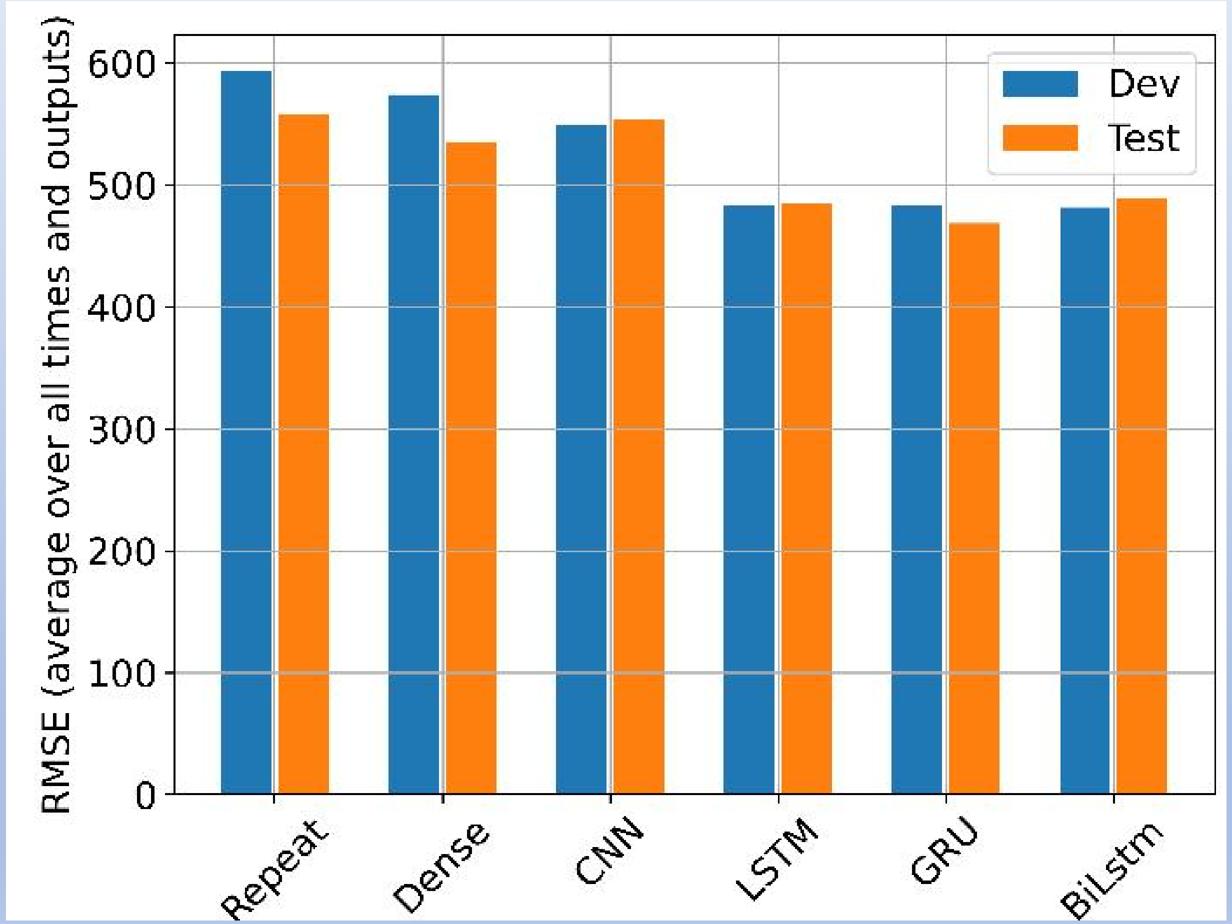
Water Consumption Prediction (value) – Models



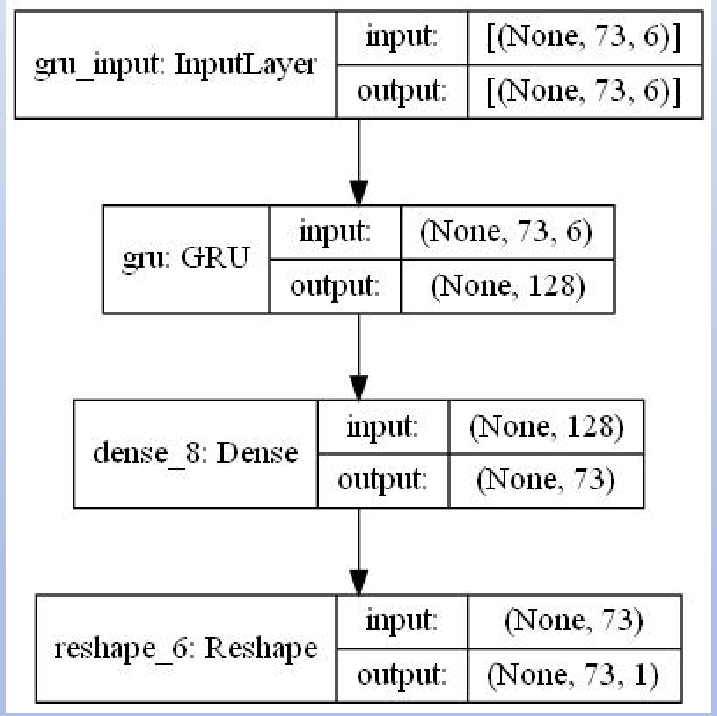
- **LSTM**



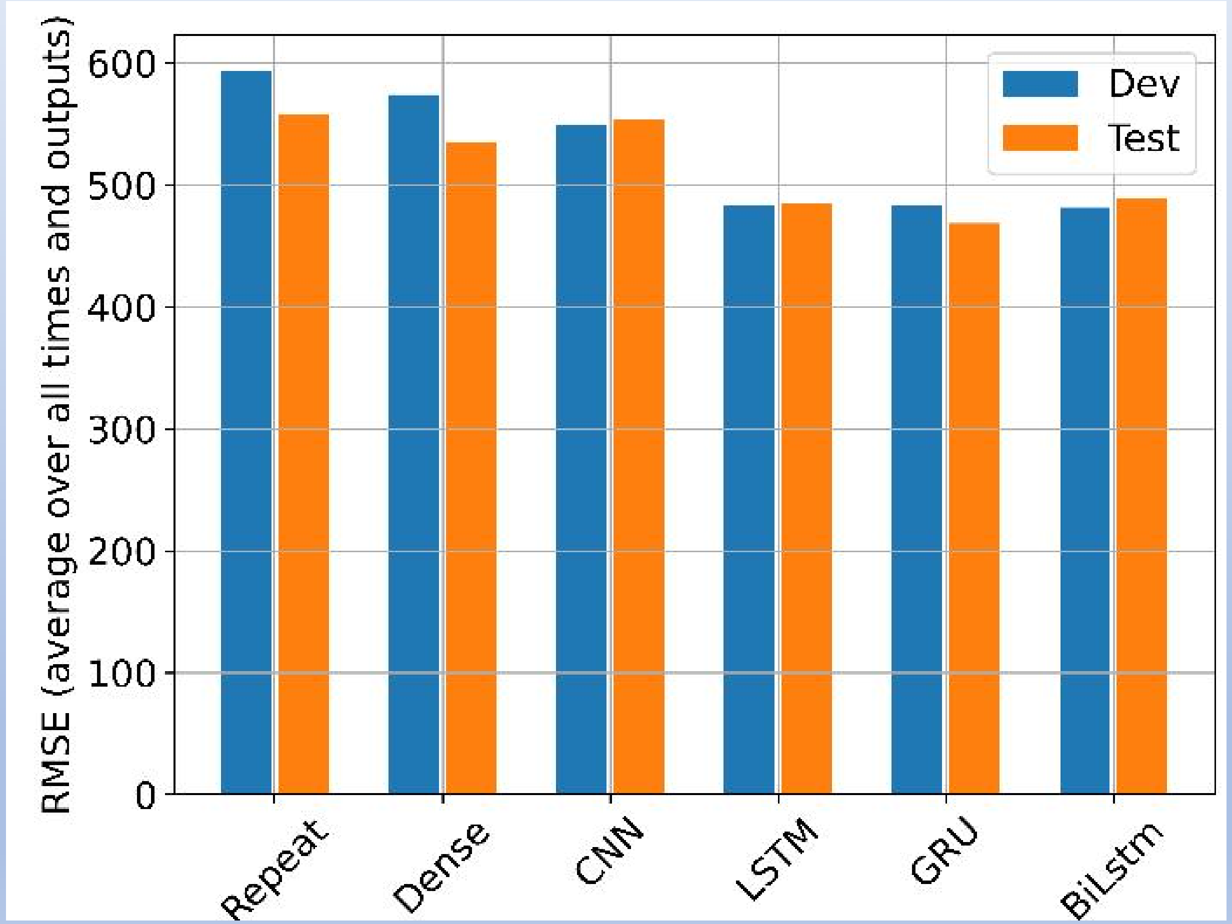
Water Consumption Prediction (value) – Models



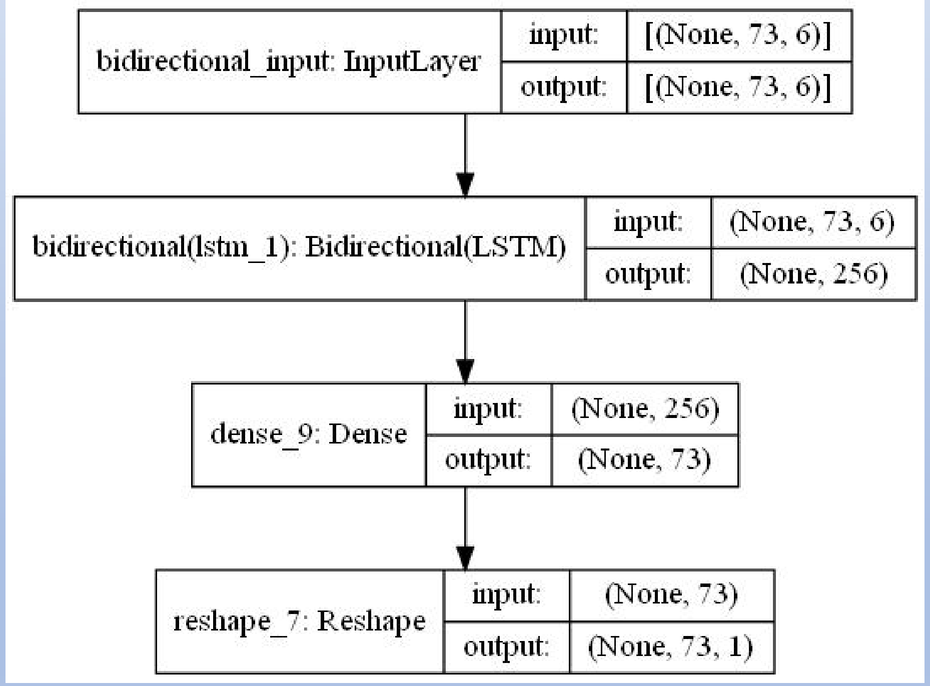
- **GRU**



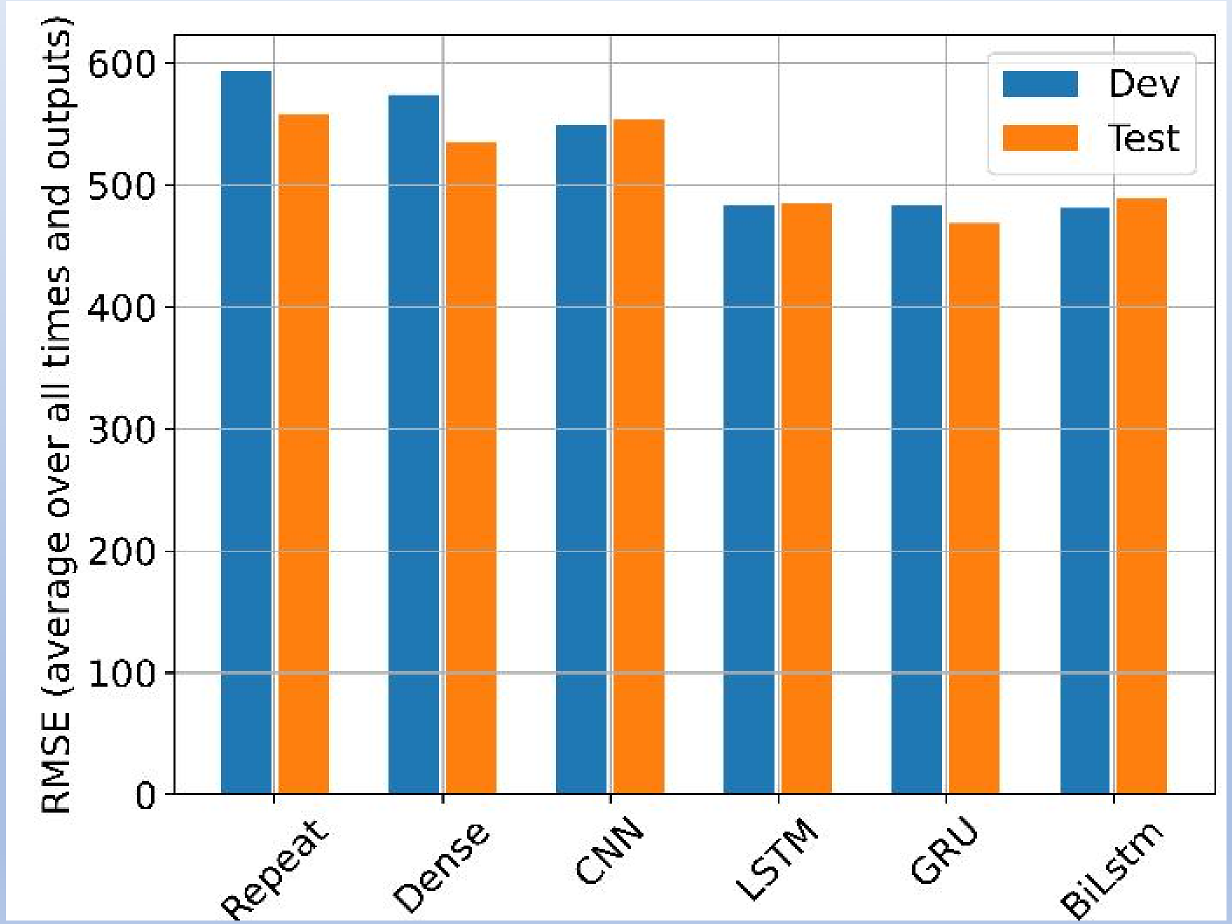
Water Consumption Prediction (value) – Models



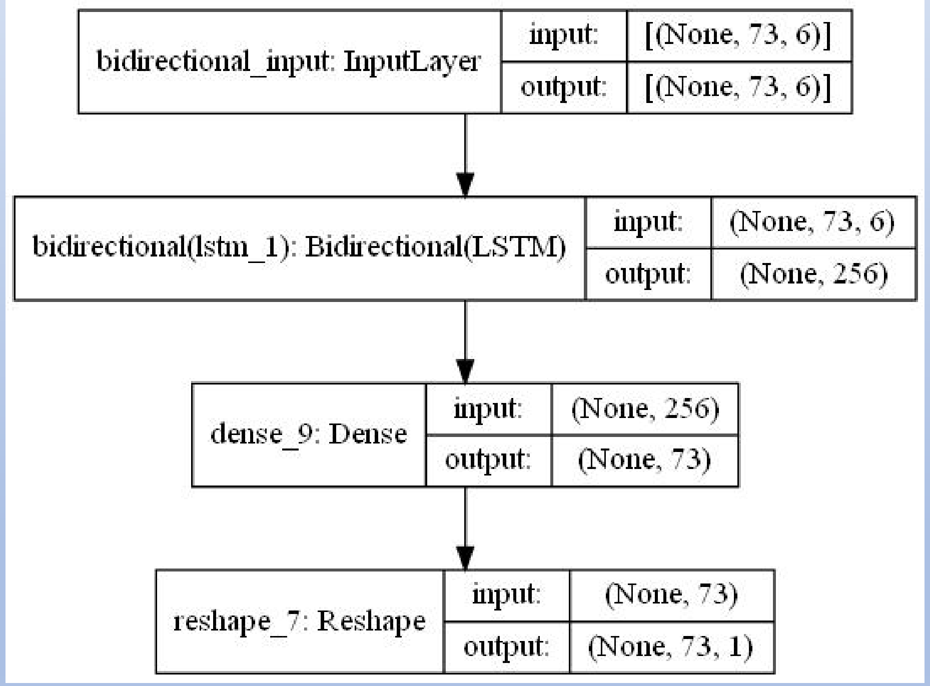
- **Bi_LSTM**



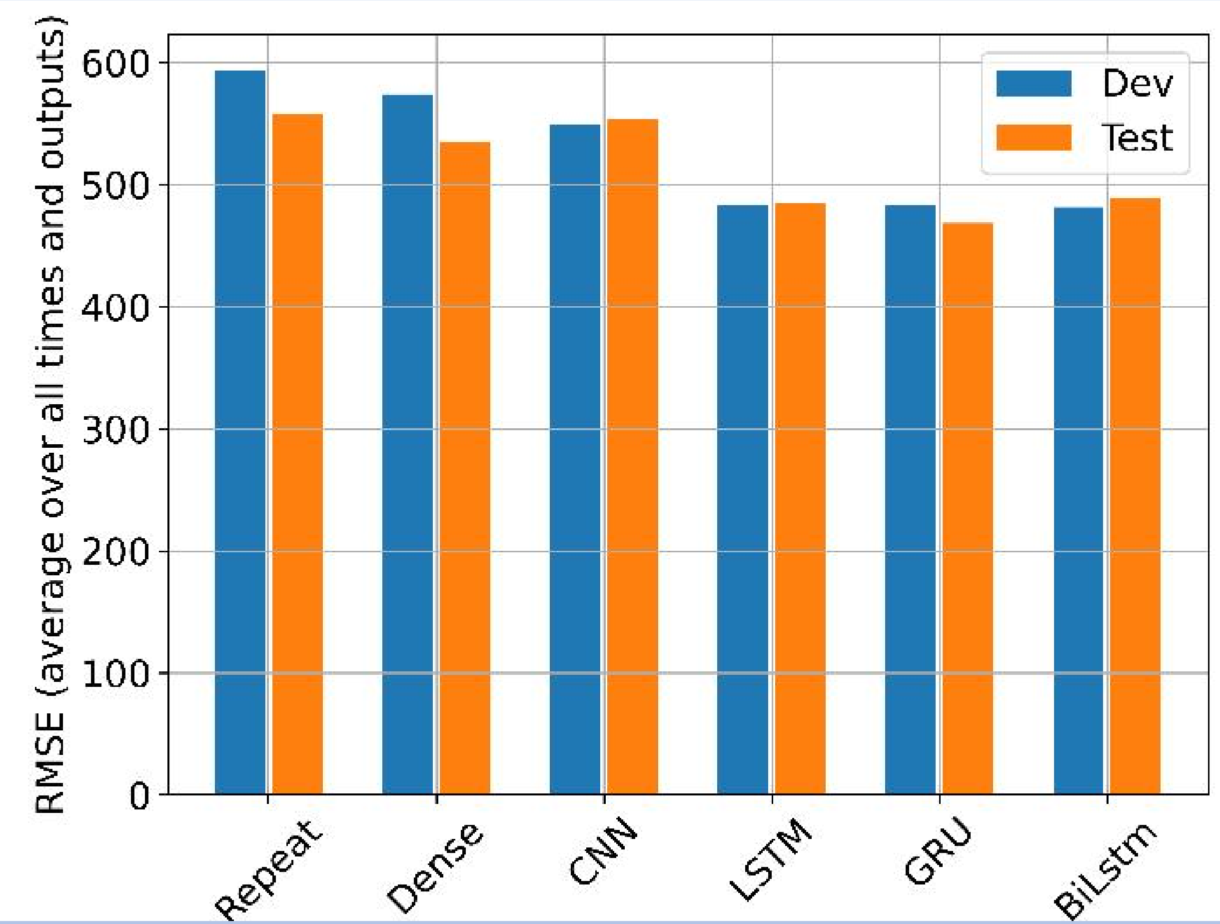
Water Consumption Prediction (value) – Models



- **Bi_LSTM**

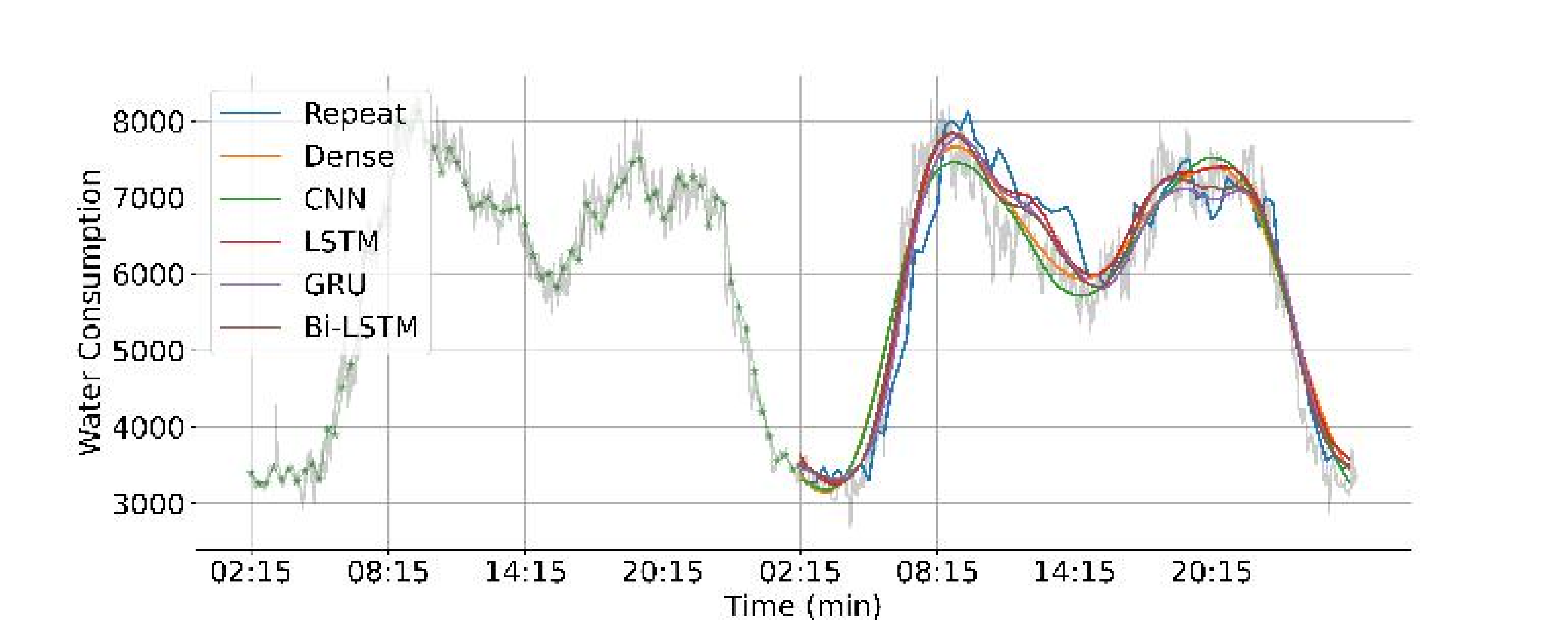


Water Consumption Prediction (value) – Models

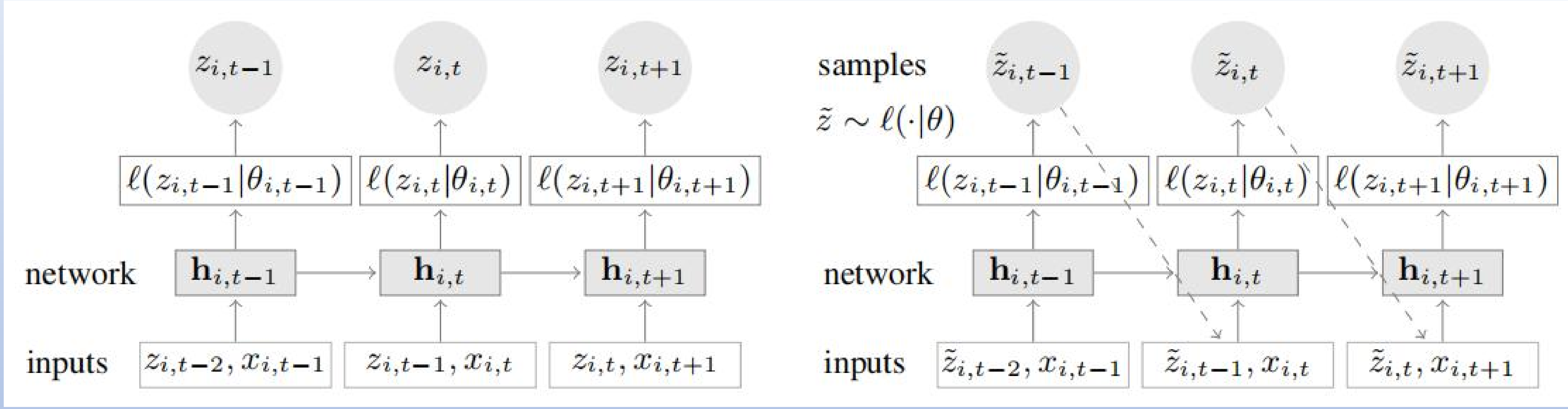


Model	Repeated	Dense	CNN
train	703	818	742
dev	594	574	549
test	558	535	554
Model	**LSTM**	**GRU**	**BILSTM**
train	636	656	628
dev	483	483	481
test	485	468	489

Water Consumption Prediction (value) – Models



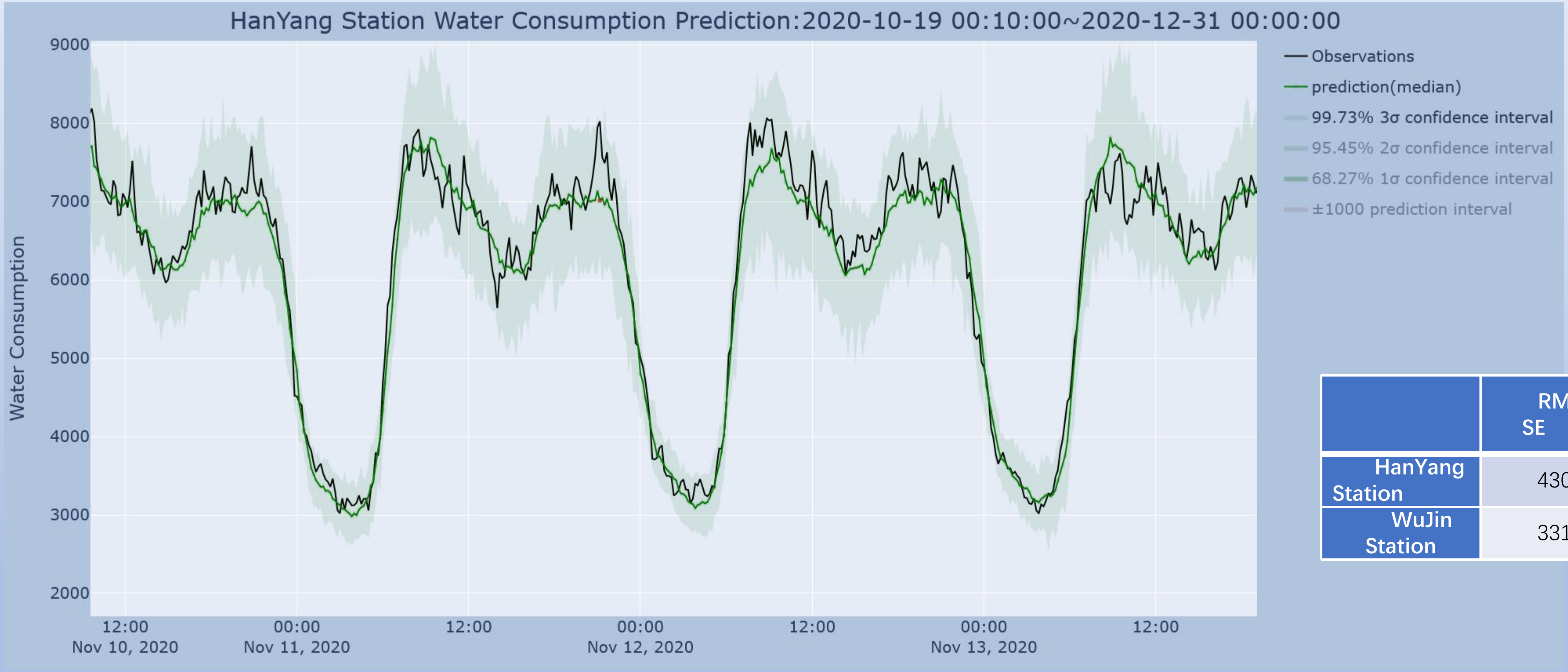
*Water Consumption Prediction (*Probabilistic*) – DeepAR*



- The model consists of the parameters of the RNN $h(\cdot)$ as well as the parameters of $\theta(\cdot)$
- $\mu_{i,t} = w_\mu^T h_{i,t} + b_\mu$
- $\sigma_{i,t} = \log(1 + \exp(w_\sigma^T h_{i,t} + b_\sigma))$
- log-likelihood

$$\mathcal{L} = \sum_{i=1}^{N} \sum_{t=t_0}^{T} l(z_{i,t}|\theta(h_{i,t}, \Theta))$$

Water Consumption Prediction (Probabilistic) – DeepAR



	RMSE	MAPE	% more than ±1000
HanYang Station	430	5.30%	3.60%
WuJin Station	331	6.00%	1.32%

Conclusion

- According to the experimental results of this paper, it can be found that the revised recurrent neural network model represented by GRU can well deal with the problem of time series data prediction. In terms of water consumption prediction (value) in this paper, the water consumption sequence of the previous day is used to predict the water consumption sequence of the next day. Using GRU model, the root mean square error (RMSE) of the test set is only 80% of the model where the previous day's water consumption is seen as equal as the next day's water consumption.
- The Deep Autoregressive (DeepAR) model based on the recurrent neural networks shows satisfactory performance in probabilistic prediction. According to the experimental results of this paper, based on the DeepAR model, using the water consumption five days ago to predict the water consumption interval after one day, the mean absolute percentage error (MAPE) in test set can be controlled within 6%, which can be used for some pipe bursts alarming.