AGU PUBLICATIONS

Geophysical Research Letters

Supporting Information for

Spatiotemporal Attentive Gated Recurrent Unit: A Novel Method to Forecast O3

Yang Li[1], Xingguo Chen[1], Xiaoyan Xu[2], and Min Shao[3]

[1]Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

[2]School of the Environment, Nanjing University, Nanjing 210046, China

[3]School of Environment, Nanjing Normal University, Nanjing 210023, China

Contents of this file

Table S1 to S3
Figure S1 to S3

Introduction

Supplementary information includes tables, which involve longitude and latitude of each station, details of experimental design and specific results of experimental comparison in manuscript, and figures, which contain the details of proposed model and derivative model and of comparison results between proposed and derivative model.

Monitoring station	Longitude	Latitude
ATZX	118.737	32.009
CCM	118.749	32.057
MGQ	118.803	32.108
PK	118.626	32.088
RJL	118.803	32.031
SXL	118.778	32.072
XLDXC	118.907	32.105
XWH	118.795	32.078
ZHM	118.777	32.014

Table S1. The longitude and latitude of each monitoring station (unit: degree).

Experiment category	Experiment name
Seq2Seq	Seq2Seq_LSTM
	Seq2Seq_GRU
Seq2Seq+Attention	Seq2Seq_LSTM+Attention
	Seq2Seq_GRU+Attention
Spatiotemporal attentive	STAGRU
	STALSTM

Table S2. Experimental design. Seq2Seq based models are the combination of Encoder-Decoder framework and LSTM or GRU, and Seq2Seq+Attention applies single attention mechanism on the basis of Seq2Seq based models. Spatiotemporal attentive based method includes STAGRU and STALSTM. The settings of all models are consistent.

Table S3. Specific results of testing RMSE, R2 and SMAPE in all monitoring stations.

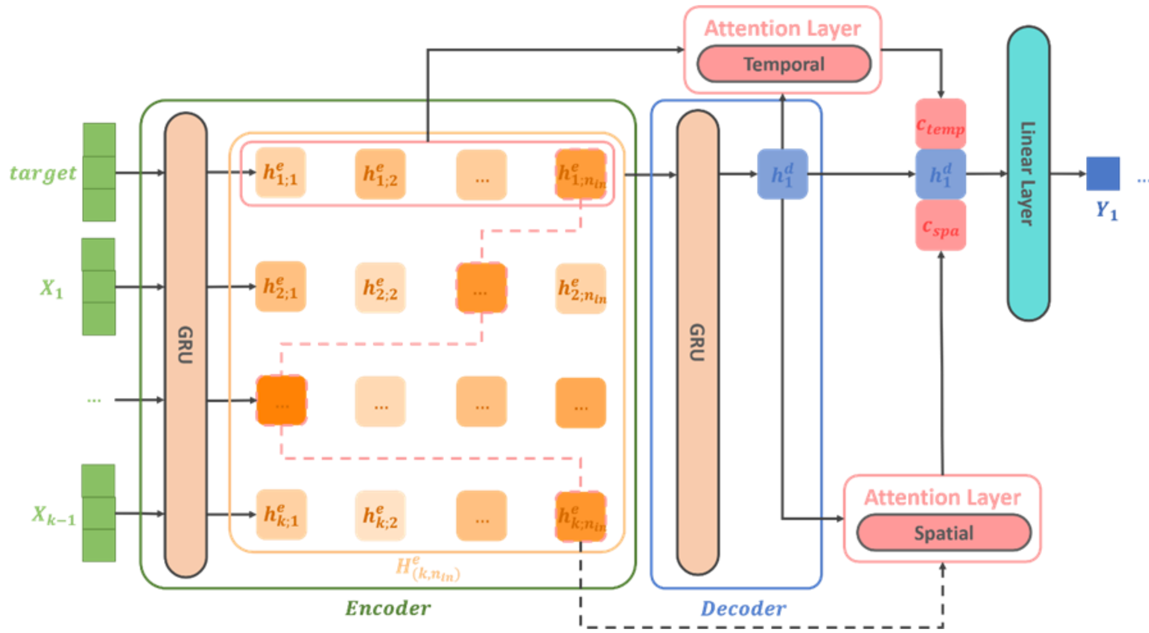

Figure S1. The structure of STAGRU. target represents the past sequence of target station, and $\boldsymbol{X_i}$ represents the past sequence of surrounding station $\boldsymbol{i}$. Sending those data to encoder, the GRU component encodes each moment into hidden state, and a hidden state matrix is produced, named $\boldsymbol{H^e_{(k,n_{in})}}$. Note that the row and column of $\boldsymbol{H^e_{(k,n_{in})}}$ depends on the number of stations and the length of past sequence. The last encoding hidden state of target sequence is fed into decoder. With that hidden state, decoder produces decoding hidden state for each predicting step. In each predicting step, temporal-context vector that derive from the encoding hidden states of target station and spatial-context vector that derive from the encoding hidden states with the highest attention weight in each monitor station are applied to make a prediction. Specifically, the temporal-context vector $\boldsymbol{c_{temp}}$ and the spatial-context vector $\boldsymbol{c_{spa}}$ are concatenated with current decoding hidden state $\boldsymbol{h^d}$, then the concatenation is sent to a linear layer to forecast $\boldsymbol{Y}$.

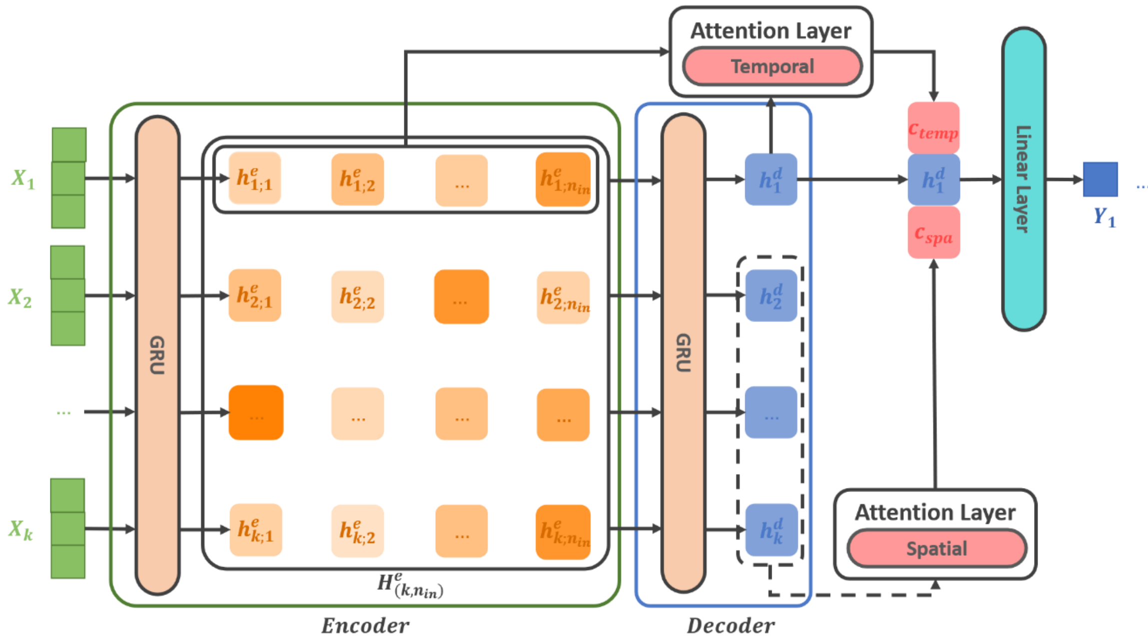

Figure S2. The structure of STAGRU-Decoding. The main difference is in the spatial information learning. STAGRU-Decoding learn the spatiality from the decoding hidden states of other surrounding stations. This operation will be applied for each predicting step of each station. In this manner, there are no target station here. All stations are forecasted synchronously.

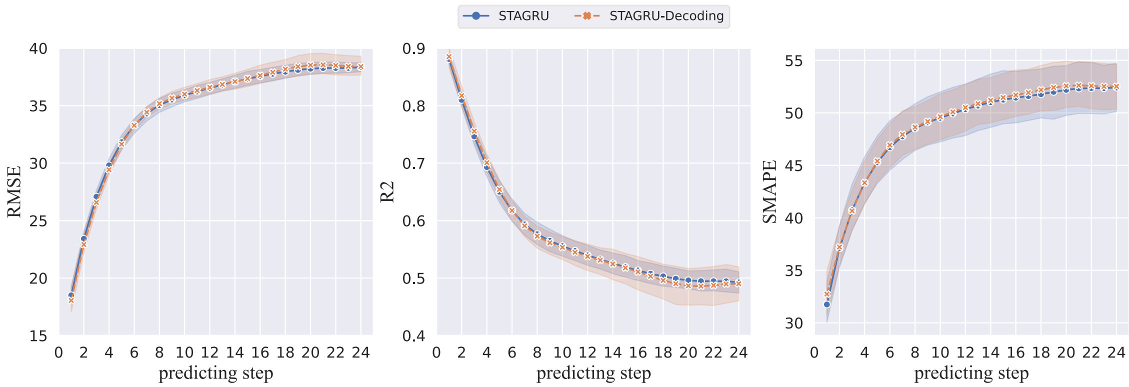

Figure S3. Comparison of STAGRU and STAGRU-Decoding.