Supplement of Geosci. Model Dev., 15, 8439–8452, 2022
https://doi.org/10.5194/gmd-15-8439-2022-supplement


Geoscientific Model Development
Open Access

*Supplement of*

# Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement

**Haochen Sun et al.**

*Correspondence to:* Xingcheng Lu (xingchenglu2011@gmail.com)

**List S1**

- WRF
    - PSFC (surface pressure)
    - U10 (10-meter U wind)
    - V10 (10-meter V wind)
    - T2 (2-meter temperature)
    - Q2 (2-meter water vapor mixing ratio)
- CMAQ
    - CO (carbon monooxide)
    - SO2 (sulfur dioxide)
    - NO2 (nitrogen dioxide)
    - O3 (ozone)
    - The sum of ASO4J, ASO4I, ANO3J, ANO3I, ANH4J, ANH4I, AXYL1J, AALKJ, AXYL2J, AXYL3J, ATOL1J, ATOL2J, ATOL3J, ABNZ1J, ABNZ2J, ABNZ3J, ATRP1J, ATRP2J, AISO1J, AISO2J, ASQTJ, AORGCJ, AORGPAJ, AORGPAI, AECJ, AECI, A25J, A25I, ANAJ, ANAI, ACLJ, AISO3J, AOLGAJ, AOLGBJ (fine suspended particulate matters)

**Text S1**

An encoder LSTM with input dimension $d$ and hidden dimension $h$ is parametrized by a weight matrix $W \in \mathbb{R}^{4h\times(h+d)}$ and a bias vector $\mathbf{b} \in \mathbb{R}^{4h}$. It takes the input of a time series with length $T$ and dimension $d$, $\{\mathbf{x}_t\}_{t=1}^{T}$ and computes the hidden states $\mathbf{h}_t \in \mathbb{R}^h$ and cell states $\mathbf{c}_t \in \mathbb{R}^h$ as well as the intermediate variables input gate $\mathbf{i}_t \in \mathbb{R}^h$, forget gate $\mathbf{f}_t \in \mathbb{R}^h$, output gate $\mathbf{o}_t \in \mathbb{R}^h$ and gate gate $\mathbf{g}_t \in \mathbb{R}^h$ for each step $1 \leq t \leq T$ recurrently as follows:

$$\begin{pmatrix}\mathbf{i}_t\\ \mathbf{f}_t\\ \mathbf{o}_t\\ \mathbf{g}_t\end{pmatrix} \leftarrow \begin{pmatrix}\sigma\\ \sigma\\ \sigma\\ \tanh\end{pmatrix}\left(W\begin{pmatrix}\mathbf{h}_{t-1}\\ \mathbf{x}_t\end{pmatrix} + \mathbf{b}\right) \quad (1)$$

$$\mathbf{c}_t \leftarrow \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \quad (2)$$

$$\mathbf{h}_t \leftarrow \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (3)$$

Note that following the convention, $c_0$ and $h_0$ are set to 0. $\sigma(z) = (1 + e^{-z})^{-1}$ is the sigmoid function, and $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ is the hyperbolic tangent function, the activation functions are applied in an element-wise manner. $\odot$ denotes the element-wise product.

The hidden state of the last time-step $\mathbf{h}_T$ is taken as the output of the encoder LSTM (i.e., the encoding of the input time-series) and subsequently fed into the decoder LSTM.

**Text S2**

A decoder LSTM with hidden dimension $h$ and $T$ time-steps is parametrized by a weight matrix $W \in \mathbb{R}^{4h\times h}$ and bias vector $\mathbf{b} \in \mathbb{R}^{4h}$. It takes the encoding of a time-series $\mathbf{h}_0 \in \mathbb{R}^h$ as the input and computes the hidden states $\mathbf{h}_t \in \mathbb{R}^h$ and cell states $\mathbf{c}_t \in \mathbb{R}^h$, via the intermediate variables input gate $\mathbf{i}_t \in \mathbb{R}^h$, forget gate $\mathbf{f}_t \in \mathbb{R}^h$, output gate $\mathbf{o}_t \in \mathbb{R}^h$ and gate gate $\mathbf{g}_t \in \mathbb{R}^h$ for each step $1 \leq t \leq T$ recurrently as follows:

$$\begin{pmatrix}\mathbf{i}_t\\ \mathbf{f}_t\\ \mathbf{o}_t\\ \mathbf{g}_t\end{pmatrix} \leftarrow \begin{pmatrix}\sigma\\ \sigma\\ \sigma\\ \tanh\end{pmatrix}(W\mathbf{h}_{t-1} + \mathbf{b}) \quad (4)$$

$$\mathbf{c}_t \leftarrow \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \quad (5)$$

$$\mathbf{h}_t \leftarrow \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (6)$$

Note that similar to Text S1, $\mathbf{c}_0$ is set to 0. The hidden states of each time-step $\{\mathbf{h}_t\}_{t=1}^{T}$ is taken as the output of the decoder LSTM and is subsequently fed into the dense layer.

**Table S1 Metrics used for model evaluation.**

| Metric | Formula |
|---|---|
| Mean bias error (MBE) | $\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i - y_i$ |
| Mean absolute error (MAE) | $\frac{1}{N}\sum_{i=1}^{N} \lvert y_i - \hat{y}_i \rvert$ |
| Root mean square error (RMSE) | $\sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$ |
| Symmetric mean absolute percentage error (SMAPE) | $\frac{100\%}{N}\sum_{i=1}^{N} \frac{\lvert y_i - \hat{y}_i \rvert}{(\lvert y_i \rvert + \lvert \hat{y}_i \rvert)/2}$ |
| Pearson correlation coefficient (R) | $\frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}}$ |

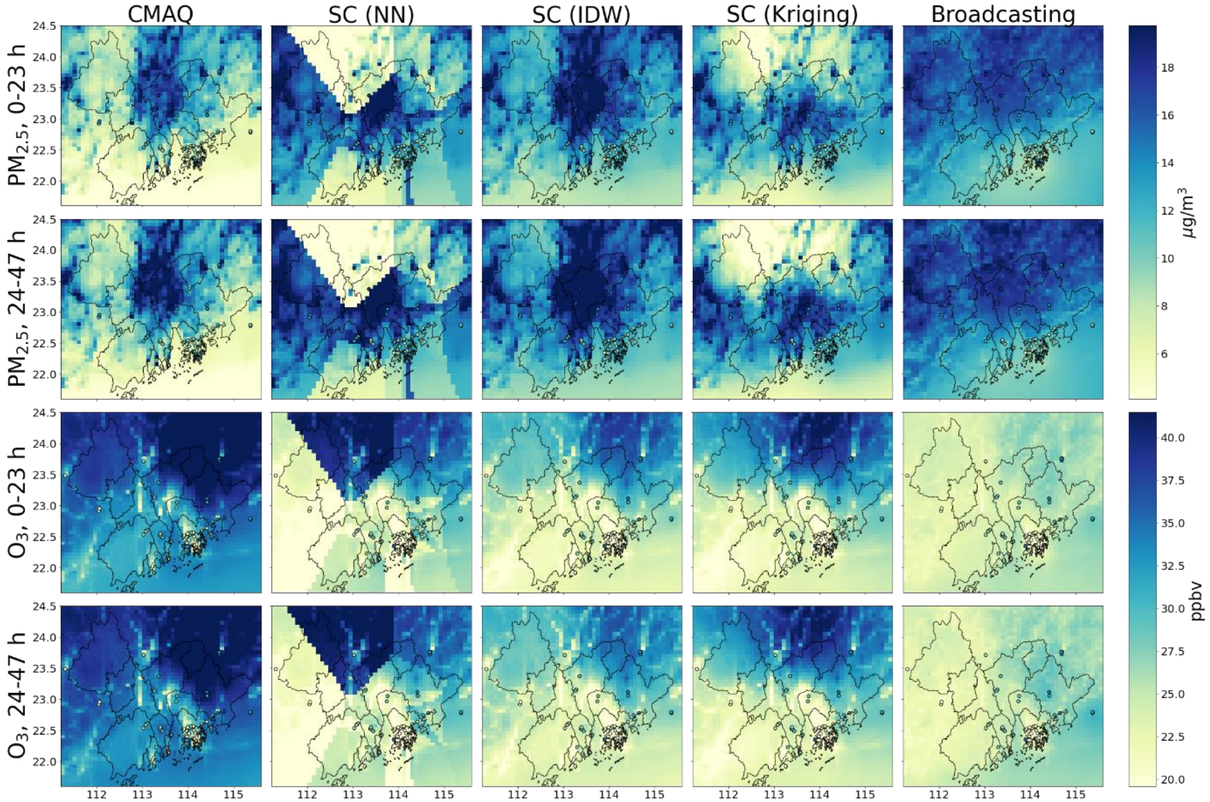


**Figure S1 The regional forecast results in May 2021.**

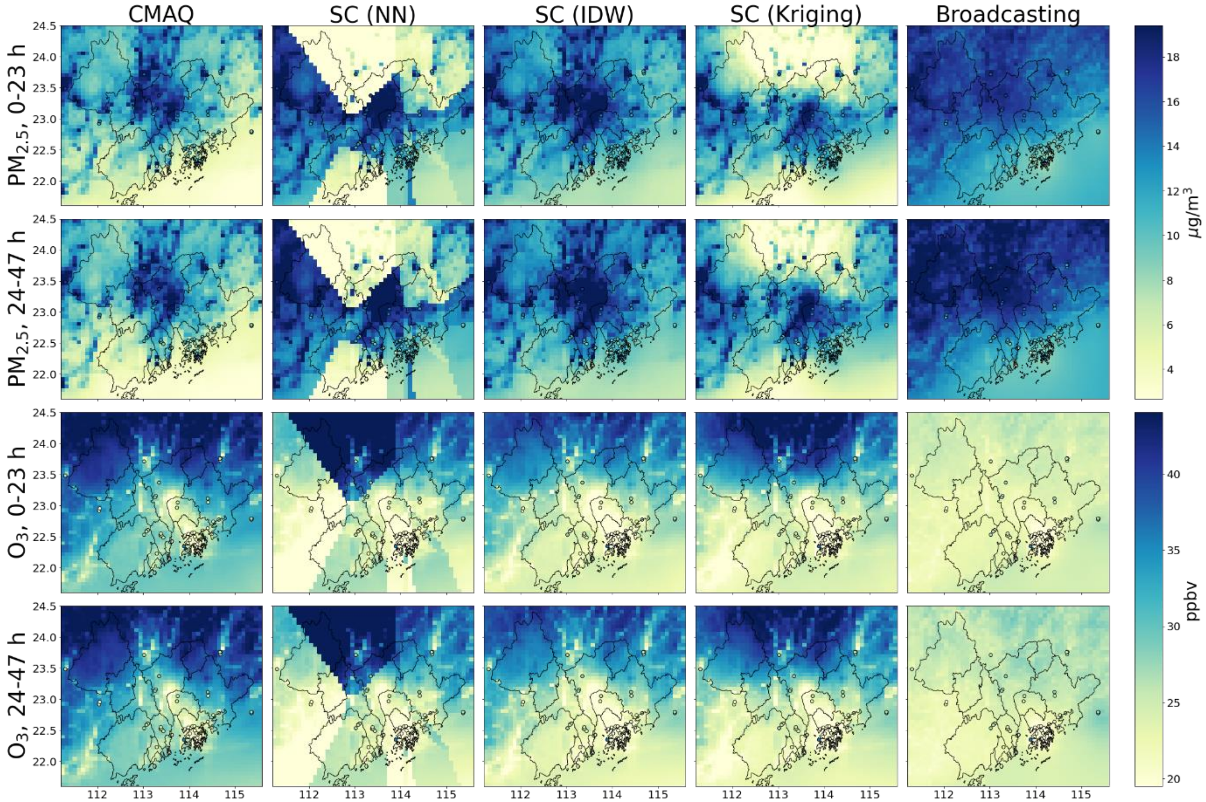


**Figure S2 The regional forecast results in August 2021.**

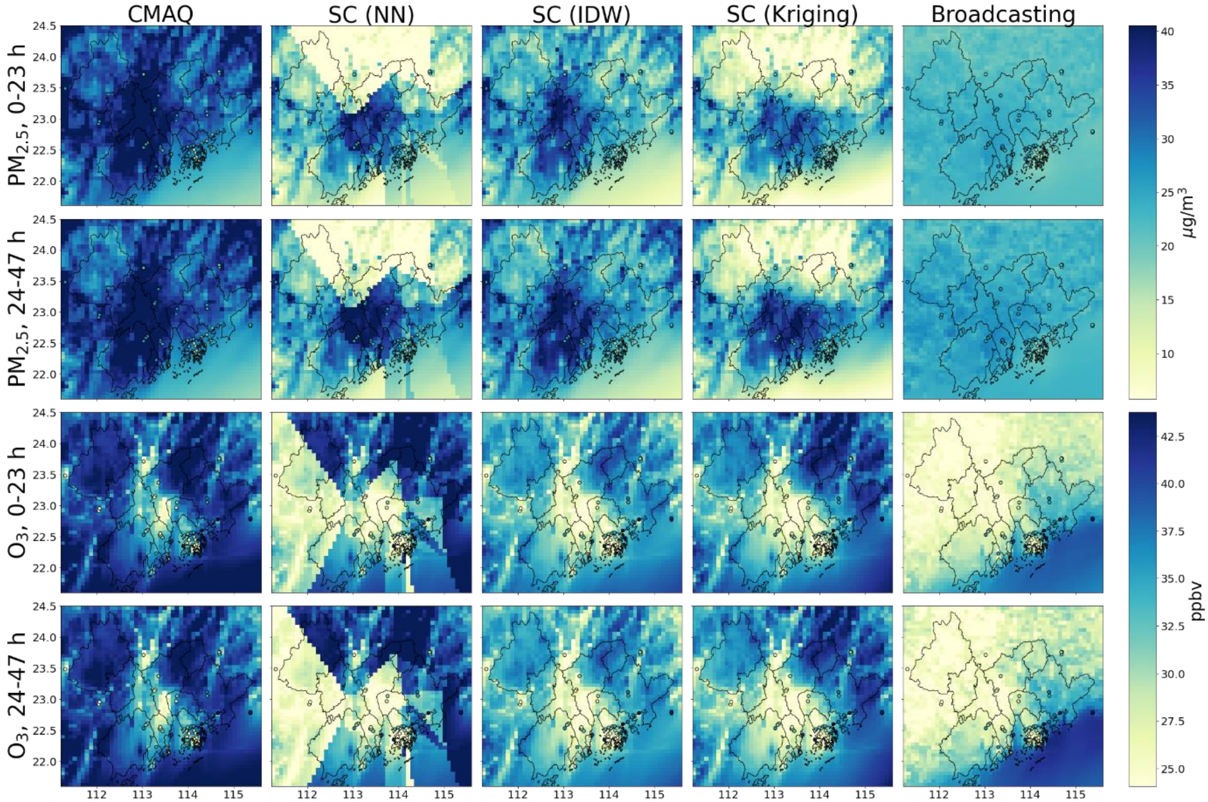


**Figure S3 The regional forecast results in November 2021.**