List S1

- **WRF**
  - PSFC (surface pressure)
  - U10 (10-meter U wind)
  - V10 (10-meter V wind)
  - T2 (2-meter temperature)
  - Q2 (2-meter specific humidity)

- **CMAQ**
  - CO (carbon monoxide)
  - SO2 (sulfur dioxide)
  - NO2 (nitrogen dioxide)
  - O3 (ozone)
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 $b \in \mathbb{R}^{4h}$. It takes the input of a time series with length $T$ and dimension $d$, $\{x_t\}_{t=1}^T$ and computes the hidden states $h_t \in \mathbb{R}^h$ and cell states $c_t \in \mathbb{R}^h$ as well as the intermediate variables input gate $i_t \in \mathbb{R}^h$, forget gate $f_t \in \mathbb{R}^h$, output gate $o_t \in \mathbb{R}^h$ and gate gate $g_t \in \mathbb{R}^h$ for each step $1 \leq t \leq T$ recurrently as follows:

\[
\begin{pmatrix}
    i_t \\
    f_t \\
    o_t \\
    g_t
\end{pmatrix}
\leftarrow
\begin{pmatrix}
    \sigma \\
    \sigma \\
    \sigma \\
    \text{tanh}
\end{pmatrix}
(W h_{t-1} + b) \tag{1}
\]

\[
c_t \leftarrow f_t \odot c_{t-1} + i_t \odot g_t \tag{2}
\]

\[
h_t \leftarrow o_t \odot \text{tanh}(c_t) \tag{3}
\]

Note that following the convention, $c_0$ and $h_0$ are set to 0. $\sigma(z) = \left(1 + e^{-z}\right)^{-1}$ is the sigmoid function, and \(\text{tanh}(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}\) 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 $h_T$ 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 $b \in \mathbb{R}^{4h}$. It takes the encoding of a time-series $h_0 \in \mathbb{R}^h$ as the input and computes the hidden states $h_t \in \mathbb{R}^h$ and cell states $c_t \in \mathbb{R}^h$, via the intermediate variables input gate $i_t \in \mathbb{R}^h$, forget gate $f_t \in \mathbb{R}^h$, output gate $o_t \in \mathbb{R}^h$ and gate gate $g_t \in \mathbb{R}^h$ for each step $1 \leq t \leq T$ recurrently as follows:

\[
\begin{pmatrix}
    i_t \\
    f_t \\
    o_t \\
    g_t
\end{pmatrix}
\leftarrow
\begin{pmatrix}
    \sigma \\
    \sigma \\
    \sigma \\
    \text{tanh}
\end{pmatrix}
(W h_{t-1} + b) \tag{4}
\]

\[
c_t \leftarrow f_t \odot c_{t-1} + i_t \odot g_t \tag{5}
\]

\[
h_t \leftarrow o_t \odot \text{tanh}(c_t) \tag{6}
\]

Note that similar to Text S1, $c_0$ is set to 0. The hidden states of each time-step $\{h_t\}_{t=1}^T$ is taken as the output of the decoder LSTM and is subsequently fed into the dense layer.
<table>
<thead>
<tr>
<th>Metric</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean bias error (MBE)</td>
<td>$\frac{1}{N} \sum_{i=1}^{N} \hat{y}_i - y_i$</td>
</tr>
<tr>
<td>Mean absolute error (MAE)</td>
<td>$\frac{1}{N} \sum_{i=1}^{N}</td>
</tr>
<tr>
<td>Root mean square error (RMSE)</td>
<td>$\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$</td>
</tr>
<tr>
<td>Symmetric mean absolute percentage error (SMAPE)</td>
<td>$\frac{100%}{N} \sum_{i=1}^{N} \frac{</td>
</tr>
<tr>
<td>Pearson correlation coefficient (R)</td>
<td>$\frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}<em>i - \bar{\hat{y}})}{\sqrt{\sum</em>{i=1}^{N}(y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}$</td>
</tr>
</tbody>
</table>
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