Real-Time Spatiotemporal Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analysis

Pratyush Muthukumar[*], Emmanuel Cocom[*], Jeanne Holm[**], Dawn Comer[**], Anthony Lyons[**], Irene Burga[**], Christa Hasenkopf[***], and Mohammad Pourhomayoun[*]

[*]Department of Computer Science, California State University Los Angeles, [pmuthuk2, ecocom, mpourho]@calstatela.edu

[**]City of Los Angeles, [jeanne.holm, dawn.comer, anthony.lyons, irene.burga]@lacity.org, ***OpenAQ, christa@openaq.org

ABSTRACT

Air pollution is responsible for the early deaths of 7 million people every year in the world. The first and the most important step in mitigating the air pollution risks is to understand it, discover the patterns and sources, and predict it in advance. Real-time air pollution prediction requires a highly complex model that can solve this spatiotemporal problem in multiple dimensions. Using a combination of spatial predictive models (deep Convolutional Neural Networks) and temporal predictive models (deep Long Short-Term Memory), we utilized the Convolutional LSTM structure that learns correlations between various points of location and time. We created a sequential encoder-decoder network that allows for accurate air pollution prediction 10 days in advance using data of 10 days in the past in the county of Los Angeles on a Nitrogen Dioxide metric. Through a 5D tensor reformatting of air quality satellite image data, we provide a prediction for Nitrogen Dioxide in various areas of Los Angeles over various time periods.

INTRODUCTION

Air pollution is a silent killer. It is responsible for the early deaths of 7 million people every year, around 600,000 of whom are children [1]. It means that every 5 seconds, somebody around the world dies prematurely from the effects of air pollution [1]. With the percentage of the global population living in urban areas projected to increase from 54% in 2015 to 68% in 2050 and in the U.S. up to 89%, the prevention of a significant increase in air pollution-related loss of life requires comprehensive mitigation strategies, as well as forecast systems, to limit and reduce the exposure to harmful urban air [2][3].

In this paper, we developed predictive models based on advanced machine learning algorithms to discover and classify patterns in urban air quality and predict air pollution in different areas. In designing the predictive models, we considered both temporal and spatial patterns in the data. The air quality data at a specific location is highly correlated to the past data at that location (temporal correlation) [4]. Also, it is highly correlated to air quality of adjacent areas because the air pollutants can transmit through the atmosphere from one area to other areas around it (spatial correlation).

The spatiotemporal machine learning problem has been studied recently in various contexts including weather-related applications [5][6][7]. Much of the previous research focuses on increasing either spatial or temporal correlations, but it is considerably more difficult to introduce an accurate measurement of both spatial correlation and temporal correlation in a highly complex model.

The Convolutional Long Short-Term Memory (ConvLSTM) model is a complex machine learning model used for inputs that comprise of sets of frames of data, which allows for unaltered video inputs. The ConvLSTM model is a variation of the traditional Long Short-Term Memory Network, a time-series Recurrent Neural Network. Recall the Fully Connected LSTM (FC-LSTM) structure with characteristic input, forget, and memory gates:

The key equations for the FC-LSTM are as follows:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \\
i_f &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}x_{h-1} + W_{co} \circ c_t + b_o) \\
h_t &= o_t \circ \tanh(c_t),
\end{aligned}
$$

where $\circ$ denotes the Hadamard product [8].

Notice that in traditional LSTM models, the input vector must be a one-dimensional vector, where each element of the vector is fed into different states of the model. Accordingly, the output of the LSTM is also a one dimensional vector parameterized by time. Traditional LSTMs do not account for the spatial correlations in an environment, and in order to introduce these spatial correlations, we induced convolution throughout each of the gates and cell/hidden states.

There are two options to induce convolution in a traditional ConvLSTM, depending on the sequence of convolution. Either the convolution can intermediately occur on the input tensor before being fed into a traditional FC-LSTM, or the convolution can accept input tensors as input and perform convolution operations when calculating the values of the cell and hidden states throughout the model. The first option refers to performing the convolution operation on a multidimensional input, transforming it into a single dimensional input, prior to its input in the FC-LSTM structure. Research in this field denotes the first option as a Convolutional Neural Network - Long Short Term Memory (CNN-LSTM), as the individual

structures are not changed, and the second option as a Convolutional LSTM (ConvLSTM), as the traditional LSTM's cell operations must be changed, referring to the addition of the convolution operation in each of the states of the FC-LSTM[9] [10]. If we allow convolution to replace the Hadamard products in the traditional LSTM structure, we can now input multidimensional input tensors to the structure and receive a multidimensional output tensor as output. The key equations for the ConvLSTM then are defined as

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} * c_{t-1} + b_i) \\
i_f &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} * c_{t-1} + b_f) \\
c_t &= f_t * c_{t-1} + i_t * \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}x_{h-1} + W_{co} * c_t + b_o) \\
h_t &= o_t * \tanh(c_t),
\end{aligned}
$$

where $*$ denotes the convolution operation [11].

METHODS

Dataset

Our input data was sourced from the U.S. Geological Survey's EarthExplorer database records of the Sentinel 2 satellite, launched on June 23, 2015 [12]. The Sentinel 2 satellite launched by the European Space Agency in March 2015 images and records atmospheric and terrain data through 13 spectral bands based on the wavelength of the emitted light. Sentinel 2 operates along a 290-km orbital swath [13]. Of these 13 bands, we selected two spectral bands that provided data on the air pollution in the greater Los Angeles area. The first band with a 442.7 nm central wavelength measured the coastal aerosol levels allowing us to view fine particulate matters including dust, smoke, and general particulate matter. We also chose a finer spectral band at 945.1 nm central wavelength to measure specifically the Nitrogen Dioxide levels in the atmosphere. A sample input is shown in Figure 1.

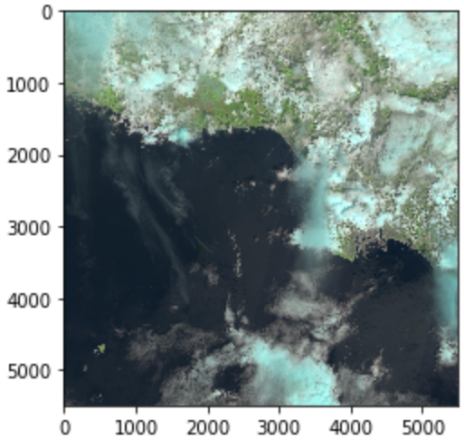

Figure 1. Sample Raw Data (Source: USGS EarthExplorer database of satellite imagery of Los Angeles taken on April 29, 2019 by ESA's Sentinel 2 satellite)

The blue structures correspond to strictly Nitrogen Dioxide air pollution, while the white, cloud-like structures correspond to general particulate matter.

Data Preprocessing

In order to feed our data into the ConvLSTM model, we transformed 225 GeoTIFF high resolution images into a 5D tensor. The 225 GeoTIFF images corresponded to 1642 days of data, with each image being the imaging of the T11SLT 100km $\times$ 100km^2 tile, which is roughly the western 75% of Los Angeles County. Each of the 225 GeoTIFF images were taken 2 days apart, the orbit time of the Sentinel 2 satellite.

For our ConvLSTM input, we decided to focus only on the blue cloud-like structures that imaged Nitrogen Dioxide. To do so, we first imported our dataset into a JPEG format through the OpenCV python package. Since the size of the data is very large, we resampled the data into two batches each with different lower resolutions.One resampling was a 400px by 400px JPEG image dataset of all 225 images and the other was a 40px by 40px JPEG image dataset of all 225 images. For both of the lower resolution resamplings, we applied a mask that only showed the (0,60,60) to (225,255,255) RGB color scheme, roughly corresponding to only light blue shades. We set all other non light blue shades to the (0,0,0) RGB color scheme or black. An example masked image is shown in Figure 2.

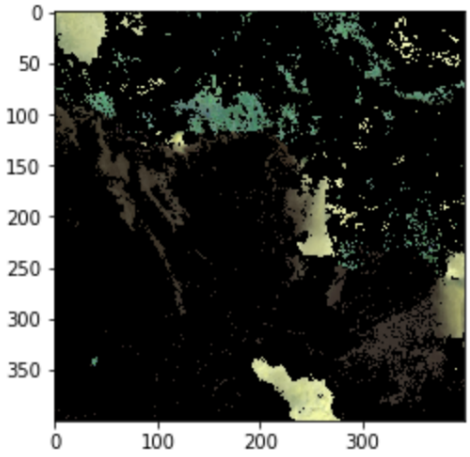

Figure 2. Masked Image

In this way, we had created four resamplings of the orignal image dataset: two masked 40px by 40px datasets and two masked 400px by 400px datasets. For two of the datasets with a mask, one 40px by 40px and one 400px by 400px, we binarized the data, so that all light blue pixels became 1, and all black pixels became 0. Thus, two of our datasets were binary arrays. For the remaining two image datasets, we kept the masked light blue and black color scheme intact. In this way, we had two datasets of masked color images (RGB) and two datasets of bit arrays. The color image datasets allowed us to use a ReLU activation function in the output layer, while the binary array datasets allowed us to use a sigmoid activation function.

After resampling the data into various groups based on binary image, masked image, or image resolution criterion, we began parsing our 225 image datasets into readable formats for our ConvLSTM models. In the current form, our image dataset was in the form of (225,400,400,3) for 400px by 400px masked images, (225,40,40,3) for 40px by 40px masked images, (225,400,400,1) for 400px by 400px binary images, and (225,40,40,1) for 40px by 40px binary images. We then took all of our datasets and batched every five image frame as one sample.

An overview of the input data preprocessing is described in Figure 3.

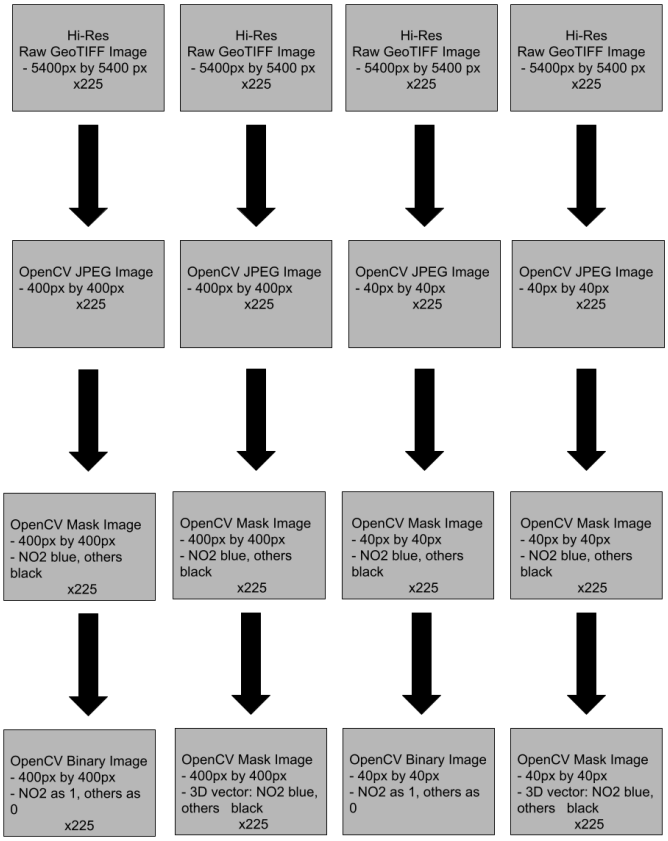

Figure 3. Data Preprocessing High-Level Process

Input Data Labelling

An important determinant of the success and practicality of a time series model is based on the selection of the training and testing sets from a continuous stream of time-indexed data. In our model, the label for a training set input is the frame that directly follows the current frame. In this way, the model is picking up on the pattern to base a current input image on the goal of accurately predicting a future image which in this case means predicting the next 10 days in the future.

In our model, we use five input frames as a data sample to predict five future frames. In fact, since each frame was separated by 2 days, we use 10 days of previous satellite imagery data in order to predict 10 days in the future of satellite imagery data. For the input labels, we first shifted over all data by five frames, with the first 5 frames corresponding to the 6th through 10th frames and the last 5 frames corresponding to the 226th through 230th frame. The 226th through 230th frames were filled with an averaging of the 30 days prior. Thus, we encoded sequential samples in training/testing inputs indexed by time and training/testing labels indexed by time, but did not overlap the training/testing inputs and the training/testing labels. We were able to perform a continuous prediction from the model's learned patterns. A visual representation of our labelling is shown in Figure 4:

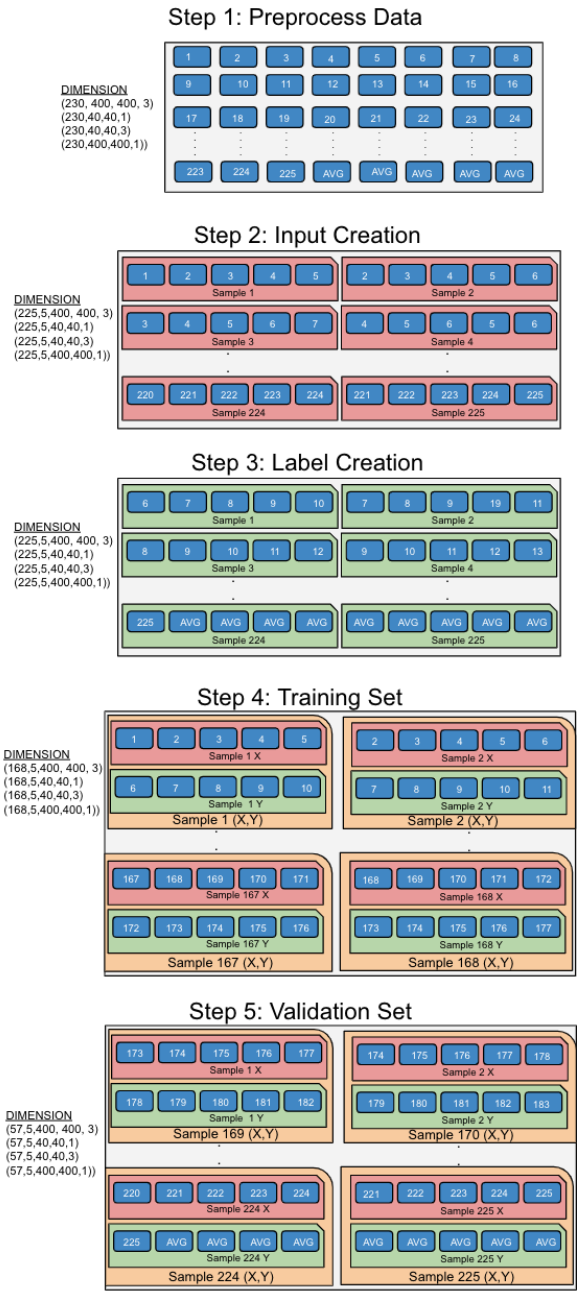

Figure 4. Training/Validation Input and Label Creation Process for sequential training and testing data without overlapping inputs and labels.

RESULTS

Aerial Satellite Image Model

We developed ConvLSTM models that learned correlations through spatial and temporal bounds. We predicted 5 frames in the future from 5 frames in the past, with each frame being 2 days apart. For Figures 5-7, we visualized the prediction of 10 days worth of Nitrogen Dioxide levels through data from the previous 10 days. Figures 5-7 display the predictions on the masked 40px by 40px resampled dataset. We have recolored the figures to accentuate the nature of the prediction, however its original form remains an RGB image with a light blue and black color scheme. We used our data to essentially predict early March 2020's Nitrogen Dioxide air pollution levels from late February 2020's Nitrogen Dioxide pollution data. For figures 5-7, the left column displays our predictions, while the right column displays the ground truth.

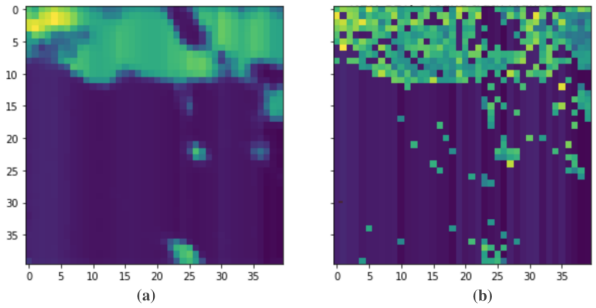

Figure 5. Frame 1 Prediction: 2nd day in the future prediction of Nitrogen Dioxide air pollution in Los Angeles County from previous 10 days of data, (a) Prediction, (b) Ground Truth

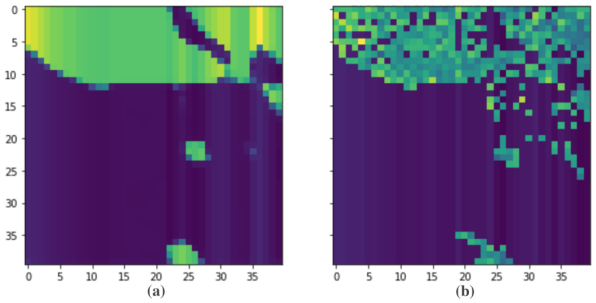

Figure 6. Frame 2 Prediction: 10th day in the future prediction of Nitrogen Dioxide air pollution in Los Angeles County from previous 10 days of data, (a) Prediction, (b) Ground Truth

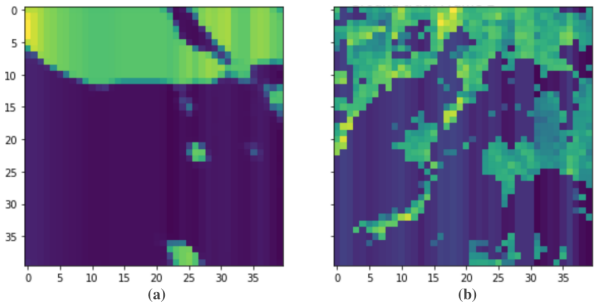

Figure 7. Last Frame Prediction: 10th day in the future prediction of Nitrogen Dioxide air pollution in Los Angeles County from previous 10 days of data, (a) Prediction, (b) Ground Truth

As expected, the prediction of the near future is more accurate than farther in the future. In the context of the problem, this is because the Nitrogen Dioxide levels of tomorrow are more correlated to the past week's Nitrogen Dioxide Levels than the future week's Nitrogen Dioxide levels. As we move to 10+ days in the future, the data from the previous 10 days are no longer strongly correlated to accurately model Nitrogen Dioxide particulate matter in the greater Los Angeles area.

Error Analysis

We used the Structural Similarity Index Measurement (SSIM) as a error measure to quantify the model's accuracy. SSIM is a very popular tool used for accurate assessment of weather prediction algorithms due to its ability to judge models on the similarities in the structure of a prediction [14]. Possible outputs of the SSIM metric range from 0 to 1, with 0 being completely dissimilar and 1 being exactly identical. With true pixel value p and predicted value $\hat{p}$, the SSIM of a single pixel is

$$\text{SSIM}(p,\hat{p}) = \frac{2\mu_p\mu_{\hat{p}} + c_1}{(\mu_p^2 + \mu_{\hat{p}}^2 + c_1)\sigma_p^2 + \sigma_{\hat{p}}^2 + c_2},$$

where c_1 and c_2 are constants relating to the relative noise of an image [15]. For a complete image I, the SSIM is

$$SSIM(I) = \sum_{p \in I} \frac{2\mu_p\mu_{\hat{p}} + c_1}{(\mu_p^2 + \mu_{\hat{p}}^2 + c_1)\sigma_p^2 + \sigma_{\hat{p}}^2 + c_2}$$

[15]. We then ran our SSIM analysis metrics on our binary input and masked images with our first frame having the most accurate SSIM value in the two samples or ten frames of data we predicted. The future frames fell in accuracy by around $10-15\%$ due to the ground truth being less correlated to the previous 10 days of Nitrogen Dioxide levels as compared to the first few frames of the ground truth.

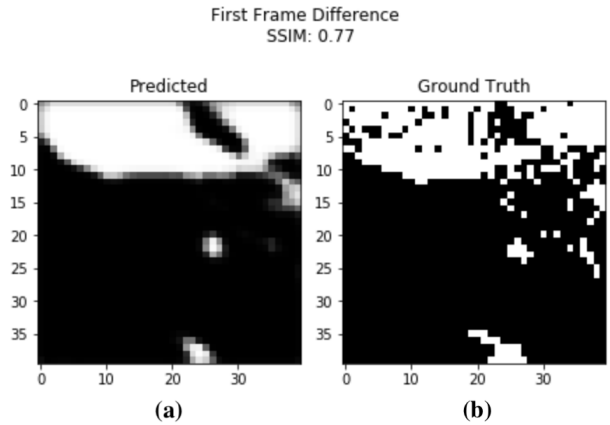

Figure 8. First Frame SSIM Measurement: 77% structurally similar from SSIM error metric of 2nd day in the future prediction of Nitrogen Dioxide air pollution in Los Angeles County from previous 10 days of binary input data. (a) Prediction, (b) Ground Truth

Table 1. SSIM Values for First Sample (First 5 Frames): structural similarity percentages of ten days in the future Nitrogen Dioxide predictions in LA County from previous ten days of data.

Sample 1	
	SSIM
Frame 1	0.77
Frame 2	0.70
Frame 3	0.63
Frame 4	0.56
Frame 5	0.51

CONCLUSION

In this paper, we developed predictive models based on advanced machine learning algorithms to discover and classify patterns in urban air quality, and specifically predict Nitrogen Dioxide in greater Los Angeles area. In designing the predictive models, we took into account both temporal and spatial patterns in the data (i.e. the air quality correlation to the past and future data as well as the correlation to the adjacent locations).

To discover and learn both temporal and spatial patterns, we developed a Convolutional Long Short-Term Memory (ConvLSTM) model, which is a complex machine learning approach used for inputs that comprise of sets of frames of multi-dimensional data. Our model was able to use the Nitrogen Dioxide air pollution data in the greater Los Angeles area of 10 days in order to predict Nitrogen Dioxide pollution anywhere in Los Angeles 10 days in the future.

This work can be used to alert researchers on the patterns of movement for Nitrogen Dioxide at any given time period in the next five years within the greater Los Angeles area.

FUTURE WORK

In the future, we would also like to analyze ground-based sensors with a wide variety of pollutant and atmospheric sensors including PM2.5, carbon monoxide, ozone, temperature, and windspeed.

This study could expand to larger areas past Los Angeles county and include satellite image data and ground-based sensor data in to order include more information when predicting air pollution.

REFERENCES

[1] With a premature death every five seconds, air pollution is violation of human rights, says un expert – united nations sustainable development. Website. `https://www.un.org/sustainabledevelopment/`.

[2] Earl Swigert. Unicef: An urban world. Website. `https://www.unicef.org/sowc2012/urbanmap`.

[3] 2018 revision of world urbanization prospects | multimedia library - united nations department of economic and social affairs. Website. `https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html`.

[4] Nino Künzli, Michael Jerrett, Wendy J Mack, Bernardo Beckerman, Laurie LaBree, Frank Gilliland, Duncan Thomas, John Peters, and Howard N Hodis. Ambient air pollution and atherosclerosis in los angeles. *Environmental health perspectives*, 113(2):201–206, 2005.

[5] Yijun Lin, Nikhit Mago, Yu Gao, Yaguang Li, Yao-Yi Chiang, Cyrus Shahabi, and José Luis Ambite. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 359–368, 2018.

[6] Chiou-Jye Huang and Ping-Huan Kuo. A deep cnn-lstm model for particulate matter (pm2. 5) forecasting in smart cities. *Sensors*, 18(7):2220, 2018.

[7] Sandip Roy, Yan Wan, Christine Taylor, and Craig Wanke. A stochastic network model for uncertain spatiotemporal weather impact at the strategic time horizon. In *10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, page 9348, 2010.

[8] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[9] Seongchan Kim, Ji-Sun Kang, Minho Lee, and Sa-kwang Song. Deeptc: Convlstm network for trajectory prediction of tropical cyclone using spatiotemporal atmospheric simulation data. 2018.

[10] Rafaela C Nascimento, Yania M Souto, Eduardo Ogasawara, Fabio Porto, and Eduardo Bezerra. Stconvs2s: Spatiotemporal convolutional sequence to sequence network for weather forecasting. *arXiv preprint arXiv:1912.00134*, 2019.

[11] Yipeng Liu, Haifeng Zheng, Xinxin Feng, and Zhonghui Chen. Short-term traffic flow prediction with conv-lstm. In *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pages 1–6. IEEE, 2017.

[12] USGS. Usgs earthexplorer satellite imagery database. Website. `https://earthexplorer.usgs.gov/`.

[13] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[14] Benjamin Klein, Lior Wolf, and Yehuda Afek. A dynamic convolutional layer for short range weather prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4840–4848, 2015.

[15] Mehul P Sampat, Zhou Wang, Shalini Gupta, Alan Conrad Bovik, and Mia K Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE transactions on image processing*, 18(11):2385–2401, 2009.