Supporting information for:

Classification and quantification of low-visibility events using deep learning over eastern China

Yuting Liang[1], Shitong Zhao[1], Dantong Liu[1*], Changhao Wu[1,2,3], Li Yi[4], Xingcan Jia[5]

[1]School of Earth Sciences, Zhejiang University, Hangzhou, China

[2]Yunnan University, Kunming, China

[3]Yunnan International Joint Laboratory of Monsoon and Extreme Climate Disasters, Kunming, China

[4]Ocean University of China, Qingdao, China

[5]Institute of Urban Meteorology (IUM), Chinese Meteorological Administration, Beijing, China

Contents of this file

Supporting information, including data preprocessing and standardization, network architecture and attention mechanisms, hybrid loss function definition, and model training configuration (Texts S1–S4), illustration of the study domain and sampling method (Figure S1), schematic of the CNN model configuration (Figure S2), seasonal spatial patterns of $PM_{2.5}$ concentration (Figure S3); seasonal responses of simulated fog and haze area to $PM_{2.5}$ concentration (Figure S4) and summary of data sources, model performance metrics, and comparison with baseline models (Tables S1–S3)

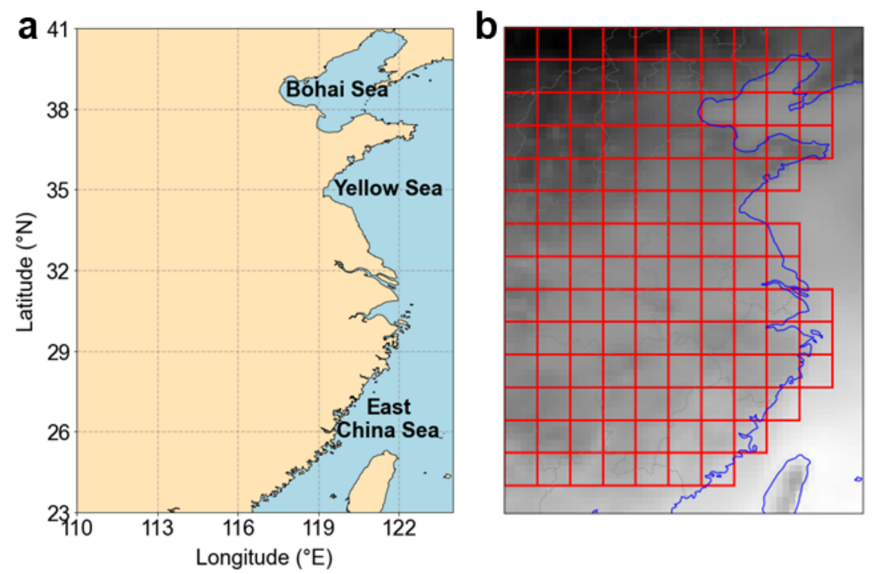

Figure S1 Study domain and sampling method. (a) The geographical study domain includes the coastal areas of the Bohai, Yellow, and East China Seas. (b) Schematic of the window sampling method, where the domain is overlaid with the sampling grid (red).

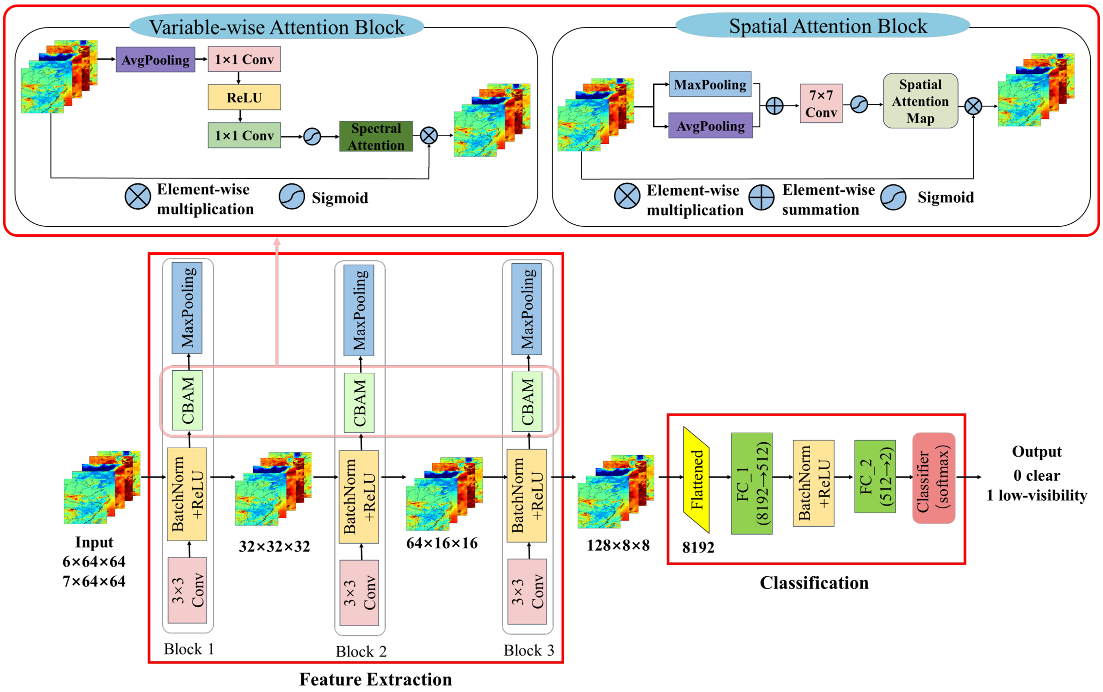

Figure S2 Configuration of the CNN model. The model is composed of a feature extraction module with three sequential blocks, with each block containing a 3×3 convolutional layer (Conv), batch normalization (BN), rectified linear unit (ReLU), convolutional block attention module (CBAM), and Max-Pooling. The CBAM contains two attention mechanisms: a channel attention module to adaptively re-weight the importance of each feature channel, and a spatial attention module to highlight the most salient spatial regions within the feature maps. The classification module first flattens the feature maps into a vector, which is then processed through a fully connected layer (FC1), followed by a BN layer and a ReLU activation, a second fully connected layer (FC2), and a final softmax classifier to output the probabilities for the 'low-visibility' and 'clear' classes.

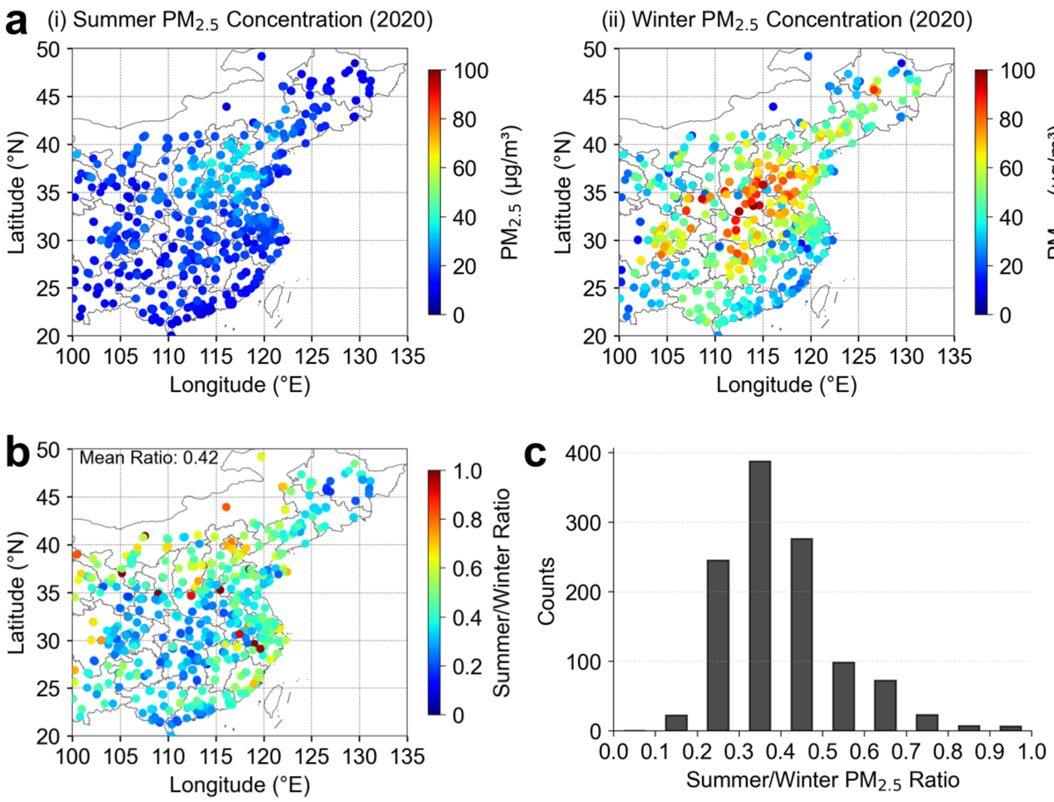

Figure S3 Seasonal spatial patterns of $PM_{2.5}$ concentration in 2020. (a) Spatial distributions of mean $PM_{2.5}$ concentration (μg/m³) for (i) summer and (ii) winter, and the resulting summer-to-winter ratio map. (c) Histogram showing the frequency distribution of the summer-to-winter $PM_{2.5}$ ratio from all stations.

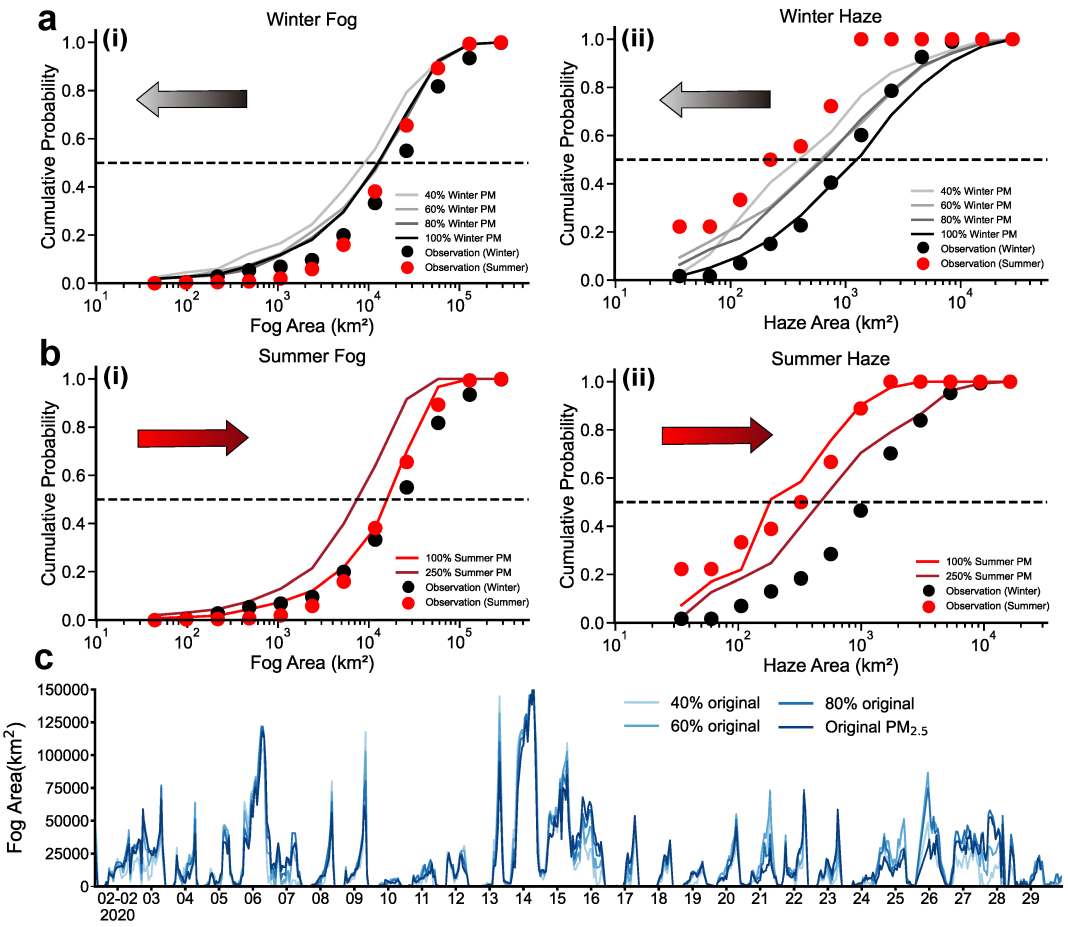

Figure S4 Seasonal responses of simulated fog and haze area to $PM_{2.5}$ concentration. (a) Cumulative probability distributions for simulated winter (i) fog and (ii) haze area at original (100%) and reduced (80%, 60%, 40%) $PM_{2.5}$ concentrations, compared against observed winter and summer values. (b) Cumulative probability distributions for simulated summer (i) fog and (ii) haze area at original (100%) and increased (250%) $PM_{2.5}$ concentrations, compared against observed winter and summer values (black dots). (c) Time series of simulated total fog area (km²) under varying $PM_{2.5}$ reduction scenarios for February 2020.

Table S1 Data used in this study. This study uses several datasets as model inputs: ECMWF global meteorological reanalysis, and FY-4A satellite data, and hourly $PM_{2.5}$ measurements from the China National Environmental Monitoring Centre (CNEMC). The ground-truth binary mask is derived from Chinese meteorological observations sourced from the U.S. National Climatic Data Center (NCDC/NOAA). The annual population distribution data for 2020 at 1 km resolution was sourced from WorldPop. All datasets cover the 2019–2020 period.

Name	Unit	Data type	Data sources	Time span
2m temperature	K	Hourly/0.25°	ECMWF	2019–2020
2m dewpoint temperature	K			
brightness temperature (3.72 μm)	K	15-min/4km	FY-4A/NSMC	2019–2020
brightness temperature (10.8 μm)	K			
solar zenith angle	°	15-min/4km	FY-4A/NSMC	2019–2020
$PM_{2.5}$ measured data	μg/m³	Hourly/Station	CNEMC	2019–2020
ground-based visibility observation data	km	Hourly/Station	NCDC	2019-2020
population distribution	km	Annual/1km	WorldPop	2020

Note:

(ECMWF: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels,

FY-4A/NSMC: http://data.nsmc.org.cn/DataPortal/cn/data/detail.html,

CNEMC: http://air.cnemc.cn:18007/,

NCDC: ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/,

WorldPop: https://www.worldpop.org/)

Table S2 Summary of model performance metrics (Accuracy, Recall, Precision, and F1-score) for the two models, evaluated for both ‘low-visibility’ and ‘clear’ classes.

	Sat-PM	Sat
Accuracy	0.93	0.91
clear Recall	0.95	0.94
clear Precision	0.96	0.95
clear F1-score	0.96	0.95
low-visibility Recall	0.81	0.69
low-visibility Precision	0.75	0.73
low-visibility F1-score	0.78	0.71

Note: Model performance was evaluated based on a binary classification of ‘low-visibility’ (visibility ≤ 2 km) and ‘clear’ (visibility > 2 km), with ‘low-visibility’ treated as the positive class. The metrics are derived from a 2×2 contingency table where:

a = True Positive (TP): The number of correctly identified low-visibility events.

b = False Positive (FP): The number of clear events incorrectly identified as low-visibility events.

c = False Negative (FN): The number of clear events incorrectly identified as low-visibility events (i.e., missed fog events).

d = True Negative (TN): The number of correctly identified low-visibility events.

$$\text{Accuracy} = \frac{\mathrm{a} + \mathrm{d}}{\mathrm{a} + \mathrm{b} + \mathrm{c} + \mathrm{d}}$$

$$\text{clear Recall} = \frac{\mathrm{b}}{\mathrm{b} + \mathrm{d}}$$

$$\text{clear Precision} = \frac{\mathrm{d}}{\mathrm{c} + \mathrm{d}}$$

$$\text{low-visibility Recall} = \frac{\mathrm{a}}{\mathrm{a} + \mathrm{c}}$$

$$\text{low-visibility Precision} = \frac{\mathrm{a}}{\mathrm{a} + \mathrm{b}}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision} + \text{Recall}}$$

Table S3 Performance comparison of the Sat model with baseline configurations. The table lists key performance metrics (Accuracy, Recall, and Precision) for the proposed Sat model against three ablation study baselines: a standard CNN (no attention), a CNN with only channel attention, and a CNN with only spatial attention. Metrics are provided for both ‘low-visibility’ and ‘clear’ classes.

Model	Accuracy	low-visibility Recall	low-visibility Precision	clear Recall	clear Precision
Sat (Ours)	90.95%	69%	73%	94%	95%
CNN (no attention)	89.71%	71%	67%	93%	94%
CNN (channel attention)	89.87%	70%	68%	94%	94%
CNN (spatial attention)	89.25%	67%	67%	94%	94%

Text S1 Data Preprocessing Details

The data for this study were compiled from satellite remote sensing, meteorological reanalysis, and ground-based observations for the period 2019–2020, with a detailed summary provided in Table S1. Satellite inputs included brightness temperatures and the solar zenith angle from the FY-4A satellite (NSMC). For meteorology, we used 2-meter temperature and dewpoint temperature from the hourly, 0.25° resolution European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 global reanalysis. Ground-based observations from individual monitoring stations consisted of hourly $PM_{2.5}$ concentrations from the China National Environmental Monitoring Centre (CNEMC) and visibility data from the U.S. National Climatic Data Center (NCDC/NOAA). To ensure temporal consistency, all data were standardized to UTC+8. And to ensure spatial consistency across multi-source data, all inputs were resampled to the FY-4A grid (4 km resolution). Missing values were filled via nearest-neighbor interpolation. The input samples were cropped to a spatial size of 142 km × 142 km (Figure S1b) to align with the model input requirements.

To characterize the moisture conditions required for fog formation, we utilize 2-m temperature and dew point depression from ERA5 reanalysis (Hersbach et al., 2020). A dew point depression approaching 0 K serves as a direct proxy for near-surface saturation. These surface-level metrics are critical for distinguishing fog from low stratus clouds, as only fog requires saturation at the 2-m level while low stratus clouds maintain a non-saturated surface layer.

Satellite inputs comprise FY-4A brightness temperatures (BT) from shortwave (3.72 μm) and thermal (10.8 μm) infrared channels (Yang et al., 2017), alongside the solar zenith angle (SOZ) (Bessho et al., 2016). The 10.8 μm channel captures the emission temperature of optically thick layers and effectively distinguishes ground-contacting fog from colder low stratus when referenced against ERA5 surface temperatures. During daytime, the 3.72 μm channel differentiates aerosol types via particle-size-dependent scattering, where fog droplets (2–10 μm) induce efficient Mie scattering and high BTs (Pincus, 2004) while submicron haze and absorptive fog-haze exhibit suppressed reflectance. This spectral behavior is encapsulated in the Brightness Temperature Difference (BTD, $BT_{3.72} - BT_{10.8}$) (Hunt, 1972), which provides a distinct negative emissivity signal for nighttime fog but becomes physically complex during the day. While coarse aerosols produce positive daytime BTDs (Ackerman, 1997), fog often presents ambiguous near-zero values as strong solar reflectance counteracts the thermal emissivity deficit (Ellrod, 1995; Rosenfeld and Lensky, 1998). To resolve these diurnal non-linearities, the model integrates SOZ and spectral data via deep learning to robustly classify fog, haze, and fog-haze across varying illumination regimes.

Text S2 Network Architecture and Attention Mechanism

As illustrated in Figure S2, the model processes input tensors with shapes of (6, 64, 64) for the Sat configuration and (7, 64, 64) for the Sat-PM configuration. The feature extraction module learns hierarchical features through three sequential convolutional blocks. Each block contains a 3×3 convolutional layer (Conv), batch normalization (BatchNorm) (Ioffe and Szegedy, 2015), a rectified linear unit (ReLU) activation (Krizhevsky et al., 2012), and the dual-branch attention module for adaptive refinement. A max-pooling layer concludes each block, progressively reducing the spatial resolution from 64×64 to 8×8 while expanding the channel depth (32→64→128). Following extraction, the classification module flattens the final 3D feature map (128 × 8 × 8) into a 1D vector of 8,192 features. This vector is processed by two fully connected (FC) layers, which reduce the dimensionality from 8,192 to 512, and finally to two final output neurons, corresponding to the prediction scores for class 0 (clear) and class 1 (low-visibility). Finally, a softmax activation function converts these scores into class probabilities, and the class with the higher probability is selected as the model final prediction.

The variable-wise (channel) attention computes an importance weight vector for input parameters using global average pooling (Hu et al., 2018):

$$M_c(F) = \sigma(W_1(\delta(W_0 \cdot \mathrm{AvgPool}(F))))$$

where σ is the Softmax function, F is the input feature map, and $W_0 \in \mathbb{R}^{(C/L)\times C}$ and $W_1 \in \mathbb{R}^{C\times(C/L)}$ represent the weights of two convolutional layers that form a bottleneck structure, with C being the channel count and L the compression ratio. Simultaneously, the spatial attention module employs parallel max-pooling and average-pooling operations (Woo et al., 2018), followed by a 7×7 convolutional kernel to enable the network to adaptively focus on the most informative spatial areas, formulated as:

$$M_s(F) = \sigma(\mathrm{Conv}_{7\times 7}([\mathrm{MaxPool}(F),\mathrm{AvgPool}(F)]))$$

where σ is the softmax function, $\mathrm{Conv}_{7\times 7}$ represents the 7×7 convolution operation, and the inputs from average and max pooling are concatenated.

Text S3 Hybrid Loss Function

The Hybrid Loss function is defined as the weighted sum of Balanced Focal Loss and Dice Loss (Lin et al., 2017; Milletari et al., 2016). The two are combined with equal weighting ($\alpha=0.5$) to ensure the model learns both accurately and precisely:

$$\text{Hybrid Loss} = \alpha \cdot \text{BalancedFocalLoss} + (1 - \alpha) \cdot \text{DiceLoss}$$

The Focal Loss addresses class imbalance by down-weighting easy negatives:

$$\text{BalancedFocalLoss} = -\beta y_{\text{true}}(1 - p)^{\gamma} \log{(p)} - (1 - \beta)(1 - y_{\text{true}})p^{\gamma} \log{(1 - p)}$$

where p is the probability for the ground-truth class, the focusing parameter γ is set to 3.0, and the balancing coefficient β is set to 0.55. The Dice Loss optimizes the intersection over union:

$$\text{Dice Loss} = 1 - \frac{2\sum(y_{\text{pred}} \cdot y_{\text{true}}) + \epsilon}{\sum y_{\text{pred}} + \sum y_{\text{true}} + \epsilon}$$

where y_{true} is the ground truth, and y_{pred} is the model prediction.

Text S4 Training Configuration

The model was implemented in PyTorch. The model was trained using the AdamW optimizer with an initial learning rate of 0.0004 (Loshchilov and Hutter, 2019) and a weight decay of 0.0005. The learning rate was dynamically managed by a scheduler that halved the rate when the validation loss plateaued for five consecutive epochs. An early stopping mechanism with a patience of ten epochs was implemented. To improve generalization, a progressive dropout strategy was applied to the fully connected layers: a higher rate (0.6) was used on the first layer to regularize high-dimensional features, while a lower rate (0.4) was applied to the second layer to preserve critical classification information (SrivastavaNitish et al., 2014).

References

Ackerman, S. A.: Remote sensing aerosols using satellite infrared observations, J. Geophys. Res.: Atmos., 102, 17069–17079, https://doi.org/10.1029/96JD03066, 1997.

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama, H., Uesawa, D., Yokota, H., and Yoshida, R.: An introduction to himawari-8/9— japan’s new-generation geostationary meteorological satellites, J. Meteorol. Soc. Jpn., II, 94, 151–183, https://doi.org/10.2151/jmsj.2016-009, 2016.

Ellrod, G. P.: Advances in the detection and analysis of fog at night using GOES multispectral infrared imagery, Weather Forecasting, 10, 606–619, https://doi.org/10.1175/1520-0434(1995)010%3C0606:AITDAA%3E2.0.CO;2, 1995.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P. de, Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, https://doi.org/10.1002/qj.3803, 2020.

Hu, J., Shen, L., and Sun, G.: Squeeze-and-excitation networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7132–7141, 2018.

Hunt, G. E.: Radiative properties of terrestrial clouds at visible and infra-red thermal window wavelengths, https://doi.org/10.1002/qj.49709942013, 1972.

Ioffe, S. and Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on Machine Learning, International Conference on Machine Learning, shortConferenceName: ICML, 448–456, 2015.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P.: Focal loss for dense object detection, Proceedings of the IEEE International Conference on Computer Vision, 2980–2988, 2017.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, https://doi.org/10.48550/arXiv.1711.05101, 4 January 2019.

Milletari, F., Navab, N., and Ahmadi, S.-A.: V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), 2016 Fourth International Conference on 3D Vision (3DV), 565–571, https://doi.org/10.1109/3DV.2016.79, 2016.

Pincus, R.: A first course on atmospheric radiation, EOS Trans. Trans. Am. Geophys. Union, 85, 341–341, https://doi.org/10.1029/2004EO360007, 2004.

Rosenfeld, D. and Lensky, I. M.: Satellite-based insights into precipitation formation processes in continental and maritime convective clouds, 1998.

SrivastavaNitish, HintonGeoffrey, KrizhevskyAlex, SutskeverIlya, and SalakhutdinovRuslan: Dropout, J. Mach. Learn. Res., https://doi.org/10.5555/2627435.2670313, 2014.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: CBAM: convolutional block attention module, Proceedings of the European Conference on Computer Vision (ECCV), 3–19, 2018.

Yang, J., Zhang, Z., Wei, C., Lu, F., and Guo, Q.: Introducing the new generation of chinese geostationary weather satellites, fengyun-4, https://doi.org/10.1175/BAMS-D-16-0065.1, 2017.