

Enhancing ozone nowcasting over East Asia using a data-to-data translation approach with observations from a geostationary environment monitoring spectrometer



Seong-Taek Oh^a, Jeong-Eun Park^a, Jaehoon Jeong^b, Sungwook Hong^{a,c,*}

^a Department of Environment, Energy, and Geoinformatics, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, Republic of Korea

^b National Institute of Environmental Research, Ministry of Environment, 42 Hwangyong-ro, Seogu, Incheon, 22689, Republic of Korea

^c DeepThoTh Co., Ltd., 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, Republic of Korea

ARTICLE INFO

Keywords:

Ozone
GEMS
Nowcasting
Data-to-Data transition
Satellite remote sensing

ABSTRACT

The degradation of air quality caused by excessive anthropogenic ozone (O_3) concentrations negatively affects ecosystems and the atmosphere. To monitor this issue, the Geostationary Environment Monitoring Spectrometer (GEMS) aboard the Geostationary Korea Multi-Purpose Satellite (GK)-2B provides ozone products over Asia-centered circular regions based on the ozone algorithm derived from representative polar-orbit atmospheric environmental satellites. In this study, we developed a few-hour GEMS O_3 nowcasting model using data-to-data (D2D) translation with a conditional generative adversarial network. This model is based on hourly GEMS O_3 time-series products and can be used to predict ozone concentrations. The D2D model underwent training and testing employing paired input and output datasets of GEMS O_3 , with data collected from March 22, 2020, to June 21, 2020, and from March 22, 2021, to June 18, 2021, respectively. The resulting D2D ozone nowcasting model was used to determine ozone concentrations in time zones where GEMS O_3 products were unavailable. Test results of the D2D model demonstrated excellent statistical scores, including a bias of 2.162 Dobson Units (DU, where 1 DU corresponds to 2.687×10^{16} molecules/cm²), root-mean-square error (RMSE) of 5.606 DU, a correlation coefficient (CC) of 0.994 for 1-h prediction, a bias of 1.421 DU, RMSE of 5.903 DU, and CC of 0.992 for 2-h prediction, and a bias of 1.169 DU, RMSE of 6.797 DU, and CC of 0.988 for 3-h prediction. Despite the dataset pairing and number limitations, the D2D prediction model accurately forecasted GEMS O_3 within 3 h in East Asia.

1. Introduction

Over the past few decades, human activities have significantly changed the composition of the atmosphere, leading to increased emissions of trace gases including CO₂, CO, NO_x, CH₄, and hydrocarbons (ESA, 2021). These emissions, particularly those of NO_x and hydrocarbons, elevate surface ozone levels, resulting in a degradation of global air quality. Stratospheric ozone is crucial in shielding the biosphere from harmful solar ultraviolet (UV) radiation, but excessive tropospheric O₃ levels caused by chemical reactions involving NO_x, CO, CH₄, and other hydrocarbons can be toxic to ecosystems, animals, and humans (ESA, 2021). Tropical forests and savannas are significant sources of

non-methane hydrocarbon (NMHC) emissions. However, their determination remains challenging (Guenther et al., 1995, 2006). Natural sources account for over 50% of NMHC emissions, with lightning being a major contributor (Ott et al., 2007). NO_x emissions from both anthropogenic and natural sources (Crutzen and Andreae, 1990; Thompson et al., 1996; Ziemke et al., 2009), including biomass and savanna burning, lead to emissions of ozone precursors. Biomass-burning in South America and Africa lead to notable seasonal and interannual variations in CO, NO_x, CH₄, and hydrocarbon emissions. During extreme El Niño-Southern Oscillation events in 1997 and 2006 (Chandra et al., 2009; Siegert et al., 2001), the burning of more extensive tropical rainforests in Brazil and Indonesia led to the spread of ozone plumes as far as India (Thompson et al., 2001).

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.

* Corresponding author. Department of Environment, Energy, and Geoinformatics, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul, 05006, Republic of Korea.

E-mail addresses: gavin5@sju.ac.kr (S.-T. Oh), wjd196@sju.ac.kr (J.-E. Park), jaehoon80@korea.kr (J. Jeong), sestiya@gmail.com, [\(S. Hong\).](mailto:sestiya@sejong.ac.kr)

<https://doi.org/10.1016/j.apr.2024.102054>

Received 9 September 2023; Received in revised form 15 January 2024; Accepted 15 January 2024

Available online 20 January 2024

1309-1042/© 2024 Turkish National Committee for Air Pollution Research and Control. Production and hosting by Elsevier B.V. All rights reserved.

Abbreviations	
ANN	Artificial Neural Networks
AOD	Aerosol Optical Depth
CC	Correlation Coefficient
CGAN	Conditional Generative Adversarial Networks
CNN	Convolutional Neural Network
CMAQ	Community Multi-scale Air Quality
DL	Deep Learning
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
D2D	Data-to-Data
FW	Full West
FC	Full Central
GEMS	Geostationary Environment Monitoring Spectrometer
GK	Geostationary Korea Multi-Purpose Satellite
GOME	Global Ozone Monitoring Experiment
L	Level
MBE	Mean Bias Error
ML	Machine Learning
NMHC	Non-Methane Hydrocarbon
NN	Neural Network
OMI	Ozone Monitoring Instrument
RF	Random Forest
RMSE	Root Mean Square Error
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
TROPOMI	Tropospheric Monitoring Instrument
TOMS	Total Ozone Mapping Spectrometer
UTC	Universal Time Coordinated
UV	Ultraviolet
UVI	Ultraviolet index
VIS	Visible
WOUDC	World Ozone and Ultraviolet Radiation Data Center
WRF	and Weather Research and Forecasting

Ozone is a greenhouse gas that is crucial for generating OH, the primary oxidizing agent responsible for reactions with all trace gases in the troposphere, through UV radiation and high-humidity photolysis (Forster et al., 1997; Lacis et al., 1990). Ozone can undergo long-range transport, posing potential threats to regions that are distant from emissions (ESA, 2021; Thompson et al., 2001).

During the daytime, ozone photochemically produces atmospheric NO₂ by converting NO into NO₂ (and vice versa) in the presence of sunlight (Van Geffen et al., 2022), as follows (Correa, 1993):



where $h\nu$ is photon energy of solar ultraviolet radiation and O (¹D) is excited singlet oxygen atom.

Satellite observations are indispensable for investigating large-scale temporal and spatial variabilities in ozone concentrations because of the restricted spatial extent of ground-based ozonesonde locations (Thompson et al., 2003a, 2003b). Several satellites with environmental sensors, including the Global Ozone Monitoring Experiment (GOME) (Munro et al., 1998), the Ozone Monitoring Instrument (OMI) (Liu et al., 2010), Stratospheric Aerosol and Gas Experiment, and Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (Fishman et al., 2008), and the Tropospheric Monitoring Instrument (TROPOMI) (Veeckind et al., 2012) have played a pivotal role in monitoring tropospheric ozone profiles. Notably, compared to ground-based ozonesonde observations, satellite-based tropospheric ozone observations have considerable spatiotemporal variability owing to the presence of the stratospheric ozone layer.

The Geo-Kompsat (GK-2B) satellite was launched on February 18, 2020, and became the first geostationary satellite explicitly designed for atmospheric environmental monitoring. The GK-2B satellite carries two sensors, the Geostationary Environment Monitoring Spectrometer (GEMS) and the Geostationary Ocean Color Imager, for environmental and oceanic purposes, respectively. The GEMS sensor, a collaborative development between the Korea Aerospace Research Institute and Ball Aerospace and Technology Corp., is a hyperspectral instrument with the capacity to acquire 1000 bands spanning the UV and visible (VIS) spectral domains, covering a wavelength range of 300–500 nm. The GEMS level(L)-2 baseline products include NO₂, SO₂, O₃, HCHO, CHO-CHO, aerosols, clouds, ultraviolet index (UVI), and surface reflectivity. The GEMS L2 products were developed and tested using L1b data from OMI, TROPOMI, airborne Geostationary Trace gas and aerosol sensors,

and radiance spectral data.

Several machine learning (ML) and deep learning (DL) techniques have successfully surpassed traditional limitations in handling big data. Artificial neural networks (ANN) are valuable tools for uncovering the hidden characteristics of data in multiple research fields (Lee et al., 2017), including data-driven research (Chen and Lin, 2014), and satellite remote sensing (Tao et al., 2018). Recently, DL techniques, including convolutional neural networks (Chen et al., 2019; Xu and Ge, 2020), convolutional long short-term memory (Shi et al., 2015), and conditional generative adversarial networks (CGAN) (Denton et al., 2015; Kim et al., 2020b; Kim et al., 2017; Mao et al., 2017; Odet al., 2017; Radford et al., 2015; Santos et al., 2017; Zhu et al., 2017) have been applied to satellite remote sensing research (Kim, et al., 2019a; Kim and Hong, 2019b; Li and Wand., 2016; Park et al., 2020). For example, a geographically and temporally weighted generalized regression neural network model was employed to assess spatiotemporal variations in the relationship between NO₂ and its influencing factors across the urban agglomeration of Wuhan, China (Li et al., 2020), while a deep Convolutional Neural Network (CNN) exhibited superior results in estimating TROPOMI NO₂ over Texas compared with multiple linear regression methods, support vector machines, and random forest (RF) techniques (Ghahremanloo et al., 2021). A neural network (NN) model trained using more than 360,000 radiative transfer simulations was employed to estimate NO₂ concentrations using a four-month dataset of TROPOMI observations over China (Li et al., 2022) and a CNN approach using simulated tropospheric column density of NO₂ from the Community Multiscale Air Quality model was employed to estimate TROPOMI NO₂ levels across the United States (Lops et al., 2023). Finally, a coupled ML model with a recurrent NN and Long Short-Term Memory using TROPOMI NO₂ products, ground monitoring data, and NASA weather data was presented to predict emissions of NO₂, SO₂, CO, and other pollutants originating from a single power plant in rural Alabama (Alnaim et al., 2022).

Multiple ML and DL methodologies have been employed in conjunction with Korean geostationary satellites to investigate atmospheric conditions within the Far East Asia region. For example, the RF technique was employed to estimate PM10 and PM2.5 levels across South Korea, utilizing satellite-derived aerosol optical depth (AOD) and numerical model-derived meteorological and emission variables (Park et al., 2019) and an ANN approach was used to recalibrate inaccurate pixels within the GEMS dataset (Lee et al., 2023). However, in Asia, ML- and DL-based studies of O₃ estimation or prediction using GEMS data have rarely been conducted.

The National Institute of Environmental Research, a research agency

operated by the South Korean government, is currently forecasting O₃ in South Korea using meteorological models, including Weather Research and Forecasting (WRF) and air quality models, such as Community Multi-scale Air Quality (CMAQ). These forecasts were performed four times daily (5 a.m., 11 a.m., 5 p.m., and 11 p.m. Korean Standard Time), starting on April 15, 2015. Thus, the primary objective of this study was to provide information for GEMS O₃ prediction during the temporal gap between WRF and CMAQ-based O₃ predictions. Consequently, this study employed a CGAN model implemented using Pix2Pix (Lee et al., 2017) and GEMS L2 O₃ products. The proposed O₃ nowcast model, as presented in this study, underwent comprehensive training and testing using pairs of GEMS L2 O₃ data at specific times with a stack of multiple GEMS O₃ data at different times. The resulting CGAN-predicted GEMS O₃ products compensate for the lack of GEMS O₃ monitoring data and provide supplementary information for atmospheric environmental forecasters and analysts.

2. Data

This study used GEMS O₃ products as input and output datasets, acquired by the GEMS sensor located 128°E longitude above the Earth's equator. The GEMS sensor has a narrow spectral response function of 0.6 nm full-width and observes the Asia region, spanning from 5°S to 45°N and 75°E to 145°E, with a temporal resolution of 30 min. The spatial resolutions of the GEMS sensor are 3.5 km × 8 km for aerosols and 7 km × 8 km for trace gases, respectively, when observing Seoul, Republic of Korea (Kim et al., 2020a; Lasnik et al., 2014).

2.1. Study area

The GK-2B/GEMS satellite has four distinct observational modes, each characterized by a varying spatial coverage. In this investigation, we employed the full west (FW) and full central (FC) modes, which comprehensively encompass East Asia and incorporate the entire Korean Peninsula within their observation areas. Fig. 1 depicts the GEMS observational modes available during daytime for each month. The red box highlights the daytime datasets that are accessible, spanning six consecutive temporal data points, serving the purpose of a 3-h nowcast in this study within the GEMS FW observation mode throughout the 12 months.

This study focused on the Korean Peninsula, along with parts of Japan and China, within the latitudinal range of 17°N to 43°N and the longitudinal range of 110°E to 133°E, using the GEMS FW mode. Fig. 2 illustrates the study area on April 5, 2022, at 05:45 UTC. The two red points in the study area indicate ground-based ozone datasets in Seoul and Hong Kong.



Fig. 2. Study area. The red spots indicate the location of the ozonesonde deployment in the WOUDC.

The proposed D2D nowcast model employed GEMS O₃ data as the input dataset, which were resampled into a numerical array format of 948 × 260 pixels from the original GEMS L2 O₃ data with an array size of 10,400 × 12,000 over Asia. The GEMS O₃ data in Dobson units (DU) were normalized from 0 to 1.

2.2. GEMS O₃ data

The GEMS O₃ products, expressed in DU (Hong et al., 2017; Park et al., 2018), provide high-quality O₃ concentration data for East Asia derived from the differential optical absorption spectroscopy (DOAS) technique (Platt and Stutz, 2008) using simple and multiple spectral optical ozone depths fitted within the 432–450 nm range (Platt and

Schedule of the GEMS observation											
Area no.	1	2	3	4	5	6	7	8	9	10	11
UTC	23:00	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00
Jan			HE	HK	FC	FW	FW	FW			
Feb			HE	HK	FC	FW	FW	FW	FW		
Mar		HE	HK	FC	FC	FW	FW	FW	FW		
Apr	HE	HK	FC	FC	FC	FW	FW	FW	FW	FW	
May	HE	HK	FC	FC	FW	FW	FW	FW	FW	FW	
Jun	HE	HK	FC	FC	FW	FW	FW	FW	FW	FW	
Jul	HE	HK	FC	FC	FW	FW	FW	FW	FW	FW	
Agu	HE	HK	FC	FC	FW	FW	FW	FW	FW	FW	
Sep	HE	HK	FC	FC	FW	FW	FW	FW	FW	FW	
Oct		HE	HK	FC	FC	FW	FW	FW	FW		
Nov			HE	HK	FC	FW	FW	FW			
Dec			HE	HK	FC	FW	FW	FW			

Fig. 1. Schedule of the GEMS observation depending on observational modes for daytime of each month (NESC, 2021). The acronyms HE, HK, FC, and FW denote the four distinct observational modes of GEMS: 'Half East,' 'Half Korea,' 'Full Central,' and 'Full West' modes, respectively.

Stutz, 2008). The DOAS technique has been extensively employed to assess total ozone content through ground-based and satellite-based observations, including GOME, OMI, and SCIAMACHY sensors (Hong et al., 2017). Nevertheless, the accuracy of GEMS O₃ products, including other satellite sensors for atmospheric environmental monitoring, is affected by various variables, including solar zenith angle, aerosol optical depth (AOD), peak altitude, single-scattering albedo, aerosol composition, and surface reflectance (Hong et al., 2017). Table 1 summarizes the characteristics of GEMS O₃ products (Kim et al., 2020a).

2.3. Ozonesonde datasets for validation

The World Ozone and Ultraviolet Radiation Data Center (WOUDC) (McElroy and Hare, 1999) is a globally recognized data center that offers a wide range of ozone data, including ozonesonde data, in partnership with over 500 institutions. The ozonesonde dataset from the WOUDC provides altitude-resolved ozone concentrations and total column ozone (TCO) based on the vertical ozone distribution in the atmosphere. This study selected WOUDC ozonesonde data as validation data for the data-to-data (D2D) translation-based ozone nowcasting model, focusing on the Seoul and Hong Kong regions (Fig. 2).

2.4. TROPOMI total column ozone datasets for validation

TROPOMI TOC is a secondary ozone product of the Sentinel-5 Precursor satellite, providing high-precision ozone measurements with a spatial resolution of 7 km × 3.5 km and daily global coverage (ESA, 2021). The ozone retrieval algorithm used in the TROPOMI O₃ products closely resembled that of the GEMS. Both algorithms are based on the ozone mapping spectrometer algorithm. Consequently, a strong correlation between the GEMS O₃ products and TROPOMI O₃ products is expected (Baek et al., 2022). Moreover, because of the characteristics of TROPOMI O₃ products, these products exhibit a strong correlation with ground-based ozone observation data (Garane et al., 2019). This correlation can serve as indirect validation data, complementing the insufficient ground validation data in this study.

3. Methods

3.1. Data-to-data translation using Pix2Pix

The D2D nowcast model incorporates Pix2Pix (Lee et al., 2017) with a CGAN structure (Isola et al., 2017) to obtain GEMS ozone data from past ozone data as input datasets. The D2D model utilizes preprocessed datasets transformed into numerical arrays with values ranging from 0 to 1, derived from paired input ozone concentration data (measured in DU). The D2D model generates nowcast ozone datasets through a post-processing step to convert the output datasets (numerical arrays ranging from 0 to 1) back into the original ozone concentration data in the DU. Therefore, the D2D model facilitated data-to-data translation between pairs of ozone data for consecutive time intervals.

The main characteristics of Pix2Pix in the D2D model is the use of paired image data during training. The Pix2Pix architecture uses a U-net

Table 1
Characteristics of GEMS O₃ product.

O ₃ product	Features
Importance	O ₃ /Aerosol precursor
Minimum (DU)	200
Maximum (DU)	500
Nominal (DU)	300
Accuracy	3 %
Window	317.5 nm, 331.2 nm, 331.2 nm, 340 nm, 380 nm
Spatial Resolution at Seoul;	7 km × 8 km
Solar Zenith Angle (°)	70
Algorithm	TOMS

encoder-decoder network (Isola et al., 2017) for the generator network (*G*) and a PatchGAN (Isola et al., 2017) discriminator network (*D*). The U-net architecture in the generator captures and preserves fine-grained details of input and output data. The PatchGAN discriminator categorized the authenticity of individual patches within an output as real or fake. The utilization of the PatchGAN discriminator enhances the quality of the generated data.

The generator was trained to produce outputs that closely resemble the actual inputs through a training approach involving both adversarial loss (*L_A*) and reconstruction loss (*L₁*). This combination of losses ensured that the generated output was closely aligned with the ground-truth data. Thus, the loss function (*L_{D2D}*) of the D2D model is expressed as follows (Isola et al., 2017):

$$L_{D2D} = \min_{G} \max_{D} \{L_A\} + \lambda \bullet L_1 \quad (3)$$

where $\min_{G} \max_{D} \{\cdot\}$ describes the minimum-maximum function between *G* and *D* models. λ denotes the tradeoff value between *L_A* and *L₁*. Importantly, λ was set to be 1 in this study.

The *L_A* consists of the cross-entropies (*E*) of *G* and *D* as follows [55, 56]:

$$L_A = E[\log(D(Y_P, Y_O))] + E[\log(1 - D(Y_O, G(X_O, Y_O)))] \quad (4)$$

where *X_O* and *Y_O* denote the pairs of observed input and output datasets, respectively. Subscript *P* denotes the predicted results of the D2D model. This study assigned the past time series of the GEMS O₃ dataset as *X_O* at a specific time, and a stack of three time series of GEMS O₃ datasets as *Y_O* for O₃ prediction for 1, 2, and 3 h O₃ predictions. During model training, the first cross-entropy (the first term in Eq. (4)) was iteratively minimized in the *G*. In contrast, the second cross-entropy (the second term in Eq. (4)) was iteratively maximized to determine the probability of discriminating between the prepared and predicted O₃ datasets. The logarithmic functions within the cross-entropies counteracted the insufficient gradient reduction in the initial stages of the D2D model training process (Nguyen et al., 2017).

The *L₁* loss (Michelsanti and Tan, 2017), used to minimize the distance between the predicted output dataset (*Y_P*) and given output dataset (*Y_O*) was calculated as follows (Lee et al., 2017):

$$L_1 = E(\|Y_O - Y_P\|_1) \quad (5)$$

In this study, *Y_P* and *Y_O* denote the D2D model prediction and given GEMS O₃ datasets, respectively.

Fig. 3 provides a more detailed representation of the CGAN structure in the D2D model and the iterative process involving the generator (*G*), discriminator (*D*), and loss functions (*L₁* and *L_A*). The input data are denoted as a past pair (*X_O*) and a future pair (*Y_O*). The D2D model generates the output data *Y_P*.

3.2. Learning scheme for nowcasting O₃

The D2D-nowcast ozone model was trained and tested using pairs of input datasets (*X_O, Y_O*), which were presented as a stack of time series of GEMS O₃ datasets. This study used a past time series of GEMS O₃ products as input data, which consisted of a stack of *t* – 2 to *t* hour GEMS O₃ and a stack of *t* + 1 to *t* + 3 h GEMS O₃ with an array size of 1024 × 1024 × 3. The mathematical expressions for these datasets are as follows:

$$X_O = \{GEMS(t-2), GEMS(t-1), GEMS(t)\} \quad (6)$$

$$Y_O = \{GEMS(t+1), GEMS(t+2), GEMS(t+3)\} \quad (7)$$

where GMES denotes the GEMS O₃ products. The subscript *O* denotes the prepared GEMS O₃ datasets at a certain time (*t*) from *t* – 2 to *t* + 3, with 1 h intervals.

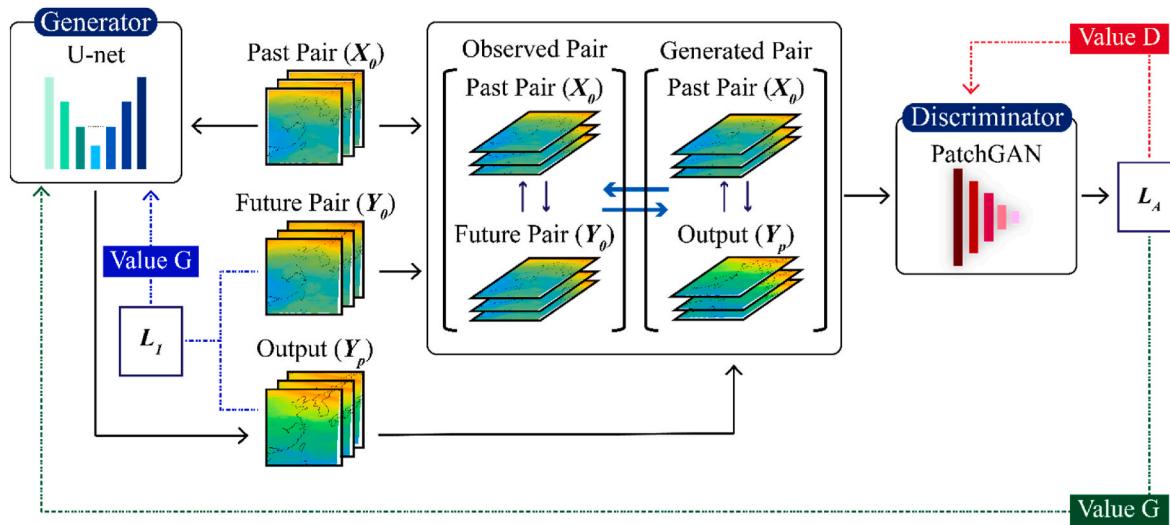


Fig. 3. The CGAN and training process in the D2D model. Here, G , D , L_A , and L_1 represent the generator, discriminator, and loss functions in the CGAN, respectively. X_O and Y_O represent the past and future GEMS datasets, respectively. Y_p indicates a prediction of GEMS O₃ from 1 to 3 h, trained to closely resemble Y_O .

3.3. Converted datasets for D2D translation

In the preprocessing stage, the original GEMS O₃ datasets, initially measured in DU, underwent min–max normalization within 0–1 to handle differences between each O₃ dataset, as follows (Park et al., 2020):

$$\dot{X}_O = \frac{X_O - X_{O,min}}{X_{O,max} - X_{O,min}} \quad (8)$$

$$\dot{Y}_O = \frac{Y_O - Y_{O,min}}{Y_{O,max} - Y_{O,min}} \quad (9)$$

where \dot{X}_O and \dot{Y}_O are the normalized GEMS datasets, respectively. $X_{O,min}$, $X_{O,max}$, $Y_{O,min}$, and $Y_{O,max}$ represent the minimum and maximum O₃

values in the GEMS datasets for X_O and Y_O , respectively.

The constructed D2D-nowcasting model generates a stack of predicted GEMS O₃ data (Y_p) within 0–1. For post-processing, the Y_p was denormalized into the numerical array dataset (Y_p) of the predicted O₃ levels at $t + 1$, $t + 2$, and $t + 3$ h.

Fig. 4 is an example of the paired datasets used to develop the D2D ozone nowcasting model. Fig. 4(a) shows a sequence of three consecutive time series for the GEMS O₃ products utilized as both input ($t - 2$ h, $t - 1$ h, and t) and output ($t + 1$ h, $t + 2$ h, and $t + 3$ h) data in a $1024 \times 1024 \times 3$ array. The reference time t , was June 29, 2021, at 02:45 UTC. Fig. 4(b) provides an overview of the temporal evolution of ozone, emphasizing temporal differences.

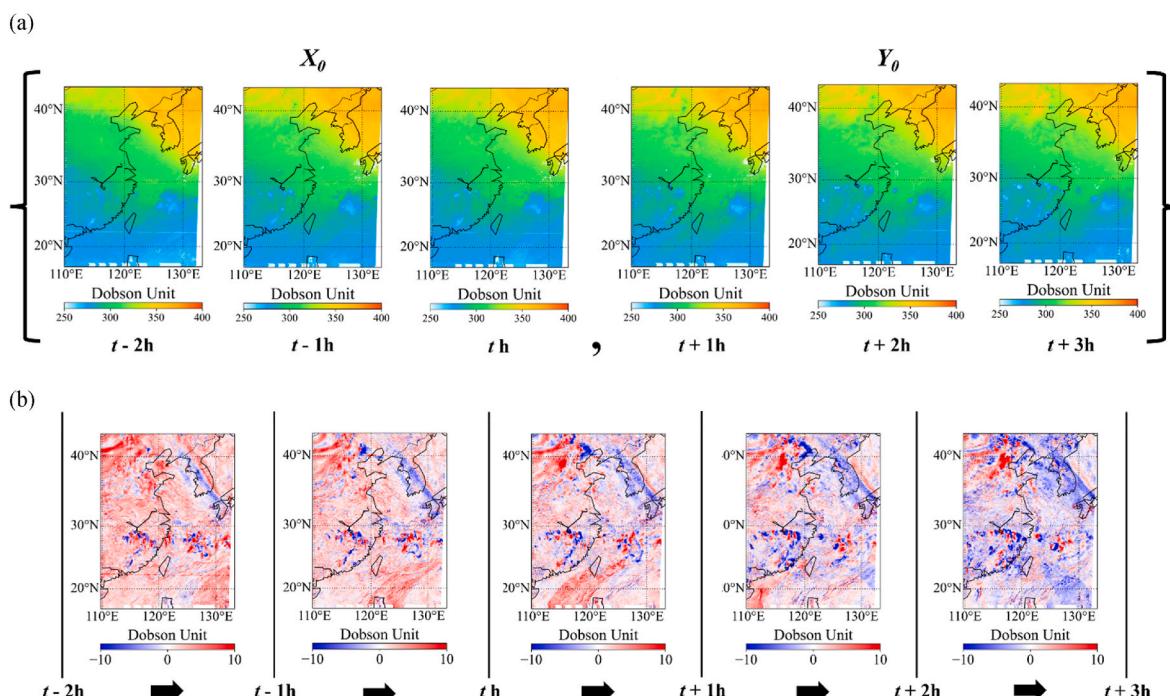


Fig. 4. (a) Instances of GEMS O₃ product pairs at time instances $t - 2$ h, $t - 1$ h, and t and $t + 1$ h, $t + 2$ h, and $t + 3$ h, while (b) showcases the corresponding temporal differences. The reference time, denoted as t , corresponds to June 29, 2021, at 02:45 UTC.

3.4. D2D model training process

In the training phase, the D2D model was repeatedly trained to nowcast GEMS O₃ data at $t + 1$, $t + 2$, and $t + 3$ h. Simultaneously, the nowcast GEMS O₃ data were iteratively compared and discriminated against other pairs of actual GEMS O₃ products at $t + 1$, $t + 2$, and $t + 3$ h. During iteration steps, the D2D nowcasting model was chosen as the final constructed D2D model, demonstrating superior correlation coefficient (CC) and root-mean-square-error (RMSE) values when comparing the nowcasted and observed GEMS O₃ products. Fig. 5 shows the configuration of the D2D nowcasting model used in this study. The arrows denote the stack of paired input data, the training process with the generator and discriminator, and post-processing.

This study used 598 pairs of GEMS O₃ data with hourly intervals between 00:45 and 07:45 UTC from March 1, 2021, to October 30, 2021, for D2D model training. The D2D model was tested using a dataset consisting of 83 pairs, all captured within an identical interval spanning March 1, 2022, to April 26, 2022. Additionally, this study employed a dataset comprising 515 pairs of GEMS O₃ products for nowcasting GEMS O₃ during the timeframe not covered by observations, spanning April 27, 2022, to October 30, 2022.

The D2D model was implemented using TensorFlow in conjunction with Python 3.6.0, a prevalent open-source machine learning framework. The implementation was conducted on a Linux Ubuntu 18.04.5 platform, operated across two NVIDIA Titan-RTX GPUs and an Intel Xeon CPU. The D2D model presented in this study was trained for approximately 6 h within the hardware and software as described.

3.5. Statistical validation

This study validated the D2D model with GEMS O₃ products using the following statistical scores: CC, bias, RMSE, relative mean bias error (rMBE), and relative RMSE (rRMSE) (Engerer and Mills, 2015):

$$CC = \frac{\sum_{i=1}^n (C_{D,i} - \bar{C}_D) \sum_{i=1}^n (C_{G,i} - \bar{C}_G)}{\sqrt{\sum_{i=1}^n (C_{D,i} - \bar{C}_D)^2} \sqrt{\sum_{i=1}^n (C_{G,i} - \bar{C}_G)^2}} \quad (10)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (C_{G,i} - C_{D,i}) \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_{G,i} - C_{D,i})^2} \quad (12)$$

$$rMBE = \frac{1}{n} \sum_{i=1}^n (C_{G,i} - C_{D,i}) / \bar{C}_G \quad (13)$$

$$rRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_{G,i} - C_{D,i})^2} / \bar{C}_G \quad (14)$$

where n denotes the total number of pixels in the GEMS O₃ dataset. The index of the pixels is denoted by i , and ranges from 1 to n . The ozone concentration is denoted as C . The subscripts D and G indicate the D2D and GEMS models, respectively. $C_{G,i}$ represents the ozone concentration of the i -th pixel in the GEMS data, and $C_{D,i}$ represents the C of the i -th pixel in the D2D-predicted GEMS O₃ data. The average C values of the GEMS- and D2D-predicted O₃ data are represented as \bar{C}_G and \bar{C}_D , respectively.

This study presents the rMBE and rRMSE metrics (Engerer and Mills, 2015) to assess the performance of the D2D model. These metrics categorize the performance of the model, as summarized in Table 2.

4. Results

4.1. D2D model

Fig. 6 shows the loss functions of the generator (G), discriminator (D), and L_1 as functions of iterations during D2D model training. The green curve depicting the loss value of G increased rapidly at approximately 8000 iterations, whereas it subsequently decreased swiftly and tended to stabilize minimally after approximately 40,000 iterations. In contrast, the blue curve illustrating the loss value of D exhibited a behavior opposite to that of G . The red curve representing L_1 loss consistently decreased to zero. These results indicate that the D2D model effectively captured time-series variation in GEMS O₃, making it suitable for nowcasting GEMS O₃ observations.

Fig. 7 depicts the CC and RMSE changes between the GEMS-observed and D2D-nowcast O₃ products during the iteration of the model training phase. This figure shows three distinct graphs for $t + 1$, $t + 2$, and $t + 3$ h, respectively. In this study, the most effective D2D model was determined by considering the optimal CC and RMSE values. Significant variation was observed at approximately the 8,000th training iteration, (Fig. 6). However, after reaching approximately the 10,000th iteration, the CC value, which represents the correlation between the GEMS O₃

Table 2
Four performance categories.

Performance	rMBE	rRMSE
Excellent	<2 %	<5 %
Good	2–5 %	5–10 %
Average	5–10 %	10–15 %
Poor	≥10 %	≥15 %

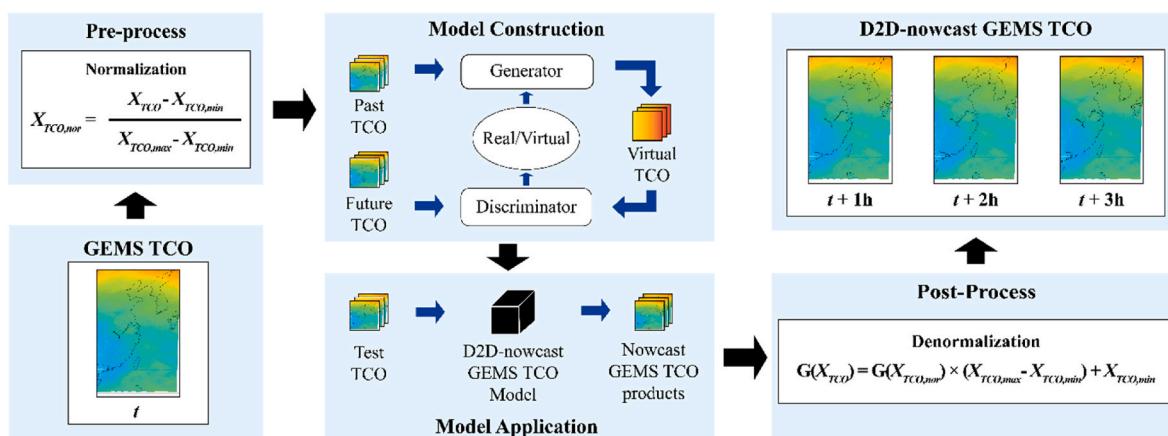


Fig. 5. D2D model structure. TCO signifies total column ozone. X_{TCO} denotes the data pixels of the GEMS TCO dataset.

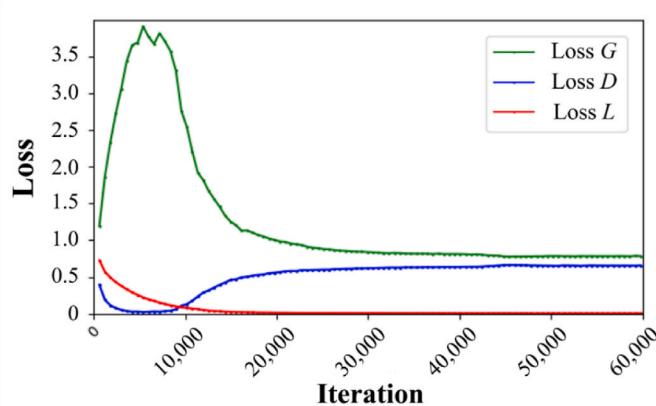


Fig. 6. Loss functions of the generator, discriminator, and L_1 as a function of iteration during the D2D model training.

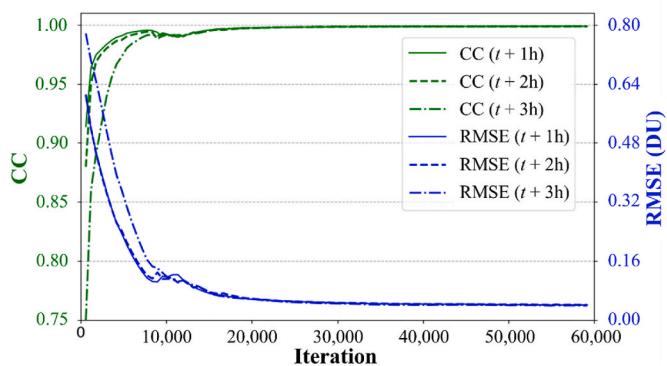


Fig. 7. Variations of CC and RMSE of GEMS and D2D ozone products for the $t + 1$ h, $t + 2$ h, and $t + 3$ h prediction models, respectively.

observation and the D2D prediction, increased steadily. Simultaneously, the RMSE value, which indicates the difference between GEMS O₃ observations and D2D predictions, decreased consistently. These results confirm that the D2D model accurately learns the GEMS O₃ observations for predictions at $t + 1$, $t + 2$, and $t + 3$ h.

This study identified an optimal D2D model for nowcasting GEMS O₃ by training for 57,408 iterations. For example, on April 10, 2022, at 04:45 UTC (equivalent to 13:45 Korean Standard Time), the D2D O₃ nowcast model exhibited a remarkable CC value of 0.991 and an RMSE value of 11.252 DU. Table 3 summarizes the statistical results of the adapted D2D model with CC, bias, RMSE, rMBE, and rRMSE values, serving as a comprehensive reference for comparison between the GEMS-observed and D2D-nowcast O₃ products.

Fig. 8 illustrates the results of the D2D-nowcast model, which compares the GEMS O₃ product and D2D-predicted O₃ at different lead times. Fig. 8(a) displays the GEMS O₃ product on March 17, 2022, from 04:45 UTC to 06:45 UTC, while Fig. 8(b) shows the D2D-predicted O₃ one, two, and 3 h ahead of Fig. 8(a). Fig. 8(c) shows the discrepancy

between Fig. 8(a) and (b). The D2D-nowcasted O₃ was qualitatively accurate one, two, and 3 h ahead of the GEMS O₃ over the Korean Peninsula, eastern and southeastern China, and Northwestern Pacific region. However, it exhibited some inconsistencies with GEMS O₃ for the flow of high O₃ density over Beijing and northern China. However, the D2D-nowcasted O₃ exhibited excellent agreement with the GEMS O₃, although accuracy tended to decrease slightly from one to 3 h. This result suggests that the proposed D2D model is suitable for nowcasting GEMS O₃ within a 3-h time frame.

Fig. 9 shows the scatterplots between the GEMS and D2D-predicted O₃ values, with CC = 0.994, bias = 2.162 DU, RMSE = 5.606 DU, rMBE = 0.007, rRMSE = 0.018 and a slope of 0.946 for 1-h predictions; CC = 0.992, bias = 1.421 DU, RMSE = 5.903 DU, rMBE = 0.005, rRMSE = 0.019 and a slope of 0.945 for 2-h predictions; CC = 0.988, bias = 1.169 DU, RMSE = 6.797 DU, rMBE = 0.004, rRMSE = 0.022 and a slope of 0.944 for 3-h predictions. These results demonstrate the D2D O₃ prediction model's accuracy within 3 h. The D2D model showed less bias than the accuracy of the GEMS O₃ product, with an average of 0.533 %. The RMSE of the D2D-predicted model was 1.967 % compared to the maximum value for the GEMS O₃ products (Table 1). These statistics demonstrate that the proposed method, which combines the GEMS O₃ product with the CGAN-based D2D model, can achieve higher accuracy than previous studies utilizing the TROPOMI O₃ product (Wang et al., 2022).

4.2. D2D model validation with ozonesonde measurement

Fig. 10 visually represents a comparison between the GEMS total column O₃ nowcast by the constructed D2D model and the O₃ observed at the ozonesonde points in Pohang and Hong Kong in 2021. Fig. 10(a) is a scatterplot between the GEMS total columns O₃ and ozone in Seoul, with CC = 0.546, bias = -40.067 DU, RMSE = 43.803 DU, rMBE = 0.144, rRMSE = 0.157, and a slope of 0.790. Similarly, Fig. 10(b) is a scatterplot between the D2D-nowcasted O₃ and ozone concentration at Pohang, with CC = 0.547, bias = -41.132 DU, RMSE = 44.731 DU, rMBE = 0.148, rRMSE = 0.161 and a slope of 0.830. Fig. 10(c) is a scatterplot between the GEMS Total Column O₃ and ozone in Hong Kong, with CC = 0.655, bias = 5.250 DU, RMSE = 16.205 DU, rMBE = -0.019, rRMSE = 0.059, and a slope of 1.011. Finally, Fig. 10(d) is a scatterplot between the D2D-generated O₃ and ozone concentration at Hong Kong, with CC = 0.514, bias = 3.729 DU, RMSE = 17.886 DU, rMBE = -0.014, rRMSE = -0.066 and a slope of 0.864.

There was a relatively low correlation between D2D-predicted O₃ data and ozonesonde data, with a CC below 0.655, compared to the correlation between the GEMS and D2D-predicted O₃ data. Both the GEMS and D2D-nowcasted O₃ data underestimated ground-based O₃ observations in Pohang, South Korea, but they exhibited good agreement in Hong Kong. Furthermore, the paired O₃ datasets from Pohang were more scattered than those from Hong Kong. These differences may be attributed to the regional characteristic of higher O₃ concentrations in Pohang than in Hong Kong. Additionally, these variations may result from a comparison involving a limited number of collocated datasets between ground-based and GEMS O₃ observations, a distinction between point and gridded observations, a lack of validation data, or differences in observation height between ground-based and satellite data (Boynard et al., 2009). Nevertheless, the GEMS and D2D-predicted O₃ values in Pohang and Hong Kong exhibited excellent agreement, indicating that the D2D model successfully learned the features of the GEMS O₃ observations. Consequently, any discrepancies between GEMS and ground O₃ observations can be addressed by improving the GEMS O₃ product algorithm. Therefore, these results emphasize that the accuracy of the D2D model relies on that of the GEMS O₃ products. These results highlight the potential of the D2D model to nowcast virtual O₃ data and assist O₃ forecasters.

Table 3
Statistical results of the Adapted D2D Prediction model for O₃.

Statistical scores	Results
CC	0.991
BIAS (DU)	1.584
RMSE (DU)	6.102
rMBE (%)	0.533
rRMSE (%)	1.967

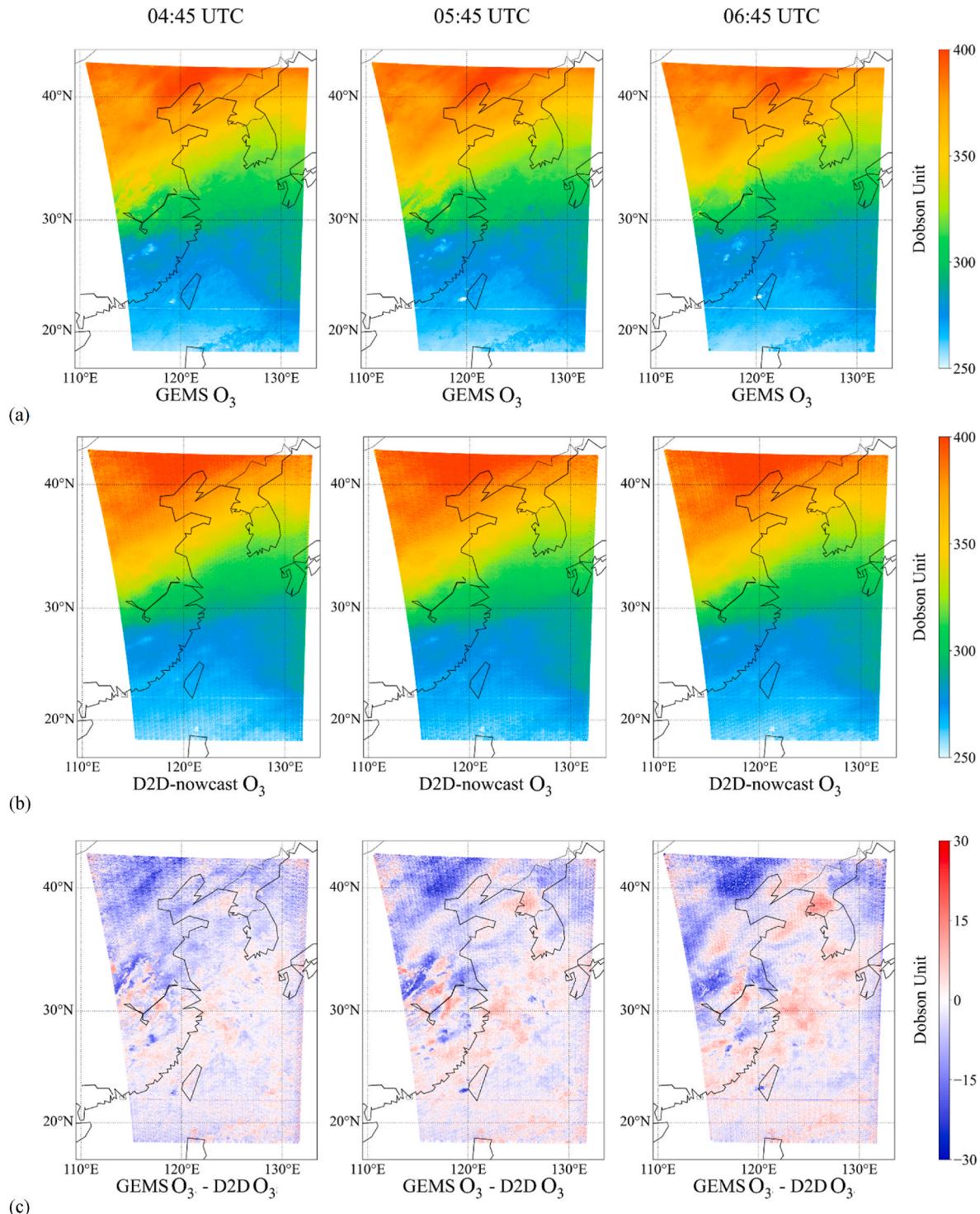


Fig. 8. Comparison between the GEMS O_3 product and the D2D-nowcast GEMS O_3 data at different time intervals on March 17, 2022. (a) GEMS O_3 product and (b) D2D-nowcast GEMS O_3 data from 04:45 UTC to 06:45 UTC, and (c) the discrepancy between the GEMS O_3 product and the D2D-nowcast GEMS O_3 data for the corresponding period. In each figure, the x-axis represents longitude, and the y-axis represents latitude.

4.3. D2D model validation with TROPOMI O_3 product

Fig. 11 are scatterplots comparing GEMS and D2D-nowcast O_3 products for 1-h predictions against TROPOMI O_3 products. Fig. 11(a) delineates the scatter plot between the GEMS and TROPOMI O_3 products, yielding a CC of 0.970, a bias of -4.231 DU, and an RMSE of 6.234 DU. Thus, significant agreement between GEMS and TROPOMI ozone datasets were apparent. Fig. 11(b) is the scatterplot between D2D-

nowcast and TROPOMI ozone, with a CC of 0.914, a bias of -3.714 DU, and an RMSE of 8.102 DU, indicative of reasonable agreement.

The typical temporal difference between the TROPOMI and GEMS ozone measurements was approximately 15 min. The GEMS O_3 product employs the TROPOMI ozone algorithm to align closely with TROPOMI's outcomes of TROPOMI, as shown in Fig. 11. These results demonstrate the consistency between the GEMS and D2D-nowcast O_3 products for 1-h predictions and TROPOMI O_3 products, in addition to

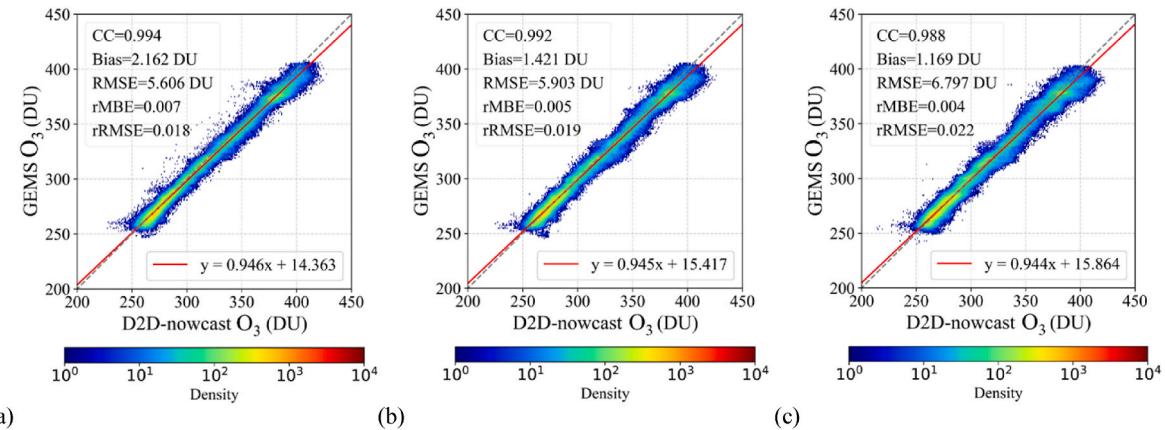


Fig. 9. Scatterplots between the GEMS-observed and D2D-predicted products. The time is the same as Fig. 8 at (a) 04:45 UTC for $t+1$ h, (b) 05:45 UTC for $t+2$ h, and (c) 06:45 UTC for $t+3$ h.

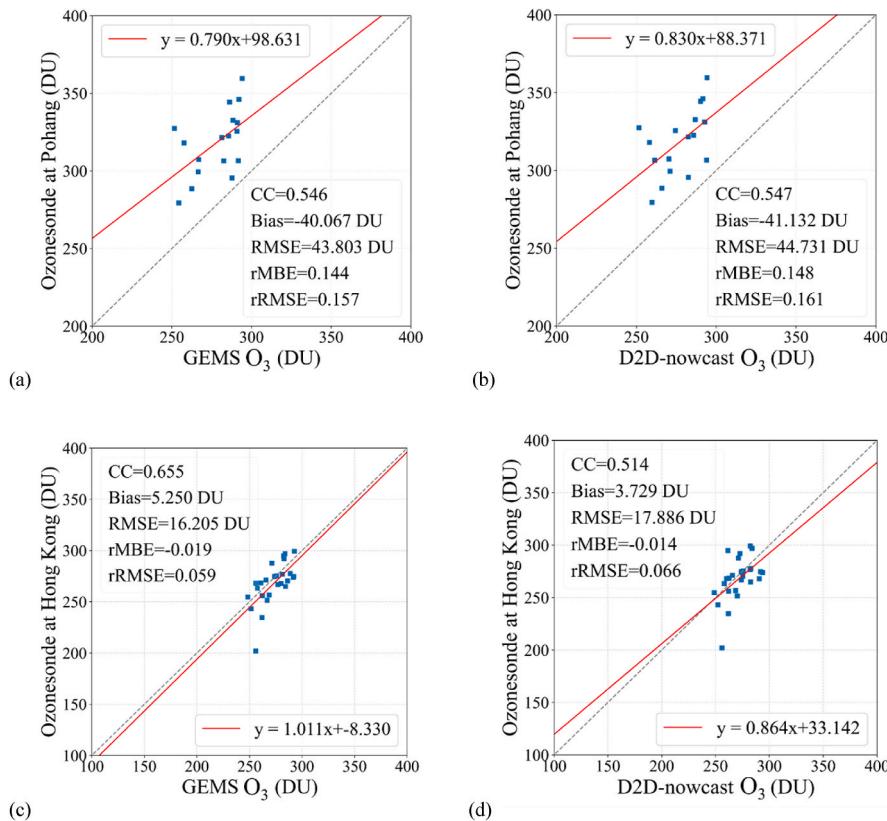


Fig. 10. Scatterplots between the GEMS-observed and 1-h D2D-nowcast O_3 data with ozonesonde measurements. The comparison results of O_3 measurements in Pohang, South Korea, at (a) and (b), and in King's Park, Hong Kong, at (c) and (d) are displayed.

the coherence and precision of D2D-nowcast O_3 data compared to GEMS O_3 products.

5. Discussion

This study presents a D2D method for nowcasting spatiotemporal variations in GEMS O_3 products up to 3 h in advance in time zones where GEMS observations are unavailable. Previous studies have not yet focused on nowcasting GEMS O_3 in Asia using ML and DL techniques. In general, ML- and DL-based methods for ozone prediction have concentrated on the importance of diverse factors, including physical properties, emission sources and rates, precursor observations, model-based data, and other predictive variables (Kang et al., 2021; Wang et al.,

2022; Li and Cheng., 2021). Our study achieved highly accurate nowcasting of GEMS O_3 by leveraging past GEMS L2 O_3 observations using the D2D method without relying on additional ancillary data. The proposed D2D model offers a valuable contribution to the atmospheric environmental community by providing a useful method for predicting multiple atmospheric gases, including O_3 .

This study also provides for the potential application to nowcasting of global O_3 concentrations, using two additional geostationary air quality satellites: NASA's EVI-1, covering North America, and ESA's Sentinel-4 covering, Europe (Chen and Lin, 2014) based on the developed D2D model using GEMS over Asia. This study contributes to our understanding and prediction of the global concentration and distribution of O_3 .

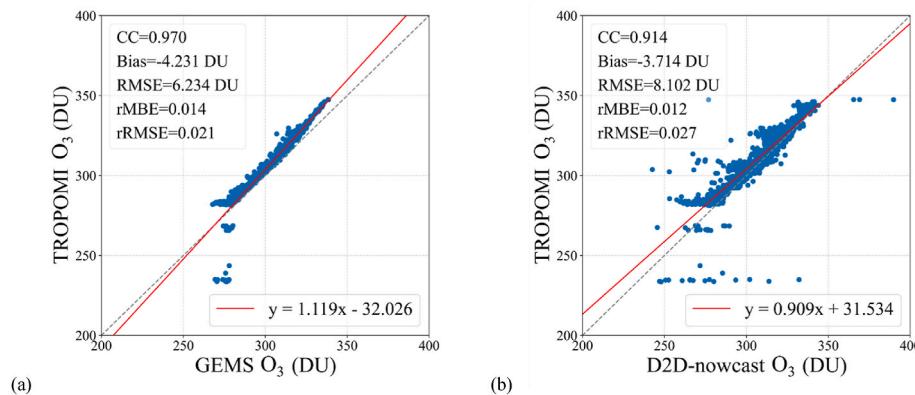


Fig. 11. Scatterplots for the annually-averaged ozone data (a) between GEMS and TROPOMI and (b) between the 1-h D2D-nowcasted and TROPOMI ozone data for 2021.

This study exhibited a few limitations. First, the validation of the D2D-nowcast model was challenging because of the restricted availability of ground-based O₃ observation sites in the study area. Second, the proposed D2D model requires six consecutive GEMS datasets for prediction. However, GEMS can only observe a maximum of eight times per day owing to the characteristics of a geostationary satellite. Third, the limited amount of data available from the relatively new GEMS environmental satellites restricted the reliability of the study, which could be enhanced using long-term GEMS data in future research. Fourth, an uncertainty of this study is related to the paired input and output datasets tailored to the study area, depending on the spatial coverage limitations of GEMS observation modes, because the FW mode of GEMS observation depends on daily solar effects and seasonal effects from three daily observations in winter and six daily observations in summer. Therefore, this study selected a stack of three successive observations as the maximum for nowcasting daily GEMS O₃ data. Finally, the accuracy of the proposed D2D model was sensitive to normalization preprocessing of the GEMS O₃ data during the dataset preparation step. For example, linear normalization tends to underestimate high O₃ concentrations, whereas exponential normalization tends to overestimate GEMS O₃ products at high ozone concentrations.

6. Conclusions

GEMS, the first geostationary satellite-based sensor with hyperspectral bands from UV to VIS wavelengths, has been instrumental in the spatiotemporal monitoring of various environmental parameters, including global O₃ concentration and transport in Asia, as well as aerosols, clouds, NO₂, SO₂, CHOCHO, HCHO, UVI, and surface reflectivity products. However, the effectiveness of GEMS observations is limited by solar effects, resulting in spatial and temporal gaps in monitoring coverage. This study analyzed GEMS O₃ using GK-2B/GEMS L2 TCO data covering East Asia, including the Korean Peninsula. The proposed D2D model was developed using the CGAN technique with paired GEMS O₃ product datasets. The D2D nowcasting model for the GEMS demonstrated excellent statistical agreement with GEMS O₃ observations within each 3-h prediction period.

The proposed D2D model can provide valuable GEMS O₃ nowcasting data from one to 3 h in advance during the 4-h temporal gap between the

numerical model-based O₃ predictions. Thus, the D2D model has the advantage of creating a synergistic effect on detailed O₃ predictions using the numerical prediction model. In addition, the D2D model has the advantage of nowcasting GEMS O₃ data within 3 h near twilight time-zones when GEMS data are unavailable. The proposed D2D approach can also be extended to other geostationary and polar-orbit satellites for monitoring ozone data, making it useful for environmental experts, public users, and decision makers.

Funding

This study was supported by the grant from the National Institute of Environment Research (NIER), funded by the Ministry of Environment (MOE) of the Republic of Korea (2023-01-02-095), and supported by “Development of Advanced Science and Technology for Marine Environmental Impact Assessment” (KIMST-20210427) and “Development of risk managing technology tackling ocean and fisheries crisis around Korean Peninsula by Kuroshio Current” (RS-2023-00256330) from the Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries.

CRediT authorship contribution statement

Seong-Taek Oh: Formal analysis, Investigation, Methodology, Validation, Visualization, Roles, Writing – original draft, Writing – review & editing. **Jeong-Eun Park:** Formal analysis, Methodology, Validation. **Jaehoon Jeong:** Data curation, Funding acquisition, Methodology, Project administration. **Sungwook Hong:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Roles, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Table A.1
List of acronyms and corresponding descriptions

Acronym	Description
ANN	Artificial Neural Networks
AOD	Aerosol Optical Depth
CC	Correlation Coefficient
CGAN	Conditional Generative Adversarial Networks
CNN	Convolutional Neural Network
CMAQ	Community Multi-scale Air Quality
DL	Deep Learning
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
D2D	Data-to-Data
FW	Full West
FC	Full Central
GEMS	Geostationary Environment Monitoring Spectrometer
GK	Geostationary Korea Multi-Purpose Satellite
GOME	Global Ozone Monitoring Experiment
L	Level
MBE	Mean Bias Error
ML	Machine Learning
NMHC	Non-Methane Hydrocarbon
NN	Neural Network
OMI	Ozone Monitoring Instrument
RF	Random Forest
RMSE	Root Mean Square Error
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
TOMS	Total Ozone Mapping Spectrometer
TROPOMI	Tropospheric Monitoring Instrument
UTC	Universal Time Coordinated
UV	Ultraviolet
UVI	Ultraviolet Index
VIS	Visible
WOUDC	World Ozone and Ultraviolet Radiation Data Center
WRF	Weather Research and Forecasting

References

- Alnaim, A., Sun, Z., Tong, D., 2022. Evaluating machine learning and remote sensing in monitoring NO₂ emission of power plants. *Rem. Sens.* 14, 729. <https://doi.org/10.3390/rs14030729>.
- Baek, K., Kim, J.-H., Bak, J., Haffner, D.P., Kang, M., Hong, H., 2022. Evaluation of Total Ozone Measurements from Geostationary Environmental Monitoring Satellite (GEMS). *EGUsphere*, pp. 1–23. <https://doi.org/10.5194/egusphere-2022-1402>.
- Boynard, A., Clerbaux, C., Coheur, P.F., Hurtmans, D., Turquety, S., George, M., Hadji-Lazaro, J., Keim, C., Meyer-Arnek, J., 2009. Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations. *Atmos. Chem. Phys.* 9 (16), 6255–6271. <https://doi.org/10.5194/acp-9-6255-2009>.
- Chandra, S., Ziemke, J., Duncan, B., Diehl, T., Livesey, N., Froidevaux, L., 2009. Effects of the 2006 El Niño on tropospheric ozone and carbon monoxide: implications for dynamics and biomass burning. *Atmos. Chem. Phys.* 9, 4239–4249. <https://doi.org/10.5194/acp-9-4239-2009>.
- Chen, C., Chen, X., Cheng, H., 2019. On the over-smoothing problem of cnn based disparity estimation. *IEEE/CVF Int. Conf. Computer Vision (ICCV)* 8997–9005. <https://doi.org/10.1109/ICCV.2019.00909>.
- Chen, X.-W., Lin, X., 2014. Big data deep learning: challenges and perspectives. *IEEE Access* 2, 514–525. <https://doi.org/10.1109/ACCESS.2014.2325029>.
- Correa, S.M., 1993. A review of NO_x formation under gas-turbine combustion conditions. *Combust. Sci. Technol.* 87, 329–362. <https://doi.org/10.1080/00102209208947221>.
- Crutzen, P.J., Andreae, M.O., 1990. Biomass burning in the tropics: impact on atmospheric chemistry and biogeochemical cycles. *Science* 250, 1669–1678. <https://doi.org/10.1126/science.250.4988.1669>.
- Denton, E.L., Chintala, S., Fergus, R., 2015. Deep generative image models using a laplacian pyramid of adversarial networks. In: *Advances in Neural Information Processing Systems*, vol. 28. <https://doi.org/10.48550/arXiv.1506.05751>.
- Engerer, N., Mills, F., 2015. Validating nine clear sky radiation models in Australia. *Sol. Energy* 120, 9–24. <https://doi.org/10.1016/j.solener.2015.06.044>.
- ESA (European Space Agency), 2021. SSP/TROPOMI Tropospheric Ozone ATBD. <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-5p/products-algorithms>. (Accessed 28 August 2023).
- Fishman, J., Bowman, K.W., Burrows, J.P., Richter, A., Chance, K.V., Edwards, D.P., Martin, R.V., Morris, G.A., Pierce, R.B., Ziemke, J.R., 2008. Remote sensing of tropospheric pollution from space. *Bull. Am. Meteorol. Soc.* 89, 805–822. <https://doi.org/10.1175/2008BAMS2526.1>.
- Forster, F., P, M., Shine, K.P., 1997. Radiative forcing and temperature trends from stratospheric ozone changes. *J. Geophys. Res. Atmos.* 102, 10841–10855. <https://doi.org/10.1029/96JD03510>.
- Garane, K., Koukouli, M., Verhoeft, T., Lerot, C., Heue, K., Fioletov, V., Balis, D., Bais, A., Bazureau, A., Dehn, A., Goutail, F., Granville, J., Griffin, D., Hubert, D., Keppens, A., Lambert, J., Loyola, D., McLinden, C., Pazmino, A., Pommereau, J., Redondas, A., Romahn, F., Valks, P., Rozendaal, M., Xu, J., Zehner, C., Zerefos, C., Zimmer, W., 2019. TROPOMI/SSP total ozone column data: global ground-based validation and consistency with other satellite missions. *Atmos. Meas. Tech.* 12 (10), 5263–5287. <https://doi.org/10.5194/amt-12-5263-2019>.
- Ghahremanloo, M., Lops, Y., Choi, Y., Yeganeh, B., 2021. Deep learning estimation of daily ground-level NO₂ concentrations from remote sensing data. *J. Geophys. Res. Atmos.* 126, e2021JD034925 <https://doi.org/10.1029/2021JD034925>.
- Guenther, A., Hewitt, C.N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W., 1995. A global model of natural volatile organic compound emissions. *J. Geophys. Res. Atmos.* 100, 8873–8892. <https://doi.org/10.1029/94JD02950>.
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P.I., Geron, C., 2006. Estimates of global terrestrial isoprene emissions using MEGAN (model of emissions of gases and aerosols from nature). *Atmos. Chem. Phys.* 6, 3181–3210. <https://doi.org/10.5194/acp-6-3181-2006>.
- Hong, H., Kim, J., Jeong, U., Han, K.-s., Lee, H., 2017. The effects of aerosol on the retrieval accuracy of NO₂ slant column density. *Rem. Sens.* 9, 867. <https://doi.org/10.3390/rs9080867>.
- Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. *IEEE Conf. Computer Vision Pattern Recognit (CVPR)* 1125–1134. <https://doi.org/10.1109/CVPR.2017.632>.
- Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.K., Kim, S., 2021. Estimation of surface-level NO₂ and O₃ concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* 288, 117711 <https://doi.org/10.1016/j.envpol.2021.117711>.
- Kim, J., Jeong, U., Ahn, M.-H., Kim, J.H., Park, R.J., Lee, H., Song, C.H., Choi, Y.-S., Lee, K.-H., Yoo, J.-M., 2020a. New era of air quality monitoring from space: geostationary Environment Monitoring Spectrometer (GEMS). *Bull. Am. Meteorol. Soc.* 101, E1–E22. <https://doi.org/10.1175/BAMS-D-18-0013.1>.
- Kim, J.-H., Ryu, S., Jeong, J., So, D., Ban, H.-J., Hong, S., 2020b. Impact of satellite sounding data on virtual visible imagery generation using conditional generative adversarial network. *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 13, 4532–4541. <https://doi.org/10.1109/JSTARS.2020.3013598>.

- Kim, K., Kim, J.-H., Moon, Y.-J., Park, E., Shin, G., Kim, T., Kim, Y., Hong, S., 2019. Nighttime reflectance generation in the visible band of satellites. *Rem. Sens.* 11, 2087. <https://doi.org/10.3390/rs11182087>.
- Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks. International Conference on Machine Learning. PMLR, pp. 1857–1865. <https://doi.org/10.48550/arXiv.1703.05192>.
- Kim, Y., Hong, S., 2019b. Deep learning-generated nighttime reflectance and daytime radiance of the midwave infrared band of a geostationary satellite. *Rem. Sens.* 11, 2713. <https://doi.org/10.3390/rs11222713>.
- Lacis, A.A., Wuebbles, D.J., Logan, J.A., 1990. Radiative forcing of climate by changes in the vertical distribution of ozone. *J. Geophys. Res. Atmos.* 95, 9971–9981. <https://doi.org/10.1029/JD095iD07p09971>.
- Lasnik, J., Stephens, M., Baker, B., Randall, C., Ko, D., Kim, S., Kim, Y., Lee, E., Chang, S., Park, J., 2014. Geostationary Environment Monitoring Spectrometer (GEMS) over the Korea Peninsula and Asia-Pacific Region. AGU Fall Meeting Abstracts, pp. A51A–A3003, 2014.
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G.B., Seo, J.B., Kim, N., 2017. Deep learning in medical imaging: general overview. *Korean J. Radiol.* 18, 570–584. <https://doi.org/10.3348/kjr.2017.18.4.570>.
- Lee, Y., Ahn, M.-H., Kang, M., Eo, M., 2023. Spectral replacement using machine learning methods for continuous mapping of the Geostationary Environment Monitoring Spectrometer (GEMS). *Atmos. Meas. Tech.* 16, 153–168. <https://doi.org/10.5194/amt-16-153-2023>.
- Li, C., Wand, M., 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. European Conf. Computer Vision 702–716. <https://doi.org/10.48550/arXiv.1604.04382>.
- Li, C., Xu, X., Liu, X., Wang, J., Sun, K., van Geffen, J., Zhu, Q., Ma, J., Jin, J., Qin, K., 2022. Direct retrieval of NO₂ vertical columns from UV-Vis (390–495 nm) spectral radiances using a neural network. *J. Remote Sensing*, 9817134. <https://doi.org/10.34133/2022/9817134>.
- Li, T., Cheng, X., 2021. Estimating daily full-coverage surface ozone concentration using satellite observations and a spatiotemporally embedded deep learning approach. *Int. J. Appl. Earth Obs. Geoinf.* 101, 102356 <https://doi.org/10.1016/j.jag.2021.102356>.
- Li, T., Wang, Y., Yuan, Q., 2020. Remote sensing estimation of regional NO₂ via space-time neural networks. *Rem. Sens.* 12, 2514. <https://doi.org/10.3390/rs12162514>.
- Liu, X., Bhartia, P., Chance, K., Spurr, R., Kurosawa, T., 2010. Ozone profile retrievals from the ozone monitoring instrument. *Atmos. Chem. Phys.* 10, 2521–2537. <https://doi.org/10.5194/acp-10-2521-2010>.
- Lops, Y., Ghahremanloo, M., Pouyaei, A., Choi, Y., Jung, J., Mousavinezhad, S., Salman, A.K., Hammond, D., 2023. Spatiotemporal estimation of TROPOMI NO₂ column with depthwise partial convolutional neural network. *Neural Computing and Applications*, pp. 1–12. <https://doi.org/10.48550/arXiv.2204.05917>.
- Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S., 2017. Least squares generative adversarial networks. IEEE international conference on computer vision (ICCV) 2794–2802. <https://doi.org/10.1109/ICCV.2017.304>.
- McElroy, C., Hare, E., 1999. World Ozone and Ultraviolet Radiation Data Centre. https://woudc.org/archive/Publications/Meeting_Reports/Umkehr_Sub-Committee. (Accessed 28 August 2023).
- Michelsanti, D., Tan, Z.-H., 2017. Conditional Generative Adversarial Networks for Speech Enhancement and Noise-Robust Speaker Verification. <https://doi.org/10.48550/arXiv.1709.01703> arXiv preprint arXiv:1709.01703.
- Munro, R., Siddans, R., Reburn, W.J., Kerridge, B.J., 1998. Direct measurement of tropospheric ozone distributions from space. *Nature* 392, 168–171. <https://doi.org/10.1038/32392>.
- NESC (National Environmental Satellite Center), 2021. Annual Observation Schedule for GEMS. <https://nesc.nier.go.kr/ko/html/satellite/operation.do>. (Accessed 28 August 2023).
- Nguyen, V., Yago Vicente, T.F., Zhao, M., Hoai, M., Samaras, D., 2017. Shadow detection with conditional generative adversarial networks. IEEE Int. Conf. Computer Vision (ICCV) 4510–4518. <https://doi.org/10.1109/ICCV.2017.483>.
- Ott, L.E., Pickering, K.E., Stenckhoff, G.L., Huntrieser, H., Schumann, U., 2007. Effects of lightning NO_x production during the 21 July European Lightning Nitrogen Oxides Project storm studied with a three-dimensional cloud-scale chemical transport model. *J. Geophys. Res. Atmos.* 112, 1–18. <https://doi.org/10.1029/2006JD007365>.
- Park, J., Hong, H., Lee, H., 2018. A Novel Algorithm for Tropospheric Nitrogen Dioxide (NO₂) Vertical Profile Retrieval from Space-Borne UV-Vis Hyper-Spectral Sensors: Ideas and Application to OMI. AGU Fall Meeting Abstracts. A53G-2552.
- Park, J.-E., Kim, G., Hong, S., 2020. Green band generation for advanced baseline imager sensor using Pix2Pix with advanced baseline imager and advanced Himawari imager observations. *IEEE Trans. Geosci. Rem. Sens.* 59, 6415–6423. <https://doi.org/10.1109/TGRS.2020.3032732>.
- Park, S., Shin, M., Im, J., Song, C.-K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J., Lee, D.-W., 2019. Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. *Atmos. Chem. Phys.* 19, 1097–1113. <https://doi.org/10.5194/acp-19-1097-2019>.
- Platt, U., Stutz, Z., 2008. Differential optical absorption spectroscopy. In: Platt, U., Stutz, Z. (Eds.), *Differential Absorption Spectroscopy*. Springer, Heidelberg, Germany, pp. 135–174.
- Raford, A., Metz, L., Chintala, S., 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. <https://doi.org/10.48550/arXiv.1511.06434> arXiv preprint arXiv:1511.06434.
- Santos, C.N.d., Wadhawan, K., Zhou, B., 2017. Learning Loss Functions for Semi-supervised Learning via Discriminative Adversarial Networks. <https://doi.org/10.48550/arXiv.1707.02198> arXiv preprint arXiv:1707.02198.
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28, 802–810. <https://doi.org/10.48550/arXiv.1506.04214>.
- Siegert, F., Ruecker, G., Hinrichs, A., Hoffmann, A., 2001. Increased damage from fires in logged forests during droughts caused by El Nino. *Nature* 414, 437–440. <https://doi.org/10.1038/35106547>.
- Tao, Y., Hsu, K., Ihler, A., Gao, X., Sorooshian, S., 2018. A two-stage deep neural network framework for precipitation estimation from bispectral satellite information. *J. Hydrometeorol.* 19, 393–408. <https://doi.org/10.1175/JHM-D-17-0077.1>.
- Thompson, A., Pickering, K., McNamara, D., Schoeberl, M., Hudson, R., Kim, J., Browell, E., Kirchhoff, V.W.J.H., Nganga, D., 1996. Where did tropospheric ozone over southern Africa and the tropical Atlantic come from in October 1992? Insights from TOMS, GTE TRACE A, and SAFARI 1992. *J. Geophys. Res. Atmos.* 101, 24251–24278. <https://doi.org/10.1029/96JD01463>.
- Thompson, A.M., Witte, J.C., Hudson, R.D., Guo, H., Herman, J.R., Fujiwara, M., 2001. Tropical tropospheric ozone and biomass burning. *Science* 291, 2128–2132. <https://doi.org/10.1126/science.291.5511.2128>.
- Thompson, A.M., Witte, J.C., McPeters, R.D., Oltmans, S.J., Schmidlin, F.J., Logan, J.A., Fujiwara, M., Kirchhoff, V.W., Posny, F., Coetzee, G.J., 2003a. Southern hemisphere additional ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology 1. Comparison with total ozone mapping spectrometer (TOMS) and ground-based measurements. *J. Geophys. Res. Atmos.* 108, 8238. <https://doi.org/10.1029/2001JD000967>.
- Thompson, A.M., Witte, J.C., Oltmans, S.J., Schmidlin, F.J., Logan, J.A., Fujiwara, M., Kirchhoff, V.W., Posny, F., Coetzee, G.J., Hoegger, B., 2003b. Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology 2. Tropospheric variability and the zonal wave-one. *J. Geophys. Res. Atmos.* 108, 8241. <https://doi.org/10.1029/2002JD002241>.
- Van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoest, T., Lambert, J.-C., Sneep, M., Ter Linden, M., Ludewig, A., Boersma, K.F., 2022. Sentinel-5P TROPOMI NO₂ retrieval: impact of version v2. 2 improvements and comparisons with OMI and ground-based data. *Atmos. Meas. Tech.* 15, 2037–2060. <https://doi.org/10.5194/amt-15-2037-2022>.
- Veefkind, J.P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H., De Haan, J., Kleipool, Q., 2012. TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Rem. Sens. Environ.* 120, 70–83. <https://doi.org/10.1016/j.rse.2011.09.027>.
- Wang, W., Liu, X., Bi, J., Liu, Y., 2022. A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environ. Int.* 158, 106917 <https://doi.org/10.1016/j.envint.2021.106917>.
- Xu, H.-h., Ge, D.-y., 2020. A novel image edge smoothing method based on convolutional neural network. *Int. J. Adv. Rob. Syst.* 17, 1729881420921676 <https://doi.org/10.1177/1729881420921676>.
- Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE Int. Conf. Computer Vision (ICCV) 2223–2232. <https://doi.org/10.1109/ICCV.2017.244>.
- Ziemke, J., Chandra, S., Duncan, B., Schoeberl, M., Torres, O., Damon, M., Bhartia, P., 2009. Recent biomass burning in the tropics and related changes in tropospheric ozone. *Geophys. Res. Lett.* 36, L15819 <https://doi.org/10.1029/2009GL039303>.