Spring 2024

DEEP LEARNING APPLICATIONS ON IONOSPHERIC STUDIES

Yang Pan

Mingwu Jin
University of Texas at Arlington

Yue Deng
University of Texas at Arlington

DEEP LEARNING APPLICATIONS ON IONOSPHERIC STUDIES

by

YANG PAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctoral of Science in Physics at
The University of Texas at Arlington
May 2024

Arlington, Texas

Thesis Committee:

Yue Deng, Co-Chair
Mingwu Jin, Co-Chair
Ramon E. Lopez
Frederick D. Wilder
Amir Shahmoradi

ABSTRACT

DEEP LEARNING APPLICATIONS ON IONOSPHERIC STUDIES

Deep Learning Applications on Ionospheric Studies

Yang Pan, Ph.D.

The University of Texas at Arlington, 2024

Supervising Professors: Mingwu Jin, Yue Deng

Machine learning techniques, particularly deep learning techniques, have been vigorously pursued to tackle space physics problems and achieved some impressive results recently. The growth of deep learning technologies in different domains enables innovative solutions to those problems compared to conventional methods. Filling data gaps in instrumental observations is among the demanding issues, which benefits space physicists to study ionospheric phenomena with complete data coverage. Global total electron content (TEC) and regional ionospheric electron density (*Ne*) are among important physical parameters in ionospheric studies. Due to the limited coverage of global navigation satellite system (GNSS) ground receivers and sporadic operations of the Millstone Hill incoherent scatter radar (ISR), the global TEC maps and regional *Ne* observations suffer huge amount of data gaps. In this dissertation, we utilize the advanced deep learning methods, generative adversarial networks (GANs) and neural architectural search (NAS), to fill the data gaps in TEC maps and *Ne* patterns. We have conducted comprehensive experiments to demonstrate their superior performances over traditional methods. Through these studies, it becomes increasingly evident that the great potential of deep learning will play a key role in future research of ionosphere and the broader realm of space physics.

ACKNOWLEDGMENTS

I thank my supervisor Dr. Yue Deng for her continuous support and guidance through projects to projects, and Dr. Mingwu Jin, especially for the communications and his support in machine learning aspects. Meanwhile, the special thanks also belong to Dr. Shun-Rong Zhang, who provides the support regarding details of incoherent scatter radar. Thanks to my parents, Shiqiao Pan and Guimei Yang, they have not only taught me how to hehave, but have been supporting me through all the hardships in my life. And my special thanks give to Dr. Thomas C. Bower, M.D., and Dr. Alberto Pochettino, M.D., and the team in Mayo Clinic, Rochester, Minnesota for their great help in sustaining my physical well beings. And this credit goes to my godfather, Dr. Chao Feng, for his clarifications on my puzzlements. For the blessed wisdom shall guide us on this lifelong spiritual journey.

LIST OF FIGURES

LIST OF TABLES

Table of Contents

1 INTRODUCTION

Atmosphere of Earth is typically defined as layers of gases surrounding the Earth [*Hart*, 1978], which majorly consist of gases as Nitrogen and Oxygen under constant and complicated chemical reactions [*Schlager et al.*, 2012]. The temperature serves a valuable criterion to stratify the atmosphere into multiple layers, including troposphere, stratosphere [*Spar*, 1970], mesosphere, thermosphere, and exosphere as shown in Figure 1.1a [*Kelley*, 2003]. Meanwhile, the upper atmosphere region between 50 to 1,000 km above sea level resides ionosphere, and its name indicates that ionized atoms and molecules are the ingredients. The ionosphere is stratified into multiple layers, including D, E and F as shown in Figure 1.1b, based on electron density distribution [*Appleton*, 1956].

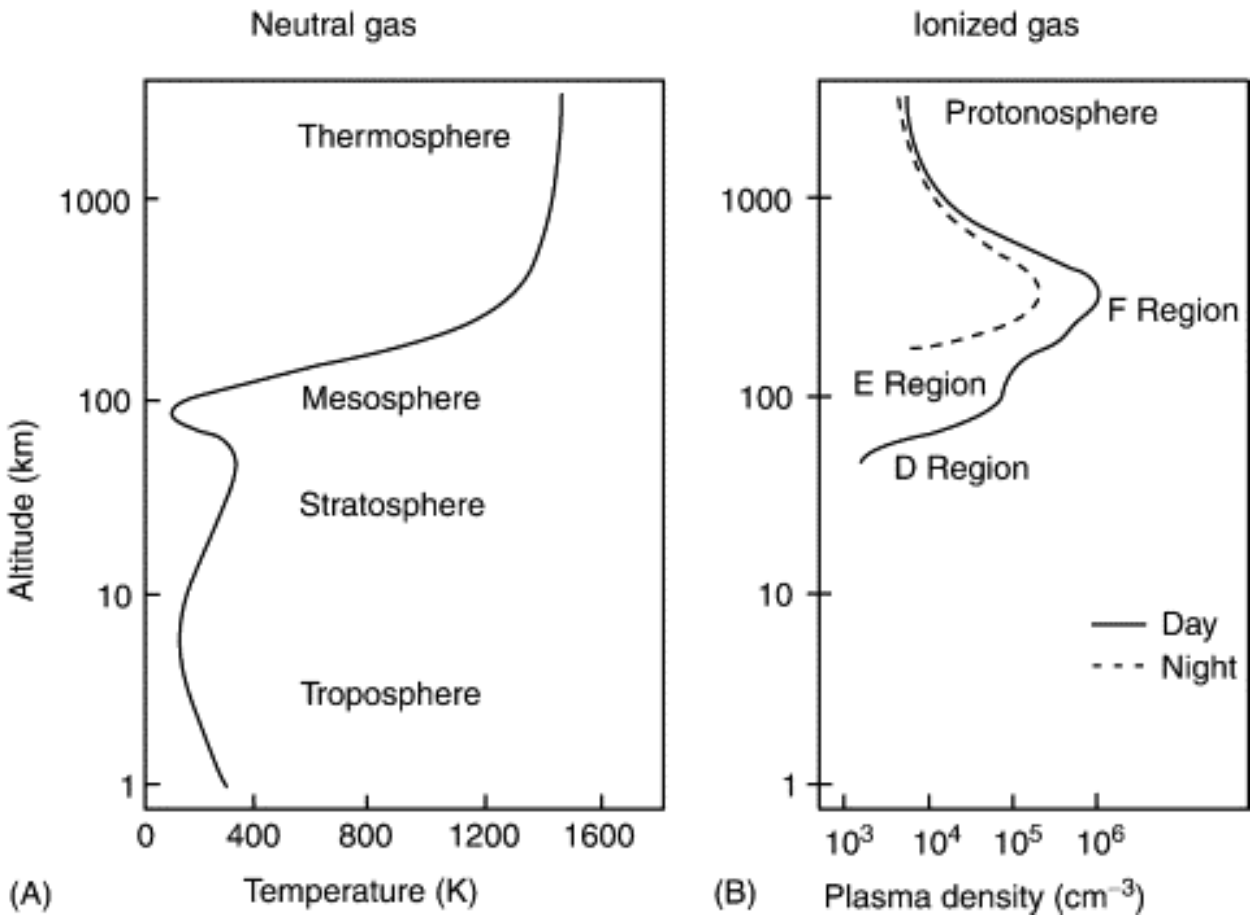


*Figure 1.1 Stratification on (a) atmosphere and (b) ionosphere on temperature and plasma density respectively. [*Kelley*, 2003].*

1.1 Ionospheric Stratification

The source of ionized particles comes primarily from extreme ultraviolet (EUV), x-ray solar radiation, and energetic particle precipitation. Meanwhile, the recombination is the process in which charged particles are formed back into neutral molecules. The ionization and recombination processes [*Bradbury*, 1938; *Mitra and Jones*, 1954] make the ionosphere a quite dynamic environment.

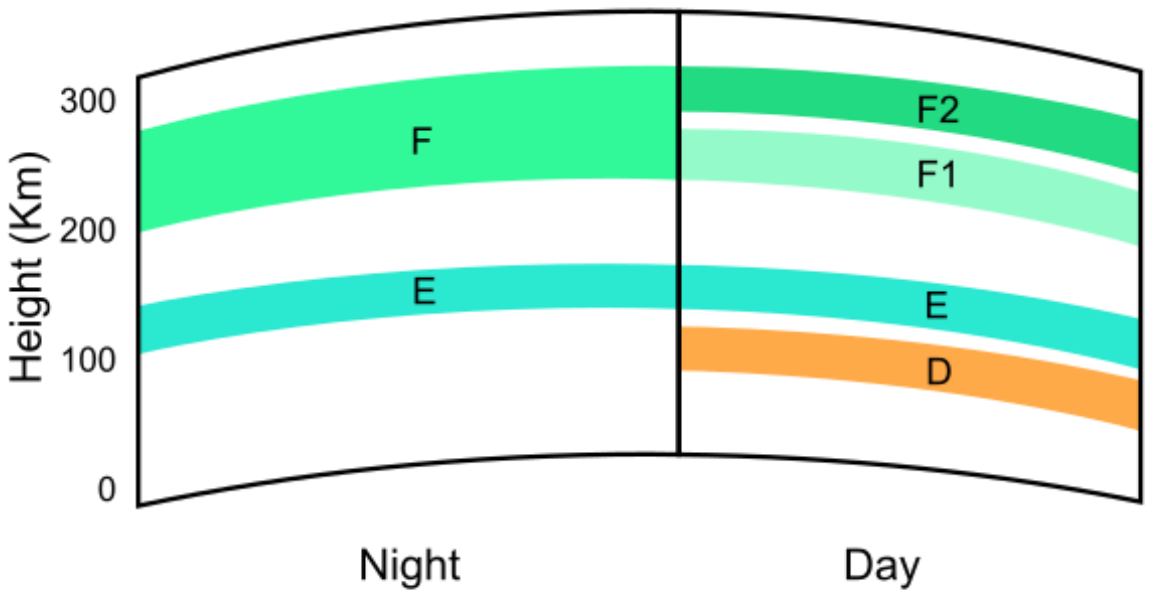


Figure 1.2 Ionospheric layers

[http://www.weather.nps.navy.mil/~psguest/EMEO_online/module3/module_3_2.html].

Normally, the Earth's atmosphere can be generally divided into multiple layers: D, E, and F regions (Figure 1.2) [*Cummer et al.*, 1998; *Schunk and Nagy*, 1978]. The lower ionospheric layer, the harder EUV reaches, which makes the ionization rate differ from layer to layer. Besides, the gas composition varies at different altitudes, which strongly influences the plasma density. The status of ionosphere can significantly impact the radio communication. During radio communication, the radio wave drives electron away. The neutral molecules residing within will collide with the wave, consuming the wave energy. Typically, the lower frequency, the higher chance of collision, resulting a higher absorption rate than the higher frequency waves. Therefore, the radio waves behave differently when propagating through the ionosphere due to their different frequencies [*Godyak et al.*, 1999; *Rawer*, 2013].

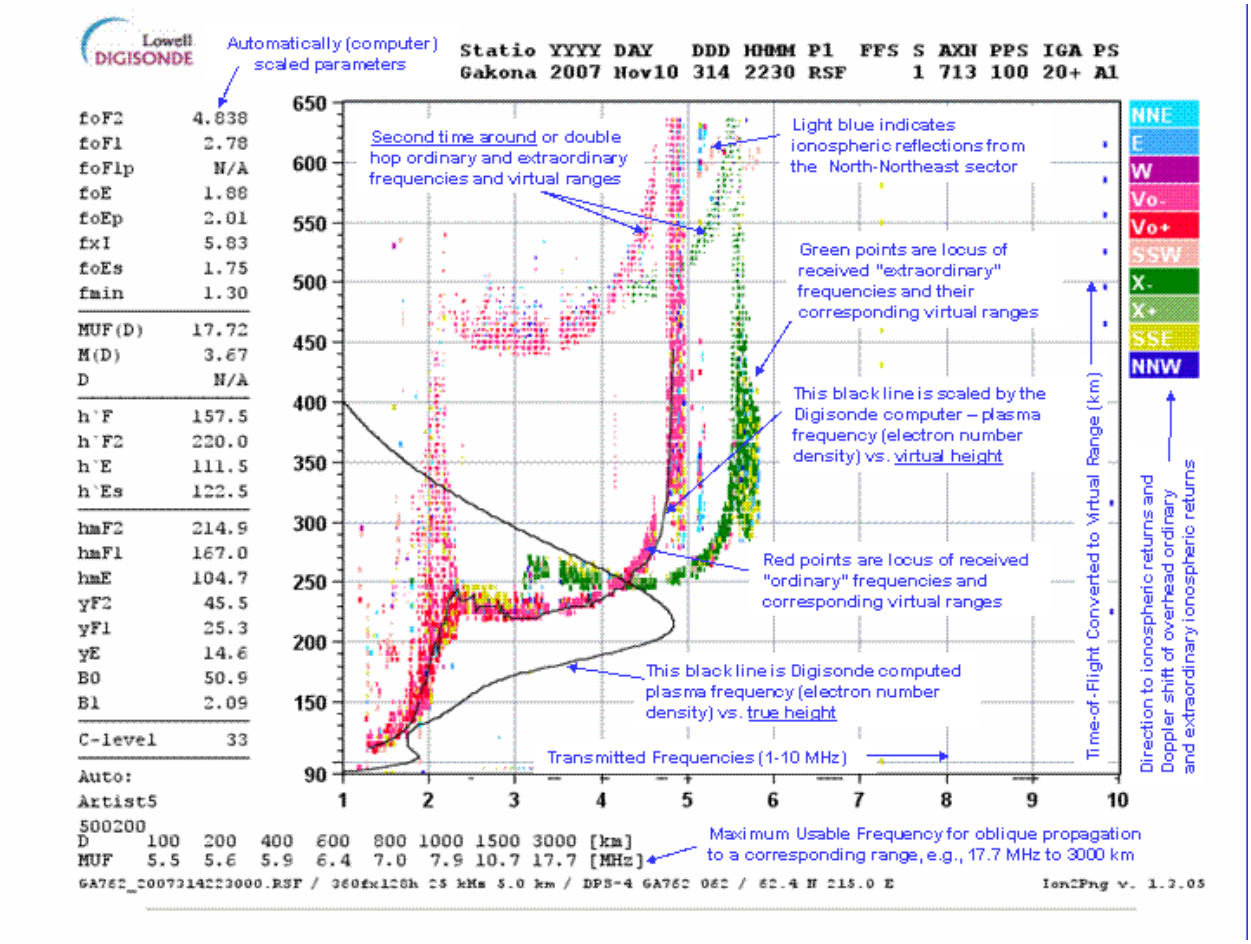


Figure 1.3 Lowell Digisonde

[https://web.archive.org/web/20130404234733/http://www.haarp.alaska.edu/haarp/dsonde.html].

Noticeably, when the broadcast wave exceeds a certain frequency, called critical frequency, the wave will penetrate the layer without reflection [*Theimer and Taylor*, 1961]. And an ionosonde broadcasts a range of high frequencies (HF) [*Davies and Baker*, 1966] and measures the critical frequencies of different layers, as different traces shown on the ionogram (Figure 1.3) [*Judd*, 1987].

To briefly introduce those layers, the D region is the bottom ionospheric region with few free electrons, which is ranged about 70 to 90 km. The E region is the middle region (90 to 160 km) and accounts for the trans-Atlantic radio wave signal in 1902 by the experiment of Guglielmo Marconi [*Marconi*, 1922]. While the upper most F region (over 160 km) has the greatest

concentration of electrons in three layers, which is well reflected by Figure 1.1b. The curves on right hand side are the electron density profiles (EDPs), and the density peak at F region is noted as NmF2. The corresponding F2 region critical frequency (FoF2) [*Kuznetsov et al.*, 1998], with hmF2, are two of the key parameters to induce empirical EDP models such as International Reference Ionosphere (IRI) [*D. Bilitza*, 2001; *Dieter Bilitza et al.*, 2022]. Besides, the upcoming Geospace Dynamics Constellation (GDC) mission [*Pfaff*, 2016] plans to send instrument around an altitude of 350km, which focuses on studies of the F region. Moreover, this layer is crucial to the long-distance HF radio communications. It is worth mentioning that the reduced solar radiation during nighttime accounts for the disappearance of D region and F1 region (Figure 1.2). The integral of electrons along altitudinal path in ionosphere plays another important role in addition to those stratified regions.

1.2 Total Electron Content

Total electron content (TEC) is defined as the integration of electron substances along the path in ionosphere. Global Navigation Satellite System (GNSS) [*Hegarty and Chatre*, 2008] is a constellation of satellites providing communications between orbiting instruments and receivers.

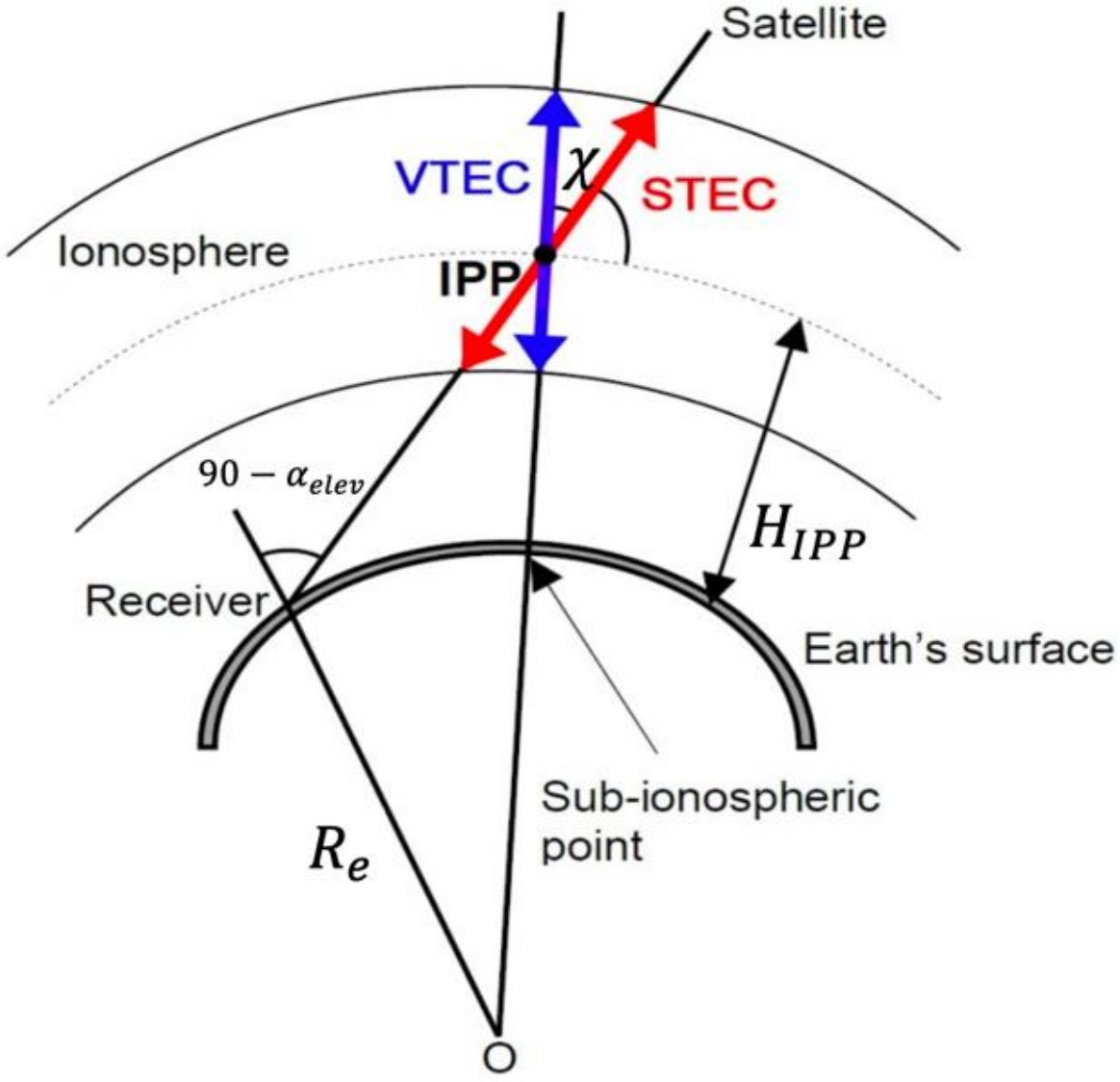


Figure 1.4 TEC measurement between satellite and ground receiver [Janssen, 2012].

Slant TEC (sTEC) is defined as the TEC along the arbitrary path of communication between satellite and receiver through ionosphere. As shown in Figure 1.4 *[Janssen, 2012]*, the part of line of sight from satellite to ground receiver within ionosphere is defined as slant TEC (STEC). Within this graph, ionospheric piercing point (IPP) is the intersection point of the line of sight and the ionospheric thin shell, based on the assumption of models. Electron contents inducing

signal phase delay can be noted in Equation *1.1* [*Ruffini et al.*, 1998], where L_i as the phase delay and path length s_i based on a specific frequency.

$$L_i = s_i - \frac{40.3}{f_i^2} \cdot STEC_i \quad (1.1)$$

STEC can be calculated using dual-frequency measurement, in which two different signal bands (*L1* at 1575.42 MHz and *L2* at 1227.60 MHz) work together to derive STEC as shown in Equation *1.2* [*Emery and Camps*, 2017; *Ruffini et al.*, 1998], where f_1 and f_2 are the frequencies mentioned above, while L_1 and L_2 are the phase delays of f_1 and f_2, respectively. And the vertical TEC (VTEC) is derived directly from STEC by applying cosine based on the zenith angle χ at IPP (Equation *1.3*, *[Janssen, 2012]*).

$$STEC = \frac{1}{40.3}\left(\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\right)(L_1 - L_2) \quad (1.2)$$

$$VTEC = STEC \cdot \cos\chi \quad (1.3)$$

With the coverages of GNSS ground receivers around the world, a global TEC map shows important information of ionosphere in a global scale as demonstrated in Figure 1.5. Upon the right-hand side color bar lies the unit for TEC, in which 1 TEC unit (TECU) equals to 10^{16} *electrons/m*^{2}. Besides, it can be easily seen that the better coverage of land regions over the oceanic counterparts. This is due to the limited number of GNSS ground receivers on oceans [*Vierinen et al.*, 2016].

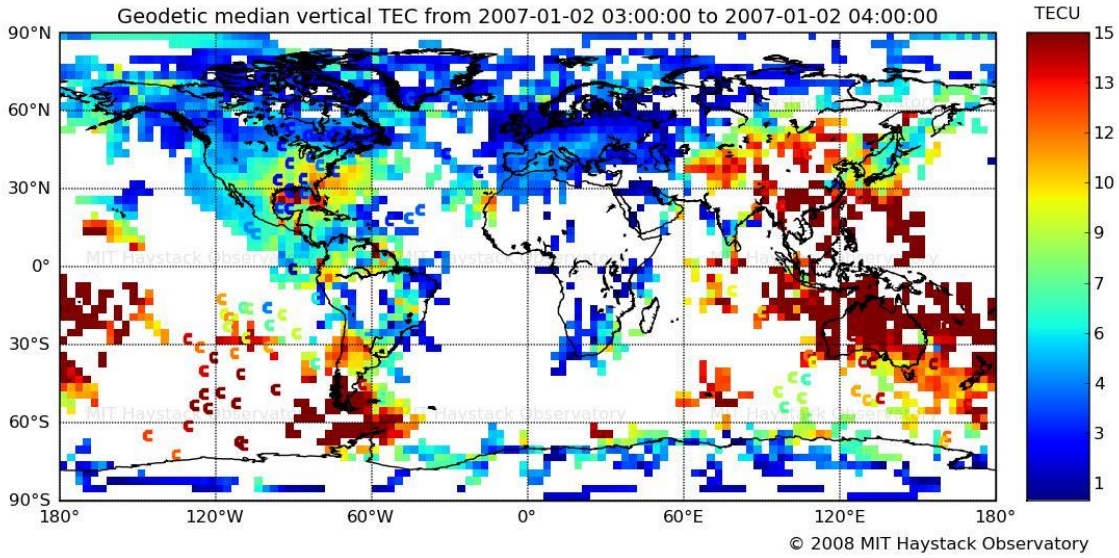


Figure 1.5 Global GNSS TEC map [Anthea Coster and Komjathy, 2008].

To study ionospheric activities, especially the travelling ionospheric disturbances (TIDs) [*AJ Coster et al.*, 2003; *Lyons et al.*, 2019; *Nishimura et al.*, 2020; *Tsugawa et al.*, 2007; *S-R Zhang et al.*, 2019], the complete global TEC map provides valuable information to derive the disturbances including the origin, wavelength, propagation speed, and direction, etc. Figure 1.6 shows an example of detecting TIDs over north American continent at 2017-09-07 23 UT during a geomagnetic storm. The differential TEC (dTEC), shown in the left panel of Figure 1.6, is calculated by subtracting background TEC over a time window, such as 30-minute. With the representative wave-like shape and repetition on dTEC, we identify the arrow-pointing region T1 in Figure 1.6 is identified as TID [*S-R Zhang et al.*, 2022].

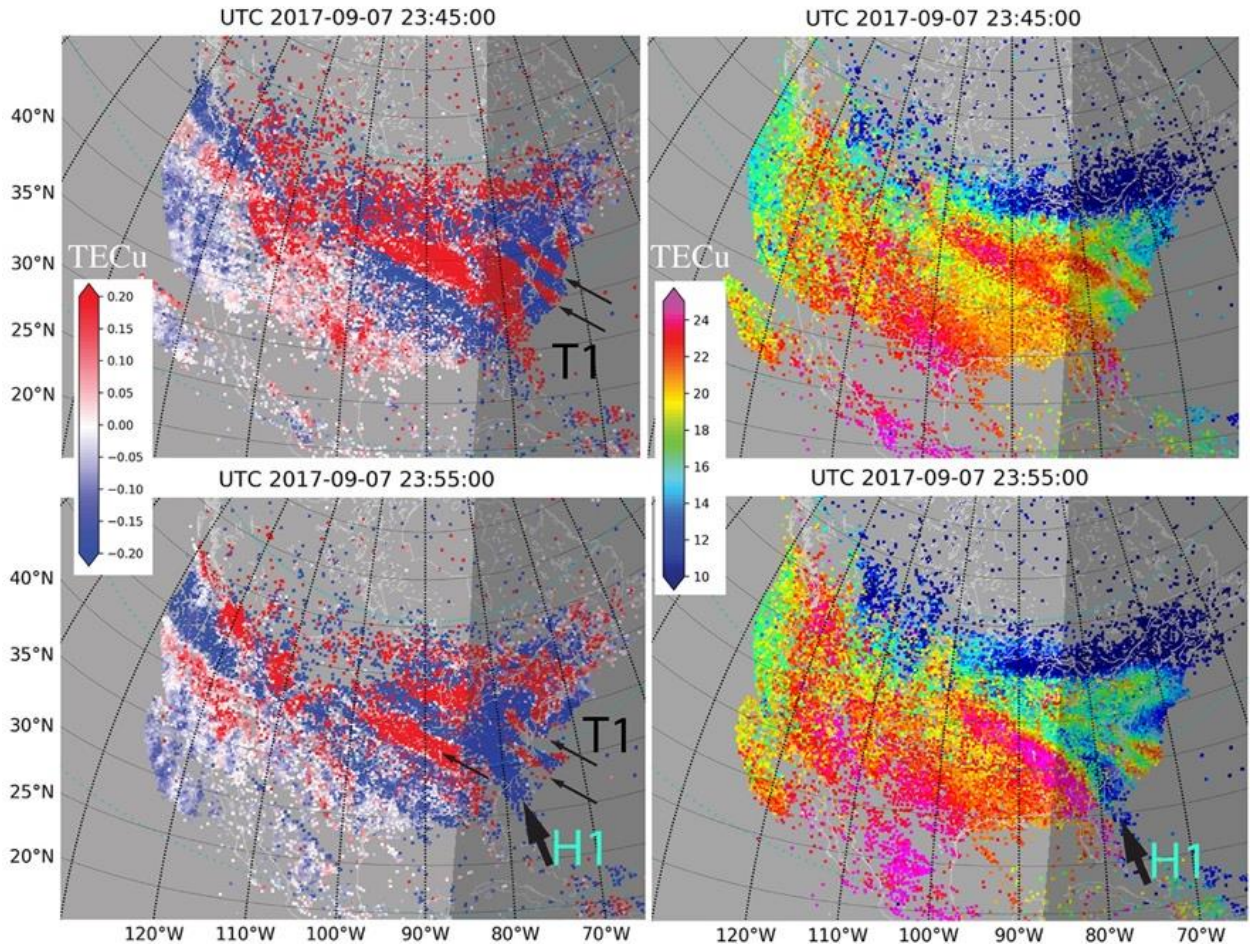


Figure 1.6 Detection of TID based on TEC map [S-R Zhang et al., 2022], left: dTEC maps, right: of TECu maps.

The GNSS-TEC is noted as Massachusetts Institute of Technology TEC (MIT-TEC) in the following sections, unless otherwise stated. The TEC map based on observation is limited and incomplete, however, resources providing complete global TEC maps exist. International GNSS service (IGS) has been providing IGS global TEC (IGS-TEC) data since June 1998, with the help of cooperation among five members of IGS Ionosphere Associate Analysis Centers (IAACs) *[M. Hernández-Pajares et al., 2009]*. The global IGS-TEC maps are produced based on the harmonic coefficients and appear smooth as shown in Figure 1.7. IGS-TEC is a reliable complete TEC map source for us to apply machine learning models, which we will explain in detail in later chapters.

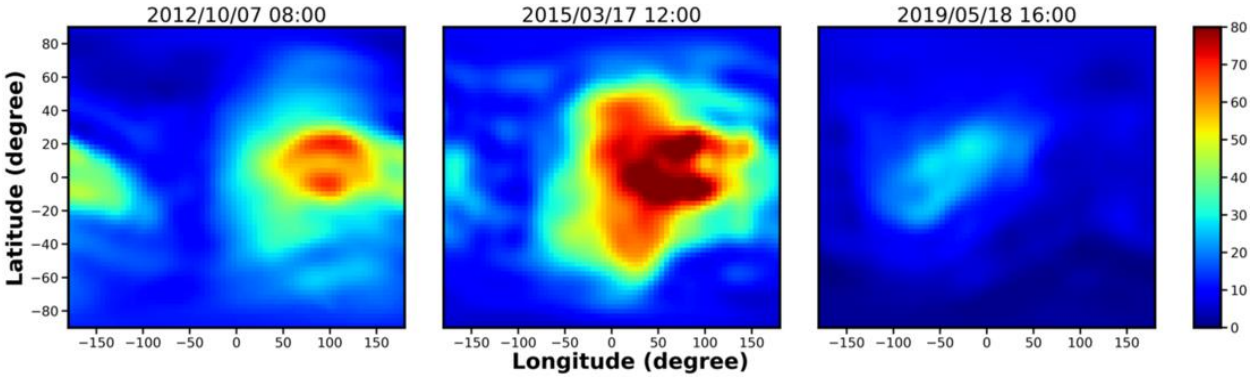


Figure 1.7 The IGS-TEC maps at three time points: 2012-10-07 08UT, 2015-03-17 12UT, and 2019-05-18 16UT [Pan et al., 2020].

1.3 Ionospheric Electron Densities

As seen in Figure 1.1b, the electron density profile (EDP) exhibits a certain pattern. Different from TEC measurement, the EDP includes detailed information of electron densities at varied heights. Incoherent scatter radar (ISR) is a powerful ground antenna measuring ionospheric parameters including electron density, ion and electron temperature, drift velocity, etc [*Wannberg et al.*, 1997]. Electron density shows a strong region-based characteristics, and the Millstone Hill (42.6°N, 71.5°W, dip 71.6°), ISR (MLH-ISR) provides valuable data for investigations at mid-latitude. The example of EDP provided by MLH-ISR (available at site: http://cedar.openmadrigal.org/) is shown in Figure 1.8. Zenith stands for the zenith radar, which is fixed and vertical-directing antenna along with the other radar called Millstone Hill Steerable Antenna (MISA). As the name indicated, MISA is steerable and scans a certain region. Both two radars consume a large amount of resource, majorly electricity, to operate. Therefore, Figure 1.8 has shown data gaps of radar off-hours from 00 UT to 20 UT. To describe electron density quantitatively, a logarithm ($log_{10}Ne$) is applied, where Ne is the notation for electron density.

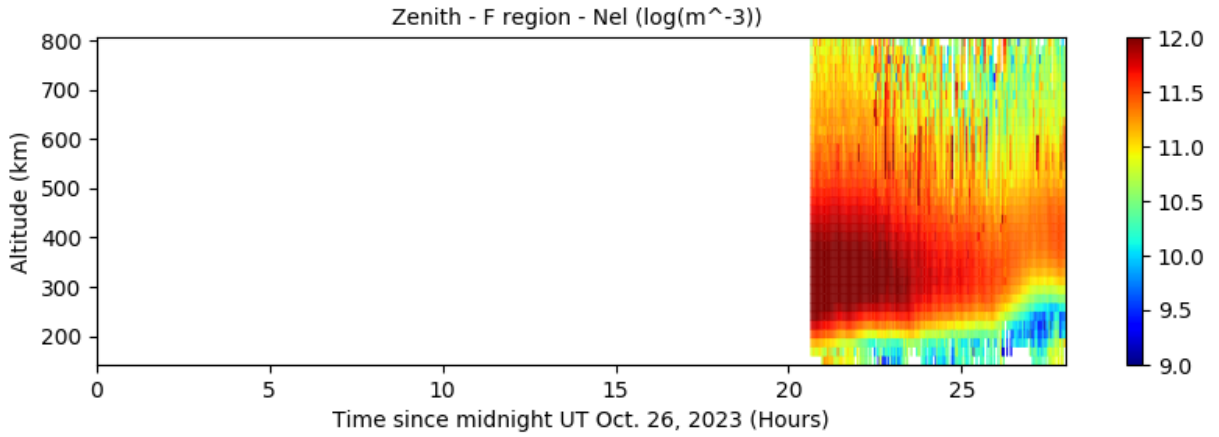


Figure 1.8 MLH-ISR recorded vertical profile of 2023-10-26 [http://cedar.openmadrigal.org/].

In addition to MIT-TEC observations, empirical models also exist to provide complete global electron density profiles. International Reference Ionosphere (IRI) [*D. Bilitza*, 2001; *Dieter Bilitza et al.*, 2022] considers ionosonde observations and is parametrized to the solar and magnetic activities, such as sunspot number, F10.7 index, Ap3 index. Noticeably, IRI has been updated regularly and the latest versions are IRI-2016 and IRI-2020 [*Dieter Bilitza et al.*, 2022]. By filling the data gaps of ISR record electron density profiles, we could provide more useful information for studies on regional ionosphere than empirical models. In Chapter 5, we will introduce the studies on the prediction of ionospheric electron density using neural networks.

1.4 Machine Learning in Ionospheric Studies

"Let data speak for themselves.", which might summarize the growing artificial intelligence (AI) techniques in recent decades. The various domains include the image processing, natural language processing (chatGPT, [*Ray*, 2023]), vehicle automation [*Vishnukumar et al.*, 2017], etc. Research is no exclusion in getting benefits from the development of AI technologies as we witness increasing number of publications using machine learning (ML) algorithms to

address topics in space physics. Conventionally, the physical mechanism for each phenomenon is the priority, where large amount of experimental data could be presented by formulas and be used to refine these formulas. However, in most realistic sceneries, the system might be too complicated to be described by simple equations. Such as the electron densities in ionosphere depends not only on ionization and recombination processes, but also the interaction with mesosphere and thermosphere (mesosphere, ionosphere, and thermosphere coupling, or MIT-coupling [*Laštovička et al.*, 2014; *Qian et al.*, 2014]). Besides, the background neutral densities have a say in affecting the ionospheric electron densities. The IRI [*D. Bilitza*, 2001; *Dieter Bilitza et al.*, 2022] is the empirical model suiting description of the ionospheric electron density on a monthly scale, while lacking the diurnal (day-to-day) variations. Machine learning methods are majorly data-driven, where data themselves include rich information of non-linearity. Machine learning methods use the penalty-mechanism to guide its learning process to reach the optimal model that describes the data distribution, which is particularly useful for forecasting or filling some missing physical parameters.

Various machine learning methods have been applied in space physics. As one of the representative algorithms, deep neural network (DNN [*Goodfellow et al.*, 2016]) extracts the characteristics of large amount of input data by processing them with multiple neural network (NN) layers. Each layer is composed of multiple neurons fulfilling the weighted summation of inputs followed by a non-linear activation as shown in Equation *1.4*, where w_i stands for weight and x_i is the input from previous layer, while b_i is the bias. The non-linear activation function is noted as *f()*, some frequently used examples include, sigmoid, tanh, and ReLU [*Dubey et al.*, 2022; *Narayan*, 1997].

$$y = f(\sum_i w_i \cdot x_i + b_i) \tag{1.4}$$

As an example of full-connected DNN is shown in Figure 1.9, the weighted summation with bias passing through an activation function is highlighted and zoomed in the red box. A DNN consists of three types of layers: input layer, hidden layers (noted "deep" if more than two), and output layer. The input parameters are processed through these layers, eventually reach the output node, i.e. electron density in this work.

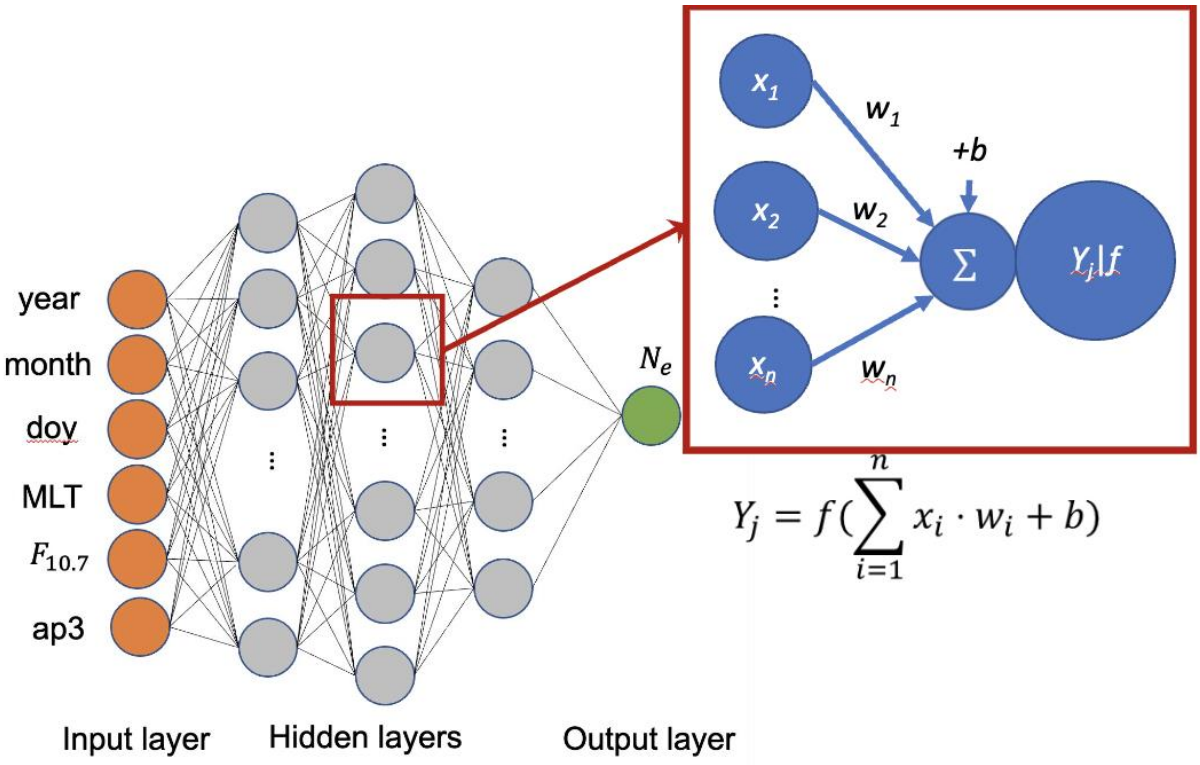


Figure 1.9 Architecture of deep neural network, the weighted summation section is zoomed.

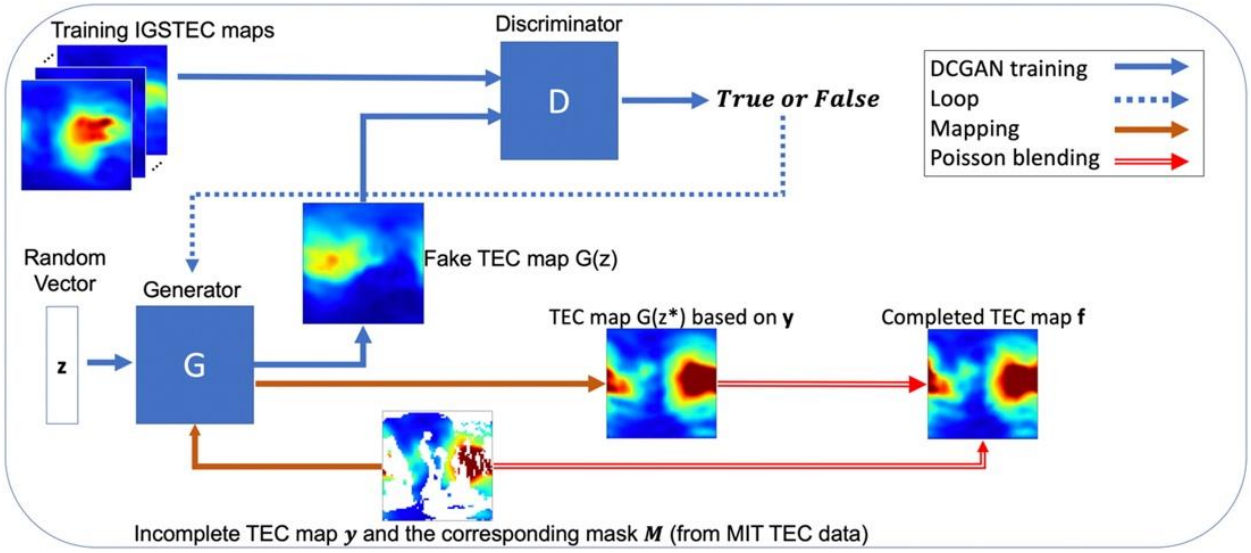


Figure 1.10 Deep convolutional generative adversarial network – Poisson blending (DCGAN-PB) architecture [Pan et al., 2020].

For global total electron content maps a more advanced network called generative adversarial network (GAN) is used. The GAN was proposed to generate artificial data closely resembling the real data [*I Goodfellow et al.*, 2014]. As shown in Figure 1.7, the global TEC maps could be treated similarly to the images of human face, especially when the map is plotted in local time rather than longitude as the day-side solar radiation contributes as the major source for ionization. GAN contains two essential modules, the generator and discriminator. We develop a deep learning model to generate complete global TEC maps with deep convolutional GAN with Poisson blending (DCGAN-PB) [*Pan et al.*, 2020], and the architecture is displayed in Figure 1.10. Briefly, the generator produces fake complete TEC map samples and tends to fool the discriminator, which judges the source of the sample either real or fake. The gambling game pushes both generator and discriminator more and more precise until the trained generator can produce fake samples, so that the discriminator can no longer differentiate them from the real samples. Finally,

the trained generator is used to perform image completion(/inpainting) tasks. Moreover, the particular structure of GAN models does affect the inpainting results. Therefore, we further develop an advanced GAN model to improve over DCGAN-PB. Both TEC map completion GAN models and the electron density prediction model will be presented in detail in the following chapters.

2 METHODOLOGIES

Our focused domain lies in the ionosphere on two major parameters: total electron content (TEC) and electron density (*Ne*). And by walking through the background of ionosphere, the mechanism under the hood to decide TEC and *Ne* is dynamic and complicated, which leads to the stage set for surging machine learning applications.

In the following sections, the methodologies will be introduced and described in detail. Thereafter, the results will be shared and discussed to evaluate the machine learning performance.

2.1 Generative Adversarial Networks

Two essential modules of generative adversarial networks (GANs), the generator and the discriminator, were briefly mentioned in section 1.4. Figure 1.10 has shown the architecture of deep convolutional GAN with Poisson blending (DCGAN-PB). The generator (denoted as G) coins fake graphs (denoted as $G(z)$), and z is the input as a random vector in this specific case). While the discriminator (denote as D) takes the input with no knowledge on its source. Either a fake sample produced by generator or a real case from the training dataset will be sent to the discriminator. And the discriminator works to identify its source and gets penalized when making wrong judgements. Meanwhile, the generator tries to coin more deceitful samples to fool the discriminator. Two modules (D and G) are trained in a competitive way until the discriminator is unable to distinguish the artificial data from the real data, which means that the generator evolves the capability to fake sample resembling the real ones. Then the trained generator serves the important inpainting step as shown in Figure 1.10: an incomplete TEC map y provides the ground truth pixels for the coined sample $G(z)$. The inpainting error, normally root mean squared error (RMSE), evaluates between

the coin sample $G(z)$ and incomplete case y on the overlapping locations. Then same procedure is done with another initial vector z', which leads to a new corresponding error. Generator gets constant feedback from the inpainting error and adjusts the initial vector until a desired number of epochs is exhausted. Based on z^*, the coined sample $G(z^*)$ resembles the ground truth with the lowest possible inpainting error and serves as the cap over the ground truth. Finally, all the procedures form into a complete TEC map f. For the detailed architecture of DCGAN-PB, we followed that of a stable DCGAN learning [*Radford et al.*, 2015] with certain modifications, and the architecture detail is included in APPENDIX A.

$$z^* = \min_{\boldsymbol{z}} \mathcal{L}_C(\boldsymbol{z}|y, M) + \mathcal{L}_p(\boldsymbol{z}) = \min_{\boldsymbol{z}} \|W \odot (G(\boldsymbol{z}) - \boldsymbol{y})\|_1 + \lambda \log(1 - D(G(\boldsymbol{z}))) \quad 2.1$$

$$W_i = \begin{cases} \sum_{j \in \varkappa_i} \frac{1 - M_j}{N} & if\ M_i = 1 \\ 0 & if\ M_i = 0 \end{cases} \quad 2.2$$

Typically, the machine learning models own a loss function to give feedback on model performance and guide the optimization. With the notations introduced above, the loss function of DCGAN-PB is described at Equation *2.1*. As described when introducing the architecture (Figure 1.10), the optimal vector $\boldsymbol{z}^*$ is searched by starting from an initial random vector $\boldsymbol{z}$ and an incomplete TEC map $\boldsymbol{y}$. For the two loss functions, $\mathcal{L}_C$ denotes for context loss and $\mathcal{L}_p$ is the prior loss. Within the context loss calculation $\mathcal{L}_C(\boldsymbol{z}|y, M)$, M stands for the binary mask corresponding to $\boldsymbol{y}$, where the pixel locations of data gaps are marked as zeros and the remaining as ones with available TEC value. While for the weights $\boldsymbol{W}$, i and j are the pixel indices and $\varkappa_i$ is the neighborhood of pixel i with N as the number of neighbor pixels. This design highlights the observed data points closer to the missing data region, while the pixels with missing data or

surrounded by observed data have zero contribution to the weight. Besides, the prior loss describes the discriminator loss for DCGAN training.

$$\min_{f} \iint_{\Omega} |\nabla f - \nabla(G(\boldsymbol{z}^*))|^2 \qquad 2.3$$

Especially for DCGAN-PB, Poisson blending is applied as a post-processing step when the coined sample $G(\boldsymbol{z}^*)$ is capped over the incomplete TEC map. Let Ω notes the gap region of missing data, the ∇ is the gradient operator, and the f is the ground truth outside Ω. Equation *2.3* tries to bridge gradients between the surrounding ground truth pixels and filled pixels of $G(\boldsymbol{z}^*)$. The smoothness is achieved and the mosaic-looking artifacts are largely reduce, which is to be explained in the results section in detail.

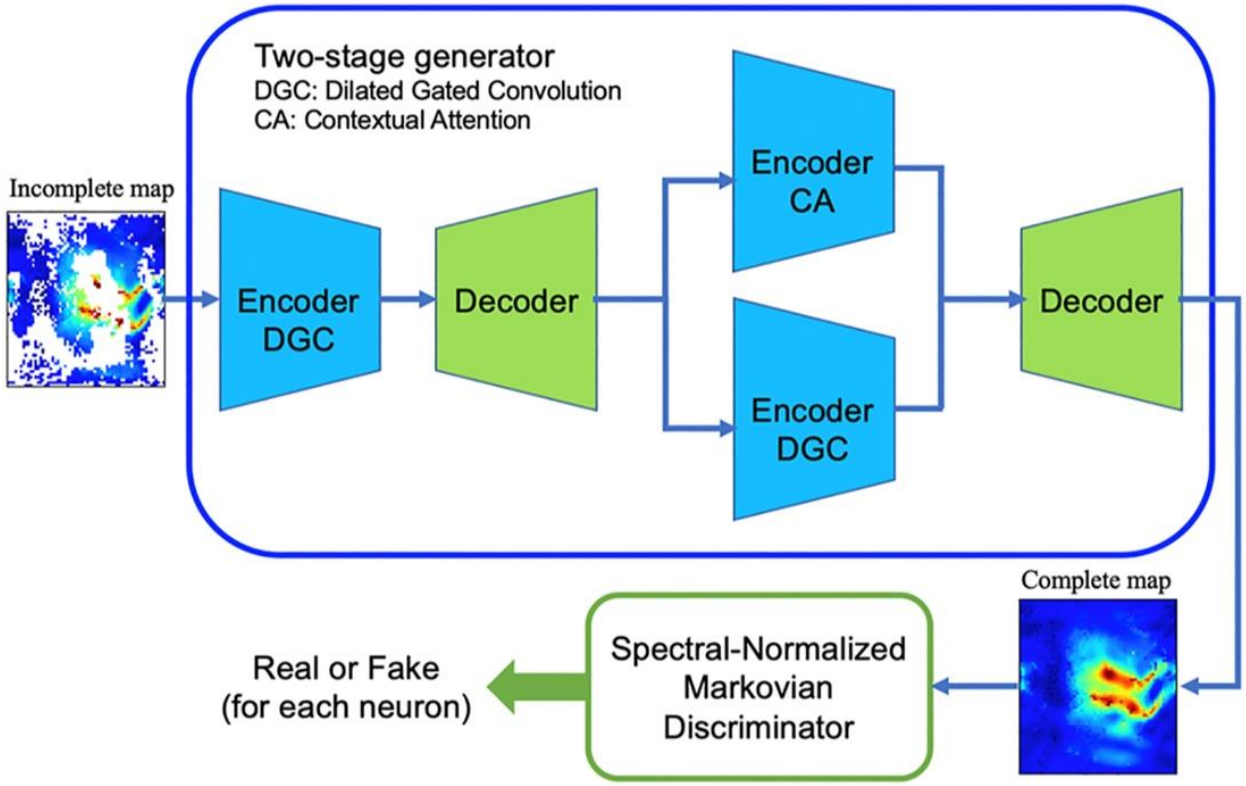


Figure 2.1 The overall model architecture of SNP-GAN [Pan et al., 2021].

Another GAN application, spectrally normalized patch GAN (SNP-GAN) presents a more advanced GAN model with its advantages. As shown in Figure 2.1, the two-stage generator is highlighted with details. The most noticeable part is the "end-to-end" generator, in which the input and output are both TEC map (incomplete global TEC map as the input in SNP-GAN), instead of the vector to map flow in DCGAN-PB (Figure 1.10). Encoder-decoder is a commonly used module to extract necessary characteristics in the latent space, which is defined as the data of different dimensions within the module. Dilated gated convolution (DGC) applies the kernel of sparse version with background pixels in between, and the global features could be extracted compared to the typical kernel. In which the typical kernel extracts regional feature of the image and is confined to the kernel size. If we define missing data of the incomplete map as foreground and the ground truth as background, contextual attention (CA) learns to obtain feature information from the background to fill the foreground.

Before describing the loss functions of SNP-GAN, a complete global map (x) might be covered with a mask (m) with the element-wise product $x \odot m$ and resulting a incomplete global map (z). The coarse network is defined as the first-stage of the generator, which points to the output $G_{coarse}(z, m)$ right after the first encode-decoder module. While the counter-apart part of the generator $G(z, m)$ includes the second encoder-decoder module where two encoders (DGC and CA) are concatenated. The generator loss $\mathcal{L}_G$ is defined below:

$$\mathcal{L}_G = \alpha_1 \times \mathbb{E}_{x \sim \mathbb{P}_x(x)}[|x - G_{coarse}(z, m)| + |x - G(z, m)|] + \alpha_2 \times (-\mathbb{E}_{z \sim \mathbb{P}_z(z)}[D(G(z, m))]) \quad (2.4)$$

the entropy is denoted as $\mathbb{E}$ and $\mathbb{P}$ is the distribution space. While the two weights α_1 and α_2 are adjustable hyper-parameters. Then for the discriminator loss $\mathcal{L}_D$:

$$\mathcal{L}_D = 0.5 \times \mathbb{E}_{x\sim\mathbb{P}_x(x)}\left[ReLU\left(1 - D(x)\right)\right] + 0.5 \times \mathbb{E}_{z\sim\mathbb{P}_z(z)}[ReLU(1 + D(G(z,m)))] \quad 2.5$$

where the rectified linear unit (ReLU) activation function is defined as:

$$ReLU(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases} \quad 2.6$$

2.2 Neural Network Aided with Neural Architecture Search

The neural networks (NNs) have long been applied to model on ionospheric electron densities as we have mentioned in the introduction section, which is one of the most powerful machine learning methods for regression and classification. Usually, the neural network consists of the input layer, the hidden layer(s), and the output layer. Each hidden layer is made of multiple nodes, so called neurons. Each neuron performs a non-linear activation of the weighted sum of outputs from the previous layer. When the number of the hidden layers is equal to or greater than two, the NN is called the deep neural network (DNN) otherwise the single-layer neural network (SLNN). Given the input and output variables x and y, respectively, a DNN model makes prediction as $\boldsymbol{y} = f(\Theta, \boldsymbol{x}|\Lambda)$, where Θ is the trainable parameters (i.e. weights and biases connecting neurons) and Λ is the hyperparameters defining the network structure and training conditions (such as the number of layers, the number of neurons in each layer, dropout, optimizer, learning rate, etc.). If Λ is fixed and the training data are X^{train} and Y^{train}, Θ can be optimized by the following training:

$$\Theta^* = arg\min_{\Theta} loss\left(y^{train}, f\left(\Theta, x^{train}|\Lambda\right)\right), for\left(x^{train}, y^{train}\right) \in \{X^{train}, Y^{train}\} \quad 2.7$$

where "loss" is the loss function measuring the overall difference between the observations and the model predictions on the training data.

However, Equation 2.7 only optimizes on Θ for a fixed network, i.e., fixed Λ. Based on the task and data, the performance of DNN is also dependent on the hyperparameters Λ. Manually tuning these hyperparameters could become tedious and time consuming, and lead to unsatisfactory results such as over-complicated models, long training time, or large test errors if not tuned properly [*Thomas Elsken et al.*, 2017; *T. Elsken et al.*, 2019]. The search algorithms were developed to obtain the optimal solution automatically in a pre-defined hyperparameter space.

Automatic machine learning (AutoML) has become a viral research topic as machine learning is widely applicable in many fields [*Hutter et al.*, 2019]. It enables researchers in the field other than machine learning to build their models more efficiently. Neural architecture search (NAS) [*T. Elsken et al.*, 2019] is one subject of AutoML and aims to search to the best NN for a given task and dataset, whose flow chart is summarized in Figure **2.2**. Reinforcement learning [*Baker et al.*, 2016; *Zoph and Le*, 2016] was first proposed for NAS, followed by gradient methods [*H Cai et al.*, 2018a; *Luo et al.*, 2018], evolutionary algorithms [*Desell*, 2017a; b; *Guo et al.*, 2020; *Real et al.*, 2017; *Suganuma et al.*, 2017], and network morphism [*H Cai et al.*, 2018b; *Thomas Elsken et al.*, 2017; *Jin et al.*, 2019]. NAS aims to find the optimal network structure through the following alternative optimization,

$$\Lambda^* = arg\min_{\Lambda} cost\left(\boldsymbol{y}^{val}, f(\Theta^*, \boldsymbol{x}^{val}|\Lambda)\right), for\ (\boldsymbol{x}^{val}, \boldsymbol{y}^{val}) \in \{X^{val}, Y^{val}\} \quad 2.8$$

$$\Theta^* = arg\min_{\Theta} loss\left(\boldsymbol{y}^{train}, f\left(\Theta, \boldsymbol{x}^{train}\middle|\Lambda^*\right)\right), \text{for}\left(\boldsymbol{x}^{train}, \boldsymbol{y}^{train}\right) \in \{X^{train}, Y^{train}\} \quad 2.9$$

where the data are divided into the training set $\{X^{\text{train}}, Y^{\text{train}}\}$ and the validation set $\{X^{\text{val}}, Y^{\text{val}}\}$. While "*cost*" is the cost function measuring the model prediction error on the validation data $\{X^{\text{val}}, Y^{\text{val}}\}$, and "*loss*" is the loss function measuring the model fitting error on the training data $\{X^{\text{train}}, Y^{\text{train}}\}$ with a fixed Λ^*.

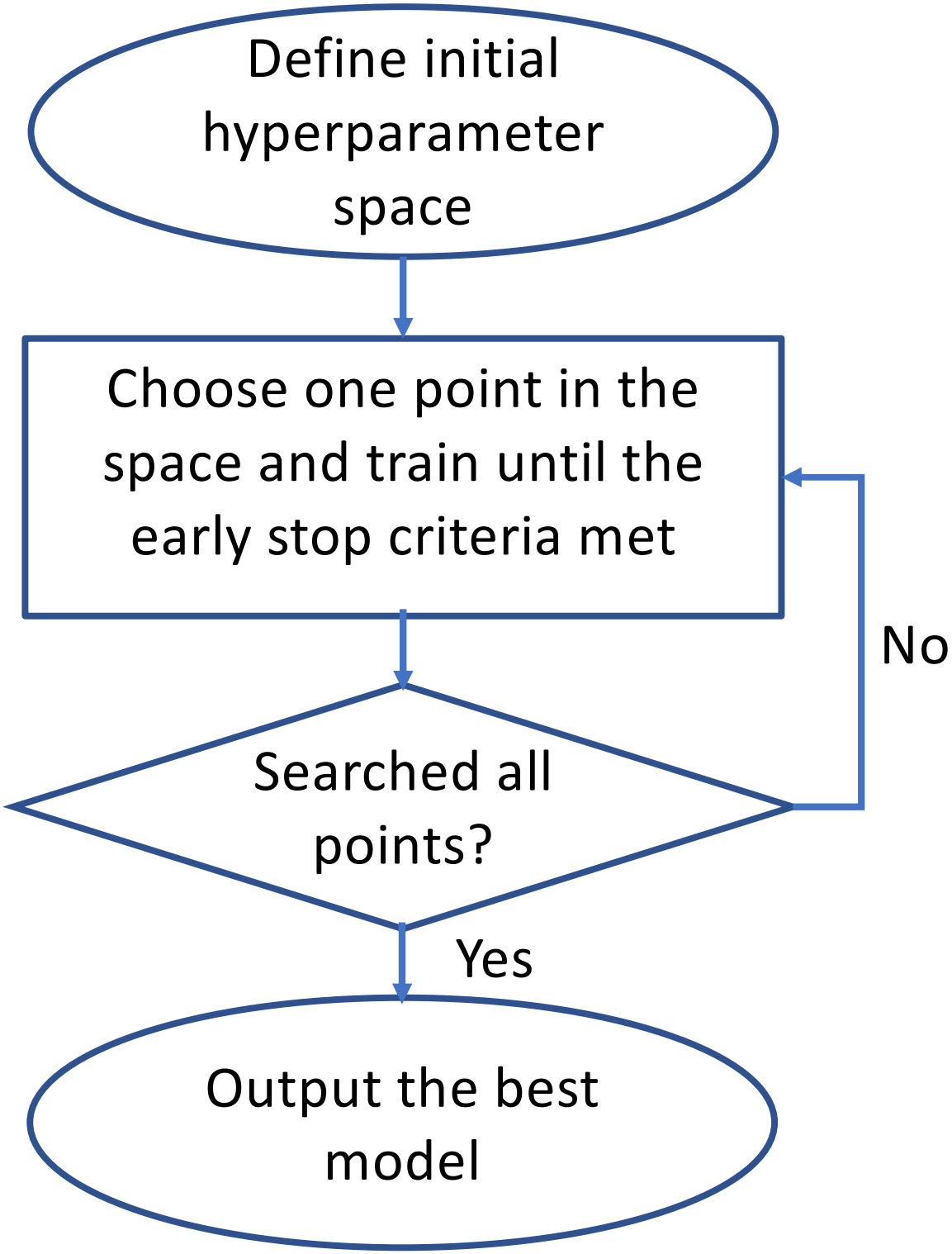


Figure 2.2 *Flow chart of Neural Architecture Search (NAS).*

AutoKeras [*Jin et al.*, 2019] with a high-level user interface is a NAS method based on network morphism, which modifies the NN using the morphism operations, such as inserting a layer or adding a skip-connection. To search the optimal network structure, a hierarchical tree structure is used, whose basic component is the node. For instance, the mother node is an abstract idea of the NN configuration, which is followed by a child node consisting of dense layers, activation layers, normalization layers, etc. The other child nodes include learning rate and training optimizer. Each child node can serve as the parent node for the nodes connected at the next level, and a tree structure is conducted. Finally, the leaf is an end node without any child node. The hyperparameter space defined in Table **2.1** is the result of large number of empirical searches with different combinations. The neuron number in each layer no greater than 64 has already offered decent result for both SLNN and DNN. For DNN, the layer number is refrained to no more than 4 based on the literatures (such as the global 3D *Ne* model to predict density maps mainly for 200 to 400 km altitude [*W Li et al.*, 2021]) and our preliminary trials. The most noticeable is the learning rate search polls. To achieve stable and converging models, larger learning rates fit the SLNNs while DNNs prefers comparatively smaller ones. The reason is that a more complicated neural network structure requires more fine tuning, and hence a smaller learning rate will have a higher chance of leading to a more stable model as judged by the loss curves. However, a lower learning rate does not guarantee a smaller converged loss value. Thus, manual tuning on learning rates becomes undesirable with consideration on the efficiency. Besides, Adam optimizer [*Kingma and Ba*, 2014] is fixed as the training optimizer for all the models which is not explicitly mentioned in the table.

Table 2.1 *Hyperparameter space of NAS. The candidates in each hyperparameter poll are the optimal results of multiple trials. For instance, the single layered architecture prefers a larger learning rate than the deep neural architecture.*

Hyperparameter	Range
Number of layers	SLNN: [1] DNN: [2, 3, 4]
Neuron number	[16, 18, 20, …, 64]
Learning rate	SLNN: 9e-04, 8e-04, …, 1e-04
	DNN: 5e-04, 4e-04, …, 5e-05

Three representative search algorithms in AutoKeras for NAS are: random search, greedy search (GS), and Bayesian optimization (BO). A trial is defined as a round of optimization of Equation *2.8* with a single set of hyperparameter configuration when the early stopping criterion, i.e., no significant improvement of the objective function, is met. Besides, the maximum allowed number of trials is defined as 200 at the beginning. For those three search algorithms, random search randomly picks a hyperparameter configuration without repetition for each trial until the number of trials is reached. Apparently, the random search suffers the computational inefficiency (longer computation time and inferior model performance) in searching the hyperparameter space as there is no mechanism to guide the search directions compared with greedy search [*Thomas Elsken et al.*, 2017; *G Li et al.*, 2020] and Bayesian optimization [*Dey et al.*, 2020]. The greedy search (GS) selects a node with a probability inversely proportional to the number of leaves of that node. The other hyperparameters in the search space will be picked randomly first, then as the previous best trial to form a trial configuration. Therefore, the advantage for the greedy search over the random search is that the search can always return to the best trial when the new configuration does not offer better performance. Each trail of the Bayesian optimization (BO) [*Dey et al.*, 2020; *Snoek et al.*, 2012] consists of a loop of update, generation, and observation. A neural network kernel function is defined to measure the edit-distance between two network structures,

which will enable the Gaussian process-based update of the network architecture. Upper-confidence bound is used for the cost function, whose optimization leads to generation of the next network architecture Λ^*. The observation is to obtain the optimal weights Θ^* for the new network architecture as shown in Equation *2.9*. These three steps repeat until the pre-defined trial number is reached. More details of AutoKeras can be found in [*Jin et al.*, 2019]. To compare the computational efficiency of three search algorithms, some initial runs were conducted by defining a maximum number of trials of 200 and a maximum number of epochs of 200. The random search always used up all 200 trials with MAE inferior to GS and BO, while GS and BO usually reached the optimal solution in less than 100 trials. In addition, both GS and BO averaged around 80 epochs per trial while the random search peaked over 150 epochs per trial by average. Under the current dataset and 200 maximum trials, the BO optimal model of first 50 trials almost led to twice the MAE of GS and a much more complicated architecture ([1024, 128, 16, 1024, 256]) than that of GS ([16, 64, 32]). Thereafter, the greedy search (GS) algorithm is used for all the following experiments.

3 TEC Map Completion using DCGAN and Poisson Blending

3.1 Introduction

Total electron content (TEC) is an important parameter characterizing the ionospheric plasma number density [*Mannucci et al.*, 1998]. Especially, the dynamic TEC value can be used to identify traveling ionospheric disturbance, which can be caused by geomagnetic storm events. TEC also influences the communication between satellites and the ground stations and has been included as a parameter in the space weather forecasting [*Afraimovich and Astafyeva*, 2008; *Azeem et al.*, 2015; *A J Coster et al.*, 2017; *Jakowski et al.*, 2002; *Lyons et al.*, 2019; *McGranaghan et al.*, 2018; *McGranaghan et al.*, 2017; *S-R Zhang et al.*, 2017]. The measurement of TEC values requires the receivers on the Earth surface. However, due to the limited coverage, for example, barely no receivers on the oceans that takes about 70% of the Earth surface, the observational data of TEC are incomplete for a global map. For example, if the observed TEC values are mapped onto a 2-D geographical map, a huge amount of missing data in the ocean parts will occur. Massachusetts Institute of Technology (MIT)-TEC is a large public database recording huge amount of TEC observations since 2000, currently involving 6,000+ global navigation satellite system (GNSS) receivers. On the other hand, international GNSS service (IGS) has been providing IGS Global TEC (IGS TEC) data since June 1998, with the help of cooperation among five members of IGS Ionosphere Associate Analysis Centers (IAACs) [*M. Hernández-Pajares et al.*, 2009]. The complicated processing, including analysis and validation algorithms, leads to the final IGS TEC maps. First, all five IAACs apply their distinct algorithms to fit the observations from a few hundred GNSS sites to form the 2-D TEC maps. Then the results are compared based on different parameters, such as vertical TEC (VTEC) performance, slant TEC (STEC) performance, and delay code biases (DCBs) estimations, to get the best-fitted map. A large amount of manual

work is needed to finally produce a single complete global TEC map. To alleviate the large amount of manual work needed in TEC map completion, automatic processing methods are in demand to provide a timely and reliable solution.

Artificial intelligence (AI) has been a blooming research topic for the last decade. It shows great potential on replacing human beings in challenging and time-consuming works, such as language translation, gaming, and driverless automobile [*Stallkamp et al.*, 2012]. AI techniques have been adapted in space science, mostly for predication and pattern recognition. For example, a support vector machine (SVM) model was developed to predict high-latitude ionospheric phase scintillation [*McGranaghan et al.*, 2018], and a deep learning model was used to extract auroral key local structures from large amount of auroral images to lessen intense labor of human experts [*Q Yang et al.*, 2019]. The recent fast growth of AI has been spurred by deep learning algorithms. Deep neural network (DNN) is one of the most representative deep learning algorithms, which disperses and discovers the properties of a huge amount of input data by processing them in multiple neural network (NN) layers [*Ciregan et al.*, 2012]. Each layer is filled with neurons as a weighted sum and usually a nonlinear activation of outputs from a previous layer. Finally, the output layer makes either classification or estimation based on the problem in hand. Deep convolutional neural network (DCNN), one important category of DNN, uses convolutional operations in hidden layers to extract hierarchical feature patterns in data. It can take 1-D or multidimensional arrays, such as 2-D images, as the input and adds convolution operation in each layer of DNN [*Dumoulin and Visin*, 2016; *Krizhevsky et al.*, 2012]. DCNN requires little hand-engineering of data, for example, images, compared to conventional machine learning methods because it learns the best filters to extract most useful features in data through the training, which is particularly suitable for automatic completion of TEC maps. In 2014, the generative adversarial

network (GAN) has been proposed [*I J Goodfellow et al.*, 2014], which can generate artificial data closely resembling the real data. GAN includes two DNNs: the generator and the discriminator. The generator produces the fake data from the random initialization to mimic the real data, while the discriminator is used to distinguish the fake data from the real data. The back propagation is used to improve the generator and the discriminator performance in a competition between two DNNs. The successful training of GAN leads to the artificial data that are hardly distinguishable from the real data [*Yeh et al.*, 2017] and, therefore, a model capable of representing the physical situation. The deep convolutional GAN (DCGAN) [*Radford et al.*, 2015] has been utilized to successfully generate fake human face images [*Kim*, 2016]. An improved version of DCGAN has been applied to recover the artificially masked human face images and showed decent recovery results [*Brandon*, 2016]. One of the reasons that neural works perform so well in their applications is the large database. Usually, we are targeting at a class of thing, for instance, a bike or road signs, and they have their properties, the distribution of pixels. The more available data are, the better neural networks would get by optimizing the weights repeatedly. Moreover, neural networks are highly customizable. With customized number of convolutional layers, it can dig out the hidden features. Deep learning is the process of coding and recording the empirical like human beings.

Recently, we proposed a regularized DCGAN (RDCGAN) for data completion of TEC maps [*Chen et al.*, 2019]. The RDCGAN work is aimed to complete the MIT TEC maps using DCGAN. Since there were no complete MIT TEC maps for training, RDCGAN was developed with an additional reference discriminator (i.e., RDCGAN has two discriminators and one generator) where the generated TEC maps are regularized by the complete IGS TEC maps to improve the TEC completion performance. In this work, we aim to learn IGS TEC completion process through DCGAN, which is considerably simpler than RDCGAN. This work is not a direct

extension of RDCGAN work but rather a parallel development of machine learning methods for automatic space physics data processing.

Although DCGAN can generate the missing TEC values that have the correct content aligned with the surrounding available observations, those generated TEC values may have some baseline shift and not join the surrounding observations continuously. Thus, directly overlying the DCGAN generated TEC values on the missing regions in the original TEC map leads to the "mosaic" artifact. To address this issue, we adapt Poisson blending [*Gangnet and Blake*, 2003] after DCGAN, which blends the values according to the gradients around the gaps instead of direct use of the recovered values from DCGAN. Poisson blending has wide applications such as the reconstruction of 3-D surfaces from point samples robustly with fine details [*Kazhdan et al.*, 2006]. As shown by our experiments, this postprocessing method can lead to significantly improved completion performance. We also investigate the influence of training data size and missing data pattern on the proposed DCGAN model and compare DCGAN with conventional image inpainting methods.

3.2 Data and Experiments

All IGS-TEC data are downloaded from https://cdaweb.gsfc.nasa.gov/pub/data/gps/ [*Manuel Hernández-Pajares*, 2004] for the time period from 1 June 1998 to present. Each IGS TEC map is averaged over a time interval of 2 hour in Universal Time (UT) and resized from a 73×71 matrix to a 64×64 matrix by using 2-D cubic interpolation [*Shepard*, 1968], corresponding to a spatial resolution of $5.6°$ in longitude and $2.8°$ in latitude. The abnormal data out of the range of usual TEC values (>999 absolute TEC units [TECu]) are discarded. The total number of available IGS TEC maps for training and test is 91,579.

3.2.1 Random Mask Experiment

In the first experiment, we demonstrate the impact of the amount of training data and Poisson blending on the completion performance of DCGAN using random masking. DCGAN was first trained using (1) 2 years of IGS TEC maps (January 2010 to December 2011) for the same time period as our previous work [*Chen et al.*, 2019] and (2) about 12 years of IGS TEC maps (June 1998 to December 2010) for one solar cycle. Three TEC maps at 2012-10-07 08:00 UT, 2015-03-17 12:00 UT, and 2019-05-18 16:00 UT were selected as test data, representing medium, high, and low solar activities, as shown in Figure 1.7. Three types of random masking are tested as shown in Figure 3.2: (1) 15% missing data with 2×2 minimum gap size (the minimum gap size means the smallest size of any missing data area; each pixel denotes $5.6°$ in longitude [x] and $2.8°$ in latitude [y]); (2) 40% missing data with 2×2 minimum gap size; and (3) 15% missing data with 4×4 minimum gap size. In addition to the completed TEC maps for those three representative times shown in Figure 1.7, the root mean squared error (RMSE) using randomly selected 360 IGS TEC maps in the year of 2012, 2015, and 2019 is used as a quantitative measure in Equation *3.1*:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\left(f_i - \widetilde{f}_i\right)^2}{N}} \qquad 3.1$$

where f_i is the original TEC value, $\widetilde{f}_i$ is the recovered TEC value, and N is the total number of missing TEC values. In addition to RMSE, we applied Bland–Altman (BA) plot [*Altman and Bland*, 1983] as another criterion to quantify TEC completion performance. The BA plot is used to

evaluate the agreement between the original TEC values and the filled TEC values using different methods. The x and y coordinates on the BA plot for the ith missing TEC value are calculated as:

$$(x_i, y_i) = \left(\frac{f_i + \widetilde{f}_i}{2}, f_i - \widetilde{f}_i\right), \text{for } i = 1,2,\dots,N \qquad 3.2$$

thus, the BA plot highlights the distribution of filled TEC values from the true ones with the bias and 95% confidence intervals labeled as horizontal lines.

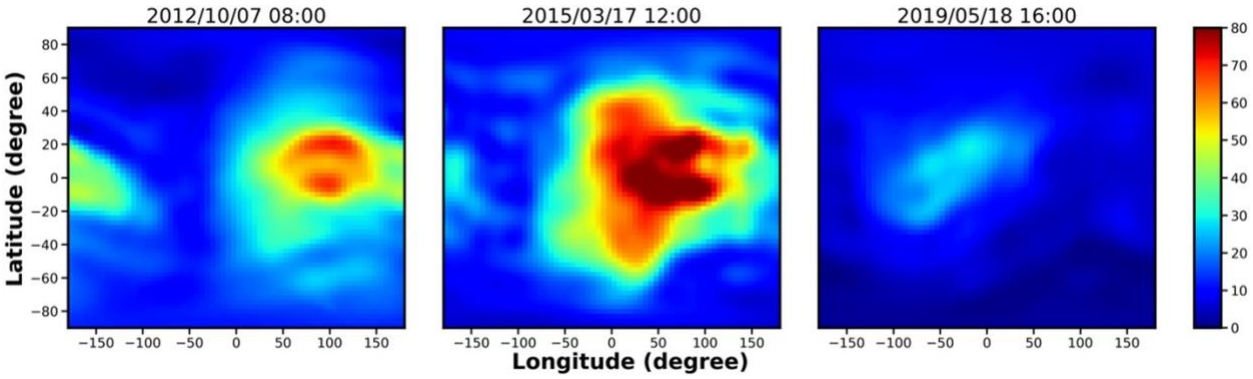


Figure 3.1 The test IGS TEC maps at three time points: 2012-10-07 08:00 UT (medium solar activity), 2015-03-17 12:00 UT (high solar activity), and 2019-05-18 16:00 UT (weak solar activity) [Pan et al., 2020].

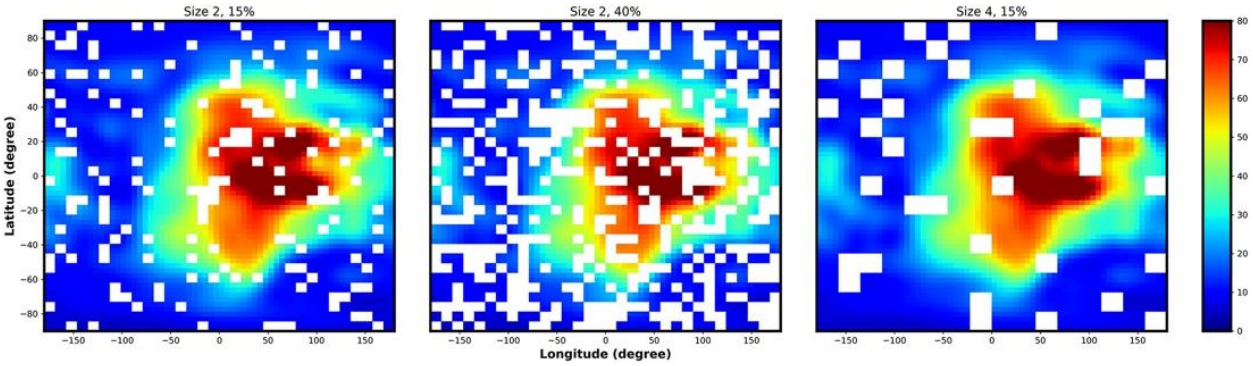


Figure 3.2 The different random masking methods (overlaid on the TEC map at 2015-03-17 12:00 UT), from left to right: (1) 15% missing data with 2 × 2 minimum gap size; (2) 40% missing data with 2 × 2 minimum gap size; and (3) 15% missing data with 4 × 4 minimum gap size [Pan et al., 2020].

3.2.2 MIT-TEC Mask Experiment

In the second experiment, we extract a more realistic mask using the MIT-TEC data (obtained from the Madrigal database, http://www.openmadrigal.org), so-called MIT-TEC mask as shown in Figure 3.3. The missing data, mostly in the oceans and polar regions, are about 52% of the total data (white regions in Figure 3.3). It represents a more realistic and challenging case for TEC map completion as the missing part takes a large portion of the map and is continuous. In this case, IGS-TEC data about 18 years were used for DCGAN training and Poisson blending ("DCGAN-PB") to get the final completed maps.

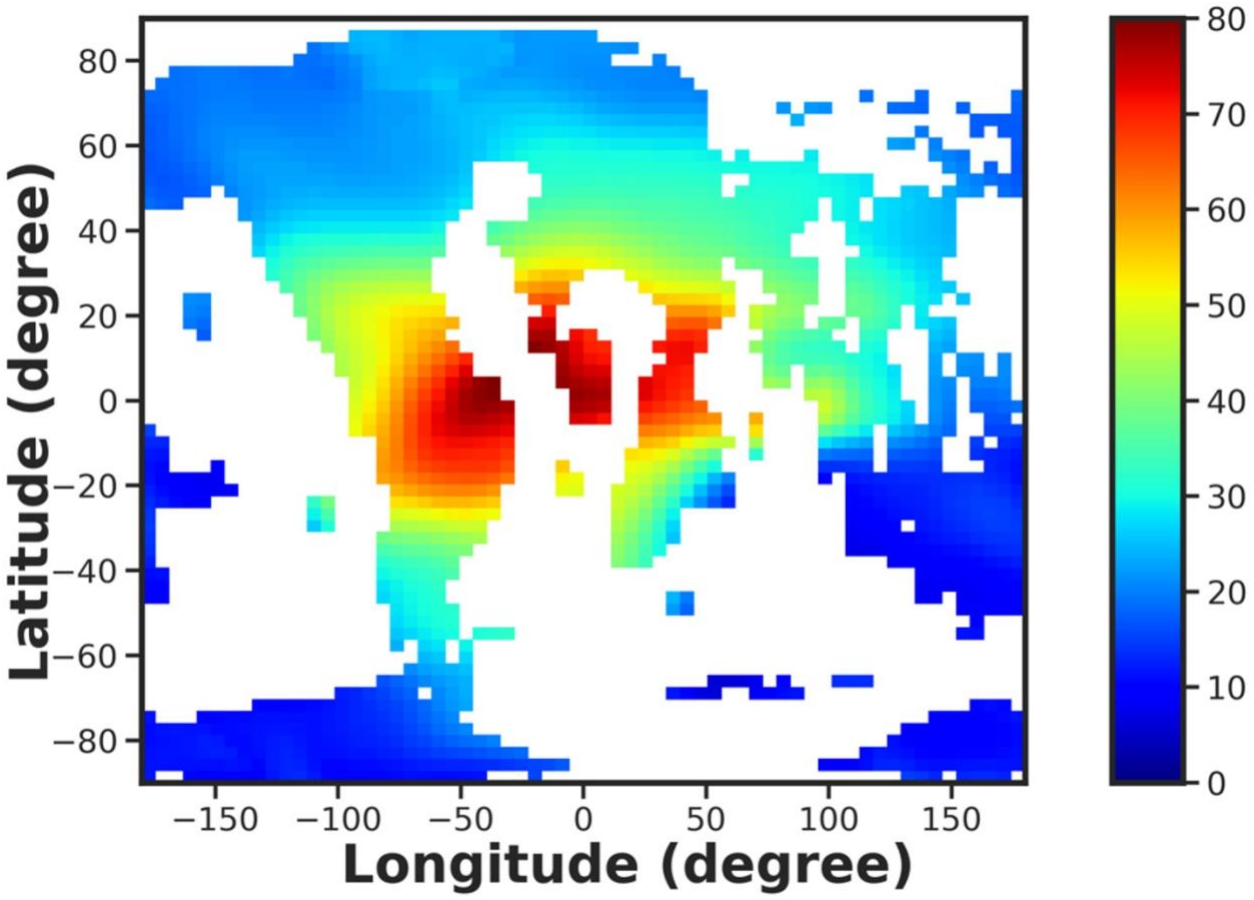


Figure 3.3 The MIT-TEC mask overlaid on the TEC map at 2000-07-14 14:00 UT [Pan et al., 2020].

For comparison, two traditional image inpainting techniques, TELEA [*Telea*, 2004] and Naïve-Stokes (NS) [*Bertalmio et al.*, 2001], are also applied to complete TEC maps. Starting from the peripheral intensities of the data gap, these two inpainting methods recover the gap value accordingly from external to internal of the gap through applying the weights on the surrounding ground truths of the pixel to be filled. Both methods have shown decent image inpainting performance [*Bertalmio et al.*, 2001; *Telea*, 2004].

Since it is more meaningful for space weather application to evaluate the performance of different methods under high solar activities than low solar activities, three representative time points are selected: 2000-07-14 14:00 UT, 2003-11-01 10:00 UT, and 2015-03-17 12:00 UT. In the following section, the completed TEC maps and RMSE from our DCGAN-PB method, TELEA and NS have been compared. In order to avoid the appearance of test cases in the training set, a ten-fold cross-validation of DCGAN-PB is applied. In each fold, TEC maps of 18 years are used for training and the remaining 2 years for testing. We select three cross-validation sets to reconstruct TEC maps at those three time points.

3.2.3 Cross-Validation Experiment

A criterion for a success ML model is the variety of the training dataset. As the solar radiation dominates the ionization process, the solar activity contributes significantly to the TEC in a global scale. One solar cycle has approximately eleven years, and the database is selected with care based on that fact. In both GAN models, we take the IGS-TEC data from 1999 to 2018, a total of 20-year, as the whole dataset. Different durations of training data are taken to test the effect of Poisson blending. Cross-validation is the strategy we applied, and the 20-year has been classified

into high solar activity and low solar activity based on the annual mean F10.7 of 100 solar flux unit (sfu) as shown in Table 3.1. Besides, in each test year 50 IGS-TEC maps are selected to calculate the errors.

Table 3.1 High/low solar activity years and assignment of 10 sets for tenfold cross validation [Pan et al., 2020].

Set index	1	2	3	4	5	6	7	8	9	10
High (F10.7>100 sfu)	1999	2002	2014	2001	2000	2015	2003	2004	2013	2012
Low (F10.7≤100 sfu)	2005	2018	2009	2006	2008	2011	2017	2016	2010	2007

3.3 Results

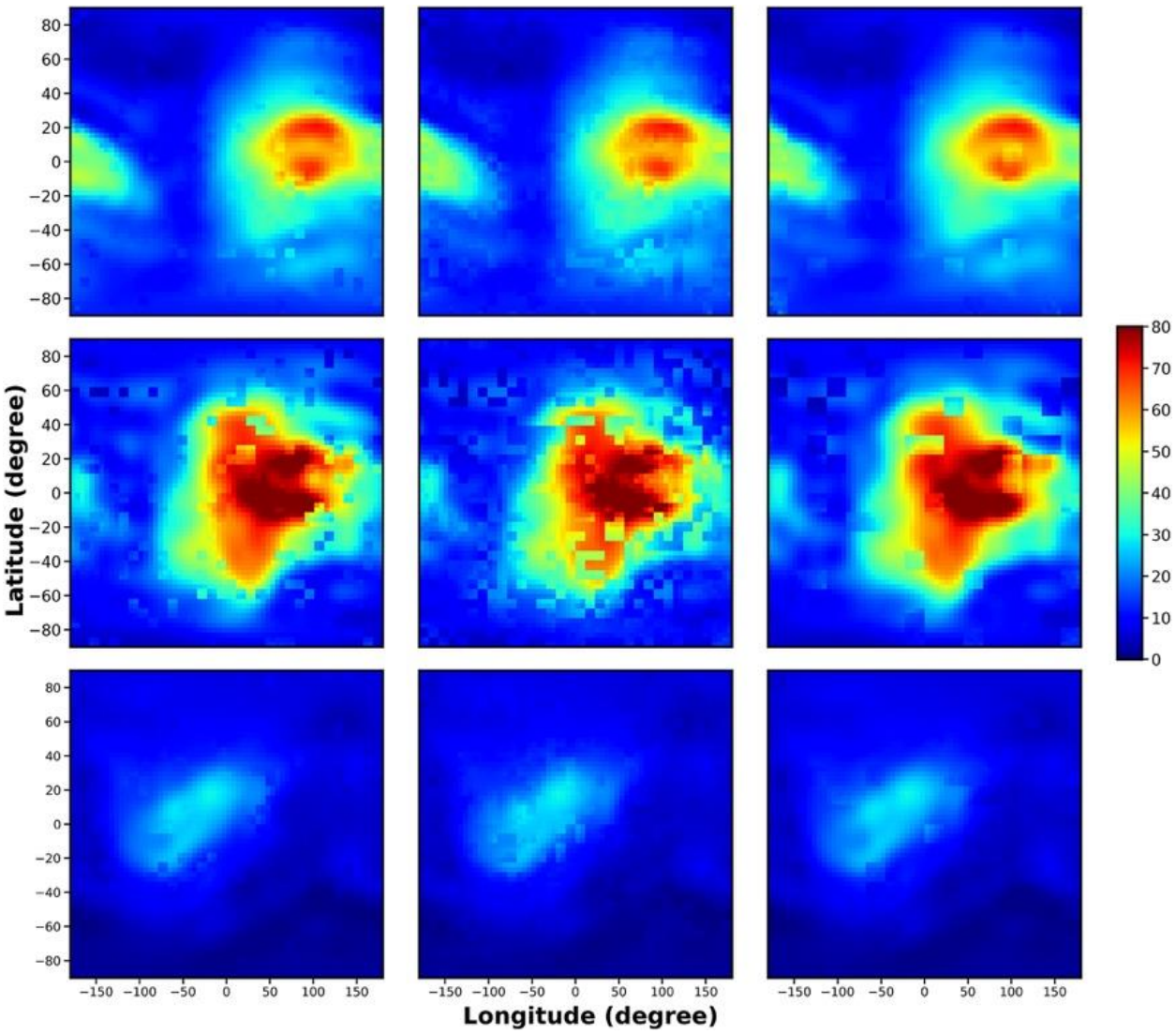
3.3.1 Random Mask Results – Influence of Training Data and Poisson Blending

Figure 3.4 shows the completed TEC maps of DCGAN without PB using 2-year training data (2010–2011) under different solar activities shown in Figure 3.1 and different random masks shown in Figure 3.2. DCGAN fills the TEC data gaps well in some places, for example, the high TEC value region in 2012 maps (the top row). In addition, the larger portion of missing data, the worse the completed map. However, the completed maps suffer the mosaic artifacts, for example, the middle row for high solar activity in 2015, as the TEC values generated from DCGAN in the missing data region have some baseline shift from the surrounding region. One possible reason for the worse performance in 2015 than in 2012 is that the training data (2010–2011) contain most low solar activity maps, thus leading to a trained model lack of generalization of high solar activities.

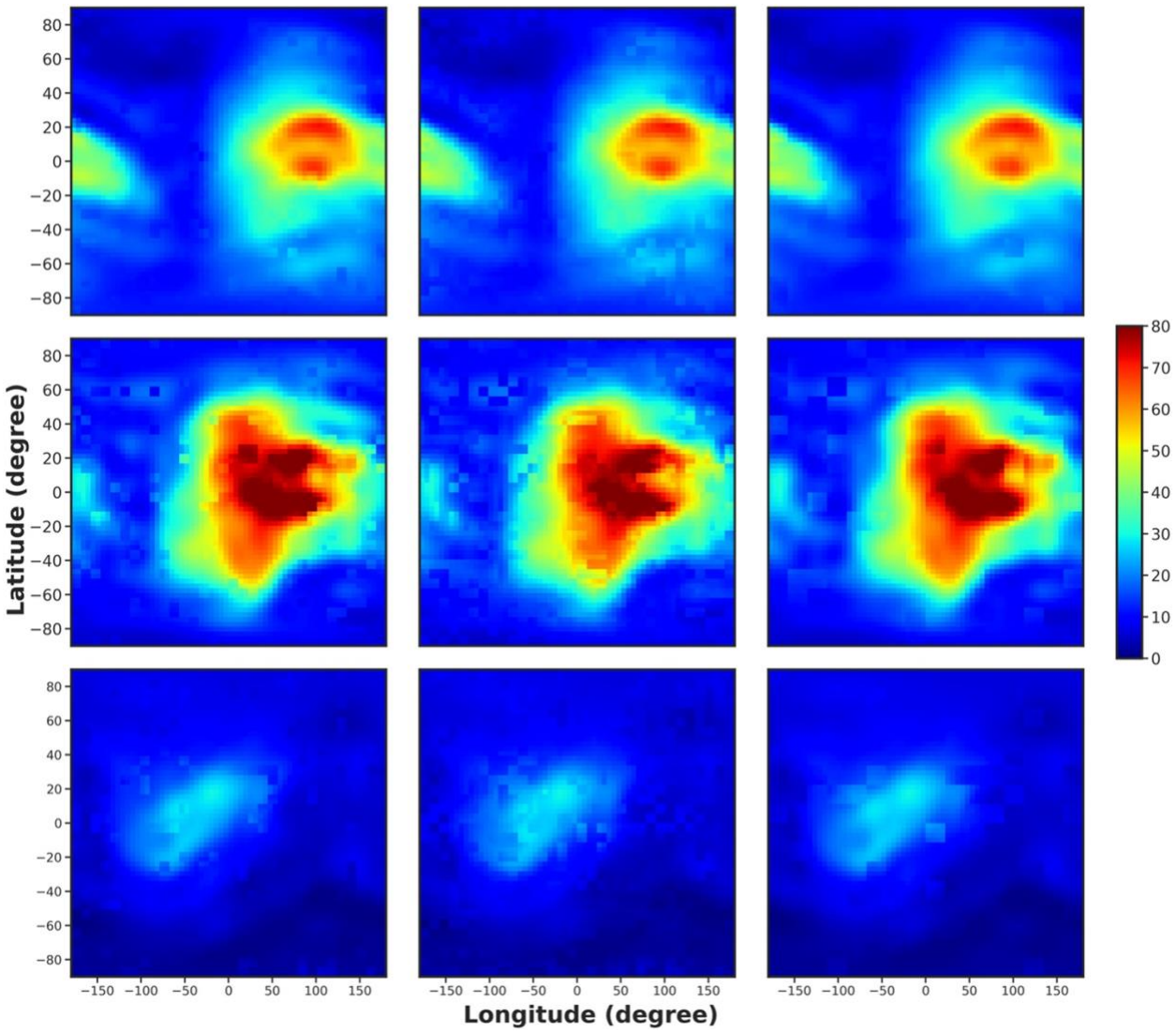
As we increase the training data from 2 years to about 12 years, the completed TEC maps are notably improved as shown in Figure 3.4b. Since one solar cycle is about 11 years, the training data including all the cases of TEC maps such as high solar activity years (with more large

geospace storms and F10.7 > 100 sfu) and low solar activity years (F10.7 ≤ 100 sfu) lead to a much-improved completion performance as shown in comparison of Figure 3.4(b&a). Similarly, the completed maps with less missing data and smaller gaps are better than those with more missing data and larger gaps. Nevertheless, the mosaic-looking noises still exist in the completed TEC maps. The mosaic artifacts can be seen in Figure 3.4(a&b), where unrealistic changes of TEC values are obvious in the missing data region. These artifacts likely generate abnormal disturbances if feeding into global circulation models (GCMs), which may affect space weather forecasts depending on the specific method used for data assimilation. However, the detailed quantitative investigation is out of the scope of this work and will be conducted in our future work.

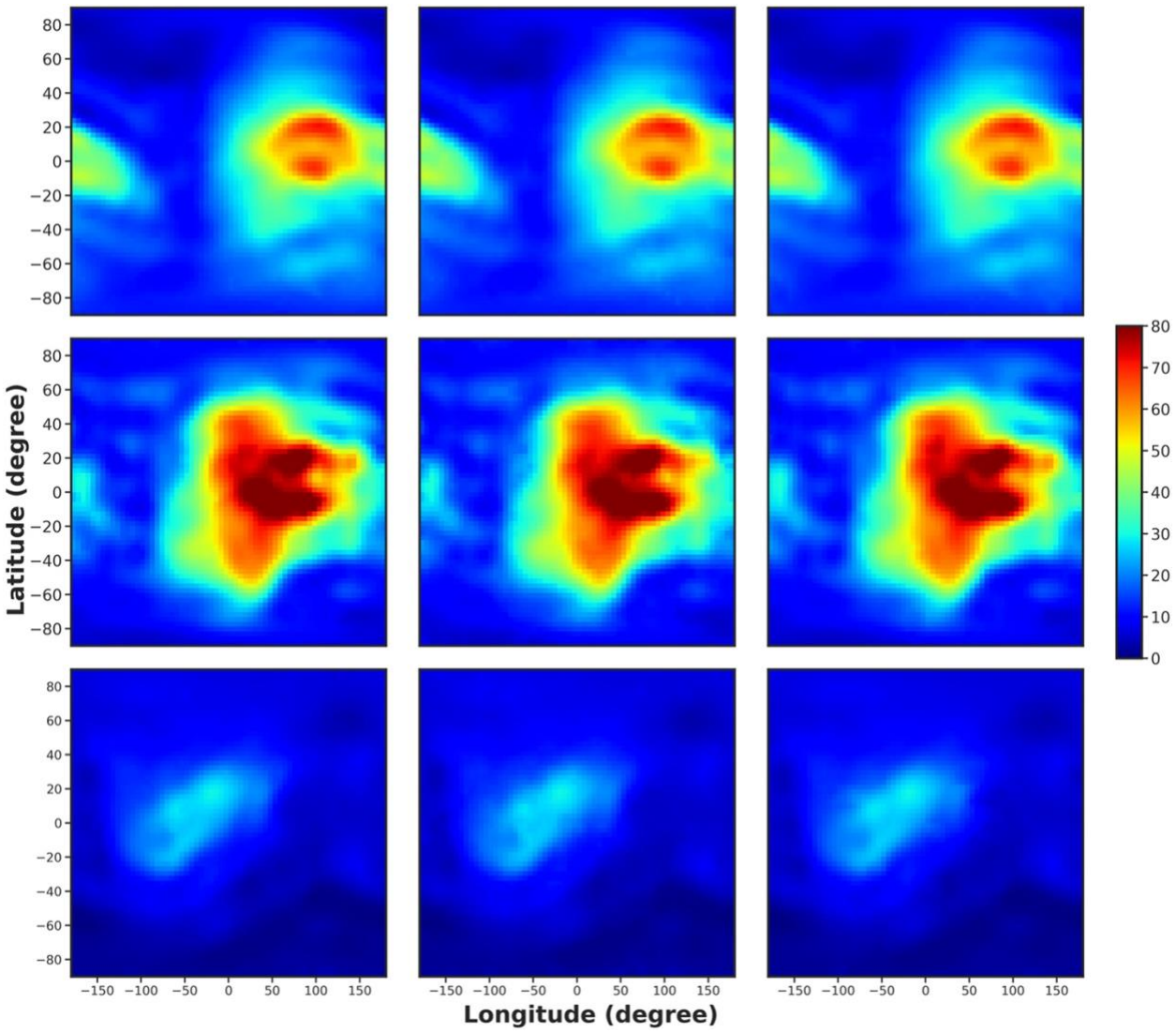
With postprocessing of PB, the mosaic-looking artifacts are effectively removed as shown in Figure 3.4c. The completed TEC maps reassemble the original IGS TEC map under all conditions, which demonstrates the superior image completion performance of DCGAN with large training data and PB.



(a) 2-year training data without Poisson blending



(b) 12-year training data without Poisson blending



(c) 12-year training data with Poisson blending

Figure 3.4 The completed TEC maps from DCGAN without and with Poisson blending. The 2-year training data (a, 2010-2011) compares with 12-year training data (b&c, 1998-2010). From top to bottom of each subplot: 2012-10-07 08 UT, 2015-03-17 12 UT, and 2019-05-18 16 UT. And three different mask styles from left to right: (1) 15% missing data with 2×2 minimum gap size; (2) 40% missing data with 2×2 minimum gap size; 15% missing data with 4×4 minimum gap size [Pan et al., 2020].

To quantitatively evaluate the performance of three completion procedures, the root mean squared error (RMSE) values are summarized in Figure 3.5. In general, more training data lead to smaller RMSE values (~50% decrease), that is, better recovered TEC values, and the additional PB further brings down RMSE values dramatically (more than 50% decrease). It is interesting to

note the abnormality that RMSE with 2-year training data in 2019 is slightly better than that with 12-year training data. The possible reason is that the solar activities were low in 2010, 2011, and 2019. DCGAN trained with 2-year data (2010–2011) works better to fill TEC maps with low solar activities in 2019. Nevertheless, the more general model trained with 12-year data (1998–2010) still yields low RMSE values in the low solar activity case (<0.73 TECu with PB). While the 2-year model produces high RMSE in the medium (2012) and high (2015) solar activity years, the 12-year model reduces these errors by half. This demonstrates the importance of training data on deep learning-based data completion.

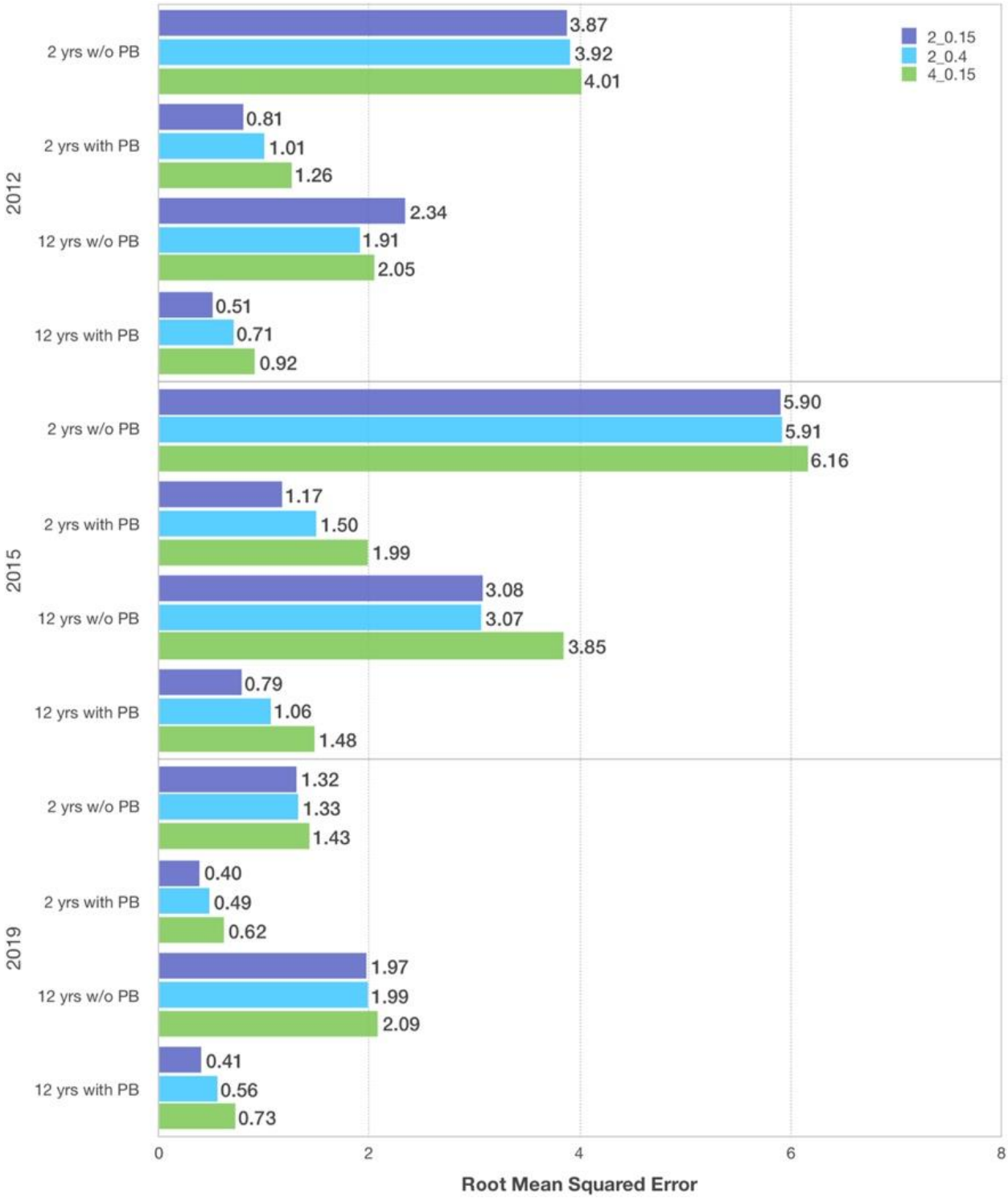


Figure 3.5 Root mean squared error (RMSE) of the recovered TEC values compared to the original values from IGS-TEC maps for randomly selected 360 maps in the year of 2012, 2015, and 2019. The unit is TECu. "2 yrs": DCGAN training with 2-year data; "12 yrs": DCGAN training with 12-year data; "2_0.15": 15% missing data with 2×2 minimum gap size; "2_0.4": 40% missing data with 2×2 minimum gap size; "4_0.15": 15% missing data with 4×4 minimum gap size; "PB": Poisson blending [Pan et al., 2020].

3.3.2 MIT Mask Results – Realistic Comparison with Two Other Completion Methods

The original IGS TEC maps at 2000-07-14 14:00 UT, 2003-11-01 10:00 UT, and 2015-03-17 12:00 UT are shown in the first row of Figure 3.6. The MIT-TEC mask shown in Figure 3.3 is applied to them, and TELEA, NS, and DCGAN-PB are utilized to fill the data gap to reconstruct the global map. As shown in Figure 3.6, DCGAN-PB outperforms the other two methods and produces the completed TEC maps closest to the original IGS-TEC maps, particularly notable in the high TEC value regions (hot colored areas). Quantitative results of RMSE are shown in Figure 3.7. The average RMSE for TELEA and NS is about 7–9 TECu, while DCGAN-PB effectively suppresses the error to 3–4 TECu. The performance for the MIT mask (Figure 3.7) is worse than that for the random masks (Figure 3.5) because the MIT mask has much larger gaps and more missing data than the random masks, thus posing a harder completion task. To further look into the distributions of recovered TEC values from different completion methods, the recovered TEC values ($\widetilde{f}_i$) versus the original values (f_i) are plotted out in Figure 3.8. The green line denotes the perfect recovery, that is, $\widetilde{f}_i = f_i$. In general, all three methods produce the data points around this line, reflecting a decent recovery of missing TEC values. The plots of DCGAN-PB are less spread than those for the other two methods. After linear regression, the fitted lines using these points are drawn in red. The deviations from the green line (perfect recovery) are noticeable for TELEA and NS, while DCGAN-PB leads to almost overlapped red and green lines. One possible reason PB works well is that the IGS TEC maps are based on some sphere-harmonic functions and thus vary smoothly. Our proposed DCGAN-PB outperforms TELEA and NS in terms of both bias and variance of recovered TEC values.

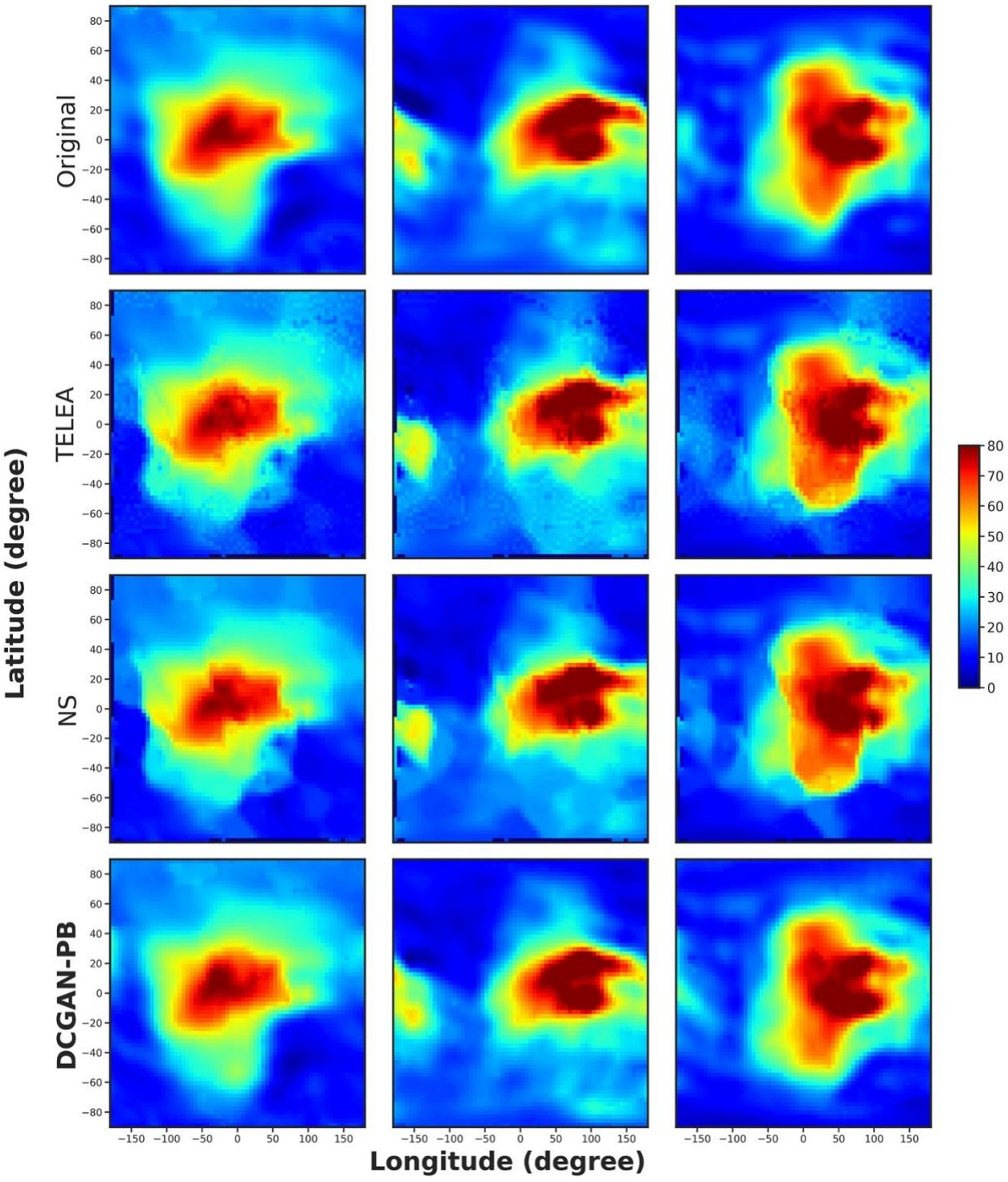


Figure 3.6 Top: the original IGS TEC maps at 2000-07-14 14:00 UT, 2003-11-01 10:00 UT, and 2015-03-17 12:00 UT on (from left to right). The completed TEC maps (with MIT-TEC mask) from two conventional inpainting methods (TELEA and NS) and DCGAN-PB method are shown in the second, third, and bottom rows [Pan et al., 2020].

Figure 3.7 RMSE of the recovered TEC values compared with the original IGS TEC maps from three different TEC map completion methods (TELEA, NS, and DCGAN-PB) for three high solar activity cases [Pan et al., 2020].

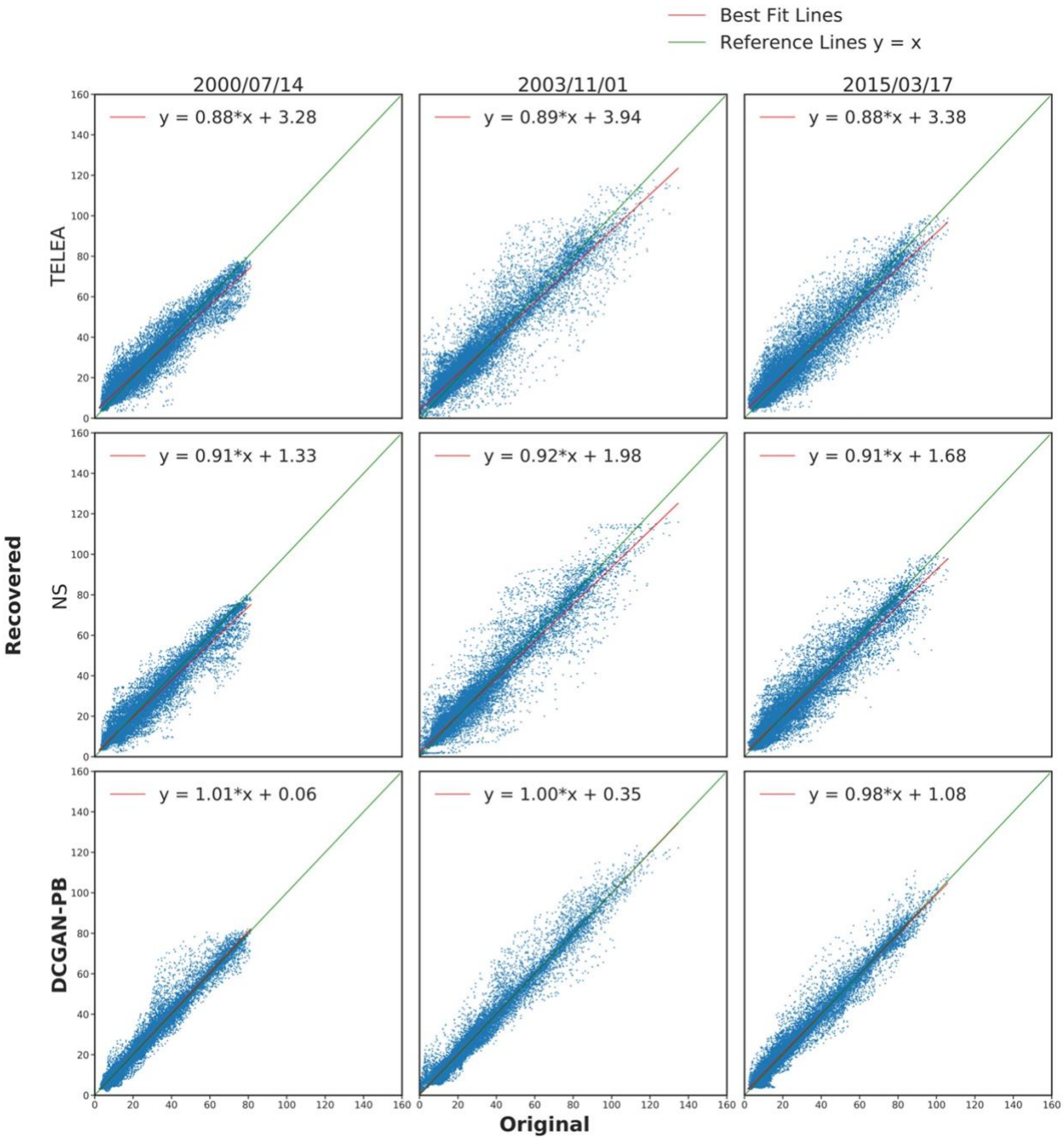


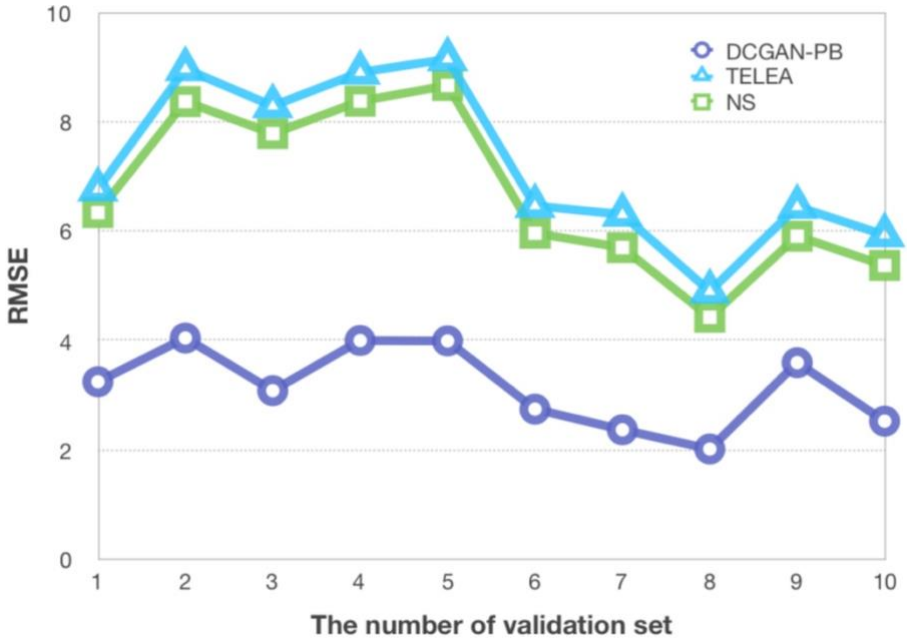
Figure 3.8 The recovered TEC values (vertical) versus the original TEC values (horizontal) for TELEA, NS, and DCGAN-PB methods [Pan et al., 2020].

3.3.3 Ten-Fold Cross-Validation Results with MIT-TEC Mask

The ten-fold cross-validation (CV) results are shown in Figure 3.9 and Figure 3.10. The former displays the RMSE separately for high (a) and low (b) solar activity years. Similar to results in Figure 3.7, DCGAN-PB leads to smaller RMSE values (less than 4.04 TECu for high solar activity years and about two TECu or less for low solar activity years) than TELEA and NS in all low and high solar activity years. The fluctuation of RMSE values for DCGAN-PB is remarkably smaller than those from TELEA and NS as well. The statistical values (mean $\pm$ standard deviation [SD]) of RMSE from 10 sets for TELEA, NS, and DCGAN-PB are 7.21 ± 1.49 TECu, 6.69 ± 1.49 TECu, and 3.16 ± 0.74 TECu for high solar activity years and 3.60 ± 0.91 TECu, 3.11 ± 0.92 TECu, and 1.65 ± 0.32 TECu for low solar activity years, respectively. DCGAN-PB reduces the average RMSE by about 50% and the variation by more than 50%. Figure 3.10 shows the BA plots of all the test cases of cross-validation for TELEA, NS, and DCGAN-PB. The mean difference is −0.41 TECu for TELEA, 0.62 TECu for NS, and −0.31 TECu for DCGAN-PB. The 95% confidence interval (± 1.96 SD) is [−11.79, 10.97] TECu for TELEA, [−9.79, 11.03] TECu for NS, and [−5.32, 4.71] TECu for DCGAN-PB. DCGAN-PB has the least mean difference of −0.31 TECu (about 25% improvement over TELEA) and the least confidence interval of 10 (more than 50% reduction compared to TELEA and NS). Note that the large differences of TELEA and NS are caused by the edge pixels as shown in Figure 3.6. If we removed these edge pixels, the 95% intervals are [−9.38, 7.60] TECu for TELEA, [−7.24, 7.58] TECu for NS, and [−4.74, 4.31] TECu for DCGAN-PB, which still demonstrates the significant improvement of DCGAN-PB over the other two methods. The BA plots are consistent with RMSE results and demonstrate the superior completion performance of DCGAN-PB. The advantage of the proposed method lies on exploiting and extracting the fundamental features from large training dataset, while the traditional inpainting

methods are limited by information from one single image. Based on our cross-validation results, if the global average RMSE around three TECu (around 10% relative error) and 95% of filling values falling in ± 5 TECu deviation can satisfy the particular application requirement, our DCGAN-PB model may provide sufficient accuracy. Nevertheless, the caveat is that this type of deep learning models is still at the early stage of development and needs more substantial investigation in order to be deployed for space weather forecast applications.

(a) High solar activity years



(b) Low solar activity years

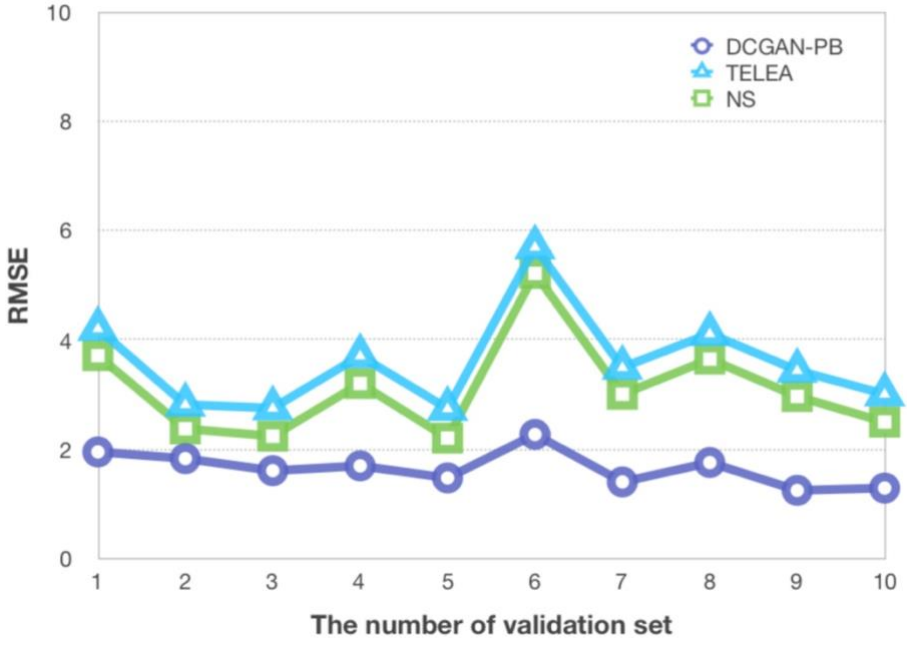


Figure 3.9 RMSE for 10 cross-validation sets by TELEA, NS, and DCGAN-PB methods for (a) high solar activity years and (b) low solar activity years [Pan et al., 2020].

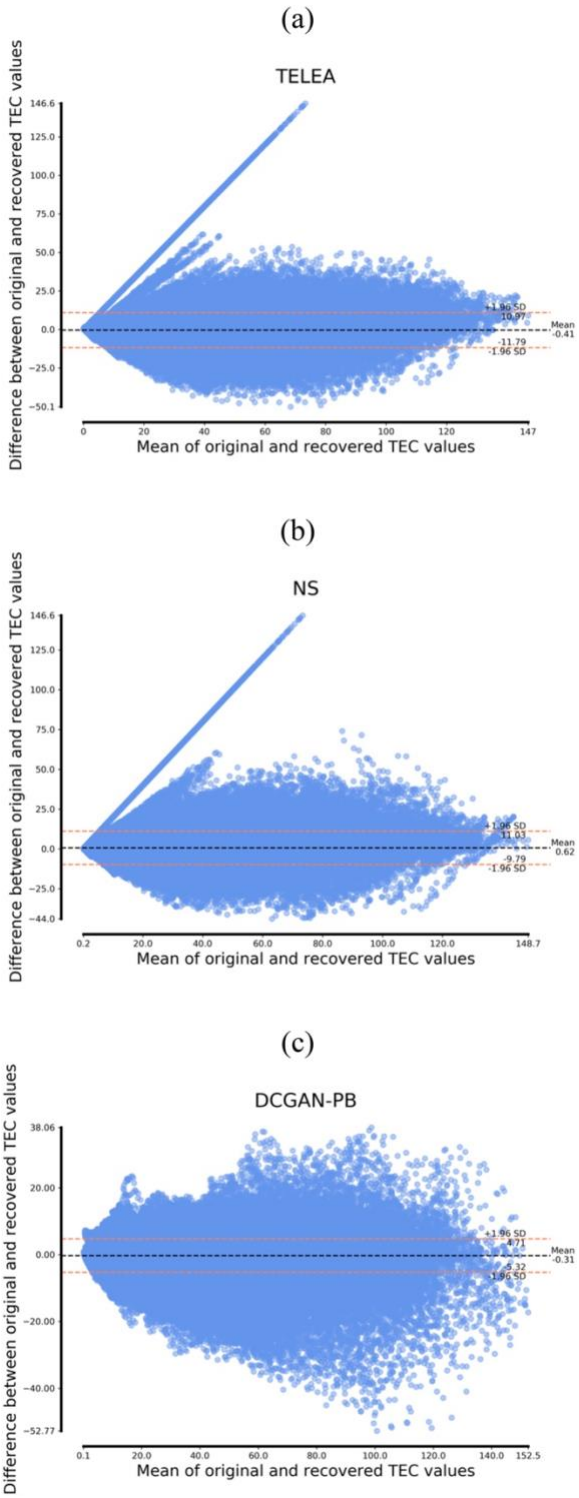


Figure 3.10 The BA plot of three different automatic TEC completion methods among all the test cases in the tenfold cross-validation: (a) TELEA; (b) NS; and (c) DCGAN-PB [Pan et al., 2020].

3.3.4 Discussions

Our initial Poisson blending results have some extreme abnormal values at the edges of the TEC maps. As mentioned by Mahmoud Afifi [*Afifi and Hussain*, 2015], the traditional Poisson blending technique suffers potential smudges if the blended region is at the edges of the image since there are no image values to use beyond the image domain. We have overcome this issue by assuming a cyclic pattern of the TEC maps in the longitude direction (so the left and right edges of TEC maps are essentially connected together).

The test data are separated from the training data for more than 6 months, and the memory of ionosphere is not an issue since the memory of ionosphere is typically less than 1 month. For cross-validation results, there are test data within a month from the training data. The memory effect may influence the final result. We conducted the additional calculation of RMSE of DCGAN-PB by excluding test data in January and December to represent results without forward and backward memory effect. The BA plots for the high solar activity years and the low activity years show the good agreement between two sets of results with mean differences of −0.07 TECu for the high years and −0.01 TECu for the low years and all individual differences falling inside 95% interval, except for one low solar year close to the interval edge. The memory effect seems to have negligible impact on our results.

It is difficult to provide the theoretical uncertainties for deep learning models although the generative models hold the potential to reason the uncertainty, which is beyond the scope of this work. Nevertheless, the mean and SD of RMSE of filled TEC values of DCGAN-PB from the cross-validation experiment are 3.16 ± 0.74 TECu and 1.65 ± 0.32 TECu, for the high solar activity year and the low solar activity year, respectively. Based on additional Bland-Altman

analysis, the bias of DCGAN-PB is about −0.31 TECu, and the 95% of the differences are between −5.32 to 4.71 TECu, much improved over the other two automatic completion methods.

In our current work, we used 2-hr temporal average and spatial resolution of 5.6° in longitude and 2.8° in latitude for TEC maps. Therefore, the short-lived and small-scale features, such as polar cap patches, that are shorter than the temporal resolution and smaller than the spatial resolution in this work would be difficult to recover. Since our goal is to fill the gaps in the global large-scale TEC maps, the investigation of transient and small-scale phenomena is out of the scope of this study. Nevertheless, the deep learning methods, such as DCGAN, may be worth investigating for those phenomena using regional high temporal and spatial resolution TEC data.

The λ value used in the current DCGAN-PB model (Equation *2.1*) is 0.05. To investigate the influence of the context loss, we have chosen the 2-year model to test with different λ values (0.0005–0.9) since the mosaic artifacts are prominent in this case. The RMSE results for 2012-10-07 08:00 UT and 2015-03-17 12:00 UT only change slightly with different λ values (usually around or less than 10%, except for 2012-10-07 with 4×4 minimum gap size). However, even for the largest differences on RMSE on 2012-10-07, the filled TEC maps using different λ values are hard to discern any visual difference of the mosaic artifacts. Therefore, the λ parameter balancing the contribution of the context loss and the prior loss works well in the tested range of 0.0005–0.9.

For DCGAN training, it is time consuming and needs parameter tuning for satisfactory results. To save time, in this work, we use a relatively small TEC matrix 64 by 64. The training time with 18-year TEC images with 200 epochs is about 1 day 4 hr using a single NVIDIA Titan V GPU card. The completion takes about 12 s without PB and 39 s with PB for 4,000 iterations.

Furthermore, the tuning of DCGAN structure and hyperparameters is empirical; thus, no optimal settings are guaranteed. We tried different learning rates and found that the learning rate of 0.00002 yielded an expected trend of training losses of three models (2, 12, and 18 years). The 2-year model converges quickly and has a dip at 100 epochs. Thus, we used 95 epochs for this model. For the 12- and 18-year models, although the training losses move in the right direction, it seems that the fluctuation of each batch update is much larger than the 2-year model and more epochs than 100 are needed to further improve the training performance. Therefore, 200 epochs are used in this work for the 12- and 18-year model training. We also investigated the errors (RMSE) versus epochs for the 2-year model. The test error follows the trend of the training error with slight larger values. A slight upward trend at the end of the test error curve was observed, which indicates the over-training might start to occur. The optimal choice of the learning rate and the number of epochs along with other hyperparameters are worth thorough investigation in future studies.

In this work, we focused on the impact of the training size on TEC map completion performance. Our results indicated that the larger training data size, the better the learning performance for the current network structure. We also empirically found that the learning rate of 0.00002 and the λ value of 0.05 yielded a good completion performance for the proposed DCGAN-PB model. Since the model training is very time consuming, the hyperparameter tuning in this work is coarse and may not be optimal. The optimization of the network architecture and hyperparameters may be addressed by the new techniques, such as AutoGAN to dissect GAN into modules (e.g., conv layers and pooling layers) and to test the performance with different combinations and complexities in a smaller scale [*Gong et al.*, 2019; *Salimans et al.*, 2018]. It would be our future work to apply these new techniques to optimize our DCGAN model.

Nevertheless, our results demonstrate promising performance of the current model that is superior to the traditional methods.

The success of the learning methods is built on the assumption that the training and test data are independently sampled from the identical distribution. However, the TEC value distribution in each map is highly dependent on the solar and geomagnetic conditions. In general, the larger the training set is, the greater the chance the training sample would be from the same distribution as the test sample. Indeed, for the three test TEC maps (2012-10-07 08:00 for medium solar activity, 2015-03-17 12:00 UT for high solar activity, and 2019-05-18 16:00 UT for weak solar activity), we conducted the Kolmogorov-Smirnov (KS) two-sample test between them and the training maps. For 2-year training data, the null hypothesis cannot be rejected in 6, 0, and 0 cases for three test maps (i.e., the cases with the same distribution between the training map and the test map), respectively, while for 12-year training data, the numbers increase to 11, 1, and 2. If the Bonferroni correction was applied, the numbers for the KS null hypothesis are 180, 9, and 0 for 2-year training data and 933, 221, and 295 for 12-year training data. These results demonstrate that the larger the training dataset is, the greater chance it could match the distribution of the test data.

The comparison with simulation results based on first-principle physics models, such as GITM [*Ridley et al.*, 2006] and TIE-GCM [*Richmond et al.*, 1992], would be worth future investigation although the simulation results from these models usually have differences from the ground truth. Such a simulation and comparison effort are beyond the scope of this work. However, the physics-based models may be combined with deep learning to improve the task-based performance. For example, the physics-guided neural networks (PGNNs) add the output of a physics model of temperature and water density/depth relationship as an input in addition to

common environmental parameters for lake temperature prediction [*Daw et al.*, 2022]. PGNN with prior physics knowledge showed better generalization than neural networks alone. Along this direction, the proposed deep learning framework may be improved by adding physical constraints (in Equation 1) for better completion performance.

DCGAN-PB along with our previous work R-DCGAN [*Chen et al.*, 2019] provides an automatic option for the completion of TEC maps through deep learning of the existing filled TEC maps. The filled TEC maps can not only provide references and inputs for physics models for space weather prediction applications, but also be used to study traveling ionosphere disturbances (TIDs) during the geomagnetic events. The comparison of deep learning-based methods and first-principle physics models will strongly improve our understanding and prediction of ionospheric disturbance. Furthermore, the developed deep learning tools can greatly expedite the filling process for new TEC observations, which is important to space weather forecast as the timely output of the complete maps is needed for (nearly) real-time applications in GPS and high-frequency (HF) communications. However, the caution must be taken to avoid the artifacts for learning-based methods, such as mosaic artifacts introduced by DCGAN. If these artifacts affect space weather forecasts to an unacceptable level, the techniques to remove these artifacts, such as Poisson blending in DCGAN-PB, have to be applied.

DCGAN-PB has a simpler network structure than R-DCGAN [*Chen et al.*, 2019], where the former is to mimic IGS TEC completion process and the latter is to fill MIT-TEC maps using IGS TEC as a reference. Although there is a slight difference on two models' purposes, both of them can be used to fill TEC maps. A thorough comparison is worth investigating in near future. If the similar performance is achieved by two models, the simpler model would be preferred since

the runtime is shorter and the tunable parameters are fewer that lead to a fast and robust implementation.

4 TEC Map Completion through a Deep Learning Model: SNP-GAN

4.1 Introduction

Total electron content (TEC) is the quantitative measurement of column electron content in the ionosphere. The study of TEC is important for various applications in space weather, such as identification of the geospace storm effect on plasma density gradients, traveling ionospheric disturbances (TIDs), and etc. [*AJ Coster et al.*, 2003; *Lyons et al.*, 2019; *Nishimura et al.*, 2020; *Tsugawa et al.*, 2007; *S-R Zhang et al.*, 2019]. The measurement of TEC is done using Global Positioning System (GPS) [*Makela et al.*, 2000]. Meanwhile, a statistical interpolation method, Kriging (as known as Gaussian process regression), was used in the Wide Area Augmentation System (WAAS) to estimate the vertical delay at each ionospheric grid point [*Sparks et al.*, 2010]. There are two databases providing global TEC data. The Massachusetts Institute of Technology (MIT) computes GPS receiver biases, and has been collecting and processing TEC observations (MIT-TEC) with high spatial and temporal resolutions since 1998 from over 6,000 observations all over the world [*Vierinen et al.*, 2016]. Due to the limited coverage of receiver stations, which communicate with satellites to measure the TEC data [*Mannucci et al.*, 1998], it presents the raw incompleteness of MIT-TEC maps which curb the studies over the ocean region. On the other hand, the International Global Navigation Satellite System Service (IGS) uses TEC data over 500 observations and complicated data processing methods to fill the data gaps, largely on the ocean part, to obtain complete global IGS-TEC maps. It involves large amount of manual efforts in the processing and communications between IGS Ionosphere Associate Analysis Centers [*M. Hernández-Pajares et al.*, 2009]. Due to its small number of receivers and smoothing processes, mesoscale (100–1,000 km) structures are often lost in complete IGS-TEC maps. An automatic data

filling method for the raw MIT-TEC maps is desired to provide another useful source of complete TEC maps in addition to IGS-TEC.

Since the TEC maps can be treated as 2D images, the recent advance in the image inpainting using deep learning methods can be exploited to fill missing data in the raw TEC maps. Conventionally, the image inpainting methods aim to utilize the relationship among pixels in a single image or a limited available set of images in hand [*Barnes et al.*, 2009; *Bertalmio et al.*, 2001; *Telea*, 2004]. One issue of these conventional solutions is the limitation on available data to infer the missing part. Usually, they rely on the rule that the closer pixels have stronger relationship than those far apart. However, the performance deteriorates greatly for the data gaps that are large and irregular [*Bertalmio et al.*, 2001; *Telea*, 2004]. Image completion based on Generative Adversarial Networks (GANs) [*I J Goodfellow et al.*, 2014] and deep convolutional GAN (DCGAN) [*Radford et al.*, 2015] has achieved impressive performance, for example, in human face reconstruction [*Brandon*, 2016]. It is worth mentioning that the generator in DCGAN is trained to generate images that are similar to real ones from a randomly initialized vector, and the incomplete input image only provides the context ground truth for the final iterative optimization for completion. Another issue with image inpainting is how to deal with the irregular shape of the missing data. Among several methods addressing this issue, partial convolutions (PConv) [*G Liu et al.*, 2018] masked and renormalized the convolution to be conditioned on only valid pixels. An automatic mechanism was included to update the mask for the next layer as part of the forward pass. PConv was further improved by learnable bidirectional (forward and reverse) attention maps, where the latter allowed the decoder of U-net to concentrate on filling missing data in the irregular gap [*Xie et al.*, 2019]. The spectrally normalized patch-based GAN (SNP-GAN) [*Yu et al.*, 2019] generalizes partial convolution through gated convolution. Since global and local GANs are not

applicable to irregular gaps, a patch-based GAN loss was proposed by applying spectrally normalized discriminator on dense image patches so that a multiple-stage architecture including a coarse network and a refinement network [*Miyato et al.*, 2018] could be utilized. Different from the previous DCGAN models, these deep GAN models dealing with irregular image gaps can achieve the end-to-end image completion using the incomplete image and its corresponding gap mask as the model input. Moreover, a contextual attention module used in the generator, which borrows the available background information, can lead to better inpainting performance.

Deep learning-based image inpainting has been adapted to handle the issue of incomplete TEC map or to improve the global TEC prediction ability recently. With an additional reference discriminator using IGS-TEC maps, R-DCGAN shows good completion performance for MIT-TEC maps with data gaps [*Chen et al.*, 2019]. In another work, Poisson blending following DCGAN (DCGAN-PB) has achieved excellent inpainting results of IGS-TEC maps for both small and dispersive gaps and large and continuous gaps [*Pan et al.*, 2020]. The pix2pix image translation model [*Isola et al.*, 2017] was adopted to develop a DeepIRI model [*Ji et al.*, 2020] to obtain improved International Reference Ionosphere (IRI) TEC maps. Using 11-year IGS-TEC training data, a U-net generator of DeepIRI can greatly improve the global TEC maps generated by IRI, particularly ionospheric peak structures. Recently, a method incorporating the temporal information in TEC map completion has been proposed [*Sun et al.*, 2022]. The so-called Video Imputation with SoftImpute, Temporal smoothing and Auxiliary data (VISTA) method was built on the conventional matrix completion algorithms with an additional temporal term to enforce the temporal consistence and/or an auxiliary data constraint to enforce the similarity to interpolation of spherical harmonic functions. VISTA was shown to better preserve the TEC structures than the conventional matrix completion methods based on sparsity and low-rank assumptions.

DCGAN-based completion methods have been very successful for the TEC maps completion. However, since they aim to build a model (generator) to generate TEC maps from the initial random vector without referring the incomplete TEC map context, the generated map has to be transformed to the closest encoding in the latent image manifold using the context and prior losses based on the incomplete map [*Yeh et al.*, 2017]. This transform not only takes substantial time due to iterative optimization, but also leads to additional uncertainties of final filled maps. In this work, we adapt SNP-GAN for TEC map completion to address the issues with DCGAN-based completion methods. The comparison with DCGAN-PB through cross-validation shows that SNP-GAN can achieve better quantitative completion results and faster completion than DCGAN-PB. Furthermore, we conduct case studies of geomagnetic storm events in March 2013 and February 2016 using the trained SNP-GAN model to reconstruct the MIT-TEC maps, which shed light on the potential of SNP-GAN for TEC map completion to study large and mesoscale TEC structures. It is worth noting that since IGS-TEC is one of widely used community sources for complete TEC maps, we used it as the training data for deep learning models that can fill the incomplete MIT-TEC maps. These deep learning models learn the data filling processes from IGS-TEC and apply them to fill the missing data in MIT-TEC, while keeping the original available data intact. Therefore, no IGS-TEC data were used as ground truth during MIT-TEC map completion.

4.2 Data and Experiments

All the data sets used in this study are MIT-TEC (http://cedar.openmadrigal.org/) and IGS-TEC (https://cdaweb.gsfc.nasa.gov/pub/data/gps/). The raw MIT-TEC maps have a spatial resolution of 1° in longitude by 1° in latitude with a temporal resolution of 5-min, and the IGS-TEC maps have the resolution of 5° in longitude by 2.5° in latitude with an optional temporal

resolution of 2-hr, 1-hr, or 15-min. A 2D interpolation was applied to the raw IGS-TEC maps with 2-hr cadence to reshape them into 64×64 images for training of low-resolution deep learning models and with 15-min cadence to reshape them into 128×128 images for training of high-resolution deep learning models. The MIT-TEC maps were re-binned (to the mean) to a spatial resolution matching that of trained models, unless otherwise stated. As the trained models (such as the SNP-GAN model architecture in Figure 2.1) do not take the temporal relation into consideration, the time cadence of MIT-TEC was re-binned into 10 min for quantitative calculation and kept 5 min for qualitative map viewing. We focused on the low-resolution models in Results and provided some preliminary results of the high-resolution models in Discussion. The IGS-TEC maps from 1999 to 2018 were used for training of different GAN-based inpainting models, while the test was conducted on IGS-TEC maps excluded from the training data set for the cross-validation experiment or on MIT-TEC maps for the case studies. For quantitative evaluation, we used the MIT-TEC masks in longitude (equivalent to local time, LT) by latitude (MIT-TEC LT masks) (see Figure 4.1 for example) for cross-validation of IGS-TEC maps and a random brush mask for the case studies of MIT-TEC maps. The root mean squared error (RMSE) of the IGS-TEC cross-validation (CV) test sets and the MIT-TEC test sets serves as the quantitative criterion to evaluate the reconstruction performance.

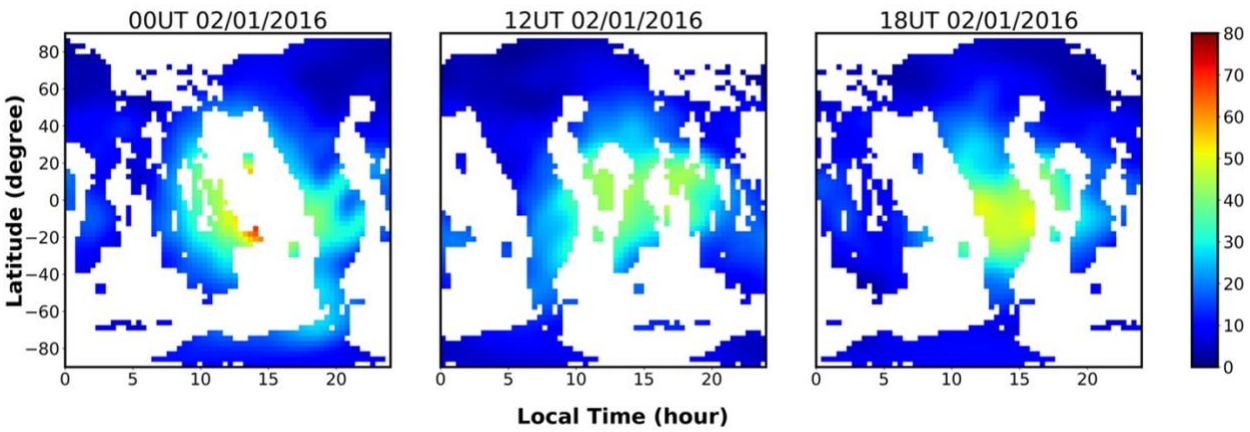


Figure 4.1 The masked IGS-TEC maps ($5.6^{\circ} \times 2.8^{\circ}$) at 00:00, 12:00, and 18:00 UT on 2016-02-01 in local time versus latitude. The blank regions in each map represent the MIT-TEC LT masks [Pan et al., 2021].

4.2.1 Cross-Validation (CV) on IGS-TEC Maps

Similar to our previous work [*Pan et al.*, 2020], we use a 10-fold cross-validation with pairs of low and high solar activity years (low solar activity: F10.7≤100 sfu; high solar activity: F10.7 >100 sfu) to systematically evaluate the inpainting performance for different models. Specifically, the IGS-TEC data from 18 out of 20 years (i.e., 90% of all data) were randomly selected for training, and the rest two (i.e., 10% of all data), one low solar activity year and one high solar activity year, were used for test. In the training data, 80% were used for training and 20% were used for validation. The years of the 10-fold CV test sets are shown in Table 3.1. Fifty TEC maps were randomly selected for each test year to calculate the RMSE values. The MIT-TEC LT masks were used (as shown in Figure 4.1) to artificially make incomplete TEC maps from the complete IGS-TEC maps for the calculation of RMSE. More details about the cross-validation settings and other styled mask experiments can be found in [*Pan et al.*, 2020].

The SNP-GAN model was trained using the following parameters: an effective batch size (EBS) of 16×8 (the number of batches $\times$ the number of GPU cores) and a fixed learning rate of

0.0001. The EBS and learning rate for DCGAN-PB were 64×1 and 0.00002, respectively. Both models were first run 2000 epochs on a validation set to determine the optimal number of epochs. DCGAN-PB reached the best RMSE values around 200 epochs, while SNP-GAN did at 2,000 epochs. Since SNP-GAN has a more complicated architecture than DCGAN-PB, it took more time to be well trained. It is worth mentioning that the arrangement of TEC map in local time by latitude for training leads to better performance than that in longitude by latitude. Although longitude and local time have a one-to-one correspondence, for any given UT, the TEC distribution arranged in local time shows a more consistent pattern. For example, the TEC peaks are almost always in the center of the map and the low TEC is on each side of the map throughout different UTs. If longitude is used, this pattern gradually shifts as UT advances. This spatial consistency with local time offers advantage to the GAN model to learn the underlying data distribution more effectively. This is similar to human face image, which always has the nose in the center. Thus, the position of this structure provides a more or less consistent and stable TEC pattern as UT progresses for effective inpainting network training. The deep learning models in this work were all trained in local time by latitude, unless otherwise stated.

4.2.2 Case Studies on Two Storm Events: March 2013 and February 2016

In order to study the performance of SNP-GAN during geomagnetic storms, two storm events (2013-03-16 to 2013-03-18 namely "2013 storm," and 2016-01-30 to 2016-02-04, namely "2016 storm") were selected. The MIT-TEC maps during these two storm events were binned in time for a temporal resolution of 10 min, which leads to a total of 216 TEC maps for calculation of RMSE. The models trained on 18 years' IGS-TEC maps from 1999 to 2018 excluding 2013 and 2016 (corresponding to test sets, "cv_9" and "cv_8" shown in Table 3.1, respectively) were used

in the case studies. Not only were the test data (MIT-TEC) different from the training data (IGS-TEC), but also the time periods of all the test data were excluded from the training data. Therefore, the training process was not contaminated by any information of the test data.

To quantitatively evaluate the performance of the SNP-GAN model, we conducted two experiments. The first experiment is to reconstruct the IGS-TEC maps that were intentionally masked with the MIT-TEC LT masks as shown in Figure 4.1. The second experiment is to reconstruct the MIT-TEC maps. For MIT-TEC maps, there are no complete maps available. In order to quantitatively evaluate the data completion performance of different methods on MIT-TEC, we intentionally removed some available data points from MIT-TEC maps to save them for the calculation of RMSE. 10% and 20% random brush masks take away approximately 10% and 20% data from MIT-TEC maps, which provide a sufficient number of missing data points with the pattern mimicking the large gaps in the original MIT-TEC maps for evaluation.

The trained DCGAN-PB and SNP-GAN were used to reconstruct the MIT-TEC maps (at the spatial resolution of $5.625°$ in longitude by $2.8125°$ in latitude and the temporal resolution of 5 min) from 2016-01-30 to 2016-02-04. Note that the finest temporal resolution in IGS-TEC data is 15 min, which is lower than the 5-min temporal resolution of the MIT-TEC data. The goal of this part of study is to investigate whether the deep learning-based methods can recover both the global maps and the localized mesoscale structures from MIT-TEC data.

4.3 Results

4.3.1 Cross-Validation on IGS-TEC Maps

The RMSE values averaged over 50 randomly selected TEC maps of each CV test year are plotted in Figure 4.2 ([a] for high solar activity years and [b] for low solar activity years). For each test year, SNP-GAN achieves lower RMSE values than DCGAN-PB, except for the ninth CV test set of the low solar activity years. The average RMSE of SNP-GAN is 1.99 TECU for high solar activity years and 1.09 TECU for low solar activity years. To evaluate the relative performance between the two models, we introduce "relative RMSE" (R-RMSE) where the RMSE value is divided by the average TEC value of all ground truth pixels in the masked regions. The R-RMSEs are 8.06% and 9.91% for SNP-GAN, respectively. The average RMSE of DCGAN-PB are 3.16 TECU for high solar activity years and 1.65 TECU for low solar activity years, and the corresponding R-RMSEs are 12.64% and 15.03%, respectively. RMSE shows that SNP-GAN has an improvement of 37% for high solar activity years and 34% for low solar activity years compared to DCGAN-PB (with the same improvements of R-RMSE). From the two-sample t-test between RMSEs in the 10-fold cross-validation of two models, we get a p-value of 0.000478 for high solar activity years and a p-value of 0.000187 for low solar activity years, which indicates that SNP-GAN outperforms DCGAN-PB in reducing RMSE values with statistically significance.

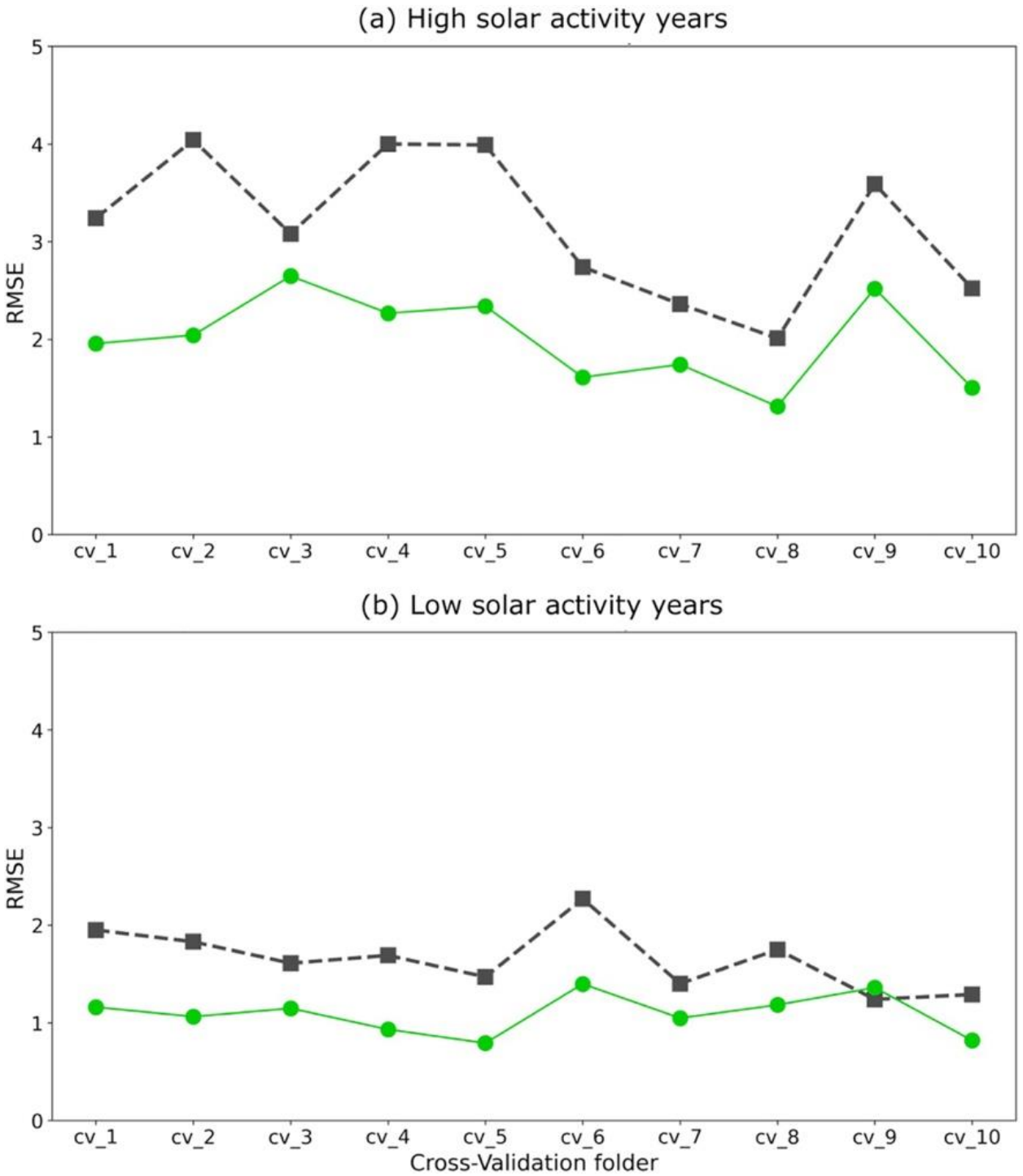


Figure 4.2 Average RMSE (TECU) of each CV test set of high solar activity years in 10-fold cross-validation of IGS-TEC data for (a) high solar activity years and (b) low solar activity years. (DCGAN-PB: black squares with dashed lines; SNP-GAN: green dots with solid lines) [Pan et al., 2021].

4.3.2 Case Studies

The interplanetary magnetic field (IMF) provides important information of the geomagnetic activity level and the southward IMF B_z could trigger magnetic reconnection, and thus ionospheric storms. We plotted the IMF during the storm times in 2013 and 2016 in Figure 4.3. The moments of TEC maps in the following figures are marked as the vertical green lines in Figure 4.3: 2013-03-17 10:00 UT and 2016-01-30 00:00 UT for the IGS-TEC map completions using the MIT-TEC LT mask in Figure 4.4, respectively; 2013-03-17 00:00 UT and 2016-02-03 00:00 UT for the MIT-TEC completions using the random brush masks in Figure 4.5, respectively; 2016-01-31 16:07:30 UT for the global MIT-TEC TEC peak structure recovery in Figure 4.6; and 2016-02-03 01:37:30 UT for the mesoscale structure recovery in Figure 4.7. Note that the test data set does cover the cases from quiet time to strong storms. While Figure 4.3 only shows the IMF conditions, the IMF B_z is strongly related to Kp index. During 2013 storm event, the maximum Kp is larger than 6, which indicates an intense storm. While in 2016 event, the maximum Kp is between 3-6, which means a moderate storm. We select several time points during these two events shown in Figure 4.3, which reflect both storms (large negative B_z or high Kp index) and the relative quiet time (near zero B_z or low Kp index).

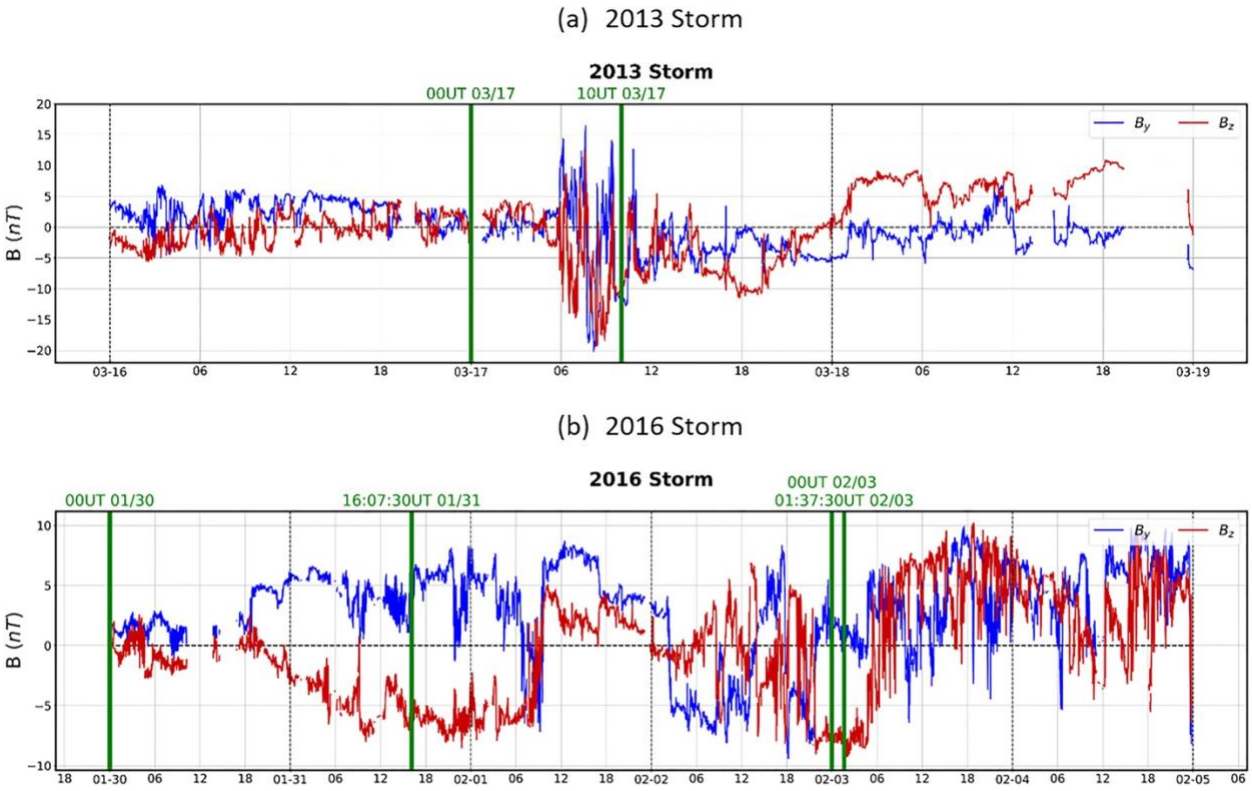


Figure 4.3 The IMF conditions of (a) 2013 storm and (b) 2016 storm. The green vertical lines represent time points for the following figures: 2013-03-17 10:00 UT for Figure 4.4a, 2016-01-30 00:00 UT for Figure 4.4b, 00:00 UT on 2013-03-17 for Figure 4.5a, 2016-02-03 00:00 UT for Figure 4.5b, 2016-01-31 16:07:30 UT for Figure 4.6, and 2016-02-03 01:37:30 UT for Figure 4.7 [Pan et al., 2021].

4.3.2.1 Reconstruction of Masked IGS-TEC Maps

Two representative inpainting results are shown in Figure 4.4a for 2013 storm (at 2013-03-17 10:00 UT) and Figure 4.4b for 2016 storm (during a quiet period at 2016-01-30 00:00 UT). SNP-GAN can recover the TEC maps much better than DCGAN-PB, especially for the high TEC values in Figure 4.4a. If we define regions with TEC values greater than 50 TECU to be "TEC peaks," which is near the Equatorial Ionization Anomaly (EIA) region, the DCGAN-PB result shows superfluous TEC peak structures in the LT interval from 23 LT to 0 LT, while the SNP-GAN result preserves the TEC peak structures that closely resemble the original maps. The quantitative results are shown in Table 4.1 IGS-TEC completion. SNP-GAN achieves much smaller RMSE

values than DCGAN-PB. The improvement is about 55.6% and 56.7% for 2013 storm and 2016 storm, respectively.

Table 4.1 RMSE Results (TECU) for 2013 Storm and 2016 Storm Using DCGAN-PB and SNP-GAN for IGS-TEC Completion With the MIT-TEC LT Mask and MIT-TEC Completion With Random Brush Masks 10% and 20% [Pan et al., 2021].

Data and mask type	Event	DCGAN-PB	SNP-GAN
IGS-TEC completion with the MIT-TEC LT masks	2013 storm	4.8061	2.1342
	2016 storm	3.2414	1.4020
MIT-TEC completion with random brush masks 10%	2013 storm	6.2374	6.0961
	2016 storm	3.7849	3.7312
MIT-TEC completion with random brush masks 20%	2013 storm	6.6980	6.0637
	2016 storm	3.8634	3.8046

4.3.2.2 Reconstruction of MIT-TEC Maps

The reconstructed MIT-TEC maps (with additional 10% and 20% random brush masks) at 2013-03-17 00 UT are shown in Figure 4.4, where a peak TEC structure (>50 TECU) can be seen in the IGS-TEC map. The SNP-GAN inpainting results for both 10% and 20% random brush masks apparently recover the double-band TEC structure (>50 TECU), while the DCGAN-PB inpainting results show too much connection between two bands. Note that the 20% mask removed some peak pixels, which lead to the discontinuity of the southward peak around 18LT in both DCGAN-PB and SNP-GAN results. This also leads to larger RMSE values for the result of the 20% mask than that of the 10% mask. The RMSE values of SNP-GAN in this case substantially outperforms that of DCGAN-PB.

Another example at 2016-02-03 00UT is shown in Figure 4.5. The solar activity in this case is relatively low compared to the previous case (Figure 4.4), thus leading to a smaller area of high TEC values on the dayside. The masked ground pixels are mostly distributed in low and middle latitude regions, where two bright TEC spots are near the equator. Both DCGAN-PB and SNP-GAN results look similar for the 20% brush mask case, which is also reflected by the similar

RMSE performance. However, the 10% mask erased partially the two TEC bright spots, while the 20% mask was on the regions with low TEC values. As a matter of fact, the average TECU of the masked pixels in the 10% mask is approximately twice as much as that of the 20% mask. Consequently, the RMSE values are worse for the 10% mask than the 20% mask. However, the deterioration is much smaller for SNP-GAN (RMSE ratio between the 10% mask and the 20% mask is 1.10) compared to DCGAN-PB (RMSE ratio between the 10% mask and the 20% mask is 1.76). SNP-GAN preserves more consistent completion performance than DCGAN-PB even if some high TEC values are missing.

The average RMSE values through the whole storm time in 2013 and 2016 are shown in Table 4.1 for random brush masks 10% and 20% masks, respectively. As for 10% mask, the quantitative performance is similar between DCGAN-PB and SNP-GAN for both 2013 storm and 2016 storm for 10% mask. As the masked pixels increased from 10% to 20% where more high TEC values were brushed out, the SNP-GAN reconstruction leads to reductions of 9.47% and 1.52% compared to DCGAN-PB for 2013 storm and 2016 storm, respectively. This demonstrates that SNP-GAN outperforms DCGAN-PB mainly on recovering the structures with high TEC values, while keeps a similar performance on smooth background TEC values. It is also noted that when the number of brushed out pixels increased from 10% to 20%, RMSE of SNP-GAN barely increased (indeed deceased slightly for 2013 storm) compared to DCGAN-PB, which demonstrates the consistent inpainting performance of SNP-GAN for different sizes of missing data gaps in general.

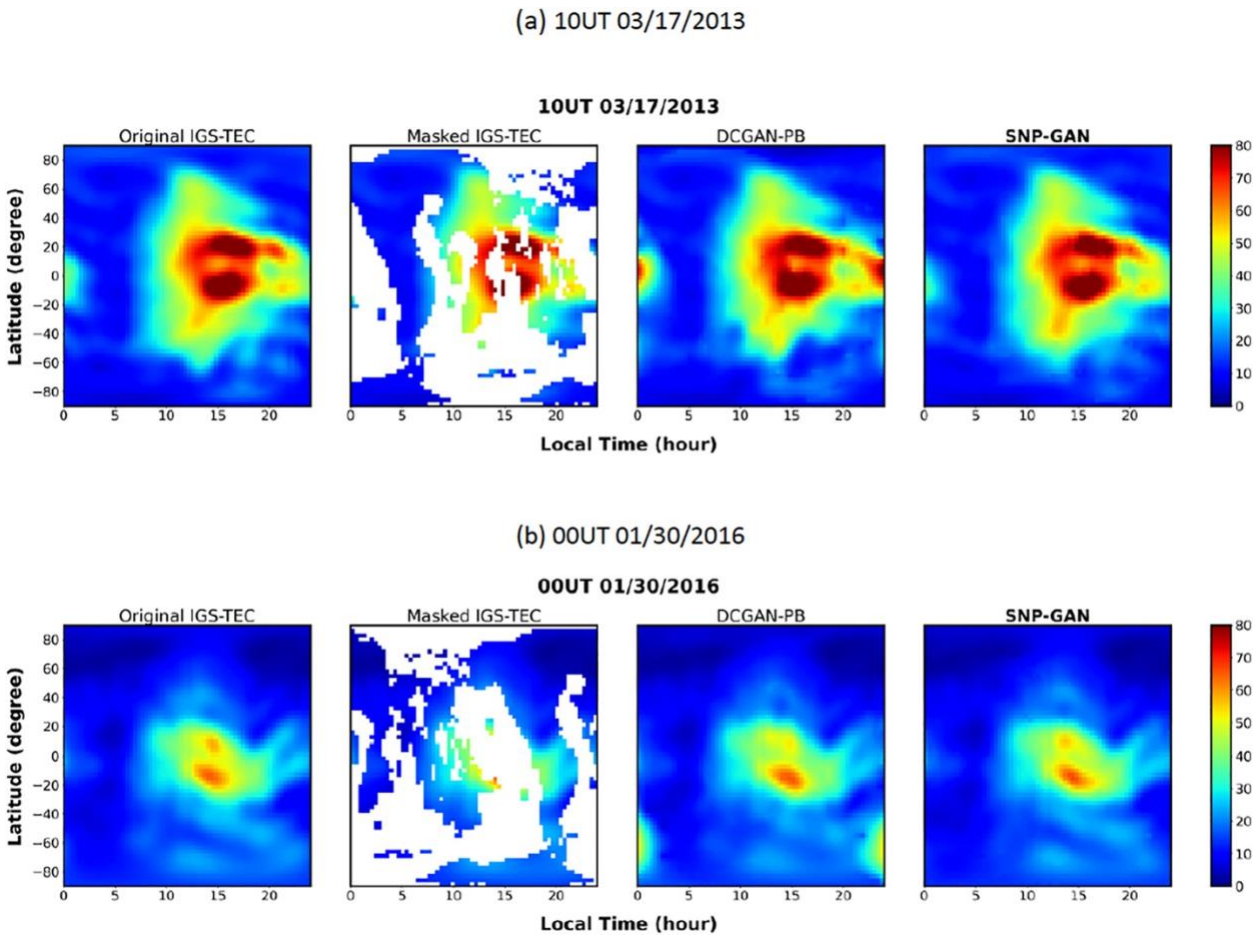
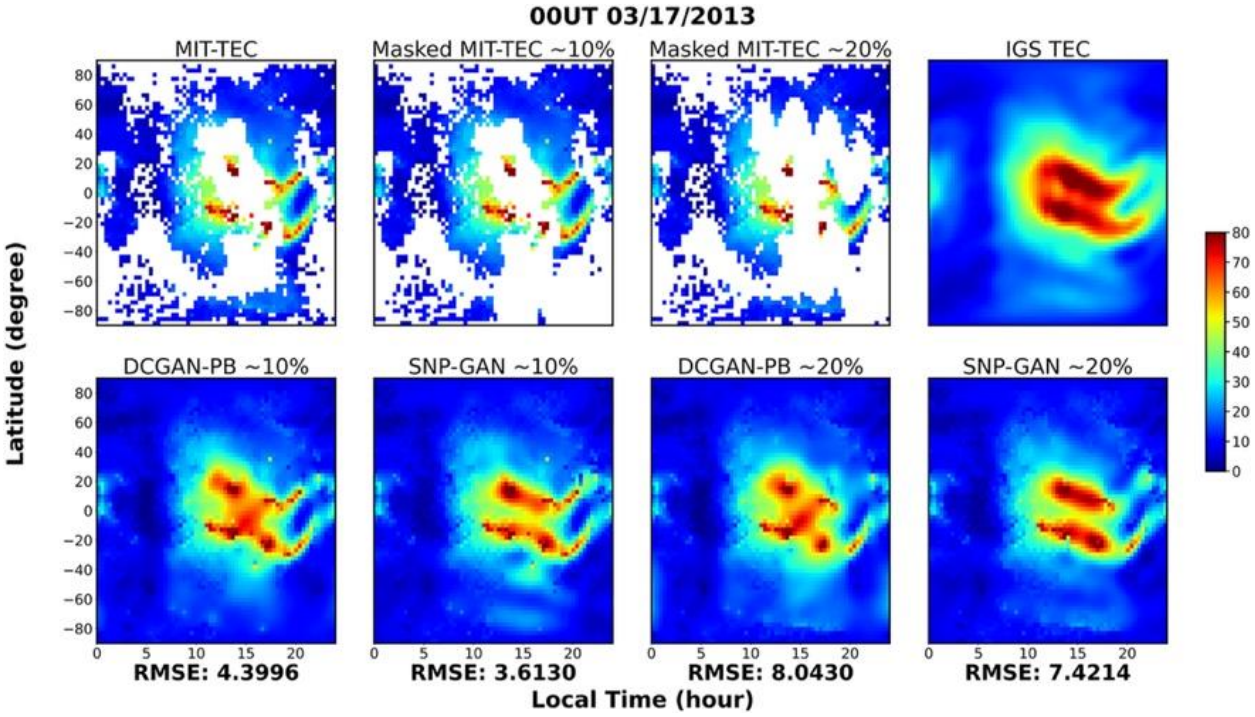
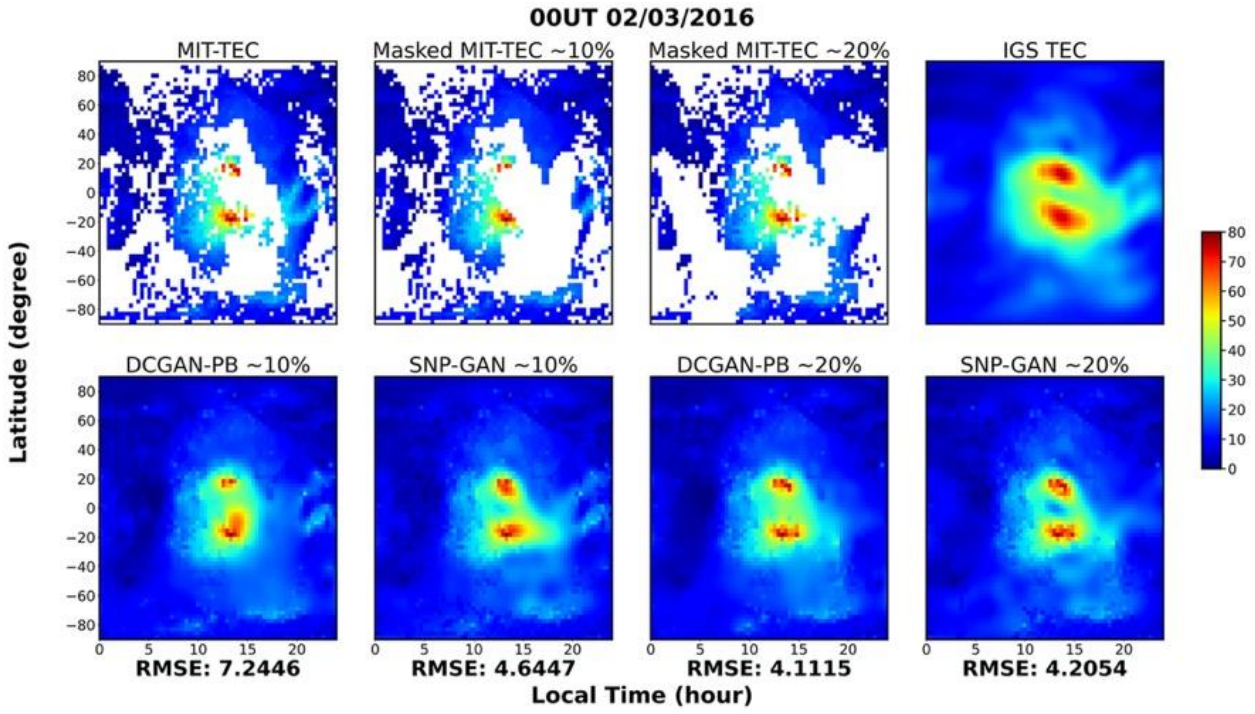


Figure 4.4 Reconstructed IGS-TEC maps at (a) 2013-03-17 10:00 UT and (b) 2016-01-30 00:00 UT using DCGAN-PB and SNP-GAN [Pan et al., 2021].



(a) 2013-03-17 00 UT



(b) 2016-02-03 00 UT

Figure 4.5 Reconstructed MIT-TEC maps using DCGAN-PB and SNP-GAN with 10% and 20% random brush masks [Pan et al., 2021].

4.3.2.3 Reconstruction of TEC Structures at Different Scales

In addition to the global maps (local time v.s. latitude) in Figure 4.6, the polar maps were also provided from 50°N to the geographic north pole in Figure 4.7.

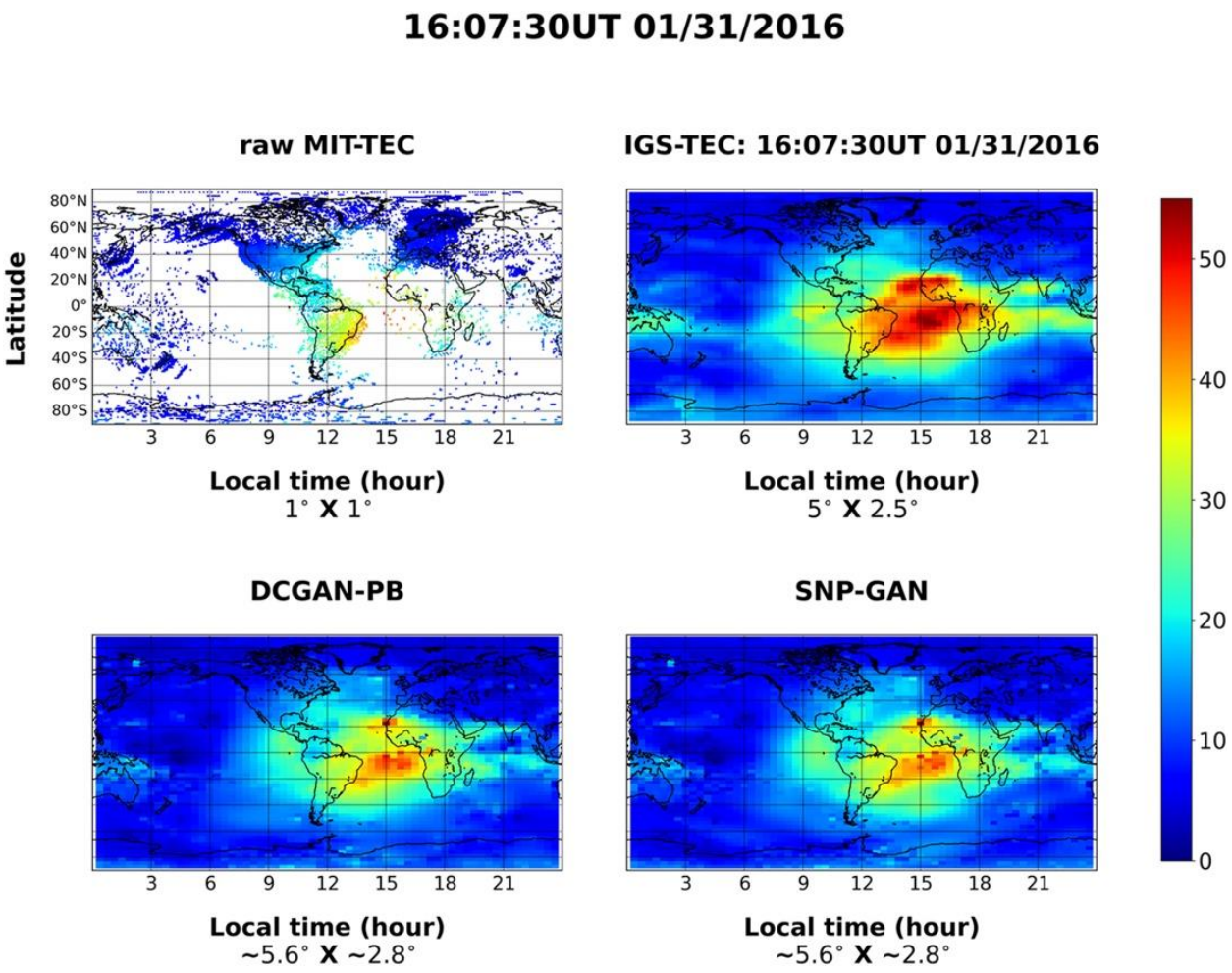
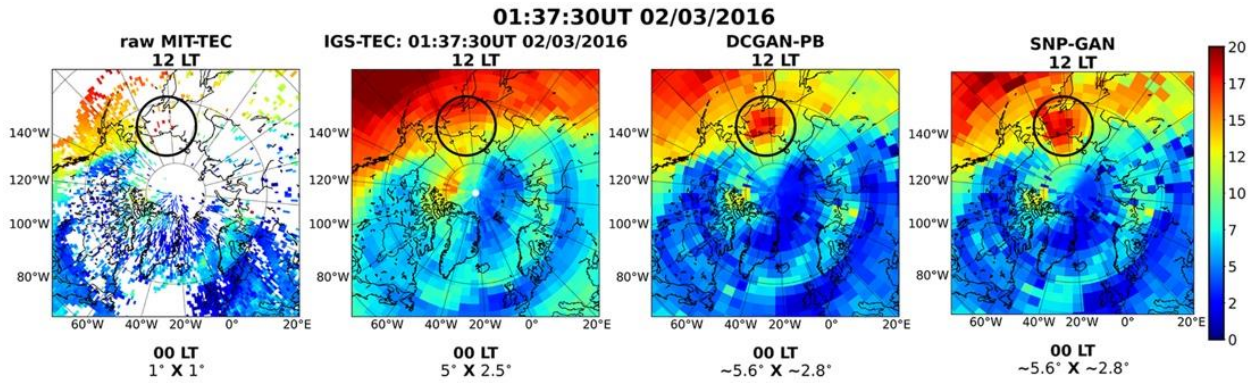


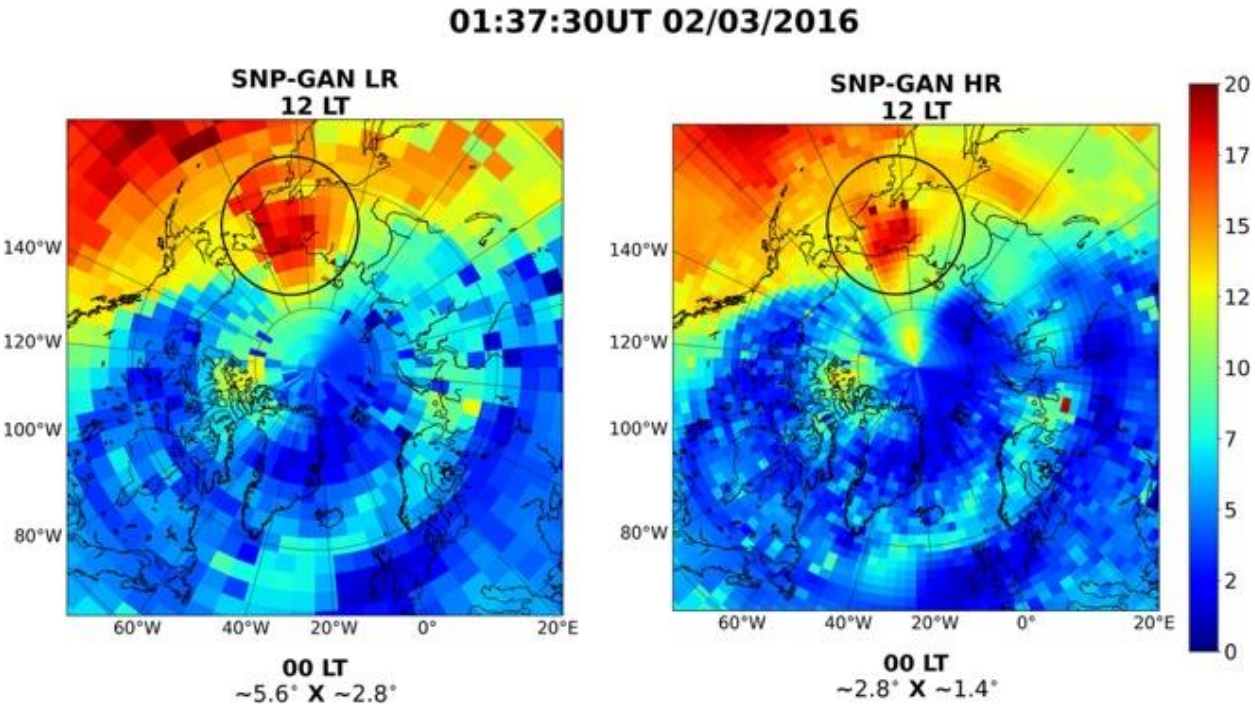
Figure 4.6 Global TEC maps at 2016-01-31 16:07:30 UT. TEC peaks above the Atlantic ocean can be seen in all completed maps [Pan et al., 2021].

The reconstruction results at 2016-01-31 16:07:30 UT under southward IMF condition are shown in Figure 4.6. As can be seen, the global IGS-TEC map shows a clear TEC peak above the Atlantic Ocean area. Due to the limited receiver stations, the data coverage over the ocean is very

sparse and the peak is missing in the MIT-TEC map. However, both DCGAN-PB and SNP-GAN successfully reconstruct the large-scale structure which fell into the data gap of the MIT-TEC map. Both DCGAN-PB and SNP-GAN reconstruct the TEC structures on the Atlantic Ocean.



(a) Low-resolution



(b) Low-resolution and high-resolution

Figure 4.7 North polar TEC maps at 2016-02-03 01:37:30 UT. A cusp-like structure is visible in: (a) raw MIT-TEC, DCGAN-PB map, SNP-GAN map; (b) SNP-GAN map low-resolution, and high-resolution [Pan et al., 2021].

Figure 4.7a shows the polar distribution of TEC at 2016-01-31 16:07:30 UT. A cusp-like structure is clearly visible west to Bering Strait (as circled) for TEC maps reconstructed by DCGAN-PB and SNP-GAN. As indicated by several isolated bright spots with high TEC values in the original (incomplete) MIT-TEC map, a TEC peak structure may exist. However, such a cusp-like structure cannot be observed from the IGS-TEC map, which is likely due to the heavy smoothing used in the IGS-TEC processing methods. Both DCGAN-PB and SNP-GAN seem to be good models to combine both the complete IGS-TEC maps and the MIT-TEC raw data. On one hand, they learn the IGS-TEC data distribution for data filling. On the other hand, they constrain the data filling process to the MIT-TEC data input to avoid over-smoothing. Therefore, those bright spots in the original incomplete MIT-TEC map are not smeared by the deep learning inpainting models, which shows their great potential of identifying mesoscale structures. Furthermore, SNP-GAN seems to be in a good balance that keeps both the global structures of the IGS-TEC map and the local structures of the MIT-TEC map.

4.4 Discussion

SNP-GAN outperforms DCGAN-PB by a large margin in the 10-fold cross-validation of masked IGS-TEC data, which demonstrated the strong end-to-end learning capability of SNP-GAN. For MIT-TEC data, SNP-GAN is also more effective on recovering TEC structures with large TEC values than DCGAN-PB. DCGAN-PB seems to work well on smoothing background TEC values, largely due to the postprocessing of Poisson blending. However, we found that Poisson blending was not able to improve SNP-GAN inpainting performance, likely due to its end-to-end inpainting training. The advantage of SNP-GAN is not only on the accuracy of reconstruction of missing TEC data, but also on saving the computation time of reconstruction.

While DCGAN-PB needs an iterative optimization to map the generated TEC map to the missing TEC data context, the "end-to-end" inpainting network architecture of SNP-GAN allows much more efficient inpainting. It takes approximately 4 days 7 hr for 2,000 epochs training of the DCGAN-PB model on one NVIDIA Tesla V100 GPU, while 3 days 3 hr for the same number of epochs of SNP-GAN training on 8X NVIDIA Tesla V100 GPUs. By averaging over 108 TEC maps used in the case studies (in 2013 and 2016), the SNP-GAN model takes approximately 1.6 s to load the trained model and one set of the input data (an incomplete TEC map and its mask) and about 0.74 s to complete a TEC map on one NVIDIA Tesla V100 GPU. For comparison, the DCGAN-PB model takes much longer time, 144 s, to complete a TEC map on the same GPU including the similar loading time and the iterative post optimization. That is more than 60 times (=144 s/2.34 s) speed up of the completion process for SNP-GAN over DCGAN-PB. Once the SNP-GAN model was trained, it only takes less than 1 s to complete one TEC map. Therefore, it is feasible to produce real-time global TEC maps using SNP-GAN given that the data from thousands of stations could be collected instantly. However, it is challenging to acquire the real-time MIT-TEC data from over 6,000 receivers distributed worldwide. Usually, there is a 7-day delay for the availability of such a data. It is worth noting that there is a real-time TEC map project led by NOAA, which is confined to the North America region.

Different from our previous work on IGS-TEC maps, the current work investigates innovatively on the training and test sets of different types. Although DCGAN-PB and SNP-GAN were trained on the IGS-TEC maps, the MIT-TEC maps were used as the test data since the complete MIT-TEC maps were not available. Both quantitative and qualitative results on the completion of MIT-TEC maps show the effectiveness of deep-learning based inpainting methods. The case studies further provide some insights of these models that can produce sensible complete

TEC maps from MIT-TEC data by keeping both global and local TEC features. From Figure 4.6, when the missing part is on the Atlantic ocean in the global view, both models produced decent qualitative looking by recovering the TEC peaks over the ocean part with finer TEC structures recovered by SNP-GAN. For the north polar view where the cusp-like structures occur, the deep-learning models are capable of identifying these structures, whereas the IGS-TEC maps miss this critical information.

In addition to "year-wise" split of data for training, we did extra experiments using "month-wise" split as follows: training/validation data are 1999–2018 except 2016, where April to December except May are for training and May for validation. 2016-01-30 to 2016-02-04 are used for test (in order to compare with results using year-wise training in Table 4.1). First, we used the validation set to choose the best learning rate (LR) out of 1×10^{-3}, 1×10^{-4}, and 1×10^{-5}. The LR of 1×10^{-4} yields the best validation performance, that is, smallest RMSE on the validate set. Then, we chose the model trained using this LR to test on the storm event period in 2016. The results are 4.5063 TECU, 3.7504 TECU, and 4.0017 TECU, respectively, for three cases in Table 4.1 for 2016 storm. As can be seen, the year-wise SNP-GAN results are much better than the month-wise results in the IGS-TEC completion. While the differences are small for the MIT-TEC completion in random brush masks, the year-wise SNP-GAN model still yield the least RMSE values. Although the RMSE values are close for two models, the visual inspection of completed TEC maps reveals that the year-wise model produces more consistent results than the month-wise model. Since the year-wise model includes the data containing both seasonal changes in a year and yearly changes in a solar cycle, it works better than the month-wise model, whose training data are lack of certain months.

To better utilize the high spatial resolution of MIT-TEC data, we interpolated the IGS-TEC maps to 128×128 maps with spatial resolution of $2.8125°$ in longitude by $1.40625°$ in latitude to train SNP-GAN. The trained high-resolution SNP-GAN model was then used to complete the MIT-TEC map at $2.8125°$ in longitude by $1.40625°$ in latitude. The high-resolution model result at 2016-01-31 16:07:30 UT is shown in Figure 4.7(b) along with the low-resolution model result ($5.6125°$ in longitude by $2.8125°$ in latitude). Although the cusp-like structure can be seen in both low-resolution and high-resolution SNP-GAN models (circles in Figure 4.7(b)), the high-resolution SNP-GAN model produces a narrower and sharper cusp-like peak and better resolves the transition between different regions than the low-resolution model. Note that the training of high-resolution SNP-GAN is much more time consuming than the low-resolution one, with approximately 87 hours for the former and 50 hours for the latter on 8X Nvidia V100 GPUs, respectively. Since the hyperparameters and the masks are not fully optimized in this study, some small abrupt TEC value changes can be observed in the completed TEC maps. Nevertheless, these preliminary results with high-resolution SNP-GAN demonstrate that although IGS-TEC maps have a limited spatial resolution, a SNP-GAN model trained on up-sampled IGS-TEC maps can still take advantage of the high-resolution MIT-TEC data to produce the complete maps with rich structures and smooth transition between different regions.

To validate the existence of the cusp-like structure in Figure 4.7. In situ measurements by satellites may be utilized. However, we checked DMSP satellites orbits and could not find an overlap of the location. In order to approve the mesoscale structure is not artificial, we conducted some additional reconstructions with conventional image inpainting methods including linear interpolation, TELEA and NS (Pan et al., 2020) corresponding to Figure 4.7a (low-resolution) and Figure 4.7b (high-resolution). The results demonstrate that all methods yield the cusp-like structure.

The linear interpolation performs decently in the polar region, while TELEA and NS have an artificial spear-like structure to the north pole. In addition, TELEA suffers severe mosaic-like artifacts in the high-resolution map. SNP-GAN seems to produce a more appealing look of this structure. We also conducted a quantitative evaluation of linear interpolation for experiments in Table 4.1. The linear interpolation did a nice job for IGS-TEC maps, which are smooth, but were inferior to two deep learning models for MIT-TEC maps, which contain more subtle structures. The visual inspection of reconstructed maps also confirmed that the more subtle structures can be recovered better by the deep learning methods than linear interpolation. A quantitative comparison between TELEA/NS and the deep learning-based inpainting (DCGAN-PB) can be found in our previous work [*Pan et al.*, 2020]. We also included videos of TEC maps reconstructed by SNP-GAN that show the structure continuously evolves along the time for more than one hour instead of an abrupt change. All these provide supporting evidence that the structure is not an artificial structure due to the methodology, instead represents a real phenomenon. Certainly, more systematic validation of the mesoscale structure is needed. But it is out of the scope of this study, which focuses on introducing a new method for TEC map completion and its potential. In the future, we will select events during which the data coverage in the moon-midnight location time section is great. The correlation of the mesoscale structure from SNP-GAN with the satellite observations will then be examined in a comprehensive way covering various kinds of events in different seasons, which will be an excellent topic for future investigation.

The confidence level of the filled TEC values is an important metric in order to adapt the proposed deep-learning completion methods. Since these models are highly nonlinear, the assumption of Gaussian distribution may not be a good approximation, thus leading to erroneous confidence levels. One empirical solution is to start with different random seeds for each missing

pixel and to run multiple training and completion. Therefore, an empirical distribution can be obtained to access the confidence level. However, such an investigation is substantial and out of the-scope of this work.

5 Neural Network Models for Ionospheric Electron Density Prediction: A Neural Architecture Search Study

5.1 Introduction

The incoherent scatter radar (ISR), located at Millstone Hill (42.6°N, 71.5°W, dip 71.6°) can provide direct measurements of ionospheric parameters, such as electron density (*Ne*), plasma temperature, and line of sight ion velocity. The altitudinal (range) variation of these parameters is measured continuously over time by the ISR. However, most traditional ISRs operate for campaign purposes but not on a daily basis. Figure 5.1 shows an example of *Ne* around 350 km at the Millstone Hill station in 2012, where a lot of data are missing. Therefore, a model that can fill the observational data gaps for these parameters under real solar/geomagnetic conditions would be desired for various space weather and ionospheric research purposes.

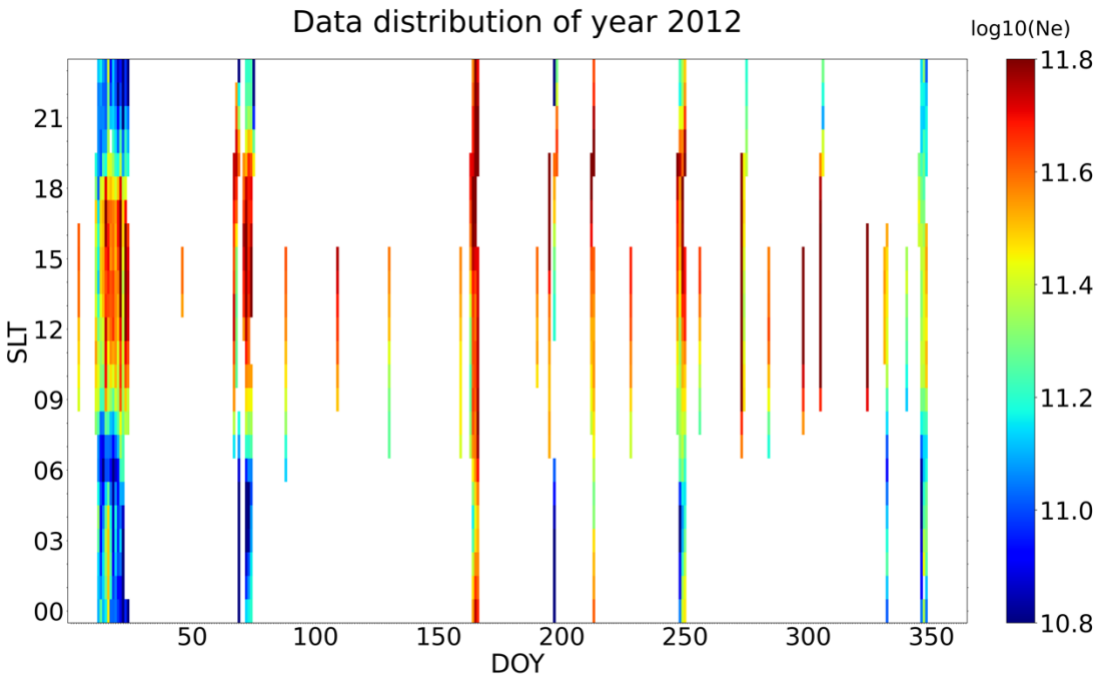


Figure 5.1 The ISR records of Ne in the logarithmic scale around 350 km altitude in 2012. Horizontal axis: day of year (DOY); vertical axis: solar local time (SLT); the intensity represents logarithmic electron density ($log_{10}Ne$), while the blank space represents missing records. Most of the region is in blank, indicating the irregularity of ISR's operation.

Conventionally, the empirical models were developed to provide this information. For example, a global model, international reference ionosphere (IRI) *D. Bilitza* [2001] and IRI-2016 [*D. Bilitza et al.*, 2017], takes primarily ionosonde observations to generate 3D distributions of ionospheric parameters. The ISR ionospheric model (ISRIM) [*Holt et al.*, 2002] has been built for multiple ISRs around the world developed initially for Millstone Hill ISR observations in the time and vertical domains [*Holt et al.*, 2002]. Additional regional models beyond local vertical variations were also developed near Millstone Hill as well as in the North America longitudes. These statistical models took a binning and fitting approach to construct an empirical model in space and time [*S R Zhang and Holt*, 2007]. In each bin, the sequential least-squares fit is based on the normalized F10.7 and Ap3 indices, especially with the new introduced parameter F10.7p [*L Liu et al.*, 2006; *Richards et al.*, 1994] for better linear fitting [*S R Zhang and Holt*, 2007]. However, ISRIM was designed to provide ionospheric climatology where altitudinal and temporal variations are represented by smooth analytical models. The artificial neural network (ANN) models may be trained to better fill the data gaps or to predict these parameters.

The neural network regression models have been developed for space weather research (see for example [*S Wing et al.*, 2005]). A single hidden layer ANN with 18 neurons was used to derive ionospheric models in order to evaluate the long-term trends of *Ne* for the Defense Meteorological Satellite Program (DMSP) data [*Y Cai et al.*, 2019; *Yue et al.*, 2018]. The deep neural network (DNN) was used to model *Ne* to reconstruct the dynamics in the plasmasphere [*Bortnik et al.*, 2016]. To offer the short-term variations, a three-dimensional dynamic electron density (DEN3D) model [*X Chu et al.*, 2017; *X N Chu et al.*, 2017] is also developed for plasmasphere using DNN with enhanced number of drivers of F10.7 and AL apart from SYM-H.

Several global ANN models have been proposed to predict ionospheric *Ne*. The ANN-based ionospheric models (ANNIM-2D and ANNIM-3D) have been proposed using a single-layer NN (SLNN) and more than 10-year data from the global positioning system radio occultation (GPS-RO) missions [*Gowtam et al.*, 2019; *Sai Gowtam and Tulasi Ram*, 2017; *Tulasi Ram et al.*, 2018], such as CHAllenging Minisatellite Payload (CHAMP), Gravity Recovery And Climate Experiment (GRACE), and Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC). The ground-based Digisonde Global Ionosphere Radio Observatory (GIRO) (with 864 spatial grids for ANNIM-3D) is also on the database list for ANN-based models. Another global model (with 864 sub-models) was also proposed using COSMIC data [*Habarulema et al.*, 2021], where each sub-model adapted a SLNN. A three-hidden-layer DNN was used for a global 3D model ("ANN-TDD") based on COSMIC, Fengyun-3C and Digisonde data [*W Li et al.*, 2021]. The most recent work combined DNN with IRI ("ANN-IRI") to improve *Ne* prediction compared to pure data-driven ANNs, particular in the lower ionosphere [*D Yang and Fang*, 2023]. These pioneer models reproduce the large-scale ionospheric phenomena and generally outperform the monthly-average model of IRI-2016 during the quiet time. However, firstly, the radio occultation measured *Ne* assumes the spherical symmetry which is the major source of errors when retrieving from vertical profiles [*Lei et al.*, 2007]. Secondly, the aforementioned NN models usually have a worse prediction performance during the storm time than IRI-2016 with the STORM option on (specifically tailored for predictions during the storm time). One reason is that the storm events are comparatively taking up a smaller percentage in all the data used for the model training (i.e. not focusing on storm time behaviors), thus leading to inferior *Ne* prediction of these NN models during the storm time. Furthermore, these NN models usually chose the network structures and hyperparameters manually. Not only is the manual tuning tedious (e.g. thousands of experiments

were used to find a good 3-hidden-layer network structure [*W Li et al.*, 2021]), but also these models could only achieve sub-optimal prediction performance.

To address this issue with NN models for *Ne* prediction, we use an automatic optimization algorithm, so called neural architecture search (NAS) to optimize a single hidden layer NN (SLNN) and a deep NN (DNN) model and compare their performance. As our goal is to introduce NAS for optimization of NN models of *Ne* prediction, we used Millstone Hill ISR data at a fixed altitude (~350 km) from 2003 to 2018 since the data around this altitude are abundant and likely relevant to the low-earth-orbit (LEO) missions, such as CHAMP and the upcoming Geospace Dynamics Constellation (GDC, https://lws.larc.nasa.gov/gdc/) mission with an expected orbit altitude of 350-400 km.

5.2 Data and Experiments

The Millstone ISR *Ne* data at the fixed altitude of ~350 km from 2003-2018 were used for training and test of different NN prediction models. The input variables are year, day number of year (DOY), solar local time (SLT, hour), daily F10.7 index (solar flux unit or sfu), and 3-hourly Ap index (Ap3), in which the cyclic sine and cosine are applied on DOY (DOY_s and DOY_c in Equation *5.1*) and SLT (SLT_s and SLT_c in Equation *5.2*) to reflect the periodic changes of these two input variables as suggested by previous studies [*Athieno et al.*, 2017; *Habarulema et al.*, 2021] as well as more stable training. All the input parameters in our models are normalized within the interval [0, 1] (as commonly done in NN modeling), which is different from the previous works [*Athieno et al.*, 2017; *Habarulema et al.*, 2021]. If not specifically elaborated, the output variable

Ne stands for the logarithmic electron density (i.e. *Ne* is equivalent to $\log_{10}Ne$, particularly for the numerical values) in the following sections.

$$DOY_s = [\sin(2\pi \times \frac{DOY}{365}) + 1]/2, DOY_c = [cos(2\pi \times \frac{DOY}{365}) + 1]/2 \quad 5.1$$

$$SLT_s = \left[\sin\left(2\pi \times \frac{SLT}{24}\right) + 1\right]/2, SLT_c = [cos(2\pi \times \frac{SLT}{24}) + 1]/2 \quad 5.2$$

Table 5.1 Data setting and the conditions to clean ISR data. The ISR data has the greatest number of observations near height of 350km, which indicates the data availability is of our major consideration. The filters on two F10.7 and Ap3 would rule out high intensity geophysical events.

Parameter	**Valid Range**	
Years	Training	2003 to 2018 except the val&test sets
	Validation	[2010, 2015]
	Test	[2007, 2012]
F10.7	≤ 300 sfu	
Ap3	≤ 80	
Altitude	~350 km	
Ne	$[\log_{10}(5 \times 10^9), \log_{10}(3 \times 10^{12})]$ el/m^3	

A total of 16 years of ISR data from 2003 to 2018 were used. We calculated the annual F10.7 average and used a threshold 90 solar flux units (sfu) to divide 16 years into high solar activity (HSA) and low solar activity (LSA) years. The validation and test sets were chosen as a pair of HSA year (2015 for validation and 2012 for test) and LSA year (2010 for validation and 2007 for test) to have a similar data distribution to the training data. Year 2010 and 2015 were selected as validation set, while year 2007 and 2012 were reserved as test set. Remaining 12 years of data were used for training. We first cleaned the ISR data following the conditions in Table 5.1.

Specifically, the data corresponding to high solar activity and intense earth magnetic conditions (with F10.7 over 300 sfu and Ap3 greater than 80 units), which take about only 2% of whole dataset, were discarded following the previous work [*Y Cai et al.*, 2019]. The *Ne* values were also confined to the range of $[5 \times 10^9, 3 \times 10^{12}]$ el/m^3. Furthermore, the noisy data that show isolated peaks/troughs or irregular time intervals in daily patterns were discarded. Finally, though the most frequent ISR observation cadence is 4-minute, the remaining data were binned to a one-hour interval. One hour cadence was chosen to balance short-term variability in data and temporal resolution of the model. We also assured that the training, validation, and test sets followed the similar distribution of that of the overall *Ne*. After all these preprocessing of data, the training/validation/test set include 8,052/1,461/1,970 data records, respectively.

We used the mean absolute error (MAE), root mean squared error (RMSE), and relative error (RE) of the test data as the quantitative measures for the prediction performance. The Bland-Altman plots were used to interrogate the agreement between model output and ground truth *Ne*. We also quantitatively compared the predicted annual and day-to-day variations for all models supplemented by rankings of a daily variation prediction.

5.3 Results

In this section, the best network structure for the NAS models and the search for the best learning rates for all the models are presented first. Then, the prediction performance is evaluated statistically using MAE, RMSE, RE, and Bland-Altman plot. Next, we compare the NN models with an empirical model in a climatological study. Finally, we analyze the prediction performance

in a resolved temporal scale. The day-to-day electron density pattern prediction is shown for different models with a ranking study.

5.3.1 Determination of Optimal Number of Epochs through Validation Loss Dips

In Table **5.2**, the number of hidden layers and the number of neurons in each layer are shown. For the NAS models, these numbers were determined by the best validation loss (MAE of the validation data) from eight independent randomly initialized AutoKeras trainings. Since the early stop was used in NAS, a fine tune of learning rate was conducted using the training and validation loss curves where each tuning run consists of 8,000 epochs, after the network structures were determined. The training and validation loss curves for the best learning rate of each model (the last row of Table **5.2**) are shown in Figure A.1. As demonstrated, the validation loss curve floats slightly above the faster converging training loss and keeps decreasing until reaching the black dot. As the increase of the validation loss indicates the possibility of the model overfitting, we chose the number of epochs for each method as the one that achieved the minimal validation loss in the first 8,000 epochs. Therefore, the optimal number of epochs differ among methods. Although the fluctuations are observed in the validation loss curve, around the minimum validation loss the model performance is similar.

Table 5.2 *The hyperparameters for four NN models, which are the optimal results of each category in architecture, learning rate, and validation loss dip epoch.*

	SLNN	DNN	SLNN-NAS	DNN-NAS
# of layers and neurons	[18]	[24, 22, 20]	[52]	[60, 32]
Learning rate	5e-04	9e-05	1.6e-04	7.7e-05
# of epochs	2195	4444	2116	6046

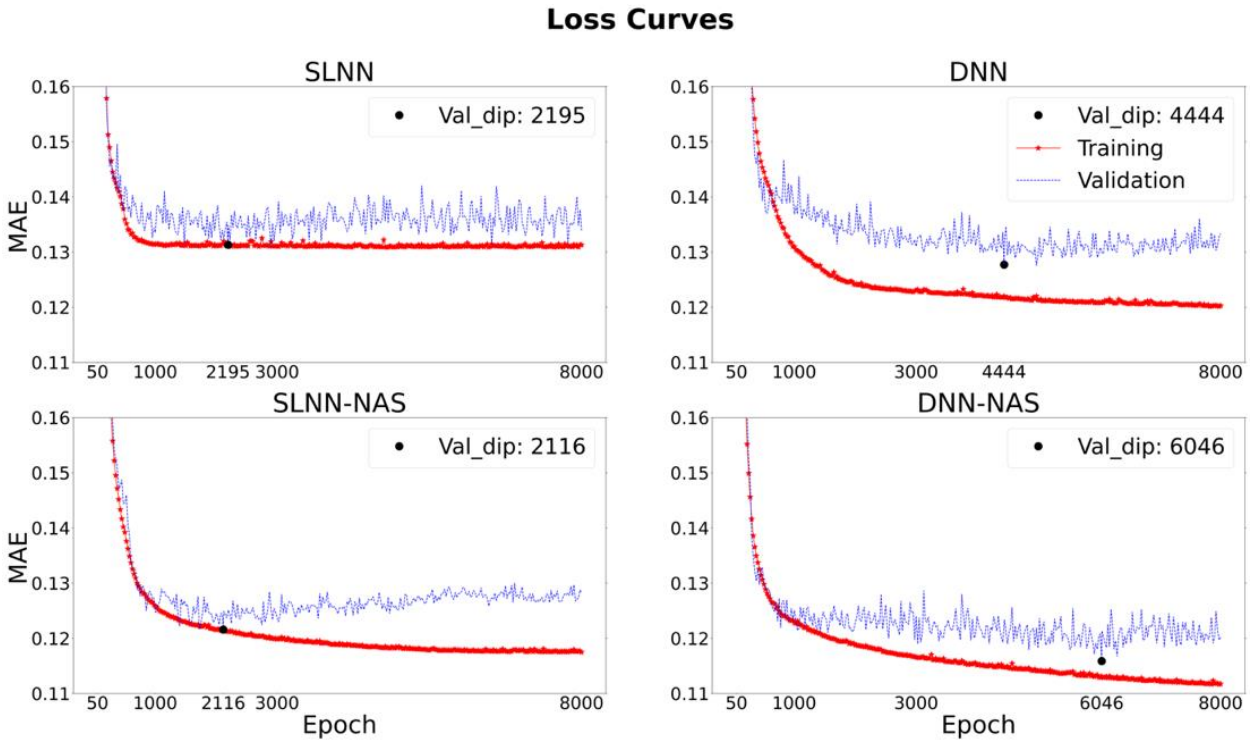


Figure 5.2 *The training (red stars) and validation (blue dashes with more variations) loss curves of the four NN models (the optimal number of epochs marked as the black dot). The two DNN models take more epochs to evolve the optimal results due to more complexity than SLNNs, while the NAS guided models lead to better model generality (lower possible validation loss).*

5.3.2 Overall Prediction Performance

Their quantitative metrics, MAE, RMSE, and RE, on the test data are shown in Table 5.3 below:

Table 5.3 Prediction errors for four models in mean absolute error (MAE), root mean square error (RMSE), and relative error (RE) percentage.

	SLNN	*DNN*	*SLNN-NAS*	*DNN-NAS*
MAE	0.1399	0.1312	0.1307	0.1250
RMSE	0.1908	0.1805	0.1821	0.1784
RE (%)	1.2667	1.1872	1.1844	1.1327

Two NAS models have lower prediction errors than their counterparts with fixed architectures. For example, NAS results in 6.6% reduction on MAE of *Ne* for SLNN and 4.7%

reduction for DNN, respectively. DNN-NAS achieves the best prediction performance, i.e. lowest MAE, RMSE, and RE. Its improvement over SLNN is more than 10% on MAE and RE.

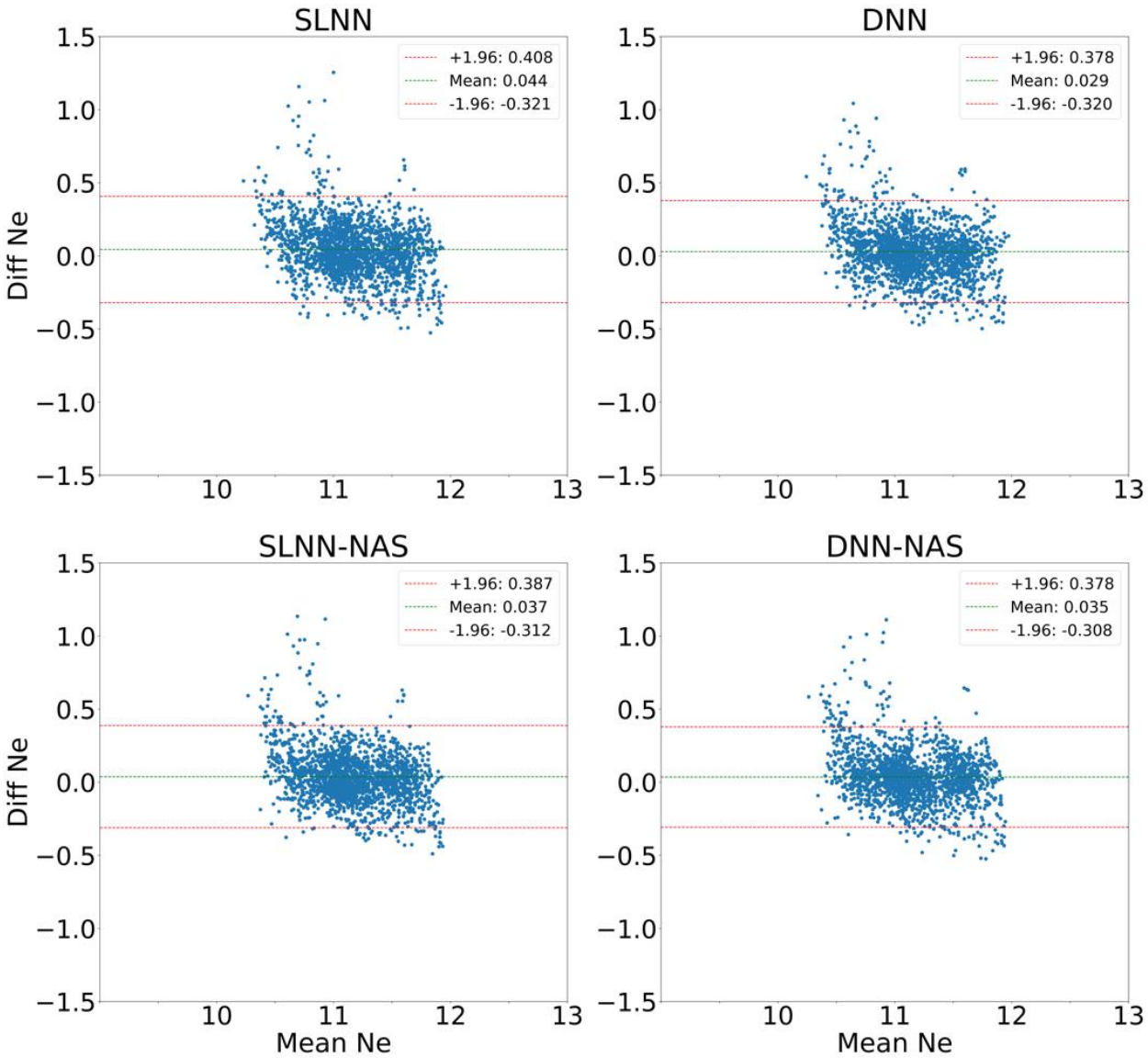


Figure 5.3 *BA-plots of the four optimal models (SLNN, DNN, SLNN-NAS, and DNN-NAS), in which the calculations are based on the test set. DNN tends to have the lowest averaged difference (green line in the upper right subplot) and the DNN-NAS owns the narrowest limits of agreements (distance between two red lines in the lower right subplot). The Y-axis is the Ne difference between the model prediction and the observation. The X-axis is the average of the model prediction and the observation.*

The Bland-Altman (BA) plots in Figure **5.3** show the agreement between each model prediction and the ground truth *Ne* from ISR observation. All four models introduce a small positive bias ranging from 0.029 to 0.044 (several tenth of a percent of the average $\log_{10}(Ne)$). This is likely due to the large positive differences, i.e. the points above 1.96 SD line, particularly when the electron densities are low. These NN models seem to substantially overestimate *Ne* occasionally at the sharp dip of observations, such as 11 UT as indicated in Figure **5.5**(b). Significant model development effort is needed to address this issue in future. Statistically, SLNN shows the least agreement with the largest bias and the widest 95% limits of agreement (± 1.96 SD). SLNN-NAS is better than SLNN, but still worse than DNN and DNN-NAS. DNN-NAS has a slightly larger bias but a narrower 95% limits of agreement than DNN. Again, DNN-NAS achieves the best agreement between the prediction and the ground truth since DNN-NAS adapts an optimal network structure and other hyperparameters, such as learning rate.

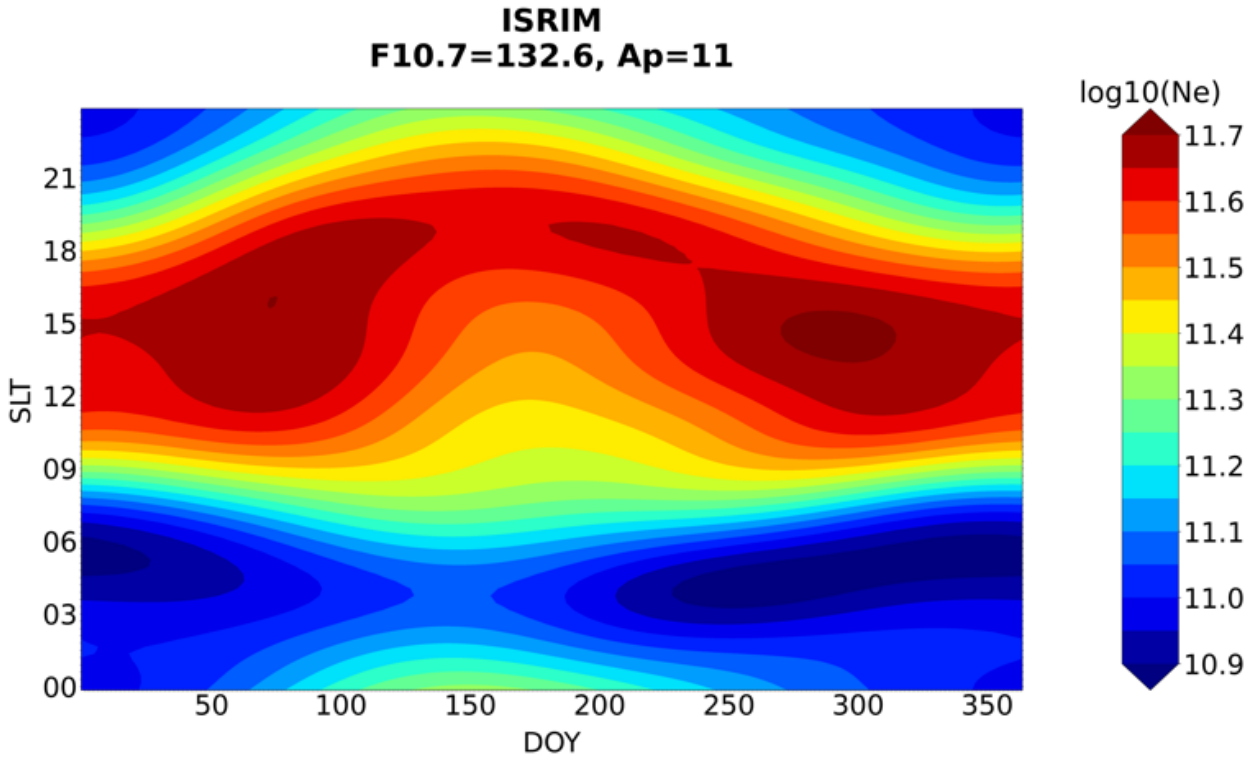
5.3.3 Climatological Analysis

The climatological study can verify whether the NN models can keep track of *Ne* characteristics at a long temporal scale. For comparison, the ISRIM [*Holt et al.*, 2002; *S R Zhang and Holt*, 2007; *S R Zhang et al.*, 2005] was used, which is an open-source online tool for *Ne* climatological study under different conditions (altitude, geodetic latitude, F10.7, and Ap3). The annual *Ne* patterns from ISRIM (Figure **5.4** (a)) and four NN models (Figure **5.4** (b) and (c)) in 2012 are all plotted for 24 hours × 365 days (or 366 for the leap years). Although the NN models were trained using 1-hour data, they are flexible to take any time inputs, i.e. capable of automatic interpolation. As the best temporal resolution of ISRIM is 18 min, we used the best available time cadence 4-min of ISR data for NN models to achieve the smooth looking of these climatological

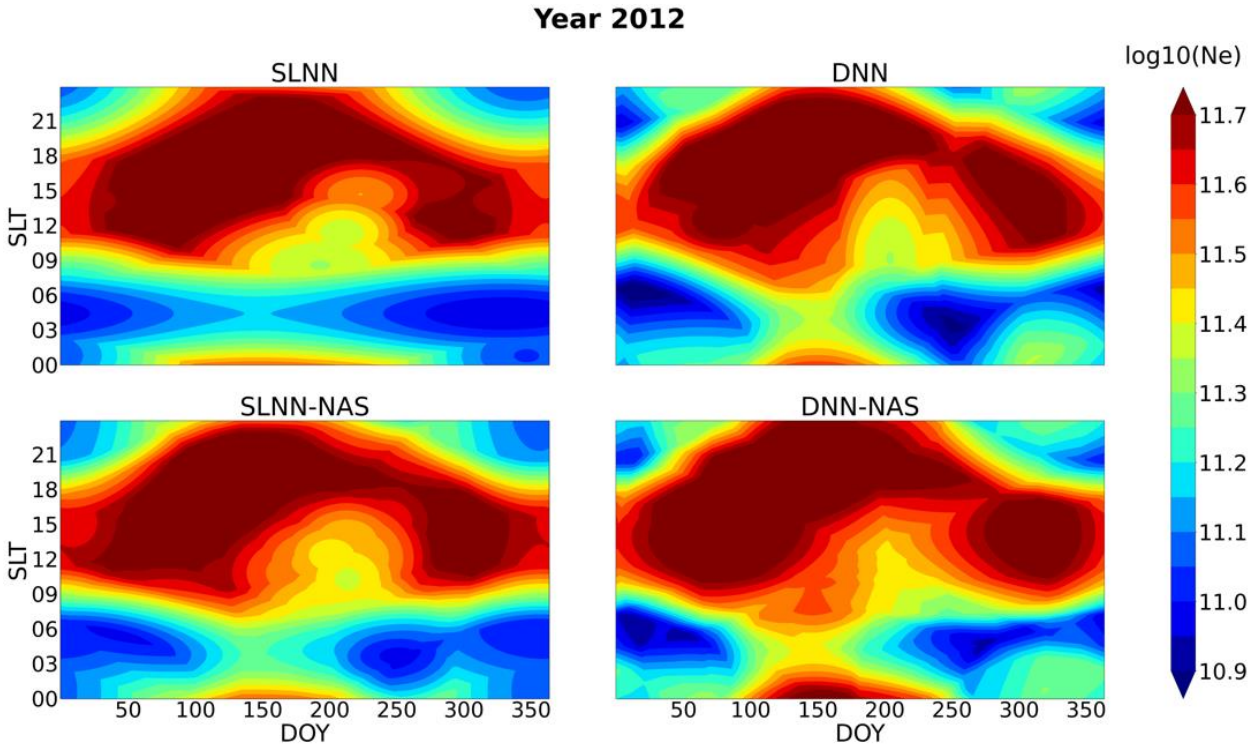
Ne maps without introducing any artificial structures. Note that as ISRIM used the fixed altitude, F10.7, and Ap3, and the four NN models were run with the same fixed values to obtain Figure **5.4** (b). All NN models reproduce an asymmetric semi-annual pattern of *Ne* as shown in ISRIM, which resembles as a saddle-like structure with *Ne* concentration peaks in Spring and Fall. The two SLNN models show more choppy edges on the crests, which could imply the incapability of the simple architecture to fully catch the data characteristics. DNN-NAS seems to have two more appealing crests, while the other three NN models suffer a star like artifact at the center. The 14 isolated

thread-like enhancements in Figure **5.4**(c) could be the indication of 27-day mid-latitude topside ionospheric electron variation [*Rich et al.*, 2003].

(a) ISRIM climatological pattern of medium solar activity.



(b) semi-annual patterns of climatological study.



(c) semi-annual patterns based on external geophysical indices.

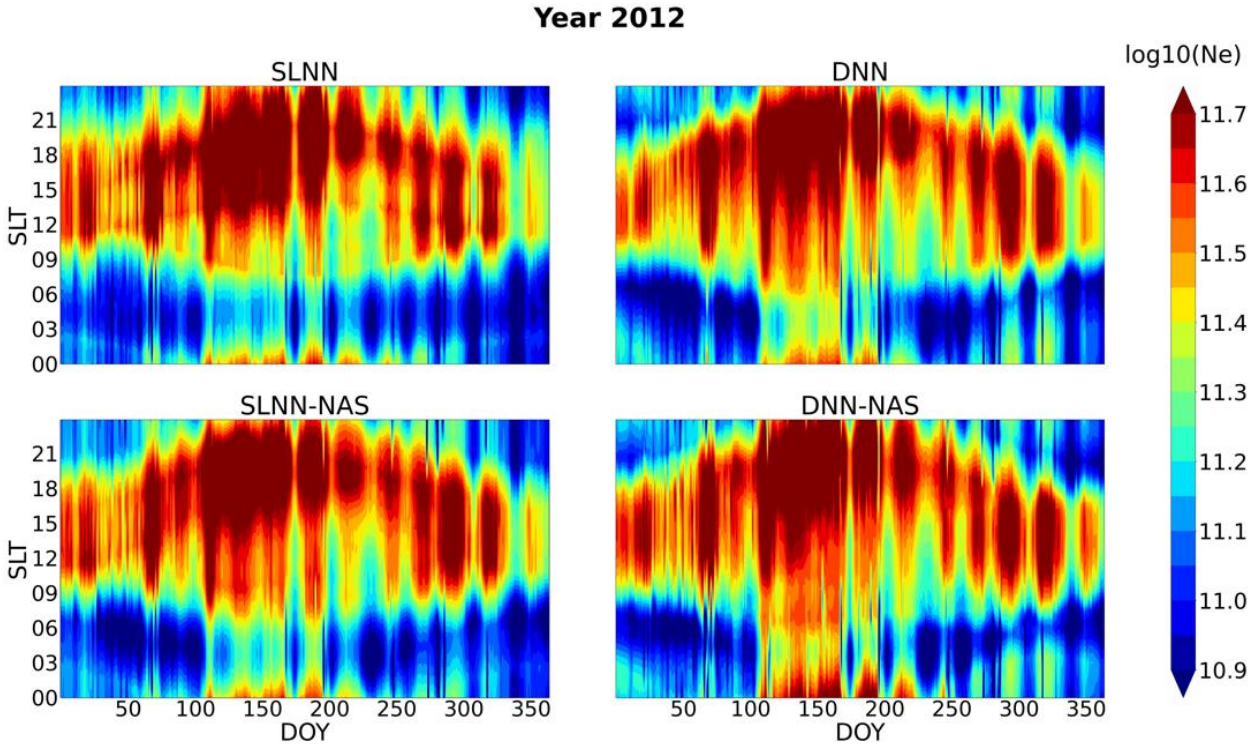


Figure 5.4 *Annual electron density patterns of year 2012 from different sources: (a) ISR empirical model (ISRIM), (b) four model predictions based on the fixed F10.7 and Ap3, (c) four model predictions based on the time-varying F10.7*

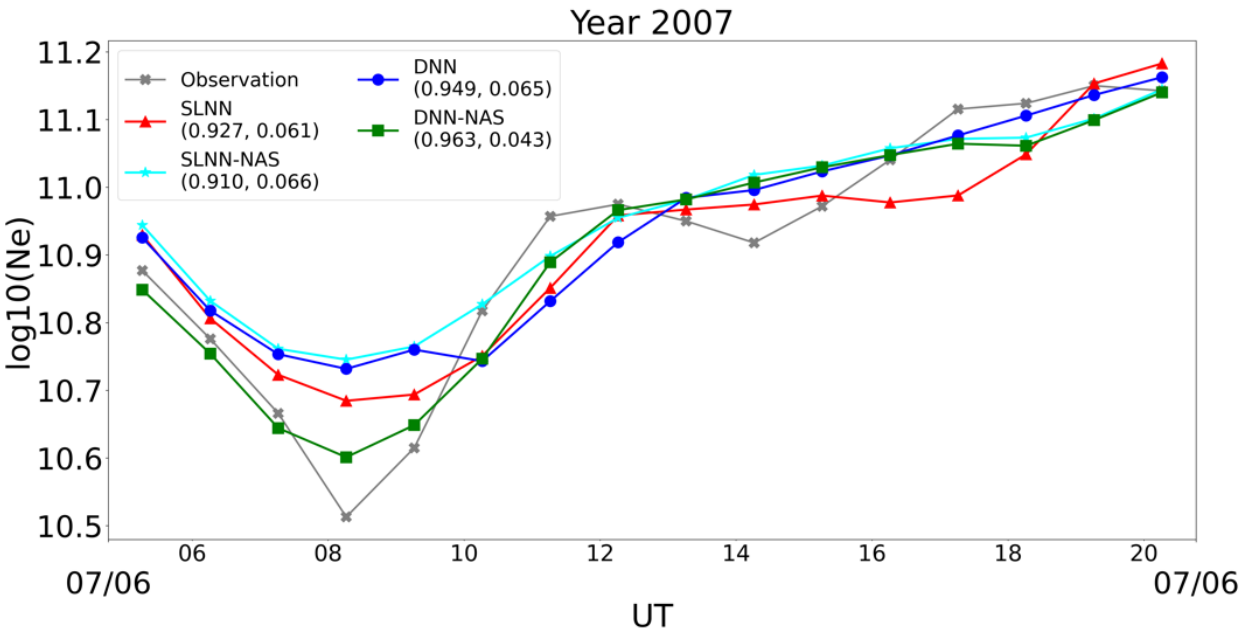
and Ap3. Based on the nature of neural network models, the input can be arbitrary values. We set the evenly distributed temporal information to get the time related drivers (year, DOY, and SLT), while comparison between (a) and (b) serves as the comparison on the climatological study, while (c) demonstrates a more realistic case of Ne annual pattern with time-varying F10.7 and Ap3 inputs.

5.3.4 Daily *Ne* Pattern Prediction

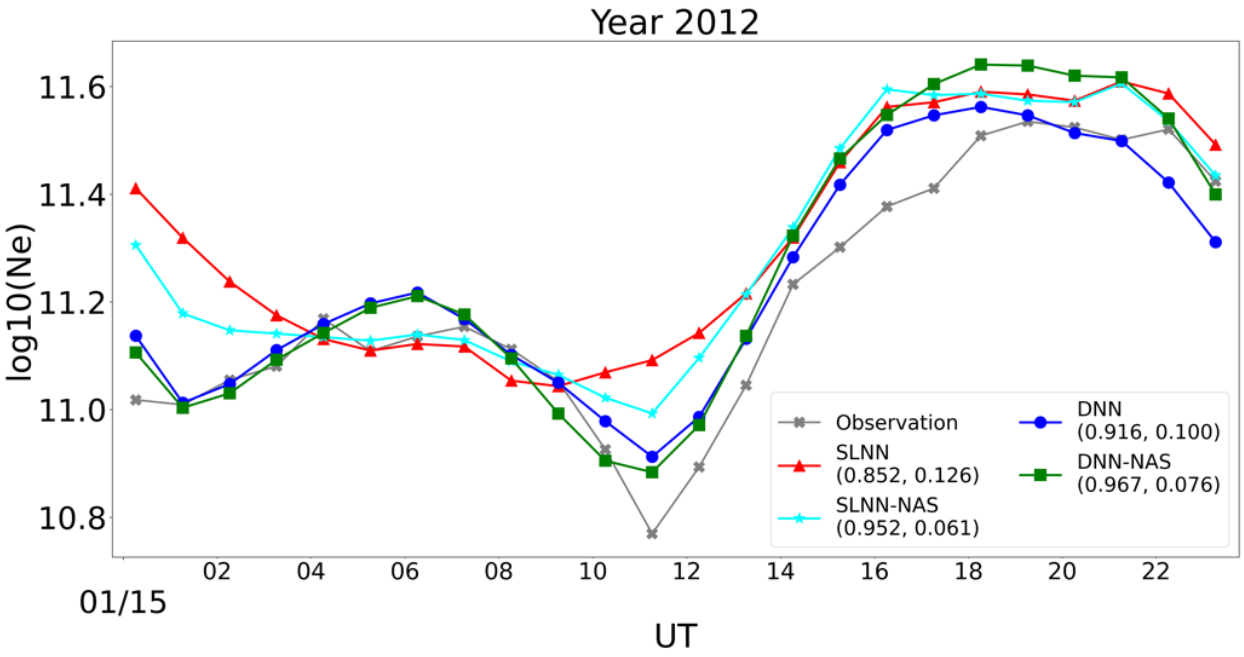
To evaluate the model performance, the daily *Ne* patterns were compared to illustrate how well the models predict in a resolved temporal scale varying from annual to daily. All the drivers (year, cyclic DOY and cyclic SLT, F10.7, and Ap3) served to get the model output. The two geophysical indices were obtained from OpenMadrigal database of MIT Haystack Observatory if not available in ISR. Since the days with full hourly *Ne* coverage are limited in the ISR data, we have identified a total of 128 days in the test data with a decent full-day hourly coverage. Three examples of hourly changes of *Ne* in a day (2007-07-06, 2012-01-15, and 2012-08-01) are shown in Figure **5.5** with observations and different model outputs. Note that 2007-07-06 and 2012-08-01 do not have a full 24-hour coverage. To quantify the agreement between the prediction and the observation, Pearson correlation coefficients (CCs) and MAEs are calculated and shown in Figure **5.5**. The higher CC values indicate the better trend match (with the removal of the mean and normalization) and the lower MAEs indicate less discrepancies between prediction and observation. Between 06-10 UT, Figure **5.5**(a) (2007-07-06) overestimated the trough and Figure **5.5**(c) (2012-08-01) underestimated the trough. For a particular day, the NN models could either underestimate (Figure **5.5**(c)) or overestimate (Figure **5.5**(b)). As shown in Figure **5.3** of the BA plots, although there are large positive prediction differences for low *Ne* observations, the overall biases of all four NN models are as small as several tenth of a percent of the average $\log_{10}(Ne)$. In general, all NN models follow the observation patterns (gray cross) well, but DNN-NAS achieves the largest CC (and the smallest MAE except for 2012-01-15). For 2012-01-15 in Figure **5.5** (b),

the dip is later than the other two cases since the sun rises later in winter than in summer. Figure **5.5** shows that DNN-NAS predicts the observations better than the other three models, which are the dominant cases in all 128 days with a good daily coverage.

(a) 2007-07-06



(b) 2012-01-15



(c) 2012-08-01

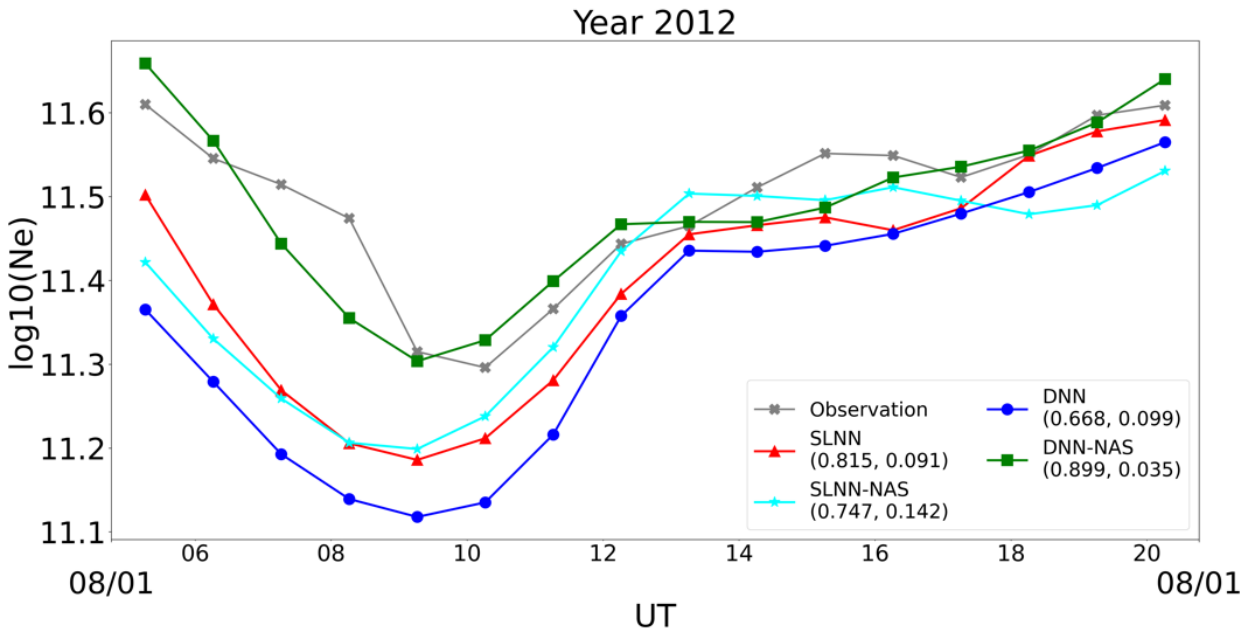


Figure 5.5 *Daily Ne pattern prediction on three different days: (a) 2007-07-06, (b) 2012-01-15, and (c) 2012-08-01. Gray cross: the ISR observation; red triangle: SLNN; cyan star: SLNN-NAS; blue circle: DNN; green square: DNN-NAS. The two parameters (Pearson correlation coefficients and MAE) help evaluate how well model outputs predict the observed diurnal Ne pattern. Generally, all model outputs follow the observed diurnal pattern well, while DNN-NAS predicts the best.*

We calculated CC and MAE for all 128 daily patterns from the test set and ranked four models. The number of ranks for four models are shown in Table **5.4**. Specifically, 1-4 ranks are corresponding to the decreasing CC or increasing MAE. For example, rank 1 represents the largest CC or the least MAE, which corresponds the best prediction of daily pattern. And rank 4 represents the smallest CC or the largest MAE, which corresponds the worst prediction of daily pattern. As can be seen, DNN-NAS has a dominatingly good prediction performance with 61 (48%) for CC (rank #1) and 54 (42%) for MAE (rank #1).

Table 5.4 *The ranks for daily pattern predictions. Among the 128 days in the test set, the Pearson correlation coefficients (CCs) and mean absolute errors (MAEs) are calculated and sorted from best (highest CC or lowest MAE). The DNN-NAS shows the greatest number of rank 1 cases.*

		SLNN	DNN	SLNN-NAS	DNN-NAS
CC	Rank 1	25	16	26	61 (48%)
	Rank 2	26	35	41	26
	Rank 3	32	48	31	17
	Rank 4	45	29	30	24
MAE	Rank 1	17	30	27	54 (42%)
	Rank 2	34	32	33	29
	Rank 3	29	32	44	23
	Rank 4	48	34	24	22

To assess the performance of NN models during several continuous days, the duration with decent observation coverage is selected for a further comparison. Two indices as drivers (F10.7 and Ap3) are shown in the upper panel of Figure **5.6**. A cubic interpolation (up to 3rd order of polynomials) is applied to the Ap3 index for the reference purpose. In Figure **5.6**, both model predictions and observations show a strong correlation to Ap3. For this specific week, the Ap3 index varies from quiet time (near 20) to moderate active values more than 40. It has been noticed that *Ne* drops below 10.9 at the post-midnight of 2012-09-05 (as indicated by the arrow), which seems to be caused by the Ap3 peak at 00 UT of 2012-09-05. (Another Ap3 peak on 2012-09-03 caused a similar but smaller *Ne* dip at the post-midnight of 2012-09-04 although ISR data were not available during this period.) All NN models seem to track these changes well. For SLNN model prediction, the narrower *Ne* variation matches ISR observation well with least MAE. DNN-NAS seems to track the observation best, especially the *Ne* dip on 2012-09-05, resulting the highest

CC. Please refer to Table **5.4**. Although DNN-NAS has the overall the greatest number of best performances on MAE and CC, the other methods also could have ranked higher in some cases.

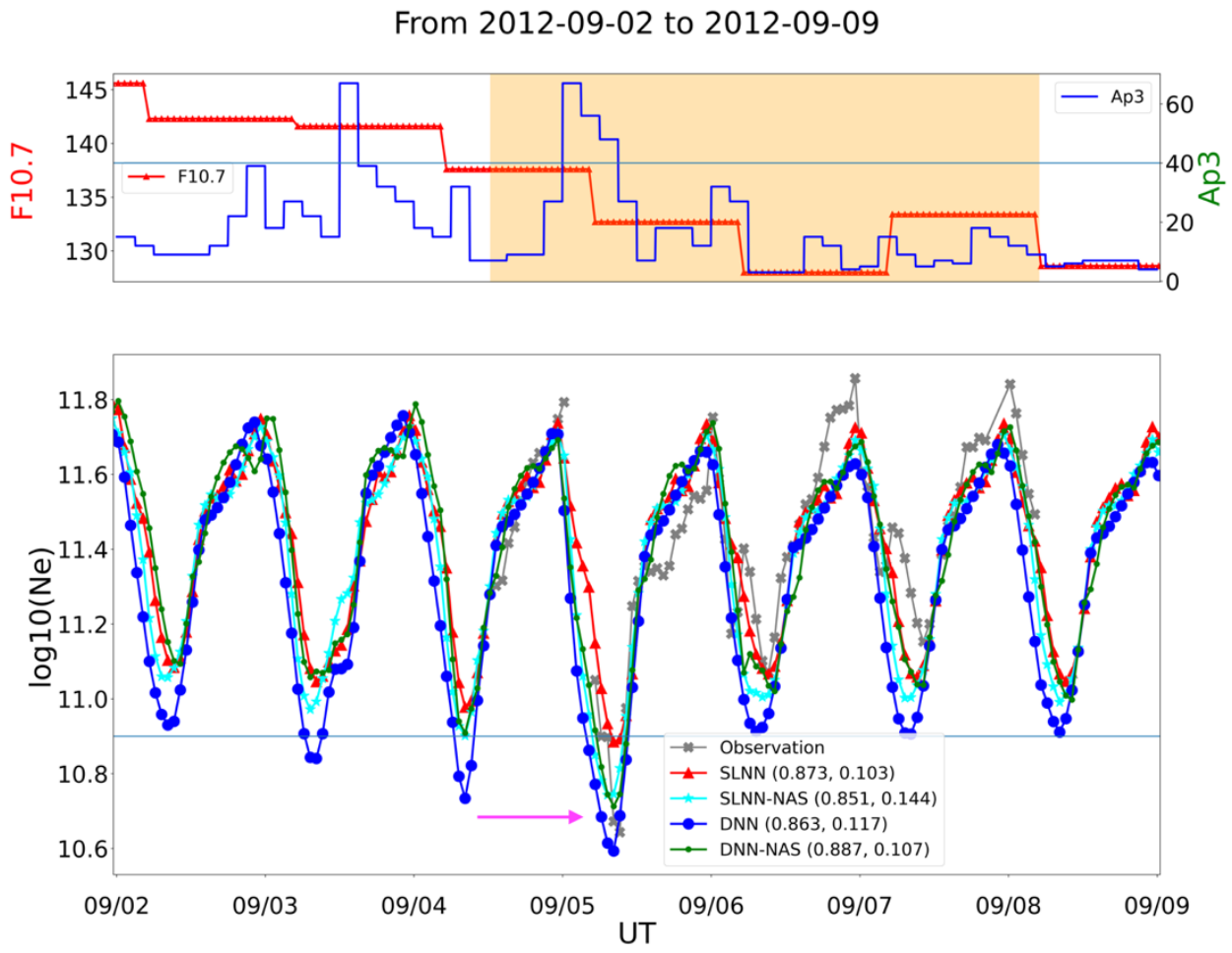


Figure 5.6 Ne patterns during 2012-09-02 to 2012-09-09. The two geophysical drivers are drawn in the upper panel with the shaded region indicating the continuous ISR observation available. Four model outputs are of different markers followed with CCs and MAEs (based on observational values) in parentheses. Clearly, we see the Ap3 serves as the major driver effect to the model outputs as the predictions dip down when Ap3 reaches its peak at early time of 2012-09-05. The two horizontal lines at 40 for Ap3 and 10.9 for $\log_{10}$(Ne) serve as references.

5.4 Discussion

In this study, we have shown that NAS helps find an optimal neural network setting to reduce the *Ne* prediction errors for both SLNN and DNN. Furthermore, NAS could make the

process more efficient with little manual interventions. Generally, we started with a large and sparse search poll of assigned hyperparameters. Based on the behavior of loss curves, the search poll was refined to reach the optimal neural architecture. The multiple GPU cores facilitated this hierarchic search. The manual determination of the optimal network structure is a daunting work. For example, with a fixed number of three hidden layers, thousands of full trainings were performed to obtain the number of neurons in each layer [*W Li et al.*, 2021]. Even the simple selection of the optimal learning rate could involve a substantial amount of manual work as we did for the two manual models. The NAS provides an efficient way to identify the optimal hyperparameters for NN models. For the current simple application of NAS for *Ne* prediction at the fixed geophysical location and altitude, the search process is fast (about 33 minutes on NVIDIA A6000, 22 minutes for NAS search and 11 minutes for additional epochs). However, the converging status of training and validation curves is absent in the early-stopping search. Considerate amount of manual work is still required to run additional epochs based on the NAS guided architectures and analyze the loss curves. Thus, we would assume more advanced NAS application could further reduce the tedious work spent in optimizing the neural networks.

Overfitting remains a general concern with machine learning models. As shown in Figure **5.2**, the training loss could be continuously reduced. As a matter of fact, when we used a complex NN model, the fitting error can approach a very low value at the cost of reducing model generalization to an acceptable level with high prediction errors. Thus, the validation dips chosen in Figure **5.2** alleviate this issue. Furthermore, NAS uses an early-stopping criterion for an efficient search. For highly nonlinear problems, NAS could trap in a local minimum. We used multiple random initializations for NAS to avoid this problem. DNN-NAS stands out in the overall quantitative measurements, climatological study, and prediction rankings of daily patterns.

All NN models predict *Ne* well during the moderate event in the daytime section (Figure **5.6**). This is consistent with previous studies of *Ne* prediction using NN models and due to a couple of reasons. First, the training data are confined to the condition (Ap3 $\leq$ 80 in Table 5.1), which causes the NN models to be prone to these cases. Secondly, the physical drivers are not fine enough in time, e.g. F10.7 is a daily average and Ap3 is 3-hour average. We conducted an additional training of DNN-NAS without the restriction on Ap3 (i.e, Ap3 could be larger than 80 which covers intense storm periods), namely DNN-NAS*. The comparison between DNN-NAS and DNN-NAS* is shown in Figure 5.7. The shade region is approximately from 05 UT to 15 UT on 2012-03-09. Though DNN-NAS has overall better CC and MAE, DNN-NAS* showed a much larger CC and lower MAE than DNN-NAS in the shade region. However, both models struggle to track the *Ne* dip around 08UT on 2012-03-10. As the ISR data with Ap3 $\geq$ 80 are only account for less than 2% of the total data, it is not a surprise that DNN-NAS* only improved over DNN-NAS in certain regions and suffered performance loss in other regions [*Gowtam et al.*, 2019; *Habarulema et al.*, 2021; *W Li et al.*, 2021]. In future work, either a separate model for major geomagnetic events or a general model with different weights on these events should be built with more event data to address this challenging problem.

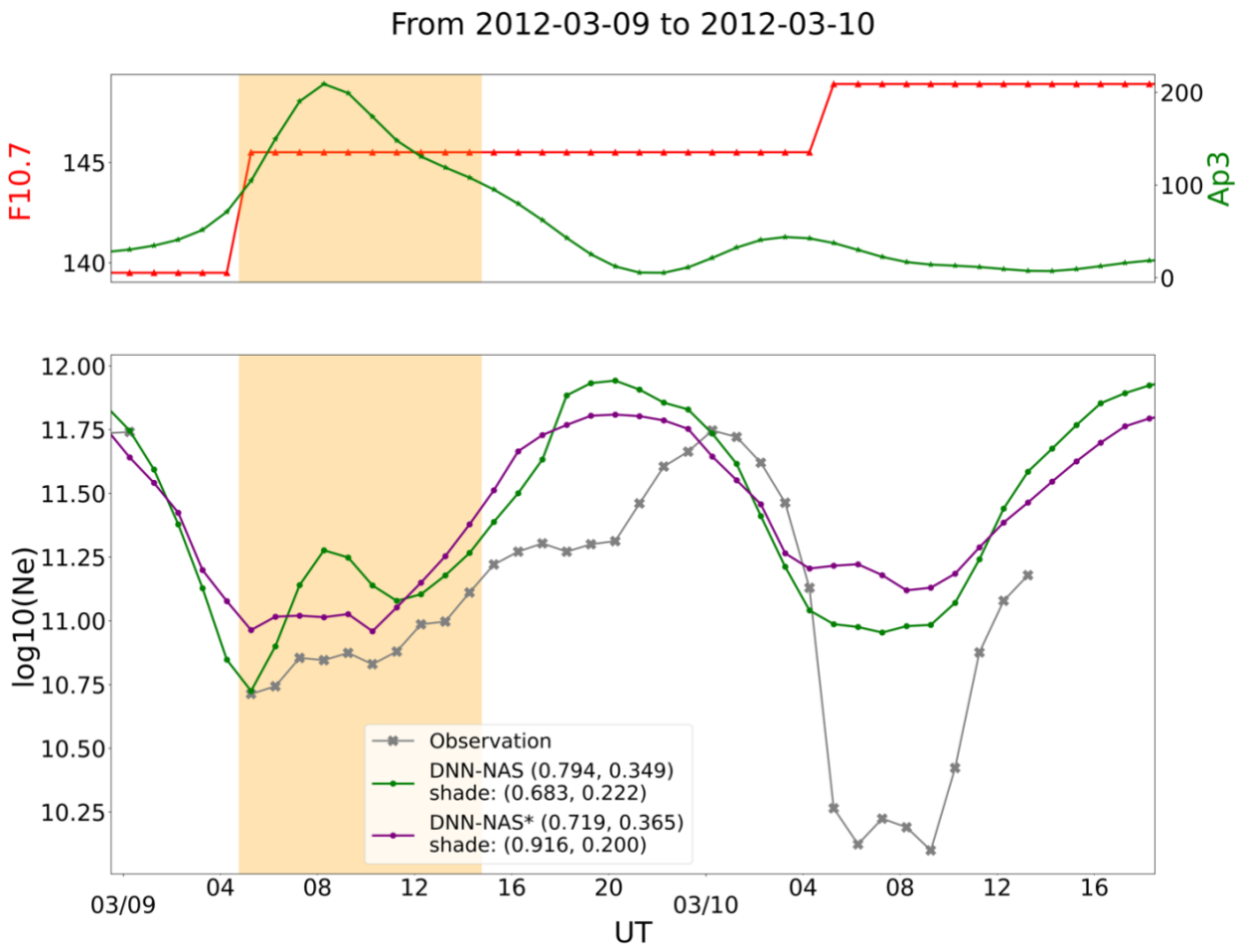


Figure 5.7 DNN-NAS trained with Ap3≤80 and DNN-NAS trained without the restriction on Ap3, the DNN-NAS models trained with and without filter on Ap3 have the prediction results in green and purple color. The CC and MAE calculated on the observational data are in the parentheses (the whole curve after the model name and the shade region after "shade").*

One possible reason for the limited performance improvement of NAS models over fixed NN models may be the lack of sufficient training data. To address this issue, we applied cubic B-spline to the vertical profiles of *Ne* with 15-minute cadence. After removing the abnormal data points, a total of nearly 43,000 data points around 350 km were used for training/validation/test (where the test set was changed to 2007 and 2016 in order to balance the amount of validation and test data), about 4 times of data points with 1-hr cadence (11,483). In the following sections, we

call the data with 1-hr cadence as the 1-hr dataset (2007 and 2012 as test data) and that with 15-minute cadence as the 15-min dataset (2007 and 2016 as test data). However, the models trained on the 15-min dataset led to similar findings as those using the 1-hr dataset, i.e. NAS led to only marginal improvement of *Ne* prediction. Therefore, the lack of sufficient training data may not be the primary reason for the limited improvement of NAS models. There are a couple of caveats for modeling of the interpolated data study. First, the interpolation may introduce artifacts that deviate from the true data. This part of study merely investigated how the amount of data affect NAS, complexity, and overfitting. The real models are preferred to be built on the original data without simple interpolations. Secondly, the interpolation does not inherently introduce data variability. Since the peak electron density around 350 km altitude is known to exhibit significant *Ne* variations with changing ionospheric conditions, it may be challenging to predict these large variations without sufficient data (e.g. extra observations in addition to ISR) or more advanced deep learning techniques.

We further conducted a complexity analysis of NN by increasing the number of network weights of a SLNN (denoted as "Complexity") on the 15-min dataset. The validation loss is plotted with the change of the complexity of SLNN in Figure 5.8. As can be seen, the loss function drops quickly at the beginning and converges to a steady level slightly below 0.125 after the complexity reaches 128 network parameters, i.e. 16 neurons in the hidden layer. Therefore, for the ISR *Ne* data, fully connected NN seems to reach its performance limit at a simple structure. This explains why NAS could only achieve limited improvement over fixed NN models, which are already complex enough to model the data in hand.

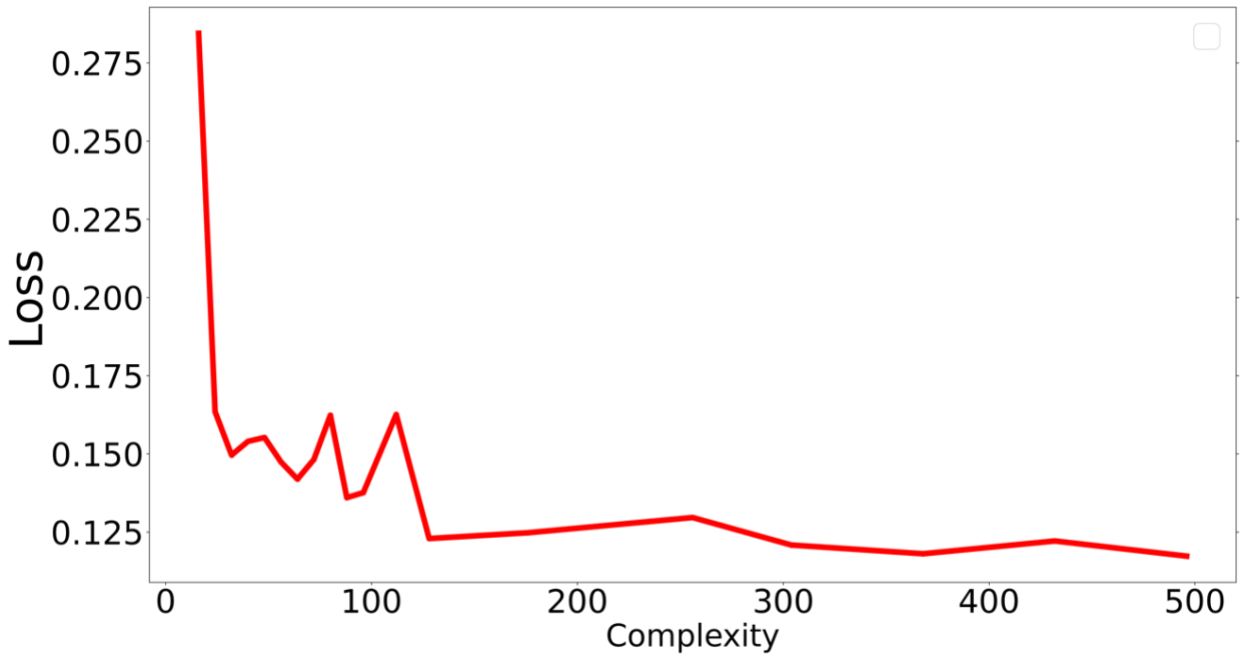
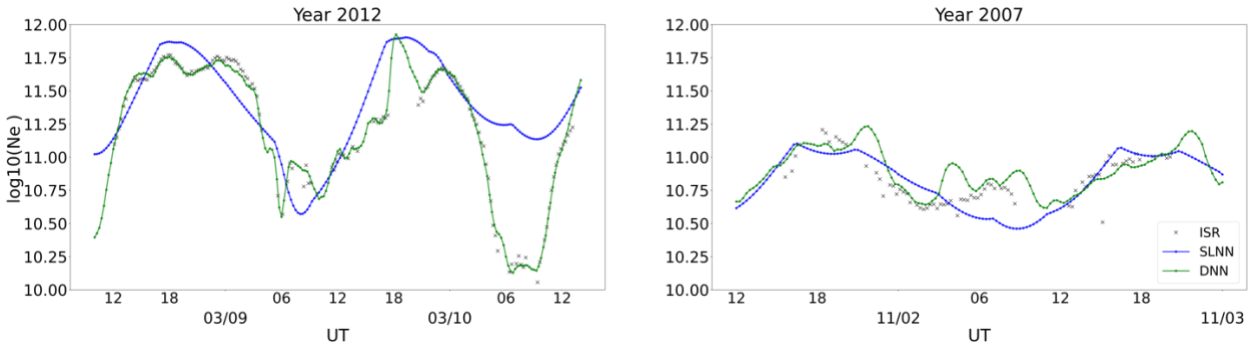


Figure 5.8 Prediction performance changes along with the model complexity. The complexity is defined as the total number of trainable weights of the NN model. The mean absolute error of the validation set serves as the loss function, where the less loss indicates the better performance.

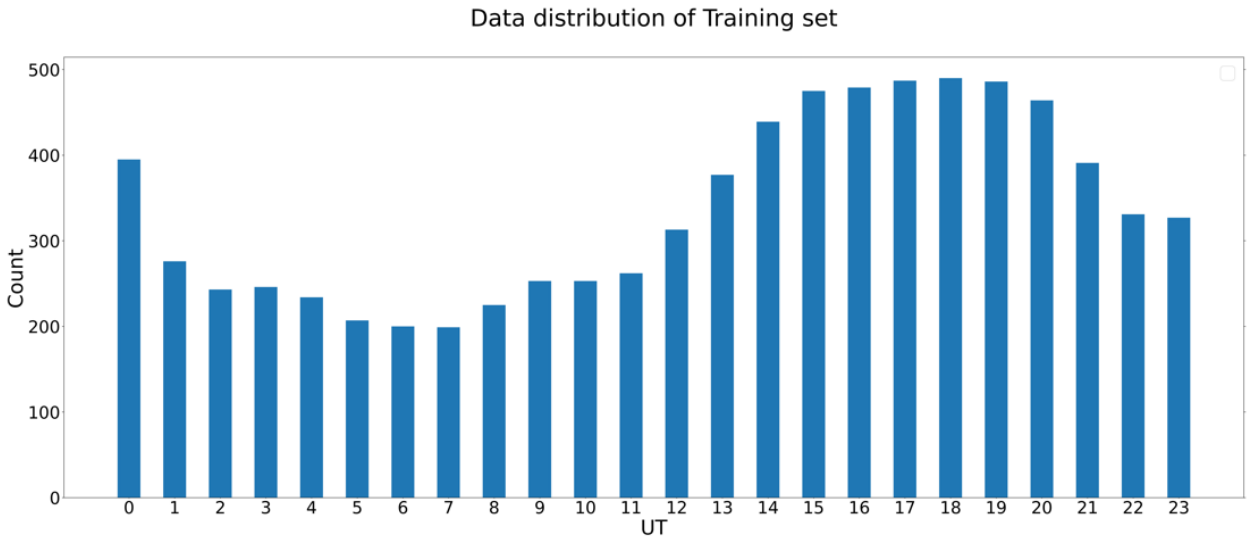
It is also worth noting that DNN-NAS could achieve much better fitting, but at the cost of losing the generality. For an overtrained DNN-NAS model ([512, 512, 512, 512, 32]), the training MAE (after 7,900 epochs of training) is as low as 0.0529, compared to 0.1261 of SLNN for the 15-min dataset. However, the test MAE dropped to 0.1587, compared to 0.1285 of SLNN, which indicates the loss of generality of DNN-NAS. In Figure 5.9(a) for the training data (2012 for the 15-min dataset), the overtrained DNN-NAS can fit the complicated structures on March 9th and dip on March 10th of the observations, while SLNN fails to catch these structures. However, in Figure 5.9(b) for the test data (2007 for the 15-mind dataset), DNN-NAS shows some abnormal oscillations as the signs of overfitting.



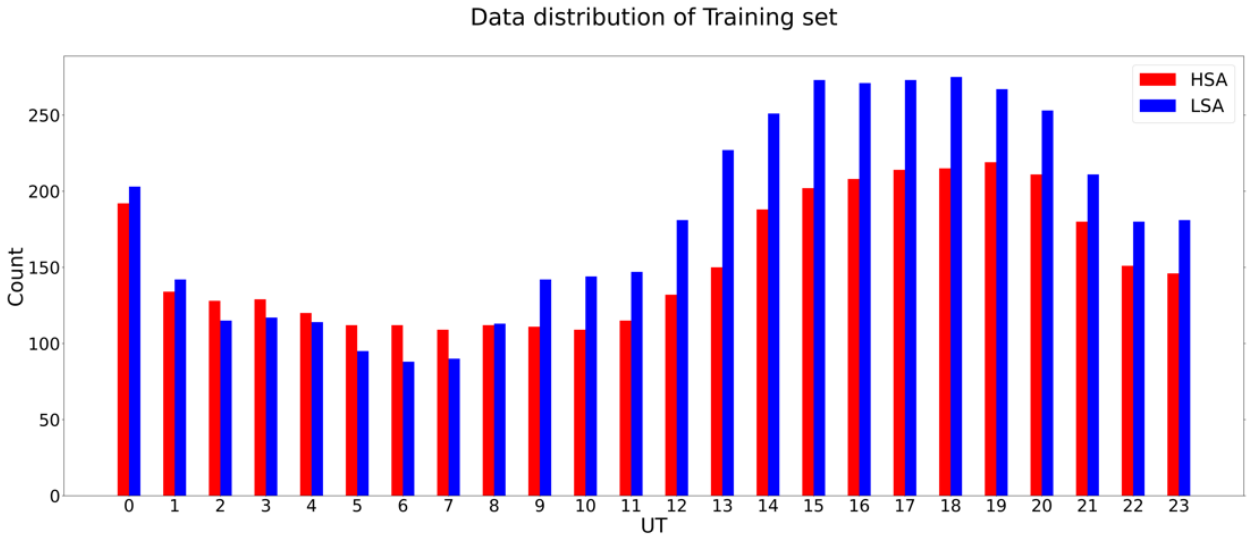
(a) (b)

Figure 5.9 Overfitting of DNN (architecture: [512, 512, 512, 512, 32], green) (a) fitting and (b) prediction. SLNN (18 hidden neuron, blue) is served as a benchmark. DNN can fit the ISR data more closely than SLNN as shown in (a). However, DNN leads to an unrealistic wavy pattern for prediction as shown in (b).

Furthermore, the histogram of the training data across UT hours is shown in Figure 5.10(a). As can be seen, there is a trough between 0-12 UT (nighttime at the Millstone Hill). Usually, NN models are trained better with more data. This data imbalance (37% for 00-12 UT and 63% for 12-24 UT) may be a partial reason that the *Ne* prediction performance of all NN models is worse in 00-12 UT (nighttime) than 12-24 UT (daytime) as shown in Figure **5.5** in addition to the different dynamics during nighttime and daytime. We also plotted the histogram of training data in high solar activity (HSA) years (> 90 sfu) and low solar activity (LSA) years (≤ 90 sfu). They follow the same pattern of the total data, while the amount LSA data (54% of the total) is more than that of HSA (46% of the total). Again, this is consistent with Figure **5.5**, where the prediction in the summer of 2007 (LSA year, Figure **5.5** (a)) is better than that in the summer of 2012 (HSA year, Figure **5.5** (c)). In addition, we calculated the prediction errors for two time periods of 00-12 UT and 12-24 UT (Table 5.5) as well as HSA and LSA years (Table 5.6). These results further confirm that the larger the data amount for training, the better the prediction performance.



(a)



(b)

Figure 5.10 The training data distribution across UT: (a) the total data; and (b) the data for high solar activity (HSA) years (>90 sfu) and for low solar activity (LSA) years (<90 sfu). The amount of data is more in 12-24 UT (63% of the total) than in 00-12 UT (37% of the total).

Table 5.5 Prediction performance of different NN models for 00-12 UT and 12-24 UT. The prediction performance of 12-24 UT is much better than that of 00-12 UT because more data in 12-24 UT were available for training (63% vs 37%).

		SLNN	*DNN*	*SLNN-NAS*	*DNN-NAS*
MAE	00-12 UT	0.1727	0.1584	0.1597	0.1439
	12-24 UT	0.1082	0.1021	0.1005	0.0973
RMSE	00-12 UT	0.2329	0.2136	0.2172	0.2004
	12-24 UT	0.1449	0.1389	0.1365	0.1325
RE (%)	00-12 UT	1.5946	1.4593	1.4740	1.3280
	12-24 UT	0.9617	0.9078	0.8933	0.8653

Table 5.6 Prediction performance of different NN models for high solar activity (HSA) years and low solar activity (LSA) years. The prediction performance of LSA is generally better than that of HSA because more available data in LSA years were available for training (54% vs 46%).

		SLNN	*DNN*	*SLNN-NAS*	*DNN-NAS*
MAE	HSA	0.1350	0.1307	0.1297	0.1189
	LSA	0.1269	0.1250	0.1197	0.1139
RMSE	HSA	0.1847	0.1762	0.1753	0.1627
	LSA	0.1808	0.1656	0.1673	0.1598
RE (%)	HSA	1.2013	1.1283	1.1282	1.0270
	LSA	1.1932	1.0996	1.0929	1.0460

Finally, the current study confines to *Ne* prediction at a fixed latitude and altitude in order to investigate the effectiveness of different NN models. 3D NN models have been proposed using the ionospheric radio occultation measurements in previous studies [*Gowtam et al.*, 2019; *Habarulema et al.*, 2021; *Sai Gowtam and Tulasi Ram*, 2017; *Tulasi Ram et al.*, 2018]. The static nature of fully connected NN is also accountable for the limited prediction performance of this study (in line with previous studies) as electron density change is a dynamic process, influenced by different geomagnetic parameters or other factors at different space and time scales. For example, the increase of Ap3 affects neutral density, which can cause the electron density change

over the next few hours rather than the instant change. Though the geophysical indices serve as the drivers in many developed models [*X Chu et al.*, 2017; *X N Chu et al.*, 2017; *Habarulema et al.*, 2021; *W Li et al.*, 2021], the atmospheric neutral components at Millstone Hills, which have shown strong correlations with electron density, may not be accurately described by the current input parameters of the NN models (F10.7 and Ap3). Technically, the more advanced generative models with the time histories of the input parameters may lead to much more improved prediction than the fully connected NN models without memory mechanism. Besides, this study examined the feasibility of applying NAS in identifying an optimal network structure of future works on either building electron density vertical profile based on ISR or other electron density models. Combined with aforementioned technical advancement, electron density prediction offered by deep learning could be significantly improved. And new drivers may be needed to accommodate the resolved temporal resolution, such as adding the 81-day average F10.7 (F10.7p) for the historical information or the geomagnetic AE index, and the physical processes, such as neutral composition, in our future work. Last but not least, information theory can help identify and select the drivers and their time histories that are relevant for predicting the output parameter, e.g., solar wind parameters [*S. Wing et al.*, 2016; *Simon Wing et al.*, 2022a; *Simon Wing et al.*, 2022b].

6 Summary and Future Work

6.1 SUMMARY

In this work, we have demonstrated the applications of deep learning on addressing a couple of ionospheric research topics: 1) Completion of the global total electron content (TEC) map using generative adversarial networks (GANs); 2) prediction of the regional ionospheric electron densities at a fixed height. Our results show the superior performance of these deep learning techniques.

For TEC map completion, DCGAN with Poisson blending (DCGAN-PB) was first proposed to learn and complete the IGS-TEC maps using random and MIT-TEC masks. DCGAN-PB can learn from thousands of TEC maps in different conditions and extract useful features. Then it overcomes the challenges by effectively recovering the missing data in a large area, which is not covered by GNSS observation network. The performance of DCGAN is strongly influenced by the amount of training data and PB can significantly improve the final completed maps.

The more advanced SNP-GAN model can further improve TEC map completion performance over the DCGAN-PB model both quantitatively and qualitatively. The end-to-end structured generator of SNP-GAN leads to better TEC completion performance and shorter inpainting time than DCGAN-PB. Both deep learning models can maintain the large-scale structures as shown in the IGS-TEC maps and show the potential on recovering the mesoscale structure from sparse MIT-TEC data, which is absent in the IGS-TEC maps. In summary, SNP-GAN will be a tool of choice for TEC map completion thanks to its unique capability of revealing critical TEC structures and much improved completion efficiency.

As our focus has shifted from the global to the regional, and from TEC to electron density, the regional electron density recorded by the incoherent scatter radar suffers a lot of temporal and spatial gaps. We thrived to pursue an automatic way to optimize the neural network structure for best *Ne* prediction performance. Among the neural network models for *Ne* prediction at a fixed height in the topside ionosphere using 16-year ISR observations at Millstone Hill, we demonstrate that neural architecture search (NAS) is able to achieve this goal. In addition to modeling efficiency, NAS derived DNN models also lead to better prediction performance than manually tuned SLNN (more than 10% improvement on MAE and RE) and rank the highest for daily *Ne* pattern prediction based on CC and MAE. The climatological *Ne* patterns from different NN models reveal the two crests in Spring and Fall seasons in general. We also investigated the reason for limited improvement of NAS due to the network complexity and overfitting besides the lack of memory mechanism of the fully connected NN.

Among all the three applications above, a common practice is to investigate the cases under different conditions. For the split of training, validation, and test sets, we have classified the high solar activity (HSA) and low solar activity (LSA) based on the annual averaged F10.7. It is agreed that in HSA the errors tend to be larger than those in LSA, which can be explained by, not only either the absolute values of TEC/*Ne* are larger, but the environment is more dynamic during HSA years. A more active Sun brings solar winds, which later increases the chance of geomagnetic events such as storms. Besides, the amount of data plays a major role in training the deep learning model. Eighteen years of IGS-TEC data lead to much improved GAN models for TEC completion than only two years of data, while the Millstone ISR data with the observations concentrated during daytime and LSA renders better Ne prediction during these time periods as shown in Figure 5.10b,

where the period 12-24 UT contains around 63% observed data. A more balanced dataset or individual models under different conditions shall bring benefits to further improve these models.

6.2 FUTURE WORKS

For the GAN models on global TEC map, the upscaled resolution did provide a clearer view and better chance for us to identify the meso-scale structure as shown in Figure 4.7. Besides, the difficulty to train the model in a stable way is one of the major drawbacks among GANs. Therefore, a more advanced GAN capable of stably processing large amount of TEC map in a higher resolution shall be one of the main directions to proceed our further investigations. Meanwhile, the state-of-the-art Transformers-based models, , such as the chat generative pre-trained transformer (a summary on chat-GPT [*Ray*, 2023]), and the diffusion models (such as DALL-E [*Ramesh et al.*, 2021]) may be utilized in ionospheric research for much improved model capability.

Based on the review of latest IRI models [*Dieter Bilitza et al.*, 2022], not only does these models use common physical parameters such as the sunspot and solar index F10.7, but also they keep updating by referring to update-to-date external data source. Though IRI are designed to provide ionospheric parameters at a monthly scale, the deep learning models, such as recurrent neural network, may utilize IRI-2020 (latest version) as an input to further improve the prediction performance. The complete dataset shall largely benefit the ionospheric models aimed to either regional or global.

Among all three studies, we are tackling the issue of incomplete data using advanced numerical methods, such as deep learning. Another way to expand our capability of filling these

data gaps is through combining different types of observations. During the collection of MIT-TEC data, the satellite to ground receiver communication is greatly bound to the limited coverage of ground receivers, which results in incomplete global TEC map, particularly over the oceans. Meanwhile, the radio occultation (RO) is the technique which utilizes the delay and bend angle of signal propagation within ionosphere to obtain *Ne* information. The low earth orbiting (LEO) satellites can make the RO measurements. For example, RO data from COSMIC satellites provide a better global TEC coverage. In addition, CubeSat is a more recent technique, which consists of many lower cost satellites in LEO. However, these radio communication quality depends on the coherent signals reflected off sea ice, inland water bodies, and calm ocean surface [*Wang and Morton*, 2022]. To improve the quality of these radio signal communications, machine learning methods have also been applied for automatic anomaly detections [*Yunxiang Liu and Morton*, 2022]. For achieve better and reliable coverage on ionospheric parameters, it would be valuable to investigate the deep learning methods in the radio communications as another major future work.

APPENDICES

APPENDIX A

The architecture of the generator and discriminator of DCGAN-PB is shown in Figure A.1. The number on the top denotes the number of channels for convolution and that at the bottom denotes the dimension of the output matrix.

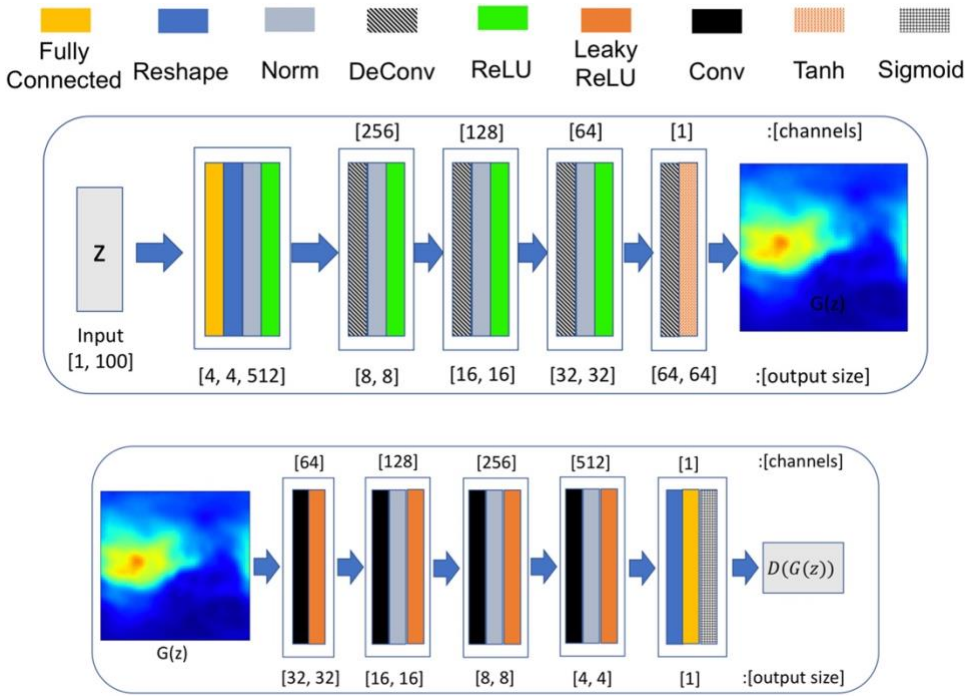


Figure A.1 The architecture of generator (upper) and discriminator (bottom) used in DCGAN-PB. (Norm: batch normalization; DeConv: de-convolution; ReLU: rectified linear unit; Conv: convolution; Tanh: hyperbolic tangent activation function; Sigmoid: sigmoid activation function.) [Pan et al., 2020].

For the generator (Figure A.1), the input layer projects and reshapes a random vector z (1×100) into a $4 \times 4 \times 512$ high-level feature matrix before passing through the batch normalization (Norm) (for stability of training) and the rectified linear unit (ReLU) (to introduce the nonlinearity). The three hidden layers have the similar structures with a deconvolution (DeConv) layer, a Norm layer, and a ReLU layer to convert the high-level features to details in

TEC maps. The output layer is composed of another DeConv layer and a hyperbolic tangent activation function (Tanh) to get the final TEC map. Note that four DeConv layers are fractionally strided convolution instead of real deconvolution, which converts the original $4 \times 4 \times 512$ feature matrix into the final 64×64 TEC map. Thus, the generator produces a 64×64 TEC map ($G(z)$) from an 1×100 random vector (z).

For the discriminator (Figure A.1), the TEC map, either the real IGS TEC map or the generated TEC map (shown in Figure A.1), is fed into four convolutional layers to extract the high-level features from the TEC map. Note that the input layer uses only a leaky ReLU and three hidden layers have both the Norm layer and leaky ReLU for a stable training. Leaky ReLU is used here to avoid the zero-activation problem of ReLU. Finally, the output layer reshapes the $4 \times 4 \times 512$ feature matrix into a $1 \times 4{,}096$ vector, and then a fully connected layer followed by a sigmoid function is used to produce a value between 0 and 1. Thus, the discriminator reads a TEC map and yields a single value $D(G(z))$, based on which the prediction is made to tell whether it is the true TEC map (>0.5) or the fake one (<0.5).

References

Afifi, M., and K. F. Hussain (2015), MPB: A modified Poisson blending technique, *Computational Visual Media*, *1*(4), 331-341, doi: 10.1007/s41095-015-0027-z.
Afraimovich, E. L., and E. I. Astafyeva (2008), TEC anomalies—Local TEC changes prior to earthquakes or TEC response to solar and geomagnetic activity changes?, *Earth, Planets and Space*, *60*(9), 961-966, doi: 10.1186/BF03352851.
Altman, D. G., and J. M. Bland (1983), Measurement in Medicine: The Analysis of Method Comparison Studies, *Journal of the Royal Statistical Society: Series D (The Statistician)*, *32*(3), 307-317, doi: https://doi.org/10.2307/2987937.
Appleton, S. E. (1956), Regularities and irregularities in the ionosphere, *Vistas in Astronomy*, *2*, 779-790, doi: https://doi.org/10.1016/0083-6656(56)90001-0.
Athieno, R., P. T. Jayachandran, and D. R. Themens (2017), A neural network-based foF2 model for a single station in the polar cap, *Radio Science*, *52*(6), 784-796, doi: 10.1002/2016rs006192.
Azeem, I., J. Yue, L. Hoffmann, S. D. Miller, W. C. Straka III, and G. Crowley (2015), Multisensor profiling of a concentric gravity wave event propagating from the troposphere to the ionosphere, *Geophysical Research Letters*, *42*(19), 7874-7880, doi: https://doi.org/10.1002/2015GL065903.
Baker, B., O. Gupta, N. Naik, and R. Raskar (2016), Designing neural network architectures using reinforcement learning, *arXiv preprint arXiv:1611.02167*.
Barnes, C., E. Shechtman, A. Finkelstein, and D. B. Goldman (2009), PatchMatch: a randomized correspondence algorithm for structural image editing, *ACM Trans. Graph.*, *28*(3), Article 24, doi: 10.1145/1531326.1531330.
Bertalmio, M., A. L. Bertozzi, and G. Sapiro (2001), Navier-stokes, fluid dynamics, and image and video inpainting, paper presented at Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 8-14 Dec. 2001.
Bilitza, D. (2001), International Reference Ionosphere 2000, *Radio Science*, *36*(2), 261-275, doi: Doi 10.1029/2000rs002432.
Bilitza, D., M. Pezzopane, V. Truhlik, D. Altadill, B. W. Reinisch, and A. Pignalberi (2022), The International Reference Ionosphere Model: A Review and Description of an Ionospheric Benchmark, *Reviews of Geophysics*, *60*(4), e2022RG000792, doi: https://doi.org/10.1029/2022RG000792.
Bilitza, D., D. Altadill, V. Truhlik, V. Shubin, I. Galkin, B. Reinisch, and X. Huang (2017), International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions, *Space Weather*, *15*(2), 418-429, doi: 10.1002/2016sw001593.
Bortnik, J., W. Li, R. M. Thorne, and V. Angelopoulos (2016), A unified approach to inner magnetospheric state prediction, *Journal of Geophysical Research-Space Physics*, *121*(3), 2423-2430, doi: 10.1002/2015ja021733.
Bradbury, N. E. (1938), Ionization, negative-ion formation, and recombination in the ionosphere, *Terrestrial Magnetism and Atmospheric Electricity*, *43*(1), 55-66, doi: https://doi.org/10.1029/TE043i001p00055.
Brandon, A. (2016), Image completion with deep learning in TensorFlow, *bamos/github. io:[сайт].–2016.–9 авг.–URL: https://bamos. github. io/2016/08/09/deep-completion/(дата обращения: 09.01. 2020)*.

Cai, H., L. Zhu, and S. Han (2018a), ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware, arXiv:1812.00332, doi: 10.48550/arXiv.1812.00332.
Cai, H., T. Chen, W. Zhang, Y. Yu, and J. Wang (2018b), Efficient architecture search by network transformation, paper presented at Proceedings of the AAAI Conference on Artificial Intelligence.
Cai, Y., X. Yue, W. Wang, S. Zhang, L. Liu, H. Liu, and W. Wan (2019), Long-Term Trend of Topside Ionospheric Electron Density Derived From DMSP Data During 1995–2017, *Journal of Geophysical Research: Space Physics*, *124*(12), 10708-10727, doi: 10.1029/2019ja027522.
Chen, Z., M. Jin, Y. Deng, J.-S. Wang, H. Huang, X. Deng, and C.-M. Huang (2019), Improvement of a Deep Learning Algorithm for Total Electron Content Maps: Image Completion, *Journal of Geophysical Research: Space Physics*, *124*(1), 790-800, doi: https://doi.org/10.1029/2018JA026167.
Chu, X., et al. (2017), A neural network model of three-dimensional dynamic electron density in the inner magnetosphere, *Journal of Geophysical Research: Space Physics*, *122*(9), 9183-9197, doi: 10.1002/2017ja024464.
Chu, X. N., J. Bortnik, W. Li, Q. Ma, V. Angelopoulos, and R. M. Thorne (2017), Erosion and refilling of the plasmasphere during a geomagnetic storm modeled by a neural network, *Journal of Geophysical Research-Space Physics*, *122*(7), 7118-7129, doi: 10.1002/2017ja023948.
Ciregan, D., U. Meier, and J. Schmidhuber (2012), Multi-column deep neural networks for image classification, paper presented at 2012 IEEE Conference on Computer Vision and Pattern Recognition, 16-21 June 2012.
Coster, A., and A. Komjathy (2008), Space Weather and the Global Positioning System, *Space Weather*, *6*(6), doi: https://doi.org/10.1029/2008SW000400.
Coster, A., J. Foster, and P. Erickson (2003), Monitoring the ionosphere with GPS, *GPS world*, *14*(5), 40-45.
Coster, A. J., L. Goncharenko, S.-R. Zhang, P. J. Erickson, W. Rideout, and J. Vierinen (2017), GNSS Observations of Ionospheric Variations During the 21 August 2017 Solar Eclipse, *Geophysical Research Letters*, *44*(24), 12,041-012,048, doi: https://doi.org/10.1002/2017GL075774.
Cummer, S. A., U. S. Inan, and T. F. Bell (1998), Ionospheric D region remote sensing using VLF radio atmospherics, *Radio Science*, *33*(6), 1781-1792, doi: https://doi.org/10.1029/98RS02381.
Davies, K., and D. M. Baker (1966), On frequency variations of ionospherically propagated HF radio signals, *Radio Science*, *1*(5), 545-556, doi: 10.1002/rds196615545.
Daw, A., A. Karpatne, W. D. Watkins, J. S. Read, and V. Kumar (2022), Physics-guided neural networks (pgnn): An application in lake temperature modeling, in *Knowledge Guided Machine Learning*, edited, pp. 353-372, Chapman and Hall/CRC.
Desell, T. (2017a), Large Scale Evolution of Convolutional Neural Networks Using Volunteer Computing, arXiv:1703.05422, doi: 10.48550/arXiv.1703.05422.
Desell, T. (2017b), Large scale evolution of convolutional neural networks using volunteer computing, paper presented at Proceedings of the Genetic and Evolutionary Computation Conference Companion.
Dey, S., S. C. Kanala, K. M. Chugg, and P. A. Beerel (2020), Deep-n-Cheap: An Automated Search Framework for Low Complexity Deep Learning, in *Proceedings of The 12th Asian Conference on*

Machine Learning, edited by P. Sinno Jialin and S. Masashi, pp. 273--288, PMLR, Proceedings of Machine Learning Research.
Dubey, S. R., S. K. Singh, and B. B. Chaudhuri (2022), Activation functions in deep learning: A comprehensive survey and benchmark, *Neurocomputing*, *503*, 92-108, doi: https://doi.org/10.1016/j.neucom.2022.06.111.
Dumoulin, V., and F. Visin (2016), A guide to convolution arithmetic for deep learning, arXiv:1603.07285, doi: 10.48550/arXiv.1603.07285.
Elsken, T., J.-H. Metzen, and F. Hutter (2017), Simple and efficient architecture search for convolutional neural networks, *arXiv preprint arXiv:1711.04528*, doi: https://doi.org/10.48550/arXiv.1711.04528.
Elsken, T., J. H. Metzen, and F. Hutter (2019), Neural Architecture Search: A Survey, *J Mach Learn Res*, *20*(1), 1997-2017, doi: https://doi.org/10.48550/arXiv.1808.05377.
Emery, W., and A. Camps (2017), Chapter 6 - Remote Sensing Using Global Navigation Satellite System Signals of Opportunity, in *Introduction to Satellite Remote Sensing*, edited by W. Emery and A. Camps, pp. 455-564, Elsevier.
Gangnet, M., and A. Blake (2003), Poisson image editing, paper presented at Acm Siggraph.
Godyak, V. A., R. B. Piejak, and B. M. Alexandrovich (1999), Effective electron collision frequency and electrical conductivity of radio frequency plasmas, *Journal of Applied Physics*, *85*(6), 3081-3083, doi: 10.1063/1.369646.
Gong, X., S. Chang, Y. Jiang, and Z. Wang (2019), Autogan: Neural architecture search for generative adversarial networks, paper presented at Proceedings of the IEEE/CVF International Conference on Computer Vision.
Goodfellow, I., Y. Bengio, and A. Courville (2016), *Deep learning*, MIT press.
Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative adversarial nets, *Advances in neural information processing systems*, *27*, doi: https://doi.org/10.48550/arXiv.1406.2661.
Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014), Generative Adversarial Networks, arXiv:1406.2661, doi: 10.48550/arXiv.1406.2661.
Gowtam, V. S., S. Tulasi Ram, B. Reinisch, and A. Prajapati (2019), A New Artificial Neural Network-Based Global Three-Dimensional Ionospheric Model (ANNIM-3D) Using Long-Term Ionospheric Observations: Preliminary Results, *Journal of Geophysical Research: Space Physics*, *124*(6), 4639-4657, doi: 10.1029/2019ja026540.
Guo, Z., X. Zhang, H. Mu, W. Heng, Z. Liu, Y. Wei, and J. Sun (2020), Single path one-shot neural architecture search with uniform sampling, paper presented at European conference on computer vision, Springer.
Habarulema, J. B., D. Okoh, D. Buresová, B. Rabiu, M. Tshisaphungo, M. Kosch, I. Häggström, P. J. Erickson, and M. A. Milla (2021), A global 3-D electron density reconstruction model based on radio occultation data and neural networks, *Journal of Atmospheric and Solar-Terrestrial Physics*, *221*, 105702, doi: 10.1016/j.jastp.2021.105702.
Hart, M. H. (1978), The evolution of the atmosphere of the earth, *Icarus*, *33*(1), 23-39, doi: https://doi.org/10.1016/0019-1035(78)90021-0.
Hegarty, C. J., and E. Chatre (2008), Evolution of the Global Navigation SatelliteSystem (GNSS), *Proceedings of the IEEE*, *96*(12), 1902-1917, doi: 10.1109/JPROC.2008.2006090.

Hernández-Pajares, M. (2004), IGS ionosphere WG status report: performance of IGS ionosphere TEC maps-position paper, paper presented at IGS workshop, Bern.
Hernández-Pajares, M., J. M. Juan, J. Sanz, R. Orus, A. Garcia-Rigo, J. Feltens, A. Komjathy, S. C. Schaer, and A. Krankowski (2009), The IGS VTEC maps: a reliable source of ionospheric information since 1998, *Journal of Geodesy*, *83*(3), 263-275, doi: 10.1007/s00190-008-0266-1.
Holt, J. M., S. R. Zhang, and M. J. Buonsanto (2002), Regional and local ionospheric models based on Millstone Hill incoherent scatter radar data, *Geophysical Research Letters*, *29*(8), 48-41-48-43, doi: 10.1029/2002gl014678.
Hutter, F., L. Kotthoff, and J. Vanschoren (2019), *Automated machine learning: methods, systems, challenges*, Springer Nature.
Isola, P., J. Y. Zhu, T. Zhou, and A. A. Efros (2017), Image-to-Image Translation with Conditional Adversarial Networks, paper presented at 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017.
Jakowski, N., S. Heise, A. Wehrenpfennig, S. Schlüter, and R. Reimer (2002), GPS/GLONASS-based TEC measurements as a contributor for space weather forecast, *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*(5), 729-735, doi: https://doi.org/10.1016/S1364-6826(02)00034-2.
Janssen, V. (2012), Likely impact of the approaching solar maximum on GNSS surveys: Be alert but not alarmed.
Ji, E.-Y., Y.-J. Moon, and E. Park (2020), Improvement of IRI Global TEC Maps by Deep Learning Based on Conditional Generative Adversarial Networks, *Space Weather*, *18*(5), e2019SW002411, doi: https://doi.org/10.1029/2019SW002411.
Jin, H., Q. Song, and X. Hu (2019), Auto-keras: An efficient neural architecture search system, paper presented at Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining.
Judd, F. C. (1987), *Radio Wave Propagation (HF Bands): Radio Amateur's Guide*, Heinemann.
Kazhdan, M., M. Bolitho, and H. Hoppe (2006), Proceedings of the fourth Eurographics symposium on Geometry processing, edited, The Eurographics Association.
Kelley, M. C. (2003), IONOSPHERE, in *Encyclopedia of Atmospheric Sciences*, edited by J. R. Holton, pp. 1022-1030, Academic Press, Oxford.
Kim, T. (2016), DCGAN-tensorflow, edited, DCGAN-tensorflow.
Kingma, D. P., and J. Ba (2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, doi: https://doi.org/10.48550/arXiv.1412.6980.
Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, *25*.
Kuznetsov, V. V., V. V. Plotkin, G. V. Nesterova, and I. I. Nesterova (1998), Universal variation of the <I>F</I>₂-layer critical frequency and solar activity, *Earth, Planets and Space*, *50*(1), 57-61, doi: 10.1186/BF03352086.
Laštovička, J., G. Beig, and D. R. Marsh (2014), Response of the mesosphere-thermosphere-ionosphere system to global change - CAWSES-II contribution, *Progress in Earth and Planetary Science*, *1*(1), 21, doi: 10.1186/s40645-014-0021-6.
Lei, J., et al. (2007), Comparison of COSMIC ionospheric measurements with ground-based observations and model predictions: Preliminary results, *Journal of Geophysical Research: Space Physics*, *112*(A7), doi: 10.1029/2006ja012240.

Li, G., G. Qian, I. C. Delgadillo, M. Muller, A. Thabet, and B. Ghanem (2020), Sgas: Sequential greedy architecture search, paper presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Li, W., D. Zhao, C. He, Y. Shen, A. Hu, and K. Zhang (2021), Application of a Multi-Layer Artificial Neural Network in a 3-D Global Electron Density Model Using the Long-Term Observations of COSMIC, Fengyun-3C, and Digisonde, *Space Weather*, *19*(3), e2020SW002605, doi: 10.1029/2020sw002605.
Liu, G., F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro (2018), Image Inpainting for Irregular Holes Using Partial Convolutions, paper presented at Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018//.
Liu, L., W. Wan, B. Ning, O. Pirog, and V. Kurkin (2006), Solar activity variations of the ionospheric peak electron density, *Journal of Geophysical Research: Space Physics*, *111*(A8).
Luo, R., F. Tian, T. Qin, E. Chen, and T.-Y. Liu (2018), Neural Architecture Optimization, arXiv:1808.07233, doi: 10.48550/arXiv.1808.07233.
Lyons, L. R., Y. Nishimura, S.-R. Zhang, A. J. Coster, A. Bhatt, E. Kendall, and Y. Deng (2019), Identification of Auroral Zone Activity Driving Large-Scale Traveling Ionospheric Disturbances, *Journal of Geophysical Research: Space Physics*, *124*(1), 700-714, doi: https://doi.org/10.1029/2018JA025980.
Makela, J. J., S. A. González, B. MacPherson, X. Pi, M. C. Kelley, and P. J. Sultan (2000), Intercomparisons of total electron content measurements using the Arecibo Incoherent Scatter Radar and GPS, *Geophysical Research Letters*, *27*(18), 2841-2844, doi: https://doi.org/10.1029/2000GL000023.
Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister, and T. F. Runge (1998), A global mapping technique for GPS-derived ionospheric total electron content measurements, *Radio Science*, *33*(3), 565-582, doi: https://doi.org/10.1029/97RS02707.
Marconi, G. (1922), Radio telegraphy, *Journal of the American Institute of Electrical Engineers*, *41*(8), 561-570, doi: 10.1109/JoAIEE.1922.6591020.
McGranaghan, R. M., A. J. Mannucci, B. Wilson, C. A. Mattmann, and R. Chadwick (2018), New Capabilities for Prediction of High-Latitude Ionospheric Scintillation: A Novel Approach With Machine Learning, *Space Weather*, *16*(11), 1817-1846, doi: https://doi.org/10.1029/2018SW002018.
McGranaghan, R. M., A. Bhatt, T. Matsuo, A. J. Mannucci, J. L. Semeter, and S. Datta-Barua (2017), Ushering in a New Frontier in Geospace Through Data Science, *Journal of Geophysical Research: Space Physics*, *122*(12), 12,586-512,590, doi: https://doi.org/10.1002/2017JA024835.
Mitra, A. P., and R. E. Jones (1954), Recombination in the lower ionosphere, *Journal of Geophysical Research (1896-1977)*, *59*(3), 391-406, doi: https://doi.org/10.1029/JZ059i003p00391.
Miyato, T., T. Kataoka, M. Koyama, and Y. Yoshida (2018), Spectral Normalization for Generative Adversarial Networks, arXiv:1802.05957, doi: 10.48550/arXiv.1802.05957.
Narayan, S. (1997), The generalized sigmoid activation function: Competitive supervised learning, *Information Sciences*, *99*(1), 69-82, doi: https://doi.org/10.1016/S0020-0255(96)00200-9.
Nishimura, Y., S. R. Zhang, L. R. Lyons, Y. Deng, A. J. Coster, J. I. Moen, L. B. Clausen, W. A. Bristow, and N. Nishitani (2020), Source Region and Propagation of Dayside Large-Scale

Traveling Ionospheric Disturbances, *Geophysical Research Letters*, *47*(19), e2020GL089451, doi: https://doi.org/10.1029/2020GL089451.
Pan, Y., M. Jin, S. Zhang, and Y. Deng (2020), TEC Map Completion Using DCGAN and Poisson Blending, *Space Weather*, *18*(5), e2019SW002390, doi: https://doi.org/10.1029/2019SW002390.
Pan, Y., M. Jin, S. Zhang, and Y. Deng (2021), TEC Map Completion Through a Deep Learning Model: SNP-GAN, *Space Weather*, *19*(11), e2021SW002810, doi: https://doi.org/10.1029/2021SW002810.
Pfaff, R. F., Jr. (2016), An Overview of the Scientific and Space Weather Motivation for the "Notional" Geospace Dynamics Constellation Mission, in *AGU Fall Meeting Abstracts*, edited, pp. SA23C-01.
Qian, L., A. G. Burns, B. A. Emery, B. Foster, G. Lu, A. Maute, A. D. Richmond, R. G. Roble, S. C. Solomon, and W. Wang (2014), The NCAR TIE-GCM, in *Modeling the Ionosphere–Thermosphere System*, edited, pp. 73-83.
Radford, A., L. Metz, and S. Chintala (2015), Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, arXiv:1511.06434, doi: 10.48550/arXiv.1511.06434.
Ramesh, A., M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever (2021), Zero-Shot Text-to-Image Generation, *ArXiv*, *abs/2102.12092*, doi: https://doi.org/10.48550/arXiv.2102.12092.
Rawer, K. (2013), *Wave propagation in the ionosphere*, Springer Science & Business Media.
Ray, P. P. (2023), ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, *Internet of Things and Cyber-Physical Systems*, *3*, 121-154, doi: https://doi.org/10.1016/j.iotcps.2023.04.003.
Real, E., S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin (2017), Large-scale evolution of image classifiers, paper presented at International Conference on Machine Learning, PMLR.
Rich, F. J., P. J. Sultan, and W. J. Burke (2003), The 27-day variations of plasma densities and temperatures in the topside ionosphere, *Journal of Geophysical Research: Space Physics*, *108*(A7).
Richards, P., D. Torr, B. Reinisch, R. Gamache, and P. Wilkinson (1994), F 2 peak electron density at Millstone Hill and Hobart: Comparison of theory and measurement at solar maximum, *Journal of Geophysical Research: Space Physics*, *99*(A8), 15005-15016.
Richmond, A. D., E. C. Ridley, and R. G. Roble (1992), A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophysical Research Letters*, *19*(6), 601-604, doi: https://doi.org/10.1029/92GL00401.
Ridley, A. J., Y. Deng, and G. Tóth (2006), The global ionosphere–thermosphere model, *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(8), 839-864, doi: https://doi.org/10.1016/j.jastp.2006.01.008.
Ruffini, G., A. Flores, and A. Rius (1998), GPS tomography of the ionospheric electron content with a correlation functional, *IEEE Transactions on Geoscience and Remote Sensing*, *36*(1), 143-153, doi: 10.1109/36.655324.
Sai Gowtam, V., and S. Tulasi Ram (2017), An Artificial Neural Network-Based Ionospheric Model to Predict NmF2 and hmF2 Using Long-Term Data Set of FORMOSAT-3/COSMIC Radio

Occultation Observations: Preliminary Results, *Journal of Geophysical Research: Space Physics*, *122*(11), 11,743-711,755, doi: https://doi.org/10.1002/2017JA024795.
Salimans, T., H. Zhang, A. Radford, and D. Metaxas (2018), Improving GANs Using Optimal Transport, arXiv:1803.05573, doi: 10.48550/arXiv.1803.05573.
Schlager, H., V. Grewe, and A. Roiger (2012), Chemical Composition of the Atmosphere, in *Atmospheric Physics: Background – Methods – Trends*, edited by U. Schumann, pp. 17-35, Springer Berlin Heidelberg, Berlin, Heidelberg.
Schunk, R. W., and A. F. Nagy (1978), Electron temperatures in the F region of the ionosphere: Theory and observations, *Reviews of Geophysics*, *16*(3), 355-399, doi: https://doi.org/10.1029/RG016i003p00355.
Shepard, D. (1968), A two-dimensional interpolation function for irregularly-spaced data, in *Proceedings of the 1968 23rd ACM national conference*, edited, pp. 517–524, Association for Computing Machinery.
Snoek, J., H. Larochelle, and R. P. Adams (2012), Practical bayesian optimization of machine learning algorithms, *Advances in neural information processing systems*, *25*.
Spar, J. (1970), Troposphere and Stratosphere: Climate of the Free Atmosphere. D. F. Rex, Ed. Elsevier, New York, 1969. x + 450 pp., illus. $52.50. World Survey of Climatology, vol. 4, *Science*, *168*(3934), 959-960, doi: doi:10.1126/science.168.3934.959.b.
Sparks, L., J. Blanch, and N. Pandya (2010), Kriging as a means of improving WAAS availability, paper presented at Proceedings of the 23rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2010).
Stallkamp, J., M. Schlipsing, J. Salmen, and C. Igel (2012), Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, *Neural Networks*, *32*, 323-332, doi: https://doi.org/10.1016/j.neunet.2012.02.016.
Suganuma, M., S. Shirakawa, and T. Nagao (2017), A genetic programming approach to designing convolutional neural network architectures, paper presented at Proceedings of the genetic and evolutionary computation conference.
Sun, H., Z. Hua, J. Ren, S. Zou, Y. Sun, and Y. Chen (2022), Matrix completion methods for the total electron content video reconstruction, *The Annals of Applied Statistics*, *16*(3), 1333-1358, 1326.
Telea, A. (2004), An image inpainting technique based on the fast marching method, *Journal of graphics tools*, *9*(1), 23-34, doi: https://doi.org/10.1080/10867651.2004.10487596.
Theimer, O., and L. S. Taylor (1961), On the index of refraction in the ionosphere, *Journal of Geophysical Research (1896-1977)*, *66*(10), 3157-3162, doi: https://doi.org/10.1029/JZ066i010p03157.
Tsugawa, T., Y. Otsuka, A. J. Coster, and A. Saito (2007), Medium-scale traveling ionospheric disturbances detected with dense and wide TEC maps over North America, *Geophysical Research Letters*, *34*(22), doi: https://doi.org/10.1029/2007GL031663.
Tulasi Ram, S., V. Sai Gowtam, A. Mitra, and B. Reinisch (2018), The improved two-dimensional artificial neural network-based ionospheric model (ANNIM), *Journal of Geophysical Research: Space Physics*, *123*(7), 5807-5820, doi: https://doi.org/10.1029/2018JA025559.
Vierinen, J., A. J. Coster, W. C. Rideout, P. J. Erickson, and J. Norberg (2016), Statistical framework for estimating GNSS bias, *Atmos. Meas. Tech.*, *9*(3), 1303-1312, doi: 10.5194/amt-9-1303-2016.

Vishnukumar, H. J., B. Butting, C. Müller, and E. Sax (2017), Machine learning and deep neural network — Artificial intelligence core for lab and real-world test and validation for ADAS and autonomous vehicles: AI for efficient and quality test and validation, paper presented at 2017 Intelligent Systems Conference (IntelliSys), 7-8 Sept. 2017.
Wang, Y., and Y. J. Morton (2022), Ionospheric Total Electron Content and Disturbance Observations From Space-Borne Coherent GNSS-R Measurements, *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-13, doi: 10.1109/TGRS.2021.3093328.
Wannberg, G., et al. (1997), The EISCAT Svalbard radar: A case study in modern incoherent scatter radar system design, *Radio Science*, *32*(6), 2283-2307, doi: 10.1029/97RS01803.
Wing, S., J. R. Johnson, E. Camporeale, and G. D. Reeves (2016), Information theoretical approach to discovering solar wind drivers of the outer radiation belt, *Journal of Geophysical Research-Space Physics*, *121*(10), 9378-9399, doi: 10.1002/2016ja022711.
Wing, S., J. R. Johnson, D. L. Turner, A. Y. Ukhorskiy, and A. J. Boyd (2022a), Untangling the solar wind and magnetospheric drivers of the radiation belt electrons, *Journal of Geophysical Research: Space Physics*, *127*(4), e2021JA030246.
Wing, S., D. L. Turner, A. Y. Ukhorskiy, J. R. Johnson, T. Sotirelis, R. Nikoukar, and G. Romeo (2022b), Modeling radiation belt electrons with information theory informed neural networks, *Space Weather*, *20*(8), e2022SW003090.
Wing, S., J. Johnson, J. Jen, C. I. Meng, D. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin, and K. Takahashi (2005), Kp forecast models, *Journal of Geophysical Research: Space Physics*, *110*(A4).
Xie, C., S. Liu, C. Li, M. M. Cheng, W. Zuo, X. Liu, S. Wen, and E. Ding (2019), Image Inpainting With Learnable Bidirectional Attention Maps, paper presented at 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 27 Oct.-2 Nov. 2019.
Yang, D., and H. X. Fang (2023), A Low-Latitude Three-Dimensional Ionospheric Electron Density Model Based on Radio Occultation Data Using Artificial Neural Networks With Prior Knowledge, *Space Weather-the International Journal of Research and Applications*, *21*(1), e2022SW003299, doi: ARTN e2022SW003299
10.1029/2022SW003299.
Yang, Q., D. Tao, D. Han, and J. Liang (2019), Extracting Auroral Key Local Structures From All-Sky Auroral Images by Artificial Intelligence Technique, *Journal of Geophysical Research: Space Physics*, *124*(5), 3512-3521, doi: https://doi.org/10.1029/2018JA026119.
Yeh, R. A., C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do (2017), Semantic image inpainting with deep generative models, paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition.
Yu, J., Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang (2019), Free-Form Image Inpainting With Gated Convolution, paper presented at 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 27 Oct.-2 Nov. 2019.
Yue, X. A., L. H. Hu, Y. Wei, W. X. Wan, and B. Q. Ning (2018), Ionospheric Trend Over Wuhan During 1947-2017: Comparison Between Simulation and Observation, *Journal of Geophysical Research-Space Physics*, *123*(2), 1396-1409, doi: 10.1002/2017ja024675.
Yunxiang Liu, and Y. J. Morton (2022), Improved Automatic Detection of GPS Satellite Oscillator Anomaly using a Machine Learning Algorithm, *NAVIGATION: Journal of the Institute of Navigation*, *69*(1), navi.500, doi: 10.33012/navi.500.

Zhang, S.-R., P. J. Erickson, L. P. Goncharenko, A. J. Coster, W. Rideout, and J. Vierinen (2017), Ionospheric Bow Waves and Perturbations Induced by the 21 August 2017 Solar Eclipse, *Geophysical Research Letters*, *44*(24), 12,067-012,073, doi: https://doi.org/10.1002/2017GL076054.
Zhang, S.-R., P. J. Erickson, A. J. Coster, W. Rideout, J. Vierinen, O. Jonah, and L. P. Goncharenko (2019), Subauroral and Polar Traveling Ionospheric Disturbances During the 7–9 September 2017 Storms, *Space Weather*, *17*(12), 1748-1764, doi: https://doi.org/10.1029/2019SW002325.
Zhang, S.-R., et al. (2022), Traveling Ionospheric Disturbances in the Vicinity of Storm-Enhanced Density at Midlatitudes, *Journal of Geophysical Research: Space Physics*, *127*(8), e2022JA030429, doi: https://doi.org/10.1029/2022JA030429.
Zhang, S. R., and J. M. Holt (2007), Ionospheric climatology and variability from long-term and multiple incoherent scatter radar observations: Climatology in eastern American sector, *Journal of Geophysical Research: Space Physics*, *112*(A6), doi: https://doi.org/10.1029/2006JA012206.
Zhang, S. R., J. M. Holt, A. P. Van Eyken, M. McCready, C. Amory-Mazaudier, S. Fukao, and M. Sulzer (2005), Ionospheric local model and climatology from long-term databases of multiple incoherent scatter radars, *Geophysical Research Letters*, *32*(20), doi: https://doi.org/10.1029/2005GL023603.
Zoph, B., and Q. V. Le (2016), Neural architecture search with reinforcement learning, *arXiv preprint arXiv:1611.01578*.