

# Automatic recognition of tweek atmospherics and plasma diagnostics in the lower ionosphere with the machine learning method

Mao Zhang¹, GaoPeng Lu¹,²,³\*, HaiLiang Huang¹, ZhengWei Cheng⁴, YaZhou Chen⁵, Steven A. Cummer⁶, JiaYi Zheng¹, and JiuHou Lei¹

<sup>1</sup>School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China;

<sup>2</sup>Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;

<sup>3</sup>Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044. China:

<sup>4</sup>State Key Laboratory of Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;

<sup>5</sup>National Key Laboratory on Electromagnetic Environment Effects, Army Engineering University, Shijiazhuang Campus, Shijiazhuang 050003, China;

<sup>6</sup>Electrical and Computer Engineering Department, Duke University, Durham, NC 27708, USA

### **Key Points:**

- An automatic method was developed to recognize tweek atmospherics and diagnose the lower ionosphere with the machine learning method.
- The differences in the ionospheric reflection height (h) and equivalent electron densities at reflection heights ( $N_e$ ) between the automatic and manual methods were  $-0.07 \pm 2.73$  km and  $0.03 \pm 0.92$  cm<sup>-3</sup>, respectively.
- The automatic method was capable of recognizing higher harmonic tweek sferics as well.

**Citation:** Zhang, M., Lu, G. P., Huang, H. L., Cheng, Z. W., Chen, Y. Z., Cummer, S. A., Zheng, J. Y., and Lei, J. H. (2023). Automatic recognition of tweek atmospherics and plasma diagnostics in the lower ionosphere with the machine learning method. *Earth Planet. Phys.*, 7(3), 407–413. http://doi.org/10.26464/epp2023039

**Abstract:** Tweek atmospherics are extremely low frequency and very low frequency pulse signals with frequency dispersion characteristics that originate from lightning discharges and that propagate in the Earth–ionosphere waveguide over long distances. In this study, we developed an automatic method to recognize tweek atmospherics and diagnose the lower ionosphere based on the machine learning method. The differences (automatic – manual) in each ionosphere parameter between the automatic method and the manual method were  $-0.07 \pm 2.73$  km,  $0.03 \pm 0.92$  cm<sup>-3</sup>, and  $91 \pm 1,068$  km for the ionospheric reflection height (h), equivalent electron densities at reflection heights ( $N_e$ ), and propagation distance (d), respectively. Moreover, the automatic method is capable of recognizing higher harmonic tweek sferics. The evaluation results of the model suggest that the automatic method is a powerful tool for investigating the long-term variations in the lower ionosphere.

Keywords: machine learning method; tweek atmospherics; reflection height; D-region ionosphere

# 1. Introduction

Lightning strokes spawned by thunderstorms radiate electromagnetic waves from a few hertz to tens of megahertz. The vast majority of this energy is in the extremely low frequency (ELF, 3–3,000 Hz) and very low frequency (VLF, 3–30 kHz) bands. These radio emissions are reflected at the bottom of the lower ionosphere and propagate in the Earth–ionosphere waveguide with a low

attenuation rate (2–3 dB/1,000 km), which allows them to travel over considerable distances. This waveguide mode propagation causes an appreciable dispersion near the cutoff frequency of the Earth–ionosphere waveguide around 1.8 kHz. Such dispersed sferics are referred to as "tweeks" (Yamashita, 1978) because of the distinct chirping sound they make on a loudspeaker or earphone. The properties of tweeks have been investigated and explained through the theoretical considerations of many authors (Outsu, 1960; Yano et al., 1989; Ryabov, 1992; Yedemsky et al., 1992).

Tweeks have been utilized to estimate the ionospheric reflection height (h), equivalent electron densities ( $N_e$ ) at reflection heights, and propagation distance (d) (Kumar et al., 1994, 2008, 2009; Ohya et al., 2003, 2012). Compared with the active measurement of

First author: M. Zhang, zhangmao@mail.ustc.edu.cn Correspondence to: G. P. Lu, gaopenglu@gmail.com Received 19 DEC 2022; Accepted 07 APR 2023. Accepted article online 24 APR 2023. ©2023 by Earth and Planetary Physics. radar (Igarashi et al., 2000) or rocket sounding (Maeda, 1971), the passive measurement of VLF/ELF tweeks is relatively convenient and economical. Moreover, because of the large data set of VLF/ ELF, the inversion technique is a powerful tool for deriving the long-term variation at the bottom of the lower ionosphere. In contrast, the explosively enlarging data set of VLF/ELF tweeks makes it nearly impossible to process tweek records manually. To analyze a 30-year tweek data set, Ohya et al. (2008) developed an automatic procedure to estimate the reflection height of tweeks, based on a traditional computer vision algorithm by curve fitting, setting the intensity threshold, and calculating the image gradient. However, their method involved careful adjustment of the parameters, which were closely related to the signal source and the observational system. Thus, a more robust automatic method that can fit for different observational systems is required to deal with the massive tweek atmospherics.

In this research, we propose a new method to automatically recognize tweek atmospherics and investigate the lower ionosphere characteristics, based on the machine learning method. After training on a specific data set, the method can fit for different observational systems. In Section 2, the theoretical background of tweek atmospherics and the VLF data set are explained. In Section 3, the automatic recognition and investigation method is described. The performance of the model is evaluated in Section 4, and the results are summarized in Section 5.

# 2. Theoretical Background and Data

### 2.1 Theoretical Background

The electromagnetic field in the Earth–ionosphere waveguide can be decomposed into a sequence of independent field structures (modes) that propagate with different group velocities. The reflection height h and the nighttime electron density  $N_{\rm e}$  can be derived by analyzing the dispersive features of tweek atmospherics. The reflection height of different harmonics is given as follows (Yamashita, 1978; Ohya et al., 2003):

$$h = \frac{nc}{2f_{\rm cn}},\tag{1}$$

where c is the velocity of light in the free space, and  $f_{\rm cn}$  is the cutoff frequency for each mode (n) obtained from the spectrograms. Theoretically, the ionospheric electron density  $N_{\rm e}$  at the reflection height h can be derived from the Appleton–Hartree formula. Specifically, the electron density  $N_{\rm e}$  can be calculated by the following equation (Ohya et al., 2003):

$$N_{\rm e} = 1.241 \times 10^{-2} f_{\rm c} f_{\rm H} \left({\rm el/m}^3\right),$$
 (2)

where  $f_c$  is the mean cutoff frequency  $(f_{cn}/n)$  and  $f_H$  is the electron gyrofrequency.

The electron gyrofrequency  $f_H$  is estimated as follows:

$$f_{\rm H} = \frac{e\mu_0 H}{2\pi m},\tag{3}$$

where  $e = 1.602 \times 10^{-19} \text{C}$  is the elementary charge,  $m = 9.109 \times 10^{-31} \text{kg}$  is the mass of an electron,  $\mu_0 = 4\pi \times 10^{-7} \text{H/m}$  is permeability in the vacuum, and H is the local geomagnetic field strength, which is calculated with the International Geomagnetic

Reference Field 13 (IGRF-13) model by inputting the VLF station location and the observation time. For the first-order mode, the frequency dispersion of tweek atmospherics is given as (Ohya et al., 2003)

$$f(t-t_0) = \frac{f_c(t-t_0)}{\sqrt{(t-t_0)^2 - (d/c)^2}},$$
(4)

where  $t_0$  is the occurrence time of a lightning discharge, and d is the horizontal propagation distance. Once the time–frequency curve of the tweek signal is extracted from the spectrograms, the parameters  $f_c$ ,  $t_0$ , and d can be derived by curve fitting with the least squares method. However, because of the background noise and various shapes of the tweek signals, it is challenging to design a method that can cope with various types of dynamic spectra and automatically extract the time–frequency curve of the tweek signal based on a traditional computer vision algorithm. Thus, we used the machine learning method with a convolutional neural network to automatically extract the time–frequency curves from the VLF spectrograms.

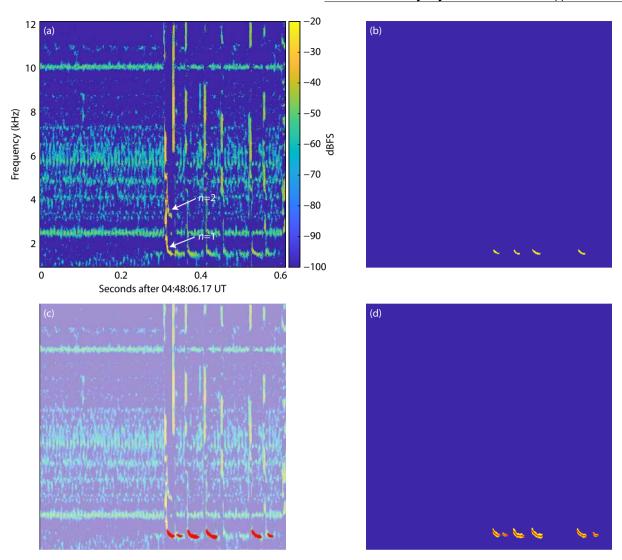
### 2.2 Data

The sferic data were recorded by VLF receivers located at (35.98°N, 79.10°W) with a sampling rate of 100 kHz (Lu GP et al., 2013). Although both azimuthal  $(B_{\sigma})$  and radial  $(B_{r})$  components of the horizontal magnetic fields can be calculated from the two measured orthogonal signals, we used only  $B_{\phi}$  to avoid a low signal-to-noise ratio (Cummins et al., 1998; Cheng ZG et al., 2006). To make our method fit the VLF data set, we selected 500 VLF segments randomly from the data set in 2004-2012. Moreover, these 500 segments were randomly divided into two subsets, a training set and a validation set, at a ratio of 4:1, that is, 400 segments for the training set and 100 segments for the validation set. The validation set did not participate in the training process and was used to tune the hyperparameters and evaluate the model performance. Every segment lasted for 0.6 s and was converted into the dynamic spectrum by the short-time Fourier transform with a time window of 0.01 s. Thus, each data segment in this work was a two-dimensional array  $I(t_i, f_i)$  of a dynamic spectrum where indices  $t_i$  and  $f_i$  corresponded to time and frequency, respectively. The slowly varying background noise of the VLF signals could be further reduced by subtracting the average intensity at each frequency channel with the following expressions (Zhang PJ et al., 2018):

$$I'\left(t_{i},f_{j}\right) = \left|I\left(t_{i},f_{j}\right) - \frac{\sum_{n=a}^{b}I\left(t_{n},f_{j}\right)}{b-a}\right|,\tag{5}$$

where  $I(t_i, f_j)$  and  $I'(t_i, f_j)$  denote the signal intensity before and after the noise reduction, respectively, and a and b are the starting and ending time indexes of the data segment.

Figure 1a shows an example of the time–frequency diagram of nighttime VLF signals received by the station. This segment was divided into the validation set and thus did not participate in the training process. The diagram of obvious tweek atmospherics from 0.3 to 0.6 s shows the frequency dispersion at the first-order cutoff frequency near 1.7 kHz. The second-order harmonic tweek sferics can also be seen at approximately 0.31 s (indicated by the



**Figure 1**. (a) A typical spectrogram of tweek atmospherics observed at (35.98°N, 79.10°W) on January 13, 2012. (b) The grayscale map (reproduced to be a pseudo-color map) of the spectrogram labeled manually with ElSeg software. The yellow region denotes the tweek dispersion curve, and the blue region denotes the background. (c) Automatic recognition of the tweek dispersion curve by the model. The background is shaded, and the recognized tweek signals are shown in red. (d) The grayscale map of the automatic recognition result. The yellow region denotes the tweek dispersion curve, and the blue region denotes the background. The red solid lines denote the extracted tweek dispersion curves.

white arrow in Figure 1a). However, because of the larger attenuation rate of the higher harmonics, the cutoff frequency of the second-order tweek sferics could not be seen clearly from the diagram.

### 3. Method

To calculate the lower ionosphere parameters with Equation (4), the time–frequency curve of tweek signals needed to be extracted from the spectrogram. This curve could be derived through two steps: (1) automatically recognizing and drawing the contour line of tweek signals, and (2) extracting the time–frequency line based on the contour plot of tweek signals.

### 3.1 Automatic Recognition

Step 1 is an image segmentation problem for small targets. To meet the accuracy requirements, the deep convolutional network

was used in our method.

# 3.1.1 Model and training

In this work, the object-contextual representation (OCR) approach (Yuan YH et al., 2019) was applied to automatically label the outline of tweek signals. The High-Resolution Representation net (HRNet) was chosen to be the backbone network (Sun K et al., 2019). The combined HRNet + OCR approach (abbreviated as ocrnet) achieved the first place on the Cityscapes leaderboard in 2020 and thus was expected to meet the accuracy requirements of Step 1. The PaddleSeg toolkit, which features high efficiency and a complete flow, was used to apply the model (Liu Y et al., 2021). PaddleSeg has a number of built-in models, including the HRNet and OCR approaches. Thus, it was convenient to apply or modify our method for tweek recognition.

As a supervised learning algorithm, the ocrnet approach needs

labels attached to the original images. Thus, all 500 spectrograms in our data set were labeled manually, including the training set and the validation set. To accelerate the labeling process, the interactive segmentation annotation software ElSeg (Hao YY et al., 2021) was used. Once the appropriate pretrained model (provided in the PaddleSeg toolkit) was imported, ElSeg would help label part of the image in response to our manual operations. Users do not need to manually label all the contour lines of the tweek signals with ElSeg, which can greatly accelerate the labeling process.

Figure 1b shows a grayscale map of the spectrogram labeled with ElSeg software. The yellow region and the blue region denote the tweek dispersion curve and the background, respectively. As shown in the figure, the label does not contain all the tweek curves because the manual criteria can be flexible and complicated. Loose criteria can increase the recognition rate of the tweek signals. However, not all the tweek signals have a high signal-to-noise ratio. Thus, loose criteria may result in more background noise and further reduce the accuracy of the inversion parameters of the ionosphere. In contrast, strict criteria can guarantee that the selected tweek signals have a high signal-to-noise ratio. However, strict criteria will lead to the automatic model having a lower recognition rate. In fact, the selection criterion is a trade-off for every researcher, depending on his or her requirements for the automatic model.

After the labeling process, both the original and labeled images were fed to the ocrnet for training. The training process was performed on a single Nvidia RTX A6000 graphics processing unit. The pretrained ocrnet model (provided in the PaddleSeg toolkit) was used to accelerate the training process.

# 3.1.2 Loss function

The tweek signal areas were obviously much smaller than the background (Figure 1b), which caused a class-imbalance problem. A simple way to deal with the class-imbalance problem was to adjust the loss function of the neural network. Most segmentation methods based on the deep network rely on logistic regression, which optimizes the cross-entropy loss (Goodfellow et al., 2016):

$$Loss_{CE} = -\sum_{classes} y_{true} log(y_{pred}),$$
 (6)

where "classes" includes the background and the objects (in this work, the class contains the background and the tweek signal area), and  $y_{\rm pred}$  is the network probability estimate of the ground truth probability of each pixel. In cross-entropy loss, each pixel is treated equally; thus, the result will be greatly influenced by classes with more pixels (i.e., the background in the spectrogram). In this work, the Lovas hinge loss Loss<sub>LH</sub> (Berman et al., 2018) was used to handle the class-imbalance problem. The total loss of the neural network was a mixture of the aforementioned two types of loss:

$$Loss_{total} = \alpha \cdot Loss_{CE} + \beta \cdot Loss_{LH}, \tag{7}$$

where coefficients  $\alpha$  and  $\beta$  were chosen as 0.9 and 0.1, respectively.

# 3.2 Information Extraction

 $After the \ training \ process, the \ model \ could \ automatically \ recognize$ 

and draw the contour of tweek signals, as shown in Figures 1c, d (the red areas in Figure 1c and the yellow areas in Figure 1d). Note that the model not only labeled the four tweek dispersion curves labeled manually, as shown in Figure 1b, but also recognized two less obvious tweek dispersion curves that were ignored in manual labeling, suggesting the model has high recall. (A detailed evaluation of the model is presented in Section 4.)

Thus far, we have obtained the contour plot of tweek dispersion curve contours. The next step was to extract the time–frequency line of tweek signals. The only difference between the manual and automatic methods was the labeled area in the last step (Figures 1b, d). The subsequent steps, including the time–frequency line extraction and the least-squares curve fitting, were the same in the manual and automatic methods. The labeled tweek dispersion curve contour was denoted as  $(t_l, f_m)_{D_n}$ , which could be derived from the grayscale images (see Figures 1b, d) by finding the connected domain. The subscript  $D_n$  denoted the connected domain of the nth tweek signal contour. The extracted time–frequency line  $(\bar{t}_n - \bar{f}_n)$  could be calculated as follows:

$$\bar{t}_n = \{t_l\}_{D_{n'}} \tag{8}$$

$$\bar{f}_n = \left\{ \exp\left(\frac{\sum_m I(t_l, f_m) \times Inf_m}{\sum_m I(t_l, f_m)}\right) \right\}_{D_n}, \tag{9}$$

where  $I(t_l,f_m)$  is the signal intensity at the connected domain of the nth tweek signal contour. With Equations (8) and (9), the time–frequency lines of the automatically recognized six tweek signals could be derived and are shown as the red lines in Figure 1d. By substituting the time–frequency lines into Equation (4), the reflection height h, the nighttime electron density  $N_{\rm ev}$  and the horizontal propagation distance d could be derived with the least squares method. The results of the six tweek signals (shown in Figure 1d from left to right) are presented in Table 1. For comparison, the ionosphere parameters derived from the manually operated labels are presented in the right columns of Table 1. Moreover, the relative error, also presented in Table 1, was calculated as follows:

error = 
$$\sqrt{\left(\frac{h^a - h^b}{h^b}\right)^2 + \left(\frac{N_e^a - N_e^b}{N_e^b}\right)^2 + \left(\frac{d^a - d^b}{d^b}\right)^2}$$
, (10)

where the parameters with superscripts a and b denote the results from the automatic model and the manually operated labels, respectively. Note that the relative error between the automatic model and the manually operated model was small (< 2%) except for the first tweek signal (13.6%), which was mainly caused by the error in the propagation distance.

# 4. Evaluation

### 4.1 Precision and Recall

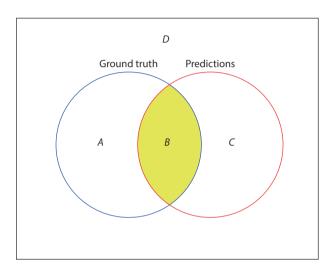
In image segmentation, precision and recall are performance metrics that describe the purity and completeness of the automatic predictions relative to the manually operated labels. Figure 2 shows a schematic of the definition of precision and recall in image segmentation. Precision and recall were derived as follows:

$$precision = \frac{B}{B \cup C'}$$
 (11)

**Table 1.** Nighttime ionospheric reflection height h, electron density  $N_{er}$  and propagation distance d calculated from the tweek signals in Figure 1.

Tweeka	<i>h</i> (km)	$N_{\rm e}$ (cm <sup>-3</sup> )	d (km)	h <sup>b</sup>	$N_{\rm e}^{ m b}$	$d^{b}$	Error <sup>c</sup>
1	87.8	28.7	2,074	88.4	28.5	2,400	0.136
2	90.3	27.9	2,310	-	-	-	-
3	90.4	27.8	2,310	89.4	28.2	2,310	0.018
4	89.4	28.2	2,311	89.0	28.3	2,340	0.014
5	89.0	28.3	2,000	89.2	28.2	2,000	0.004
6	90.4	27.8	2,320	_	_	_	_

<sup>&</sup>lt;sup>a</sup>Number sequence of the tweek signals in Figure 1d. <sup>b</sup>Calculated from the manually operated labels in Figure 1b. A dash denotes missing values. <sup>c</sup>Calculated from Equation (10).



**Figure 2.** Schematic of the definition of precision and recall. The blue-edged circle denotes the ground truth area (i.e., the manually operated labels). The red circle denotes the predictions of the model (i.e., the automatically labeled areas). The whole evaluation area is divided into four subareas (*A*, *B*, *C*, *D*).

$$recall = \frac{B}{A \cup B'}$$
 (12)

where B denotes the intersection of the automatically labeled areas and the manually labeled areas,  $B \cup C$  denotes the automatically labeled areas (i.e., the yellow areas in Figure 1d), and  $A \cup B$  denotes the manually labeled areas (i.e., the yellow areas in Figure 1b).

Recall describes the completeness of the model, and precision describes the purity of the model. However, because not all the tweek signals were labeled in manual operations, the precision calculated with Equation (11) was bound to be low. Thus, we modified the precision by restricting the calculation to the manually labeled tweek signals. For example, when calculating the precision of the segment in Figure 1, the original definition of precision required the calculation of all six tweek contour areas. The modified precision involved the calculation of only the four tweek contour areas that were also labeled in manual operations. The recall and the modified precision were calculated as 0.916 and 0.346, respectively. This high recall suggests that the model was able to recognize almost all the manually labeled tweek signals and that the automatically labeled areas tended to cover

the manually labeled areas. The modified precision seemed to be low. However, it should be noted that the low precision did not always lead to high error when calculating ionosphere parameters because the manually labeled areas constituted the core regions of the automatically labeled areas, and the signal intensity was generally low in the extended areas. Thus, the time–frequency lines of tweek signals calculated from the automatically labeled areas with Equation (9) were less influenced by the extended areas. For instance, the precision of tweek signals shown in Figure 1 was 0.309, whereas the relative error was 0.4%–13.6%.

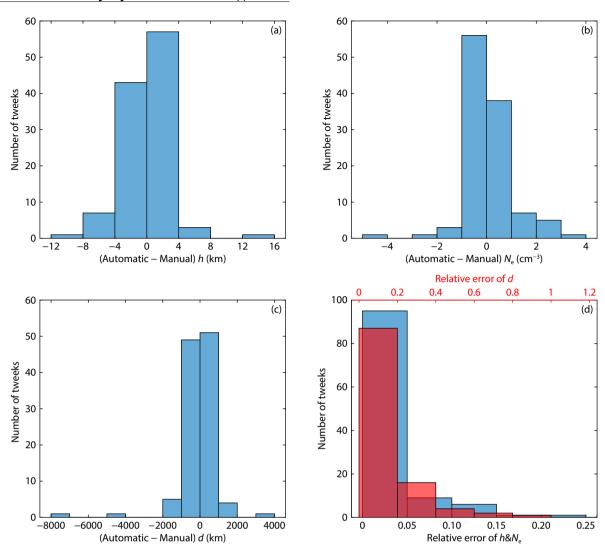
### 4.2 Comparison with Manual Operations

To further evaluate the performance of the automatic method, we compared the ionosphere parameters derived by the automatically labeled tweeks and the manually labeled tweeks, using the 100 spectrograms in the validation set. The automatic method and the manual method labeled 122 first-order tweeks concurrently. For 10 of these 122 tweeks, the time–frequency lines could not be appropriately fitted to Equation (4). For the remaining 112 tweeks, the manual method gave reasonable values for h,  $N_{\rm e}$ , and d of 87.00  $\pm$  4.88 km, 29.21  $\pm$  1.61 cm<sup>-3</sup>, and 2,692  $\pm$  1,559 km (mean value  $\pm$  standard deviation), respectively. For comparison, the automatic method gave values for h,  $N_{\rm e}$ , and d of 86.93  $\pm$  4.98 km, 29.24  $\pm$  1.70 cm<sup>-3</sup>, and 2,600  $\pm$  1,270 km, respectively.

Figures 3a-c show results for the comparison of h,  $N_{\rm e}$ , and d, respectively, between the automatic and manual methods ( $\delta$  = automatic - manual). Of the total 112 tweeks, 89.3% (100), 84.0% (94), and 89.3% (100) fell in the range of  $-4 \text{ km} < \delta h < 4 \text{ km}$ ,  $-1 \text{ cm}^{-3} < \delta N_e < 1 \text{ cm}^{-3}$  and  $-1,000 \text{ km} < \delta d < 1,000 \text{ km}$ , respectively. Figure 3d shows the relative error calculated with Equation (10). Because the error was mainly caused by the propagation distance, we calculated the relative error of the reflection height plus the electron density (error<sub>h&N</sub>, shown as the blue bars) and propagation distance (error<sub>d</sub>, shown as the red bars) separately. Of the 112 tweeks, 84.8% (95) and 77.7% (87) fell in the range of error<sub>h&N</sub> < 5% and error<sub>d</sub> < 20%, respectively. Moreover, the mean values and the standard deviations of the differences between the automatic method and the manual method for  $\delta h$ ,  $\delta N_e$ , and  $\delta d$ were  $-0.07 \pm 2.73$  km,  $0.03 \pm 0.92$  cm<sup>-3</sup>, and  $91 \pm 1,068$  km, respectively.

### 4.3 Automatic Recognition of Higher Harmonics

In this work, the automatic model was also able to recognize the higher harmonic tweek sferics, even though the training set was



**Figure 3**. Comparison results between the automatic method and the manual method. (a) Differences  $\delta h$  (= automatic – manual) of the reflection height h. Of the 112 tweeks, 89.3% (100) fall in the range of –4 km <  $\delta h$  < 4 km. (b) Differences  $\delta N_e$  (= automatic – manual) of the electron density  $N_e$ . Of the 112 tweeks, 84.0% (94) fall in the range of –1 cm<sup>-3</sup> <  $\delta N_e$  < 1 cm<sup>-3</sup>. (c) Differences  $\delta d$  (= automatic – manual) of the propagation distance d. Of the 112 tweeks, 89.3% (100) fall in the range of –1,000 km <  $\delta d$  < 1,000 km. (d) Relative error of the reflection height and the electron density (error<sub> $\hbar \delta N_e$ </sub>, blue bars, calculated from the first two terms on the right side of Equation (10)) and the propagation distance (error<sub>d</sub>, red bars, calculated from the last term on the right side of Equation (10)). Of the 112 tweeks, 84.8% (95) and 77.7% (87) fall in the range of error<sub> $h \delta N_e$ </sub> < 5% and error<sub>d</sub> < 20%, respectively.

selected arbitrarily and contained only a few higher harmonic samples (30 second-order harmonics, 4 third-order harmonics, and 2 fourth-order harmonics in 400 images). Figure 4 shows two examples of the automatically recognized higher harmonic tweek sferics, which suggests further application of the automatic model with higher harmonic tweek sferics.

# 5. Conclusions

In this work, we propose a new method to automatically recognize tweek atmospherics and investigate the lower ionosphere parameters (the reflection height h, electron density  $N_{\rm e}$ , and propagation distance d) based on the machine learning method. The machine learning model achieved a recall of 0.916 on the validation set. The differences between the automatic method and the manual method for each ionosphere parameter were  $-0.07 \pm 2.73$  km,  $0.03 \pm 0.92$  cm<sup>-3</sup>, and  $91 \pm 1,068$  km for h,  $N_{\rm e}$ , and d, respectively.

Moreover, the automatic model was capable of recognizing higher harmonic tweek sferics. The evaluation results of the model suggest that the automatic method is a powerful tool for investigating long-term variations in the lower ionosphere.

# Acknowledgments

This work was supported by the Chinese Academy of Sciences (CAS) Project of Stable Support for Youth Team in Basic Research Field (Grant No. YSRR-018), the National Key R&D Program of China (Grant No. 2019YFC1510103), the National Natural Science Foundation of China (Grant Nos. 41875006 and U1938115), the Chinese Meridian Project, and the International Partnership Program of CAS (Grant No. 183311KYSB20200003). The PaddleSeg and ElSeg toolkits are available at https://github.com/PaddlePaddle/PaddleSeg. The International Geomagnetic Reference Field (IGRF-13) is available at https://github.com/space-physics/igrf.

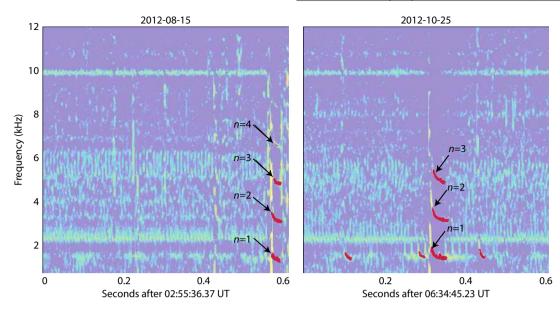


Figure 4. Two spectrograms of higher harmonic tweek sferics in the validation set, automatically recognized by the model.

### References

Berman, M., Triki, A. R., and Blaschko, M. B. (2018). The Lovasz–Softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4413–4421). Salt Lake City, UT, USA: IEEE. https://doi.org/10.1109/CVPR.2018.00464

Cheng, Z. G., Cummer, S. A., Baker, D. N., and Kanekal, S. G. (2006). Nighttime D region electron density profiles and variabilities inferred from broadband measurements using VLF radio emissions from lightning. J. Geophys. Res.: Space Phys., 111(A5), A05302. https://doi.org/10.1029/2005JA011308

Cummins, K. L., Krider, E. P., and Malone, M. D. (1998). The US National Lightning Detection Network™ and applications of cloud-to-ground lightning data by electric power utilities. *IEEE Trans. Electromagn. Compat.*, 40(4), 465–480. https://doi.org/10.1109/15.736207

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press

Hao, Y. Y., Liu, Y., Wu, Z. W., Han, L., Chen, Y. Z., Chen, G. W., Chu, L. T., Tang, S. Y., Yu, Z. Y., . . . Lai, B. H. (2021). EdgeFlow: achieving practical interactive segmentation with edge-guided flow. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 1551–1560). Montreal, BC, Canada: IEEE. https://doi.org/10.1109/ICCVW54120.2021.00180

Igarashi, K., Murayama, Y., Nagayama, M., and Kawana, S. (2000). D-region electron density measurements by MF radar in the middle and high latitudes. *Adv. Space Res.*, 25(1), 25–32. https://doi.org/10.1016/S0273-1177(99)00893-5

Kumar, S., Dixit, S. K., and Gwal, A. K. (1994). Propagation of tweek atmospherics in the earth–ionosphere wave guide. *IL Nuovo Cimento C*, 17(3), 275–280. https://doi.org/10.1007/8F02509168

Kumar, S., Kishore, A., and Ramachandran, V. (2008). Higher harmonic tweek sferics observed at low latitude: estimation of VLF reflection heights and tweek propagation distance. *Ann. Geophys.*, 26(6), 1451–1459. https://doi. org/10.5194/angeo-26-1451-2008

Kumar, S., Deo, A., and Ramachandran, V. (2009). Nighttime D-region equivalent electron density determined from tweek sferics observed in the South Pacific Region. Earth, Planets Space, 61(7), 905–911. https://doi.org/10.1186/ BF03353201

Liu, Y., Chu, L. T., Chen, G. W., Wu, Z. W., Chen, Z. Y., Lai, B. H., and Hao, Y. Y. (2021). PaddleSeg: a high-efficient development toolkit for image segmentation. Preprint arXiv:2101.06175. https://doi.org/10.48550/arXiv.2101.06175

Lu, G. P., Cummer, S. A., Li, J. B., Zigoneanu, L., Lyons, W. A., Stanley, M. A., Rison, W., Krehbiel, P. R., Edens, H. E., ... Samaras, T. (2013). Coordinated observations of sprites and in-cloud lightning flash structure. *J. Geophys. Res.: Atmos.*, 118(12), 6607–6632. https://doi.org/10.1002/jgrd.50459

Maeda, K. I. (1971). Study on electron density profile in the lower ionosphere. *J. Geomagn. Geoelectr.*, 23(2), 133–159. https://doi.org/10.5636/jgg.23.133

Ohya, H., Nishino, M., Murayama, Y., and Igarashi, K. (2003). Equivalent electron densities at reflection heights of tweek atmospherics in the low-middle latitude D-region ionosphere. *Earth, Planets Space*, *55*(10), 627–635. https://doi.org/10.1186/BF03352469

Ohya, H., Shiokawa, K., and Miyoshi, Y. (2008). Development of an automatic procedure to estimate the reflection height of tweek atmospherics. *Earth, Planets Space, 60*(8), 837–843. https://doi.org/10.1186/BF03352835

Ohya, H., Tsuchiya, F., Nakata, H., Shiokawa, K., Miyoshi, Y., Yamashita, K., and Takahashi, Y. (2012). Reflection height of daytime tweek atmospherics during the solar eclipse of 22 July 2009. *J. Geophys. Res.: Space Phys.*, 117(A11), A11310. https://doi.org/10.1029/2012JA018151

Outsu, J. (1960). Numerical study of tweeks based on waveguide mode theory. *Proc. Res. Inst. Atmos.*, 7, 58–71.

Ryabov, B. S. (1992). Tweek propagation peculiarities in the earth-ionosphere waveguide and low ionosphere parameters. *Adv. Space Res., 12*(6), 255–258. https://doi.org/10.1016/0273-1177(92)90067-8

Sun, K., Xiao, B., Liu, D., and Wang, J. D. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5686–5696). Long Beach, CA, USA: IEEE. https://doi.org/10.1109/ CVPR.2019.00584

Yamashita, M. (1978). Propagation of tweek atmospherics. *J. Atmos. Terr. Phys.*, 40(2), 151–153, 155–156. https://doi.org/10.1016/0021-9169(78)90019-3

Yano, S., Ogawa, T., and Hagino, H. (1989). Waveform analysis of tweek atmospherics. J. Atmos. Electr., 9(1), 31–42. https://doi.org/10.1541/jae.9.31

Yedemsky, D. Y., Ryabov, B. S., Shchokotov, A. Y., and Yarotsky, V. S. (1992). Experimental investigation of the tweek field structure. *Adv. Space Res.*, 12(6), 251–254. https://doi.org/10.1016/0273-1177(92)90066-7

Yuan, Y. H., Chen, X. K., Chen, X. L., and Wang, J. D. (2019). Segmentation transformer: object-contextual representations for semantic segmentation. Preprint arXiv:1909.11065. https://doi.org/10.48550/arXiv.1909.11065

Zhang, P. J., Wang, C. B., and Ye, L. (2018). A type III radio burst automatic analysis system and statistic results for a half solar cycle with Nançay Decameter Array data. Astron. Astrophys., 618, A165. https://doi.org/10.1051 /0004-6361/201833260