



## RESEARCH ARTICLE

10.1029/2021SW002981

### Key Points:

- Monitoring the electron density along different  $L$ -shells by measuring the dispersion relation via whistlers
- Convolutional neural network based method to accurately segment the whistler traces from spectrogram images
- Seasonal variation of plasmaspheric equatorial electron density based on more than 1 million analyzed whistler traces over 9 years

### Correspondence to:

J. Lichtenberger,  
lityi@sas.elte.hu

### Citation:

Pataki, B. Á., Lichtenberger, J., Clilverd, M., Máthé, G., Steinbach, P., Pásztor, S., et al. (2022). Monitoring space weather: Using automated, accurate neural network based whistler segmentation for whistler inversion. *Space Weather*, 20, e2021SW002981. <https://doi.org/10.1029/2021SW002981>

Received 10 NOV 2021

Accepted 4 FEB 2022

# Monitoring Space Weather: Using Automated, Accurate Neural Network Based Whistler Segmentation for Whistler Inversion

Bálint Ármin Pataki<sup>1</sup> , János Lichtenberger<sup>2,3</sup> , Mark Clilverd<sup>4</sup> , Gergely Máthé<sup>1</sup> , Péter Steinbach<sup>5</sup> , Szilárd Pásztor<sup>2</sup>, Lilla Murár-Juhász<sup>2</sup>, Dávid Koronczay<sup>2</sup>, Orsolya Ferencz<sup>2</sup>, and István Csabai<sup>1</sup> 

<sup>1</sup>Department of Physics of Complex Systems, ELTE Eötvös Loránd University, Budapest, Hungary, <sup>2</sup>Space Research Group, Department of Geophysics and Space Sciences, ELTE Eötvös Loránd University, Budapest, Hungary, <sup>3</sup>Institute of Earth Physics and Space Science, Sopron, Hungary, <sup>4</sup>British Antarctic Survey (UKRI-NERC), Cambridge, UK, <sup>5</sup>ELKH-ELTE Research Group for Geology, Geophysics and Space Sciences, Budapest, Hungary

**Abstract** It is challenging, yet important, to measure the—ever-changing—cold electron density in the plasmasphere. The cold electron density inside and outside of the plasmopause is a key parameter for radiation belt dynamics. One indirect measurement is through finding the velocity dispersion relation exhibited by lightning induced whistlers. The main difficulty of the method comes from low signal-to-noise ratios for most of the ground-based whistler components. To provide accurate electron density and  $L$ -shell measurements, whistler components need to be detectable in the noisy background, and their characteristics need to be reliably determined. For this reason, precise segmentation is needed on a spectrogram image. Here, we present a fully automated way to perform such an image segmentation by leveraging the power of convolutional neural networks, a state-of-the-art method for computer vision tasks. Testing the proposed method against a manually, and semi-manually segmented whistler data set achieved <10% relative electron density prediction error for 80% of the segmented whistler traces, while for the  $L$  shell, the relative error is <5% for 90% of the cases. By segmenting more than 1 million additional real whistler traces from Rothera station Antarctica, logged over 9 years, seasonal changes in the average electron density were found. The variations match previously published findings, and confirm the capabilities of the image segmentation technique.

**Plain Language Summary** When lightning strikes on the Earth, electromagnetic waves are generated, that can travel along the magnetic field lines of the Earth and can be observed in the other hemisphere. As the waves are in the 1–30 kHz range, they can be measured with radio antennas and exhibit a whistling sound, thus their name, whistlers. As the electromagnetic waves travel, they are distorted due to the velocity dispersion in the ionosphere and the magnetosphere. As the rate of dispersion depends on the physical parameters of the above-mentioned regions, the precise measurement of the whistlers can be a method to monitor the electron density on different magnetic field lines. The cold electron density inside and outside of the plasmasphere is a key parameter for radiation belt dynamics, that can affect satellites negatively, their precise measurement can be fruitful in many applications. Whistlers have been measured and collected for decades, however, due to the noisy nature of the detection data, the precise, automated estimation of the physical parameters through whistler traces has been a challenging task. We trained a convolutional neural network based model to accurately mark the whistlers on a frequency-time spectrogram. The trained model can find and segment the whistlers, which allows accurate physical parameter estimation, within 10% error in electron density for 80% of the time.

## 1. Introduction

Whistlers have been regarded as cheap and effective tools for plasmasphere diagnostics since the early years of whistler research when the nose whistler was discovered (Helliwell et al., 1956) and their propagation theory was developed (Smith et al., 1960). The plasmopause was discovered through whistler analysis (D. L. Carpenter, 1963). Whistlers can also affect energetic particle populations in the radiation belts (Abel & Thorne, 1998) and trigger other whistler mode waves (D. L. Carpenter, 1978; Hosseini et al., 2019).

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Whistler inversion procedures are based on the whistler travel time integral, which is derived from the Appleton-Hartree dispersion relation for parallel propagation (Helliwell, 1965; Lichtenberger, 2009):

$$T(f) = \frac{1}{c} \int_{path} \mu ds, \quad \mu^2 = 1 + \frac{f_p^2}{f(f_H - f)}, \quad (1)$$

where  $f_p$  is the plasma frequency and  $f_H$  is the gyrofrequency. The integral is taken along the propagation path (field line). The magnetic field model used is centered dipole model, the field-aligned electron density distribution model used is the (Ozhogin et al., 2012). The inversion procedure is based on comparing the travel times of the model whistler with the travel times of the measured whistler. The travel time of the measured whistler is obtained from scaled  $f$ - $t$  pairs (a series of frequency and time coordinates of a whistler trace measured on a conventional spectrogram). Several algorithms have been developed for such a purpose (e.g., Bernard, 1973; Lichtenberger, 2009; Park, 1972; Tarcsai, 1975). However, all algorithms require the  $f$ - $t$  pairs as input, and because annotating the  $f$ - $t$  point of whistler traces is a very labor-intensive task, they have never been used as operational tools. The only exception is the Virtual Trace Transformation method used in the Automatic Whistler Detector and Analyzer (AWDA) systems (Lichtenberger et al., 2008, 2010), where a multiple path whistler propagation model is used to estimate the equatorial electron densities directly from sferic-filtered spectrograms.

Recently (Harid et al., 2021) developed a mask-scoring regional convolutional neural network (MSRCNN) for the automated extraction of whistlers, but the extracted  $f$ - $t$  pairs were not used to obtain the equatorial electron densities through whistler inversion. In this study, we present a method which is similarly based on machine learning, leveraging the PointRend (Kirillov et al., 2020) architecture. The new method extends previous analysis by identifying the equatorial cold electron density for each component of a whistler, and the  $L$ -shell on which it was propagating. For training the neural network a set of model whistlers were generated with realistic frequency-dependent amplitudes consistent with real whistler traces. Measured VLF background noise was added to make the simulations even more realistic. The precise knowledge of the model whistler parameters made it possible to analyze the stability of the inversion procedure (Section 2.3). The developed method was then applied to whistlers recorded at Rothera, Antarctica, between 2008 and 2016 (Section 3). Large-scale statistics based on more than 700,000 inverted whistler traces are presented in Section 4.3.

## 2. Materials and Methods

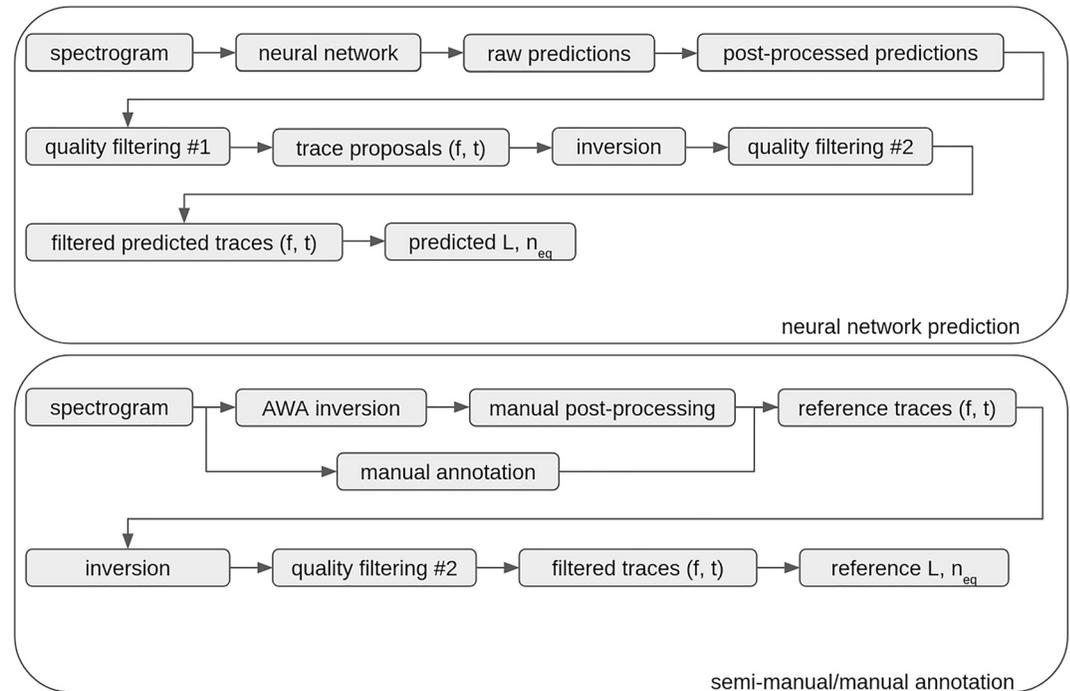
### 2.1. Whistler Data

During the almost two decades of operation, receivers of the Automatic Whistler Detector and Analyzer Network (Lichtenberger et al., 2008) have collected millions of whistler events. Among them—as the whistler activity is the highest here—we selected Rothera station whistler data (British Antarctic Survey, UK, Lat.  $-67.57$ deg, Long.  $-68.12$ deg) to initially develop and test the neural network based whistler inversion method. Based on the success of this work, data from the other stations will be processed in the near future. The Automatic Whistler Detector system started to operate at Rothera in May 2008. We used a subset of the data recorded from the period 2008–2016. In the manuscript whistler events and whistler traces are used, as a whistler event is a single, continuous measurement (usually lasting 1–2 s), which usually contains several whistler traces.

### 2.2. Modeling

Object detection and segmentation is a common task within the computer vision branch of machine learning. In object detection, each individual object is marked with a bounding box on the image, allowing the ability to count and locate them. Segmentation techniques can be separated into two forms. In semantic segmentation, each pixel is assigned to a class or background, but when objects belonging to the same class overlap, the individual objects are not resolved. In instance segmentation each pixel is assigned to a specific object, thus overlapping objects from the same category are separated while producing a theoretically pixel-accurate segmentation.

Instance segmentation is convenient to use in whistler analysis because in this manner individual whistler traces can be detected separately, regardless how close they are to each other. However, the nature of the whistler traces—long and thin objects, in close proximity—makes the task difficult for the most common instance segmentation neural network architectures that have been developed for everyday images where the contrast

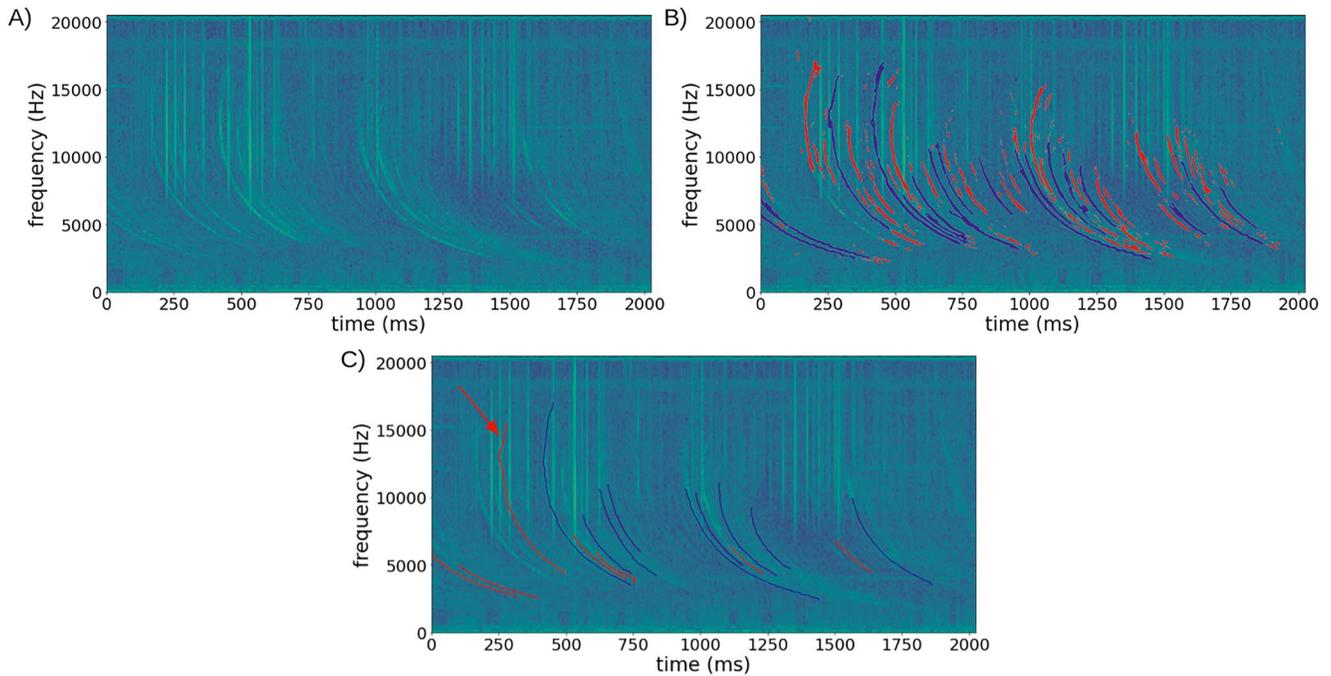


**Figure 1.** Flowchart of the testing and the prediction generation process. Top: the neural network generates the raw predictions, prediction masks, on the spectrogram. Next, the prediction masks are postprocessed, all predictions are projected on top of each other. Then, the too small blobs are removed, and from the rest, the  $(f, t)$  trace proposals are generated, which are to be inverted. Based on the inversion results, the low-quality traces are discarded, the final result are the predicted traces with the predicted physical parameters. Bottom: the spectrograms are fed to the AWA algorithm, and when the AWA could not invert the trace, the trace annotation is done via manual postprocessing. For traces, that the AWA could not invert, a fully manual annotation was necessary. The obtained reference traces were next inverted and filtered, producing the filtered traces and the reference physical parameters. The quality filtering #1 refers to the blob filtering: time-frequency bandwidth and area covered thresholds; while the quality filtering #2 refers to the inversion based filters: residual threshold and the highest marked frequency compared to the nose frequency. For the test set, the trace proposals were matched to the reference traces, and finally the predicted and the reference physical parameters were compared for the matched traces.

between objects and background is typically higher. For example, the Mask R-CNN (He et al., 2017) model, which is one of the most popular architectures, uses  $14 \times 14$  or  $28 \times 28$  pixel resolution when predicting the segmentation mask, which is insufficient to resolve individual whistler traces. While for objects in the common data sets, such as MS COCO (Lin et al., 2014) or Pascal VOC (Everingham et al., 2015), and in everyday objects, this level of mask accuracy is sufficient, for whistlers pixel-level accuracy is desired. Only a few architectures focus on pixel-level instance segmentation. The Mask Scoring R-CNN (Huang et al., 2019) architecture tries to improve the prediction mask accuracy by adding a MaskIOU head, which predicts the intersection over union for the generated mask, and its score can be used to select the most accurate masks. The PointRend (Kirillov et al., 2020) architecture tackles the problem as an image rendering task. The initial low-resolution prediction mask is refined in regions of the image where high-frequency features are present, resulting in the refinement of the mask boundaries. While making high-resolution prediction masks on a whole regular pixel grid would require excessive computing power and memory, with the PointRend model high-resolution predictions are achievable.

A PointRend model with a ResNet50 backbone was trained using the MMDetection framework for 30 epochs on the  $1,024 \times 1,024$  px simulated spectrograms. Each pixel covered 1 ms in time and 19.5312 Hz in frequency, thus a single image spanned 1.024 s and a frequency range of 0–20,000 Hz. The MMDetection configuration file, which contains all the details used for the training process is available in the provided code repository.

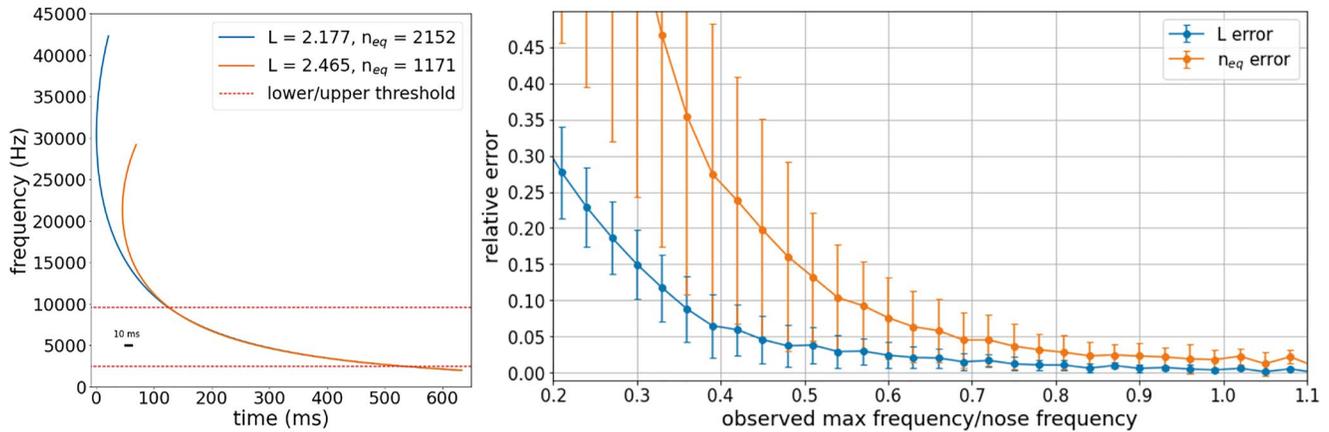
Then, arbitrarily long spectrograms were split into  $1,024 \times 1,024$  px (containing 1,024 ms) patches using a 500 ms sliding window. Predictions for the patches were made separately, which were then concatenated to regain the original arbitrarily long spectrogram. The prediction process is shown in Figure 1. For each pixel that was at the overlap of two patches, the logical AND was calculated of the individual pixel predictions. For



**Figure 2.** The prediction and filtering process. (a) The spectrogram that was patched into  $1,024 \times 1,024$  px ( $1,024$  ms) regions with a sliding window of  $500$  ms. (b) All the raw mask predictions are projected onto each other. Each contiguous region, here described as a blob, is considered as a potential whistler trace. The blobs not passing the quality filters (at least  $200$  px<sup>2</sup> area, at least  $100$  px time, and  $100$  px frequency span) were discarded, shown as red. (c) The blobs which passed the quality filter were converted into  $f$ - $t$  point pairs to calculate the average time for each frequency bin for the blobs. Finally, the  $f$ - $t$  pairs were inverted and filtered based on their frequency span and residuals. The traces shown in red did not pass the quality filter, while the traces shown in blue did. The trace highlighted with the red arrow would pass the nose frequency requirement, but it had a fit residual that was too high.

the pure instance segmentation setup, the predicted mask often “jumped” between two or more traces, making the resulting prediction impractical. To overcome this, individual instance segmentation prediction masks were combined into a single prediction for each image. The combination method was a pixel-wise logical AND operation, meaning, that if a pixel was predicted to be on a whistler in any of the individual instance segmentation mask predictions, that pixel was marked positive. However, when a pixel was predicted as background for all the individual mask predictions, that pixel was marked negative. Thus resulted in a new, combined binary mask prediction (see Figure 2b), where the predicted pixels are highlighted by either red or blue colors. In the next step, the projected predictions which formed a continuous, connected region, that is, a blob, were identified. Blobs that covered either an area smaller than  $200$  px<sup>2</sup> or spanned less than  $1,950$  Hz ( $100$  px) in frequency or  $100$  ms ( $100$  px in time bins were discarded [highlighted with red on Figure 2b]). Next, the frequency-time ( $f$ - $t$ ) pairs were generated from the blobs for each frequency bin value, the mean of the times in the blob for each given frequency was calculated. Finally, the  $f$ - $t$  pairs were inverted and those considered low-quality traces were discarded, shown red in Figure 2c). A trace was marked low-quality if its marked highest frequency bin was lower than the  $60\%$  of its nose frequency (the frequency where the gradient of the whistler trace diverges), the value of which was calculated from the inversion. Traces where the mean absolute residual of the fit exceeded a given threshold ( $0.005$ ) were also discarded. The residual restriction filtered out cases when the predicted  $f$ - $t$  pairs do not fit any realistic whistler shapes well—see the trace highlighted with the red arrow in Figure 2.

The average time needed to generate predictions for a  $2$  s long radio measurement is currently  $9.2$  s on a regular PC (Nvidia GeForce GTX 1070 8GB GPU, Intel Core i7-6850K CPU). The  $9.2$  s is made up of  $0.6$  s spent on spectrogram generation (CPU),  $3.3$  s for generating the neural network predictions (GPU) and  $5.3$  s for the post-processing, such as prediction prefiltering and inversion of an average of  $7.5$  traces (the neural network produces on average  $7.5$  traces that need to be inverted for each  $2$  s long spectrograms). Further speedup is possible by both software and hardware upgrade, although this was not a goal for the present study. However, it is worth noting that the method could process radio measurements continuously without the need for a powerful CPU/GPU cluster.



**Figure 3.** Left: when only the lower part of the whistler trace is visible on the spectrogram, that is, the region between the red dashed lines, whistlers with significantly different physical parameters could be nearly identical. However, when higher frequencies are shown, it is clear that the traces are different, which suggests that the highest frequency on the predicted trace is connected to the reliability of the physical parameter estimation. *Note.* The frequency is shown for 45 kHz for clearer visualization, while all the data used in the study reached only 20 kHz. Right: relative error of the physical parameters for the matched, simulated whistler traces. As the whistler trace is spread out to higher and higher frequencies, in nose frequency units, the error decreases. This suggests that the highest detected frequency compared to the nose frequency is a good measure for the physical parameter fit reliability.

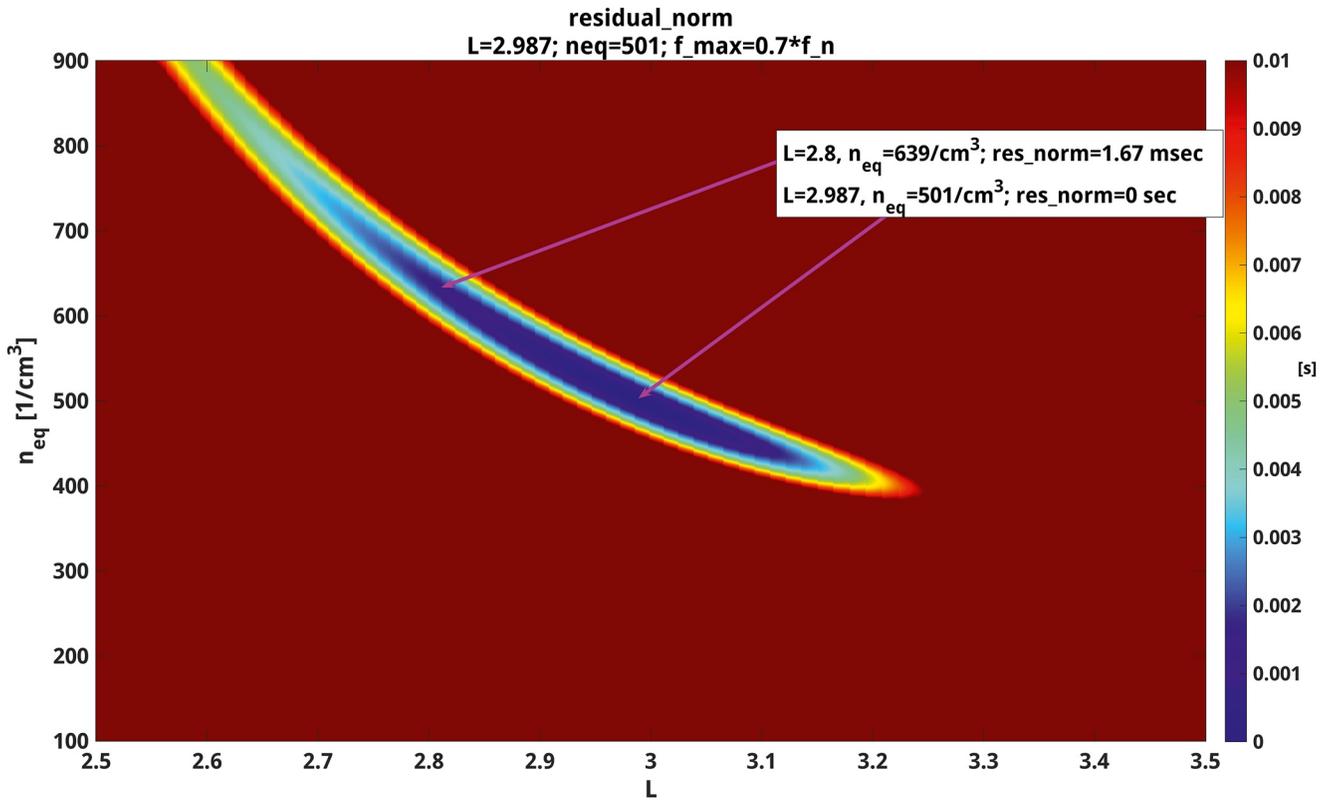
### 2.3. Inversion Stability

During the inversion method (Lichtenberger, 2009), as for any curve fitting method, the robustness of the fitted parameters is highly dependent on the layout of the data points. For example, when fitting a parabola on noisy points, measured far away from the vertex only on the positive side, the vertex of the fitted parabola might fall far away from the actual vertex. However, when the measurement points fall close to the vertex and on both sides of it, the fit becomes much more reliable. For whistlers, the same difficulty applies. The traces can be up to 5–15 px (5–15 ms) wide on the used spectrogram resolution (1 ms), thus the predicted  $f$ - $t$  points could have significant noise. The goodness of a whistler trace fit can be defined as the accuracy of the physical parameters  $L$  and  $n_{eq}$ . Thus, it is crucial to explore which cases result in reliable physical parameter estimates.

The nose frequency of the whistler is defined as the frequency that has the fastest travel speed, thus provides the earliest signal of the whistler. However, it is frequently above the upper frequency limit of the recorded signal. Also, due to noise and the weakness of the signal, the nose frequency cannot be always read from the spectrogram, but can be obtained from the inversion. In our experiments 2,000 whistler traces were generated by Lichtenberger (2009) model covering the physically relevant parameter grid of  $L$  and  $n_{eq}$  for the Rothera station.

The traces were trimmed to the 2,500–17,500 Hz range, which is the lowest/highest frequency band of the whistler traces for vast majority of whistlers at Rothera station. Next, the traces were trimmed further to have an upper limit of  $x$  times their nose frequency for various  $x$  values. Finally, the trimmed traces were paired in a way that if any two traces could be time-shifted to each other such that the difference of the traces at all frequency bins was less than 5 ms (5 px)—which is smaller than the width of the traces on the spectrogram—then the traces were marked as paired. For the paired traces the  $L$  and  $n_{eq}$  relative errors are shown on Figure 3 as a function of the above-mentioned  $x$  value. The result can be interpreted as if the simulated  $f$ - $t$  pairs do not have at least 0.6 times the nose frequency of both traces, the physical parameters of the inversion can have errors larger than 3% for  $L$  and 8% for  $n_{eq}$ . Thus, for the evaluation, results are discarded when the predicted trace did not reach 60% of its nose frequency—calculated from the fit.

As the usual whistler trace has a width of 10 ms on the spectrograms we used in this study, any curve fitting method based on highlighted points on the spectrogram is vulnerable when the highlighted points, the predictions, do not reach high enough frequencies compared to the nose frequency. An example is shown in Figure 3, where two simulated whistlers are closer than 10 ms at all frequency bins between 2,500 and 9,600 Hz, however, their physical parameters differ significantly ( $n_{eq}$  is 1,171 and 2,152 for them). The higher frequency threshold (9,600 Hz) means 31% and 45% of the nose frequency for the two whistlers. It is clear, that if the higher frequency threshold were to be increased, the traces would not be paired anymore. However, for real whistler measurements,



**Figure 4.** Norm of residuals plot computed for a model whistler trace with  $L = 2.987$  and  $n_{eq} = 501/cm^3$  parameters. The  $f$ - $t$  pairs of the model whistler were computed only for  $f < f_{nose}$  frequencies. Inversions (Lichtenberger, 2009) were run using these  $f$ - $t$  pairs on an  $L = 2.5 - 3.5$  and  $n_{eq} = 100 - 900/cm^3$  grid. The color scale represents the norm of residuals. The norm of residuals slowly varies along a hyperbolic curve on the grid, forming a shallow trench. Within the trench, the norm of residuals is  $< 10msec$ . These values are comparable to those obtained from successful real inversions, while the estimated  $L$  and  $n_{eq}$  parameters differ from the accurate one up to 50% in  $L$  and 100% in  $n_{eq}$ .

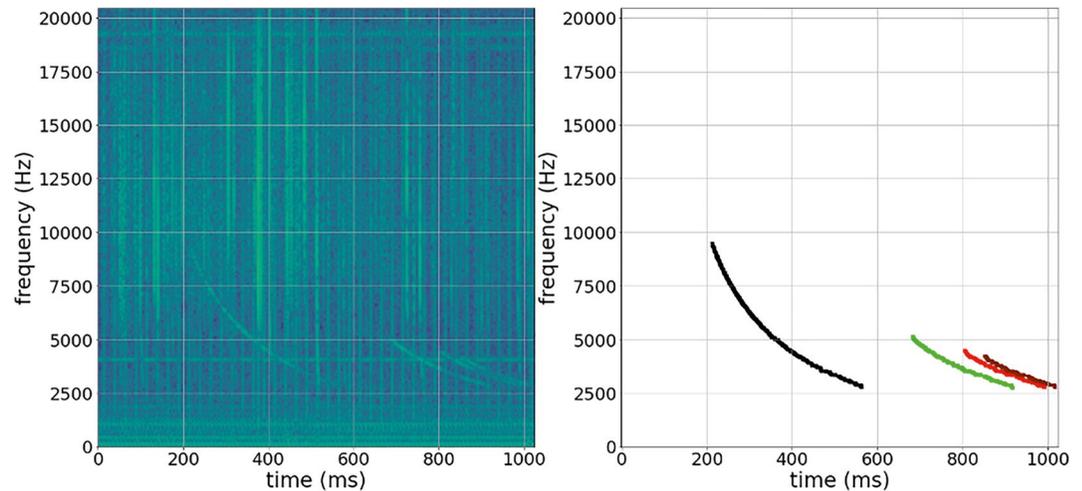
often the trace is non-visible for higher frequencies. The  $\chi^2$  surface of a fit is shown in Figure 4, which shows that the  $n_{eq}$  parameter is much more sensitive than  $L$  in terms of relative error for similar residuals.

### 3. Data

#### 3.1. Training Data Set

A total of 9,069 whistler traces from the Rothera station were selected from the AWA (Lichtenberger et al., 2010) preprocessing outputs. The traces were manually segmented—due to the preprocessing, the manual part of the segmentation consisted of the selection of properly preprocessed traces, marking the minimal and maximal frequency and the width of the traces—resulting in a tight annotation contour around the traces which were then separated from the background, preserving both its shape and amplitude. Then, the traces were converted from spectrogram to a time signal. Next, the obtained time signal traces were placed next to each other, resulting in a whistler event. All the generated whistler events had members (traces) ranging between 1 and 14. At the last step, the generated whistler events were added to background noise, that was measured at the Rothera station. At the last step, the time signals were converted to realistic spectrograms, see Figure 5, generating an annotated data set which can be used to train a neural network.

The training data set contained 9,096, 1,024 ms long  $1,024 \times 1,024$  px spectrograms, each covering a 20,000 Hz frequency range. In total, the 9,096 train images contained 57,072 whistler traces. Each training sample consists of a background noise added spectrogram, which is the input image, see Figure 5, and one binary segmentation mask for each whistler trace within the event. The individual whistler masks were generated by placing them (without the other members within the whistler event) on an all zero background and transforming it to a spectrogram, which was thresholded to obtain the binary segmentation mask.



**Figure 5.** Left: the input spectrogram image example from the training data set. Right: The four binary masks highlighted, one for each whistler traces.

Finally, the input image was transformed to a three color channel image by placing the pixel-wise square root, the raw pixel, and the pixel-wise square of the image in the three color channels. By this operations, three transformed images were provided for the neural network model, which expected images to have three color channels.

### 3.2. Test Data Set

The performance of the trained model was measured on real spectrograms instead of generated/transformed whistlers, as the network might learn confounding factors made during the transformation. For model selection 30 validation spectrograms were used, while for testing the resulting model 70 spectrograms were used. The validation data set was used to select the most accurate model via hyperparameter tuning, while the test data set was used to measure the performance of the model that was selected based on the validation data set. The hyperparameter search consisted of using three different annotation mask width (thin, medium, and thick) for the training data set, during training. And for the inference, the prediction filtering hyperparameters of the model (intersection over union threshold for the non-max suppression, max number of predictions, proposals per image, and the score threshold). Due to the special postprocessing of the prediction masks, a rather permissive hyperparameter setup was found to work the best. In the manuscript, only the results of the test data set are shown.

Annotating a whistler trace is practically reading the  $(f, t)$  coordinates of the trace on a spectrogram. A whistler trace is not a continuous line, rather a set of patches. The annotation procedure, therefore, requires an experienced annotator as the obvious candidate point to represent the  $(f, t)$  coordinates (i.e., the one with the highest energy) often biased by the noise background. The annotation was done by five experienced researchers from the Space Research Group.

The test data set consisted of 70, 2 s long spectrograms with good quality whistlers (whistlers, that are potentially “invertible”), that were selected from AWA (Lichtenberger et al., 2010) preprocessing outputs. The whistler traces on the spectrograms were annotated semi-manually (using the output of Automatic Whistler Analyzer (Lichtenberger et al., 2010)) and manually.

The semi-manual annotation uses the successful AWA (Lichtenberger et al., 2010) inversions. The successfully inverted traces appear as vertical ones on the Virtual Trace Transformation spectrogram. In this case, only the upper and lower frequencies were selected manually, the  $(f, t)$  pairs, as well as the corresponding powers between these two frequencies, were obtained from the spectrogram using the  $L$  and  $n_{eq}$  parameters from AWA. For traces, where there was no successful AWA inversion available, but the trace was high quality enough, the human annotator manually selected the points along the trace by mouse clicks, which points were interpolated later linearly.

The annotation consisted of 338 traces, out of those 76 were discarded as they did not pass the quality requirements defined in the Inversion stability section, leaving 262 traces.

## 4. Results

### 4.1. Evaluation

The main goal of accurately segmenting whistlers is to produce accurate physical parameter estimates. Thus, during the evaluation process, the focus is on the accuracy of the predicted physical parameters as a measure of the proposed model.

The presented method found 523 traces on the 70 test spectrograms, 179 of the traces were discarded due to low quality, while 344 traces passed the quality filter. The predicted traces were then matched with the annotated ones. If a predicted and an annotated trace fell within 10 ms of each other in more than 50 frequency bins those traces were considered matched. Out of the 344 predicted traces, 130 were matched, and 214 were not. This provides us with a subset of 130 traces with which to undertake direct comparison between the results of the neural network model and those of the semi-manually/manually annotated group.

The manual annotation is subjective in the sense, that it is physically not possible to mark every single trace, as often many of them are very weak, see Figure 2a.

After visual inspection of the neural network test set prediction, it was confirmed, that many some of the predictions are related to traces, which were not annotated manually due to having a low signal-to-noise ratio, however, the neural network could still find, and accurately mark them. On the other hand, there were also many traces, for which the neural network was not able to generate a high-quality prediction, which would pass all the quality filters, while the same was done by the semi-manual/manual annotation process.

Based on the results, roughly 40% of the neural network predicted traces (130/344) could be matched to half of the semi-manually/manually selected traces (130/262). The results can be interpreted as the model is capable of finding and producing quality predictions for half of the whistler traces, while the semi-manual/manual annotation process might have had a similar rate.

The final goal is accurate plasmaspheric electron density monitoring through whistler parameter estimation, not the detection and counting of all whistlers, and all traces, thus missing some is not necessarily a problem. In Figure 6, the physical parameters calculated from the inversion of the annotated and predicted traces are shown for the 130 matched traces. For  $L$  the prediction method has less than 5% error in 90% of the traces, while for  $n_{eq}$  has less than 10% and 20% relative error in 80% and 90% of the traces. The results also suggest that the predictions are reliable, the 214 non-matched predictions are most likely not false positives, but actual whistler traces that were missed during the annotation process.

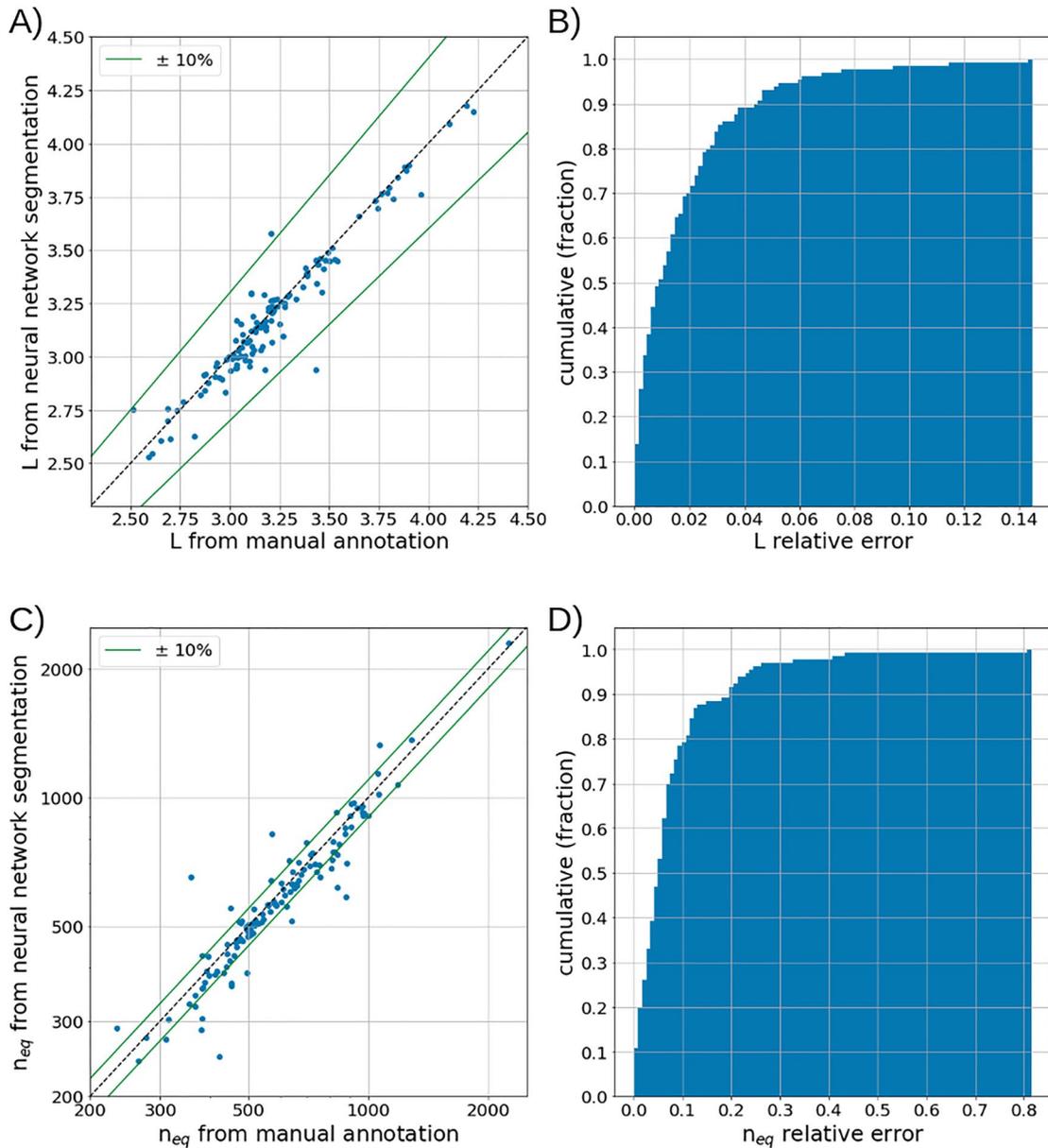
### 4.2. Selection Bias

As only half of the annotated traces were found by the neural network method, the question of selection bias is relevant. In Figure 7, the kernel density estimate is shown for the physical parameters for the annotated traces grouped as matched (130) and non-matched (132). A Kolmogorov-Smirnov test was performed on the cumulative distributions and the resulting  $p$ -value was 0.011 for  $L$  and 0.290 for  $n_{eq}$ . The neural network is more likely to select traces with  $L$  values around 3.2–3.5, while less likely to detect traces with higher  $L$  values. No such distinction was found for  $n_{eq}$ .

The observable frequency bandwidth of a whistler trace is lower as  $L$  is increasing. Thus making a prediction, which passes all the quality requirements, for these traces is substantially more difficult, which is the most likely explanation for the neural network quality-filtered predictions being slightly biased with respect to the  $L$  values.

### 4.3. Large-Scale Trends

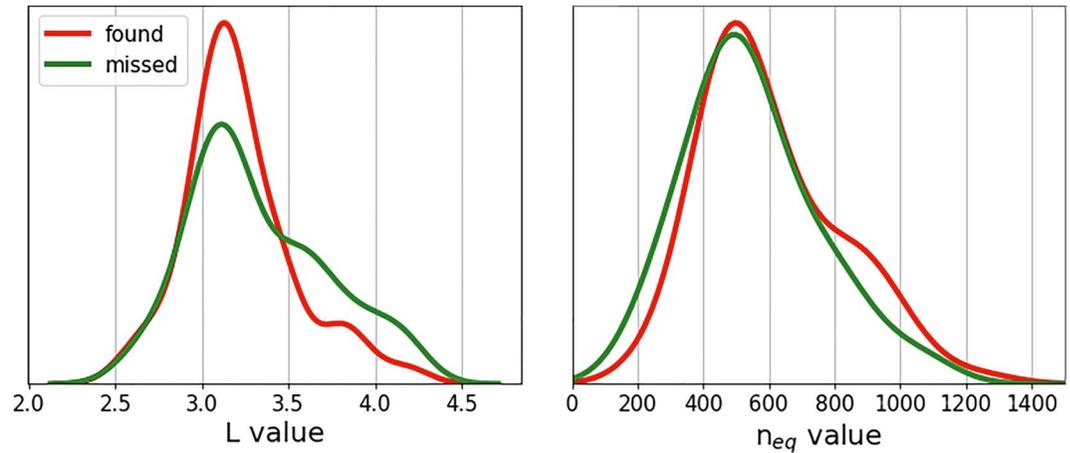
The developed neural network model was applied to 74.5 hr of radio measurement data recorded at Rothera, Antarctica between 2008 and 2016. The 74.5 hr consisted 165,371 spectrograms (whistler events) with an average length of 1.6 s. The spectrograms were selected through criteria published in (Lichtenberger et al., 2010). The



**Figure 6.** The physical parameters estimated from the inversion of the 130 matched whistler traces in the test set. The inversion was performed on the annotation and on the predictions as well. Panels (a and c) show the obtained physical parameters by the neural network model as a function of the physical parameters obtained by semi-manual annotation. Panels (b and d) shows the cumulative histogram of the relative errors.

model produced  $f$ - $t$  pairs for 1,048,267 whistlers, out of which 712,626 passed the quality filters. An example whistler event is shown in Figure 8.

The detected whistler number distribution with respect to the  $L$  shell and the electron density (top chart of Figure 9) clearly shows, that while for smaller  $L$  values (2.5–3.5) a wide range of  $n_{eq}$  were recorded, for higher  $L$  values (3.9–4.6) only smaller electron densities were found. This inverse proportionality was found also by (D. Carpenter & Anderson, 1992). The seasonality of the average electron density value across different  $L$  shells is shown in the bottom chart on Figure 9: in the May–September period, the average electron density is significantly lower than the October–April period, which harmonizes with former findings on a smaller data set (Clilverd et al., 1991) from Antarctica. The agreement of the seasonal trend also supports the reliability of the presented method. As the presentation of large-scale statistics is not the main goal of the present study, contrary to the presentation of the method, deeper analysis of electron density variations is a future research area.



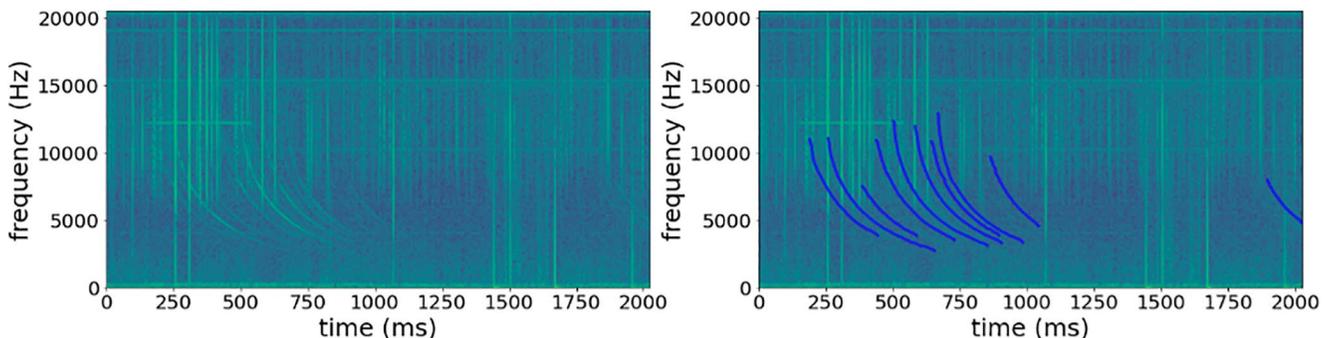
**Figure 7.** Kernel density estimates of the physical parameters for the matched and nonmatched traces on the validation set. A Kolmogorov-Smirnov test was performed on the cumulative distributions, which resulted in a  $p$ -value of 0.011 for  $L$  and 0.290 for  $n_{eq}$ , suggesting that the model randomly miss traces in terms of electron density, but there is a small systematic bias for  $L$  value.

## 5. Conclusions

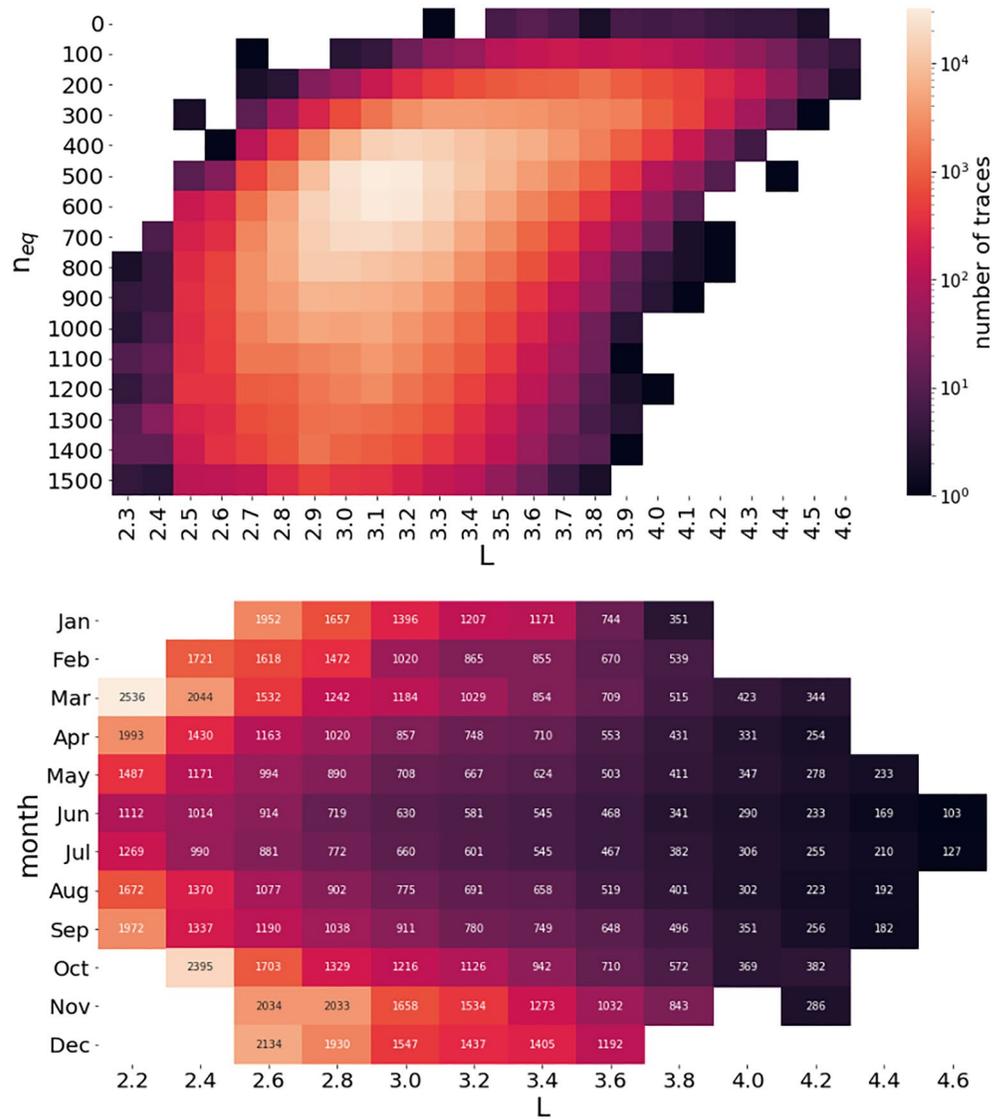
Whistlers have a unique duality. On one hand, the equipment needed for recording them is relatively cheap and widely available. While on the other hand, the large-scale, accurate processing of the recorded traces present serious challenges due to the low signal-to-noise ratio and necessity of a bandwidth of several kHz to obtain useful information. There has been significant effort put into researching whistler processing with various signal and image processing methods. Many of these works extracted important information from whistler traces, while the full, accurate automation remains to be solved.

In the last decade, the traditional image processing methods were displaced in many computer vision tasks by convolutional neural networks. While convolutional neural networks were known in the late 1980's (LeCun et al., 1989), their widespread use did not occur until the early 2010's (Krizhevsky et al., 2012), when large enough training data sets and powerful enough compute engines (GPUs) became widely available. Since then, neural networks have become a popular and state of the art tool for many tasks, such as image classification, image segmentation, speech recognition, machine translation, and many others.

Neural networks have been applied to radio measurements to detect and count whistlers. Konan et al. (2020) used a YOLO object detection neural network and a sliding deep convolutional neural network to detect whistlers. Mochalov and Mochalova (2018) used a ResNet50 neural network architecture to classify spectrograms according to whether a whistler is present in them, while Conti et al. (2015) used a time delay neural network for the same task. However, more information can be extracted than just whistler count distributions. Whistler



**Figure 8.** An example of whistlers, from the unannotated large-scale data set, which were properly detected by the model. Left: the spectrogram, which is the input of the neural network model. Right: the detected whistlers are marked with blue on top of the spectrogram.



**Figure 9.** Whistler statistics obtained from more than 700,000 traces recorded at the Rothera station. Top: The number of found whistlers as a function of  $L$  and  $n_{eq}$ . Bottom: the heatmap shows the average  $n_{eq}$  values for each month split up to 0.2  $L$  bins. The number in each cell is the actual average  $n_{eq}$  value. Where there is no number, it means, there were no whistlers analyzed within that bin (e.g.,  $L = 2.2$  in December).

propagation  $L$  shells and their electron densities are coded in the whistler shape, elucidating physical parameters which cannot be extracted by detection alone. Harid et al. (2021) segmented whistler traces on spectrogram images using an MSRCNN, although the result was used only for counting them, not for actually inverting the whistlers and estimating their physical parameters. The main added value of our convolutional neural network model is the automated inversion of whistler traces, thus providing electron density estimates over a range of  $L$  shells.

The presented study proposes a convolutional neural network based method, that can segment, mark individual frequency-time points along a whistler trace on a spectrogram. Such points can be used to estimate the electron density along a determined  $L$  shell. The current method was developed and tested using data from the Rothera station, Antarctica. As other stations might have different background noise, and different whistler characteristics retraining the model using the given stations' data might be needed. The method also has limitations. First of all, many of the recorded whistlers are not detected, which is a minor issue when there are many whistler observations at a station. The other, more important limitation is, that the method can give only the statistical

errors of the physical parameters for a whistler trace. The individual, single trace tailored error estimation is not available. Developing a method that can estimate the errors of the physical parameters for each trace remains a future research direction.

## Data Availability Statement

The training data with the annotation masks is available at Zenodo with <https://doi.org/10.5281/zenodo.5570921>, the code used for training the neural network is accessible at [https://github.com/patbaa/whistler\\_segmentation](https://github.com/patbaa/whistler_segmentation), while the large-scale test data set is not shared publicly due to its extensive data volume.

## Acknowledgments

This research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Artificial Intelligence National Laboratory Program MILAB. M. Clilverd would like to acknowledge the UKRI-NERC support for the Rothera whistler measurements under the National Capability—Single Science (Space Weather Observatory) grant.

## References

- Abel, B., & Thorne, R. M. (1998). Electron scattering loss in Earth's inner magnetosphere: I. Dominant physical processes. *Journal of Geophysical Research*, 103(A2), 2385–2396. <https://doi.org/10.1029/97ja02919>
- Bernard, L. C. (1973). A new nose extension method for whistlers. *Journal of Atmospheric and Terrestrial Physics*, 35, 871–880. [https://doi.org/10.1016/0021-9169\(73\)90069-x](https://doi.org/10.1016/0021-9169(73)90069-x)
- Carpenter, D., & Anderson, R. (1992). An ISEE/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research*, 97(A2), 1097–1108. <https://doi.org/10.1029/91ja01548>
- Carpenter, D. L. (1963). Whistler evidence of a 'knee' in the magnetospheric ionization density profile. *Journal of Geophysical Research*, 68(6), 1675–1682. <https://doi.org/10.1029/jz068i006p01675>
- Carpenter, D. L. (1978). Whistlers and VLF noises propagating just outside the plasmapause. *Journal of Geophysical Research*, 83(A1), 45–57. <https://doi.org/10.1029/ja083ia01p00045>
- Clilverd, M. A., Smith, A., & Thomson, N. (1991). The annual variation in quiet time plasmaspheric electron density, determined from whistler mode group delays. *Planetary and Space Science*, 39(7), 1059–1067. [https://doi.org/10.1016/0032-0633\(91\)90113-o](https://doi.org/10.1016/0032-0633(91)90113-o)
- Conti, L., Innocenzi, P. M., & Placidi, L. (2015). Time delay neural network for recognition of whistler waves.
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1), 98–136. <https://doi.org/10.1007/s11263-014-0733-5>
- Harid, V., Liu, C., Pang, Y., Alvina, A. J., Golkowski, M., Hosseini, P., & Cohen, M. (2021). Automated large-scale extraction of whistlers using mask-scoring regional convolutional neural network. *Geophysical Research Letters*, 48(15), e2021GL093819. <https://doi.org/10.1029/2021gl093819>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961–2969). <https://doi.org/10.1109/iccv.2017.322>
- Helliwell, R. A. (1965). *Whistlers and related ionospheric phenomena*. Stanford University Press.
- Helliwell, R. A., Crary, J. H., Pope, J. H., & Smith, R. L. (1956). The "nose" whistler – A new high latitude phenomenon. *Journal of Geophysical Research*, 61, 139. <https://doi.org/10.1029/jz061i001p00139>
- Hosseini, P., Golkowski, M., & Harid, V. (2019). Remote sensing of radiation belt energetic electrons using lightning triggered upper band chorus. *Geophysical Research Letters*, 46(1), 37–47. <https://doi.org/10.1029/2018gl081391>
- Huang, Z., Huang, L., Gong, Y., Huang, C., & Wang, X. (2019). Mask scoring R-CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6409–6418). <https://doi.org/10.1109/cvpr.2019.00657>
- Kirillov, A., Wu, Y., He, K., & Girshick, R. (2020). PointRend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9799–9808). <https://doi.org/10.1109/cvpr42600.2020.00982>
- Konan, O. J., Mishra, A. K., & Lotz, S. (2020). Machine learning techniques to detect and characterise whistler radio waves. arXiv preprint arXiv:2002.01244. <https://arxiv.org/abs/2002.01244>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- Lichtenberger, J. (2009). A new whistler inversion method. *Journal of Geophysical Research*, 114. <https://doi.org/10.1029/2008JA013799>
- Lichtenberger, J., Ferencz, C., Bodnár, L., Hamar, D., & Steinbach, P. (2008). Automatic Whistler Detector and Analyzer System: Automatic whistler detector. *Journal of Geophysical Research*, 113. <https://doi.org/10.1029/2008JA013467>
- Lichtenberger, J., Ferencz, C., Hamar, D., Steinbach, P., Rodger, C. J., Clilverd, M. A., & Collier, A. B. (2010). The Automatic Whistler Detector and Analyzer (AWDA) system: Implementation of the analyzer algorithm. *Journal of Geophysical Research*, 115. <https://doi.org/10.1029/2010JA015931>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). Microsoft COCO: Common objects in context. In *European Conference on Computer Vision* (pp. 740–755). [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Mochalov, V., & Mochalova, A. (2018). Algorithms and results of streaming whistler recognition. In *E3S web of conferences* (Vol. 62, p. 02001). <https://doi.org/10.1051/e3sconf/20186202001>
- Ozhogin, P., Tu, J., Song, P., & Reinisch, B. W. (2012). Field-aligned distribution of the plasmaspheric electron density: An empirical model derived from the IMAGE RPI measurements. *Journal of Geophysical Research*, 117(A6). <https://doi.org/10.1029/2011ja017330>
- Park, C. G. (1972). *Methods to determine electron concentrations in the magnetosphere from nose whistlers* (Technical Report No. 3454-1). Radioscience Laboratory, Stanford Electronics Laboratories, Stanford University.
- Smith, R. L., Helliwell, R. A., & Yabroff, I. (1960). A theory of trapping of whistlers in field-aligned columns of enhanced ionization. *Journal of Geophysical Research*, 65, 815. <https://doi.org/10.1029/jz065i003p00815>
- Tarcsai, G. (1975). Routine whistler analysis by means of accurate curve fitting. *Journal of Atmospheric and Terrestrial Physics*, 37, 1447. [https://doi.org/10.1016/0021-9169\(75\)90075-6](https://doi.org/10.1016/0021-9169(75)90075-6)