



# Seasonal Variation of the D-Region Ionosphere Modelled using Machine Learning Based VLF Remote Sensing

David K. Richardson<sup>\*(1)</sup>, and Morris B. Cohen<sup>(1)</sup>

(1) School of Electrical and Computer Engineering, Georgia Institute of Technology

## Abstract

This work improves upon a previously developed neural network modelling process that predicted waveguide parameters for the D-region ionosphere on two days [1]. The previous model was limited by manually determining the ideal set of transmitters (Tx) and receivers (Rx) and by computation time. An automatic quality assessment tool was developed to automatically evaluate the optimal network for each day [2]. We also obtained a 14x improvement in model training time by leveraging GPUs and improving the parallelization of the training process. These advancements allowed us to model 328 days across up to 21 paths. With this larger sample size, we show the model is capable of following expected seasonal trends. The model has also been adapted to be used with nighttime data, and is showing promising early results.

## 1 Introduction

The D-region ionosphere covers an altitude range (60-90 km) which is difficult to measure using conventional means. For lower altitude measurements, it is possible to use high-altitude balloons or aircraft to carry instrumentation; however, these methods do not extend high enough. Above the D-region, it is possible to use satellites to provide direct measurements, but atmospheric drag prevents satellites from safely flying within the D-region. It is possible to use sounding rockets to measure electron density, but they can only provide a single vertical path over a short period of time [3]. Due to the cost of a single flight and the limited coverage, it is prohibitively expensive to use sounding rockets for large scale D-region measurements.

The use of very low frequency (VLF) waves to infer properties of the D-region ionosphere has been explored by multiple previous studies [1], [4], [5]. Due to their frequency, VLF waves experience almost total reflection from the D-region. Combined with the Earth's ground being reflective at VLF frequencies, this forms a spherical waveguide, known as the Earth-Ionosphere Waveguide, which allows VLF waves to propagate large distances around the globe. As the waves propagate, they are affected by the ionosphere's properties at the reflection points.

Narrowband VLF transmitters are one source of VLF radiation which can be used to make predictions about the

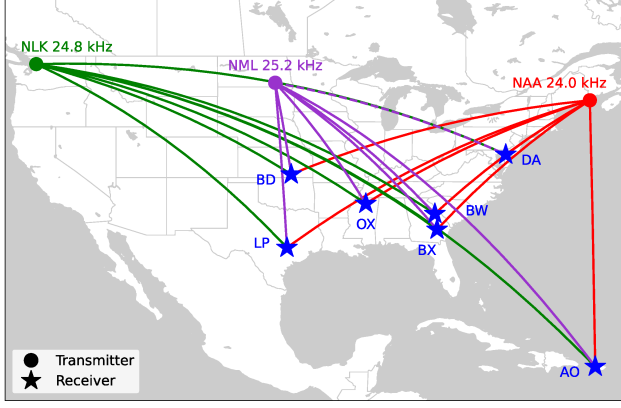
D-region along a transmitter-receiver path [4]. These predictions typically only use a small number of receivers, and perform a search over a variety of electron density profiles to select which profile best fits the measurements. This technique is limited to small numbers of transmitter-receiver paths due to searching through a large number of profiles for each path. Recently, Gross developed a method for modelling arbitrarily large transmitter-receiver networks using machine learning techniques. Their work provides a general framework for predicting a two parameter model for electron density profiles, but was only applied to two days [1].

This work leverages the machine learning modelling technique to predict the electron density profiles for 328 daytime days and 1 nighttime day. By dramatically increasing the number of predictions, we are able to understand how the model responds to seasonal variations in the ionosphere. The model seeding process was improved to account for differences between days not seen in the original paper. Additionally, the technique was extended to provide initial results for nighttime predictions, a traditionally difficult period to estimate.

## 2 Dataset

In this work, we use data recorded by seven VLF/low frequency (LF) AWESOME receivers located in the South-Eastern United States (US) and Puerto Rico. The design of the receiver is described in full by Cohen, et. al [6]. These receivers monitor three Navy VLF transmitters, ranging in frequency from 24.0-25.2 kHz, in the Northern US. The locations of the transmitters and receivers are shown in Figure 1. A significant limitation of the previous model was the manual evaluation of which transmitters and receivers were available on a given day. To resolve this, an automatic quality assessment tool was developed to determine the optimal set of transmitters and receivers for all days in the dataset [2]. This tool utilizes a set of heuristics to quickly determine whether there are any issues with the measurements, such as transmitter or receiver downtime, high noise levels, low signal levels, etc. We can now determine each day's ideal set of transmitters and receivers for the entire dataset in a matter of hours with no manual intervention.

The amplitude and phase measurements are critical to pre-

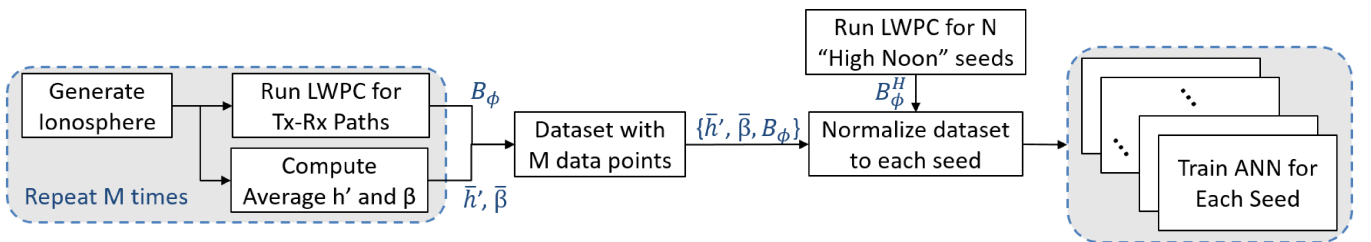


**Figure 1.** A map of the transmitters and receivers used in this work. The great circle paths are indicated by the lines between receivers and transmitters.

dicting electron density ( $N_e$ ), but they do not provide sufficient information to train a machine learning model. In order to train a model to predict  $N_e$ , a dataset containing both an input, amplitude and phase of the VLF waves, and the desired output, parameters describing  $N_e$  vs altitude, is required. For this work, we used a two parameter model ( $h'$  and  $\beta$ ) for  $N_e$  described by (1) [5].

$$N_e(h) = 1.43 \times 10^{13} e^{-0.15h'} e^{\beta(h-h')} m^{-3} \quad (1)$$

Since large scale measurements of D-region ionosphere  $N_e$  are currently impractical, an alternative method of obtaining training data was required. In a recent paper, Gross described a technique for generating synthetic ionospheres [1]. By leveraging this technique, in combination with a modelling software developed by the Navy called long wave propagation capability (LWPC), we generated 200,000 training samples containing  $B_\phi$ , average  $h'$ , and average  $\beta$  for each transmitter-receiver path. The data generation process is illustrated in the leftmost dashed box in Figure 2. To complete the nighttime models, we applied the same technique to generate an additional 200,000 training samples in representative  $h'$  and  $\beta$  ranges. For the daytime synthetic data,  $h'$  and  $\beta$  ranged from 67-77 km and 0.35-0.49  $\text{km}^{-1}$  respectively. For the nighttime synthetic data,  $h'$  and  $\beta$  ranged from 78-88 km and 0.58-0.72  $\text{km}^{-1}$  respectively. These ranges are based on Thomson's D-region observation and modelling efforts [4], [7].



**Figure 2.** Block diagram describing the model training process.

### 3 Model Training

Since each day has a potentially different set of available transmitters and receivers, we train a separate model for each network. While this increases the total training time required, it does not require the model to be robust to missing input features. The problem of missing features is a recognized problem in the field of machine learning, so designing our methods to always have the entire feature set is critical to a well functioning model [8]. To maximize our efforts, networks which occurred more frequently were prioritized during the training process. This prioritization allows us to account for 50% of the total days with only 43 separate network configurations. Since there are 1028 unique network configurations, this is a significant reduction in required computational time.

During the training process we are using the synthetic dataset described previously. Since the dataset does not have any direct relationship to the real world, we introduce a seed  $h'$  and  $\beta$  (and the corresponding  $B_\phi$ ) to use as a normalization factor. In this work, the seed value is used to make an assumption about the conditions of the ionosphere at high noon, or the time at which the sun is most directly overhead. Without knowing the high noon conditions, we are unable to determine the best seed value prior to training the model. Instead, we train an individual neural network for each seed value and select the best performing seed on a day by day basis. This seeding technique was first explored by Gross [1]. As a consequence of the seeding process, we also remove sources of consistent error such as receiver gain or transmitter power varying between days. To improve the computation time, multiple GPUs were used in parallel to reduce the training time from approximately 14 days to approximately 24 hours. The entire model training process can be seen in Figure 2. To extend this technique to nighttime predictions, we replace the high noon seed with a similar midnight seed corresponding to the point at which the sun was least overhead. This allows us to utilize the same daytime modelling process, only changing the training data and range of seed values.

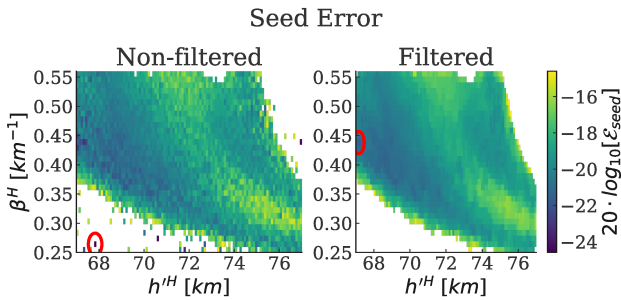
### 4 Optimal Seed Selection

For a given day, we know both the number of available paths,  $N$ , and the duration of the prediction in seconds,  $K$ . To remain consistent with the notation used by Gross, we

will denote the output of LWPC given an input  $h'$ ,  $\beta$  as  $L(h', \beta)$ . Here, LWPC is used to convert the model's predicted  $h'$  and  $\beta$  into a usable magnetic flux density. To determine the optimal seed, we first define the seed error as the RMSE of all the predictions, (2). The seed error is then used to determine how well or poorly that seed's model performed.

$$\mathcal{E}_{\text{seed}}(h'_{\text{seed}}, \beta_{\text{seed}}) = \left[ \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \left| \frac{B_{\phi n,k}}{B_{\phi n}^{\text{Noon}}} - \frac{L(h'_{\text{seed}}, \beta_{\text{seed}})}{L(h'_{n,k}, \beta_{n,k})} \right|^2 \right]^{\frac{1}{2}} \quad (2)$$

The seed error must be computed for each seed on a given day. We expect that seeds close together should result in similar model performance since that represents only a small change in reference value. For most days, this is the case; however, on certain days, we see an isolated well performing seed which indicates the model for that seed is likely over fitting and producing non-realistic values. This effect can be seen in Figure 3, where the red circles indicate the minimum seed error before and after a 3x3 median filter was applied. In the cases without an isolated seed, the 3x3 median filter does not have a drastic effect on the optimal seed. Therefore, it is safe to select the seed which corresponds with the lowest seed error after the 3x3 median filter is applied as the optimal seed for that day.



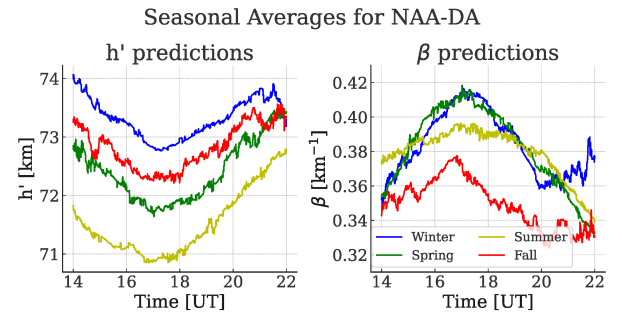
**Figure 3.** Plot of seed error with the red circles indicating the global minimum. The left plot shows an isolated global minimum which likely represents non-realistic predictions that overfit to the data. A 3x3 median filter was applied to remove the isolated minimum.

## 5 Results

### 5.1 Daytime Seasonal Averages

To better understand the model's performance,  $h'$  and  $\beta$  were predicted for 328 days. Due to shorter days and lower levels of ionizing radiation during the winter, we expect the  $h'$  value, which estimates ionosphere height, to be higher. The model predictions were averaged seasonally for the NAA-DA path to illustrate that the model is predicting these trends as shown in Figure 4. NAA-DA was selected because it was the most frequently available path in the 328 days used, but similar trends can be seen on other paths not

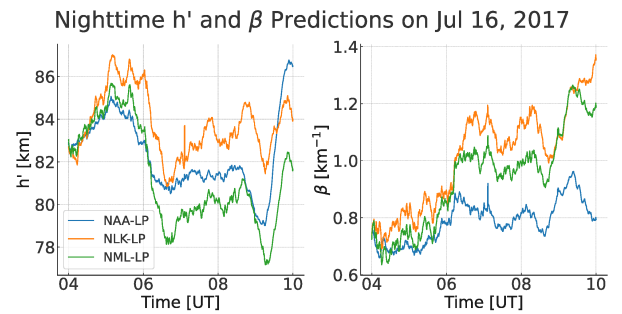
shown here. While the model may make some mistakes, the agreement between the model's seasonal averages and our expectations indicates we can trust the model's predictions in aggregate. In addition to showing the seasonal shift, Figure 4 also illustrates that, on average, the model predicts a mostly parabolic curve for both  $h'$  and  $\beta$ . This agrees with the expectation that  $h'$  reaches a minimum at high noon and increases away from that point and vice versa for  $\beta$ . By verifying the model's daytime predictions agree with known trends, we can be confident that the model is making reasonable predictions. This work demonstrates the first large scale application of using synthetic ionosphere data to train machine learning models for the D-region ionosphere.



**Figure 4.** Plot demonstrating the seasonal variation in model predictions for NAA-DA. The increase in  $h'$  during the winter and fall align with our expectations of ionospheric seasonal variation.

### 5.2 Nighttime Predictions

During the night, the ionosphere moves up in altitude. The reduction in solar radiation makes the D-region ionosphere more chaotic during the night. These two factors make predicting electron density from VLF waves a difficult problem. In this work, we present nighttime  $h'$  and  $\beta$  predictions for July 16, 2017. As seen in Figure 5, the range of  $h'$  values are reasonable; however, we see some unexpectedly large values for  $\beta$  during the second half of the night. While these are only early results, they indicate the model is likely better able to predict the height of the D-region ionosphere compared to the rate  $N_e$  changes with altitude.



**Figure 5.** Nighttime predictions for three paths on July 16, 2017. The  $h'$  predictions show reasonable values that are within our expectations. The  $\beta$  values extend above what we expect to be reasonable beginning shortly after 6 UT.

## 6 Discussion and Conclusion

Previous work on predicting D-region  $N_e$  has often been limited to a few days and a limited number of transmitter-receiver paths [1], [4]. Additionally, traditional VLF remote sensing is susceptible to calibration errors and terrain artifacts [4], [9], [10]. By normalizing to the noon-time value, we improve the model's robustness to these effects. In this work, we demonstrated the use of a machine learning model for predicting waveguide parameters of the D-region ionosphere across 328 days and up to 21 paths. We show the model's agreement with expected seasonal variations indicating the model is capable of coping with a range of ionospheric conditions. The daytime modelling technique was also expanded to work for nighttime conditions. Since nighttime data is lacking a high noon reference value, we instead shift to using midnight as a reference value, although other options may be explored in the future. By only changing the reference time, we can apply the same technique to both daytime and nighttime measurements, greatly simplifying the overall model. Initial results from the nighttime predictions indicate the model is able to predict reasonable values for  $h'$  with some unexpected behavior in  $\beta$  predictions. Predictions during nighttime conditions are traditionally more difficult and less well understood, so it is reasonable for the model to struggle in this regime.

As the duration of predictions increases, transmitter phase drift may become an issue as this is not accounted for through the high noon seed. Fortunately, polarization is invariant to the absolute transmitter phase which may allow us to circumvent the problem. With this large dataset, it may now be possible to compare single day measurements to a quiet day case to detect events such as gravity and acoustic waves. While not shown here, model predictions during Hurricane Irma hint at this being possible in the near future. By transitioning to a four parameter  $N_e$  profile and smaller spatial features in the training data, the model will likely be better equipped to handle some of the smaller perturbations during these gravity and acoustic waves [11]. As a by product, this may also improve the model's performance during the more chaotic nighttime environment.

## 7 Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1148903. This work has been supported by the Defense Advanced Research Projects Agency (DARPA) through US Department of the Interior award D19AC00009 to the Georgia Institute of Technology.

## References

- [1] N. C. Gross and M. B. Cohen, "Vlf remote sensing of the d region ionosphere using neural networks," *Journal of Geophysical Research: Space Physics*, vol. 125, no. 1, Jan. 2020, ISSN: 2169-9402. DOI: 10.1029/2019ja027135. [Online]. Available: <http://dx.doi.org/10.1029/2019JA027135>.
- [2] D. Richardson. (2020). "Lf," [Online]. Available: <https://lf.readthedocs.io/en/latest/> (visited on 09/16/2020).
- [3] A. C. Aikin, J. A. Kane, and J. Troim, "Some results of rocket experiments in the quietdregion," *Journal of Geophysical Research*, vol. 69, no. 21, pp. 4621–4628, Nov. 1964, ISSN: 0148-0227. DOI: 10.1029/jz069i021p04621. [Online]. Available: <http://dx.doi.org/10.1029/JZ069i021p04621>.
- [4] N. Thomson, "Experimental daytime vlf ionospheric parameters," *Journal of Atmospheric and Terrestrial Physics*, vol. 55, no. 2, pp. 173–184, Feb. 1993, ISSN: 0021-9169. DOI: 10.1016/0021-9169(93)90122-f. [Online]. Available: [http://dx.doi.org/10.1016/0021-9169\(93\)90122-F](http://dx.doi.org/10.1016/0021-9169(93)90122-F).
- [5] J. R. Wait and K. P. Spies, *Characteristics of the Earth-Ionosphere Waveguide for VLF Radio Waves*. Boulder, Colorado: National Bureau of Standards, Dec. 1964.
- [6] M. B. Cohen, R. K. Said, E. W. Paschal, J. C. McCormick, N. C. Gross, L. Thompson, M. Higginson-Rollins, U. S. Inan, and J. Chang, "Broadband long-wave radio remote sensing instrumentation," *Review of Scientific Instruments*, vol. 89, no. 9, p. 094 501, Sep. 2018, ISSN: 1089-7623. DOI: 10.1063/1.5041419. [Online]. Available: <http://dx.doi.org/10.1063/1.5041419>.
- [7] N. R. Thomson and W. M. McRae, "Nighttime ionosphericdregion: Equatorial and nonequatorial," *Journal of Geophysical Research: Space Physics*, vol. 114, no. A8, n/a–n/a, Aug. 2009, ISSN: 0148-0227. DOI: 10.1029/2008ja014001. [Online]. Available: <http://dx.doi.org/10.1029/2008JA014001>.
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [9] N. R. Thomson, "Re-radiation of vlf radio waves from mountain ranges," *Journal of Atmospheric and Terrestrial Physics*, vol. 51, no. 4, pp. 339–349, Apr. 1989, ISSN: 0021-9169. DOI: 10.1016/0021-9169(89)90084-6. [Online]. Available: [http://dx.doi.org/10.1016/0021-9169\(89\)90084-6](http://dx.doi.org/10.1016/0021-9169(89)90084-6).
- [10] N. R. Thomson, C. J. Rodger, and M. A. Clilverd, "Daytime d region parameters from long-path vlf phase and amplitude," *Journal of Geophysical Research: Space Physics*, vol. 116, no. A11, n/a–n/a, Nov. 2011, ISSN: 0148-0227. DOI: 10.1029/2011ja016910. [Online]. Available: <http://dx.doi.org/10.1029/2011JA016910>.
- [11] J. McCormick, "D-region tomography: A technique for ionospheric imaging using lightning generated sferics and inverse modeling," Ph.D. dissertation, Georgia Institute of Technology, 2019.