Chorus waves modeled by an artificial neural network: The importance of imbalanced regression

Xiangning Chu[1], Jacob Bortnik[2], Wen Li[3], Xiao-Chen Shen[3], Qianli Ma[2,3], Donglai Ma[2], David Malaspina[1], Sheng Huang[3]

[1] Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, USA [2] Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA, [3] Center for Space Physics, Boston University, Boston, MA, USA

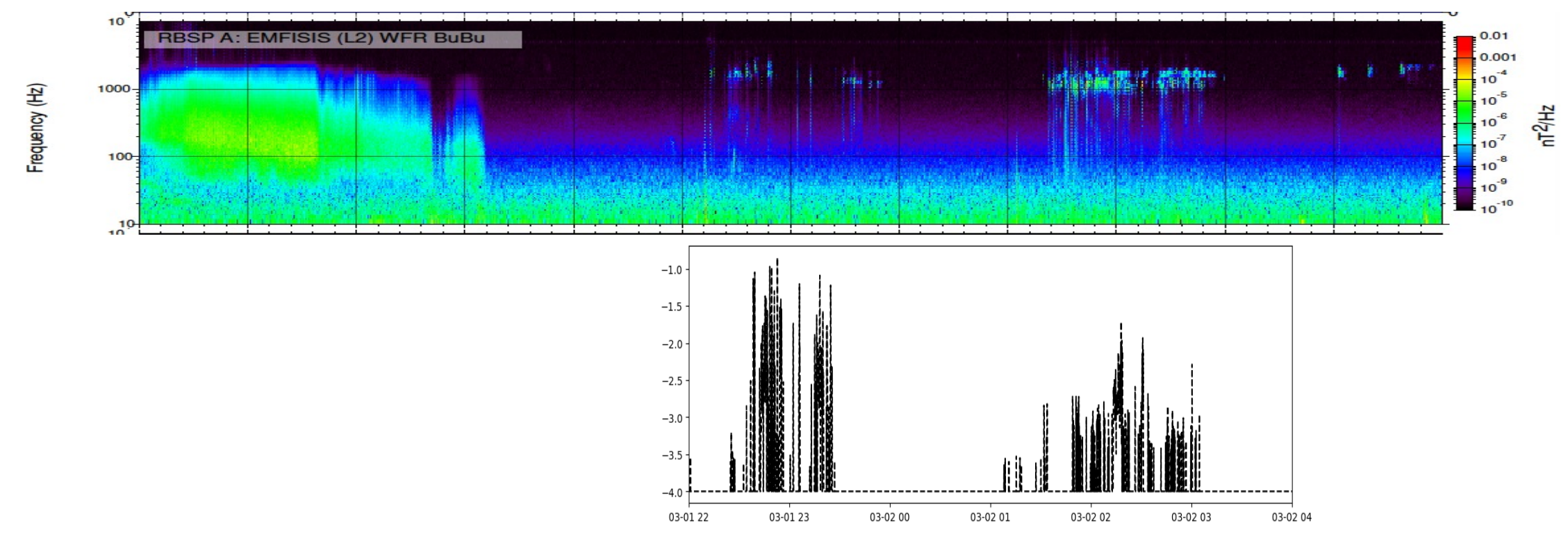
Space Weather Workshop 2023
Boulder, CO

Introduction

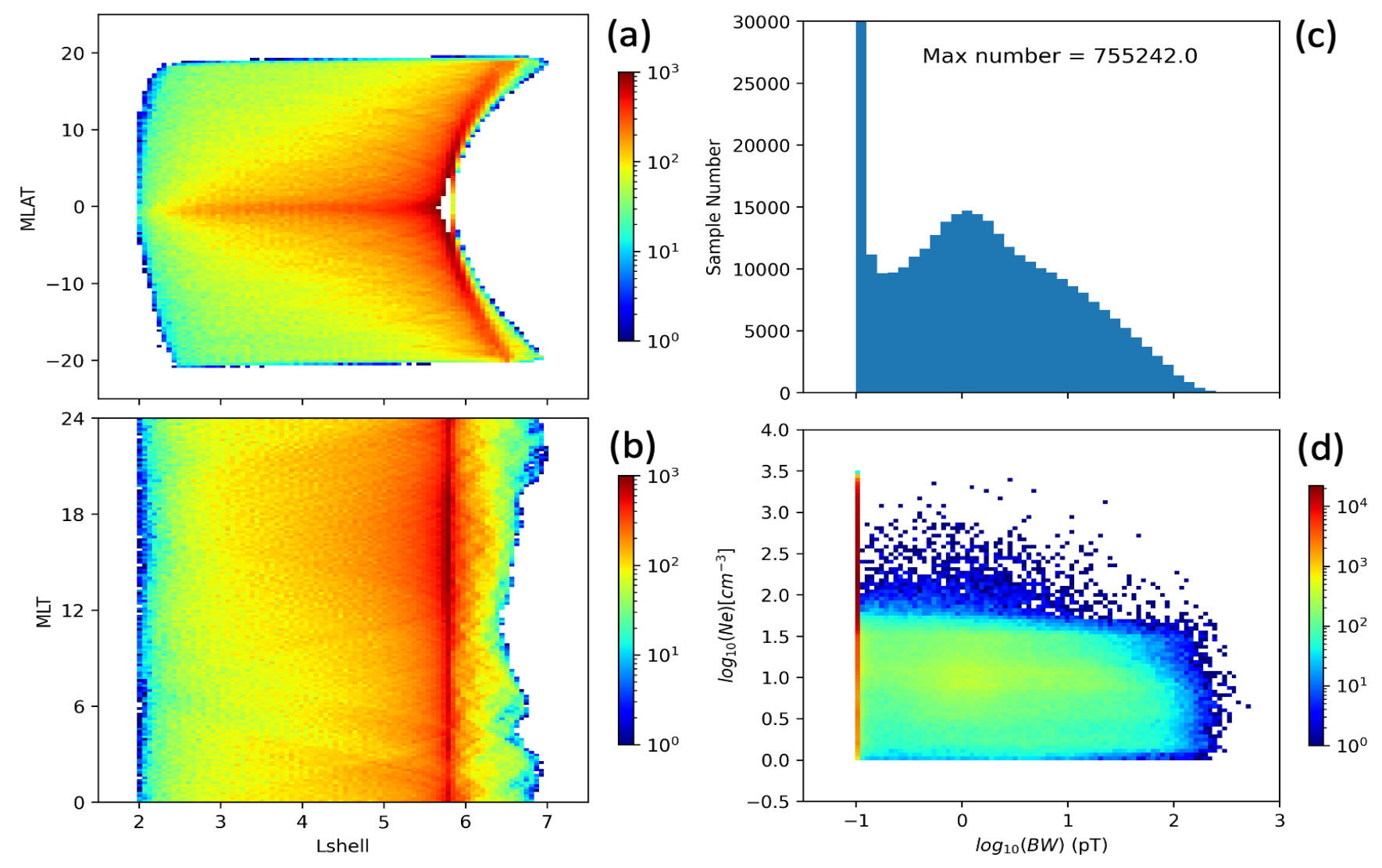
- The relativistic electron acceleration and loss processes that occur under various geomagnetic conditions are of key interest in the Earth's radiation belt research.
- Whistler-mode chorus waves play an important role in the local acceleration in accelerating seed electrons.
- The shortcomings of current Fokker-Planck simulations:
- The plasma and wave environment specified using in-situ spacecraft measurements have three disadvantages.
  - First, they are limited by the magnetic local times (MLT) of the spacecraft orbits, while the plasma and wave distributions are usually highly asymmetric in MLT.
  - The empirical statistical models, on the other hand, usually provide statistical averaged distributions and cannot resolve event-specific or time-dependent variations in the plasma and wave environments.
  - The chorus waves data set is highly imbalanced, suffering from the 'too-often-too-quiet' problems. In other words, statistical models usually underestimates the amplitude of strong waves, which is interesting and important.
- The shortcomings of current statistical-averaged chorus models
  - Data imbalance is a ubiquitous problem inherent in the real world, including machine learning and space physics.
  - The large volume of quiet-time data dominates the statistical-averaged models, either linear regression or neural network models.
  - The strong activity, which is more interesting to space weather, are underestimated significantly.
  - Imbalanced regression is less investigated in machine learning.
- In this study, we developed an imbalanced regressive (IR) neural network (NN) chorus model.
  - By including the MPB index and applying an imbalanced regression technique developed in our group,
  - For the first time, we have an IR chorus model that can correctly predict the amplitude of the strong chorus waves.
  - The distribution and evolution of the chorus waves are investigated using the IR-NN chorus model.

Imbalanced Regression (IR) and dataset

- The chorus wave amplitude is obtained from the EMFISIS instruments onboard the Van Allen Probes.
- The LB chorus waves are identified using the following criteria: (1) they occur outside the plasmapause, (2) within the frequency range of 0.05-0.5 fce, (3) they have planarity > 0.6, and (4) ellipticity > 0.7 (see detailed description in Li et al. (2016) and Shen et al., (2019)), (5) for observations with no chorus waves, the wave amplitude is filled by 0.1 pT as the lower threshold.
- The measurements are taken between L=[2, 7], and MLAT=[-20º, 20º].
- The chorus waves are sporadic, with more quiet samples than strong waves
- The histogram of Bw shows a highly imbalanced dataset. There are 10 times more quiet time samples (< 1pT) than chorus wave samples.
- The chorus waves are well organized by the plasmapause.



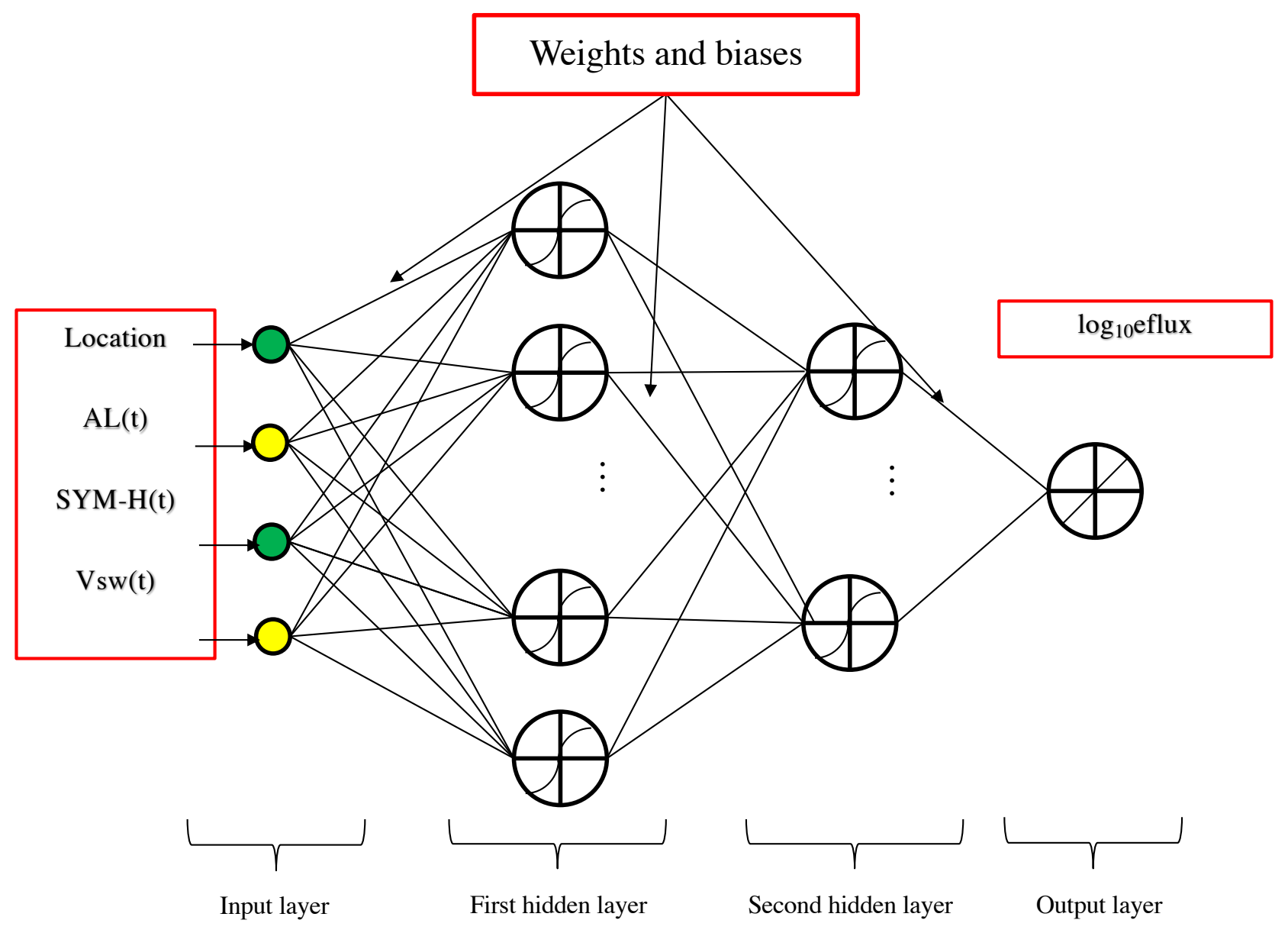
The wave spectrum from the waveform receiver (WFR) on the EMFISIS wave instrument, and the lower-band chorus wave amplitude.



Statistical properties of the LB chorus wave amplitude $\log_{10}(B_w)$. The numbers of data samples as a function of (a) L shell and MLAT, (b) L shell and MLT, (c) wave amplitude, and (d) plasma density and wave amplitude.

Data and Model description

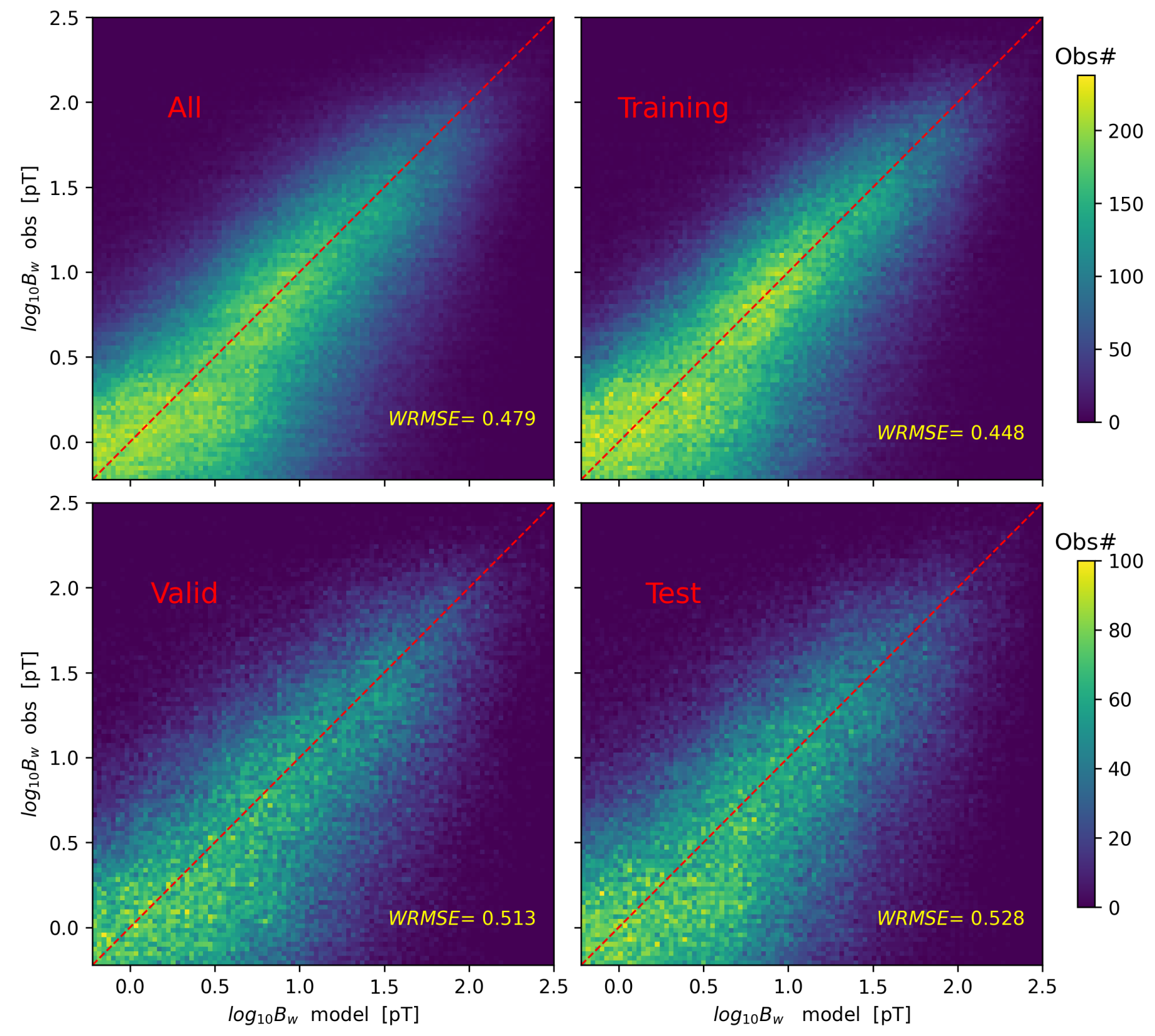
- Chorus wave amplitudes obtained from EMFISIS onboard Van Allen Probes from January 2013 to September 2019
- There are 66 million data points in total.
- The input parameters, including solar wind parameters and geomagnetic indices, are obtained from the OMNI dataset, and the MPB index
- A fully-connected neural network is employed with time series of input parameters and predicts the logarithm of the electron fluxes.
- The neural network is generalized using a few methods: normalization of input parameters, regularization, batch normalization, dropouts, and early stopping.
- Imbalanced regression techniques developed and applied.
- Feature selection and hyperparameter optimization:
  - The SME index is the best and only index that we need.



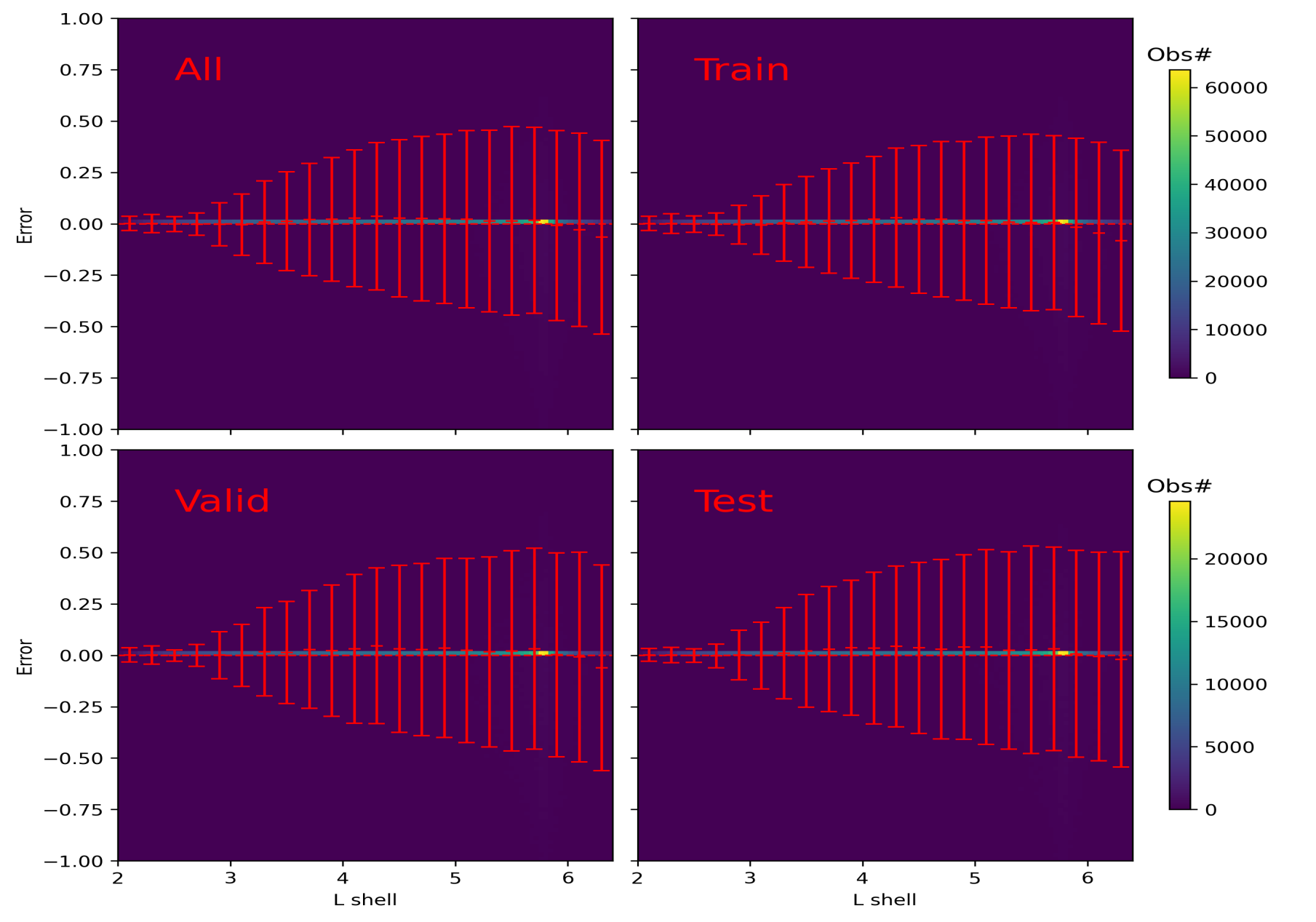
The architecture of the neural network model.

Model performance

- Applying imbalanced regression technique yields a non-biased distribution at high wave powers.
- The IR-NN chorus model reproduces the observations without much over- or under-estimation, regardless of the chorus amplitude $\log_{10}(B_w)$.
- The observation-model data pairs now lie along the diagonal line, suggesting that the model can correctly capture the wave amplitude.
- The model error is well organized by the electron density, showing a sharp transition at the plasmapause.



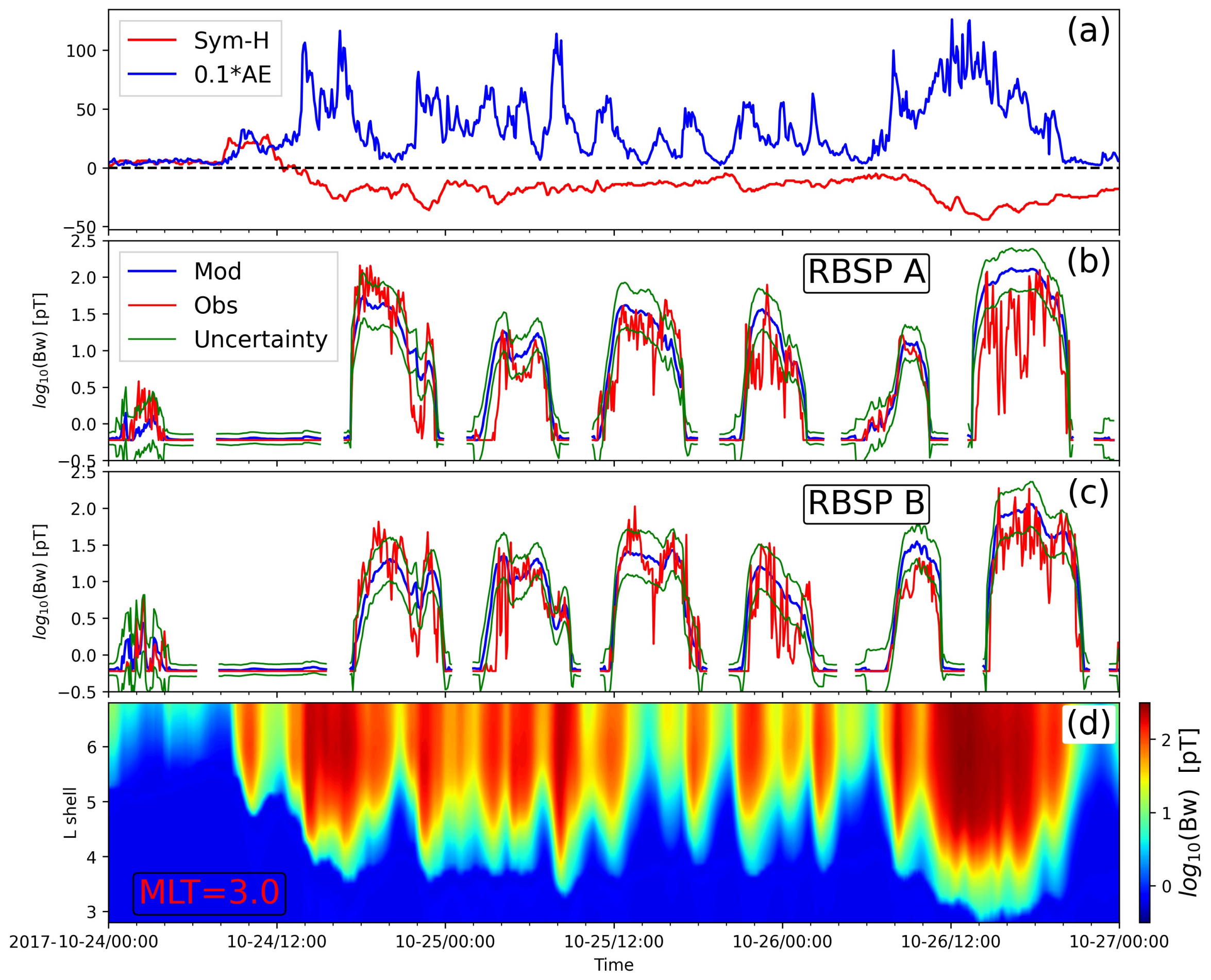
The two-dimensional distribution of the model predicted and observed LB chorus wave amplitude for four datasets (all, training, validation, and test).



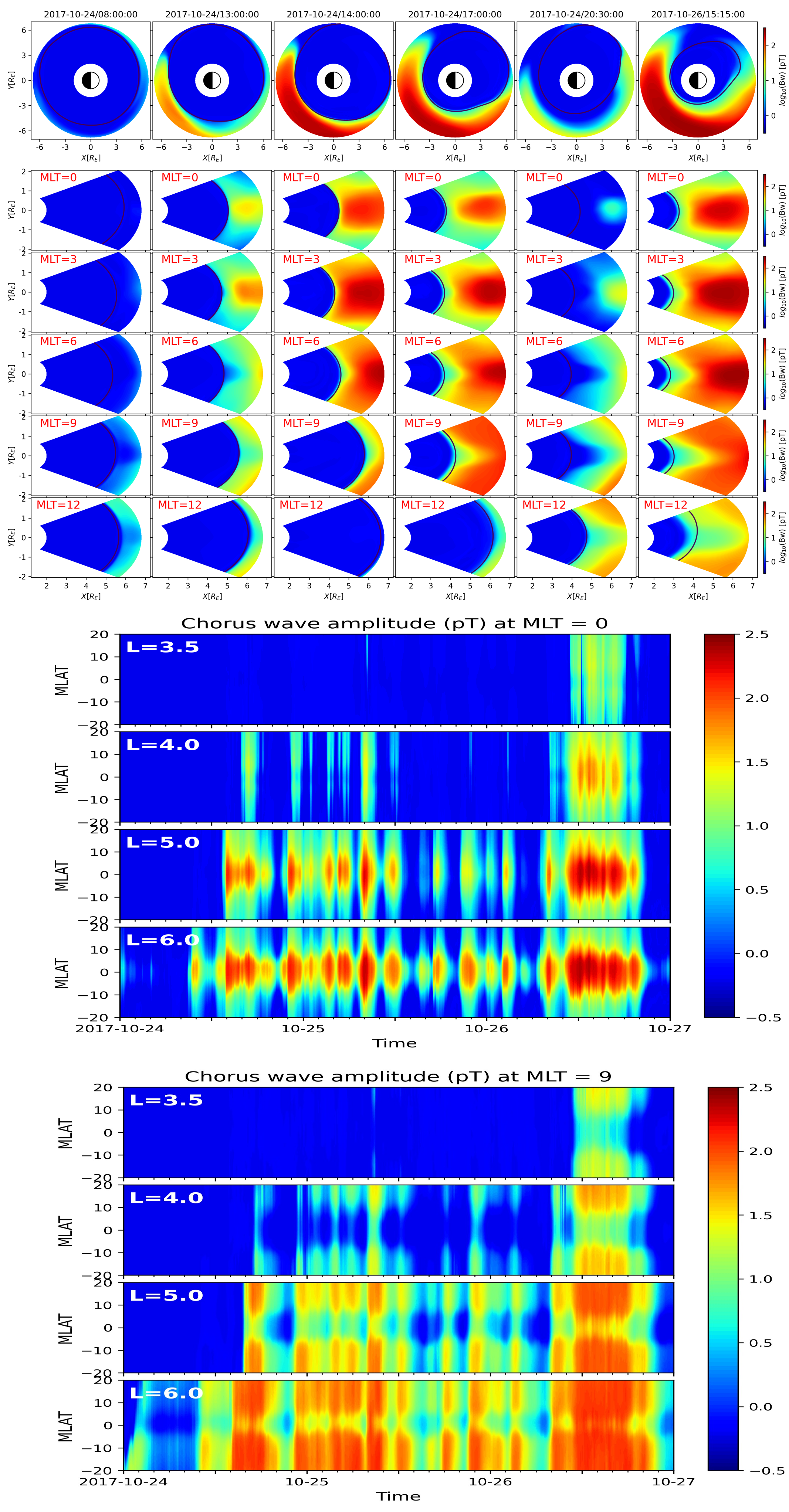
The error distribution as a function of the L shell for the four datasets (all, training, validation, and test).

Event Analysis

- The chorus model is validated using out-of-sample measurements along Van Allen Probes' trajectory during two storm events.
- The variations of chorus waves along the trajectory are well captured.
- The modeled amplitude of strong chorus waves matches observations.
- The quiet time amplitudes are also well predicted by the chorus model.



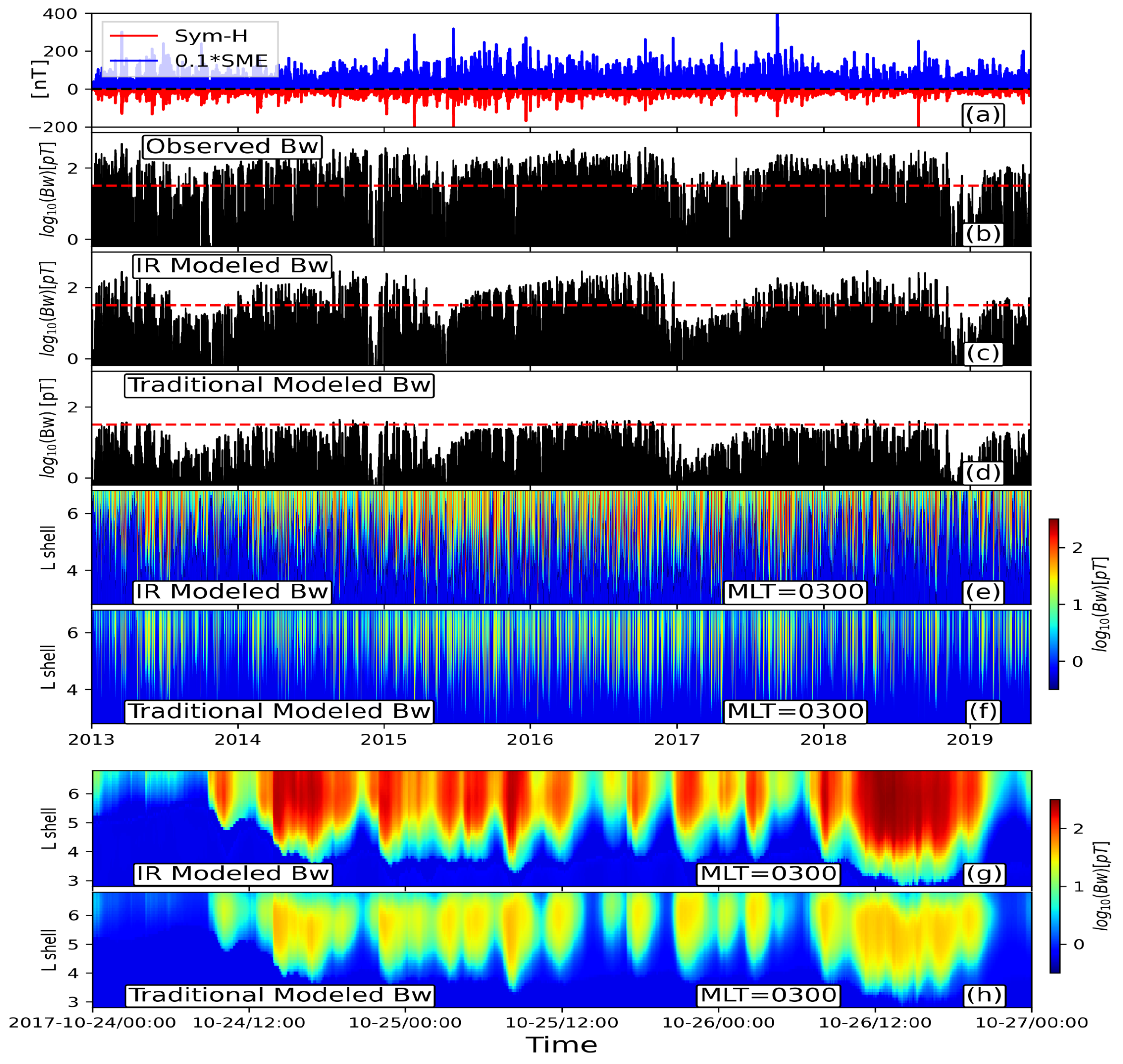
Comparison between the observed chorus amplitude and IR-NN chorus modeled results during October 24-27, 2017.



The evolution of chorus wave amplitude on the equatorial planes (top row) and meridian planes (bottom five rows) at different MLTs (MLT=0, 3, 6, 9, and 12). The temporal evolution of chorus wave amplitude as a function of MLAT at different L shells (L=3.5, 4.0, 5.0, and 6.0) near midnight (MLT=3, left) and near noon (MLT =9, right).

Why do we care?

- From a perspective of radiation belt physics, an IR-NN chorus model is essential.
  - Note that the IR chorus model could predict the strong chorus waves (300 pT).
  - The traditional NN model is capped at about 30 pT, and cannot predict strong chorus waves.
  - The traditional NN model significantly underestimate the most important chorus waves by a factor of > 10.
  - The acceleration of relativistic electrons depends on the averaged wave power across all local times.
  - The acceleration of electrons by the chorus waves may be underestimated by a factor of 100.
  - Correctly predicting the wave amplitude of strong chorus waves is essential to correctly simulate the acceleration of relativistic electrons.
- From perspective of space weather, an imbalanced regressive technique is essential:
  - All space physics/weather data are imbalanced.
  - Large-to-extreme events are rare compared to quiet times.
  - Prediction of space weather are dominated by quiet times, usually significantly underestimate the amplitude of geomagnetic events.
  - E.g., DST, KP, auroral electrojet indices (AE/AU/AL), GIC, radiation belt fluxes, TEC scintillations.
  - Correctly predict the amplitude of these events is important in sending out alerts with confidence.
- From the perspective of machine learning.
  - Data imbalance is a ubiquitous problem inherent in the real world.
  - Imbalanced regression is less investigated in general, although imbalanced classification attracted much attention.



The comparison between the observed chorus amplitude and those modeled by the imbalanced regressive and traditional neural network (NN) models.

Discussion

- We develop a neural network model of lower-band chorus waves using an imbalanced regressive technique.
- Our IR chorus model can correctly predict the amplitude of the strong chorus waves, for the first time.
- A pilot model is developed to provide the model uncertainties.
- The equatorial evolution of the chorus waves is consistent with the electron drift path of substorm injections.
- The chorus waves peak at the equator (plasma sheet) in the source MLT near midnight. They show a minimum at the equator toward noon, with two off-equator amplitude peaks in two hemispheres.
- Imbalanced regression methods require more attention since most datasets in space physics, space weather, and real world are imbalanced.
- This work is supported by NASA ECIP award 80NSSC19K0911, NASA award 80NSSC20K0196, 80NSSC22K1023, 80NSSC18K1227, XSEDE bridges allocation TG-PHY190033 . WL, XS, and QM acknowledge NASA grants 80NSSC19K0845, 80NSSC20K0698, and NSF grant AGS-1847818.