JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD043577

Special Collection:
Advances in Machine Learning for Earth Science: Observation, Modeling, and Applications

Toward Real-Time Assessment of Infrasound Event Detection Capability Using Deep Learning-Based Transmission Loss Estimation

A. Janela Cameijo[1,2], A. Le Pichon[1], Y. Sklab[3], S. Arib[4], Q. Brissaud[5], S. P. Näsholm[5,6], C. Listowski[1], and S. Aknine[2]

[1]CEA, DAM, DIF, Arpajon, France, [2]Université Lyon 1, LIRIS, Villeurbanne, France, [3]Sorbonne Université, IRD, UMMISCO, Bondy, France, [4]Laboratoire ETIS, CY Cergy Paris Université, CNRS, Cergy-Pontoise, France, [5]Solutions Department, NORSAR, Kjeller, Norway, [6]Department of Informatics, University of Oslo, Oslo, Norway

Abstract Accurate modeling of infrasound transmission loss is crucial for assessing the performance of the International Monitoring System, which monitors compliance with the Comprehensive Nuclear-Test-Ban Treaty by detecting atmospheric explosions. This modeling supports the design and maintenance of the operating monitoring network. State-of-the-art propagation modeling tools enable transmission loss to be finely simulated using atmospheric models. However, the computational cost prohibits the exploration of a large parameter space in operational monitoring applications. To address this, recent studies made use of a deep learning algorithm capable of making transmission loss predictions almost instantaneously. However, the use of nudged atmospheric models leads to an incomplete representation of the medium, and the absence of temperature as an input makes the algorithm incompatible with long-range propagation. In this study, we address these limitations by using both wind and temperature fields as inputs to a neural network, simulated up to 130 km altitude and 4,000 km distance. We exploit convolutional and recurrent layers to capture spatially and range-dependent features embedded in realistic atmospheric models, improving the overall performance. The neural network reaches an average error of 4 dB compared to full parabolic equation simulations and provides epistemic and data-related uncertainty estimates. Its evaluation on the 2022 Hunga Tonga-Hunga Ha'apai volcanic eruption demonstrates its prediction capability using atmospheric conditions and frequencies not included in the training. This represents a significant step toward near real-time assessment of International Monitoring System detection thresholds of explosive sources.

Plain Language Summary Accurate modeling of infrasound transmission loss is essential in a wide range of applications, such as improving atmospheric data assimilation for numerical weather prediction, assessing attenuation maps of sources of interest, or estimating the spatial and temporal variability of the International Monitoring System infrasound network performance. However, the high computational cost of numerical modeling solvers makes them impractical in near real-time analysis. To address this, we develop a convolutional recurrent neural network able to predict ground-level transmission losses for a propagation range of 4,000 km and for five frequencies ranging from 0.1 to 1.6 Hz in 0.045 s. The proposed method exploits range-dependent atmospheric specifications that combine horizontal wind speed and temperature fields, including small-scale atmospheric perturbations. In comparison with the state-of-the-art neural network (Brissaud et al., 2023, https://doi.org/10.1093/gji/ggac307), the proposed model achieves an average error of the same magnitude while extending the propagation range at a global scale and providing estimates of epistemic and data-related uncertainty. The model is evaluated on the 2022 Hunga Tonga-Hunga Ha'apai volcanic eruption and demonstrates its performance in a general setting by providing accurate predictions at new sampling regions, new dates, and new source frequencies.

Key Points:

- We provide a convolutional recurrent neural network estimating in near real-time ground-level infrasound transmission loss
- The neural network exploits spatial and range-dependent features in atmospheric models and makes predictions with an uncertainty estimate
- The network can be used as a tool for near real-time estimation of infrasound event detection capability at a global scale

Correspondence to:
A. J. Cameijo,
alice.cameijo@cea.fr

Citation:
Cameijo, A. J., Le Pichon, A., Sklab, Y., Arib, S., Brissaud, Q., Näsholm, S. P., et al. (2025). Toward real-time assessment of infrasound event detection capability using deep learning-based transmission loss estimation. *Journal of Geophysical Research: Atmospheres*, *130*, e2025JD043577. https://doi.org/10.1029/2025JD043577

Received 7 FEB 2025
Accepted 12 SEP 2025

Author Contributions:
Conceptualization: A. Janela Cameijo, A. Le Pichon, Q. Brissaud, S. P. Näsholm
Methodology: A. Janela Cameijo, A. Le Pichon, Y. Sklab, S. Arib
Resources: A. Le Pichon
Software: A. Janela Cameijo
Supervision: A. Le Pichon
Validation: A. Le Pichon, Y. Sklab, S. Arib, Q. Brissaud, S. P. Näsholm, C. Listowski, S. Aknine
Writing – original draft: A. Janela Cameijo

1. Introduction

Many high-energy atmospheric phenomena, whether natural (meteoroids, earthquakes, or volcanoes) or human-made (aircraft, chemical or nuclear explosions), generate acoustic waves at inaudible frequencies (≤20 Hz), called infrasound. Such waves can propagate thousands of kilometers through various atmospheric waveguides and be refracted back to the surface, allowing for detection on a global scale. For that reason, they are continuously recorded by the ground-based infrasound stations of the International Monitoring System (IMS). In order to monitor compliance with the Comprehensive Nuclear-Test-Ban Treaty, the IMS was designed to allow the

Writing – review & editing:
A. Janela Cameijo, A. Le Pichon, Y. Sklab, Q. Brissaud, S. P. Näsholm, C. Listowski, S. Aknine

detection of any explosions with a yield of one kiloton of TNT (Christie et al. (2022); Marty (2019)). The International Data Center uses automatic processes to identify such explosions among the mass of data recorded by the IMS stations, providing an estimate of the wavefront parameters useful for source location and characterization (back azimuth, apparent velocities, amplitudes, frequencies; Mialle et al. (2019)). The IMS infrasound recordings play a key role in the development of new advanced processing methods leveraging deep learning algorithms (e.g., Bishop et al. (2022); Albert and Linville (2020)).

Transmission loss (TL) is the cumulative decrease in acoustic energy as waves propagate. Typically, TL is given in decibels and in terms of amplitude at range divided by the amplitude at a fixed reference distance. Accurate modeling of TL is essential for interpreting the IMS infrasound station measurements and evaluating event detection thresholds (Green and Bowers (2010); Le Pichon et al. (2009)). Such evaluation represents a key step toward optimizing the design of the IMS network to effectively monitor infrasonic sources worldwide. As infrasound recordings can provide information about various atmospheric properties, such as winds and temperatures, accurate modeling of TL can also help to better infer these properties at altitudes where direct atmospheric measurements are scarce, through inverse methods (Amezcua et al. (2024); Assink (2012); Blixt et al. (2019); Letournel et al. (2024); Le Pichon et al. (2005); Smets and Evers (2014); Vera Rodriguez et al. (2020)).

The computational cost of existing numerical propagation modeling tools, such as finite-difference codes (Brissaud et al. (2016); de Groot-Hedlin et al. (2011)), spectral element methods (Brissaud et al. (2017); Martire et al. (2022)), normal modes (Waxler et al., 2021), or parabolic equation solvers (Waxler et al., 2021), does not allow the exploration of a wide parameter space (variations in the atmospheric state, in frequencies, or in source location) for near real-time TL estimation. This makes these tools inconvenient for operational frameworks with real-time calculation requirements, such as the monitoring of explosive sources by IMS stations. To overcome prohibitive computational time, less expressive methods such as ray tracing can be used. This approach estimates TLs at a global scale, integrating complex atmospheric data (e.g., multiple wind components, strong vertical stratification, horizontal dependencies) as well as Earth's topographical relief. However, it is limited by the presence of shadow zones (Gutenberg (1939) chapter 8; Pierce (2019)) for which the predicted TL is underestimated. Moreover, this method is a high-frequency approximation that neglects diffraction-related effects that occur in finite-frequency wave propagation. Le Pichon et al. (2012) propose an alternative approach, with empirical propagation equations relying on simplified assumptions. The model is based on multiparametric regression fitting to the output of parabolic equation solvers to predict TL at ground level. However, this was derived from idealized atmospheric specifications neglecting range-dependent variations and focusing only on stratospheric wind speeds at 50 km altitude. As a result, the model struggles to predict TL when the structure of the atmospheric waveguides varies along the propagation path, in the absence of stratospheric waveguides, or at high frequencies ($\geq$1 Hz).

Machine learning techniques are currently explored in the field of acoustic wave propagation modeling in the atmosphere. Brissaud et al. (2023) enhance the work of Le Pichon et al. (2012) by developing a deep learning algorithm predicting infrasound TL from realistic range-dependent atmospheric fields simulated at a regional scale. The method exploits a convolutional neural network (LeCun et al., 2015) to estimate ground-level TL at around 5 dB error compared to parabolic equation-based simulations. A key aspect is the negligible computation times at the inference stage (around 0.05 s per prediction) compared to physical solvers. The TL is estimated for a given frequency between 0.1 and 3.2 Hz. While these results are promising, the propagation range of 1,000 km is a limitation when performing a global-scale investigation of TL, for example, when studying long-range propagation of microbaroms (Vorobeva et al., 2021) or when predicting the IMS detection thresholds, where the average distance between infrasound stations is around 2,000 km. Moreover, the input data used to train the model are built using an interpolation of various atmospheric specifications with different resolutions. This leads to an incomplete resolution of the atmospheric variability. A simplification is finally made by considering only wind speeds as inputs of the neural network.

Pettit and Wilson (2020) introduced a physics-aware neural network to model acoustic wave propagation through the atmospheric boundary layer ($\leq$2 km altitude) and recover underlying physical parameters. The propagation range is kept within 1 km from the source, and only audible frequencies are considered (50 and 150 Hz). The proposed model is a physics-informed neural network. It is a data-driven algorithm accounting for physical constraints by integrating an additional regularization term in its loss function. This approach provides a good

approximation of the TL fields at a local scale but fails to recover the underlying physical parameters not provided as inputs (e.g., friction velocity, surface heat flux). Another study developed a fully connected neural network to model two-dimensional acoustic TL from a set of predefined inputs describing a turbulent atmosphere (Hart et al., 2021). Errors of about 7 dB were obtained over a 10-km distance, but the atmospheric modeling is reduced to 13 classes, which does not allow for a detailed characterization of the medium.

In this study, we enhance the work of Le Pichon et al. (2012) and address the limitations of Brissaud et al. (2023) by exploiting realistic range-dependent wind speed and temperature fields simulated up to 4,000 km distance and by optimizing the deep learning architecture. The key optimization relies on the use of recurrent layers in addition to convolutional layers. The convolutional layers capture the spatially local features embedded in the input atmospheric specifications. The recurrent ones capture the range dependencies, as infrasound waves travel only in the forward direction so that the attenuation at a range d does not depend on the atmospheric state beyond. The proposed neural network is trained on a large set of 4,000-km-long parabolic equation simulations. Predictive uncertainties are quantified by considering the uncertainty related to the model architecture and to the input data. We retrospectively apply our model on the atmospheric conditions of 15 January 2022 to emulate a near real-time TL map around the Hunga Tonga-Hunga Ha'apai (now referred to as Tonga) volcano, whose major eruption generated infrasound that propagated over several thousand kilometers (Vergoz et al., 2022). The network is also evaluated on new source frequencies, representing a concept validation and a first step toward the evaluation of the minimum detectable explosive energy by the IMS (Le Pichon et al., 2009).

The current paper is organized as follows. In Section 2, the data used to train the neural network are presented. Section 3 details the architecture of the convolutional recurrent neural network. Section 4 evaluates the model performances on testing and generalization data. Section 5 quantifies the epistemic and the data uncertainties. Section 6 presents the Tonga case study, and Section 7 reviews and summarizes our findings.

2. Data for Transmission Loss Estimation

2.1. Inputs: Realistic Range-Dependent Atmospheric Slices

Infrasound propagation is sensitive to the atmospheric medium, in particular to horizontal and vertical gradients of sound speed, which is directly related to temperature and wind (de Groot-Hedlin et al., 2009). Waves traveling through the atmosphere can be refracted down to the ground by positive gradients or upward by negative ones. Hence, infrasound can propagate through different waveguides, depending on the wind and the temperature in the troposphere (0–12 km altitude), stratosphere (12–60 km altitude), and mesosphere (60–90 km altitude). Above 90 km altitude, in the thermosphere, the temperature greatly increases, leading to a permanent thermospheric waveguide. Within the geometric ray-tracing approximation, the effective sound speed ratio c_{ratio} indicates the presence or the absence of such waveguides:

$$c_{\text{ratio}}(z) = \frac{c_{\text{eff}}(z)}{c_{\text{eff}}(z=0)} \quad ; \quad c_{\text{eff}}(z) = u_0(z) + c(z), \tag{1}$$

where $u_0(z)$ is the horizontal component of the wind speed in the direction of propagation and $c(z) = \sqrt{\gamma RT(z)}$ is the adiabatic sound speed with γ the adiabatic index, R the specific gas constant for air, and T the absolute temperature. For sources at ground level, the condition $c_{\text{ratio}}(z) \geq 1$ indicates the presence of a waveguide at the altitude z, which refracts the wave back toward the surface. Infrasound is then reflected by the ground and propagates through successive reflections between the surface and the upper atmosphere. The c_{ratio} approximation holds only for high frequencies, flows with small Mach numbers, and shallow propagation angles. It does not take into account the influence of crosswinds or diffraction. However, as shown in Assink (2012), infrasound TL is not very sensitive to such winds.

The atmospheric input data used to train the neural network are realistic slices $A_{z,d}$ of c_{ratio}. The slices are vertical-horizontal planes covering the altitude $z \in [0, 130]$ km and the distance from the source $d \in [0, 4000]$ km (see Figure 1). The temperatures and wind speeds used to calculate the c_{ratio} are extracted from the Whole Atmosphere Community Climate Model (WACCM) forecast products (Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, University Corporation for Atmospheric Research; Gettelman et al. (2019)). We use the sixth version of this product, providing a horizontal resolution of 100 × 140 km and a

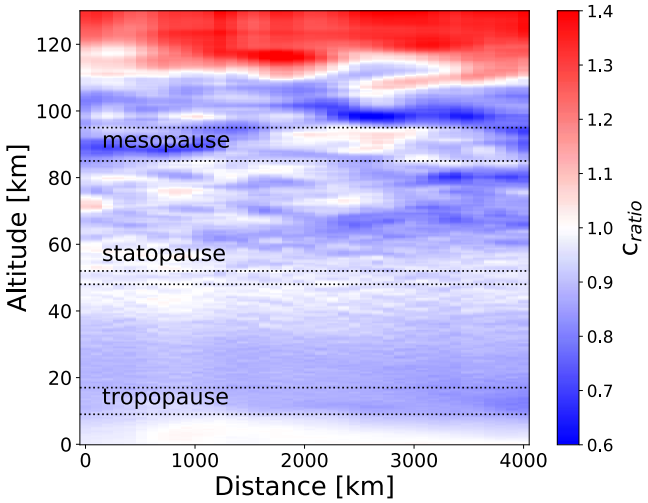

Figure 1. Two-dimensional c_{ratio} field constructed using temperatures and horizontal wind speeds extracted from the WACCM model. This example highlights the waveguide due to the strong increase of the temperature above the mesopause as well as a weaker waveguide in the stratosphere ($c_{ratio} \geq 1$ in red and white, respectively).

model top above 130 km, distributed across 88 vertical levels. The horizontal winds are calculated by projecting the zonal and meridional components along the wave propagation path from the source to the receiver. In contrast, Brissaud et al. (2023) used the ERA5 high-resolution reanalysis product from the European Center for Medium-Range Weather Forecasts to extract wind speeds up to the mesosphere. This model possesses 137 vertical levels but reaches a maximum altitude of 80 km (Hersbach et al., 2020). To represent the atmosphere beyond this limit, Brissaud et al. (2023) used two climatologies (Drob et al. (2015); Picone et al. (2002)). However, in comparison to WACCM, this hybrid approach does not capture the interannual variability in the upper layers of the atmosphere. This can lead to poorly modeled wave propagation in the absence of stratospheric waveguides.

Because of its coarse spatial resolution, WACCM cannot resolve the full gravity wave (GW) spectrum, and the effect of these waves must be parameterized, as it is the case in most state-of-the-art numerical weather prediction systems (Gettelman et al. (2019); Alexander (2021)). However, such parameterizations represent the deposition of GW energy in the larger-scale flow without accounting for the vertical propagation of the waves (Plougonven, 2020). Notably, the wind perturbations along the GW propagation paths are not accounted for when GW is not resolved by the dynamical core. As these oscillations can have a major impact on infrasound propagation paths (Chunchuzov & Kulichkov, 2019), attention should be paid to account for their effect in propagation simulations and, in particular, across the IMS infrasound network (Listowski et al., 2024). In the current work, we opt for an idealized albeit already used method in infrasound studies (Brissaud et al. (2023); Green and Bowers (2010); Le Pichon et al. (2012)). It involves adding perturbations with realistic amplitudes obtained from a vertical GW spectrum model (Gardner et al., 1993). Figure 2 panel a) shows an example of a vertical c_{ratio} profile disturbed by 10 realizations of the GW model. Panel b) illustrates a range-dependent wind field disturbed by GWs, which can be superimposed on the c_{ratio} slices. See Appendix A for additional details on the process developed to obtain such perturbations. This approach assumes a continuum of sources and does not account for the filtering of GWs by the large-scale flow upon vertical propagation. However, it provides the needed small scales of realistic amplitudes to feed the training database with relevant atmospheric features otherwise lacking from the WACCM model. Moreover, the stochastic generation of GW through the randomly chosen phases (see Appendix A) leads to an ensemble approach. In doing so, we define an uncertainty for the atmospheric field, which is used in Section 5, where the sensitivity of the neural network to the atmospheric slices is investigated.

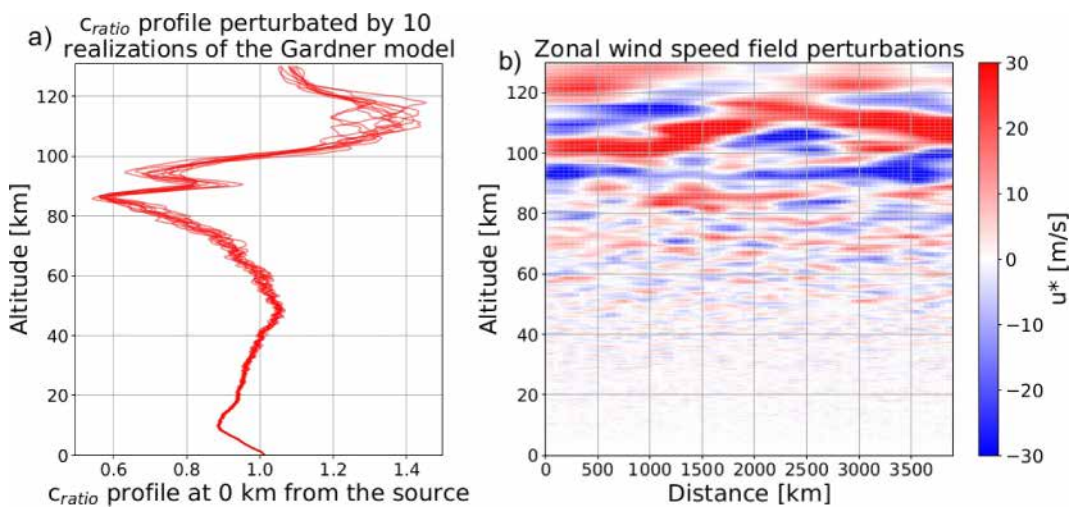

Figure 2. (a) Vertical c_{ratio} profile perturbed by 10 realizations of the GW model. (b) Two-dimensional zonal wind disturbances field induced by the GW. Such perturbations are accounted for in the atmospheric slices.

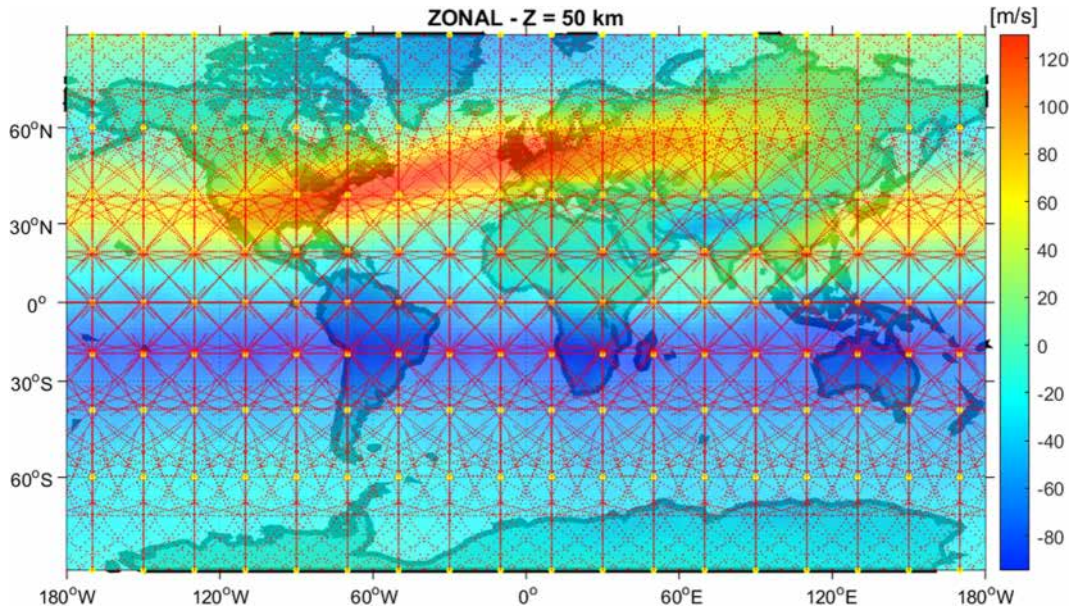

Figure 3. Distribution of 4,000-km-long atmospheric slices on the globe.

Aiming to represent a large quantity of atmospheric conditions, we sample the Earth with a set of 162 points arranged on a regular grid of 20°resolution on 15 January 2021. From each point, we collect atmospheric slices $A_{z,d}$ of c_{ratio} along eight directions: north, north-east, east, south-east, south, south-west, west, and north-west (see Figure 3). Each slice is perturbed by ten two-dimensional GW fields and is projected along two azimuths, 90° and 270°. We therefore obtain $162 \times 8 \times 10 \times 2 = 25,920$ atmospheric slices $A_{z,d}$. The choice of spatial resolution results from a trade-off between achieving global sampling and maintaining a reasonable computational cost for building the database. Propagating along eight directions and projecting onto two azimuths enable the multiplication of atmospheric conditions in a second step, at minimal additional computational cost. We chose the date of January 15 because it corresponds to a period of the year with extreme dynamical events, such as sudden stratospheric warmings (Baldwin et al., 2021). Moreover, in 2021 that day, the winds reached zonal speeds of -50 m.s^{-1} in the tropics and more than 150 m.s^{-1} at midlatitudes in the Northern Hemisphere.

2.2. Outputs: Infrasound Transmission Losses

The TL expresses the amplitude and phase variation as a wave propagates through the atmosphere. Along its path, a part of the acoustic energy loss is due to intrinsic wave attenuation mechanisms in the atmosphere, combined with the geometric spread. The TLs to be estimated are simulated using the numerical solver *ePape* from the NCPAprop package developed at the National Center for Physical Acoustics (Waxler et al., 2021). *ePape* simulates, for a single frequency and in a vertical-horizontal plane, the amplitude of the long-range pressure produced by a unit point source in relation to the level at a reference distance of 1 km from that source. The TL is obtained by calculating the modulus of this simulated pressure field and is commonly expressed in decibels:

$$\text{TL} = 20 \times \log_{10}\left(\frac{P}{P_0}\right), \tag{2}$$

where P is the pressure at a given distance and P_0 the pressure at the reference distance of 1 km.

ePape uses the atmospheric absorption coefficients defined by Sutherland and Bass (2004) and the parabolic equation method (PE) to simulate the pressure field. The PE method is well adapted to simulate wave propagation in a range-dependent medium, such as the one modeled by the atmospheric slices $A_{z,d}$. It accounts for diffraction effects and scattering by acoustic impedance variations in the atmospheric model (Waxler & Assink, 2019). The PE method assumes the atmospheric medium as a locally stratified domain and considers only signals with a frequency content large compared to the Brunt-Väisälä frequency (0.05 Hz). Inputs to the *ePape* solver are

atmospheric states represented by the zonal, meridional, and vertical winds in $m.s^{-1}$, the temperature in Kelvin, the density in $g.cm^{-3}$, and the pressure in mbar. The influence of the winds on the propagation is modeled using the c_{ratio} approximation described in Section 2.1.

Each of the 25,920 atmospheric slices $A_{z,d}$ is associated with five ground-level TLs ℓ_f simulated at frequencies $f = 0.1, 0.2, 0.4, 0.8$ and 1.6 Hz. The maximum propagation range is fixed at 4,000 km from the source. This enables the study of events on a global scale (such as the Tonga volcano eruption), as well as the evaluation of the detection capabilities of the IMS network. The expensiveness of the PE method only appears during the database creation stage. The proposed neural network will predict ground-level TL from new atmospheric scenarios almost instantaneously.

3. A Convolutional Recurrent Neural Network-Based Solution for TL Estimation

3.1. Architecture of the Model

To accurately and rapidly estimate ground-level TL, we develop a supervised neural network designed to emulate the output of the numerical solver *ePape*. We denote this network with $F_\theta(A_{z,d}, f)$. Supervised neural networks are powerful learning systems capable of approximating nonlinear functions by optimizing a set of parameters θ to minimize a loss function. They map input data, such as $(A_{z,d}, f)$, to output data, such as ℓ_f, by learning hierarchical representations that capture the complex relationships between them.

The proposed deep learning architecture relies on convolutional neural networks and recurrent neural networks (CNNs and RNNs; LeCun et al. (2015)). CNNs are well adapted for extracting spatial patterns in multidimensional data. They detect and encode local spatial correlations by applying convolutional filters that extract features. These features are progressively transformed and compressed into a latent representation of reduced spatial dimensions through pooling operations. This latent representation captures increasingly abstract and informative patterns of the input, making CNNs particularly effective for identifying spatial dependencies in the atmospheric slices $A_{z,d}$. RNNs process sequential data by maintaining a state vector, which serves as a memory of past information in the sequence. At each time step, the state vector is updated by combining the current input with the information from previous steps, allowing the network to model temporal or range dependencies. However, traditional RNNs struggle with long-range dependencies due to vanishing or exploding gradients. To address this, gate recurrent units (GRUs; Cho (2014)) are applied in the current work. GRUs employ gating mechanisms to control the flow of information, enabling the network to capture both short- and long-range dependencies efficiently. In the context of this study, GRUs are well-adapted for processing range-dependent atmospheric slices, where sequential correlations between adjacent c_{ratio} profiles play a critical role in estimating TLs.

To detect and encode relevant features in the atmospheric slices $A_{z,d}$, $F_\theta(A_{z,d}, f)$ stacks three "convolutional feature extraction blocks" (CFE Blocks), an "alignment block", a "recurrent block", and three "dense feature transformation blocks" (DFT Blocks). Each CFE block contains a two-dimensional convolutional layer, a pooling layer, and a dropout layer. The convolutional layers (in purple in Figure 4) have an increasing number of filters across the CFE Blocks (64, 128, and 256, respectively). These filters progressively extract more complex and abstract features from the input data. We incorporate nonlinearity in $F_\theta(A_{z,d}, f)$ by using the smooth and differentiable hyperbolic tangent activation function. Hyperbolic tangent transforms output values between -1 and 1, which limits the risk of gradient explosion. The max-pooling layers (in yellow in Figure 4) perform dimensionality reduction on the convolved arrays $A_{z,d}$. They limit the number of parameters to be learned and create invariance to small shifts and distortions. Finally, dropout layers (in blue in Figure 4) randomly drop 35% of the connections between neurons in each forward pass. They introduce a regularization that limits overfitting by making the network less dependent on specific neurons. At the end of the three CFE blocks, the atmospheric slices $A_{z,d}$ are encoded into latent structures of dimension (height $\times$ width $\times$ depth) $= (55 \times 5 \times 256)$, the *height* being the altitude and the *width* the distance from the source.

Each latent structure is then passed to the alignment block to be transposed and reshaped into $(5 \times (55 \times 256))$ arrays. This allows them to be exploited by two GRU layers (in green in Figure 4), which require inputs in the form of (*time steps* $\times$ *features*). The recurrent layers improve the performance of the architecture by enabling the range dependencies that exist in the encoded range-dependent atmospheric slices to be captured. The source frequency f of the TL to be predicted is added at the end of the recurrent block.

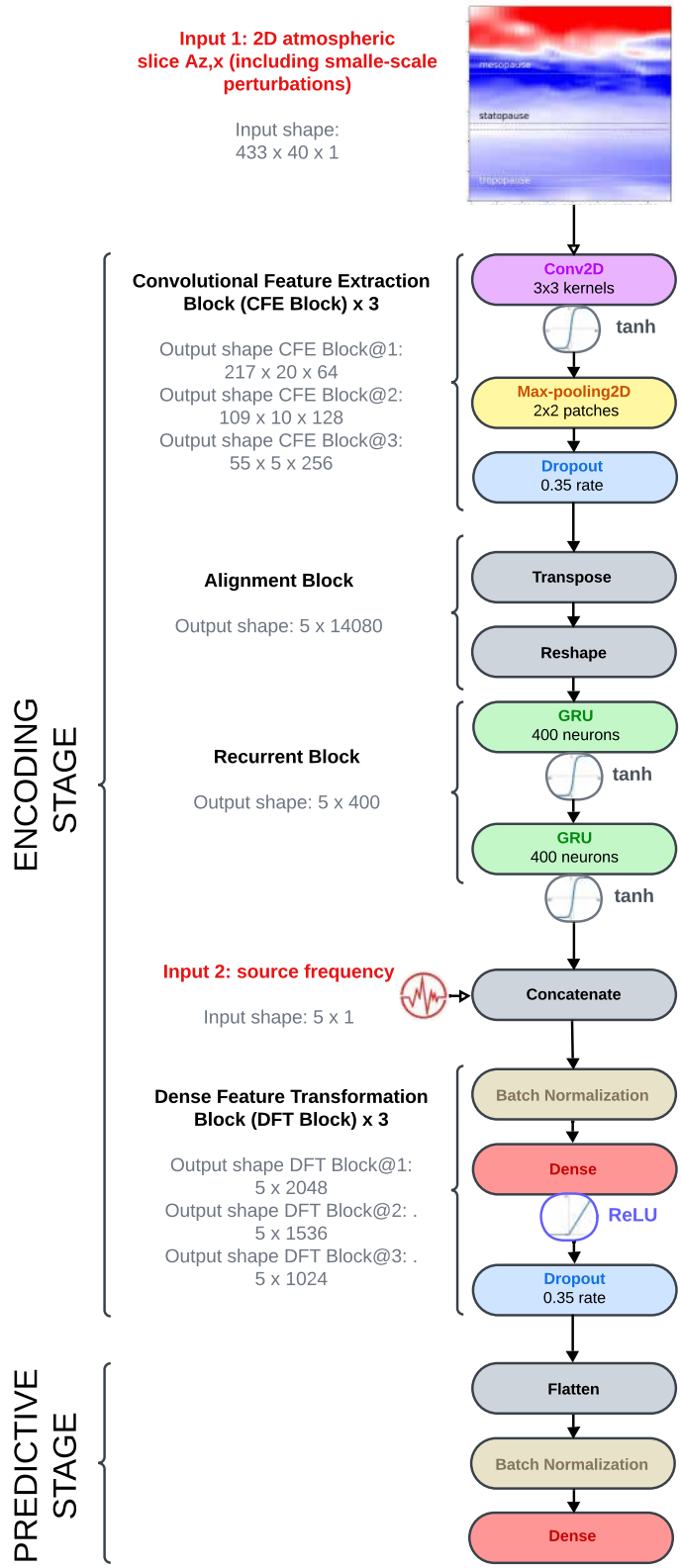

Figure 4. Architecture of $F_\theta(A_{z,d}, f)$. The atmospheric model $A_{z,d}$, including GW perturbations, and the source frequency f are first encoded by three CFE blocks, an alignment block, a recurrent block, and three DFT blocks. Then, the ground-level TL is predicted over 4,000 km.

The last transformations applied on the encoded atmospheric slices and the source frequency are performed by three DFT blocks. Each DFT block contains a batch normalization layer, a fully connected layer, and a dropout layer. Batch normalization (in brown in Figure 4) keeps the data standard deviation close to 1, which stabilizes the training and increases the robustness. The fully connected layers (in red in Figure 4) have a decreasing width across the DFT blocks, containing respectively 2048, 1,536, and 1,024 neurons. As normalization operations are performed before each fully connected layer, we use the rectified linear unit function as the activation function (ReLU; Glorot et al. (2011)) without risking gradient explosion even if it does not hold the data distribution between −1 and 1 as the hyperbolic tangent function does.

After these encoding stages, $F_\theta\left(A_{z,d},f\right)$ flattens and normalizes the data to pass them to an 800-neurons fully connected layer with the linear activation function to estimate the ground-level TL ℓ_f linearly interpolated over 800 points.

3.2. Learning Procedure

To train $F_\theta\left(A_{z,d},f\right)$, we select some of the atmospheric slices $A_{z,d}$ and associated PE simulations ℓ_f as training and validation samples. Five GW realization fields out of 10 are superimposed on each slice, keeping half of the perturbation fields unseen to evaluate the model on them later. Within the same objective, we remove 12 sampling points, all eight directions of propagation associated, among the initial 162 ones. About 70% and 20% of the remaining data (randomly selected) form the training and the validation sets. As a result, the training set contains $K_{\text{train}} = 0.7 \times [(162 - 12) \times 8 \times 5 \times 2 \times 5] = 42{,}000$ samples. We perform a cross-validation with 10 independent selections of the training and validation samples. This allows training $F_\theta\left(A_{z,d},f\right)$ 10 times and keeping the model with the best hyperparameters.

$F_\theta\left(A_{z,d},f\right)$ is implemented in Python using the TensorFlow library (Abadi et al., 2015). The training was performed on a high-performance computing cluster equipped with Nvidia A100 GPUs (40 GB memory each). On average, training the model takes approximately 23 min. See Appendix B for additional details.

After training, $F_\theta\left(A_{z,d},f\right)$ achieves near-instantaneous inference. It predicts a 4,000-km-long ground-level TL in 0.045 s regardless of the source frequency f (Dell Inc. Intel(R) Core(TM) i9-13900 48 CPUs 77.8 GB RAM on RedHat 9.5). By gathering multiple samples in the same batch, it can reduce this time to approximately 0.08 s for 100 simultaneous predictions. The estimated computation time saving compared to the PE method is three to four (for $f \geq 3$ Hz) orders of magnitude for global-scale applications requiring several thousand simulations at various frequencies (as in the assessment of detection capabilities of the IMS network).

4. Experimental Evaluations

4.1. Performances on the Testing Data Set

We evaluate $F_\theta\left(A_{z,d},f\right)$ on the 10% remaining atmospheric slices limited to five GW perturbation fields (see Section 3.2). These slices correspond to test data, that is, to atmospheric conditions extracted at the same sampling points as those used for learning, but along different directions. At this stage, 12 sampling points among the 162 ones are kept aside to use them later for the generalization. As a result, the testing set contains $K_{\text{test}} = 0.1 \times [(162 - 12) \times 8 \times 5 \times 2 \times 5] = 6{,}000$ samples.

Figure 5 shows three examples of TL for various frequencies and atmospheric conditions. Panel a) corresponds to a downwind scenario (horizontal average of c_{ratio} maxima between 30 and 60 km altitude larger than one), which favors long-range propagation of infrasound. Panel b) illustrates a typical upwind scenario with the horizontal average of c_{ratio} maxima between 30 and 60 km lower than one. In both cases, $F_\theta\left(A_{z,d},f\right)$ approximates the TL as a function of distance as well as the behavior far from the source. The model also captures quite well the first shadow zone visible in panel a), when the propagation conditions are favorable in the stratosphere. One advantage of $F_\theta\left(A_{z,d},f\right)$ is its ability to represent rapid changes in propagation regimes across the distance, leading to strong variations in the TL to be predicted along the propagation path. This ability is missing in the semianalytical attenuation model developed by Le Pichon et al. (2012). However, panel c) nuances this observation, with an example of prediction misfitting the PE simulation after 2,500 km of propagation. We observe in all cases that $F_\theta\left(A_{z,d},f\right)$ does not reproduce all small-scale spatial variations present in the PE simulations. We explain this

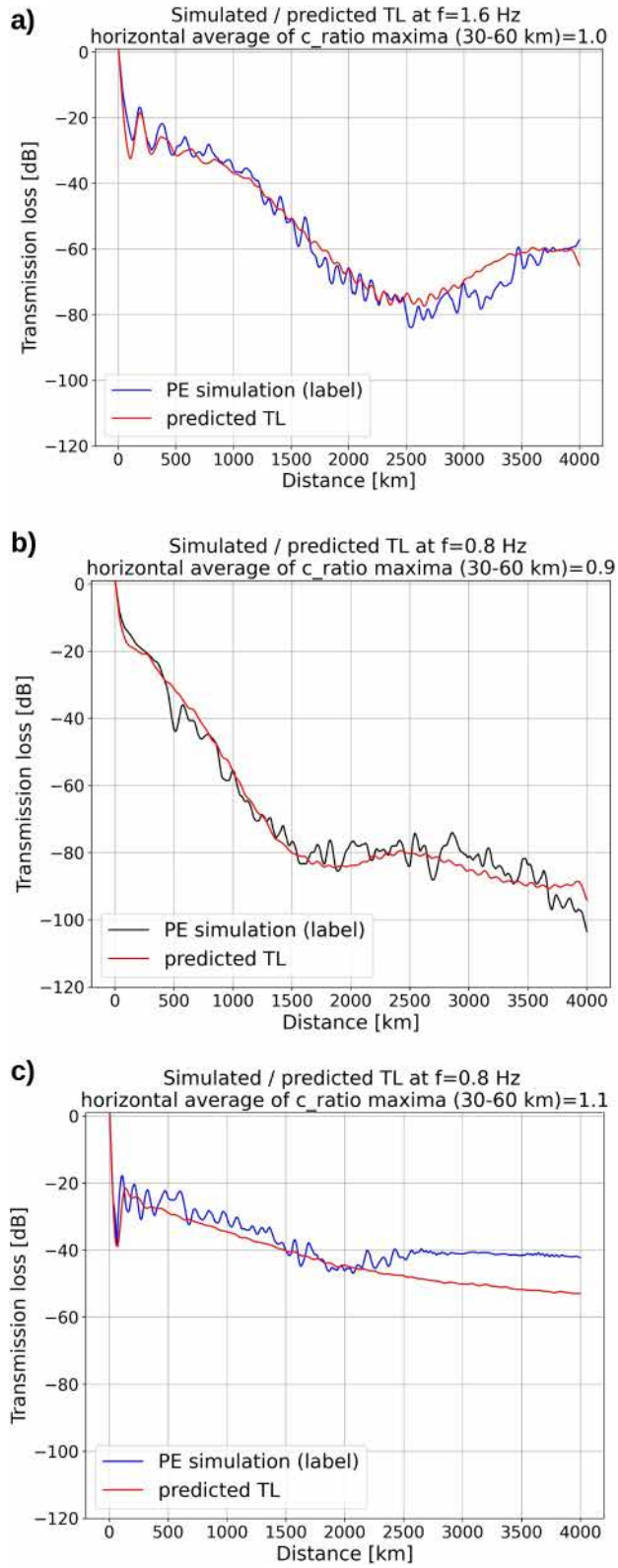

Figure 5. Examples of ground-level TLs predicted by $F_\theta(A_{z,d},f)$ compared to PE simulations. The maximum of c_{ratio} between 30 and 60 km altitude is averaged over 4,000 km. (a) Results at 1.6 Hz with $c_{\text{ratio}} = 1$ (downwind scenario). (b) Results at 0.8 Hz with $c_{\text{ratio}} = 0.9$ (upwind scenario). (c) Poorer result at 0.8 Hz.

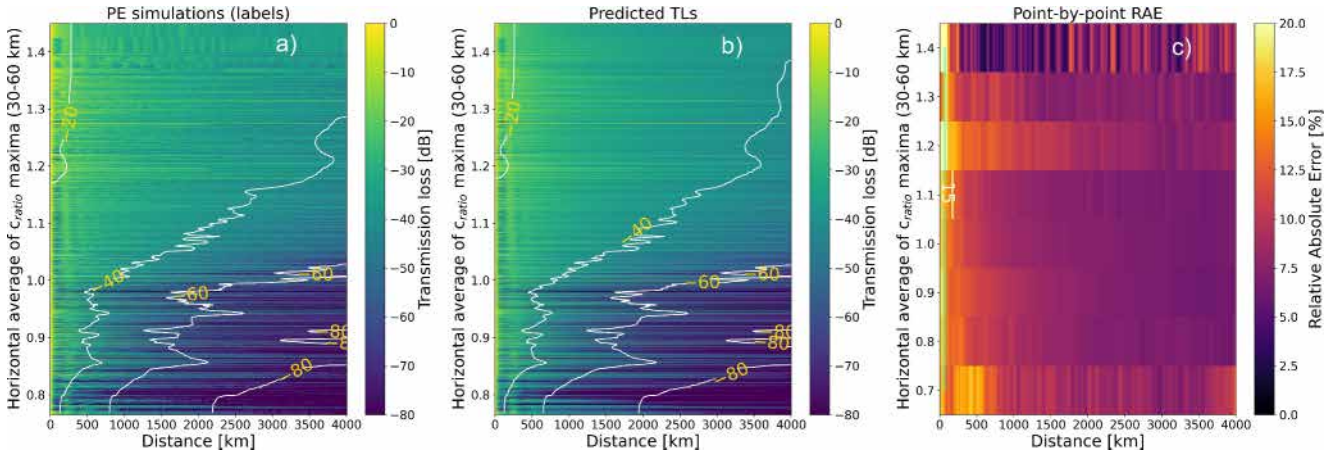

Figure 6. Comparison between 6,000 PE simulations (panel a) and predictions (panel b) associated with a pointwise error in percentage (panel c).

"low-pass filter" property by the well-known difficulty of deep learning algorithms in predicting high-frequency features (referred to as "spectral bias"; Rahaman et al. (2019)) and by the use of convolutional layers, which tend to smooth the local patterns. This bias could be mitigated either by preprocessing the input data in the frequency domain (e.g., standardizing the source frequency) or by modifying the architecture to better capture high-frequency features (e.g., by increasing the model's depth and width or by adding residual connections).

A broader evaluation of the testing data has been carried out to confirm the observation derived from Figure 5. Figure 6 compares the 6,000 predictions made by $F_\theta(A_{z,d},f)$ to the PE simulations. In the three panels of the figure, scenarios are horizontally arranged, sorted on the y-axis by the horizontal average of c_{ratio} maxima between 30 and 60 km. This representation allows investigating the ability of $F_\theta(A_{z,d},f)$ to capture the influence of the stratospheric waveguide on wave propagation (de Groot-Hedlin et al., 2009). The comparison between the PE simulations (panel a)) and the predictions (panel b)) confirms the capacity of the model to recover the overall TL while smoothing out fine-scale spatial variations.

We introduce two metrics to quantify the errors: the relative absolute error (MRAE) and the root mean square error (RMSE), both averaged along the propagation path. The MRAE captures the difference in percentage between predictions $\hat{\ell}_f$ and PE simulations ℓ_f:

$$\text{MRAE} = \frac{1}{D}\sum_{d=1}^{D}\left(\frac{|(\ell_f)_d - (\hat{\ell}_f)_d|}{|(\ell_f)_d|}\times 100\right); \quad D = 800 \quad \text{points}, \tag{3}$$

where d is the distance from the source. Such a metric penalizes errors close to the source equally just as much as those at larger ranges. This avoids overweighting errors occurring at large distances, where infrasound is often very attenuated, preventing signals to be detected above the background noise at IMS stations. The RMSE is computed to make comparisons with results published by Brissaud et al. (2023). The RMSE evaluates the difference between two TLs in decibels.

For each sample of the testing set, the pointwise relative absolute error along the propagation path shows regions with errors reaching 15% in the first 250 km (see Figure 6 panel c)). These higher error areas correspond to regions where most TL variation occurs within the first acoustic shadow zone and the first stratospheric return. The MRAE computed on the whole testing set follows a distribution peaking at an error of 6.2% (see Figure 7 panel a)). The median of the overall distribution is 7% with a 95%-percentile of 15.4%.

The distribution of the RMSE has a mean of 4.3 dB over the 4,000-km-long propagation paths. A slight degradation of the performance is noted for upwind scenarios (see panel b) of Figure 7). The predominant class of the error distribution is for downwind scenarios, with an RMSE between 2.5 and 3.5 dB, compared to 3.5–4.5 dB for the upwind scenarios. These results are consistent with Brissaud et al. (2023), which obtained an average

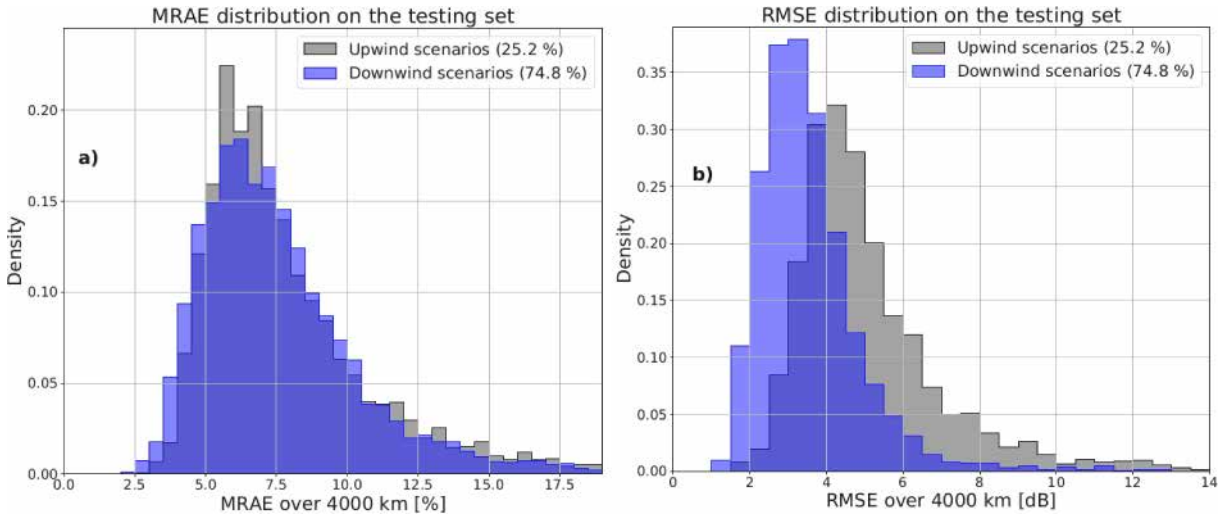

Figure 7. Distributions of the MRAE (panel a) and of the RMSE (panel b) along the 4,000-km-long propagation paths for the 6,000 samples of the testing set. A distinction is made between upwind and downwind scenarios.

RMSE of 5 dB regardless of the initial wind conditions over 1,000 km. This highlights the ability of our new model to maintain its performance against a realistic atmospheric medium covering a longer propagation range.

As in Brissaud et al. (2023), our analysis reveals a degradation of the performance with increasing frequency. Figure 8 shows the MRAE distributions obtained by $F_\theta(A_{z,d}, f)$ for the five frequencies $f = 0.1, 0.2, 0.4, 0.8$, and 1.6 Hz. The MRAE distributions restricted to the TL of higher frequencies ($f \geq 0.8$ Hz) are wider than the one associated with the TL of lower frequencies (e.g., 14.4% of 95%-percentile at 0.1 Hz and 20.3% at 1.6 Hz). We explain this by the larger sensitivity of higher frequencies to small-scale variations, which are more challenging to predict with our current data set size.

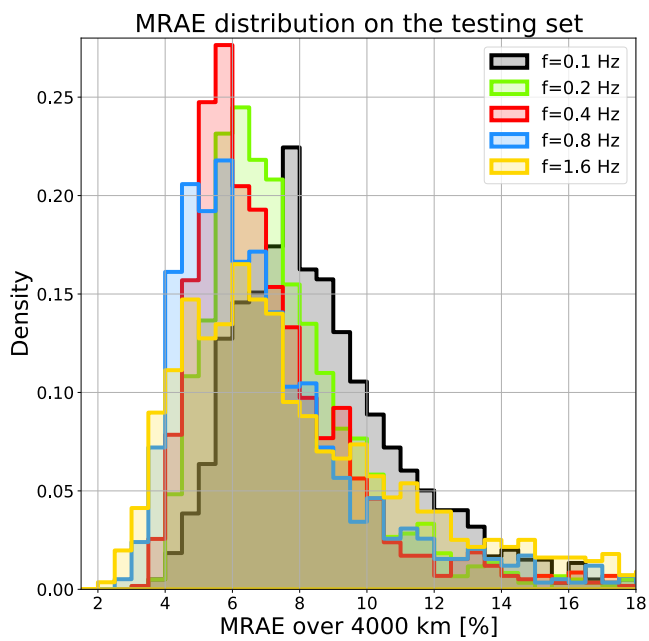

Figure 8. MRAE distributions along the 4,000-km-long propagation paths for the five source frequencies. One can see the degradation of the performance with increasing frequency.

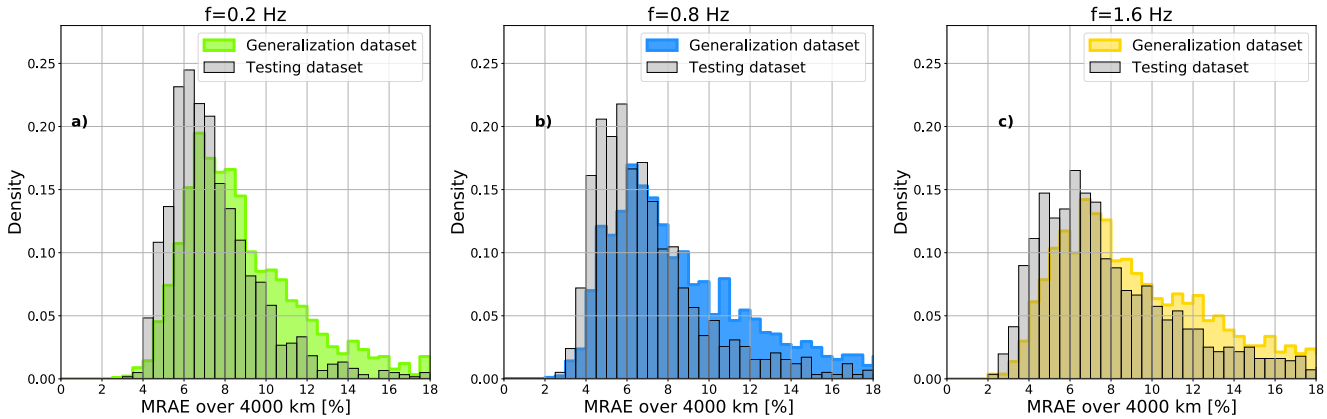

Figure 9. MRAE distributions at 0.2, 0.8, and 1.6 Hz on the testing and generalization sets.

4.2. Performance Assessment Using the Generalization Set

To further evaluate the model performance, we construct a generalization set using 12 sampling points previously left apart and keep all 10 GW perturbation fields per atmospheric slice (see Section 3.2). The generalization samples correspond to unseen locations on the Earth, sampled on the same date as the training data, with additional small-scale variation fields. As a result, the generalization set contains $K_{\text{gen}} = 12 \times 8 \times 10 \times 2 \times 5 = 9.600$ samples.

Analyzing the variability of the atmospheric conditions present in the generalization set is important for a comprehensive performance evaluation of $F_\theta(A_{z,d},f)$. The comparison of the mean c_{ratio} fields computed on the generalization and the training sets reveals differences three times larger than the highest differences reached when making the same comparison on the testing and the training sets, particularly in the thermosphere. We explain these observations by the use of a larger number of GW perturbation fields in the generalization set, whose effects are mainly visible in the upper layers of the atmosphere. See Appendix C for additional details.

As the differences between the generalization and the training sets are larger than between the testing and the training ones, we observe a slight degradation of the performance on the generalization data. We note a wider distribution of the MRAE over 4,000 km and a larger median (of approximately 8.5% on the generalization samples instead of 7% on the testing ones). Figure 9 illustrates this result by comparing the MRAE distribution on the testing and the generalization sets at 0.2, 0.8, and 1.6 Hz. A similar shift in the generalization distributions is noted for all source frequencies.

5. Epistemic Uncertainty and Data Sensitivity

Quantifying the uncertainties is crucial for many applications, such as monitoring explosive sources using the IMS. In this section, we introduce two ways of quantifying the uncertainties associated with the predictions of $F_\theta(A_{z,d},f)$ by distinguishing epistemic uncertainty from data uncertainty.

A way to quantify the epistemic uncertainty (i.e., related to the model's architecture) is to use Bayesian methods. A Bayesian neural network $G_\phi(I)$ (Tishby and Solla (1989); Buntine (1991)) can predict a new sample (I^*,O^*) that incorporates epistemic uncertainty by considering its parameters ϕ as random variables following a posterior distribution $p(\phi,(I,O))$ given observed samples (I,O). The posterior distribution is based on prior belief on ϕ and on the use of Bayes' theorem (Gawlikowski et al., 2023). The predictive posterior distribution for a new input I^* is obtained by marginalizing over ϕ, which implies integrating over all possible values of ϕ weighted by their posterior probabilities:

$$p(O^*|I^*,(I,O)) = \int p(O^*|I^*,\phi) \times p(\phi,(I,O))d\phi \tag{4}$$

However, calculating the posterior distribution $p(\phi,(I,O))$ and performing marginalization can be intractable for complex models such as neural networks. In this work, we use the Monte Carlo dropout method (Gal & Ghahramani, 2015) to approximate $p(O^*|I^*,(I,O))$ and indirectly the posterior distribution. I corresponds to an input sample $(A_{z,d},f)$ and O to a PE simulation ℓ_f. We modify the architecture of $F_\theta(A_{z,d},f)$ so as to define its dropout layers as a set of random variables following a Bernoulli distribution. The units of these modified layers can be activated not only during training but also during a prediction. At the inference stage, the model is then no more deterministic but stochastic. The distribution associated with each new input $(A_{z,d},f)^*$ is obtained by averaging the predictions of the set formed by such Bayesian networks. This method allows us to provide a distribution of predictions for any new input while being inexpensive in terms of training time.

The second source of uncertainty reflects the inherent noise of the data that feed the neural network. We use the test-time augmentation method to quantify the sensitivity of $F_\theta(A_{z,d},f)$ to the variations of its input data. Several versions of each input $(A_{z,d},f)$ are generated by applying augmentation techniques and making predictions for each of them, which can then be averaged to obtain an output distribution. This method has the advantage of not being resource-intensive because the architecture is not modified and does not require additional training samples. The augmentation applied on each input implies superimposing 10 realizations of the GW perturbation fields on the atmospheric slices $A_{z,d}$. Therefore, the simulated and predicted TLs are the result of the average of 10 simulations/predictions originating from the same slice.

We calculate the uncertainties on the generalization set. Figure 10 shows an example of TLs associated with a sampling point of the generalization set, with all eight directions of propagation considered. The predicted standard deviation (in orange) is estimated by considering both epistemic and data uncertainties and is compared to the data uncertainty obtained on the PE simulations (in blue). One can see larger uncertainties (up to 3 dB) for propagation directions associated with upwind atmospheric conditions (see panels c) and d)). Nevertheless, even in the directions where the uncertainty is the largest, it does not recover all variations of the PE simulations. This is explained by the aforementioned disability of $F_\theta(A_{z,d},f)$ to capture all TL variations induced by the GW perturbations. Panel i) represents the pointwise relative absolute error averaged along the eight directions. This confirms the increase of error in the first 250 km, reaching 15.6%. See Appendix D for additional details.

6. Tonga Case Study: Example of Generalization in Time, Space, and Frequency

A direct application of our model is the near real-time prediction of TL maps around a source of interest. To emulate this use case, we choose the Tonga volcano as the source, which erupted on 15 January 2022. This event is widely documented (Matoza et al. (2022); Podglajen et al. (2022); Vergoz et al. (2022)). The eruption generated atmospheric waves that were detected by all the operational IMS infrasound stations.

We collect 360 atmospheric slices around the Tonga volcano to form a new generalization set: the *Tonga-set*. Like the previous generalization set presented in Section 4.2, these samples correspond to unseen locations on the Earth. In addition, the date of sampling is different. The mean c_{ratio} fields computed on the training set and the *Tonga-set* reveal regions with a difference of up to 0.1 of c_{ratio} above the mesopause (see Figure 11). This is six times larger than the highest difference reached when comparing the training to the previous generalization set. The distributions of minimal, average, and maximal values of c_{ratio} in the troposphere, stratosphere, mesosphere, and thermosphere give more details on these differences (see Appendix E). As an illustration, the training set contains slices with a minimal c_{ratio} between 0.65 and 0.73 in the troposphere and between 0.4 and 0.6 in the stratosphere, which are situations totally absent in the *Tonga-set*. On the contrary, the proportion of slices with a minimal c_{ratio} between 0.7 and 0.8 in the stratosphere or with a mean c_{ratio} between 1.0 and 1.15 in the thermosphere is overrepresented in the *Tonga-set*. Similarly, the *Tonga-set* overrepresents slices with a c_{ratio} maximum of 1.6 above 90 km altitude. Such differences impact the generalization capabilities of $F_\theta(A_{z,d},f)$, which depends on the similarities between the training and the evaluation sets.

At 01:00:00 UTC on 15 January 2022, atmospheric models predict regions with both downwind and upwind propagation conditions. Stratospheric guiding is visible westward (200°–315° azimuth) with mean c_{ratio} reaching 1.2 at 50 km altitude. On the contrary, for eastward propagation, the mean c_{ratio} reaches 0.6 in the stratosphere. At lower altitudes, we observe the existence of a weak tropospheric waveguide at the top of the boundary layer in all directions. In addition, mesospheric waveguides are predicted for the direction 135°.

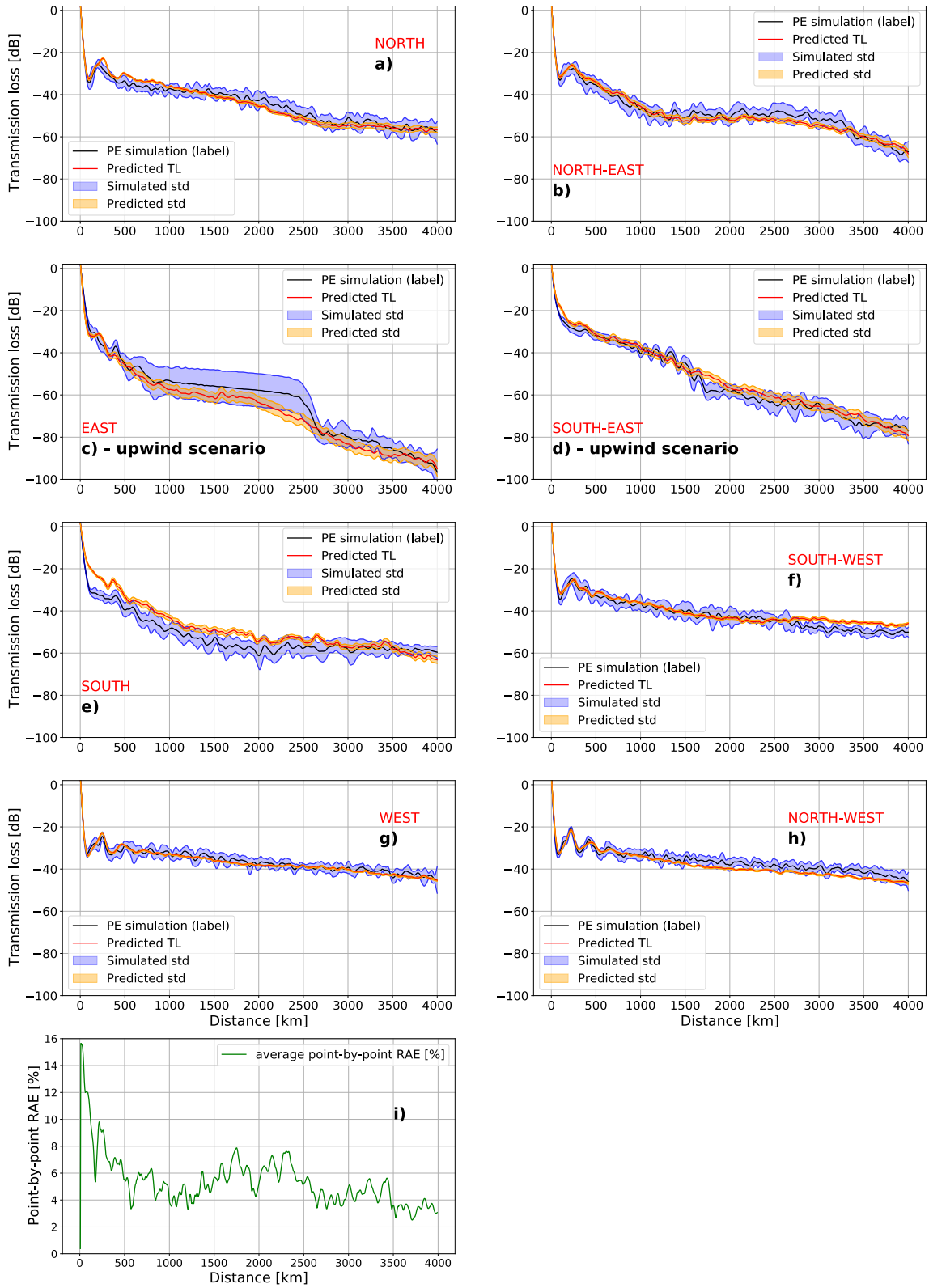

Figure 10. TLs associated with a sample of the generalization set at $f = 0.2$ Hz, all directions of propagation considered. The predicted standard deviation (in orange) is computed by combining the epistemic and the data uncertainties. It is compared to the simulated data uncertainty obtained on the PEs (in blue).

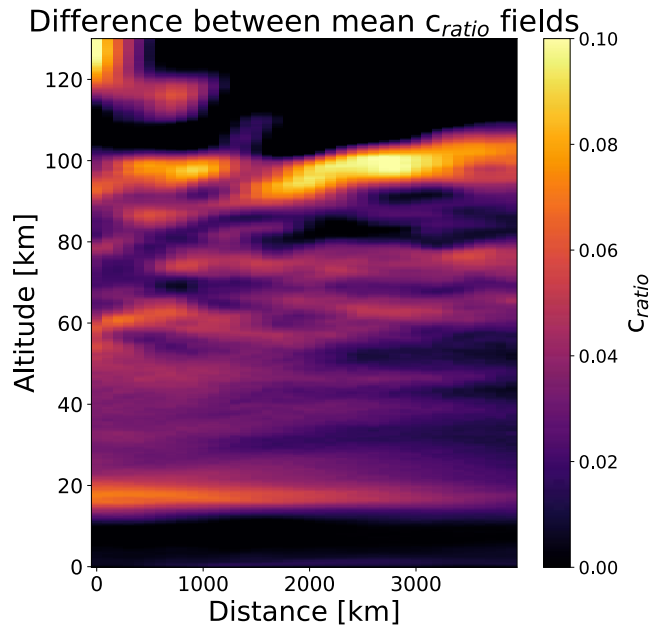

Figure 11. Differences between the mean c_{ratio} fields computed on all the samples of the training set and the *Tonga-set*.

At the inference stage and with a batch of 360 samples, $F_\theta(A_{z,d},f)$ predicts the 4,000-km-long ground-level TLs all around the volcano in less than 0.3 s, regardless of the frequency (Dell Inc. Intel(R) Core(TM) i9-13900 48 CPUs 77.8 GB RAM on RedHat 9.5). Despite the differences observed between the training set and the *Tonga-set*, the median MRAE computed on this last set is 7.9% only. This illustrates the ability of $F_\theta(A_{z,d},f)$ to accurately predict TL maps for an event absent of its training and validation sets. Figure 12 shows simulated and predicted TL maps (panels a) and b)) as well as the pointwise error (panel c)) at $f = 0.8$ Hz. The TL reaches -125 dB eastward, where infrasound waves are not refracted to the ground in the stratosphere but propagate into the thermosphere (upwind scenarios). This direction is associated with a global increase of the errors, particularly along 135°. These areas with higher errors are explained by the aforementioned differences observed above the mesopause when comparing the training set and the *Tonga-set*.

Appendix F shows the distribution of the MRAE as a function of frequency and mean c_{ratio}. This confirms poorer performances in upwind cases and with increasing frequency. At 1.6 Hz, the 95%-percentile of the MRAE rises up to 26.4%, which is by far the highest error we observe. Such sensitivity to downwind/upwind conditions is larger than what was observed previously for the testing set (see Section 4.1). This is explained by the lack of representation of slices with a minimal stratospheric c_{ratio} between 0.7 and 0.8 in the training set compared to the *Tonga-set*, preventing $F_\theta(A_{z,d},f)$ from correctly learning how to predict the TL in that situation. In the absence of stratospheric waveguides (upwind), infrasound waves reach the mesopause, where the training set and the *Tonga-set* differ significantly.

For the 360 predictions, the epistemic and the data uncertainties are quantified using a combination of the Monte Carlo dropout method with the test-time augmentation technique. The total uncertainty of a given prediction corresponds to the standard deviation obtained from an initial atmospheric slice disturbed by 10 GW perturbation fields and whose associated TLs are simulated/predicted 10 times by the set of Monte Carlo dropout $F_\theta(A_{z,d},f)$s. The total uncertainty increases with increasing frequency and for upwind conditions. Figure 13 shows the total uncertainty map obtained by $F_\theta(A_{z,d},f)$ at $f = 1.6$ Hz. The standard deviation reaches more than 10 dB eastward, where the 95%-percentile of the MRAE rises up to 26.4%. This highlights again that high uncertainties are estimated in regions of large errors.

To further deepen the generalization capabilities of $F_\theta(A_{z,d},f)$, we consider unseen source frequencies f: 0.3, 0.6, 1.0, 1.2, and 1.4 Hz. Appendix F displays the distributions of the MRAE according to these frequencies and to the mean c_{ratio}. The strong correlation between the error distributions obtained on the set of new frequencies

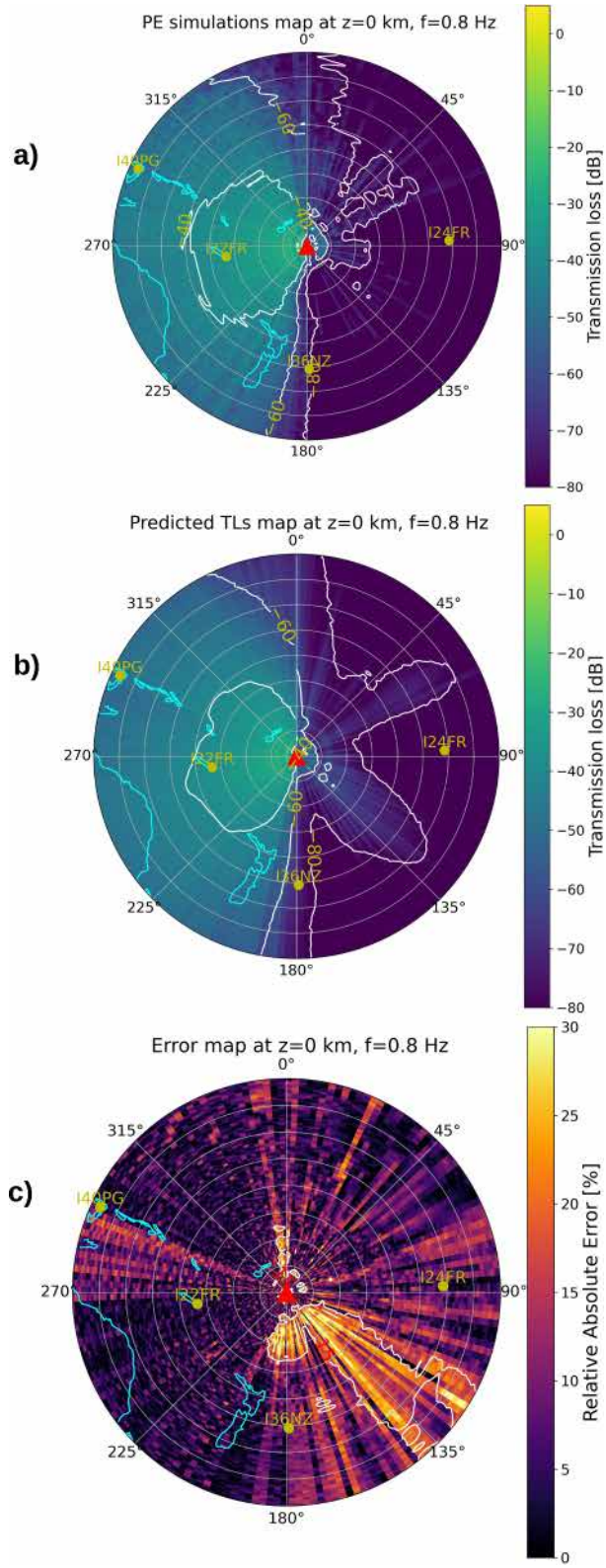

Figure 12. Simulated and predicted TLss around the Tonga volcano (panels a and b), and pointwise error in percentage (panel c) at $f = 0.8$ Hz. The IMS stations located within a 4,000-km radius of the volcano are marked with dots (IS22, IS24, IS36, and IS40).

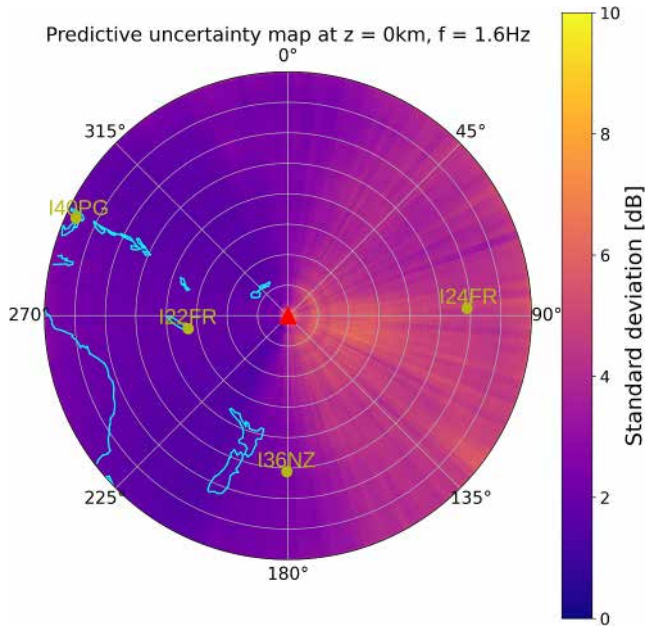

Figure 13. Epistemic and data uncertainty map at 1.6 Hz.

compared with the original distributions is highlighted. For each distribution, the errors remain in the same order of magnitude. Higher errors (between 20% and 30%) are spread on the same range of mean c_{ratio} values (mainly for upwind cases). However, the increase of error with increasing frequency is visible as early as f becomes larger than or equal to 0.6 Hz. The *Tonga-set* study thus shows the overall ability of $F_\theta(A_{z,d},f)$ to generalize to new atmospheric model regimes, new dates, and new source frequencies.

7. Conclusion

We have shown the ability of a deep learning algorithm $F_\theta(A_{z,d},f)$ to estimate almost instantaneously infrasound ground-level TLs over a distance of 4,000 km, at $f \in [0.1, 0.2, 0.4, 0.8, 1.6]$ Hz. $F_\theta(A_{z,d},f)$ is a convolutional recurrent neural network inspired by the CNN developed in Brissaud et al. (2023). Our addition of recurrent layers after the convolution layers allows it to extract both spatially local and range-dependent features embedded in realistic atmospheric models. However, as in Brissaud et al. (2023), we cannot assert that $F_\theta(A_{z,d},f)$ respects the inherent range-dependence of the atmospheric models. Indeed, despite the use of recurrent layers, the filters in the convolutional layer use local information before and after each given distance to encode spatial features.

The Sutherland-Bass coefficients used in *ePape* allow the inclusion of the absorption of acoustic energy by the atmosphere. However, these coefficients do not account for second-order effects due to changes in viscosity and specific heat ratio with deviations in the atmospheric composition above 90 km altitude (Sutherland and Bass (2004)). Moreover, it is important to note that the used PE method cannot model nonlinear propagation effects, which become more prominent above the mesopause. To limit the computational cost, the propagation problem was simplified in Cartesian coordinates, assuming a flat terrain with infinite ground impedance at sea level, and by preselecting only five frequencies. Given these limitations, which may introduce bias in the simulations, numerical explorations with fully resolved time- and range-dependent wave propagation techniques accounting for nonlinear propagation effects would provide more realistic simulation results.

$F_\theta(A_{z,d},f)$ was trained and validated on 42,000 samples $(A_{z,d},f)$. Once done, it estimates a TL in approximately 0.045 s, regardless of the frequency. This constitutes a major improvement compared to other numerical propagation modeling tools such as PE solvers, with a computation time saving between three and four orders of magnitude. The performance of the network was evaluated on 6,000 testing samples. A median MRAE of 7% and a mean RMSE of 4.3 dB were obtained. The robustness of the model to the atmospheric conditions was

highlighted, as well as its "low-pass filter" property. A slight decrease in performance with increasing frequency has been noticed.

$F_\theta(A_{z,d},f)$ was then evaluated on atmospheric conditions that differ more from the training one. Generalization data correspond to atmospheric slices built from previously randomly drawn points on the Earth. As expected, the model reached poorer performances (median MRAE 8.5%). The generalization data were also used to estimate epistemic and data-related uncertainties. This revealed an increase in the uncertainties with increasing error, particularly for upwind atmospheric scenarios. The combination of the two uncertainties partly recovers the whole variability of the simulated TL considering multiple GW perturbation fields. Still, they allow for the qualitative incorporation of a part of the uncertainties associated with the atmospheric models used as input data (Listowski et al., 2024). There are other techniques that could be explored to quantify the uncertainties in more detail, such as deep ensemble approaches (Lakshminarayanan et al., 2017).

We used $F_\theta(A_{z,d},f)$ to predict near real-time TL maps for infrasound generated by the eruption of the Tonga volcano. The slices extracted around the volcano represented unseen atmospheric conditions in terms of location and date. The median MRAE value was 7.9%. We demonstrated the ability of the model to predict the TL maps at five new source frequencies.

This whole work represents a first step toward the near real-time assessment of the global detection capabilities of the IMS infrasound network. Accurate TL estimations are essential for assessing the minimum detectable explosive energy (e.g., Le Pichon et al. (2009); Blom et al. (2018)). The detection capability of the IMS network is also strongly affected by the time-varying station noise level above which signals can be reliably detected (e.g., Marty et al. (2021)). Computing global detectability maps at an hourly temporal resolution requires large numbers of TL estimations across multiple frequencies. While the computational costs of PE solvers are prohibitive, this remains feasible using empirical attenuation relations (Green and Bowers (2010); Le Pichon et al. (2009); Le Pichon et al. (2012)) or our neural network. However, these challenges are amplified when examining statistical approaches through the exploration of ensembles of initial atmospheric conditions (e.g., Hedlin and Drob (2014); Vanderbecken et al. (2020)). Since conventional numerical propagation methods are not yet capable of exploring a vast parameter space for global and real-time operational applications, cost-effective machine learning-based models present a promising approach to reduce computational burden while maintaining high prediction accuracy.

Future works include the specialization of the neural network on a regional scale using fine-tuning. An expected outcome is an increase in performance when predicting TL on global and regional reference events (e.g., the Tonga volcano eruption, the Lebanese explosion in Beirut–Pilger et al. (2021), the Finnish explosions Hukkakero–Vorobeva et al. (2023), and the Negev desert controlled explosion–Fee et al. (2013)). Increasing the amount of input data could further enhance the neural network performance. Appendix G shows promising preliminary results, obtained by doubling the size of the training database by sampling the Earth's atmosphere both on 15 January 2021, and 15 August 2021. Conducting explainability studies of the predictions could also improve the neural network performance. Such studies highlight useful characteristics in the input data for understanding the model's decision-making process and know how to build relevant training databases. In the context of data assimilation, such neural architectures also offer valuable insights for developing adjoint methods (Letournel et al., 2024). This application takes direct advantage of the nature of the neural networks, since numerical propagation methods such as PE solvers do not allow the determination of all the partial derivatives needed to calculate the adjoint without being computationally expensive. Finally, a topic of interest is the modeling of two-dimensional outputs, allowing TL predictions at receivers in altitude. The predicted two-dimensional TL could be exploited for further atmospheric investigations using stratospheric balloon observations of large explosive sources (Albert et al. (2023); Bowman and Krishnamoorthy (2021); Podglajen et al. (2022); Silber et al. (2023)).

Appendix A: Construction of Two-Dimensional Range-Dependent GW Perturbation Fields

The vertical GW spectrum as a function of the vertical wavenumber m essentially consists of two regimes. The first part, $m < m^*$, is the nonsaturated part driven by the GW source, with an increase until the maximum of spectral energy at $m = m^*$, where m^* is called the dominant wavenumber. At larger wavenumbers, $m > m^*$, this

is the saturated part of the spectrum (Smith et al., 1987) with a characteristic −3 slope in the logarithmic scale (see Figure 1 of Gardner et al. (1993)). Typical values for m^* are a few to about 10 km as documented in Gardner et al. (1993) and in the literature (Allen and Vincent (1995); Hostetler and Gardner (1994); Chu (2018)).

The method used throughout this study consists in using the analytical expression of the vertical GW spectrum presented in Gardner et al. (1993) with the vertical profile of m^* to build vertical profiles of perturbations for a given set of 10-km thick overlapping atmospheric layers (50% overlap). The analytical expression of the vertical GW spectrum is given by Gardner et al. (1993) (their Equation 7):

$$F_u(m) = \begin{cases} 2\pi \dfrac{\alpha N^2}{m_*^3}\left(\dfrac{m}{m_*}\right)^S & m \le m_* \\ 2\pi \dfrac{\alpha N^2}{m^3} & m_* \le m \le m_b \\ 2\pi \dfrac{\alpha N^2}{m_b^3}\left(\dfrac{m_b}{m}\right)^{\frac{5}{3}} & m_b \le m; \end{cases} \tag{A1}$$

where the nonsaturated regime ($m \le m_*$), the saturated regime ($m_* \le m \le m_b$), and the turbulence regime ($m_b \le m$) are represented. The term m_b is defined as the buoyancy wavenumber identifying the transition between saturated GW and turbulence, and N as the buoyancy frequency reflecting the vertical stability of the atmosphere. The term α is a constant below 1 accounting for superposition effects when multiple waves interact, generating instabilities that lower the threshold of wave saturation (Gardner et al., 1993).

The vertical profile of m^* is given by Gardner et al. (1993) (their Equation 43):

$$m_* \sim \exp\left[-\frac{z}{(q+s)H}\right]; \tag{A2}$$

where z is the altitude, H the atmospheric scale height, and s and $-q$ the spectral indices of the m-spectrum in the source regime and the saturation regime, which are key parameters that influence both the shape and the magnitude of the horizontal wavenumber spectrum.

For each of the vertical GW spectra that are modeled for a set of altitudes z, the inverse Fourier transform is derived given a randomly chosen phase. This provides vertical perturbation profiles, which respectively characterize each of the atmospheric layers centered on each z. Then, all these vertical profiles are weighted and linearly combined to produce a full perturbation profile for the whole atmosphere, retaining perturbation amplitudes characteristic of each altitude. The Gaussian weights used in the combination are centered at midlayers and normalized to ensure the total weight addressed to a given atmospheric layer is 100%, thus avoiding an excess of redistributed energy in any of the layers. This method is applied for a given number of randomly chosen phases, producing an ensemble of several realizations of GW perturbation profiles as shown in Figure 2 panel a).

In the last stage, randomly chosen vertical GW profiles (among those derived above) are horizontally and linearly combined using the horizontal correlation length documented in Gardner et al. (1993) (their Section 4 and Table 4). The final product is a two-dimensional range-dependent GW perturbation field as shown in Figure 2 panel b). Such fields are superimposed on the c_{ratio} slices.

Appendix B: Details on the Learning Process

To optimize memory, we interpolate all atmospheric slices $A_{z,d}$ and ground-level TLs. The atmospheric slices are interpolated to a regular grid (height × width × depth) = (433 × 40 × 1). The *height* corresponds to the altitude, with $z \in [0, 129.9]$, and the *width* corresponds to the distance from the source, with $d \in [0, 3900]$, keeping only a vertical c_{ratio} profile every 100 km. The resulting 0.3 km step in altitude is sufficient to preserve the effects induced on the atmospheric slices by the small-scale GW perturbations. Indeed, as stated in Appendix A, the GW vertical spectra upon which the perturbations are built peak at a critical wavenumber m^* corresponding to a few kilometers in wavelength. The exponential decrease of energy at smaller and larger wavenumbers, respectively, ensures that the relevant part of the GW spectrum will be well accounted for at each altitude with a

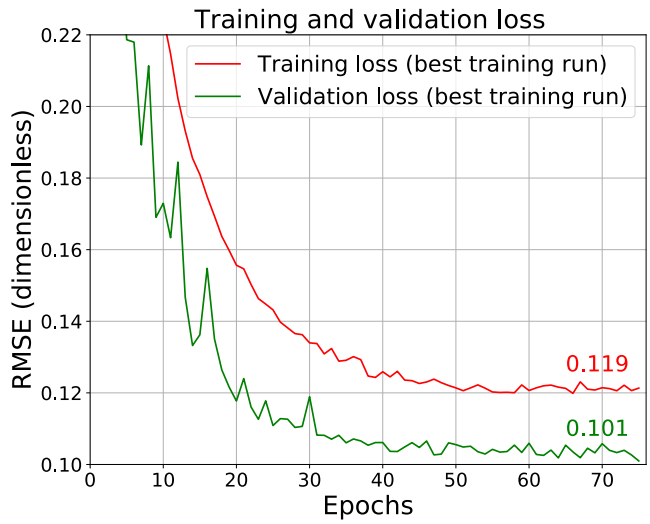

Figure B1. Training and validation loss of $F_\theta(A_{z,d},f)$ during the best run.

0.3-km vertical step. The ground-level TL is interpolated with a regular step of 5 km, leading to a ℓ_f vector of 800 points. This allows unifying the TL calculated using the *ePape* solver, which has different spatial resolutions according to the source frequency f.

Both the atmospheric slices $A_{z,d}$ and ground-level TLs ℓ_f are standardized before feeding $F_\theta(A_{z,d},f)$. The standardization involves removing the mean and scaling to unit variance. This preprocessing step allows us to speed up learning and minimize the risks of explosion/disappearance of the gradient when dealing with an important number of parameters ($F_\theta(A_{z,d},f)$ contains approximately 27 million parameters).

Following the work of Brissaud et al. (2023), the difference between the predictions $\hat{\ell}_f$ and the PE simulations ℓ_f is quantified by the root mean square error (RMSE) loss function. This metric is widely used in regression problems. Since the atmospheric slices $A_{z,d}$ and the expected outputs ℓ_f are standardized, the RMSE is dimensionless. The training loss of the best run of $F_\theta(A_{z,d},f)$ (see Section 3.2) converges to 0.119 after 76 iterations, as shown by the green curve in Figure B1.

Training is conducted over a maximum of 150 epochs, with an early stopping if the loss function on validation data remains stable beyond 20 iterations. To improve the convergence, the parameters θ are initialized in the first convolutional layer using the Glorot initializer (Glorot & Bengio, 2010), and the input data $(A_{z,d},f)$ are divided into batches of size 32. The Adam optimizer (Kingma, 2014) is used with an initial learning rate $\eta = 1 \times 10^{-4}$, which is reduced by a factor of 10 after the tenth epoch if no improvement in validation loss is observed.

Appendix C: Details on the Comparisons Between the Training and the Generalization Sets

This appendix aims at detailing the comparison realized between the training and the generalization sets presented in Section 4.2. Each panel of Figure C1 focuses on a specific atmospheric layer. The altitudes of each layer are defined as presented in Section 2, with the troposphere being defined between 0 and 12 km, the stratosphere between 12 and 60 km, the mesosphere between 60 and 90 km, and the thermosphere between 90 and 130 km altitude. For each layer, the minimal, average, and maximal c_{ratio} distributions are plotted for both the training and the generalization sets. As an example, the minimal distribution is obtained by distributing all the minimal values of c_{ratio} among 20 bins of c_{ratio} ranging from 0.1 to 2.0 with a step of 0.1. As the training and the generalization sets contain different numbers of scenarios, all results are plotted as a density.

For each layer, we observe small departures of the distributions of the generalization atmospheric slices $A_{z,d}$ compared to the ones obtained on the training set. These departures can take the form of some situations slightly overrepresented in the generalization set, like in the mesosphere with a minimal c_{ratio} of 0.6 or a maximal c_{ratio} of

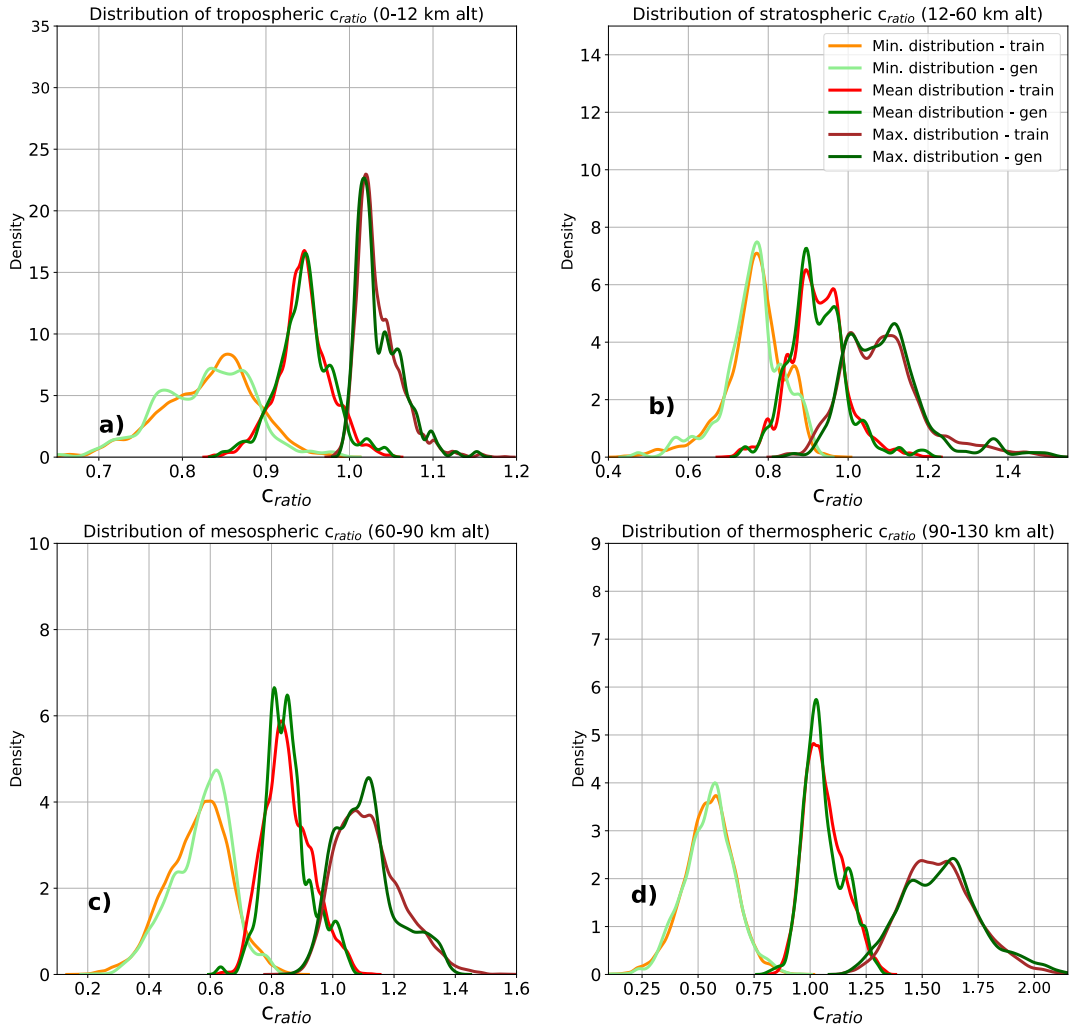

Figure C1. Distributions of minimal, average, and maximal values of c_{ratio} in the troposphere, stratosphere, mesosphere, and thermosphere computed on the training and the generalization sets.

1.15. We explain this by the differences induced in the way we extract the generalization slices compared to the training ones. Generalization samples correspond to 12 sampling points randomly set apart from the 162 initial ones, associated with 10 GW perturbation fields. Generalization data correspond thus to unseen locations on the Earth sampled on the same date as the training ones, with additional small-scale variations. However, it is important to notice that the variability induced by the GW perturbation fields remains minor and is mainly visible above the stratosphere (see Figure 2 panel b)). Indeed, the minimal and maximal zonal wind speed disturbances induced by these fields are on average $-25/25$ m.s^{-1}.

Appendix D: Epistemic and Data-Related Uncertainties on the Generalization Set

This appendix is a complement to Section 5 and provides details on the epistemic and data-related uncertainties obtained on the generalization set.

The epistemic uncertainty associated with all the predictions of the generalization set is relatively small, with an average of 1.98 dB for all source frequencies. The largest standard deviations (between 3 and 5 dB) are reached for upwind atmospheric scenarios with $f \geq 0.8$ Hz. Figure D1 highlights the increase of epistemic uncertainty with increasing frequency, with an average of 1.6 dB for $f \leq 0.4$ Hz up to 2.18 dB at 0.8 Hz and 2.8 dB at 1.6 Hz. We also observe an increase of epistemic uncertainty in the first 250 km from the source, particularly at higher

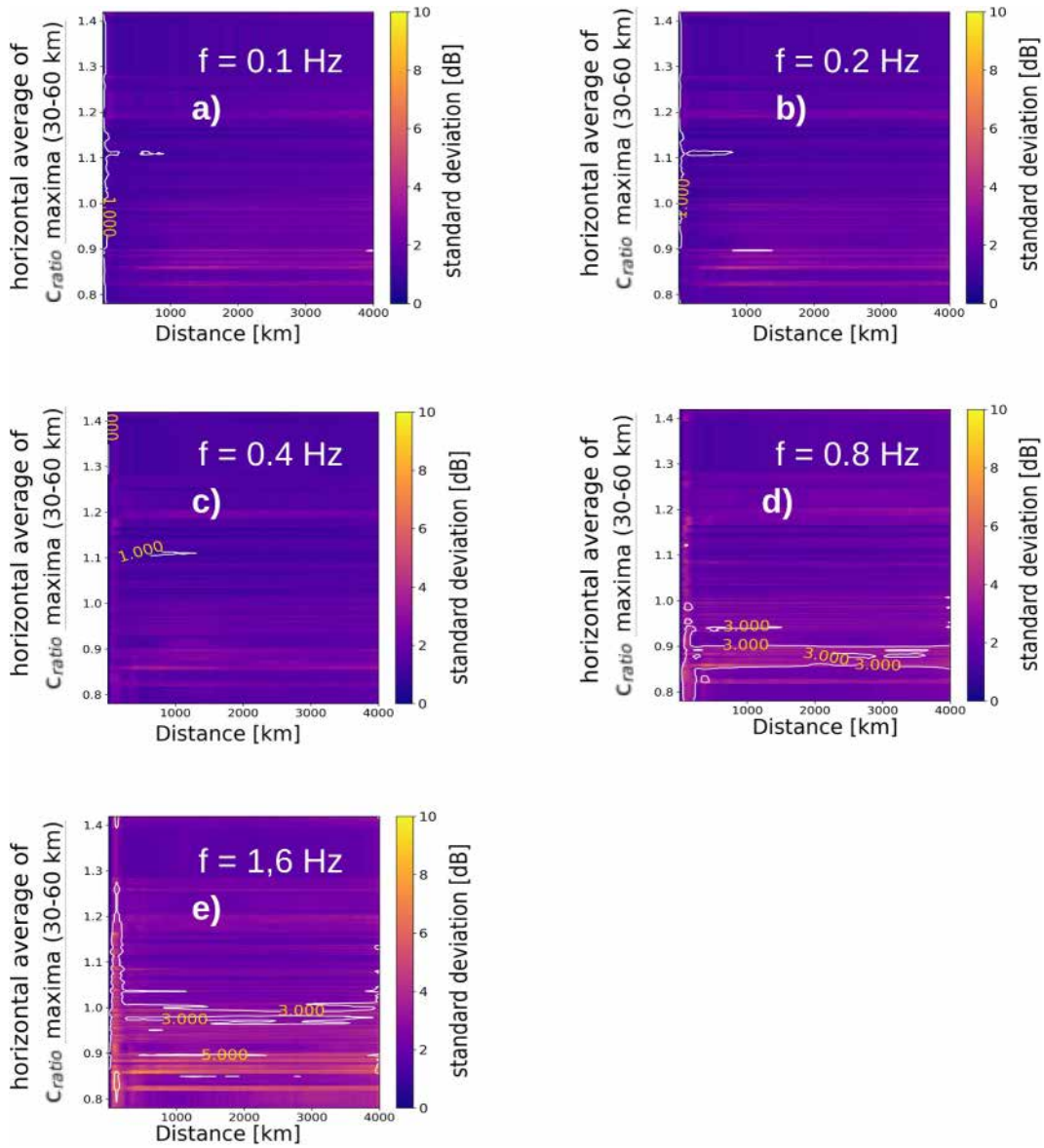

Figure D1. Uncertainty related to the model architecture, quantified on the generalization set using the Monte Carlo dropout method.

frequencies. All of this demonstrates the correlation existing between the higher uncertainty estimates and the areas with higher predictive errors (see Section 4.1).

We performed various numbers of inferences to estimate the epistemic uncertainty using the Monte Carlo dropout method (10 as presented in Section 5, 50, 100, and 150). Each time, the average epistemic uncertainty remains below 2 dB across all source frequencies. This may seem low but is quite expected since the Monte Carlo dropout method relies on the assumption that dropout layers introduce sufficient stochasticity during forward passes to effectively sample the model's parameters. However, this assumption may lead to a less accurate approximation of the Bayesian inference and resulting epistemic uncertainty. Such limitation can thus provide an additional explanation for the inability of $F_\theta(A_{z,d}, f)$ to recover all variations of the PE simulations, as observed in Section 5.

Regarding data-related uncertainty, the standard deviation estimated on the generalization predictions is on average equal to 1.16 dB along 4,000 km, against 3.22 dB on the PE simulations. This indicates that the uncertainty induced on the TL by perturbing the initial c_{ratio} slices with 10 GW perturbation fields is not fully

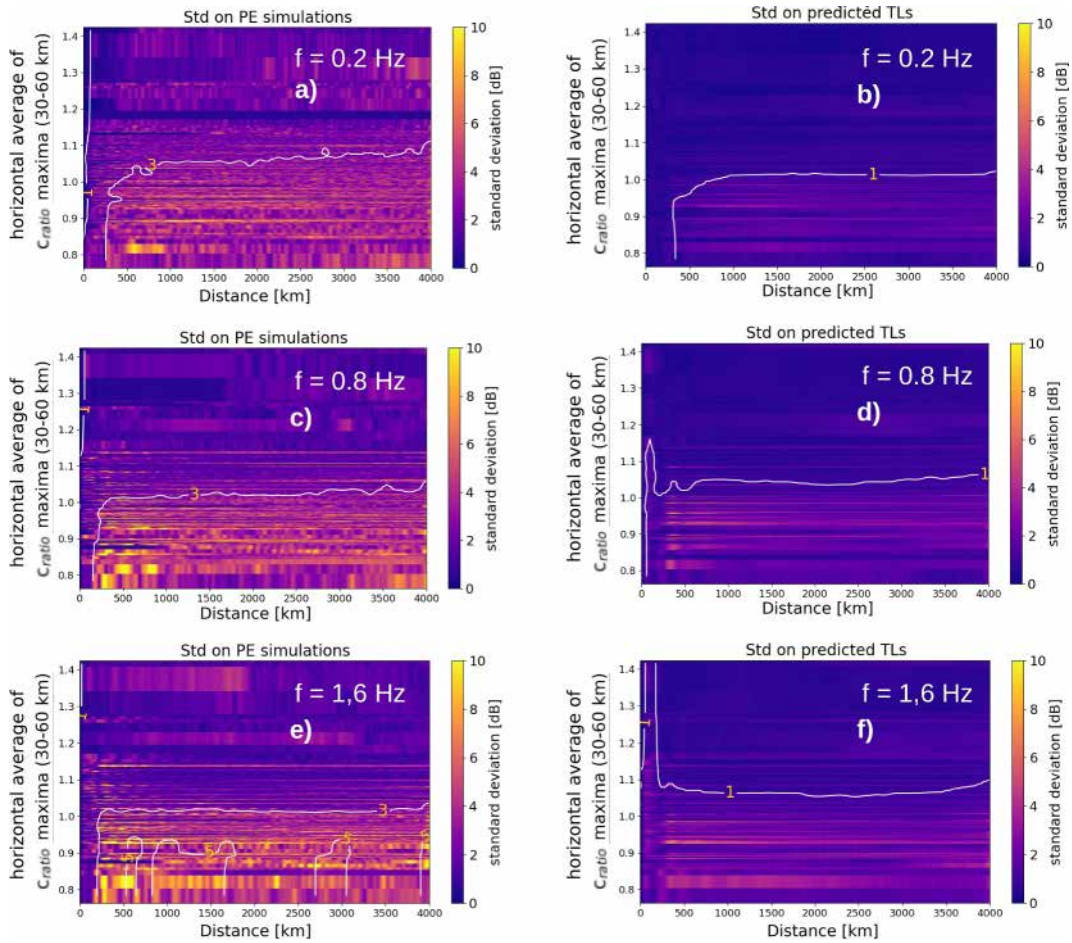

Figure D2. Uncertainty related to the input data, quantified on the generalization set using the test-time augmentation technique.

captured by $F_\theta(A_{z,d},f)$. However, even if the model acts as a low-pass filter, the data uncertainty associated with the predictions follows the global increase of uncertainty observed on the PE simulations in the absence of stratospheric waveguides (upwind scenarios). Figure D2 shows moreover that there is a slight increase in the average data-related uncertainty with increasing frequency, from 1 dB at 0.1 Hz to 1.5 dB at 1.6 Hz. Even if such a pattern is not visible in the PE simulations, we can link it with the increase of predictive errors with increasing frequency (see Section 4.1).

Appendix E: Details on the Comparisons Between the Training and the *Tonga* Set

This appendix aims to detail the comparison realized between the training set and the *Tonga-set* introduced in Section 6. The process used to obtain the distributions presented in Figure E1 is similar to the one described in Appendix C1. The results presented in this Appendix can be linked with Figure 11, which gives another point of view on the differences between the two sets.

The atmospheric slices of the *Tonga-set* being extracted on a never-before-encountered location on Earth and on a new date, we observe strong differences between their distributions and the one obtained on the training data. These differences are visible for all the minimal, average, and maximal c_{ratio} distributions for all considered atmospheric layers. Some departures correspond to over/under-represented atmospheric conditions in a given set, as previously observed in Appendix C1. For example, the *Tonga-set* contains a strong overrepresentation of slices with a minimal c_{ratio} of 0.75 in the stratosphere or with an average c_{ratio} of 1.04 above 90 km altitude. In addition to that, some distributions exhibit a horizontal shift compared to the training ones, such as the maximal distribution

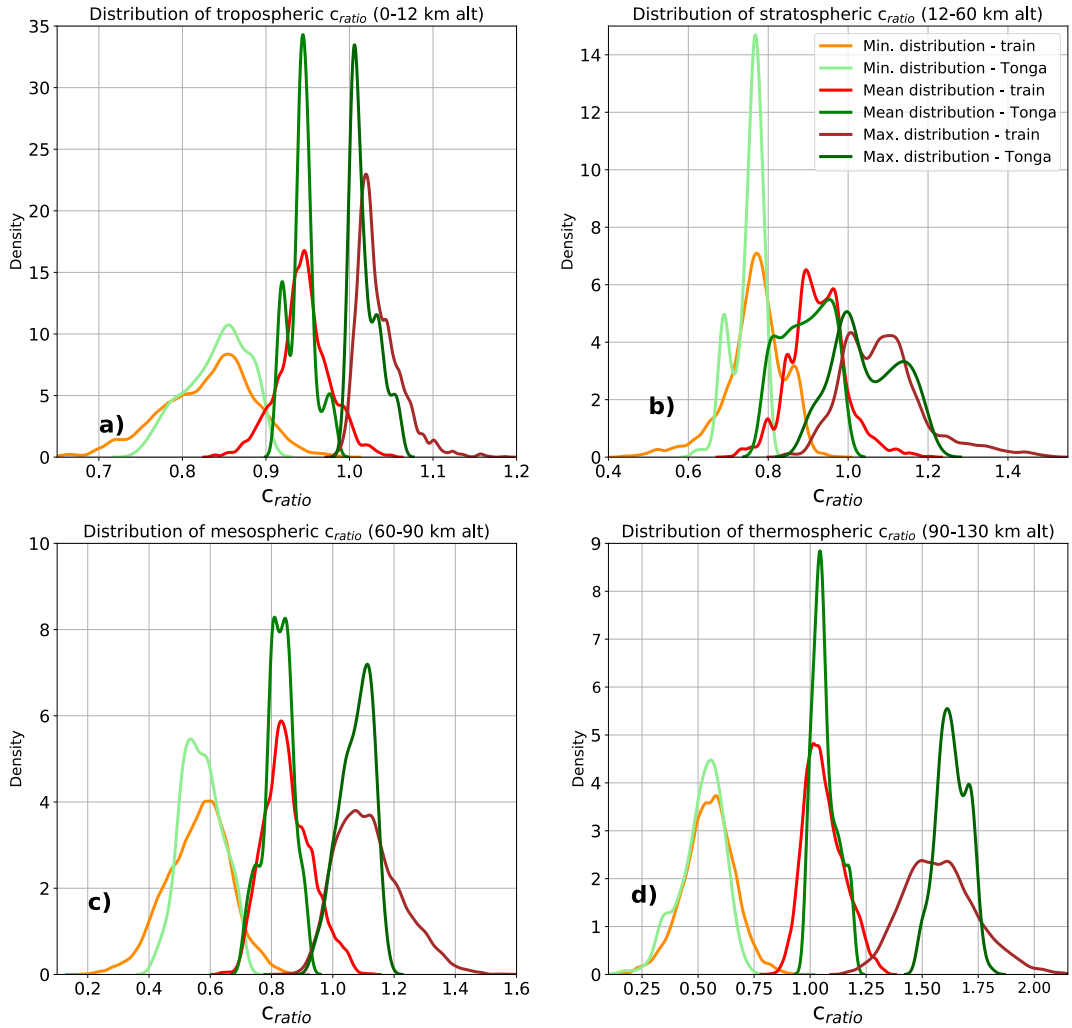

Figure E1. Distributions of minimal, average, and maximal values of c_{ratio} in the troposphere, stratosphere, mesosphere, and thermosphere computed on the training set and the *Tonga-set*.

in the troposphere or the maximal one in the thermosphere. It is important to note that all these differences can interfere in the same atmospheric slice.

Appendix F: MRAE Distribution According to the Atmospheric Conditions and Source Frequencies on the *Tonga-Set*

The current Appendix shows the distribution of the MRAE according to the source frequency and the mean c_{ratio} in the troposphere, stratosphere, mesosphere, and thermosphere on the *Tonga-set*. The mean c_{ratio} values are obtained by averaging the c_{ratio} contained within the range of altitudes of each atmospheric layer and along the 4,000-km-long propagation paths.

Figure F1 highlights the conditions associated with larger generalization errors ($\geq$20%), particularly in the absence of stratospheric or mesospheric waveguides at $f \geq 0.8$ Hz. Another region of higher errors is highlighted in panel a) for maximal mean c_{ratio} values in the troposphere. We link this region with notable differences in the training set and the *Tonga-set* atmospheric slice distributions (see horizontal shift of the *Tonga* maximal distribution in the troposphere, Appendix E1).

We further analyze the generalization capabilities of $F_\theta\left(A_{z,d},f\right)$ by evaluating it on five never-before-encountered source frequencies: 0.3, 0.6, 1.0, 1.2, and 1.4 Hz. Figure F2 details the distribution of the MRAE according to

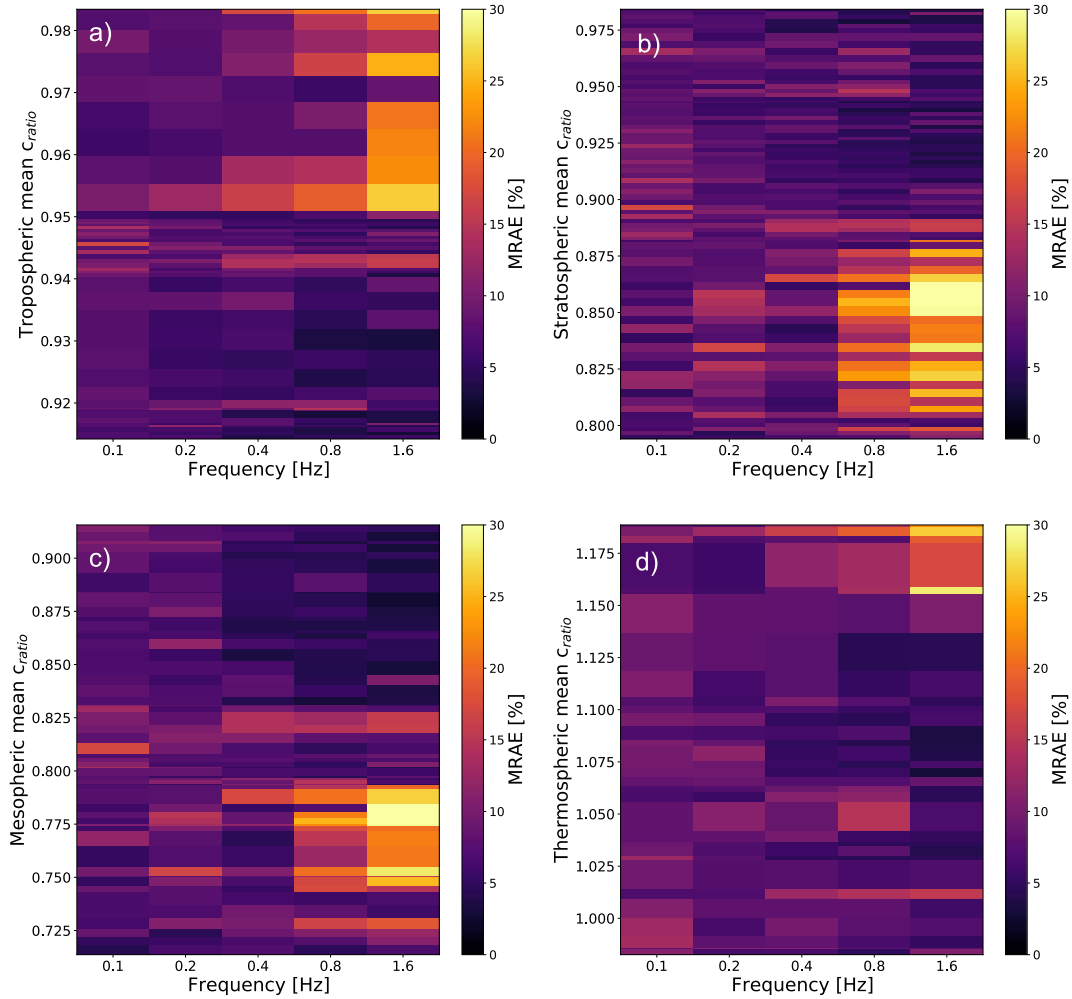

Figure F1. MRAE distribution according to the atmospheric conditions on the *Tonga-set* for the five usual source frequencies $f = 0.1, 0.2, 0.4, 0.8, 1.6$ Hz.

these new frequencies and the mean c_{ratio} per atmospheric layer on the *Tonga-set*. It highlights areas of larger generalization errors (≥20%), which can be strongly linked with the ones obtained by $F_\theta(A_{z,d}, f)$ on the same atmospheric slices, for the five usual source frequencies (see Figure F1). These regions with higher errors spread on the exact same mean c_{ratio} values but started as early as 0.6 Hz.

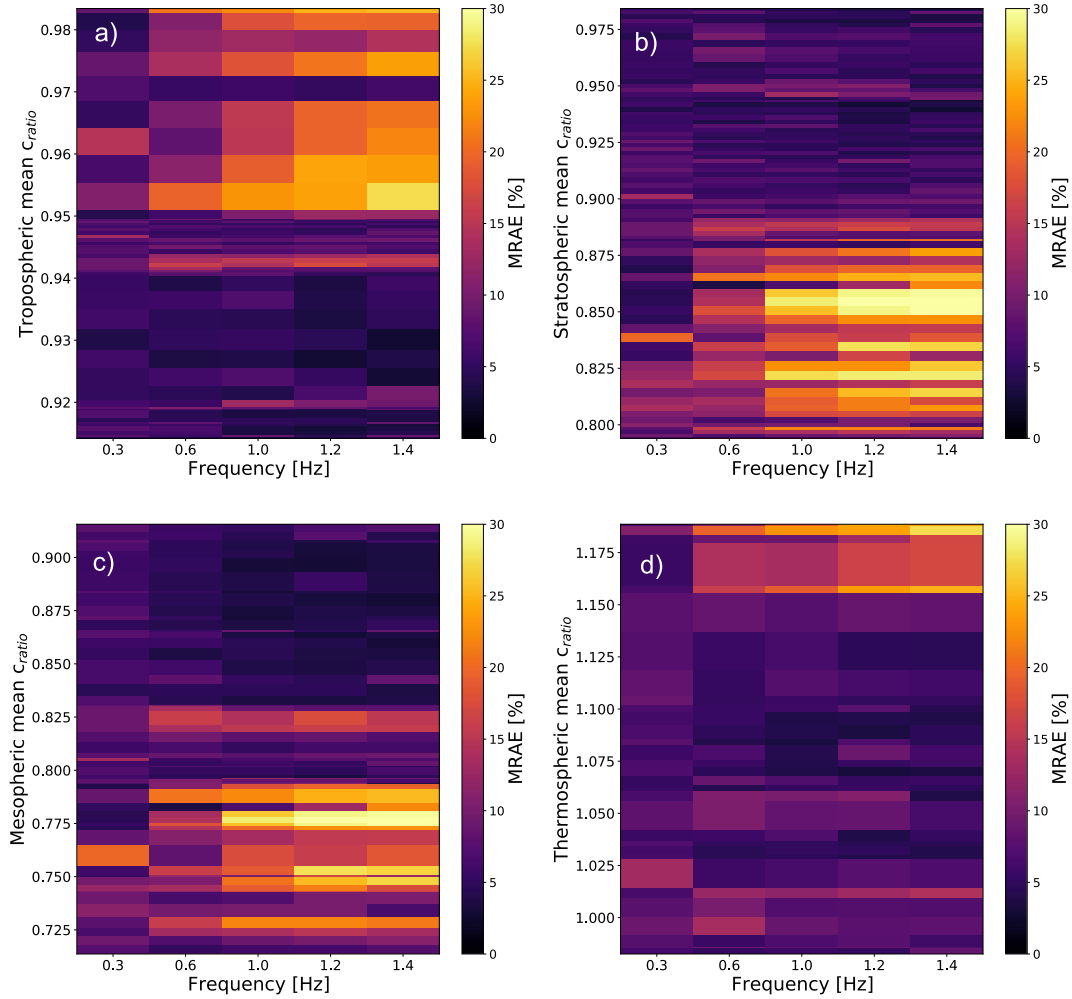

Figure F2. MRAE distribution according to the initial atmospheric conditions on the *Tonga-set*, for five never-before-encountered source frequencies $f = 0.3, 0.6, 1.0, 1.2, 1.4$ Hz.

Appendix G: Impact of the Time Coverage of the Training Database on the Generalization Capabilities

We create an expanded database by sampling the Earth's atmosphere not only on 15 January 2021, but also on 15 August 2021. We build realistic atmospheric slices and run PE simulations following the processes described in Section 2. The neural network is trained on this expanded database as presented in Section 3. By doing so, we aim to investigate the effect of increasing the time resolution of the training database on the model's performance. The model presented in this study is referred to as the "generic model", and the newly trained one as the "expanded model."

The generic and the expanded models are both evaluated on a new generalization set composed of 360 atmospheric slices around the military site of Hukkakero (Finland), extracted on 20 August 2019 and 14 August 2020. This site was selected because of its repeated explosions during a summer month in the northern hemisphere and is a benchmark event in the infrasound community (Vorobeva et al., 2023).

Figure G1 panel a) shows five boxplots (for the five source frequencies) summarizing the MRAE distributions along the 4,000-km-long propagation paths obtained by applying the generic model on this new generalization set. Compared with the previous evaluations, a degradation of the performance is noted. For all source frequencies, the median error is around 10%. We attribute this to the dates of 14 August 2019 and 20, 2020 being well outside the training domain of the generic model. On the other side, the expanded model shows improved

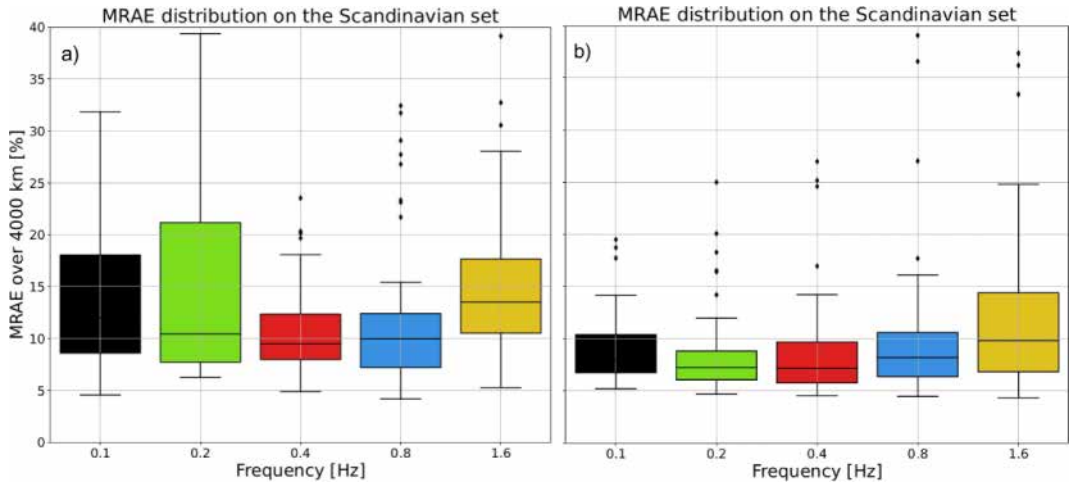

Figure G1. MRAE distributions along the 4,000-km-long propagation paths obtained by applying the generic and expanded models on the new generalization set.

performance (see panel b), with a significant reduction in median and quantile errors, especially below 0.4 Hz. This seems to indicate that increasing the time resolution of the training database by sampling the Earth at multiple dates (days, months, years) leads to better generalization capabilities of $F_\theta(A_{z,d},f)$.

Data Availability Statement

WACCM data (Atmospheric Chemistry Observations and modeling, National Center for Atmospheric Research, University Corporation for Atmospheric Research; Atmospheric Chemistry Observations & Modeling, National Center for Atmospheric Research, University Corporation for Atmospheric Research, (2020)) were obtained via the NCAR Research data archive and are accessible at https://doi.org/10.5065/G643-Z138. The authors are grateful to the National Center for Physical Acoustics (NCPA) at the University of Mississippi for the PE modeling tool *ePape* (Waxler et al., 2021), which is available on GitHub at https://github.com/chetzer-ncpa/ncpaprop-release or via https://doi.org/10.5281/zenodo.5562713. The TensorFlow library for Python (Abadi et al., 2015) can be downloaded from https://doi.org/10.5281/zenodo.4724125. The code to preprocess the input and output data as well as testing the neural network is available on GitHub at https://github.com/Alice-Cameijo/DeepLearning_Infrasound or via https://doi.org/10.5281/zenodo.15319091.

Acknowledgments
The authors are grateful for the valuable insights provided by Pierre Andraud, PhD candidate at the French Atomic Energy Commission, and Andreas Steinberg, research scientist at the German Federal Institute for Geosciences in Hannover, on the uncertainty quantification in the field of deep learning. The authors would also like to thank Edouard Forestier, engineer at Naval Group, and the team of researchers at NORSAR for their previous works on the use of convolutional and recurrent neural networks for TL estimations. We are grateful to Sarah Albert, Pierrick Mialle, and the third anonymous reviewer for their careful reading of the manuscript and their valuable comments, which allowed to improve the paper. We would like to thank the Defense Innovation Agency for its financial support which made this project possible. Quentin Brissaud, geophysicist and data-scientist at NORSAR, and Sven Peter Näsholm, Associate Professor at the University of Oslo and senior research geophysicist at NORSAR, acknowledge support from the Research Council of Norway basic research program FRIPRO through the project *Airborne inversion of Rayleigh waves* under Contract 335903. This study was facilitated by previous research realized within the framework of the ARISE and ARISE2 projects (Blanc et al. (2018, 2019)), funded by the European Commission FP7 and Horizon 2020 programmes (Grant Numbers 284387 and 653980).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., & Devin, M. (2015). Tensorflow: Large-scale machine learning [software]. In *Proc. 12th USENIX symp. on operating syst. des. implementation* (pp. 2–4). https://doi.org/10.5281/zenodo.4724125

Albert, S., Bowman, D., Silber, E., & Dannemann Dugick, F. (2023). The atmosofar channel: First direct observations of an elevated acoustic duct. *Earth and Space Science*, *10*(10), e2023EA003149. https://doi.org/10.1029/2023ea003149

Albert, S., & Linville, L. (2020). Benchmarking current and emerging approaches to infrasound signal classification. *Seismological Research Letters*, *91*(2A), 921–929. https://doi.org/10.1785/0220190116

Alexander, M. J., Liu, C. C., Bacmeister, J., Bramberger, M., Hertzog, A., & Richter, J. H. (2021). Observational validation of parameterized gravity waves from tropical convection in the whole atmosphere community climate model. *Journal of Geophysical Research: Atmospheres*, *126*(7), e2020JD033954. https://doi.org/10.1029/2020jd033954

Allen, S. J., & Vincent, R. A. (1995). Gravity wave activity in the lower atmosphere: Seasonal and latitudinal variations. *Journal of Geophysical Research*, *100*(D1), 1327–1350. https://doi.org/10.1029/94jd02688

Amezcua, J., Näsholm, S. P., & Vera Rodriguez, I. (2024). Using satellite data assimilation techniques to combine infrasound observations and a full ray-tracing model to constrain stratospheric variables. *Monthly Weather Review*, *152*(8), 1883–1902. https://doi.org/10.1175/mwr-d-23-0186.1

Assink, J. D. (2012). *Infrasound as upper atmospheric monitor*. Ph. D. Thesis.

Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, University Corporation for Atmospheric Research. (2020). Whole atmosphere community climate model (waccm) model output [Dataset]. *Boulder CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. https://doi.org/10.5065/G643-Z138

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., et al. (2021). Sudden stratospheric warmings. *Reviews of Geophysics*, *59*(1), e2020RG000708. https://doi.org/10.1029/2020rg000708

Bishop, J. W., Blom, P. S., Webster, J., Reichard-Flynn, W., & Lin, Y. (2022). Deep learning categorization of infrasound array data. *Journal of the Acoustical Society of America*, *152*(4), 2434–2445. https://doi.org/10.1121/10.0014903

Blanc, E., Ceranna, L., Hauchecorne, A., Charlton-Perez, A., Marchetti, E., Evers, L. G., et al. (2018). Toward an improved representation of middle atmospheric dynamics thanks to the ARISE project. *Surveys in Geophysics*, *39*(2), 171–225. https://doi.org/10.1007/s10712-017-9444-0

Blanc, E., Pol, K., Le Pichon, A., Hauchecorne, A., Keckhut, P., Baumgarten, G., et al. (2019). Middle atmosphere variability and model uncertainties as investigated in the framework of the ARISE project. In *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 845–887).

Blixt, E. M., Näsholm, S. P., Gibbons, S. J., Evers, L. G., Charlton-Perez, A. J., Orsolini, Y. J., & Kværna, T. (2019). Estimating tropospheric and stratospheric winds using infrasound from explosions. *Journal of the Acoustical Society of America*, *146*(2), 973–982. https://doi.org/10.1121/1.5120183

Blom, P. S., Dannemann, F. K., & Marcillo, O. E. (2018). Bayesian characterization of explosive sources using infrasonic signals. *Geophysical Journal International*, *215*(1), 240–251. https://doi.org/10.1093/gji/ggy258

Bowman, D. C., & Krishnamoorthy, S. (2021). Infrasound from a buried chemical explosion recorded on a balloon in the lower stratosphere. *Geophysical Research Letters*, *48*(21), e2021GL094861. https://doi.org/10.1029/2021gl094861

Brissaud, Q., Martin, R., Garcia, R. F., & Komatitsch, D. (2016). Finite-difference numerical modelling of gravitoacoustic wave propagation in a windy and attenuating atmosphere. *Geophysical Journal International*, *206*(1), 308–327. https://doi.org/10.1093/gji/ggw121

Brissaud, Q., Martin, R., Garcia, R. F., & Komatitsch, D. (2017). Hybrid Galerkin numerical modelling of elastodynamics and compressible Navier–Stokes couplings: Applications to seismo-gravito acoustic waves. *Geophysical Journal International*, *210*(2), 1047–1069. https://doi.org/10.1093/gji/ggx185

Brissaud, Q., Näsholm, S. P., Turquet, A., & Le Pichon, A. (2023). Predicting infrasound transmission loss using deep learning. *Geophysical Journal International*, *232*(1), 274–286. https://doi.org/10.1093/gji/ggac307

Buntine, W. L. (1991). Bayesian backpropagation. *Complex Systems*, *5*, 603–643.

Cho, K. (2014). On the properties of neural machine translation: Encoder-decoder approaches.

Christie, D., Veloso, J. V., Campus, P., Bell, M., Hoffmann, T., Langlois, A., et al. (2022). Detection of atmospheric nuclear explosions: The infrasound component of the international monitoring system. *Kerntechnik*, *66*(3), 96–101. https://doi.org/10.1515/kern-2001-0058

Chu, X., Zhao, J., Lu, X., Harvey, V. L., Jones, R. M., Becker, E., et al. (2018). Lidar observations of stratospheric gravity waves from 2011 to 2015 at McMurdo (77.84° S, 166.69° E), Antarctica: 2. potential energy densities, lognormal distributions, and seasonal variations. *Journal of Geophysical Research: Atmospheres*, *123*(15), 7910–7934. https://doi.org/10.1029/2017jd027386

Chunchuzov, I., & Kulichkov, S. (2019). Internal gravity wave perturbations and their impacts on infrasound propagation in the atmosphere. In *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 551–590).

de Groot-Hedlin, C. D., Hedlin, M. A., & Walker, K. (2011). Finite difference synthesis of infrasound propagation through a windy, viscous atmosphere: Application to a bolide explosion detected by seismic networks. *Geophysical Journal International*, *185*(1), 305–320. https://doi.org/10.1111/j.1365-246x.2010.04925.x

de Groot-Hedlin, C. D., Hedlin, M. A., & Drob, D. P. (2009). Atmospheric variability and infrasound monitoring. In *Infrasound monitoring for atmospheric studies* (pp. 475–507).

Drob, D. P., Emmert, J. T., Meriwether, J. W., Makela, J. J., Doornbos, E., Conde, M., et al. (2015). An update to the Horizontal Wind Model (HWM): The quiet time thermosphere. *Earth and Space Science*, *2*(7), 301–319. https://doi.org/10.1002/2014ea000089

Fee, D., Waxler, R., Assink, J., Gitterman, Y., Given, J., Coyne, J., et al. (2013). Overview of the 2009 and 2011 sayarim infrasound calibration experiments. *Journal of Geophysical Research: Atmospheres*, *118*(12), 6122–6143. https://doi.org/10.1002/jgrd.50398

Gal, Y., & Ghahramani, Z. (2015). Bayesian convolutional neural networks with bernoulli approximate variational inference.

Gardner, C. S., Hostetler, C. A., & Franke, S. J. (1993). Gravity wave models for the horizontal wave number spectra of atmospheric velocity and density fluctuations. *Journal of Geophysical Research*, *98*(D1), 1035–1049. https://doi.org/10.1029/92jd02051

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., et al. (2023). A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, *56*(Suppl 1), 1513–1589. https://doi.org/10.1007/s10462-023-10562-9

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The whole atmosphere community climate model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12380–12403. https://doi.org/10.1029/2019jd030943

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323). JMLR Workshop and Conference Proceedings.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings.

Green, D. N., & Bowers, D. (2010). Estimating the detection capability of the International Monitoring System infrasound network. *Journal of Geophysical Research*, *115*, D18. https://doi.org/10.1029/2010jd014017

Gutenberg, B. (1939). The velocity of sound waves and the temperature in the stratosphere in southern california. *Bulletin of the American Meteorological Society*, *20*(5), 192–201. https://doi.org/10.1175/1520-0477-20.5.192

Hart, C. R., Wilson, D. K., Pettit, C. L., & Nykaza, E. T. (2021). Machine-learning of long-range sound propagation through simulated atmospheric turbulence. *Journal of the Acoustical Society of America*, *149*(6), 4384–4395. https://doi.org/10.1121/10.0005280

Hedlin, M. A., & Drob, D. P. (2014). Statistical characterization of atmospheric gravity waves by seismoacoustic observations. *Journal of Geophysical Research: Atmospheres*, *119*(9), 5345–5363. https://doi.org/10.1002/2013jd021304

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hostetler, C. A., & Gardner, C. S. (1994). Observations of horizontal and vertical wave number spectra of gravity wave motions in the stratosphere and mesosphere over the mid-pacific. *Journal of Geophysical Research*, *99*(D1), 1283–1302. https://doi.org/10.1029/93jd02927

Kingma, D. P. (2014). Adam: A method for stochastic optimization.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems* (Vol. 30).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Le Pichon, A., Ceranna, L., & Vergoz, J. (2012). Incorporating numerical modelling into estimates of the detection capability of the IMS infrasound network. *Journal of Geophysical Research*, *117*, D5.

Le Pichon, A., Herry, P., Mialle, P., Vergoz, J., Brachet, N., Garcés, M., & Ceranna, L. (2005). Infrasound associated with 2004–2005 large Sumatra earthquakes and tsunami. *Geophysical Research Letters*, *32*, 19. https://doi.org/10.1029/2005gl023893

Le Pichon, A., Vergoz, J., Blanc, E., Guilbert, J., Ceranna, L., Evers, L., & Brachet, N. (2009). Assessing the performance of the International Monitoring System's infrasound network: Geographical coverage and temporal variabilities. *Journal of Geophysical Research*, *114*, D8. https://doi.org/10.1029/2008jd010907

Letournel, P., Listowski, C., Bocquet, M., Le Pichon, A., & Farchi, A. (2024). Evaluating numerical weather prediction models in the middle atmosphere using coherent oceanic acoustic noise observations. *Journal of Geophysical Research: Atmospheres*, *129*(23), e2024JD042034. https://doi.org/10.1029/2024jd042034

Listowski, C., Stephan, C. C., Le Pichon, A., Hauchecorne, A., Kim, Y. H., Achatz, U., & Bölöni, G. (2024). Stratospheric gravity waves impact on infrasound transmission losses across the International Monitoring System. *Pure and Applied Geophysics*, 1–33. https://doi.org/10.1007/s00024-024-03467-3

Martire, L., Martin, R., Brissaud, Q., & Garcia, R. F. (2022). SPECFEM2D-DG, an open-source software modelling mechanical waves in coupled solid–fluid systems: The linearized navier–stokes approach. *Geophysical Journal International*, *228*(1), 664–697. https://doi.org/10.1093/gji/ggab308

Marty, J. (2019). The IMS infrasound network: Current status and technological developments. In *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 3–62).

Marty, J., Doury, B., & Kramer, A. (2021). Low and high broadband spectral models of atmospheric pressure fluctuation. *Journal of Atmospheric and Oceanic Technology*, *38*(10), 1813–1822. https://doi.org/10.1175/jtech-d-21-0006.1

Matoza, R. S., Fee, D., Assink, J. D., Iezzi, A. M., Green, D. N., Kim, K., et al. (2022). Atmospheric waves and global seismoacoustic observations of the January 2022 Hunga eruption, Tonga. *Science*, *377*(6601), 95–100. https://doi.org/10.1126/science.abo7063

Mialle, P., Brown, D., Arora, N., & colleagues from, I. D. C. (2019). Advances in operational processing at the international data centre. *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits*, 209–248. https://doi.org/10.1007/978-3-319-75140-5_6

Pettit, C. L., & Wilson, D. K. (2020). A physics-informed neural network for sound propagation in the atmospheric boundary layer. In *Proceedings of meetings on acoustics* (Vol. 42, p. 022002). AIP Publishing. https://doi.org/10.1121/2.0001383

Picone, J. M., Hedin, A. E., Drob, D. P., & Aikin, A. C. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research*, *107*(A12), SIA–15. https://doi.org/10.1029/2002ja009430

Pierce, A. D. (2019). *Acoustics: An introduction to its physical principles and applications*. Springer.

Pilger, C., Gaebler, P., Hupe, P., Kalia, A. C., Schneider, F. M., Steinberg, A., et al. (2021). Yield estimation of the 2020 Beirut explosion using open access waveform and remote sensing data. *Scientific Reports*, *11*(1), 14144. https://doi.org/10.1038/s41598-021-93690-y

Plougonven, R., de la Cámara, A., Hertzog, A., & Lott, F. (2020). How does knowledge of atmospheric gravity waves guide their parameterizations? *Quarterly Journal of the Royal Meteorological Society*, *146*(728), 1529–1543. https://doi.org/10.1002/qj.3732

Podglajen, A., Le Pichon, A., Garcia, R. F., Gérier, S., Millet, C., Bedka, K., et al. (2022). Stratospheric balloon observations of infrasound waves from the 15 January 2022 Hunga eruption, Tonga. *Geophysical Research Letters*, *49*(19), e2022GL100833. https://doi.org/10.1029/2022gl100833

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., & Courville, A. (2019). On the spectral bias of neural networks. In *International conference on machine learning* (pp. 5301–5310). PMLR.

Silber, E. A., Bowman, D. C., & Ronac Giannone, M. (2023). Detection of the large surface explosion coupling experiment by a sparse network of balloon-borne infrasound sensors. *Remote Sensing*, *15*(2), 542. https://doi.org/10.3390/rs15020542

Smets, P. S., & Evers, L. G. (2014). The life cycle of a sudden stratospheric warming from infrasonic ambient noise observations. *Journal of Geophysical Research: Atmospheres*, *119*(21), 12–084. https://doi.org/10.1002/2014jd021905

Smith, S. A., Fritts, D. C., & Vanzandt, T. E. (1987). Evidence for a saturated spectrum of atmospheric gravity waves. *Journal of the Atmospheric Sciences*, *44*(10), 1404–1410. https://doi.org/10.1175/1520-0469(1987)044<1404:efasso>2.0.co;2

Sutherland, L. C., & Bass, H. E. (2004). Atmospheric absorption in the atmosphere up to 160 km. *Journal of the Acoustical Society of America*, *115*(3), 1012–1032. https://doi.org/10.1121/1.1631937

Tishby, & Solla (1989). Consistent inference of probabilities in layered networks: Predictions and generalizations. In *International 1989 joint conference on neural networks* (pp. 403–409). IEEE.

Vanderbecken, P., Mahfouf, J.-F., & Millet, C. (2020). Bayesian selection of atmospheric profiles from an ensemble data assimilation system using infrasonic observations of May 2016 Mount Etna eruptions. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD031168. https://doi.org/10.1029/2019jd031168

Vera Rodriguez, I., Näsholm, S. P., & Le Pichon, A. (2020). Atmospheric wind and temperature profiles inversion using infrasound: An ensemble model context. *Journal of the Acoustical Society of America*, *148*(5), 2923–2934. https://doi.org/10.1121/10.0002482

Vergoz, J., Hupe, P., Listowski, C., Le Pichon, A., Garcés, M. A., Marchetti, E., et al. (2022). IMS observations of infrasound and acoustic-gravity waves produced by the January 2022 volcanic eruption of Hunga, Tonga: A global analysis. *Earth and Planetary Science Letters*, *591*, 117639. https://doi.org/10.1016/j.epsl.2022.117639

Vorobeva, E., Assink, J., Espy, P. J., Renkwitz, T., Chunchuzov, I., & Näsholm, S. P. (2023). Probing gravity waves in the middle atmosphere using infrasound from explosions. *Journal of Geophysical Research: Atmospheres*, *128*(13), e2023JD038725. https://doi.org/10.1029/2023jd038725

Vorobeva, E., De Carlo, M., Le Pichon, A., Espy, P. J., & Näsholm, S. P. (2021). Benchmarking microbarom radiation and propagation model against infrasound recordings: A vespagram-based approach. In *Annales Geophysicae* (Vol. 39(3), pp. 515–531). https://doi.org/10.5194/angeo-39-515-2021. Copernicus Publications.

Waxler, R., & Assink, J. (2019). Propagation modelling through realistic atmosphere and benchmarking. In *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 509–549).

Waxler, R., Hetzer, C., Assink, J., & Velea, D. (2021). Chetzer-ncpa/ncpaprop-release: NCPAprop (v2. 1.0v2. 1.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.5562713