Predicting infrasound transmission loss using deep learning

Quentin Brissaud[1,1], Sven Peter Näsholm[2,2], Antoine Turquet[1,1], and Alexis Le Pichon[3,3]

[1]NORSAR
[2]University of Oslo,NORSAR
[3]CEA

November 30, 2022

Abstract

Modelling the spatial distribution of infrasound attenuation (or transmission loss, TL) is key to understanding and interpreting microbarometer data and observations. Such predictions enable the reliable assessment of infrasound source characteristics such as ground pressure levels associated with earthquakes, man-made or volcanic explosion properties, and ocean-generated microbarom wavefields. However, the computational cost inherent in full-waveform modelling tools, such as Parabolic Equation (PE) codes, often prevents the exploration of a large parameter space, i.e., variations in wind models, source frequency, and source location, when deriving reliable estimates of source or atmospheric properties – in particular for real-time and near-real-time applications. Therefore, many studies rely on analytical regression-based heuristic TL equations that neglect complex vertical wind variations and the range-dependent variation in the atmospheric properties. This introduces significant uncertainties in the predicted TL. In the current contribution, we propose a deep learning approach trained on a large set of simulated wavefields generated using PE simulations and realistic atmospheric winds to predict infrasound ground-level amplitudes up to 1000 km from a ground-based source. Realistic range dependent atmospheric winds are constructed by combining ERA5, NRLMSISE-00, and HWM-14 atmospheric models, and small-scale gravity-wave perturbations computed using the Gardner model. Given a set of wind profiles as input, our new modelling framework provides a fast (0.05 s runtime) and reliable (~5 dB error on average, compared to PE simulations) estimate of the infrasound TL.

Predicting infrasound transmission loss using deep learning

Quentin Brissaud [https://orcid.org/0000-0001-8189-4699],[1] Sven Peter Näsholm [https://orcid.org/0000-0001-9107-4002],[1,2] Antoine Turquet [https://orcid.org/0000-0003-1920-935X],[1] and Alexis Le Pichon [https://orcid.org/0000-0001-6531-069X][3]

[1] *NORSAR, Gunnar Randers vei 15, Kjeller, Norway*

[2] *Department of Informatics, University of Oslo, P.O. Box 1080, NO–0316 Oslo, Norway*

[3] *CEA, DAM, DIF, F-91297 Arpajon, France*

(Dated: May 6, 2022)

SUMMARY

Modelling the spatial distribution of infrasound attenuation (or transmission loss, TL) is key to understanding and interpreting microbarometer data and observations. Such predictions enable the reliable assessment of infrasound source characteristics such as ground pressure levels associated with earthquakes, man-made or volcanic explosion properties, and ocean-generated microbarom wavefields. However, the computational cost inherent in full-waveform modelling tools, such as Parabolic Equation (PE) codes, often prevents the exploration of a large parameter space, i.e., variations in wind models, source frequency, and source location, when deriving reliable estimates of source or atmospheric properties – in particular for real-time and near-real-time applications. Therefore, many studies rely on analytical regression-based heuristic TL equations that neglect complex vertical wind variations and the range-dependent variation in the atmospheric properties. This introduces significant uncertainties in the predicted TL. In the current contribution, we propose a deep learning approach trained on a large set of simulated wavefields generated using PE simulations and realistic atmospheric winds to predict infrasound ground-level amplitudes up to 1000 km from a ground-based source. Realistic range dependent atmospheric winds are constructed by combining ERA5, NRLMSISE-00, and HWM-14 atmospheric models, and small-scale gravity-wave perturbations computed using the Gardner model. Given a set of wind profiles as input, our new modelling framework provides a fast (0.05 s runtime) and reliable ($\sim$ 5 dB error on average, compared to PE simulations) estimate of the infrasound TL.

Keywords: Infrasound, Wave propagation, Machine learning, Numerical modelling

1. INTRODUCTION

Surface and subsurface sources (e.g., explosions, microbaroms, earthquakes) excite low-frequency acoustic waves, i.e., infrasound, that can travel large distances in the Earth's atmosphere. The refraction and reflection of infrasound waves back to the surface due to vertical and horizontal gradients of atmospheric winds and temperatures enable their detection at ground arrays. Because infrasound waves carry information about the source, they have traditionally been used to retrieve location and yield estimates of nuclear explosions (*Evers and Haak*, 2010). Recently, the detection and modelling of infrasound phases have also enabled the inversion of critical seismic source and subsurface parameters such as focal mechanism (*Shani-Kadmiel et al.*, 2021), focal depth (*Averbuch et al.*, 2020; *Lai et al.*, 2021), ground motions (*Hernandez et al.*, 2018), or seismic velocity structures (*Brissaud et al.*, 2021).

Accurately predicting the spatial distribution of infrasound attenuation, i.e., Transmission Loss (TL), is key to build robust estimates of source and subsurface characteristics. Parabolic Equations (PE) (*Waxler et al.*, 2021) or finite difference codes (*de Groot-Hedlin*, 2008; *Brissaud et al.*, 2016) are typically used to compute accurate estimates of acoustic amplitudes in realistic wind structures. However, owing to the prohibitive computational cost of full-waveform numerical modelling tools, most infrasound studies rely on empirical equations to relate infrasound amplitudes to source parameters. Widely-used regression equations include models to estimate the explosion yield from peak infrasound amplitudes (e.g., *Golden et al.*, 2012) and empirical equations relating pressure at the source and observed infrasound amplitudes (*Le Pichon et al.*, 2012). In particular, the construction of empirical equations ignores or greatly over-simplifies atmospheric wind structures. For instance, in *Le Pichon et al.* (2012), the authors assume a single range-independent Gaussian stratospheric duct to optimize their regression model. Yet, vertical and horizontal wind gradients at various altitudes can drastically affect the TL at the ground (*de Groot-Hedlin et al.*, 2010).

Empirical models rely on over-simplistic representations of the wind structures because the mapping between source frequency, atmospheric specifications, and TL is highly nonlinear and poorly constrained. In order to bridge the gap between computationally expensive numerical models and over-simplistic empirical equations, supervised Machine-Learning (ML) models trained over synthetic or recorded datasets can offer an accurate and inexpensive alternative to existing modelling tools (*Michalopoulou et al.*, 2021). Previous studies have employed ML models to predict TL: *Pettit and Wilson* (2020) built a Physics-Informed Neural Network (PINN) trained over synthetic PE simulation results to predict attenuation maps (along range and altitude) in the atmospheric boundary layer. PINN introduces regularization terms in the cost function to account for physics-based constraints. This model provides an inexpensive alternative to existing modelling tools but shows low accuracy as it struggles with adjusting the weights of the physics-informed parameters in the objective function. Additionally, atmospheric specifications are encoded using only wind profiles, and this approach was not adapted to long-range propagation. *Hart et al.* (2021) used a fully connected neural network to predict two-dimensional (2D) attenuation in a turbulent atmosphere from a set of predefined input parameters describing the turbulent field. This model shows a relatively low error (< 7 dB) but relies on over-simplified wind models with a set of 13 inputs to describe the velocity field which are not representative of long-range propagation.

Relating wind structures to TLs is key to accurately reproduce full-waveform simulations. Instead of using pre-defined parameters to describe the wind velocity field, Convolutional-Neural Networks (CNN, *Krizhevsky et al.* (2012)) provide an excellent solution to identify patterns of interest within input wind models. Such patterns are extracted using a set of filters described by a number of coefficients that are optimized during the ML training process. Such network is generally followed by a set of fully-connected layers relating the encoded information by the CNN and the output. In this contribution we propose a new ML model trained over synthetic PE simulations to build ground TL in realistic range-dependent wind models that both shows a low computational cost compared to existing modelling tools, and high accuracy over long-range propagation.

2. BUILDING A TRANSMISSION-LOSS DATASET

Building a synthetic TL dataset requires a modelling tool and a set of atmospheric models. Similar to *Le Pichon et al.* (2012), we generate TL profiles using the open-source (PE) solver ePape, provided by the US National Center for Physical Acoustics (NCPA, *Waxler et al.*, 2021). To provide realistic bounds for the atmospheric models, we collect 1048 slices of 1000 km length up to 80 km altitude from ERA5 re-analysis models, discretized over 137 altitude levels (*ECMWF*, 2018) with a horizontal resolution of 1 degree. The choice of 1000 km slice length enables the analysis of a wide variety of regional observations (e.g., *Ceranna et al.*, 2009; *Fee and Matoza*, 2013) while keeping the computational time low to build the training dataset. The spatial step of 1 degree is picked as a trade-off between the resolution to capture ERA5 spatial variability and the computational time to both download atmospheric models and run simulations. Since ERA5 models are limited to around 80 km altitude, we use two empirical models to retrieve atmospheric properties up to 120 km altitude: HWM-14 to obtain zonal and meridional winds (*Drob et al.*, 2015), and NRLMSISE-00 to retrieve temperatures (*Picone et al.*, 2002). ERA5 and HWM-14/NRLMSISE-00 atmospheric models are stitched together using a cubic interpolation over the altitude range of 75 to 85 km. Because atmospheric properties vary with latitude, longitude, and time of the year, ERA5 profiles are uniformly sampled between latitudes -40 to 70 degrees, longitudes -150 to 165 degrees, and between years 2010 to 2020 (see Fig. 1a).

ERA5 models lack resolution to capture fine-scale wind and temperature fluctuations such as gravity-wave breaking above the troposphere (*Chunchuzov et al.*, 2015; *Chunchuzov and Kulichkov*, 2019). To account for unresolved wind perturbations at higher altitudes, infrasound studies typically consider the Gardner model to add gravity-wave perturbations to the original wind profiles (*Gardner et al.*, 1993). Therefore, we account for small-scale perturbations by considering four Gardner realizations for each atmospheric slice in addition to the original slice (see green stage in Fig. 2a). Similar to *Norris and Gibson* (2002), we generate Gardner perturbations by considering four altitude levels 84, 70, 45, and 21 km, at which we sample standard deviations uniformly within the range of, respectively, 1–25, 1–18, 1–10, and 1–5 m/s. Finally, because the direction of propagation within an atmospheric slice, i.e., upwind or downwind propagation, greatly alters the TLs at the ground, we augment our dataset of atmospheric models by running simulations in both scenarios by changing the sign of the projected winds (see yellow stage in Fig. 2a). Our final dataset includes 41920 simulations.

Effective velocity ratios $\overline{c}_{\text{eff}}$, i.e., ratio between the maximum effective velocity in a given atmospheric layer and the surface, provide useful insight into the likelihood of infrasound refractions along the direction of propagation. Similarly to *Le Pichon et al.* (2012), we compute $\overline{c}_{\text{eff}}$ as $\overline{c}_{\text{eff}} = \max_{z \in \text{layer}}\{c_{\text{eff}}(z)\}/c_{\text{eff}}(z = 0)$, where $c_{\text{eff,layer}}(z) = c(z) + w(z)$ is the effective velocity, where c (m/s) is the adiabatic sound velocity, w (m/s) the along-path wind velocity, z (m) the altitude, and layer $= (z_{\text{start}}, z_{\text{end}})$ is given by the altitude bounds z_{start} and z_{end} (m) for a given atmospheric layer. The distribution of effective velocity ratios $\overline{c}_{\text{eff}}$ computed from our final atmospheric model dataset for three different altitude regimes, shown in Fig. 1b, is close to a Gaussian distribution, centered around $\overline{c}_{\text{eff}} = 1$. This indicates that our dataset includes models with and without strong high-altitude ducts. The distribution of tropospheric effective velocity ratios is narrower than for higher-altitude layers. This owes to the small number of occurrences of tropospheric wave ducts in our dataset. In addition to vertical variations of atmospheric properties, lateral variations can play a significant role for long-range infrasound propagation. We quantify the range of lateral variations by computing the maximum lateral standard deviation of wind velocities in a given atmospheric layer $\text{std}_{\text{layer}}$ (m/s) such that $\text{std}_{\text{layer}} = \max_{z \in \text{layer}}(\text{std}_{x \in \text{range}}\{w(x, z)\})$, where std is the standard deviation, $w(x, z)$ (m/s) is the along-path wind at a given range x (m) and altitude z (m), range $= (0, 1000)$ km is the total atmospheric slice range. In contrast to large vertical variations of wind velocities, most ERA5 models show small lateral variations of wind velocities ($\text{std}_{\text{layer}} < 15$ m/s, see Fig. 1c). The largest lateral wind variations occur above the stratosphere since winds at these high altitudes are generally the strongest on Earth (*Blanc et al.*, 2018).

TL profiles are then computed over 1000 km from the source for a source at ground level using 7 Padé coefficients and the Sutherland-Bass attenuation model (*Sutherland and Bass*, 2004) using NCPA's ePape PE simulator (*Waxler et al.*, 2021). We extract 10 atmospheric profiles along each 1000 km slice, i.e., 100 km horizontal discretization, from

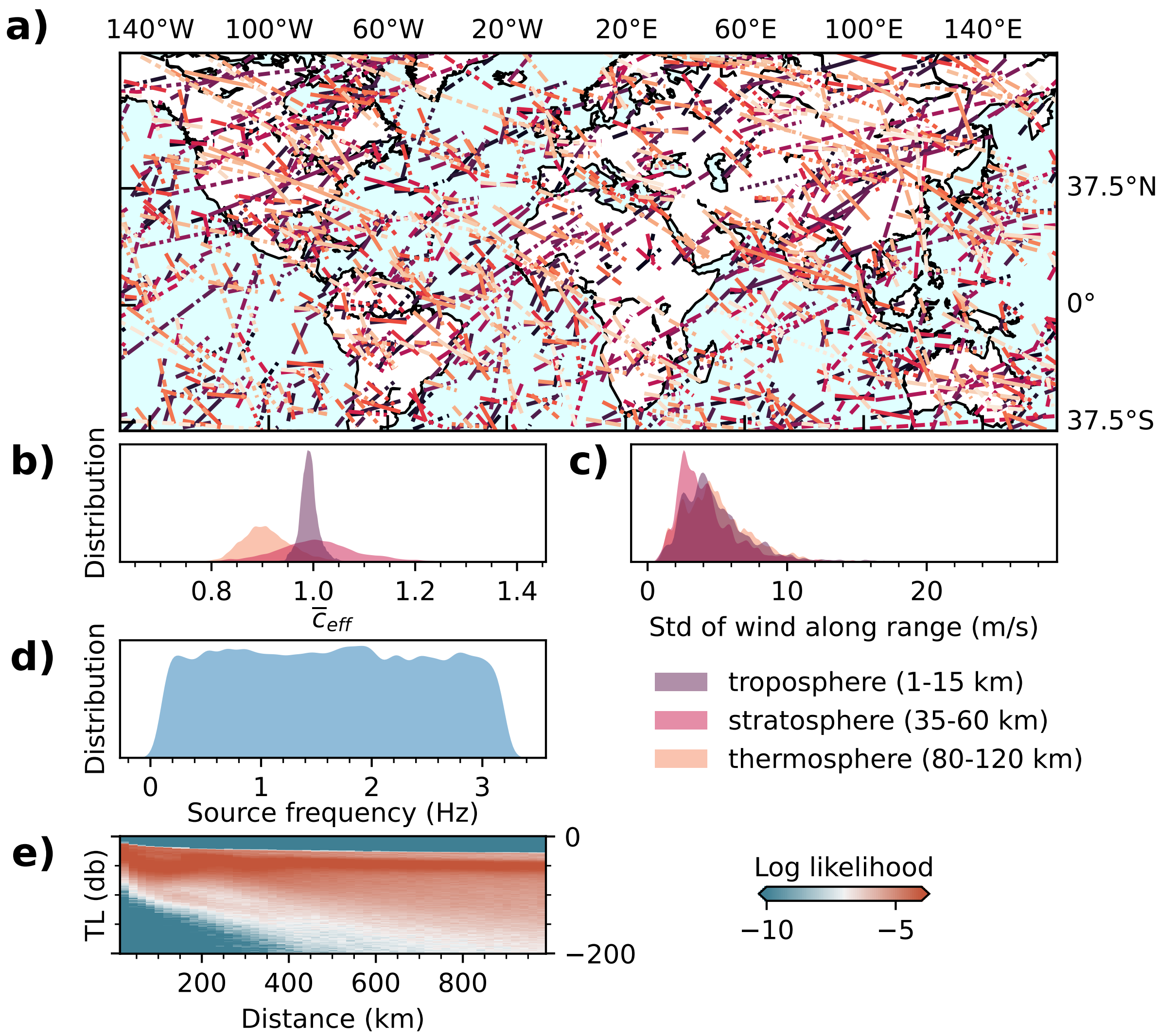

Figure 1. Atmospheric model and TL datasets. (a) distribution of 1000 km long atmospheric slices extracted from the ERA5 dataset. Slices are given different colors and line styles (dashed and solid lines) to facilitate the visualization of their distribution around the globe. (b) Distribution of effective velocity ratio $\overline{c}_{\mathrm{eff}}$ between the ground and various atmospheric layers: troposphere (purple) between 1 and 15 km altitude, troposphere (purple) between 35 and 60 km altitude, and thermosphere (purple) between 80 and 120 km altitude. (c) Distribution of standard deviations of wind velocities along range for various atmospheric layers. (d) Distribution of input source frequencies used in PE simulations to build the entire TL dataset. (e) TL distribution represented as log likelihood (computed from Gaussian Kernel density estimates) vs distance determined from our entire TL dataset.

the ERA5 dataset. Signals from sources of interest (earthquakes, volcanoes, large explosions) typically show dominant frequencies below 5 Hz. Therefore, similar to *Le Pichon et al.* (2012), we sample 5 source frequencies from a uniform distribution between 0.1 to 3.2 Hz for each atmospheric slice (see Fig. 1d and Fig. 2a). PE assumes slow lateral variations in the atmospheric properties over the scale of one wavelength. To ensure smoothly varying atmospheric properties, we must only consider models that do not include lateral variations over the scale of the largest wavelength considered, which means $\lambda \approx 3.5$ km at 0.1 Hz. Because we are using a 100 km horizontal discretization, interpolation of atmospheric properties within the NCPA software will generate smooth-enough models to fulfil the PE assumptions. The resulting distribution of TL profiles is shown in Fig. 2e. Most profiles show TL values > -100 db at large distances from the source owing to the presence of Gardner perturbations in most models which refract significantly the infrasound energy back to the surface (*Chunchuzov et al.*, 2015).

PE neglect nonlinear terms and cross-winds. Nonlinearities affect primarily the amplitude and frequency content of thermospheric phases for large-amplitude pressure sources (*Sabatini et al.*, 2019). Therefore, uncertainties on the predicted amplitudes must be accounted for when investigating high-yield surface sources. When large-amplitude sources are considered, PE simulations will severely overpredict the amplitude of refracted phases at the ground. While cross-winds have a significant impact on the apparent backazimuth observed from refracted phases at stations located at large distances from the source, their influence on infrasound amplitudes is generally considered insignificant (*Hernandez et al.*, 2018; *Shani-Kadmiel et al.*, 2021).

3. DESIGNING A TRANSMISSION-LOSS MODEL

PE based simulations are often used to provide a mapping between 2D range-dependent profiles (temperature, winds, and pressure), frequency, and transmission loss profiles under the effective-velocity approximation. Our goal is to retrieve the same TL estimates as provided directly by PE, but at a significantly reduced computational cost. This is achieved using an alternative nonlinear map between the atmospheric specification and frequency inputs and the TL output using a neural network which is pre-trained on an extended set of PE simulations. Variations of surface-to-surface TL with range for a given source frequency between different atmospheric models are primarily controlled by lateral and vertical wind variations. To reduce the ML architecture complexity, we assume a nonlinear mapping to exist between frequency, 2D wind, and TL and that this adequately approximates the full PE solution.

We implement this mapping between winds and ground TLs using a supervised deep learning algorithm. A deep learning neural network maps a set of inputs, e.g., wind profiles and frequency, into a set of outputs, e.g., TL profiles. For a given network architecture, supervised learning consists of the optimization of hierarchically organized nonlinear functions. The optimization process iteratively updates the non-linear function parameters by comparing training outputs and outputs predicted by the deep learning model. The most generic network consists of a succession of fully-connected layers where each layer is composed of a set of nonlinear functions described by a weight, a bias, and an activation function. For fully-connected layers, the output of each function from a first layer is used as input to each function of the next layer. Such architecture does not assume any relationships between the inputs and outputs of successive layers. This generic layer configuration leads to lower predictive power, as it requires an extended number of parameters to optimize and ignores spatial correlations in the input data.

Accounting for spatial correlations, i.e., relationships between neighboring inputs such as local wind gradients, are key to extract physically-meaningful patterns from continuous input data (e.g., images or timeseries) and improve network performances (*d'Ascoli et al.*, 2019). To leverage spatial correlations, Convolutional Neural Networks (CNN) use a series of operations, namely, digital filtering, pooling, and activation (see blue stage in Fig. 2b) to extract patterns at different scales across 1D or 2D input data (*Krizhevsky et al.*, 2012). In 2D, the digital filtering step consists of the convolution product between a series of kernel and the input image which outputs a filtered image. During the training of a CNN, the optimization process will update the values, or parameters, that compose the kernels (e.g., 25 parameters for a 5×5 kernel). Pooling consists of the downsampling of the inputs by typically computing averages or determining the maximum of the filtered image. This downsampling step reduces the number of parameters to train and makes the model more robust to variations in the position of the features (i.e., wind patterns here) in the input image. CNNs generally outperform fully-connected networks for both regression and classification tasks owing to their efficient pattern extraction stage (*d'Ascoli et al.*, 2019).

The infrasound refraction process can be seen as the cumulative effect of successive wind heterogeneities, i.e., wind patterns, along the propagation path bending the rays back to the surface (*Chunchuzov et al.*, 2015). CNNs are excellent choices when extracting wind patterns and encoding the nonlinear relationship between wind patterns and ground TLs. We therefore use a CNN architecture by representing each along-path wind model, used as input of PE simulators, as a one-channel 2D image, i.e., gray 2D image, where the x-axis is the source range, the y-axis the altitude, and the wind amplitude the contrast. Since the relationship between frequency and TL for complex wind structures is poorly constrained, we approximate this undefined mapping by using fully-connected layers, which make no assumptions about the input spatial correlations. The final architecture (Fig. 2b) consists of two layers of 2D convolutions using 5×5 kernels (i.e., smallest filters with size 100×15 km) followed by Batch normalization and Average Pooling. The encoded winds are then concatenated with the source frequency input, and three fully-connected layers. Average pooling consists of taking the average of the output of each convolution which is employed to both reduce the dimensionality and learn translation invariance over the input representation. Batch normalization (*Ioffe and Szegedy*, 2015) re-centers and re-scale the input of each layer over each mini-batch during the training process. Normalizing batches reduces the variations of distributions in inputs at each layer, speeds up training, and produces more reliable models. Both Batch normalization and Average Pooling layers are used to make the ML model more robust to new data. The last fully connected layer consists of the output layer that represents the normalized TL profile between 0 to 1000 km. All weights are initialized using a uniform Glorot initializer (*Glorot and Bengio*, 2010).

To facilitate the recognition of patterns in input data, winds are vertically downscaled and horizontally upscaled from a 10×1000 2D image, i.e., 10 profiles discretized over 1000 points along the altitude, to a 50×40 2D image. To limit the range of input and output values, input profiles and output TLs are then normalized by removing the mean and scaled to unit

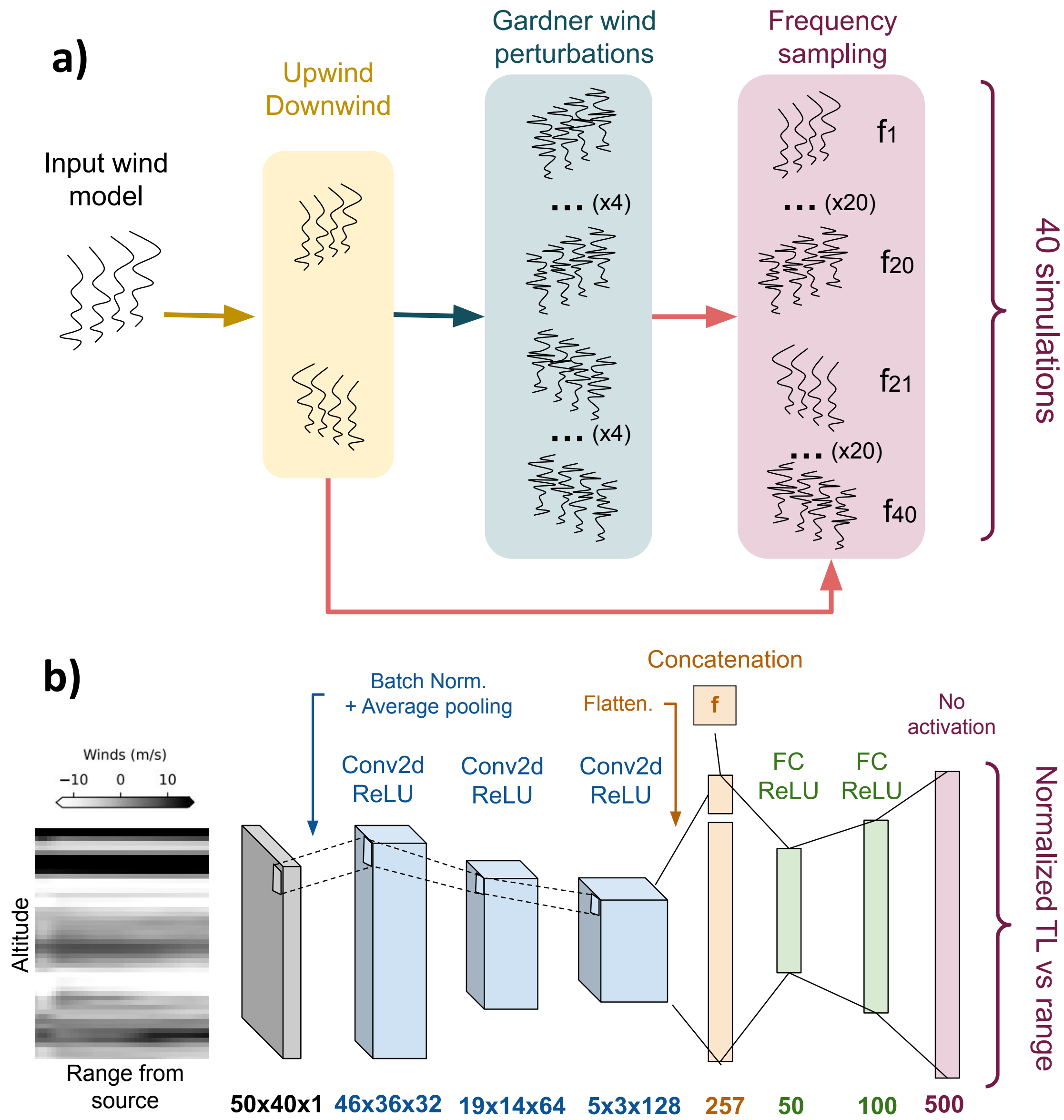

Figure 2. Ground-truth dataset creation and ML architecture. (a) Procedure to augment our atmospheric model dataset. First upwind and downwind scenarios are considered for each wind slice. The difference between upwind and downwind scenarios corresponds simply to flipping the sign of the projected winds onto the slice. Then, 5 random set of Gardner perturbations are generated for both upwind and downwind scenarios. Finally, 4 input frequencies are considered for each perturbed wind model. A total of 40 wind models are generated for each atmospheric slice extracted from the ERA5 dataset. (b) Cartoon depicting a deep learning network workflow for TL predictions. We use 2D representation of wind amplitudes (grey) with size 50×40 as inputs for our ML model. In the first stage (blue) we use three 2d Convolutional layers (Conv2d) to encode the wind information as a vector of size 256. In the second stage (orange), we concatenate this wind encoding with the input source frequency. In the third stage (green), we build a mapping between input frequency and encoded wind representation using two Fully-Connected (FC) layers to finally produce a normalized TL vs range of size 500 (red). This normalized TL can be transformed back to dB by using the

variance. Both mean and variance are computed over the training dataset only. The output layer corresponds to the normalized TL profile linearly interpolated over 500 points within the range 0 to 1000 km. We train the neural network using an Adam optimizer (*Kingma and Ba*, 2015) with a starting learning rate of 10^{-4}. ReLu activation functions are used throughout the network expect for the output layer where we do use any activation function. The ML architecture is implemented in Python using the TensorFlow library (*Abadi et al.*, 2015). More details about architecture optimization are provided in Appendix A.

4. VALIDATION OF MACHINE-LEARNING PREDICTIONS

To optimize our ML model, we split our full dataset between 85% training data and 15% validation data. Strong correlations in TL are expected between PE simulations using wind models corresponding to perturbed versions of the same original unperturbed wind model along a given atmospheric slice. Therefore, before training, all simulations corresponding to the same original atmospheric slice (see the first stage in Fig. 2a) are added to same set (either training or validation) to make our model more robust to new data. To facilitate convergence, we adaptatively update the learning rate when the Root Mean-Square-Error (RMSE) does not decrease over the course of 3 epochs, i.e., training steps. RMSE is computed as $\mathrm{RMSE} = \sqrt{(1/N)\sum_{i=1,N} |\mathrm{TL}^i_{\mathrm{PE}} - \mathrm{TL}^i_{\mathrm{ML}}|^2}$, where $i \in (1, N)$ is the simulation index in the test dataset, N the size of the test dataset, $\mathrm{TL}_{\mathrm{PE}}$ is the TL profile predicted with PE, and $\mathrm{TL}_{\mathrm{ML}}$ is the TL profile predicted with ML. To avoid over-fitting the training data, we use early stopping if the MSE does not decrease over the course of 12 epochs. Finally, to speed up the training process and improve generalization, we use mini-batches of size 32.

We evaluate the performances the ML architecture by training our model over five folds, i.e., five different splits between training and testing datasets. The ML model converges within 65 epochs for our best fold with a validation RMSE (over normalized TL profiles) twice larger than the training RMSE (see Fig. 3a). Once trained, the ML model has a computational cost of around 0.05 s (Dell T5610 Intel Xeon E5-2630 v2 2.6 GHz 6 CPUs 64GB RAM on CentOS 7) for all input frequencies. Over the same frequency range, PE simulation cost increases significantly with frequency, up to 100 s at 3.2 Hz (see Fig. 3b),

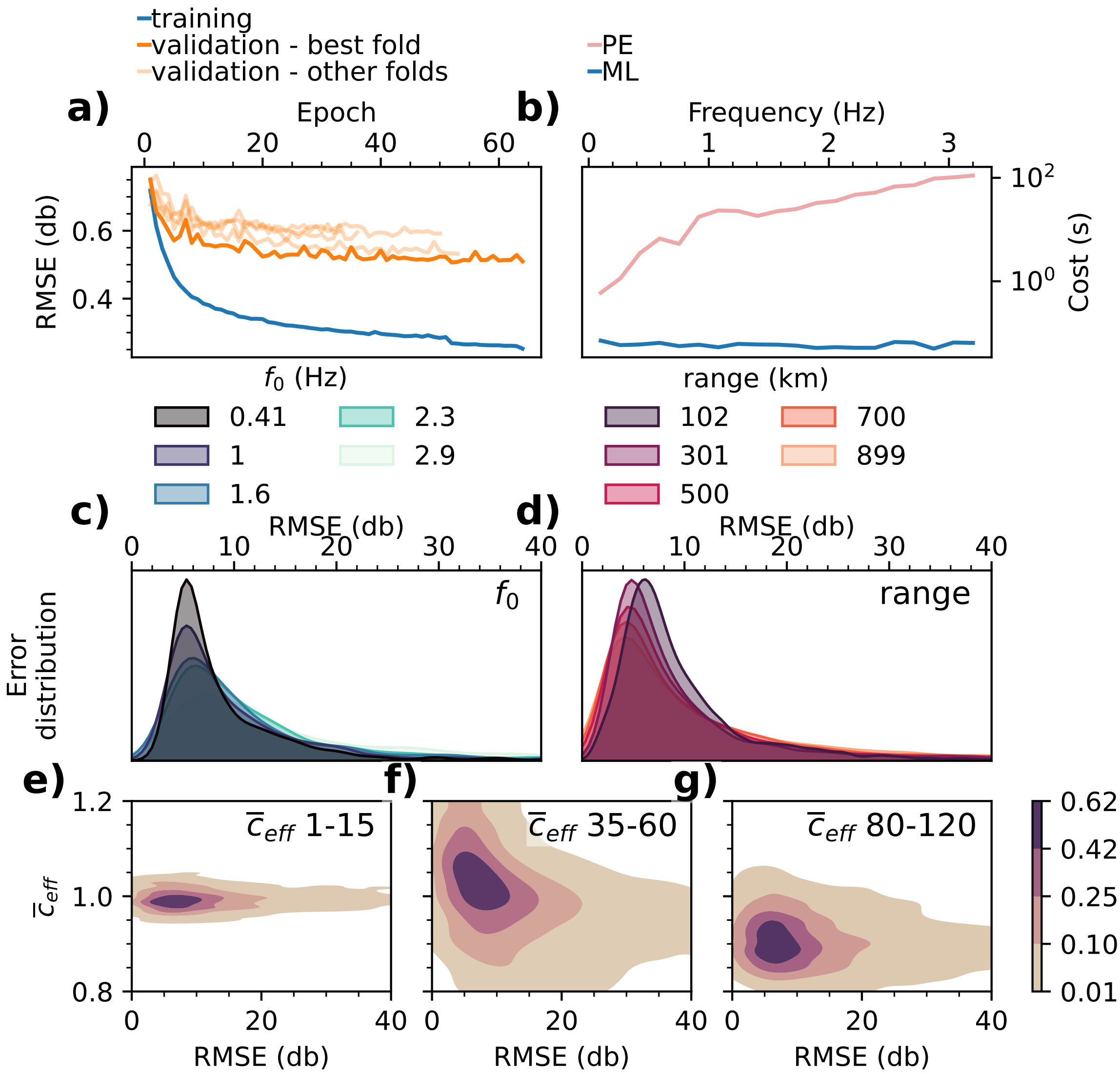

Figure 3. Training and validation of the ML model. (a) Evolution of Root-Mean Square Errors (RMSE) with training epoch for different training (blue) and validation (orange) folds. The fold with best final accuracy is shown as a thick orange line. (b) Computational cost of PE simulations (red) and ML predictions (blue) vs input source frequencies. (c) Distribution of RMSE over the testing dataset for various input frequencies. (d) Distribution of RMSE over the testing dataset for various ranges from the source. (e-g) Distribution of RMSE over the testing dataset for various values of effective velocity ratio $\overline{c}_{eff}$ in (e) the troposphere, (f) the stratosphere, and (g) the thermosphere.

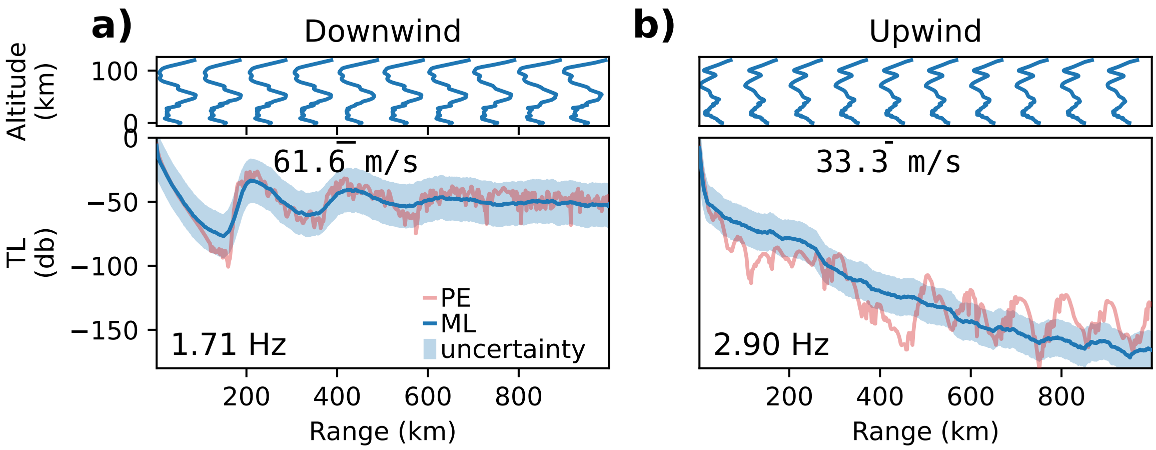

Figure 4. TL predicted by PE simulations (red) and ML model (blue) along with the ML uncertainty (light blue) for a (a) downwind and an (b) upwind scenario. Top, corresponding range-dependent effective velocity models. The ML uncertainty u is computed, in a given frequency range $\mathbf{f}$, as the standard deviation of TL errors vs range from the source over the testing dataset such that $u(r, \mathbf{f}) = \mathrm{std}\{|\mathrm{PE}(r, f) - \mathrm{ML}(r, f)|\}$, where r is the range, f is the frequency, PE is the TL predicted using Parabolic Equations, and ML is the TL predicted using Machine Learning.

which is 2000 times larger than the cost for a ML prediction. In Figs 3c and 3d, we show that the RMSE of our ML model follows a bell-shaped distribution centred between 5 to 9 dB with both variations in distance from the source and source frequency. This distribution of errors indicates that our ML implementation is stable for the range of frequencies and distances considered in our dataset. Larger errors tend to occur for high frequencies (> 2 Hz) and close to the source (< 200 km). Higher frequencies are more sensitive to small-scale wind variations which leads to more complex distributions of TL with range. This added complexity in high-frequency TLs leads to larger errors in ML predictions. Most TL variations occur within 200 km from the source with the presence of the first acoustic shadow zone and first stratospheric return which explains the larger errors observed close to the source. Errors are also stable with variations in effective velocity ratios in different atmospheric layers (Figs 3efg).

We observe in Figs 4a and 4b that ML predictions match well the average variations of TL with range from the source. In particular, the ML model captures accurately the TL gain associated with the different stratospheric returns and the TL asymptotic behaviour at large distances from the source. However, the ML model does not fully reproduce high-frequency TL variations, which owe to vertical and horizontal atmospheric model variations including small-scale changes in effective wind velocities. The ML model therefore provides a low-passed solution of the true TL profile. Our model is unable to learn the entire mapping between atmospheric model heterogeneities and TLs primarily due to both the downsampling of wind profiles and the lack of training data. Yet, large uncertainties are present in currently available atmospheric models, in particular above the troposphere where small-scale wind and temperature perturbations are generally unresolved. Therefore, these high-frequency TL oscillations generally fall within the uncertainty range associated with available atmospheric model resolutions. Along with ML predictions, we can determine an estimate of the ML uncertainty u by computing the standard deviation of TL errors vs range in a given frequency range $\mathbf{f}$, as the standard deviation of TL errors vs range from the source over the testing dataset such that $u(r, \mathbf{f}) = \mathrm{std}\{|\mathrm{PE}(r, f) - \mathrm{ML}(r, f)|\}$, where r is the range, f is the frequency, PE is the TL predicted using Parabolic Equations, and ML is the TL predicted using Machine Learning. The frequency dependence of the uncertainty curves u (see error distribution vs frequency in Fig 3c) is accounted for by computing the errors in five frequency ranges $\mathbf{f}$ equally distributed between 0.1 to 3.2 Hz. We observe that errors between our ML predictions and the PE simulations generally fall within the ML uncertainty range (blue shaded region in Figs 4a and 4b). As suggested by the distributions shown in Figs 3c and 3d, the uncertainty range remains stable with variations in frequency and range from the source.

5. ANALYTICAL VS ML PREDICTIONS OF GROUND TLS

Stratospheric winds are one of the dominant factors to explain the refraction of acoustic waves at large distances from the source (*de Groot-Hedlin et al.*, 2010). A widely used empirical regression equation, introduced in *Le Pichon et al.* (2012), referred in the rest of the paper as LP12, has provided estimates of TL over large distances from a variety of surface sources (*Hernandez et al.*, 2018; *Vorobeva et al.*, 2021; *De Carlo et al.*, 2021). However, the original model was optimized over a set of idealized synthetic and range-independent models where the main feature was a stratospheric duct of various strength, modelled using a Gaussian wind profile centered at 50 km altitude added to the U.S. Standard Atmosphere.

Estimates of LP12 uncertainties over idealized range-independent profiles (*Tailpied et al.*, 2021) show low errors compared to PE simulations (< 10 dB) when strong winds are ducting the signal in the stratosphere. However, in the case of upwind propagation, the accuracy decreases significantly, especially at high frequencies where the errors can be up to 70 dB. Yet, uncertainties introduced by this empirical model for realistic range-dependent wind models are still mostly unconstrained. Comparisons with our PE simulation dataset offer the opportunity to investigate the uncertainties associated with highly heterogeneous wind models for both LP12 and our ML model.

A typical approach to investigate the influence of stratospheric winds on refracted infrasound is to represent the variations of TLs with variations in stratospheric effective velocity ratios, i.e., stratospheric wind strength, and range from the source for different frequencies (*Le Pichon et al.*, 2012). Yet, in contrast to the dataset used for the optimization of LP12, effective velocity ratios in our dataset are not equally distributed since we use the atmospheric model products and not idealized profiles. To provide meaningful comparisons with LP12, we build uniformly-spaced 2D TL maps by performing a linear interpolation of the ML- and PE-predicted TLs between $0.85 \leq \overline{c}_{\mathrm{eff,\ 35–60\ km}} \leq 1.2$, where $\overline{c}_{\mathrm{eff,\ 35–60\ km}}$ is the effective velocity ratio between 35 to 60 km altitude. Linearly-interpolated TL maps are shown in Fig. 5. Comparison between Figs 5a and 5b as well as between Figs 5e and 5f shows that the PE-based TL is well-reproduced by ML for the two frequencies considered. As mentioned earlier, our ML model tends to smooth out the rapid oscillations in TL predicted by PE simulations. Yet, average errors shown in Figs 5c and 5g are stable around 5 dB for all values of $\overline{c}_{\mathrm{eff,\ 35–60\ km}}$.

We also observe that LP12, represented as isocontours in Figs 5b and 5f, is able to capture the main features of the TL maps, namely the first acoustic shadow zone and first stratospheric return within 250 km from the source, and the high attenuation for low stratospheric effective velocity ratios ($\overline{c}_{\mathrm{eff,\ 35–60\ km}} < 1$). The good agreement between numerical simulations and LP12 (Figs 5b and 5f) suggests that average TLs are most sensitive to stratospheric winds when a strong duct is present. LP12 also captures well the high-$\overline{c}_{\mathrm{eff,\ 35–60\ km}}$ trends of median

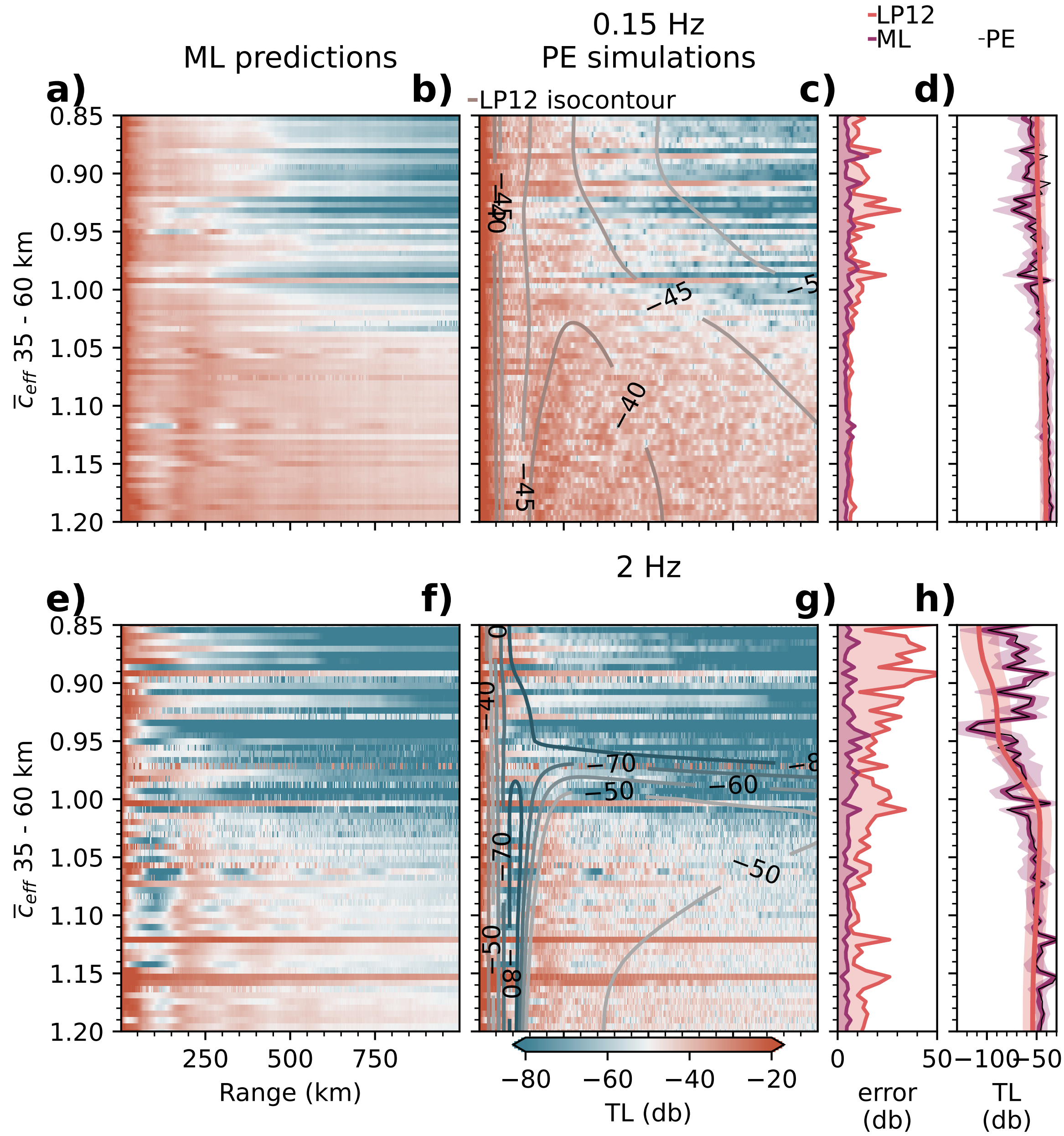

Figure 5. Comparisons of TL maps produced by PE, ML, and LP12 models. (a,b) and (e,f) TL maps vs range and effective velocity ratio $\overline{c}_{\text{eff}}$ between $35-60$ km altitude for a source frequency at 0.15 Hz (a,b) and at 2 Hz (e,f) as predicted by (a,e) the ML model, (b,f) PE simulations, and (b,f isocontours) Le Pichon model. (c,g) RMSE in dB between the interpolated TL maps from the PE simulations and the ML model (purple) and Le Pichon model (LP12, red) at (c) 0.15 Hz and (g) 2 Hz. (d,h) Median TL in dB vs $\overline{c}_{\text{eff, 35–60 km}}$ computed from the interpolated TL maps from the PE (black), the ML (purple), and LP12 (red) models at (f) 0.15 Hz and (h) 2 Hz.

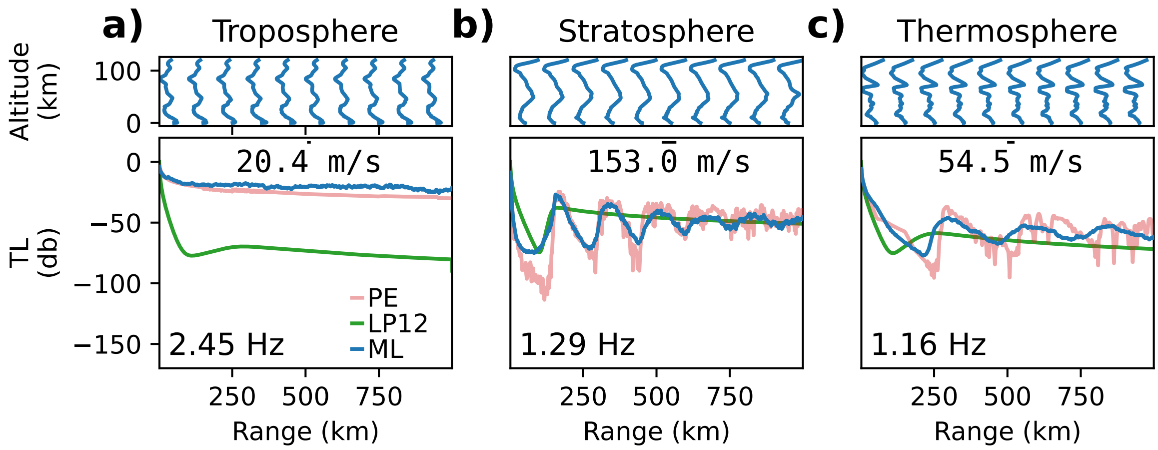

Figure 6. TL predicted by PE simulations (red), LP12 (green), and ML model (blue) for a wind model with a (a) tropospheric duct, (b) stratospheric duct, (c) thermospheric duct, and (d) strong upwind conditions in the stratosphere. (a-d) top, effective velocity profiles used for PE predictions.

TLs (Figs 5d and 5h). However, errors between LP12 and PE simulations increase significantly for low stratospheric effective velocity ratios ($\bar{c}_{\mathrm{eff},\ 35\text{–}60\ \mathrm{km}} < 1$).

LP12 systematically underpredicts TLs for low effective velocity ratios at high frequencies (Fig. 5g), which is consistent with a previous assessment of the empirical model (*Tailpied et al.*, 2021). This owes primarily to the presence of wind ducts outside the stratosphere that are not accounted for in the polynomial parametrization of the empirical model). LP12's errors are particularly strong at high frequencies (*Chunchuzov et al.*, 2015) and close to the source where small wind variations can make acoustic energy return to the ground (*Chunchuzov et al.*, 2015).

The influence of various ducting conditions on ML and LP12 predictions are further illustrated in Figure 6. LP12 captures well the first acoustic shadow zone as well as the asymptotic TL trend at large distance from the source (Figure 6b). However, the error between PE and LP12 increases significantly when a tropospheric or a thermospheric duct are present (Figure 6ac). In particular, tropospheric ducted arrivals generally show strong acoustic amplitudes at ground arrays and can represent up to 20% of the energy radiated from the source (*Drob et al.*, 2003). Accounting for tropospheric ducting is therefore critical for accurate attenuation assessments in the range of distances from the source (< 1000 km) considered here. However, these ducts generally exist only up to a range of $\sim$ 750 km and generally do not affect longer-range propagation at a global scale (*Drob et al.*, 2003). Still, under certain circumstances the tropospheric ducting can be extended as demonstrated, e.g., both in simulations and data recordings at beyond 1600 km range from the Sayarim infrasound calibration experiments (*Fee et al.*, 2013), as well as up to 1000 km from the Antares rocket explosion (*Vergoz et al.*, 2019). For example, a strong tropospheric tailwind direction jet can enhance the tropospheric waveguide.

6. CONCLUSIONS AND DISCUSSION

In this contribution we have proposed an ML-based approach to rapidly ($\sim$ 0.05 s runtime) and reliably ($\sim$ 5 db dB error on average, compared to PE simulations) predict estimates of ground TL from surface sources up to 1000 km. The trained ML model takes as input a range-dependent atmospheric specification and a wave frequency to generate a TL estimate. Errors compared to full PE simulations remain low for increasing source frequency at close range of the source. Our ML model can reproduce complex TL where guided tropospheric waves and multiple stratospheric returns are present. Comparisons with the regression equation introduced in *Le Pichon et al.* (2012) indicate that considering only the influence of stratospheric winds between 35 and 60 km altitude enables one to reproduce the main features of the variations of TL with effective velocity ratio (LP12's errors remain below 10 dB at low frequency for $\overline{c}_{\mathrm{eff}} > 1$). However, by neglecting the impact of tropospheric and high-altitude winds, LP12 can lead to significant errors (RMSE $\sim$ 50 dB) while the ML model is able to capture accurately the TL for highly heterogeneous wind structures.

Several techniques could be used to further improve the accuracy of our ML model. Running additional simulations will increase the size of the training dataset which will reduce the RMSE but will not affect the computational cost of ML predictions once trained. Building on *Raissi et al.* (2019); *Pettit and Wilson* (2020), physical constraints imposed by the PEs and its boundary conditions could be integrated into the cost function to facilitate the convergence of our ML model. Because we trained our algorithm over atmospheric models extracted only from the ERA5 and the NRLMSISE-00/HWM-14 climatological models, biases might be present in the structure of the input wind fields used for training due to the specific system of equations solved to produce ERA5 models. Acquiring atmospheric models from additional sources (e.g., MERRA dataset as presented in *Kumar et al.* (2015)), could make the ML model more robust to arbitrary wind models. In addition to atmospheric models, small-scale gravity-wave models could be enhanced by considering more realistic range-dependent perturbations (*Drob et al.*, 2013; *Lalande and Waxler*, 2016).

Transfer Learning (TrLe) can be used to improve the performances of CNNs over small datasets (*Zhuang et al.*, 2020). CNN parameters are generally initialized using somewhat arbitrary distributions (such as the uniform Glorot initializer (*Glorot and Bengio*, 2010)) that are not tailored to specific classification or regression problems. Because the optimization process is sensitive to the initial parameter distributions (misfits typically show large numbers of local minima), arbitrary distributions do not guarantee convergence. The idea behind TrLe is to exploit invariances in the feature extraction process across different datasets and different tasks (e.g., filters learned to extract edges in dogs vs cats classification can also be used to detect cars) to facilitate the convergenve of the optimization process. TrLe consists of initializing a ML model using the parameters of another ML model pre-trained over a different dataset and possibly for a different task. Here, we tested TrLe by assuming that there are some invariances between our wind feature extraction problem and traditional image-segmentation problems such as multi-class classification of real images (e.g., ImageNet *Deng et al.*, 2009). We tested TrLe by replacing our CNN encoding stage (blue in Fig. 2b) by both a VGG16 (*Simonyan and Zisserman*, 2015) or a ResNet50 (*He et al.*, 2016) network and trained our network using their pre-trained weights and removing pooling layers. However, TrLe's performances were worse (RMSE 9) than with the model presented in Fig. 2b owing to the significant differences between both the set of images used for training in VGG16 or ResNet50 and our wind inputs and the problem of image detection vs TL prediction. .

Our ML model was trained over a set of simulations generated by a PE modelling tool (*Waxler et al.*, 2021) which has strong assumptions about infrasound propagation (see Section 2). In particular, a limitation of PE simulations is in the fact that it ignores cross-winds which might have a strong impact on the acoustic wavefronts at large distances from the source. Yet, this question remains largely unanswered in the literature and further research is needed to provide robust assessments of cross-winds influence on the acoustic wavefield. Nonetheless, ML predictions are expected to be significantly improved if the synthetic dataset were generated using more accurate modelling tools solving the linearized Navier-Stokes equations such as Finite-Differences (FD, *Brissaud et al.* (2016); *Sabatini et al.* (2019)) or Spectral Element Methods (SEM, *Brissaud et al.* (2017); *Martire et al.* (2021)). However, the computational cost associated with such method is much greater than for PE simulations and generating a large synthetic dataset would require extensive computational resources. This cost could be somewhat alleviated since, by resolving the full three-dimensional wavefield, multiple TLs could be extracted from one FD or SEM simulation by considering different azimuths from the source. Once trained over computationally expensive FD or SEM simulations, we can anticipate the cost of one ML simulation to be on the same order than presented here (< 0.1 s) which makes ML even more attractive than when trained over PE simulations. As FD or SEM tools can incorporate topography, an encoded representation of topographic variations (e.g., one-dimensional CNN) could be concatenated to the frequency and encoded winds to provide more accurate predictions.

This work paves the way for the monitoring and characterization of infrasound sources. Recent studies (*Vorobeva et al.*, 2021; *De Carlo et al.*, 2021) have shown that infrasound generated by colliding ocean waves, called microbaroms, may provide important constraints on stratospheric winds. To validate their theoretical model connecting ocean sources and observations, these studies rely on the empirical model presented in *Le Pichon et al.* (2012). Extending the current ML model to longer ranges (> 1000 km) would be critical for global acoustic event analysis, but would also allow an enhanced modelling of microbarom amplitudes, hence also facilitating the development of global infrasound-based near-realtime atmospheric model diagnostics. Similarly, fast and accurate TL predictions would enable the efficient reconstruction of microbarom soundscapes (*den Ouden et al.*, 2021), which would enhance our understanding of global infrasonic background noise levels. The localization of infrasound sources is generally performed using only the arrival times and backazimuth observed at ground arrays and neglects amplitude (e.g., *Blom et al.* (2018)). The absence of amplitude inputs in the optimization process owes to the high computational cost of full-waveform modelling approaches. The inexpensive ML model introduced here could enable the exploration of variations of relative amplitudes between stations with the choice of source location. Computationally inexpensive ML modeling would therefore be a great asset for near-real-time monitoring of natural hazards, such as volcanoes, and explosions for the Comprehensive Nuclear-Test-Ban treaty verification.

Finally, because ML models provide an analytical relationship between input wind models and ground TLs, our ML tool could be used to investigate the sensitivity of infrasound amplitudes with variations in wind models. Sensitivity kernels could be built using explanatory techniques such as Layer-wise Relevance Propagation (*Bach et al.*, 2015) which propagates the ML predictions backwards in the neural network to determine what part of the input data, i.e., wind model, was used to build a given output, i.e., TL. The construction of wind sensitivity kernels could then be employed to further constrain wind structures in infrasound-based wind inversions (*Vera Rodriguez et al.*, 2020). While we restricted our model to absolute TL predictions, i.e., predictions of the norm of the complex TL, both real and imaginary parts of the TL could be independently predicted. Predicting complex TL would enable one to reconstruct the full infrasound time series from any source time function input (e.g., *Arrowsmith et al.* (2012)).

AUTHOR CONTRIBUTIONS

Quentin Brissaud (QB) and Sven Peter Näsholm (SPN) initiated this work and elaborated the plan for the study. QB performed the wave propagation simulations and implemented the ML training and validation. Antoine Turquet (AT) implemented the Gardner's model in Python. Alexis Le Pichon (ALP) generated the LP12 TL profiles (*Le Pichon et al.*, 2012) which are presented in Fig. 5. QB created the figures, which were further elaborated in collaboration with all co-authors. QB wrote the initial manuscript draft and all co-authors contributed in review, revisions, and editing previous to submission.

ACKNOWLEDGMENTS

The authors would like to thank Alexander Binder and Adín Ramírez Rivera, Associate Professors at the Department of Informatics, University of Oslo, for insightful discussion in the implementation and optimization of Convolutional Neural Networks and transfer learning. The authors are also grateful to Claus Hetzer, Research and Development Engineer at University of Mississippi, for providing details about the theoretical background behind the Parabolic Equations modelling software ePape. The authors are thankful for insightful reviews received from Jelle Assink and an anonymous referee. These have helped us significantly improve the manuscript.

This work was mainly funded from a NORSAR institute grant. It was also supported by the project *Middle Atmosphere Dynamics: Exploiting Infrasound Using a Multidisciplinary Approach at High Latitudes* (MADEIRA), funded by the Research Council of Norway basic research programme FRIPRO/FRINATEK under Contract No. 274377. This study was facilitated by previous research performed 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 Nos. 284387 and 653980)

DATA AVAILABILITY STATEMENT

The ERA5 operational data were accessed from the ECMWF MARS archive using the Climate Data Store API (*ECMWF*, 2018), which is accessible to ECMWF Member and Co-operating States. We are grateful to the National Center for Physical Acoustics (NCPA) at the University of Mississippi for making the Parabolic Equation modelling tool ePape publicly available through GitHub at *Waxler et al.* (2021). The TensorFlow library for Python can be downloaded from the TensorFlow repository (`https://doi.org/10.5281/zenodo.4724125`). The ML model Python implementation, and the corresponding PE TL profiles will be released upon publication on a GitHub repository.

REFERENCES

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker,

V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng (2015), TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org.

Arrowsmith, S. J., R. Burlacu, K. Pankow, B. Stump, R. Stead, R. Whitaker, and C. Hayward (2012), A seismoacoustic study of the 2011 january 3 circleville earthquake, *Geophysical Journal International*, *189*(2), 1148–1158, doi:10.1111/j.1365-246X.2012.05420.x.

Averbuch, G., J. D. Assink, and L. G. Evers (2020), Long-range atmospheric infrasound propagation from subsurface sources, *J. Acoust. Soc. Am.*, *147*(2), 1264–1274, doi: 10.1121/10.0000792.

Bach, S., A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek (2015), On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PloS one*, *10*(7), e0130,140, doi:10.1371/journal.pone.0130140.

Blanc, E., L. Ceranna, A. Hauchecorne, A. Charlton-Perez, E. Marchetti, L. Evers, T. Kvaerna, J. Lastovicka, L. Eliasson, N. Crosby, P. Blanc-Benon, A. Le Pichon, N. Brachet, C. Pilger, P. Keckhut, J. Assink, P. M. Smets, C. Lee, J. Kero, T. Sindelarova, N. Kämpfer, R. Rüfenacht, T. Farges, C. Millet, S. Näsholm, S. Gibbons, P. Espy, R. Hibbins, P. Heinrich, M. Ripepe, S. Khaykin, N. Mze, and J. Chum (2018), Toward an improved representation of middle atmospheric dynamics thanks to the ARISE project, *Surveys in Geophysics*, *39*(2), 171–225, doi:10.1007/s10712-017-9444-0.

Blanc, E., K. Pol, A. Le Pichon, A. Hauchecorne, P. Keckhut, G. Baumgarten, J. Hildebrand, J. Höffner, G. Stober, R. Hibbins, P. Espy, M. Rapp, B. Kaifler, L. Ceranna, P. Hupe, J. Hagen, R. Rüfenacht, N. Kämpfer, and P. Smets (2019), Middle atmosphere variability and model uncertainties as investigated in the framework of the arise project, in *Infrasound Monitoring for Atmospheric Studies*, pp. 845–887, Springer.

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

Brissaud, Q., R. Martin, R. F. Garcia, and D. Komatitsch (2016), Finite-difference numerical modelling of gravitoacoustic wave propagation in a windy and attenuating atmosphere, *Geophys. J. Int*, *206*(1), 308–327.

Brissaud, Q., R. Martin, R. F. Garcia, and D. Komatitsch (2017), Hybrid galerkin numerical modelling of elastodynamics and compressible navier–stokes couplings: applications to seismo-gravito acoustic waves, *Geophys. J. Int*, *210*(2), 1047–1069, doi:10.1093/gji/ggx185.

Brissaud, Q., S. Krishnamoorthy, J. M. Jackson, D. C. Bowman, A. Komjathy, J. A. Cutts, Z. Zhan, M. T. Pauken, J. S. Izraelevitz, and G. J. Walsh (2021), The first detection of an earthquake from a balloon using its acoustic signature, *Geophysical Research Letters*, *48*(12), e2021GL093,013, doi:10.1029/2021GL093013.

Ceranna, L., A. Le Pichon, D. Green, and P. Mialle (2009), The buncefield explosion: a benchmark for infrasound analysis across central europe, *Geophysical Journal International*, *177*(2), 491–508, doi:10.1111/j.1365-246X.2008.03998.x.

Chunchuzov, I., and S. Kulichkov (2019), Internal gravity wave perturbations and their impacts on infrasound propagation in the atmosphere, in *Infrasound Monitoring for Atmospheric Studies*, pp. 551–590, Springer, doi:10.1007/978-3-319-75140-5_16.

Chunchuzov, I., S. Kulichkov, V. Perepelkin, O. Popov, P. Firstov, J. Assink, and E. Marchetti (2015), Study of the wind velocity-layered structure in the stratosphere, mesosphere, and lower thermosphere by using infrasound probing of the atmosphere, *Journal of Geophysical Research: Atmospheres*, *120*(17), 8828–8840, doi:10.1002/2015JD023276.

d'Ascoli, S., L. Sagun, G. Biroli, and J. Bruna (2019), Finding the needle in the haystack with convolutions: on the benefits of architectural bias, *Advances in Neural Information Processing Systems*, *32*.

De Carlo, M., P. Hupe, A. Le Pichon, L. Ceranna, and F. Ardhuin (2021), Global microbarom patterns: a first confirmation of the theory for source and propagation, *Geophysical Research Letters*, *48*(3), e2020GL090,163, doi:10.1029/2020GL090163.

de Groot-Hedlin, C. (2008), Finite-difference time-domain synthesis of infrasound propagation through an absorbing atmosphere, *The Journal of the Acoustical Society of America*, *124*(3), 1430–1441, doi:10.1121/1.2959736.

de Groot-Hedlin, C. D., M. A. Hedlin, and D. P. Drob (2010), Atmospheric variability and infrasound monitoring, in *Infrasound Monitoring for Atmospheric Studies*, pp. 475–507, Springer, doi:10.1007/978-1-4020-9508-5_15.

den Ouden, O. F., P. S. Smets, J. D. Assink, and L. G. Evers (2021), A bird's-eye view on ambient infrasonic soundscapes, *Geophysical Research Letters*, *48*(17), e2021GL094,555, doi:10.1029/2021GL094555.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009), Imagenet: A large-scale hierarchical image database, in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, doi:10.1109/CVPR.2009.5206848.

Drob, D. P., J. Picone, and M. Garcés (2003), Global morphology of infrasound propagation, *Journal of Geophysical Research: Atmospheres*, *108*(D21), doi:10.1029/2002JD003307.

Drob, D. P., D. Broutman, M. A. Hedlin, N. W. Winslow, and R. G. Gibson (2013), A method for specifying atmospheric gravity wavefields for long-range infrasound propagation calculations, *Journal of Geophysical Research: Atmospheres*, *118*(10), 3933–3943, doi: https://doi.org/10.1029/2012JD018077.

Drob, D. P., J. T. Emmert, J. W. Meriwether, J. J. Makela, E. Doornbos, M. Conde, G. Hernandez, J. Noto, K. A. Zawdie, S. E. McDonald, J. D. Huba, and J. H. Klenzing (2015), An update to the horizontal wind model (HWM): The quiet time thermosphere, *Earth and Space Science*, *2*(7), 301–319, doi:https://doi.org/10.1002/2014EA000089.

ECMWF (2018), Era5 model level, https://cds.climate.copernicus.eu/.

Evers, L. G., and H. W. Haak (2010), The characteristics of infrasound, its propagation and some early history, in *Infrasound monitoring for atmospheric studies*, pp. 3–27, Springer, doi:10.1007/978-1-4020-9508-5_1.

Fee, D., and R. S. Matoza (2013), An overview of volcano infrasound: From hawaiian to plinian, local to global, *Journal of Volcanology and Geothermal Research*, *249*, 123–139, doi:10.1016/j.jvolgeores.2012.09.002.

Fee, D., R. Waxler, J. Assink, Y. Gitterman, J. Given, J. Coyne, P. Mialle, M. Garces, D. Drob, D. Kleinert, et al. (2013), Overview of the 2009 and 2011 sayarim infrasound calibration experiments, *Journal of Geophysical Research: Atmospheres*, *118*(12), 6122–6143.

Gardner, C. S., C. A. Hostetler, and S. J. Franke (1993), Gravity wave models for the horizontal wave number spectra of atmospheric velocity and density fluctuations, *Journal of Geophysical Research: Atmospheres*, *98*(D1), 1035–1049, doi:10.1029/92JD02051.

Glorot, X., and Y. Bengio (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.

Golden, P., P. Negraru, and J. Howard (2012), Infrasound studies for yield estimation of HE explosions, *Tech. rep.*, Southern Methodist University, Dallas Texas.

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

He, K., X. Zhang, S. Ren, and J. Sun (2016), Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, doi: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016

Head, T., M. Kumar, H. Nahrstaedt, G. Louppe, and I. Shcherbatyi (2021), scikit-optimize/scikit-optimize, doi:10.5281/zenodo.5565057, last accessed on 29 October 2021.

Hernandez, B., A. Le Pichon, J. Vergoz, P. Herry, L. Ceranna, C. Pilger, E. Marchetti, M. Ripepe, and R. Bossu (2018), Estimating the ground-motion distribution of the 2016 mw 6.2 amatrice, italy, earthquake using remote infrasound observations, *Seismol. Res. Lett.*, *89*(6), 2227–2236, doi:10.1785/0220180103.

Ioffe, S., and C. Szegedy (2015), Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, p. 448–456, JMLR.org, doi:10.5555/3045118.3045167.

Kingma, D. P., and J. Ba (2015), Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun, doi: arXiv:1412.6980v9.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012), Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, *25*, 1097–1105, doi:10.1145/3065386.

Kumar, G. K., K. K. Kumar, G. Baumgarten, and G. Ramkumar (2015), Validation of MERRA reanalysis upper-level winds over low latitudes with independent rocket sounding data, *Journal of Atmospheric and Solar-Terrestrial Physics*, *123*, 48–54, doi: 10.1016/j.jastp.2014.12.001.

Lai, V. H., Z. Zhan, Q. Brissaud, O. Sandanbata, and M. S. Miller (2021), Inflation and asymmetric collapse at kīlauea summit during the 2018 eruption from seismic and infrasound analyses, *Journal of Geophysical Research: Solid Earth*, p. e2021JB022139, doi:10.1029/2021JB022139.

Lalande, J.-M., and R. Waxler (2016), The interaction between infrasonic waves and gravity wave perturbations: Application to observations using uttr rocket motor fuel elimination events, *Journal of Geophysical Research: Atmospheres*, *121*(10), 5585–5600, doi: https://doi.org/10.1002/2015JD024527.

Le Pichon, A., L. Ceranna, and J. Vergoz (2012), Incorporating numerical modeling into estimates of the detection capability of the ims infrasound network, *Journal of Geophysical Research: Atmospheres*, *117*(D5), doi:10.1029/2011JD016670.

Martire, L., R. Martin, Q. Brissaud, and R. Garcia (2021), Specfem2d-dg, an open source software modeling mechanical waves in coupled solid-fluid systems: the linearised navier-stokes approach, *Geophysical Journal International*, doi:10.1093/gji/ggab308.

Michalopoulou, Z.-H., P. Gerstoft, B. Kostek, and M. A. Roch (2021), Introduction to the special issue on machine learning in acoustics, *The Journal of the Acoustical Society of America*, *150*(4), 3204–3210, doi:10.1121/10.0006783.

Norris, D., and R. Gibson (2002), Inframap enhancements: environmental/propagation variability and localization accuracy of infrasonic networks, in *Proceedings of the 24th Seismic Research Review—Nuclear Explosion Monitoring: Innovation and Integration*, pp. 809–813.

Pettit, C. L., and D. K. Wilson (2020), A physics-informed neural network for sound propagation in the atmospheric boundary layer, in *Proceedings of Meetings on Acoustics 179ASA*, vol. 42, p. 022002, Acoustical Society of America, doi:10.1121/2.0001383.

Picone, J., A. Hedin, D. P. Drob, and A. Aikin (2002), NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, *Journal of Geophysical Research:*

Space Physics, *107*(A12), SIA–15, doi:10.1029/2002JA009430.

Raissi, M., P. Perdikaris, and G. E. Karniadakis (2019), Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, *378*, 686–707, doi:10.1016/j.jcp.2018.10.045.

Sabatini, R., O. Marsden, C. Bailly, and O. Gainville (2019), Three-dimensional direct numerical simulation of infrasound propagation in the earth's atmosphere, *Journal of Fluid Mechanics*, *859*, 754–789, doi:10.1017/jfm.2018.816.

Shani-Kadmiel, S., G. Averbuch, P. Smets, J. Assink, and L. Evers (2021), The 2010 haiti earthquake revisited: An acoustic intensity map from remote atmospheric infrasound observations, *Earth and Planetary Science Letters*, *560*, 116,795, doi:10.1016/j.epsl.2021.116795.

Simonyan, K., and A. Zisserman (2015), Very deep convolutional networks for large-scale image recognition, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun.

Sutherland, L. C., and H. E. Bass (2004), Atmospheric absorption in the atmosphere up to 160 km, *The Journal of the Acoustical Society of America*, *115*(3), 1012–1032, doi:10.1029/2006JD007806.

Tailpied, D., A. L. Pichon, and B. Taisne (2021), Assessing uncertainties in infrasound network performance modelling: application to the Euro-Mediterranean and Southeast Asian region, *Geophysical Journal International*, doi:10.1093/gji/ggab399, ggab399.

Vera Rodriguez, I., S. P. Näsholm, and A. Le Pichon (2020), Atmospheric wind and temperature profiles inversion using infrasound: an ensemble model context, *J. Acoust. Soc. Am.*, *148*(5), 2923–2934, doi:10.1121/10.0002482.

Vergoz, J., A. L. Pichon, and C. Millet (2019), The Antares explosion observed by the USArray: An unprecedented collection of infrasound phases recorded from the same event, in *Infrasound monitoring for atmospheric studies*, pp. 349–386, Springer.

Vorobeva, E., M. D. Carlo, A. L. Pichon, P. J. Espy, and S. P. Näsholm (2021), Benchmarking microbarom radiation and propagation model against infrasound recordings: a vespagram-based approach, *Annales Geophysicae*, *39*, 515–531, doi:10.5194/angeo-39-515-2021.

Waxler, R., C. Hetzer, J. Assink, and D. Velea (2021), chetzer-ncpa/ncpaprop-release: Ncpaprop v2.1.0, doi:10.5281/zenodo.5562713, last accessed on 29 October 2021.

Zhuang, F., Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He (2020), A comprehensive survey on transfer learning, *Proceedings of the IEEE*, *109*(1), 43–76, doi: 10.1109/JPROC.2020.3004555.

Appendix A: Hyper-parameter optimization

The ML model is described by a set of hyper-parameters that must be optimized in order to obtain the best regression performance. First, we optimized the ML architecture, i.e., the number of CNN and dense layers as well as number of CNN filters, using a Bayesian optimization with Gaussian Processes as implemented in the scikit-optimize Python library (*Head et al.*, 2021). In addition to architecture optimizations, we investigated the variations in RMSE with the choice of training parameters (batch size and validation dataset size) as well as inputs image size. Such variations are shown in Fig. 7. There are generally negligible error differences between each model. As a trade-off between training time and error we choose batches of size 32, a dataset of size 20%, and input images of size 20×4.

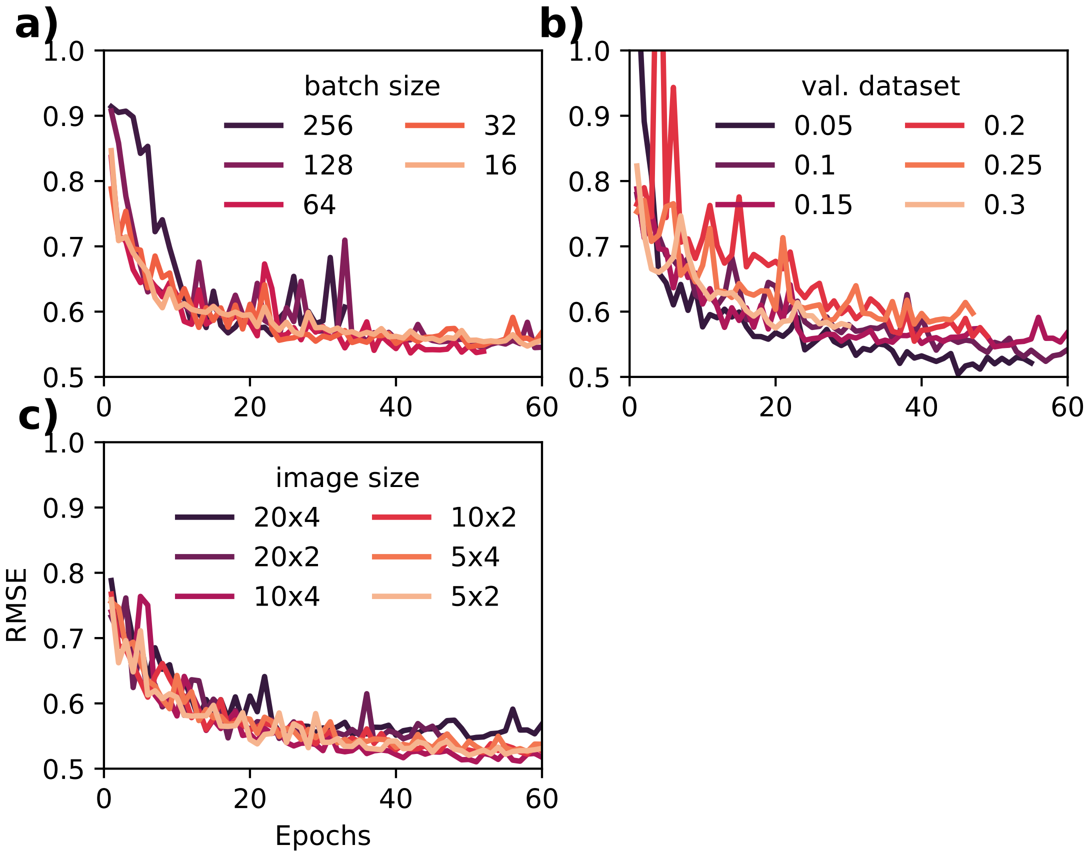

Figure 7. Optimization of training and input hyperparameters. RMSE vs epochs during training for variations in (a) batch size, (b) validation dataset size, and (c) input image size from a baseline model with: batch size 32, 15% validation dataset size, and 20×4 input size.