Machine learning and marsquakes: a tool to predict atmospheric-seismic noise for the NASA InSight mission

A E Stott, R F Garcia, A Chédozeau, A Spiga, N Murdoch, B Pinot, D Mimoun, C Charalambous, A Horleston, S D King, et al.

Geophysical Journal International

Geophys. J. Int. (2023) **233,** 978–998
Advance Access publication 2022 November 25
GJI Seismology

https://doi.org/10.1093/gji/ggac464

Machine learning and marsquakes: a tool to predict atmospheric-seismic noise for the NASA InSight mission

A.E. Stott [1], R.F. Garcia,[1] A. Chédozeau,[1] A. Spiga,[2] N. Murdoch,[1] B. Pinot,[1] D. Mimoun,[1] C. Charalambous,[3] A. Horleston,[4] S.D. King,[5] T. Kawamura,[6] N. Dahmen,[7] S. Barkaoui,[6] P. Lognonné[6] and W. B. Banerdt [8]

[1] *Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO),* 31400 *Toulouse, France. E-mail: alexander.stott@isae-supaero.fr*
[2] *Laboratoire de Météorologie Dynamique/IPSL, Sorbonne Université, CNRS, Ecole Normale Supérieure, PSL Research University, Ecole Polytechnique,* 75005 *Paris, France*
[3] *Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, UK*
[4] *School of Earth Sciences, University of Bristol, BS*8 1*RJ Bristol, UK*
[5] *Department of Geosciences, Virginia Tech, Blacksburg,* 24061 *VA, USA*
[6] *Institut de Physique du globe de Paris, Université de Paris, CNRS,* 75005 *Paris, France*
[7] *Institute of Geophysics, ETH Zurich,* 8092 *Zurich, Switzerland*
[8] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena,* 91109 *CA, USA*

Accepted 2022 November 7. Received 2022 October 7; in original form 2022 April 19

SUMMARY

The SEIS (seismic experiment for the interior structure of Mars) experiment on the NASA InSight mission has catalogued hundreds of marsquakes so far. However, the detectability of these events is controlled by the weather which generates noise on the seismometer. This affects the catalogue on both diurnal and seasonal scales. We propose to use machine learning methods to fit the wind, pressure and temperature data to the seismic energy recorded in the 0.4–1 and 2.2–2.6 Hz bandwidths to examine low- (LF) and high-frequency (HF) seismic event categories respectively. We implement Gaussian process regression and neural network models for this task. This approach provides the relationship between the atmospheric state and seismic energy. The obtained seismic energy estimate is used to calculate signal-to-noise ratios (SNR) of marsquakes for multiple bandwidths. We can then demonstrate the presence of LF energy above the noise level during several events predominantly categorized as HF, suggesting a continuum in event spectra distribution across the marsquake types. We introduce an algorithm to detect marsquakes based on the subtraction of the predicted noise from the observed data. This algorithm finds 39 previously undetected marsquakes, with another 40 possible candidates. Furthermore, an analysis of the detection algorithm's variable threshold provides an empirical estimate of marsquake detectivity. This suggests that events producing the largest signal on the seismometer would be seen almost all the time, the median size signal event 45–50 per cent of the time and smallest signal events 5−20 per cent of the time.

Key words: Planetary interiors; Seismic noise; Planetary seismology.

1 INTRODUCTION

The NASA InSight lander has recorded seismic and meteorological data for over three Earth years since arriving on Mars on the 26th of November 2018. This seismic data contain several important marsquakes (InSight Marsquake Service 2022) which have enabled Stähler *et al.* (2021), Khan *et al.* (2021) and Knapmeyer-Endrun *et al.* (2021) to place constraints on the core, mantle and crust of Mars using seismological tools. The task of identifying marsquakes falls on the Marsquake Service (MQS) who systematically search the data as it is downlinked to identify marsquakes, as outlined in Clinton *et al.* (2021) and Ceylan *et al.* (2021). The first quake in the MQS catalogue is from sol 105, where a sol is a Mars day and sol 0 is the landing day. Since then there have been 951 quakes in the almost thousand sols of the mission so far. However, between sols 500 and 800 only 27 of these 951 events were observed. This is largely due to the environmental noise injection into the seismic data. The period corresponds to the Northern Hemisphere winter and is marked by the persistent presence of local turbulence throughout the Martian sol, as characterized by Chatain *et al.* (2021). Moreover, even during

the summer sols marsquakes are predominantly identified only in the quiet periods just after sunset when the weather conditions are extremely calm.

This demonstrates that atmospheric injection inhibits the discrimination of data of a seismic origin. In doing so, a complete catalogue of Martian events cannot be collated. This means events which would add further information on the Martian structure are missed and further constraints cannot be placed on the seismicity of Mars. Moreover, autocorrelation techniques have been implemented to examine the Martian structure. These approaches are either based on the coda of marsquakes or on identifying the ambient diffuse seismic wavefield (Compaire *et al.* 2021; Schimmel *et al.* 2021). As a result, the separation of the seismometer data into atmospherically induced and seismic origin components is of interest.

To limit atmospheric injection into the seismic data, the SEIS (seismic experiment for the interior structure of Mars) instrument was placed on the ground away from the lander and covered with a wind and thermal shield (WTS, Lognonné *et al.* 2019). Stott *et al.* (2021) verified that the WTS removed the direct forcing pathway on SEIS and so reduced wind-induced signals. However, a problematic injection of atmospherically generate noise into the seismic record persists. The impact of wind, pressure and temperature on the seismic data was predicted in the pre-landing noise model of Mimoun *et al.* (2017). At the lowest frequencies $<10^{-3}$ Hz temperature effects dominate (Lognonné *et al.* 2020). Garcia *et al.* (2020) showed pressure coupling through compliance for certain periods of the sol below 0.5 Hz. For marsquakes however, wind noise dominates, with the lander providing a key source of vibrations as outlined in Murdoch *et al.* (2017, 2018) and Stott *et al.* (2021). The *in situ* sensitivity of the seismic data to the wind and pressure variation was examined in Charalambous *et al.* (2021), which described the relationship between the seismic energy level (i.e. envelope of the seismic data), wind speed and pressure fluctuations.

Atmospherically generated noise, therefore, forms the most significant component of seismic noise on Mars. Machine learning has been applied to separate such noise features within seismic data on Earth, for example, in Johnson *et al.* (2020) and Seydoux *et al.* (2020). These unsupervised learning approaches aim to cluster data segments with common attributes. The deep scattering network approach in Seydoux *et al.* (2020) was implemented for the InSight data in Barkaoui *et al.* (2021), which identified types of glitches (Scholz *et al.* 2020) and atmospheric noise. In planetary scenarios, machine learning approaches have been applied for event detection on the Moon. Knapmeyer-Endrun & Hammer (2015) and Civilini *et al.* (2021), respectively, developed a hidden Markov model and convolutional neural network approach to detect patterns related to moonquakes. These are a supervised learning classification style problem.

The aim of this work is to analyse the seismic data, and therefore marsquakes, with respect to the atmospheric contribution. To that end, we formulate a supervised learning regression style problem, to predict the seismic energy from the atmospheric data. This integrates the wind speed, wind direction, pressure and temperature data to allow for information on the variability in the relationship due to atmospheric conditions to be taken into account. We implement two types of machine learning algorithms for this curve fitting task, Gaussian process (GP) regression and a multilayer perceptron (MLP) neural network (NN). The resulting machine learning models produce an estimation of the atmospheric noise level in the SEIS data. This prediction can be used as a tool to analyse known marsquakes and identify new ones.

2 DATA AND PRE-PROCESSING

2.1 The seismic data

The InSight events are catalogued by the MQS according to a developed taxonomy. The event categories (as described in Clinton *et al.* 2021) are split as:

(i) Low frequency (LF)—energy below 2.4 Hz on all three components.
(ii) Broad band (BB)–energy predominantly below 2.4 Hz on all three components with some excitation above and including the 2.4 Hz mode.
(iii) 2.4 Hz—energy in all three components localized to the 2.4 Hz mode.
(iv) High frequency (HF)—energy in all three components predominantly at 2.4 Hz and above.
(v) Very high frequency (VF)—energy in all three components predominantly at 2.4 Hz and above but the horizontal energy is greater than the vertical for higher frequencies. They have a particular spectral characteristic.
(vi) Super high frequency (SF)—very short duration, exciting energy above 2.4 Hz with horizontal energy larger than vertical.

The 2.4-Hz mode is proposed to be a resonance of geophysical origin and is excited by seismic events at these frequencies (Dahmen *et al.* 2021b; Giardini *et al.* 2020; Hobiger *et al.* 2021; Compaire *et al.* 2021). On top of event excitation it is also observed ambiently (Dahmen *et al.* 2021b; Compaire *et al.* 2021) with a variable amplitude when the noise level is low enough.

For our purposes, we can split them into a lower frequency (LF and BB events) and higher frequency category (2.4 Hz, HF and VF events). In this work, we do not examine the SF events as these are already detected by an automated algorithm (Dahmen *et al.* 2021a). On top of these categories, the events are split into the following classes:

(i) A—multiple clear phases with clear polarization.
(ii) B—multiple clear phases with no polarization or with polarization but not clearly identifiable phases.
(iii) C—signal is clear but clear phase picking is difficult.
(iv) D—signal is very weak, impossible to pick phases or may not be a seismic event.

As described in Clinton *et al.* (2021). The aim of this class system is to indicate their usability for seismic analysis, as much as their general quality.

Fig. 1 shows the probabilistic power spectral density (PPSD) of the VBB (the very broad-band seismometer on SEIS) data taken over sols 257–267. This is the histogram of the PSD over this period and is normalized to give the probability of the PSD taking that value for a given frequency. To examine the lower frequency group we use the velocity channel filtered between 0.4–1 Hz. This region has a relatively flat/white spectrum as shown in Fig. 1.

For the higher frequency category we focus on the 2.2–2.6 Hz range to isolate the energy in the 2.4 Hz mode, which is always excited by events in this group. Isolating this mode means that we do not have to be concerned with the increasing sensitivity to wind with increasing frequency, shown in Fig. 1.

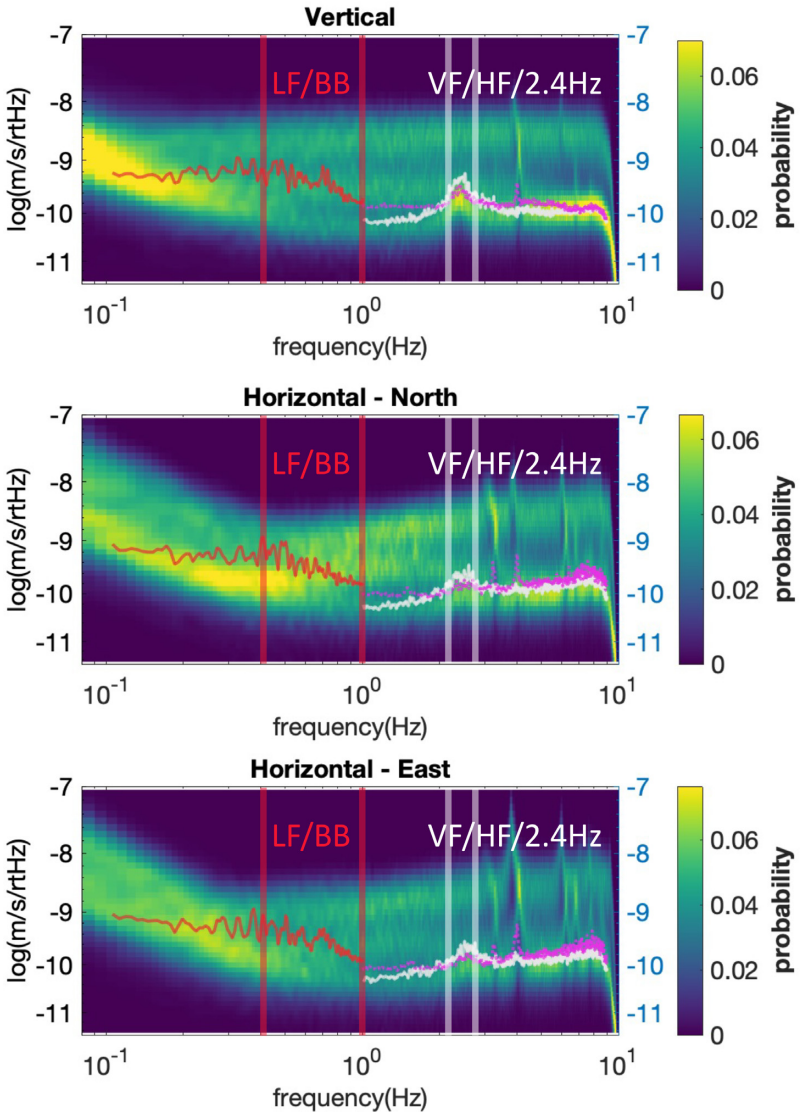

Figure 1. The PPSD of the (top to bottom) vertical, north and east components of the VBB over sols 257–267. The red PSD overlay is the LF S0173a event, the white is the HF S0424c and the magenta is the VF S0475a event. The red lines indicate the 0.4–1 Hz band used for the LF envelope and the white lines the 2.2–2.6 Hz bandwidth for the HF envelope.

2.2 Pre-processing and data cleaning

Based on the above considerations for how atmospheric energy can generate seismic signals, the data used for the machine learning models are:

(i) seismic energy level of the VBB Z, N and E axes in the 0.4–1 and 2.2–2.6 Hz bandwidths);
(ii) wind speed boom 1 and 2;
(iii) wind direction boom 1 and 2;
(iv) air temperature from boom 1 and 2;
(v) pressure and its envelope (in the 0.1–4 Hz frequency band).

Note boom 1 and 2 refers to the two different wind sensors of TWINS, which each have complementary orientations designed to cover wind from all directions (Banfield *et al.* 2019). The machine learning model therefore has eight inputs (2 wind speeds, 2 wind directions, 2 air temperatures, the static pressure and the RMS envelope of the pressure signal) and three outputs (the seismic energy level, envelope, for Z, N and E).

Prior to the implementation of machine learning algorithms, the InSight data must be pre-processed and cleaned. InSight data are available (both seismic and meteorological) in miniseed (Service 2019). We use the ObsPy Python package (see Beyreuther *et al.* 2010; Megies *et al.* 2011; Krischer *et al.* 2015) for processing the data.

There are several data issues and peculiarities that need to be taken into account. For example, the available data for each sensor over the mission are not always at the same data rate, due to data downlink volume considerations. Moreover, there are periods of data dropout on some sensors either from anomalies or, more recently, power considerations when some sensors had to be turned off. The pre-processing must blend together this data taken at different sampling rates and identify the periods when data from all sensors was available. The data channels used are outlined in Table A1.

The first step is to identify the sols which have coincident data from all sensors, that is, find the sols where there is at least one of the channels in Table A1 for each sensor. Once this has been obtained, the seismic data pipeline is to:

(i) load data;
(ii) remove the sensor response;
(iii) rotate to the vertical (Z), north (N) and east (E) coordinate frame;
(iv) filter to 0.4–1 Hz and 2.2–2.6 Hz bandwidth;
(v) compute the signal envelope over 10s;
(vi) downsample to 1 sample per second (sps).

The pressure data are similarly filtered to obtain the envelope and both the envelope and original data are downsampled to 1 sps. The wind speed, direction and air temperature are also converted to the common sample rate of 1 sps, which is chosen as it is the most common sample rate for this data. Each data vector is saved sol by sol so that the processing can be done in batches and is readily extended to new data.

After this initial treatment, several artefacts must be removed from the data. These include:

(i) large erroneous peaks;
(ii) other peaks due to glitches and other transients;
(iii) data gaps.

The glitches are a common problem to tackle with the InSight seismic data and most are removed through the ISAE deglitching technique proposed in Scholz *et al.* (2020). However, some remain which appear as spikes in the envelope. A further processing step is made to remove these peaks significantly above the local median value of the data within a window. The data gaps are identified from the miniseed files. The method to discount these periods of data is through a mask vector, which is a Boolean logical inclusion or exclusion of the data at that point. This makes it straightforward to modify without deletion and manipulate the data, in terms of keeping the original time sample.

The next step is to apply the transformations for input into the machine learning algorithms. The seismic envelope is converted to its base 10 logarithm to give the seismic energy level and, along with the pressure envelope, is smoothed with a moving median and mean filter each of length 15 s. The median filter is applied first to remove excess spikes and then smoothed by the mean filter. Each variable is then normalized to have an amplitude between −1 and 1. The final data are stored along with the mask vectors for each sol. After this full pre-processing, ∼30 million samples are available (at 1 sps up to sol 1000) for use with the machine learning algorithms.

In order to handle the large memory and computational requirement for both the pre-processing and training machine learning models the code was run on the CALMIP computing cluster facility. The choice to operate on the data one sol at a time when possible enabled parallelization of the code for improved calculation time and reduced the need for very large matrices to be stored or operated on.

3 ATMOSPHERICALLY GENERATED NOISE

3.1 Theoretical perspectives

Prior to landing, Mimoun *et al.* (2017) constructed a noise model for SEIS to determine the level of injection from each possible noise source. In the bandwidths where seismic events are observed, the main effect is due to the wind. The forcing mechanism is given by drag/lift

$$F = \frac{1}{2} C_{D/L} A \rho_A U^2 \tag{1}$$

where $C_{D/L}$ and A are the drag/lift coefficient and surface area of whichever body the wind forces, ρ_A is the air density and U the wind speed.

The pathways through which this will generate vibrations sensed by SEIS are either by forcing on the lander or on the WTS, whereby the vibrations are coupled through the regolith to SEIS feet. Murdoch *et al.* (2017) developed a physical model for lander-induced vibrations into the SEIS data which was integrated into the overall noise model of Mimoun *et al.* (2017). Stott *et al.* (2021) verified that lander-induced vibrations are the dominant noise source at least for high wind speeds. The lander and SEIS assembly also have vibrational modes excited by the wind which were modelled in Murdoch *et al.* (2018) and analysed in Dahmen *et al.* (2021b) and Hurst *et al.* (2021).

On top of wind-induced sources, pressure forcing can cause ground deformation (Garcia *et al.* 2020; Kenda *et al.* 2020). Moreover, Charalambous *et al.* (2021) demonstrated that the pressure envelope between 0.1 and 4 Hz is well correlated to the seismic energy level. The VBB is also directly sensitive to temperature (Mimoun *et al.* 2017), however, this does not fluctuate on the scale of events so we only consider the atmospheric temperature to take into account changes of air density. The magnetic field variations so far have not been seen to be very well correlated with the seismic data in event bandwidths and so are not included.

3.2 The observed variation of the atmospheric data with the seismic data

The input–output relationships for the pre-processed data are shown in Fig. 2 for three periods (of three sols) over the mission centred around sol 258, 495 and 690. The colour coding indicates the time of the sol where dark colours indicate night, turning progressively lighter through blue during the morning and then darker through red in the afternoon to the night. It can be seen that the relationship between the seismic data and each of these variables changes over both the Martian sol and season, as was demonstrated in Charalambous *et al.* (2021). This is straightforwardly observed by the variation with pressure in Fig. 2 which covers a different range of pressures for each section.

The plots on the diagonals of the scatter plot matrix in Fig. 2 for the three seismic components show the KDE (kernel density estimate) of the probability density function (pdf) for each of the three different periods of data. For the sol 258 section, there are three modes and a wide range of values covered. In the sol 495 section, the range narrows and there are two modes before a single mode is shown in the sol 690 section. This is matched by the relationships to each of the atmospheric data, which generally appear to collapse and become less structured as the conditions are more similar over the entire sol. However, different regimes are still observable. For example, in the sol 690 section the dark nighttime zone in the wind speed boom 2 and pressure envelope is still distinct from the other times of sol. The slope for this nighttime period is still greater than for the other times of sol but shallower than the same coloured zone in the other seasons.

In short, the goal of the machine learning is to fit the curves presented in Fig. 2, accounting for both the highlighted diurnal and

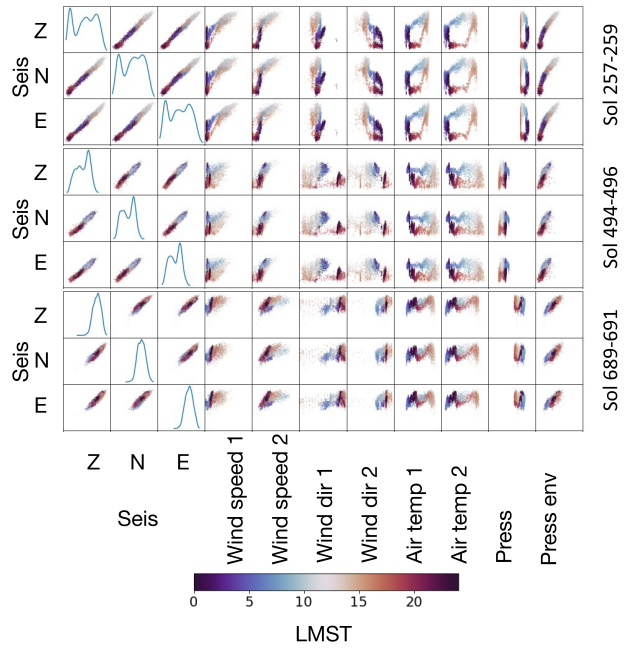

Figure 2. A matrix of scatter plots between the atmospheric data (model inputs) and the three axis seismic noise levels (model outputs). This is plotted for 3 sols at three different seasons over the mission. The diagonal is the KDE of the pdf for the seismic data. The colour indicates the time of the sol starting with black at midnight LMST (local mean solar time) going to lighter blue to midday before turning darker red over the afternoon.

seasonal variation. Some cause of this variability can be understood through consideration of the drag/lift eq. (1) as outlined above. First of all, this is dependent on the air density which is, in turn, dependent on pressure and air temperature. Moreover, drag/lift coefficient is dependent on the atmospheric conditions, for example, whether the wind is in a turbulent or laminar flow. As a result, we must consider the atmospheric stability at InSight over the Martian day and seasons, presented in Spiga *et al.* (2021), Banfield *et al.* (2020) and Chatain *et al.* (2021).

During the daytime on Mars there is strongly convective turbulence. This provides the highest wind speeds and strongest overall seismic response. On the other hand, the Martian atmosphere is often more stably stratified at night leading to laminar flows and when night time turbulence occurs it is generally weaker. Although the night time signals are not usually as strong as the daytime, the seismic response is more sensitive yielding a higher velocity for a given wind speed than during the day. This is shown in Fig. 2 where the darker colours at night generally show a steeper slope to the wind speed than the lighter daytime colours. There can also be a quiet period just after sunset when the planetary boundary layer collapses and there is little wind speed. During this period there is very little wind or turbulence and so this period is where most seismic events are found (Clinton *et al.* 2021).

Chatain *et al.* (2021) investigated the seasonal variability in turbulence across the Martian sol at InSight. They showed that a buoyancy source is the dominant factor for the daytime turbulence whereas the evening/night time turbulence is driven by wind shear. Earlier in the mission, the stable atmosphere at night also creates a buoyancy sink which can inhibit the shear source, causing a turbulent flow to move towards a laminar one. As the Northern Hemisphere winter approached, the wind shear source begins to dominate the buoyancy sink, leading to increased evening/night time turbulence and the quiet period completely disappears.

The seismic response is therefore understandable in terms of the physical relationship, however, this also illuminates that there are many components to take into consideration. As a result, it is difficult to determine a physical model which is able to predict the seismic signal sufficiently for event analysis. To that end, we implement a machine learning model to account for the complexities outlined here. This also allows other possible sources/forcing to play a part in the prediction, which may not be immediately identifiable.

4 IMPLEMENTATION OF THE MACHINE LEARNING MODELS

The next step is to implement machine learning models to fit the input-output relationships demonstrated above in Fig. 2. The atmospheric variables (the 2 wind speeds, directions, air temperatures, pressure and pressure envelope) are the inputs and the envelope of the three seismometer axes are the outputs. We consider two main methods to do this, GP regression and NNs. At a base level each of these approaches can be considered equivalent implementations to obtain the prediction. We will first present their individual applications to the data and then compare their performances.

4.1 Gaussian process regression

4.1.1 Background

A Gaussian process (GP), as defined in Williams & Rasmussen (2006), is a set of random variables where any finite selection has a joint Gaussian distribution. GP regression applies this probabilistic approach around the regression model

$$y = f(\mathbf{x}) + \eta \tag{2}$$

where the output of the model y is predicted from the input vector $\mathbf{x}$ through the function $f(\cdot)$ plus some white noise $\eta \sim \mathcal{N}(0, \sigma^2)$. The function $f(\cdot)$ is considered to be Gaussianly distributed itself, that is, we consider the distribution of functions that fit the data. In this way GP regression is considered non-parametric, as opposed to having a specific form defined by parameters, for example, a quadratic function.

In the GP framework the distribution of the function is given as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{3}$$

where $m(\mathbf{x})$ and $k(\mathbf{x})$ are termed the mean and covariance function of the output $f(\mathbf{x})$. These functions are selected depending on the characteristics of the problem at hand. The mean function can be used to encode known functional information directly. The covariance function sets the prior form for the functions that fit the data, that is, it determines the characteristics of the functions that fit the data. The covariance function must be positive-semi-definite and is described by a chosen kernel function to give the level of similarity between the outputs for two arbitrary inputs $\mathbf{x}$ and $\mathbf{x}'$.

From the definition, any set of samples from a GP has a multivariate normal distribution. So for a set of inputs, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denoted as X the vector of function outputs $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)]$ has the distribution

$$\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K}) \tag{4}$$

where $\mathbf{m} = [m(\mathbf{x}_1), \ldots, m(\mathbf{x}_n)]$ and the matrix $\mathbf{K}$ has entries $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. For simplicity the mean is herein set to be zero. Now

consider the set of inputs X along with the corresponding outputs $\mathbf{y}$. To train the GP regression model we must obtain the conditional distribution (the posterior) as

$$\mathbf{f}|X, \mathbf{y} \tag{5}$$

For a set of new, test inputs, denoted as X^*, the prediction from the GP regression can be inferred from the conditional distribution

$$\mathbf{f}|X, \mathbf{y}X^* \tag{6}$$

where the joint distribution of the vectors $\mathbf{f}$ and $\mathbf{f}^*$ follows the multivariate normal distribution. This can be evaluated assuming the white Gaussian noise term η. Note that this is a distribution and so the prediction is taken as the expectation. The covariance can also be extracted and so the evaluation of prediction confidence intervals is straightforward.

The performance of a GP regression model depends on the selection of the mean and covariance functions, that is, the prior distribution of the function. The mean function is usually set to zero. The covariance function is often set by a kernel.

The behaviour kernel and mean functions are controlled by hyperparameters. For example, a common kernel function is the squared exponential

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2l^2}) \tag{7}$$

where the lengthscale hyperparameter l defines the level of similarity between two points at a distance $|\mathbf{x} - \mathbf{x}'|$. To evaluate the appropriate hyperparameter value for a given problem an optimization is performed to maximize the log marginal likelihood of the training data given as

$$\log(P(\mathbf{y}|X)) \tag{8}$$

In this work, we use the GPy library for Python (GPy since 2012) to implement GP regression models. The prediction requires matrix inversion and so has high computational complexity. This makes them more appropriate for small data sets.

4.1.2 Implementation

A global model was fitted using 3000 data samples for the training data, randomly selected from across the entire available data. The time separation of these was checked and found that there is no gap longer than 2 sols between consecutive samples confirming a representative sampling across the mission. The prediction of this model was then calculated for each sol separately and stored. Four models with different kernels were trained on the LF data to identify the best choice. These were (1) an Exponential kernel, (2) radial basis function (RBF), (3) Exponential added to an MLP kernel and (4) Exponential added to an MLP kernel multiplied by another exponential kernel. For further details on these kernels, see Williams & Rasmussen (2006).

To evaluate model performance we use the root mean square error (RMSE) metric given by

$$\mathrm{RMSE} = \sqrt{1/N \sum_{n=0}^{N-1} (y_n - \hat{y}_n)^2}$$

where N is the number of data points, y_n and $\hat{y}_n$ are the observed and predicted values at time instance n. The RMSE was calculated on 20 per cent of the whole mission's data (test data not used for training) for each model and is shown in Table 1. The Exponential + MLP kernel marginally the lowest error and so was selected for use in the rest of this study.

Table 1. RMSE of GP global model for different kernels.

GP Kernel type	Test RMSE
Exp.	0.075
RBF	0.078
Exp. + MLP	0.074
Exp. + MLP × Exp.	0.075

This global model used only 3000 samples to calculate the fit, which is a very small portion of the total ∼ 30 million available. This is because GP regression requires significant computational and memory resources in optimization and prediction. As a result, using a larger portion of the data for training quickly becomes intractable. This global model provides a good estimate of the relationship dynamics, however, it may not sufficiently represent some parts of the atmospheric–seismic relationship for marsquake analysis. To this end, we also propose a local model where a GP regression model is obtained independently for each sol. The training data for this model are 3000 samples randomly selected from the preceding and following sols. In this way, the local conditions are best accounted for. This local method does require significantly more resources as a new model is trained for each sol, when only one for the entire mission was previously required. Herein we will refer to the GP model trained on data sampled from across the entire mission as the global GP model and the GP trained on data only from the surrounding sols as the local GP model.

4.2 Neural networks

4.2.1 Background

A neural network (NN) takes each element in an input feature vector and maps them to an output value through a series of connected 'neurons'. Each 'neuron' multiplies each connection by a weight before summing, applying a bias (also a type of weight) and passing through an activation function (Murphy 2012). The activation function can take several forms, for example, the ReLu (rectified linear) function given as

$$f(x) = \max(0, x) \tag{9}$$

where the output is zero for inputs below zero and linear for outputs above. This choice is useful for regression problems and is less susceptible to vanishing gradient problems than other activation function choices (Goodfellow *et al.* 2016). The full NN is made up of sequentially applied layers each consisting of several neurons taking the output of the previous layer as an input until the output layer is reached.

The number of neurons in each layer is the breadth of the network and the number of layers is the depth. Large numbers of layers gives rise to deep learning (Goodfellow *et al.* 2016), useful for approximating extremely complex functions. NNs are able to approximate a wide class of functions (universal approximation theory, Cybenko 1989) given an appropriate architecture, that is, the number of layers, the number of neurons on each layer and the choice of activation functions.

The training of an NN to a specific task requires the weights and bias values for each neuron to be optimized. This is achieved through stochastic gradient descent algorithms to find the minimum

of the loss function, in our case the mean square error of the predicted output and the target output. These algorithms evaluate the gradient (using backpropagation) for a subset of training data and iteratively update the weights to step towards the minimum of the loss function. The subset of training data used for an evaluation of the gradient is known as a batch. Splitting the data into batches is helpful for computational efficiency but also provides a regularization effect. The resulting step in updating the weights dependent on the gradient is controlled by a learning rate, trading off eventual accuracy and convergence speed. A complete pass of the training data (all the batches) is known as a training epoch. Several epochs may be completed to finalize the model weights. Comparing the loss for the training data to a separate validation/development data set (not contained within the training data) for each epoch indicates where a model may be overfitting or have reached convergence on a good solution. A detailed explanation of these aspects can be found in Murphy (2012) and Goodfellow *et al.* (2016) for example.

The choice of topology of the network, batch size, learning rate and number of epochs are determined empirically through a process known as tuning. During tuning several different networks are trained for an array of different hyperparameters (i.e. batch size, number of epochs and learning rate). For each topology/hyperparameter, the training loss and validation/development loss for a trained NN are evaluated to find the optimal choice in terms of lowest error while not overfitting (small difference between training and validation loss) and parsimony. In general, the NN performance improves for larger data sets.

4.2.2 Implementation

An MLP NN architecture was implemented using the Keras python package of Chollet *et al.* (2015). Each neuron in the hidden layers was defined by the ReLu activation function in eq. (9), apart from the final output layer where a linear combination was used. A dropout layer was added to the first layer of the NN for regularization. The dropout layer sets a given node in a layer to zero with the dictated frequency during training which reduces the networks ability to memorize samples, reducing the chances of overfitting (Goodfellow *et al.* 2016).

The total data set (for both the LF and HF energy data) was split into a training, validation and test data set. The training set is used for the optimization of the network weights while the validation set is used as a separate assessment the performance over training to detect overfitting and select the optimal network topology and hyperparameters (batch size and level of dropout). The test set then gives the final assessment of performance, ensuring separate data has been used to what has been used for optimization to give a true performance of the network.

Prior to separation into the training, test and validation data set the data were split into one hour long chunks. This splitting by hour avoids closely correlated consecutive samples being present in both training and validation/test data sets. The test data set was taken as 20 per cent of these chunks at random from across the mission and set aside. Of the remaining chunks 80 per cent were randomly selected from across the mission to train the NN (64 per cent of the total data) and the remaining 20 per cent were taken as a separate validation set (16 per cent of the total data).

Figs A1 and B1 show the training and validation error for a range of models with a range of neurons, numbers of hidden layers, batch sizes and level of dropout for both LF and HF energy predictions. The training for each model used the ADAM optimizer (Kingma & Ba 2014, a variant of the stochastic gradient decent algorithm) with a starting learning rate of 0.001 and done over 500 epochs. The best choice of network is that with the lowest validation loss and smallest gap between validation and training loss. For our application, the final choice of network for the LF prediction was made to be a network with 6 hidden layers, 30 neurons on each layer with a dropout of 0.005 and batch size of 512. On the other a model of 4 hidden layers with 30 neurons on each layer with a dropout of 0.005 and batch size of 512 was chosen for the HF prediction. Note that there are however several networks with similar training and validation losses and therefore performance.

Fig. 3 shows the training and validation loss over the training epochs of the selected networks. The LF model is very slowly decreasing and the RMSE has not improved significantly for several epochs and so is sufficiently trained. The HF model training loss improves very slightly beyond the test loss after ∼220 epochs and so training can be halted. Note that the validation loss varies over epochs of the order of performance difference between the similar networks, highlighting that each is a valid choice. At this level, the effects of the sensors performance can be a factor. In particular the wind sensor has limited resolution at low and high wind speeds. These selected models which are used for the subsequent results on application to the data are referred to as the NN/neural network model.

4.3 Analysis of the fits

In this section, we will discuss the performance of the GP and NN models trained above to predict both the LF (0.1–1 Hz) and HF (2.2–2.6 Hz) seismic energy. Table 2 shows the RMSE computed for each of the three models from 20 per cent of the overall data. This gives a comparable view of the overall performance for each model. Note that the local GP model had several sols removed from the error calculation as they had erroneous predictions due to sampling data artefacts in the training data. The NN approach provides the lowest error, followed by the local and then global GP models. This is the case for both LF and HF seismic energy predictions. On top of this, the LF model has slightly better performance than the HF model.

In order to analyse how the performance of models varies over the Martian sol and seasons we will examine over three exemplar sols: 258, 495 and 690 which are the central sols for the periods in Fig. 2. For the LF bandwidth, Figs 4, 5 and 6 show the predictions compared to the observations and residual errors of the LF bandwidth for the GP global, GP local and NN models respectively, while Figs 7, 8 and 9 show the same for the HF bandwidth. The residuals are shaded with their smoothed values over 100s and 1000s and their PDF estimates are shown on the side panel. First note that there is no significant difference in the prediction of the vertical (Z) and horizontal (N and E) components for any model and for all sols and all perform reasonably at this scale. Table 3 shows the RMSE for each sol for each model. For each sol the RMSE is lowest for the NN, followed by the local GP and then global GP, as was the case with the error calculated over the whole data set. This is aside from the local GP prediction on sol 690 which has equivalent/slightly improved error than the NN for the LF and HF predictions, respectively.

One major advantage of the GP models over the NNs is that the prediction is given as a distribution, and so confidence intervals are obtained intrinsically. These confidence intervals (plotted at 95 *per cent*) are generally much tighter to the prediction for the

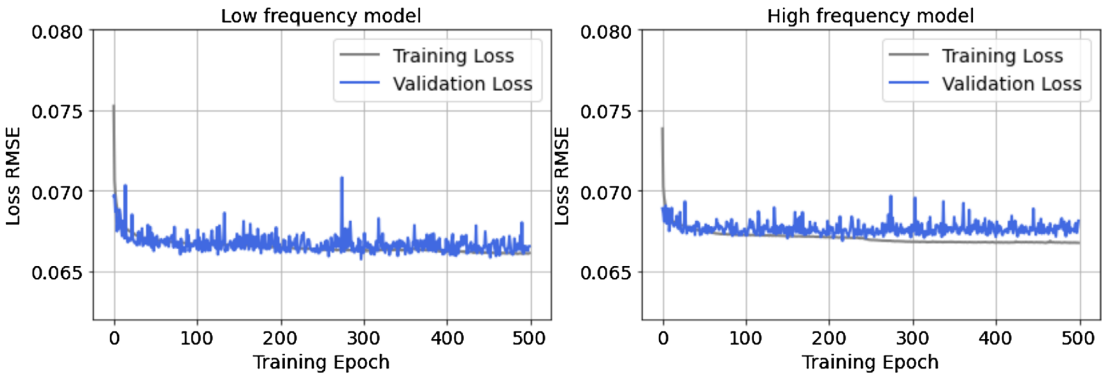

Figure 3. The validation and training RMSE of the selected NN models for LF and HF over 500 training epochs.

Table 2. Test RMSE of ML models.

Model	Test RMSE	
	LF	HF
GP global	0.074	0.081
GP local	0.072	0.075
Neural network	0.068	0.069

local GP than the global GP model, indicating the local model to have better constrained the relationship.

4.3.1 Analysis of diurnal and seasonal variation

Recall that there are are three common regimes over a Martian sol: daytime turbulence, the evening quiet period and night time turbulence. These regimes are clear on sol 258 but change over the mission, most notably the quiet period is only briefly visible on sol 495 and has disappeared by sol 690, when the night time turbulence has become more intense. The pdf of the residuals is generally Gaussian-like and consistent for each sol, indicating there is no major change in the prediction quality over the seasons. However, certain features can be noticed. The night time (turning into early morning) turbulence shows sharp transient level changes, especially on sol 258 and sol 690. These rapid transient shifts are not always instantly predicted and cause a level change in the residuals highlighted by the light coloured smoothed trends. These transients may be because the seismometer records effects at further distances than the meteorological data. Moreover, the daytime convective turbulence has greater variation than the other turbulent periods. This variation is largest on sol 495, where the residuals are larger during the daytime and so their distribution is not completely stationary. Note that sol 495 also has the largest RMSE for the LF prediction.

The prediction of the HF energy level during the quiet, low noise, periods is generally more constant in appearance than for the LF. This suggests that the HF estimates do not predict all the general ambient variation in this regime. This is the likely cause of the higher RMSE generally seen for the HF compared to the LF predictions. One possible reason is that the atmospheric sensors are not sensitive enough for this prediction, as the higher frequencies are generally more sensitive to the wind (Charalambous *et al.* 2021). On the other hand, the HF prediction covers the 2.4 Hz mode. As introduced, this mode is proposed to be of geophysical origin and is amplified by events in this bandwidth making it useful to detect marsquakes. However, the mode also exhibits ambient modulation over the sol and season, the generation of which has yet to be pinned down. The fact that these machine learning models fail to predict the overall ambient variation of the 2.4 Hz suggests that it is not generated with local atmospheric effects. This effect may explain the lower RMSE for the sol 690 prediction as there is no quiet period and hence no ambient variation of the 2.4 Hz which requires prediction.

5 QUAKE ANALYSIS AND DETECTION

In this section, we take the seismic noise energy predictions/estimates of the introduced machine learning models and show how we can use it to examine the quality of and detect events. The best performing and most robust model is that of the NN and so we will use its results unless stated otherwise.

5.1 Signal-to-noise ratio of marsquakes

The residual (model prediction subtracted from the observed data) can be used to examine the seismic signals independence from atmospheric injections. As the prediction is in the logarithmic domain, this subtraction yields the signal-to-noise ratio (SNR) of the observed energy to the predicted noise level at a particular point in time. We can then calculate the SNR for a particular event is given as the maximum of this residual signal during the event, based on the prediction from NN models. This yields an approach to assess event quality in terms of its separation from the atmospherically generated noise. The SNR was calculated in this way for every event in the MQS catalogue (with available data) in both the LF (0.4–1 Hz) and HF (2.2–2.6 Hz) bandwidths. A table of these is provided in Supporting Information. Note that this SNR is in terms of signal energy.

5.1.1 Example marsquake analysis with machine learning prediction

To show how the machine learning prediction can be used to analyse a particular event, Fig. 10 shows two events from the low-frequency group. The S0395a event (left) is a BB quality C and the S0325a event (right) is an LF quality B. These are two clearly observed events and the seismic energy during both events clearly deviates from the proposed NN model predicted noise levels and the LF SNR for both events is around 3. The S0395a (left) event takes place in a fairly quiet period, with a clearly visible event-like envelope in

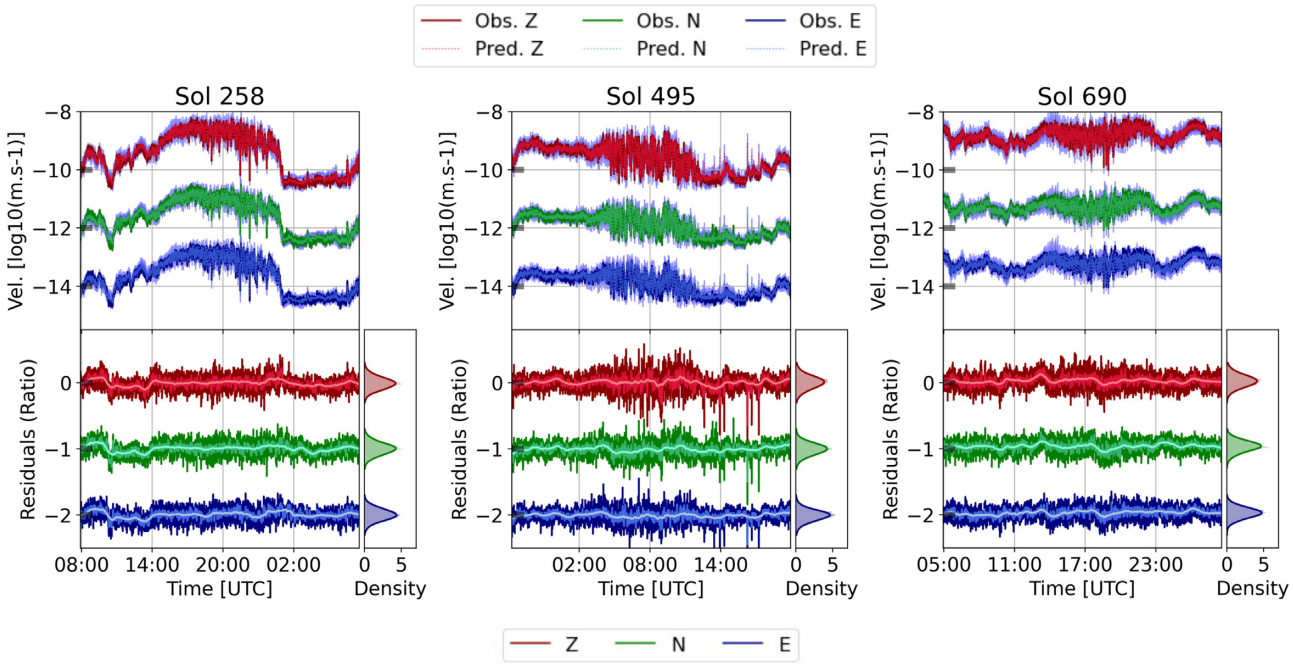

Figure 4. Global GP LF prediction and observations. The N and E traces are offset on the *y*-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to $-10\ \log(\mathrm{m\,s^{-1}})$ or 0.

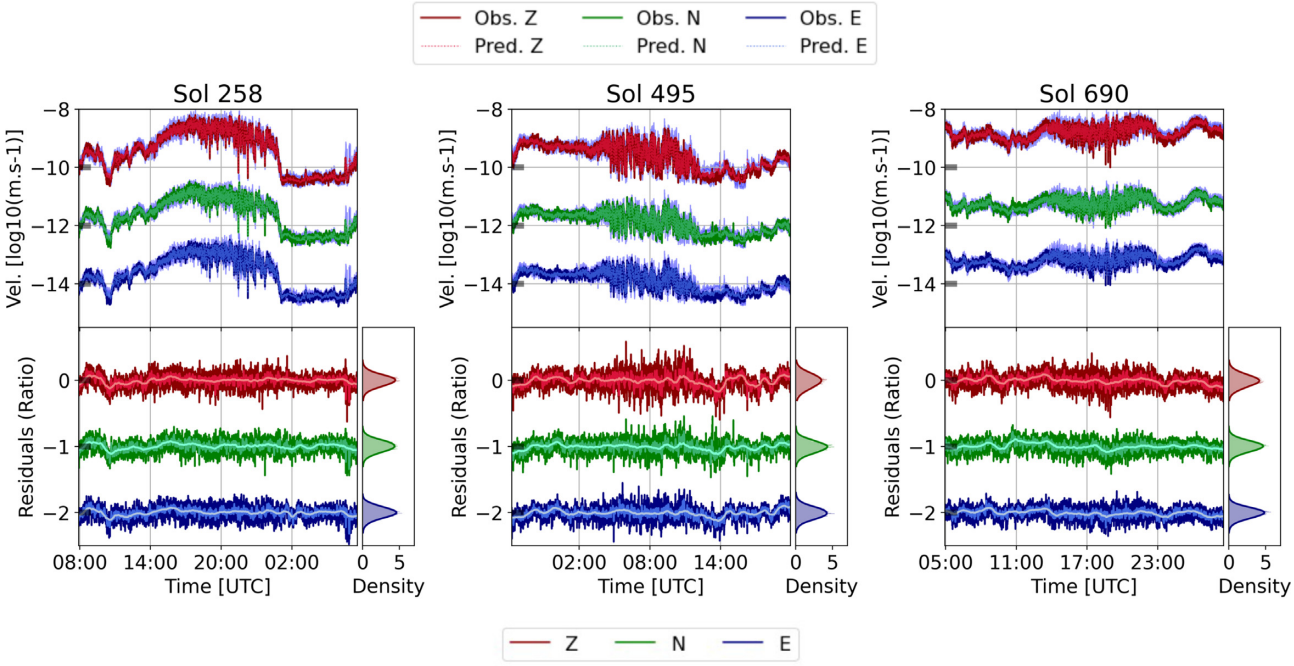

Figure 5. Local GP LF prediction and observations. The N and E traces are offset on the *y*-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to $-10\ \log(\mathrm{m\,s^{-1}})$ or 0.

the observed energy and SNR. As it is a BB event there is some energy at high frequencies. It can be seen that there are two distinct peaks in the HF energy but only one in the LF energy. As it is a C quality only one phase pick is given in the catalogue. Furthermore, the LF SNR highlights some energy prior to the event start not indicated by the catalogue. This can also be seen in the spectrogram. This energy emerges from the lower frequencies, similar to atmospheric noise, however, the machine learning has indicated it is separable. On the other hand, the S0325a (right) event takes place in a fairly noisy period, however, two phase arrivals are apparent in the observed LF energy. This is better highlighted in the LF SNR and gives the event its quality B rating meaning its distance can be determined.

Fig. 11 shows two high-frequency type events. The S0542a event

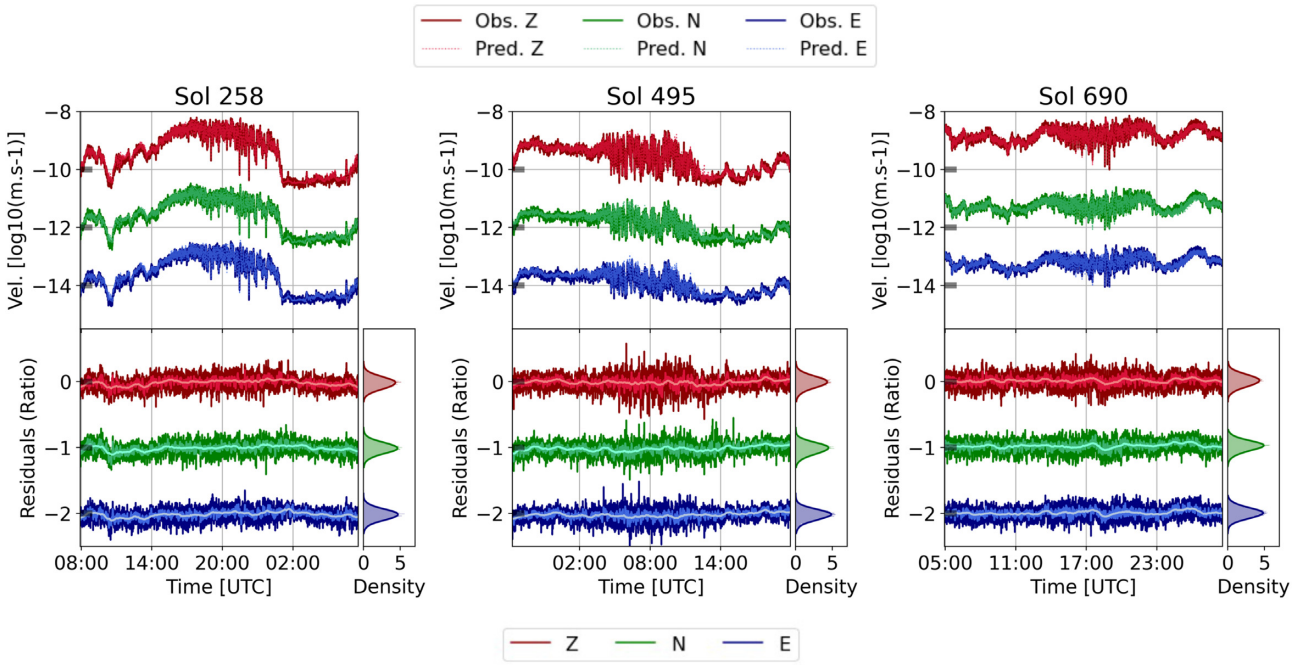

Figure 6. NN LF prediction and observations. The N and E traces are offset on the y-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to -10 $\log(\mathrm{m\,s^{-1}})$ or 0.

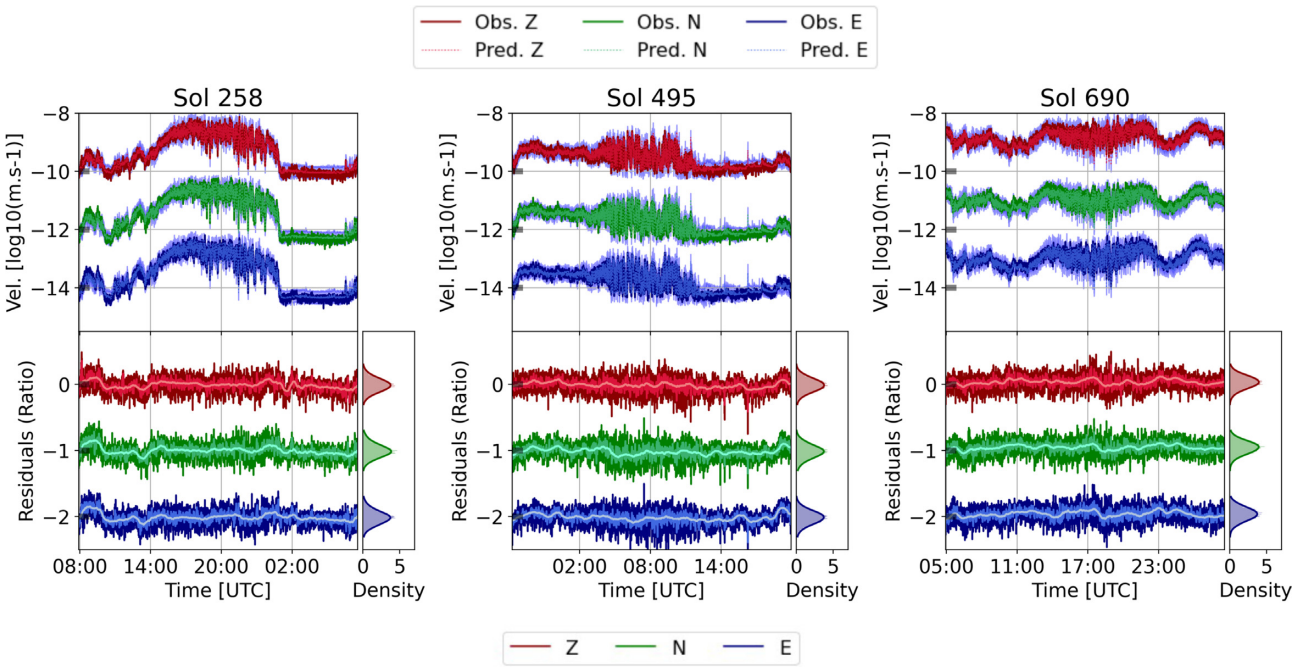

Figure 7. Global GP HF prediction and observations. The N and E traces are offset on the y-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to -10 $\log(\mathrm{m\,s^{-1}})$ or 0.

(left) is a VF quality C event while the S0534a event (right) is a 2.4 Hz type quality C. These events both occur in an extremely noisy period as the season progressed into winter. After the S0542a event, only two events were identified until sol 653. In spite of the noise, the NN model prediction of S0542a (left) is able to establish the signal clearly with a peak SNR of nearly 8. During this S0542a event (left) there is a large gust which causes the SNR to drop and then recover. This enables the assessment of where the energy of the event is suitable to use and where it may be contaminated by noise. The comparison to the model predictions also clearly separates the energy of the the S0534a event (right) from a very noisy background. As S0534a (right) is a 2.4 Hz type event, only energy at this mode is clearly observable in the spectrogram. However, the predictions

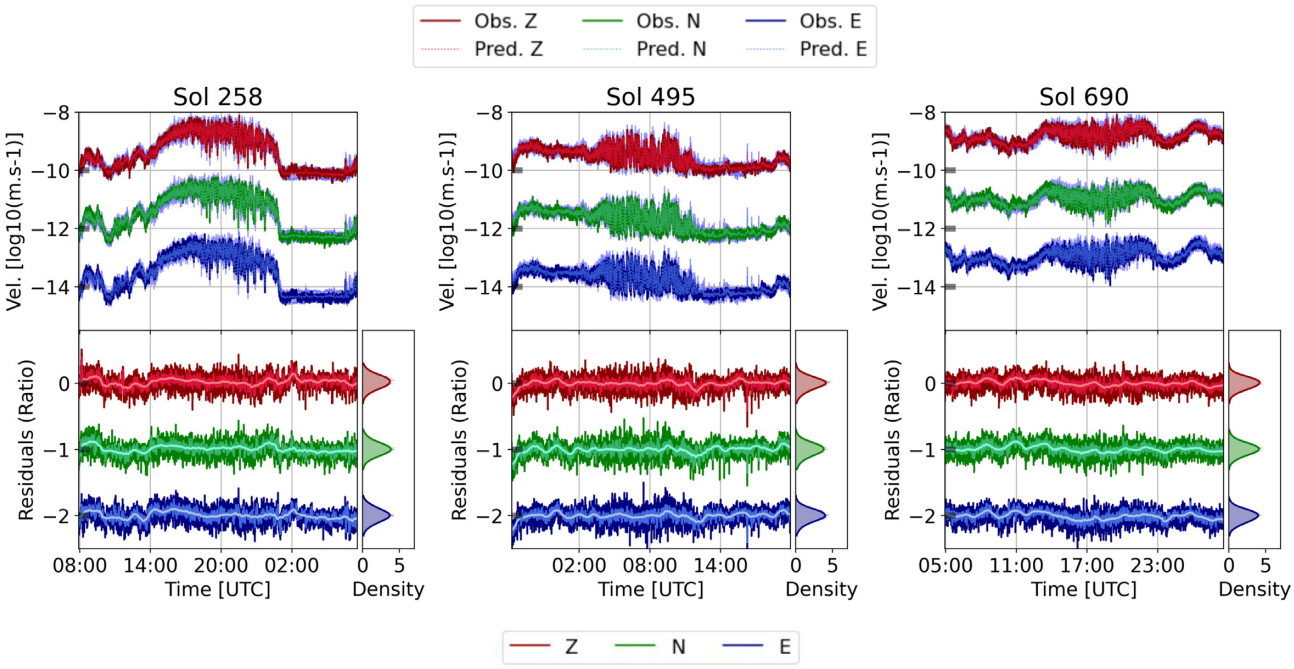

Figure 8. Local GP HF prediction and observations. The N and E traces are offset on the y-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to -10 $\log(\mathrm{m\,s^{-1}})$ or 0.

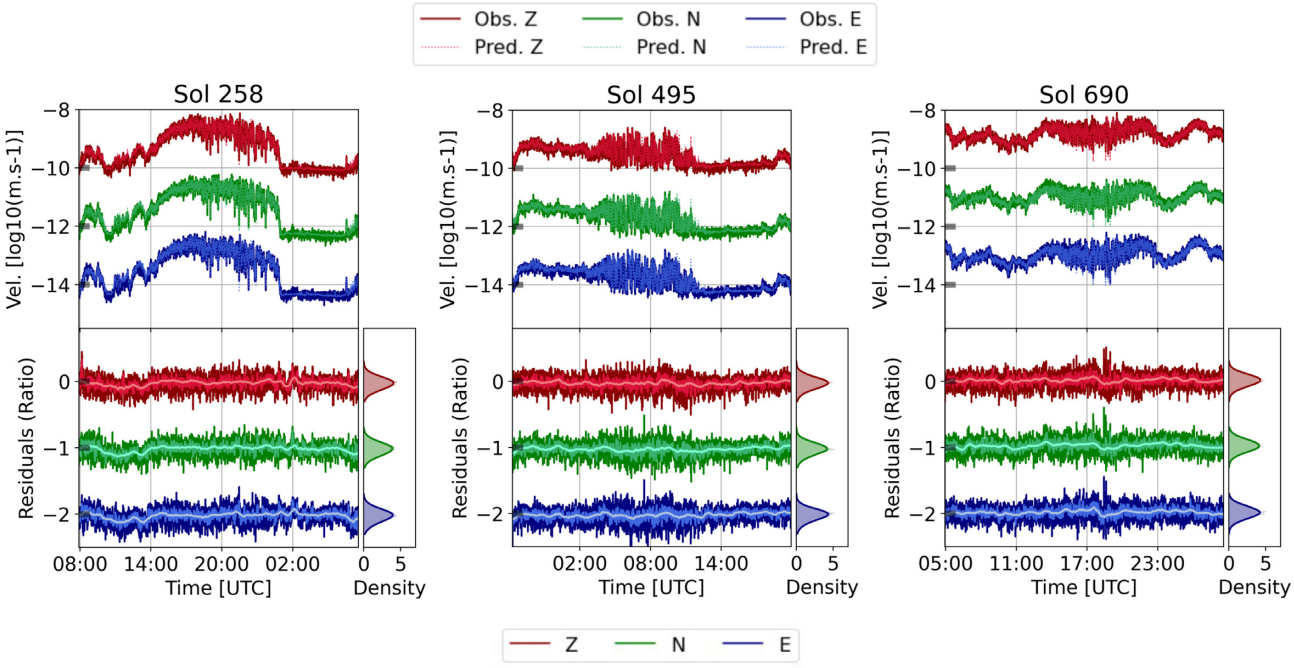

Figure 9. NN HF prediction and observations The N and E traces are offset on the y-axis by 2 $\log(\mathrm{m\,s^{-1}})$ and 1 for the prediction/observation in the top row and residual in the bottom row, respectively. The black markers represent successive relative points equivalent to -10 $\log(\mathrm{m\,s^{-1}})$ or 0.

Table 3. RMSE of ML models for three sols.

Model	Sol 258		Sol 495		Sol 690	
	LF	HF	LF	HF	LF	HF
GP global	0.068	0.077	0.077	0.078	0.70	0.077
GP local	0.066	0.075	0.074	0.074	0.064	0.068
Neural network	0.064	0.073	0.068	0.071	0.063	0.068

indicate there is some LF energy for S0534a (right) with an SNR of up to 2.

The amplitude velocity residuals is shown in the fourth panel down for each figure, where the amplitude of the predicted noise is removed from the observed velocity amplitude, thereby decorrelating the current atmospherically generated noise from the event envelope. Both high-frequency type events in Fig. 11 show a slope break at the HF residuals, indicating a possible secondary

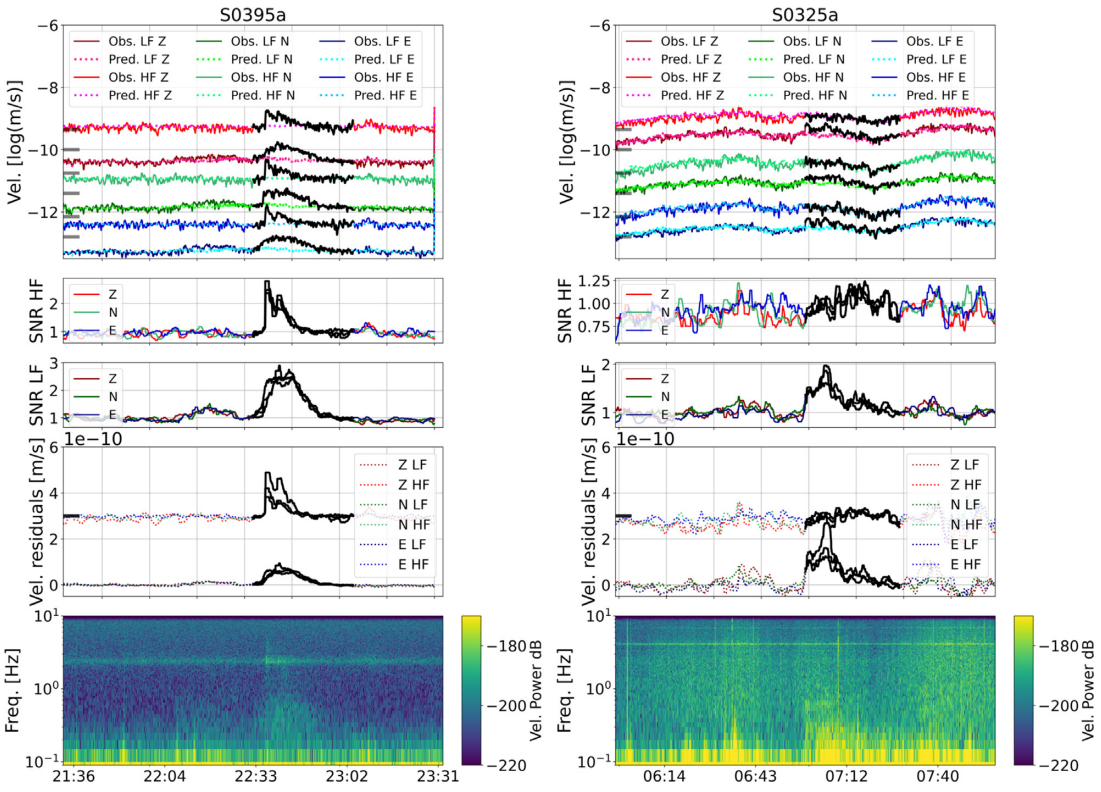

Figure 10. The NN model predictions and SNR analysis for the S0395a BB quality C (left) and S0325a LF quality B (right) events. The top panel of each event has the LF predictions/observations offset by -1.4 log(m s^{-1}) and -2.8 log(m s^{-1}) for the N and E axes with the HF predictions/observations offset from the respective LF predictions by 0.65 log(m s^{-1}). The black marks indicate the relative level of -10 log(m s^{-1}) for each trace. The second bottom panel has the HF velocity residual offset by 3×10^{-10} m s^{-1} from the LF, indicated by the black mark.

phase pick which is not given in the MQS catalogue as it cannot be clearly seen in the raw data.

This analysis of these events demonstrate how the machine learning predictions of atmospherically generated noise can be used to extract useful details from marsquakes. This provides a framework which yields information helpful for the detailed and consistent cataloguing of marsquakes.

5.2 Marsquake catalogue analysis

Fig. 12 shows the SNR for every event (with available data) throughout the first thousand sols. This does exclude most events later in the mission as power constraints prevented the wind and pressure sensor to be on at all times. The top and bottom panels show the SNR for each event (all types) for in the HF (2.2–2.6 Hz) and LF (0.4–1 Hz) bandwidths, respectively, based on the proposed NN model to predict the corresponding bandwidth seismic energy. It can be seen that the SNR generally modulates with the overall noise level throughout the mission (Clinton *et al.* 2021), increasing up to around sol 330 then decreasing until the noise was too high for many events to be observed. This is in accordance with the level of local turbulence demonstrated in Spiga *et al.* (2021) and Chatain *et al.* (2021).

Notice that many events from the HF grouping of events (2.4 Hz, HF and VF) have reasonable LF SNRs. This indicates the presence of separable LF energy within these events, in contrast to their taxonomy outlined above. This demonstrates that the above observation of LF energy in the S0534a event (in Fig. 11) is not unique to that event and is a consistent feature which occurs regularly for all the high frequency types of event.

To examine this further, Fig. 13 shows the residual velocity amplitude (the subtraction of the prediction from the observed in the amplitude domain) against each events magnitude (as derived in Böse *et al.* 2021) and distance for both the LF and HF NN model predictions. There is a good correlation between the magnitude and the amplitude residual in both the LF and HF bandwidths for all the HF event types, indicated by the Pearson correlation coefficient, ρ, in the legend where it was significant enough to reject the null hypothesis of no correlation, that is, the p-value was below 0.05 for 95 per cent confidence interval. The magnitude of an event is dependent on both observed amplitude and distance (Böse *et al.* 2021). The high-frequency group of events (2.4 Hz, HF and VF) are generally closer and so the amplitude is the dominant parameter in determining the magnitude. The low -frequency event types (LF and BB) are more effected by distance which contributes to the lack of statistically significant correlation for the LF events magnitude and residuals in the LF bandwidth, along with having fewer data points. The effect of attenuation (due to their distance) on HF energy prohibits a correlation for the LF and BB events at the high frequencies.

In contrast to magnitude, there is no clear correlation with distance and the amplitude residual for the high-frequency event groups. This shows the presence of LF energy is typical in all high-frequency type events as a function of their size and it is not an observational bias depending on location. As a result, this suggests that both the high- and low-frequency event classes are

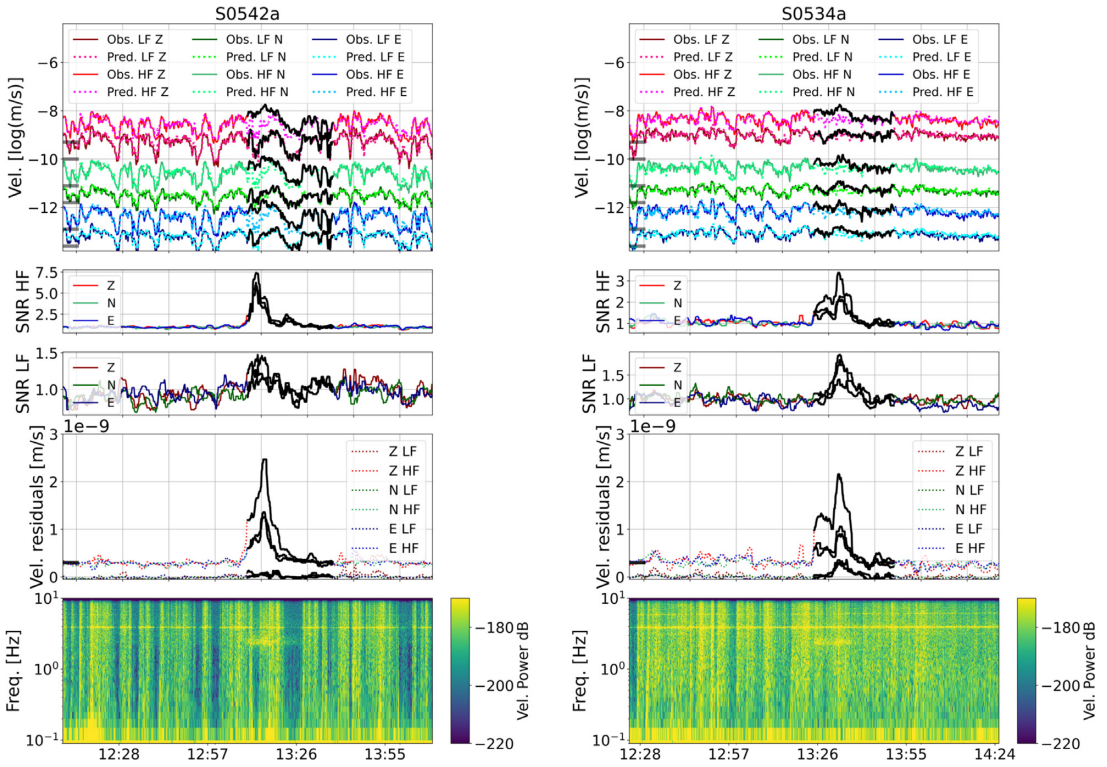

Figure 11. The NN model predictions and SNR analysis for the S0542a VF quality C (left) and S0534a 2.4 Hz quality C (right) events. The top panel of each event has the LF predictions/observations offset by -1.8 log(m s^{-1}) and -3.2 log(m s^{-1}) for the N and E axes with the HF predictions/observations offset from the respective LF predictions by 0.8 log(m s^{-1}). The black marks indicate the relative level of -10 log(m/s) for each trace. The second bottom panel has the HF velocity residual offset by 3×10^{-10} m s^{-1} from the LF, indicated by the black mark.

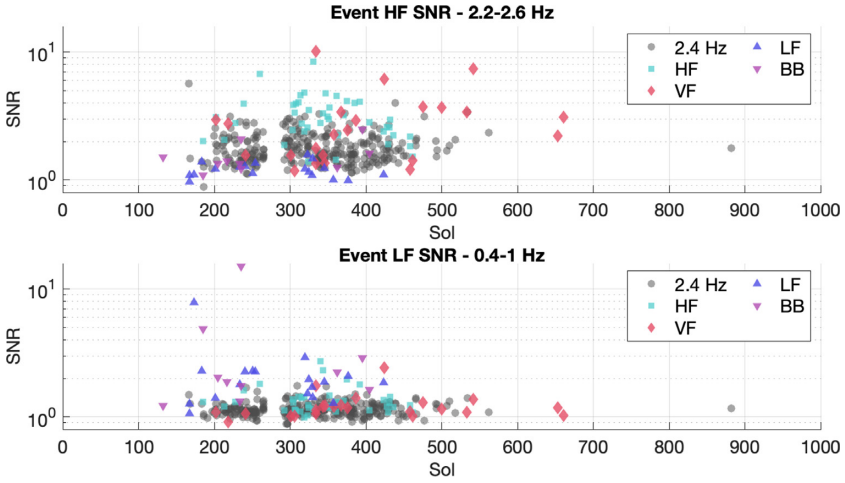

Figure 12. The SNR for the observed seismic energy to the predicted noise levels calculated for the LF 0.4–1 Hz and HF 2.2–2.6 Hz bandwidths. This is shown for each event class over the mission.

not disparate groupings and that they instead follow a continuum in frequency content. This in turn demonstrates the ability for the NN model to decorrelate the noise from the observed events.

5.3 Quake detection

New marsquakes can also be directly detected through identifying where the predicted noise level diverges from the observed signal. To do so we must calculate a threshold the residual energy must reach to suggest a seismic event. We propose the following algorithm to calculate a variable threshold as:

(i) calculate exponential base 10 of the data to return to the velocity amplitude domain and subtract the prediction from the observation to obtain a residual,

(ii) filter the residual with a moving median over 100s and subtract the same signal filtered with a moving median over 2000s,

(iii) take the absolute value of the signal and calculate the moving median over 300s for the LF model and 100s for the HF model,

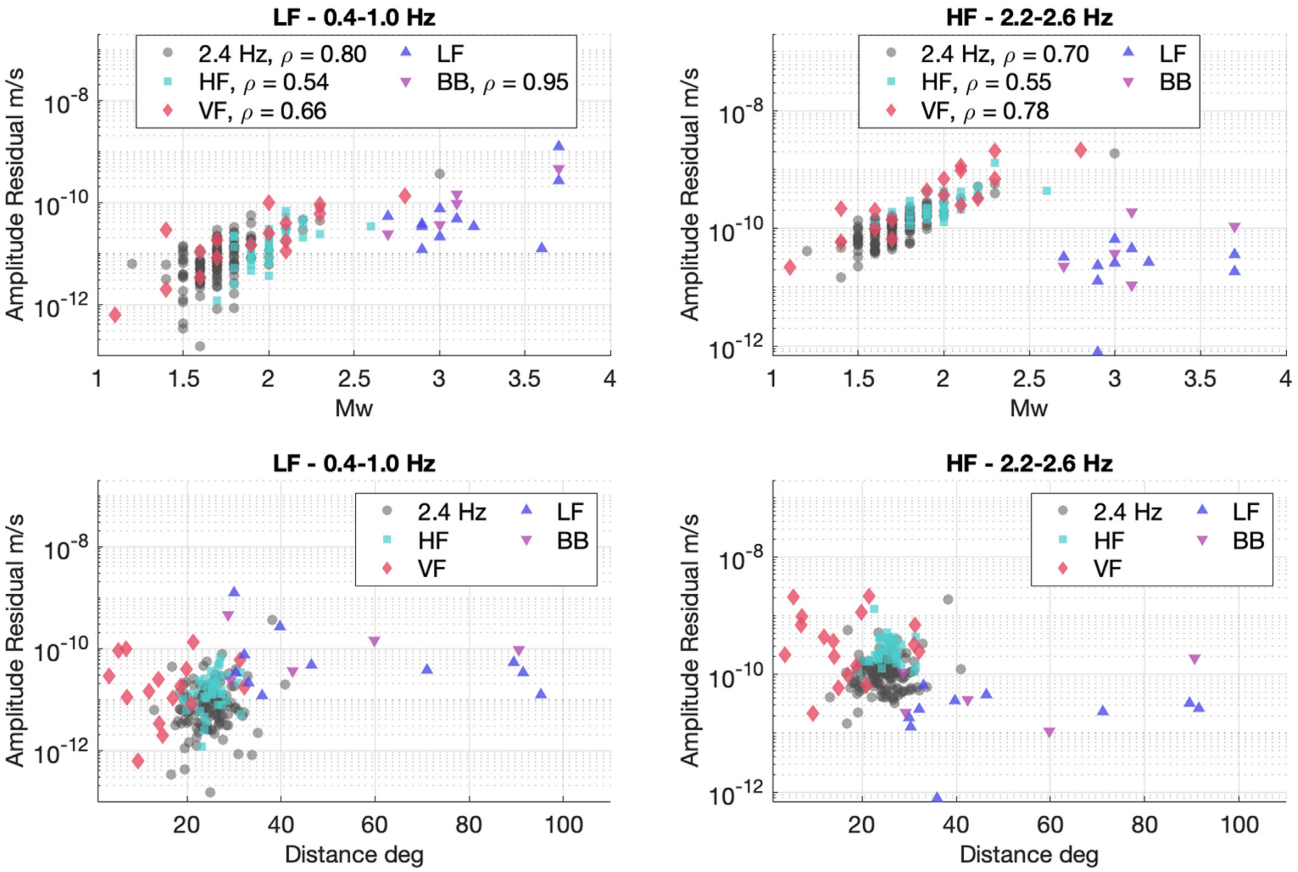

Figure 13. The residual velocity (observed velocity amplitude − NN model predicted velocity amplitude) against the marsquake magnitude and distance for both LF and HF models. The legend indicates the Pearson correlation coefficient (ρ) for a relationship where it is significant (p-value < 0.05).

Table 4. New LF type events.

Event name	UTC time	Type	SNRZ LF
M0190a	2019-06-10T07:44:46	LF	1.59
M0205a	2019-06-25T18:45:03	BB	1.38
M0214a	2019-07-04T23:42:47	BB	1.68
M0248a	2019-08-08T20:32:23	LF	2.28
M0336a	2019-11-07T09:54:51	BB/HF	1.75
M0345a	2019-11-16T16:31:10	LF	2.12
M0395a	2020-01-06T20:46:37	LF	1.44
M0434a	2020-02-15T21:41:37	BB/HF	1.72
M0440a	2020-02-22T06:39:31	LF	1.37
M0452a	2020-03-05T07:11:54	LF	2.53
M0547a	2020-06-10T22:48:27	LF	1.79

(iv) calculate the threshold over a period of 2 hr as the mean value plus three standard deviations of the signal in the window and advance the window at 0.5 hr intervals over the sol,

(v) detect outliers where the absolute error is above this threshold,

(vi) select outliers only on all three components,

(vii) select periods where the detection has positive amplitude as a marsquake will have additional signal.

Calculating the threshold in this way is attempting to find an anomalous, outlying, sample within the distribution over the 2 hr searched.

The detection algorithm performed on the LF, 0.4–1 Hz, band found 25 of 32 LF and BB quakes (available with all atmospheric data) in the MQS catalogue (InSight Marsquake Service 2022). Of those not detected, one is the S0405c event which occurs at the very

Table 5. New HF type events.

Event name	UTC time	Type	SNRZ HF
M0166a	2019-05-16T09:46:01	2.4 Hz	1.3
M0205b	2019-06-25T02:49:38	2.4 Hz	1.68
M0250a	2019-08-10T03:34:41	2.4 Hz	2.21
M0254a	2019-08-15T04:08:50	2.4 Hz	1.64
M0256a	2019-08-17T00:46:51	2.4 Hz	1.33
M0258a	2019-08-19T01:57:51	2.4 Hz	1.66
M0291a	2019-09-21T10:58:54	2.4 Hz	2.02
M0303a	2019-10-04T10:27:15	2.4 Hz	1.53
M0312a	2019-10-13T16:10:53	2.4 Hz	1.64
M0313a	2019-10-14T13:43:33	2.4 Hz	1.52
M0313b	2019-10-14T18:09:01	2.4 Hz	1.69
M0314a	2019-10-15T18:38:26	2.4 Hz	1.65
M0317a	2019-10-18T16:53:42	2.4 Hz	1.57
M0331a	2019-11-02T05:04:35	2.4 Hz	1.53
M0333a	2019-11-04T08:23:55	2.4 Hz	1.83
M0336a	2019-11-07T09:50:17	HF/BB	1.52
M0338a	2019-11-08T17:37:39	HF/BB	1.81
M0338b	2019-11-09T07:52:22	2.4 Hz	1.7
M0349a	2019-11-20T13:29:37	2.4 Hz	1.74
M0351a	2019-11-22T16:59:05	2.4 Hz	1.57
M0354a	2019-11-25T16:50:08	2.4 Hz	1.64
M0365a	2019-12-07T00:10:51	2.4 Hz	1.79
M0431a	2020-02-12T19:41:13	2.4 Hz	1.34
M0434a	2020-02-15T21:43:03	BB/HF	1.53
M0446a	2020-02-28T05:36:02	2.4 Hz	1.93
M0458a	2020-03-10T21:22:01	2.4 Hz	1.7
M0466a	2020-03-19T19:37:39	2.4 Hz	1.53
M0516a	2020-05-10T02:52:42	2.4 Hz	1.77

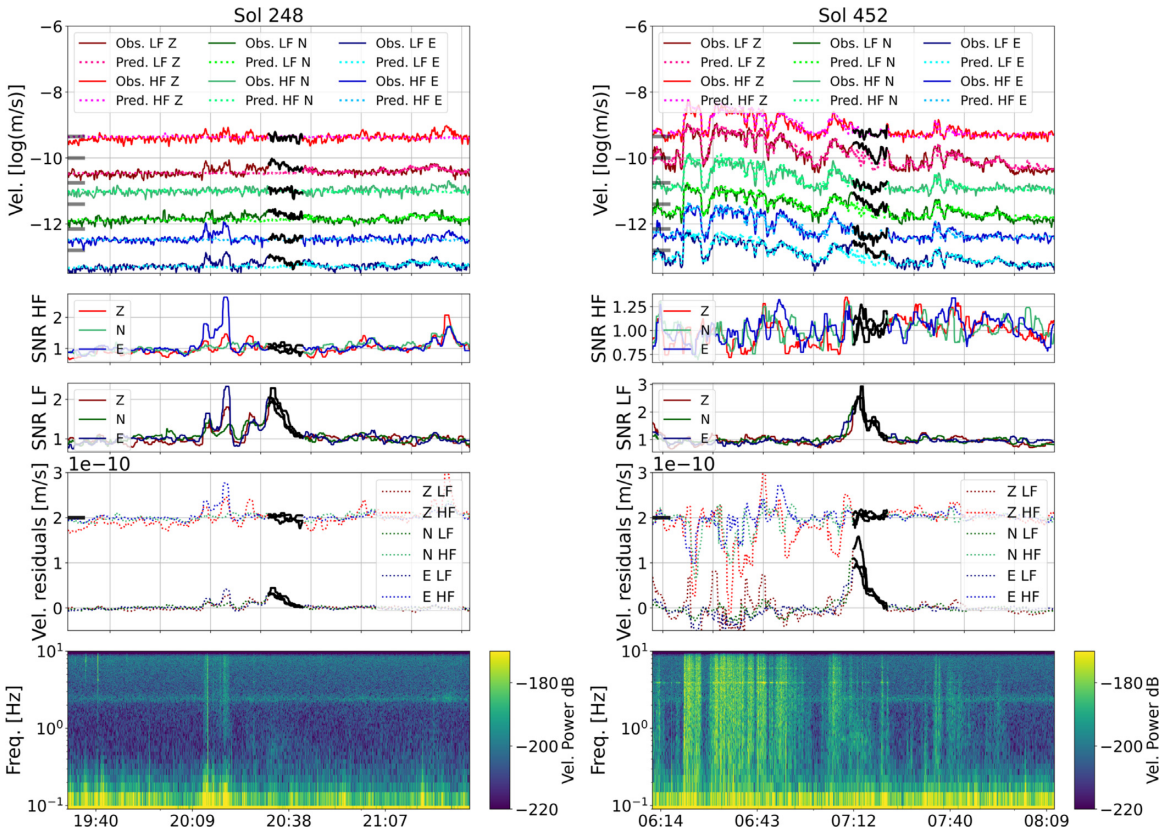

Figure 14. Examples of new LF group event detections. The top panel of each event has the LF predictions/observations offset by -1.4 log(m s^{-1}) and -2.8 log(m s^{-1}) for the N and E axes with the HF predictions/observations offset from the respective LF predictions by 0.65 log(m s^{-1}). The black marks indicate the relative level of -10 log(m s^{-1}) for each trace. The second bottom panel has the HF velocity residual offset by 2×10^{-10} m s^{-1} from the LF, indicated by the black mark.

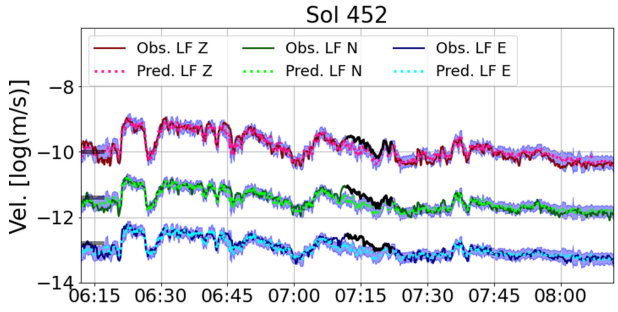

Figure 15. The new LF event detected with the local GP model. The horizontal components are offset by -1.4 log(m s^{-1}) and -2.8 log(m s^{-1}), respectively, with the black marker indicating the relative level of -10 log(m s^{-1}).

start of the noisy period and has several major glitches nearby. As a result, the residual is very noisy and so difficult to detect the quake as an outlier even though it is observed with an SNR of 2 to the predicted noise level. Another event not detected is S0132a which is quality D and other techniques (Charalambous *et al.* 2021) have also shown it to be heavily contaminated. The remaining events not detected all have SNR below 1.3 for all components and are all class D or C events. On inspection the energy in these events is mainly below 0.4 Hz so not appropriately represented by the chosen envelope. From the remaining LF detections, 33 can be attributed to high-frequency types of marsquakes already present in the MQS catalogue. As demonstrated above, this group does often contain LF energy.

Following this, there are 530 remaining new LF detections. These were further filtered to find more likely marsquake candidates by requiring a minimum detection duration of 2 min and an SNR greater than 1.2 on the vertical and 1.1 on the horizontals, leaving 250 new LF detections which were manually reviewed. The manual review followed a similar process to that of MQS (Clinton *et al.* 2021) and discarded detections owing to the following:

(i) known aseismic signals such as glitches/donks not fully filtered from the data, a full coverage is given in (Ceylan *et al.* 2021),
(ii) a discontinuity in the data caused by a change in the wind retrieval algorithm,
(iii) anomalous spikes caused either by a short gust of wind during a quiet period or extremely strong rapid gusts of wind during the daytime which cause a larger residual and not fully filtered out by the detection algorithm.

After removing such false detections we propose 11 new marsquakes in the lower frequency classes which are presented in Table 4. These each contain confirmatory characteristics of a marsquake whereby, some bandlimited energy is identifiable in the spectrogram separate to wind excitation and the envelope follows the decay characteristics of an event. These new events are given a name according to the MQS naming convention with the prefix M instead of S, followed by sol number and then alphabetical suffix to denote which event it is per sol. A further nine possible detections are listed in Table B1 which are not straightforward to confirm but

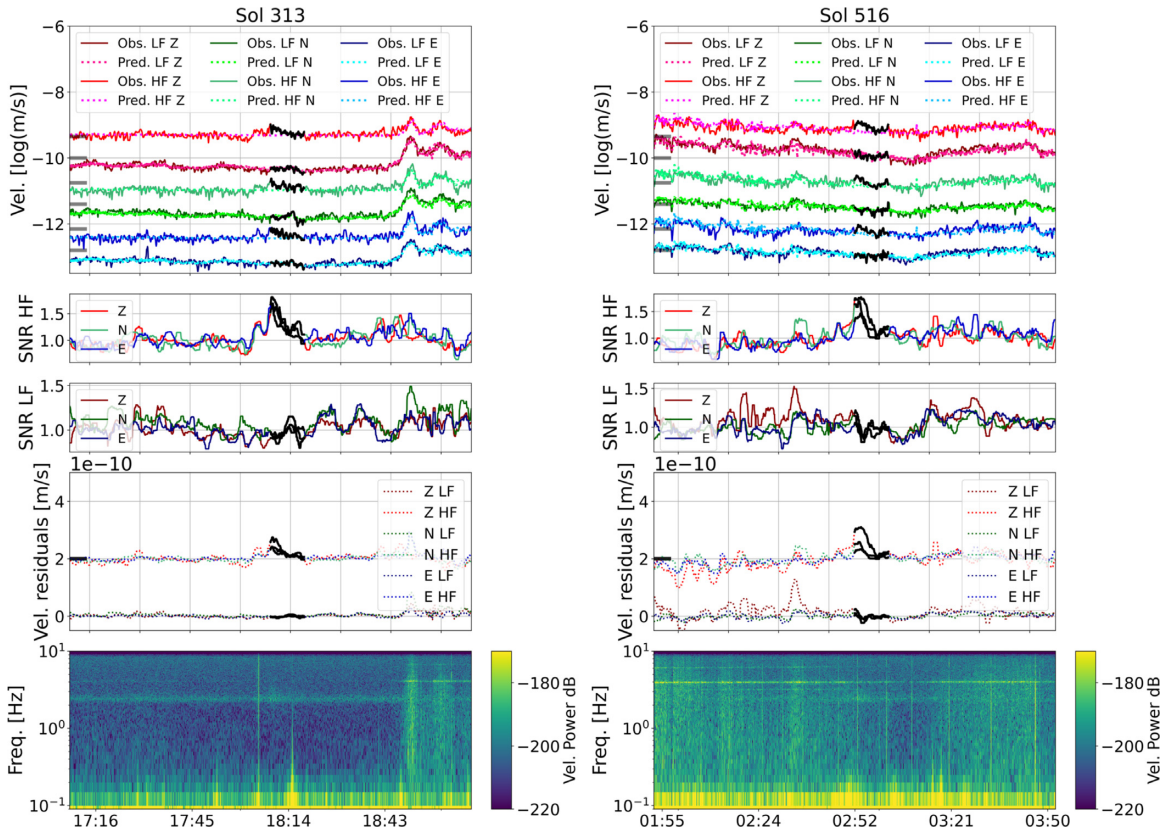

Figure 16. Examples of new HF group event detections. The top panel of each event has the LF predictions/observations offset by -1.4 log(m s^{-1}) and -2.8 log(m s^{-1}) for the N and E axes with the HF predictions/observations offset from the respective LF predictions by 0.65 log(m s^{-1}). The black marks indicate the relative level of -10 log(m s^{-1}) for each trace. The second bottom panel has the HF velocity residual offset by 3×10^{-10} m s^{-1} from the LF, indicated by the black mark.

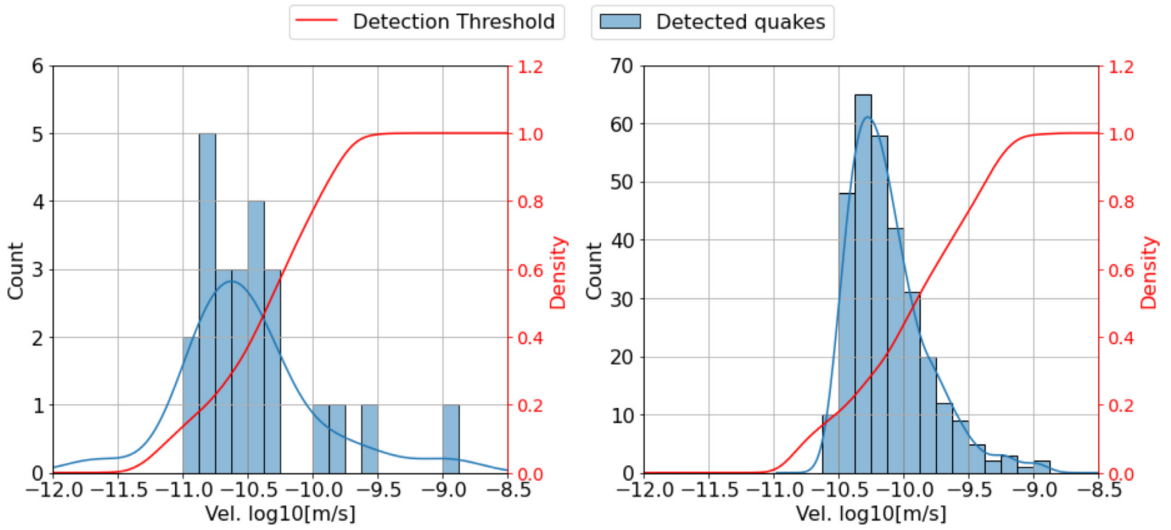

Figure 17. The distribution of detection threshold and events. The red curve is the cumulative distribution of the detection algorithm threshold and the blue is the histogram/KDE of the level at which the low-frequency (left) and high-frequency (right) groups of events were detected as being outliers by the proposed detection algorithm.

have some features in line with those of putative marsquakes and are of interest for catalogue completeness.

The detection algorithm was also ran on the 2.2–2.6 Hz HF envelope. This found 314 out of 377 of the HF event types in the current MQS catalogue (InSight Marsquake Service 2022). Most undetected events are quality D with some C and only three quality B events missed. A further 10 of the new detections are associated with the low-frequency grouping of events already in the MQS catalogue. The remaining new HF detections were filtered similarly to the LF detections above, requiring a minimum duration of 2 min and an SNR greater than 1.2 on the vertical and and on one of the horizontal components while the other horizontal component should

have an SNR of at least 1.1. This yields 498 new detection candidates to examine. A manual review (following the same procedure as above) yielded an extra 28 new marsquakes presented in Table 5 and a further 31 possible detections in Table C1. Of these new HF detections two are shared with the new LF detections.

A comparison of the machine learning model prediction during the MQS catalogue events not detected by the proposed algorithm shows they are seen by the machine learning model prediction. However, they are simply not observed as significant outliers by the algorithm. A particular consideration for HF event detection is that the machine learning model does not well predict ambient modulation of the 2.4 Hz mode. As a result, if the ambient excitation increases and then decreases fast enough it may yield a detection. This poses a problem for the discrimination of low-amplitude HF type events from these ambient variations and so the proposed new detections are based on the subjective criteria to have event-like qualities.

Most of the newly detected events are typically found in either noisy periods typically with episodic gusts or as very low amplitude events in quieter periods. Both of these types of events are difficult to identify by eye and so difficult to be caught by the MQS. This highlights how the machine learning predictions can be used to aid marsquake detection and provide new information to add confidence to a difficult detection. Fig. 14 shows two such archetype new LF group events detected on Sol 248 (left) and 452 (right). The sol 248 (left) event is an example of a low-amplitude event in a quiet period. On the other hand, the sol 452 (right) event is an example of a detection in a noisy period, however, it is clearly separate from the environment and also exceeds the confidence interval of the local GP model, shown in Fig. 15. Fig. 16 shows this archetype for two new HF events detected on sol 313 (left) and 516 (right). The sol 313 event (left) is an example low-amplitude 2.4 Hz type event during a relatively quiet period. It can be seen in the top panel (left) that the noise prediction for the 2.4 Hz energy during this period is flat and the ambient variation is not predicted. In this case the event has a double peak in the HF SNR/velocity residuals and event-like decay which lead to a confirmed detection. The sol 516 event (right) is in a noisy period with nearby broad-band artefacts but there is clear divergent HF energy. The spectrogram shows an increase in energy at the 2.4 Hz mode confirming the event.

5.4 Seismic detectivity

The threshold obtained in the detection algorithm effectively dictates the energy level required of a marsquake to be detected at that point in time. Tracking this threshold over the mission enables an estimate the detectability of a marsquake similar to one in the catalogue over the mission.

Fig. 17 shows, for both the LF (0.4–1 Hz) and HF (2.2–2.6 Hz) bandwidths, the histogram (in red) of the signal level each marsquake reached to break the threshold is also shown in Fig. 17. Note that the event level histograms are positively skewed, indicating that smaller marsquakes are likely hidden below the noise level. This can be compared to the cumulative distribution of the detection threshold obtained over the mission predictions, shown as the blue curve. This cumulative distribution indicates the percentage of time in the data set for which the detection threshold is at or below such a level. As a result, the comparison between this distribution and detection level histogram then yields the amount of the time where a similar event could be observed. For example, the largest LF and HF type events reached a level of $\sim 10^{-9}$ m s^{-1} for which the probability of the detection threshold being below is $\sim$1 and so they could be seen almost the entire time. This means that events in the data similar to those causing the largest signals seen on SEIS (S0173a and S0235b) have likely to not have been missed. The median amplitude event occurs at $10^{-10.6}$ m s^{-1} for LF and $10^{-10.2}$ m s^{-1} for HF, for which the noise level is at or below for $\sim$45 per cent and $\sim$50 per cent of the time, respectively. Finally, the smallest events are at $10^{-10.9}$ m s^{-1} for the LF and $10^{-11.0}$ m s^{-1} for HF for which the noise levels are at or below for $\sim$20 per cent and $\sim$5 per cent of the time, respectively. Note, however, that the detection level is not a true noise level, it is the level for the residual signal smoothed and detrended as in the detection algorithm. As a result, we can only empirically interpret the detectability in relation to the events observed rather than state the true noise level.

6 CONCLUSION

This work implemented GP regression and an NN to predict the seismic energy, as seen by SEIS on InSight, generated from the wind and other atmospheric data. This was done for an LF (0.4–1 Hz) and HF (2.2–2.6 Hz) bandwidth in order to examine the two main frequency bands marsquakes are observed. We showed how the relationship between seismic energy and each input atmospheric variable varies across the Martian sol and seasons, owing to the different forcing from different turbulent conditions. We implement machine learning as a regression style problem for this task, using mathematical complexity to account for a myriad of such factors.

The NN model was found to have the best performance for both LF and HF regions across the entire data set in terms of RMSE. As such it is the most useful for marsquake analysis. However, the GPs confidence interval is useful to understand the uncertainty around a prediction and may be of more interest for future work to unpick the impact of the different atmospheric data variables and inform physical models.

The NN model's prediction was used to first analyse the energy envelope of several marsquakes and derive an SNR (in terms of energy) across the marsquake catalogue. This demonstrated how it can be used for specific event analysis. Examining the SNR for both LF and HF predictions found that the high-frequency group (HF, VF and 2.4 Hz events) consistently contain LF energy with increasing event magnitude. This demonstrates that the marsquake events family is more of a continuum, rather than disparate groupings.

An algorithm was proposed to automatically detect marsquakes. This has led to the detection of 39 new marsquakes and a further 40 possible candidates. Tracking the detection threshold from the algorithm over the data set yields an empirical way to determine the detectivity of marsquakes of a certain size. The cumulative distribution of the threshold gives the amount of time in the mission where the noise level is at or below this level. Comparing this threshold to the catalogue of marsquakes showed that it is unlikely any marsquakes similar to those generating the largest amplitudes observed on SEIS have been missed, whereas the median amplitude marsquake could be observed $\sim$45 per cent and $\sim$50 per cent of the time for LF and HF events, respectively. Through aiding catalogue completeness and analysing marsquake detectivity, this work can be used to help place constraints on the seismicity of Mars.

The detection algorithm could be developed for implementation on board missions to optimize data downlink, such as that proposed by Civilini *et al.* (2021). Our proposed algorithm can rank the detection in terms of SNR and so the threshold at which a detection would be downlinked is flexible depending on downlink resources. This

is an advantage of the signal processing based approach. However, the detection step could be made more robust with a further machine learning based implementation. Other future advances could consider the use of a machine learning model in conjunction with a physical model. In this way, it could utilize the InSight data as prior information to help predict noise levels for future missions as in Mimoun *et al.* (2017).

A notable feature of the machine learning model predictions is that it does not predict the ambient variations in the 2.4 Hz mode, although it does tend to predict the average level. This could suggest its driving source is not solely from local atmospheric effects, however although it could be from regional effects or due to particular atmospheric conditions. Future work could define a cut-off point to separate marsquake excitation from ambient variations which will aid catalogue completeness and also isolate the signal sources for the 2.4 Hz excitation, leading to more information to determine the modes origins.

This work highlights the use of machine learning for consistent treatment of large-scale data in planetary missions. Owing to this consistent treatment, information can be deduced and catalogued repeatably. This represents a sensible approach for (extremely valuable) planetary data and can be considered in geophysical data on the Earth.

ACKNOWLEDGMENTS

This paper is InSight contribution number 264. We acknowledge NASA, CNES, their partner agencies and institutions (UKSA, SSO, DLR, JPL, IPGP-CNRS, ETHZ, IC, MPS-MPG) and the flight operations team at JPL, SISMOC, MSDS, IRIS-DMC and PDS for providing SEED SEIS data. We would like to thank the editor Christoph Sens-Schönfelder and two anonymous reviewers. This work was supported by funding from the French Space Agency CNES and ANR (ANR-19-CE31-0008-08). We also acknowledge the help of the CALMIP computing cluster. AH is funded by the UK Space Agency under grant numbers ST/R002096/1 and ST/W002523/1.

DATA AVAILABILITY

SEIS data are referenced at http://dx.doi.org/10.18715/SEIS.INSIGHT.XB_2016. The Mars Quake service (MQS) catalogue of events used in this contribution is the Mars Seismic Catalogue, InSight Mission, V9 acknowledging ETHZ, IPGP, JPL, ICL, ISAE-Supaero, MPS and the University of Bristol. It is available at http://doi.org/10.12686/a14. The list of confirmed and possible new marsquake detections is available on Github with DOI https://doi.org/10.5281/zenodo.7405279 along with SNRs for MQS catalogue events.

REFERENCES

Banfield, D. *et al.*, 2019. Insight auxiliary payload sensor suite (apss), *Space Sci. Rev.,* **215**(1), 1–33.

Banfield, D. *et al.*, 2020. The atmosphere of mars as observed by insight, *Nat. Geosci.,* **13**(3), 190–198.

Barkaoui, S. *et al.*, 2021. Anatomy of continuous mars seis and pressure data from unsupervised learning, *Bull. seism. Soc. Am.,* **111**(6), 2964–2981.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. Obspy: a python toolbox for seismology, *Seismol. Res. Lett.,* **81**(3), 530–533.

Böse, M. *et al.*, 2021. Magnitude scales for marsquakes calibrated from insight data, *Bull. seism. Soc. Am.,* **111**(6), 3003–3015.

Ceylan, S. *et al.*, 2021. Companion guide to the marsquake catalog from insight, sols 0–478: data content and non-seismic events, *Phys. Earth planet. Inter.,* **310,** 106597, doi:10.1016/j.pepi.2020.106597.

Charalambous, C. *et al.*, 2021. A comodulation analysis of atmospheric energy injection into the ground motion at insight, mars, *J. geophys. Res.: Planets,* **126,** e2020JE006538, doi:10.1029/2020JE006538.

Chatain, A., Spiga, A., Banfield, D., Forget, F. & Murdoch, N., 2021. Seasonal variability of the daytime and nighttime atmospheric turbulence experienced by insight on mars, *Geophys. Res. Lett.,* **48**(22), e2021GL095453, doi:10.1029/2021GL095453.

Chollet, F. *et al.*, 2015. *Keras,* https://keras.io.

Civilini, F., Weber, R., Jiang, Z., Phillips, D. & Pan, W.D., 2021. Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning, *J. geophys. Int.,* **225**(3), 2120–2134.

Clinton, J.F. *et al.*, 2021. The marsquake catalogue from insight, sols 0–478, *Phys. Earth planet. Inter.,* **310,** 106595, doi:10.1016/j.pepi.2020.106595.

Compaire, N. *et al.*, 2021. Autocorrelation of the ground vibrations recorded by the seis-insight seismometer on mars, *J. geophys. Res.: Planets,* **126**(4), e2020JE006498, doi:10.1029/2020JE006498.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function, *Math. Control Signal Syst.,* **2**(4), 303–314.

Dahmen, N.L. *et al.*, 2021a. Super high frequency events: a new class of events recorded by the insight seismometers on mars, *J. geophys. Res.: Planets,* **126**(2), e2020JE006599, doi:10.1029/2020JE006599.

Dahmen, N.L. *et al.*, 2021b. Resonances and lander modes observed by insight on mars (1–9 hz), *Bull. seism. Soc. Am.,* **111**(6), 2924–2950.

Garcia, R.F. *et al.*, 2020. Pressure effects on the seis-insight instrument, improvement of seismic records, and characterization of long period atmospheric waves from ground displacements, *J. geophys. Res.: Planets,* **125**(7), e2019JE006278, doi:10.1029/2019JE006278.

Giardini, D. *et al.*, 2020. The seismicity of mars, *Nat. Geosci.,* **13**(3), 205–212.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning,* MIT Press.

GPy, since, 2012. *GPy: A Gaussian Process Framework in Python,* available at: http://github.com/SheffieldML/GPy.

Hobiger, M. *et al.*, 2021. The shallow structure of mars at the insight landing site from inversion of ambient vibrations, *Nat. Commun.,* **12**(1), 1–13.

Hurst, K. *et al.*, 2021. Resonances of the insight seismometer on mars, *Bull. seism. Soc. Am.,* **111**(6), 2951–2963.

InSight Marsquake Service, 2022. *Mars Seismic Catalogue, InSight Mission; V9 2022-01-01, ETHZ, IPGP, JPL, ICL, ISAE-Supaero,* MPS, Univ. Bristol, doi:10.12686/a14.

Johnson, C.W., Ben-Zion, Y., Meng, H. & Vernon, F., 2020. Identifying different classes of seismic noise signals using unsupervised learning, *Geophys. Res. Lett.,* **47**(15), e2020GL088353, doi:10.1029/2020GL088353.

Kenda, B. *et al.*, 2020. Subsurface structure at the insight landing site from compliance measurements by seismic and meteorological experiments, *J. geophys. Res.: Planets,* **125**(6), e2020JE006387, doi:10.1029/2020JE006387.

Khan, A. *et al.*, 2021. Upper mantle structure of mars from insight seismic data, *Science,* **373**(6553), 434–438.

Kingma, D.P. & Ba, J., 2014. *Adam: A Method for Stochastic Optimization,* preprint, arXiv:1412.6980.

Knapmeyer-Endrun, B. & Hammer, C., 2015. Identification of new events in apollo 16 lunar seismic data by hidden markov model-based event detection and classification, *J. geophys. Res.: Planets,* **120**(10), 1620–1645.

Knapmeyer-Endrun, B. *et al.*, 2021. Thickness and structure of the martian crust from insight seismic data, *Science,* **373**(6553), 438–443.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. Obspy: a bridge for seismology into the scientific python ecosystem, *Comput. Sci. Discov.,* **8**(1), 014003, doi:10.1088/1749-4699/8/1/014003.

Lognonné, P. *et al.*, 2019. Seis: Insight's seismic experiment for internal structure of mars, *Space Sci. Rev.,* **215**(12), doi:10.1007/s11214-018-0574-6.

Lognonné, P. *et al.*, 2020. Constraints on the shallow elastic and anelastic structure of mars from insight seismic data, *Nat. Geosci.,* **13**(3), 213–220.

Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. Obspy–what can it do for data centers and observatories?, *Ann. Geophys.,* **54**(1), 47–58.

Mimoun, D., Murdoch, N., Lognonné, P., Hurst, K., Pike, W.T., Hurley, J., Nébut, T. & Banerdt, W.B., 2017. The noise model of the seis seismometer of the insight mission to mars, *Space Sci. Rev.,* **211**(1), 383–428.

Murdoch, N., Mimoun, D., Garcia, R.F., Rapin, W., Kawamura, T., Lognonné, P., Banfield, D. & Banerdt, W.B., 2017. Evaluating the wind-induced mechanical noise on the insight seismometers, *Space Sci. Rev.,* **211**(1), 429–455.

Murdoch, N., Alazard, D., Knapmeyer-Endrun, B., Teanby, N. & Myhill, R., 2018. Flexible mode modelling of the insight lander and consequences for the seis instrument, *Space Sci. Rev.,* **214**(8), 1–24.

Murphy, K.P., 2012. *Machine Learning: A Probabilistic Perspective,* MIT Press.

Schimmel, M. *et al.*, 2021. Seismic noise autocorrelations on mars, *Earth Space Sci.,* e2021EA001755, doi:10.1029/2021EA001755.

Scholz, J.-R. *et al.*, 2020. Detection, analysis, and removal of glitches from insight's seismic data from mars, *Earth Space Sci.,* **7**(11), e2020EA001317, doi:10.1029/2020EA001317.

Service, InSight Mars SEIS Data, 2019. *Seis Raw Data, Insight Mission*. IPGP, JPL, CNES, ETHZ, ICL, MPS, ISAE-Supaero, LPG, MFSC, doi:10.18715/SEIS.INSIGHT.XB_2016.

Seydoux, L., Balestriero, R., Poli, P., Hoop, M.d., Campillo, M. & Baraniuk, R., 2020. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning, *Nat. Commun.,* **11**(1), 1–12.

Spiga, A. *et al.*, 2021. A study of daytime convective vortices and turbulence in the martian planetary boundary layer based on half-a-year of insight atmospheric measurements and large-eddy simulations, *J. geophys. Res.: Planets,* **126**(1), e2020JE006511, doi:10.1029/2020JE006511.

Stähler, S.C. *et al.*, 2021. Seismic detection of the martian core, *Science,* **373**(6553), 443–448.

Stott, A.E. *et al.*, 2021. The site tilt and lander transfer function from the short-period seismometer of insight on mars, *Bull. seism. Soc. Am.,* **111**(6), 2889–2908.

Williams, C.K. & Rasmussen, C.E., 2006. *Gaussian Processes for Machine Learning,* Vol. **2,** MIT Press, Cambridge, MA.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

MachineLearningandMarsquakes_Detections.zip

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX A: DATA CHANNELS

APPENDIX B: NEURAL NETWORK TUNING

APPENDIX C: LIST OF POSSIBLE NEW DETECTIONS

Table A1. Data channels.

Data type	Channels				
	20 sps	10 sps	1 sps	0.5 sps	0.1 sps
VBB velocity U, V and W	02.BHU	03.BHU			
Wind speed boom 1 and 2			10.LWS	10.VWS	13.VWS
Wind direction boom 1 and 2			10.LWD	10.VWD	13.VWD
Air temperature boom 1 and 2			10.LKO	10.VKO	13.VKO
Pressure	12.BDO	13.BDO	10.LDO		

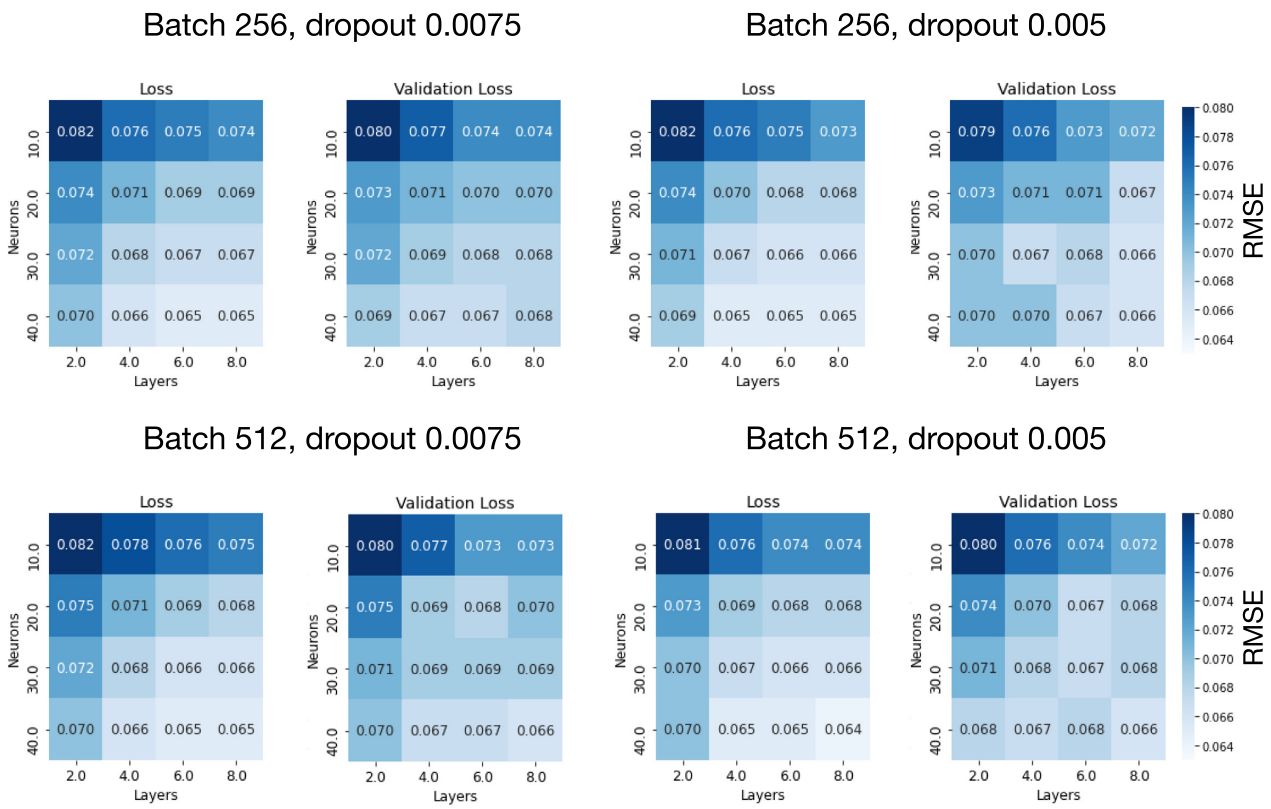

Figure B1. NN training losses for the selection of the LF seismic energy prediction model. The training and validation loss are shown for a range NN topologies, batch size and dropout rates. Each square contains the average RMSE for either the test and training loss from the last 10 epochs after 500 epochs of training.

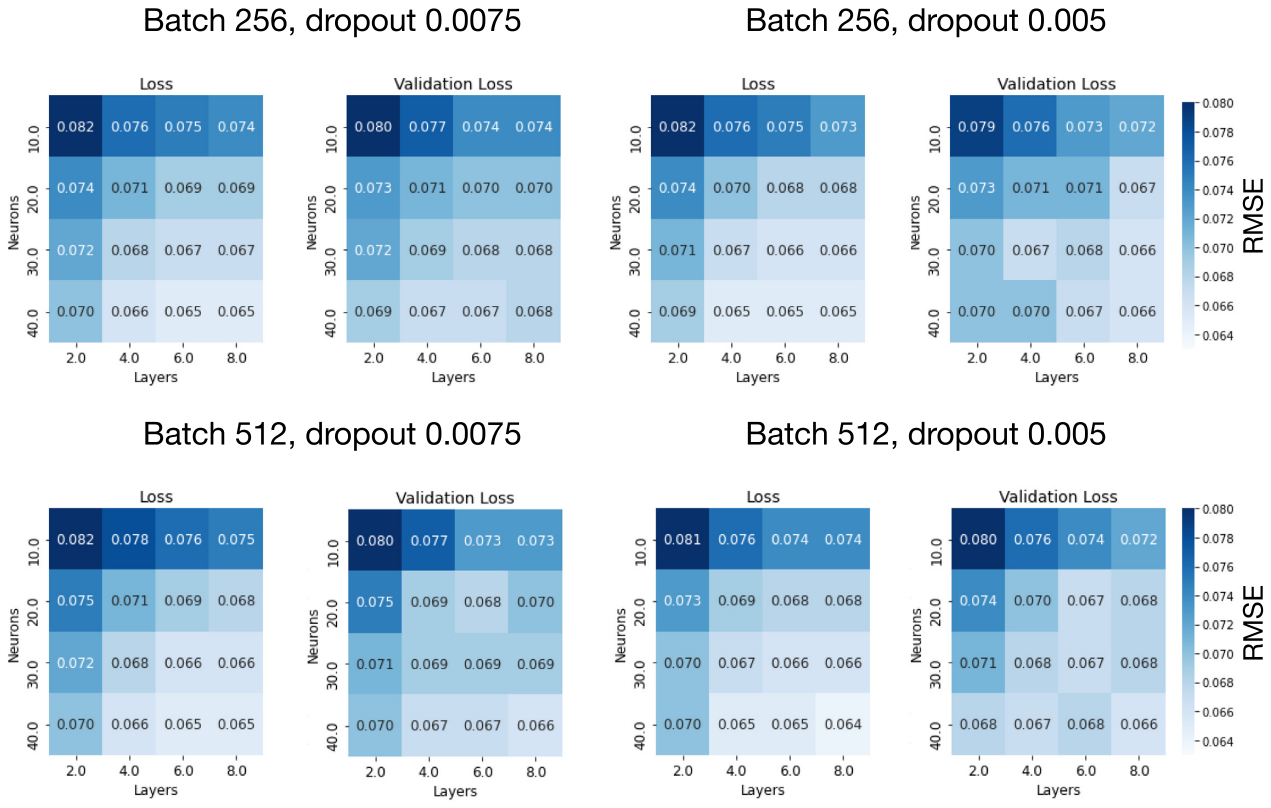

Figure B2. NN training losses for the selection of the HF seismic energy prediction model. The training and validation loss are shown for a range NNs topologies, batch size and dropout rates. Each square contains the average RMSE for either the test and training loss from the last 10 epochs after 500 epochs of training.

Table C1. Possible LF-type events.

Event sol	UTC time	SNRZ LF
220	2019-07-11T02:47:37	1.37
232	2019-07-23T11:40:07	1.23
255	2019-08-15T22:55:39	1.91
260	2019-08-21T02:57:43	1.44
356	2019-11-27T21:58:22	1.33
379	2019-12-21T10:38:33	1.44
428	2020-02-09T19:48:02	1.49
692	2020-11-06T19:22:59	1.5
882	2021-05-21T06:49:52	1.64

Table C2. Possible HF-type events.

Event sol	UTC time	SNRZ HF
177	2019-05-27T04:58:12	1.59
193	2019-06-13T09:38:01	1.43
199	2019-06-18T19:19:02	1.56
199	2019-06-18T21:39:21	1.44
204	2019-06-24T02:03:10	1.48
223	2019-07-14T06:12:19	2.0
235	2019-07-26T14:48:25	1.26
258	2019-08-19T03:40:06	1.37
301	2019-10-01T16:43:31	1.41
312	2019-10-13T14:02:38	1.38
320	2019-10-21T21:37:39	1.58
329	2019-10-31T03:00:42	1.37
332	2019-11-03T03:57:15	1.45
335	2019-11-06T08:10:31	1.46
336	2019-11-07T04:51:20	1.82
337	2019-11-08T06:43:48	1.48
342	2019-11-13T09:21:07	1.48
350	2019-11-21T15:14:01	1.34
354	2019-11-25T21:14:48	1.35
355	2019-11-26T18:57:03	1.52
356	2019-11-27T20:47:56	1.5
381	2019-12-22T19:11:19	2.09
391	2020-01-02T17:50:11	1.39
398	2020-01-09T21:00:15	1.61
403	2020-01-14T09:25:40	1.54
429	2020-02-10T20:19:53	1.45
441	2020-02-23T01:40:34	1.61
455	2020-03-08T12:14:32	1.6
457	2020-03-10T12:52:55	1.58
497	2020-04-20T14:39:15	1.77
502	2020-04-25T07:18:44	1.72