Updated global reference models of broadband coherent infrasound signals for atmospheric studies and civilian applications

Samuel K. Kristoffersen[1], Alexis Le Pichon[2], Patrick Hupe[3], and Robin S. Matoza[4]

[1]CEA, DAM, IDF
[2]CEA, DAM, DIF
[3]BGR
[4]University of California, Santa Barbara

November 26, 2022

Abstract

The International Monitoring System (IMS) infrasound network has been established to detect nuclear explosions and other signals of interest embedded in the station specific ambient noise. The ambient noise can be separated into coherent infrasound (e.g. real infrasonic signals) and incoherent noise (such as that caused by wind turbulence). Previous work statistically and systematically characterizing coherent infrasound recorded by the IMS. This paper expands on this analysis of the coherent ambient infrasound by including updated IMS datasets up to the end of 2020, for all 53 of the currently certified IMS infrasound stations using an updated configuration of the Progressive Multi-Channel Correlation (PMCC) method. This paper presents monthly station dependent reference curves for the back azimuth, apparent speed, and root-mean squared amplitude, which provide a means to determine the deviation from nominal monthly behaviour. In addition, a daily Ambient Noise Stationarity (ANS) factor based on deviations from the reference curves is determined for a quick reference to the data quality compared to the nominal situations. Newly presented histograms provide a higher resolution spectrum, including the observations of the microbarom peak, as well as additional peaks reflecting station dependent environmental noise. The aim of these reference curves is to identify periods of sub-optimal operation (e.g. non-operational sensor) or instances of strong abnormal signals of interest.

Updated global reference models of broadband coherent infrasound signals for atmospheric studies and civilian applications

S. K. Kristoffersen[1], A. Le Pichon [1], P. Hupe[2], and R. S. Matoza[3]

[1]CEA, DAM, DIF, 91297 Arpajon, France
[2]BGR, B4.3, Hannover, Germany
[3]Department of Earth Science and Earth Research Institute, University of California, Santa Barbara, CA, USA

Key Points:

- A new and updated analysis of the coherent ambient infrasound is presented
- Monthly reference back azimuth, horizontal velocity, and RMS amplitude curves for each station are provided with descriptions
- Updated PMCC algorithm allows for additional peaks in the coherent RMS amplitude histogram spectra to be resolved

Corresponding author: S.K. Kristoffersen, `samuel.kristoffersen@cea.fr`

Abstract

The International Monitoring System (IMS) infrasound network has been established to detect nuclear explosions and other signals of interest embedded in the station specific ambient noise. The ambient noise can be separated into coherent infrasound (e.g. real infrasonic signals) and incoherent noise (such as that caused by wind turbulence). Previous work statistically and systematically characterizing coherent infrasound recorded by the IMS. This paper expands on this analysis of the coherent ambient infrasound by including updated IMS datasets up to the end of 2020, for all 53 of the currently certified IMS infrasound stations using an updated configuration of the Progressive Multi-Channel Correlation (PMCC) method. This paper presents monthly station dependent reference curves for the back azimuth, apparent speed, and root-mean squared amplitude, which provide a means to determine the deviation from nominal monthly behaviour. In addition, a daily Ambient Noise Stationarity (ANS) factor based on deviations from the reference curves is determined for a quick reference to the data quality compared to the nominal situations. Newly presented histograms provide a higher resolution spectrum, including the observations of the microbarom peak, as well as additional peaks reflecting station dependent environmental noise. The aim of these reference curves is to identify periods of sub-optimal operation (e.g. non-operational sensor) or instances of strong abnormal signals of interest.

1 Introduction

The International Monitoring System (IMS) includes a global network of infrasound arrays designed to detect atmospheric nuclear explosions and to monitor compliance with the Comprehensive Nuclear Test-Ban Treaty (CTBT) (Campus & Christie, 2009). However, the network is capable of detecting many additional atmospheric infrasonic sources, including but not limited to, seismic activity (de Groot-Hedlin & Hedlin, 2019), volcanoes (Matoza et al., 2019), and atmospheric convection (e.g. thunderstorms) (Waxler & Assink, 2019).

The determination of the sources of signals is greatly dependent on the propagation conditions of the atmosphere (Norris et al., 2009; Kulichkov, 2009) and the local environmental and instrumental effects on the operation. The atmospheric propagation conditions can affect the locations from which an infrasound signal of interest can be detected, and, therefore, the determination of the source location. Due to refraction and reflection of infrasound, the atmosphere can act as a waveguide, allowing for the observation of infrasound source events from large distances (Garcés et al., 1998; Waxler & Assink, 2019; Norris et al., 2009). This type of wave-guide is formed when the effective speed of sound becomes larger than that of the source location (Gabrielson, 2006). The effective speed of sound is the speed of sound as a function of both the atmospheric temperature, and the local wind speed, such that

$$c_{eff} = \sqrt{\gamma R T(\vec{r})} + u_0(\vec{r}), \tag{1}$$

where γ is the ratio of the specific heat at constant pressure (c_p) and the specific heat at constant volume (c_V), R is the universal gas constant, $\vec{r}$ is the location in the atmosphere, T is the temperature in K, and $u_0(\vec{r})$ is the local wind speed at $\vec{r}$. Therefore, seasonal changes in the global circulation patterns can have significant effects on infrasound propagation.

Instrumental parameters, which differ from station to station, can also have a substantial impact on infrasound observations. The size and shape of the sensor array affects the precision, accuracy and detectable frequency of measured infrasound signals. A larger aperture allows for higher precision measurements, however, there is also an increase in the sensitivity to noise and other sources of uncertainty and an increase in the risk of aliasing, which reduces the capability to measure at higher frequencies. The solution to minimize these effects is to include a small sub-array, which increases the maximum observable frequency by reducing the risks of aliasing while maintaining a high precision (Marty, 2019).

Since the identification and location of infrasound sources is of critical importance for this infrasound monitoring network, signals of interest must be distinguishable from noise and clutter, where noise refers to incoherent signal and clutter refers to repetitive coherent infrasonic signals (Evers & Haak, 2001; Ceranna et al., 2019) which are not a signal of interest. It is therefore important to be able to determine typical seasonal conditions at each station to identify anomalous conditions and measurements. Providing this informal data improves the capability of the IMS infrasound network to identify signals of interest versus coherent (spatially correlated) and incoherent (spatially uncorrelated) noise. This also provides indications when data quality is poor, and expedites the identification of potential environmental and/or instrumental damage in need of operator intervention.

To determine the wave parameters from the infrasound signals, the Progressive Multi-Channel Correlation (PMCC) method was used (Cansi, 1995; Cansi & Le Pichon, 2008). This algorithm estimates the back azimuth, horizontal (apparent) velocity, and amplitude for logarithmically increasing frequency bands between 0.01 and 4 Hz. Probability Density Functions (PDF) of the Power Spectral Density (PSD) of the raw signals provide an additional tool by which the infrasound signals can be assessed (McNamara & Buland, 2004).

Previous works have provided measures of the incoherent noise (Bowman et al., 2005; Brown et al., 2012; Marty et al., 2021) and measures of the ambient coherent infrasound (Matoza et al., 2013); both globally and at each station. The goal of this work is to provide an update of the previous ambient coherent infrasound signals at each station. This includes updated data (all available stations up to 2021), an updated PMCC routine with 26 (instead of 15) frequency bands (Hupe et al., in review) and 2D histograms for presenting the ambient coherent infrasound signals. These new data formats provide a means of identifying potentially anomalous signals, which are both station and month specific. These anomalous signals could be due to instrumental errors, environmental conditions, or due to geophysically interesting phenomena.

Additional background information on the IMS network, the PMCC and PDF data analysis routines are discussed in Section 2. Section 3 provides the monthly reference metrics for each station, and discuss how these data can be accessed and used. This section also introduces some examples of anomalous PMCC observations identified using these metrics. Finally, Section 4 provides an overview and discussion of these results and the reference data.

2 Data and Methods

The IMS is a large network for monitoring nuclear testing using seismic, infrasound, hydroacoustic, and radionuclide observations. There are 60 (planned or constructed) stations which comprise the infrasound network of the IMS (see Figure 1). The construction of this network began in 2001, with 37 stations being built in the first 5 years (Marty, 2019), and currently consists of 53 certified stations. This global network provides excellent coverage of infrasound events, with more than 18 years of data available for many of the stations (see Figure 1). In the last three to four years the data availability has been

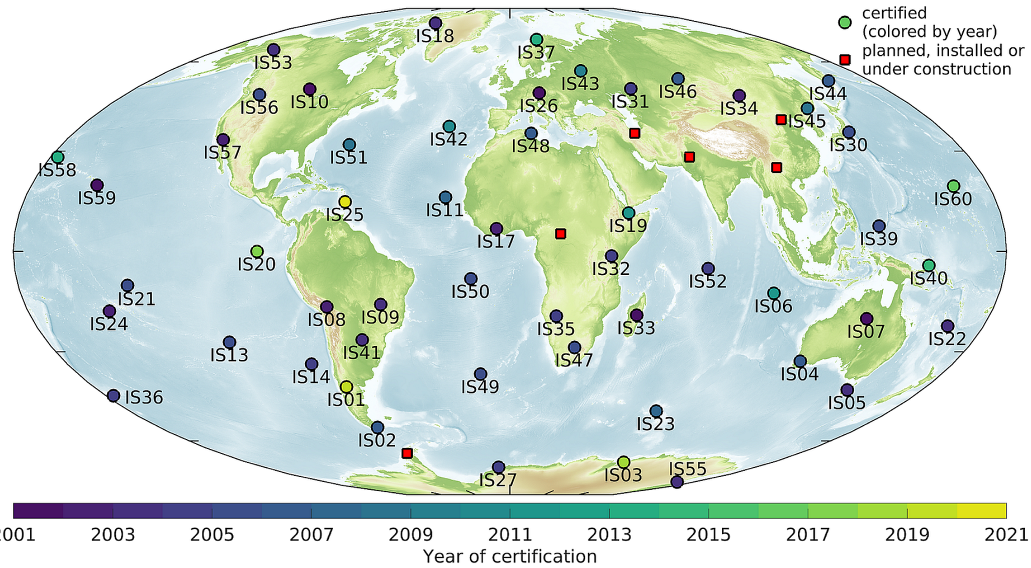

Figure 1. Map of the IMS infrasound stations, the circles are certified stations, and the squares are stations that are either planned or under construction; only certified stations (circles) were used for this study. The colour of the circles denotes the year of certification (Hupe et al., in review).

quite consistent for all of the currently constructed stations. With 55 stations currently completed or under construction, the network is nearing completion.

For this research, data from 53 certified IMS infrasound stations for up to eighteen years per station were used (Figure 1). Data availability varies from station-to-station, predominantly due to when each station was commissioned, with seven stations out of the 60 station network still to be certified and/or built. Typically, once a station is certified, there are few data gaps. Many of these data gaps are related to station upgrades, such as the installation of additional sensors, or the improvement of the wind-noise reduction system (Le Pichon et al., 2012).

2.1 European Centre for Medium-Range Weather Forecasts (ECMWF) Model

The High Resolution (HRES) ECMWF model provides wind and temperatures (along with many other quantities not used in this study) at a horizontal resolution of 0.25° (28 km at the surface) with a temporal resolution of 6 hours. There are 137 vertical levels from the surface up to 0.01 hPa. The local speed of sound and wind speeds at each station are determined by taking the surface level, and the horizontal grid value which contains the station. The stratospheric winds and temperatures are simply chosen from the appropriate vertical levels (40 to 60 km) of the model data.

This method of calculating the winds, speed of sound, and effective speed of sound was chosen out of simplicity, and due to the ease of access to this data versus acquiring all of the individual stations' data. In addition, this allows for a self-consistency between the local (ground) effective speed of sound, and that calculated for the stratosphere, which provides a more robust measure of the ratio of the effective speed of sounds. The drawbacks of this approach are that the data is for a relatively large grid and it does not account for the local environmental conditions, such as the reduced wind observed in a forested area, or due to other artificial wind fences etc. It was determined that these trade-offs were worthwhile for this study, but should be kept in mind when interpreting the data.

2.2 Power Spectral Density

The Power Spectral Density (PSD) is acquired by taking a normalized periodogram estimate (McNamara & Buland, 2004), such that

$$P = \frac{2}{NF_s} |Y|^2 , \tag{2}$$

where P is the power, N is the length of the time-series, F_s is the sampling frequency (Hz), and Y is the FFT of the time-series. Following Welch's method (McNamara & Buland, 2004), both the PSD and the corresponding Probability Density Function (PDF) are determined. The PDF shows the relative probability of signal at each power and frequency.

2.3 Progressive Multi-Channel Correlation

The PMCC method calculates the time-delay of arrival (TDOA) of the signals between all pairs of sensors to detect a signal of interest (SOI) in a noisy time-series (Cansi, 1995). The TDOA is determined for each combination of sensor pairs in the full sensor array.

A measure of the self-consistency of the TDOA is made using the combinations of three sensors, and all of the relative time-delay of arrivals, providing the closure relation,

$$r_{ijk} = \Delta t_{ij} + \Delta t_{jk} + \Delta t_{ki}, \tag{3}$$

where Δt_{ij} is the time-delay of arrival between the ith and jth elements, respectively (Cansi & Klinger, 1997; Cansi & Le Pichon, 2008). By summing these TDOA's, the total result, for a perfect plane wave with no noise, would be 0. The first set of three sensors is typically chosen such that the array size is small (i.e. the sensors are close together). By starting with a small array, the risk of aliasing issues, and errors caused by decoherence and noise are reduced, at the expense of measurement precision. Next, the consistency is determined, such that

$$C_n = \sqrt{\frac{6}{n\,(n-1)\,(n-2)} \sum_{i>j>k} r_{ijk}^2}, \tag{4}$$

where n is the number of sensors in a sub-array which is an element of the entire sensor array (Cansi & Klinger, 1997; Cansi & Le Pichon, 2008). If the consistency is less than a certain threshold, $C_{threshold}$, then a detection is flagged, and more sensors are progressively added to the sub-array while maintaining the consistency below $C_{threshold}$. For the progressively larger arrays, the additional sensors are added with the restraint that the TDOA using the new sensor corresponds to the maximum that is nearest to the computed TDOA using the previous sub-array. As more sensors are included, the aperture of the sub-array, and the precision, both increase.

Once the total sub-array for a measurement and the consistency are determined, the horizontal (apparent) speed and the azimuth are determined from the TDOA arrival values and the array geometry. The slowness, $\vec{S}$, can be determined from the TDOA's.

This process is repeated for each frequency band, and time window by applying a Chebyshev bandpass filter to the time-series. An value of back azimuth and velocity is determined for each time and frequency band. These PMCC pixels are then grouped based on similar values of back azimuth, speed, frequency, and time by considering a weighted distance between PMCC pixels (Cansi & Klinger, 1997),

$$d(p_1, p_2) = \sqrt{\left(\frac{t_2 - t_1}{t}\right)^2 + \left(\frac{f_2 - f_1}{f}\right)^2 + \left(\frac{v_2 - v_1}{v}\right)^2 + \left(\frac{\theta_2 - \theta_1}{\theta}\right)^2}. \tag{5}$$

Weights can be adapted for each parameter. Then, considering the pixels in the group, the corresponding values of back azimuth, speed, frequency and time are determined by taking the average of each of those contained in this grouping.

2.4 Ambient Noise Stationarity Factor

When observing the PMCC results, it is important to be able assess the data quality, and therefore the likelihood that an observation is due to a real event, or is an anomalous outlier. To do this, an Ambient Noise Stationarity (ANS) factor was developed for each day in the available data set for all the available stations. This factor was developed empirically using the available station data.

For each station, the data were aggregated by month, and the number of detections as a function of the back azimuth, horizontal speed or RMS amplitude. These form the monthly reference curves described in more detail in Section 3.3. The ANS factor for each of the three quantities of interest (back azimuth, horizontal speed, and RMS amplitude) was determined by finding the daily histograms of these quantities, and taking the square of the Pearson correlation coefficient between the monthly reference curves and the daily data (Manders et al., 1992; Mohapatra & Weisshaar, 2018). The resulting ANS factor is a value between 0 and 1, with larger values indicating the data better fits with the typical distribution for that station and month of year. The total ANS factor was calculated using the product of these three intermediate ANS factors, such that

$$Q = Q_{Az} Q_v Q_{Amp} \tag{6}$$

where Q_{Az} is the azimuth ANS factor, Q_v is the speed ANS factor, and Q_{Amp} is the RMS amplitude ANS factor. The product was chosen since, if any single ANS factor is small, this will result in the total ANS factor being small.

2.5 PMCC Metrics

To determine instances of abnormal results, it is necessary to determine metrics by which to measure and compare the PMCC outputs. Figure 2 shows an example of the style of figure used for this study, this example is for IS26 (Germany) in 2015.

The first panel contains the back azimuth (y-axis) of the observed detections as a function of the time (x-axis) and frequency (colour). The gray-scale plot on the same panel is the ratio of the maximum effective speed of sound in the stratosphere (40 to 60 km) to the speed of sound at the surface (determined from the ECMWF model data). The scale of c_{eff} ratio is a gray gradient from 0 to 1, above which all of the values are white. This choice of scale was to provide a reference for when propagation would be per-

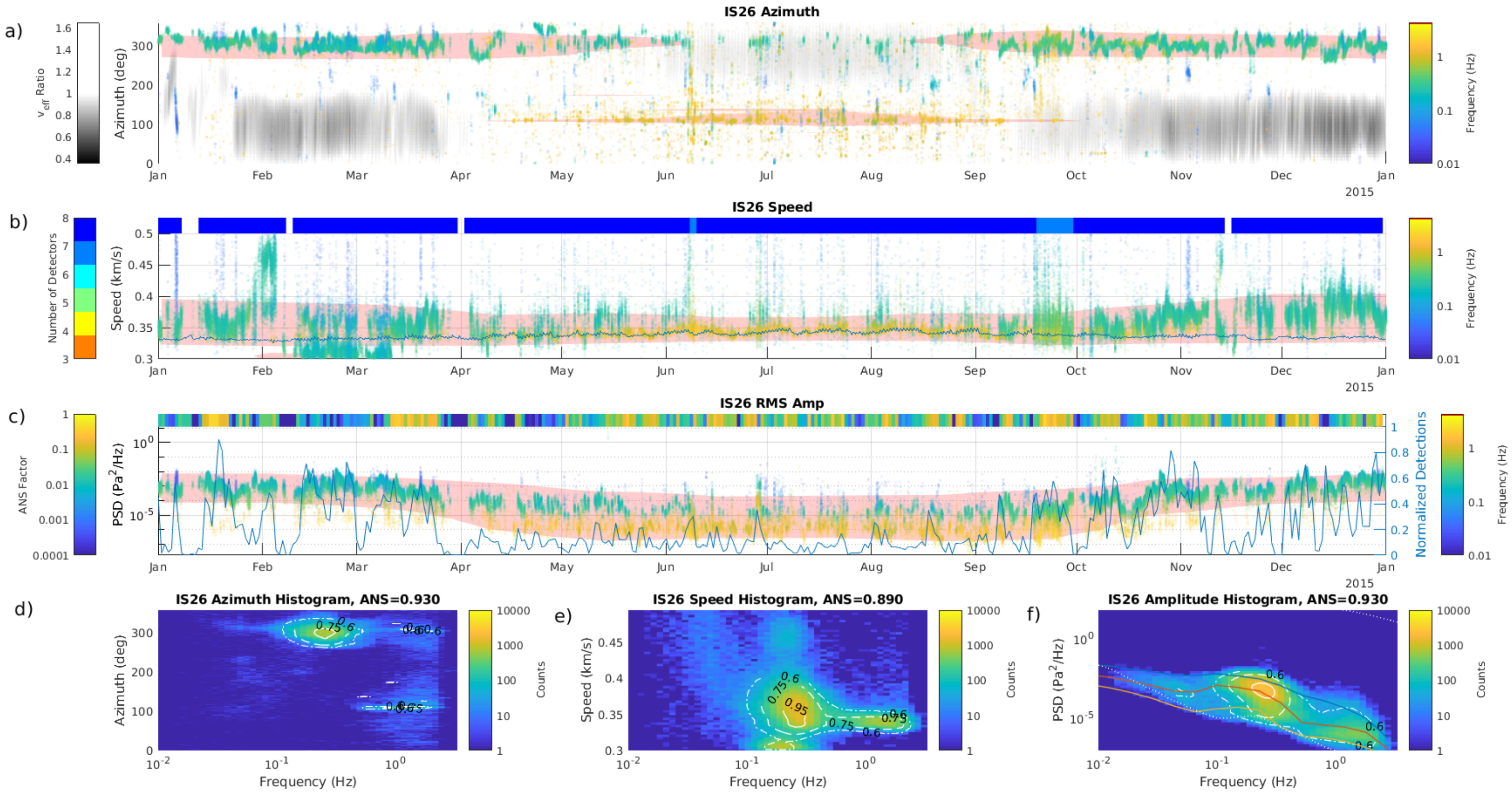

Figure 2. a) is the back azimuth (degrees), b) is the speed, and c) is the PSD of the RMS amplitude (amplitude squared divided by the frequency band), the colour denoting the frequency. The histograms on the bottom show the number of counts (colour) as function of the frequency in Hz (x-axis) and (from left to right) d) the back azimuth, e) speed, and f) PSD for the entire year. Additional information regarding this figure is found in the text.

mitted in the stratosphere. When the c_{eff} ratio is greater than approximately 1, propagation would not be permitted through the stratosphere resulting in reflection. Therefore, any instance when the gray-scale is white, propagation from great distances via the stratosphere waveguide would be permitted. Locally generated sources with incident angles close to horizontal would still be possible; these signals would have horizontal speeds close to the speed of sound at the ground. The red shading represents the regions in which the number of detections are greater than 0.75 of the maximum number of detections over all the years; this provides a guide to where the detections are most likely to be expected.

The second panel contains a similar plot, but for horizontal speed (y-axis) as a function of time (x-axis) and frequency (colour). The coloured bar at the top shows the number of available sensors per day. This provides a quick reference to know when a sensor was not functioning, and could explain potential anomalies. As can be seen for this particular year and station, aside from a few short blips, the full complement of 8 sensors was operational during the entire year. The blue line shows the speed of sound at ground level (from ECMWF model data). This line effectively shows the minimum value for the velocity measurements. Note that the high frequency speeds (yellow), which would typically be from local sources, follow the local speed of sound. Instances of speeds less than the local speed of sound (e.g. during Febraury and March) would be non-physical, and therefore require additional study. Similar to the first panel, the red contours represent regions where the number of detections is greater than 0.75 of the maximum for the entire data-set.

The third panel contains a similar plot to the first two panels, but for the PSD of the root mean squared (RMS) amplitude, that is the amplitude squared divided by the frequency band for that detection. Each PMCC family is the average over a range of frequencies. The RMS amplitude is on the y-axis, time is on the x-axis, and frequency is the colour. The blue line shows the number of detections per day; this is another useful parameter, as a sudden increase or decrease in the number of daily detections could be indicative of environmental or instrumental issues. Again, the red contours represent the 0.75 of the maximum level for detections. The coloured bar at the top is a daily ANS factor, described in Section 2.4. The colour-scale for this ANS factor is chosen such that blues and greens should be considered suspect or bad data (or perhaps something real but abnormal has occured), and yellows and oranges can be considered to be more typical, with yellow being the highest data quality.

The next three panels are of 2D histograms, one for the back azimuth (left), the second for the horizontal speed (middle), and the third for the RMS amplitude (right). The colour of the histograms is the number of counts of azimuth (speed, or amplitude) for a particular frequency (in Hz) for the given year (2018 in this case). The isocontours are the 0.95, 0.75, and 0.6 levels of the normalized (by the maximum value) histograms for the entire data-set. The ANS factor was determined for each year by comparison with the entire data set in a similar manner to that described in Section 2.4. The coloured lines on the amplitude histogram are the 95th, 50th and 5th percentiles; the solid white line is the median modelled noise (Bowman et al., 2005), and the dotted white lines are the upper and lower modelled noise limits.

In addition to being an example template, Figure 2 also provides some examples of results which could be considered anomalous, and warrant further investigation. Looking at the velocity histogram, there are considerable amounts high speed observations in the low to mid-frequency range. From the second panel, there are sporadic increases in the spread of observed velocities (particularly during times when there are fewer than the maximum number of available sensors), and large, mid-frequency velocity observations at the beginning of February. There are also a considerable number of sub-speed of sound observations during February and early March, which are non-physical, and thus would require additional investigation to determine the cause.

3 Results

The main results of this study are the calibration curves and data, which provide a reference by which the yearly and monthly station data can be compared. These curves and data allow for the identification of anomalous data or events, which can then be studied in further detail to identify the cause of the the atypical observations. These could be due to instrumental errors, local environmental conditions, or something of geophysical interest. In any of these cases, it is useful to identify these anomalies, whether to remove potentially erroneous data, or to identify some atypical geophysical phenomenon.

3.1 Station Reference Data

The first type of reference curve, which has been calculated for each station for each available year up to 2020, is generated from the histograms of the back azimuth, the horizontal speed, and the RMS amplitude. These type of figures are provided in the summary plots for each station for each year (see, for example, Figure 2). The reference contours can be used to determine if a year's data deviates substantially from the average yearly data.

The second type of reference curve is the monthly plot for each station. These reference curves are similar to the histograms shown in panels d), e) and f) of Figure 2 (showing the back azimuth, horizontal speed and the RMS amplitude), but are exclusive to each month for each station. Thus, there are 12 reference curves (one for each month) per available station.

The data for the histograms and reference curves are also saved, for comparisons with monthly PMCC data products for every station (see section 3.3 for a more detailed explanation). The importance of the monthly data products is the seasonal variability which is observed in the PMCC detections.

Given the updated data-sets, new versions of the ambient coherent signals for all of the available stations are produced. An update of the coherent signal and the incoherent noise, similar to what is shown in Matoza et al. (2013), is provided in Figure 3. The prominent microbarom peak is clear in the coherent signal spectrum (the orange colour in panel a), and a similar leveling off at low frequency, as seen in Matoza et al. (2013). There is also a second peak observed at high frequency, which is accompanied by a precipitous decrease in amplitude at the highest frequencies. The effects of the local wind speed on the detections, panel b) in Figure 3, result in the observed relative increase of detections with increasing frequency. This panel shows the fraction of the detections observed above a given wind speed threshold; the wind speeds were determined using the ECMWF model data at 1013.25 hPa. As expected, the cumulative number of detections decreases with increasing wind speed, with the number of detections at lower frequency decreasing more rapidly with increasing wind speed. This is due to natural 'red' spectrum of the wind noise, and the filtering systems used to reduce wind noise being much more effective at higher frequencies (Marty, 2019).

Figure 4 shows the ambient coherent noise histogram observed over all of the IMS infrasound stations. It is of interest that these updated curves more clearly show the mi-

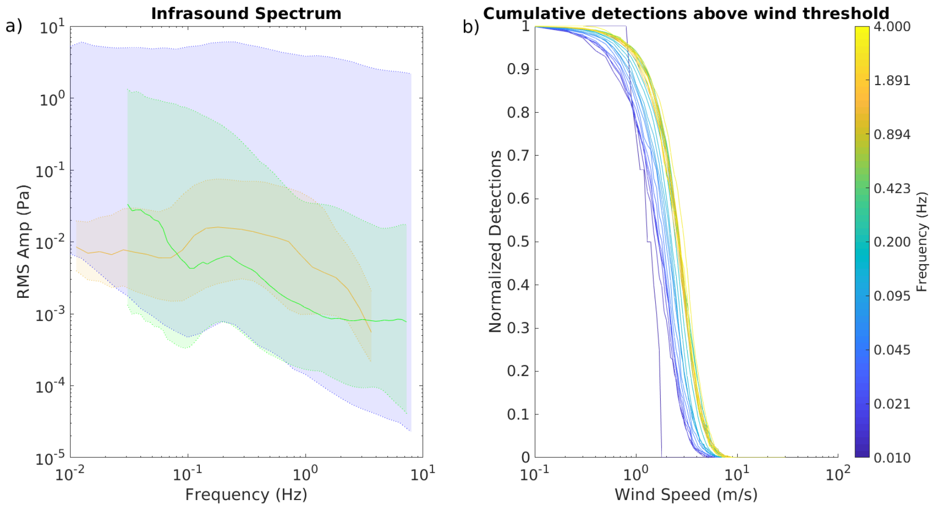

Figure 3. a) RMS Amplitude spectra, orange is the coherent signal determined from the PMCC pixels, green is the Bowman et al. (2005) incoherent noise model and blue is the Marty et al. (2021) incoherent noise model for all available IMS infrasound stations. The solid lines denote the median, while the shaded areas cover from the 5th to the 95th percentiles. b) cumulative detections for all stations observed above a given wind speed (x-axis); the colours denote the frequency of the detections. Note: the Marty et al. (2021) model does not contain a median measurement.

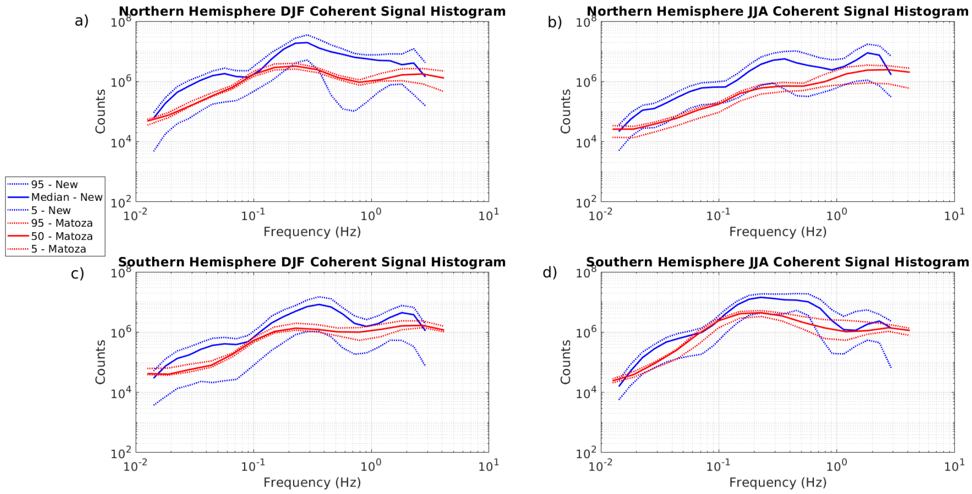

Figure 4. Number of coherent pixels detected per season per hemisphere for all IMS infrasound stations in each frequency band. a) Northern hemisphere winter, b) Northern hemisphere summer, c) southern hemisphere summer, d) southern hemisphere winter. The solid lines are the median number of counts per season, and the dashed lines are the 95 and 5 percentiles. The blue lines are for the new PMCC processing and updated data-set, and the red lines are for the previous results presented by Matoza et al. (2013).

crobarom peak at 0.2 Hz, and a prominent peak at around 2 Hz, which were not observed in the all-station results acquired by Matoza et al. (2013). As these 2 Hz observations are seen at all of the IMS stations, the source must be either global, or there are enough local sources in this band that every station observes some sources in this band (such as surf, atmospheric convection, volcanoes etc.). Similar to Matoza et al. (2013), very low observation rates are observed at the lowest frequency bands, as well as at the highest frequency band. This could be due to the relatively broad (2nd order) filtering applied to each band, which would allow more signal from outside bands to be observed. This results in higher (lower) frequency pixels observed in the lowest (highest) frequency band, moving the mean (observed) frequency up (down) to the next frequency band. Another possible explanation for the low frequency behaviour, as described by Matoza et al. (2013), is that the low-frequency, coherent signal is only observable when the incoherent wind noise is sufficiently low.

Finally, Figures 5 and 6 show the coherent RMS amplitude (in Pa) for every available station (sorted by latitude) as a function of the frequency. Each Figure is subdivided into two panels; Figure 5 shows the DJF and MAM, and Figure 6 shows JJA and SON. The seasons (with respect to the northern hemisphere) are determined such that December, January, and February comprise winter; March, April, and May comprise spring; June, July, and August comprise summer; and September, October, and November comprise autumn. Strong seasonal variations are observed between winter and summer. Larger amplitudes are observed during the winter season, particularly in the mid-frequency band (0.08 Hz to 0.5 Hz). The spring and summer appear to be more transitional between the summer and winter, with no noticeable differences between the northern and southern hemispheres. Additionally, the equatorial stations do not exhibit this seasonal behaviour, the amplitude remaining relatively consistent throughout the year.

3.2 Assessing the ANS Factor

The ANS factor (see Section 2.4) provides a measure of the data's deviation from nominal monthly behaviour. An example of the daily ANS factor values are shown in panel 3 of Figure 2. A known event, in this case the Sarychev eruption (Matoza et al., 2011), which occurred from June 11-16, 2009, can be used to demonstrate the utility of the ANS factor. Figure 7 shows the PMCC results around the Sarychev eruption at IS30. The eruption is observed at high-frequency between 30° and 35° back azimuth. There is a noticeable decrease in the ANS factor for several days, starting on June 12 corresponding to the eruption. Other instances of low ANS factor occur around June 5 and June 20. It is likely these dips in ANS factor (especially around June 20) are due to the reduction in the number of detections which result in the significant deviation from the nominal amplitude behaviour (see Figure 8). It should be noted that the Sarychev eruption corresponds with a significant increase in the number of detections, which means the decrease in ANS factor is more significant, since it would require a larger amount of anomalous behaviour to explain this decrease in ANS factor. Figure 8 shows that there is a decrease in all three measurements' ANS factors on June 12, with an especially large decrease in the azimuth ANS factor. These result in the local minimum of the total ANS factor. Although, this shows the ability of the ANS factor to identify anomalous events, it demonstrates that one needs to consider many factors, including the number of detections.

Another example of the use of the ANS factor around the Sarychev eruption is from IS44, see Figure 9. Again, there is a decrease in the ANS factor coincident with the erup-

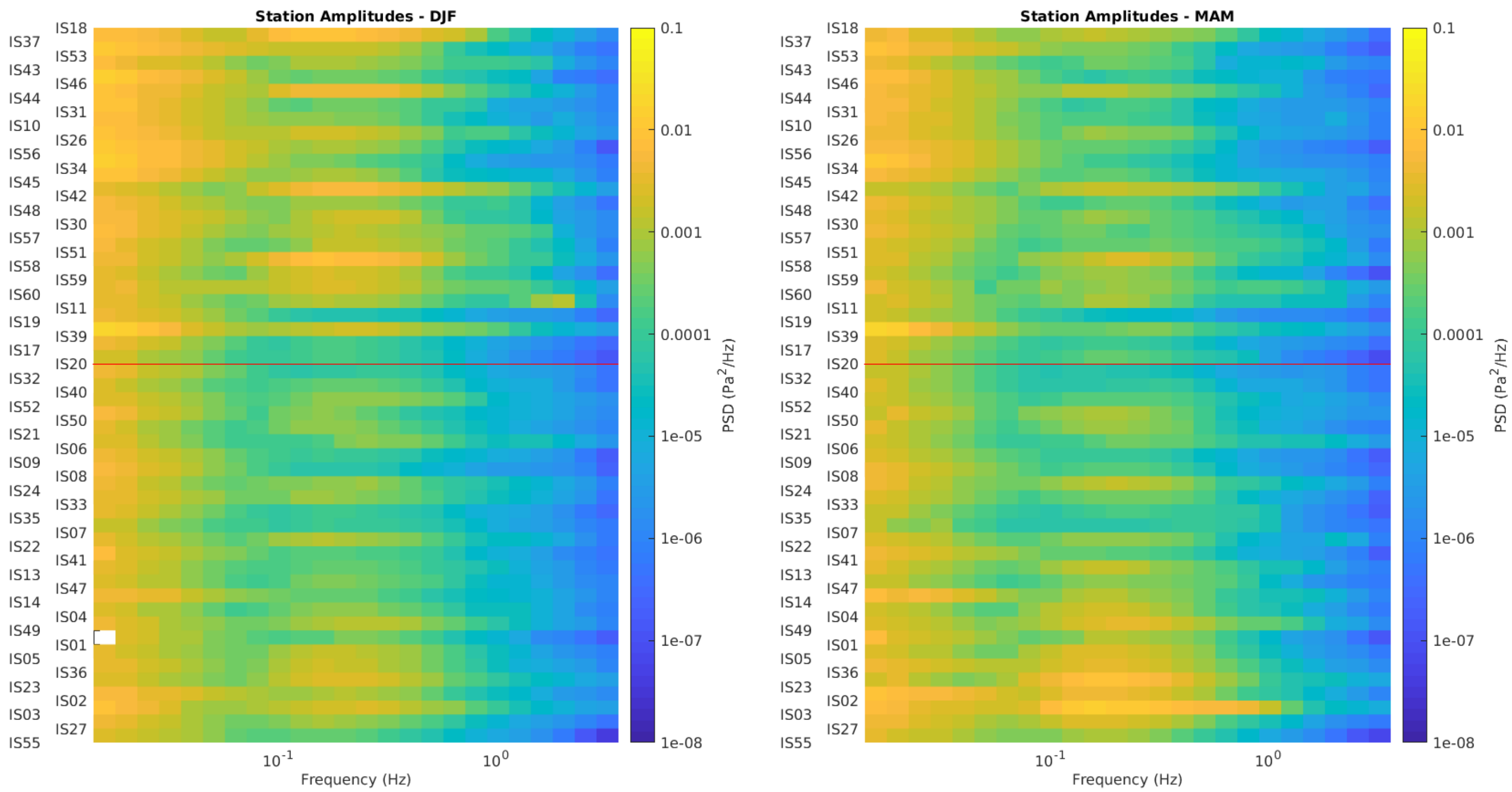

Figure 5. RMS amplitude of PMCC pixels (coherent signal) observed at 53 certified IMS stations as a function of frequency in Hz. The stations, shown on the y-axes, are sorted by latitude. The two panels show the median amplitudes for each season: a) DJF, b) MAM. The colour shows the power spectral density of the RMS amplitude in $\mathbf{Pa^2/Hz}$. The red line delineates the northern and southern hemisphere stations, with the northern stations above the line, and the southern stations below it.

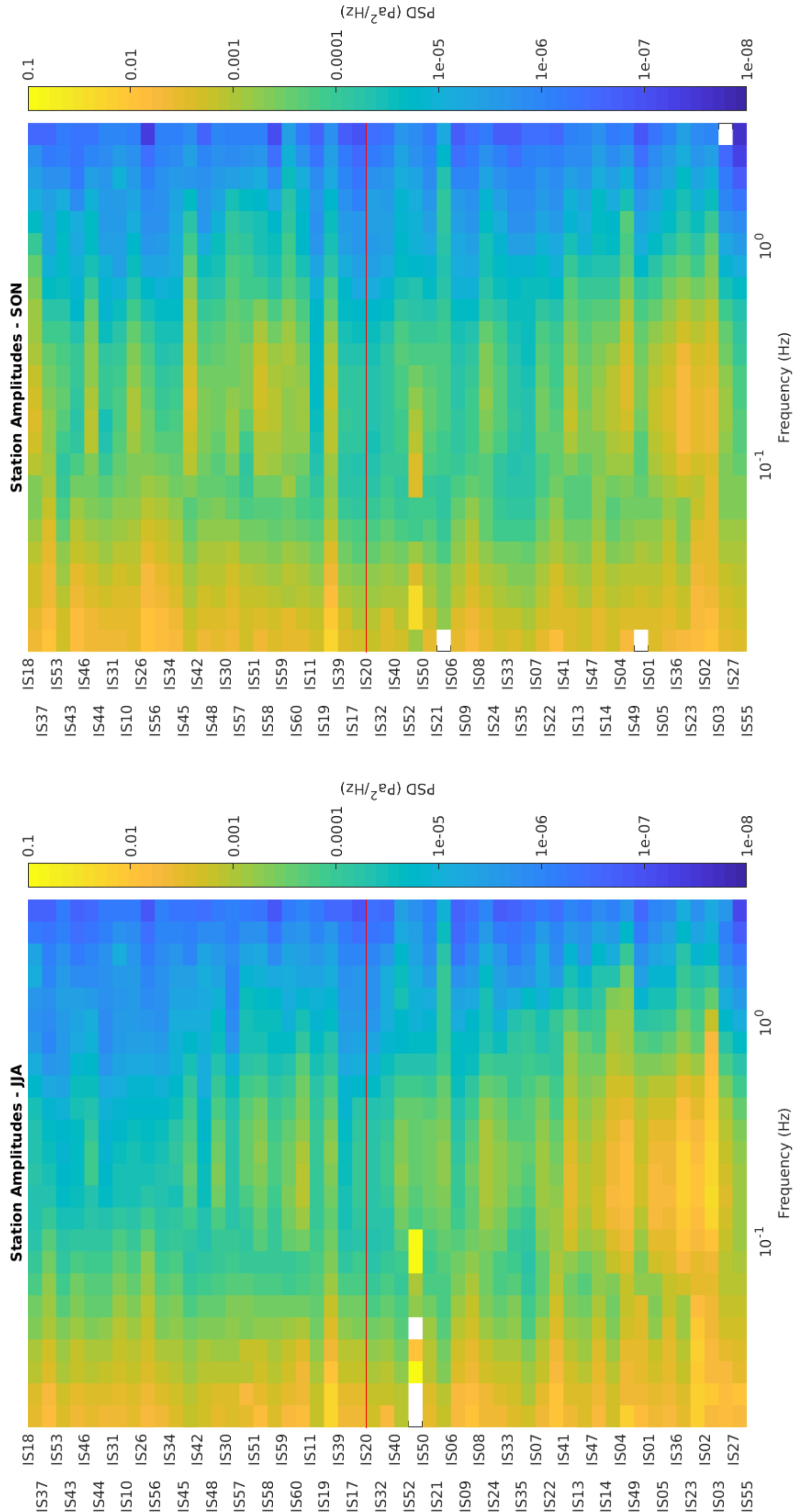

Figure 6. Same as Figure 5, but for a) JJA and b) SON.

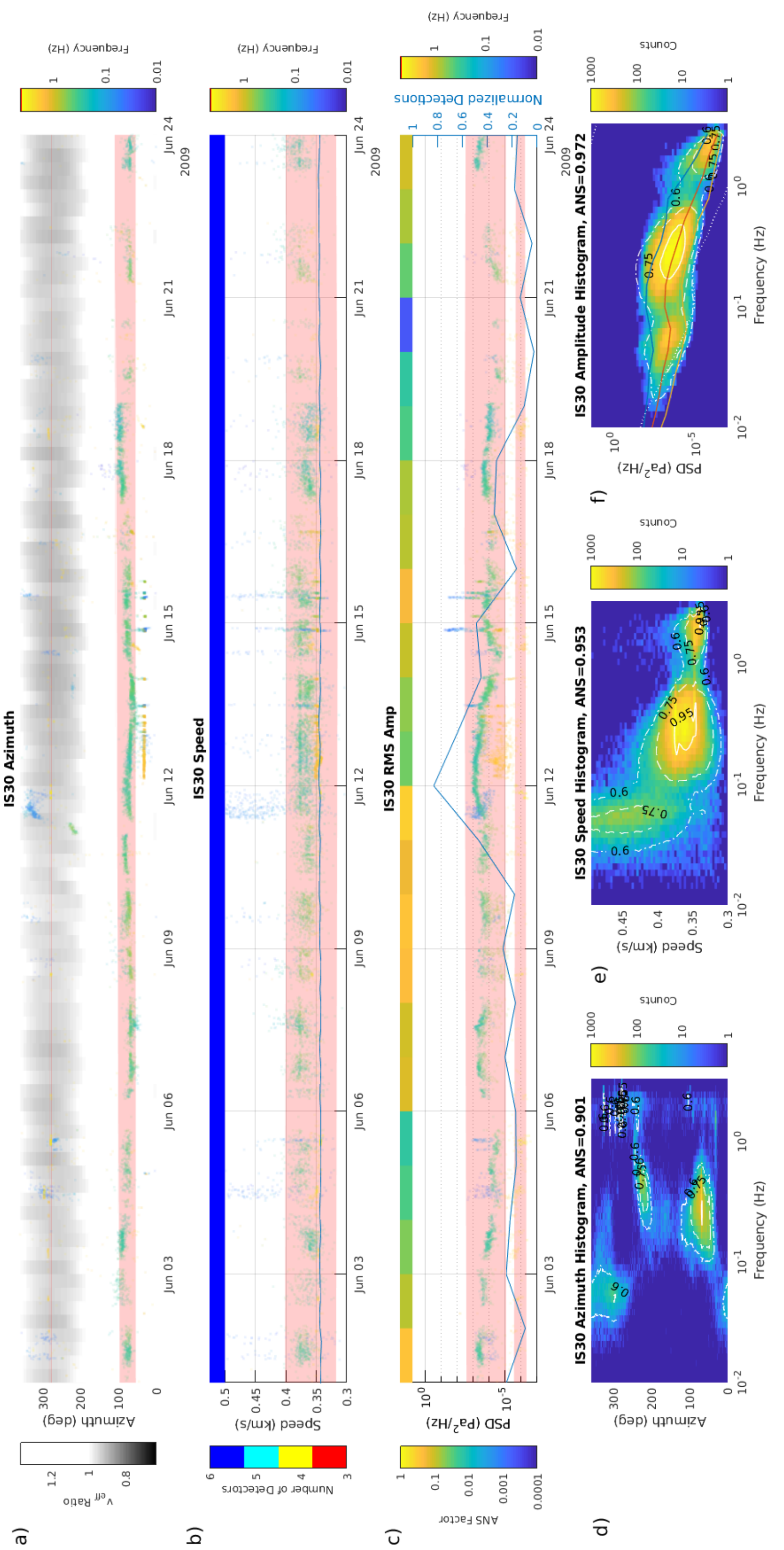

Figure 7. Same as Figure 2, but for IS30 during the time-period around the Sarychev eruption (June 11-16, 2009).

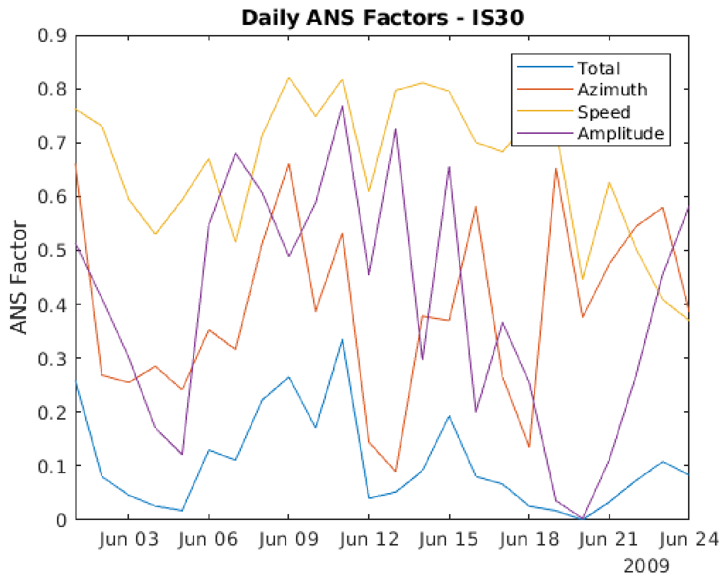

Figure 8. The total ANS factor (blue) and the azimuth (red), speed (yellow), and RMS amplitude (purple) ANS factors for IS30 around the Sarychev eruption.

tion (also note the large number of high frequency detections around 200° azimuth). There is also an increase in the number of observations during the eruption. There are two other times of low ANS factor (June 6-7 and June 9-10), which both correspond to reductions in the number of detections similar to that observed for IS30. It is also of note that very low ANS factor observed from June 6 to 7 corresponds to a period of only 3 operational sensors (see Figure 9, panel 2). This shows the ANS factor can be used to identify periods when the data quality is poor, such as when there are only the minimum of 3 operational sensors.

The ANS factor, therefore, acts as a good indicator of poor data quality or anomalous observations, but it is necessary to consider other factors when further investigating instances of low ANS. In particular, the number of detections is useful in separating cases of poor data quality from anomalous events. Finally, the azimuth, speed, and RMS amplitude observations provide further details to differentiate poor data quality from anomalous events. An anomalous event is likely to have a concentrated azimuth, velocity, and/or RMS amplitude. This is seen in the very tight azimuthal observations of the Sarychev eruption, with other low ANS factor events not showing similarly tight groupings in any of the parameters.

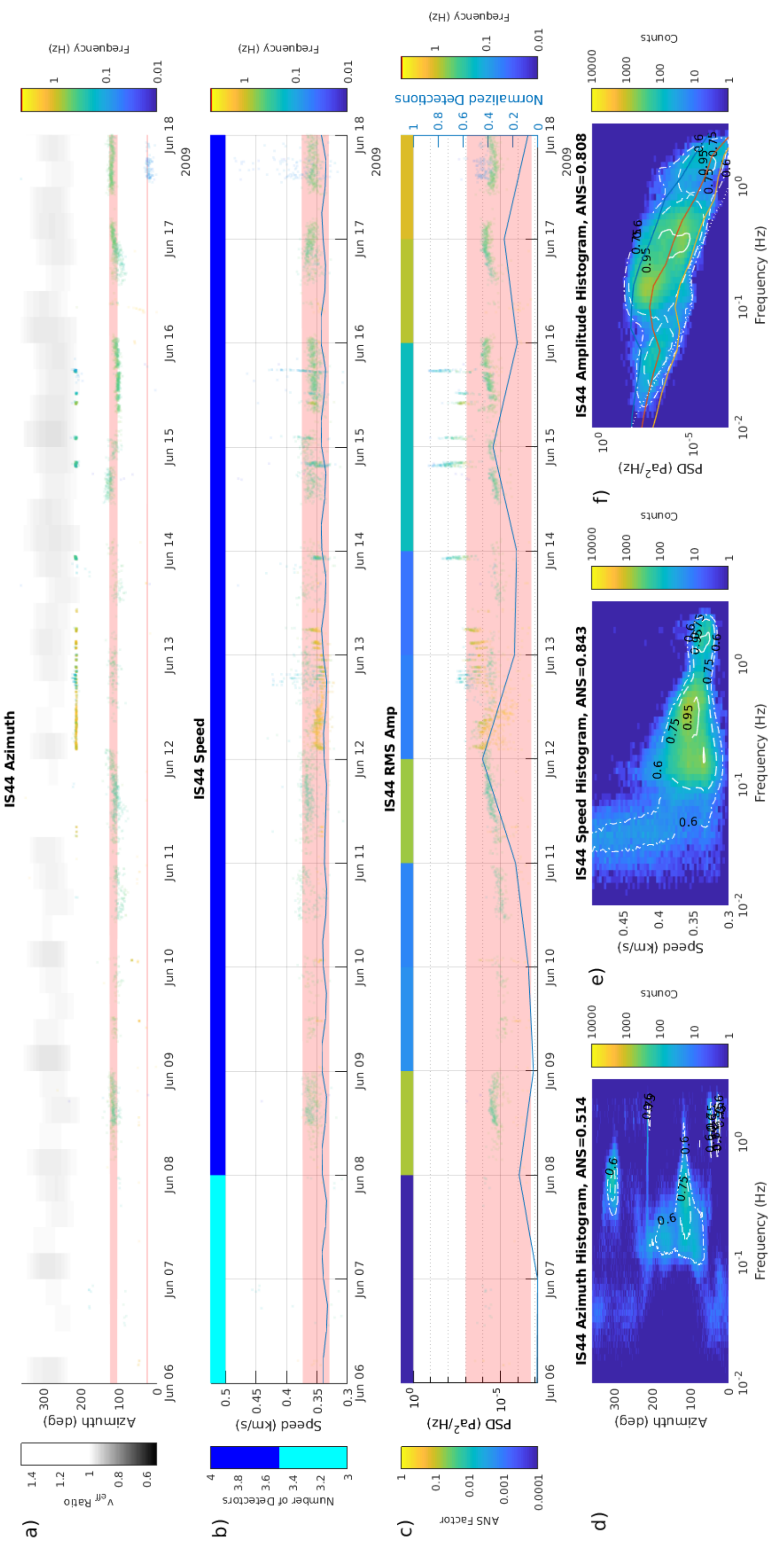

Figure 9. Same as Figure 7, but for IS44.

3.3 Reference Database: User manual

The monthly station reference data are available in netcdf files (Kristoffersen et al., 2022). These files are saved in directories for each station, as ISXX_Hist.nc, where XX denotes the station (e.g. 01), see the following diagram for an description of the data directories.

```
Data Directory
 ├──────────── IS01
 │              └──── IS01_Hist.nc
 ├──────────── ISXX
 │              └──── ISXX_Hist.nc
 ├──────────── ⋮
 └──────────── IS60
                └──── IS60_Hist.nc
```

In each of the ISXX_Hist.nc files, there are 7 variables, 2 each for the back azimuth, the horizontal speed, and the RMS amplitude, and one universal frequency vector for the other six variables. A description of these variables is provided in Table 1. The back azimuth is in degrees, the horizontal speed is in km/s and the RMS amplitude is converted to a power spectral density (in Pa^2/Hz) that is accomplished by squaring the RMS amplitude and dividing by the corresponding frequency band of the PMCC family. The variables N_{az}, N_v, and N_{amp} are the arrays for the histograms for back azimuth, horizontal speed, and RMS amplutide, respectively. These variables are 3-dimensional, with the first two dimensions corresponding to the x and y variables and the third dimension is the month, such that 1 is January, 2 is February and so on.

The histogram counts are the log base-10 of the normalized histogram counts, which were normalized by maximum of the log base-10 histogram counts, such that

$$N_{az} = \frac{\log_{10}(N_{az-total})}{\max(\log_{10}(N_{az-total}))} \tag{7}$$

where $N_{az-total}$ is the histogram of the back azimuth (with the same procedure used for the horizontal speed and the RMS amplitude). After normalization, the reference levels were determined by finding the 0.95, 0.75, and 0.6 thresholds. These thresholds are the contours provided in Figure 2.

Large peaks in the histogram data found outside of these reference curves suggest that there are potential anomalies in the observations, justifying further investigation. This indicates detections that do not conform with typical seasonal trends for that sta-

Back Azimuth				
Histogram	Dimension	Size	Parameter	Variable
N_{az}	1	49	Frequency (Hz)	f
	2	360	Azimuth (°)	az
	3	12	Month	N/A
Horizontal Speed				
Histogram	Dimension	Size	Parameter	Variable
N_v	1	49	Frequency (Hz)	f
	2	59	Speed (km/s)	v
	3	12	Month	N/A
RMS Amplitude				
Histogram	Dimension	Size	Parameter	Variable
N_{amp}	1	49	Frequency (Hz)	f
	2	59	RMS Amplitude (Pa^2/Hz)	amp
	3	12	Month	N/A

Table 1. Description of the reference histogram data contained in the .nc data files. The N_x variables contain the normalized histogram counts, for the corresponding x and y variables. The third dimension is the month, with 1=January, 2=February etc.

tion. It should be noted that these curves do not provide the cause of atypical observations, which could be due to either instrumental issues, or real (atypical) geophysical phenomena. Therefore, further investigation would be necessary to identify said causes.

4 Discussion and Conclusion

We provide an update, and improvements to the coherent ambient infrasound observations presented by Matoza et al. (2013). These include a more recent data-set, with data up to the end of 2020, for all of the available stations, 53 versus 39 for Matoza; and an updated PMCC algorithm configured with 26 frequency bands, which represents an improvement over the 15 used in Matoza.

Reference curves for the amplitude, back azimuth, and horizontal (apparent) velocity are provided. These can be used to identify abnormal trends in the data-sets for any of the currently operable stations. These trends could be due to either instrumental/environmental conditions which result in poor data quality, or are due to geophysically interesting phenomena worth additional study. Although the reference curves cannot determine what the cause of deviations from the average observational conditions are, it is useful to identify situations which may go otherwise unnoticed.

In addition to these reference curves, the data ANS factor provides a quick reference to determine the daily data deviation from nominal behaviour. This ANS factor is determined from the monthly reference curves, and therefore should account for seasonal variability (at least on a monthly scale). There are many infrasound data quality/event identification tools such as the Modular Utility for STatistical kNowledge Gathering (MUSTANG) (Casey et al., 2018) and NETwork processing - Vertically Integrated Seismic Analysis (NET-VISA) (Mialle et al., 2019; Bras et al., 2020). However, the ANS factors, which quantify the deviation from the reference curves, provide an additional means to identify anomalous events or times of poor data quality. As was mentioned regarding the reference curves, a low ANS value does not necessarily correspond to poor data, but rather deviations from the typical behaviour, which also includes repetitive clutter. Consequently, these lower ANS values could also indicate a strong transient event of geophysical interest, and should be considered in conjunction with the number of PMCC detections. Overall, these quality metrics are a useful supplement to the open-access infrasound data products provided by Hupe et al. (in review). In addition, the ANS factor could be an additional metric for event identification by the aforementioned data quality tools (e.g. MUSTANG, NET-VISA). NET-VISA is currently considered to be fully integrated in the processing environment of the International Data Centre (IDC) of the CTBTO (Bras et al., 2020). The ANS factors could additionally support the discrimination between infrasound clutter and events of interest in the IDC workflow, with the potential to reduce the number of false event hypothesis resulting from clutter at different stations (Mialle et al., 2019).

The normalized detections above a wind speed threshold demonstrate a relative increase in the number of observed detections with increasing frequency, as expected due to the ‘red’ colour of the wind noise and increased efficacy of the WNRS at higher frequencies. There is a slight decrease (followed by a continued increase) of the relative number of detections above about 0.02 Hz. This effect is likely due to the high amplitude associated with the microbarom peak at this frequency. This higher amplitude would result in a relative increase in the SNR, which could explain the local maximum of detections around the microbarom peak.

The new presented coherent histograms have some similarities to the Matoza curves, as would be expected, such as the microbarom peak, and the decrease in detections with decreasing frequency towards lower frequencies. As discussed in Matoza et al. (2013), this is likely due to the decreased efficacy of the WNRS, and therefore reduced SNR at low frequencies. The low SNR would result in relatively fewer observations of IS events. The newer results, however, do show two additional peaks, one between about 0.01 and 0.1 Hz, and the second at around 2 Hz. The 2 Hz peak appears to be more prominent during the summer months, while the low frequency peak does not show a significant seasonal variability.

5 Future Work

Currently, PMCC algorithm assumes only one single source to be detected in each pixel. Although the detection of multiple signals at different frequencies is possible, there exists the chance that if there are two or more signals in the same frequency band, that the PMCC approach could miss these secondary signals. Further studies could involve a similar analysis using sensors able to separate coherent signals overlapping in the same time-frequency domain such as the CLEAN algorithm (den Ouden et al., 2020) or the vespagram approach (Vorobeva et al., 2021). By performing a similar analysis on these multi-source results, and generating the same types of reference curves, such data analysis approaches would allow building reference curves for these additional infrasound analysis methods. Given the capability of these approaches to resolve multiple sources for each pixel, both direct comparisons of the reference curves for these results, as well as comparisons of the relative ANS factors could be performed. The ANS factor comparisons would provide an estimate as to which approach produces more anomalous data, and the direct comparison of the reference curves would allow for the potential identification of differences in the output data of interest (azimuth, apparent speed, and RMS amplitude) for these methods.

Acknowledgments

IMS data are available from the CTBTO for scientific purposes through the virtual Data Exploitation Centre (vDEC): https://www.ctbto.org/specials/vdec/. ECMWF products, including the atmospheric model analysis, are available via www.ecmwf.int/en/forecasts/accessing-forecasts under CC-BY 4.0 License. Figure 1 is reproduced from (Hupe et al., in review) under CC BY 4.0 License. The monthly reference histograms are available as supporting information to this study (Kristoffersen et al., 2022). This project 19ENV03 Infra-AUV has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme. R.S.M. acknowledges NSF grant EAR-1847736. P.H. acknowledges the CTBTO Preparatory Commission for providing access (via vDEC) to the IMS infrasound network data. All authors thank the CTBTO and the IMS station operators for guaranteeing the high quality of the infrasound data.

Disclaimer: The views expressed in this manuscript are those of the authors and do not necessarily reflect the view of the Preparatory Commission of the CTBTO.

References

Bowman, J. R., Baker, G. E., & Bahavar, M. (2005). Ambient infrasound noise. *Geophysical Research Letters*, *32*(9). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL022486` doi: https://doi.org/10.1029/2005GL022486

Bras, R. L., Arora, N. S., Kushida, N., Mialle, P., Bondár, I., Tomuta, E., ... Taylor, T. (2020). Net-visa from cradle to adulthood. a machine-learning tool for seismo-acoustic automatic association. *Pure and Applied Geophysics*, 1-22.

Brown, D., Ceranna, L., Prior, M., Mialle, P., & Le Bras, R. J. (2012, 09). The idc seismic, hydroacoustic and infrasound global low and high noise models. *Pure and Applied Geophysics*, *171*. doi: 10.1007/s00024-012-0573-6

Campus, P., & Christie, D. R. (2009). Worldwide observations of infrasonic waves. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies* (pp. 185–234). Dordrecht: Springer Netherlands. Retrieved from `https://doi.org/10.1007/978-1-4020-9508-5_6` doi: 10.1007/978-1-4020-9508-5_6

Cansi, Y. (1995). An automatic seismic event processing for detection and location: The P.M.C.C. method. *Geophysical Research Letters*, *22*(9), 1021-1024. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95GL00468 doi: https://doi.org/10.1029/95GL00468

Cansi, Y., & Klinger, Y. (1997, 01). An automated data processing method for mini-arrays. *News Lett*, *11*.

Cansi, Y., & Le Pichon, A. (2008). Infrasound event detection using the progressive multi-channel correlation algorithm. In D. Havelock, S. Kuwano, & M. Vorländer (Eds.), *Handbook of signal processing in acoustics* (pp. 1425–1435). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-0-387-30441-0_77 doi: 10.1007/978-0-387-30441-0_77

Casey, R., Templeton, M. E., Sharer, G., Keyson, L., Weertman, B. R., & Ahern, T. (2018, 02). Assuring the Quality of IRIS Data with MUSTANG. *Seismological Research Letters*, *89*(2A), 630-639. Retrieved from https://doi.org/10.1785/0220170191 doi: 10.1785/0220170191

Ceranna, L., Matoza, R., Hupe, P., Le Pichon, A., & Landès, M. (2019). Systematic array processing of a decade of global ims infrasound data. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 471–482). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-75140-5_13 doi: 10.1007/978-3-319-75140-5_13

de Groot-Hedlin, C., & Hedlin, M. (2019). Detection of infrasound signals and sources using a dense seismic network. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 669–700). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-75140-5_21 doi: 10.1007/978-3-319-75140-5_21

den Ouden, O. F. C., Assink, J. D., Smets, P. S. M., Shani-Kadmiel, S., Averbuch, G., & Evers, L. G. (2020, 01). CLEAN beamforming for the enhanced detection of multiple infrasonic sources. *Geophysical Journal International*, *221*(1), 305-317. Retrieved from https://doi.org/10.1093/gji/ggaa010 doi: 10.1093/gji/ggaa010

Evers, L. G., & Haak, H. W. (2001). Listening to sounds from an exploding meteor and oceanic waves. *Geophysical Research Letters*, *28*(1), 41-44. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000GL011859 doi: https://doi.org/10.1029/2000GL011859

Gabrielson, T. B. (2006). Refraction of sound in the atmosphere. *Acoustics Today*, *2*, 7-17.

Garcés, M. A., Hansen, R. A., & Lindquist, K. G. (1998, 10). Traveltimes for infrasonic waves propagating in a stratified atmosphere. *Geophysical Journal International*, *135*(1), 255-263. Retrieved from https://doi.org/10.1046/j.1365-246X.1998.00618.x doi: 10.1046/j.1365-246X.1998.00618.x

Hupe, P., Ceranna, L., Le Pichon, A., Matoza, R. S., & Mialle, P. (in review). The coherent infrasound wavefield: new IMS broadband bulletin products for atmospheric studies and civilian applications. *Earth Syst. Sci. Data Discuss.* doi: 10.5194/essd-2021-441

Kristoffersen, S., Le Pichon, A., Hupe, P., & Matoza, R. (2022). *Global reference histograms of the IMS infrasound broadband detection lists [dataset].* BGR product data center. doi: 10.25928/bgrseis_hist-ifsd

Kulichkov, S. (2009). On the prospects for acoustic sounding of the fine structure of the middle atmosphere. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies* (pp. 511–540). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-1-4020-9508-5_16 doi: 10.1007/978-1-4020-9508-5_16

Le Pichon, A., Ceranna, L., & Vergoz, J. (2012). Incorporating numerical modeling into estimates of the detection capability of the ims infrasound network. *Journal of Geophysical Research: Atmospheres*, *117*(D5). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD016670 doi: https://doi.org/10.1029/2011JD016670

Manders, E., Stap, J., Brakenhoff, G., van Driel, R., & Aten, J. (1992, 11). Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *Journal of Cell Science*, *103*(3), 857-862. Retrieved from https://doi.org/10.1242/jcs.103.3.857 doi: 10.1242/jcs.103.3.857

Marty, J. (2019). The IMS infrasound network: Current status and technological developments. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound* *monitoring for atmospheric studies.* (chap. 1). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-75140-5_1

Marty, J., Doury, B., & Kramer, A. (2021). Low and high broadband spectral models of atmospheric pressure fluctuation. *Journal of Atmospheric and Oceanic Technology*, *38*(10), 1813 - 1822. Retrieved from `https://journals.ametsoc.org/view/journals/atot/38/10/JTECH-D-21-0006.1.xml` doi: 10.1175/JTECH-D-21-0006.1

Matoza, R., Fee, D., Green, D., & Mialle, P. (2019). Volcano infrasound and the international monitoring system. In A. Le Pichon, E. Blanc, & A. Hauchercorne (Eds.), *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 1023–1077). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-75140-5_33` doi: 10.1007/978-3-319-75140-5_33

Matoza, R., Landès, M., Le Pichon, A., Ceranna, L., & Brown, D. (2013, 01). Coherent ambient infrasound recorded by the international monitoring system. *Geophysical Research Letters*, *40*. doi: 10.1029/2012GL054329

Matoza, R., Le Pichon, A., Vergoz, J., Herry, P., Lalande, J.-M., Lee, H.-i., ... Rybin, A. (2011). Infrasonic observations of the june 2009 sarychev peak eruption, kuril islands: Implications for infrasonic monitoring of remote explosive volcanism. *Journal of Volcanology and Geothermal Research*, *200*(1), 35-48. Retrieved from `https://www.sciencedirect.com/science/article/pii/S037702731000377X` doi: https://doi.org/10.1016/j.jvolgeores.2010.11.022

McNamara, D. E., & Buland, R. P. (2004, 08). Ambient Noise Levels in the Continental United States. *Bulletin of the Seismological Society of America*, *94*(4), 1517-1527. Retrieved from `https://doi.org/10.1785/012003001` doi: 10.1785/012003001

Mialle, P., Brown, D., & Arora, N. (2019). Advances in operational processing at the international data centre. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 209–248). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-75140-5_6` doi: 10.1007/978-3-319-75140-5_6

Mohapatra, S., & Weisshaar, J. (2018, 11). Modified pearson correlation coefficient for two-color imaging in spherocylindrical cells. *BMC Bioinformatics*, *19*. doi: 10.1186/s12859-018-2444-3

Norris, D., Gibson, R., & Bongiovanni, K. (2009). Numerical methods to model infrasonic propagation through realistic specifications of the atmosphere. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies* (pp. 541–573). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-1-4020-9508-5_17 doi: 10.1007/978-1-4020-9508-5_17

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

Waxler, R., & Assink, J. (2019). Propagation modeling through realistic atmosphere and benchmarking. In A. Le Pichon, E. Blanc, & A. Hauchecorne (Eds.), *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits* (pp. 509–549). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-75140-5_15 doi: 10.1007/978-3-319-75140-5_15