A DESCRIPTION OF THE INFRASOUND DATA PROCESSING AT THE FRENCH CEA/DASE

Alexis Le Pichon, Julien Vergoz, Nicolas Brachet, Pierre Gaillard, Yves Cansi, and Jocelyn Guilbert

Sponsored by Commissariat à l’Energie Atomique

ABSTRACT

Although the infrasound network of the International Monitoring Network (IMS) specified in the Comprehensive Nuclear-Test-Ban Treaty (CTBT) is currently not fully established, infrasound stations have provided continuous dataflow from the early 2000s. From this starting date, the *Département Analyse Surveillance de l’Environnement* (DASE) of *Commissariat à l’Energie Atomique* (CEA) has been involved in research and development of various infrasound algorithms leading to a complete processing flowchart which starts at the station level (signal detection and categorization) and ends at the network level (event building and yield estimate). Applied to the continuously growing IMS network, this processing has demonstrated its capability for detecting and locating a wide variety of infrasonic sources on a global scale.

We present an updated description of all parts of this data processing, including the most recent evolutions of the different algorithms. More precisely, it includes the following:

- An enhancement of the station processing tool (PMCC algorithm), and
- A precise description of the algorithm dedicated to association and location of detections.

OBJECTIVES

The infrasound network sending data to the International Data Centre (IDC) has been organized under the assumption that each individual station is a mini-array. This implies that the data processing has to be organized using a two-step scheme, including first station processing dedicated to detection, and second network processing leading to event characterisation (namely, location and preliminary yield estimation). We present an overview of the algorithms used at the French CEA/DASE pursuing the same objectives.

RESEARCH ACCOMPLISHED

Station Processing

Detection Algorithm

Progressive Multi-Channel Correlation (PMCC) (Cansi, 1995) is an antenna technique that is commonly used by the scientific community for detecting coherent signals recorded on infrasound arrays. The PMCC detector was originally developed by the CEA/DASE. In 2000, the software was installed in the operational environment of the French National Data Centre (NDC) to process in real-time seismic and infrasound data. Four years later, the same algorithm was implemented in the operational system of the International Data Centre (IDC) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). Since then, collaboration exchanges have continued between the IDC and DASE, and several changes have been made to enhance the PMCC source code and tune the configuration parameters. The detector has exhibited good performance in terms of detection sensitivity and robustness in an operational environment.

Recent studies performed at the CEA/DASE have shown that the IDC version (DFX/Geotool-PMCC) and the CEA/DASE version (WinPMCC) of PMCC software benefit from the implementation of an adaptive processing window duration and log-spaced frequency bands. This tested configuration enables better detection and characterization of all received signals in their wave-front parameter space (e.g., frequency-azimuth space, frequency-trace-velocity space). The following sections describe the major enhancements that have been implemented in the latest version of WinPMCC software.

Frequency-Dependent PMCC Parameters

The duration of the sliding time-window, the frequency band spacing and filter parameters (filter order and ripple) are key parameters in the PMCC processing. For standard processing of IMS-type infrasound arrays, the window length required ranges from tens to hundreds of seconds. In previous releases of WinPMCC, the window length remained constant during a processing run. Hence, previous PMCC processing at the CEA/DASE was performed in multiple separate, independent runs with different target signal frequencies. Typical configurations were 30s windows for high frequency runs (0.1-4Hz) and 60s windows for low frequency runs (0.02-0.5Hz).

The PMCC algorithm has been modified to process low and high frequency signals simultaneously. The new version is highly configurable and allows the duration of the processing window and detection parameters to be dependent on the frequency band. For example, the new single processing run may be performed with log-spaced frequency bands between 0.01 and 5 Hz and window lengths may vary linearly with the period (cf. Figure 1).

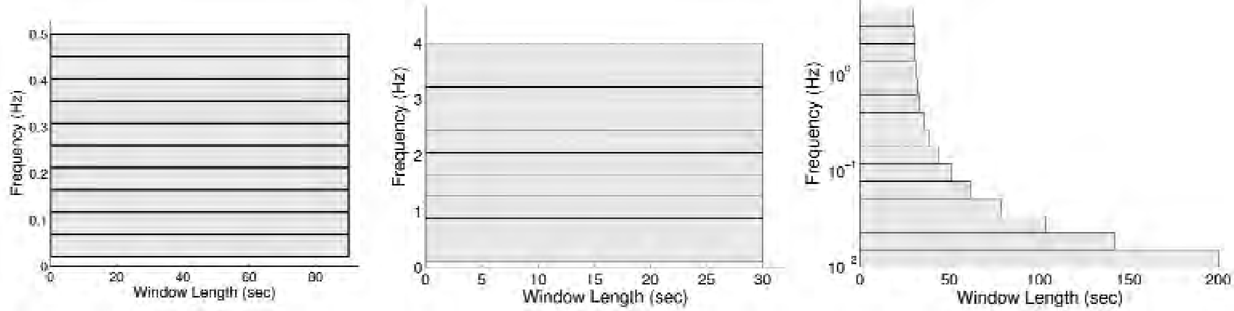

Figure 1. Examples of two 10-band standard configurations for low- and high-frequency processing (0.02-0.5 Hz and 0.1-4 Hz, left and middle respectively), replaced by a single configuration consisting of 15 bands with log-spaced filter parameters (0.01-5 Hz) and a variable window length.

In the new PMCC version, configuration parameters controlling the detection and detection clustering (family) processes can also be adjusted as a function of frequency.

PMCC Measurement Errors

Work is underway at CEA/DASE to determine more rigorously the azimuth and speed uncertainties. The current algorithm estimates the uncertainties based on statistical analysis of the distribution of PMCC detection pixels in the azimuth-speed space. The new code that is being considered performs the calculation of infrasound measurement errors as a function of physical parameters, i.e., dependant on the array geometry and the wave properties (cf. Figure 2). The new code associates a "geometrical" error to each individual PMCC pixel (cf. Szuberla and Olson, 2004). These errors are later used to compute the mean and standard deviation values of the PMCC families. This is done using normal laws, which naturally gives low weight to pixels with larger errors (i.e., detected by a limited number of sensors or anisotropic geometry) and favoring the results obtained on the best sub-networks.

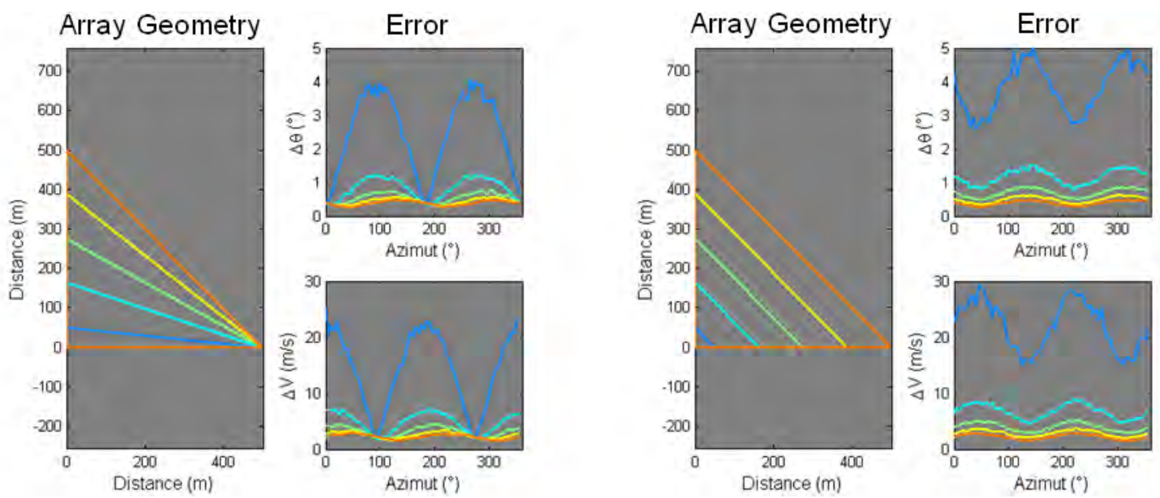

Figure 2. Errors related to the geometry (flatness on the left, aperture on the right) of an array composed of 3 sensors with a sampling rate of 20 Hz.

Re-processing of Historical IMS Infrasound Data

In order to measure the impact of the implemented changes in WinPMCC algorithm, the CEA/DASE has re-computed detection bulletins for a given set of stations over several years. Statistical analyses were made to compare the results produced by the old system (using 2 separate PMCC runs at Low and high frequencies) and the new system (single PMCC run with adaptive window length).

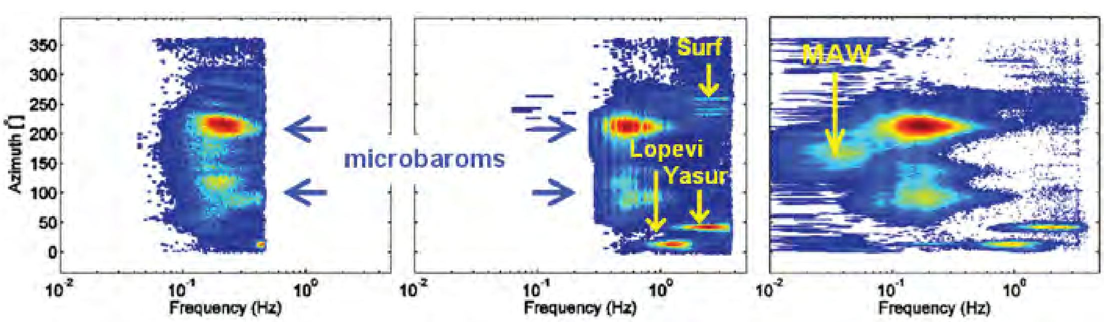

Figure 3. Comparison between processing results using two 10-band low- (0.02-0.5 Hz) and high-frequency (0.1-4 Hz) configurations (left and middle, respectively) and a 15-band log-spaced configuration (0.01-5 Hz, right). Warm colors correspond to increasing number of detections in frequency-azimuth space.

Figure 3 shows a comparison between the automatic detections produced by the CEA/DASE before and after using the optimized filtering and detection configuration parameters. Statistics were obtained over 7 years of detection data at IS22, New Caledonia. Compared with the previous processing, a better signal discrimination is expected thanks to a more accurate estimation of signal frequencies. Furthermore, signals of interest at lower frequencies (20-50 s) appear now in the detection bulletin, e.g., Mountain Associated Waves (MAW) from New Zealand in the back azimuth 150 degrees.

Figure 4 presents the results of a multi-year re-processing of data at IS22 using the new log-spaced configuration. They reveal systematic seasonal signal variations correlated with mid-latitude stratospheric wind reversals.

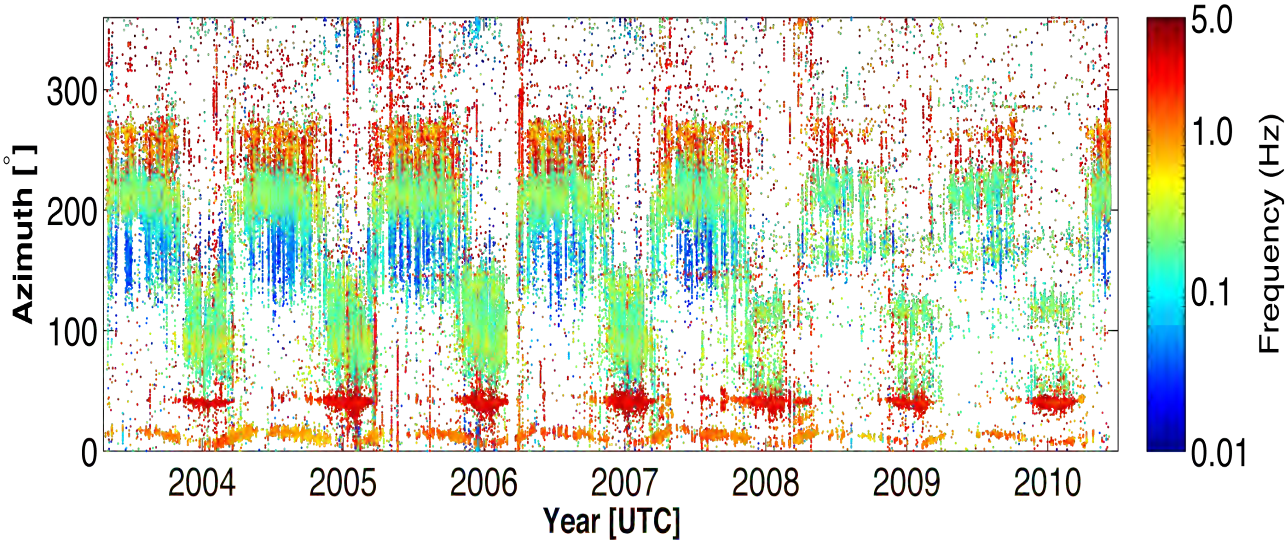

Figure 4. PMCC detections at IS22 from 2003 to 2010 using the 15-band log-scaled filter parameters (0.01-5 Hz) and variable window length illustrated in Figure 1. From 2008 to 2011, the reduced detection capability is explained by a missing central array element.

Categorization Algorithm

Before the association step, it is important to efficiently remove detections that are not of interest for fusion (categorized as "noise"). Detections falling in the noise category are typically those associated to long duration phenomena—e.g., microbaroms, industrial noise, or surf noise—or high-frequency transient signals related to small local infrasound sources—e.g., thunderstorms, small quarry blasts.

Categorization is a fundamental step in the infrasound chain of processing as it has a direct impact on the event building. Any detection incorrectly tagged as noise will subsequently be discarded by the association process, and may lead to missing an event. On the other hand, if too many noise detections are retained, the number of combinations in the association algorithm dramatically increases and the probability of false associations is very high. CEA/DASE has been using a categorization procedure that is similar to the one that is implemented at the IDC (Brachet et al., 2010). However, the method sometimes exhibits some limitations, most of the time in categorizing microbaroms. On-going studies are carried out at the CEA/DASE in order to enhance the algorithm.

The IDC categorization method applies a "cleaning" strategy at two levels:

(1) Single detection review: local sources are removed (detections with a dominant frequency greater than 2 Hz; signal duration below 50s; horizontal trace velocities outside the typical range for long range propagation of infrasound, 0.3 to 0.45 km/s)

(2) Metafamily review: Metafamilies are clusters of detections that have similar characteristics considering the following parameters: azimuth, trace velocity, frequency and time. Threshold values may be configured according to the sensitivity of each array to their respective environment. Long duration clusters, typically more than 1800s, are likely related to recurrent sources of signals and therefore categorized as noise.

(3) Detecting outliers inside noise clusters: detections with atypical F-stat value (outside a certain range determined by the average F-stat value in the cluster +/- standard deviation).

As a result about 80% of the detections are categorized as noise and therefore discarded in the subsequent network processing stage.

A new algorithm has been recently developed at the CEA/DASE in order to improve the metafamily review (second step described above, consisting of tracking the detection background to better remove it). The first tests are encouraging. This method filters out 85 to 95% of the detections (mainly microbaroms and local industrial activity) from the bulletins, and the remaining detections are kept as inputs for network processing. Whereas the IDC procedure uses fixed threshold values adapted to each station, this new statistical approach is able to adapt automatically to the data. This allows one to significantly reduce the number of configuration parameters. The main assumption relies on the normal distribution and the independence of the azimuth and frequency detection parameters associated to the noise sources. The algorithm consists of three steps:

(1) Clustering the detection in the Azimuth/Frequency space. A recursive algorithm is used with a simple 2D tree principle. At initialisation, a root cell is associated with a bounding rectangle that contains all the detections. At splitting step, detections of the current cell are partitioned to one side or the other of a splitting line that is orthogonal to the coordinate axes. Estimation of the density over each dimension is performed with the Parzen window method. The splitting dimension and the splitting value are chosen according to the lowest density gap of multimodal densities. The two resulting cells are the children of the original cell and the splitting process is repeated recursively until a stopping condition is met. The stopping conditions are fulfilled if both densities are unimodal (the data are considered as normally distributed so that no more split is needed) or if the number of detections in the considered cluster becomes less than a threshold (typically 10).

(2) Searching for long duration clusters: clusters of long duration (typically 1800s) are labelled as “noise.”

(3) Detecting outliers inside noise clusters: For each noise cluster, outliers are removed with respect to some waveform parameters (amplitudes and F-Stat parameters are proposed). A detection is defined as an outlier in the current cluster and a given parameter if it lies outside the range $\mu \pm k\sigma$ (where μ is the average and σ the standard deviation, k is typically chosen equal to 3 which represents a confidence of ~99% assuming a normal distribution). This step is very important because it allows categorizing as “non-noise” a transient signal which would arrive from the same direction and a comparable dominant frequency as a “noise” cluster, but with a different amplitude or signal to noise ratio (related to F-stat).

Network Processing

Methodology for Building Events

The fusion procedure is based on a uniform atmosphere assuming a constant celerity equal to 0.3 km/s, typical of infrasonic waves propagating in the ground to stratosphere waveguide (Brown et al., 2002). It has been widely proved that this assumption is fulfilled in most cases (Le Pichon et al., 2010), whatever the period of the year. The event construction follows a systematic exploration of all possible associations obtained by cross bearing between categorized detection bulletins, satisfying a geophysical criterion based on the used velocity model.

The quality of each association is quantified by two scores: the “geophysical” and “signal” scores. The geophysical score is a measure of the gap between the calculated celerities and the initial velocity model. The signal score quantifies the consistency of the signal features relatively to the pre-location obtained by cross bearings, in terms of detected mean frequency (Brachet et al., 2010). On-going studies try to implement other signal features such as peak to peak amplitudes and detection durations but still need to be validated.

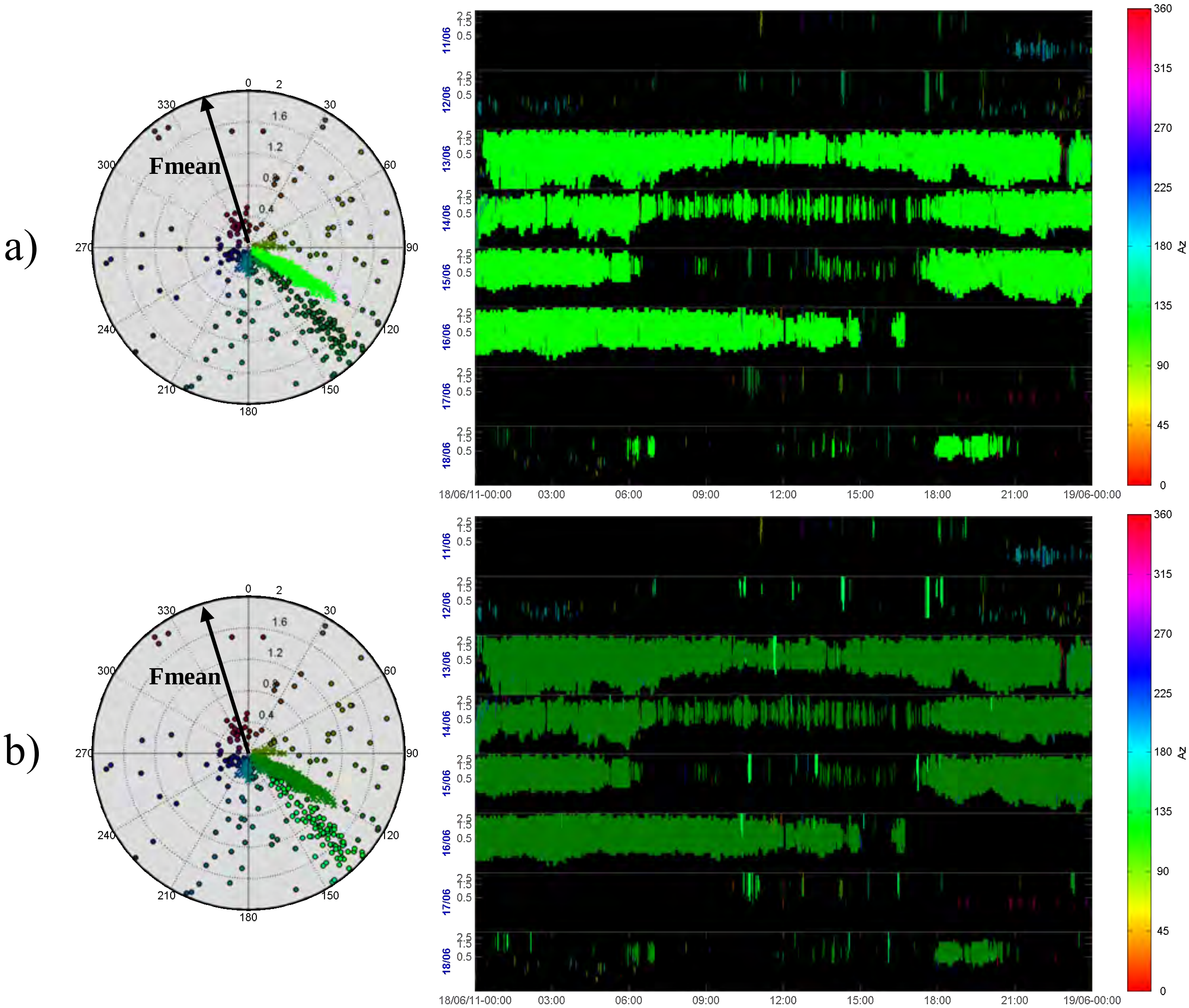

Figure 5. Example of categorization on 8 days of PMCC detections at IS48 (Tunisia), from 11 to 18 June 2011. In each panel a) and b) on the left part, detections are shown in Azimuth/Frequency space as radar-plots (radius is log-scaled mean frequency) and in Time/Frequency space (Y-axis is log-scaled mean frequency) on the right part. Colours represent detected back-azimuth according to the colour bar on the right. On radar plots, circles are detections categorized as "non noise" while crosses are those categorized as "noise". a) During this time period, the Dubbi volcano (Ethiopia) erupts, which generates continuous detections during 4 days. As expected, all these detections are clustered (bright green detections in panel a)) and categorized as noise due to its very long duration. b) Another source (associated to mining activity) also generates more sporadic detections coming almost from the same direction as Dubbi volcano with a higher but overlapping frequency band (bright green detections in panel b)). The new procedure allows characterizing statistically both sources and separate them correctly (for example, on June 13, 2011, a detection of source b) is identified inside a stream of volcanic detections). Finally, all detections of source a) are removed while detections of source b) are kept for network processing.

The pre-location is calculated assuming a point-like source and using multi-station measurements where both onset time and back-azimuth are taken into account. The location is initiated by cross bearings, and iteratively modified using a non linear iterative least squares inversion scheme (*Coleman and Li*, 1996). In case of conflict (one arrival at one station can be associated to more than one arrival at the other stations), the number of stations and/or the score (in case of equality) decides between all candidates. In order to take into account errors in the origin time estimates due to location uncertainty (because of azimuth deviation added to measure uncertainties) and additional propagation types (e.g., stratospheric and thermospheric), the scores thresholds are enhanced to decrease the missed-event rate.

Interactive Review of Automatic Events

In order to review automatic infrasound events, the CEA/DASE has developed the "Netinfra" software. This prototype tool allows one to interactively check, modify, remove and validate infrasound events and the associated detections. Netinfra is an all-in-one application that integrates various visualization tools in the same window, which makes the analyst's review work easy and efficient. The graphical user interface is composed of four panels (MAP, INFORMATION, DETECTIONS and WAVEFORMS, and ASSOCIATION) which dynamically communicate with action buttons. The analyst has access to a very wide range of information, which covers the original signal waveforms and detections, the results of the phase association and event location, and an estimate of the detection capability map at the time the event is built (cf. section "Detection capability"). The display of the direction and intensity of the prevailing atmospheric winds is also very useful information for the analyst which helps to validate or reject detection associations. On-going studies will allow us to use the 3D atmospheric models coupled to 3D ray tracing tool (WASP 3D) to interactively build propagation tables of celerities and azimuth deviations, associate the corresponding arrivals, and systematically perform an enhanced location of built events.

Figure 6 presents an example of infrasound interactive review in the Netinfra GUI:

- PMCC metafamilies (clusters of station detections associated to the same source) are represented by:
    - Points in the ASSOCIATION panel (in which time is absolute),
    - Sets of rectangles in the DETECTIONS/WAVEFORMS panel (in which time is relative to the first displayed detection of interest on each station) and
    - Lines representing their back-azimuth in the MAP panel.

- Results of the categorization are accessible trough colour-coded points (e.g., red metafamilies are those categorized as noise clusters).

- Results of the fusion procedure are accessible through colour-coded lines (e.g., the colour of lines between two metafamilies represents the "geophysical" score which characterizes the association).

- The detection capability is accessible in the MAP panel in the background colour.

- Stratospheric (45-55 km altitude) and/or tropospheric (8-12 km altitude) winds calculated from ECMWF-91levels can be overlaid on the map. The intensity and direction of the winds help the analysts to decide whether station detections are associated to an event.

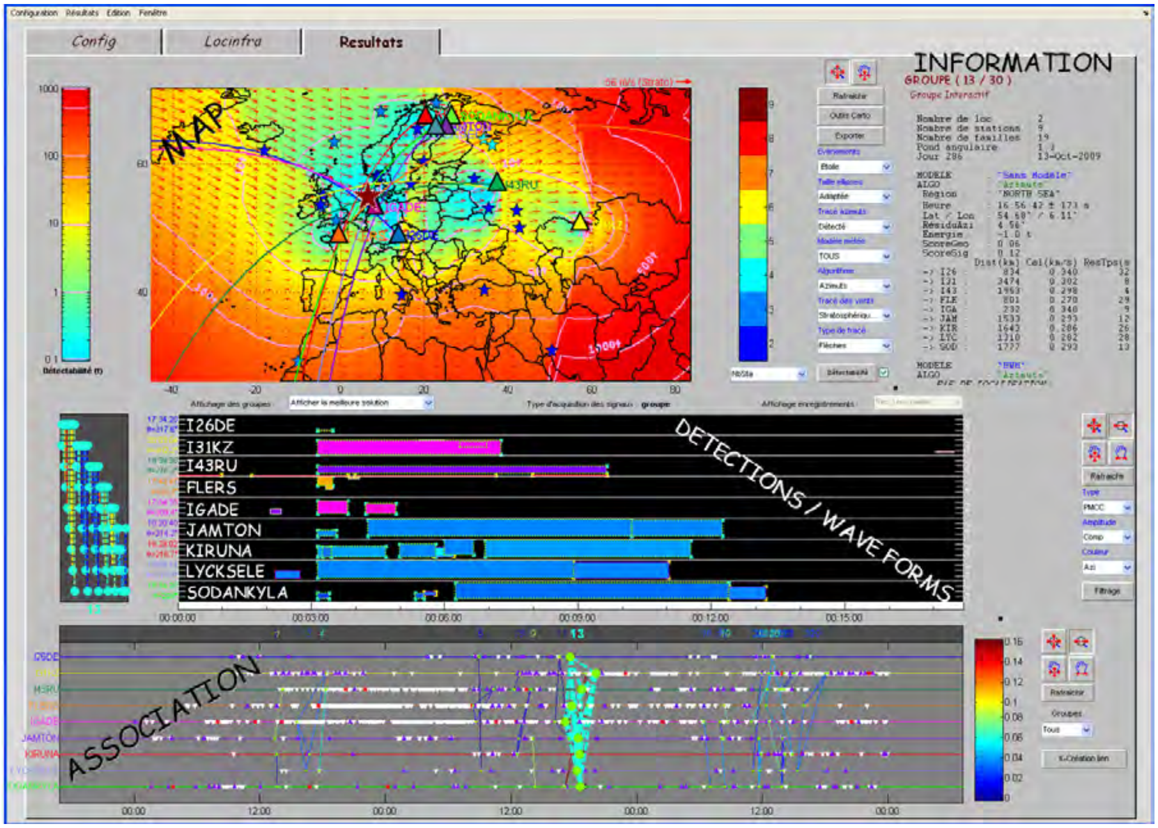

Figure 6. Example of event review with the Netinfra tool. The analyst has chosen an event above North Sea detected by 9 stations (bolide over the Netherlands, October 13, 2009, denoted as a dark red star in the MAP panel). Associated metafamilies and detections are displayed in the DETECTION/WAVEFORMS panels as highlighted dashed bordered rectangles, as back-azimuth lines in the MAP panel and as a set of line/points in the ASSOCIATION panel. As expected, stratospheric winds (denoted as red arrows on the map), favour eastward propagation, explaining that 8 of the 9 stations are located east of the event. Simple propagation effects can be observed (duration extension of the detections with distance, frequency decrease with distance) with typical celerities ranging from 280 to 340 m/s (for the 2 closest stations on which tropospheric arrivals are observed). The frequency content of the detection at FLERS station is a bit high (>2Hz) for such distance (800 km), in upwind conditions, with a celerity of 0.27 km/s. Such an association is probably spurious and the analyst may decide to remove the association of FLERS detection from the event before validating it.

Detection Capability

Recent global scale observations of infrasonic waves recorded by the IMS network confirm that its detection capability is highly variable in space and time. In order to model the detection capability of an infrasound network, it is necessary to predict the signal amplitude at any location, and further evaluate whether the signal is detectable above the noise level at the receivers. Different approaches incorporating background noise and various yield-scaling relationships have been proposed (e.g., Clauter and Blandford, 1997; Trost, 1997). Significant advances were achieved by Stevens et al., 2002 by considering attenuation relations derived from recordings of historical atmospheric nuclear and chemical explosions (e.g., Whitaker, 1995). However, conclusions from these studies may be misleading because they do not include an accurate description of the time varying stratospheric winds. Using the state-of-the-art specifications of the stratospheric wind and time-dependent station noise models, recent simulations predicted that explosions equivalent to ~500 t of TNT would be detected by at least two stations at any time of the year over the earth's surface (Le Pichon et al., 2009; Green and Bowers, 2010). Such simulations are useful to check and improve the geophysical consistency of events resulting from the association procedure, by identifying the stations most likely to detect the event and aiding the identification of mis-associated arrivals. Furthermore, they may help to optimize infrasound networks with respect to both number and configuration of stations to monitor infrasonic sources of interest.

Global Scale Studies

The methodology used to estimate the network detection capability is developed in Le Pichon et al., 2009. Considering the full IMS infrasound network, the predicted seasonal variations of the detection thresholds roughly follow a double peak pattern (cf. Figure 7). The best performances are predicted around January and July when the zonal winds are the strongest. During the equinox periods, around April and October, higher thresholds are explained by the zonal wind reversals. The largest seasonal variations of the thresholds are observed at high-latitude regions where the stratospheric winds dominate, as opposed to the equatorial region where small variations of the thresholds are predicted.

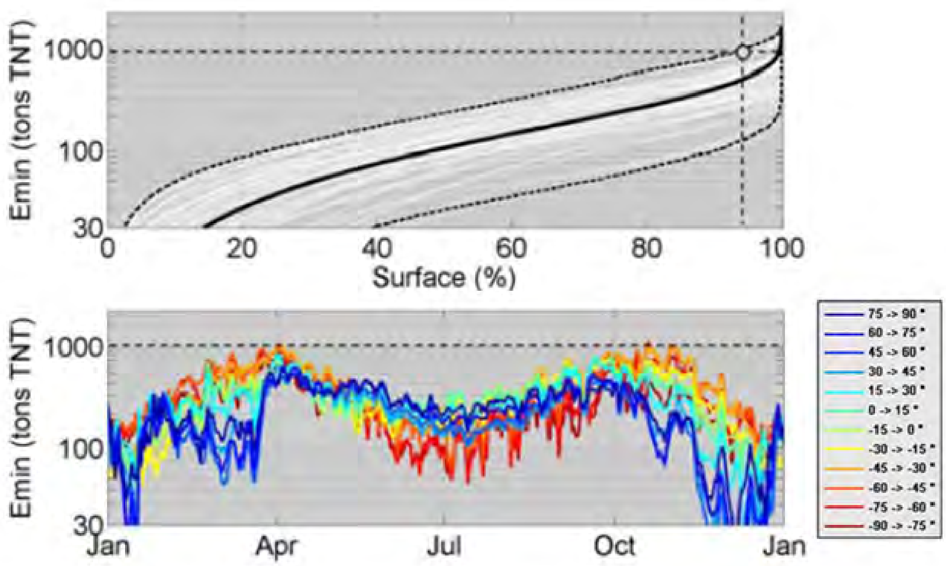

Figure 7. Summary of the IMS network two-station coverage detection capability in the 0.2-2 Hz frequency band. Left: Earth's coverage versus minimum detectable energy. Each white curve refers to one day of the year. The black solid and two dotted curves indicate the median, and the 95% confidence interval of the temporal distribution of the surface coverage. The dashed black lines indicate the detectable energy for a 95% Earth's coverage. Right: temporal fluctuations of the minimum detectable energy in 12 latitude dependent regions with a 95% Earth's coverage. The dashed black lines indicate the 1 kt limit. Simulations are carried out using the ECMWF model with site dependent wind noise models considering the full IMS network.

Regional Scale: A Useful Tool for the Analyst

The methodology has been implemented in operations and the analyst can access detection capability maps with the Netinfra tool (cf. Network Processing section of this paper and Figure 6). Automatic Network processing is daily performed for European stations (mixing both IMS and national arrays) and it is very helpful for the analyst to efficiently target false events, typically those located in geographical areas where detection capability threshold is high.

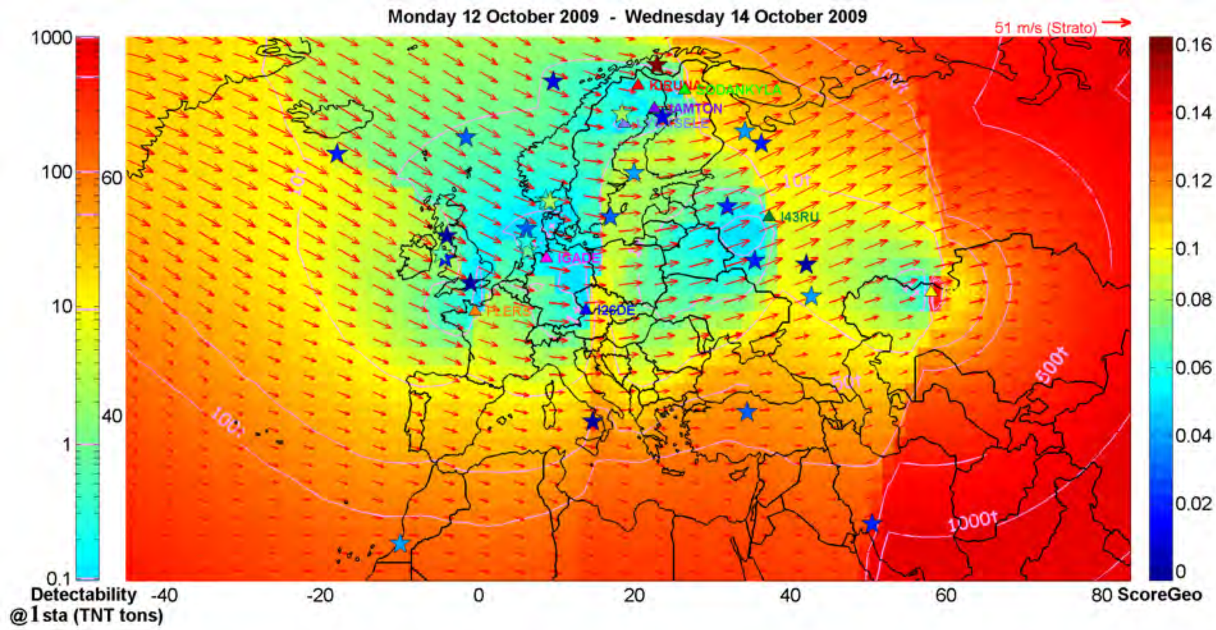

Figure 8. Example of map automatically generated by the operational Network Processing tool, in Europe area (corresponding to the Figure 6 Netinfra configuration). Each star is an automatic event which color represents its "geophysical" score (cf. network processing chapter). Background color is the detection capability at one station according to the color bar on the left, ranging from .1t to 1kt. Events in areas where detection capability threshold is high need to be reviewed in priority by the analyst because they are likely false events.

CONCLUSIONS AND RECOMMENDATIONS

The CEA/DASE has been following an operational approach for processing infrasound data which is very similar to the one used at the IDC. The latest evolutions on PMCC software – including adaptive window lengths and log-spaced filter parameters - produce satisfactory results which enhance the performance of the detector. Work is also underway at CEA/DASE to update the categorization algorithm. The introduction of the statistical analysis seems to improve the grouping of metafamilies and classification of noise vs. signal detections. The recent development of Netinfra has demonstrated that the full chain of processing for infrasound data has become a reality. This prototype tool is being used at the network processing level to build infrasound automatic events. It is also a powerful graphical interface which gathers all kinds of information that efficiently guide the analyst during the interactive review of infrasound events.

REFERENCES

Brachet, N., D. Brown, R. Le Bras, P. Mialle, and J. Coyne (2010). Infrasound Monitoring for Atmospheric Studies, Monitoring the Earth's Atmosphere with the Global IMS Infrasound Network, 77–118, Springer, ISBN:978-1-4020-9507-8, Dordrecht.

Brown D. J., C. N. Katz, R. Le Bras, M. P. Flanagan, J. Wang, and A. K. Gault (2002). Infrasonic signal detection and source location at the Prototype Data Centre, *Pure and App. Geophys*. 159: 1081–1125.

Cansi, Y. (1995). An automatic seismic event processing for detection and location: the PMCC method, *Geophys. Res. Lett*. 22: 1021–1024.

Coleman, T. F., and Y. Li (1996): An interior, trust region approach for nonlinear minimization subject to bounds, *SIAM J. Control Optim*. 6: 418-–445.

Clauter, D. A. and R. R. Blandford (1997). Capability modelling of the proposed International Monitoring System 60-Station infrasonic network, in *Proceedings of the Infrasound workshop for CTBT monitoring*, LA-UR-98–56, Santa Fe, New Mexico.

Green, D. N., and D. Bowers (2010). Estimating the Detection Capability of the International Monitoring System infrasound network, *J. Geophys. Res*. 115: D18116, doi: 10.1029/2010JD014017.

Le Pichon, A., J. Vergoz, E. Blanc, J. Guilbert, L. Ceranna, L. Evers, and N. Brachet (2009). Assessing the performance of the international monitoring system infrasound network: geographical coverage and temporal variabilities, *J. Geophys. Res*. 114: doi: 10.1029/2008JD010907.

Le Pichon, A., J. Vergoz, Y. Cansi, L. Ceranna, and D. Drob (2010). Infrasound Monitoring for Atmospheric Studies, Contribution of Infrasound Monitoring for Atmospheric Remote Sensing, 629–646, Springer, ISBN: 978-1-4020-9507-8, Dordrecht.

Trost, L. (1997). High altitude wind effects on infrasound network performance, In *Proceedings of the Infrasound Workshop for CTBT Monitoring*, Santa Fe, New Mexico.

Stevens, J. L., I. I. Divnov, D. A. Adams, J. R. Murphy, and V. N. Bourchik (2002). Constraints on Infrasound Scaling and Attenuation Relations from Soviet Explosion Data, *Pure Appl. Geophys*. 159: 1045–1062.

Szuberla, C. and J. V. Olson (2004). Uncertainties associated with parameter estimation in atmospheric infrasound arrays, *J. Acoustic. Soc. Am*. 115.

Whitaker, R. W. (1995). Infrasonic monitoring, in *Proceedings of the 17th annual Seismic Research Symposium*, Scottsdale, AZ, 997–1000.