The NoEMi (Noise Frequency Event Miner) framework

T. Accadia¹, F Acernese²,a, M Agathos³,a, P Astone⁴,a, G Ballardin⁵,
F Barone²,a, M Barsuglia⁶,a, A Basit⁷,a,b, Th S Bauer³,a, M Bebronne¹,
M Bejger⁸,c, M G Bekier⁸,a, M Bitossi⁷,a, M A Bizouard⁹,a, M Blom⁴,a,
F Bondi⁴,a, H Bonelli⁶,a, B Bonnand⁷,11, V Boschi²,a, L Bosi¹²,a,
B Bouhou¹, S Braccini⁷,a, C Bradaschia⁷,a, M Branchesi¹³,a,b,
T Briant¹⁴,a, A Brillet¹⁰,a, V Brissou⁹,a, T Bulik⁸,b, H J Bulten¹³,a,b,
D Buskulic¹, C Buy⁶,a, E Calloni²,a,b, B Canuel⁶,a, F Carbognani⁵,
F Cavalier⁹,a, R Cavalieri⁵,b, G Cella⁷,a, E Cesarini¹³,b, O Chabbi¹⁰,a,
E Chassande-Mottin⁶,a, A Chincarini¹⁵,a, A Chiummo³,a, F Cleva¹⁰,a,
E Coccia¹⁶,a,b, P-F Coladon¹⁴, C N Colacino⁷,a,b, J Colas⁵,a, A Colla⁴,a,b,
M Colombini¹¹,b, A Conte⁴,a,b, J-P. Coulon¹⁰,a, E Cuoco¹, S D’Antonio¹⁰,a,
V Dattilo³,a, M Davier⁹,a, R Day⁵,a, R De Rosa²,a,b, G Debreczeni¹⁷,
W Del Pozzo³,a,b, L Di Fiore²,a, A Di Lieto⁷,a,b, M Di Paolo Emilio¹⁶,a,c,
A Di Virgilio⁷,a, A Dietz¹, M Drago¹⁸,a,b, G Endrőczi¹⁷, V Fafone¹⁶,a,b,
I Ferrante⁷,a,b, F Fidecaro⁷,a, I Fiori¹, R Flaminio¹¹,a, L A Forte²,a,
J-D Fournier¹⁰,a, J Franc¹¹, S Franco⁹,a, S Frasca⁴,a,b, F Frasconi⁷,a,
M Galimberti¹¹, L Gammaitoni¹²,a,b, F Garufi²,a,b, M E Gaspár¹⁷,
G Gemme¹⁵,a, E Genin¹, A Gennai⁷,a,a, A Glazotto⁷,a, R Gouaty¹,
M Granata⁶,c, C Greverie¹⁰,a, G M Guidi¹⁵,a,b, J-F Hayau¹⁰,a,
A Heidmann¹¹,a, H Heitmann¹⁰,a, P Héloïse⁹,a, G Hemming⁵,a,
P Jaranowski⁵,d, R J G Jonker³,a, M Kasprzack³,a,b, I Kowalska⁶,b,
A Królak¹⁵,a, N Leroy⁹,a, N Létendre¹, T G F Li³,a,a, N Liguori¹⁸,a,b,
M Lorenzini¹³,a, V Loriette⁹,b, G Losurdo¹³,a, E Majorana⁴,a,
I Maksimovic⁹,b, V Malvezzi¹⁶,a, N Man¹⁰,a, M Mantovani⁷,a,c,
F Marchesoni¹²,a, F Marion¹, J Marków⁵,a, F Martelli³,a,b, A Masserot¹,
C Michel¹¹,a, L Milano²,a,b, Y Minenko¹⁶,a, M Mohan⁵,a, N Morgado¹¹,
A Morgia¹⁶,a,b, S Mosca²,a,b, B Mours¹, L Naticchioni¹⁴,a,b, F Nocera¹,
L Palladino¹⁶,a,c, C Palomba⁴,a, F Paololetti⁷,a,b, R Paololetti⁷,a,c, M Parisi²,a,b,
A Pasqualetti³, R Passaquieti⁷,a,b, D Passuello⁷,a, G Persichetti²,a,b,
P Piergiovanni¹³,a,b, M Pietka⁸,d, L Pinard¹¹, R Poggiani⁷,a,b, M Prato¹⁵,
G A Prodi¹⁸,a,b, M Punturo¹²,a, P Puppo⁴,a, D S Rabeling⁳,a,b, I Rácz¹⁷,
P Rapagnani¹⁴,a,b, V Re¹⁶,a,b, T Regimbau¹⁰,a, F Ricci⁴,a,b, F Robinet⁹,a,
A Rocchi¹⁶,a, L Rolland¹, R Romano²,a,c, D Rosińska⁸,c,f, P Ruggeri⁵,a,
B Sassolas¹¹,a, D Sentenac⁵,a, L Sperandio¹⁶,a,b, R Sturani¹³,a,b,
B Swinkels⁵,a, M Tacca⁵,a, L Taffarelle¹⁸,c, A P M ter Braack³,a,
A Tonelli⁷,a,b, M Tonelli⁷,a,b, O Torre⁷,a,c, E Tournefier¹, F Travasso¹²,a,b,
G Vajente⁷,a,b, J F van den Brand³,a,b, C Van Den Broeck³,a,
S van der Putten⁹,a,b, M Vasuth¹⁷, M Vavoulidis⁹,a, G Vedovato¹⁸,c,
D Verkindt¹, F Vetere³,a,b, A Viceré¹³,a,b, J-Y Vinet¹⁰,a, S Vitale⁵,a,
H Vucca¹²,a, R L Ward⁵,a, M Was⁹,a, K Yamamoto¹³,b,d, M Yvert¹,
1. Introduction

Michelson interferometers (ITF) are presently the most promising instruments in the search for gravitational waves (GW) [1]. Gravitational waves traversing an ITF change the light paths in its arms, and the differential length variation is observed as a power fluctuation in the dark fringe.
(DF\textsuperscript{1}) of the ITF. The largest GW interferometers to have been built so far are the French-Italian Virgo [3] detector with 3 km-long arms, the two 4 km LIGO [4] detectors located in USA, and the German, 600 m-long GEO600 [5].

The signals expected at the detector are, however, extremely weak: current ITF data are still dominated by noise. Besides the intrinsic limitations to the detector sensitivity [1], interferometers are affected by instrumental artefacts and environmental disturbances of various origins. We usually distinguish between transient noise signals (lasting up to a few seconds) and continuous disturbances (lasting from a period of hours up to the full observation time). In this paper we will focus on the latter category, which is particularly harmful for continuous wave (CW) searches, where the analysis is based on the integration of the data over long periods of time.

GW signals are expected to be produced by rapidly rotating asymmetric neutron stars (e.g. pulsars) [6]. In fact, observations show that the rotation frequency of pulsars decreases in time (\textit{Spin-Down}), and emission of GWs could explain part of the rotational energy loss. The frequency of the emitted GW is expected to be twice the spinning frequency of the star, and is Doppler-modulated because of the relative motion with respect to the detector. The search for signals is computationally cheap for "known" pulsars (whose location in the sky, frequency and frequency derivatives are known thanks to electro-magnetic observations), because it can be done by directly searching at the expected emission frequencies (Targeted search). The most famous examples of known pulsars are the Crab [7] (expected GW emission at 59.44 Hz) and Vela [8] (22.38 Hz). In this case it is important to constantly monitor the data quality at the target frequencies, ensuring that no disturbances degrade the detector sensitivity and thus the expected signal-to-noise ratio.

For the all-sky searches [9], in which as large a portion as possible of the source parameter space is explored (with the constraint given by the available computing power), the identification of continuous disturbances is even more important. Noise lines (narrow spectral features, so-called because they appear as persistent peaks in the DF power spectrum and lines in the power spectrograms) can mimic the signals produced by the pulsars and produce fake candidates in a given region of the sky. It is therefore crucial to locate and identify the noise sources ("line hunting"), remove or reduce them as soon as they appear whenever possible ("line mitigation"), or catalogue them as instrumental features in view of the off-line analysis.

GW interferometers are equipped with environmental sensors; seismometers, microphones and magnetometers are placed close to the main detector components and provide environment monitoring data. Relevant information about the noise coupling paths is provided by correlating the data of the channels\textsuperscript{2} relating to these sensors with the GW channel data.

NoEMi (Noise Frequency Event Miner) is a Virgo software tool dedicated to noise line monitoring and identification. It was initially conceived as an online noise monitor for CW search purposes, aiming to guarantee fast feedback (with a maximum of 1 day of delay) whenever new lines appeared in the data. NoEMi analyses, on a daily basis, the GW channel and a subset of the environmental channels, identifying the noise lines in the various channels and looking for coincidences between the DF and the auxiliary channels (by coincident lines we mean lines found at the same frequency in different channels). A line tracker algorithm reconstructs the lines over time, and finally records them in a database.

NoEMi builds and publishes “daily summary pages”, with plots and lists of the found noise lines, on the Virgo monitoring web pages. In particular, for the objectives of the targeted search, noise lines found at the same frequencies of the known pulsars are highlighted in the summary pages.

\textsuperscript{1} We will refer to the DF as the uncalibrated GW detection data stream, and to $h(t)$ as the calibrated strain data [2].

\textsuperscript{2} Channel refers to a time series obtained following conversion of an analogue data stream to a digital data stream.
Particular care has been taken in the development of a web interface to the NoEMi database, in order to make it easily accessible for detector characterisation and off-line noise studies. In this paper we present the NoEMi framework providing examples of its usage as a “noise monitor” and as an instrument for line identification in the most recent Virgo runs.

2. The NoEMi pipeline
NoEMi’s main components are schematised in Fig. 1. They are the Event finder, the Event analyser, the Coincidence finder and the Line tracker.

![Figure 1. The NoEMi pipeline and the associated software architecture.](image)

2.1. The Event finder
The Event finder runs Fast Fourier Transforms (FFTs) on the data, following a data-cleaning procedure, which removes the short transients [10]. For each FFT, the program, using an autoregressive (AR) technique, identifies the spectrum peaks, also called frequency events (EVF). The main Event finder configurable parameters are the length of the FFT, which determines the frequency resolution of the spectrum, and the threshold set on the peak Critical Ratio (CR), defined as the difference between the peak amplitude and the mean value of the spectrum, divided by the spectrum standard deviation. The peak parameters (frequency, CR, width, and amplitude) are recorded in a MySQL [11] database (Events DB).

2.2. The Event analyser
The Event analyser makes a statistical analysis of the EVF parameters, collecting the events of a few tens of FFTs. In particular, two distributions are computed: persistency (frequency distribution of the events, normalised by the number of FFTs in the collection) and CR versus frequency (average CR of the events in each frequency bin). The program identifies the peaks in the two histograms, setting a threshold based on their average and standard deviation, recording their maximum value and width (defined as the full width over threshold). The peaks of the two
histograms can be considered as a “projection” of the noise lines in the period covered by the collection: stationary lines generate peaks in the spectra which fall into the same frequency bin, thus forming a peak in the persistency distribution. Instead, the CR histogram helps to identify the strong non-stationary lines, the persistency of which may be too low to emerge in the first distribution. The algorithm also looks for peaks falling in or near the Doppler-shifted frequency band of the known pulsars, highlighting them in the summary web pages in order to alert the noise experts.

2.3. The Coincidence finder

The next step is the Coincidence finder: the set of peaks extracted in the DF or $h(t)$ are compared with those of the auxiliary channels, and marked as coincident in frequency if their peaks overlap.

2.4. The Line tracker

As a last step, the Line tracker algorithm reconstructs the time evolution of the noise lines, comparing the frequency of the latest peaks with those identified in the previous iterations. If more than one line is found near the peak frequency, the program chooses the best candidate, also comparing the other parameters (mean persistency, CR, amplitude). For each iteration NoEMi stores the line parameters and the list of coincident auxiliary channels in the Lines database (Lines DB), assigning them the ID that identifies the matched line. The set of parameters stored for each iteration in the Lines DB allows for the dynamical computation of the “global” line parameters, such as the mean frequency, persistency, CR, etc., or the number of iterations in which an auxiliary channel has been found in coincidence with the GW channel (thus making it possible to discard the “randomly” coincident peaks).

Fig. 2 shows examples of histograms and plots produced in each step of the analysis of the DF channel. The maximum “update frequency” of the lines depend on the length of the FFTs used in the first step of the analysis and on the number of FFTs collected to calculate the statistical distributions: low-frequency resolution requires shorter FFTs and allows more frequent iterations (leading to better line tracking) than high-resolution. In the last Virgo run (VSR4), NoEMi ran in two different configurations at the same time:

(i) “CW search-oriented” configuration: $h(t)$ channel sampled at 4096 Hz, 1 mHz frequency resolution, daily updates, “low” threshold on CR (3). This configuration enhances NoEMi’s sensitivity at low frequency (up to 2 kHz), which is the CW analysis target. Moreover, the high frequency resolution increases the identification efficiency for the sharp and stationary lines, the most harmful for the CW all-sky searches;

(ii) “detector characterisation-oriented” configuration: DF channel sampled at 20 kHz, 10 mHz frequency resolution, updates every 2 hours, “high” threshold on CR (6). This configuration helps the identification and follow-up of the non-stationary lines up to 10 kHz, thanks to its higher time resolution.

3. NoEMi user interface and Lines DB

NoEMi is integrated in the Noise Monitors (NM) API framework [12], which provides a common interface to a web server for the publication of the daily results and for the access to the databases. The summary web pages are saved on disk and a calendar applet on the page allows for the results to be displayed for any day of interest. NoEMi produces various summary pages: for each channel the daily time-frequency diagrams of the peak maps (the set of frequency events) and their distributions, and the updated lists of noise lines.

The web interface to the NoEMi databases makes it possible to run queries using specific filtering criteria: for instance (concerning the Lines DB) a user can select lines found in a given
Figure 2. a) Time-Frequency diagram of the frequency events; the red lines show the frequencies of the known pulsars. b) Persistency and CR histograms built from the events distributions. The red dashed lines show the frequencies of the known pulsars, while the green solid lines show the identified peaks, i.e. those higher than the thresholds set on the histograms (the horizontal dashed lines). c) Time-Frequency plot of the reconstructed lines; note that the time range of this plot is 40 days, while the Event Time-Frequency diagram covers 48 hours. d) Example of a reconstructed non-stationary line in VSR4. The plots refer to the DF channel.

time or frequency interval, setting thresholds on their dynamic parameters (persistency, CR, line “duration”, etc.) or even query the lines coincident with a subset of the auxiliary channels. Each line is linked to the Metadata table [13], which is filled with user-defined information once the line is identified. Thanks to the Metadata table, the Lines DB can be considered a true “Line catalogue”, dynamically updated by NoEMi and manually maintained by the noise experts. The NM API allows the users to easily associate metadata to the lines, and to use the metadata as additional filtering criteria: for instance, to distinguish between identified and not yet identified lines, or to select lines containing specific descriptions.

4. Lines identification and monitoring with NoEMi
During the Virgo runs, but also in commissioning periods, NoEMi was used as a line monitor and (with the Lines DB) as a tool for line identification.

4.1. Hardware injections
Hardware injections considered here are simulated pulsar GW signals injected in the detector by applying a calibrated force to displace the ITF end mirrors, in order to test the detection
efficiency of the analysis pipelines. Fig. 3 shows the signal of an injected pulsar seen in the data from the Events and Lines DB. The simulation takes into account the pulsar spin-down and the Doppler modulation of the pulsar frequency, due to its relative motion with respect to the detector. This example helped to check the NoEMi framework and evidences its line extraction and tracking capabilities.

**Figure 3.** Signal of the 1398 Hz hardware simulated pulsar signal injection in the VSR2 data. Left: Time-Frequency plot of the frequency events detected by NoEMi in the $h(t)$ channel. Right: Time-Frequency plot of the reconstructed line. The mean CR of the signal is 6.

4.2. Detector resonance studies

The Lines DB, filled by NoEMi, was also used in an off-line detector characterisation analysis. ITF components (suspensions, mirrors, etc.) generate thermal noise characterised by sharp resonances. Fig. 4 presents the time-frequency plots of two mirror resonances (bulk modes). Analysing and correlating the time evolution (and consequently the temperature behaviour) of the different resonances allowed us to associate them to the mirrors and suspensions to which they belong. This study was performed in the frame of the thermal noise investigations, in which the mirror and violin mechanical losses were characterised with the aim of predicting the effective thermal noise present in the detector [16]. The results of the analysis were used to optimise the mirror suspension design in view of the construction of the next generation Virgo detector [14].

**Figure 4.** Time-frequency plots made with DF data extracted from the Lines database for VSR3. We clearly see the lines from west input mirror bulk modes.
4.3. *The VSR2 “Vela killer line”*

During VSR2 a non-stationary noise line affected the sensitivity of the Virgo ITF at the frequency of a GW possible signal from the Vela pulsar (22.38 Hz). The disturbance caused a loss of sensitivity at this frequency by about 20%. Unfortunately, NoEMi was not yet working at that time, therefore the noise line was only discovered during the off-line data analysis.

NoEMi found that the noise line (actually a doublet, see Fig. 5a and 5b) was coincident with the data of an accelerometer monitoring the vibrations of the Thermal Compensation System (TCS) [15] optical benches (Fig. 5c). Further investigations identified the source of the disturbance as being in the two chiller devices, which refrigerate and circulate water to cool down the TCS lasers, located just out of the TCS room. The vibration source was identified in one mechanical component of the chillers, whose rotation frequency (22.4 Hz) matched the noise line. The vibration was probably transmitted to the TCS bench through the cooling pipes.

**Figure 5.** The disturbance at the Vela pulsar frequency seen in the Events DB (a), in the Lines DB (b) and in the Events DB of an accelerometer (c) (VSR2). Plot d) shows the effect (on the accelerometer data) of the shift away from the Vela region of the frequency of the chiller engine causing the disturbance, on the 49th day of VSR3. The blue band show the Doppler band of the Vela pulsar frequency.

During VSR3 the noise line near the Vela signal frequency was no longer visible in the DF, however, it was still in the accelerometer: therefore we suspected that the line was just buried under the ITF noise, which at the Vela frequency was 2 to 3 times worse with respect to VSR2. A small but indicative coherence was indeed found between the DF and the accelerometer data. The adopted solution, in order to move the disturbance away from the Vela region, was to install a variable frequency drive to change the rotation frequency of the chiller engines (Fig. 5d).
4.4. The VSR4 “Vela bump”
As a final example, we describe the case of another disturbance affecting the sensitivity at the Vela pulsar frequency, at the beginning of VSR4, which confirmed the efficiency of the NoEMi framework as a monitoring tool. Before the science run began, NoEMi detected a disturbance close to the Vela frequency and raised an alarm on the summary pages. In the peak map time-frequency plots the disturbance appeared as a region with a higher peak density (Fig. 6), over which a comb of lines emerged (which triggered the alarm).

![Figure 6](image1.png)

**Figure 6.** Time-frequency plot of the peak maps at the Vela frequency region, showing the noise bump (enclosed in the dashed rectangles) before and after the change of the frequencies of the calibration lines. The red line shows the frequency of the Vela pulsar.

![Figure 7](image2.png)

**Figure 7.** Left: DF spectrum at the Vela frequency region showing the comb of noise lines. Right: Effect of the calibration signal switch on the spectrum.

The noise hunting began to observe that two pairs of lines were equally spaced by 0.2 Hz on each side of two “remarkable” frequencies: 22.5 and 23.0 Hz (Fig. 7 Left). Since 0.2 Hz is the fundamental resonance of the Virgo Super-Attenuator (SA) [17], it was assumed that the line sidebands were caused by some coupling mechanism between the SA mode and the two frequencies.
However, this was not yet the end of the story, as no “carrier” line was found at either 22.5 or 23.0 Hz! Following deeper investigations, the conclusion was that the two frequencies could be obtained as the difference between the frequencies of a set of calibration and control lines (sinusoidal signals injected into the interferometers for calibration and control purposes) at 379 and 356 (356.5) Hz. To confirm the hypothesis the calibration lines were switched off: the bump and line comb disappeared (Fig. 7 right). The disturbance was then assumed to be caused by a non-linear coupling mechanism in the Digital-to-Analogue converters used for control signals.

Since the calibration signals could not be switched off for the rest of the data-taking period, the solution adopted was to shift the 356 and 356.5 Hz lines down by 5 Hz: the disturbance then moved to ≈ 28 Hz, where no known pulsar signal is expected. Thanks to the intervention, the sensitivity at the Vela frequency was increased by ≈ 50%. This allowed us to obtain, for VSR4, an integrated sensitivity gain of about 1.7 with respect to VSR2, the data from which had been used to set the first direct upper limit on continuous GW emissions from the Vela pulsar [8].

5. Conclusions
In this paper we described the NoEMi framework, the Virgo tool for noise line monitoring and cataloguing. NoEMi exploits some of the algorithms implemented for the CW analysis as its main engine and is integrated into the Noise Monitor API framework [12], which provides a web application for the publication of daily analysis results and an interface for its databases. NoEMi monitored the third and fourth Virgo scientific runs, ensuring, in particular, the data quality at the frequencies of the known pulsars; it also proved to be a useful tool in off-line noise hunting and detector characterisation. The line catalogue filled by NoEMi will be used to veto the noisy frequency bands in the CW all-sky search analysis. The demonstrated potentials of NoEMi promoted the plan to include it in the Noise Monitors of the Advanced generation of both the Virgo and LIGO detectors [14].

References
[1] Cella G and Giazotto A 2011 Interferometric gravity wave detectors Rev. Sci. Instrum. 82 101101
[2] Accadia T et al 2011 Calibration and sensitivity of the Virgo detector during its second science run Class. Quantum Grav. 28 025005
[3] Acernese F et al 2008 Class. Quantum Grav. 25 114015
[4] Abbott B et al 2009 Rep. Prog. Phys. 72 076901
[5] Grote H et al 2010 Class. Quantum Grav. 27 084003
[6] Palomba C for the LIGO Scientific collaboration and the Virgo Collaboration 2011 Searches for continuous gravitational wave signals and stochastic backgrounds in ligo and virgo data Proceedings of Moriond 2011 (to be published)
[7] Abbott P et al 2010 Searches for gravitational waves from known pulsars with science run 5 LIGO data ApJ 713 671
[8] Abadie J et al 2011 Beating the spin-down limit on gravitational wave emission from the Vela pulsar ApJ 737 93
[9] Antonucci F, Astone P, D’Antonio S, Frasca S and Palomba C 2008 Detection of periodic gravitational wave sources by Hough transform in the frequency and spin down plane Class. Quantum Grav. 25 184015
[10] Acernese F et al 2009 Cleaning the Virgo sampled data for the search of periodic sources of gravitational waves Class. Quantum Grav. 26 204002
[11] http://www.mysql.com
[12] Accadia T et al 2011 Noise monitor tools and their application to Virgo data These proceedings
[13] Carbognani F, Colla A, Cuoco E and Hemming G 2011 Lines database web interface software requirements Virgo internal note VIR-027A-11
[14] Accadia T et al 2011 Advanced Virgo Proceedings of Moriond 2011 (to be published)
[15] Accadia T et al 2011 Virgo+ Thermal Compensation System These proceedings
[16] Puppo P 2009 Virgo+:MS sensitivity curve Virgo internal note VIR-0639E-09
[17] Accadia T et al 2011 The seismic Superattenuators of the Virgo gravitational waves interferometer Journal of Low Frequency Noise, Vibration and Active Control 30 63