Examination of natural noise signals: preliminary results for aquatic environment

H Zhivomirov\textsuperscript{1,2} and I Nedelchev\textsuperscript{1,2}

\textsuperscript{1}Department of Theory of Electrical Engineering and Measurements, Technical University of Varna, Varna, Bulgaria
\textsuperscript{2}SigNautic Lab – Examination of Underwater Noises, Signals and Vibrations of Marine Vessels and Structures, Technical University of Varna, Bulgaria

E-mail: hristo_car@abv.bg

Abstract. The present article treats the signal analysis and the revealing of the properties of noise signals which are caused by cavitation, boiling, flow and motion of water (aquatic environment) in natural conditions. All examined signals are not of androgenic and/or technogenic origin and corresponds to a few natural processes, e.g., boiling water, rain, sea surf. All analyses are conducted in the Matlab\textsuperscript{\textregistered} environment and include: computation of the mean and RMS values, dynamic range, crest factor, skewness and kurtosis, autocorrelation time; test for stationarity; plot of the oscillogram, correlogram, histogram, periodogram and spectrogram. Finally, some common conclusions are drawn, based on the analyses results.

1. Introduction

The present paper treats the signal analysis and the revealing of the properties of the natural stochastic noise signals and contains some preliminary results for the corresponding processes in the aquatic environment. All examined processes, which are an object of consideration here, have natural essence, that is, they are not of androgenic and/or technogenic origin. By means of the signal analysis one draws a conclusion about the properties of the underlying process.

In the present paper, the term “noise signal” is referred to a signal produced by a stochastic process. The term “coloured noise” used bellow, refers to noise signal whose power spectral density (PSD) is proportional as \( \text{PSD} \propto f^\alpha \) for \( \alpha \in \mathbb{R} \). The definition of the colored noise signal (e.g., white noise) implies an infinity bandwidth, and hence infinity energy and so it is a purely theoretical and non-casual. In the real-world applications the noise signal is considered over a frequency range relevant to the context [1].

Often, it is incorrectly assumed the terms “Gaussian noise” and “white noise” to be used interchangeably. It must be reminded, that Gaussianity refers to the amplitude probability distribution in the time domain, while the term “white noise” refers to the independently distributed power in the frequency domain and neither property implies the other [2].

2. Signal analysis techniques

A specified set of signal analysis techniques are used in order to reveal the signals behaviour and to classify the signals properly. The signals are presented and analysed in three different domains – the classic time and frequency domains (along with time-frequency one) and so-called statistical domain – a special notional (semantic) domain, where one is interested of specific signal parameters as mean level, peak level, dynamic range, probability density function, stationarity etc.
The signals of interest are produced by audio recording of real-world analog signals (i.e., mathematically real and time limited), and digitalized with sampling frequency $f_s$, according to the Nyquist-Shannon-Kotelnikov sampling theorem.

### 2.1. Revealing of the signal behaviour into the time domain

For revealing of the signal behaviour into the time domain two types of signal representation are used – classic oscillogram and correlogram (a result diagram from autocorrelation analysis). The autocorrelation is calculated as [3]

$$R_{xx}[m] = \frac{1}{N} \sum_{n} x[n] \cdot x[n+m]$$

(1)

By means of correlogram one examines the given signal for randomness or independence into the time domain.

### 2.2. Revealing of the signal behaviour into the frequency domain

A lot more information could be drawn about the signal, using a representation into the frequency domain, especially concerning the signal “colour” i.e., the spectral slope. In order to classify a signal as, for instance, pink or red noise, one must observe the PSD of the signal (since the signal is not deterministic). In this paper, the one-sided PSD is computed using the Welch’s modified periodogram [4, 5]

$$x_i[m] = w[m] \cdot x[m + lH]$$

$$\hat{X}[k,l] = \frac{2}{f_s \sum_{m} w[m]} \left| \sum_{n=0}^{K-1} x_i[m] e^{-2 \pi i mk/K} \right|^2$$

$$\hat{X}[k] = \frac{1}{L} \sum_{l} \hat{X}[k,l]$$

(2)

where \( m \in \{1,2,\ldots,M\} \) is the local time index (i.e., an index relative to the start of the signal frame), \( M \in \mathbb{N} \) is the window length, \( l \in \{0,1,\ldots,L-1\} \) is the frame index, \( L \in \mathbb{N} \) is the total number of frames; \( H \in \mathbb{N} \) is the hop size (i.e., the advance from one signal frame to the next), \( k \in \{1,2,\ldots,K\} \) is the frequency bin index and \( K \in \mathbb{N} \) is the DFT size.

Different types of coloured noise signals are distinguished by the roll-off of their PSD. A few of them are listed in Tab. 1.

Another useful instrument for analysis and visualization take place into the time-frequency domain – the one-sided amplitude spectrogram, defined as [6]:

$$\text{Spectrogram} = \sum_{m} \left| \hat{X}[k,l] \right|^2$$

(3)

where

$$\hat{X}[k,l] = \frac{1}{M} \sum_{m=0}^{K-1} x_i[m] e^{-2 \pi i mk/K}$$

(4)

is the time-localized two-sided spectrum (i.e., Short-Time Fourier Transform, STFT) of the analysed signal frame \( l \). More about the applied STFT analysis could be found in [6].
Table 1. Information about some colored noise signals [2].

| Noise type    | PSD dependence | Spectral slope rate | Examples                                                                 |
|---------------|----------------|---------------------|-------------------------------------------------------------------------|
| Violet noise  | $\propto f^2$  | +6 dB/oct, +20 dB/dec | Signal from an acoustic thermal noise of water [1]                       |
| Blue noise    | $\propto f$    | +3 dB/oct, +10 dB/dec | Signal from a Cherenkov radiation process [1]                            |
| White noise   | flat           | 0 dB/oct, 0 dB/dec  | Signal from a white noise process (e.g., Johnson–Nyquist thermal noise) |
| Pink noise    | $\propto 1/f$  | -3 dB/oct, -10 dB/dec| Signal from statistical fluctuations of a number of natural processes [1]|
| Red noise     | $\propto 1/f^2$| -6 dB/oct, -20 dB/dec| Signal from a Brownian motion (Winner process) [1]                       |
| Black noise   | -              | > -6 dB/oct, > -20 dB/dec | Model of the frequency of the natural disasters                         |

2.3. Signal statistics

As far as the analysed signals are stochastic, their statistical properties are of great significance. One is interested in the magnitude statistics and histogram representation (including skewness and kurtosis values) of a given stochastic signal, as well as if it is stationary or not. The properties of interest are defined as in [4] and are shown in Table 2.

Table 2. Signal statistical parameters notations and definitions.

| Notation | Name            | Definition                                                                 |
|----------|-----------------|---------------------------------------------------------------------------|
| $X_{0}$  | Mean value      | $\frac{1}{N} \sum_{n} x[n]$                                              |
| $X$      | Root-Mean-Square value | $\left( \frac{1}{N} \sum_{n} x[n]^2 \right)^{1/2}$                     |
| $S$      | Skewness        | $\frac{1}{N} \sum_{n} (x[n] - X) \cdot (x[n] - X)^3$                     |
| $K$      | Kurtosis        | $\frac{1}{N} \sum_{n} (x[n] - X) \cdot (x[n] - X)^4$                     |
| $CF$     | Crest-factor    | $20 \log_{10} \left( \frac{\max(|x[n]|)}{\min(|x[n]|)} \right)$         |
| $DR$     | Dynamic range   | $20 \log_{10} \left( \frac{\max(|x[n]|)}{\min(|x[n]|)} \right)$         |

The histogram displays the number of samples there are in the signal that have each of these possible values, grouping together samples that have one and the same value and is used to reveal the probability density function (PDF) of the analysed signal and eventually to classified it as some of the know PDF-types (e.g., normal distribution or Laplace distribution).

Whether the signal (i.e., the time sequence) of interest is stationary or not (e.g., has a unit root) could be checked by the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [7] and the augmented
Dickey–Fuller test (ADF) [8]. The KPSS test is used for testing a null hypothesis that an observable time series is stationary around a deterministic trend \( i.e. \), trend-stationary against the alternative of a unit root, while ADF tests the null hypothesis that a unit root is present in a time series sample. In the paper both tests are used for better results.

3. Signal examination results
An experimental examination has been made of over twenty natural sounds recorded with a linear PCM digital audio recorder. Nine of them are listed on Tab. 3 by their name and duration. Objects of consideration are primarily sounds caused by water dynamic movement – cavitation, boiling, flow, motion. For sake of comparison, two additional sounds have been analysed – fire of wooden-burning stove and sound caused by wind in the wood. All sounds are in .wav file format, sampled at \( f_s = 44100 \) Hz and 24-bit coded.

All recorded signals are analysed in Matlab\(^\text{®} \) environment using a special developed script for noise analyses, which provides signal visualization in the time, frequency and time-frequency domains via oscillogram, periodogram and spectrogram, respectively. Also it provides visualization of the signal correlogram and histogram along with rich signal statistics: positive and negative peak values, mean and RMS values, signal dynamic range, crest factor, skewness and kurtosis computation, and measurement of autocorrelation time.

At the beginning of the analysis every signal is detrended and normalized to unity peak amplitude. The PSD is computed using the Welch’s modified periodogram with the Hamming window with length \( M = \frac{1}{500} \) of the signal length, hop size \( H = \frac{M}{4} \) and DFT size \( K = 2M \). The spectrogram is computed using STFT with Hamming window with length \( M = 1024 \) samples and the same hop and DFT size as the PSD. Some of the statistical parameters of the examined signals are listed in Tab. 4.

Due to the restricted space, only two of the most informative plots are given below on Fig. 1 ÷ Fig. 18 – the periodogram and the histogram of the signals. The periodograms’ frequency ranges are restricted to two decades – from 100 Hz to 10 kHz. For sake of comparison, the histograms (in blue colour) are accompanied with the appropriate superimposed fitted distribution line (in red colour).

Table 3. List of the explored signals.

| Signal abr. | Signal full name               | \( T, \) s |
|-------------|--------------------------------|------------|
| SF          | sink_flow                      | 10         |
| BC          | boiling_water_cavitation       | 10         |
| BW          | boiling_water                 | 10         |
| UF          | urban_fountain                | 5          |
| WF          | waterfall                      | 10         |
| SS          | sea_surf                       | 5          |
| RN          | rain                           | 10         |
| WL          | winter_in_the_leafs            | 10         |
| WS          | wood-burning_stove            | 10         |

Table 4. Some statistical parameters of the explored signals.

| Parameter      | SF | BC | BW | UF | WF | SS | RN | WL | WS |
|----------------|----|----|----|----|----|----|----|----|----|
| \( CF, \) dB   | 13 | 13 | 16 | 19 | 18 | 20 | 23 | 15 | 14 |
| \( DR, \) dB   | 107| 76 | 77 | 112| 80 | 87 | 125| 125| 125|
| \( S \)        | 0.005| 0.006| -0.020| 0.008| 0.004| -0.140| -0.016| 0.007| -0.057|
| \( K \)        | 3.00| 3.08| 3.56| 3.96| 3.22| 6.65| 4.52| 3.33| 3.33|
| Autocorrelation time, s | 0.0027| 0.0127| 0.0045| 0.0351| 0.0014| 3.1372| 0.0038| 3.2208| 1.5433|
Figure 1. PSD-periodogram of the signal SF.

Figure 2. Histogram (blue) and the fitted normal distribution (red) of the signal SF.

Figure 3. PSD-periodogram of the signal BC.

Figure 4. Histogram (blue) and the fitted normal distribution (red) of the signal BC.

Figure 5. PSD-periodogram of the signal BW.

Figure 6. Histogram (blue) and the fitted normal distribution (red) of the signal BW.
Figure 7. PSD-periodogram of the signal UF.

Figure 8. Histogram (blue) and the fitted normal distribution (red) of the signal UF.

Figure 9. PSD-periodogram of the signal WF.

Figure 10. Histogram (blue) and the fitted normal distribution (red) of the signal WF.

Figure 11. PSD-periodogram of the signal SS.

Figure 12. Histogram (blue) and the fitted Laplace distribution (red) of the signal SS.
Figure 13. PSD-periodogram of the signal RN.

Figure 14. Histogram (blue) and the fitted normal distribution (red) of the signal RN.

Figure 15. PSD-periodogram of the signal WL.

Figure 16. Histogram (blue) and the fitted normal distribution (red) of the signal WL.

Figure 17. PSD-periodogram of the signal WS.

Figure 18. Histogram (blue) and the fitted normal distribution (red) of the signal WS.

The following notes can be taken on the results:
• All signal segments are stationary, according to KPSS and ADF tests, which is also confirmed by the signals’ spectrograms;
• The crest factor of the signals varies between 13 dB and 23 dB, while most of them are between 13 dB and 20 dB (approx. 5 to 10 times the RMS value). The signals with the biggest CF are the sea_surf and the rain signals. They also have the biggest kurtosis (i.e., peakedness) of all signals, which could be observed on their histograms, too.
• The dynamic range of the signals varies between 75 dB and 125 dB. Some of the purely nature signals have the biggest DR – rein, winter_in_the_leafs, and wood-burning_stove. One of the possible explanations is their “discrete time” pattern – they could be presented as a set of (partially) non-overlapping micro-events.
• The skewness of all signals is close to zero, which means that all signals are symmetric and centred. This could be observed on their histograms, too.
• The kurtosis of the most signals is close to 3 (one of them is perfectly Gaussian – sinc_flow), which is an evident for normal (Gaussian) distribution of the most signals, which could be observed on their histograms, too. The bold exceptions are the rein and sea_surf signals. The last one obtains Laplace distribution of the signal amplitudes. Both of them have well pronounced peakedness.
• All signals obtain no autocorrelation, except the sea_surf and winter_in_the_leafs signals. These signals are long-term correlated, since in their time-domain pattern a kind of periodicity could be found (so-called “pulsations”).
• Regarding the PSDs of the signals:
  - the PSDs of the SF and BC signals are quite similar. Probably this is due to the similar nature of the underlying process – dynamic movement of air-bubbles in water environment;
  - the PSD of the BW signal is somewhat uniform, with resonances at ~3 kHz and ~6 kHz;
  - the PSD of the UF is alike a pink noise (roll-off of ~10 dB/dec) at the first frequency decade (100 Hz – 1000 Hz) and then band-limited to about 4 kHz;
  - the PSDs of the WF and RN signals are comparable with the white noise process, which is expected for this type of noise signals;
  - the PSD of the SS signal is like pink noise at the first decade and alike red noise (roll-off of ~20 dB/dec) at the second decade (1000 Hz – 10000 Hz), respectively;
  - the PSD of the WL signal is pink up to ~5 kHz, and then band-limited;
  - the PSD of the WS signal approximates red noise up to about 800 Hz and then is band-limited, probably by the volume of the stove, i.e. the noise is spatially low-pass filtered.

4. Conclusions
In the presented paper several types of signals, representing the movement of the aquatic environment, are reviewed and some preliminary results from their analysis are showed. Summaries of their nature were made, confirming the preliminary thesis: (i) the majority of signals have a normal (Gaussian) or close to normal probability distribution of the amplitudes; (ii) almost all signals have band-limited PSDs of “coloured” type – mostly like the white, pink and red noises. In addition, some signals obtained very similar PSDs, for instance SF-BC, WF-RN and SS-WL couples, although their different origin.

The obtained a priori information assists the further signal processing, for other purposes and applications. This study is a first step from further work on characterization of other noise signals – mainly in the underwater and underground environments. Those results can be used in the processing of signals from measurements in the aquatic environment, as well as in the development of some simulation models.

References
[1] Kasdin N 1995 Discrete simulation of colored noise and stochastic processes and 1/fα power law noise generation Proc. of the IEEE 5 83 802-27
[2] Zhivomirov H 2018 A method for colored noise generation RJAV 1 15 14-9
[3] Smith J 2008 *Mathematics of the Discrete Fourier Transform (DFT): with Audio Applications* (W3K Publishing) chapter 8 p 189

[4] Brandt A 2011 *Noise and Vibration Analysis: Signal Analysis and Experimental Procedures* (Chichester: John Wiley & Sons) chapter 10 p 211

[5] Smith J 2011 *Spectral Audio Signal Processing* (W3K Publishing) chapter 6 p 248

[6] Zhivomirov H 2019 On the development of STFT-analysis and ISTFT-synthesis routines and their practical implementation *TEM J.* 18 56-64

[7] Kwiatkowski D, Phillips P, Schmidt P and Shin Y 1992 Testing the null hypothesis of stationarity against the alternative of a unit root *Journal of Econometrics* 1-3 54 pp 159-178

[8] Dickey D and Fuller W 1979 Distribution of the estimators for autoregressive time series with a unit root *J. Am. Stat. Assoc.* 366(74) pp 427-431

**Acknowledgments**

This work was supported by the European Regional Development Fund within the OP “Science and Education for Smart Growth 2014 - 2020”, Project CoC “Smart Mechatronic, Eco- And Energy Saving Systems And Technologies”, № BG05M2OP001-1.002-0023