Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Implementation of empirical mode decomposition with sifting optimizations for real-time de-noising

Md. Rezwan Islam

241311084@vu.edu.bd

Varendra University

Akib Ikbal

Varendra University

Research Article

Keywords: Empirical mode decomposition, Sifting, De-noising, Real-time processing, Simulink, Electrocardiogram

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11771797/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Simulink implementation of empirical mode decomposition with sifting optimizations for real-time de-noising[1]

Abstract

Empirical mode decomposition (EMD) splits a signal into oscillatory components without a fixed basis, which makes it attractive for de-noising biological signals. In its usual form, however, it needs the whole record in memory and repeats sifting until a stopping rule is met, so processing time cannot be predicted. This paper describes a frame-based Simulink implementation of EMD with three changes aimed at real-time use: a fixed number of sifting passes, linear envelopes instead of cubic splines, and overlapping frames of which only the centre is kept. On a two-tone synthetic signal with white Gaussian noise (input signal-to-noise ratio, SNR, 3.98 dB), removing the first three intrinsic mode functions (IMFs) gave the best output SNR. Three sifting passes reached 13.66 dB and four passes 13.80 dB, while eight and twelve passes lowered it to 12.98 and 12.42 dB. Linear envelopes beat cubic splines by about 2 dB (13.66 against 11.63 dB) and avoided a large start-up transient. An overlap of 64 samples or more improved the SNR by about 0.2 dB and shortened the latency. On a 20 s electrocardiogram (ECG) segment from MIT-BIH record 100 with noise added at an input SNR of about 4 dB, removing only the first IMF raised the SNR to 8.10 ± 0.20 dB (ten noise realizations), and R peaks kept 0.95 of their height on average, but the P and T waves were not recovered. The synthetic results come from one signal and one noise realization, and execution on hardware was not measured.

Keywords

Empirical mode decomposition; Sifting; De-noising; Real-time processing; Simulink; Electrocardiogram

1. Introduction

Biological signals such as the electrocardiogram (ECG) are recorded together with muscle activity, electrode motion and mains interference. For monitoring and wearable devices the cleaning has to happen while the signal arrives, not after the recording is finished. Linear filters assume the signal is stationary, which is rarely true of a heartbeat. Empirical mode

[1] Abbreviations: ECG, electrocardiogram; EMD, empirical mode decomposition; IMF, intrinsic mode function; SNR, signal-to-noise ratio.

decomposition (EMD) avoids that assumption. It splits a signal into intrinsic mode functions (IMFs) using only the signal's own local extrema, ordered roughly from the fastest oscillation to the slowest [1]. Noise tends to collect in the first IMFs, so dropping or thresholding them often cleans a signal without smearing its real features [2,3].

EMD was designed as an offline method. Each IMF is extracted by sifting: find the maxima and minima, fit upper and lower envelopes with cubic splines, subtract their mean, and check a stopping criterion [1,4]. This needs the whole record in memory, runs a variable number of iterations for every IMF, and distorts the ends of the record. Ensemble versions add many noisy copies of the signal and multiply the cost further [5]. None of this fits a system that must finish a fixed amount of work in every sample period.

Simulink is a convenient place to test whether EMD can meet that kind of deadline. A model can be simulated, inspected block by block, and later passed to a code generator. EMD de-noising is usually studied in offline scripts, which says little about timing. Here we describe a frame-based Simulink implementation of an EMD de-noiser and report what three changes to the sifting process do to its output quality. The changes are a fixed number of sifting passes, linear envelopes, and overlapping frames with only the centre of each frame kept.

2. Materials and methods

2.1 Sifting and IMF extraction

For a frame x, the first IMF is found by repeating the sifting step

$$h_{\mathrm{j}} = h_{\mathrm{j}-1} - (u_{\mathrm{j}-1} + l_{\mathrm{j}-1}) / 2 \tag{1}$$

where u and l are the upper and lower envelopes and h_0 is the input of the stage. After M passes the stage output is the IMF $c_{\mathrm{k}} = h_{\mathrm{M}}$, and the residue passed to the next stage is

$$r_{\mathrm{k}} = r_{\mathrm{k}-1} - c_{\mathrm{k}}, \quad r_0 = x \tag{2}$$

The de-noised frame is a weighted sum of the IMFs and the final residue,

$$y = w_1 c_1 + \ldots + w_{\mathrm{K}} c_{\mathrm{K}} + r_{\mathrm{K}} \tag{3}$$

where a weight of 0 removes an IMF and a weight of 1 keeps it. We used $K = 4$ stages.

2.2 Model architecture

The model (Fig. 1) takes the input signal, adds white noise, and cuts the stream into frames of $N = 512$ samples with a Buffer block. Consecutive frames overlap by V samples. Each frame

goes to a subsystem called EMD Denoiser, which holds the four sifting stages (Fig. 2). Every stage is a MATLAB Function block that returns one IMF, a Sum block that subtracts it from the incoming signal to give the residue, a Constant holding the weight w_k, and a Product block. A final Sum block adds the four weighted IMFs and the last residue. A Selector block keeps samples $V/2 + 1$ to $N - V/2$ of each processed frame, and an Unbuffer block turns the frames back into a continuous signal. The solver was fixed-step and discrete, with a step equal to the sample period. The output lags the input by $N - V/2$ samples.

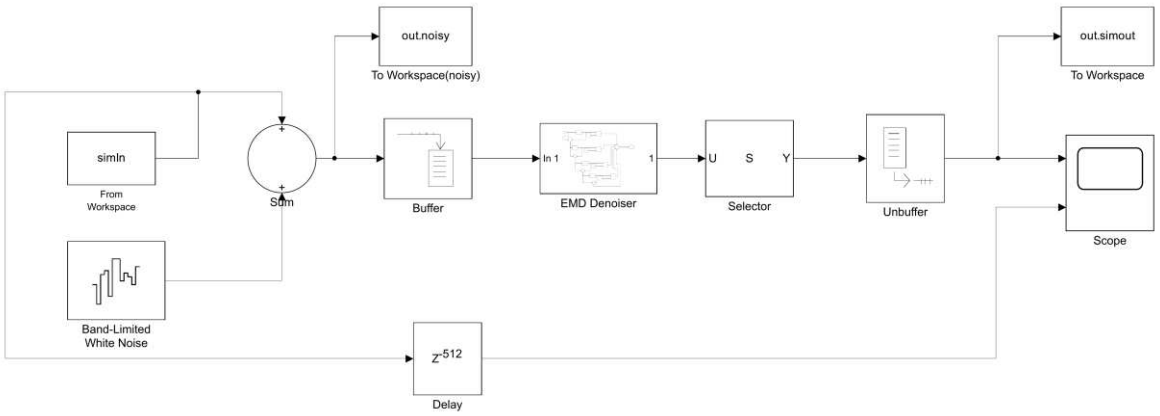


Fig. 1. Top-level Simulink model. The noisy input is cut into frames by the Buffer block, processed by the EMD Denoiser subsystem, trimmed by the Selector block and rebuilt into a continuous signal by the Unbuffer block.

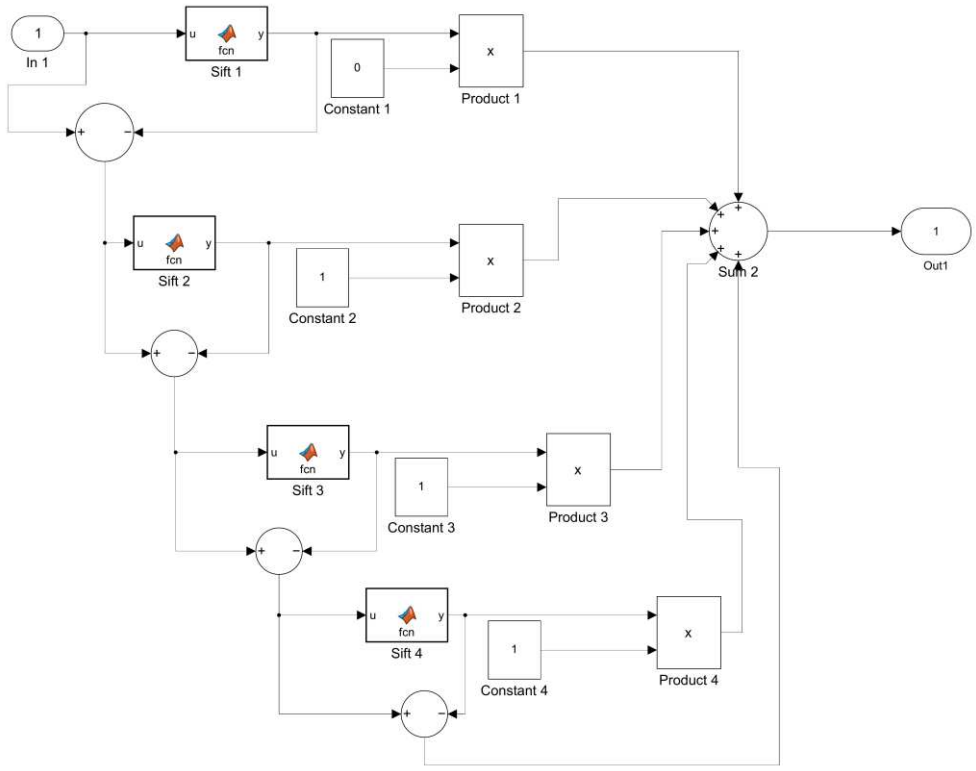


Fig. 2. Inside the EMD Denoiser subsystem: four cascaded sifting stages, each with a residue subtractor, a weight constant and a product, followed by a final sum.

2.3 Sifting optimizations

Three changes were made to the standard sifting process. First, every stage runs a fixed number of passes M instead of repeating until a stopping criterion is met, so the work per frame is known in advance. Second, the envelopes are straight lines through the maxima and through the minima, not cubic splines. The first and last samples of the frame are added to both sets of extrema so that the envelopes cover the whole frame. If a frame has fewer than two interior maxima or minima, it is returned unchanged. Third, frames overlap and only the centre of each processed frame is kept, because sifting distorts the first and last samples most. The kept centres of consecutive frames join without overlap-add. The computation per output sample is $N / (N - V)$ times that of non-overlapping frames.

2.4 Test signals

The synthetic signal was $x(t) = \sin(2\pi \cdot 5t) + 0.5 \sin(2\pi \cdot 20t)[1 + 0.5 \sin(2\pi \cdot 0.5t)]$, sampled at 1 kHz for 10 s. White Gaussian noise of variance 0.25 was added with a fixed random seed. The measured input SNR was 3.98 dB.

For the ECG test we used record 100 of the MIT-BIH Arrhythmia Database [6,7], channel MLII, sampled at 360 Hz. A 20 s segment starting at 10 s was taken, its mean removed, and the result used as the reference signal. White noise was added to give a target input SNR of 4 dB (measured: 4.05 ± 0.11 dB over ten noise realizations). The ECG model used frames of N = 512 samples without overlap (V = 0), four sifting stages, three passes per stage and linear envelopes. The recording already contains some noise, so the reference is not noise-free.

2.5 Evaluation

The output was aligned to the input using the delay estimated by cross-correlation, which matched $N - V/2$ in every run. The first and last 1000 samples were excluded. For the ECG, the delay was found by searching lags from 0 to 600 samples for the highest SNR. It was 512 samples, equal to N − V/2 with V = 0, and the SNR was computed over all aligned samples (6688 of the 7200 in the segment). The output SNR was

$$\mathrm{SNR}_{\mathrm{out}} = 10 \log_{10} [\, \Sigma\, x^2 / \Sigma\, (y - x)^2 \,] \tag{4}$$

and the gain is the difference from the input SNR. For the ECG we also report a peak ratio for each R peak: the largest de-noised sample within 5 samples (14 ms) of the reference peak, divided by the reference peak height. R peaks were taken as local maxima of the reference above 0.6 mV, more than 20 samples apart, which gave 23 beats. Because the largest of several

noisy samples is taken, this ratio is biased slightly upward. The selected weight set was run with ten noise seeds, and the mean and standard deviation (SD) are reported. All work was done in MATLAB and Simulink R2023b with the Signal Processing Toolbox and the DSP System Toolbox.

3. Results

3.1 Choice of IMFs to remove

Table 1 compares weight settings on the synthetic signal with eight passes per stage and linear envelopes. Removing the first three IMFs gave the highest output SNR, 12.98 dB, a gain of 9.00 dB over the input. Removing a fourth IMF lowered the SNR to 11.92 dB.

Table 1. Output SNR for different sets of removed IMFs (synthetic signal, eight passes, linear envelopes, input SNR 3.98 dB).

Removed IMFs	Output SNR (dB)	Gain (dB)
1	7.84	3.86
1–2	10.71	6.73
1–3	12.98	9.00
1–4	11.92	7.94

3.2 Number of sifting passes

With IMFs 1 to 3 removed, the output SNR rose steeply from 6.75 dB at one pass to 11.14 dB at two and 13.66 dB at three (Fig. 3). It peaked at 13.80 dB with four passes and then fell to 12.98 dB at eight and 12.42 dB at twelve. Three passes came within 0.14 dB of the best result with a quarter less sifting work than four. With one noise realization, a difference of that size between three and four passes cannot be called real.

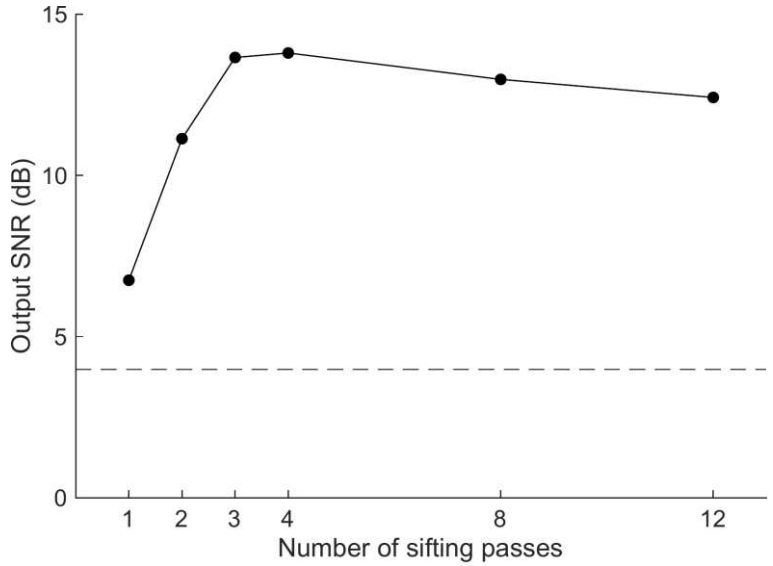


Fig. 3. Output SNR against the number of sifting passes per IMF (IMFs 1–3 removed, linear envelopes). The dashed line is the input SNR.

3.3 Envelope interpolation

With three passes, linear envelopes gave 13.66 dB and cubic spline envelopes 11.63 dB (Table 2). The spline version also produced a large transient in the first output frame, reaching about 130 for a signal whose amplitude stays below 2. We expect this comes from the spline extending across the flat initial-condition samples at the start of the first frame, but we did not test that explanation. In MATLAB R2023b on a 12th Gen Intel Core i5-12450H processor, one frame of 512 samples (four stages of three passes) took 0.372 ms with linear envelopes and 0.365 ms with cubic splines, so the choice of envelope did not change the execution time in this test. This timing is for MATLAB code, not for the compiled Simulink model or for fixed-point hardware, where a cubic spline may cost more, but we did not test that.

Table 2. Output SNR for two envelope interpolation methods (synthetic signal, three passes).

Interpolation	Output SNR (dB)	Start-up transient
Linear	13.66	None seen
Cubic spline	11.63	About 130

3.4 Frame overlap

Table 3 shows the effect of overlap with three passes and linear envelopes. Without overlap the SNR was 13.45 dB. Overlaps of 64, 128 and 256 samples gave 13.63, 13.66 and 13.63 dB, so most of the benefit was already present at 64 samples. Larger overlap shortened the latency but raised the work per output sample.

Table 3. Effect of frame overlap V (synthetic signal, N = 512, three passes, linear envelopes).

V	Output SNR (dB)	Latency (samples)	Latency (ms)	Relative work
0	13.45	512	512	1.00
64	13.63	480	480	1.14
128	13.66	448	448	1.33
256	13.63	384	384	2.00

3.5 ECG results

Table 4 lists the output SNR for eight weight settings on the ECG segment (one noise realization, input SNR 4.00 dB over the aligned window). Removing only the first IMF gave the highest SNR, 7.92 dB. Removing IMF 2 or IMF 3 as well lowered it to 5.94 and 6.46 dB, and removing IMF 2 or IMF 3 alone left the SNR below the input. This differs from the synthetic signal, where removing the first three IMFs worked best.

Table 4. Output SNR for different IMF weights on the ECG segment (record 100, 360 Hz, one noise realization, input SNR 4.00 dB).

Weights (w1 w2 w3 w4)	Output SNR (dB)	Gain (dB)
1 1 1 1	4.00	0.00
0 1 1 1	7.92	3.92
0 0 1 1	5.94	1.93
0 0 0 1	3.51	−0.49
1 0 1 1	3.74	−0.26
1 1 0 1	3.55	−0.46
1 1 0 0	2.48	−1.52
0 1 0 1	6.46	2.45

With weights 0 1 1 1 the run was repeated with ten noise seeds. The input SNR was 4.05 ± 0.11 dB, the output SNR 8.10 ± 0.20 dB and the gain 4.04 ± 0.17 dB (mean ± SD). The measured delay of 512 samples matched $N - V/2$ for $V = 0$.

Across the 23 R peaks the mean peak ratio was 0.95 (SD 0.11, median 0.95, range 0.64 to 1.10). Four peaks were attenuated to below 0.85, at no consistent position in the record. Ratios above 1 occur because the largest sample near a noisy peak was taken. The R peaks stayed in place and clearly visible (Fig. 4), but the P and T waves were not recovered, and high-frequency noise remained on the baseline.

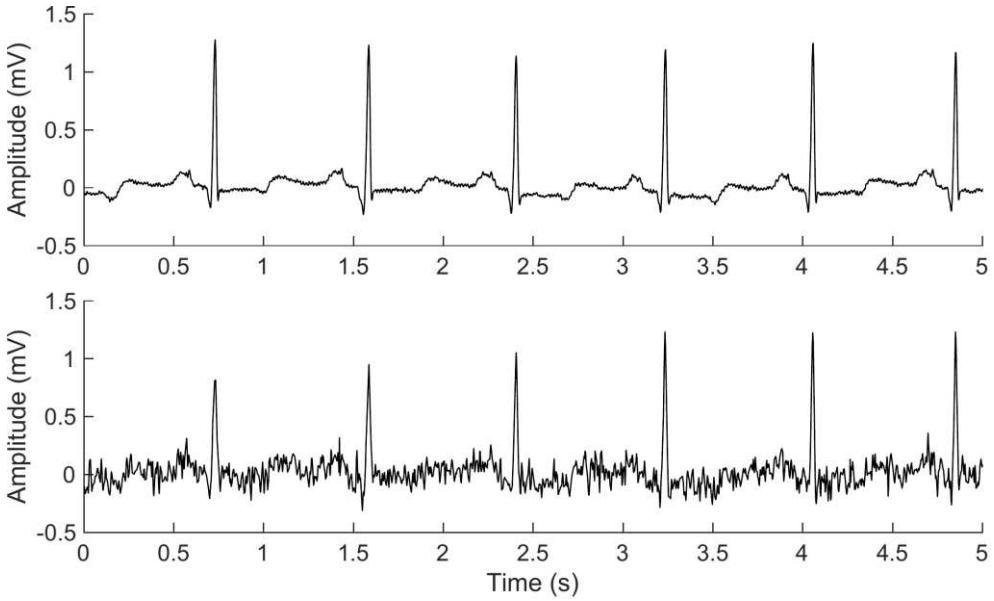


Fig. 4. Clean reference (top) and de-noised output (bottom) for the first 5 s of the record 100 segment (weights 0 1 1 1, input SNR about 4 dB).

4. Discussion

On the synthetic signal the best results came from a small, fixed number of passes. Too few passes leave the IMFs unfinished, and too many seem to over-process them: the SNR was highest at three to four passes and fell steadily afterwards. We read the drop at high pass counts as over-sifting, which flattens the amplitude variation that belongs to the signal, but the experiments here do not prove that. The practical point is that the usual practice of many passes bought nothing in this model, and a fixed count of three costs less than the iterate-until-converged approach.

Linear envelopes were better than cubic splines on this signal, and they avoided the start-up transient. Linear interpolation cannot overshoot between extrema, which is a useful property in a system that must not produce wild outputs. The effect of frame overlap was small on this signal, where the two tones vary slowly. A signal with sharper features, such as an ECG, may need more overlap.

On the ECG segment the best setting differed from the synthetic case. Removing only the first IMF gave a gain of about 4 dB, and removing the second or third IMF as well lowered the SNR. We expect this is because the QRS complex is a sharp, high-frequency event whose energy reaches the first few IMFs, so removing them costs signal as well as noise, but we did not test that explanation. R peaks kept 0.95 of their reference height on average, but a few beats fell to between 0.64 and 0.84, and the P and T waves were not recovered, so the gain in SNR does not mean the whole waveform was restored. The ECG run used no overlap. The synthetic

results suggest that overlap helps slightly and that a signal with sharp features may need more, so overlapping frames may do better here, but we did not test it.

The study has clear limits. The synthetic experiments used one signal and one noise realization, and the ECG experiments one 20 s segment of one record, so differences of a few tenths of a dB cannot be separated. The ECG noise was white noise that we added, not real recording artifact such as muscle activity or electrode motion. SNR is an average over thousands of samples and can hide short local errors. Real time here means a fixed amount of work per frame in simulation; we did not measure execution on hardware. The ECG reference contains noise of its own, and the fixed-point implementation has not been done.

5. Conclusions

A frame-based Simulink model of EMD de-noising with a fixed number of sifting passes, linear envelopes and overlapping frames gave a gain of about 9.7 dB on a noisy two-tone signal. Three to four passes per IMF worked best, more passes lowered the SNR, linear envelopes beat splines by about 2 dB, and an overlap of 64 samples captured most of the benefit of overlapping. On a 20 s ECG segment at an input SNR of about 4 dB, removing only the first IMF raised the SNR by about 4 dB and kept R peaks at 0.95 of their height on average, but the P and T waves were not recovered. The fixed-point implementation and execution on hardware remain to be done.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the first author used Claude (Anthropic) in order to help build the Simulink models, to help write and debug the MATLAB scripts, and to help draft and edit the manuscript text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A 1998;454:903-95.

[2] Boudraa AO, Cexus JC. EMD-based signal filtering. IEEE Trans Instrum Meas 2007;56:2196-202.

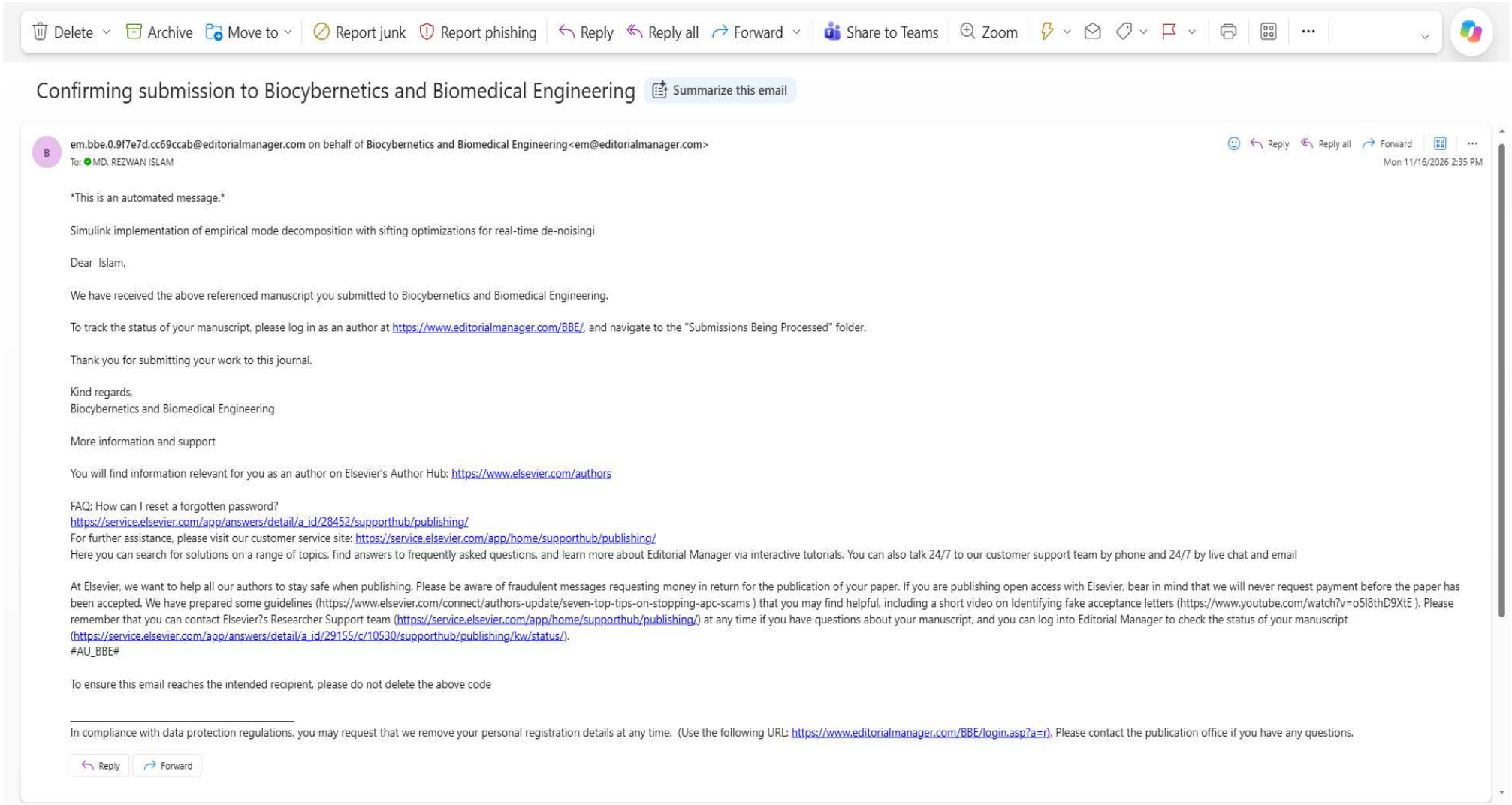
[3] Kopsinis Y, McLaughlin S. Development of EMD-based denoising methods inspired by wavelet thresholding. IEEE Trans Signal Process 2009;57:1351-62.

[4] Rilling G, Flandrin P, Gonçalves P. On empirical mode decomposition and its algorithms. In: Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP); 2003.

[5] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 2009;1:1-41.

[6] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol Mag 2001;20:45-50.

[7] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101:e215-20.


Delete
Archive
Move to
Report junk
Report phishing
Reply
Reply all
Forward
Share to Teams
Zoom
Confirming submission to Biocybernetics and Biomedical Engineering
Summarize this email
em.bbe.0.9f7e7d.cc69ccab@editorialmanager.com on behalf of Biocybernetics and Biomedical Engineering<em@editorialmanager.com>
To: MD. REZWAN ISLAM
Reply
Reply all
Forward
Mon 11/16/2026 2:35 PM
This is an automated message.
Simulink implementation of empirical mode decomposition with sifting optimizations for real-time de-noisingi
Dear Islam,
We have received the above referenced manuscript you submitted to Biocybernetics and Biomedical Engineering.
To track the status of your manuscript, please log in as an author at https://www.editorialmanager.com/BBE/, and navigate to the "Submissions Being Processed" folder.
Thank you for submitting your work to this journal.
Kind regards,
Biocybernetics and Biomedical Engineering
More information and support
You will find information relevant for you as an author on Elsevier's Author Hub: https://www.elsevier.com/authors
FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site: https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email
At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5I8thD9XtE). Please remember that you can contact Elsevier?s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).
#AU_BBE#
To ensure this email reaches the intended recipient, please do not delete the above code
In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/BBE/login.asp?a=r). Please contact the publication office if you have any questions.
Reply
Forward