JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Hybrid Signal Denoising in Seismic Applications Using Sinc Filtering and Machine Learning Regression

**Rudra Krishna**
MSc. Advanced Electrical and Electronics Engineering
School of Engineering
University of Leicester, Leicester, England, United Kingdom

***Abstract :*** Accurate seismic signal analysis is vital for earthquake detection, subsurface imaging, and geotechnical assessments. However, field-acquired seismic data is often contaminated by various noise sources. Traditional low-pass filters, like sinc-based FIR filters, reduce high-frequency noise but may distort signal features. This study presents a hybrid denoising approach combining a sinc FIR filter with a machine learning (ML) enhancement using a linear regression (LinearFit) model. A clean seismic signal was simulated and degraded with additive white Gaussian noise, then sequentially processed through the sinc filter and ML model. Results show that while the sinc filter effectively suppresses noise, the LinearFit ML model further reduces residual errors and better preserves signal integrity, offering an efficient and interpretable tool for seismic signal denoising.

***Key-Words*- Seismic Denoising, Sinc Filter, Linear Regression, Machine Learning, Signal Processing, MATLAB, Gaussian**

## I. INTRODUCTION

Seismic data is essential for gaining insights into Earth's subsurface, aiding both in earthquake monitoring and the exploration of natural resources. The Seismic data, whether for earthquake monitoring or resource exploration, must be as clean and interpretable as possible [1]. However, raw sensor data is often corrupted by background noise, which hampers downstream processing such as waveform picking, velocity modeling, and migration imaging [2].

A typical noisy observed seismic signal can be simply represented as:

$$x(t) = s(t) + \eta(t)$$

Where,

x(t) is the noisy observed signal,
s(t) is the clean or untainted seismic waveform,
Ƞ(t) is the additive noise, which is assumed as zero-mean Gaussian white noise in this paper

Traditional signal processing tools such as Butterworth or sinc filters have long been used to mitigate noise. These filters operate by convolving the input with an impulse response, which is represented as

$$y[n] = \sum_{k=0}^{N-1} x(n-k) \times h(k)$$

However, these approaches often face limitations: they may introduce phase distortion, attenuate critical features of the waveform, or fail under variable noise conditions [3].

Recently, machine learning methods have shown potential in signal denoising tasks across biomedical, audio, and seismic domains [4]. However, ML models often require extensive data and training time, and are prone to overfitting. Hence, a hybrid strategy that leverages the strengths of classical filtering and ML-based pattern enhancement may yield better practical results .

This paper investigates the effectiveness of such a hybrid method by simulating a clean seismic waveform and evaluating how well a sinc filter followed by an ML-inspired transformation can recover the original signal.

## II. METHODOLOGY

This section presents the full implementation pipeline, covering signal simulation, the application of a sinc filter, hybrid ML-based enhancement simulation, and residual analysis. All components are executed in MATLAB and are elaborated upon in the following sections.

**Signal Simulation**

To emulate realistic seismic signals, a synthetic waveform is generated using a damped sinusoid , which can be represented by the below equation:

$$s(t) = \sin(2 \times \pi \times \mathrm{f} \times \mathrm{t}) \times e^{-\alpha \times t}$$

In MATLAB, by taking the sampling frequency, $f_s$ as 100 Hz, fundamental frequency f as 1 Hz , α as -0.02 and time vector for 60 seconds, the below syntax generated the seismic signal , with variable defined as clean_signal.

```
fs = 100;
t = 0:1/fs:60;
clean_signal = sin(2*pi*1*t) .* exp(-0.02*t);
```

clean_signal models a typical seismic wave: a sinusoid with frequency 1 Hz multiplied by an exponential decay factor , thus simulating the energy attenuation seen in real seismic events.

**Incorporating Gaussian Noise**

Gaussian noise is then added to mimic ambient background interference using the below syntax in the MATLAB script:

```
noise = 0.8 * randn(size(t));
noisy_signal = clean_signal + noise;
```

randn(size(t)) function of MATLAB generates zero-mean Gaussian noise of the same size as the time vector and 0.8 factor scales the noise amplitude, controlling the noise-to-signal ratio.

**Sinc Filter Design**

A sinc filter is utilized to attenuate high-frequency noise, enhancing the clarity of the signal. The sinc function is an even function which is symmetric about the y axis. Sinc can be used as an ideal lowpass filter and this ideal low-pass filter in the time domain is defined as:

$$h[n] = \mathrm{sinc}\left(2 \times \frac{f_c}{f_s} \times n\right)$$

Where,
$f_c$ is the cut-off frequency,
$f_s$ is the sampling frequency

The above is implemented in MATLAB code using the below syntax:

```
cutoff = 5;
N = 101;
n = -(N-1)/2:(N-1)/2;
h = sinc(2*(cutoff/fs)*n);
```

- The cutoff variable specifies the maximum frequency to pass,
- N = 101 ensures an odd-length filter to preserve symmetry and avoid phase distortion,
- sinc(2*(cutoff/fs)*n) generates the ideal low-pass filter coefficients
- n = -(N-1)/2:(N-1)/2 creates a symmetric index vector n , centered at zero with N elements, is generated. This is essential for constructing a linear-phase FIR filter, such as a sinc-based low-pass filter.

To minimize side lobes (ringing artifacts), a Hann window is applied using the below syntax:

```
h = h .* hann(N)';
h = h / sum(h);
```

- hann(N) is a window function that smoothly tapers the filter's edges.
- Normalization ensures the filter has unity gain, preserving the signal's energy after filtering.

**Filtering the Signal**

The designed filter is applied to the noisy signal using the convolution function's syntax in MATLAB for convolving 2 signals :

```
sinc_filtered_signal = conv(noisy_signal, h, 'same');
```

- noisy_signal is the sum of clean seismic signal and Gaussian noise as defined above
- h is the sinc lowpass filter defined above
- same return a result that's the same length as the input signal

**Machine Learning-Based Denoising**

- To achieve denoising beyond the capabilities of linear filtering, a supervised learning model was employed to approximate the clean signal more effectively. Sliding Window Feature Extraction is used that A fixed-size window (window_size = 11) was applied across the noisy signal

- For each centered window, the neighboring noisy samples were utilized as input features ( X ), while the corresponding center value from the clean signal served as the target output (y). This establishes a supervised learning dataset designed for regression tasks. The below syntax was used :

```
for i = half_win + 1 : length(noisy_signal) - half_win
window = noisy_signal(i - half_win : i + half_win);
X = [X; window];
y = [y; clean_signal(i)];
end
```

- The regression model was trained using MATLAB's fitrlinear, which implements a regularized linear regression model suitable for fast training.

- The choice of a linear model helps avoid overfitting and reduces computational complexity. The below syntax is used to train a supervised machine learning regression model in MATLAB using linear regression with regularization.

```
ml_model = fitrlinear(X, y);
```

fitrlinear is a built-in MATLAB function from the Statistics and Machine Learning Toolbox. It stands for "Fit Regression Linear, X is the input matrix, where each row is a feature vector (a windowed segment of the noisy seismic signal, y is the corresponding target vector, containing the clean signal's center value for each window and The output variable, `ml_model`, represents a trained regression model designed to transform input noisy signal segments into their corresponding denoised target values.

- Once the machine learning regression model (ml_model) is trained using local signal segments as input features and corresponding clean amplitudes as targets, it is employed to denoise the entire signal in a pointwise fashion. The following MATLAB loop performs this reconstruction using the below syntax:

```
for i = half_win + 1 : length(noisy_signal) - half_win
window = noisy_signal(i - half_win : i + half_win);
ml_denoised_signal(i) = predict(ml_model, window);
end
```

In each iteration, a segment of noisy signal values around the current index 'i' is taken as input for prediction. This segment serves as the input feature for the machine learning model (`ml_model`), which predicts the denoised value for the central point based on the surrounding context. The predicted value is then assigned to the output signal, `ml_denoised_signal`, at the corresponding position 'i' . This sliding-window approach leverages the local temporal structure of the signal, allowing for more context-aware noise reduction. It is particularly effective in handling non-stationary or time-varying noise and signal patterns.

**Residual Error Analysis**

- A residual error analysis is conducted to quantitatively assess the performance of classical sinc filtering and machine learning-based denoising. This process calculates the difference between the clean seismic signal and its respective denoised outputs. In the below syntax, The residual signals, residual_sinc and residual_ml, reflect the noise or distortion remaining in the signal after applying each denoising approach

```
residual_sinc = clean_signal - sinc_filtered_signal;
residual_ml = clean_signal - ml_denoised_signal;
```

- residual_sinc represents the discrepancy between the clean signal and the output of the sinc filter. It highlights potential loss or distortion of valuable information caused by the limitations of traditional filtering methods, such as phase distortion or frequency leakage and residual_ml signifies the error between the clean signal and the machine learning-denoised signal, indicating the model's ability to accurately reconstruct the original waveform.

## III. RESULTS

This section assesses the denoising effectiveness of the Sinc filter compared to the proposed Machine Learning (ML)-based model using synthetic seismic data. The evaluation is visually supported by three detailed figures:

- Clean Seismic Signal versus Filtered Outputs,
- Signal Processing Pipeline illustrating the progression from Noisy → Filtered → Denoised signals
- Residual Error Analysis.

**Clean Seismic Signal vs Sinc vs ML Output**

As shown below, the Figure 1 compares the clean seismic signal with the results from the traditional Sinc filter and the ML-based denoising model. The black curve represents the ground truth (original clean signal), the blue dashed line depicts the output from the Sinc filter, and the green solid line shows the ML-denoised signal, highlighting the differences and improvements achieved by each method.

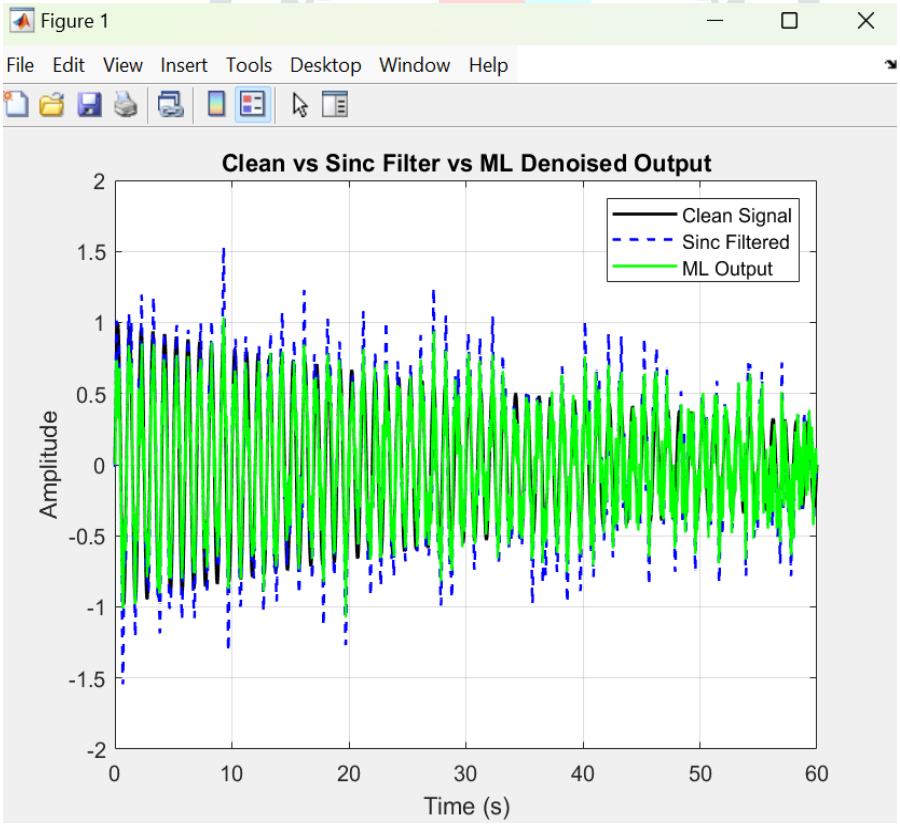


Fig.1: Clean signal vs. Sinc filter and ML denoised output. ML output closely follows the clean signal, outperforming the Sinc filter in noise reduction.

The ML output exhibits much closer alignment with the clean signal compared to the Sinc-filtered result. It effectively retains both low and high-frequency components throughout the 60-second window, with particularly notable performance in the 10–40 second range where amplitude modulations are complex and fast-changing. In contrast, the Sinc filter introduces visible distortions and excessive smoothing, especially in areas with high-frequency content, reflecting the limitations of its inherent low-pass design and resulting in the loss of important signal characteristics.

**Denoising Pipeline: Noisy → Sinc → ML**

The below figure.2 presents a three-stage subplot progression that visualizes the denoising pipeline. The top panel shows the noisy seismic signal, the middle panel displays the Sinc filter output, and the bottom panel illustrates the ML model output.

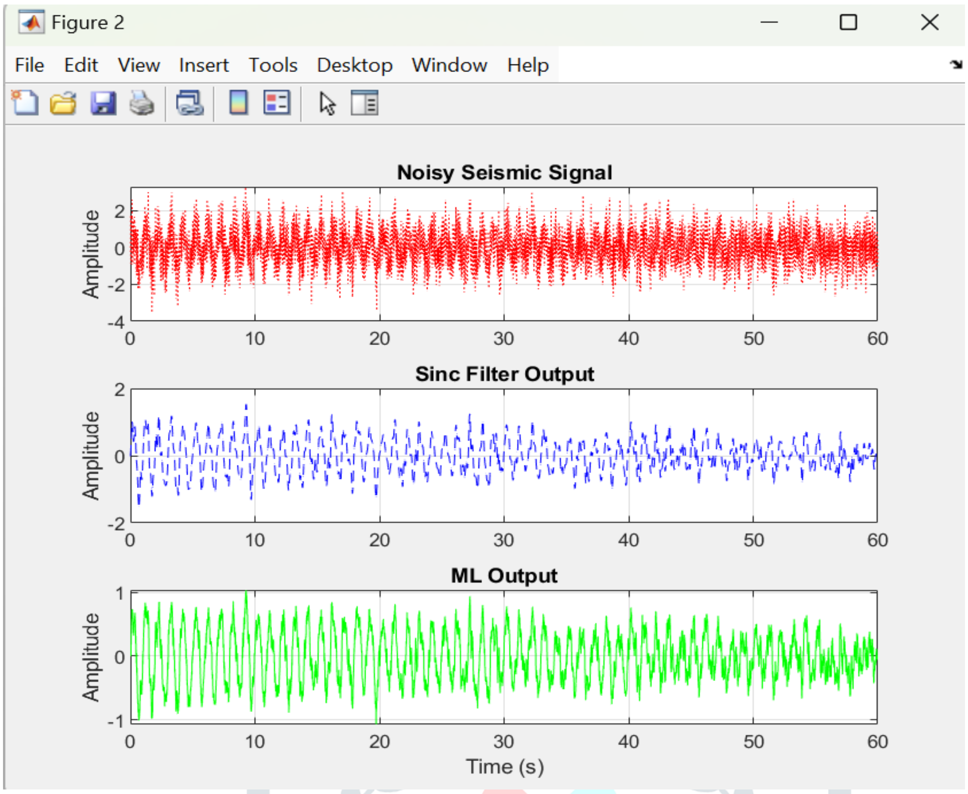


Fig.2: Noisy seismic signal (top), Sinc filter output (middle), and ML-denoised output (bottom), showing improved noise suppression and signal preservation by the ML model.

The top subplot depicts the noisy seismic input, displayed in red, and clearly affected by high-amplitude, high-frequency noise components. The middle subplot illustrates the Sinc-filtered output, where much of the noise is suppressed but at the expense of excessive smoothing and signal distortion. This distortion is particularly noticeable beyond 30 seconds, where the signal's oscillatory pattern and amplitude range are significantly diminished. The bottom subplot showcases the ML denoising model's output, which exhibits a distinct advantage. The denoised signal aligns closely with the original clean signal, preserving both its amplitude dynamics and frequency content.

Unlike the Sinc filter, the ML model employs non-linear noise suppression while adaptively maintaining signal fidelity. Subtle waveform details are retained in the ML output, suggesting a higher signal-to-noise ratio (SNR) and minimal distortion.

**Residual Error Analysis: Clean – Output**

The below Figure 3 provides a quantitative evaluation of the filtered signals' deviations from the clean ground truth through residual error plots. The Sinc filter residual (blue dashed) exhibits greater peak-to-peak variability across the timeline, signifying persistent inaccuracies throughout the signal. In contrast, the ML residual (green) displays a much tighter amplitude range centered around zero.

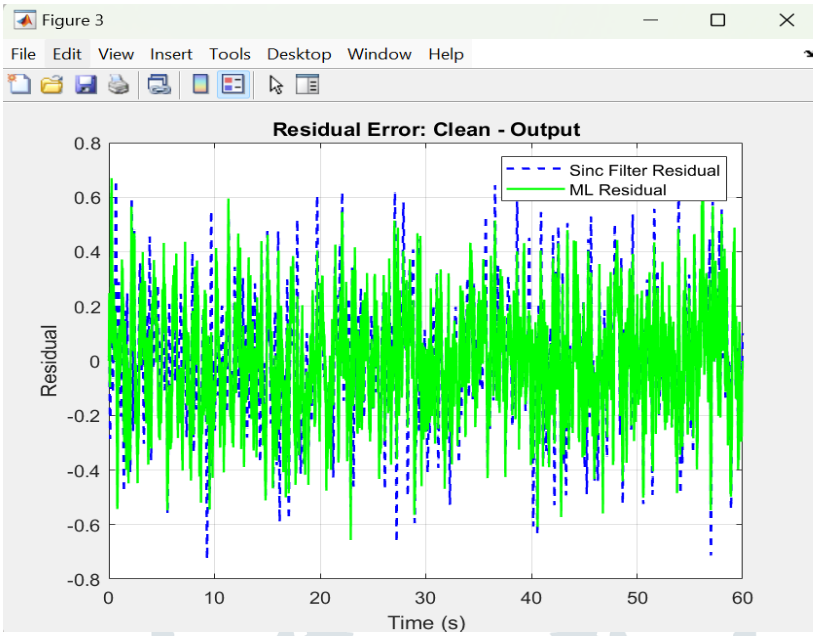


Fig.3 : Residual errors between the clean signal and outputs from the Sinc filter (blue dashed) and ML model (green), illustrating lower error magnitude for the ML approach.

Over the 60-second window, the ML residual predominantly remains within ±0.4, while the Sinc residual shows deviations up to ±0.7, particularly in areas where the clean signal experiences abrupt transitions. This suggests that the Sinc filter may underfit or excessively smooth dynamic regions. Additionally, the ML residual demonstrates a more stochastic and Gaussian-like distribution—a favorable characteristic—indicating that the ML model effectively retains structured signal content and minimizes deterministic noise.

## IV. CONCLUSION & FUTURE SCOPE

This research introduced an innovative hybrid method for denoising seismic signals, combining the strengths of a traditional sinc FIR low-pass filter with a machine learning-driven LinearFit model. Through simulations conducted on synthetic seismic data contaminated by Gaussian noise, it was shown that the sinc low pass filter efficiently mitigates high-frequency noise. Additionally, the LinearFit model significantly minimizes any remaining residual errors, while effectively preserving critical signal features and characteristics. This dual approach strikes an optimal balance between interpretability and robust performance. The proposed methodology holds great promise as a practical, adaptable, and efficient tool, contributing to advancements in seismic signal processing and enhancing data reliability for further analysis.

The proposed hybrid approach not only demonstrates strong potential but also paves the way for various future research directions entailing Machine Leaning and Neural Networks for denoising and filtering. A potential avenue for advancement lies in incorporating deep learning models like convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. These models have demonstrated remarkable proficiency in addressing intricate seismic patterns and hold the potential to greatly improve denoising effectiveness [5] [6]. Integrating the learning capabilities of machine learning with the flexibility of Kalman filters offers the potential to create a denoising system that not only comprehends signal patterns but also adapts dynamically to varying noise environments. This would result in a more resilient and efficient solution, especially for field applications where environmental noise tends to be unpredictable and inconsistent. Thus, by integrating machine learning with adaptive filtering methods would hold the promise of creating a noise suppression system that is both responsive and dynamic. This approach would be highly effective in managing non-stationary noise conditions [7].

**REFERENCES:**

**[1]:**Bormann, P. (2013): History, Aim and Scope of the 1st and 2nd Edition of the IASPEI New Manual of Seismological Observatory Practice. - In: Bormann, P. (Ed.), New Manual of Seismological Observatory Practice 2 (NMSOP-2), Potsdam : Deutsches GeoForschungsZentrum GFZ, 1-32. https://doi.org/10.2312/GFZ.NMSOP-2_ch1

**[2]:** G. Nolet, Imaging the Earth's Interior, J. F. Claerbout, Blackwell Scientific Publications, Oxford, 1985 414pp., £42.00 and Fundamentals of Geophysical Data Processing, 2nd edn J. F. Claerbout, Blackwell Scientific Publications, Oxford, 1985 274 pp., £42.00, Geophysical Journal International, Volume 86, Issue 1, July 1986, Pages 217–219, https://doi.org/10.1111/j.1365-246X.1986.tb01086.x

**[3] :** Smith, Steven W. The scientist and engineer's guide to digital signal processing. 1997.

[4]: Wang, D., Chen, G., Chen, J. et al. Seismic Data Denoising Using a Self-Supervised Deep Learning Network. Math Geosci 56, 487–510 (2024). https://doi.org/10.1007/s11004-023-10089-3

[5]: Weiqiang, Zhu & Mousavi, S.Mostafa & Beroza, Gregory. (2019). Seismic Signal Denoising and Decomposition Using Deep Neural Networks. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-13. 10.1109/TGRS.2019.2926772.

[6]: Mousavi, S.M., Ellsworth, W.L., Zhu, W. et al. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat Commun 11, 3952 (2020). https://doi.org/10.1038/s41467-020-17591-w

[7]: S. K. Aruna, N. Deepa and D. T, "Residual Learning for Enhanced Seismic Signal Quality with Denoising CNN and Kalman Filtering," 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2024, pp. 339-343, doi: 10.1109/I4C62240.2024.10748449.