Scilab case study project on Scilab

Wavelet Transform Based Preprocessing and Analysis of Noisy Raman Spectra Using Scilab

Chirag Veerwani

Indian Institute of Technology Hyderabad (IITH), Telangana, India

Signal Processing / Analytical Chemistry

April 2026

Abstract

This case study implements the wavelet transform based preprocessing framework for noisy Raman spectra classification as proposed by Pan et al. (IEEE Access, 2020) using Scilab. Real Raman spectral data of five minerals: Actinolite, Albite, Forsterite, Grossular, and Marialite are obtained from the publicly available RRUFF database and used as clean reference signals. Two noise conditions are simulated as described in the paper: Additive White Gaussian Noise (AWGN) at SNR (signal to noise ratio) levels ranging from 10 dB to 30 dB, and Baseline Background Noise (BBN) formulated as a summation of sinusoidal functions. A mixed noise condition combining both is also generated. The continuous Wavelet Transform (CWT) using the morel wavelet is applied to each noisy spectrum to generate a 2D scale map, visually demonstrating how noise and spectral peaks are represented differently in the wavelet domain. Results are presented as comparative plots of spectra under all noise conditions along with their corresponding scale maps.

1. Introduction

Raman spectroscopy is a technique widely used in analytical chemistry, geology, and materials science to identify and characterize substances. When laser light is shone on a material, some of the light gets scattered with a shift in frequency. This shift is unique to each material and acts like a fingerprint. By analysing this scattered light, one can identify which material is present.

However, in real laboratory environments, Raman spectra are never perfectly clean. Various sources of interference such as fluorescence from the sample, background radiation, and electronic noise from the detector get added to the measured spectrum. This makes it harder to see the true peaks and identify the material correctly.

Signal processing techniques are therefore needed to clean up these noisy spectra before any analysis is done. Among these, the Wavelet Transform has proven to be especially useful because it can analyse a signal at multiple scales simultaneously by capturing both sharp peaks and broad baseline drifts at the same time.

This case study implements the wavelet transform based preprocessing methodology from the IEEE Access paper by Pan et al. (2020), which proposes using the Continuous Wavelet Transform (CWT) to convert noisy 1D Raman spectra into 2D scale maps. These scale maps preserve all signal information while making noise and peaks visually distinguishable. The implementation is carried out entirely in Scilab using real mineral spectral data from the publicly available RRUFF database.

2. Problem Statement

Raman spectra collected in practice are contaminated by two main types of noise. The first is Additive White Gaussian Noise (AWGN), which is random jitter spread uniformly across the entire spectrum. This comes from electronic noise in the detector and environmental interference. The second is Baseline Background Noise (BBN), which is a smooth, slowly varying signal that shifts the entire spectrum up or down which is caused mainly by fluorescence of the sample under laser excitation.

When these noise types are present, especially at low signal-to-noise ratios, smaller Raman peaks become buried in the noise and the spectrum becomes difficult to interpret. Traditional methods like Fourier Transform filtering can remove some noise, but they treat the entire spectrum the same way and often damage the peaks in the process.

The problem this case study addresses is: how can we represent a noisy Raman spectrum in a way that clearly separates real spectral peaks from noise? The solution proposed in the reference paper and which is implemented here is to apply the Continuous Wavelet Transform using the Morlet wavelet, which converts the 1D noisy spectrum into a 2D scale map. In this scale map, real spectral peaks appear as distinct bright regions while noise appears as scattered or diffuse patterns, making the two easier to distinguish.

The specific objectives of this case study are: (1) to simulate realistic noise conditions on real Raman spectra, (2) to apply the CWT and generate scale maps for all noise conditions, and (3) to quantify the noise levels using Signal-to-Noise Ratio (SNR) calculations.

3. Basic Concepts Related to the Topic

3.1 Raman Spectroscopy

Raman spectroscopy works on the principle of inelastic scattering of light. When a laser is directed at a sample, most photons scatter elastically (Rayleigh scattering) with no change in energy. However, a small fraction of photons exchange energy with the molecular vibrations of the sample and scatter with a shifted frequency. This frequency shift, called the Raman shift, is measured in wavenumbers (cm^{-1}) and is unique to each molecule. A plot of scattered intensity versus Raman shift is called a Raman spectrum, and its peaks correspond to specific chemical bonds present in the sample.

3.2 Noise in Raman Spectra

Two types of noise are considered in this case study, following the paper by Pan et al. (2020):

Additive White Gaussian Noise (AWGN) is random noise that follows a Gaussian (bell curve) distribution and is added uniformly across the spectrum. It is characterized by the Signal-to-Noise Ratio (SNR), defined as:

$$SNR = 10 \times \log_{10} (Ps \,/\, Pn) \qquad (1)$$

where Ps is the signal power and Pn is the noise power. A higher SNR means the signal is much stronger than the noise. In the paper, SNR values range from 10 dB (noisy) to 30 dB (relatively clean).

Baseline Background Noise (BBN) is a smooth, slowly varying distortion that shifts the entire spectrum. It is caused mainly by fluorescence of the sample. In the paper, BBN is simulated as a summation of sinusoidal functions:

$$BBN(x) = A_1 \sin(f_1 x + \varphi_1) + A_2 \sin(f_2 x + \varphi_2) + A_3 \sin(f_3 x + \varphi_3) \qquad (2)$$

where A, f, and φ are the amplitude, frequency, and phase of each component respectively. Different combinations of these parameters simulate different fluorescence backgrounds.

3.3 Wavelet Transform

The Fourier Transform is a standard tool for analysing signals in the frequency domain. However, it has a major limitation because it tells you which frequencies are present in a signal, but not where in the signal they appear. For Raman spectra, which are nonstationary signals with peaks at specific wavenumber locations, this is a problem.

The Wavelet Transform overcomes this by using basis functions that are localized in both position and scale (frequency). The Continuous Wavelet Transform (CWT) of a signal x(t) is defined as:

$$X(a, b) = (1 / \sqrt{a}) \int \psi((t - b) / a) \times x(t)\, dt \qquad (3)$$

where ψ is the mother wavelet function, a is the scale parameter (controls zoom level), and b is the translation parameter (controls position). A small value of a corresponds to a fine scale that detects sharp, narrow features, while a large value of a corresponds to a coarse scale that detects broad features like baseline drift.

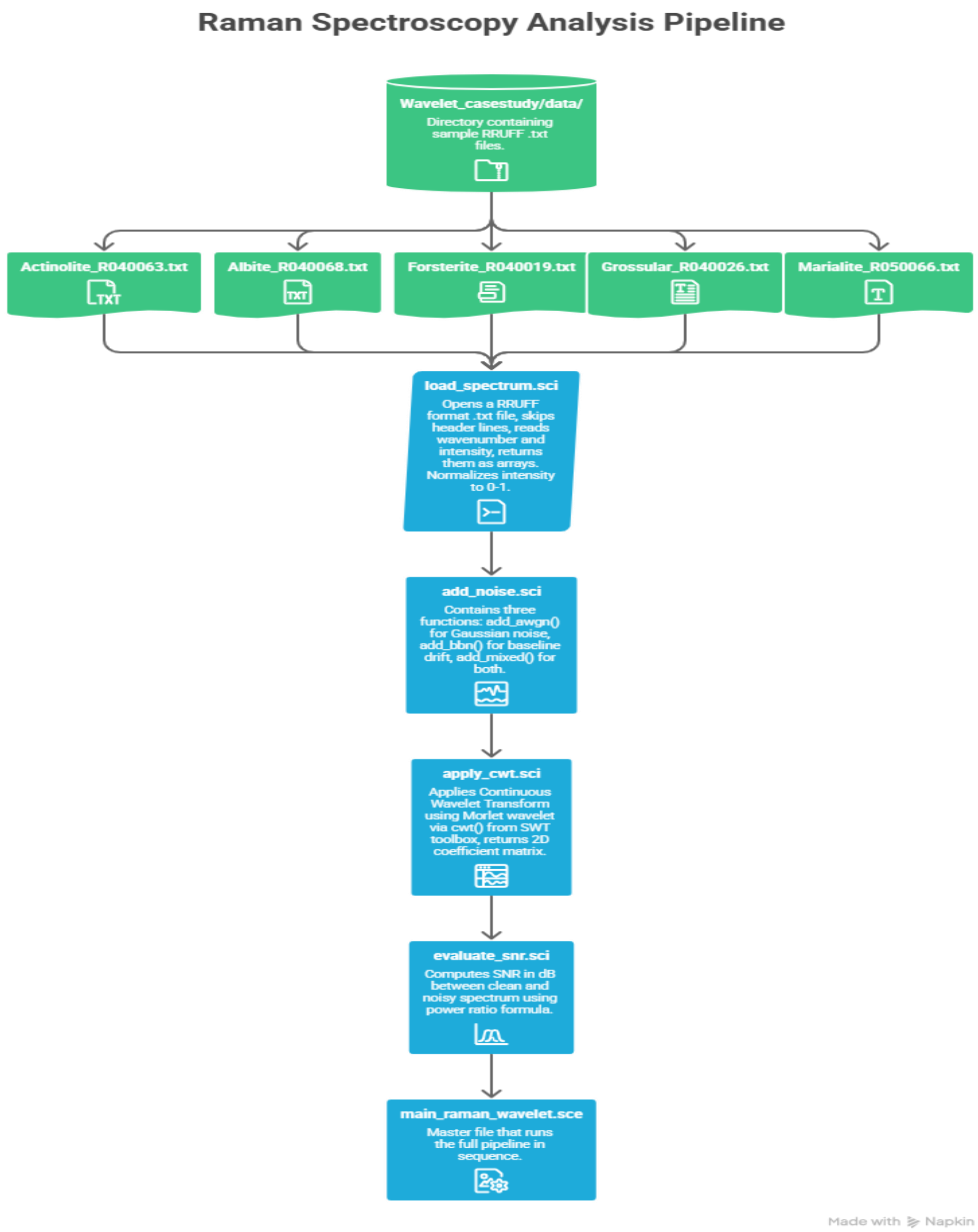
3.4 Morlet Wavelet

The paper uses the Morlet wavelet as the mother wavelet ψ. The Morlet wavelet is a sinusoidal wave modulated by a Gaussian envelope, which makes it naturally suited for detecting oscillatory features like spectral peaks. Its shape closely resembles the shape of Raman peaks, making it a good match for this application.

After applying the CWT at multiple scale levels, the output is a 2D matrix called the scalogram or scale map. The rows represent different scale levels and the columns represent position (wavenumber). Each cell in this matrix tells how strongly the signal matches the wavelet at that scale and position. When this matrix is plotted as a colour image, bright regions correspond to strong spectral features and dark regions correspond to low activity.

4. Flowchart

The flowchart below shows the step-by-step flow of the program from loading the spectrum to generating the final results.



Flowchart of the Wavelet Transform Based Raman Spectrum Preprocessing Pipeline

5. Software / Hardware Used

Operating System: Windows 11 (64-bit)

Scilab Version: 2026.0.1

Toolbox: Scilab Wavelet Toolbox (SWT) version 0.3.3, installed via ATOMS package manager. This toolbox provides the cwt() and cwtplot() functions used for applying the Continuous Wavelet Transform and visualizing the scale maps.

Data Source: RRUFF Project database (https://rruff.info) — a publicly available collection of high quality Raman spectra of well-characterized minerals. Spectra were downloaded as .txt files in the unoriented format.

Hardware: Standard laptop. No specialized hardware is required. All computation is done in software.

6. Procedure of Execution

The following steps describe how to set up and run the code from scratch.

Step 1 — Install Scilab
Download and install Scilab version 2026.0.1 from the official website (www.scilab.org). Ensure the 64-bit version is installed.

Step 2 — Install the Wavelet Toolbox
Open Scilab and run the following command in the console to install the SWT toolbox:

```
atomsInstall('swt')
```

After installation, the cwt() and cwtplot() functions become available automatically.

Step 3 — Download Spectral Data
Go to https://rruff.info and download the Raman spectra for the five minerals used in this study: Actinolite, Albite, Forsterite, Grossular, and Marialite. The spectra are available as .txt files under the unoriented Raman spectra collection. Place all downloaded files inside a folder named data/ within the project directory.

Step 4 — Set Up the Project Folder
Create a project folder with the following structure:

```
Wavelet_casestudy/
    data/
        Actinolite_R040063.txt
        Albite_R040068.txt
        Forsterite_R040019.txt
        Grossular_R040026.txt
        Marialite_R050066.txt
    load_spectrum.sci
    add_noise.sci
    apply_cwt.sci
    evaluate_snr.sci
    main_raman_wavelet.sce
```

Step 5 — Set the Working Directory

Open Scilab and set the working directory to the project folder using:

cd('C:\path\to\project_folder')

Step 6 — Select the Mineral to Analyse

Open main_raman_wavelet.sce in a text editor. Near the top of the file, find the line that loads the spectrum:

[wn, intens] = load_spectrum('data/Actinolite_R040063.txt');

To run the analysis for a different mineral, replace the filename with the corresponding file from the data/ folder. For example, to analyze Albite:

[wn, intens] = load_spectrum('data/Albite_R040068.txt');

Save the file and proceed to Step 7. Repeat this for each of the five minerals.

Step 7 — Run the Main File

Run the main file by typing the following in the Scilab console:

```
exec('main_raman_wavelet.sce')
```

The main file automatically loads all dependency .sci files in the correct order, loads the spectrum, generates all noise conditions, applies the CWT, plots the results, and prints the SNR table to the console.

Description of Each File

load_spectrum.sci — Opens a RRUFF format .txt file, skips all header lines beginning with ##, reads the wavenumber and intensity columns, and returns them as arrays. The intensity is normalized to a 0-1 range.

add_noise.sci — Contains three functions. add_awgn() adds Gaussian noise at a specified SNR level. add_bbn() adds a simulated baseline drift constructed from a sum of three sine waves. add_mixed() applies both noise types together.

apply_cwt.sci — Takes a spectrum and a set of scale values, applies the Continuous Wavelet Transform using the Morlet wavelet via the cwt() function from the SWT toolbox, and returns the 2D coefficient matrix.

evaluate_snr.sci — Computes the SNR in dB between a clean and a noisy version of a spectrum using the standard power ratio formula.

main_raman_wavelet.sce — The master file that runs the full pipeline in sequence.

7. Results

This section presents the results obtained from running the code on the Actinolite spectrum from the RRUFF database. Similar results were obtained for the other four minerals.

7.1 Note on Scale Map X-axis

An important point to note when reading the scale maps is that the x-axis shows data point indices (1 to 1077), not actual wavenumber values. Since the Actinolite spectrum spans from approximately 139 cm^{-1} to 1500 cm^{-1} across 1077 data points, the dominant peak at 670 cm^{-1} corresponds to approximately data point index 400. This is why the brightest region in the scale maps appears at $x \approx 400$, which correctly represents the 670 cm^{-1} peak.

7.2 Clean and Noisy Spectra

The figure below shows the clean Actinolite spectrum alongside its noisy versions under all three noise conditions. The AWGN plots at 30 dB, 20 dB, and 10 dB clearly show increasing levels of random jitter. The BBN plot shows the spectrum riding on a smooth sinusoidal baseline. The mixed noise plot combines both effects.

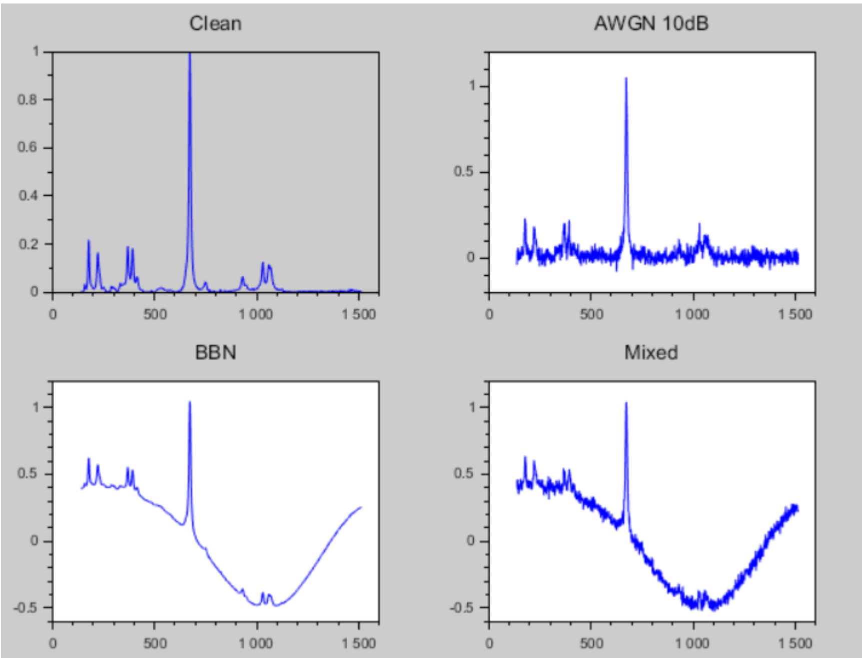


Figure 1: Clean Actinolite spectrum and noisy versions: AWGN at 10 dB, BBN only, and Mixed noise

It can be observed that at 30 dB SNR, the Raman peaks are still clearly identifiable. At 20 dB, smaller peaks start to get obscured by noise. At 10 dB, even the medium-sized peaks become difficult to identify, with only the dominant peak around 670 cm^{-1} remaining clearly visible.

7.3 AWGN at Different SNR Levels

The figure below shows the effect of AWGN at three different SNR levels side by side, making the progression of noise easier to compare.

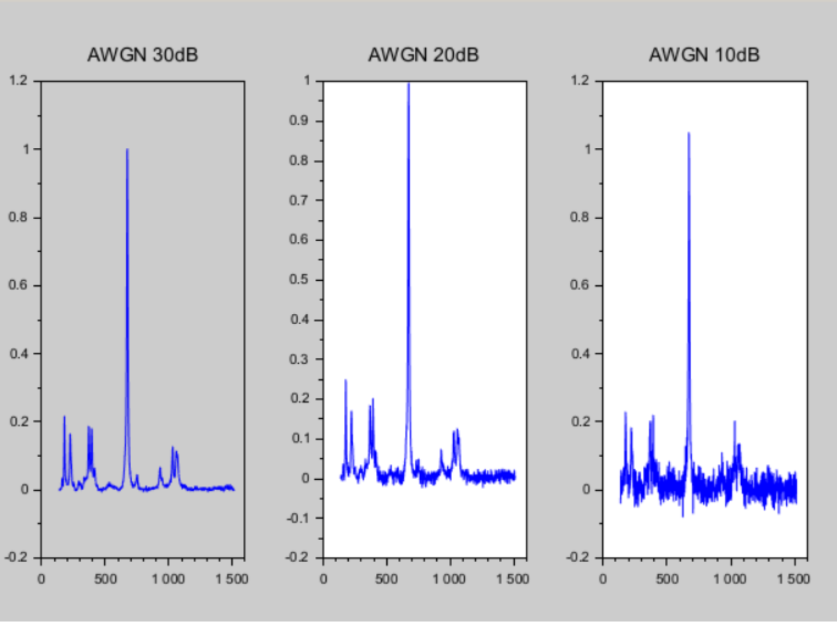


Figure 2: Actinolite spectrum with AWGN at 30 dB, 20 dB, and 10 dB SNR

7.4 CWT Scale Maps

The figure below shows the 2D CWT scale maps generated using the Morlet wavelet for the clean spectrum and the three noise conditions. Each scale map has wavenumber on the x-axis and scale level on the y-axis, with colour representing coefficient magnitude (red/yellow = high, blue = low).

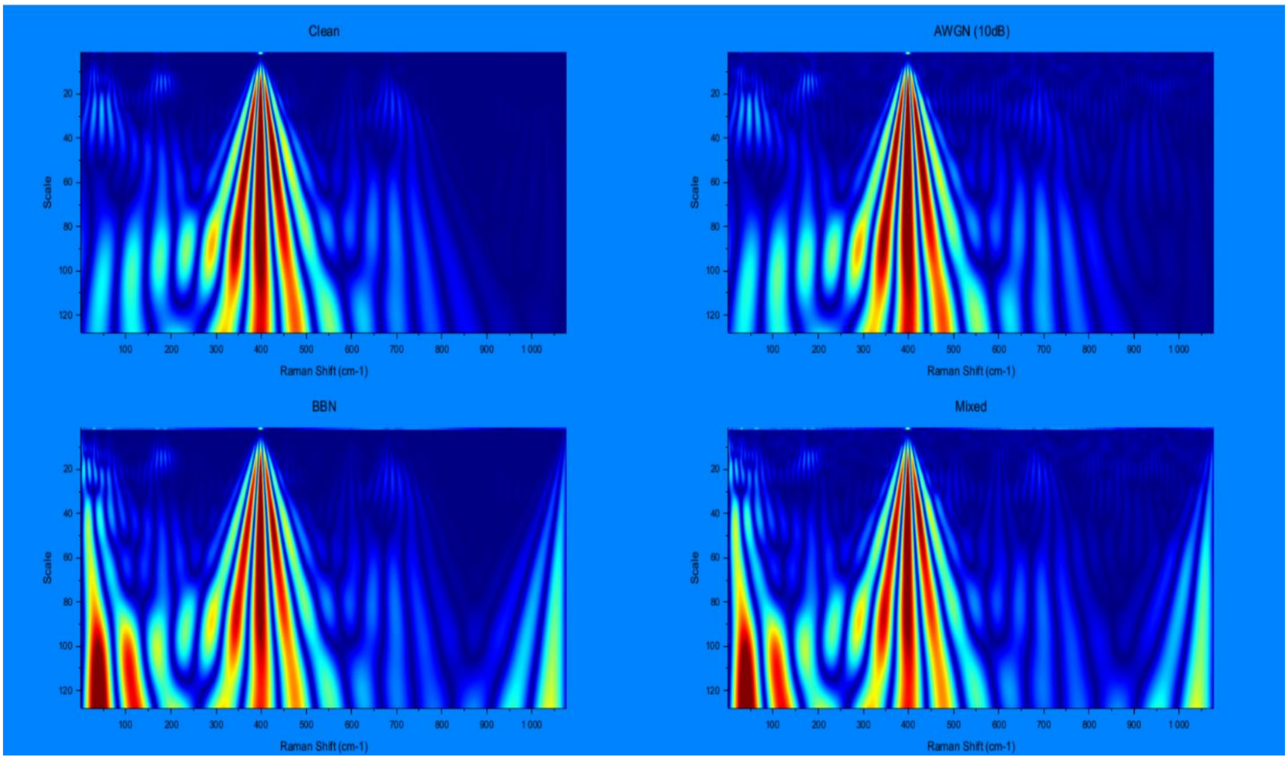


Figure 3: CWT Scale Maps — Clean, AWGN (10dB), BBN, and Mixed noise conditions

In the clean spectrum scale map, the dominant peak of Actinolite near 670 cm^{-1} appears as a prominent bright region. Smaller peaks at lower wavenumbers also appear as distinct blobs at fine scales. The background is predominantly dark blue, indicating low activity in those regions.

In the AWGN scale map, the background is no longer uniformly dark. Scattered bright speckles appear across all scales and positions, representing the random noise. However, the dominant peak blob is still visible, showing that the CWT preserves signal structure even under noise.

In the BBN scale map, the broad sinusoidal baseline causes a diffuse, large-scale brightening across the map particularly at higher scale values, which correspond to low frequency components. This is consistent with the paper's observation that baseline noise manifests at large scales in the wavelet domain.

The mixed noise scale map combines both effects: scattered speckles from AWGN and broad large-scale brightening from BBN.

7.5 SNR Computation Results

The table below shows the computed SNR values for each noise condition applied to the Actinolite spectrum.

Noise Condition	**Target SNR (dB)**	**Computed SNR(dB)**
AWGN only	10 dB	10.05 dB
AWGN only	20 dB	19.90 dB
AWGN only	30 dB	29.89 dB
BBN only	—	-11.45 dB
Mixed (AWGN 10 dB + BBN)	—	-11.49 dB

Table 1: Computed vs Target SNR for each noise condition (Actinolite spectrum)

```
--> disp('=== SNR Results (dB) ===');

  "=== SNR Results (dB) ==="

--> disp('AWGN 10dB: '); disp(snr_awgn_10);

  "AWGN 10dB: "

   10.052013

--> disp('AWGN 20dB: '); disp(snr_awgn_20);

  "AWGN 20dB: "

   19.902074

--> disp('AWGN 30dB: '); disp(snr_awgn_30);

  "AWGN 30dB: "

   29.887877

--> disp('BBN:       '); disp(snr_bbn);

  "BBN:       "

  -11.447801

--> disp('Mixed 10dB:'); disp(snr_mixed_10);

  "Mixed 10dB:"

  -11.493825
```

Scilab console: Calculating the values for the above table

The AWGN results closely match the target SNR values for example, when noise is added targeting 10 dB, the computed SNR is 10.05 dB. This confirms that the noise simulation is mathematically accurate. The BBN and mixed conditions yield negative SNR values (around -11.4 to -11.5 dB), meaning the noise power exceeds the signal power. This is expected fluorescence background in real lab measurements can easily overpower the Raman signal, which is why preprocessing methods like the wavelet transform are needed.

8. References

[1] L. Pan, P. Pipitsunthonsan, C. Daengngam, S. Channumsin, S. Sreesawet, and M. Chongcheawchamnan, "Method for Classifying a Noisy Raman Spectrum Based on a Wavelet Transform and a Deep Neural Network," IEEE Access, vol. 8, pp. 202716–202727, 2020. DOI: 10.1109/ACCESS.2020.3035884

[2] ScilabWavelet Toolbox (SWT), ATOMS, Scilab. Available: https://atoms.scilab.org/toolboxes/swt