CLE2D ver. 1.1

Full User Manual

CLE2D: A MATLAB GUI for 2D Curvilinear Lineament Extraction

CLE2D harnesses the power of advanced, unsupervised source separation techniques, including principal component analysis (for variance maximization), continuous wavelet transforms, and lineament extraction with Bayesian optimization.

This program was developed and released by:

Bahman Abbassi

Postdoctoral Fellow in Mining Engineering

Institut de Recherche en Mines et en Environnement (IRME)

Université du Québec en Abitibi-Témiscamingue (UQAT)

Adresse: IC113, 445, boul. de l’Université

Rouyn-Noranda (Québec) J9X 5E4

Emails : bahman.abbassi@uqat.ca; bahmanabbassi@gmail.com

Telephone: +1 (819) 762 0961 ext. 4311

Cell Phone: +1 (819) 290 1800

July 2024

Contents

1. Introduction 3
2. System requirements 5
3. Program interface 7
4.1. Spectral Feature Extraction 12
4.2. Principal Component Analysis (PCA) 14
4.3. Lineament extraction with Wavelet-ICA and Bayesian Hyperparameter Optimization 15
4.3.1. Hysteresis Thresholding Algorithm 16
4.3.2. Bayesian Hyperparameter Optimization 17
5. Input/output formats 19
6. Spectral feature extraction in CLE2D 20
7. Lineaments Extraction in CLE2D 27
7.1. Fast Lineaments Extraction with Raw Data Sets 27
7.2. Lineaments Extraction with Spectral Feature Extraction with PCA-DOG wavelets 34
7.3. Lineaments extraction by DOG Wavelet-PCA feature extraction with Bayesian hyperparameter optimization 39
8. References 48

1. Introduction

CLE2D is a free and independent tool specifically designed for 2D curvilinear lineament extraction. This program integrates a combination of sophisticated algorithms and techniques to provide a comprehensive solution for geoscientific data analysis. At its core, CLE2D utilizes Principal Component Analysis (PCA), Continuous Wavelet Transform (CWT), and the Hysteresis Thresholding Algorithm (HTA) for spectral feature extraction. Additionally, it employs Bayesian hyperparameter optimization to enhance the accuracy and efficiency of geological curvilinear fault extraction.

CLE2D is engineered to enable spectral source separation across multiple layers of geoscientific datasets. The program features an intuitive graphical user interface (GUI) that facilitates the visualization of both extracted spectral features and underlying curvilinear patterns. This user-friendly interface is designed to ensure that users can easily manipulate and analyze geoscientific data with precision and efficiency.

Key features of CLE2D include:

- Spectral Feature Extraction: Utilizing CWT to decompose geoscientific images into multiple scales and directions, enabling the capture of sharp image changes and detailed space-frequency information.
- Dimensionality Reduction: Employing PCA to reduce the dimensionality of high-dimensional images, retaining the most geologically pertinent information while eliminating redundant spectral data.
- Bayesian Hyperparameter Optimization (BHO): Systematically optimizing hyperparameters to enhance the extraction process, balancing precision and recall, maximizing the Fβ Score.
- Hysteresis Thresholding Algorithm (HTA): Detecting linear patterns by computing slopes and aspects of spectral features, enhancing image clarity, and mapping potential lineaments.

CLE2D is a versatile tool suitable for a range of applications in geoscientific research and exploration. It is particularly useful in the fields of:

- Mineral Exploration: Assisting in the identification and analysis of geological faults and lineaments that are critical in mineral deposit studies.
- Geophysical Data Analysis: Providing advanced methods for the interpretation of geophysical data, facilitating the detection of potential field sources.
- Geological Mapping: Enhancing the accuracy of geological maps by integrating spectral feature extraction and lineament analysis.

The program is available for free and can be easily used by anyone with MATLAB installed on their computers. In summary, CLE2D is a robust and comprehensive tool for 2D curvilinear lineament extraction, combining advanced algorithms with an easy-to-use interface to support geoscientific research and exploration. Whether you are engaged in mineral exploration, geophysical data analysis, or geological mapping, CLE2D provides the tools necessary to enhance your work with precision and efficiency.

2. System requirements

This program is designed to run on any Windows-based personal computer with at least 8 GB of random-access memory (RAM). Increasing the RAM size allows larger images to be processed at once. Since large matrices are operating in this program, the read/write speed of the storage is also essential. A solid-state drive (SSD) with a non-volatile memory express interface (NVMe) is recommended.

CLE2D is provided in MATLAB M-File format. M-Files require MATLAB 2024a version 24.1 (and later) to run. To use the program, locate the M-Files in the current folder of MATLAB and then type CLE2D in the MATLAB Command Window. This interface offers comprehensive tools for curvilinear lineament extraction with a user-friendly design, ensuring that users can easily manipulate and visualize geoscientific data with precision and efficiency.

CLE2D is designed to run efficiently on Windows-based personal computers. To ensure optimal performance and to handle the computational demands of the program, the following system requirements are recommended:

1. **Operating System**:
    - Windows-based PC.
2. **Memory (RAM)**:
    - A minimum of 8 GB of random-access memory (RAM) is required. However, increasing the RAM size is highly recommended as it allows for the processing of larger images simultaneously and enhances overall performance. Larger datasets and complex computations will benefit significantly from additional memory.
3. **Storage**:
    - Since CLE2D handles large matrices and extensive data operations, the read/write speed of the storage is crucial. A solid-state drive (SSD) with a non-volatile memory express (NVMe) interface is recommended. This configuration ensures faster data access and smoother performance during intensive computational tasks.

4. **MATLAB Requirements**:
    - CLE2D is provided in MATLAB M-File format, requiring MATLAB 2024a version 24.1 or later. Ensure that this version of MATLAB is installed on your computer to utilize CLE2D effectively. The M-Files should be located in the current folder of MATLAB, and users can start the program by typing CLE2D in the MATLAB Command Window.
5. **User Interface**:
    - CLE2D features a comprehensive and user-friendly graphical interface, which offers a wide array of tools for curvilinear lineament extraction. This design ensures that users can easily manipulate and visualize geoscientific data with precision and efficiency.

Additional Recommendations

- **Processor**: While not explicitly stated, a multi-core processor with high clock speeds will further enhance the performance, especially for more complex and computationally intensive tasks.
- **Graphics**: For advanced visualization, having a dedicated graphics card may improve the rendering speed and quality of graphical outputs.
- **Backup and Storage Management**: Given the potentially large size of data files, regular backups and efficient storage management practices are recommended to prevent data loss and ensure smooth operation.

3. Program interface

When running the CLE2D program, the user interface appears as illustrated in Figure 1. The interface comprises several key functionalities, summarized below:

- **Coordinates**
  - Max Lat/Min Lat: Set the maximum and minimum latitude.
  - Max Lon/Min Lon: Set the maximum and minimum longitude.
  - Method: Choose the method for coordinate input (Rectangular coordinates).

Coordinates can be read from a prepared text file with the following format:

First line:	Min Longitude (Min Lon)
Second line:	Max Longitude (Max Lon)
Third line:	Min Latitude (Min Lat)
Fourth line:	Max Latitude (Max Lat)

For example, in the case above:

-77.8

-77.2

53.15

53.45

- **Spacing**
  - Spacing: Define the spacing value in arcseconds for both longitude and latitude. In Quebec, each arcsecond of latitude is approximately 33 meters, and each arcsecond of longitude is about 17 meters. The program automatically adjusts these ratios for any location on Earth to ensure optimal image scaling.
  - Data Filter: Apply a filter to the input data to smooth it further.

- **Input Point Data**
  - 2D Interpolation: Select the interpolation method for the data points, which can be regularly spaced or irregularly sampled. The symbol '#' indicates the number of input data points.
  - Xn / Yn: Retrieve the number of pixels in the X and Y coordinates after interpolation.
- **Digitized Lineaments**
  - Targets Type: Choose the type of targets (digitized faults) either as point data in .csv format (fault densities) or as images of the digitized lines.
  - Spacing: Define the spacing for digitized lineaments, which should be equal to or less than the main spacing for input data.
  - Cut-off 1: Automatically set the cut-off value to confine a buffer zone for the lineaments.
  - Cut-off 2: Set a second cut-off value to narrow down the automatically generated buffer zone, which is useful if the target fault density map is too thick for upcoming optimizations.
  - Target Filter: Apply a filter to the target data.
- **Spectral Feature Extraction**

This window is divided into two main sections. First, a 2D CWT (Continuous Wavelet Transform) breaks down the desired inputs into raw spectral features. Users can adjust parameters such as the number of scales (na), scale dilation, and the angles at which the mother wavelet will operate. Each mother wavelet structure then automatically computes the scales vector, corresponding frequencies, and angles vector. Isotropic and anisotropic mother wavelets are available depending on the study's objectives. After the CWT process, the program automatically determines the number of CWT features. Second, S-PCA (Spectral Principal Component Analysis) is provided for spectral source separation. Users must decide if dimensionality reduction is necessary.

  - CWT Inputs: Choose to merge the uploaded point datasets for CWT and S-PCA.

- OrderX (n) / OrderY (m): Define the order of X and Y for the Difference of Gaussian (DOG) mother wavelet.
- Change order by scales: Option to change the orders automatically by scales, setting the values inversely proportional to the scales to detect higher frequency features.
- # of Scales (na): Set the number of wavelet scales, with Bayesian optimization available to fine-tune this parameter.
- Scale Dilation: Define the scale dilation factor, allowing access to longer scales while keeping the number of scales constant. For example, with na = 4 and Scale Dilation = 1, scales = {1, 2, 3, 4}; with Scale Dilation = 2, scales = {1, 3, 5, 7}.
- WSFR: Set the Wavelet Smoothness Filter Ratio, which is necessary to avoid interpolation artifacts in the wavelet coefficient features.
- CWT # of Angles: Set the number of directions for which the mother wavelet can surf in 2D space.
- Scales (a): Define the scale vector automatically or manually.
- CWT Angles: Define the angle vector automatically or manually.
- β: Set the angle of symmetry for CWT, indicating the wavelet's rotational invariance or symmetry.
- CWT: Run the continuous wavelet transform.
- S-PCA: Run the spectral principal component analysis after CWT.
- # CWT Fs: Set the number of CWT features.
- DR to: Specify the dimension reduction target, with Bayesian optimization available to fine-tune this parameter.

➢ **Select Features**

- Point Data: Use point data for lineament extraction.
- CWT: Use CWT results for lineament extraction.

- S-PCA: Use results of spectral principal component analysis for lineament extraction.

➢ **Lineament Extraction**

- Line Res: Define the resolution of the lineament extraction output. Larger resolutions result in crisper and smoother results but at a higher computational cost.
- SF # of Angles: Define the number of angles for calculating Aspect in the hysteresis thresholding procedure.
- Bayesian Opt MaxIter: Set the maximum iterations for Bayesian optimization.
- w: Define the width of the step filter, typically 10 percent of the largest pixel numbers, with Bayesian optimization available to fine-tune this parameter.
- VSFW: Variability of the Step Filtering Widths, set by default to 0.25, to determine the variability of w in correlation with the complexity of the extracted spectral features. Bayesian optimization helps fine-tune this parameter.
- Lineaments: Extract lineaments.
- AutoLine: Automatically fine-tune the hyperparameters (na, WSFR, DR, w, and VSFW).

➢ **Plot**

- **Plot Results**: Option to plot the results.
- **Filter**: Apply a filter to the plotted results.
- **Close All**: Close all plots.
- **Clear All**: Clear all settings.

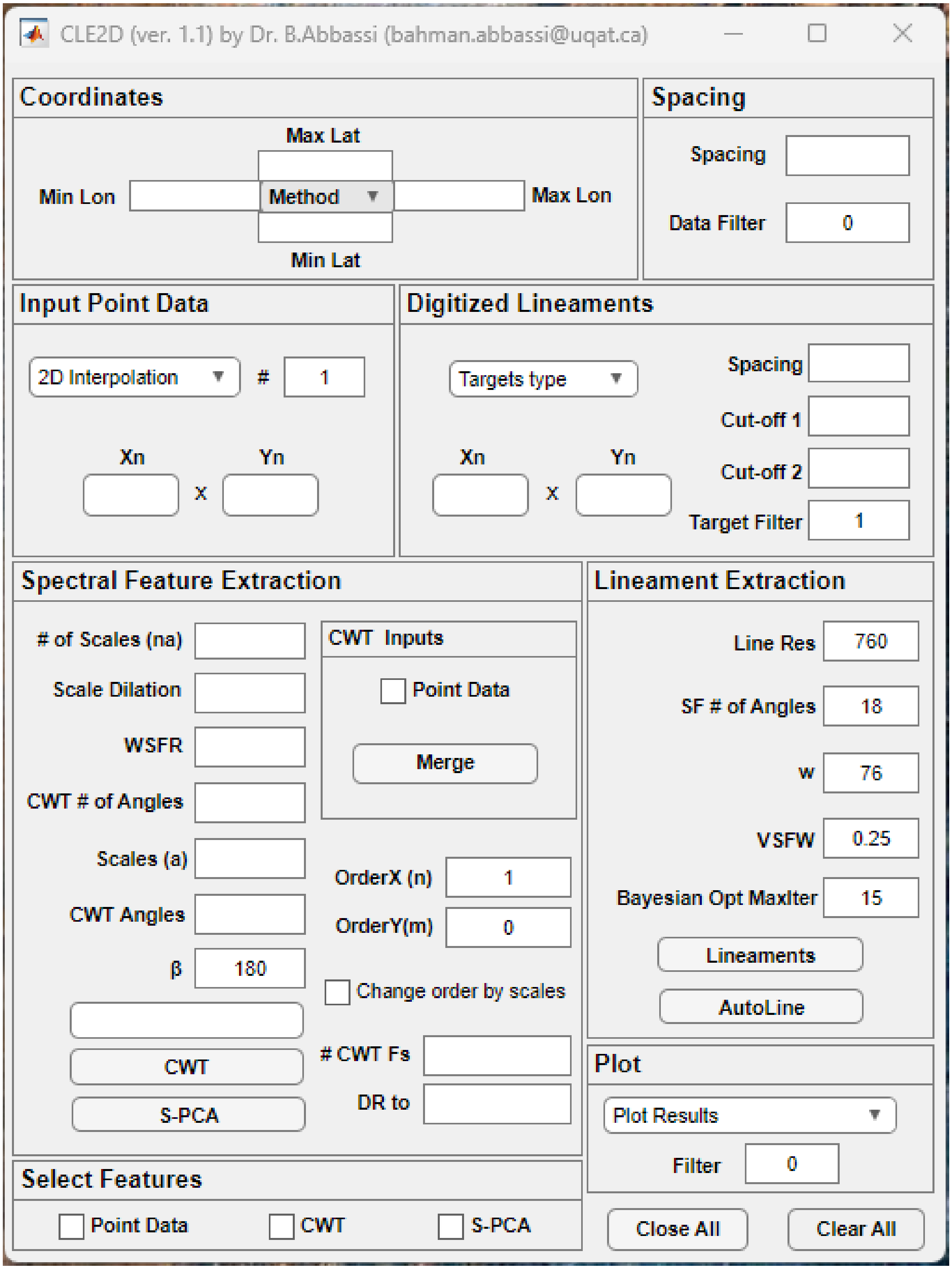

Figure 1: The CLE2D Program Interface showcasing the primary functionalities and layout for user interaction.

4. Theory

4.1. Spectral Feature Extraction

Spectral feature extraction in the frequency domain is achieved through the spectral decomposition of geo-images. Unlike the Fourier transform, which struggles with abrupt changes, wavelet transforms are localized in space and have finite durations, making them ideal for capturing sharp image changes (Antoine et al., 2004; YAWTB Team, 2024).

The Continuous Wavelet Transform (CWT) decomposes a signal/image into scaled and translated versions of a chosen wavelet (mother wavelet). First applied in geophysical data analysis in 1997, CWT is pivotal for locating potential field sources by identifying maxima lines derived from CWT coefficients.

A 2D wavelet is a natural or complex-valued oscillatory function adhering to the admissibility condition within the real plane. The Fourier transform of any pertinent wavelet function must be square-integrable and possess a zero-mean value. This ensures finite energy in the spatial and frequency domains, favoring spatial and frequency localization, multi-resolution analysis, and efficient delineation of low and high-frequency signals.

Given a 2D image *I*, the CWT coefficients can be expressed as:

$$C_I\left(a,\vec{b},\theta\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)\psi^*\left(\frac{x-b_x}{a}-\frac{y-b_y}{a},\theta\right)dxdy \qquad (1)$$

a is the scale factor, b is the translation vector, θ is the rotational angle, and ψ^* complex conjugate of the mother wavelet. The wavelet coefficients measure the similarity between the image and the scaled, translated, and rotated mother wavelet.

The CLE2D program integrates various wavelet algorithms, including:

1. Difference-of-Gaussian (Symmetric)
2. Complex Gaussian Z (Asymmetric/Symmetric)
3. Complex Gaussian X (Asymmetric/Symmetric)
4. Derivatives of Gaussian (Asymmetric)

5. Mexican Hat (Symmetric)
6. Morlet (Asymmetric)
7. Isotropic Morlet (Symmetric)
8. Cauchy (Asymmetric)
9. ES-Cauchy (Asymmetric)
10. Paul (Symmetric)
11. Wheel (Symmetric)
12. Pethat (Symmetric)
13. End-Stop (Asymmetric)
14. Gabor Mexican Hat (Asymmetric)
15. Sinc (Symmetric)
16. Fan (Symmetric)

Spectral decomposition produces high-dimensional images with patterns linked to specific scales and directions, making some decomposed spectra redundant. Dimensionality reduction is necessary to extract the most relevant features from high-dimensional data, aiding in automated pseudo-geological mapping, lineament analysis, and machine learning estimations.

The feature extraction scheme combines 2D CWT with variance, kurtosis, and negentropy maximization algorithms (PCA, kICA, and nICA) for spectral feature extraction (only PCA is present in this program). The algorithm consists of three main stages (Figure 2):

1. Preprocessing: 2D interpolation and filtering of raw datasets prepare the input images for feature extraction.
2. Spatial Feature Extraction: Employs PCA/ICA for spatial source separation and dimensionality reduction.
3. Spectral Feature Extraction: Consists of two substages:

    a. Continuous wavelet transform (CWT)

b. Spectral PCA/ICA (SPCA/SkICA/SnICA)

Effective spectral feature extraction depends on computer hardware specifications, particularly for large numbers of spectral features. Users can create more features by adjusting CWT parameters. For large numbers of CWT features, PCA/ICA methods help decorrelate and separate spectral features. Dimensionality reduction techniques can summarize features, retaining the most geologically pertinent information.

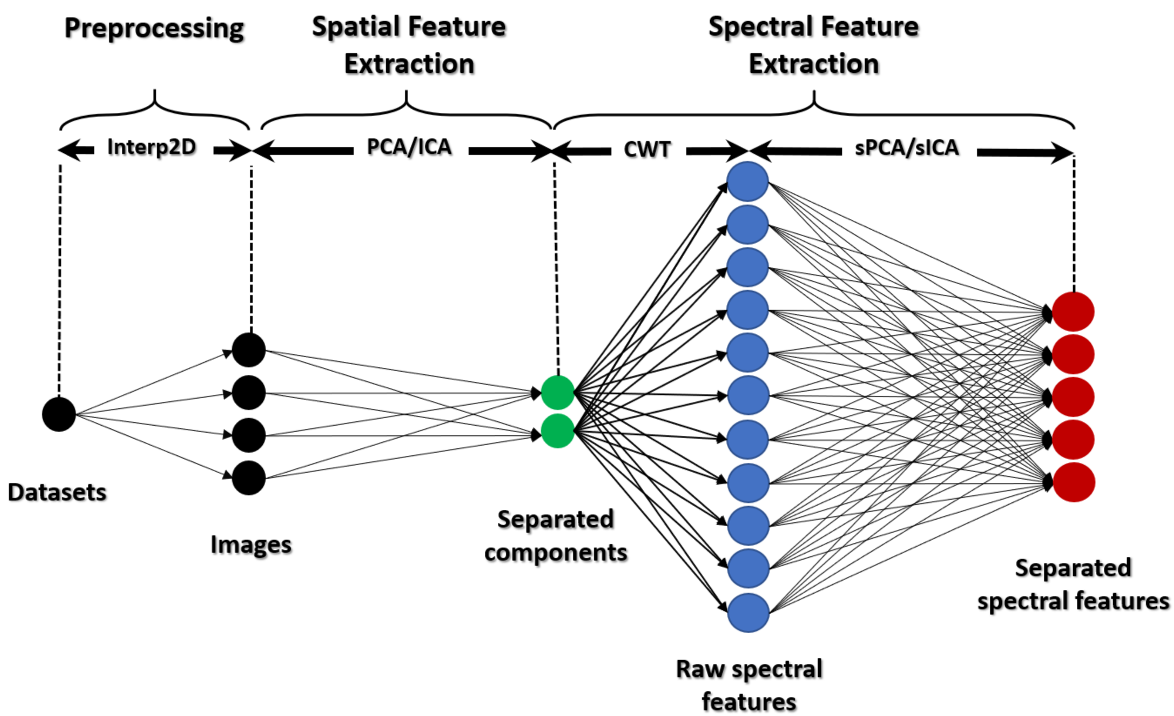

Figure 2. Schematic view of the CLE2D procedure.

4.2. Principal Component Analysis (PCA)

PCA is used to reduce the dimensionality of the data by transforming it into a set of linearly uncorrelated variables called principal components. This process involves several key steps (Hyvärinen & Oja, 2000; Moore, B., 2024):

1. Centering the Data:
   — Subtract the mean of each geo-image from the data to produce centered geo-images.

— Mathematically, if g is the original geo-image and g_m is the mean, the centered image is g - g_m .

2. Whitening the Data:
   - — Whitening transforms the centered data into unit variance and an identity covariance matrix.
   - — This is achieved through eigenvalue decomposition of the covariance matrix of the centered geo-images.
   - — The covariance matrix C is decomposed as $C= EDE^T$, where E is the matrix of eigenvectors and D is the diagonal matrix of eigenvalues.

3. Calculating the Transformation Matrix D:
   - — The transformation matrix D converts centered images into whitened images.
   - — The transformation is given by:

$$u = D(g - g_m) \tag{2}$$

   - — In this context, D can be expressed as:

$$D = ED^{-1/2}E^T \tag{3}$$

   where $D^{-1/2}$ is the diagonal matrix with the inverse square root of the eigenvalues.

4. Deriving Principal Components:
   - — The principal components are the transformed variables that maximize the variance of the data.
   - — These components are linearly uncorrelated and are ordered by the amount of variance they capture from the data.

4.3. Lineament extraction with Wavelet-ICA and Bayesian Hyperparameter Optimization

CLE2D ver. 1.1 introduces a comprehensive algorithm for extracting geological curvilinear lineaments from geophysical and geoscientific data sets. It combines several advanced methods, including Continuous Wavelet Transform (CWT), Principal Component Analysis (PCA), hysteresis thresholding algorithm (HTA) with pixel labeling, and Bayesian Hyperparameter Optimization (BHO).

CWT decomposes an image into scaled and translated versions of a wavelet, providing detailed space-frequency information. This increases dimensionality, which PCA addresses by retaining significant variance while reducing dimensionality. This facilitates the detection of linear patterns through hysteresis thresholding.

4.3.1. Hysteresis Thresholding Algorithm

The Hysteresis Thresholding algorithm detects linear patterns by computing slopes and aspects of spectral features, enhancing the image, and pixel labeling to map potential lineaments.

The Hysteresis Thresholding algorithm is as follows (Panagiotakis & Kokinou, 2015; Panagiotakis, 2024):

1. Preprocessing
   - Image Preparation: Normalize the color distribution and apply low-pass filtering to remove noise.
   - Wavelet Transform: Perform 2D CWT on the preprocessed images to decompose them into multiple scales and directions, increasing the dimensionality.
   - Dimensionality Reduction: Apply PCA to the wavelet-transformed images to reduce dimensionality while preserving significant variance.
2. Slope and Aspect Calculation
   - Compute the gradient of the principal component P at each point (x,y):

$$\nabla P = \left(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y} \right) \tag{4}$$

- Calculate the slope S and aspect A from the gradient:

$$S = \left\| \nabla P \right\|, \ A = \arctan\left(\frac{\partial P / \partial y}{\partial P / \partial x} \right) \tag{5}$$

   - Enhanced images L are created by identifying areas with high slopes, slopes of slopes, and slopes of aspects (Panagiotakis and Kokinou, 2014 and 2015):

$$L = \left(S^2 . S' . A' \right)^{1/4} \tag{6}$$

S' and A' are the "slope of the slope" and the "slope of the aspect," respectively.

3. Hysteresis Thresholding
   - Define Thresholds:
     - Low threshold T_l to identify weak lineament pixels.
     - High threshold T_h to identify strong lineament pixels.
   - Improve curvilinear patterns:
     - Convolves L with a zero mean step filter G with a width of w and an orientation angle φ:

$$I_g = L * G(w, \varphi) \tag{7}$$

     - Then the algorithm calculates the maximum of the corresponding pixel values of the images:

$$I_m = \max_{a,\varphi} I_g \tag{8}$$

   - Calculate k, the median value of the 9-pixel neighborhood of each pixel p in the enhanced image I_m.
   - Classify pixels:
     - Strong lineament pixel (C_1): If $I(p) > T_h$ and $I(p) > k$.
     - Weak lineament pixel (C_2): If $T_l < I(p) \leq T_h$ and $I(p) > k$.
     - Non-lineament pixel (C_3): If $I(p) \leq T_l$ or $I(p) \leq k$
   - Region Growing:

     Classify C_2 pixels as C_1 if connected to a C_1 pixel; otherwise, classify them as C_3.
   - Final Lineament Map:

     Combine the results to produce a binary image highlighting the potential faults.

4.3.2. Bayesian Hyperparameter Optimization

Bayesian Hyperparameter Optimization (BHO) is a crucial component in the lineament extraction process within the CLE2D program. It systematically optimizes hyperparameters to enhance the extraction of geological lineaments from geophysical data sets. BHO combines probabilistic modeling and iterative improvement to explore the hyperparameter space efficiently. The optimization considers hyperparameters related to wavelet smoothness, PCA dimensionality reduction, and hysteresis thresholding.

The primary goal is to maximize the F_β Score, which balances precision and recall in the extraction process.

The steps of the BHO algorithms are as follows (Archetti and Candelieri, 2019; MathWorks, 2024):

1. Define the Objective Function

   Formulate the objective function based on the Fβ Score
2. Initialize the Gaussian Process (GP) Model

   The Gaussian Process is a probabilistic model that estimates the objective function. Defined by a mean function $\mu(x)$ and a Matérn kernel function $k(x, x')$:

$$f_{obj}(x) \sim GP\left(\mu(x), k(x, x')\right) \tag{9}$$

3. Generate Initial Data Points

   Start with N randomly selected initial hyperparameter sets forming the initial dataset D_0.
4. Iterative Optimization Process

— Fit GP Model:

Fit the Gaussian Process to the initial dataset D_{t-1} to model the objective function.

— Optimize Acquisition Function:

Use the Expected Improvement (EI) acquisition function to balance exploration and exploitation. It guides the selection of the next hyperparameter set x_t:

$$EI(x) = \mathrm{E}\left[\max\left(f(x) - f(x^+), 0\right)\right] \tag{10}$$

where $f(x^+)$ is the current best-known objective value.

— Evaluate Objective Function:

Evaluate the objective function at x_t to obtain $f(x_t)$.

— Update dataset:

Update the dataset to include the new observation $D_t = D_{t-1} \cup \{(x_t, f(x_t))\}$

5. Convergence Check
   - Check if the optimization has converged based on predefined criteria (e.g., minimal improvement over several iterations).
   - If not converged, return to the iterative optimization step.
6. Optimal Hyperparameter Selection:

Once convergence is achieved, select the hyperparameter set with the highest F_β Score from the observations.

5. Input/output formats

CLE2D supports input data in point datasets formatted as .csv files. The data must follow an XYZ-style format where:

- X column: Represents the longitudes.
- Y column: Represents the latitudes.
- Z column: Represents the values of the geoscientific image, such as reflectance, magnetic field intensities, etc.

To use these inputs, users must first define the minimum and maximum values of the geographic coordinate system in decimal degrees for both latitude and longitude. Additionally, the spacing in arc seconds for both latitude and longitude needs to be specified. This setup ensures that the program accurately interprets the spatial extent and resolution of the input data.

Example format of a .csv file:

Table 1. An example of the input CSV file.

X (Longitude)	Y (Latitude)	Z (Geoscientific Value)
-77.8	53.15	45.0
-77.2	53.45	46.2

The number of pixels in the resulting image is automatically adjusted based on the defined project size, determined by the minimum and maximum coordinates and the specified spacing.

The outputs generated by CLE2D are primarily in MATLAB .fig format. This format includes:

- **Extracted Spectral Features**: The spectral features extracted from the input data using the 2D Continuous Wavelet Transform (CWT) and Spectral Principal Component Analysis (S-PCA).
- **Retrieved Curvilinear Lineaments**: The detected curvilinear lineaments were identified through the lineament extraction process.

These .fig files visually represent the processed data, enabling users to analyze and interpret the results directly within the MATLAB environment. The use of MATLAB's .fig format ensures compatibility with MATLAB's extensive visualization and analysis tools, facilitating further exploration and refinement of the results.

CLE2D offers a straightforward approach to handling geoscientific data with clear input and output formats, ensuring that users can efficiently manage their datasets and extract meaningful geological insights.

6. Spectral feature extraction in CLE2D

Spectral feature extraction in CLE2D involves applying the 2D Continuous Wavelet Transform (CWT) combined with source separation algorithms to decompose and extract frequency-dependent features. Dimensionality reduction can also be performed. Here is a step-by-step workflow for spectral feature extraction in CLE2D:

6.1. Read the Coordinates

Load the Rectangle Coordinates.txt file to define the project coordinates limits in the Data folder (Figure 3).

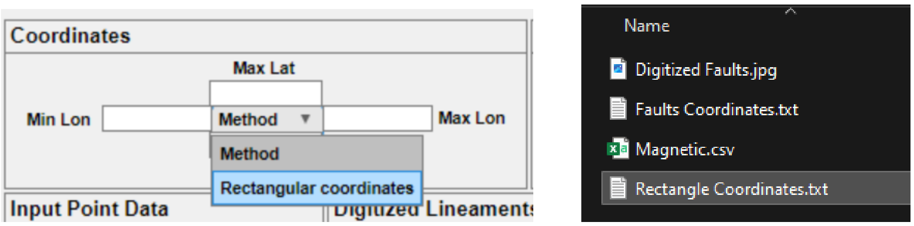

Figure 3: Loading the Rectangle Coordinates.txt file to define the project coordinates limits.

6.2. Set Spacing and Data Filter Values

Define the spacing and data filter values. If no smoothing is required, insert zero in the box. Any value greater than zero will smooth the data sets proportional to the magnitude of the data filter (Figure 4).

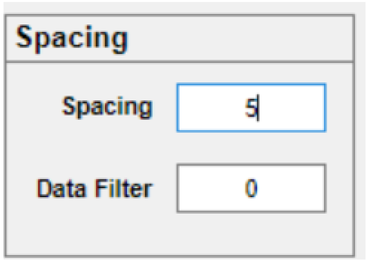

Figure 4: Setting the spacing and data filter values for spectral feature extraction.

6.3. Perform 2D Interpolation

Push the 2D Interpolation button and select the desired interpolation method. For additional data layers, insert the layer number and interpolate again. Repeat this process for all data layers (Figure 5).

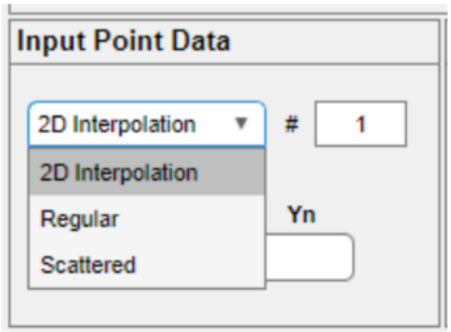

Figure 5: Performing 2D interpolation.

6.4. Configure Spectral Feature Extraction Parameters

In the Spectral Feature Extraction module, set the Number of Scales (na), Scale Dilation, and Number of Angles (Figure 6). For instance:

- Number of Scales = 3
- Scale Dilation = 2
- WSFR (Wavelet Smoothness Filter Ratio) = 1
- Number of Angles = 8

The WSFR box defines a damping factor for decomposed wavelet features, smoothing them as scales increase. Setting this parameter too low can produce unwanted interpolation artifacts, while very high values can eliminate valuable features. A useful rule of thumb is to set this parameter equal to the Scale Dilation.

This configuration means the mother wavelet expands three times at scales 1, 3, and 5 (with scale dilation being the interval between scales) in 8 directions (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°). The smoothness filter is set to 1 (WSFR = 1), indicating that the resulting wavelet features are smoothed proportionally to the scale. For example, at scale 1, the smoothness is 1; at scale 2, the smoothness is 2, and so on. This filtering helps eliminate unwanted artifacts from the interpolation of wavelet coefficients at larger scales, where the wavelet features are inherently smaller.

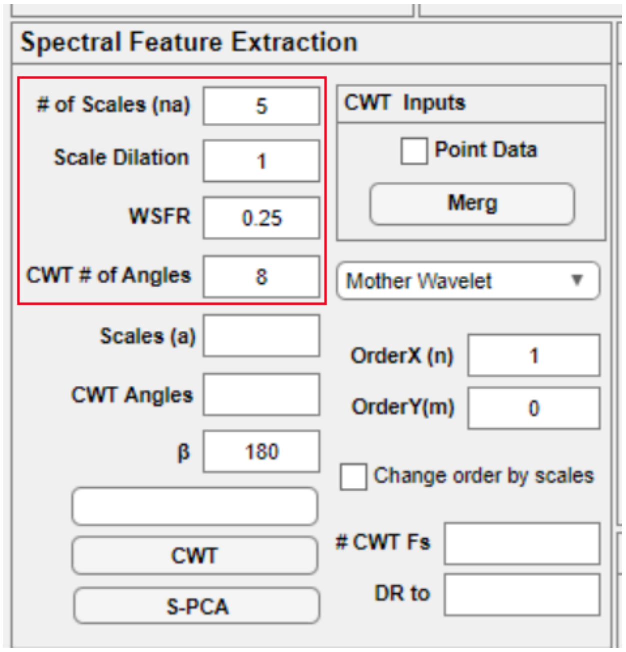

Figure 6: Configuring spectral feature extraction parameters in the Spectral Feature Extraction module.

6.5. Select Inputs for Spectral Feature Extraction

Select the desired point data for spectral feature extraction in the CWT Inputs box (Figure 7).

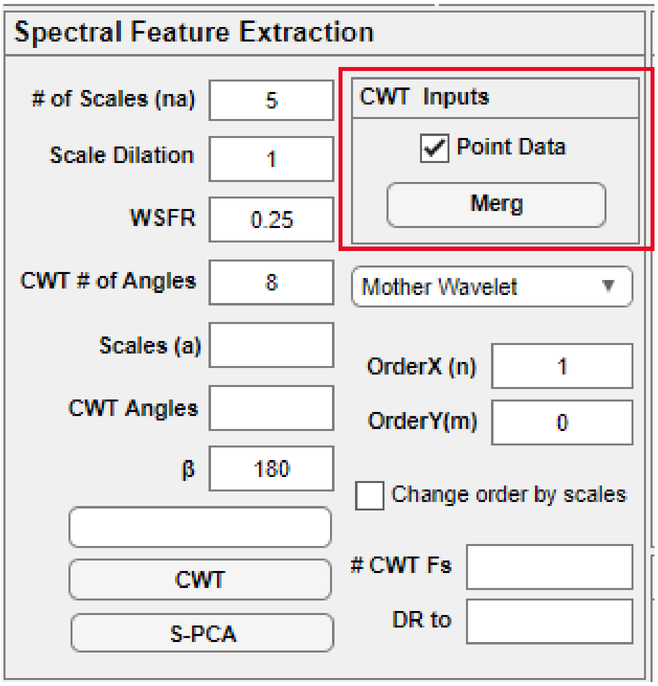

Figure 7: Selecting the desired point data for spectral feature extraction in the CWT Inputs box.

6.6. Define Orders for DOG Wavelet

For the Difference of Gaussian (DOG) wavelet, define the order of differentiations in X (n) and Y (m) directions. Optionally, choose to change the order by scales, allowing the algorithm to use higher orders of differentiations in lower scales to focus on finer edges (Figure 8).

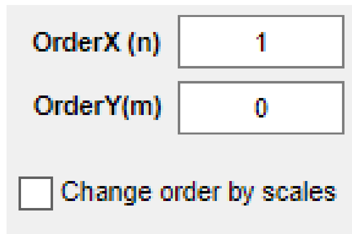

Figure 8: Defining orders for the Difference of Gaussian (DOG) wavelet.

6.7. Set the Rotational Symmetry Parameter (β)

The parameter β is related to the Order of Rotational Symmetry (ORS), which indicates how many times the wavelet can be rotated by a certain angle while maintaining its original shape. This parameter serves as an index of the wavelet's rotational invariance or symmetry. In most cases, the order of rotational symmetry for the wavelet is two, meaning the wavelet retains its shape when rotated within a symmetry range of zero to 180°. For specific cases, such as derivatives of Gaussian wavelets where m = n, the order of symmetry is four. In these instances, the wavelet only needs to be rotated within a symmetry range of zero to 90° (Figure 9).

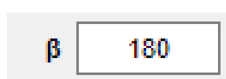

Figure 9: Setting the rotational symmetry parameter (β) for CWT.

6.8. Merge Scales and Angles

Push the Merge button to calculate the scales and angles for equal intervals, updating the Scales (a) and Angles boxes. Users can manually adjust these settings as needed (Figure 10).

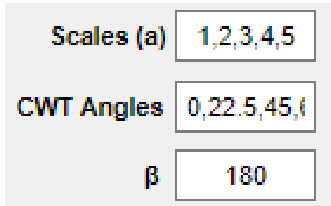

Figure 10: Merging scales and angles for equal intervals in spectral feature extraction.

6.9.Calculate Number of CWT Features

The number of CWT Features (# of CWT Fs) is automatically calculated based on the scales and angles set previously. Users can define how to reduce dimensionality in the "Dr to" box. By default, no dimensionality reduction is applied (Figure 11).

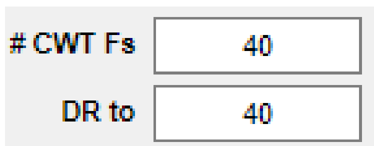

Figure 11: Calculating the number of CWT features based on the set scales and angles.

6.10. Run Continuous Wavelet Transform

Use the continuous wavelet transform algorithm by pressing the CWT button. After each run, a message confirms the successful deployment (CWT Completed). The algorithm also displays the mother wavelet in space and frequency domains for DOG wavelets (Figure 12).

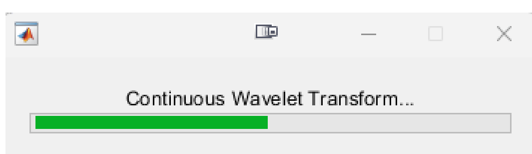

Figure 12: Running the continuous wavelet transform algorithm

After each run, a message appears that ensures the algorithm is successfully deployed (CWT Completed, Figure 13).

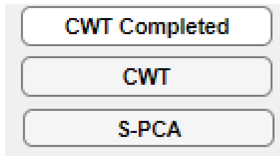

Figure 13: Confirmation message indicating the successful deployment of the CWT algorithm.

The algorithm also shows the mother wavelet in space and frequency domains for the DOG wavelet (Figure 14).

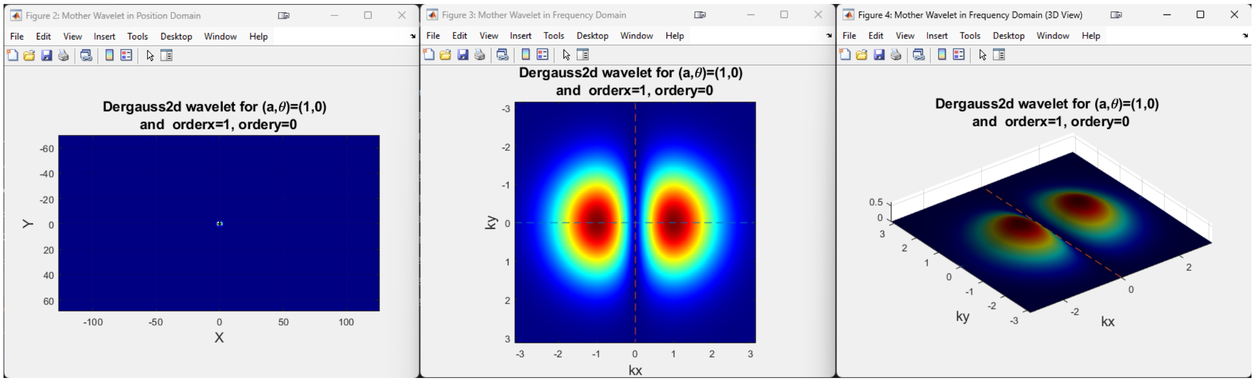

Figure 14: Displaying the mother wavelet in space and frequency domains for DOG wavelets.

6.11. Perform Dimensionality Reduction with PCA

The program employs PCA (Principal Component Analysis) for spectral feature extraction (S-PCA) to separate the decomposed wavelet features and reduce frequency-dependent overlaps. For dimensionality reduction, set the "DR to" parameter to a value less than 40. After each run, a confirmation message ("PCA Completed") will appear, indicating that the algorithm has been successfully deployed (Figure 15).

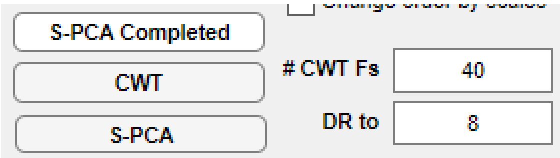

Figure 15: Using PCA for spectral feature extraction and confirming successful deployment.

6.12. Visualize Results

View the results in the Plot module under the Plot Results box. Display the CWT and S-PCA results to analyze the extracted spectral features (Figures 16 & 17).

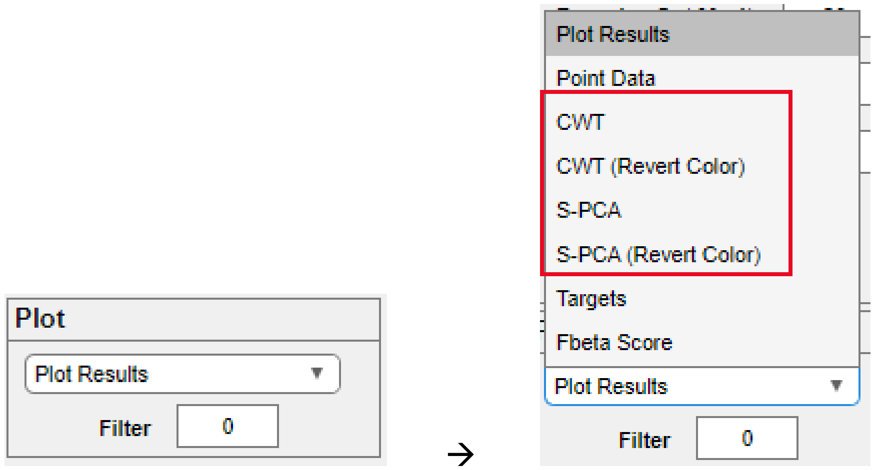

Figure 16: Displaying the results of the CWT in the Plot module.

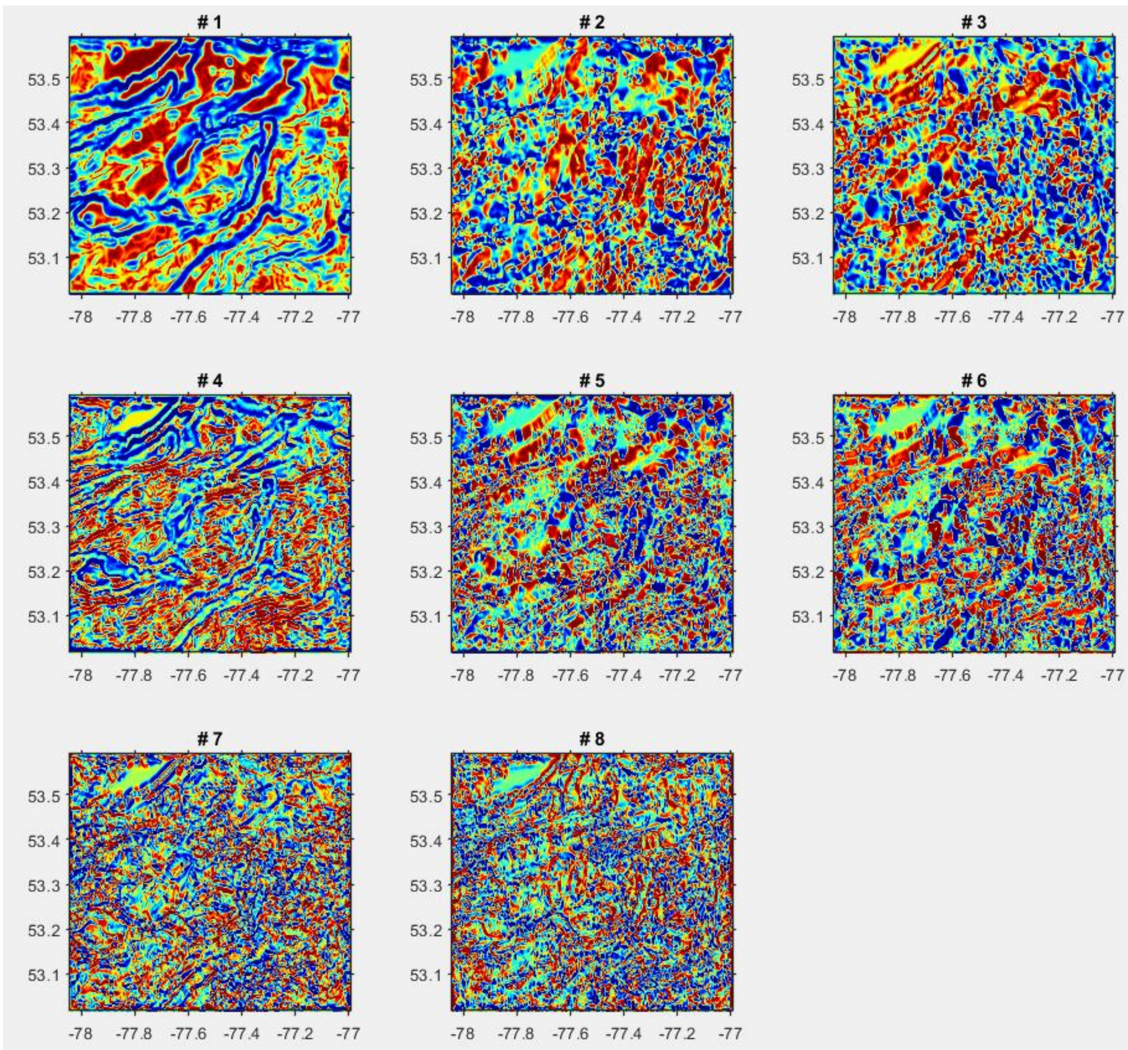

Figure 17: Displaying the results of the S-PCA in the Plot module.

7. Lineaments Extraction in CLE2D

In version 1.1 of CLE2D, the software is equipped with a novel algorithm for lineament extraction based on the extraction of curvilinear patterns using wavelet-PCA hysteresis thresholding and Bayesian hyperparameter optimization.

7.1. Fast Lineaments Extraction with Raw Data Sets

This method is the fastest way to perform lineament extraction, ideal for users who already estimate the parameters or when fast results are more desirable than highly accurate ones. The algorithm can process multiple input Point Data sets simultaneously. For simplicity, this guide demonstrates the process using a single data set of Total Magnetic Field Intensities from Western Quebec, Canada.

Here is the standard workflow for fast lineaments extraction in CLE2D:

7.1.1. Read the Coordinates

Load the Rectangle Coordinates.txt file to define the project coordinates limits in the Data folder (Figure 18).

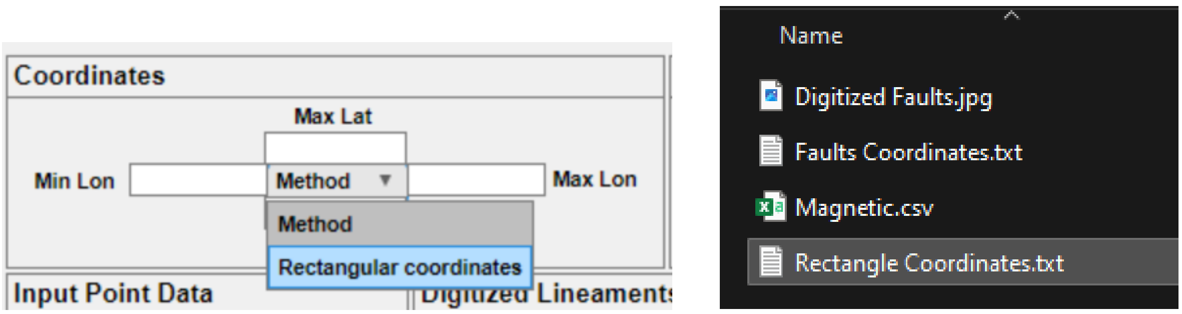

Figure 18: Loading the Rectangle Coordinates.txt file to define the project coordinates limits for lineament extraction.

7.1.2. Set Spacing and Data Filter Values

Define the spacing and data filter values. If no smoothing is needed, insert zero in the box. Any value greater than zero will smooth the data sets proportionally to the magnitude of the Data Filter (Figure 19).

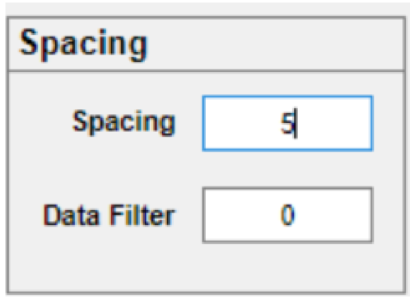

Figure 19: Defining the spacing and data filter values for lineament extraction.

7.1.3. Perform 2D Interpolation

Push the 2D Interpolation button and select the desired interpolation method. For the second layer of data, insert the number 2 and interpolate again. Repeat this for all data layers. In this example, only one layer of magnetic data sets is used (Figure 20).

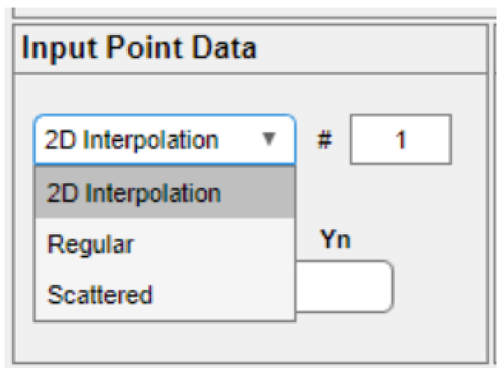

Figure 20: Performing 2D interpolation for lineament extraction.

7.1.4. Select Input Features

Select the Point Data checkbox to choose the input features for the upcoming feature selection (Figure 21).

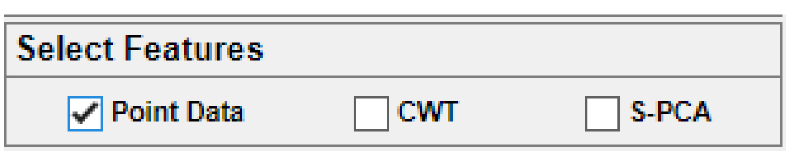

Figure 21: Selecting input features for lineament extraction.

7.1.5. Set Lineament Extraction Parameters

The next step is to set the lineament extraction parameters (Figure 22).:

- **Line Res**: Defines the resolution of the lineament extraction output in a number of pixels on the largest side of its surrounding rectangle. A larger "Line Res" results in crisper and smoother results but at a higher computational cost. Here, 320 pixels are used.
- **SF # of Angles**: Defines the number of angles for calculating Aspect in the hysteresis thresholding procedure. The default is 18, but users can adjust this number for better results.
- **Width of the Step Filter (w)**: This is typically set to 10 percent of the largest pixel numbers. For a Line Res of 320 pixels, w = 32 is used.
- **Step Filtering Widths (VSFW)**: Set by default to 0.25. VSFW determines the variability of w in correlation with the complexity of the extracted spectral features. A larger VSFW yields a more significant w for less complex features and a reduced w for more linear features. If VSFW = 0, the values of w for each lineament extraction iteration on wavelet-PCA features remain constant.
- **Bayesian Opt MaxIter**: Defines the maximum iterations for Bayesian hyperparameter optimization. The default is 15 iterations, but no optimization is performed in this case, focusing on hysteresis thresholding and pixel labeling directly on the input magnetic data.

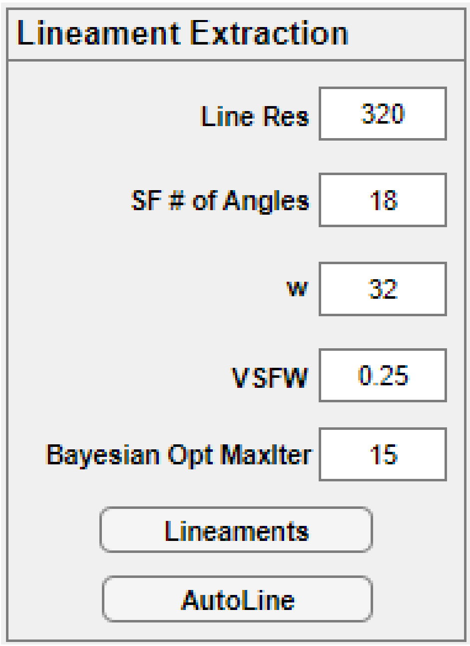

Figure 22: Setting lineament extraction parameters.

7.1.6. Upload Digitized Target Faults

Use the Digitized Lineaments box to upload digitized target faults. This step is necessary to calculate the beta score and estimate the performance of lineament extraction validated by experimental target data. Define the spacing for target interpolation, which is usually smaller or equal to the spacing of the Data Sets. Here, 20 ArcSec is used for target interpolation, which is four times coarser than raw data grids (Figure 23).

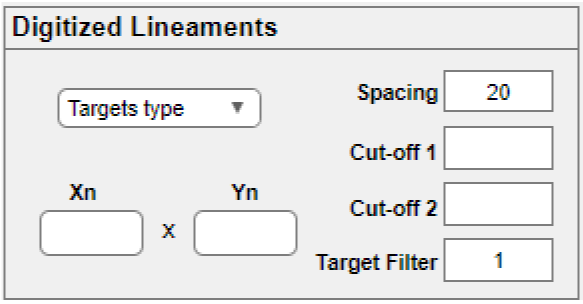

Figure 23: Uploading digitized target faults for validation.

7.1.7. Select Target Types

There are two target types (Figure 24).:

- **Point Data**: CSV files in XYZ format.
- **Image**: Requires coordinates for georeferencing. After selecting the image, a window will open to upload the coordinates in a .txt format.

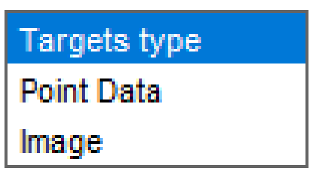

Figure 24: Selecting target types for digitized lineaments.

7.1.8. Define Cut-off Values

After selecting Point Data or Image, the Cut-off 1 box value updates automatically. Cut-off 1 determines the buffer zone around the target samples in the XY plane. The default value is set to produce the smallest possible buffer around the data points (Figure 25).

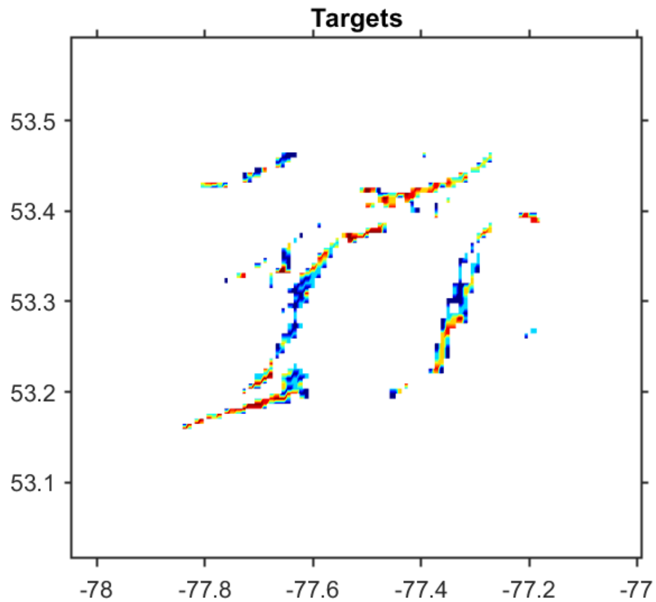

Figure 25: Defining cut-off values for the buffer zone around target samples.

Increase the Data Filter value to expand the buffer zone (Figure 26).

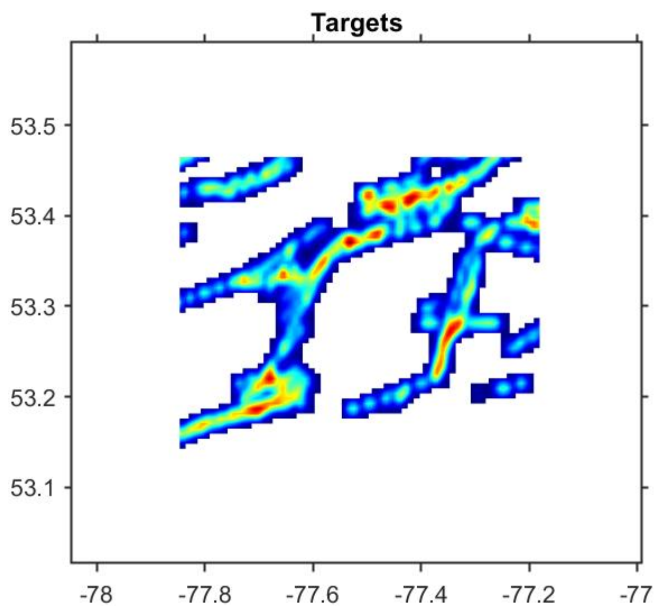

Figure 26: Expanding the buffer zone by increasing the Data Filter value. Data Filter = 1, Cut-off 1 = 0.84127.

Cut-off 2 can also contract the buffer zone. Adjusting these factors helps delineate the targets accurately and eliminates areas without target property information (Figure 27).

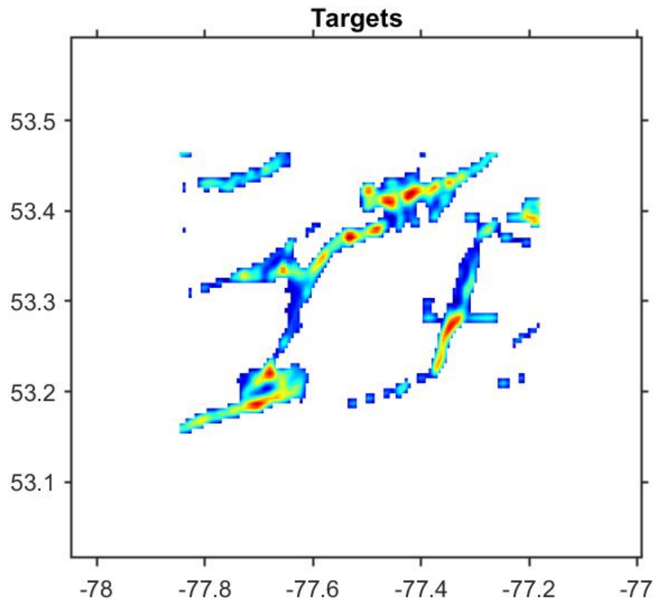

Figure 27: Contracting the buffer zone by adjusting the Cut-off 2 value. Data Filter = 1, Cut-off 1 = 0.84127, Cut-off 2 = 0.86.

7.1.9. Perform Lineament Extraction

Click the Lineaments push button to perform the extraction. From the Plot Results dropdown, select "Fbeta Score" to display the results with an F-beta score (Figure 28).

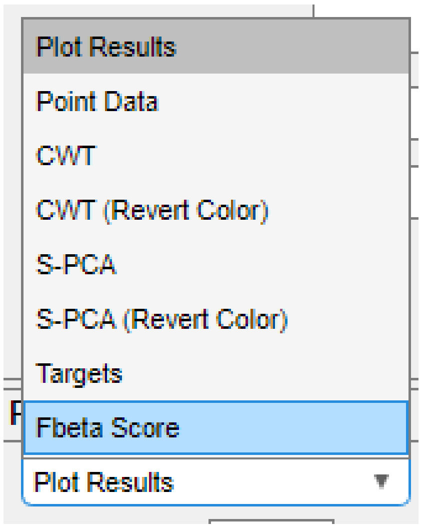

Figure 28: Performing lineament extraction and displaying the results with an F-beta score.

7.1.10. View Results

The results of the lineaments extraction on raw images are displayed as follows (Figures 29, 30 & 31):

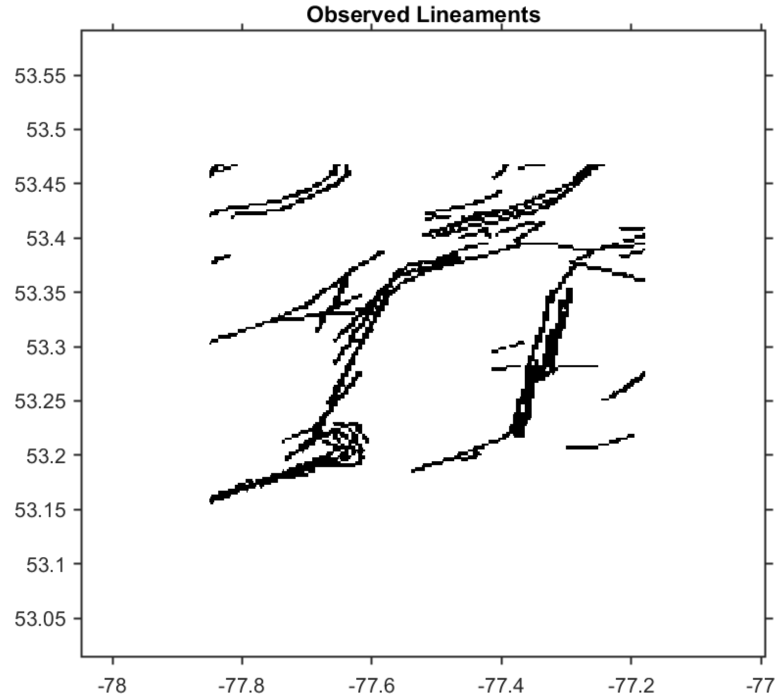

Figure 29: Digitized geological faults.

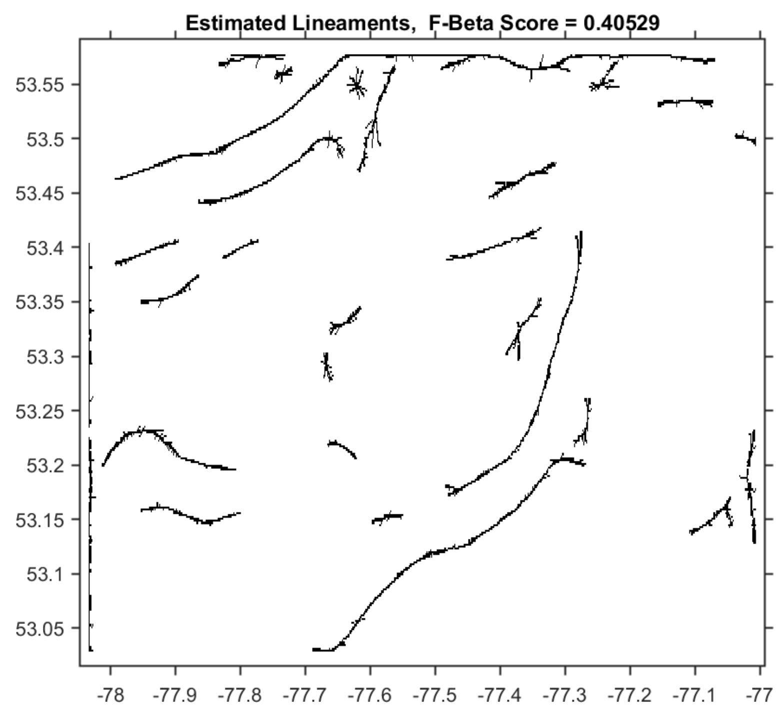

Figure 30: Extracted lineaments on raw images.

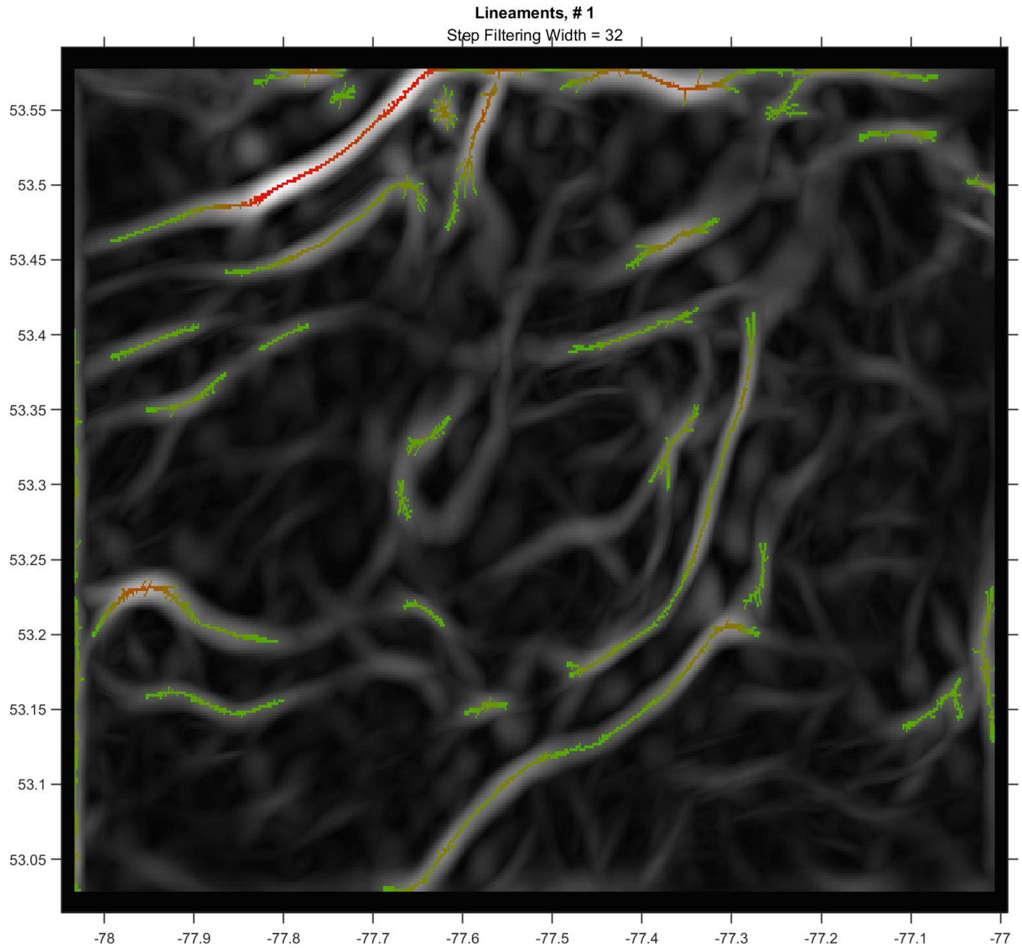

Figure 31: Detailed view of extracted lineaments on raw images.

7.2. Lineaments Extraction with Spectral Feature Extraction with PCA-DOG wavelets

This section builds on the previous method by incorporating principal component wavelet analysis (PCA) to improve edge detection for hysteresis thresholding.

7.2.1. Select Wavelet Parameters for Lineament Extraction

- Number of scales: 3.
- Scale dilation: 1.
- Wavelet Smoothness Filter Ratio (WSFR): 0.
- Eight different directions for wavelet shifting (Figure 32).

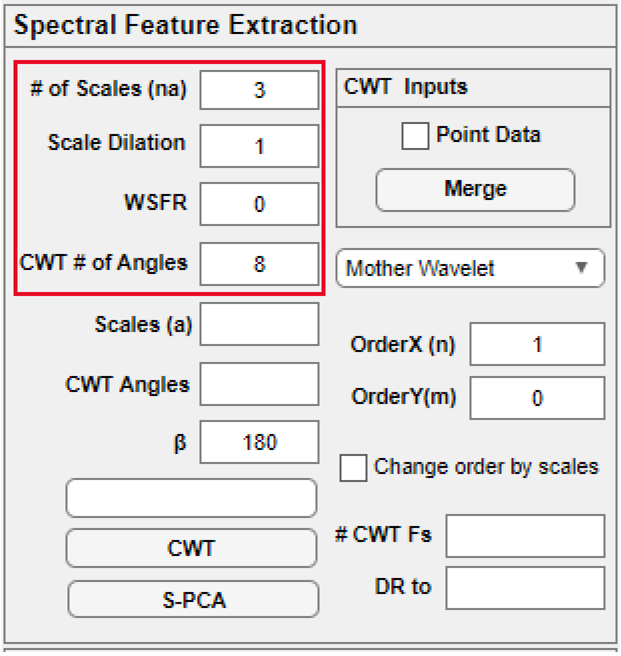

Figure 32: Setting wavelet parameters for lineament extraction using PCA-DOG wavelets.

7.2.2. Select Inputs

Check "Point Data" as the input for the CWT DOG as the mother wavelet (Figure 33).

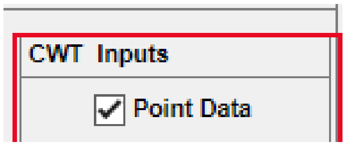

Figure 33: Selecting inputs for the CWT (DOG as the mother wavelet).

7.2.3. Merge CWT Scales and Angles

Push "Merge" to automatically calculate the CWT scales and angles, as well as the angle of rotational symmetry (Figure 34). The results will be scales = {1, 2, 3} and CWT angles = {0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°}.

Since derivatives of the Gaussian are used as the mother wavelet with n = 1 and m = 0, the resulting wavelet has a β = 180° symmetry (if n = m, β = 90°).

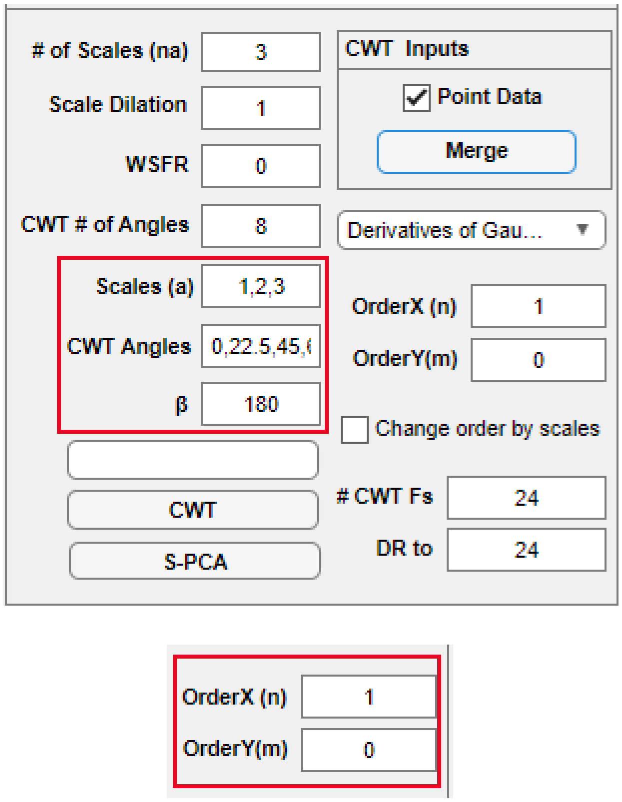

Figure 34: Merging CWT scales and angles and calculating the angle of rotational symmetry.

7.2.4. Perform DOG Wavelet Transform

Push the CWT button to perform the DOG wavelet transform. The algorithm calculates the wavelet coefficients and displays the mother wavelet in space and frequency domains (Figure 35).

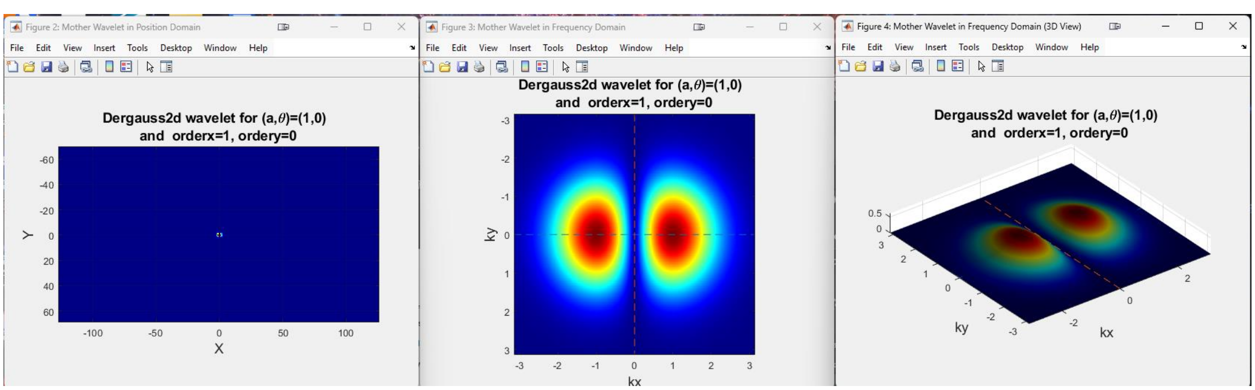

Figure 35: Performing the DOG wavelet transform and displaying the mother wavelet shape.

7.2.5. Dimensionality Reduction

Reduce the dimensionality to 5 from 24 CWT features (Figure 36).

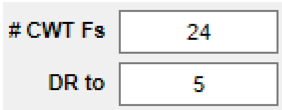

Figure 36: Reducing the dimensionality of CWT features by PCA.

7.2.6. Perform PCA on DOG Wavelet Coefficients

Click the S-PCA pushbutton to perform PCA on DOG wavelet coefficients. An "S-PCA Completed" message confirms successful completion (Figure 37).

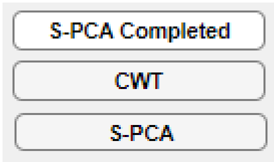

Figure 37: Performing PCA on DOG wavelet coefficients and confirming successful completion.

7.2.7. Select Features for Lineament Extraction

This time, we selected Point Data and S-PCA for lineaments extraction (Figure 38).

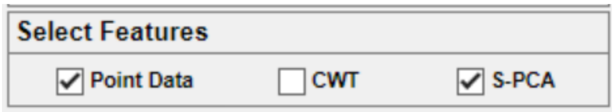

Figure 38: Selecting features for lineament extraction.

7.2.8. Extract Lineaments

Click the "Lineaments" button to perform the extraction using the same parameters as in the fast lineament extraction section. The results are presented below (Figures 39, 40 & 41).

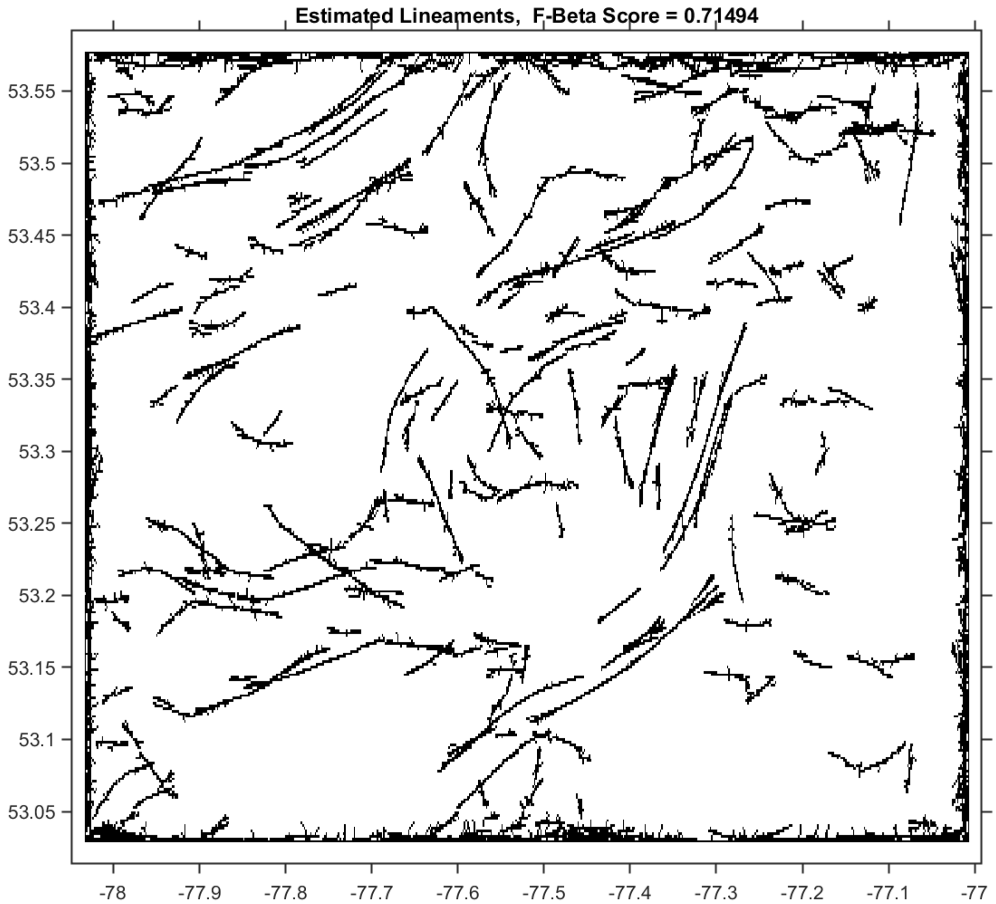

Figure 39: Extracting lineaments using DOG wavelet-PCA feature extraction.

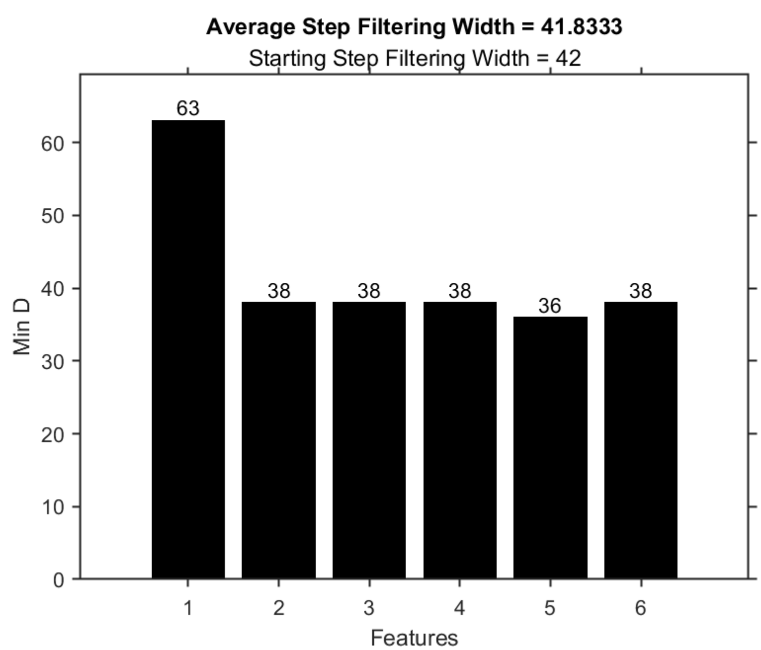

Figure 40: Step filtering widths.

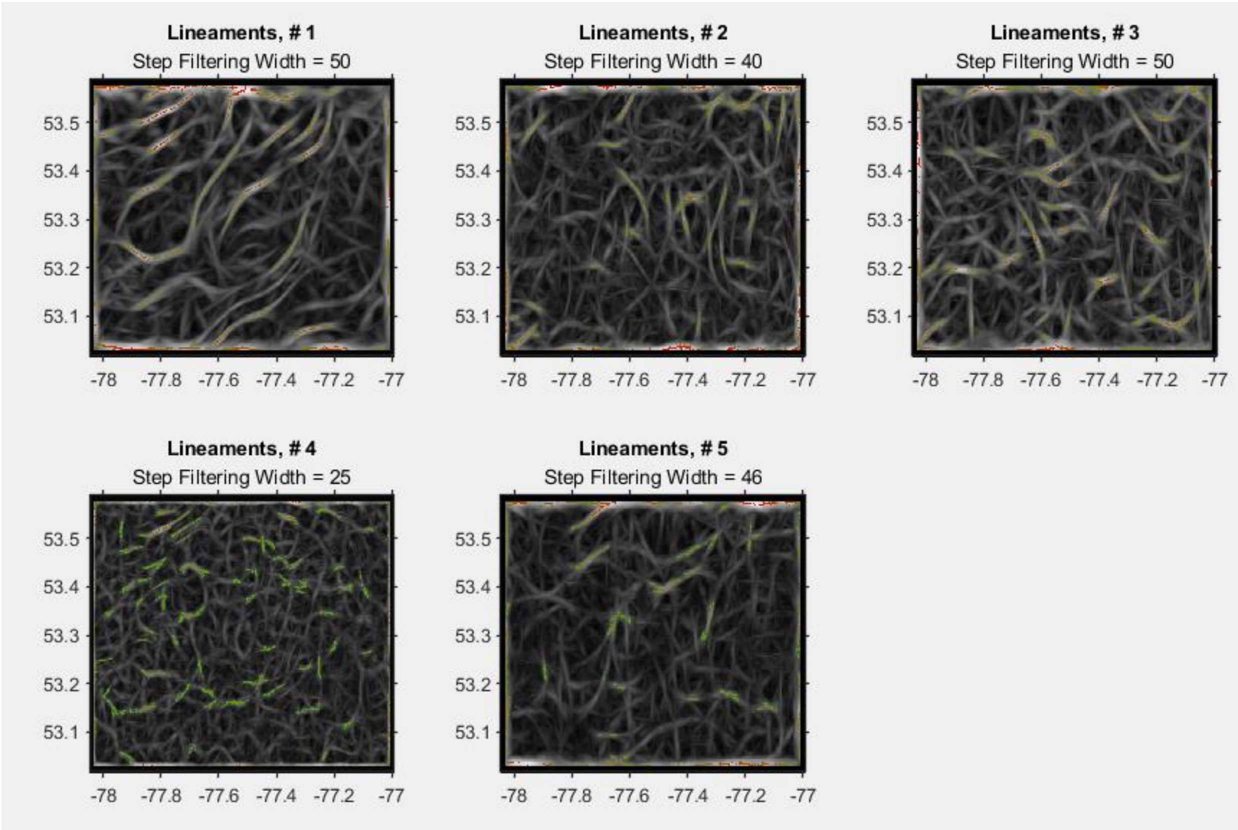

Figure 41: Detailed view of extracted lineaments using DOG wavelet-PCA feature extraction for each wavelet-principal component.

7.3. Lineaments extraction by DOG Wavelet-PCA feature extraction with Bayesian hyperparameter optimization

This section uses the "AutoLine" button for Bayesian hyperparameter optimization based on DOG Wavelet-PCA feature extraction.

7.3.1. Spectral Feature Extraction with Bayesian Optimization

The spectral feature extraction procedure here is similar to the previous section, except that for Bayesian optimization, we set the n and m changes by scales (Figure 42).

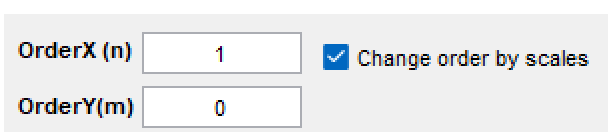

Figure 42: Setting parameters for Bayesian optimization in spectral feature extraction.

In this case, the algorithm tries to find the optimal number of scales (n_a), which in turn determines the subsequent n and m values according to Table 2 (in this case, we should tick the Change order by scales option). The following procedure is used to associate the a to n and m (Figure 43):

```
for jj = 1 : N_Angles
    for jjj = 1 : N_Scales
        if jjj == N_Scales - 0
            orderx = 1;
            ordery = 0;
        elseif jjj == N_Scales -1
            orderx = 2;
            ordery = 0;
        elseif jjj == N_Scales -2
            orderx = 1;
            ordery = 1;
        elseif jjj == N_Scales -3
            orderx = 2;
            ordery = 1;
        elseif jjj == N_Scales -4
            orderx = 2;
            ordery = 2;
        end
        if orderx == ordery
            betad = 90;
        else
            betad = 180;
        end
    end
end
```

Figure 43: Changes in the order of differentiations and β by scales.

The Optimization is set to change the values between 1-5. For example:

- For n_a = 1, we have, a = {1}, and orderx = 1; ordery = 0.

- For n_a = 2, we have, a = {1, 2}, and for a = 1, we have orderx = 2, and ordery = 0, and for a = 2, we have orderx = 1; ordery = 0.

Table 2 summarizes the values for all numbers of scales up to 5:

Table 2. Changes in the order of differentiations and β by scales.

$n_a = 5$

a	***m***	***n***	***β***
5	1	0	180
4	2	0	180
3	1	1	90
2	2	1	180
1	2	2	90

$n_a = 4$

a	***m***	***n***	***β***
4	1	0	180
3	2	0	180
2	1	1	90
1	2	1	180

$n_a = 3$

a	***m***	***n***	***β***
3	1	0	180
2	2	0	180
1	1	1	90

$n_a = 2$

a	***m***	***n***	***β***
2	1	0	180
1	2	0	180

$n_a = 1$

a	***m***	***n***	***β***
1	1	0	180

7.3.2. Select Features for Lineament Extraction

This time, we selected Point Data and S-PCA for lineaments extraction (Figure 44).

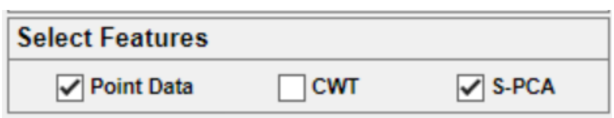

Figure 44: Selecting features for lineament extraction and running optimization.

7.3.3. Run Optimization

Click the "AutoLine" button to perform lineament extraction optimization (Figure 45).

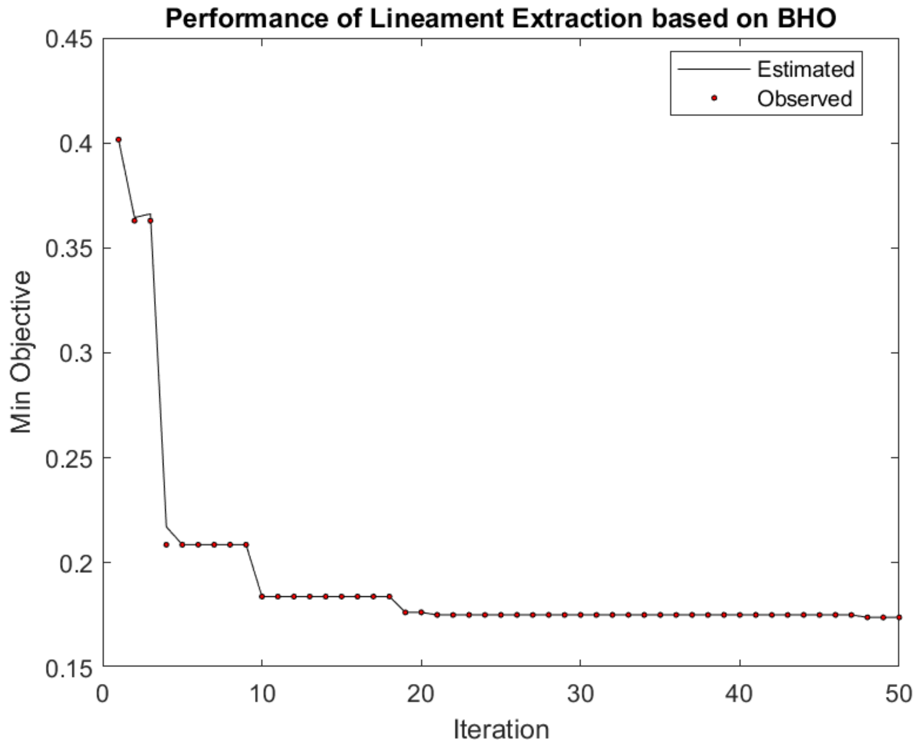

Figure 45: Performing lineament extraction optimization using Bayesian hyperparameter optimization.

7.3.4. Review and Apply Optimized Parameters

The optimized parameters show convergence to a single value or a narrow range of values.

- As a reference, use a value inside the colored box for the last 5 iterations or the median of all values. For example, the number of scales (na) is optimized to 2 in the last iteration. The program exports the optimized median values, but users can manually change them based on previous figures (Figure 46).

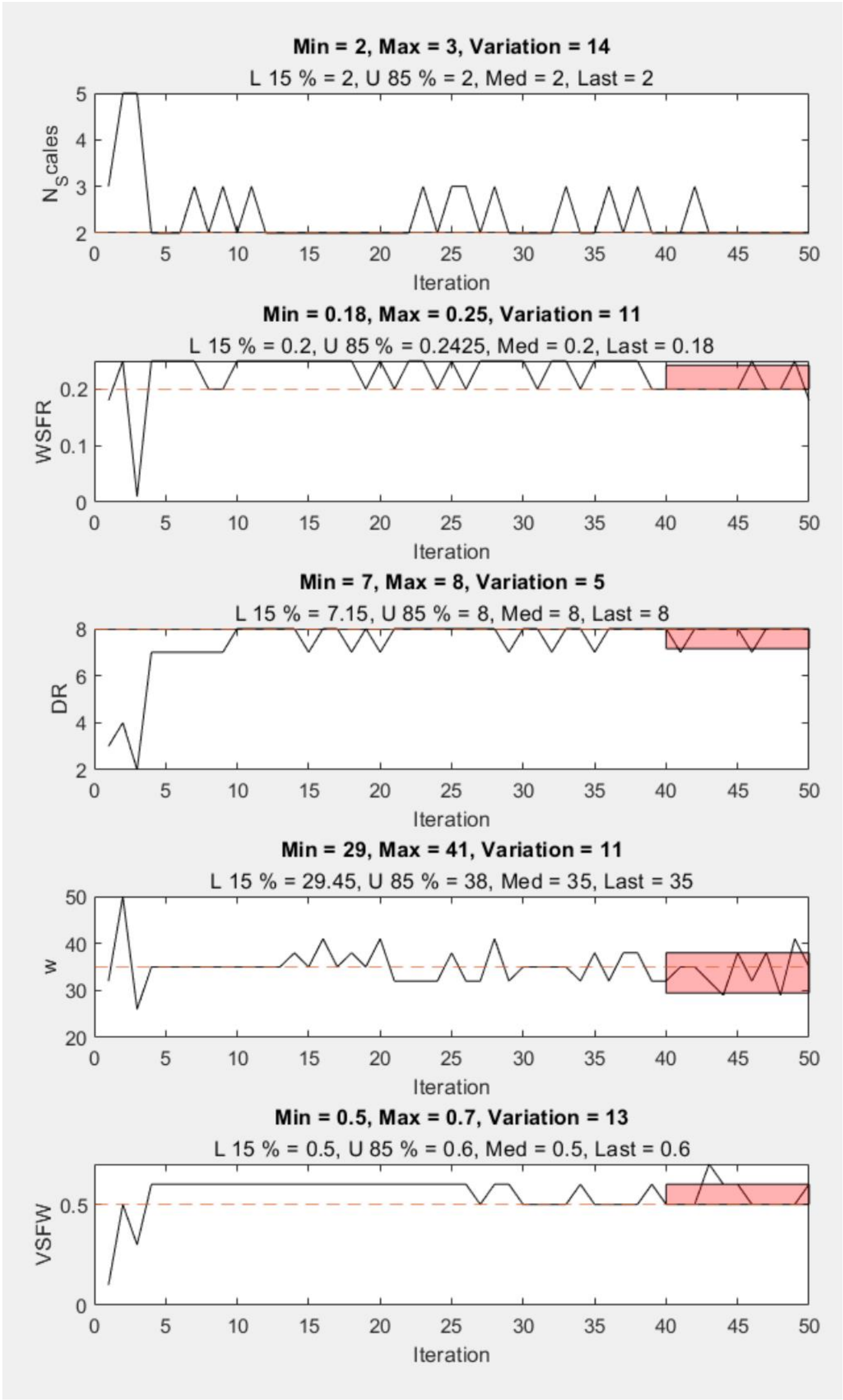

Figure 46: Reviewing and applying optimized parameters for lineament extraction.

Run the feature extraction algorithm again by pushing the Merge button to prepare inputs for CWT calculations with the newly optimized parameters, then push the CWT button (Figure 47).

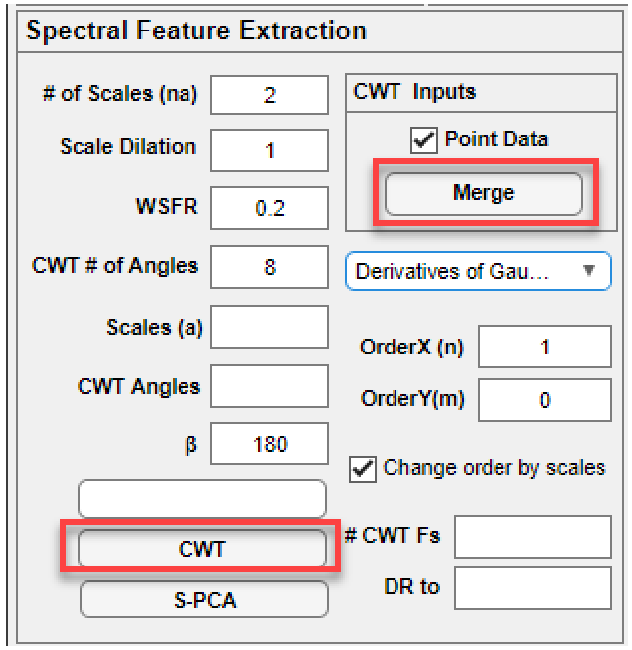

Figure 47: Running CWT with optimized parameters to extract spectral features.

7.3.5. Adjust Dimensionality Reduction

- Note that the DR to value is reset to 16. Set it back to 8 after running the CWT and before S-PCA (Figure 48).

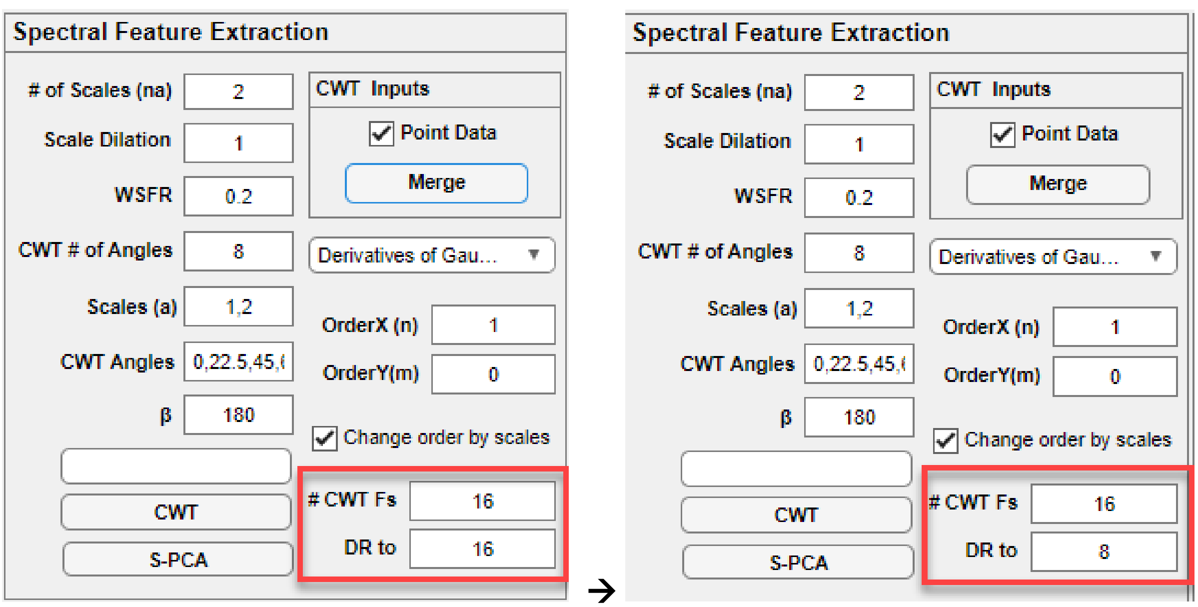

Figure 48: Adjusting dimensionality reduction after running the CWT and before S-PCA.

We verify that only Point Data and S-PCA are selected for lineaments extraction (Figure 49).

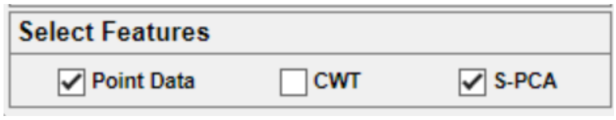

Figure 49: Verifying selection of point data and S-PCA for lineament extraction.

7.3.6. Final Extraction

Verify that only Point Data and S-PCA are selected for lineament extraction. Push the Lineaments button to extract curvilinear faults with the updated fine-tuned parameters.

The results of the lineament extraction with optimized hyperparameters are presented below (Figures 50, 51 & 52).

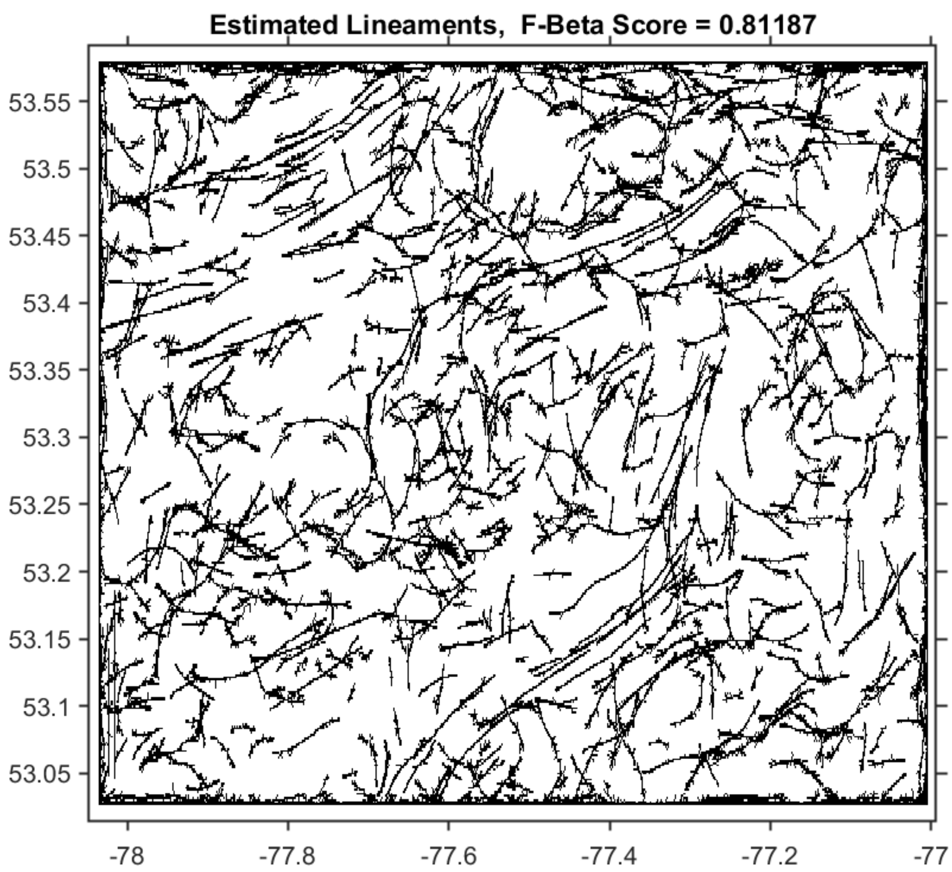

Figure 50: Results of the final lineament extraction with optimized hyperparameters.

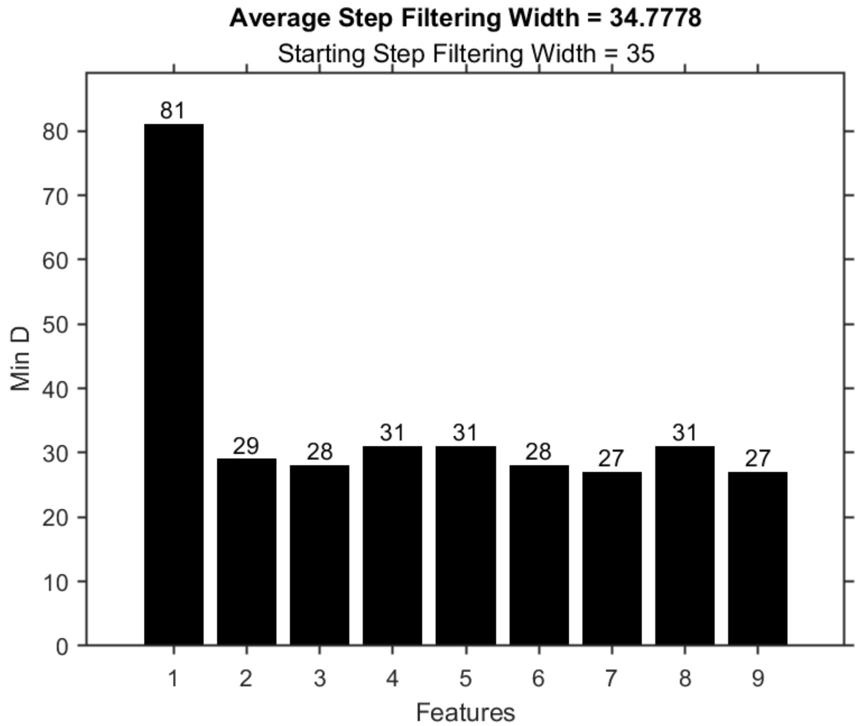

Figure 51: Variations of the step filtering widths.

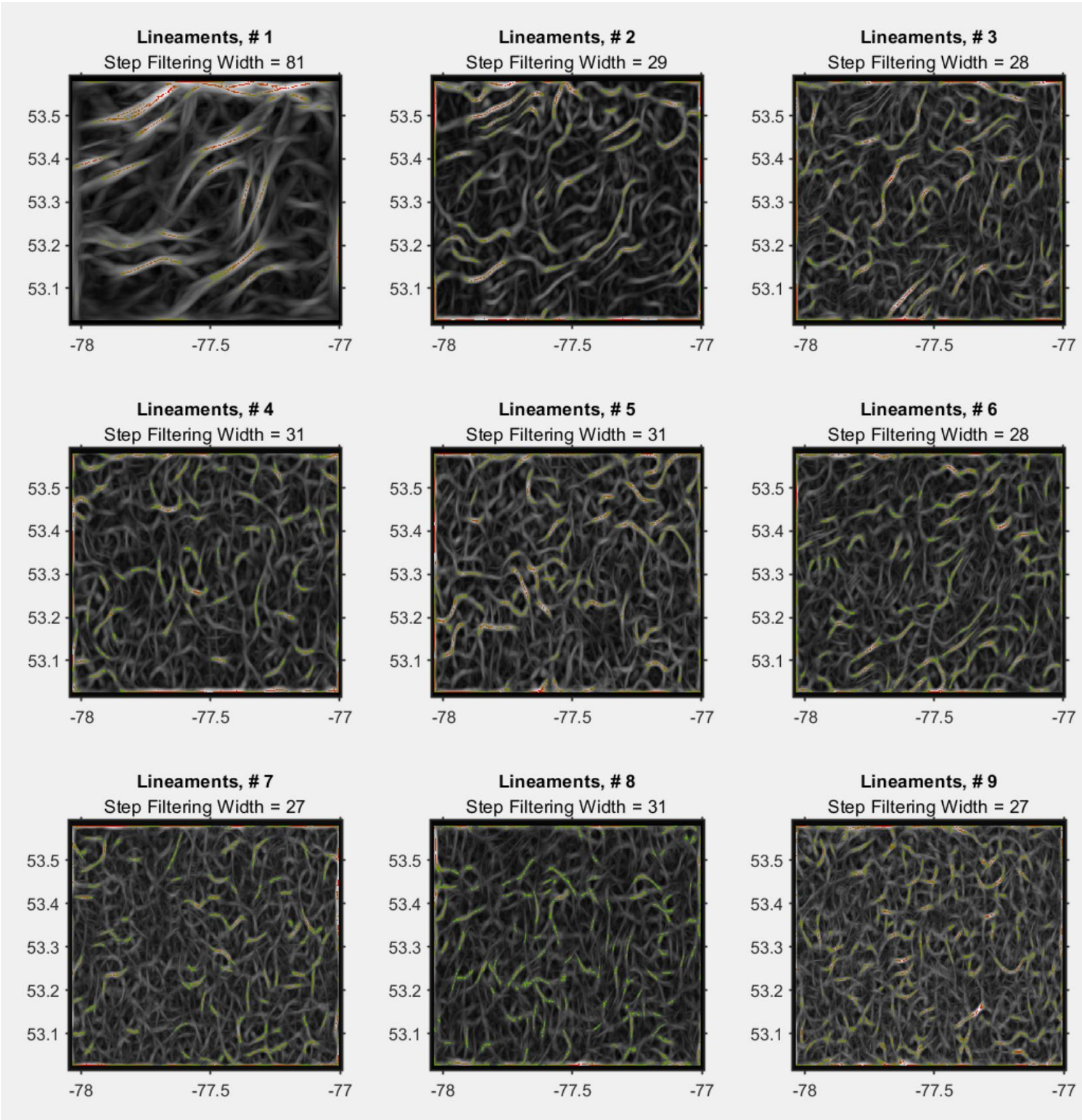

Figure 52: Detailed view of lineament extraction results with optimized hyperparameters for each wavelet principal component.

8. References

Abbassi, B., 2018. Integrated imaging through 3D geophysical inversion, multivariate feature extraction and spectral feature selection. Doctoral dissertation, Université du Québec.

Abbassi, B., Cheng, L.-Z., 2021. SFE2D: A Hybrid Tool for Spatial and Spectral Feature Extraction. In Mining Technology; Hammond, A., Donnelly, B., Ashwath, N., Eds.; IntechOpen: London, UK.

Abbassi, B., Cheng, L.-Z., Jébrak, M., Lemire, D., 2022. 3D geophysical predictive modeling by spectral feature subset selection in mineral exploration. Minerals 12(10).

Antoine, J.-P., Murenzi, R., Vandergheynst, P., Twareque Ali, S., 2004. Two-dimensional wavelets and their relatives. Cambridge University Press.

Archetti, F., Candelieri, A., 2019. Bayesian optimization and data science. Springer.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430.

MathWorks, 2024. Bayesian optimization function. MATLAB. https://www.mathworks.com/help/stats/bayesopt.html (accessed July 8, 2024).

Moore, B., 2024. PCA and ICA Package. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package (accessed July 7, 2024).

Panagiotakis, C., 2024. Filters for Curvilinear Enhancement. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/46541-filters-for-curvilinear-enhancement (accessed July 7, 2024).

Panagiotakis, C., Kokinou, E., 2015. Linear pattern detection of geological faults via a topology and shape optimization method. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(1), 3-11.

YAWTB Team, 2024. Yet Another Wavelet Toolbox. GitHub. https://github.com/jacquesdurden/yawtb (accessed July 7, 2024).