Streaming Synchrophasor Data Quality (SSDQ) – Offline v5.0

Software User's Manual

3002018686

Streaming Synchrophasor Data Quality (SSDQ) – Offline v5.0

Software User's Manual

3002018686

EPRI Project Manager:

Evangelos Farantatos

The following organization(s), under contract to EPRI, prepared this report:

Rensselaer Polytechnic Institute

Troy, NY

NOTE

ACKNOWLEDGMENTS

The following organization(s), under contract to the Electric Power Research Institute (EPRI), prepared this report:

Rensselaer Polytechnic Institute
Troy, NY

Dr. Meng Wang
Dr. Joe H. Chow

This report describes research sponsored by EPRI.

This publication is a corporate document that should be cited in the literature in the following manner:

Streaming Synchrophasor Data Quality (SSDQ) – Offline v5.0 *Software User’s Manual*. EPRI, Palo Alto, CA: 2020. < 3002018686>.

SOFTWARE DESCRIPTION

Description

The Streaming Synchrophasor Data Quality (SSDQ) Offline version 5.0 software is used to improve the quality of streaming synchrophasor data by identifying and correcting bad measurements and filling in missing measurements. The software reads data from selected data files, corrects the bad data, estimates the missing data, and exports the new datasets as Excel files. Several types of measurements, including voltage/current phasors, voltage/current magnitudes, voltage/current phasor angles, and frequencies can be processed. If the original dataset doesn't have missing entries and bad measurements, the users can create artificial scenarios by injecting bad data and erasing measurements following specific modes and percentages of bad or missing data respectively. The software can also compute the estimation performance with varying bad data modes and percentages, as well as data loss percentages and modes. In this version algorithmic and plotting enhancements have been implemented.

Benefits and Value

The value and benefits provided by the software are:

1. For an input dataset with bad measurements and missing measurements, bad data can be identified and corrected, and missing data can be estimated and filled in with high accuracy;
2. For an input dataset without bad or missing measurements, users can inject random errors to the data, erase measurements, and compare the estimation performance for varying missing and bad data modes and percentages;
3. The methods used in the software have low computational complexity and high accuracy.

Platform Requirements

The software is developed in Matlab®. Windows 8® or Windows10® 64 bit operating system is required. To run the executable file, the MATLAB Compiler Runtime (MCR) version 9.6 (R2019a) is necessary. The Windows 64-bit version of the MCR for R2019a can be downloaded from the MathWorks Web site by navigating to http://www.mathworks.com/products/compiler/mcr/index.html.

Keywords

Missing data estimation, bad data correction, PMU, streaming synchrophasor data

CONTENTS

LIST OF FIGURES

LIST OF TABLES

No table of figures entries found.

MANDATORY SOFTWARE INSTALLATION INFORMATION

Installation of EPRI Software at Client Site

EPRI develops software using a number of third party software products and tools that run on various operating systems and server platforms. Reports from the software industry suggest there are known security issues with some products and systems. EPRI recommends that, if you are using EPRI software, you review its use with your Information Technology (IT) department and their overall strategy to ensure that all recommended security updates and patches are installed as needed in your corporation. If you have any concerns please call the EPRI Customer Assistance Center (CAC) at 1-800-313-3774 (or email askepri@epri.com).

If you experience difficulties accessing the application

If you experience difficulties accessing the application after standard installation on Windows Vista, Windows 7, and Windows 8, please consult your IT department personnel to have proper access permissions setup for your use. If the problem can not be resolved, please call the EPRI Customer Assistance Center (CAC) at 1-800-313-3774 (or email askepri@epri.com).

1

USER MANUAL

This chapter provides an overview of the Streaming Synchrophasor Data Quality (SSDQ) software features.

Prerequisites and System Requirements

The software is developed in Matlab®. Windows 8® or Windows 10® 64 bit operating system is required. At least 1024 MB of RAM is required.

To run the software, the user needs first to install MATLAB Compiler Runtime (MCR) version 9.6 (R2019a).

Windows 64-bit version of the MCR for R2019a can be downloaded from the MathWorks Web site by navigating to http://www.mathworks.com/products/compiler/mcr/index.html.

For more information about the MCR and the MCR Installer, see Distribution to End Users in the MATLAB Compiler documentation in the MathWorks Documentation Center.

NOTE: You will need administrator rights to run MCR Installer.

Getting Started

To start the software, directly run the executable file. The starting interface of the software is shown in Figure 1-1. After pressing the "Accept" button, the main interface is shown as Figure 1-2.

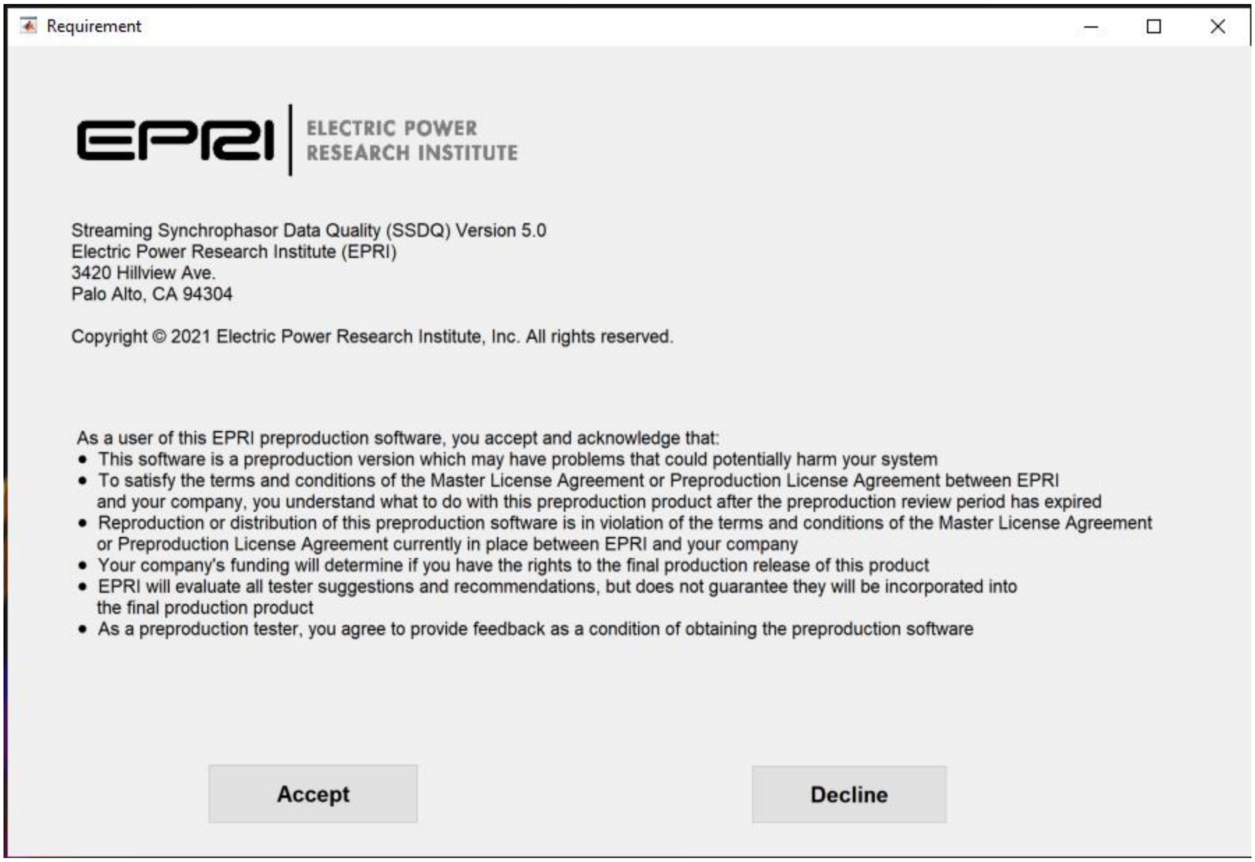

Figure 1-1: Starting interface of the SSDQ software

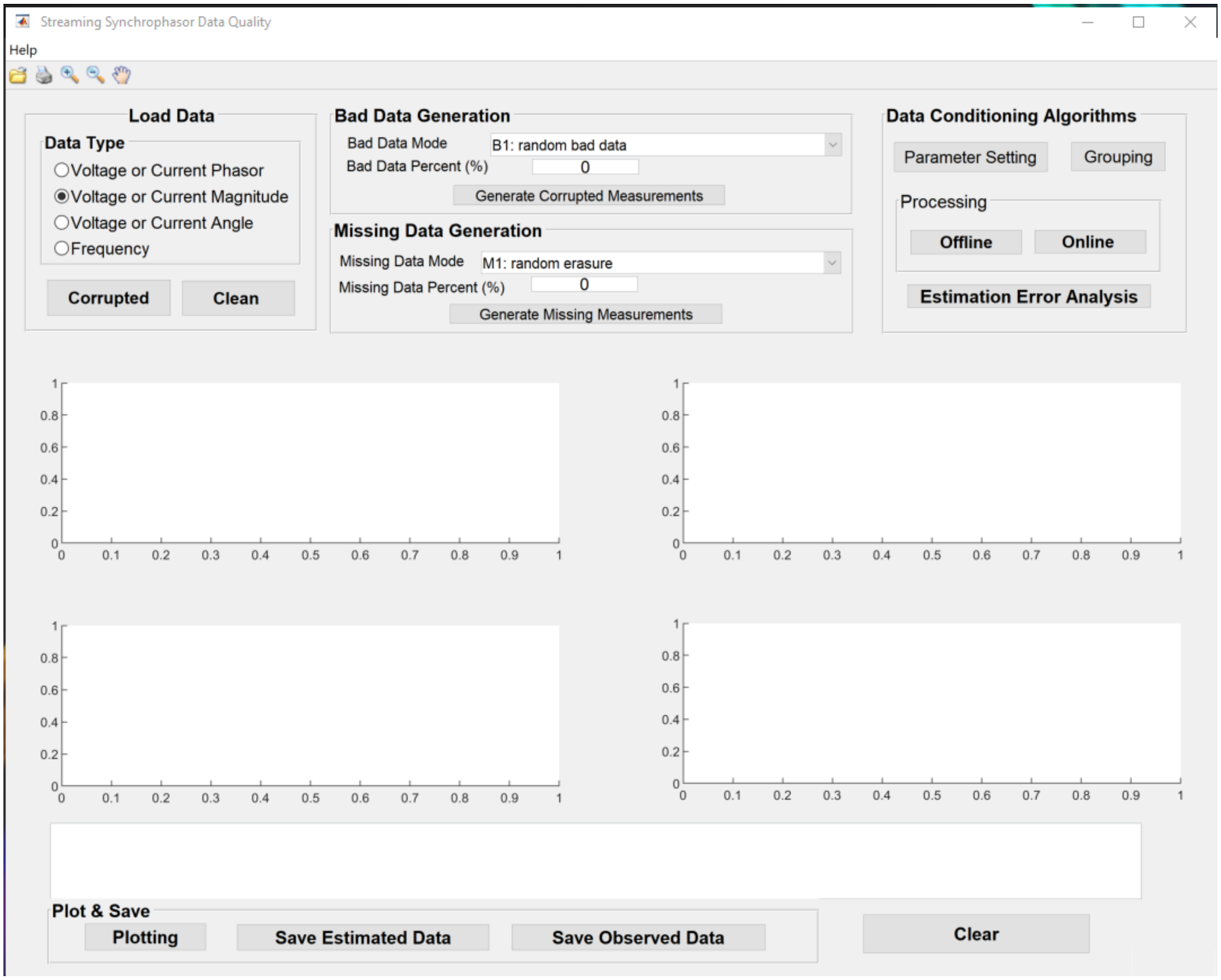

Figure 1-2: Main interface of the SSDQ software

Input Files

There are four data types that can be used as input files. The user should select the appropriate data type. The input data format is described in the appendix.

1) Voltage or Current Magnitude
2) Voltage or Current Angle
3) Frequency
4) Voltage or Current Phasor

Sample input datasets are provided in the folder "Data files". The filenames are given below.

- Corrupted_frequency.xls
- Corrupted_voltage_angle.xls

- Corrupted_voltage_magnitude.xls
- Corrupted_voltage_phasor.xls
- Full_frequency.xls
- Full_voltage_angle.xls
- Full_voltage_magnitude.xls
- Full_voltage_phasor.xls

Load Data with Missing/Bad Entries or without Missing/Bad Entries

If the input dataset has missing and/or bad entries (in the input files, the missing entries are indicated as 0), users should select the "Corrupted" option within the "Load data" panel and browse through the file. After specifying the starting and ending time of the loaded data, the corrupted data within the time interval will be plotted.

If the dataset does not have missing or bad data, the user should instead select the "Clean" option and browse through the file. After the starting and ending time of the loaded data is specified, the complete measurements will be plotted.

Generate Corrupted Data for Clean Datasets

If the input dataset is without missing or bad entries, users can specify the percentage of corrupted data and inject random errors across the entire dataset (including the first window) following one of the three modes of bad data provided by SSDQ as shown in Figure 1-3. The bad data modes are explained next.

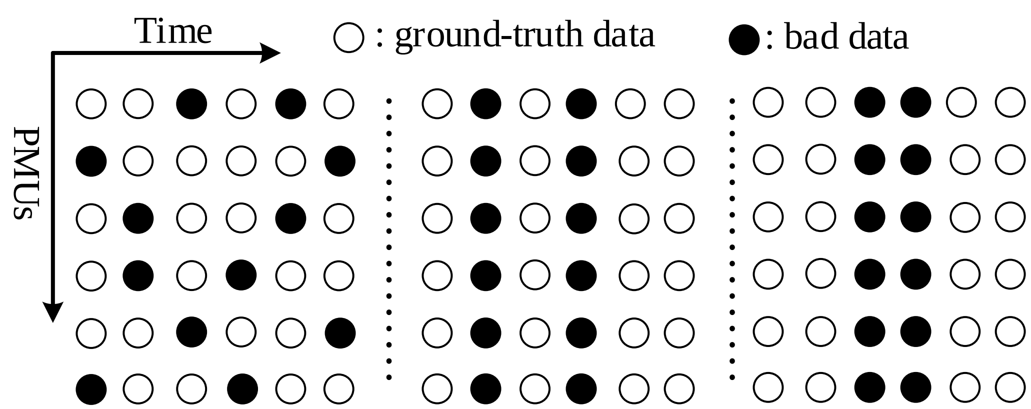

Figure 1-3: Three bad data modes

- B1: Bad data occur randomly and independently across channels and time instants. The additive errors in voltage magnitudes, angles and frequencies follow Gaussian distribution $\mathcal{N}(0, \boldsymbol{\sigma}_i{}^2)$. The standard deviation of i th channel is:

$$\boldsymbol{\sigma}_i = \frac{\sum_{k=1}^{N} |X_{ik}|}{N} \times 10\% \text{ for magnitude dataset}$$

$$\boldsymbol{\sigma}_i = \frac{\sum_{k=1}^{N} |X_{ik}|}{N} \times 100\% \text{ for angle dataset}$$

$$\boldsymbol{\sigma}_i = (\max(X_i) - \min(X_i)) \times 10\% \text{ for frequency dataset}$$

where $|X_{ik}|$ is the absolute value of k_{th} time instant in i_{th} channel and X_i is the vector of all time instants in i_{th} channel. N is the number of all the time instants in i_{th} channel.

B2: Bad data occur in all channels simultaneously at random time instants. The additive errors are independent across time but are the same for all channels at one time instant. The errors are generated following Gaussian distribution $\mathcal{N}(0, \boldsymbol{\sigma}_i{}^2)$. The standard deviation of i th channel is:

$\boldsymbol{\sigma}_i = \frac{\sum_{k=1}^{N}|X_{ik}|}{N} \times 10\%$ for magnitude dataset

$\boldsymbol{\sigma}_i = \frac{\sum_{k=1}^{N}|X_{ik}|}{N} \times 100\%$ for angle dataset

$\boldsymbol{\sigma}_i = (\max(X_i) - \min(X_i)) \times 10\%$ for frequency dataset

where $|X_{ik}|$ is the absolute value of k_{th} time instant in i_{th} channel and X_i is the vector of all time instants in i_{th} channel. N is the number of all the time instants in i_{th} channel.

- B3: This mode is composed of two types of bad data. The first type is exactly the same as described in B1 and the user can adjust the percentage of this type of bad data. The second type is described below, and the percentage of this type is fixed to be 2.5% for current version.

  The second type is the bad data occur in all channels simultaneously and consecutively for some time. The additive errors are independent across time but are the same for all channels at one instant. The errors are generated following Gaussian distribution $\mathcal{N}(\mu_i, \boldsymbol{\sigma}_i{}^2)$. The mean of i th channel is:

  $\mu_i = \frac{\sum_{k=1}^{N}|X_{ik}|}{N} \times 10\%$ for magnitude dataset

  $\mu_i = \frac{\sum_{k=1}^{N}|X_{ik}|}{N} \times 10\%$ for angle dataset

  $\mu_i = (\max(X_i) - \min(X_i)) \times 10\%$ for frequency dataset,

  where $|X_{ik}|$ is the absolute value of k_{th} time instant in i_{th} channel and X_k is the value array of all time instants in i_{th} channel. N is the number of all the time instants in i_{th} channel.

  $\boldsymbol{\sigma}_i$ is $\mu_i \times 50\%$ for all types of datasets.

Note that for B3 mode, the bad data percent has to be at least 2.5%, otherwise, a warning message would pop up as shown in Fig. 1-4.

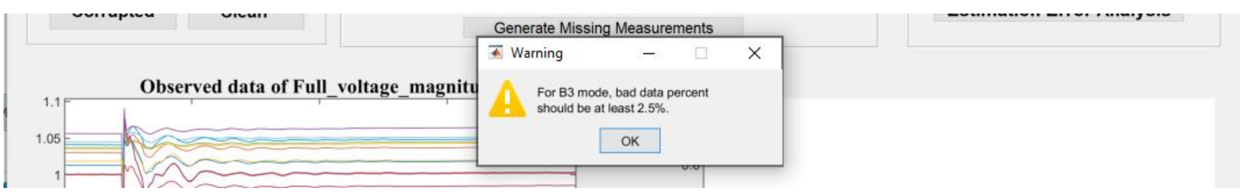

Figure 1-4: Example of warning message when the bad data percent that the user enters is less than 2.5%

Generate Missing Data for Clean Datasets

If the input dataset is without missing entries, users can specify the percentage of data loss and erase data across the entire dataset (including the first window) following one of the three modes of data loss provided by SSDQ. The data loss modes are explained next.

- M1: Data losses occur randomly across PMU channels and time instants.
- M2: Data losses occur at random time instants. At each time instant with data losses, only r out of m measurements are observed, where m is the number of PMU channels. The observed entries are random across PMU channels.
- M3: Data losses occur at consecutive time instants. Only r out of m measurements are obtained at each time instant. Moreover, these obtained measurements are from the same set of PMU channels.

One diagram to demonstrate the three modes is shown in Fig. 1-4, where r=2. In the software, we set $r \approx \frac{m}{2}$, where m denotes the number of PMU channels.

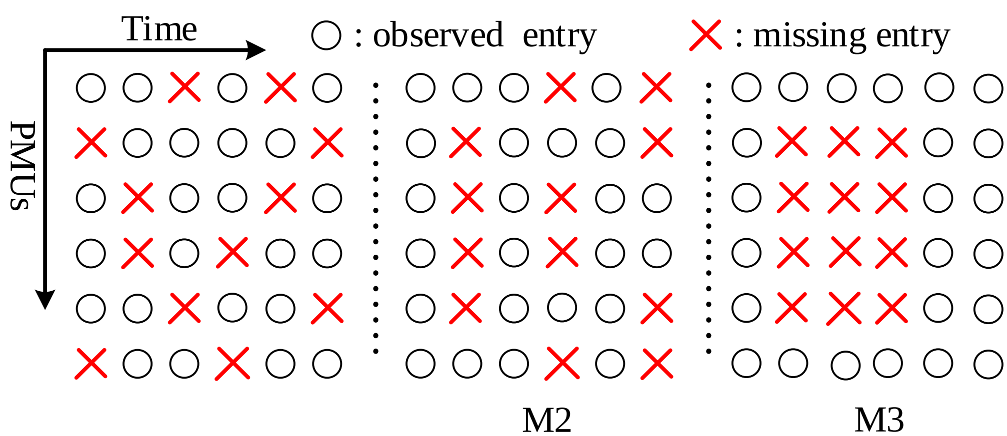

Figure 1-5: Three missing data modes

Analysis Method

The software employs the developed algorithms in both [1] and [3] to identify and correct bad data, as well as to estimate and fill in the missing entries. In online robust data estimation, the algorithm proposed in [3] is first used as an initialization algorithm to estimate the data in the first window. Next, the data in the first window after initialization is viewed as clean data, and then the algorithm proposed in [3] is utilized to estimate the remaining corrupted/missing data. In the offline robust data estimation, the algorithm proposed in [3] is used to estimate the entire dataset. When the percentage of bad data is 0, i.e., there is no bad data in the partially observed data, then the software employs the simpler version of the algorithm in [1] and [3] to estimate the unobserved entries, and the observed entries remain the same in the final output.

Parameter Setting

The default parameters of the estimation method are shown in Fig.1-5. The interface shows up after users press the "Parameter Setting" button within the "Data Conditioning Algorithms" panel. Three parameters are related to the estimation methods with streaming data:

Two parameters are related to both online and offline processing

- Number of vectors in each column of the Hankel matrix: it determines the structure of the constructed Hankel matrix.
- Block data recovery threshold: trade off parameters between the convergence and the recovery accuracy.

Four parameters are related to Online processing only

- Length of the moving data window: it determines the number of observations employed to estimate the missing entries in the current observation.
- Approximation error threshold to determine the matrix rank: it determines the dimension of the underlying subspace of the PMU data within the window.
- Ratio of approximation errors to distinguish measurements from simultaneous and consecutive bad data: it is used to differentiate synchrophasor measurements from consecutive and simultaneous bad data.
- Values of a and b: they are used to determine the time-varying threshold of bad data detection.

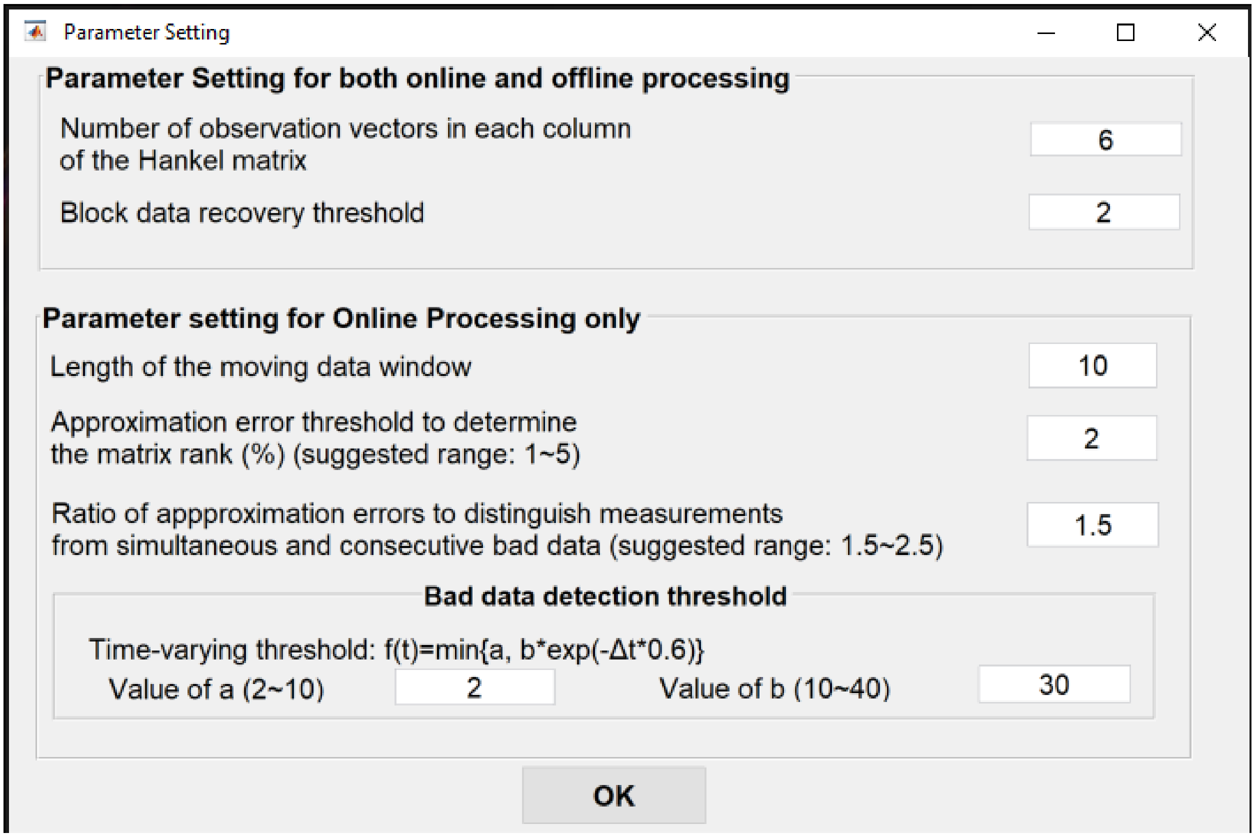

Figure 1-6: Interface of parameter setting

Grouping

After the selection of the various settings, users may choose to assign different channels into different groups by pressing "Grouping" within the "Data Conditioning Algorithms" panel, although it is not absolutely necessary. Multi-group processing separates the whole input dataset into different subgroups, and then each subgroup is processed by the algorithm separately. By assigning channels with a similar pattern into the same group, the algorithm can possibly achieve a better performance.

Fig. 1-6 shows the Grouping interface. Users can first specify the group size (the default size is 1). The selection of group size depends on how many channels are in the dataset and the current maximum group size is set as 7 manually to simplify representation, for a dataset with 11 channels, group size as 2 should be enough. Also, it is suggested to ensure that there are at least 3 channels in each group.

There are two grouping options. Users can either assign channels manually or let the K-means algorithm groups them automatically.

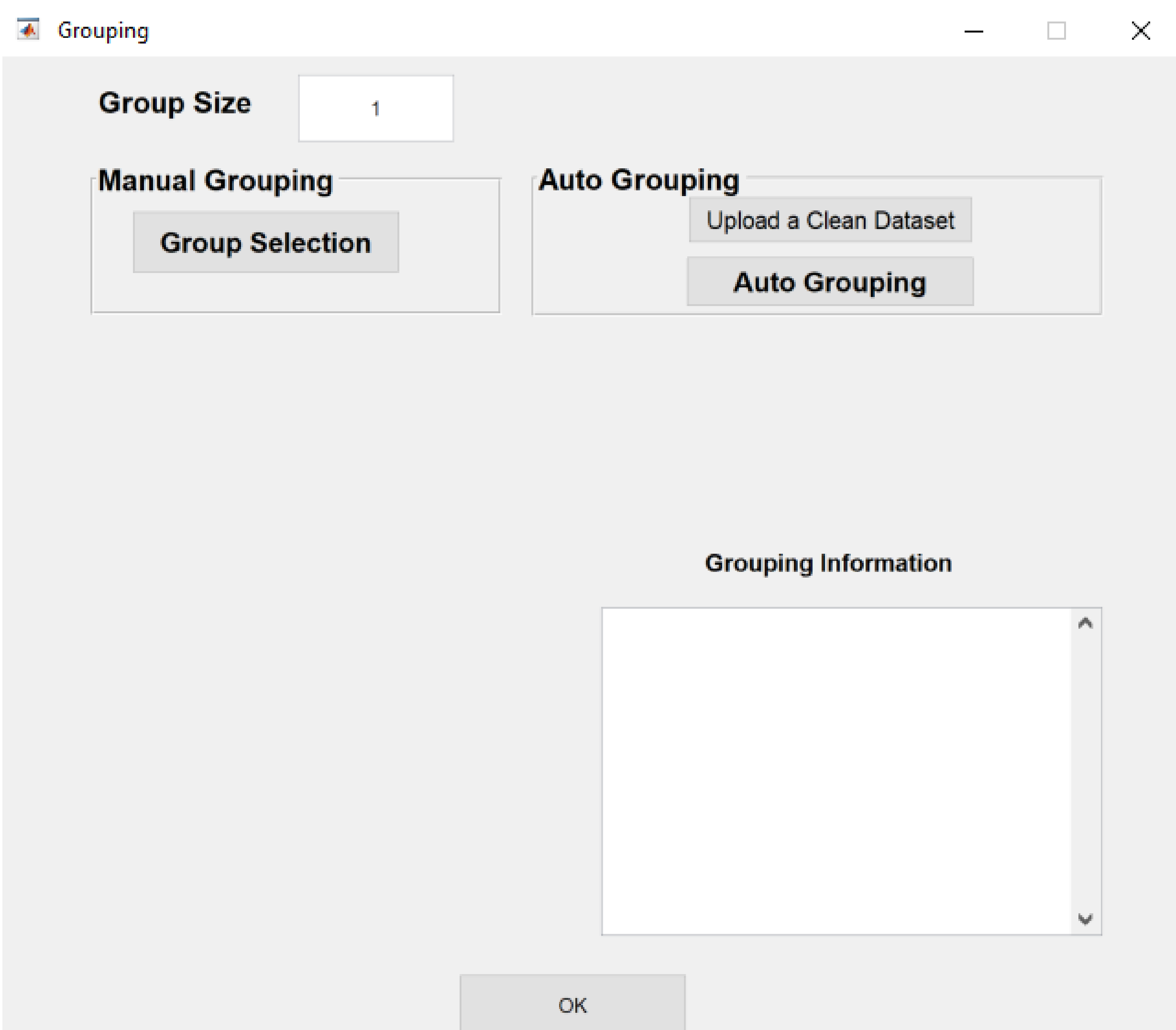

Figure 1-7: Interface of Grouping selection

For manual grouping, users can press the “Group Selection” within the “Manual Grouping” panel, and then push-buttons for each group pop up as shown in Fig. 1-7. Users can select channels for each group by pressing the respective push-button, and the channel selection panel for the specific group pops up as shown in Fig. 1-8.

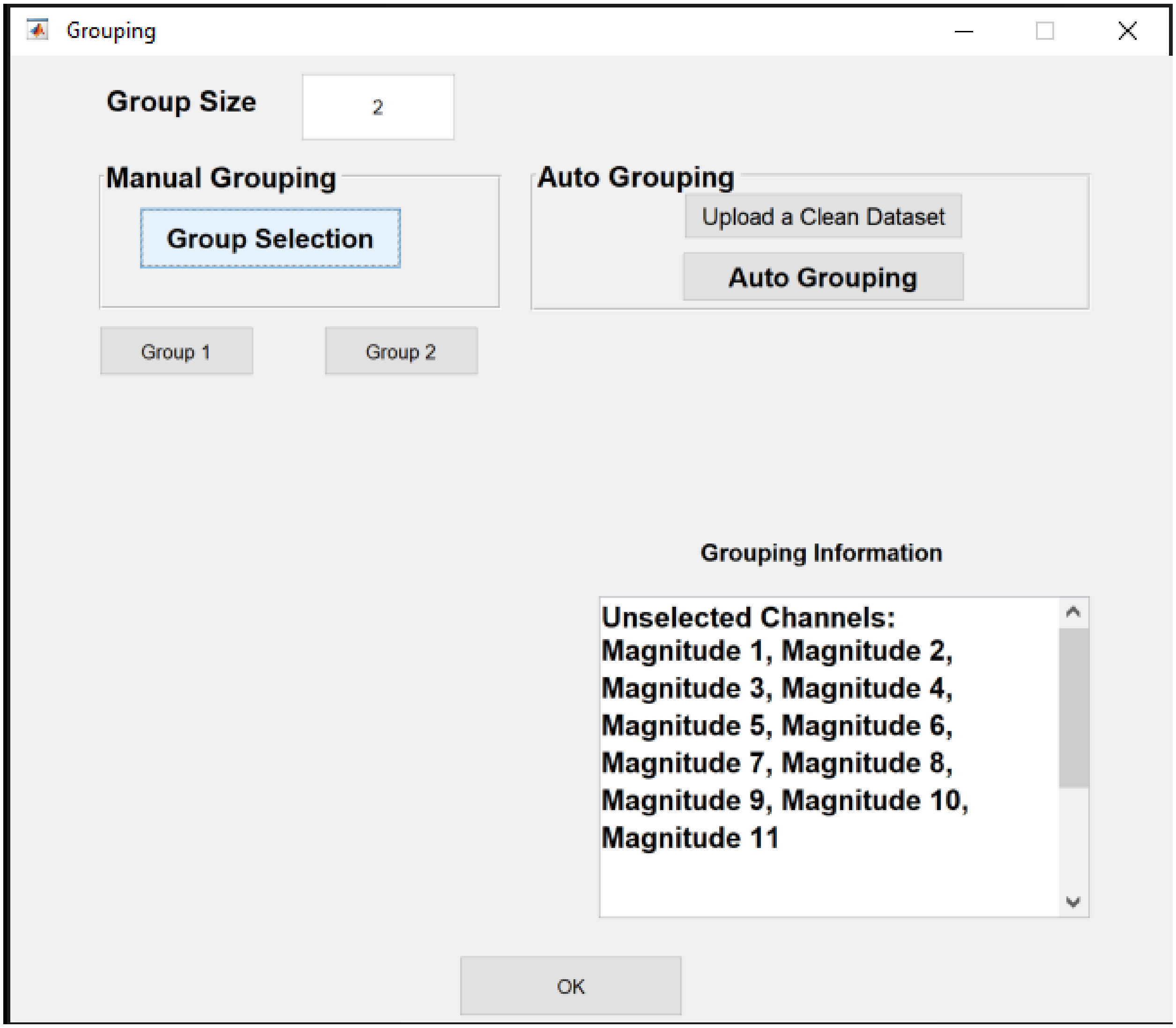

Figure 1-8: Example of Manual Selection with the group size of 2

The user can then select the channels to be put into the specific group by checking the specific box behind each channel. Furthermore, users can visualize the selected channels by pressing "Plot" for the ease of comparison. After finishing the selections, please confirm the changes by pressing "Select".

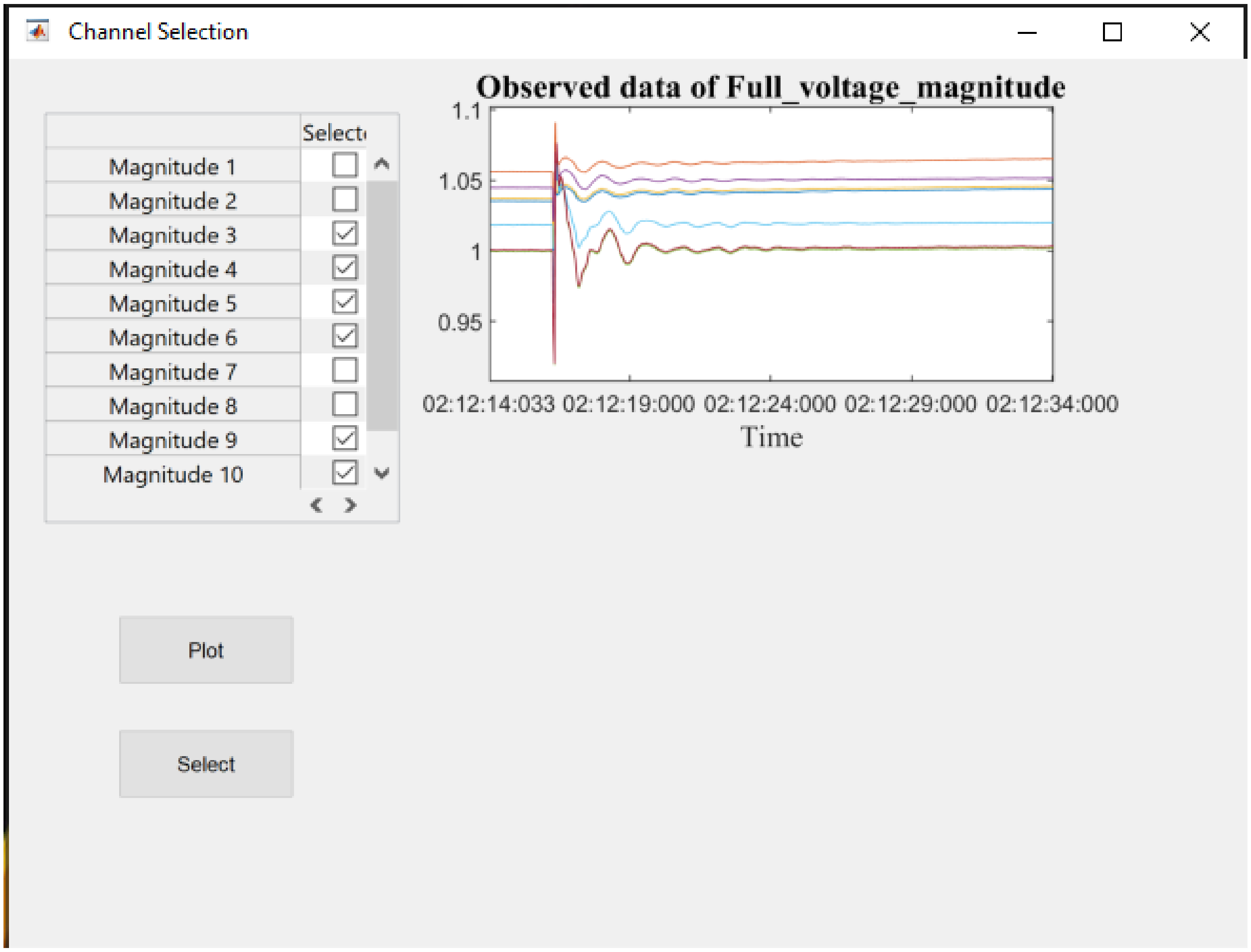

Figure 1-9: Interface of Channel Selection

For auto grouping, users can press the "Auto Grouping" after specifying the group size. Then the push-buttons for each group pop up, and the grouping process is already finished by the K-means program as shown in Fig. 1-9. In the bottom left corner of the panel, the table displays the distance between centroids of groups, and these values could be used to evaluate the similarities of different groups . Users can also adjust the grouping results by clicking on push-buttons of individual groups and making changes in the channel selection interface. Note that one channel can only appear in one group. Otherwise, an error message would appear.

If the user uploaded a clean dataset for testing, this dataset will be used as the default dataset for auto grouping. If users want to use another dataset for grouping, one could press "Upload a Clean Dataset" within Auto Grouping panel, and the dataset shall have the same number of channels as the dataset used for testing. If the user uploaded a corrupted dataset for testing, it is required to upload a clean dataset for grouping. In this case, the "Auto Grouping'' push-button is disabled until a clean dataset is uploaded.

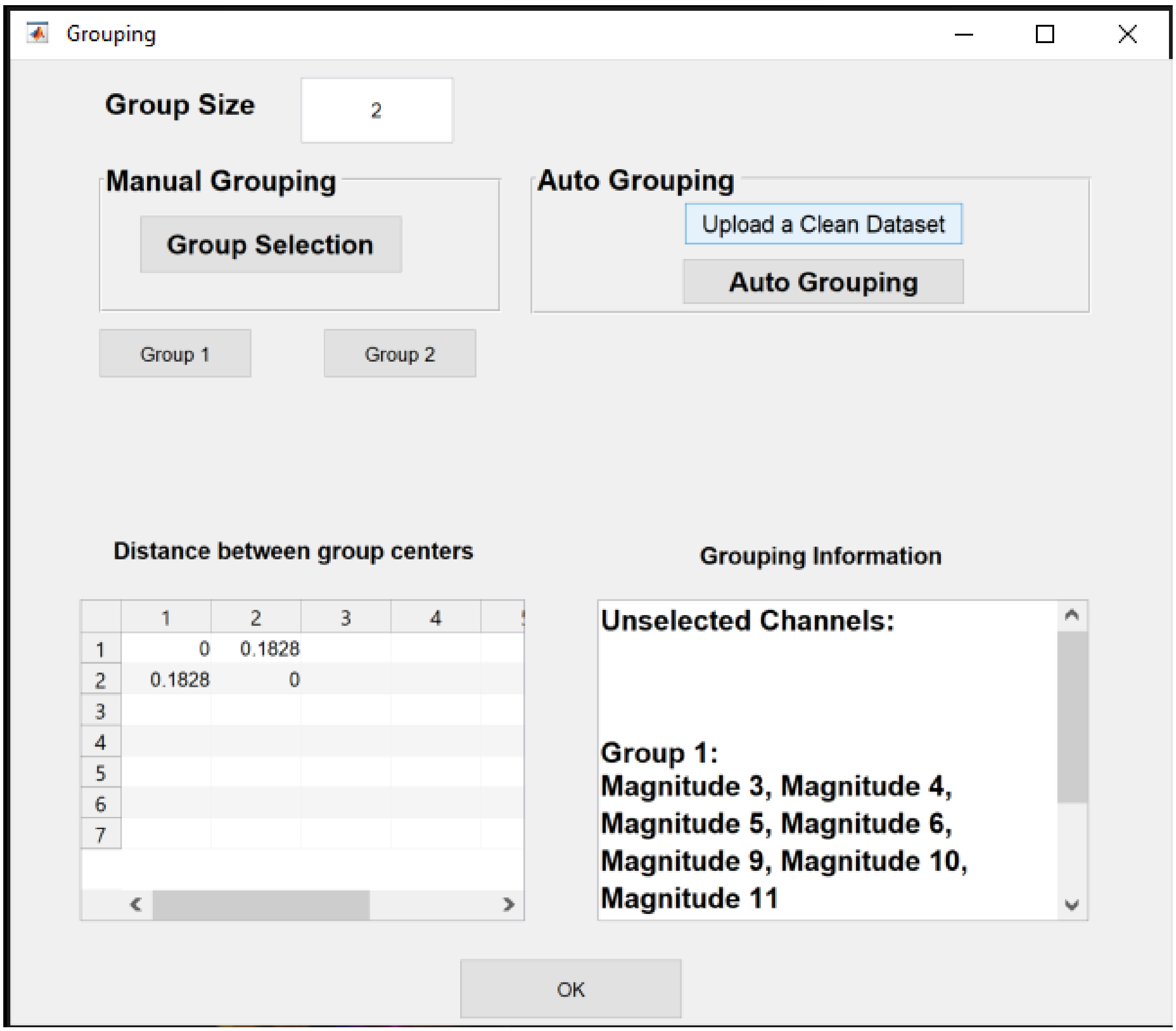

Figure 1-10: Example of Auto Grouping with the group size of 2

The “Grouping Information” panel located in the bottom right corner of the interface displays the following information:

- Unselected Channels (Not in any of groups)
- Channels names for each created group

Therefore, before pressing “OK”, users are suggested to ensure that there is no channel in the “Unselected Channels” unless they intend to.

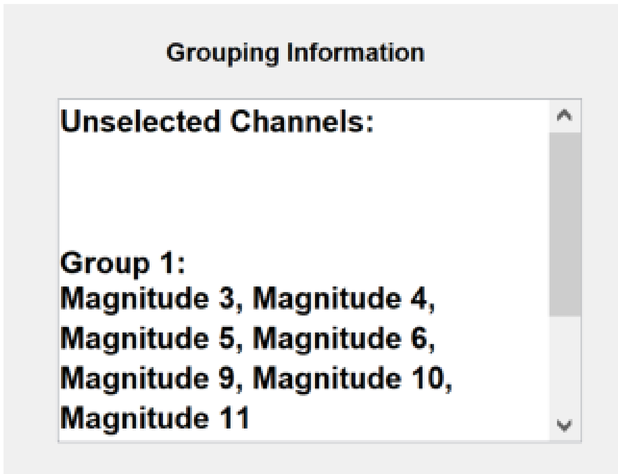

Figure 1-11: Group Information Panel

Online Robust Data Estimation

After Grouping(optional), users can press the "Online" button within the "Processing" panel to perform the analysis. In this version of the software, an initialization algorithm is added to estimate the data in the first window, and the current algorithm no longer requires the data in the first window to be clean. One example of the analysis using voltage magnitudes as input with 10% random bad data in mode B1 and 10% random erasure in mode M1 is shown below in Fig. 1-11.

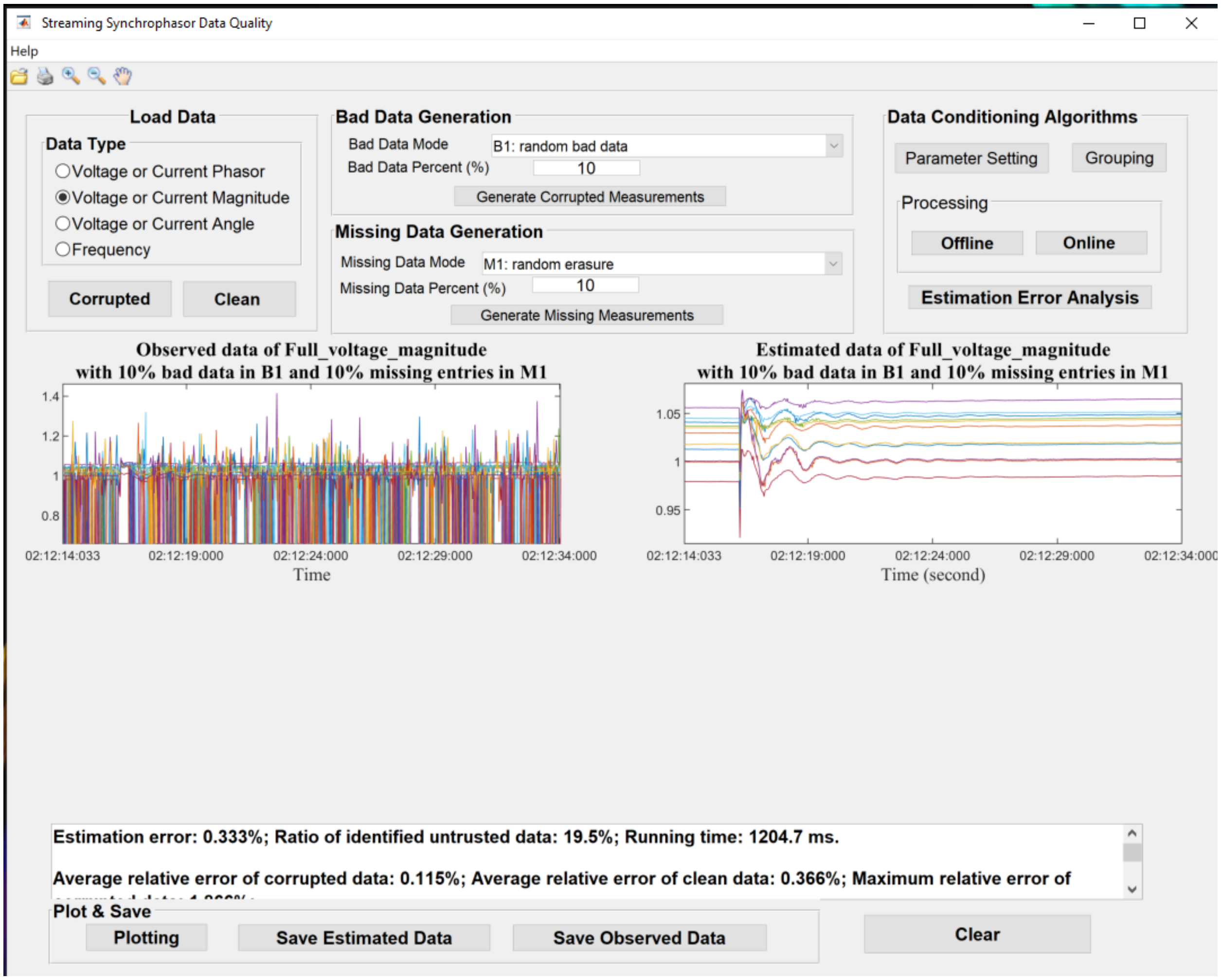

Figure 1-12: Example of the observed and estimated data with 10% bad data in B1 and 10% missing entries in M1 using online processing

Offline Robust Data Estimation

For the offline data estimation, the incorporated method is a batch mode method. The benefits of offline data processing are that it can tolerance higher percentage missing/bad data and has faster computational time. However, because offline data estimation method tries to recover the whole data simultaneously, the method lacks the analysis on the local behavior of the data. For example, it cannot identify the event data and may misinterpret the event data as bad data. One case is shown below, where the data is corrupted by 10% random bad data in mode B1 and 10% random erasure in mode M1.

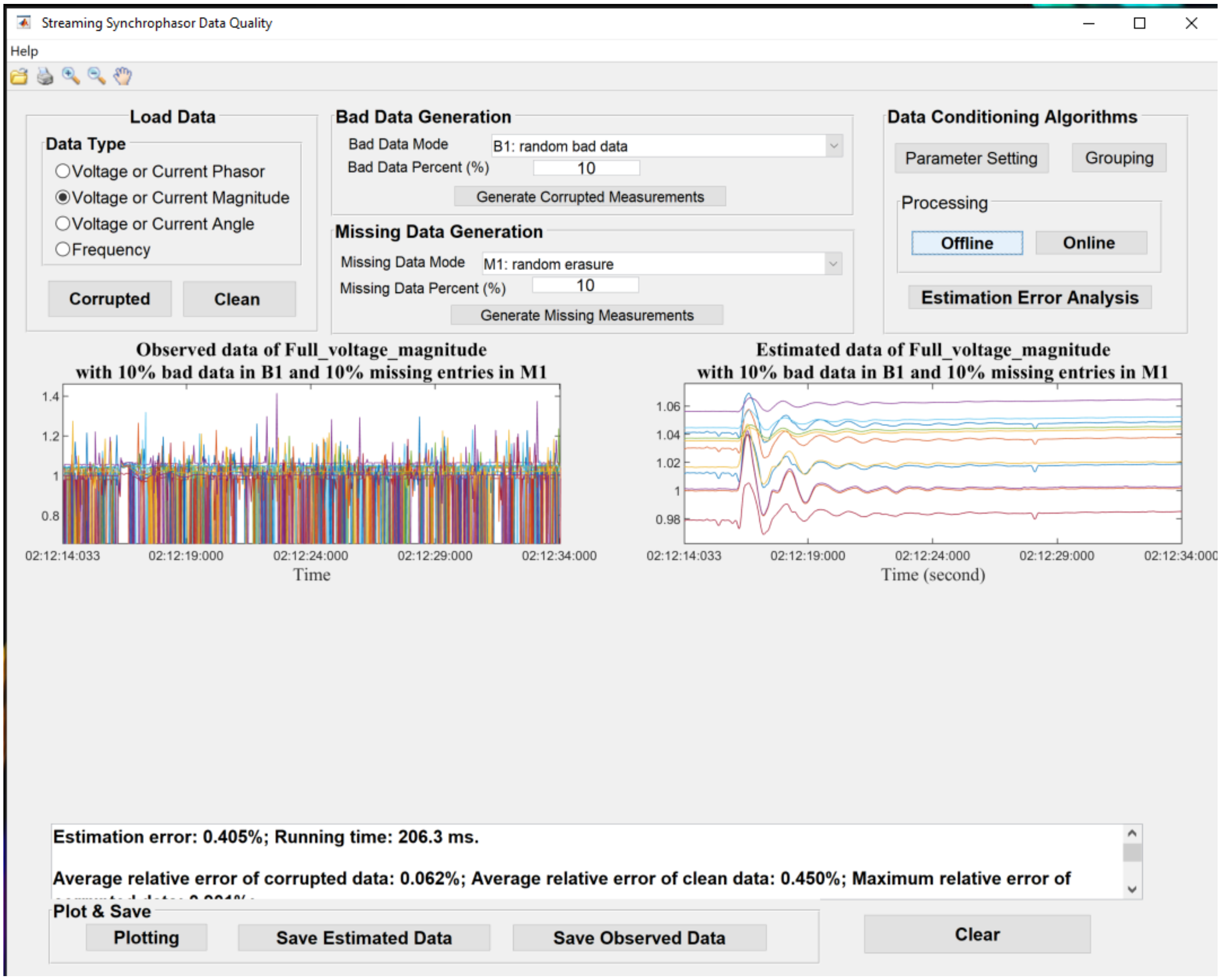

Figure 1-13: Example of the observed and estimated data with 10% missing entries in M1 and 10% bad data in B1 using offline processing

Performance Metrics

Several measurement metrics are used to evaluate the performance of processing and are shown at the bottom of main interface as shown in Fig. 1-13.

- Estimation error (overall and for each group)

  $\frac{||X-\hat{X}||_F}{||X_\Omega||_F}$, where X is the ground-truth data, $\hat{X}$ is the estimation data.

- Ratio of identified untrusted data (overall and for each group)

  Percentage of untrusted data ($|X_\delta - \hat{X}| > 0.003$, where X_δ is the observed data. $\hat{X}$ is the estimated data)

- Average relative error

  $\frac{||X_\Omega-\hat{X}_\Omega||_F}{||X_\Omega||_F}$, where X_Ω is the ground-truth data that are corrupted, $\hat{X}_\Omega$ is the estimation data that are corrupted.

- Average relative error of clean data

  $\frac{||X_{\Omega^c}-\hat{X}_{\Omega^c}||_F}{||X_{\Omega^c}||_F}$, where X_{Ω^c} is the ground-truth data that are not corrupted, $\hat{X}_{\Omega^c}$ is the estimation data that are not corrupted.

- Maximum relative error of corrupted data

  $\max_{ij\in\Omega} \frac{|X_{ij}-\hat{X}_{ij}|}{|X_{ij}|}$, where X_{ij} is the (i,j) th ground-truth data that are corrupted, $\hat{X}_{ij}$ is the (i,j) th estimation data that are corrupted.

- Running time

  time taken to process the data

Figure 1-14: Metric window

Data Conditioning Fail

Both online and offline algorithms have a tolerance of the missing data percentage or bad data percentage. When the input of the missing data percentage or bad data percentage are unreasonably high, i.e. 90% data loss percentage, the algorithm may fail as shown in Fig 1-14. Then, a message box “Data Estimation fail” will appear. The algorithm cannot handle the input % of bad and/or missing data.

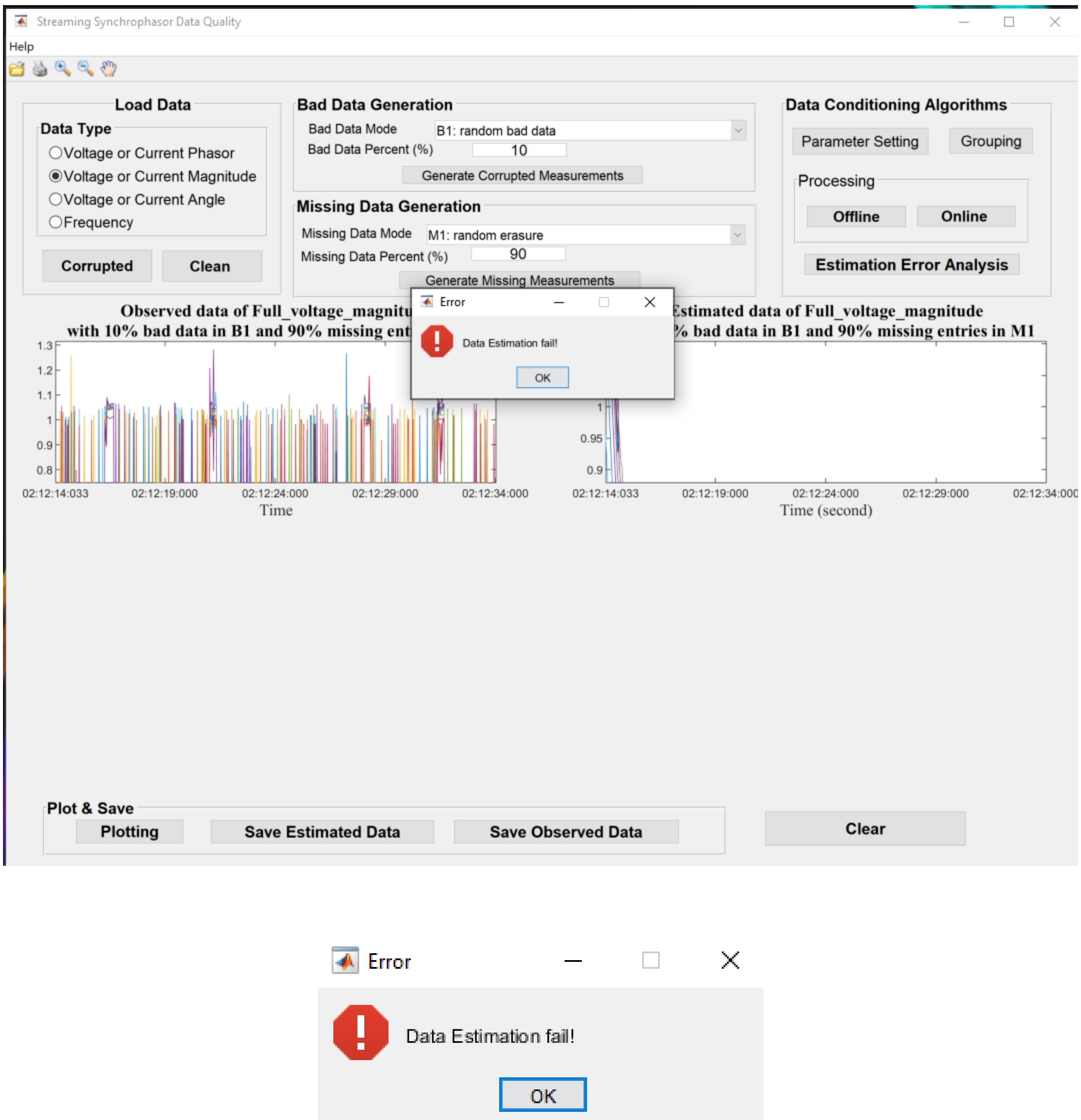

Figure 1-15: Example of data estimation fail

Plotting Selected Channels

Selected channels can be plotted, and the error analysis for these channels can be conducted by clicking the "Plotting" button. A new window will appear. Then the user can select different channels. After clicking the "Plot" button, the figures will display the original observed signal and the estimated signal for the selected channels. The box below will show each channel error analysis separately. One example is shown in Fig 1-15.

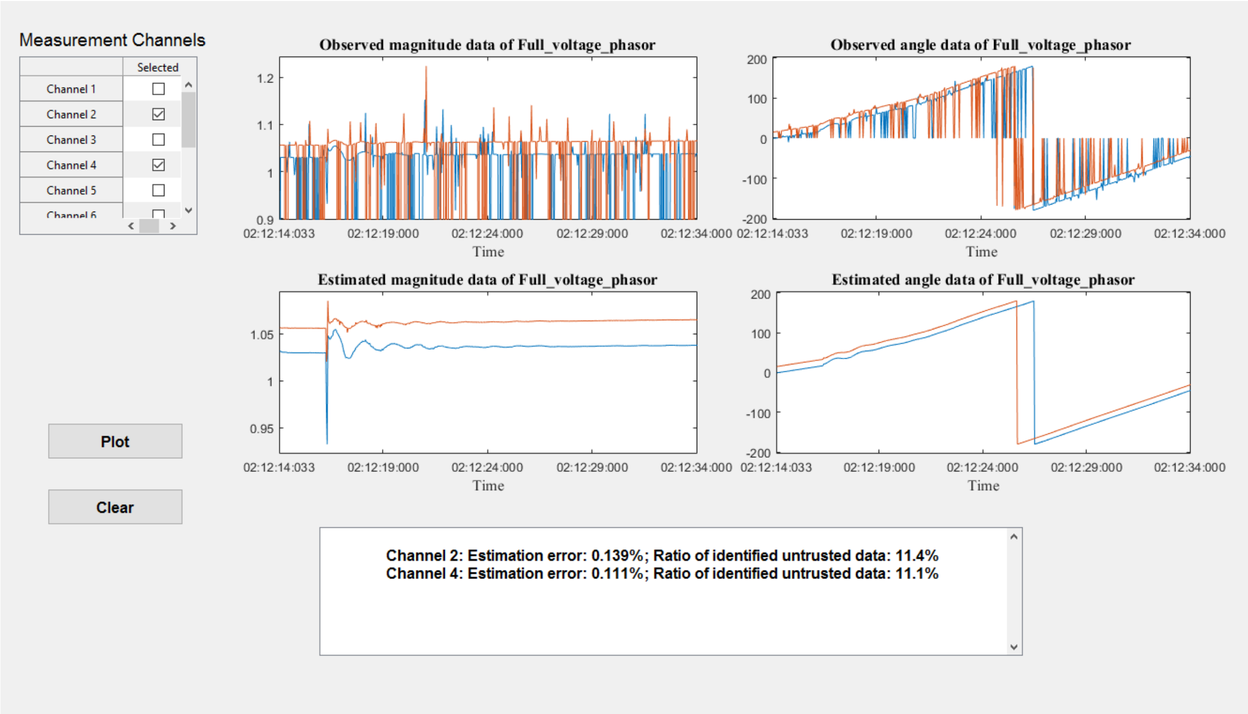

Figure 1-16: Example of plotting selected channels.

Save Observed/Estimated Data

The observed and estimated data can be saved in .xlsx files for further use, when users click the "Save Observed Data" or "Save Estimated Data" button accordingly. If the data are voltage/current phasors, then the corresponding magnitudes and angles will be exported. The format of the output file is described in the Appendix.

In addition to the observed and estimated data, the indicators of trusted data in the imported dataset will be recorded in the file as well, where "1" denotes that the corresponding data in the imported dataset is trusted, while "0" means the corresponding measurement in the original dataset is identified as missing or bad data.

Estimation Error Analysis

With this option, users can conduct multiple experiments to compute the average estimation error with varying percentages of bad and missing data. **The loaded data must be a clean dataset without any bad or missing entries.** Users can select bad/missing data modes, and specify the following parameters to analyze the estimation errors:

- Starting and ending points of bad/missing data percent: they determine the range of bad/missing data percentages in the numerical experiment.
- Step length of bad/missing data percent: together with the range of bad/missing data percentages, they determine the sequence of bad/missing data percentages.
- Number of trials for each missing percent: it determines the number of repetitions for each missing data percentage in the determined sequence.

The default parameters are shown in Fig.1-16.

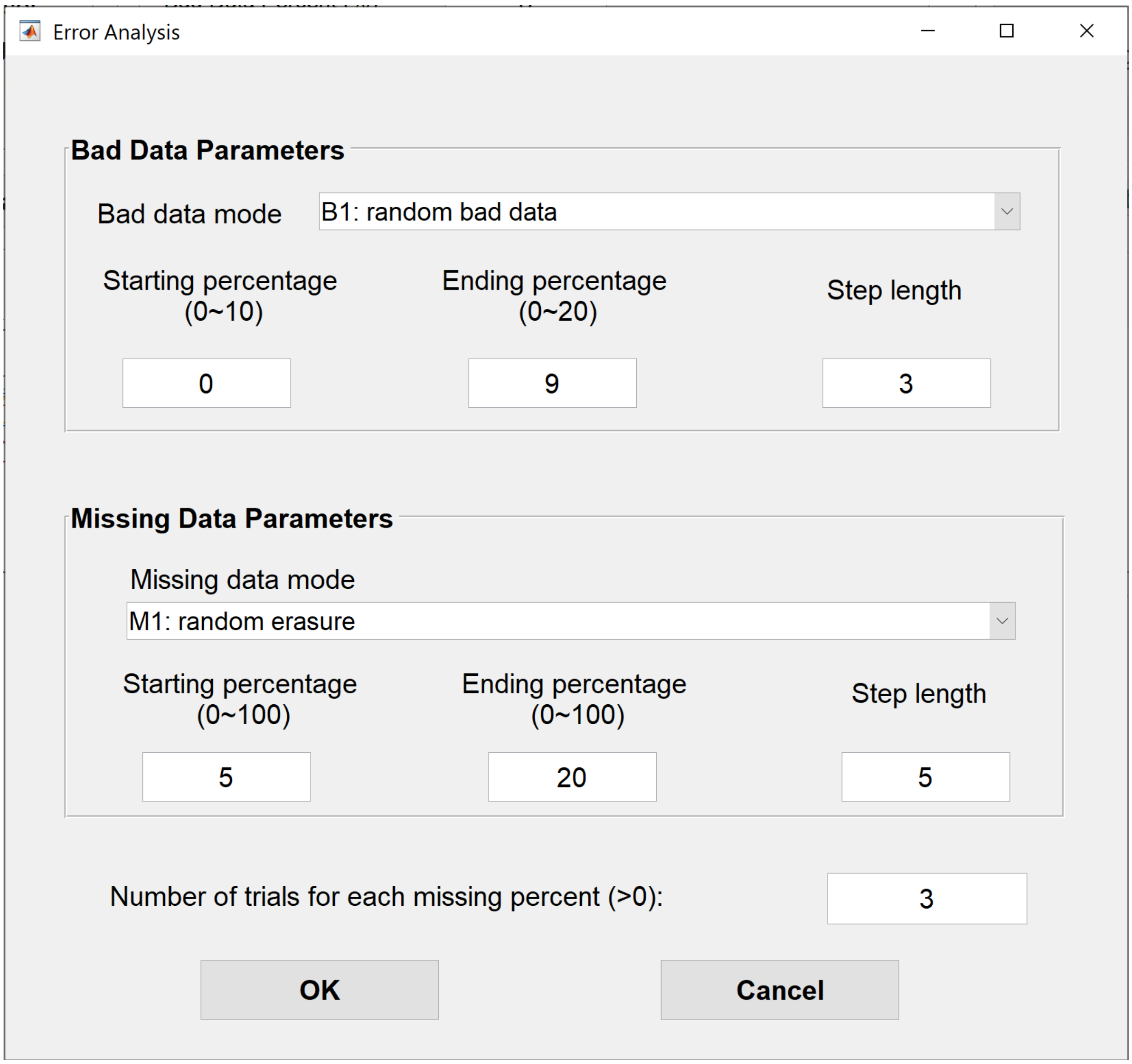

Figure 1-17: Parameter settings to compute the estimation errors

Help Button

By pressing the help menu, the interface is shown as Fig. 1-17. The relevant information is displayed.

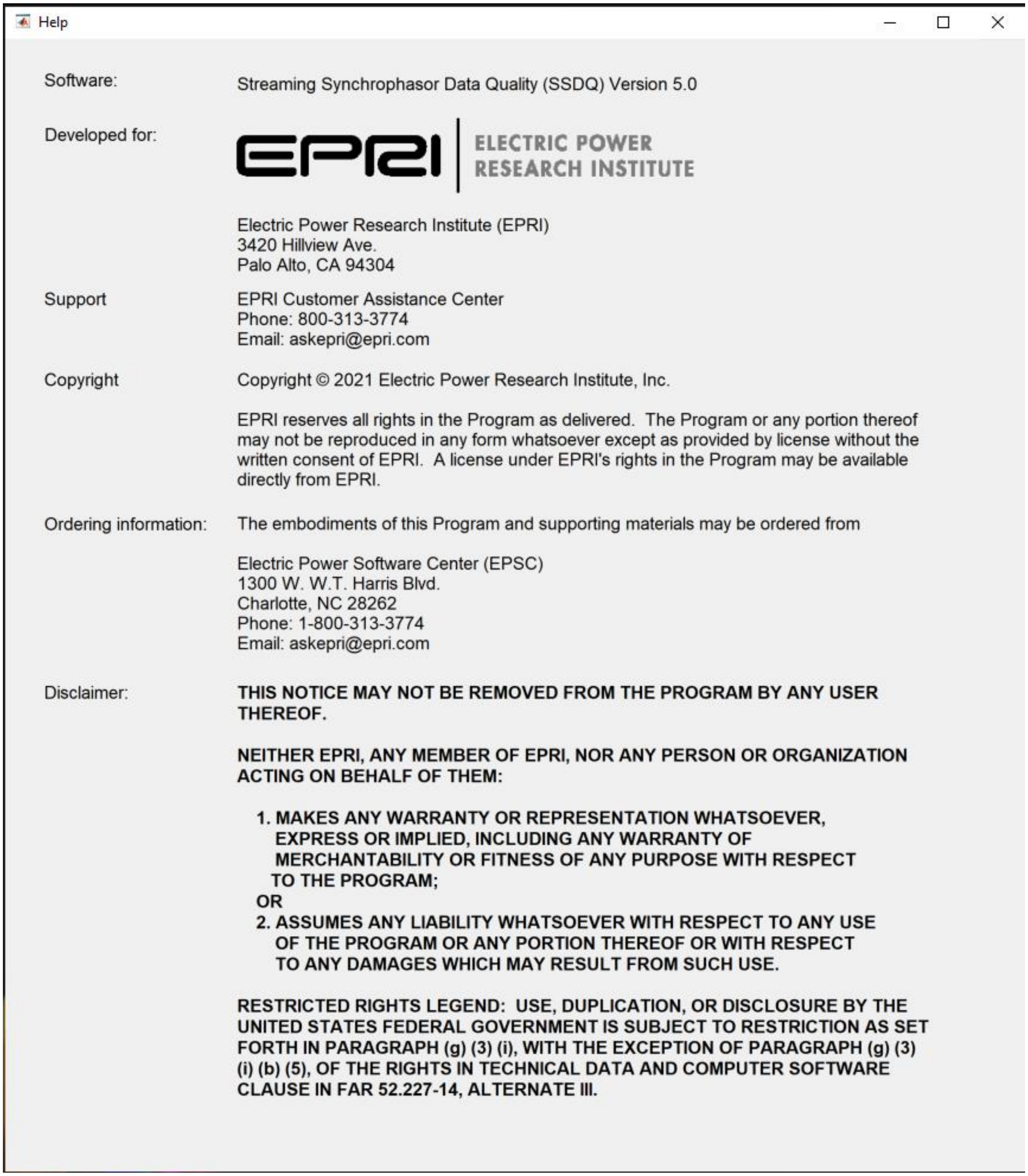

Figure 1-18: Help interface

2
TEST CASE

This chapter provides an example to describe how to use the Streaming Synchrophasor Data Quality (SSDQ) v5.0 software, with a step-by-step description of the test case.

Select the Data Type and Load Data

The first step is to select the data type. Here four selections are available: voltage/current phasor, voltage/current magnitude, voltage/current angle and frequency. Note that for different types of data, the pre-processing is different. Thus, it is necessary to select the right data type. In this example, the "voltage or current phasor" option is selected.

At the main interface, click the button of "Corrupted" or "Clean" to select one data file. The two files recording voltage/current phasors are shown in the red boxes of Fig. 2-1. In this example, we select "Clean" and select the file named "Full_voltage_magnitude". Then an interface shown in Fig. 2-2 pops up to allow users to select the starting and ending time of the data recorded in the file. If no time labels are provided in the file, the row index will be shown instead

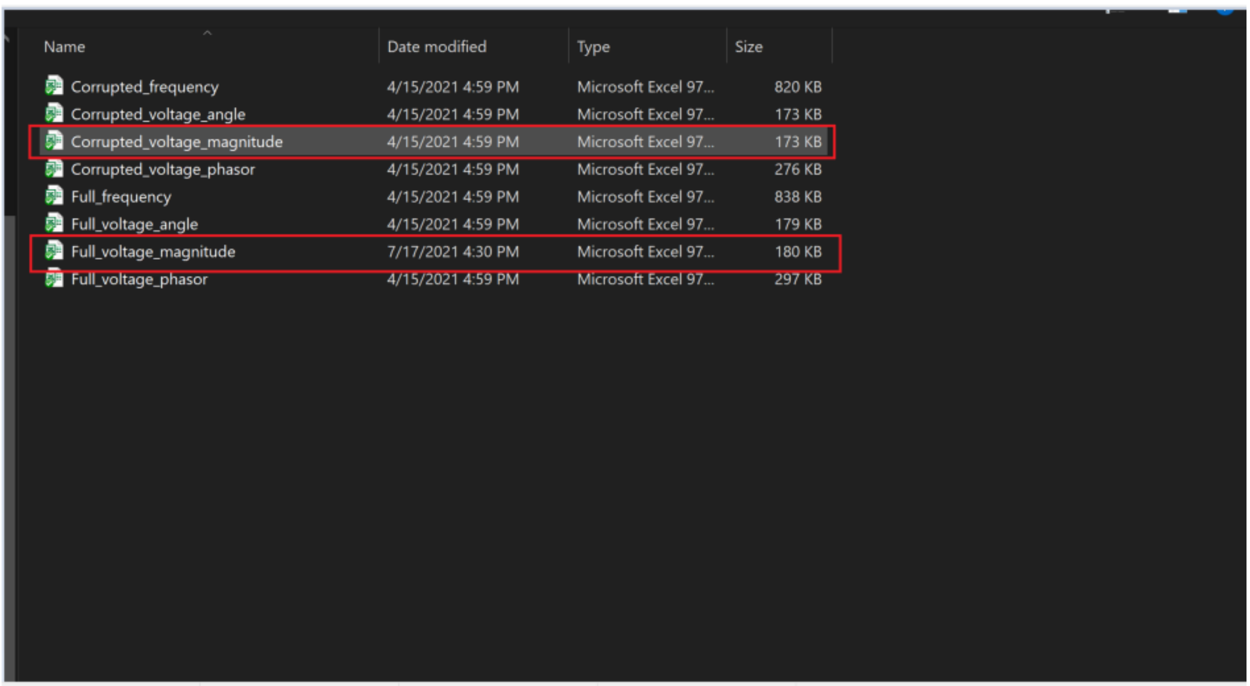

Figure 2-1: Load new dataset

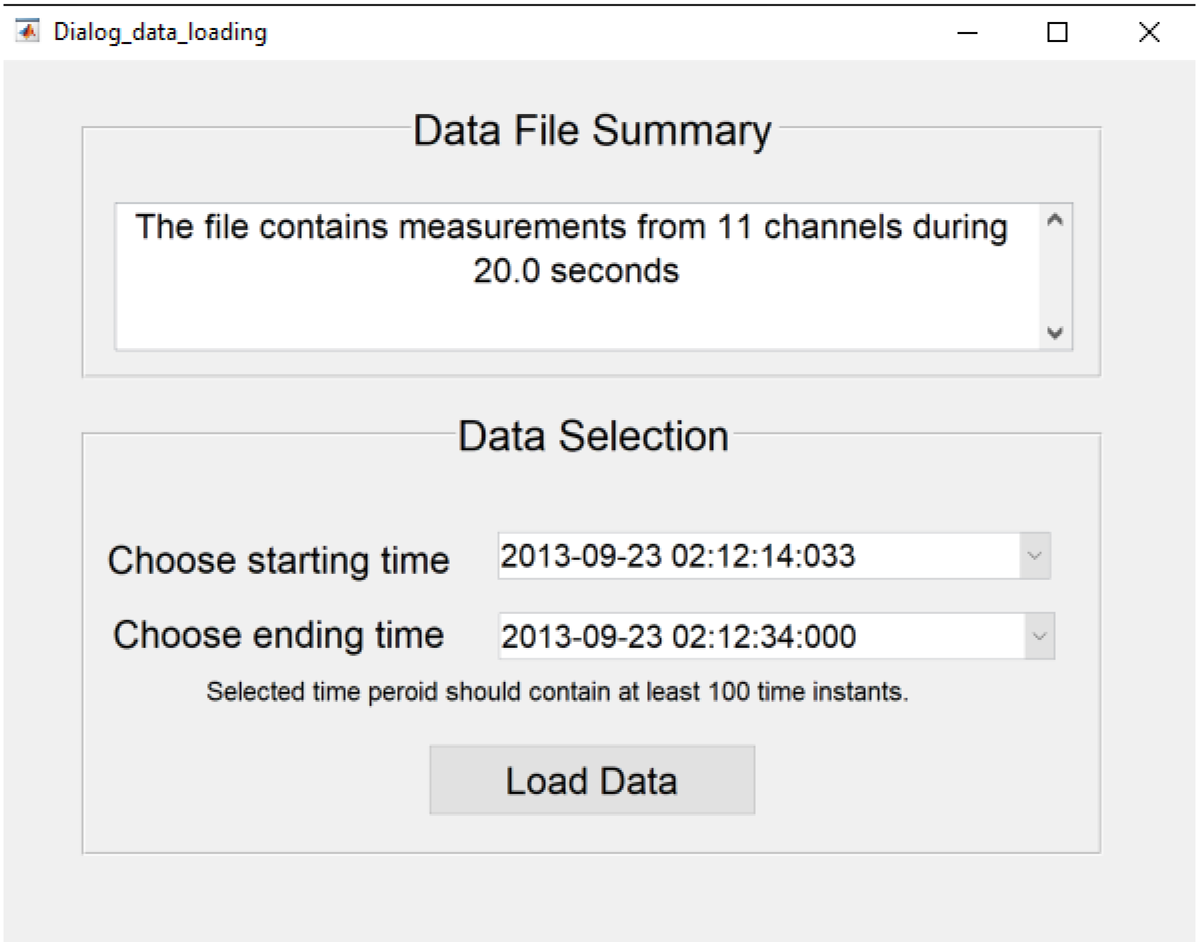

Figure 2-2: Selection of the loaded data

After specifying the starting and ending time, the corresponding measurements during the time interval will be loaded and plotted. The plotted complete measurements are shown in Fig. 2-3.

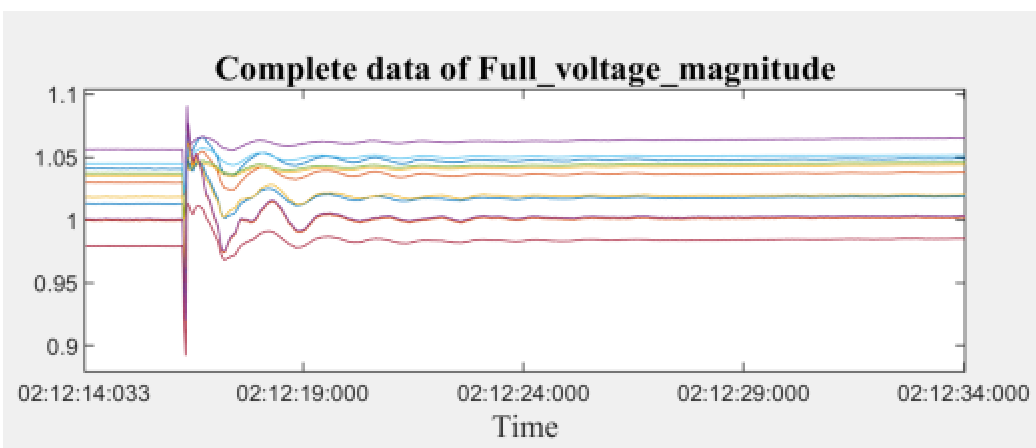

Figure 2-3: Figures of the loaded complete measurements

If we choose to load data with missing/bad entries by clicking “Corrupted” and select file “Corrupted_voltage_magnitude”, then all the corrupted measurements in the file are shown in Fig. 2-4.

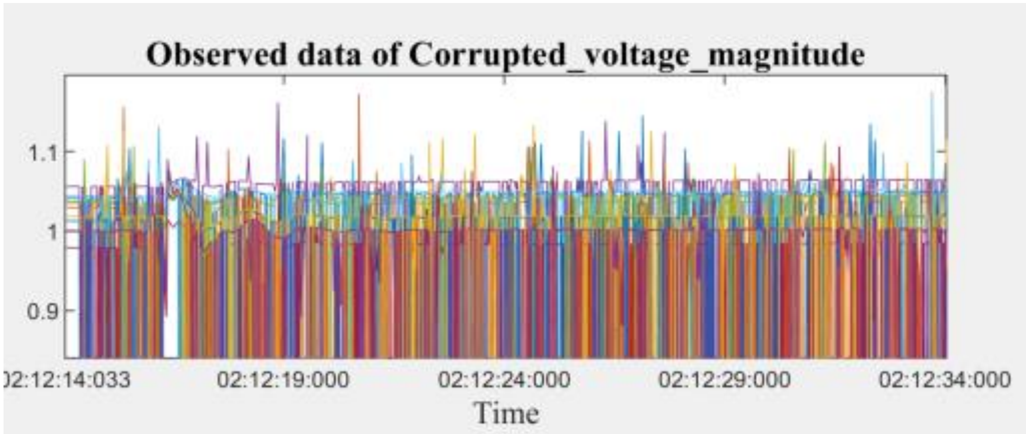

Figure 2-4: Figures of the loaded incomplete measurements

Generate the Corrupted and Missing Data

In the case that measurements that include missing/bad data are selected, the sections “Bad Data Generation” and “Missing Data Generation” are not available, since there exist corrupted and unobserved entries in the dataset.

In the case that a dataset without missing or bad entries is selected and loaded, users can:

1. Select a bad data mode and specify the bad data percentage. After users click the button “Generate Corrupted Measurements”, the corrupted data will be plotted as shown in Fig. 2-5, where in the example, there are 10% bad data in mode B1.
2. Select a missing data mode and specify the missing data percentage. After users click the button “Generate Partially Observed Data”, the partially observed data will be plotted as shown in Fig. 2-6. Here in the example, there are 10% bad data in mode B1 and 10% missing data in mode M1.

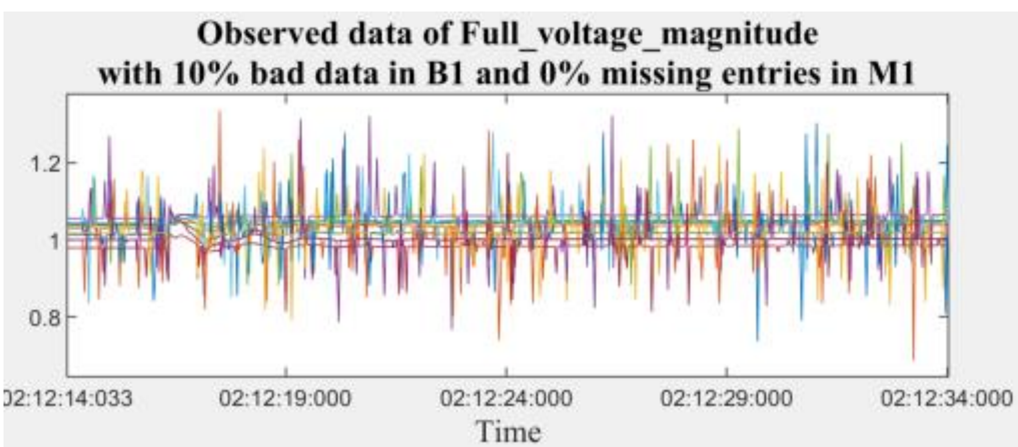

Figure 2-5: Figures of the corrupted data with 10% bad data in mode B1

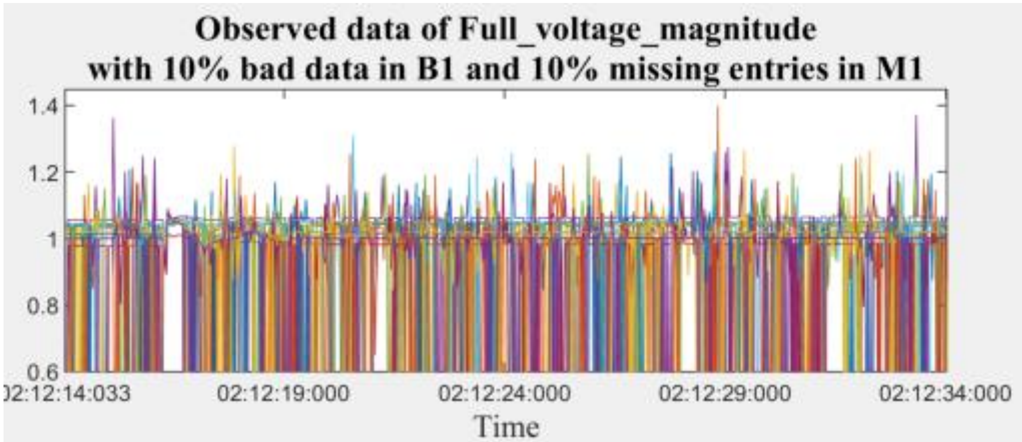

Figure 2-6: Figures of the partially observed data with 10% bad data in B1 and 10% missing data in M1

Parameter Setting of Estimation Methods

The default parameters for the estimation methods are shown in Fig. 2-7. Here we employ the default parameters to estimate the unobserved data in this example.

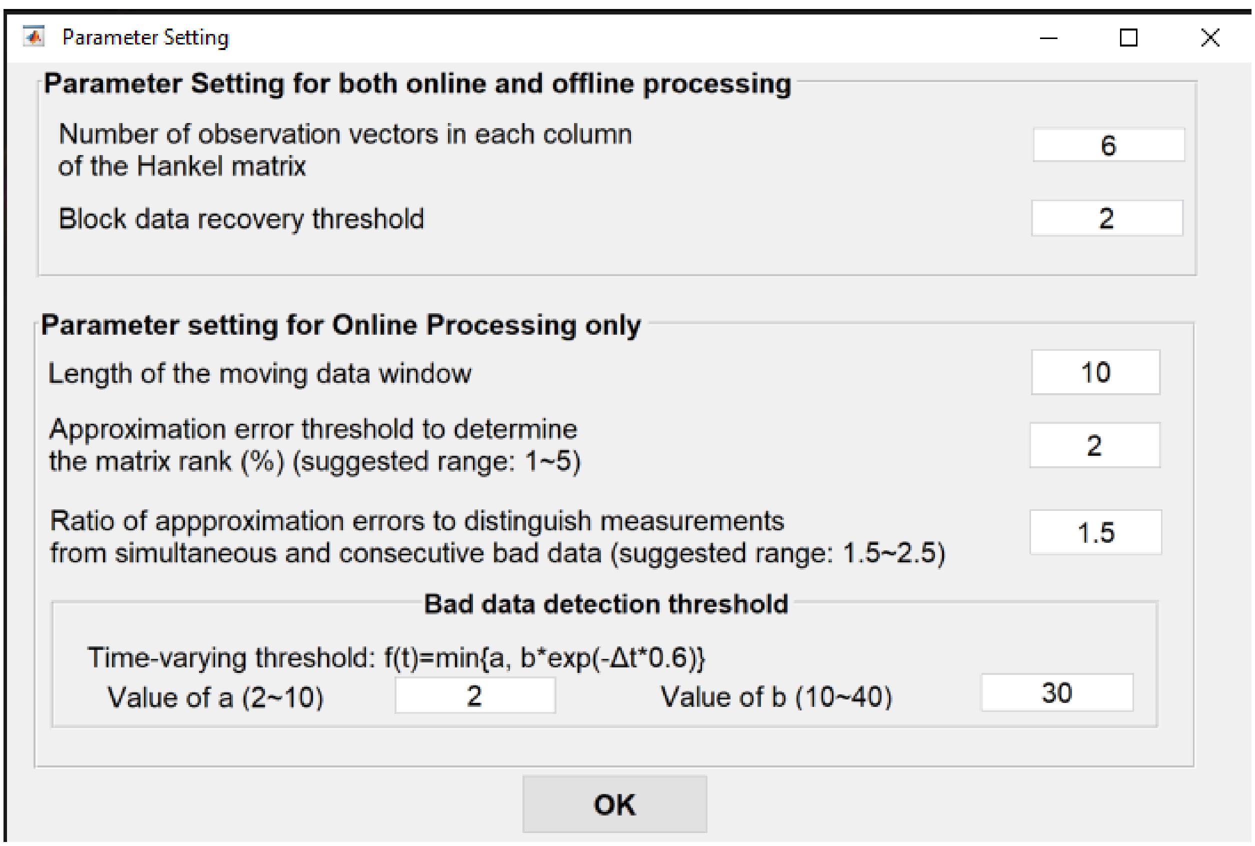

Figure 2-7: The parameter setting of estimation methods

Grouping

The user may choose to perform grouping of channels before processing, both manual selection and auto grouping can be chosen as long as all the channels are assigned into their specific group Fig. 2-8 shows the result of auto grouping with the group size of 2.

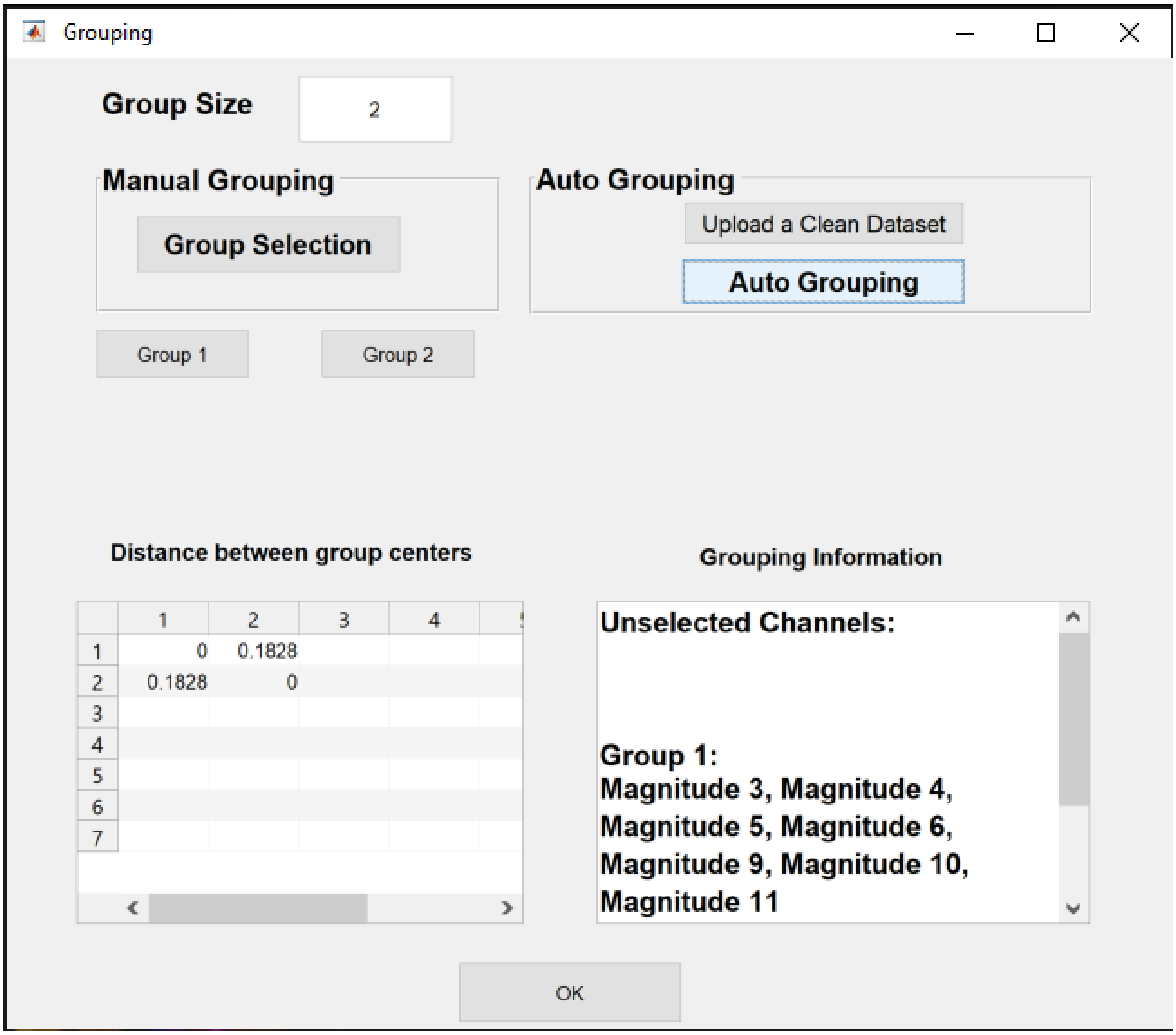

Figure 2-8: Grouping result for "Full_voltage_magnitude" (Auto-grouping)

For Corrupted dataset, if the user would like to choose auto grouping, they need to "Upload a Clean Dataset" first before grouping as shown in Fig. 2-9, because it is almost impossible to generate meaningful grouping results based on corrupted dataset.

For both options, the user is free to switch the channels in different groups by pressing the specific group push-button and changing the channels as shown in Fig. 2-10

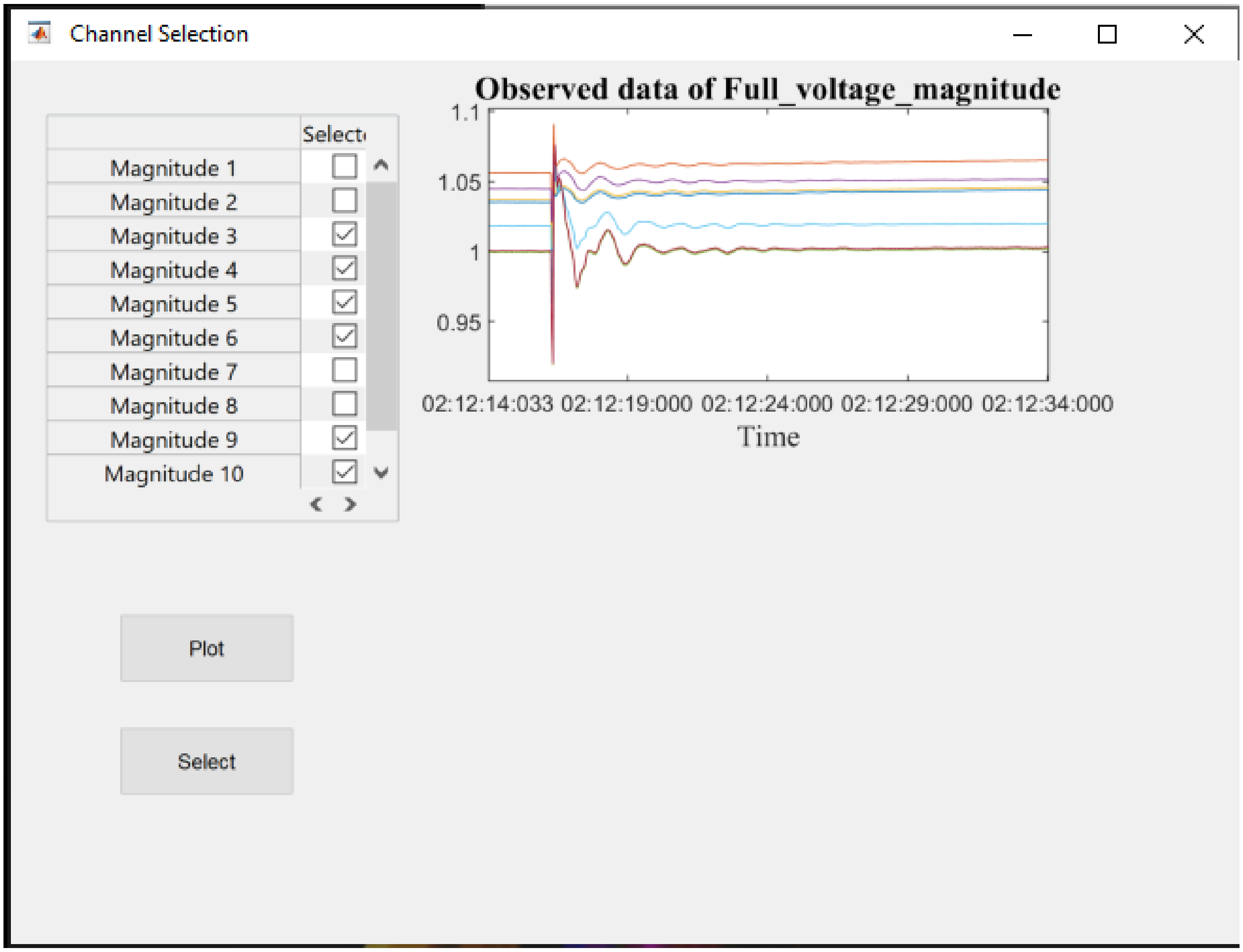

Figure 2-9: channels in Group 2 from grouping result shown in Fig. 2-8

Online Robust Data Estimation

The selected method gets implemented after users press “Online” button within the “Processing” panel. The estimated data are plotted in the interface. The estimated data from the partially observed data with corrupted entries in Fig. 2-6 are shown in Fig. 2-10.

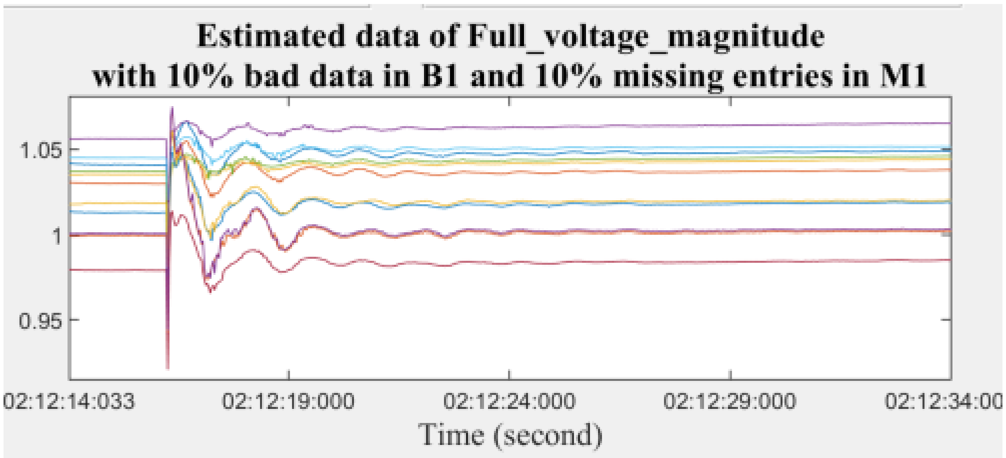

Figure 2-10: The estimated data from data shown in Fig. 2-6 through online processing

For the data shown in Fig. 2-4, the estimated data with estimation from Hankel matrix in a streaming way are shown in Fig. 2-11.

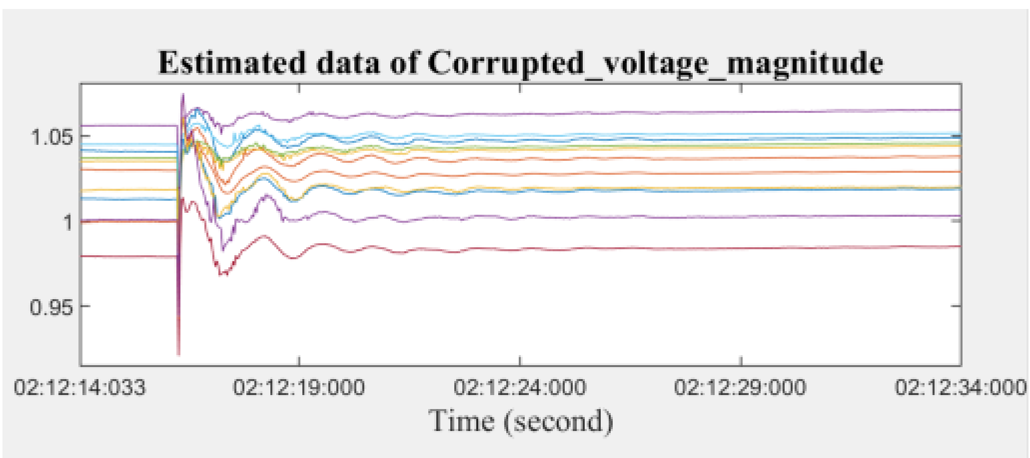

Figure 2-11: The estimated data from data shown in Fig. 2-4 through online processing

Offline Robust Data Estimation

The selected method gets implemented after users press “offline” button within the “Processing” panel. The estimated data are plotted in the interface. The estimated data from the partially observed data with corrupted entries in Fig. 2-6 are shown in Fig. 2-12.

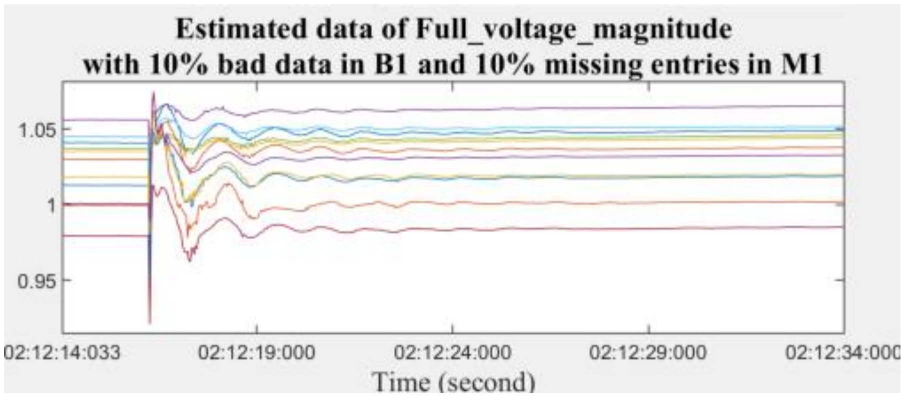

Figure 2-12: The estimated data from data shown in Fig. 2-6 through offline processing

For the data shown in Fig. 2-4, the estimated data with estimation from Hankel matrix in batch mode are shown in Fig. 2-13.

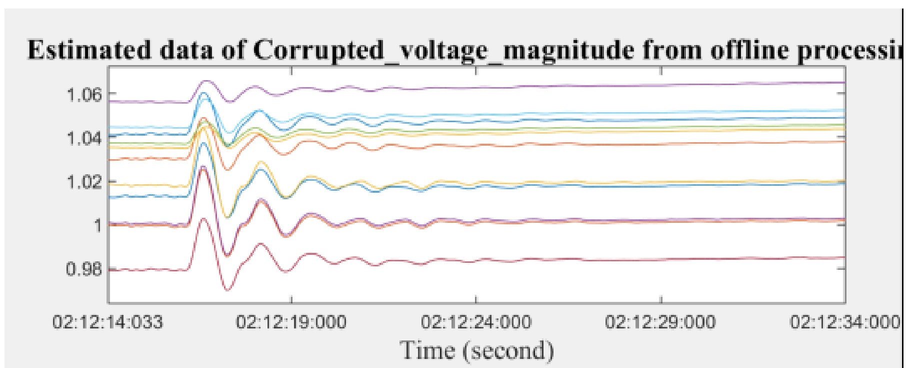

Figure 2-13: The estimated data from data shown in Fig. 2-4 through offline processing

Note that for B3 bad data, the offline algorithm may fail under the default parameter setting. In B3 bad data mode, users need to set the number of observation vectors as 30 and the window length as 40. Other parameters can be kept as default. Fig.2-14 shows the recovery results with 2.5% B3 bad data and 5% B1 bad data.

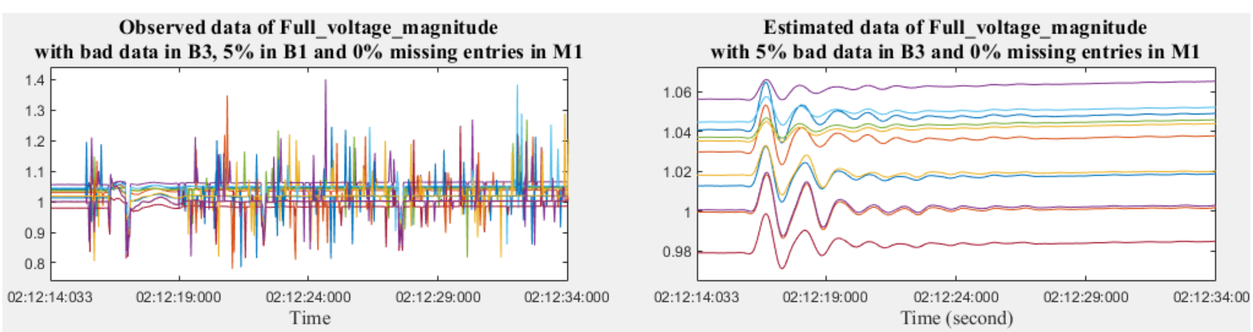

Figure 2-14: The estimated data with 2.5% B3 bad data and 5% B1 bad data

Plotting Selected Channels

For the case with 10% bad data in mode B1 and 10% missing data in mode M1, as an example, after using the "online processing" button to get the estimation results, click the "Plotting" button. A new window will appear. Then the user can select different channels for plotting. For example, Channel 2 and Channel 4 are selected in Fig. 2-15. Then by clicking the "Plot" button, the figures will display the selected channels, shown in Fig. 2-16. The box below, shown in Fig. 2-17, will show each channel error analysis separately.

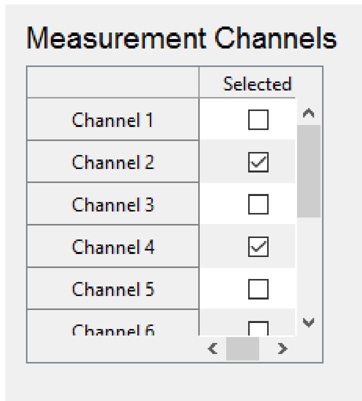

Figure 2-15: Measurement channels selection

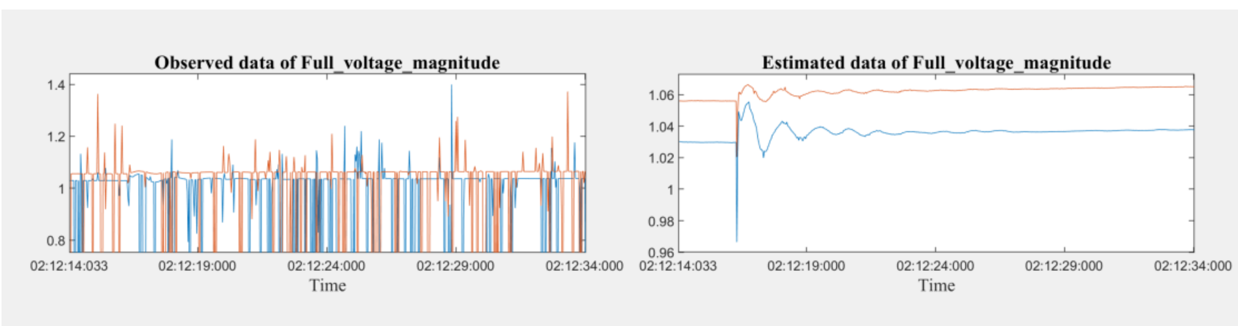

Figure 2-16: Example of plotting observed and estimated data of channel 2 and channel 4.

Channel 2: Estimation error: 0.360%; Ratio of identified untrusted data: 10.4%
Channel 4: Estimation error: 0.139%; Ratio of identified untrusted data: 10.3%

Figure 2-17: Example of error analysis of estimated data of channel 2 and channel 4.

Save Observed/Estimated Dataset in Excel File

The observed and estimated data can be saved in an excel file by pressing the "Save Observed Data" or "Save Estimated Data" button accordingly. The interface is the same as in Fig. 2-1. In this example the output file is named "Estimated_data.xls". Part of the output file is shown in Figs. 2-18 and 2-19.

1	Estimated data of Full_voltage_magnitude											
2	Time instar	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude 11
3	2013-09-2	1.041201	1.03022	1.035116	1.056072	1.037165	1.045297	0.979556	1.013114	1.000274	1.018413	1.001473
4	2013-09-2	1.041076	1.030101	1.03507	1.056026	1.037119	1.045218	0.979395	1.012954	0.99979	1.018129	1.000968
5	2013-09-2	1.040984	1.030015	1.03504	1.055996	1.037088	1.045173	0.979302	1.012863	0.999613	1.018002	1.000783
6	2013-09-2	1.040906	1.029945	1.035018	1.055972	1.037063	1.045136	0.979235	1.012799	0.999497	1.017918	1.000689
7	2013-09-2	1.040841	1.029889	1.035002	1.05595	1.037041	1.045103	0.9792	1.012766	0.999431	1.017875	1.000634
8	2013-09-2	1.040809	1.029866	1.034991	1.055935	1.037025	1.045077	0.97922	1.01279	0.999402	1.017855	1.000611
9	2013-09-2	1.040803	1.029861	1.035004	1.055947	1.037037	1.04508	0.979208	1.012778	0.9994	1.017865	1.00061
10	2013-09-2	1.0408	1.029859	1.035012	1.055955	1.037046	1.045085	0.979208	1.012779	0.999374	1.017856	1.000584
11	2013-09-2	1.040827	1.029887	1.035028	1.055972	1.037062	1.045097	0.979241	1.012813	0.999387	1.017874	1.000597
12	2013-09-2	1.040845	1.029906	1.03505	1.055993	1.037083	1.045105	0.979261	1.012834	0.99935	1.017874	1.00056
13	2013-09-2	1.040935	1.029773	1.035109	1.05603	1.036967	1.044709	0.979111	1.012648	0.999338	1.018143	1.000472
14	2013-09-2	1.040964	1.029838	1.035306	1.056175	1.037084	1.04485	0.979205	1.012751	0.999319	1.018216	1.000507
15	2013-09-2	1.04095	1.029851	1.035181	1.056057	1.037061	1.044778	0.979189	1.012752	0.99937	1.018406	1.000565
16	2013-09-2	1.041067	1.029865	1.035167	1.056108	1.037048	1.044946	0.97934	1.012916	0.999389	1.018321	1.000925
17	2013-09-2	1.041054	1.029881	1.035206	1.05624	1.037103	1.044824	0.97922	1.01281	0.999585	1.018246	1.000794
18	2013-09-2	1.040844	1.029801	1.035267	1.056196	1.037048	1.044859	0.97927	1.012728	0.999707	1.018476	1.000686
19	2013-09-2	1.041004	1.029802	1.035179	1.05597	1.037026	1.044797	0.979198	1.012819	0.999656	1.01821	1.000602
20	2013-09-2	1.040925	1.029826	1.03525	1.056174	1.037027	1.044844	0.979146	1.012781	0.999589	1.018221	1.000579
21	2013-09-2	1.040988	1.02988	1.035253	1.056088	1.037052	1.044894	0.979263	1.012769	0.999623	1.018256	1.000755
22	2013-09-2	1.040883	1.02977	1.035239	1.056135	1.037015	1.04484	0.979256	1.012781	0.999556	1.01829	1.000526
23	2013-09-2	1.040889	1.029776	1.035177	1.056091	1.037013	1.044826	0.979127	1.012668	0.999557	1.018178	1.000569
24	2013-09-2	1.040818	1.029795	1.035179	1.056162	1.03698	1.044849	0.979159	1.012726	0.999615	1.018192	1.000567
25	2013-09-2	1.040826	1.029755	1.035205	1.05628	1.036989	1.044848	0.979225	1.01266	0.999424	1.01819	1.000686
26	2013-09-2	1.040743	1.029731	1.035136	1.05608	1.037038	1.044832	0.979161	1.012617	0.999526	1.018118	1.000557
27	2013-09-2	1.040781	1.02972	1.035147	1.056042	1.037052	1.044801	0.979116	1.012591	0.999411	1.018068	1.000448
28	2013-09-2	1.040804	1.029659	1.035125	1.056179	1.037038	1.044871	0.979096	1.012597	0.999317	1.018038	1.000331

Figure 2-14: Example of the output data in the first sheet of the file

	A	B	C	D	E	F	G	H	I	J	K	L
1	Trusted data indicator of Full_voltage_magnitude											
2	Time instar	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude	Magnitude 11
3	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
4	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
5	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
6	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
7	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
8	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
9	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
10	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
11	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
12	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
13	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
14	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
15	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
16	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
17	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
18	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
19	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
20	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
21	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
22	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
23	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
24	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
25	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
26	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
27	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
28	2013-09-2:	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Sheet1 Sheet2

Figure 2-19: Example of the output indicators of trusted data in the second sheet of the file

Estimation Error Analysis

In the example, we still choose data type "Voltage or Current Magnitude", and select the data file "Full_voltage_magnitude". Note that to analyze the estimation errors with varying bad data and missing data percentages, users can only load datasets without missing/bad entries.

With the default parameters, the interface after the selected data is loaded, is shown as follows:

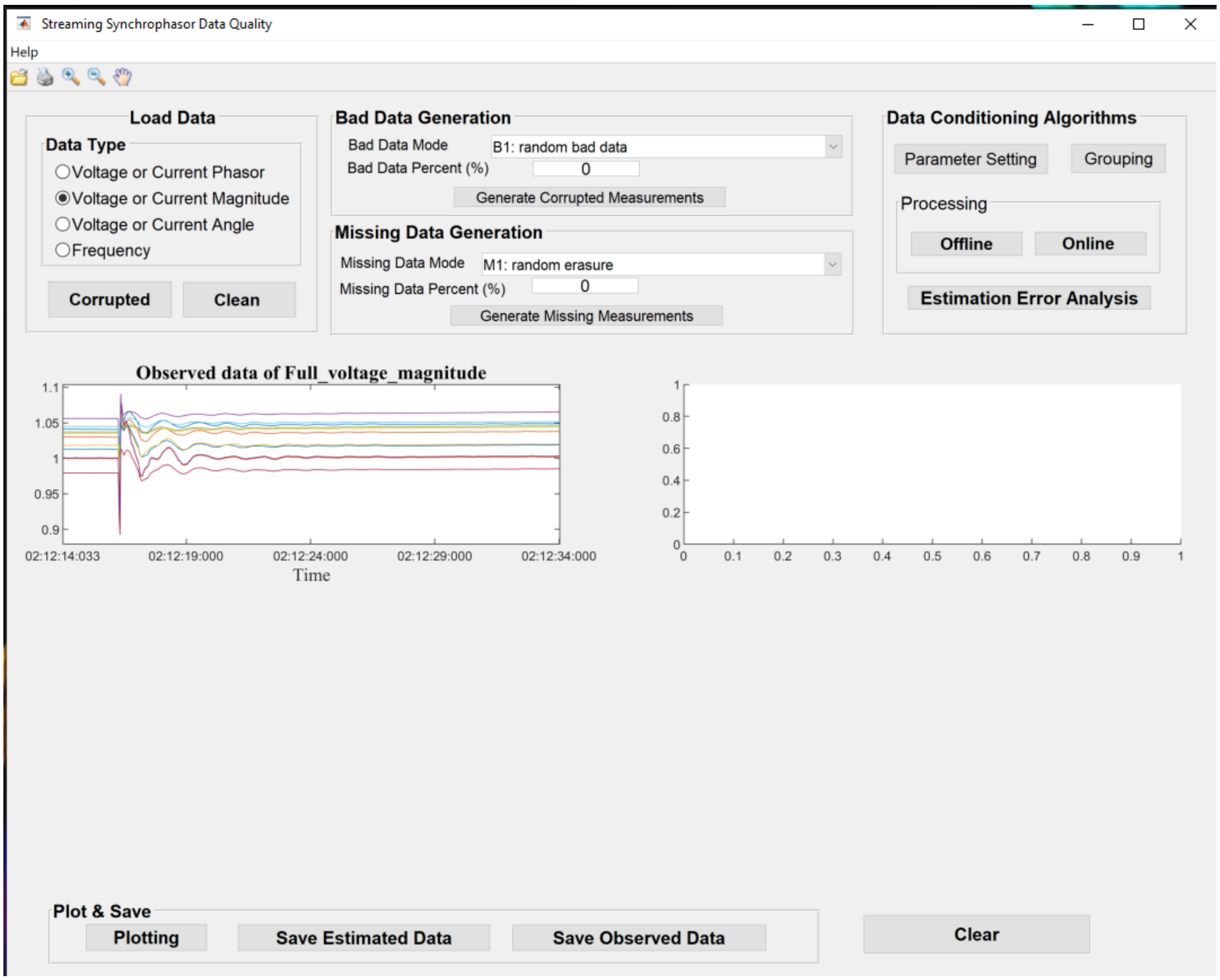

Figure 2-20: Example of estimation error analysis

After clicking the "Estimation Error Analysis" button, the following interface shows up.

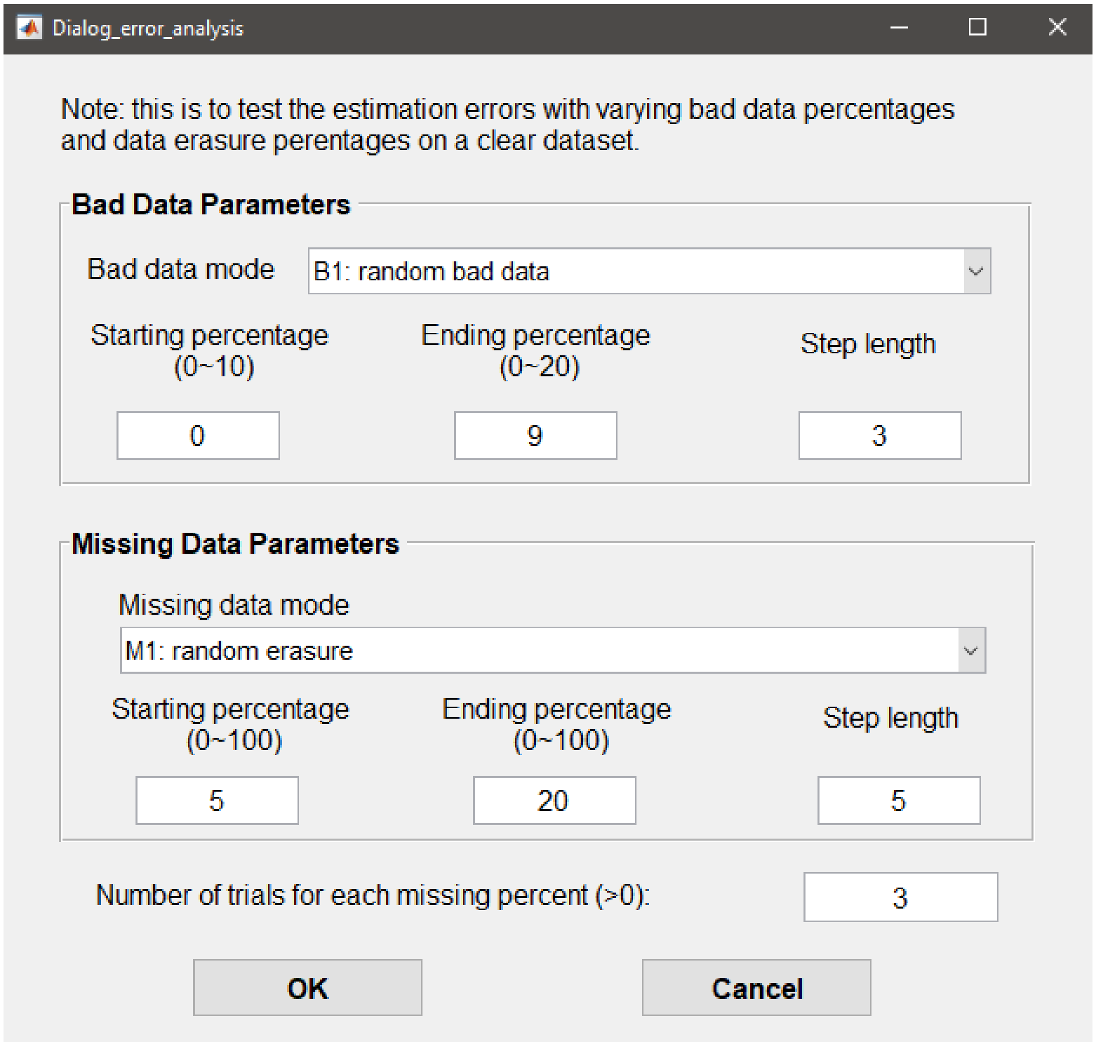

Figure 2-21: Parameter setting of estimation error analysis

With the default setting shown in Fig. 2-21, after clicking button "OK", a progress bar shows up indicating the current state of the algorithm implementation, as shown in Fig. 2-22.

Figure 2-22: The progress bar

The result is shown as follows:

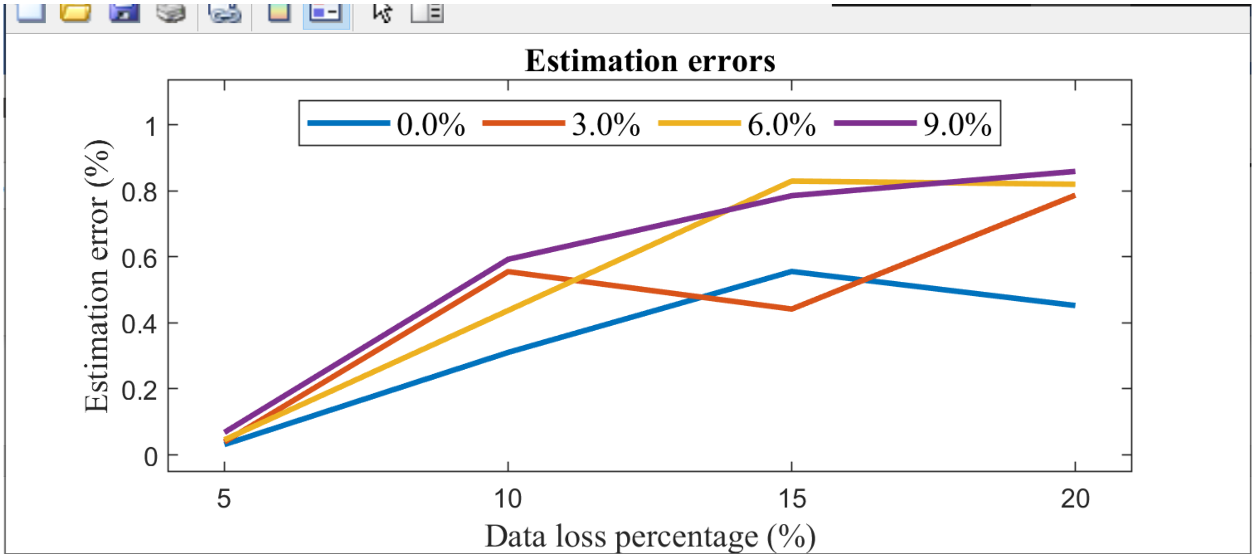

Figure 2-23: The average and maximum estimation errors in the example

Clear Function

After users click button “Clear”, no result will be shown in the interface.

3
REFERENCES

[1] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, “Model-less data quality improvement of streaming synchrophasor measurements by exploiting the low-rank Hankel structure,” accepted to *IEEE Trans. Power Systems*, 2017.

[2] S. Zhang, Y. Hao, M. Wang, and J. H. Chow, “Multi-Channel Hankel Matrix Completion through Nonconvex Optimization,’’ accepted to *IEEE Journal of Selected Topics in Signal Processing*, 2018.

[3] Zhang, Shuai, and Meng Wang. "Correction of Corrupted Columns Through Fast Robust Hankel Matrix Completion." IEEE Transactions on Signal Processing 67.10 (2019): 2580-2594.

A
APPENDIX

Input Data Format

The data in an excel file should follow a particular format. Here the default format is as follows:

- If the data is voltage/current phasor, then the format is "Magnitude of channel i; Angle of channel i; Magnitude of channel $i + 1$; Angle of channel $i + 1$ ..."
- For other cases, the format is "Measurement of channel i; Measurement of channel $\hat{\imath} + 1$ ..."
- If there are multiple sheets in the excel file, then the name of the selected sheet serves as the name of the loaded data; otherwise, the name of the loaded data is the name of the selected excel file.

Two examples are shown as follows, where the first one is part of the recorded data of voltage phasors, and the second one is part of the recorded frequency data. Each red box denotes the measurements from one PMU channel.

Time instant	Magnitude 1	Angle 1	Magnitude 2	Angle 2	Magnitude 3	Angle 3
1	348.78437	83.16725	350.3232107	81.35308	352.3998011	91.10499
2	348.764452	83.12578	350.2982865	81.31098	352.391314	91.06409
3	348.752968	83.08926	350.3029803	81.29484	352.3861698	91.02692
4	348.743858	83.05704	350.2604239	81.2815	352.3772151	90.99288
5	348.742957	83.02337	350.2490096	81.2468	352.3687973	90.95759
6	348.758528	82.98516	350.3245617	81.19529	352.3697673	90.91849

(a) Example of the recorded voltage phasors

Time instant	Freq 1	Freq 2	Freq 3	Freq 4	Freq 5
1	59.99707	59.99949	59.99699	59.997	59.99654
2	59.99715	59.99965	59.99709	59.99707	59.99666
3	59.99719	59.99975	59.99714	59.99713	59.99683
4	59.9973	59.99986	59.99722	59.99723	59.99722
5	59.99749	59.99996	59.99737	59.9974	59.99754

(b) Example of the recorded frequency data

Figure A-3-1: Examples of the format of excel data files

Output Data Format

The output data recorded in Excel files follow the same format as shown above.

Export Control Restrictions

Access to and use of EPRI Intellectual Property is granted with the specific understanding and requirement that responsibility for ensuring full compliance with all applicable U.S. and foreign export laws and regulations is being undertaken by you and your company. This includes an obligation to ensure that any individual receiving access hereunder who is not a U.S. citizen or permanent U.S. resident is permitted access under applicable U.S. and foreign export laws and regulations. In the event you are uncertain whether you or your company may lawfully obtain access to this EPRI Intellectual Property, you acknowledge that it is your obligation to consult with your company's legal counsel to determine whether this access is lawful. Although EPRI may make available on a case-by-case basis an informal assessment of the applicable U.S. export classification for specific EPRI Intellectual Property, you and your company acknowledge that this assessment is solely for informational purposes and not for reliance purposes. You and your company acknowledge that it is still the obligation of you and your company to make your own assessment of the applicable U.S. export classification and ensure compliance accordingly. You and your company understand and acknowledge your obligations to make a prompt report to EPRI and the appropriate authorities regarding any access to or use of EPRI Intellectual Property hereunder that may be in violation of applicable U.S. or foreign export laws or regulations.

The Electric Power Research Institute, Inc. (EPRI, www.epri.com) conducts research and development relating to the generation, delivery and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, affordability, health, safety and the environment. EPRI also provides technology, policy and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI's members represent approximately 90 percent of the electricity generated and delivered in the United States, and international participation extends to more than 30 countries. EPRI's principal offices and laboratories are located in Palo Alto, Calif.; Charlotte, N.C.; Knoxville, Tenn.; and Lenox, Mass.

Together…Shaping the Future of Electricity

Electric Power Research Institute
3420 Hillview Avenue, Palo Alto, California 94304-1338 • PO Box 10412, Palo Alto, California 94303-0813 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com