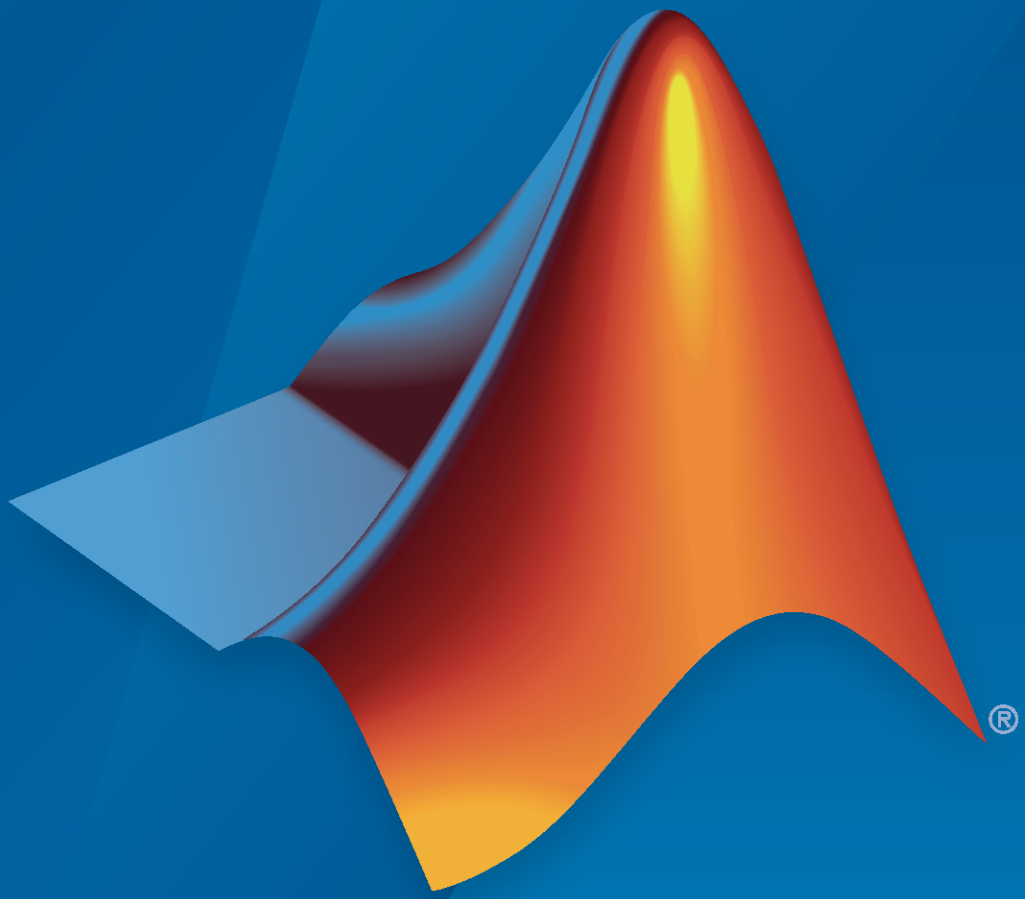
Predictive Maintenance Toolbox™

Getting Started Guide



MATLAB®

R2020b

MathWorks®

How to Contact MathWorks

Latest news: `www.mathworks.com`

Sales and services: `www.mathworks.com/sales_and_services`

User community: `www.mathworks.com/matlabcentral`

Technical support: `www.mathworks.com/support/contact_us`

Phone: 508-647-7000

The MathWorks, Inc.
1 Apple Hill Drive
Natick, MA 01760-2098

Predictive Maintenance Toolbox™ Getting Started Guide

© COPYRIGHT 2018–2020 by The MathWorks, Inc.

The software described in this document is furnished under a license agreement. The software may be used or copied only under the terms of the license agreement. No part of this manual may be photocopied or reproduced in any form without prior written consent from The MathWorks, Inc.

FEDERAL ACQUISITION: This provision applies to all acquisitions of the Program and Documentation by, for, or through the federal government of the United States. By accepting delivery of the Program or Documentation, the government hereby agrees that this software or documentation qualifies as commercial computer software or commercial computer software documentation as such terms are used or defined in FAR 12.212, DFARS Part 227.72, and DFARS 252.227-7014. Accordingly, the terms and conditions of this Agreement and only those rights specified in this Agreement, shall pertain to and govern the use, modification, reproduction, release, performance, display, and disclosure of the Program and Documentation by the federal government (or other entity acquiring for or through the federal government) and shall supersede any conflicting contractual terms or conditions. If this License fails to meet the government's needs or is inconsistent in any respect with federal procurement law, the government agrees to return the Program and Documentation, unused, to The MathWorks, Inc.

Trademarks

MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See `www.mathworks.com/trademarks` for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

Patents

MathWorks products are protected by one or more U.S. patents. Please see `www.mathworks.com/patents` for more information.

Revision History

March 2018	Online only	New for Version 1.0 (Release 2018a)
September 2018	Online only	Revised for Version 1.1 (Release 2018b)
March 2019	Online only	Revised for Version 2.0 (Release 2019a)
September 2019	Online only	Revised for Version 2.1 (Release 2019b)
March 2020	Online only	Revised for Version 2.2 (Release 2020a)
September 2020	Online only	Revised for Version 2.2.1 (Release 2020b)

Contents

1 Introduction to Predictive Maintenance Toolbox

Predictive Maintenance Toolbox Product Description **1-2**

Designing Algorithms for Condition Monitoring and Predictive Maintenance **1-3**
- Algorithms for Condition Monitoring and Prognostics **1-3**
- Workflows for Algorithm Development **1-4**
- Acquire Data **1-4**
- Preprocess Data **1-5**
- Identify Condition Indicators **1-5**
- Train Detection or Prediction Model **1-6**
- Deploy and Integrate **1-7**

2 Getting Started with Diagnostic Feature Designer

Identify Condition Indicators for Predictive Maintenance Algorithm Design **2-2**

Import and Visualize Ensemble Data in Diagnostic Feature Designer . . . **2-5**
- Load Transmission Model Data **2-5**
- Import Data into Diagnostic Feature Designer **2-6**
- Import Data **2-6**
- Visualize Data **2-9**
- Next Steps **2-12**

Process Data and Explore Features in Diagnostic Feature Designer . . . **2-14**
- Perform Time-Synchronous Averaging **2-14**
- Compute a Power Spectrum **2-17**
- Generate Features **2-19**
- Next Steps **2-30**

Rank and Export Features in Diagnostic Feature Designer **2-31**
- Rank Features **2-31**
- Choose an Alternative Ranking Algorithm **2-32**
- Delete a Set of Rankings **2-35**
- Export Features to the MATLAB Workspace **2-36**

Introduction to Predictive Maintenance Toolbox

- “Predictive Maintenance Toolbox Product Description” on page 1-2
- “Designing Algorithms for Condition Monitoring and Predictive Maintenance” on page 1-3

Predictive Maintenance Toolbox Product Description

Design and test condition monitoring and predictive maintenance algorithms

Predictive Maintenance Toolbox lets you label data, design condition indicators, and estimate the remaining useful life (RUL) of a machine.

The toolbox provides functions and an interactive app for exploring, extracting, and ranking features using data-based and model-based techniques, including statistical, spectral, and time-series analysis. You can monitor the health of rotating machines such as bearings and gearboxes by extracting features from vibration data using frequency and time-frequency methods. To estimate a machine's time to failure, you can use survival, similarity, and trend-based models to predict the RUL.

You can analyze and label sensor data imported from local files, cloud storage, and distributed file systems. You can also label simulated failure data generated from Simulink® models. The toolbox includes reference examples for motors, gearboxes, batteries, and other machines that can be reused for developing custom predictive maintenance and condition monitoring algorithms.

Designing Algorithms for Condition Monitoring and Predictive Maintenance

Predictive maintenance allows equipment users and manufacturers to assess the working condition of machinery, diagnose faults, or estimate when the next equipment failure is likely to occur. When you can diagnose or predict failures, you can plan maintenance in advance, better manage inventory, reduce downtime, and increase operational efficiency.

Developing a predictive maintenance program requires a well-designed strategy to assess the working condition of the machinery and detect incipient faults in a timely manner. Doing so requires effective use of both available sensor measurements and your knowledge of the system. You must take many factors into consideration, including:

- The observed sources of faults and their relative frequency. Such sources can be the core components of the machine (such as impeller blades and flow valves in a pump), its actuators (such as an electric motor), or its various sensors (such as accelerometers and flow meters).
- Availability of process measurements through sensors. The number, type and location of sensors, and their reliability and redundancies all affect both algorithm development and cost.
- How various sources of faults translate to observed symptoms. Such cause-effect analysis can require extensive processing of data from the available sensors.
- Physical knowledge about the system dynamics. This knowledge might come from mathematical modeling of the system and its faults and from the insights of domain experts. Understanding system dynamics involves detailed knowledge of relationships among various signals from the machinery (such as input-output relationships among the actuators and sensors), the machine operating range, and the nature of the measurements (for example, periodic, constant or stochastic).
- The ultimate maintenance goal, such as fault recovery or development of a maintenance schedule.

Algorithms for Condition Monitoring and Prognostics

A predictive maintenance program uses condition monitoring and prognostics algorithms to analyze data measured from the system in operation.

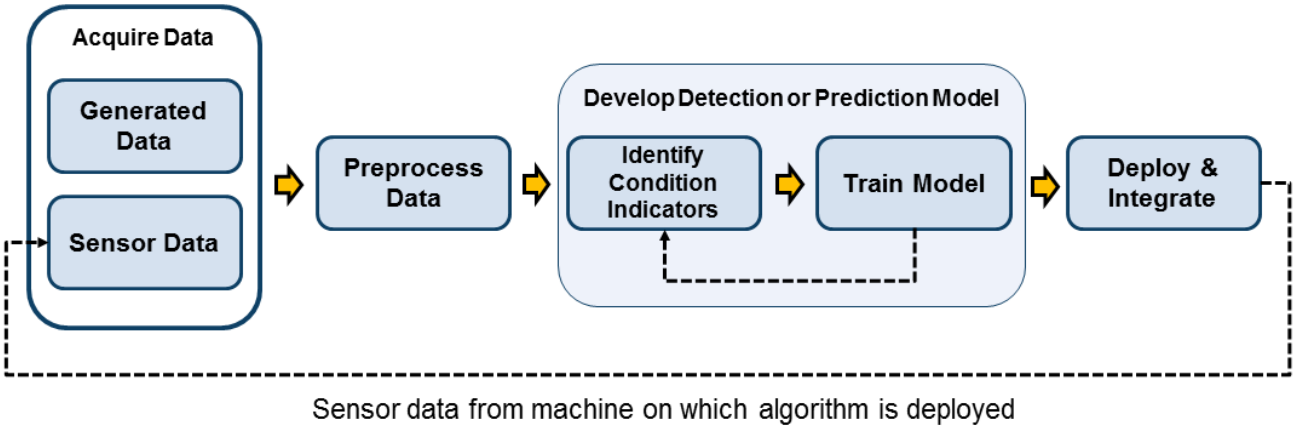
Condition monitoring uses data from a machine to assess its current condition and to detect and diagnose faults in the machine. Machine data is data such as temperature, pressure, voltage, noise, or vibration measurements, collected using dedicated sensors. A condition monitoring algorithm derives metrics from the data called condition indicators. A condition indicator is any feature of system data whose behavior changes in a predictable way as the system degrades. A condition indicator can be any quantity derived from the data that clusters similar system status together, and sets different status apart. Thus a condition-monitoring algorithm can perform fault detection or diagnosis by comparing new data against the established markers of faulty conditions.

Prognostics is forecasting when a failure will happen based on the current and past state of the machine. A prognostics algorithm typically estimates the machine's remaining useful life (RUL) or time-to-failure by analyzing the current state of the machine. Prognostics can use modeling, machine learning, or a combination of both to predict future values of condition indicators. These future values are then used to compute RUL metrics, which determine if and when maintenance should be performed. For the gearbox example, a prognostics algorithm might fit the varying peak vibration frequency and magnitude to a time series to predict their future values. The algorithm can then compare the predicted values to a threshold defining healthy operation of the gearbox, predicting if and when a fault will occur.

A predictive maintenance system implements prognostics and condition monitoring algorithms with other IT infrastructure that makes the end results of the algorithm accessible and actionable to end users who perform the actual maintenance tasks. Predictive Maintenance Toolbox provides tools to help you design such algorithms.

Workflows for Algorithm Development

The following illustration shows a workflow for developing a predictive maintenance algorithm.



Beginning with data that describes your system in a range of healthy and faulty conditions, you develop a detection model (for condition monitoring) or a prediction model (for prognostics). Developing such a model requires identifying appropriate condition indicators and training a model to interpret them. That process is very likely to be iterative, as you try different condition indicators and different models until you find the best model for your application. Finally, you deploy the algorithm and integrate it into your systems for machine monitoring and maintenance.

Acquire Data

Designing predictive maintenance algorithms begins with a body of data. Often you must manage and process large sets of data, including data from multiple sensors and multiple machines running at different times and under different operating conditions. You might have access to one or more of the following types of data:

- Real data from normal system operation
- Real data from system operating in a faulty condition
- Real data from system failures (run-to-failure data)

For instance, you might have sensor data from system operation such as temperature, pressure, and vibration. Such data is typically stored as signal or time series data. You might also have text data, such as data from maintenance records, or data in other forms. This data is stored in files, databases, or distributed file systems such as Hadoop®.

In many cases, failure data from machines is not available, or only a limited number of failure datasets exist because of regular maintenance being performed and the relative rarity of such incidents. In this case, failure data can be generated from a Simulink model representing the system operation under different fault conditions.

Predictive Maintenance Toolbox provides functionality for organizing, labeling, and accessing such data stored on disk. It also provides tools to facilitate the generation of data from Simulink models for predictive maintenance algorithm development. For more information, see “Data Ensembles for Condition Monitoring and Predictive Maintenance”.

Preprocess Data

Data preprocessing is often necessary to convert the data into a form from which condition indicators are easily extracted. Data preprocessing includes simple techniques such as outlier and missing value removal, and advanced signal processing techniques such as short-time Fourier transforms and transformations to the order domain.

Understanding your machine and the kind of data you have can help determine what preprocessing methods to use. For example, if you are filtering noisy vibration data, knowing what frequency range is most likely to display useful features can help you choose preprocessing techniques. Similarly, it might be useful to transform gearbox vibration data to the order domain, which is used for rotating machines when the rotational speed changes over time. However, that same preprocessing would not be useful for vibration data from a car chassis, which is a rigid body.

For more information on preprocessing data for predictive maintenance algorithms, see “Data Preprocessing for Condition Monitoring and Predictive Maintenance”.

Identify Condition Indicators

A key step in predictive maintenance algorithm development is identifying condition indicators, features in your system data whose behavior changes in a predictable way as the system degrades. A condition indicator can be any feature that is useful for distinguishing normal from faulty operation or for predicting remaining useful life. A useful condition indicator clusters similar system status together, and sets different status apart. Examples of condition indicators include quantities derived from:

- Simple analysis, such as the mean value of the data over time
- More complex signal analysis, such as the frequency of the peak magnitude in a signal spectrum, or a statistical moment describing changes in the spectrum over time
- Model-based analysis of the data, such as the maximum eigenvalue of a state space model which has been estimated using the data
- Combination of multiple features into a single effective condition indicator (fusion)

For example, you can monitor the condition of a gearbox using vibration data. Damage to the gearbox results in changes to the frequency and magnitude of the vibrations. The peak frequency and peak magnitude are thus useful condition indicators, providing information about the kind of vibrations present in the gearbox. To monitor the health of the gearbox, you can continuously analyze the vibration data in the frequency domain to extract these condition indicators.

Even when you have real or simulated data representing a range of fault conditions, you might not know how to analyze that data to identify useful condition indicators. The right condition indicators for your application depend on what type of system, system data, and system knowledge you have. Therefore, identifying condition indicators can require some trial and error, and is often iterative with the training step of the algorithm development workflow. Among the techniques commonly used for extracting condition indicators are:

- Order analysis
- Modal analysis
- Spectrum analysis
- Envelope spectrum
- Fatigue analysis
- Nonlinear time-series analysis
- Model-based analysis such as residual computation, state estimation, and parameter estimation

Predictive Maintenance Toolbox supplements functionality in other toolboxes such as Signal Processing Toolbox™ with functions for extracting signal-based or model-based condition indicators from measured or generated data. For more information, see “Identify Condition Indicators”.

Train Detection or Prediction Model

At the heart of the predictive maintenance algorithm is the detection or prediction model. This model analyzes extracted condition indicators to determine the current condition of the system (fault detection and diagnosis) or predict its future condition (remaining useful life prediction).

Fault Detection and Diagnosis

Fault detection and diagnosis relies on using one or more condition indicator values to distinguish between healthy and faulty operation, and between different types of faults. A simple fault-detection model is a threshold value for the condition indicator that is indicative of a fault condition when exceeded. Another model might compare the condition indicator to a statistical distribution of indicator values to determine the likelihood of a particular fault state. A more complex fault-diagnosis approach is to train a classifier that compares the current value of one or more condition indicators to values associated with fault states, and returns the likelihood that one or another fault state is present.

When designing your predictive maintenance algorithm, you might test different fault detection and diagnosis models using different condition indicators. Thus, this step in the design process is likely iterative with the step of extraction condition indicators, as you try different indicators, different combinations of indicators, and different decision models. Statistics and Machine Learning Toolbox™ and other toolboxes include functionality that you can use to train decision models such as classifiers and regression models. For more information, see “Decision Models for Fault Detection and Diagnosis”.

Remaining Useful Life Prediction

Examples of prediction models include:

- A model that fits the time evolution of a condition indicator and predicts how long it will be before the condition indicator crosses some threshold value indicative of a fault condition.
- A model that compares the time evolution of a condition indicator to measured or simulated time series from systems that ran to failure. Such a model can compute the most likely time-to-failure of the current system.

You can predict remaining useful life by forecasting with dynamic system models or state estimators. Additionally, Predictive Maintenance Toolbox includes specialized functionality for RUL prediction based on such techniques as similarity, threshold, and survival analysis. For more information, see “Models for Predicting Remaining Useful Life”.

Deploy and Integrate

When you have identified a working algorithm for handling your new system data, processing it appropriately, and generating a prediction, deploy the algorithm and integrate it into your system. Based the specifics of your system, you can deploy your algorithm on the cloud or on embedded devices.

A cloud implementation can be useful when you are gathering and storing large amounts of data on the cloud. Removing the need to transfer data between the cloud and local machines that are running the prognostics and health monitoring algorithm makes the maintenance process more effective. Results calculated on the cloud can be made available through tweets, email notifications, web apps, and dashboards.

Alternatively, the algorithm can run on embedded devices that are closer to the actual equipment. The main benefits of doing this are that the amount of information sent is reduced as data is transmitted only when needed, and updates and notifications about equipment health are immediately available without any delay. .

A third option is to use a combination of the two. The preprocessing and feature extraction parts of the algorithm can be run on embedded devices, while the predictive model can run on the cloud and generate notifications as needed. In systems such as oil drills and aircraft engines that are run continuously and generate huge amounts of data, storing all the data on board or transmitting it is not always viable because of cellular bandwidth and cost limitations. Using an algorithm that operates on streaming data or on batches of data lets you store and send data only when needed.

MathWorks® code generation and deployment products can help you with this step of the workflow. For more information, see “Deploy Predictive Maintenance Algorithms”.

References

[1] Isermann, R. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin: Springer Verlag, 2006.

See Also

Related Examples

- “Using Simulink to Generate Fault Data”
- “Rolling Element Bearing Fault Diagnosis”
- “Similarity-Based Remaining Useful Life Estimation”
- “Explore Ensemble Data and Compare Features Using Diagnostic Feature Designer”

Getting Started with Diagnostic Feature Designer

- “Identify Condition Indicators for Predictive Maintenance Algorithm Design” on page 2-2
- “Import and Visualize Ensemble Data in Diagnostic Feature Designer” on page 2-5
- “Process Data and Explore Features in Diagnostic Feature Designer” on page 2-14
- “Rank and Export Features in Diagnostic Feature Designer” on page 2-31

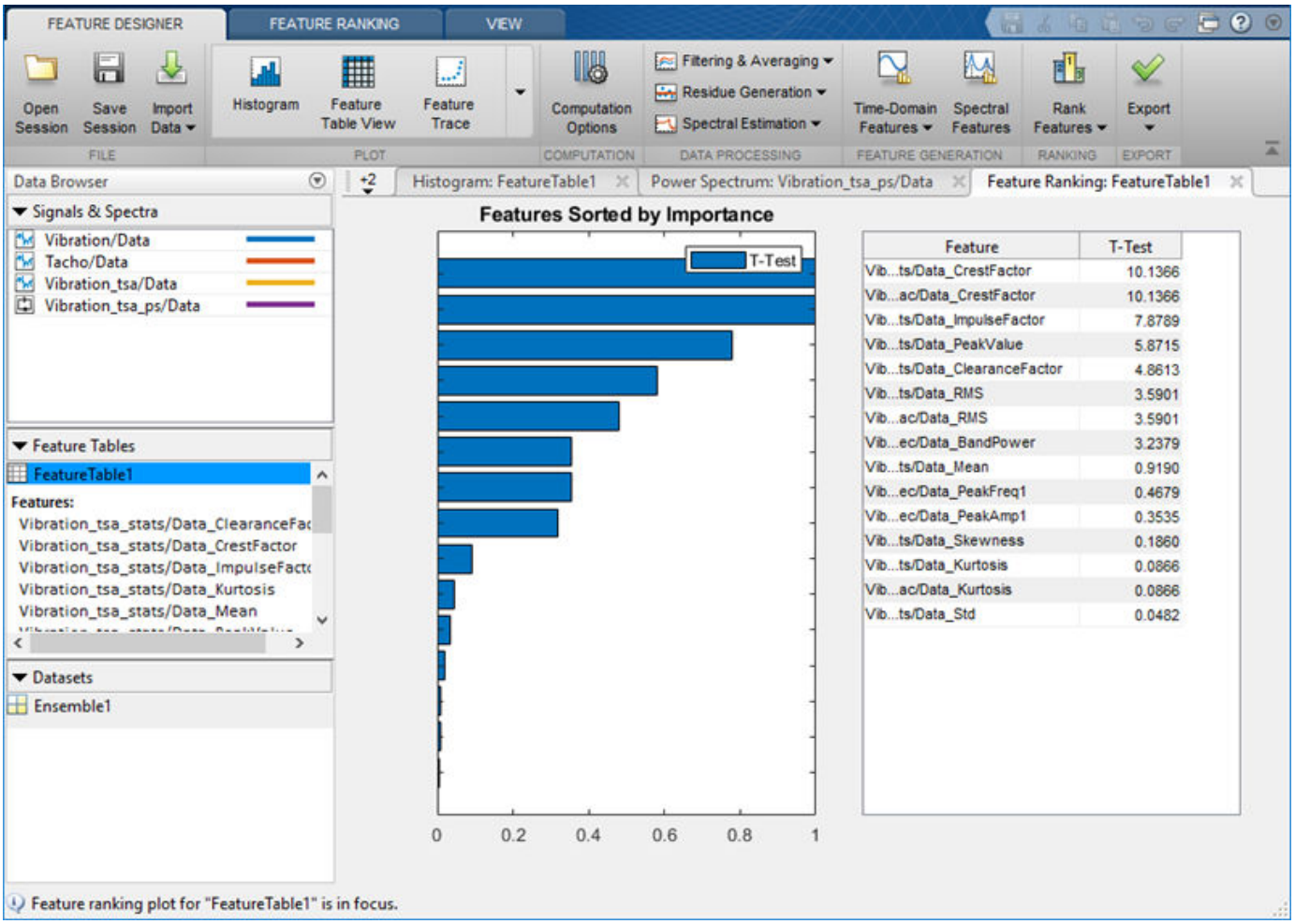
Identify Condition Indicators for Predictive Maintenance Algorithm Design

Predictive maintenance allows equipment users and manufacturers to assess the working condition of machinery, diagnose faults, or estimate when the next equipment failure is likely to occur. When you can diagnose or predict failures, you can plan maintenance in advance, better manage inventory, reduce downtime, and increase operational efficiency.

A key step in predictive maintenance algorithm development is identifying condition indicators, which are features in your system data whose behavior changes in a predictable way as the system degrades. A condition indicator can be any feature that is useful for distinguishing normal from faulty operation or for predicting remaining useful life. A useful condition indicator clusters similar system status together and sets different status apart. Examples of condition indicators include quantities derived from:

- Simple analysis, such as the mean value of the data over time.
- More complex signal analysis, such as the frequency of the peak magnitude in a signal spectrum, or the time-synchronous average of a signal from a rotating source.

In the **Diagnostic Feature Designer** app, you can develop features and evaluate potential condition indicators using a multifunction graphical interface.



The app operates on data ensembles. An *ensemble* is a collection of data sets, created by measuring or simulating a system under varying conditions. An individual data set representing one system under one set of conditions is a *member*. **Diagnostic Feature Designer** processes all ensemble members when executing one operation.

Within **Diagnostic Feature Designer**, you can interactively:

- Explore your data ensemble visually by plotting and interacting with your ensemble members together.
- Convert your data into different forms for further exploration. For example, you can create a power spectrum of your signal to evaluate its frequency-domain behavior. Or you can perform time-synchronous averaging, which filters out any noise or disturbance that is not associated with your machine rotation.
- Generate features of various types, and plot histograms that visualize the effectiveness of each feature in separating data from systems with different conditions.
- Rank the generated features by using ranking algorithms that use specific criteria to establish which features are most effective.

The following three-part tutorial takes you through the **Diagnostic Feature Designer** workflow for a transmission system model, from your initial data import to the export of your chosen features.

1. “Import and Visualize Ensemble Data in Diagnostic Feature Designer” on page 2-5
2. “Process Data and Explore Features in Diagnostic Feature Designer” on page 2-14
3. “Rank and Export Features in Diagnostic Feature Designer” on page 2-31

The full workflow for a predictive maintenance program includes multiple steps that begin with data acquisition and end with deployment and integration of a condition monitoring algorithm. For more information, see “Designing Algorithms for Condition Monitoring and Predictive Maintenance” on page 1-3.

Import and Visualize Ensemble Data in Diagnostic Feature Designer

Diagnostic Feature Designer is an app that allows you to develop features and evaluate potential condition indicators using a multifunction graphical interface.

The app operates on data ensembles. An ensemble is a collection of data sets, created by measuring or simulating a system under varying conditions. An individual dataset representing one system under one set of conditions is a member. **Diagnostic Feature Designer** processes all ensemble members when executing one operation.

This example shows how to import data into **Diagnostic Feature Designer** and visualize your imported data.

Load Transmission Model Data

This example uses data generated from a transmission system model in “Using Simulink to Generate Fault Data”. Outputs of the model include:

- Vibration measurements from a sensor monitoring casing vibrations
- Tachometer sensor, which issues a pulse every time the shaft completes a rotation
- Fault code indicating presence of a modeled fault

Load the data. The data is a table containing variables logged during multiple simulations of the model under varying conditions. Sixteen members have been extracted from the transmission model logs to form an ensemble. Four of these members represent healthy data, and the remaining 12 members exhibit varying levels of sensor drift.

```
load dfd_Tutorial dataTable
```

View this table in your workspace.

```
dataTable =

  16×3 table

       Vibration              Tacho           faultCode
    ___________________    ___________________    _________

    [6000×1 timetable]    [6000×1 timetable]        0
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        0
    [6000×1 timetable]    [6000×1 timetable]        0
    [6000×1 timetable]    [6000×1 timetable]        0
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
```

```
    [6000×1 timetable]    [6000×1 timetable]        1
    [6000×1 timetable]    [6000×1 timetable]        1
```

There are 16 rows, each representing one member. Each column has a variable name. The data variables `Vibration` and `Tacho` are each represented by a timetable, and all timetables have the same length. The third variable, `faultCode`, is the condition variable. `faultCode` has a value of `0` for healthy and `1` for degraded.

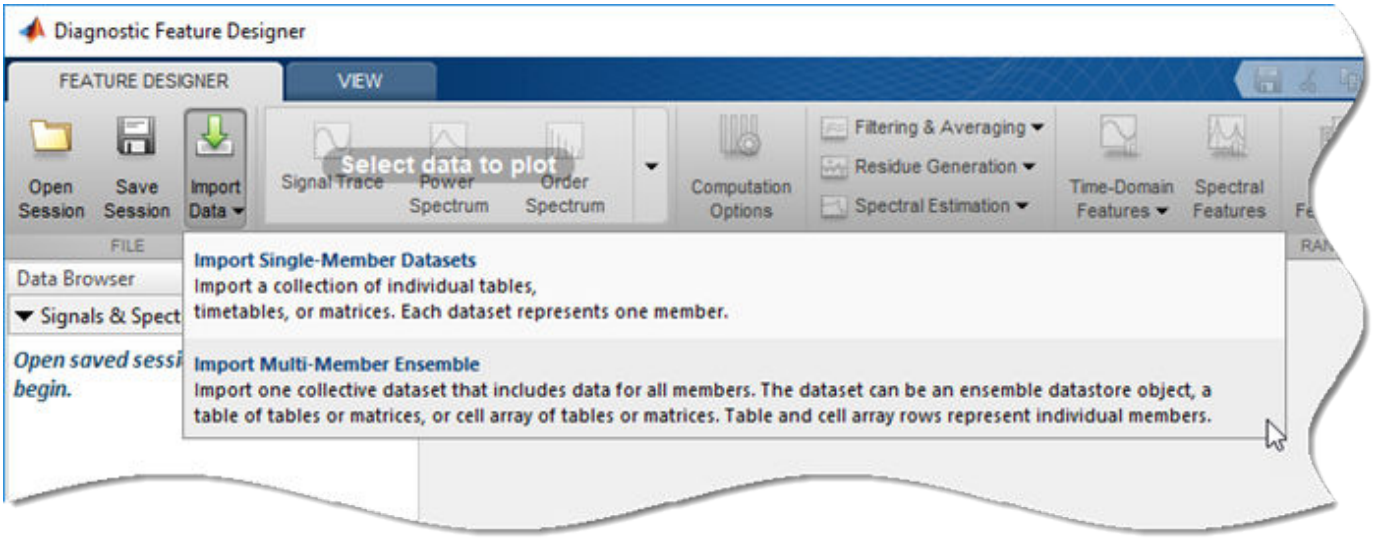
Import Data into Diagnostic Feature Designer

To open the **Diagnostic Feature Designer**, in your command window, type:

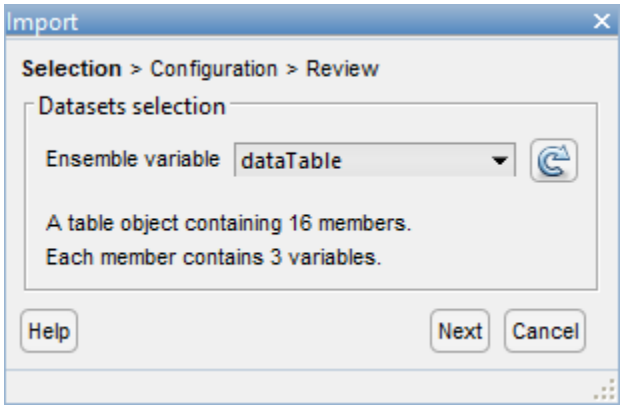
```
diagnosticFeatureDesigner
```

Import Data

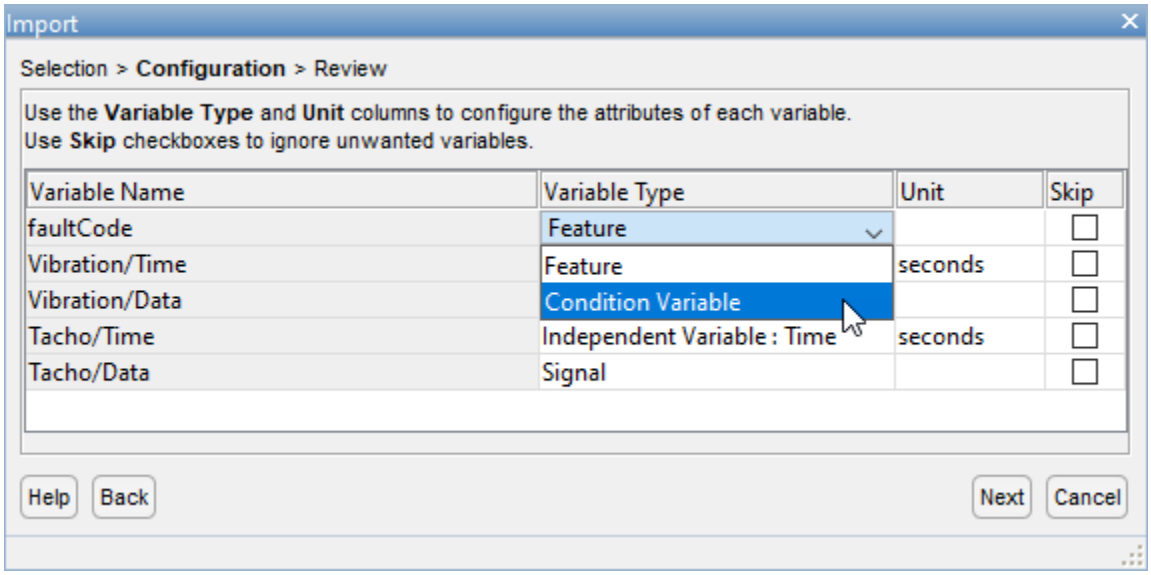
Import the dataset that you previously loaded into your workspace. In the **Import Data** menu, select **Import Multi-Member Ensemble**.



Confirm the selection for **Ensemble variable**, and click **Next**.



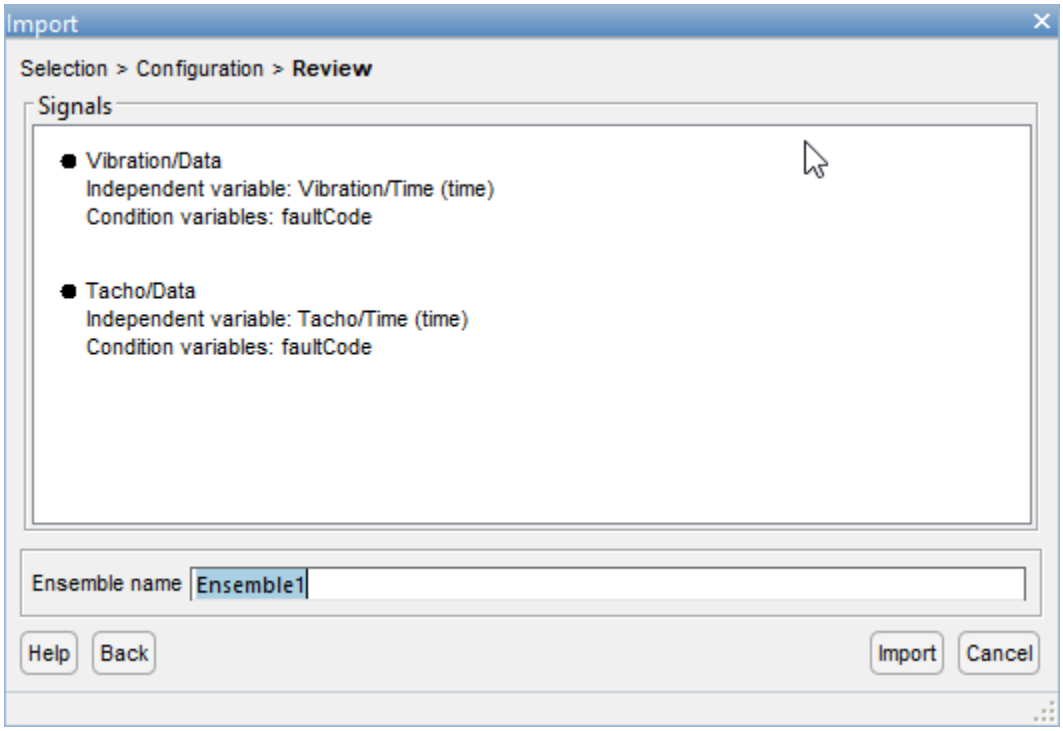
Review the list of imported variables and variable types.



The app has extracted the variable names from your member tables and embedded timetables. For example, in your source table, the `Vibration` variable is a `timetable` containing `Time` and `Data`. The variable names resulting from the import are **Vibration/Time** and **Vibration/Data**.

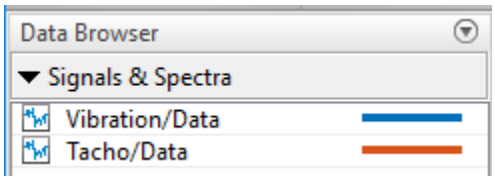
- The variables derived from the vibration and tacho signals have the correct variable types. These variables are unambiguous because they come from timetables.
- **faultCode** has the incorrect variable type `Feature` when it is a `Condition Variable`. Both features and condition variables can be numeric scalars. Update **faultCode**, and click **Next**.

Review the signals and the independent and condition variables that apply to the signals. Note the ensemble name. To complete the import process, click **Import**.

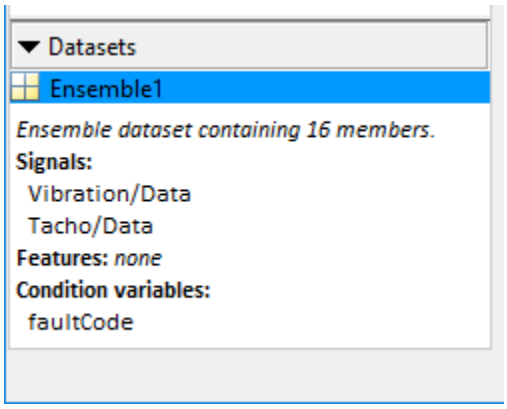


Your imported signals are now in the **Signals & Spectra** area, and your imported ensemble `Ensemble1` is in the **Datasets** area.

The color code next to a signal represents that signal in plots. The icon to the left of the signal indicates the variable type, which, for the variables you imported, is `Signal`.

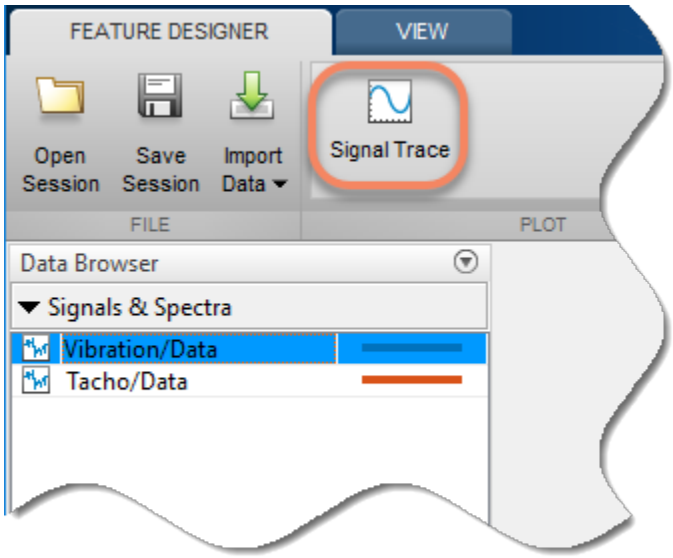


Display information about your dataset by selecting its name in the **Datasets** area.



Visualize Data

Now that your signals are loaded, plot them and view all your ensemble members together. To view your vibration signal, in the **Signals & Spectra** area, select **Vibration**. Selecting a signal variable enables the **Signal Trace** option in the plot gallery. Click **Signal Trace**.

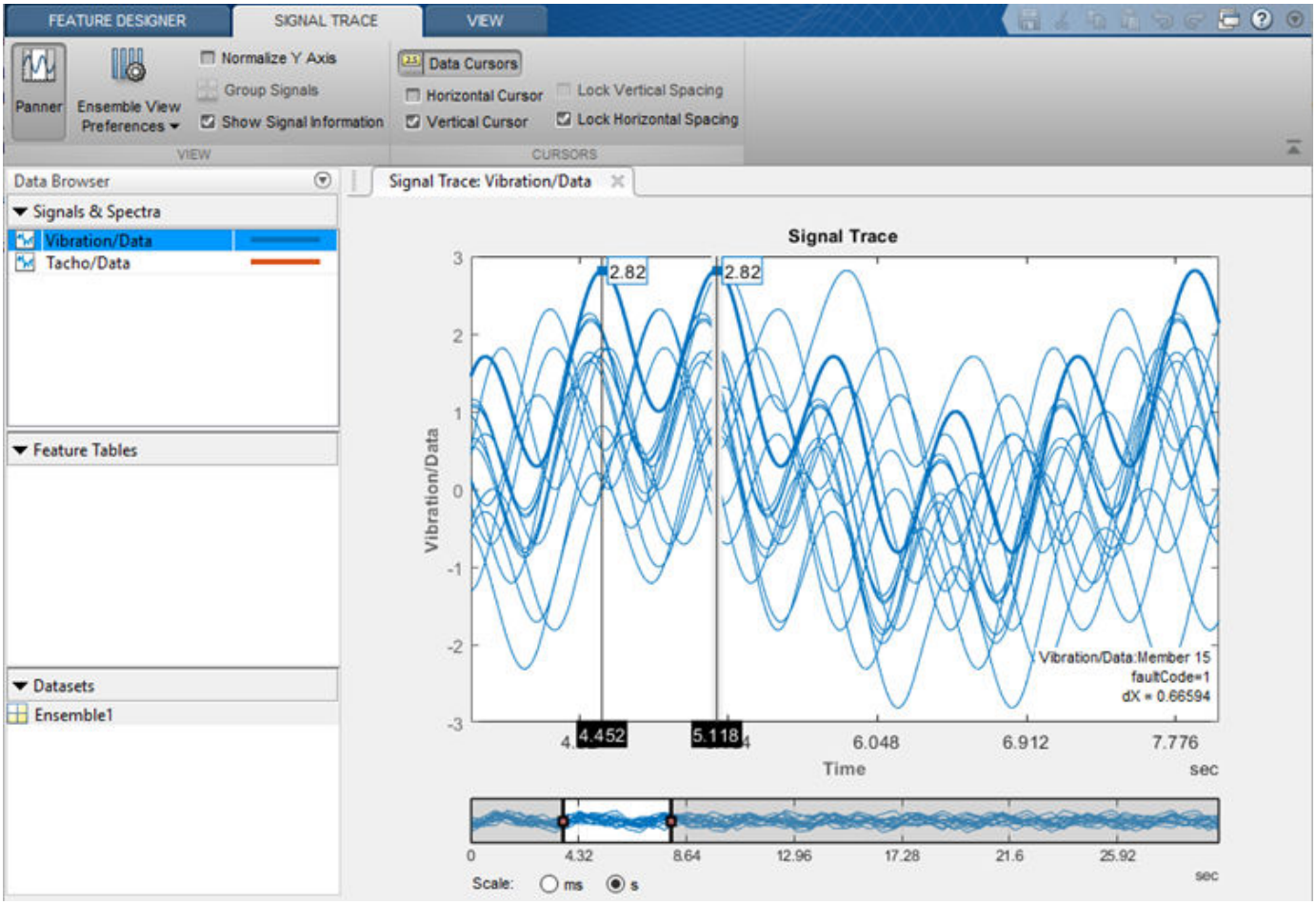


The plotting area displays a signal trace plot of all 16 members. As you move the cursor over the data, an indicator in the lower right corner identifies the member your cursor is on. A second indicator provides the fault code value for that member.

Interact with the trace plot using standard MATLAB® plot tools, such as zoom and pan. Access these tools by pointing to the top right edge of the plot. You can also use the specialized options on the **Signal Trace** tab, which appears when you select the **Signal Trace** plot.

Explore Your Data Using Signal Trace Options

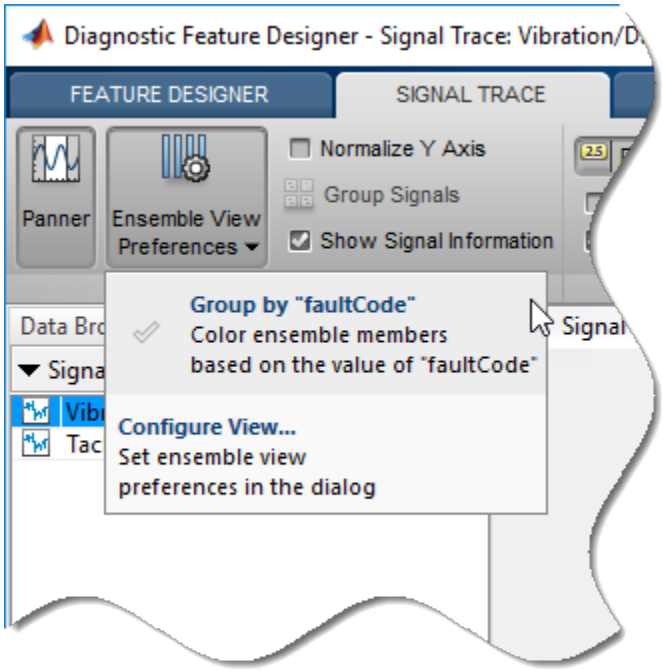
Explore the data in your plot using options in the **Signal Trace** tab.



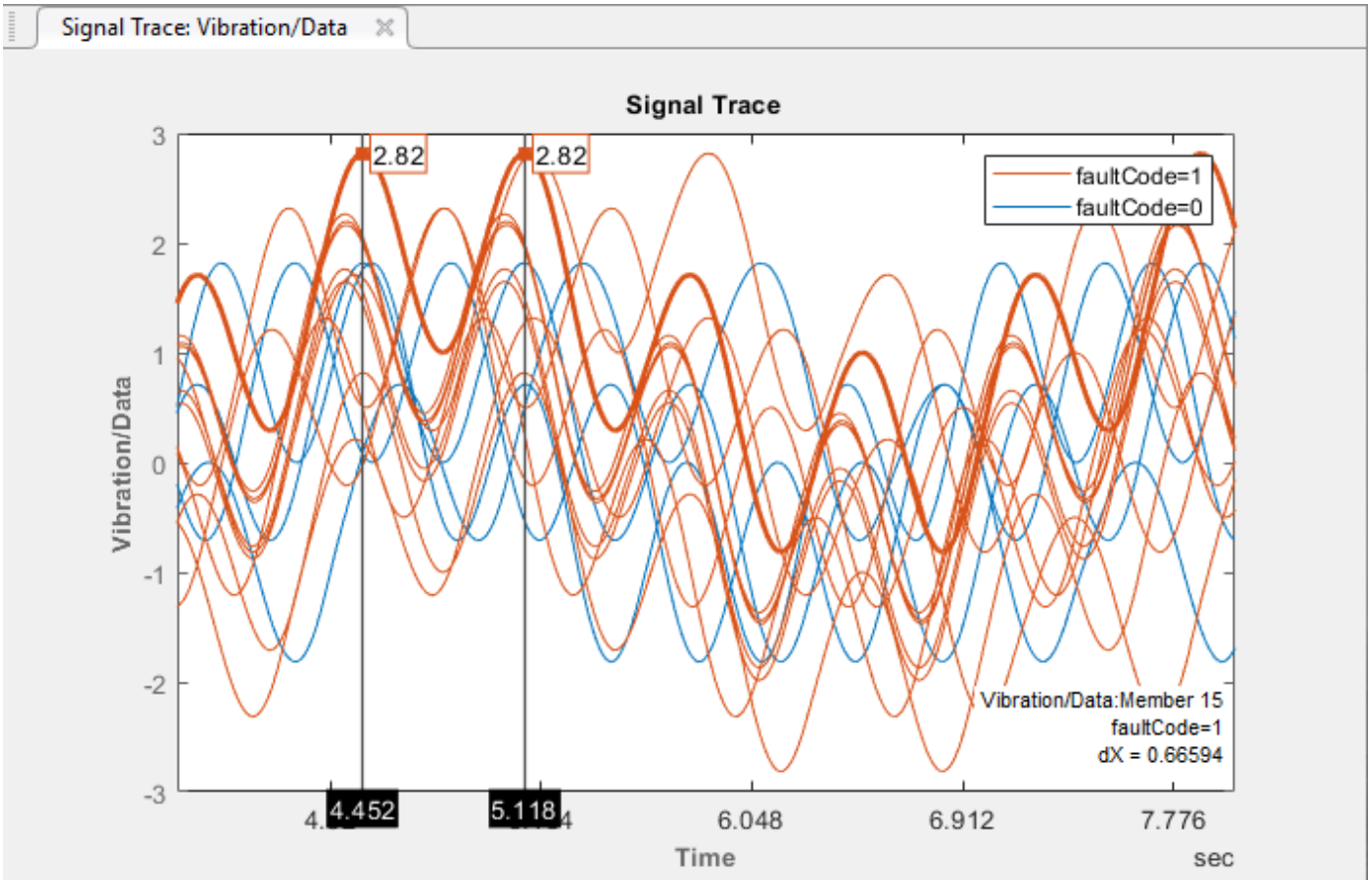
Measure the distance between peaks for the one of the members with the high peaks.

1. Zoom in on the second clusters of peaks. Click **Panner**. In the panner strip, move the right handle to about 8. Then, move the panner window so that the left handle is at about 4. You should now have the second set of peaks within the window.
2. Pause on the first high peak, and note the member number. The second high peak is a continuation of the same member trace.
3. Click **Data Cursors**, and select **Vertical Cursor**. Place the left cursor over the first high peak and the right cursor over the second peak for that member. The lower right corner of the plot displays the separation `dX`.
4. Select **Lock Horizontal Spacing**. Shift the cursor pair to the right by one peak for the same member. Is the right cursor aligned with the third member peak?
5. Restore the full window by moving the handles back to the edges of the panner.

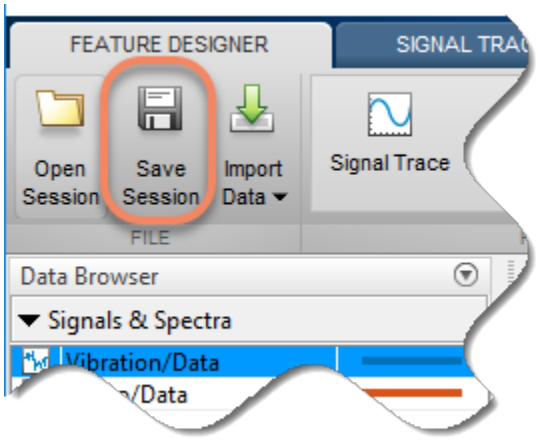
Show which members have matching `faultCode` values by using color coding. Select **Ensemble View Preferences > Group by "faultCode"**.



The resulting signal trace shows you that all the highest vibration peaks are associated with a data from faulty systems. However, not all the faulty systems have higher peaks.



Save your session data. You need this data to run the "Process Data and Explore Features in Diagnostic Feature Designer" on page 2-14 example.



Next Steps

The next step is to explore different ways to characterize your data through features. The example "Process Data and Explore Features in Diagnostic Feature Designer" on page 2-14 guides you through the feature exploration process.

See Also

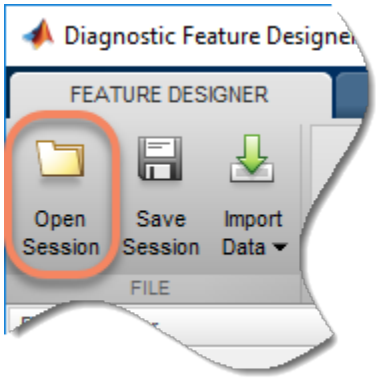
Diagnostic Feature Designer

More About

- “Process Data and Explore Features in Diagnostic Feature Designer” on page 2-14
- “Rank and Export Features in Diagnostic Feature Designer” on page 2-31
- “Explore Ensemble Data and Compare Features Using Diagnostic Feature Designer”
- “Data Ensembles for Condition Monitoring and Predictive Maintenance”
- “Condition Indicators for Monitoring, Fault Detection, and Prediction”

Process Data and Explore Features in Diagnostic Feature Designer

This example shows how to process your data in the app in preparation for feature extraction. If you want to follow along with the steps interactively, use the data you imported in “Import and Visualize Ensemble Data in Diagnostic Feature Designer” on page 2-5. Use **Open Session** to reload your session data using the file name you provided.



If you have no session data, execute the steps for loading and importing data in “Import and Visualize Ensemble Data in Diagnostic Feature Designer” on page 2-5.

A key step in predictive maintenance algorithm development is identifying condition indicators. Condition indicators are features in your system data whose behavior changes in a predictable way as the system degrades. A condition indicator can be any feature that is useful for distinguishing normal from faulty operation or for predicting remaining useful life. A useful feature clusters similar system status together and sets different status apart.

Diagnostic Feature Designer lets you design features that provide these diagnostics.

- For some features, you can generate features directly using signals you imported.
- For other features, you must perform additional signal processing, such as filtering and averaging, to have meaningful results.

The processing you perform depends both on the computational requirements of the feature and the characteristics of your systems and your system data. This example shows how to:

- Process your data in preparation for feature extraction
- Generate various types of feature
- Interpret the effectiveness of your features in histograms

Perform Time-Synchronous Averaging

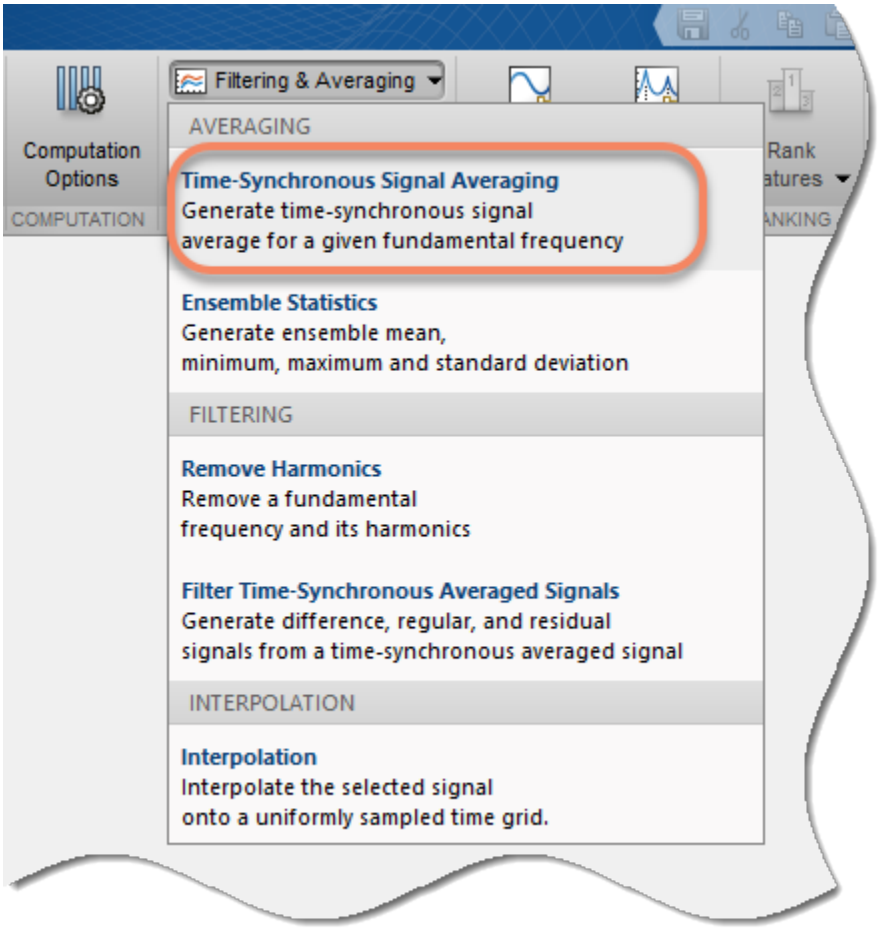
The data for this system represents a transmission system with rotating parts. The variables include tachometer outputs that precisely mark the completion of each shaft revolution. The data, therefore, is an ideal candidate for time-synchronous averaging.

Time-synchronous averaging (TSA) is a common technique for analyzing data from rotating machinery. TSA averages rotation by rotation, and filters out any disturbances or noise that is not coherent with the rotation.

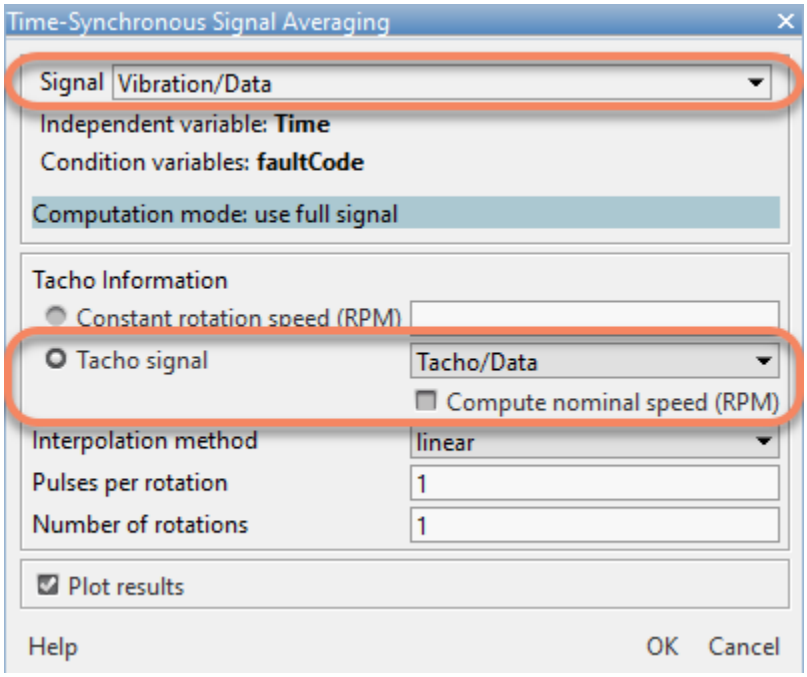
TSA is useful for isolating fault signatures that repeat each rotation, such as perturbations from gear-tooth defects. Features generated from a TSA signal rather than the original vibration signal provide clearer differentiation for rotational fault conditions. This advantage holds true even for features that are not specifically for rotating machinery.

To compute the TSA of the vibration data:

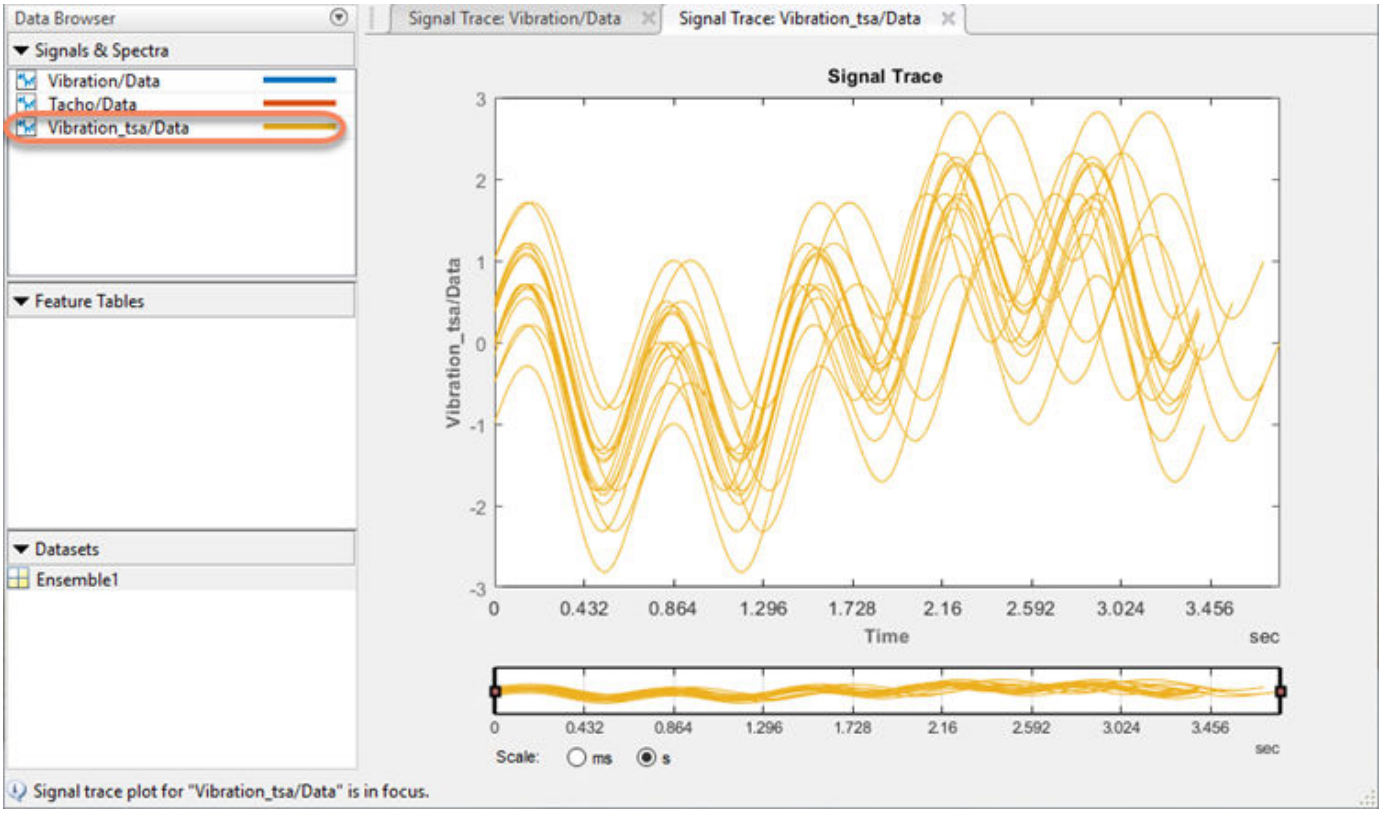
- Select **Filtering & Averaging > Time-Synchronous Signal Averaging**



- In **Signal**, select `Vibration`.
- In **Tacho Information**, select **Tacho signal** and `Tacho/Data`. Clear **Compute nominal speed (RPM)**.



Click **OK** to start the TSA computation for each of the 16 members of the ensemble. A progress bar shows the status while the computation progresses.

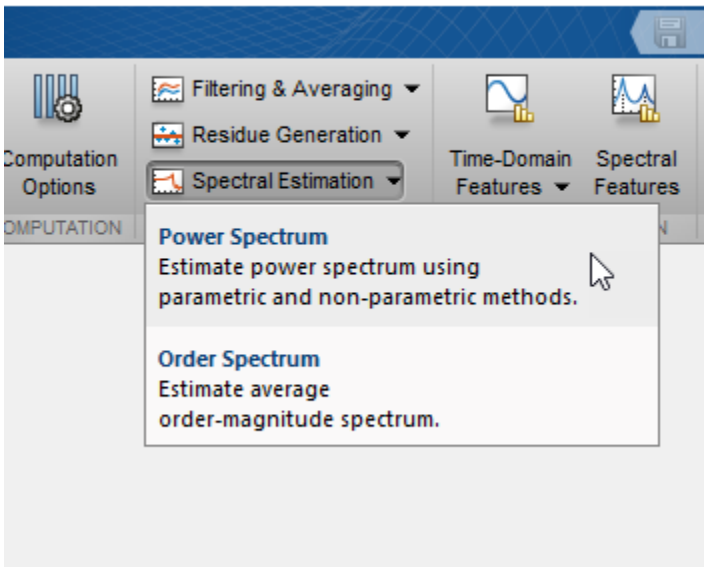


When the computation concludes:

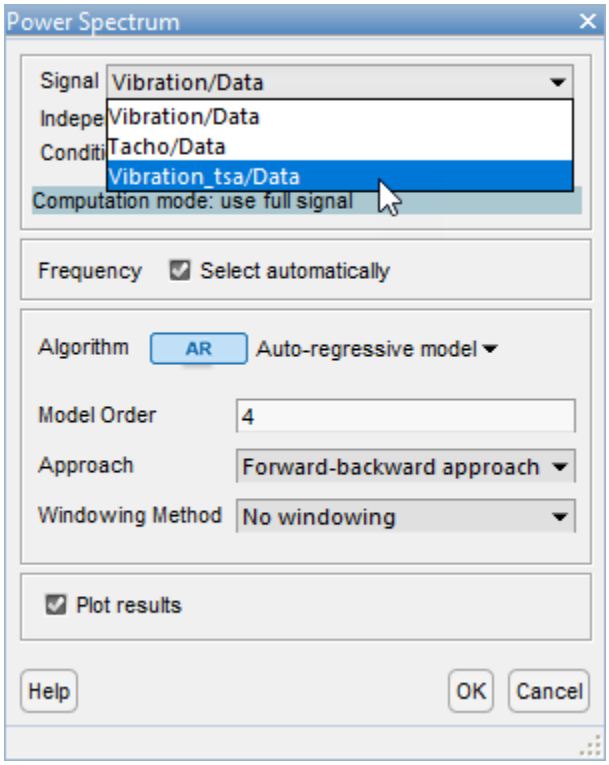
- The app adds a new signal variable `Vibration_tsa` to the imported `Ensemble1` dataset.
- The signal trace plots `Vibration_tsa`. The time axis of this trace is less than the four seconds long. The original vibration data was 30 seconds long. The shorter timespan reflects the duration of a single rotation for each member.
- The member shaft rates diverge. This divergence is evident in the increasing misalignment of the peaks during the rotation, and the fact that the member traces stop at different times.

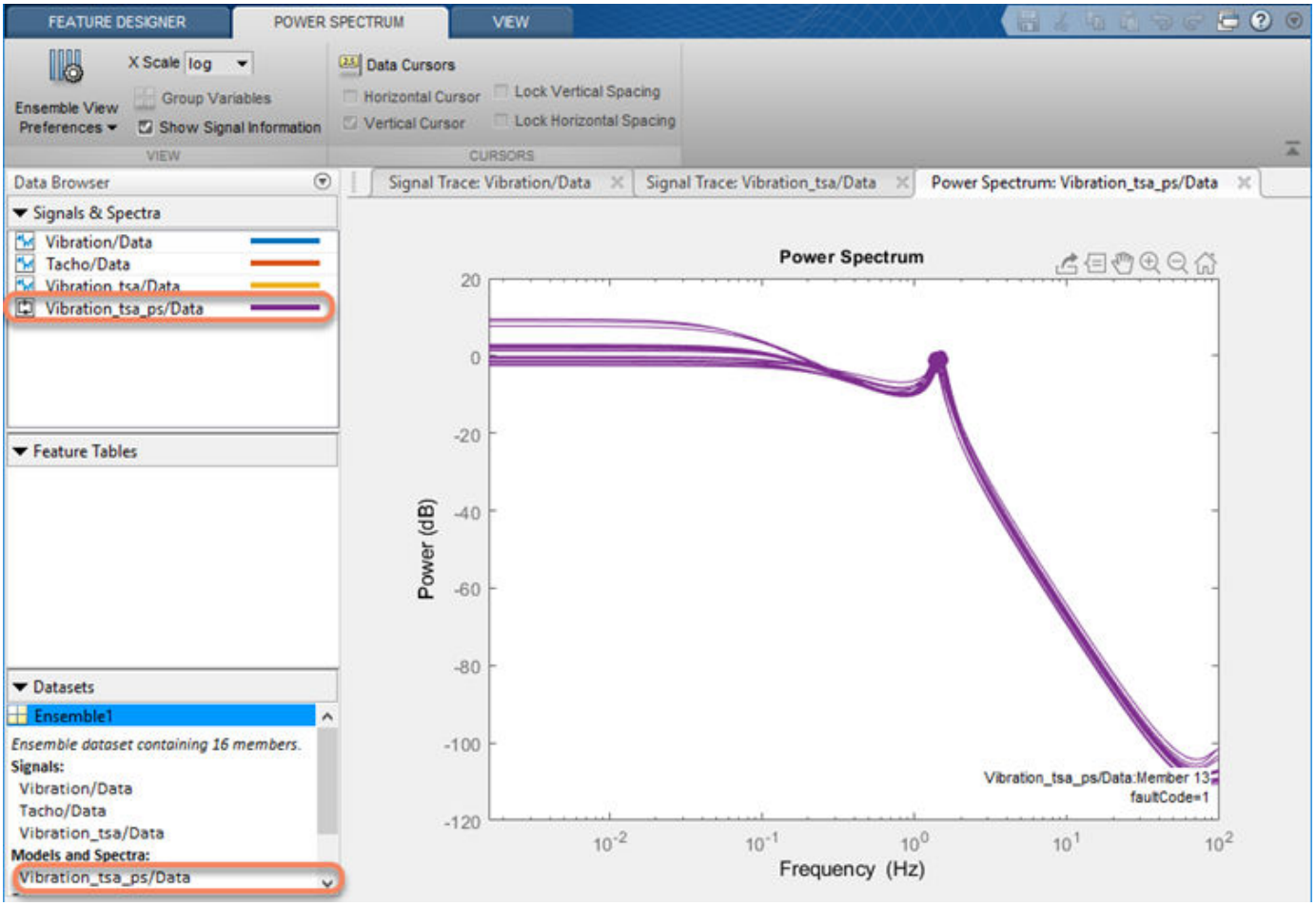
Compute a Power Spectrum

The TSA signal gives you enough information to start generating time-domain features, but you must provide a spectrum to explore spectral features. To generate a power spectrum, select **Spectral Estimation > Power Spectrum**.



From **Signal**, select your TSA signal.


Power Spectrum
Signal
Vibration/Data
Indepe
Conditi
Vibration/Data
Tacho/Data
Vibration_tsa/Data
Computation mode: use full signal
Frequency
Select automatically
Algorithm
AR
Auto-regressive model
Model Order
4
Approach
Forward-backward approach
Windowing Method
No windowing
Plot results
Help
OK
Cancel



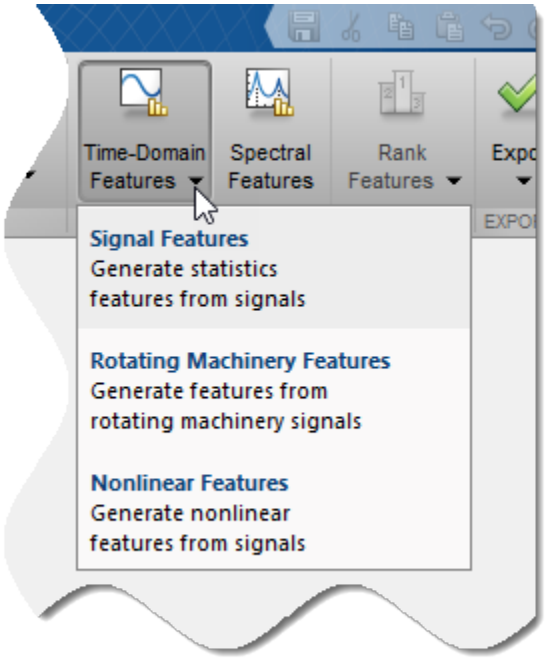
The power spectrum processing results in a new variable, `Vibration_tsa_ps`. The associated icon represents a linear model, consistent with the AR model default in the power spectrum dialog box.

A plot of the spectra appears in the plot area. As with **Signal Trace**, a **Power Spectrum** tab provides options for plotting. These options are similar to **Signal Trace**. There is no **Panner** option because **Panner** works only with time and not frequency.

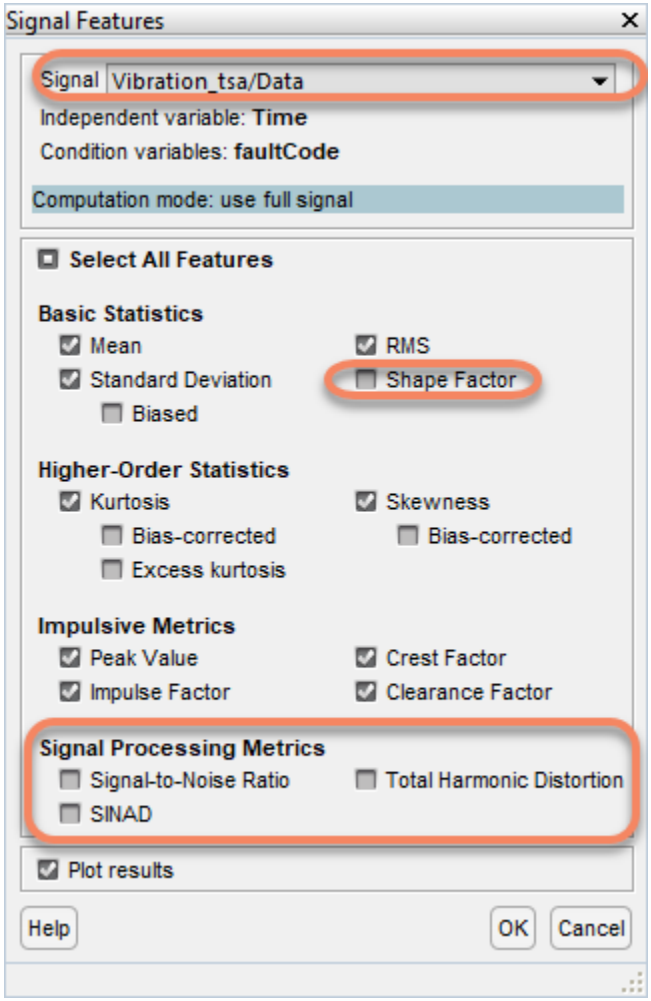
Generate Features

Signal Features

Generate features based on general data statistics, using the TSA signal as your source. Select **Time Domain Features > Signal Features**.

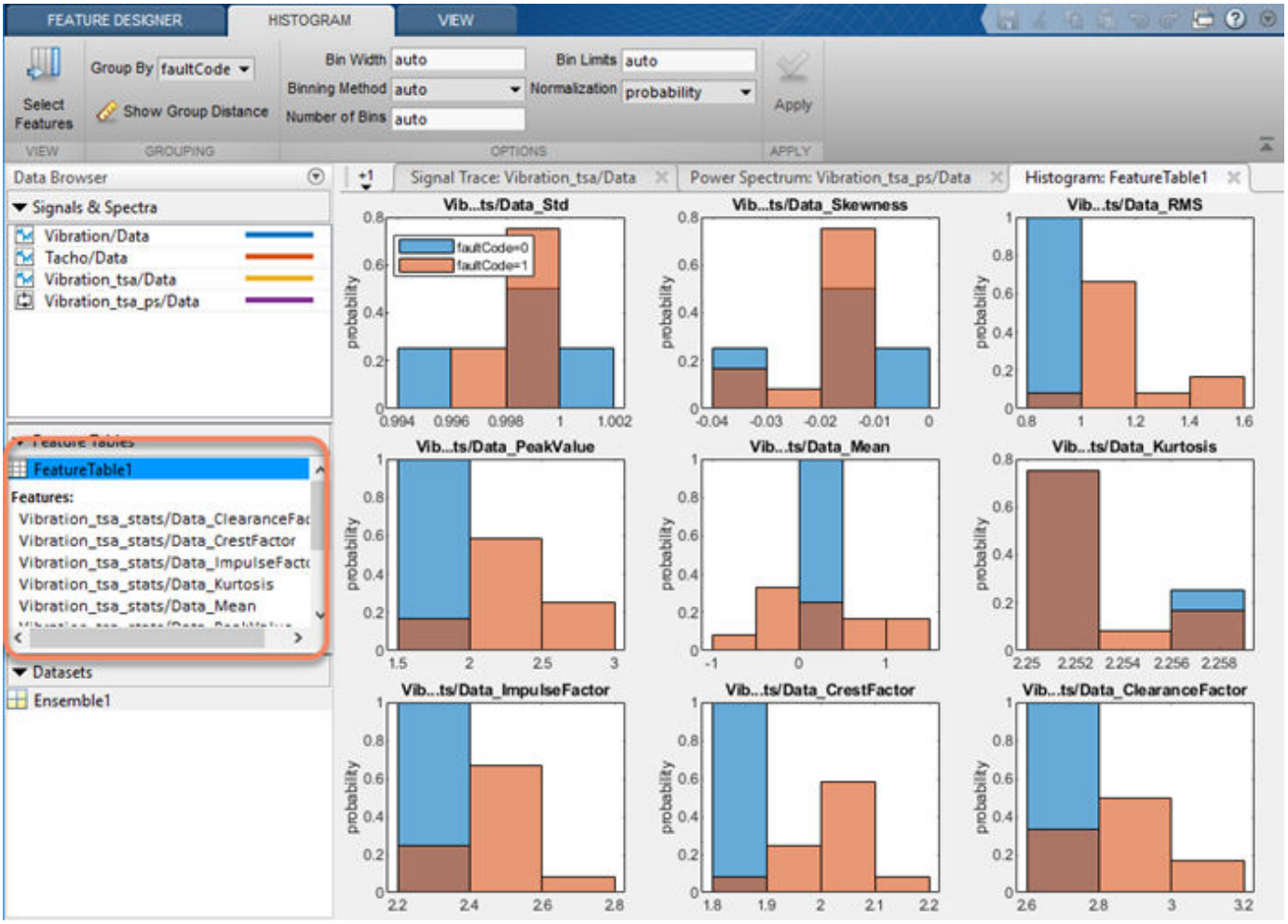


Change **Signal** to `Vibration_tsa/Data`. The default is for all the features to be selected. Clear the selections for **Shape Factor** and **Signal Processing Metrics**.



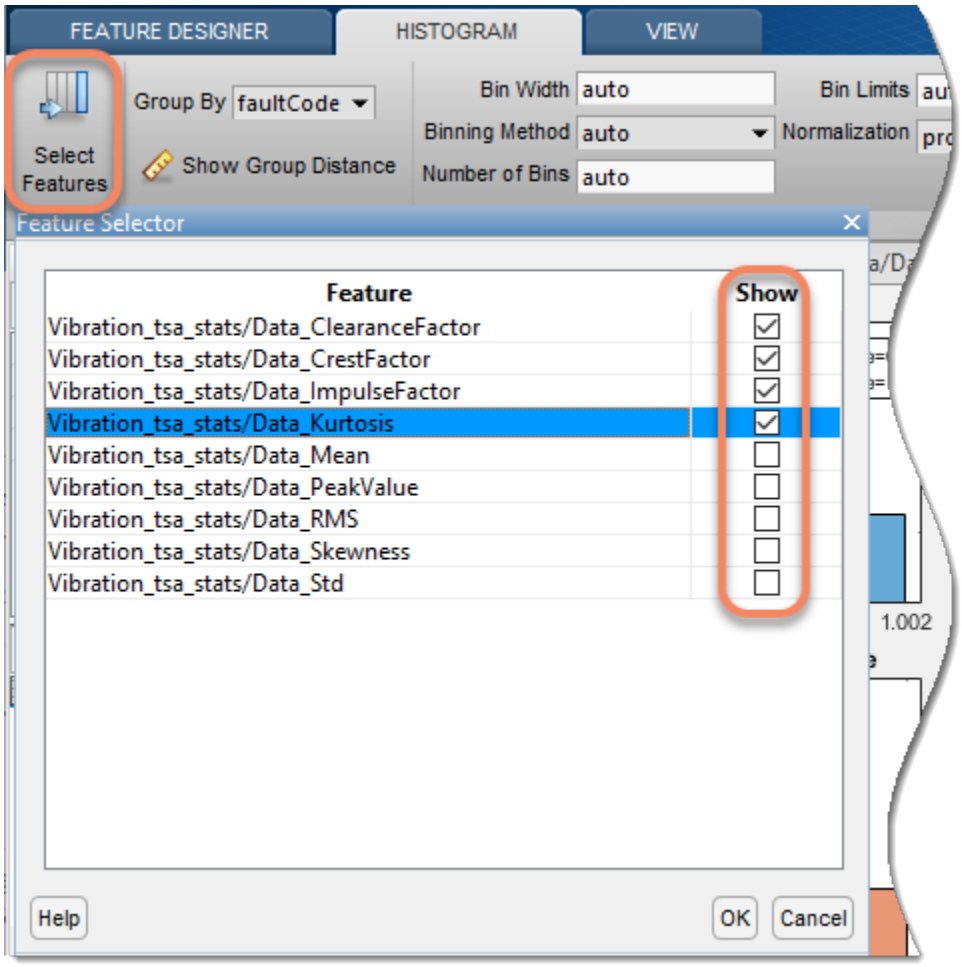
For every selected feature, the app computes a value for each ensemble member and displays the results in a histogram. Each histogram contains bins containing the number of feature values which fall within the bin range. The **Histogram** tab displays parameters that determine the content and resolution of the histograms.

The histogram groups, or color codes, the data according to the condition variable `faultCode` in **Group By**. Blue data is healthy and orange data is faulty, as indicated by the legend (color coding may appear different in your session). For feature values where the healthy and faulty labels overlap, the color appears brown due to the overlap between blue and orange.

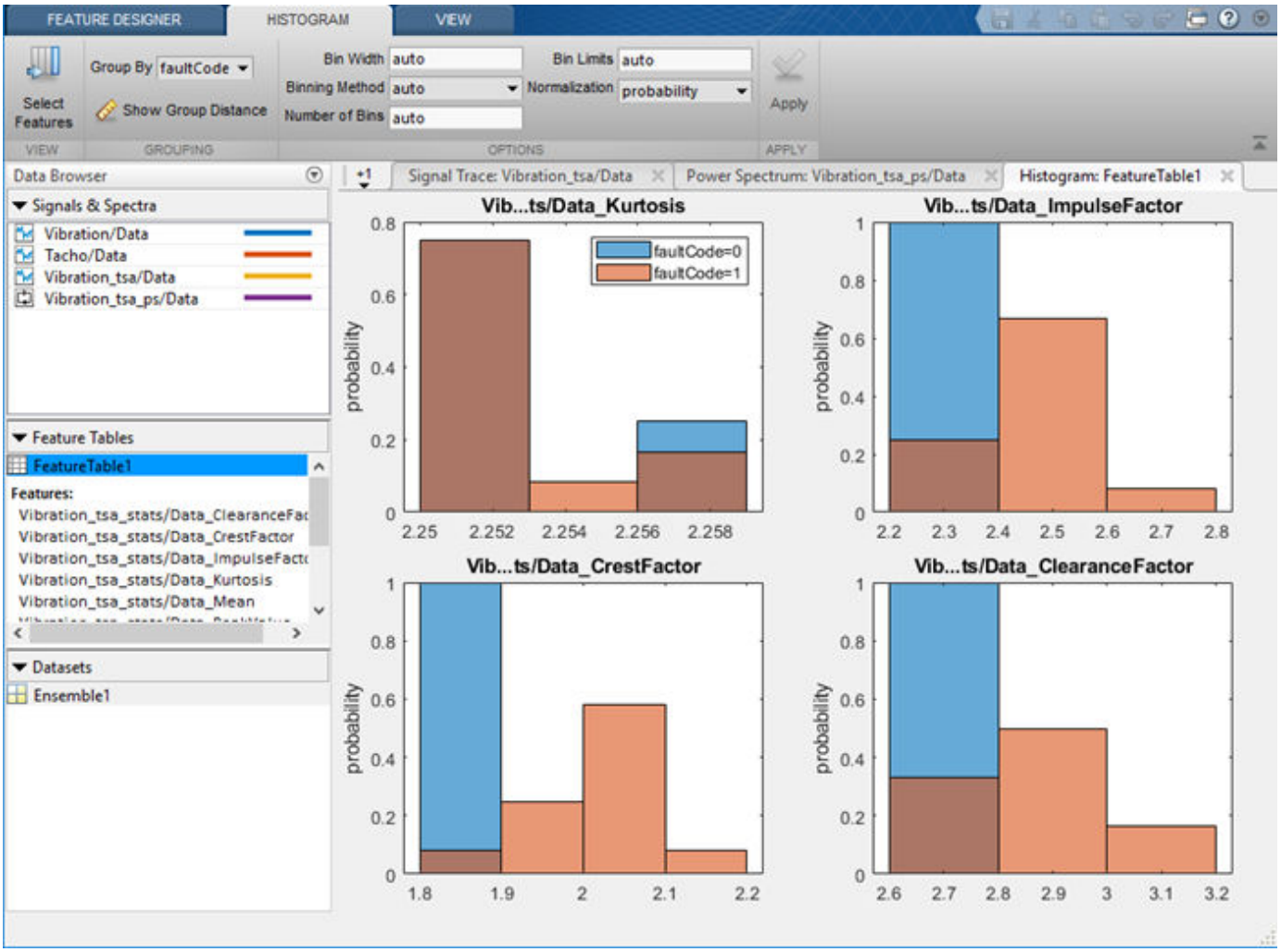


You can get a rough idea of which features are effective by assessing which ones clearly segregate blue data from orange data. `RMS` and `CrestFactor` appear effective. There are only small areas of overlap. Conversely, `Skewness` and `Kurtosis` have large amounts of overlap. These features appear ineffective for this data and this condition variable.

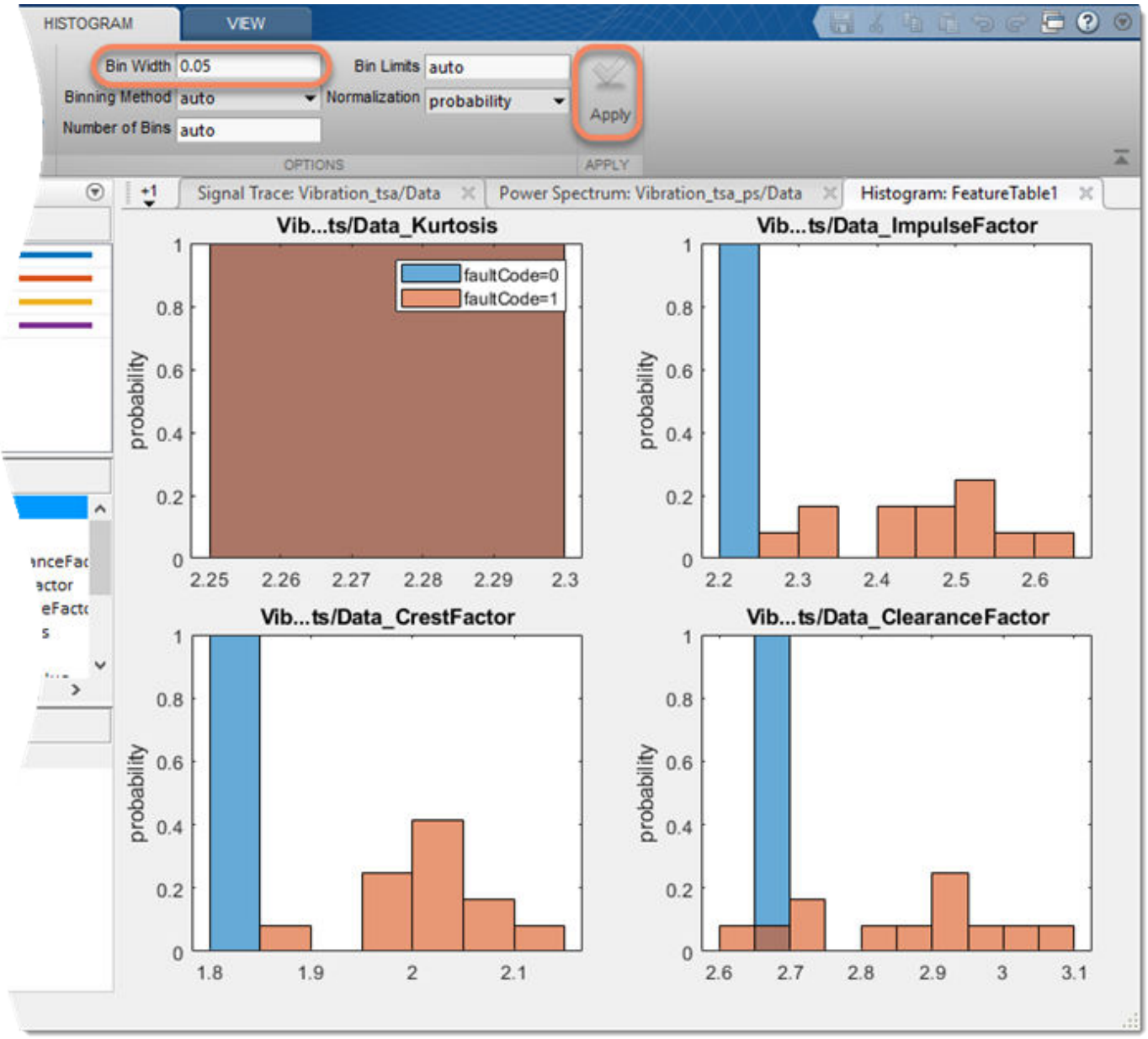
By default, the app plots the histograms for all the features in the feature table. You can focus on a subset of histograms by using **Select Features**. Use **Select Features** to limit the histogram plots to the first four in the feature table.



The histogram view now includes only the features you selected.

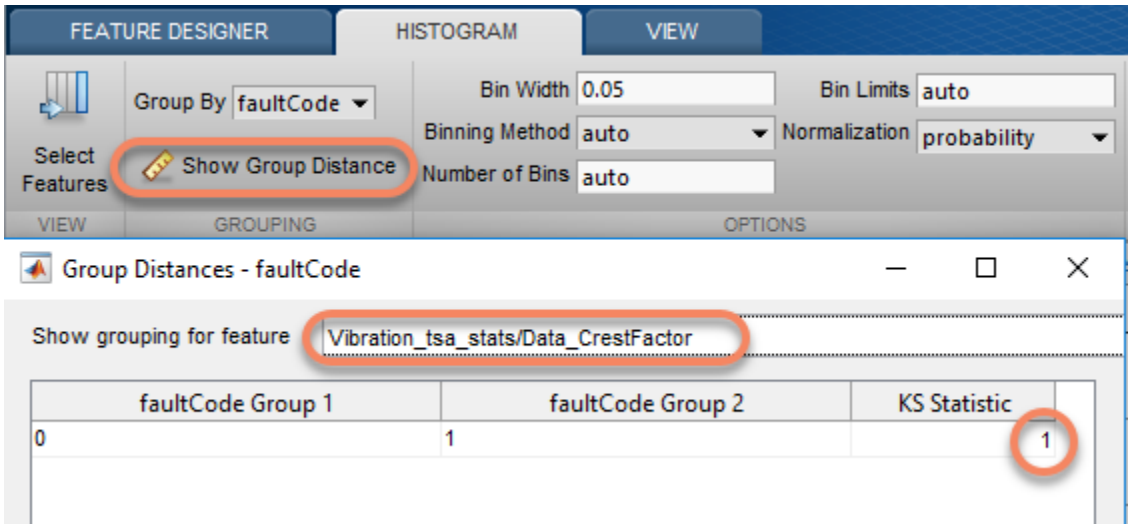


Control the appearance of the histograms using the parameters in the **Histogram** tab, which activates when you generate the histograms. The `CrestFactor` feature appears to separate healthy and unhealthy data almost completely. Investigate whether this result is sensitive to resolution. In the **Histogram** tab, the `auto` setting of bin width results in a resolution of 0.1 for `CrestFactor`. Enter a bin width 0.05, and click **Apply**.



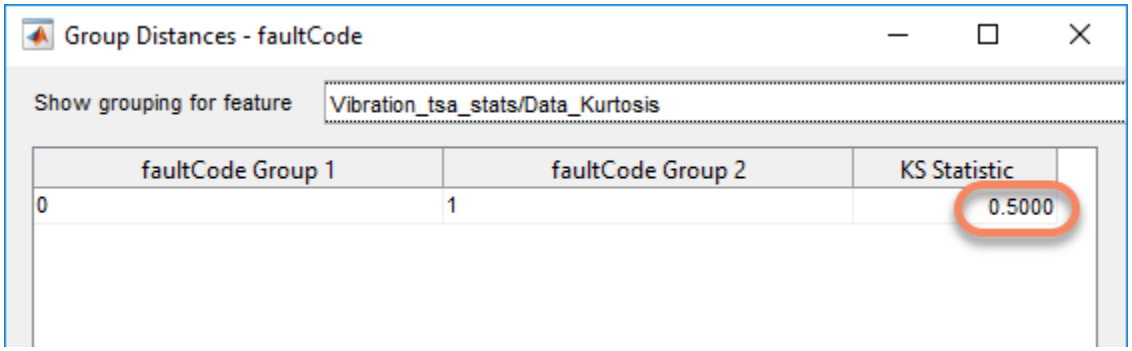
At this resolution, both `CrestFactor` and `ImpulseFactor` appear to completely separate healthy from faulty data. `ClearanceFactor` still has some mixed data, but to a lesser degree than with the larger bin width. `Kurtosis` had a smaller bin width of 0.002 with the `auto` bin width setting. Changing the bin width to 0.05 results in a single bin that contains all the `Kurtosis` data.

Histograms visualize the ability of features to separate healthy from unhealthy data. You can also get a numerical assessment using Group Distance. The group distance represents the separation between the healthy and unhealthy data distributions. Click **Group Distance**. In the dialog box, select `CrestFactor` in **Show grouping for feature**.



The group distance, represented by the KS Statistic, is 1. This probability value represents complete separation.

Next, select `Kurtosis`. The `Kurtosis` histogram showed substantial intermixing.

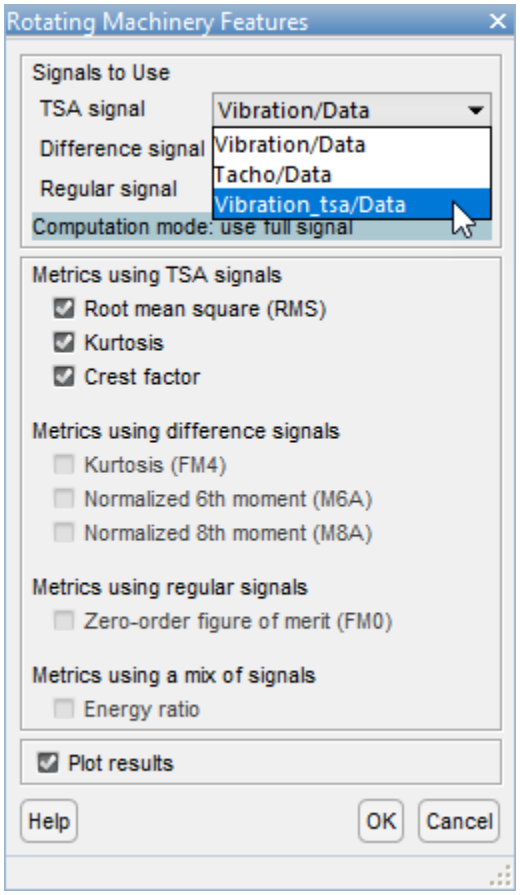


The KS Statistic in this case is 0.5, reflecting the intermixing in the histogram.

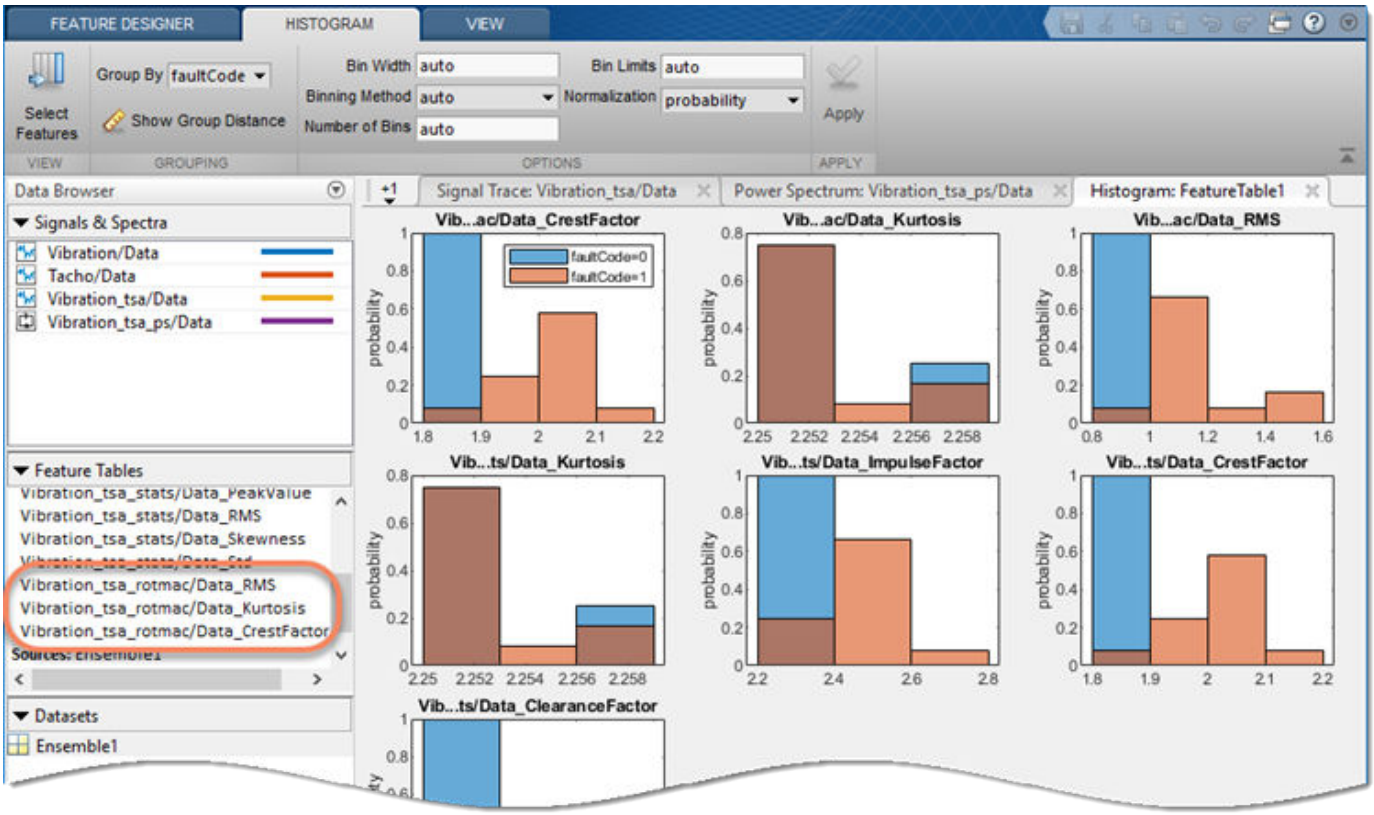
Restore **Bin Width** to `auto`.

Rotating Machinery Features

Since you have rotating machinery, compute rotating machinery features. Select **Time-Domain Features > Rotating machinery Features**. In the rotating machinery dialog box, select your TSA signal to analyze and select the TSA signal metrics.



Other feature choices in the dialog box use the filtered TSA difference and regular signals as a source. This example does not use the difference and regular signal-based features because computation for those signals assumes common shaft speed.

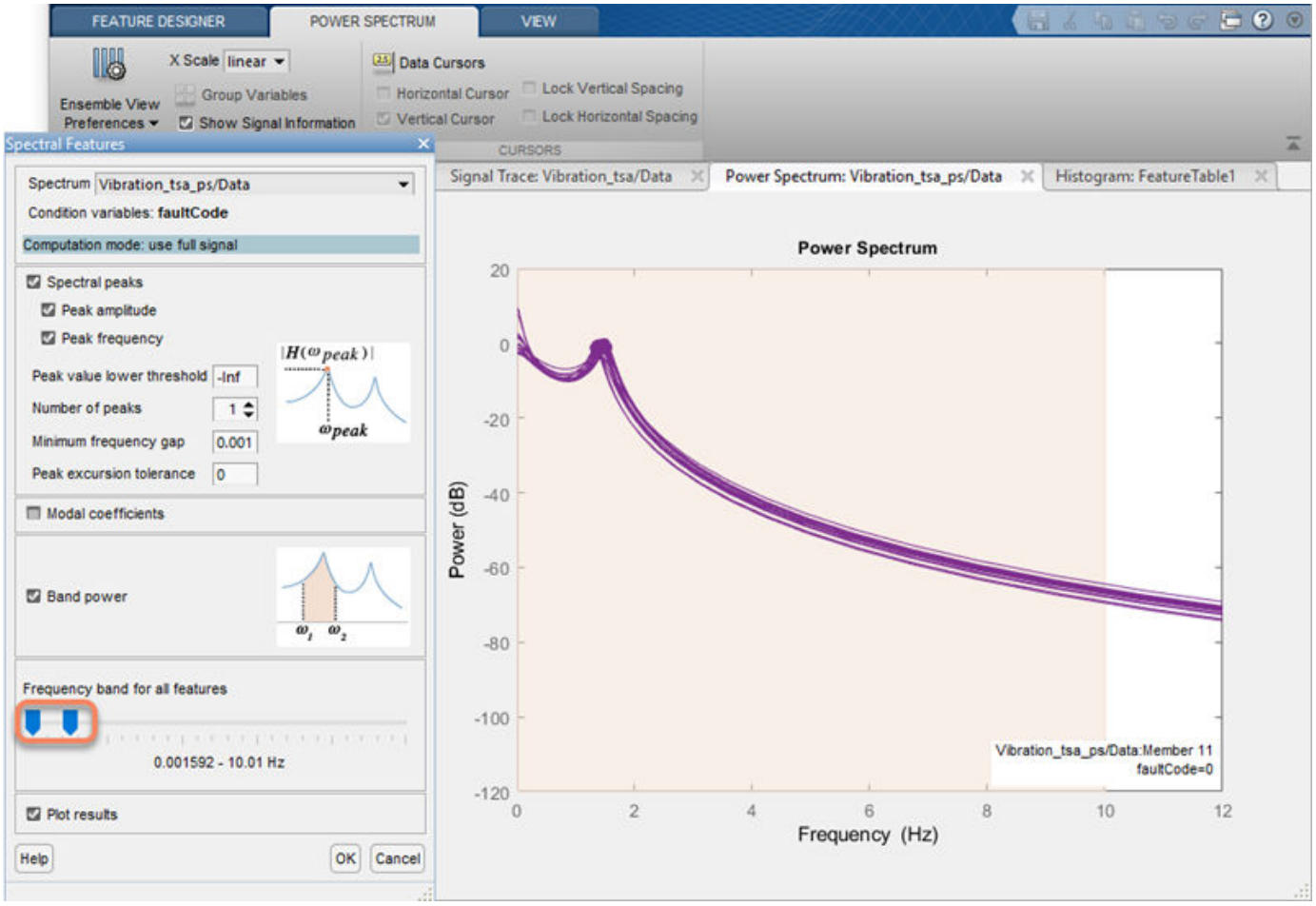


The app automatically adds the new features to the feature table and the **Select Features** list, and plots the new histograms at the top of the histogram display. `CrestFactor` and `Kurtosis` histograms are essentially the same whether they were computed as signal features or rotating machinery features, since both computations used the TSA signal as a source.

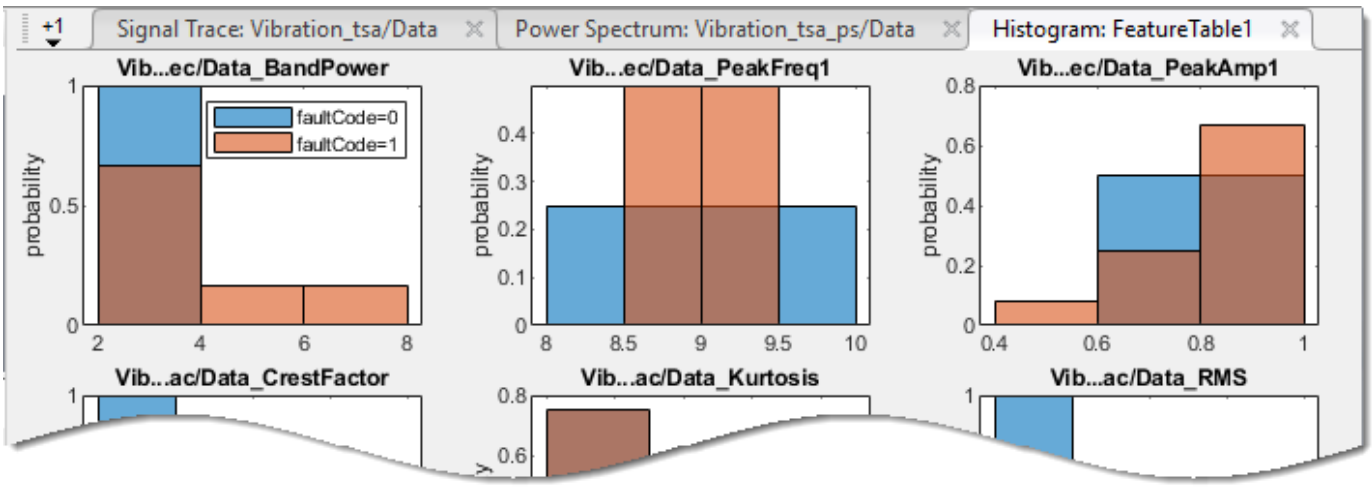
Spectral Features

Compute spectral features from the power spectrum you generated earlier. Click **Spectral Features**. In **Spectrum**, select `Vibration_tsa_ps/Data`.

Set the frequency band. The power spectrum x scale changes automatically from log to linear when you open the spectral features dialog box. When you move the frequency slider, the plot shades the region that the slider setting covers. To capture the power spectrum peaks efficiently, limit the frequency range to 10 Hz.

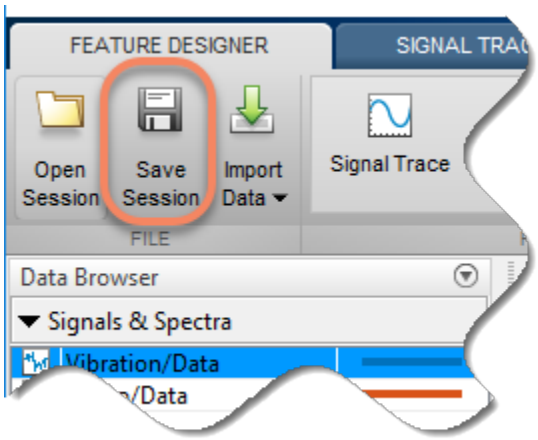


The histograms show substantial intermixing of healthy and unhealthy data in one or more of the bins for all three features.



You now have a diverse set of features.

Save your session data. You need this data to run the “Rank and Export Features in Diagnostic Feature Designer” on page 2-31 example.



Next Steps

The next step is to rank those features to determine which ones provide the best indication of system condition. For more information, see “Rank and Export Features in Diagnostic Feature Designer” on page 2-31.

See Also

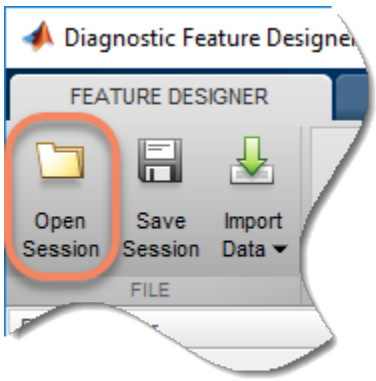
Diagnostic Feature Designer

More About

- “Import Data into Diagnostic Feature Designer” on page 2-6
- “Rank and Export Features in Diagnostic Feature Designer” on page 2-31
- “Explore Ensemble Data and Compare Features Using Diagnostic Feature Designer”
- “Interpret Feature Histograms in Diagnostic Feature Designer”
- “Data Preprocessing for Condition Monitoring and Predictive Maintenance”
- “Condition Indicators for Monitoring, Fault Detection, and Prediction”
- “Condition Indicators for Gear Condition Monitoring”

Rank and Export Features in Diagnostic Feature Designer

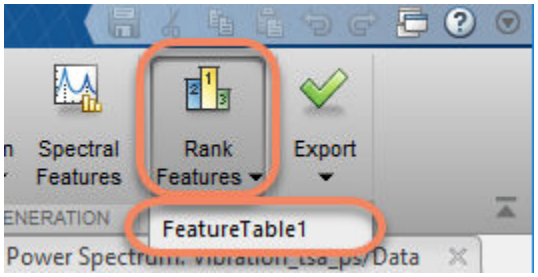
This example shows how to rank features using several classification ranking techniques, how to compare the results, and how to export features from the app. If you want to follow along with the steps interactively, use the data you imported in “Process Data and Explore Features in Diagnostic Feature Designer” on page 2-14, and uses the feature set from that example. Use **Open Session** to reload your session data using the file name you provided.



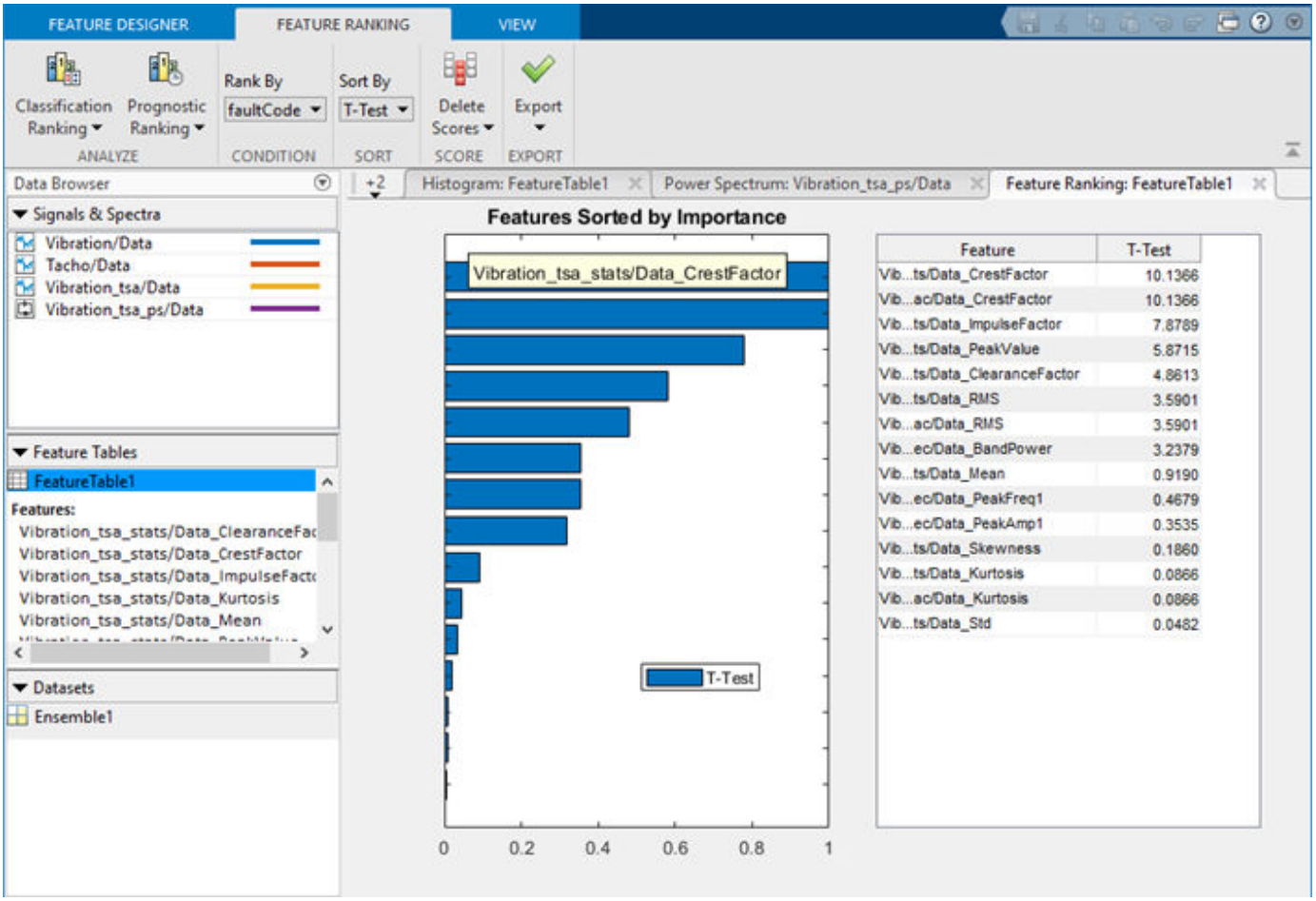
When you generate features for classification, you seek the features that perform best at distinguishing among data of different conditions. When you view the histograms, you can get an idea of the relative effectiveness of your features. In this example, you use ranking algorithms to perform this feature comparison more rigorously. Once you have chosen the features you want to retain, you export these features into the MATLAB workspace.

Rank Features

Rank your features using the default T-Test technique. Click **Rank Features**. Select `FeatureTable1`.



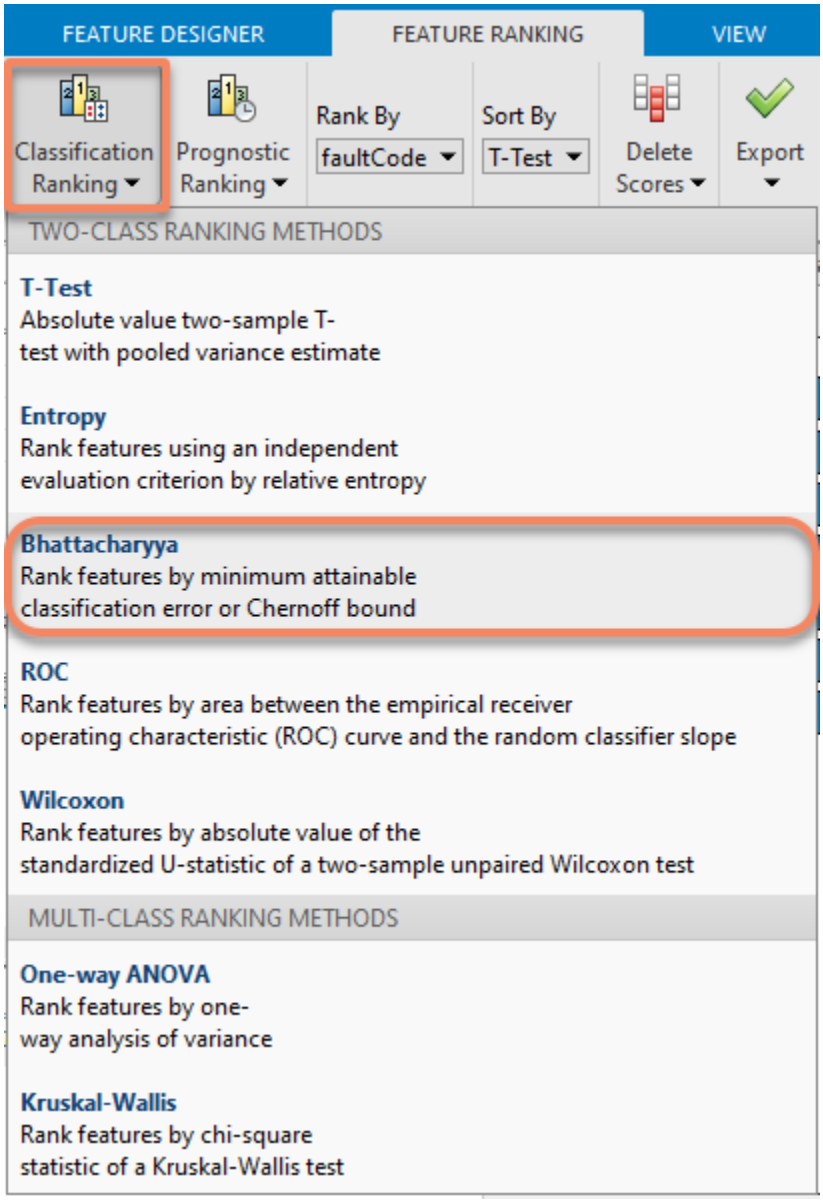
Your selection brings up a ranked list of features, displayed as both a bar chart and a numerical table.



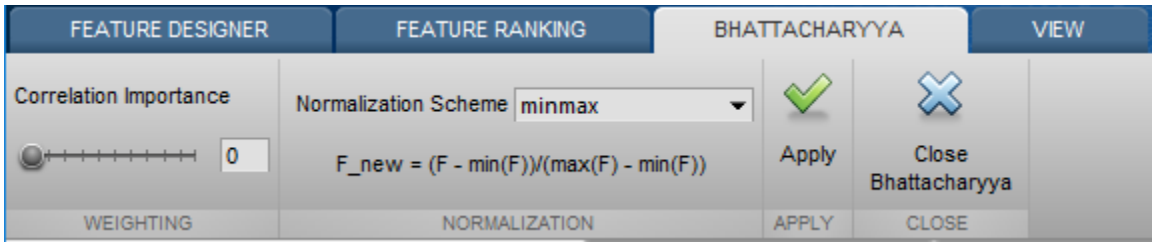
The bar chart legend shows that the initial ranking is performed using the T-Test algorithm. The chart is normalized to 1 to facilitate visual comparison. The table displays unnormalized ranking scores. Hover your mouse over a bar to get the full variable name.

Choose an Alternative Ranking Algorithm

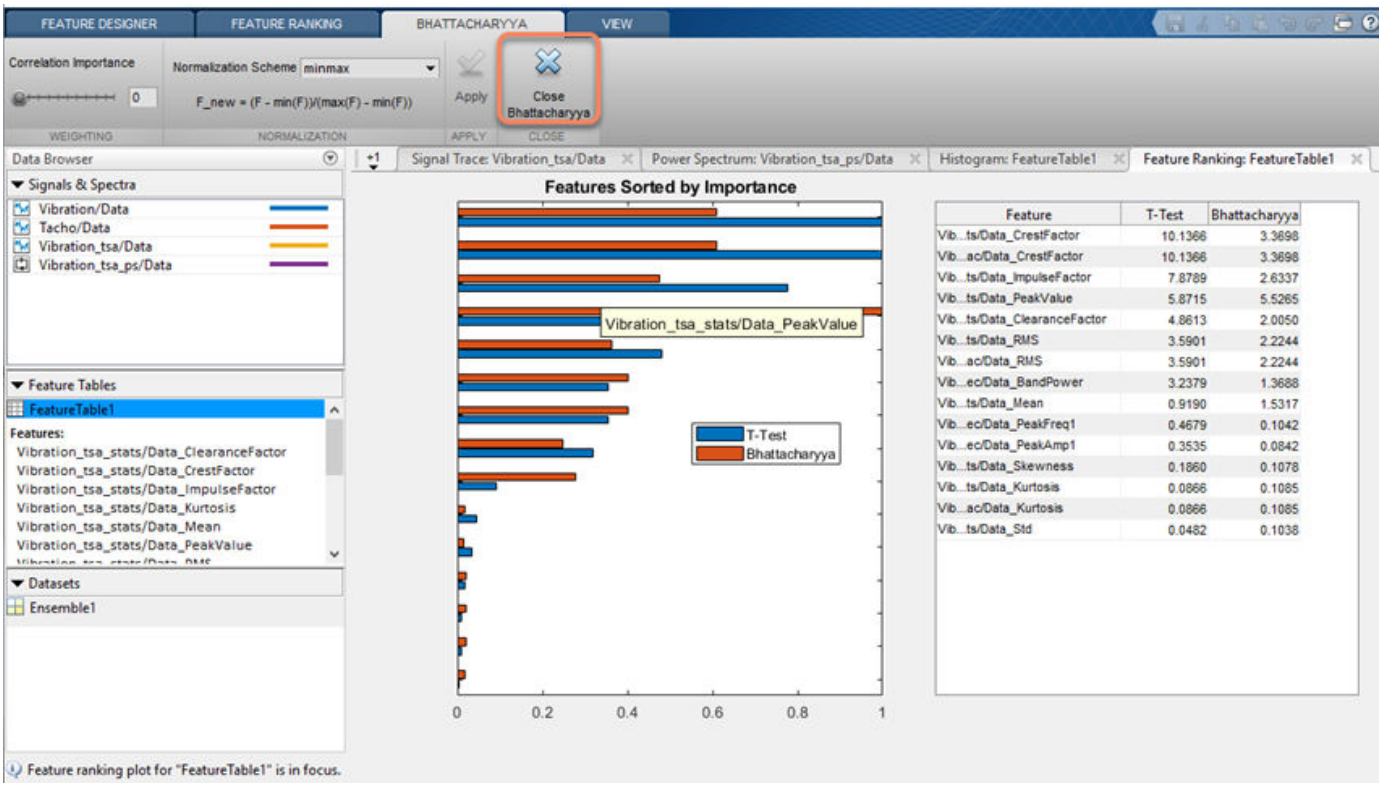
Each ranking algorithm uses different criteria to perform the ranking. In the **Feature Ranking** tab, click **Classification Ranking** to bring up a menu that summarizes each method. From that menu, select `Bhattacharyya`.



A **Bhattacharyya** tab opens with ranking specifications that are standard for all of the techniques. Click **Apply**.



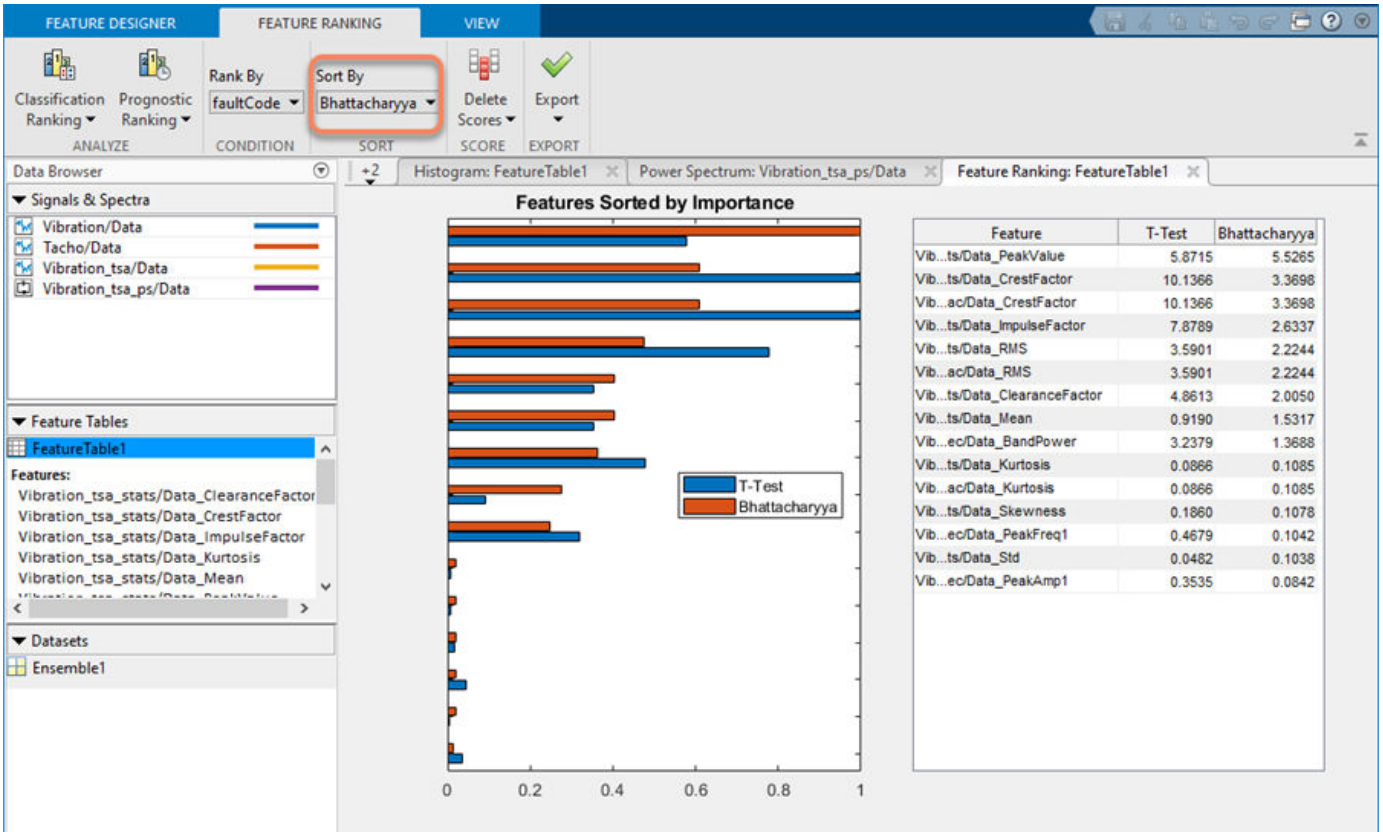
Apply updates the ranking display with the new results, displayed along with the original T-Test results.



The Bhattacharyya technique yields results that are similar to, but not identical to, the T-Test results. The highest ranking feature is `PeakValue` from the `Signal Statistics` set.

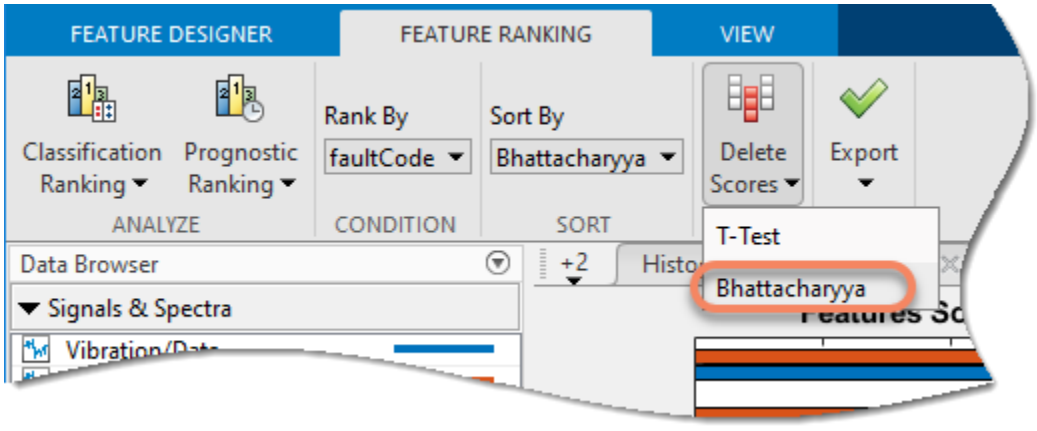
This feature was fourth in the T-test ranking. The crest factor features are still in the top three.

The ranking is still sorted by T-Test. Sort instead by Bhattacharyya. Close the **Bhattacharyya** tab and return to the **Feature Ranking** tab. Then select `Bhattacharyya` in the **Sort by** menu.

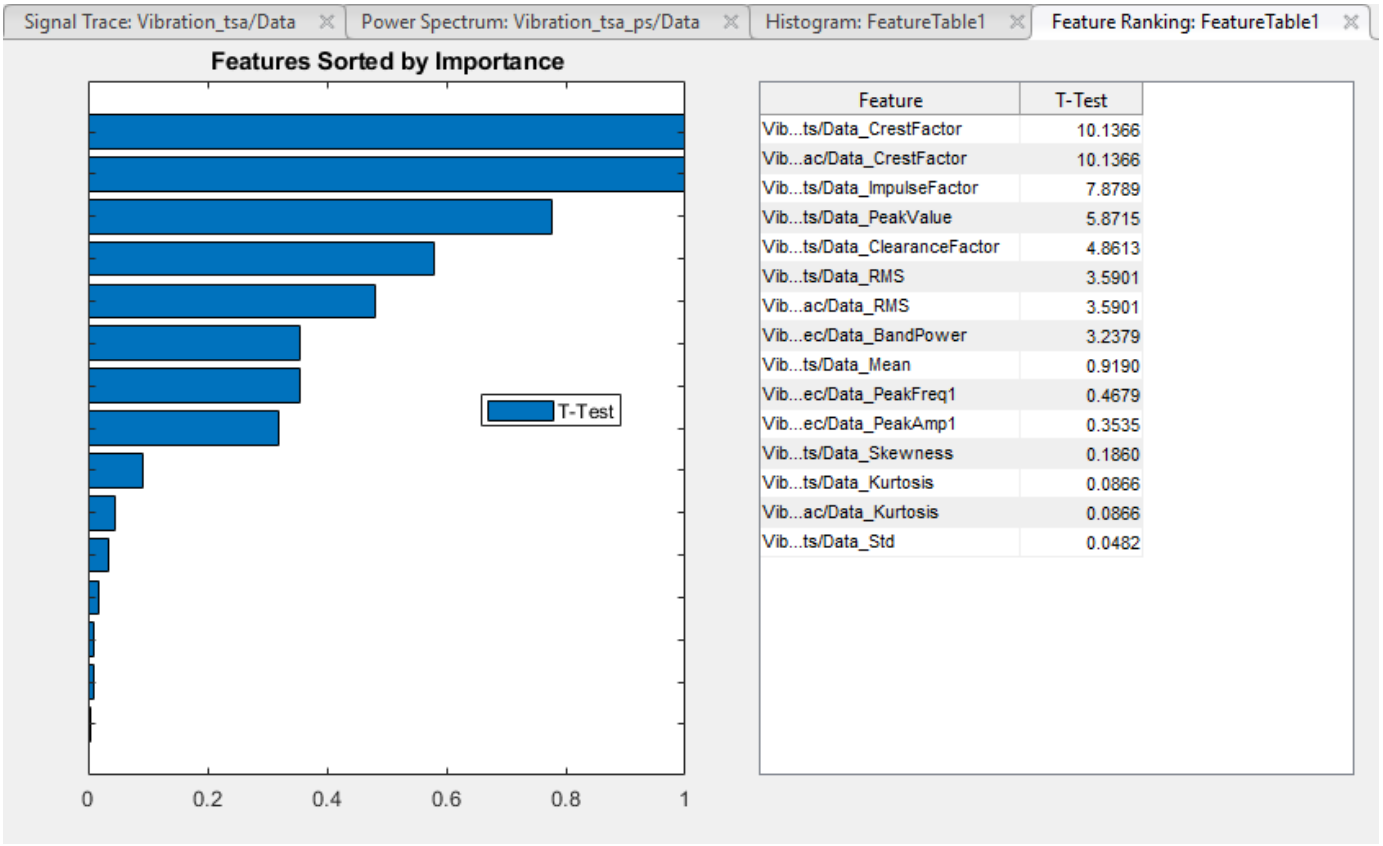


Delete a Set of Rankings

You have two sets of rankings. Now delete the Bhattacharyya results. In the **Feature Ranking** tab, select **Delete Scores > Bhattacharyya**.

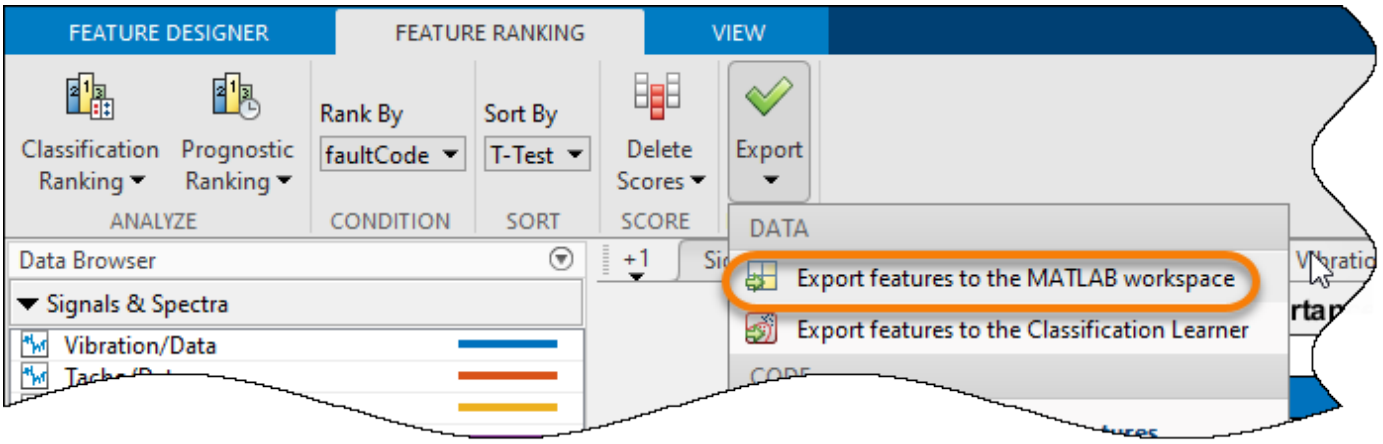


`Bhattacharyya` disappears from the ranking results.

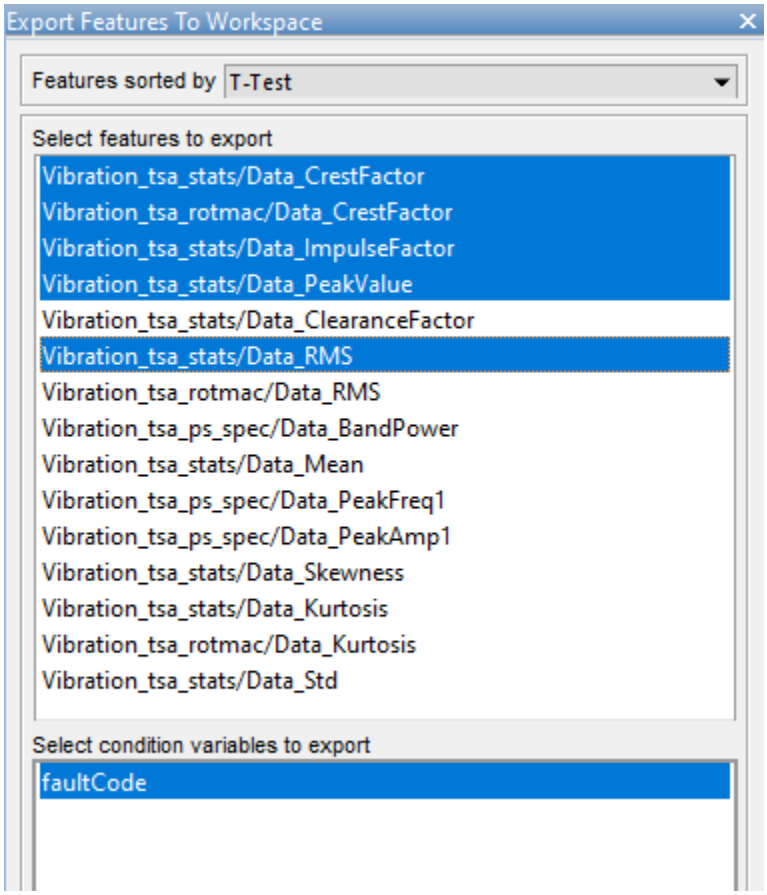


Export Features to the MATLAB Workspace

The final step in the **Diagnostic Features Designer** workflow is to export your features. In the **Feature Ranking** tab, select **Export > Export features to the MATLAB workspace**.



Select the features to export. The app preselects the top five features. Modify this selection. Clear the fifth selection and select the sixth feature using Ctrl-Click.



Your reduced feature table appears in your MATLAB workspace.

See Also

Diagnostic Feature Designer | `anova1` | `bhattacharyyaDistance` | `kruskalwallis` | `perfcurve` | `ranksum` | `relativeEntropy` | `ttest2`

More About

- “Process Data and Explore Features in Diagnostic Feature Designer” on page 2-14
- “Rank and Export Features in Diagnostic Feature Designer” on page 2-31
- “Explore Ensemble Data and Compare Features Using Diagnostic Feature Designer”
- “Perform Prognostic Feature Ranking for a Degrading System Using Diagnostic Feature Designer”
- “Data Preprocessing for Condition Monitoring and Predictive Maintenance”