Authors’ Affiliation

Shamim N. Pakzad, Lehigh University
S. Golnaz Shahidi, Lehigh University
Ruigen Yao, Lehigh University
Michael B. W. Chamberlain, University of California Berkeley

Acknowledgement

Research funding is partially provided by the National Science Foundation through Grant No. CMMI-1351537 by Hazard Mitigation and Structural Engineering program, and by a grant from the Commonwealth of Pennsylvania, Department of Community and Economic Development, through the Pennsylvania Infrastructure Technology Alliance (PITA). The authors would like to thank Mallory Nigro and Andrew Thorsen who coded the preliminary version of DIT.
# Table of Contents

| Topic               | Page ||---------------------|------|| Introduction        | 1    || Getting Started     | 2    || Using DIT           | 4    || Troubleshooting     | 16   || References          | 17   |

# Developer Information

<table>
<thead>
<tr>
<th>Developer</th>
<th>Email</th>
<th>Phone</th>
<th>Fax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shamim N. Pakzad</td>
<td><a href="mailto:snp208@lehigh.edu">snp208@lehigh.edu</a></td>
<td>(610) 758 - 6978</td>
<td>(610) 758 - 5553</td>
</tr>
<tr>
<td>S. Golnaz Shahidi</td>
<td><a href="mailto:sgs310@lehigh.edu">sgs310@lehigh.edu</a></td>
<td>(610) 758 - 4543</td>
<td></td>
</tr>
<tr>
<td>Ruigen Yao</td>
<td><a href="mailto:ruy209@lehigh.edu">ruy209@lehigh.edu</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael B. W. Chamberlain</td>
<td><a href="mailto:mchamberlain31@berkeley.edu">mchamberlain31@berkeley.edu</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Intervention

Vibration-based structural health monitoring (SHM) aims to develop methodologies to assess the condition of existing structures and identify damage in its earlier stages [1-2]. One of the main categories of SHM is data-driven damage detection methods which mainly work on the basis of extracting damage sensitive features from measured structural responses and monitor them for any potential change due to damage. There are several examples of applications of these model-free approaches in the literature [3-9]. Although these methods are proved to be efficiently effective in identifying the structural damage, their performance depends on statistical analysis to distinguish between common changes in structure’s response (e.g., due to measurement noise or environmental factors) and damage [10-12]. Damage Identification Toolsuite (DIT) was developed to facilitate applications of multiple data-driven algorithms to improve damage detection performance through incorporation of various combinations of damage features and statistical tests.
Getting Started

Product Overview

Why use Damage Identification Toolsuite (DIT)?

Use DIT if you want to...
- extract regression-based damage features from measured responses of a structure
- perform change point analysis to statistically test the significance of change in damage-sensitive features
- develop and test change point statistics from any data outside SHM

DIT allows users to...
- choose between univariate and multivariate regressions models for feature extraction
- choose between substructural and general multivariate regressions models for feature extraction
- offer multiple distance measures to compare features extracted from different states of structure’s health
- offer several methods to compress the residuals of created autoregressive models on the data
- apply various change point techniques
- visualize raw data as well as processed control statistics

About Damage Identification

What is data-driven damage detection?

Data-driven damage detection is the process of identifying damage in a structure through inspecting extracted information from monitoring data for signs of change and anomaly.
**What process does DIT use to perform damage identification?**

DIT is a package of data-driven damage detection algorithms coded in MATLAB [13]. DIT offers several damage detection features to be extracted from vibration data along with multiple control statistics and thresholding methods to improve the decision making quality regarding the structure’s condition. Figure 1 summarizes all the procedures implemented in DIT.

![Flowchart illustrating implemented algorithms in DIT](image)

**Figure 1:** Flowchart illustrating implemented algorithms in DIT

**What is change point analysis?**

Change point analysis is a tool to identify whether a change in a process is due to random variations of measurements or assignable causes; in the case of SHM, damage. DIT enables applications of several change point statistics on extracted damage sensitive features or any data collected outside SHM field.
Using DIT

Welcome

To load DIT, enter DIT into the MATLAB command window and press enter. A welcome window displaying the title of the program and information about the developers will appear. Clicking the New Session button will clear the workspace, close all existing MATLAB windows and open the Import Data page.

Figure 2: DIT welcome screen that is displayed when “DIT” is typed into the MATLAB command window
**Figure 3**: DIT Import data screen that is displayed after uploading dataset(s)

1. The **From File** button allows the user to upload .mat or .txt format dataset(s) into the GUI. The data files should be stored in DIT DATA subfolder in matrix format where every column contains the time history data from one sensor. In general if several datasets from one sensor network is available, a series of .mat or .txt files will be uploaded. The order of the files in the DIT DATA folder will be the order they appear in the GUI. The **Reset** button clears the workspace, clears the files uploaded to the GUI and resets the entire page to its base state. **Note**: Page
13 of this manual provides some general points on inputs and outputs of DIT.

2. The **Sampling Frequency** input box allows the user to input the sampling frequency of the data input. This value is used in the substructural methods of the damage feature creation page.

3. After data is uploaded to this page, the dropdown menu will appear, populated with the names of the uploaded files. Each of these files can be plotted in the axes below. Note that the x-axis represents the number of samples rather than time.

4. The **Loaded Data** panel displays the number of files currently uploaded, the name of the file that is currently selected (displayed in the axes to the right), and the size of the uploaded datasets.

5. This data segmentation option is only visible if a single file is uploaded. This option allows the user to segment a single, presumably large, dataset into a number of smaller ones. The **Segment size** input determines the length of each new dataset and the **overlap** input determines how many samples from the previously created segment are used in the next created segment.

6. The **Label Figure** panel allows the user to label the axes to the right and export the figure to file.

7. The **Select Data Range to Display** panel allows the user to specify certain limits for the x-axis in the plots of data. Clicking the **Restore Original** button will revert the zoom and display the full extent of the plotted data.

8. Clicking on this button will direct the user to the change point analysis page. This provides the option to perform change point analysis on damage features
generated outside DIT as well as any data collected without damage detection concerns.

9. Clicking this button will send the user to the damage feature page to create a variety of damage sensitive features from the uploaded data series.

**Damage Feature Creation**

![Damage Feature Creation Page](image)

**Figure 4:** DIT damage feature creation page

1. This box allows the user to select whether to apply a multivariate regression model that compares multiple sensors to each other, or a univariate regression
model that examines only a single signal at a time. If the data loaded to this page is from a single sensor reading (i.e. the uploaded dataset(s) have only one column), only the univariate option is displayed.

2. This info box displays the number of files and size of those files sent to this page from the import data page.

3. This box contains all of the general regression models along with their required parameter inputs. Details regarding these models and their application for structural damage detection can be found in [11]. Note: page 13 of this manual provides more information about input parameters of these models.

4. This box contains the two substructural regression models designed for multistory frames with linear topology. Information about development and validation of these algorithms is found in [14]. Note: page 14 of this manual provides more information about input and outputs of these models.

5. When a model is selected, Select Desired Outputs panel will display checkboxes of the various outputs that can be produced from the selected model. Figure 5 shows this panel when AR model is selected. More information about these features is available in [11, 15].
Figure 5: Output menu displayed when AR model is selected

6. Pressing this button initiates the regression process. Assuming all the required inputs have proper values, the regression will occur and the entire screen will gray out and become disabled. If not all inputs are correctly input, the relevant error message should display and tell you what is missing.

7. The Clear Page button resets the entire page back to its original state. The data that was brought in from the import data page remains stored in this page and that same data can be used again in a different regression model if desired. Note that this button clears the workspace to prevent cluttering and if you want to keep the features you have already created you should save them to file before clearing the page.

8. The Save to File button allows you to save your created outputs. If multiple outputs were created, a single .mat file containing all the output variables will be saved to the location specified by the user.
9. Clicking this button causes the change point analysis page to open. The damage features created last (the ones in the workspace) will be brought to next page and be available for a change point analysis. The damage feature page will remain open.

Change Point Analysis

![Change Point Analysis](image)

**Figure 6**: DIT change point analysis page

1. The **Reset Page** button here does not clear the workspace; rather it resets the
page and allows for the analysis of a different feature. All created change point statistics and original damage features remain in the workspace.

2. The dropdown menu in the top left allows the user to choose what damage feature to analyze. The menu is automatically populated with features created in the damage feature page or with the uploaded datasets if this page is proceeded directly from the import data page. There is also always the option ‘Upload New Feature’ along with the default selections. If this is selected, an input box to the right becomes visible that allows the user to input the name of a variable in the workspace that the user wishes to analyze.

3. This list of buttons allows the user to apply several univariate control statistics to monitor a potential change in a process. Available change point statistics are: Cumulative Sum, exponentially weighted moving average (EWMA), mean square error (MSE), modified mean square error (ModMSE), two-sample t-test, and two-sample normalised likelihood ratio (NLR) test. More information regarding the application of these statistics can be found in [10-12]

4. Clicking this button after selecting a damage feature to analyze and a change point statistic to apply will create the chosen statistic to all sensor groupings in the damage feature file. A plot of all the sensor groups will display and thresholding options will become visible afterwards. Additionally, the created change point statistic matrix will be sent to the workspace.

5. This is the plot area where the change point statistics and thresholds are displayed. In this plot, the x axis is the index of features created from an unknown state of system. Therefore, if a Reference Run number is specified in
the Damage Feature Creation page, DIT assumes a known state for features with indices lower than specified Run number and does not include their indices in this plot.

6. The Threshold Selection box provides various methods of thresholding that can be applied to the created change point statistics [16]. Input parameters for the resampling-based thresholding methods are number of bootstraps or permutations, while in significance-based thresholding method user is required to enter the desired confidence level for the threshold.

7. This dropdown menu is populated with each sensor/sensor grouping’s number. Details about what each sensor grouping number represents can be found in page 13 of this manual document. Initially, all the change statistics series for every sensor grouping is displayed, but any individual grouping scheme may be displayed instead by opening this drop down menu. It is good to note that all the thresholding methods can be applied to individual sensor groups as well as the all sensor groups.

8. This box reminds the user what change statistic is currently being plotted.

9. This panel allows the user to label the plot and export the figure to the current folder.

10. This allows the user to zoom in on any section of the plot as defined by lower and upper x-axis bounds. Clicking the Restore Original button will cause the entire plot to be displayed again.
General points on inputs and outputs of DIT

- The raw data input to the first page should take the form of an $M \times N$ matrix where $N$ is the number of channels and $M$ is the number of samples. When a single long time series is uploaded the user is provided with the option of data segmentation within the GUI.
- Every output (with a few exceptions) of the damage feature creation page will be of the same format. They will be a size $A \times B$ matrix where $A$ represents the number of datasets (or data segments) and $B$ represents the number of sensor groups (for general regression methods) or the vector of outputs (for the substructural methods).
- The exceptions to previous point are AR and ARX alpha coefficients and the uncompressed residuals from the AR model. These outputs are not scalar values and cannot be represented as a single matrix. In these cases, the format of the outputs will be $A$ cells of size $M \times N$ where $A$ represents the number of datasets (or data segments), $M$ represents the length of each output vector and $N$ represents each sensor grouping.

Input parameters of general regression models

General Regression panel allows the user to choose from Single Variate Regression (SVR) models, Collinear Regression (CR) models, Autoregressive (AR) models, and Autoregressive with exogenous term (ARX) models. Regardless of the model type,
grouping of the sensors should be specified, either by importing a .mat or .txt file containing a matrix of size M x N where M is the number of sensor groups to be processed and N is the number of sensor nodes in each group. For SVR and ARX models, N is 2, whereas for AR and CR models, N is 1 and 3, respectively. The second option is to generate the grouping scheme inside DIT by selecting the **Generate Scheme for sensors** and inserting the desired sensor numbers in the box below that. The sensor numbers should be entered comma delimited. DIT will automatically generate a grouping scheme of all the permutations of the sensors numbers provided by the user. For example, inserting [5, 20:21] in the grouping scheme box creates 6 sensor groups if SVR, CR, and ARX models are selected, and gives a grouping scheme of [5; 20; 21] if AR models are chosen. The grouping scheme will be outputted to the MATLAB workspace so the user can see each column of damage features corresponds to which sensor grouping.

**Input parameters of substructural regression models**

Substrucural Regression panel offers two options for estimating stiffness parameters of structural models with linear topology: Time Domain Regression model (TDRM), and Frequency Domain Regression model (FDRM). **Nodal Mass String** should be a comma or space delimited input. Files input to the import data page should be of size M x N where M is length of each dataset and N is number of degrees of freedom (DOF) of the structure in question. Nodal mass string input should be of length N. Default displacement reconstruction window length and
max correlation lag inputs will appear in boxes although these values can be edited. The reconstructed displacement and velocity data input boxes are optional. Two types of structures are available for these models: (1) multistory shear buildings, and (2) single span lumped mass bridge models. Outputs created by this model are interstory stiffness estimates. Each files k output will be of size \((N-1)^*2^1\) or \((N-1)^*2^2\) depending on whether the structural type is multistory shear building or bridge. In the shear building case, the first output will be the interstory stiffness between the floor and 1\(^{st}\) DOF while the ith subsequent pair of outputs refer to estimates of stiffness between DOF i and i+1. In the bridge case, the extra term refers to the stiffness estimate between DOF N and ground. The units for mass, stiffness and time histories (displacement, velocity, and acceleration) are user-defined and should be compatible across the variables and kept consistent throughout the analysis (e.g. adopting either the SI units or the English units for all variables).

Another note on the length of time series to be used for the TDRM and FDRM: generally for same estimation problem longer data should be used for the former than the latter, this is because TDRM involves the ill-posed displacement reconstruction and does not take damping into consideration, and thus tends to yield less accurate results. The specific minimum signal length depends on its frequency contents and the sampling rate, and could be determined on a case-by-case basis using convergence analysis. The general recommendation for performance is a signal length no less than 20 periods of the system fundamental mode.
## Troubleshooting

### Damage Feature Creation

**Table 1**: Empty form error messages that may appear when **Perform Regression** is clicked in **Damage Feature Creation** window.

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling Frequency Required</td>
<td>DIT will not begin the process until all necessary inputs are provided by the user. If you receive any of these error messages, resolve the problem by completing the indicated box before clicking <strong>Perform Regression</strong>.</td>
</tr>
<tr>
<td>Model order not defined</td>
<td></td>
</tr>
<tr>
<td>No structure type selected</td>
<td></td>
</tr>
<tr>
<td>Mass String required</td>
<td></td>
</tr>
<tr>
<td>Type of output is not specified</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**: Invalid value error messages that may appear when **Perform Regression** is clicked in **Damage Feature Creation** window.

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of <strong>Reference Runs</strong> for Mahalanobis Distance calculation is insufficient</td>
<td>The reference runs should be larger than the default value specified by DIT. The minimum Reference runs value for a p-order AR and ARX model is p and 2p+1 respectively.</td>
</tr>
</tbody>
</table>

### Change Point Analysis

**Table 3**: Empty form error messages that may appear in the **Change Point Analysis** window.

<table>
<thead>
<tr>
<th>Error Message</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify number of bootstraps</td>
<td>The required input parameter for the selected thresholding method should be specified.</td>
</tr>
<tr>
<td>Specify number of permutations</td>
<td></td>
</tr>
<tr>
<td>Specify confidence level</td>
<td></td>
</tr>
</tbody>
</table>
References


