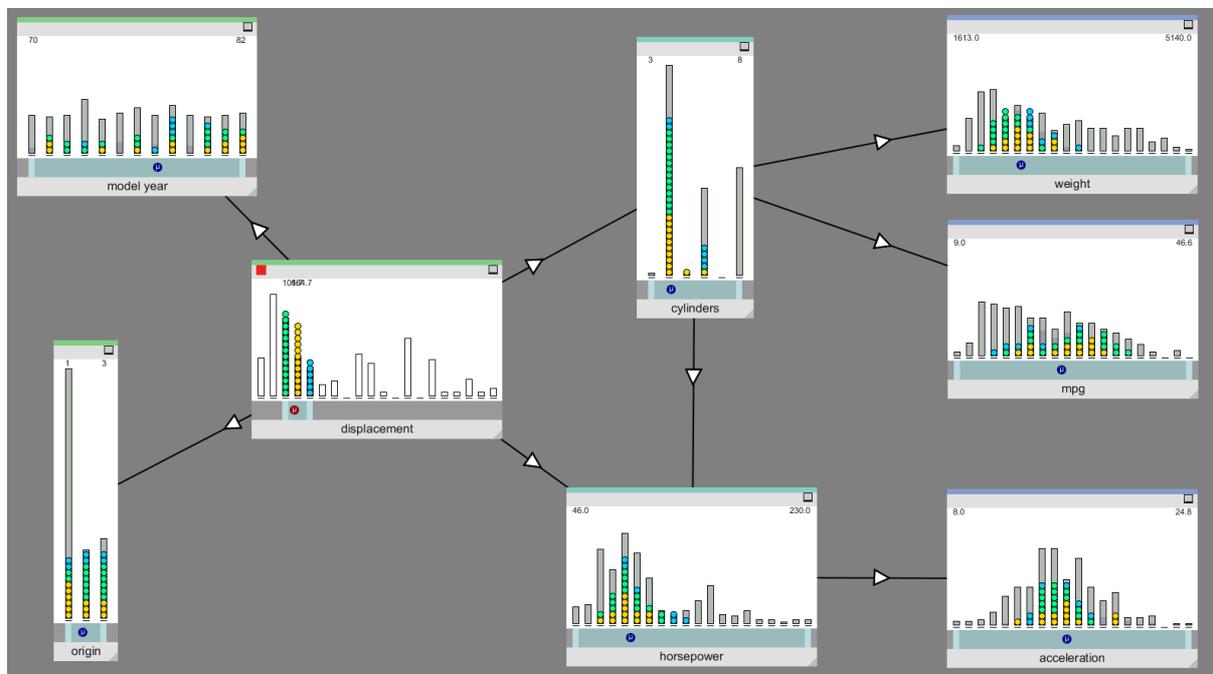


# Omnigram Explorer User Documentation

Version 1.0

Tim Taylor, 11 December 2014

---



*A Simple Interactive Tool for Exploring Complex Computational Models*

# Table of Contents

[Introduction](#)

[Installing Omnigram Explorer](#)

[Data Files and Formats](#)

[The Model Definition File \(XML format\)](#)

[<general>](#)

[<appearance>](#)

[<nodes>](#)

[<node>](#)

[<parents>](#)

[<parent>](#)

[The Model Data File \(CSV format\)](#)

[The User Interface](#)

[Nodes](#)

[What the Node Displays](#)

[Mouse interaction with Nodes](#)

[Modes of Interaction](#)

[Single Node Brushing](#)

[Multi Node Brushing](#)

[Brush Links](#)

[Show Individual Samples](#)

[Sample Modes](#)

[Causal Links](#)

[The help screen](#)

[Further suggestions for usage](#)

## Acknowledgements

*Omnigram Explorer* was developed by Tim Taylor, Alan Dorin and Kevin Korb at the Faculty of Information Technology at Monash University, in collaboration with Jodie McVernon, James McCaw, Nicholas Geard and Patricia Campbell at the Melbourne School of Population and Global Health at the University of Melbourne. The project was funded by ARC grant number DP110101758 (“*Development of an 'ageing household' model for assessing medium to long-term vaccine impact in populations*”, Dr Jodie McVernon, Dr Kevin B Korb, Dr Kathryn Glass, Dr James McCaw, Dr Emma McBryde).

Any queries regarding Omnigram Explorer should be sent to [omnigram@tim-taylor.com](mailto:omnigram@tim-taylor.com)

## Introduction

Omnigram Explorer (OE) is a tool for the interactive exploration of relationships between variables in a system. It is designed to help researchers gain a holistic, qualitative understanding of their data, which might highlight relationships that warrant further quantitative investigation. It is also a useful tool to allow non-specialists to explore the behaviour of complex systems.

At startup, the user is prompted to select a model definition file that describes the system's variables and also points to a data file. The data file contains pre-calculated samples of values for each variable. OE loads this data (or a specified subset of it) and presents it to the user in graphical form for interactive exploration. OE is agnostic about the particular form of the system and its variables: variables may be designated as inputs or outputs if desired, but this is not required.

The model definition file may also contain information about causal relationships between variables, which can then be visualised to provide hints to the OE user of interesting relationships to explore. Information about causal relationships might have been generated from existing knowledge of the system or from tools such as Bayesian causal network discovery software.

OE provides various modes of interaction that allow the user to explore the relationships in the system in different ways. A guiding principle for OE's design has been to harness the strengths of the human visual system to allow relationships between multiple variables to be visualised simultaneously. It is equally useful when exploring models with several dozen variables as it is with simpler models.

The rest of this document describes in more detail how to use the system and how to work with your own data.

## Installing Omnigram Explorer

The OE webpage can be found at <http://www.tim-taylor.com/omnigram>. From there you can obtain versions of OE for Windows, Mac and Linux. The page also has links to documentation and other related information.

From the webpage, download the software package suitable for your computer. The software is distributed as a compressed archive file (.zip for Windows and Mac, .tar.gz for Linux). When you have downloaded this file, unpack it in a suitable location on your computer. This will create a directory called `omnigram.xxx` (where `xxx` specifies the particular version you have downloaded). Within this directory you will find an executable file to run OE, named `omnigram.exe` (on Windows), `omnigram.app` (on Mac) or simply `omnigram` (on Linux).

## Data Files and Formats

When OE is run, it first asks you to select a model definition XML file to be opened. OE comes with an example XML file, `auto-mpg-loader.xml`, and associated data file, `auto-mpg.data`. These are found in the `data` subdirectory.

In order to load your own data into OE, two files are required

- A model definition file (in XML format)
- A model data file (in CSV format)

The format of these files is described in the following two subsections.

### The Model Definition File (XML format)

The overall structure of the Model Definition File is as follows:

```
<model>
  <general> ... </general>
  <appearance> ... </appearance>
  <nodes>
    <node> ... </node>
    ...
  </nodes>
</model>
```

A fragment of an example Model Definition File is shown below. As demonstrated in this example, the file may optionally begin with an XML declaration line (line 1), and comments may be inserted anywhere in the file (e.g. lines 3-9).

The required content of the file begins on line 11 with the `<model>` tag. A model should contain one compulsory `<general>` element, an optional `<appearance>` element, and one `<nodes>` element (containing one or more `<node>` elements).

Each of these components is described below.

```

1 <?xml version="1.0"?>
2
3 <!--
4 This file defines an example model suitable for loading into Omingram Explorer
5 (http://www.tim-taylor.com/omnigram).
6
7 The data used by the example is from the Auto-Mpg dataset available from the
8 UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/datasets/Auto+MPG
9 -->
10
11 <model>
12
13     <general data="auto-mpg.data" has-labels="true" label-filecol="9" live="false">
14         <label>Auto MPG</label>
15     </general>
16
17     <appearance node-bin-scale-factor="5.0" node-default-height="150" node-default-width="330"
18         min-internode-gap="20"
19         num-root-cols="2" num-inter-cols="1" num-leaf-cols="1"
20         show-causal-links="true">
21     </appearance>
22
23     <nodes>
24
25         <node id="2" datatype="discrete" min="70" max="82" role="root" filecol="7">
26             <label>model year</label>
27             <parents>
28                 <parent id="1" />
29             </parents>
30         </node>
31
32         <node id="3" datatype="discrete" min="1" max="3" role="root" filecol="8">
33             <label>origin</label>
34             <parents>
35                 <parent id="1" />
36             </parents>
37         </node>
38
39         <node id="1" datatype="continuous" min="68.0" max="455.0" role="root" filecol="3">
40             <label>displacement</label>
41         </node>

```

*Fragment of an example Model Definition XML file*

## <general>

### Description:

- Defines general features of the model
- The <general> element is required

### Attributes:

- data
  - *Required:* Yes
  - *Format:* String
  - *Description:* specifies model data file (CSV format assumed). The file must be placed in OE's data subfolder.
- num-samples
  - *Required:* No
  - *Format:* Integer

- *Description:* If specified, OE will select this number of samples at random from the data file. If not specified, all samples will be used. The maximum number of samples that OE can comfortably cope with depends on the hardware on which it is run (processor speed is the key factor). In general, it should be able to cope with 1000 samples or more without problem. See also `rng-seed`.
- `rng-seed`
  - *Required:* No
  - *Format:* Integer
  - *Description:* This is the seed number used to initialise OE's Random Number Generator. It can be useful to specify this if `num-samples` is specified: a specific value for `rng-seed` will always result in the same samples being selected.
- `has-labels`
  - *Required:* No
  - *Format:* String ("true" or "false"). Default: `false`.
  - *Description:* Specifies whether each sample (row) in the data file has an associated label (text description). See also `label-filecol`.
- `label-filecol`
  - *Required:* No
  - *Format:* Integer
  - *Description:* If `has-labels` is `true`, this column number in the data file is stored as the label (text description) for each sample (row). Currently, OE does not use these text descriptions.
- `live`
  - *Required:* No
  - *Format:* String ("true" or "false"). Default: `false`.
  - *Description:* Is the model being run live, or are we using pre-recorded data. Currently, OE only works with pre-recorded data and ignores this attribute.

#### Elements:

- `<label>`
  - *Required:* No
  - *Format:* String
  - *Description:* Text description of model (currently unused by OE).

#### **<appearance>**

#### Description:

- Allows the user to have some control, and provide hints, over the visual display of the model
- The `<appearance>` element is optional

#### Attributes:

- `node-bin-scale-factor`
  - *Required:* No
  - *Format:* Float. Default: 5.0.

- *Description:* Determines the default scale of histogram bins in nodes
- node-default-height
  - *Required:* No
  - *Format:* Integer. Default: 200.
  - *Description:* Default height (in pixels) of a node.
- node-default-width
  - *Required:* No
  - *Format:* Integer. Default: 330.
  - *Description:* Default width (in pixels) of a node.
- min-internode-gap
  - *Required:* No
  - *Format:* Integer. Default: 20.
  - *Description:* Default spacing (in pixels) between nodes.
- num-root-cols
  - *Required:* No.
  - *Format:* Integer. Default: 1.
  - *Description:* Number of columns of root (input) nodes in display. The root nodes in the file are divided equally between the specified number of columns.
- num-inter-cols
  - *Required:* No.
  - *Format:* Integer. Default: 1.
  - *Description:* Number of columns of intermediate nodes in display. The intermediate nodes in the file are divided equally between the specified number of columns.
- num-leaf-cols
  - *Required:* No.
  - *Format:* Integer. Default: 1.
  - *Description:* Number of columns of leaf (output) nodes in display. The leaf nodes in the file are divided equally between the specified number of columns.
- show-causal-links
  - *Required:* No
  - *Format:* String (“true” or “false”). Default: false.
  - *Description:* Specifies whether causal links should be displayed. These are defined by the <parent> elements in the XML file (see below).

Elements:

- None

**<nodes>**

Description:

- Defines the list of nodes (variables) in the model
- The <nodes> element is required

#### Attributes:

- None

#### Elements:

- <node> [1 or more] (see below)

#### <node>

#### Description:

- Defines the features of an individual node (variable) in the model
- The <nodes> element must contain at least one <node> element

#### Attributes:

- id
  - *Required:* No
  - *Format:* Integer. Default: 0.
  - *Description:* Specifies an ID for this node. If specified, it should be unique. These are used by the <parent> elements (see below).
- datatype
  - *Required:* Yes
  - *Format:* String (“discrete” or “continuous”).
  - *Description:* Specifies a discrete or continuous variable.
- min
  - *Required:* Yes
  - *Format:* Integer (for discrete datatype) or Float (for continuous datatype)
  - *Description:* Specifies the minimum value of this variable (this should be no more than the minimum value present in the data file).
- max
  - *Required:* Yes
  - *Format:* Integer (for discrete datatype) or Float (for continuous datatype)
  - *Description:* Specifies the maximum value of this variable (this should be no less than the maximum value present in the data file).
- role
  - *Required:* Yes
  - *Format:* String (“root”, “inter” or “leaf”)
  - *Description:* Specifies whether this node represents a root (input) variable, an intermediate variable, or a leaf (output) variable.
- filecol
  - *Required:* Yes
  - *Format:* Integer
  - *Description:* Specifies which column in the data file represents this variable’s values. This is 1-based (i.e. the first column is column 1).

#### Elements:

- <label>

- *Required:* Yes
- *Format:* String
- *Description:* Text description of the node (variable), to be used in the label bar of the node on screen (see below).
- `<parents>` [0 or 1] (see below)

## `<parents>`

### Description:

- Defines the list of parent nodes of a particular node, used for defining causal networks (where “A is a parent of B” can be read as “A exerts a causal influence on B”)
- The `<parents>` element is optional

### Attributes:

- None

### Elements:

- `<parent>` [1 or more] (see below)

## `<parent>`

### Description:

- Defines a specific parent of the current node
- The `<parents>` element must contain at least one `<parent>` element

### Attributes:

- `id`
  - *Required:* Yes
  - *Format:* Integer
  - *Description:* The node.id of the parent node.

### Elements:

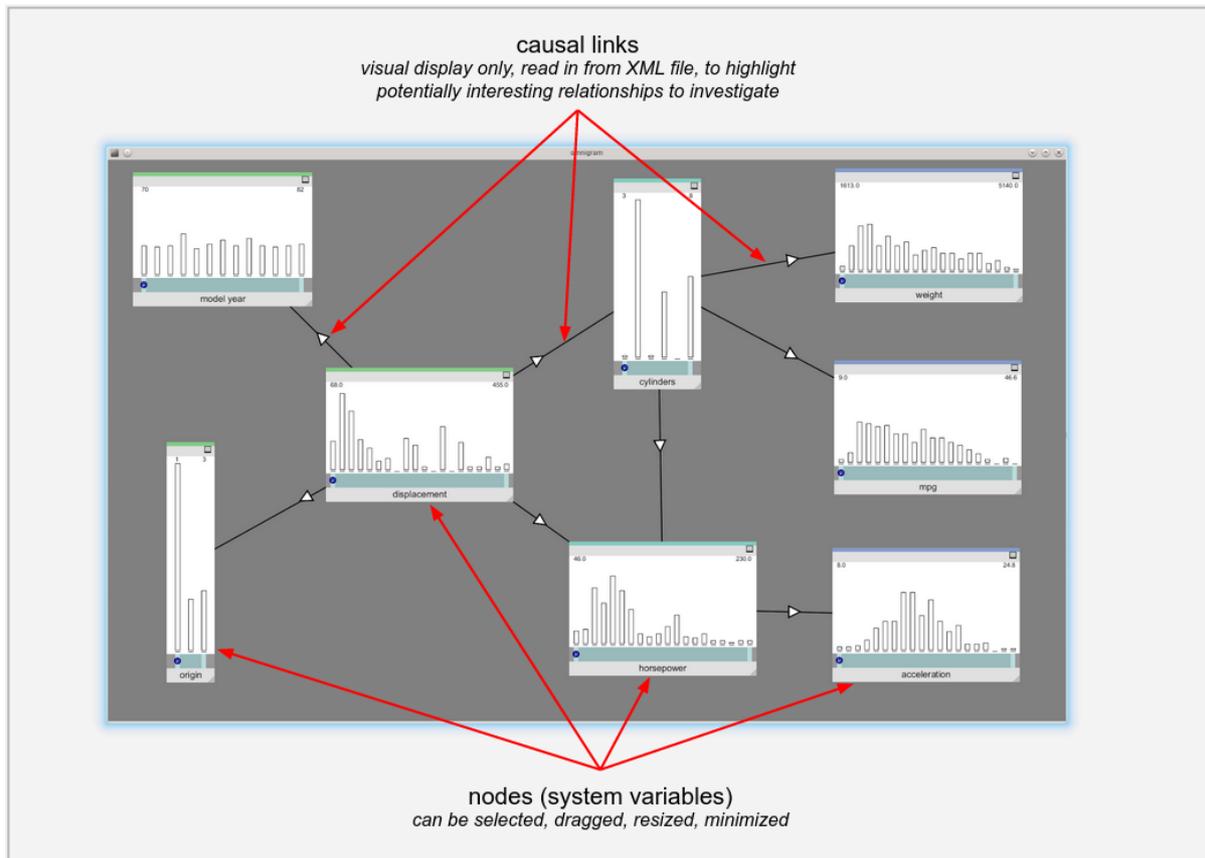
- None

## The Model Data File (CSV format)

The model data file (as specified by the `data` attribute of the `<general>` element in the XML file) contains the pre-recorded data from the model. Each row of data represents one sample, and each column represents a particular variable (as defined by the `filecol` attribute of the `<node>` elements in the XML file). The data file must be in plain comma-separated values (CSV) format, with no header lines or comments. Scientific notation (e.g. 5.2e5) cannot be used in the current implementation.

# The User Interface

The general user OE user interface is shown below.

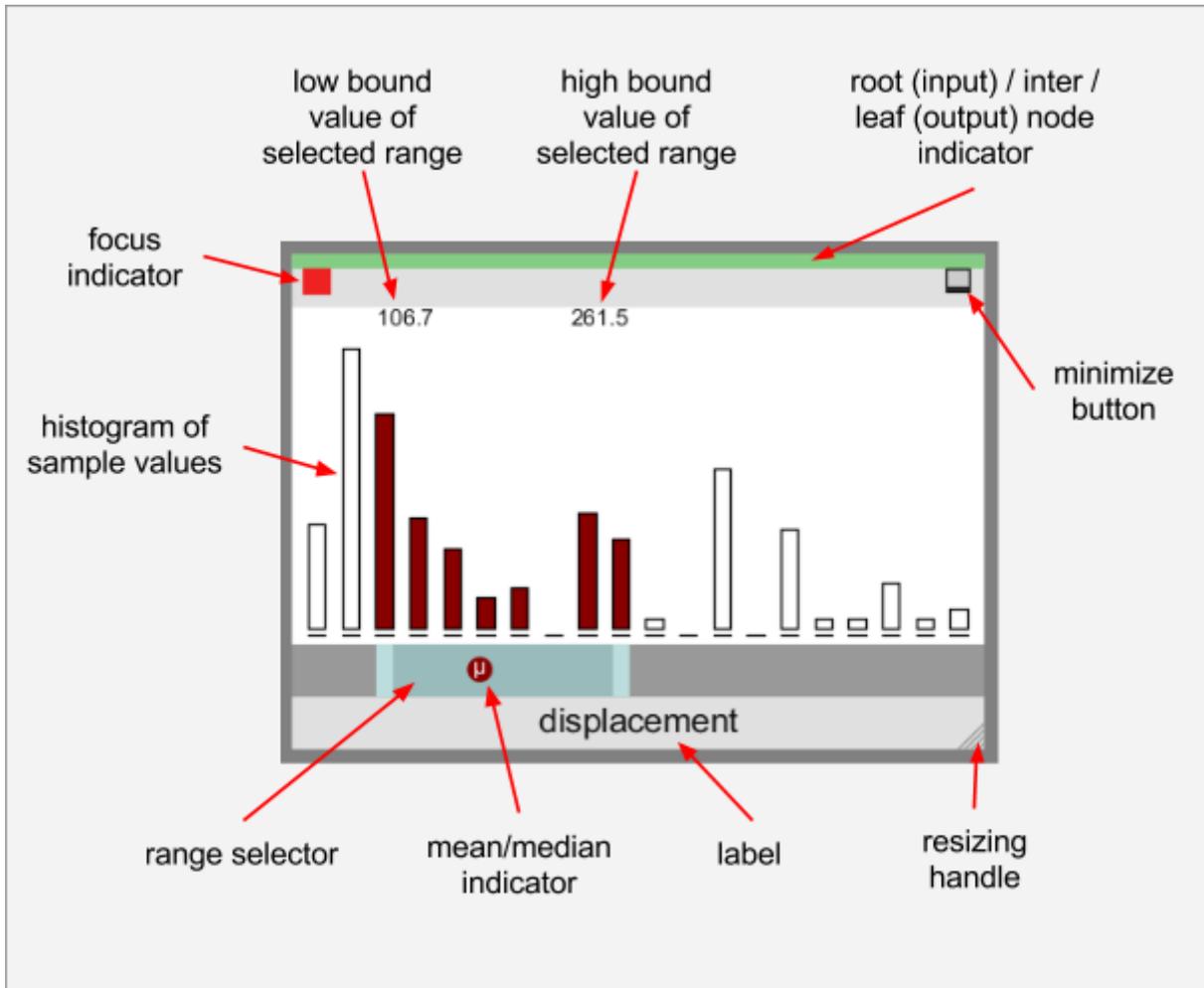


Overview of the Omnigram Explorer user interface

It is possible to magnify or shrink the overall display by pressing the '+' and '-' keys (zoom in and zoom out, respectively). This can be useful with larger models to allow all nodes to be displayed on screen.

## Nodes

The most important component of the user interface is the *Node*, which is an interactive, graphical representation of the data associated with a particular variable in the system.



*User interface details for an individual node*

### What the Node Displays

A node displays a histogram showing the distribution of values of a particular variable observed in the data read in from the data file. The range of values between the variable's minimum and maximum values (as defined by the `<node>`'s `min` and `max` attributes in the XML file) will be split into a maximum of 20 bins. For discrete variables with fewer than 20 values between their `min` and `max`, only the required number of bins are used.

A thin horizontal line is drawn at the base of each bin to indicate its location even in cases where the bin is empty (see the figure above for examples of this).

A node's histogram always shows the distribution of *all* values read in from the data file. However, a subset of these values may be highlighted in different colours, either (for focus nodes) to indicate that they lie within a range of values selected by the range selector, or (for non-focus nodes) to represent brushing in response to focus nodes (see later for further details).

A small circle is drawn on the range selector to indicate the position of the mean (or median) value of samples within the selected range (for focus nodes) or the mean/median value of brushed samples (for non-focus nodes). This display of means (represented by symbol  $\mu$ ) and medians (represented by  $x'$ ) can be toggled by pressing the 'M' key on the keyboard.

The numeric values of the lower bound and upper bound of the bins selected by the range selector are indicated just above the histogram.

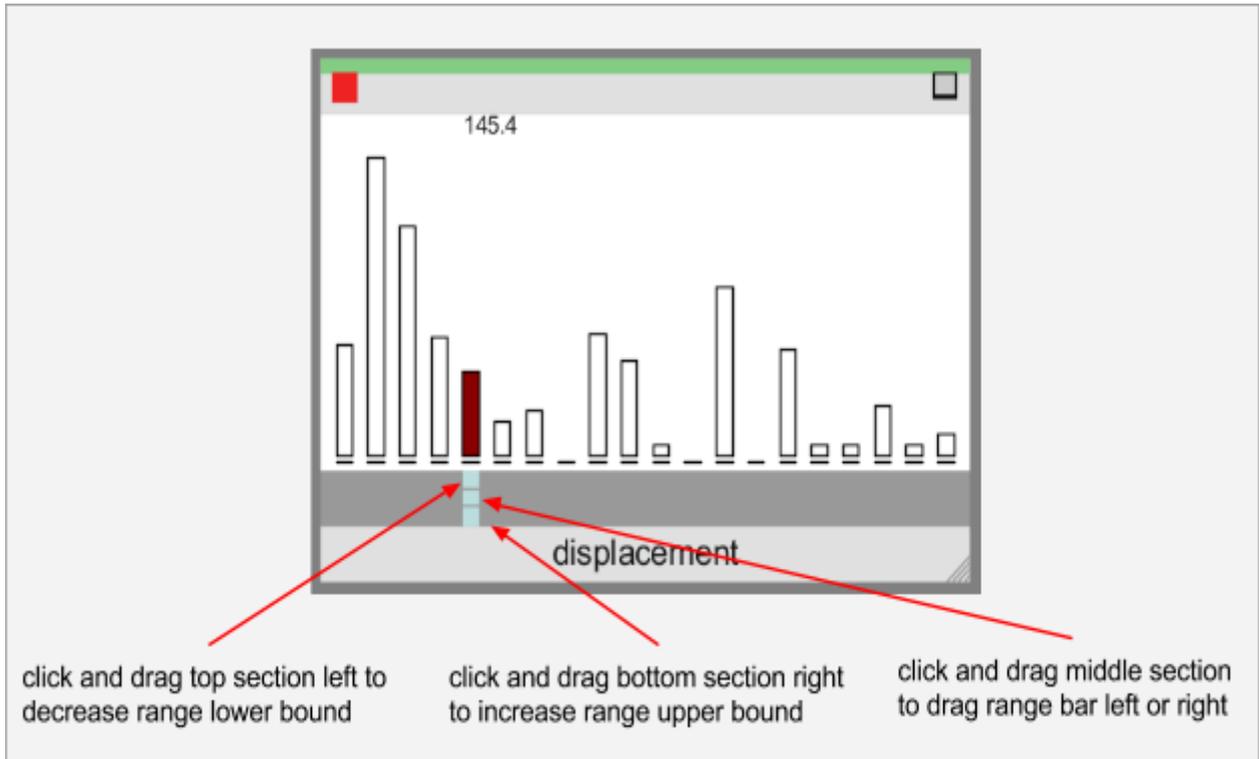
At the top of each node has a thin horizontal strip that indicates whether the node represents a root (input) variable (**green** strip), an intermediate variable (**torquoise** strip) or a leaf (output) variable (**blue** strip). The classification of each node as root, intermediate or leaf is specified in the XML file by the `role` attribute of each `<node>` element. The role influences the initial positioning of the nodes on the screen (root nodes are placed to the left, intermediate nodes in the middle, and leaf nodes to the right of the screen). Beyond that, the only significance of the role is to dictate the colour of strip shown at the top of each node, which acts as a visual reminder to the user about what type of node they are dealing with. In terms of functionality and interactions within OE, no distinction is made between node roles, and all nodes are treated similarly.

### Mouse interaction with Nodes

A variety of mouse interactions are possible with each node:

- A node may be dragged to a different position on the screen by left-clicking on either the label bar (light gray) at the bottom of the node, or the menu bar (light gray) at the top of the node, and dragging the mouse while keeping the left mouse button depressed.
- A node may be resized by left-clicking on the resizing handle (bottom right of node) and dragging the mouse while keeping the left mouse button depressed.
- A node may be minimised by left-clicking the minimise button (top right of node). It will then appear in minimised form at the bottom of the screen. The minimised node can be returned to its original size and position by clicking the maximise button on the minimised node.
- Depending on the current Mode of Interaction (see below), a node may be brought into focus by clicking on the histogram area (white). The red focus indicator (top left) appears when a node is in focus. Depending on the Mode of Interaction, several nodes, or just one, may have focus at any one time.
- The range selector lower limit may be changed by left-clicking on the left range handle (light blue/gray handle at left end of range selector) and dragging left or right.
- The range selector upper limit may be changed by left-clicking on the right range handle (light blue/gray handle at right end of range selector) and dragging left or right.
- The whole range selector bar may be dragged by left-clicking on the main body of the bar (the dark blue/gray region between the two handle) and dragging left or right.

- There is a special case for interacting with the range selector when the lower and upper limits are equal. This is illustrated in the diagram below.



*Special case range bar interaction when lower bound equals upper bound*

## Modes of Interaction

There are three basic modes of interaction in OE, which may be accessed by pressing the '1', '2' and '3' keys on the keyboard.

### Single Node Brushing

Accessed by pressing '1' on the keyboard.

In this mode, only one node can have the focus at any one time. Clicking on the histogram area of a node gives it the focus (shown by the red focus indicator in the top left of the node), and removes the focus from any other node.

The range selector of the focus node can be adjusted to select a subset of samples from the overall distribution (i.e. a subset of bins from the histogram). The bins within the selected range are shown in **dark red**, and the other bins in the focus node are shown in white.

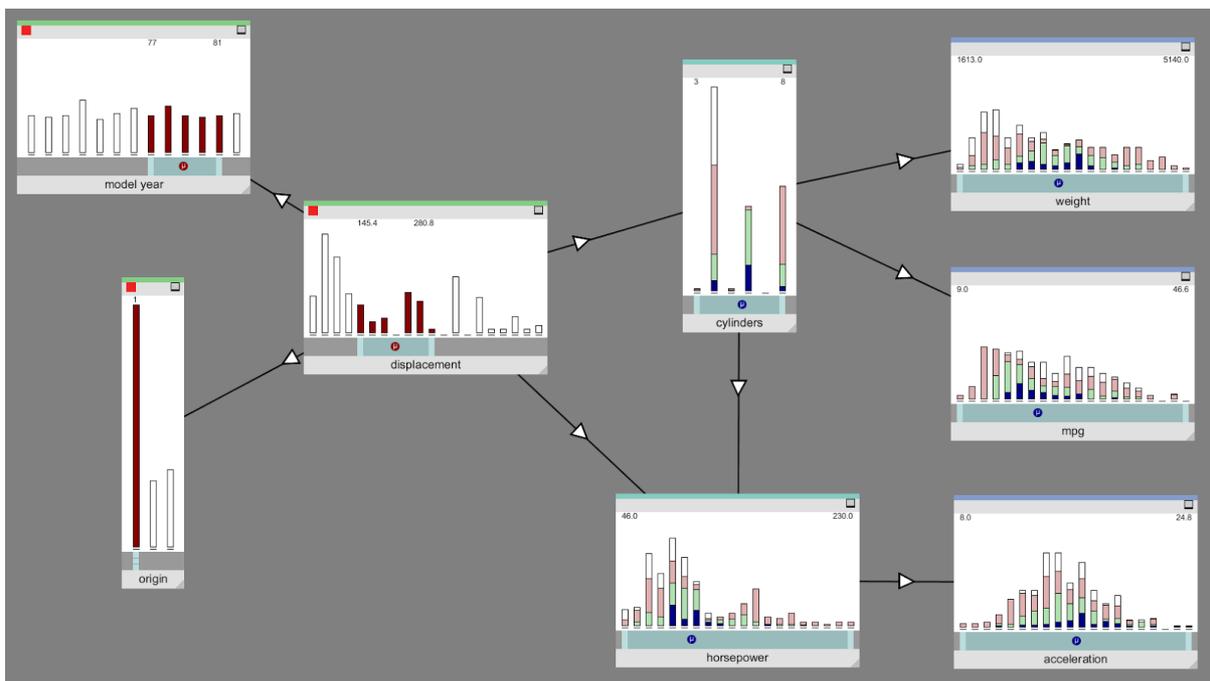
When a range of samples has been selected in the focus node in this way, all of the other nodes are updated to show where the same samples lie in the distributions of the other variables in the model. The matching samples are shown in **dark blue**. (This technique is known in the visualisation literature as *linking* and *brushing*).

The real power of linking and brushing in OE becomes apparent when you interactively change the range selection in the focus node, and watch the resulting changes in the other nodes. In particular, **selecting a fairly small subset of values with the range selector in the focus node (i.e. having a short range selector bar), dragging the bar from left to right and back, and watching how the change in values of the focus node is associated with changes in the other nodes**, is a very effective technique. The human visual system is very attuned to noticing multiple objects moving in the same direction (the Gestalt psychology laws of Continuity and Common Fate). OE makes use of this fact to allow the user to easily spot correlations between many different variables in Single Node Brushing mode.

## Multi Node Brushing

Accessed by pressing '2' on the keyboard.

Multi Node Brushing is an extension to Single Node Brushing that provides information about sensitivity of the model. An example screenshot showing Multi Node Brushing mode is shown below.



*An example of Multi Node Brushing mode*

In Multi Node Brushing mode, more than one node can have the focus at the same time. In the example shown above, the three leftmost nodes all have the focus (indicated by the red squares at the top left of each node). The range selectors in each of the focus nodes can be adjusted just like in Single Node Brushing mode, to select a subset of samples.

The information displayed in non-focus nodes in Multi Node Brushing mode is a bit more complex than in Single Node Brushing. As before, **dark blue** is used to indicate the location of samples that have been selected by the range selection in the focus nodes. In Multi Node Brushing, OE looks at the conjunction of the range selections made on the focus nodes: to take an example from the screenshot above, a blue patch on a histogram bin in the *horsepower* node indicates that there is a sample that has that horsepower value that also lies within the selected range of *model year* and lies within the selected range of *displacement* and lies within the selected range of *origin*.

You will notice from the example screenshot that non-focus nodes also show other colours. A **light green** patch indicates that there is a sample that has that value that also lies within the selected range of values for all but one of the focus nodes. So, in the example above, a blue patch on a histogram bin in the *horsepower* node might indicate that there is a sample that has that horsepower value that also lies within the selected range of *model year* and lies within the selected range of *displacement*, but *not* within the selected range of *origin*. It might also indicate that the *origin* and *model year* ranges were respected, but not the *displacement*, etc. That is, the light green colour indicates that one of the focus node ranges was not respected, but it does not tell you which one.

Similarly, a **light red** patch indicates that two of the focus node ranges were not respected. And white indicates that three or more were not respected.

The colour therefore gives some indication of the sensitivity of the model to the value ranges chosen. For example, the presence of a **light green** patch in a non-focus node indicates that by increasing the selected range of just one of the focus nodes, the patch can be turned blue, i.e. it can be made to satisfy the constraints on all of the focus nodes by loosening a single constraint.

### Brush Links

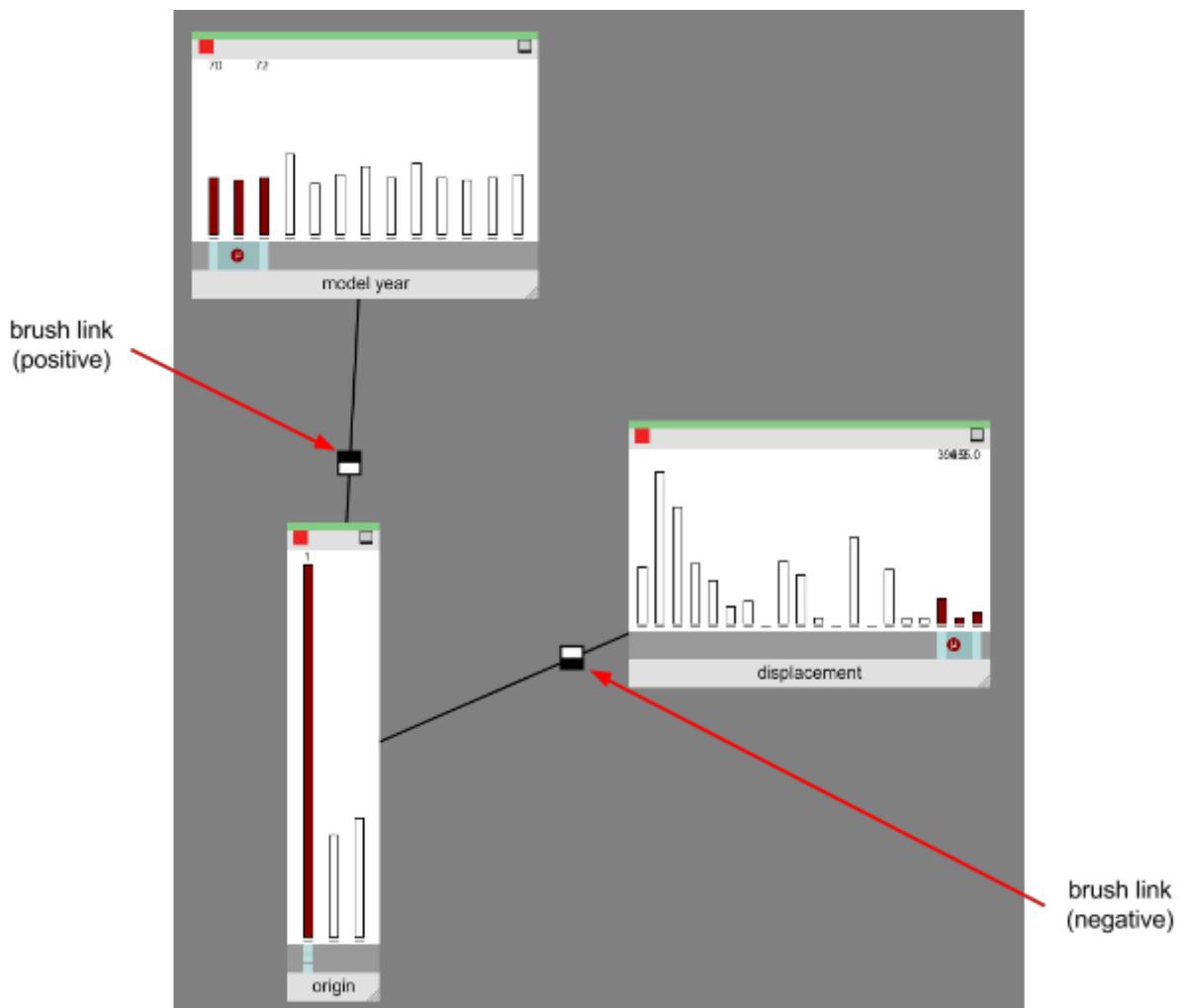
In Single Node Brushing mode, a simple, yet powerful, interaction was to drag the range selector of the focus node left and right to identify correlated changes in other nodes. The same interaction is possible in Multi Node Brushing mode for any of the focus nodes; in that case, you move the range selector bar for one of the focus nodes while leaving the range selection of the other focus nodes untouched.

In some cases, it can be interesting to see the effect of moving the range selector bars of two or more of the focus nodes at the same time. This poses a problem, as you can only click and drag one bar at a time. OE provides *brush links* to help solve this problem. A brush link is a user-specified connection between two focus nodes. When the range selector bar of one focus node (we'll call it the *primary* node) is changed, the selector bars of any other focus nodes that are directly connected by a brush link to the primary node are also changed, as directed by the brush link. The link can be positive, in which case the connected nodes are changed in the same direction as the change to the primary node, or negative, in which case the change is in the opposite direction (e.g. if the range selector

bar of the primary node is dragged to the right, the range selector bar of a negatively linked focus node moves to the left).

Brush links can be created (and deleted) interactively while OE is running. A new brush link is created as follows:

1. Ensure both nodes to be linked already have the focus (indicated by the red square at the top left of the node) before continuing. Only focus nodes can be linked. To give a node focus, just left-click on the histogram area of the node.
2. Hover the mouse pointer over one of the nodes and press the 'L' key on the keyboard. A red outline will appear around the node.
3. With the red outline on one node visible as described in the previous step, hover the mouse pointer over the second node and press the 'L' key again. A new brush link will appear between the two nodes.



*Brush links*

When first created, the brush link will be fully positive (+1.0), meaning that movement of the range selector bar in one node will result in movement in the same direction in the linked node. This is indicated by the decoration of the brush link's handle (the small square drawn in the middle of the link); a fully black upper half of the handle and fully white lower

half indicates a link of strength +1.0 (as shown on the left link in the example above). The strength can be altered by left-clicking in the link handle: click in the lower half of the handle to decrease the strength, and click in the upper half to increase the strength. The strength can be decreased right down to a minimum of -1.0, indicating a fully negative connection (represented by a fully black lower half of the handle and fully white upper half, as shown on the right link in the example above). With a negative link, when the range selector of one node is dragged to the right, the range selector of the linked node moves to the left, and vice versa.

A brush link can be deleted by hovering the mouse pointer over the link's handle and pressing the 'B' key.

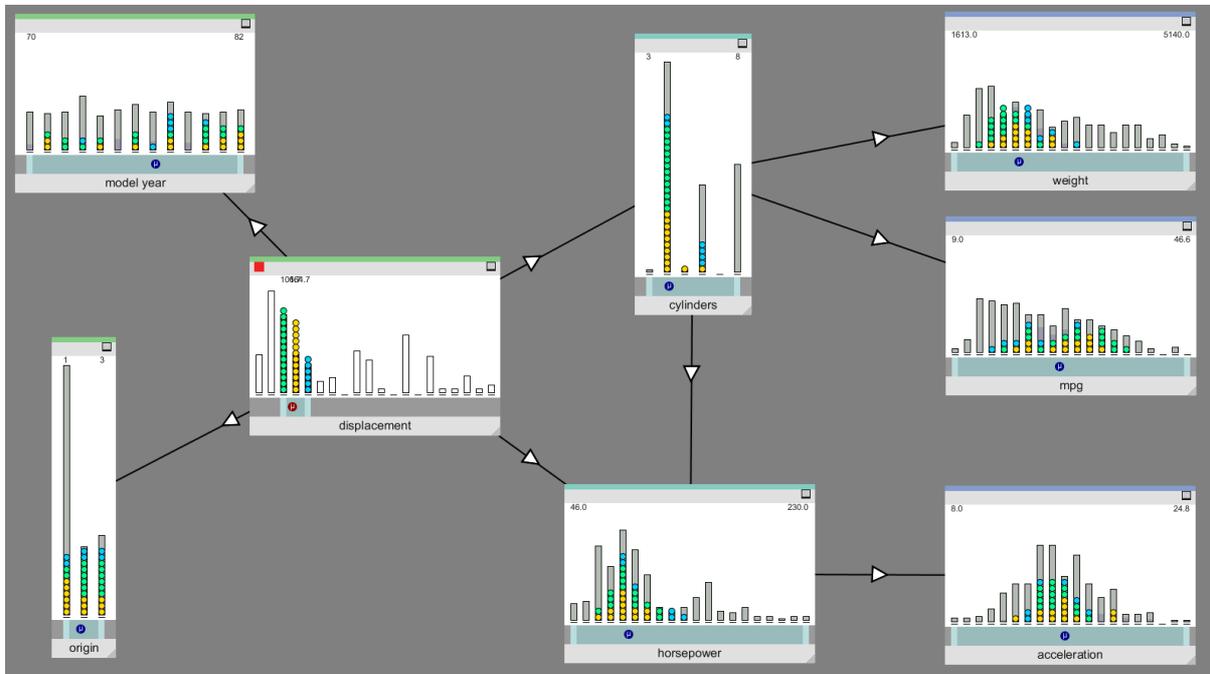
### Show Individual Samples

Accessed by pressing '3' on the keyboard.

*To begin the display of samples, one node should be chosen as the focus node (by clicking on the histogram area of the node, as usual).*

Single Node Brushing and Multi Node Brushing modes are useful for understanding the overall distribution of values observed for each node (variable) as the range of allowed values on the focus node(s) are changed.

Show Individual Samples mode provides a different way to visualise the data. The overall distribution of values for each variable is shown, as with the other two modes, but these are shown partially faded out in the background of each node, and are provided just as a reminder of the overall shape of the observed data. The real point of interest in Show Individual Samples mode is the display of individual samples from the data file. Each sample is shown as a small coloured circle placed in the position of the histogram bin corresponding to the sample value of that variable. An example screen from Show Individual Samples mode is shown below.



*An example of Show Individual Samples mode*

At any one time, a fixed number of samples is shown. The number of samples can be increased by pressing the 'X' key, and decreased by pressing the 'Z' key.

By default, OE will automatically cycle at a moderate speed through different selections of samples from the data. At each step in the cycle, one sample (the one that has been displayed for the longest time) is removed from the currently displayed set, and a new sample replaces it. The speed at which this cycling occurs can be increased by pressing the 'ARROW UP' key, and decreased by pressing the 'ARROW DOWN' key.

The user may also choose to step through the different sample selections manually rather than having them updated automatically. This is achieved by pressing the 'ARROW RIGHT' key (and pressing it again to step to the next sample selection). The previous selection can be returned to by pressing the 'ARROW LEFT' key. The user can revert to automatic cycling by pressing the 'ARROW UP' or 'ARROW DOWN' keys.

### Sample Modes

By default, in Show Individual Samples mode the samples are coloured according to which bin they belong to in the focus node (for an example, see the screenshot above, where the focus node is the *displacement* node, indicated by the red focus indicator in the top left of the node). The user can quickly see how these samples are distributed in the other nodes by looking at the distribution of colours.

An alternative colouring scheme is available, where samples are coloured at random, irrespective of which bin they appear in in the focus node. The user can toggle between these two colouring schemes by pressing the 'S' key. The random colouring scheme is

most useful when looking at a small number of samples, and it allows the user to see exactly where individual samples appear in each node according to their colour.

## Causal Links

In most of the screenshots shown above, there are arrows connecting some of the nodes. These are *causal links*. Causal links are defined in the Model Definition XML file by the optional `<parent>` elements of the `<node>`s.

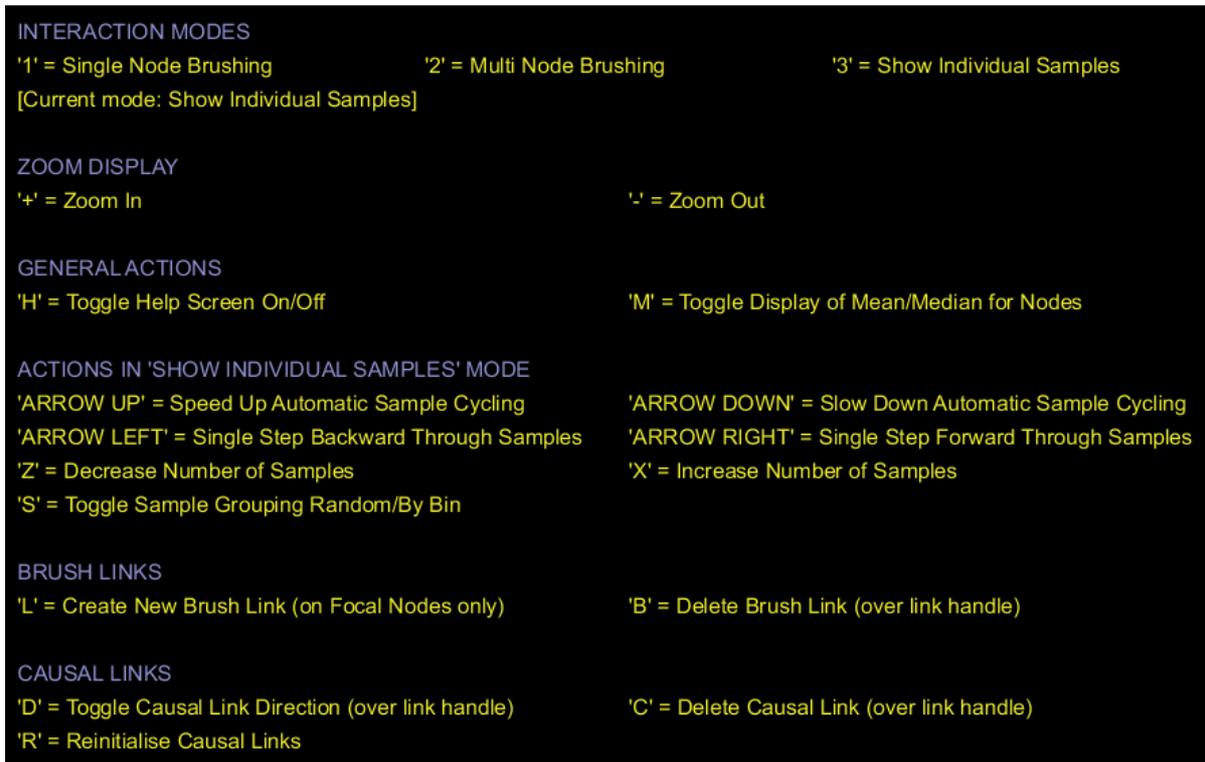
As mentioned earlier, these causal links might have been identified by knowledge of the system or from tools such as Bayesian causal network discovery software.

OE displays these links to provide hints to the user of interesting relationships to explore. Beyond this, they have no other role in OE.

A limited amount of interaction is possible with these links. The direction of a link may be reversed by hovering the mouse pointer over the link's arrowhead and pressing the 'D' key. A link may be removed by hovering the mouse pointer over the arrowhead and pressing the 'C' key. All links can be reset to the state defined in the XML file by pressing the 'R' key.

## The help screen

A help screen is available that lists all of the key presses available. This can be accessed by pressing the 'H' key.



### **Further suggestions for usage**

The previous sections have outlined the basic modes of interaction available in OE. A few additional suggestions for the effective use of the tool are listed below. These may be especially useful when working with large models with many nodes.

- If nodes have been identified that behave in a similar way (e.g. when dragging the range selector of the focus node), those nodes can be grouped together by dragging them to one part of the screen. The user can then start to partition a complex model into subsets of similarly performing variables by spatially grouping the nodes on screen.
- To further simplify a complex model if groups of similarly performing nodes have been identified, the user might choose to concentrate on one representative node from a group, and minimize all other nodes in that group.