SAND2020-6478R

Dial-A-Cluster User Manual

Patricia Crossno, Jaxon Gittinger, Warren Hunt, Matt Letter, Shawn Martin, Milosz Sielicki

Dial-A-Cluster Model

The Dial-A-Cluster (DAC) model allows interactive visualization of multivariate time series data. A multivariate time series dataset consists of an ensemble of data points, where each data point consists of a set of time series curves. The example of a DAC dataset used in this guide is a collection of 100 cities in the United States, where each city collects a year's worth of weather data, including daily temperature, humidity, and wind speed measurements.

In the DAC model, the data points are displayed using a two-dimensional scatter plot. Each data point, e.g. city, is represented as a point in the scatter plot. Interpoint distances encode similarity so that, for example, two cities are near each other if they have similar weather throughout the year. Further, DAC provides sliders that allow the user to change the relative influence of the temporal variables on the scatter plot. These changes are computed in real time so that users can see how different time series variables affect the relative similarity of different data points in the ensemble – hence the name Dial-A-Cluster.

DAC computations are performed using a weighted sum of time series distance matrices to produce a visualization via the classical multidimensional scaling (MDS) algorithm [1]. In addition to ensemble visualization, DAC provides comparisons between user selected groups in the ensemble as well as the influence of any available metadata.

DAC was developed independently outside of the Slycat™ project using Slycat's plugin architecture. Consequently, many of the user interface and representational conventions used in other Slycat™ models are missing or different in DAC. Although DAC operates on ensemble data (where an ensemble is a set of related samples defined using a set of shared variables), there is no concept of input or output variables as in other Slycat™ models, only a table of scalar metadata and multiple time-varying data sets for each ensemble member.

The DAC model consists of four linked views, as shown in **Figure 1**: (1) a *Slider* panel (left view) for adjusting the importance of each temporal variable in the similarity calculation; (2) a *Scatterplot* (center top view) showing similarities between ensemble members; (3) *Time Series Plots* (right view) for comparing three sets of selected ensemble members, with sets shown in red, blue, and green; (4) and a *Metadata Table* (bottom left view) displaying values from shared scalar variables (columns) for individual ensemble members (rows).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

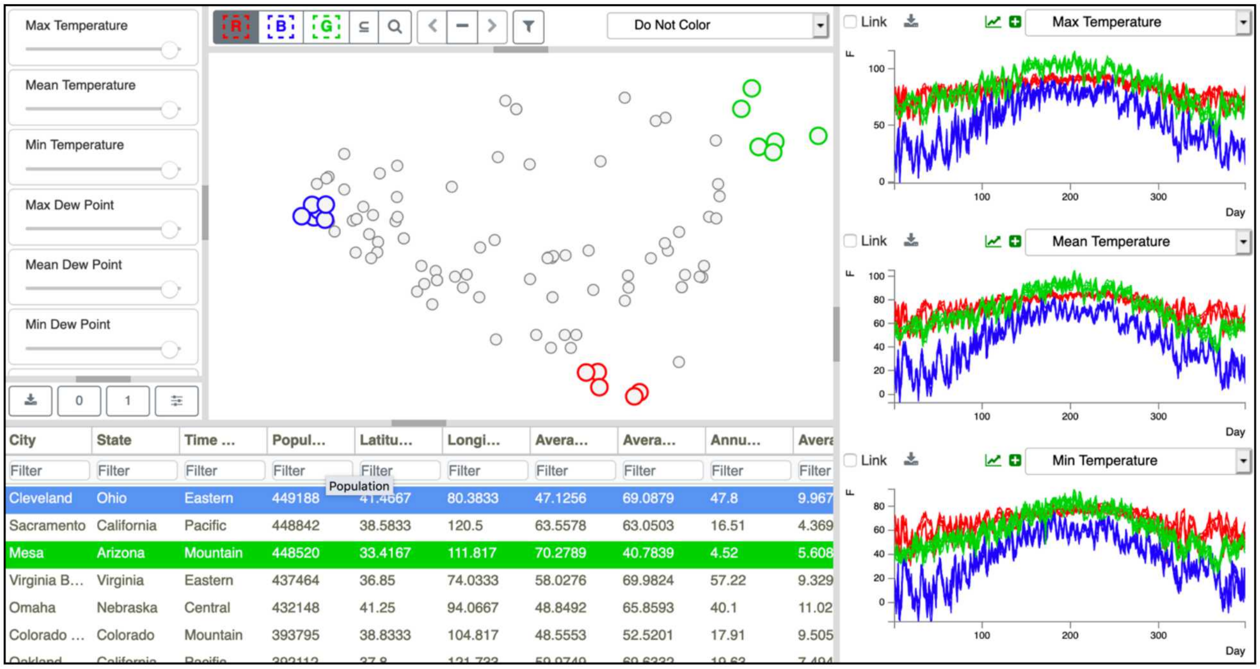


Figure 1: DAC user interface consisting of four linked views. In clockwise order, starting from the upper left corner, there is (1) a panel of sliders for adjusting temporal variable weights, (2) a scatterplot with a point per ensemble member, in which point proximity indicates member similarity, (3) time series plots for three temporal variables contrasting selected groups (red/blue/green sets) of ensemble members, and (4) a table of scalar and text metadata for each ensemble member (row).

Weather Example Data Set

In this manual, we use a weather data set to illustrate model creation and DAC in general. The weather dataset is not an ensemble of simulation outputs. Instead it is a set of measured values of various temporally-changing weather-related quantities, such as temperature, dew point, humidity, sea level pressure, visibility, wind speed, precipitation, cloud cover, and wind direction collected during 2014 for 100 cities in the United States from Weather Underground (www.wunderground.com). The number of samples in each of the time series are all identical, each consisting of a value per day for a little more than a year. In addition to the temporal variables, each city has associated meta data, including city name, state, time zone, population, latitude, longitude, average temperature, average humidity, annual precipitation, average wind speed, average visibility, and average cloud cover.

Data Formats

There are currently three data formats accepted by DAC, the Dial-A-Cluster Generic Format, the PTS CVS/META Zip Format, and a National Instruments (www.ni.com) LabView based TDMS format.

Dial-A-Cluster Generic Format

This is a multi-file format consisting of a zipped set of directories and comma separated value (CSV) files, where each directory provides a different type of information. The CSV files are grouped using a naming convention that relies on file extensions to differentiate the data types: .dac, .meta, .var, .time, and .dist. Other than the .dac file, which can have any name so long as it has a .dac extension, the names of the directories and all of the other files are fixed. At the top level, there is the .dac file and three directories named *dist*, *time*, and *var*. The *dist* directory contains files named sequentially as variable_1.dist, variable_2.dist, …, variable_*n*.dist, where *n* is the number of time series variables for each datapoint in the ensemble. The *time* directory contains files variable_1.time, variable_2.time, …, variable_*n*.time; and the *var* directory contains files variable_1.var, variable_2.var, …, variable_*n*.var. In addition, the *var* directory must contain the variables.meta file. The directory structure is shown in **Figure 2**.

Name	Size	Kind
datapoints.dac	10 KB	Document
▼ dist	--	Folder
variable_1.dist	130 KB	Document
variable_2.dist	130 KB	Document
variable_3.dist	130 KB	Document
▼ time	--	Folder
variable_1.time	5 KB	Document
variable_2.time	5 KB	Document
variable_3.time	5 KB	Document
▼ var	--	Folder
variable_1.var	517 KB	Document
variable_2.var	517 KB	Document
variable_3.var	518 KB	Document
variables.meta	636 bytes	Document

Figure 2: The directory structure for the DAC generic input format, using a dataset with only three data points.

.dac

Scalar data for the *Metadata Table* is contained in a single file with the .dac extension. The first row contains header information (variable names) for the table columns. The remaining rows are the values for these variables for each ensemble member. Each row is comma separated and ends with a new line (a CSV file). This file is at the same level as the directories *dist*, *time*, and *var*.

.dist

In the *dist* folder, there are variable_1.dist, variable_2.dist, …, variable_*n*.dist files. Each .dist file contains the all-to-all distance matrix comparing every ensemble member to each of the others, calculated using the correspond time series variable. These matrices are used to compute the visualization in the *Scatterplot* pane. The file variable_1.dist is the distance matrix calculated using the first variable defined in variables.meta. The .dist files have no header row. Given *k* ensemble members, the file will consist of *k* rows, each with *k* comma-delimited distance values. The rows and columns in these matrices are ordered according to order of the data points in the .dac file.

.time

In the *time* folder, there are variable_1.time, variable_2.time, …, variable_*n*.time files. These files contain the times when the corresponding temporal variables were sampled. For example the file variable_1.time contains the *x*-axis values for graphing the time sequence associated with the first variable in the variables.meta file. These files have no header row, consisting of just a single row of comma delimited values. The temporal units for these values are defined in the .meta file in the “Time Units” column.

.var

In the *var* folder, there are variable_1.var, variable_2.var, …, variable_*n*.var files. Here, the file variable_1.var contains the values (amplitudes in the *Time Series Plot*) for the first variable in the variables.meta. The .var files have no header row, containing a single row with comma delimited values for each data point, ordered according to the order given in the .dac fille. The units for these values are defined in the .meta file in the "Units" column.

.meta

The *var* folder also contains a file variables.meta, which defines the time series variable names shown in the *Slider* panel. The first row should consist of the headers: "Name,Time Units,Units,Plot Type". The remaining rows define each of the time series variables (one per row), providing variable name, units for labeling the x and y axes in the time series plots, and the type of plot. "Curve" is the only option currently available for "Plot Type".

Archiving in Windows

The files and subdirectories described above must be archived in a .zip file before import into the DAC model creation wizard. When creating the .zip file on Windows, do not zip from the level of an enclosing directory, but instead perform the zip with the three directories and the .dac file as the input elements. Also note that all time sequences for a given temporal variable MUST contain an identical number of samples. (Different time variables can have different number of samples, but the number of samples cannot vary within a specific variable.) The correct method for creating the archive is shown in **Figure 3**.

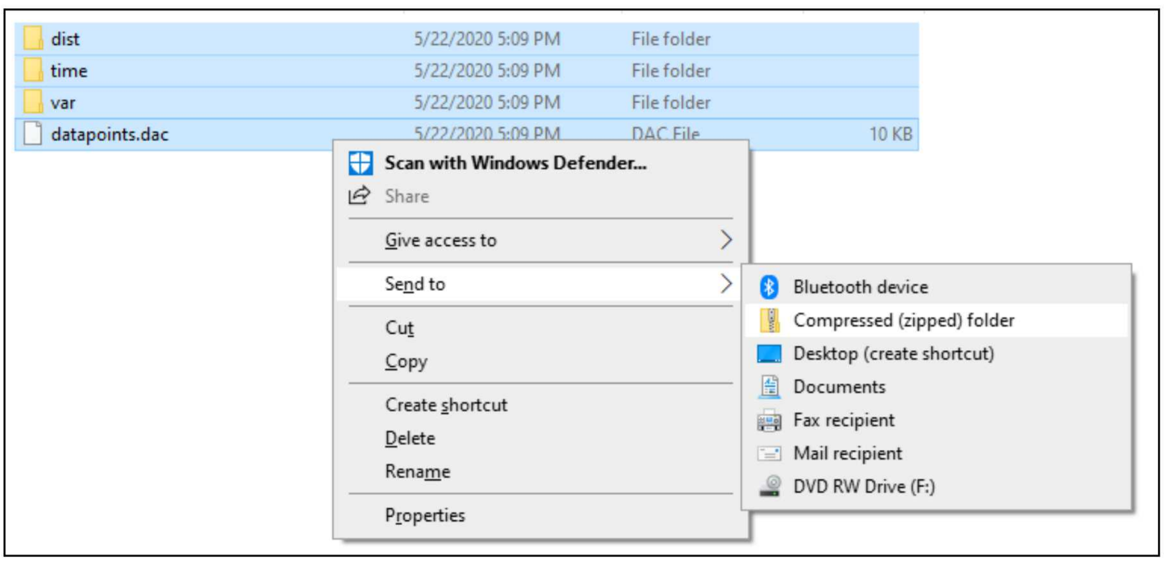


Figure 3: Archiving a DAC Generic Format directory into a .zip file on Windows.

Archiving on Mac

On a Mac, it is easiest to archive the files from the command line. This avoids the inclusion of the Mac OS files .DS_store and __MACOSX, which will trigger errors when importing into the DAC model creation wizard. From the directory containing the .dac file and the *dist*, *time*, and *var* folders, use the command:

```
$ zip -r weather-dac-gen.zip . -x ".*" -x "__MACOSX"
```

PTS CVS/META Zip Format

This data format is designed to facilitate ingestion from a Sandia-specific project. Constructing this format is complex and contains redundant copies of the same information. Consequently, it is not recommended for general use. However, we document it here for completeness. The format consists of a zip file containing two subdirectories named CSV and META.

Within the CSV subdirectory are a set of .csv files, where each .csv contains a single time series. Within the META subdirectory are corresponding sets of .ini files, where each .ini file contains the metadata for a single temporal variable for a single ensemble member (one of the .csv files in the other directory). Given that there are n temporal variables and k ensemble members, both directories will contain $n \times k$ files. There is no naming convention for these files, beyond the fact that each .csv file must have a corresponding .ini file. However, it is convenient to use a naming convention such as <member descriptor> <member index> <underscore> <variable descriptor> <variable index>. The index for the ensemble member should vary between 1 and n. The index for the temporal variable should vary between 1 and k.

So, for example, if we had data sampled at 10 cities (our ensemble members) and 20 temporal variables per city, we could name the timeseries for the 10th temporal variable for the second city as city2_temp10.csv. We would then need to name the corresponding meta variable file as city2_temp10.ini. Or instead, we could have named them more abstractly, t2_d10.csv and t2_d10.ini, respectively.

.csv

Each time series file has four columns, named “SampleNum”, “Raw”, “X”, and “Y”, respectively. SampleNum is a one-based index of the time samples in the file. Raw contains the raw data values for the quantity being sampled. X is the temporal value for the sample that will be used as the *x*-coordinate in the Time Series Plot. Y is the sampled value that will be plotted (either the original raw value or a scaled/modified value).

.ini

These files contain the metadata for each temporal variable for each sample. They also contain the scalar metadata for each sample, which is repeated for each temporal variable (since there is a file per temporal variable per sample). The metadata is described using key-value pairs in an ASCII text file. An example from our weather data set for the city of New York for the first temporal variable, Max Temperature, would look like:

```
; Fake PTS Weather data
; S. Martin 6/23/2017

[test]
City="New York"
State="New York"
Time Zone="Eastern"
Population="8213839"
Latitude="40.7833"
Longitude="72.0333"
Average Temperature (F)="52.3166"
```

Average Humidity (%)=”62.4648”
Annual Precipitation (In)="57.02"
Average Windspeed (MPH)="11.5126"
Average Visibility (Miles)="9.1206"
Average Cloud Cover (Okta)="5.27889"

[operation]
test_op_inst_id=1

[waveform]
WF_DIG_ID="1"
WF_DIG_LABEL="Max Temperature"
WF_X_UNITS="Day"
WF_Y_UNITS="F"

Comments are lines beginning with a ‘;’. The bracketed text, *[test]*, *[operation]*, and *[waveform]* are required. The key-value pairs in the *[test]* section are used in the *Metadata Table*. They can be any set of variables and values, so long as they are consistent across the samples (inconsistent values will generate warnings and/or errors during model creation – see **Model Parse Log**). The table values are in quotes. Under [operation], the *test_op_inst_id* variable provides the ensemble member number. The *test_op_inst_id* is also typically encoded in the file name, but this is optional. Under *[waveform]*, the *WF_DIG_ID* variable provides the temporal variable index. Unlike the variable names in the *[test]* section, the variable names in the *[waveform]* section cannot be changed. Only the values for those waveform variables can be modified to describe the time series quantities.

The two most important fields in the .ini file are *test_op_inst_id*, which identifies the member of the ensemble, and *WF_DIG_ID*, which identifies the variable for that member. These two fields uniquely identify the time series, regardless of the file name provided.

These CSV and META directories must then be archived using the .zip format as described in **Archiving in Windows** and

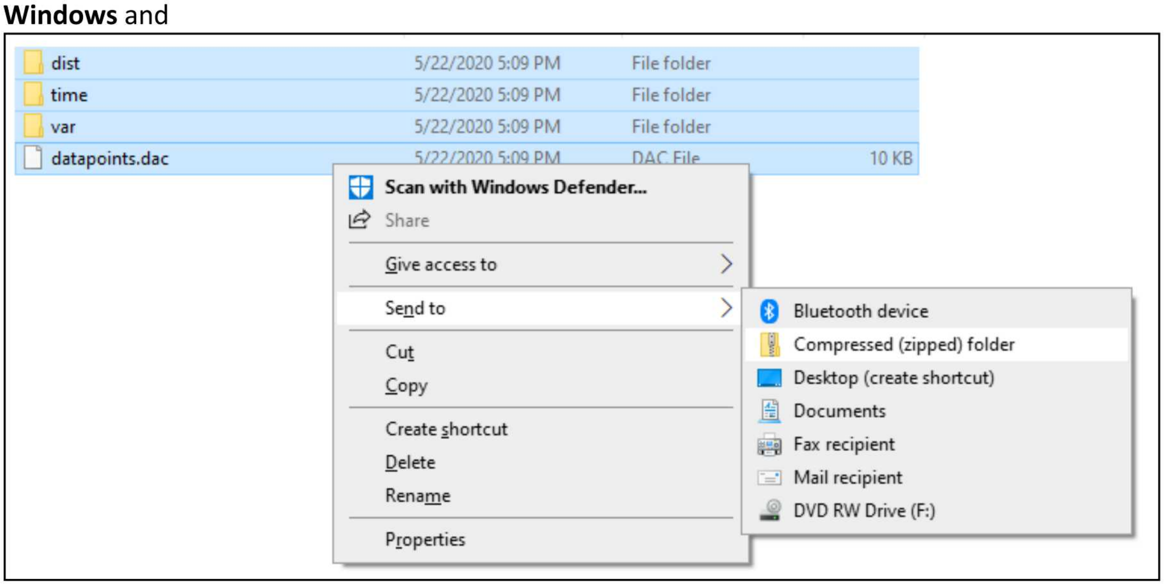


Figure 3: Archiving a DAC Generic Format directory into a .zip file on Windows.

Archiving on Mac.

TDMS Format

The TDMS format uses files created with National Instruments (www.ni.com) LabView software. The TDMS format is quite complicated and can contain a variety of information (see https://www.ni.com/en-us/support/documentation/supplemental/06/the-ni-tdms-file-format.html) . The TDMS files that can be imported into DAC are specific to a particular Sandia project. These files can be imported as individual files or as archived .zip file containing directories with TDMS files. In the case of the TDMS .zip format, DAC will identify any TDMS files (by the .tdms extension) and attempt to combine the data in those files.

Model Creation

To create a DAC model within Slycat™, click the green *Create* button in the Navbar while you are on a project page. A dropdown menu will appear as shown in **Figure 4**. Note that DAC models are only generated from data sets stored locally on your desktop machine, or on a shared disk accessible from your desktop. Unlike other Slycat™ models, there is no remote data option for ingestion of data sets stored on a cluster.

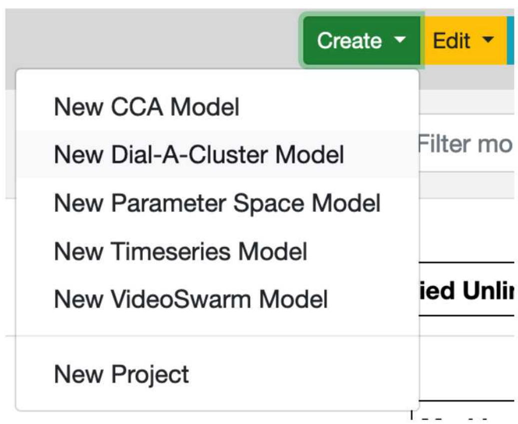


Figure 4: Model creation dropdown menu. Click on New Dial-A-Cluster Model to start the DAC model creation wizard.

Select *New Dial-A-Cluster Model*, which will start the DAC model creation wizard. The wizard takes you through a series of choices regarding the input data format, location, and model name, each of which appears in a separate tab of the wizard's popup window.

DAC Data Formats

The initial tab for selecting the input format is the *Data Format* tab shown in **Figure 5**. Select either the "Dial-A-Cluster Generic Zip Format", the "PTS CSV/META Zip Format", or the "Switchtube TDMS Format." (See **Data Formats** for an explanation of the format choices.) Click the *Continue* button on the lower right to advance to the next tab.

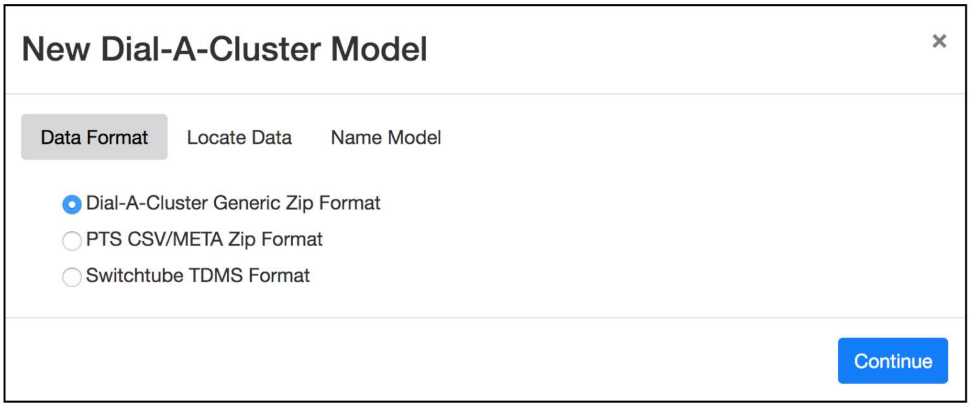


Figure 5: Data Format choices.

DAC Generic Zip Format

The *Locate Data* tab for the Dial-A-Cluster Generic Format provides a local file browser for uploading the zip file from your desktop, or shared drive (see **Dial-A-Cluster Generic Format** for details on the formatting of the generic zip file contents). First, click *Browse* in the *Locate Data* tab shown in **Figure 6**, then using the popup file browser, navigate to the location of the zip file and select it. Press *Continue* to advance to the *Name Model* tab.

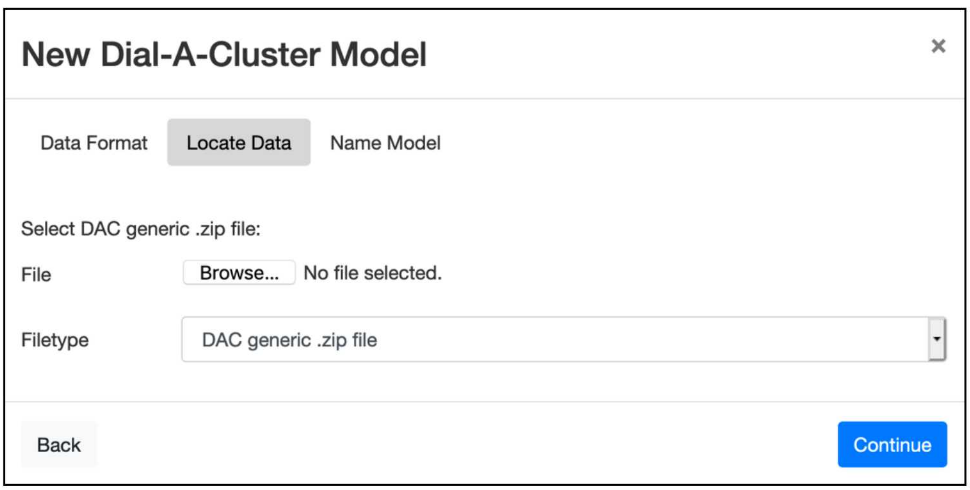


Figure 6: Locate Data tab to select the zip file used in the DAC Generic Format.

PTS CSV/META Zip Format

The PTS CSV/META Zip Format option operates in a manner similar to the DAC Generic Zip format, though the internal format of the files within the zip file are different (see **PTS CVS/META Zip Format** for details on the formatting of the PTS CSV/META zip file contents). Click *Browse* in the *Locate Data* tab shown in **Figure 7**, then using the popup file browser, navigate to the location of the zip file and select it.

The PTS version of the *Locate Data* tab also allows the user to specify two parameters for filtering the imported data. The first parameter is “Minimum number of time points in a CSV file”. Any .csv files which contain less than the specified number of time points will be ignored. The second parameter is “Minimum number of digitizers in a test”. A digitizer is synonymous with a variable in this context. This parameter will ignore any ensemble member having less than the specified number of variables. The parameters have defaults of 10 and 3, respectively.

In the *Locate Data* tab, notice the ❓ icon. If you hover over the ❓ icon, help is provided in the form of a description of the use of the mechanism used to select the TDMS suffixes using the "Include" column. In general, whenever you see the ❓ icon, hovering over that icon will provide a description of the functionality of a provided feature in the DAC user interface.

The *Continue* button will advance to the *Name Model* tab.

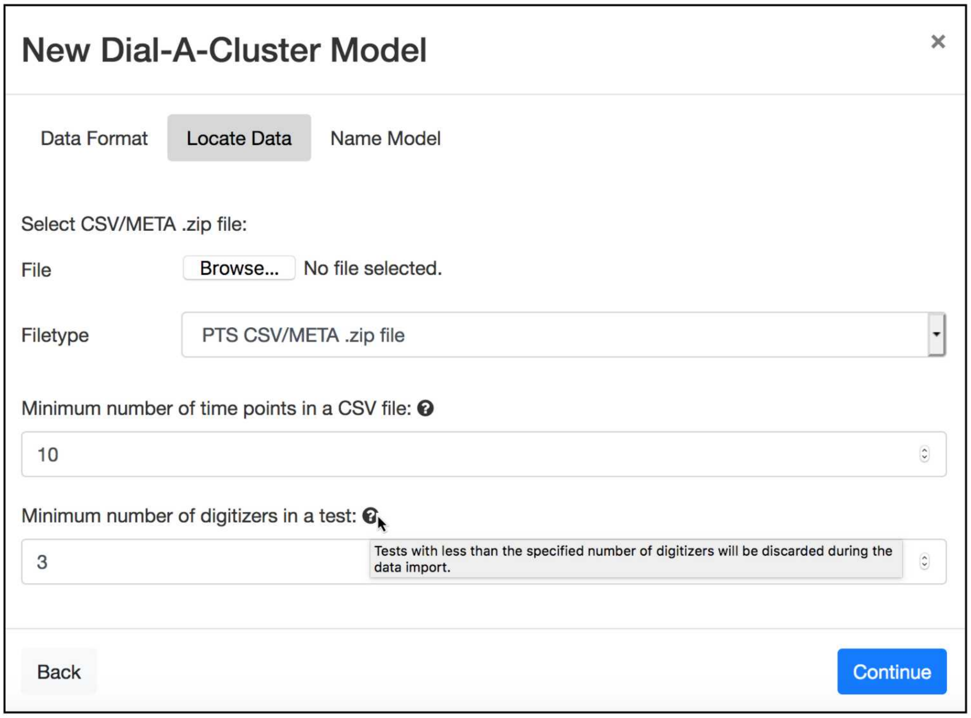


Figure 7: Locate Data tab to select the single zip file used in the PTS CSV/META Zip Format.

Switchtube TDMS Format

For the TDMS format, there are two file types available from the drop down, as shown in **Figure 8.** To select one or more .tdms files directly use the ".tmds file(s)" option. To select an archive containing TDMS files select the "TMDS .zip file" option.

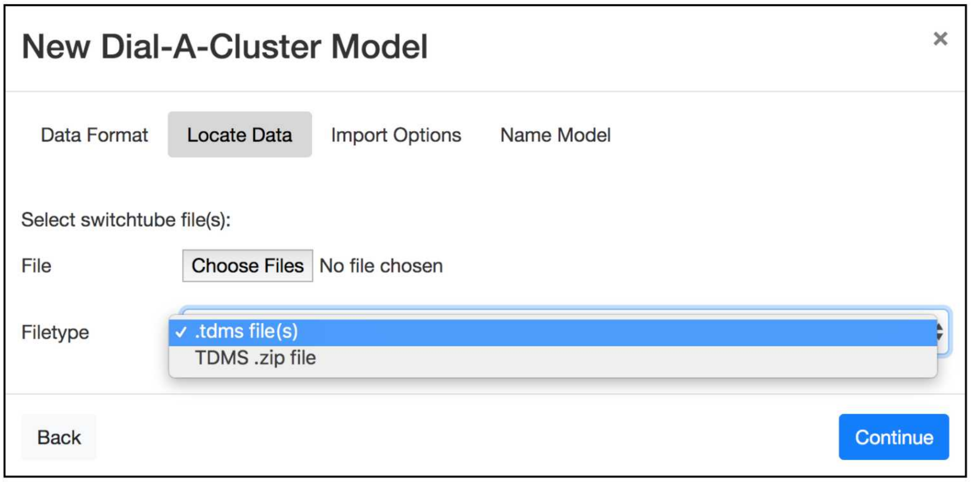


Figure 8: The TDMS file type pulldown allows the user to select either .tdms files directly, or a .zip file containing TDMS files.

In addition to the *Locate Data* tab, the TDMS format uses two additional tabs. In the case of a TDMS .zip file selection, a *Select Suffixes* tab is displayed, as shown in **Figure 9**. In the *Select Suffixes* tab, you have the option of filtering the TDMS files by a file suffix, which is used to tag a TDMS file of the form *_suffix.tdms. The user then has the option to include only TDMS files with specific suffixes. After selecting the TDMS format, the *Continue* button advances to the *Import Options* tab.

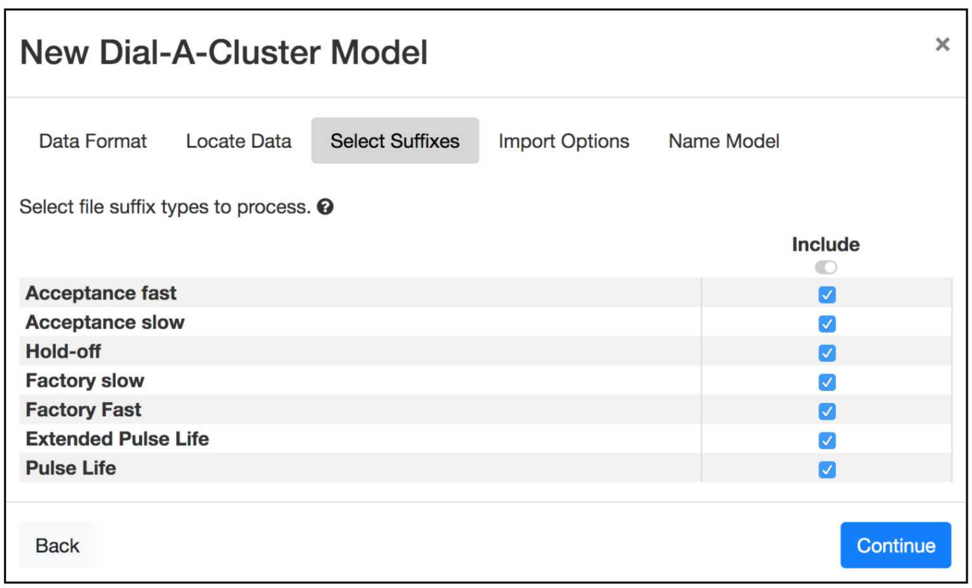


Figure 9: The Select Suffixes tab of the TDMS import wizard displayed when importing a TDMS .zip file.

For both TDMS formats, an *Import Options* tab is next displayed, as shown in **Figure 10**. The *Import Options* tab provides parameters that DAC can use to import incomplete or expected types of data. These options are similar to the options in the PTS *Locate Data* tab. The “Minimum number of time points in a channel” tells DAC to ignore any time series that have less than the specified number of time points. The “Mminimum number of channels in a test” tells DAC to ignore any tests (ensemble members) that do not have at least a certain number of channels (variables) available. These values have defaults of 10 and 2, respectively.

The type of time series expected can be specified as “General”, “Overvoltage”, or “Sprytron”. The default is “General”. These options provide information about the number and type of time series expected in the TDMS files. For TDMS files, unlike DAC Generic Files, time points do not have to match across tests. If they don’t match, the time series can still be combined using the “Intersection” or “Union” options. The “Intersection” option will intersect time points to obtain a common set of time steps, while the “Union” option will extend early or late starting time series using the first or last time series values to match longer channels. Finally, channel units and/or time steps units will be inferred from the data type (“General”, “Overvoltage”, or “Sprytron”) and/or default values generally expected from TDMS data.

Continue advances to the *Name Model* tab.

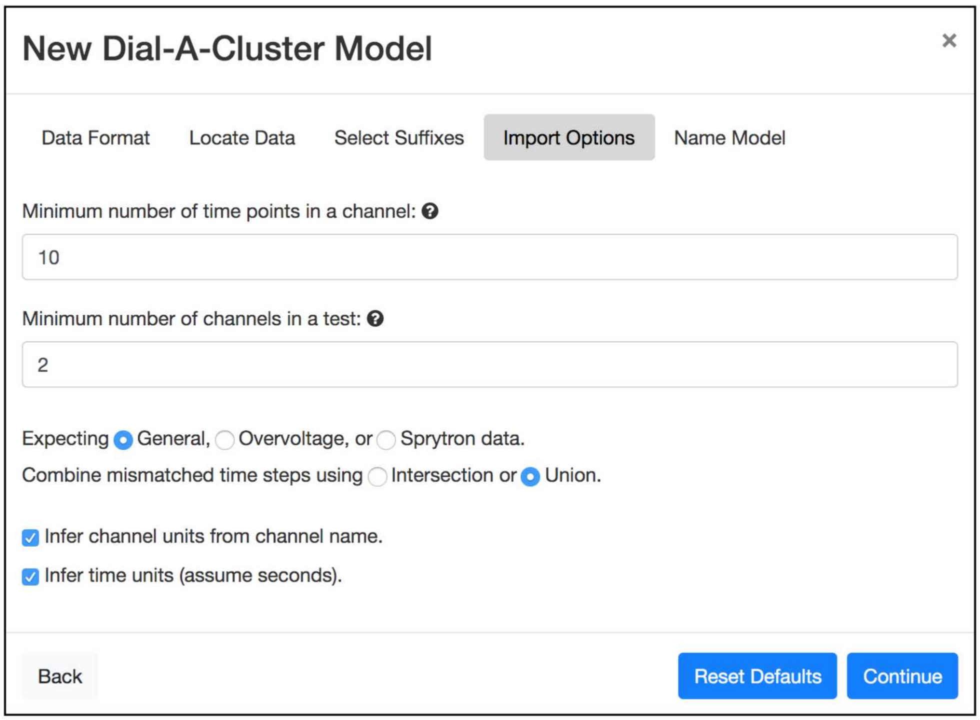


Figure 10: Options for importing DAC TDMS files.

Name Model

Once the data format and input parameters have been obtained, the DAC import wizard allows you to name the new DAC model. As shown in **Figure 11**, the *Name Model* tab has three fields: *Name*, *Description*, and *Marking*. This allows you to specify a unique and meaningful name for the model, an optional short textual description (perhaps aspects of the data used to generate the model, or parameter choices made when creating it), and optional markings. The marking selections provided in the dropdown list are site-specific for the Slycat™ server you are accessing.

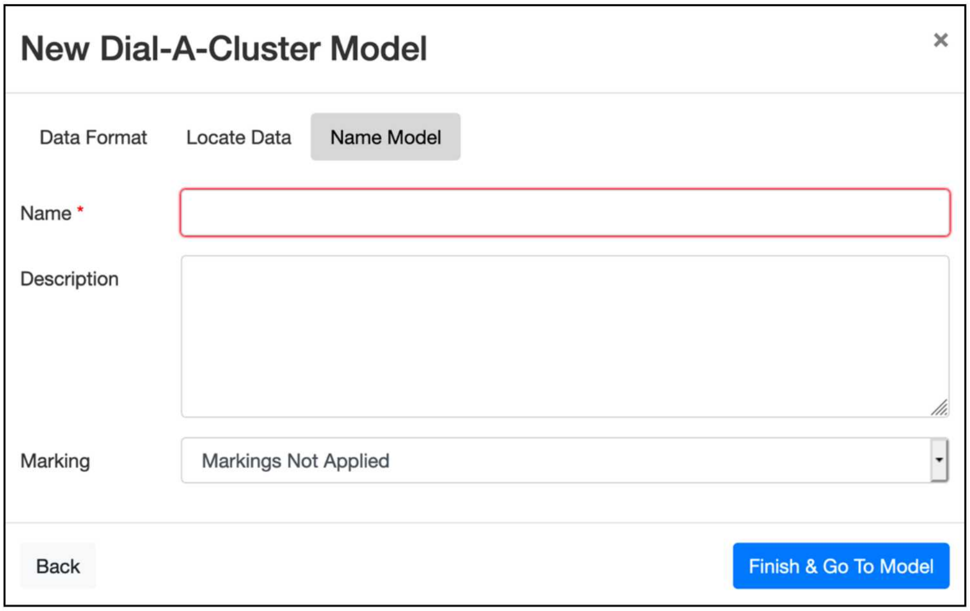


Figure 11: In the Name Model tab there are three fields: Name, Description, and a Marking dropdown list.

Once you have completed entering this information, click the *Finish & Go To Model* button in the lower right to complete the DAC model creation process. The system will now start to upload the input files.

If the model name and/or markings need to be changed after the model has been loaded, that can be performed using the Edit > Model Name & Info pulldown in the upper right of the Slycat™ window.

Loading Data and Progress Indicators

At this point, the wizard will return to the *Locate Data* tab where you selected the input files and provide a progress indicator for each of the uploads. For instance, if you selected a single DAC Generic Zip file, the progress indicator would look like the one shown in **Figure 12**. At this point you must wait for the selected file(s) to upload before closing the browser. If you close the browser, the file upload will be terminated and no data will be uploaded to the model.

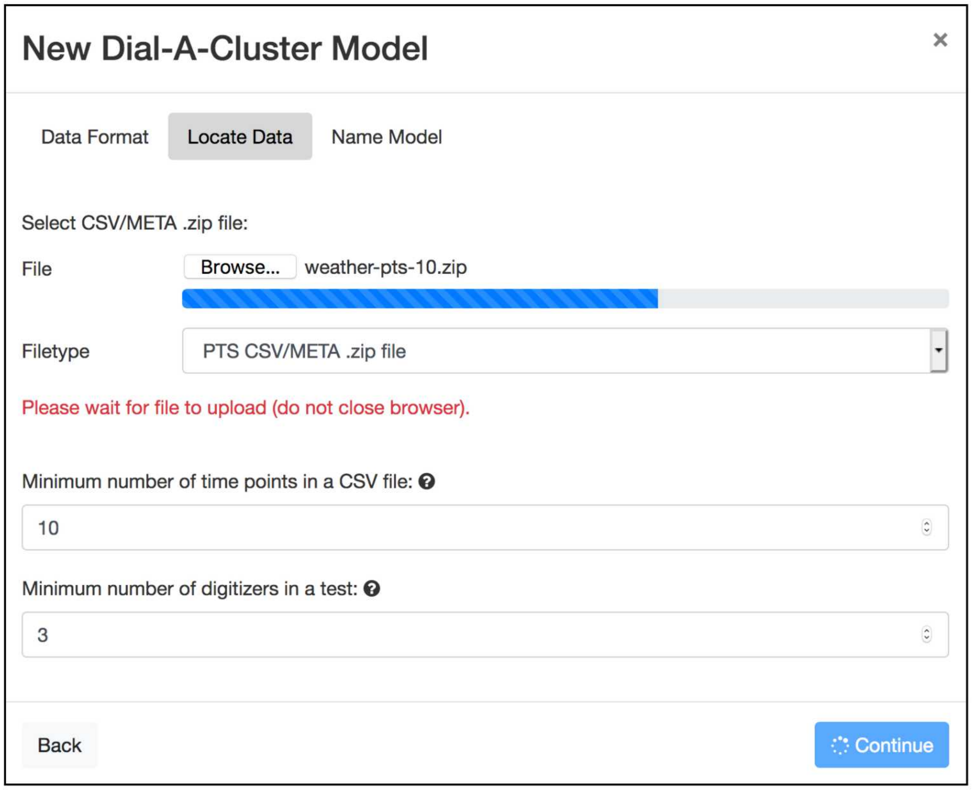


Figure 12: Uploading data in DAC model creation wizard.

Once the data is uploaded, the wizard provides a different progress indicator, as shown in **Figure 13**, to reflect the progress in processing the data set and the creation of the final model. If there are any error messages, they will appear in the space below the progress bar. If the error messages exceed the available space in the box, they will automatically scroll. To facilitate reading the error messages before they scroll out of sight, you may click within the text box to alternatively pause and restart scrolling. It is safe to close the browser during the model upload screen. For very large models, you can leave and return later. When you re-open the browser it will show progress or present the finished model.

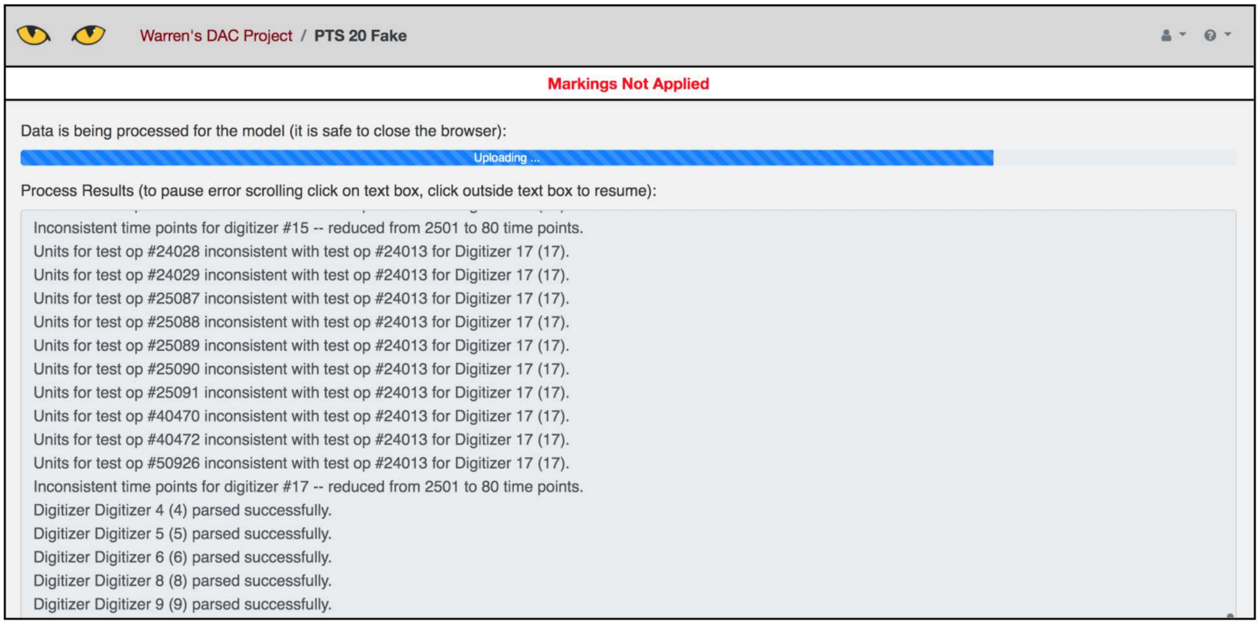


Figure 13: Data processing and model creation progress indicator.

During the processing phase, DAC will parse files and upload data to the model. For the DAC Generic Format, distance matrices are provided and uploaded directly. For PTS and TDMS data, DAC creates distance matrices quantifying time series similarity by doing an all-to-all comparison between the ensemble members. The distances in these matrices are then input to a weighted Multi-Dimensional Scaling (MDS) algorithm that projects each of the ensemble members into a two-dimensional space. Proximity of projected points indicates similarity between those samples.

Each variable's contribution (weighting) in calculating point projections can be interactively adjusted using the sliders on the left side of the DAC interface. This allows you to explore each variable's importance in forming clusters or patterns within the projected space (see the Sliders section). Values for the weights range between zero and one, with a value of zero excluding a variable, and a value of one providing full inclusion.

Once processing is complete, Slycat™ will take you to a rendering of the new DAC model. The initial view of the Weather model is shown in **Figure 14**.

Figure 14: Initial DAC model configuration for the Weather Data.

Model Parse Log

Once a model has been created, information on the success of the model import process can be viewed using the *Info* button menu pulldown, as shown in **Figure 15**.

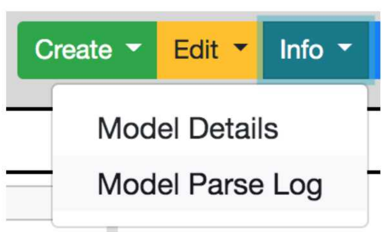


Figure 15: Accessing the Model Parse Log from the Info button pulldown.

The model parse log itself is just the complete record of the import process and any problems that may have been encountered during import. This information is displayed in the progress box as the model loaded (**Figure 13**). The parse log gives you information about which files in your ensemble were not imported, what data might have been modified, and why. The log for the Weather data is shown in **Figure 16**.

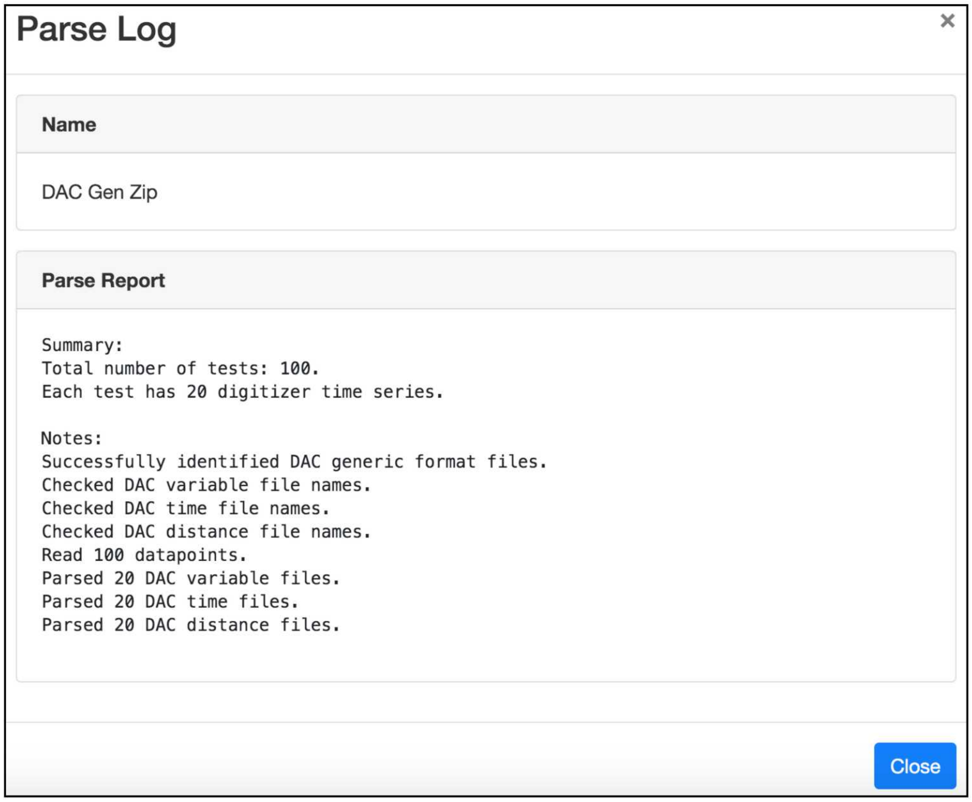


Figure 16: Model Parse Log for DAC Generic Zip Format containing the Weather data.

Model Views

There are multiple linked views in a DAC model, including a *Scatterplot* (top middle), a set of *Sliders* (upper left), three *Time Series Plots* (right side), and the *Metadata Table* (bottom left). The *Scatterplot* is the central view and it represents each ensemble member as a point. Point positions are calculated using multi-dimensional scaling (MDS) based on a weighted sum of temporal variable distances between members. Note that MDS is calculated using just the set of temporal variables shown as sliders on the left (i.e. none of the *Metadata Table* variables are involved). The *Sliders* provide interactive control of the summation weights, so analysts can explore the impact that different variables have in *Scatterplot* point clustering. Points can be selected in the *Scatterplot* or *Metadata Table* views to define up to three sets (where each set of points is color-coded in red, blue, or green). These sets are used to compare time series data from the selected points in the *Time Series Plot* views. Additionally, the interiors of the *Scatterplot* points can be independently color-coded by any of the scalar variables shown in the *Metadata Table*. Initially, points are not color-coded (white interior) and do not belong to any set (black border), as shown in **Figure 14**.

Scatterplot

Distances between points indicate similarity, with closer points being more similar. Initially, all temporal variables are fully weighted, so the MDS point projection gives equal importance to all temporal variables in calculating point proximity.

Note that dimensionality reduction from a high-dimensional space down to two dimensions generally does not completely capture the distance relationships found in the higher-dimensional space due to the limitations of the lower-dimensional embedding.

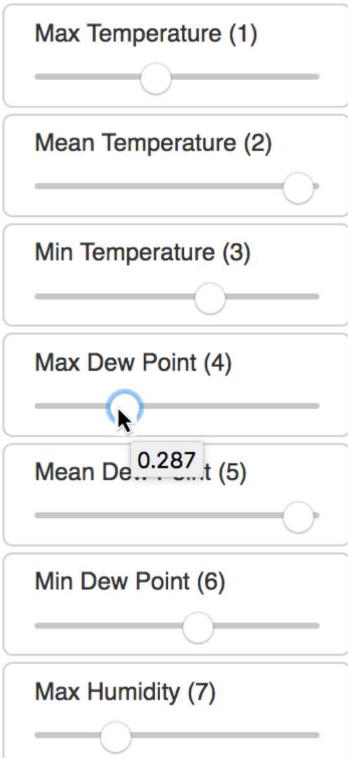


Figure 17: Temporal variables and their associated sliders, with each variable's slider set to a differing weight.

Sliders

All the temporal variables in the data set are listed in the *Slider* view (upper left view). If the full list of temporal variables cannot fit within the view, scrolling will be enabled through a scrollbar on the right side of the list. Underneath each variable is a slider for setting that variable's weight in the MDS calculation. **Figure 17** shows an example of how each slider can be set independently to provide a variety of weights on a per variable basis. Weights are valued between 0 and 1, with sliders positioned to the far left being set to 0, and sliders on the far right being set to 1. Hovering over a slider displays its actual value, also shown in **Figure 17**. Users interactively adjust a variable's weight by dragging the slider, which then triggers changes to the positions of points in the *Scatterplot*.

The controls shown in **Figure 18** appear at the bottom of the *Slider* view. From left to right, the buttons perform the following functions:

- Download all (or part of) the *Metadata Table* as a CSV file to your local machine.
- Set all sliders to 0.
- Set all sliders to 1.
- Set the slider weights to optimize the *Scatterplot* for the selected *Metadata Table* column (selected from the dropdown list in the upper right above the *Scatterplot*, as shown in **Figure 20**).

Figure 18: Slider controls are (1) download table, (2) set all sliders to 0, (3) set all sliders to 1, and (4) optimize sliders for selected Metadata Table column.

Download Table Data

Clicking the icon when no points are selected saves the contents of the entire *Metadata Table* as a CSV file. However, if you have selected points, clicking brings up the dialog shown in **Figure 19**, letting you choose between saving the full *visible* table or just the *visible* selected points.

It is important to note that the *visible* table is the table that you see after application of the table filters (see **Table Filtering** in **Metadata Table**). When you export the full *visible* table, you are export the table after the filters have been applied, but regardless of whether or not ensemble members are selected. If you want to export the entire table, you must first clear the table filters before clicking the icon. Similarly, when you export the *visible* selection, you are exporting only selections visible after the filters have been applied. And again, if you want to export all selections, regardless of the table filters, you must first clear the filters.

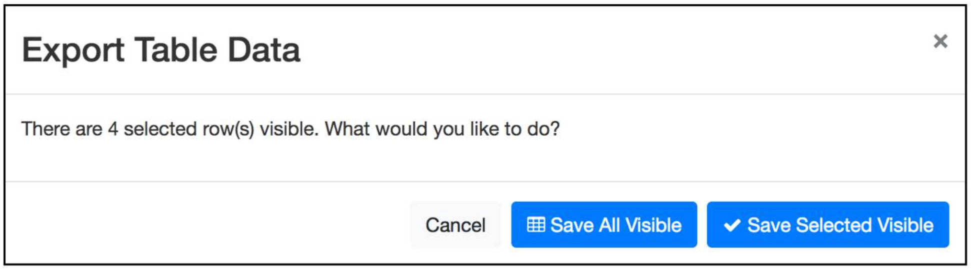


Figure 19: Export Table Data dialog.

Optimize Slider Weights

You may want to explore the relationships between temporal variables and one of the scalar variables in a column of the *Metadata Table*. First, select a *Metadata Table* column header for the comparison from the dropdown list shown in **Figure 20**. This dropdown is above the *Scatterplot* in the upper right. Then click the icon, which will adjust the slider weights to generate a *Scatterplot* projection maximizing differences in the value range of the selected variable. The resulting slider weights indicate the relative importance of each temporal variable in the phenomenon described by the set of values in that *Metadata Table* column. Note that *Metadata Table* columns with string values can be selected here, but that those strings are changed to integers starting from 1, where each integer corresponds to a string value in the column, ordered alphabetically.

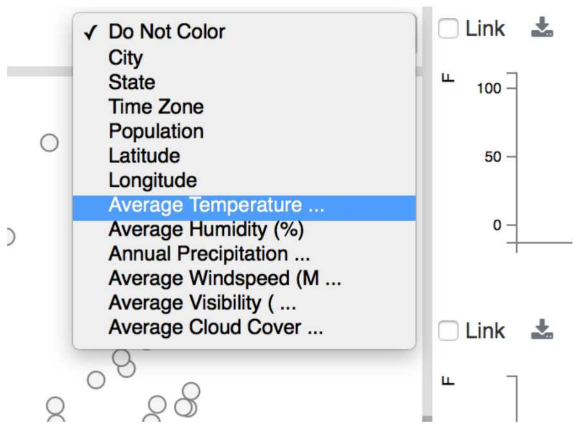


Figure 20: Weather data scalar variables for color-coding Scatterplot points. The default is 'Do Not Color'.

To provide some sense of what this function does, here is an example of the results for optimizing the point projection and slider values against the *Annual Precipitation* variable. **Figure 21** shows the *Scatterplot* where all temporal variables are fully weighted with their slider values set to *1*. The points are colored by their values for *Annual Precipitation*, where high values are dark gray and low values are white.

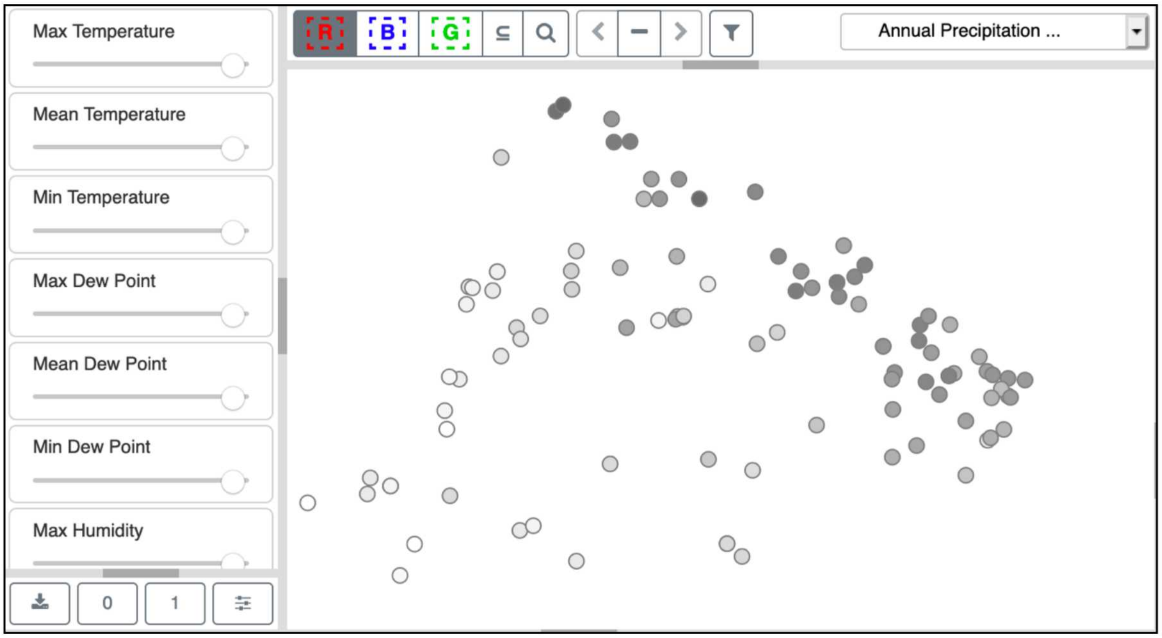


Figure 21: Slider weights (all set to one) and point positions before optimizing for Annual Precipitation.

Figure 22 shows the adjusted slider values and point projections after the optimization. From the sliders that are visible in the figure, we can see that many temporal variables do not contribute to *Annual Precipitation* (i.e. their sliders are set to zero), while others display varying levels of influence. *Mean Humidity* (not shown) exerts the strongest influence, with its slider value set to one. *Max Temperature* and *Min Dew Point* also exert influence.

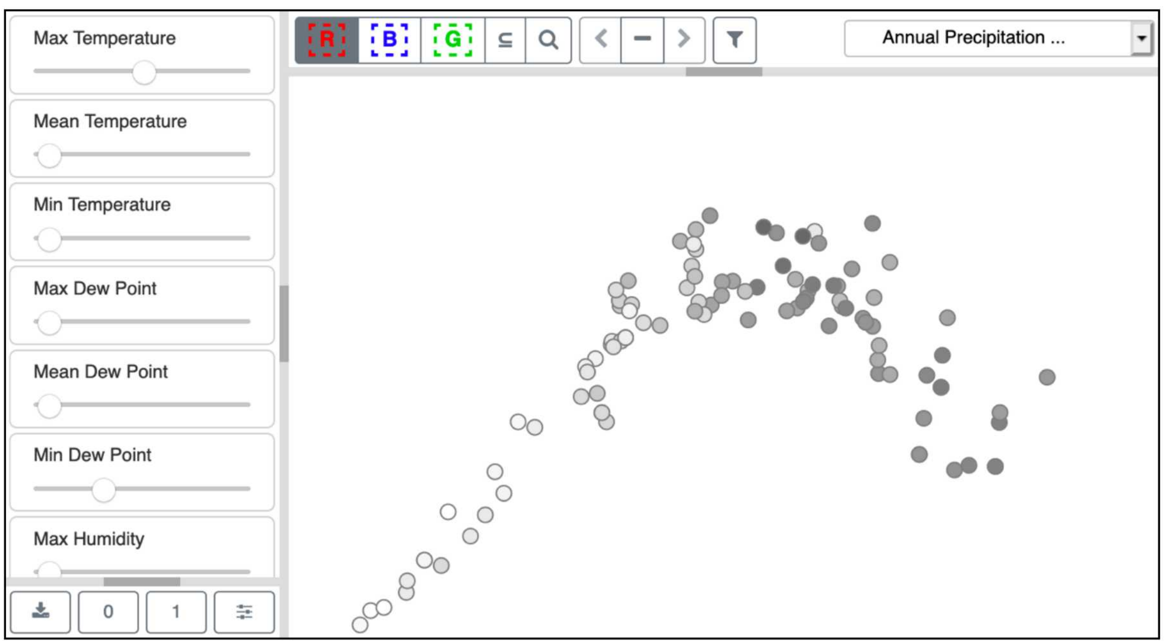


Figure 22: Slider weights and point positions after optimizing for Annual Precipitation.

Metadata Color-Coding

In addition to seeing spatial patterns through point positions, you can color-code the points by variable values from the *Metadata Table* to reveal patterns across the points. By default, the initial view is not color-coded, as shown in **Figure 14**, where *Do Not Color* is shown as the point coloring selection in the upper right corner of the *Scatterplot*. Clicking on the *Do Not Color* button opens a dropdown list of *Metadata Table* columns, as shown in **Figure 20**. The list can be used to select the column values to be used for coloring. The current selection is always shown in the field at the top and highlighted in the list.

After changing the variable to *Average Temperature*, the Scatterplot is recolored using shades of gray (see Model Preferences to change color palettes), where cities with higher average temperature values are colored darker gray and cities with lower values are drawn in lighter gray, as shown in **Figure 23.** Notice that although the MDS projection positions cities with similar values near one another, the highest and lowest values are not found on the extreme edges of the plot, since the projection is based on multiple temporal variables. Honolulu (outlined in red) and Anchorage (outlined in blue) are the points with the highest and lowest average temperatures, respectively.

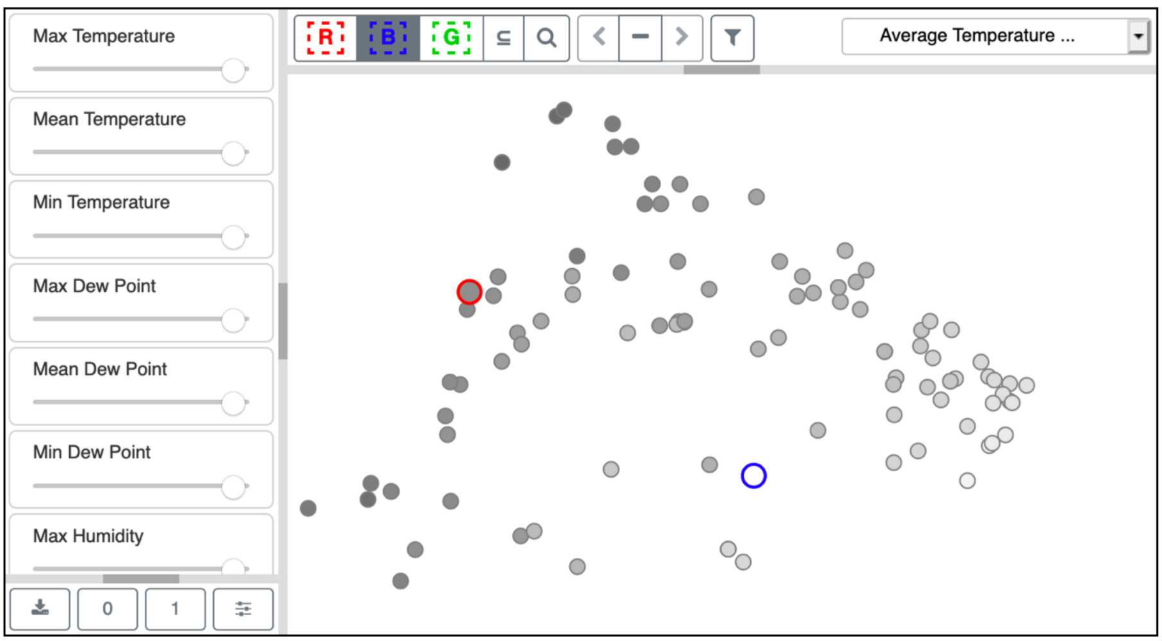
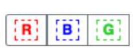


Figure 23: Scatterplot after resetting the color-coding selection to Average Temperature. Honolulu is shown outlined in red, and Anchorage is shown outlined in bue.

Selecting Point Sets 

Thus far, we have explored similarities between ensemble members through their proximity in the abstract space of the *Scatterplot*. However, exploring and comparing temporally changing variables requires comparison of the underlying time series data. DAC does this for small sets of points using plots showing variable values over time to reveal temporal patterns.

Using the *Scatterplot*, we identify potentially interesting sets of points to view as *Time Series Plots*. Up to three sets of points can be defined at one time. The points of each set are rendered with color-coded borders (red, blue, and green). Points retain their fill color, which encodes a point's value for the selected metavariable (see the previous section, **Metadata Color-Coding**).

Figure 24: Scatterplot control bar configured to select ensemble members for the red set.

To select points for the *red* selection, click on the  icon, as shown in **Figure 24**, then select a group of points. This places them in the *red* group. Clicking on the  icon (as shown in **Figure 25**) places subsequently selected points into the *blue* group. Similarly, points can be selected for the *green* group (as in **Figure 26**). Selection actions operate only on the group indicated by the currently highlighted (grayed) set icon. Points can be added to a set through either a click-and-drag rubber-band operation to encircle points, or through individual point selection. Although the initial point can simply be clicked on, adding additional points requires holding down the shift-key or the control-key while clicking or rubber-banding each new point (simply clicking on additional points will not add them to the set). Clicking on the background will clear the current point set. You can remove subsets of points by either shift- or control-clicking a previously selected point, or by holding the shift- or control-key while rubber-banding a subset of previously selected points.

Figure 25: Scatterplot control bar configured to select ensemble members for the blue set.

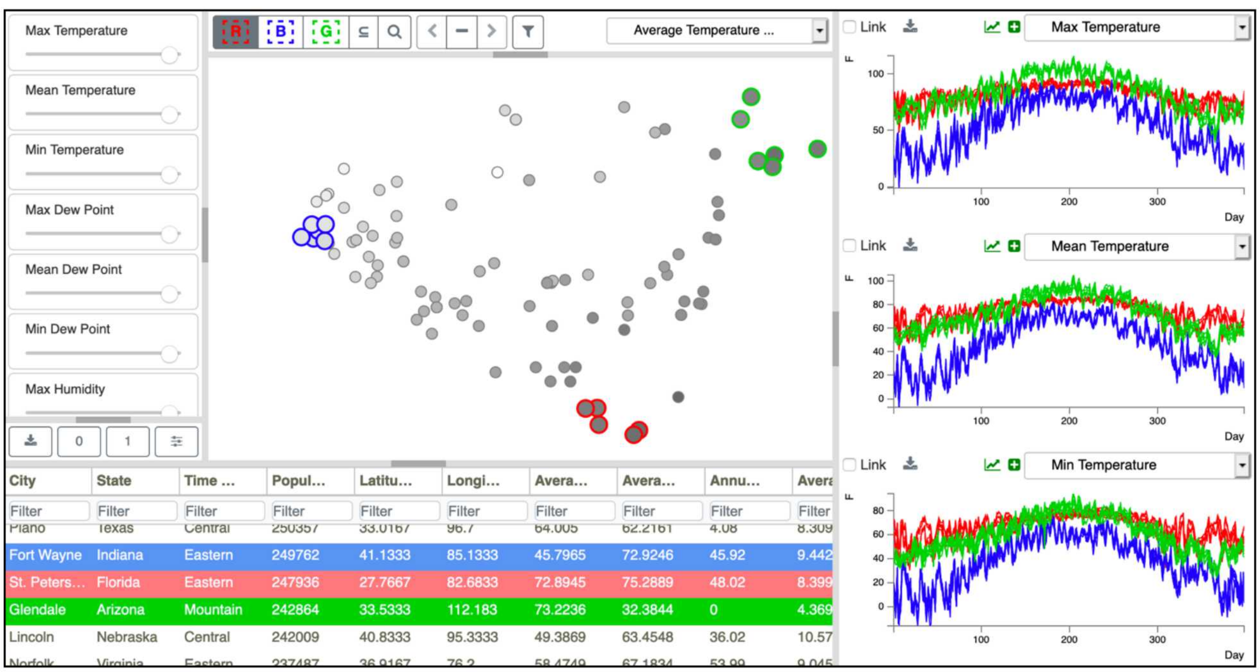


Figure 26: Cities near the extremes of the scatterplot are selected as R, B, and G sets. Looking at the Metadata Table, we see that the red group consists of hot dry cities, the green group as warm humid cities, and the blue group as colder cities. Selections are linked between the Scatterplot, Time Series Plots, and Metadata Table views.

Selections in the *Scatterplot* are linked to both the *Time Series Plots* (right view) and the *Metadata Table* (bottom left view), as shown in **Figure 26**. Plots for the selected points are drawn as correspondingly colored lines, while the selected table rows are highlighted with matching color-coded backgrounds.

Since the links between the *Scatterplot* and the *Metadata Table* are bidirectional, selections can also be made by selecting rows in the table (see **Table Row Selection**).

Comparing the precipitation distribution (**Figure 21**) with the temperature distribution (**Figure 23**), the layout of the points appears to place warm humid cities near the top, hot dry cities in the lower left, and cold wet cities on the right. Selecting points from each of these groups as our color-coded sets, we can explore this hypothesis by identifying the cities in each of the selected sets by scrolling through the *Metadata Table*, as shown in **Figure 26**. Indeed, the cities at the top are in Hawaii and Florida, the cities at the bottom left are in Arizona, and the cities on the right are in New York and the Midwest.

Point Focus Selection

Sometimes it is useful to be able to distinguish one point within a designated set. This is known as focus selection. Clicking one of the previously selected points from within a set will cause that point and its corresponding representations in all views to be highlighted in black. The point is redrawn with a bolded black outline, the corresponding line is drawn in black in all three plots, and the background of the point's corresponding row in the table is rendered in black. Selecting the point for the city of Chandler from the *red* set in **Figure 26**, produces the black highlights shown in **Figure 27**. This *black* highlighting feature is limited to a single point at a time. Focus selection operates as a toggle switch. Clicking on the point again will remove the *black* highlighting. Clicking on different point (within the colored sets) will shift the black highlight to that new point.

Focus selection works in all of the *Scatterplot*, *Metadata Table*, and *Time Series Plot* panes. In other words, you can select either a *red, green*, or *blue* selected point in the *Scatterplot*; a *red, green*, or *blue* selected row in the *Metadata Table*; or a *red*, *green*, or *blue* curve in the *Time Series Plot*. In each case, the corresponding point, table row, or plot will be highlighted in *black*. (See also **Plot Focus Selection** and **Table Row** Selection.)

Figure 27: Selection of the city of Chandler from within the red set is highlighted using a bolded black point outline in the Scatterplot, black lines in the Time Series Plots, and black background for the Metadata Table row.

Subset Analysis

If you want to restrict your analysis to a subset of the points, click the icon (as shown in **Figure 28**), then select the point subset by a click-and-drag rubber-band operation to generate a shaded rectangular region that encompasses the desired points. Releasing the mouse button completes the selection, regenerates the MDS projection based on the reduced point set, and redraws the scatterplot. The icon is now yellowed, indicating that the currently rendered scatterplot is a subset. Click anywhere in the background to return to the full point set.

Alternatively, a point or points may be excluded from the subset by holding the shift key and then clicking on the point that is to be excluded. Multiple points may be excluded by holding the shift key while using rubber-band operation.

Figure 28: Subset icon.

Only points in the current analysis subset will appear in the *Time Series Plot* views, so the number of lines may be reduced if some of the originally selected points have been excluded. In addition, set membership (in the *red*, *blue*, and *green* groups) is limited to only the points in the current analysis, so after taking a subset, set membership is reduced to include just those points. Even after returning to the full point set (by clicking on the background) the reduced selections will be in effect. Further, while the analysis is in subset mode, any rows in the *Metadata Table* that correspond to excluded points are grayed out and non-interactive.

Scatterplot Zoom

Alternatively, you may need to simply zoom in on a subset of points to see a region of the broader analysis set in more detail. Click on the icon shown in **Figure 29***: Zoom icon.*, then select a region within the *Scatterplot* by clicking and holding the mouse button until the shaded region defines the area to be zoomed. Releasing the mouse button redraws the selected region, scaling point positions to fill the scatterplot view. The icon is now yellowed, indicating that the view is zoomed. Zooming in further can be accomplished by repeating the zoom selection process. At any point, you can return to the originally scaled view by clicking anywhere in the background.

Figure 29: Zoom icon.

Unlike the analysis subset operation, set membership remains constant, irrespective of whether selected points are visible in the zoom or not. Consequently, the *Time Series Plots* and the *Metadata Table* rows do not change in response to zooming.

Ordering Time Series Plots

The three buttons to the right of the zoom icon affect the ordering of the *Time Series Plots* and are described in greater detail in **Ordering Plot Displays**.

Filter Plots

The filter button, on the far right of the controls shown in **Figure 29*: Zoom icon.*** is used to filter the *Scatterplot* and *Time Series Plots* in conjunction with the *Metadata Table* filters, as discussed in greater detail in **Table Filtering**.

Time Series Plots

As shown in **Figure 14**, the *Time Series Plots* are initially empty. Once point sets have been selected, plots for those cities of the first three temporal variables (as listed in the *Sliders* view) are generated for the selected points. An example plot is shown in **Figure 30**.

Plot Variable Selection

The field in the upper right shows the name of the temporal variable being plotted. Although only three variables may be plotted at a time, you can select which variables you wish to see. Clicking the variable label pulldown displays a list of the temporal variables available, as shown in **Figure 31**. In the list, the currently selected variable is marked with a check. Change the plot to display a different variable by selecting it from the list.

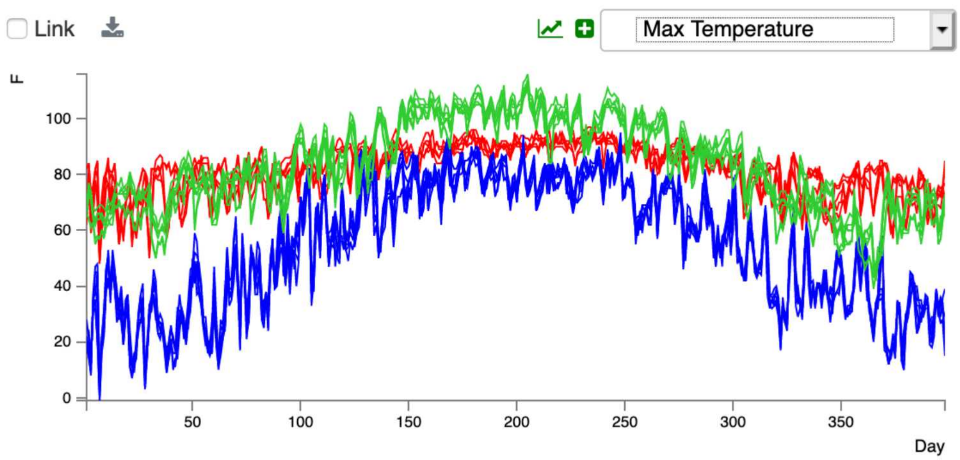


Figure 30: Max Temperature time series plots for selected city sets.

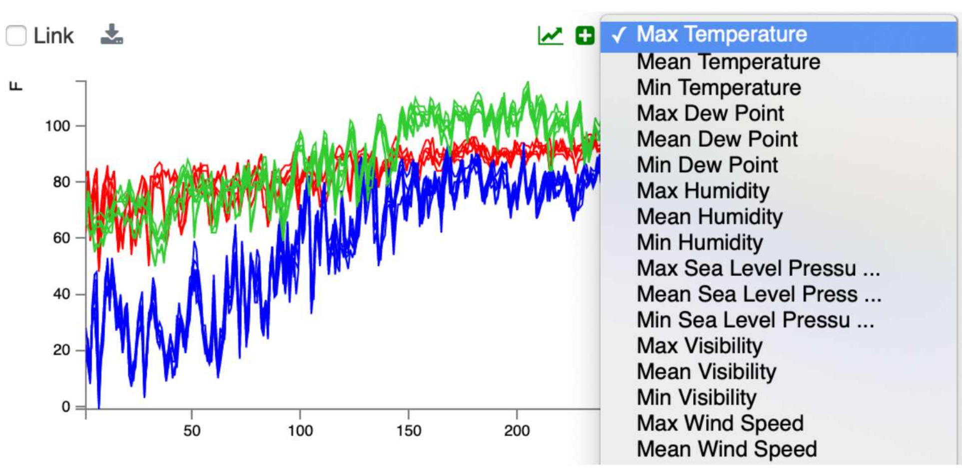


Figure 31: Dropdown temporal variable selection in Time Series Plot.

Ordering Plot Displays

In addition to manually selecting variables for the plots, the plots for all of the temporal variables can be automatically ordered to maximize the difference between the selections by clicking . The ordered plots can then be viewed three at a time, paging forward or backward using the and . buttons, respectively. If the selections change and the ordering is no longer valid, the order icon alerts the user by turning yellow .

Plot Resolution Indicators

There are two green icons to the left of the variable name pulldown. These icons provide information about the resolution of the rendering in the plot. The indicates that every selected point has been drawn as a line in the plot. Once the selected points exceed the *maximum number of plots*, the icon changes to , signaling that not all lines are being drawn. The indicates that each line is being drawn at full resolution. If the sampling in the time series data exceeds the *maximum plot resolution*, the icon changes to , alerting you that lines are not being drawn using the full number of samples.

The threshold values for both the *maximum number of plots* and the *maximum plot resolution* are user adjustable through **Model Preferences**. The default values for these variables are 33 and 500, respectively.

Plot Focus Selection

Just as we can highlight points in selected sets by clicking on them in either the *Scatterplot* or the *Metadata Table*, we can also select them from within a *Time Series Plot*. As your mouse cursor crosses into the rectangle defined by the plot axes, the cursor changes from an arrow to a cross. However, as the cursor moves over a line, it changes back to an arrow. This signals that the line under the tip of the arrow can be selected by clicking. If you click, the line is then highlighted in *black*, as is the corresponding *Scatterplot* point and the corresponding *Metadata Table* row. Focus selection is a toggle, so clicking a second time acts to de-focus the line, point, and row. (See also **Point Focus Selection** and **Table Row** Selection.)

Plot Zooming

Initially, plots are drawn to display the full range of values over both axes. However, subregions can be shown via zooming by rubber-banding within the plot. A user is shown performing a zoom operation in **Figure 32**: Zooming in a time series plot..

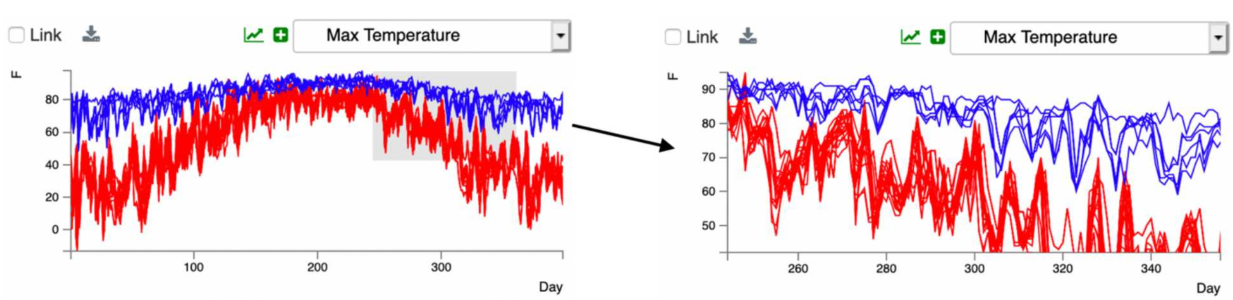


Figure 32: Zooming in a time series plot. The user zooms by rubber-banding to get the grey box on the left. After the box is complete (the mouse button is released), the plot is magnified, as see on the right.

Plot Linking

In the upper left of each time series plot there is a *Link* checkbox. These checkboxes allow the user to perform the same operation on multiple plots simultaneously. For example, if the user links all three plots, then uses zoom in the first plot, each of the other plots will be zoomed to the same scale. In addition, a vertical line indicating the cursor position, normally shown in each individual plot, is duplicated on any linked plots. Two linked plots are shown in **Figure 33**.

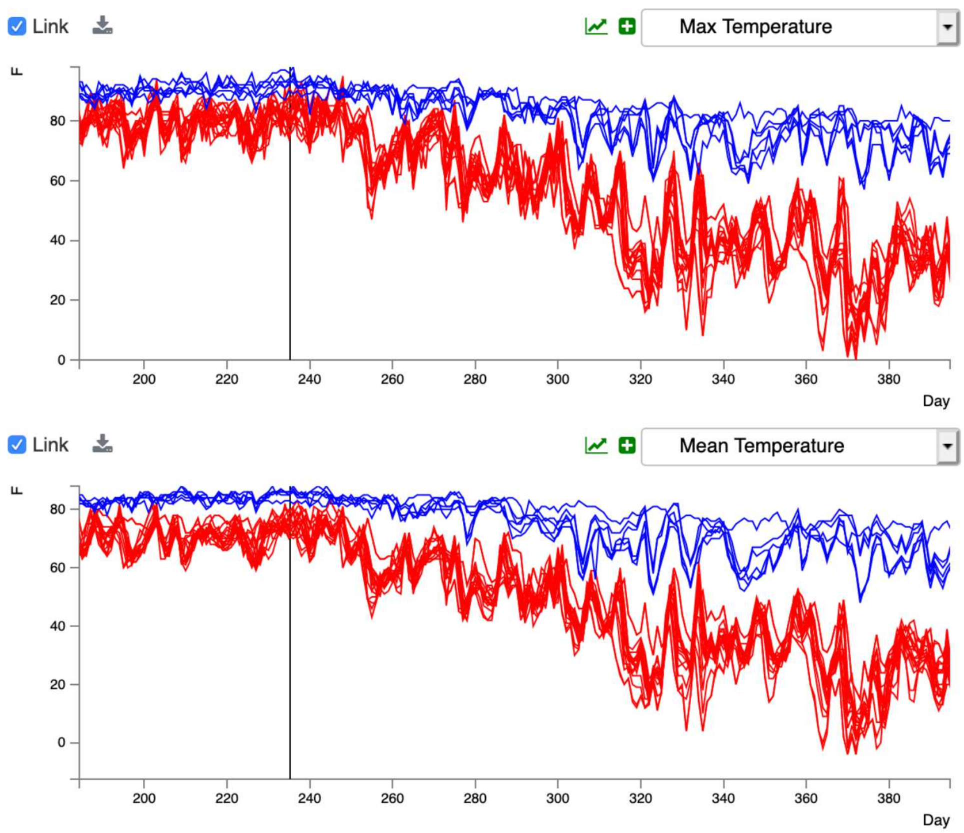


Figure 33: Linked plots, both zoomed to the same scale. The user's cursor position is shown as a vertical line, replicated in both linked plots.

In order to link plots, the time scale (*x*-axis) must be identical in the plots to be linked. If the scales are not identical, an error message will be displayed and the plots will not be linked.

Download Plot Data

Immediately to the right of the *Link* check boxes, there are download buttons for each time series plot. For each plot, this button can be used to export the actual (*x*,*y*) values in the time series plots displayed. The resulting dialog is shown in **Figure 34**.

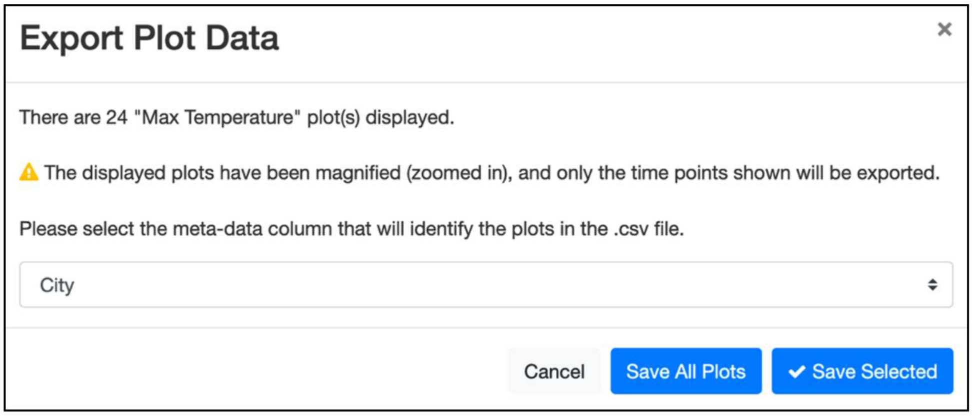


Figure 34: Exporting plot data dialog.

Plot data is exported as a CSV formatted file, and the user must select a *Metadata Table* column to use for identification of the plot in the header row. The default label is the first column in the *Metadata Table*. Additionally, there are a couple of caveats. First, if the plot scale has been zoomed, only the time points shown in the plot will be exported. If you want all time points exported, you will have to close the dialog, reset the zoom, and push the export plot button again. Second, if the plots have been filtered (see **Filter Plots** and **Table Filtering**), then only the plots displayed (selected *and* filtered) will be exported. If you want to export non-filtered plots you will have to close the dialog, clear the filters and push the export plot button again. These caveats apply to both the "Save All Plots" button and the "Save Selected" button. In both cases you are saving only what is displayed in the *Time Series Plots* display.

Metadata Table

The *Metadata Table* presents data provided through non-temporal variables, such as scalars and text. Each row represents an ensemble member, while each column displays values for an ensemble-wide shared variable.

Table Row Selection

As in the case of the *Scatterplot*, ensemble members can be selected via the *Metadata Table*. The selection color is controlled by the color selection icons above the *Scatterplot*. If none of the color selection icons are active, selections are not recognized. If a selection icon is active, there are a few different ways to select rows in the *Metadata Table*. First, if a selection is empty, simply clicking on a row will select that ensemble member. The corresponding point in the *Scatterplot* will be show in the selected color and the corresponding time series will be added to the *Time Series Plots*. This action is shown in **Figure 35**.

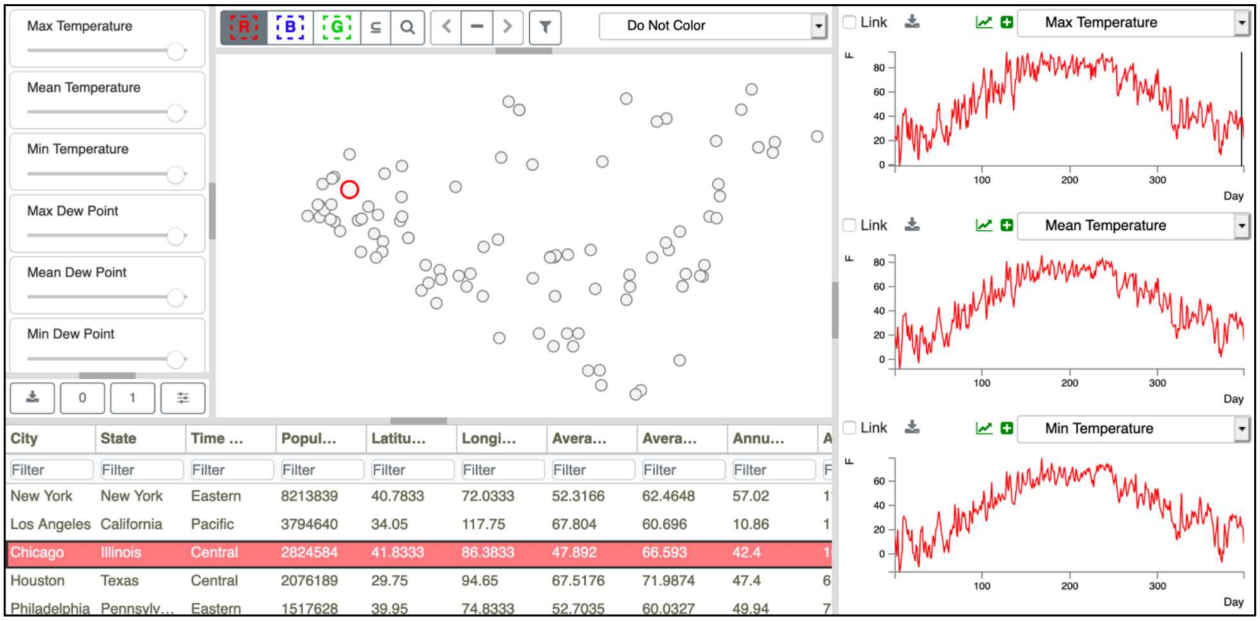


Figure 35: Single row selection from Metadata Table.

Second, if a selection is not empty, another ensemble member can be added by clicking on the new row to bring it into focus (indicated by a black box drawn around the row -- see **Table Focus Selection**), then using the shift key while clicking on the new row.

Third, to select a range from the *Metadata Table*, mark the start of the range by clicking on a row to bring it into focus, then holding the shift key while clicking on the end of the desired range.

Finally, we note that if using a subset of the dataset (see **Subset Analysis** ⊆), rows not in the subset will be highlighted in gray, and will not be available for selection or focus.

Table Focus Selection

The *Metadata Table* allows a focus selection, similar to the focus selection available for the *Scatterplot* (**Point Focus Selection**) and the *Time Series Plots* (**Plot Focus Selection**). To bring an ensemble member into focus selection, simply click on the row of interest in the *Metadata Table*. If that row is a member of one of the colored selections, it will be highlighted in black with a black outline, the corresponding point in the *Scatterplot* will be highlighted in black, and the corresponding plot in the *Time Series Plot* will be drawn in black. If that row is not a member of one of the colored selections, the row will not be highlighted in black, but will still be shown using a black outline. Any table rows not in the current subset (see **Subset Analysis** ⊆) will not be available for focus selection. A focus selection is shown in **Figure 36**.

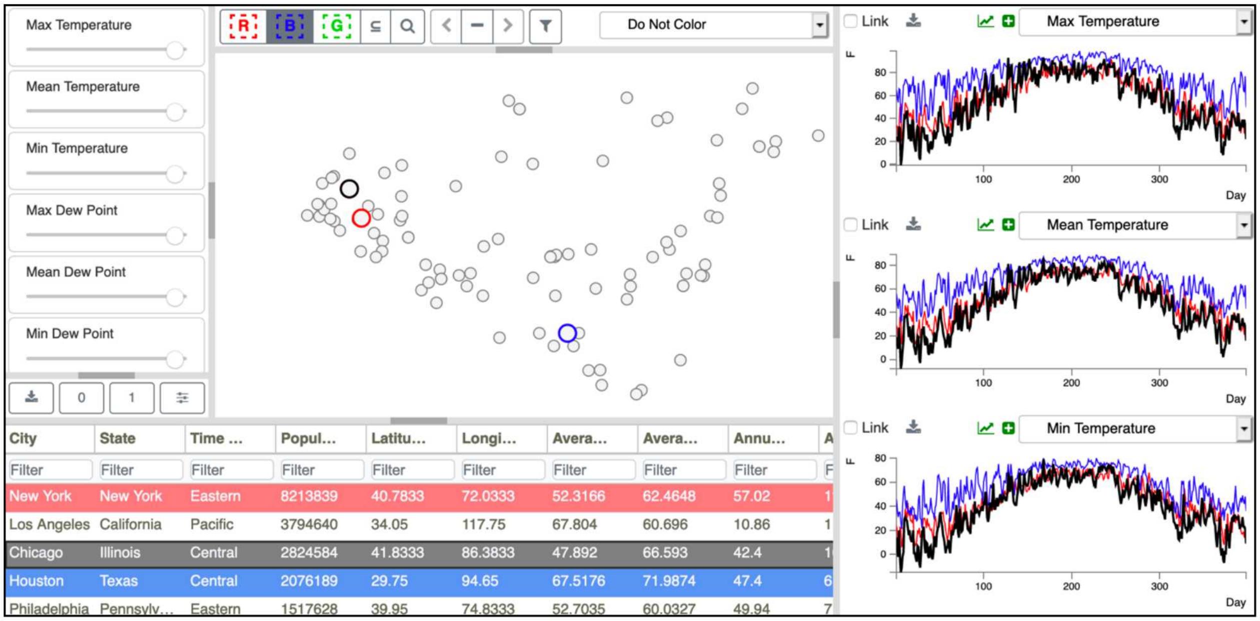


Figure 36: Focus selections are shown in black in the Metadata Table. Corresponding points in the Scatterplot and corresponding plots in the Time Series Plots are also shown in black.

Table Sorting

The rows in the *Metadata Table* can be sorted using the column headers simply by clicking on the column headers. If the full header is truncated (e.g. Latitu …), the full column name can be viewed by hovering over the column header, or column can be resized.

If a table is sorted, the table export button (see **Download Table Data**) will order the exported table rows in the sorted order.

Table Filtering

The rows in the *Metadata Table* can be filtered using the filter boxes below each column header, as shown in **Figure 37**.

City	State	Time …	Popul…	Latitu…	Longi…	Avera…	Avera…	Annu…
Filter	New	Filter	Filter	Filter	Filter	Filter	Average Temperature (F)	
New York	New York	Eastern	8213839	40.7833	72.0333	52.3166	62.4648	57.02
Albuquer…	New Mexico	Mountain	496801	35.0833	105.35	57.2211	41.5327	9.55
Buffalo	New York	Eastern	277998	42.9167	77.1667	45.1759	69.9925	45.94
Newark	New Jersey	Eastern	276200	40.7333	73.8333	52.3116	61.9623	54.92
Jersey City	New Jersey	Eastern	236808	40.714	74.071	50.8518	73.0804	47.03

Figure 37: Filtering the Metadata Table entries using the filter boxes below the column headers.

By default, the table filters use case insensitive substring search of the column entries to filter the rows of the table. However, regular expressions can also be used to filter the table. Some examples of regular expression searches are:

- ^3152 – find all serial numbers starting with 3152.
- ^3152-00(125) – find serial numbers 3152-001, 3152-002, or 3152-005.
- (Eastern|Pacific) – find cities in Eastern or Pacific time zones.

When a filter is inactive, the filter box is colored white. When a filter is active, the filter box will be colored yellow. If a filter is an incorrect regular expression, the filter will be colored red. In this case the filter is active but using case insensitive substring search. (Typically, this will cause the entire table to be filtered away.) Once the regular expression has been corrected, the filter will go back to yellow.

The table filters work in real time, so that the table is filtered as you type. Once you have filtered the table to obtain only the rows that you desire, the *Scatterplot* and *Time Series Plots* can also be filtered by pushing the filter plots button (see **Filter Plots**) above the *Scatterplot*. The points in the *Scatterplot* will then be restricted to display only those ensemble members in the filtered version of the table. In addition, the plots in the *Time Series Plots* will also be restricted to those shown in the table.

After pushing the filter plots button, it will be turned yellow to indicate that you are viewing a restricted set of the data. Any colored selections not in the filtered table will also be hidden, although the selections are not forgotten. Pushing the filter plots button again will restore the points and plots, and the selections will be unaffected.

Adding Table Columns

Generally speaking, data cannot be changed once a model has been loaded. However, columns can be added the *Metadata Table*, so that the user may enter notes or designate a given ensemble member as belonging to a specified category. Columns are added using the yellow Edit > Model Table pulldown in the upper right of the Slycat™ screen, as shown in **Figure 38**.

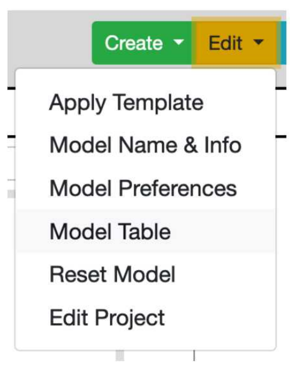


Figure 38: Edit pulldown menu with Model Table selected.

After selecting Model Table from the Edit pulldown, a wizard appears to help you add columns to the *Metadata Table*, as shown in **Figure 39**.

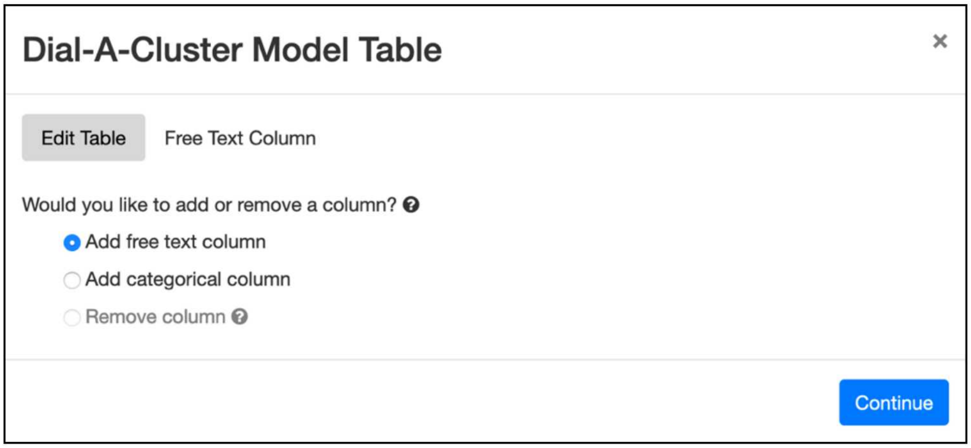


Figure 39: Initial screen for the add column wizard in Dial-A-Cluster.

DAC allows you to add two types of *Metadata Table* columns. The first type of column is free text, which lets you type notes into the table describing a particular ensemble member. To add a free text column, select “Add free text column” in the wizard and push the *Continue* button. In this case you will see the next screen in the wizard, shown in **Figure 40**.

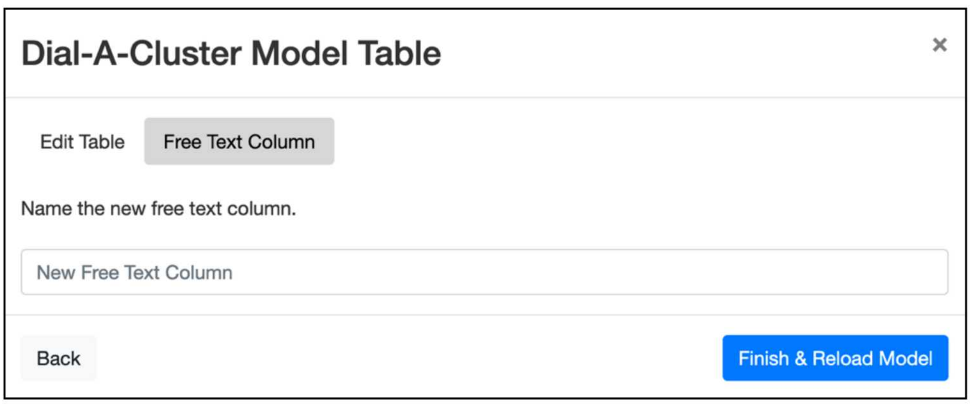


Figure 40: Add free text column wizard option.

To add a free text column, simply type the name of the new column header into the wizard using the “New Free Text Column” text box and push *Finish & Reload Model*. After the model has been reloaded, a new column will appear in the *Metadata Table* to the far right, as shown in **Figure 41**. To add notes to a row in the new column click on that row and type the notes. Once you press enter or otherwise finish the entry, it will be saved with the model as part of the table.

City	State	Time ...	Popul...	Latitu...	Longi...	Avera...	Avera...	Annu...	Avera...	Avera...	Avera...	Notes
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
New York	New York	Eastern	8213839	40.7833	72.0333	52.3166	62.4648	57.02	11.5126	9.1206	5.27889	
Los Angeles	California	Pacific	3794640	34.05	117.75	67.804	60.696	10.86	1.67839	8.94472	1.91457	
Chicago	Illinois	Central	2824584	41.8333	86.3833	47.892	66.593	42.4	10.1859	8.80151	5.46734	
Houston	Texas	Central	2076189	29.75	94.65	67.5176	71.9874	47.4	6.89196	9.27638	4.70854	If you're ever in Houston ...
Philadelphia	Pennsylv...	Eastern	1517628	39.95	74.8333	52.7035	60.0327	49.94	7.24372	8.8794	3.48492	

Figure 41: Entering free text into the Metadata Table.

The second type of column available in the add column wizard is a categorical column. A categorical column lets you define a set of categories for a particular ensemble member. Once added, the user can label the *Metadata Table* row using only those categories. To add a categorical column, select "Add categorical column" in the dialog from **Figure 39** and press *Continue*. You will see the dialog in **Figure 42**.

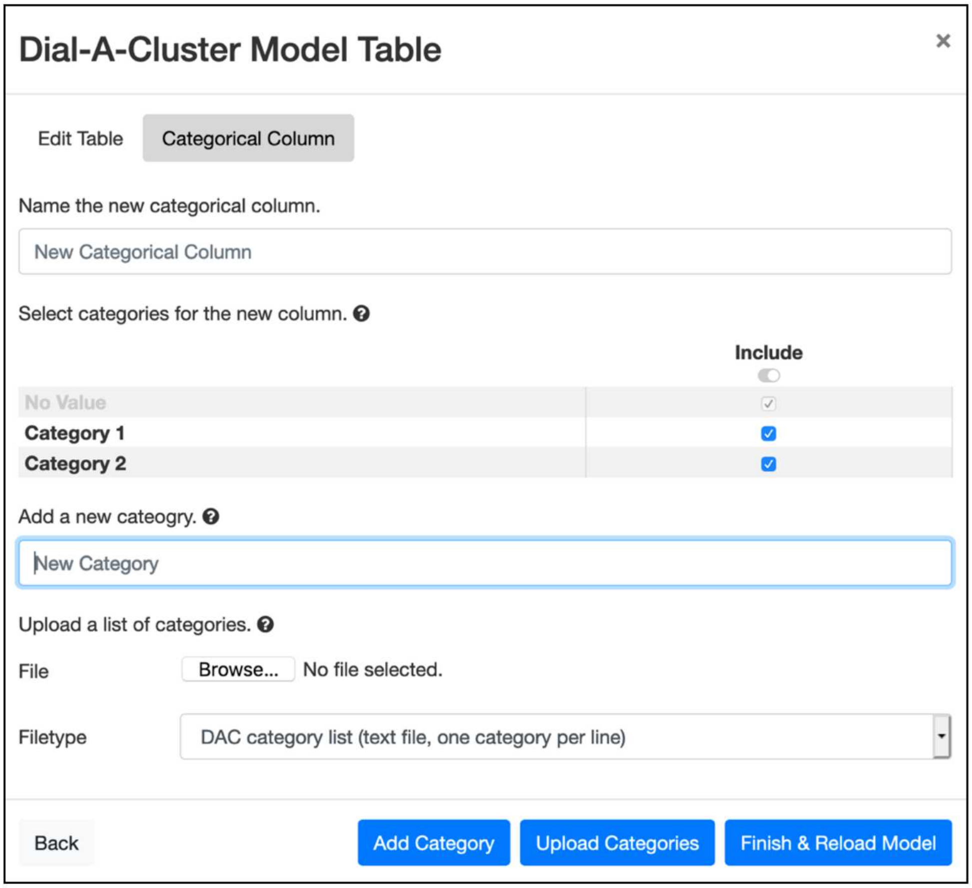


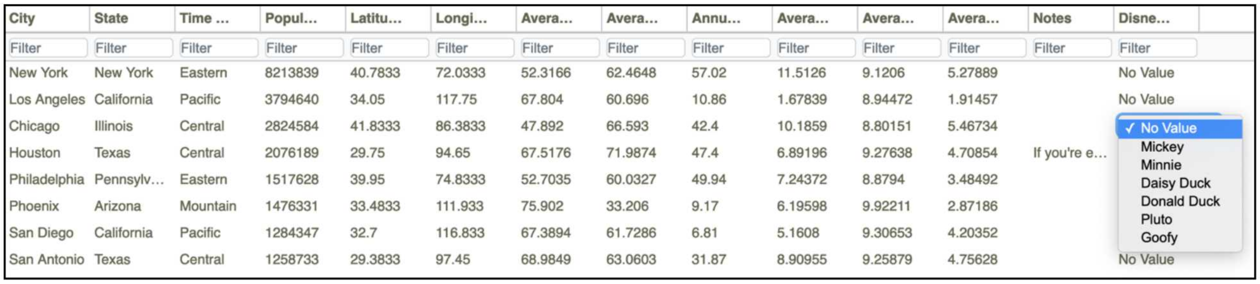
Figure 42: Wizard used to add a categorical column to the Metadata Table.

First, type the column name into the "New Categorical Column" text box. Next, add categories to your column. The categories can be added either manually by entering each category into the "New Category" text box and pushing the *Add Category* button, or uploaded from a file. In **Figure 42**, two categories have been added manually, named "Category 1" and "Category 2". These categories are shown in a list in the middle of the dialog. The "No Value" category is a required member of any category list.

To upload a list of categories, select a file using the "Browse …" button. The file should be a text file containing one category per line. Once a file has been selected, press the *Upload Categories* button. The categories from the file will be displayed in the list in **Figure 42**, where "Category 1" and "Category 2" are shown.

Once all the desired categories have been entered, you can make a final selection using the check boxes in the list in **Figure 42**. If you have entered too many categories by mistake or uploaded categories that you don't need these can now be unselected. Press the *Finish & Reload Model* button to add the new categorical column to the *Metadata Table*.

After the model has been reloaded, a new column will appear in the *Metadata Table* to the far right, as shown in **Figure 43**. To mark a row in the new column with a given category, click on that row and select the category. Once you press enter or otherwise finish the entry, it will be saved with the model as part of the table.



City	State	Time ...	Popul...	Latitu...	Longi...	Avera...	Avera...	Annu...	Avera...	Avera...	Avera...	Notes	Disne...
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
New York	New York	Eastern	8213839	40.7833	72.0333	52.3166	62.4648	57.02	11.5126	9.1206	5.27889		No Value
Los Angeles	California	Pacific	3794640	34.05	117.75	67.804	60.696	10.86	1.67839	8.94472	1.91457		No Value
Chicago	Illinois	Central	2824584	41.8333	86.3833	47.892	66.593	42.4	10.1859	8.80151	5.46734		
Houston	Texas	Central	2076189	29.75	94.65	67.5176	71.9874	47.4	6.89196	9.27638	4.70854	If you're e...	
Philadelphia	Pennsylv...	Eastern	1517628	39.95	74.8333	52.7035	60.0327	49.94	7.24372	8.8794	3.48492		
Phoenix	Arizona	Mountain	1476331	33.4833	111.933	75.902	33.206	9.17	6.19598	9.92211	2.87186		
San Diego	California	Pacific	1284347	32.7	116.833	67.3894	61.7286	6.81	5.1608	9.30653	4.20352		
San Antonio	Texas	Central	1258733	29.3833	97.45	68.9849	63.0603	31.87	8.90955	9.25879	4.75628		No Value

Figure 43: Selecting a category for a row of the Metadata Table.

Derived Models

On occasion, users may want to add additional ensemble members to an already loaded model, or even change the time series data in a model. This cannot be done in general, but it can be done in two specific cases. In the first case, a number of models can be combined to create a new model. In the second case the time series axis for different variables can be reduced, and a new model can be computed based on the restricted time series data.

Combining Models

Two or more existing models can be combined to create a new larger model. From an existing model, select Create > Combined Model from the green *Create* pulldown in the upper right of Slycat™ user interface. This pulldown option is shown in **Figure 44**. Note that this must be done from within an existing model. The pulldown menu options will be different at the project level.

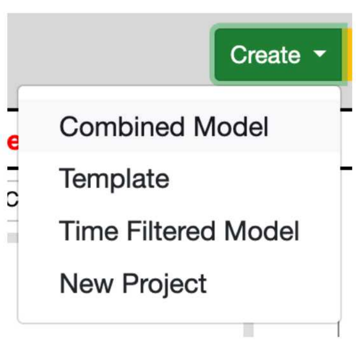


Figure 44: Combining multiple existing models to create a new model.

Selecting Create > Combined Model will open a wizard for creating a combined model, as shown in **Figure 45**. The first step in creating a combined model is choosing whether the *Scatterplot* coordinates will be re-computed from scratch or the combined model data will be projected onto the current model (the model used to launch the combined model wizard). To re-compute the *Scatterplot* coordinates using all model data equally, select "Re-compute new model" and press *Continue*. To use the current model as a basis for the *Scatterplot* coordinates, select "Project onto this model" and press *Continue*.

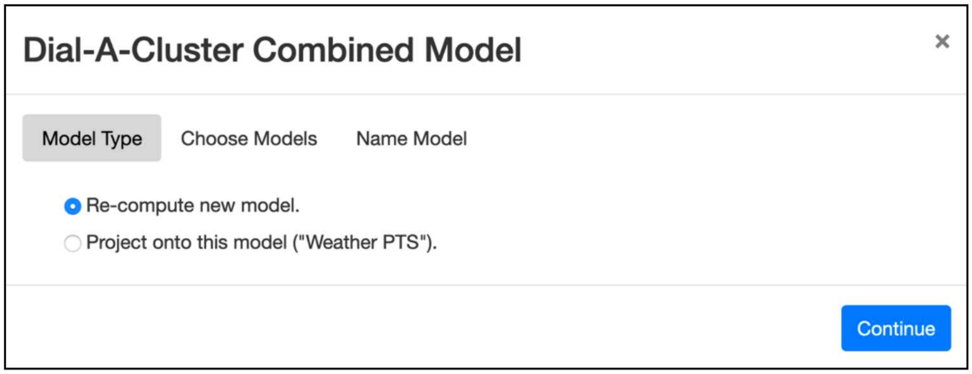


Figure 45: Initial dialog for creating a combined model.

The next dialog in the combined model wizard is shown in **Figure 46**. In this dialog, the user is presented with a list of available models in the current project. Compatible models may be selected from the list and used to create the new combined model. Compatible models must have identical *Metadata Table* headers and equal number of time series. If the user has checked the box next to "Intersect time points if mismatches are found during model computation," compatible models may have different number of time points for a given time series. Otherwise they must also have identical time points for each time series. If you select an incompatible model, the wizard will prompt you and continue only when all selected models are verified to be compatible.

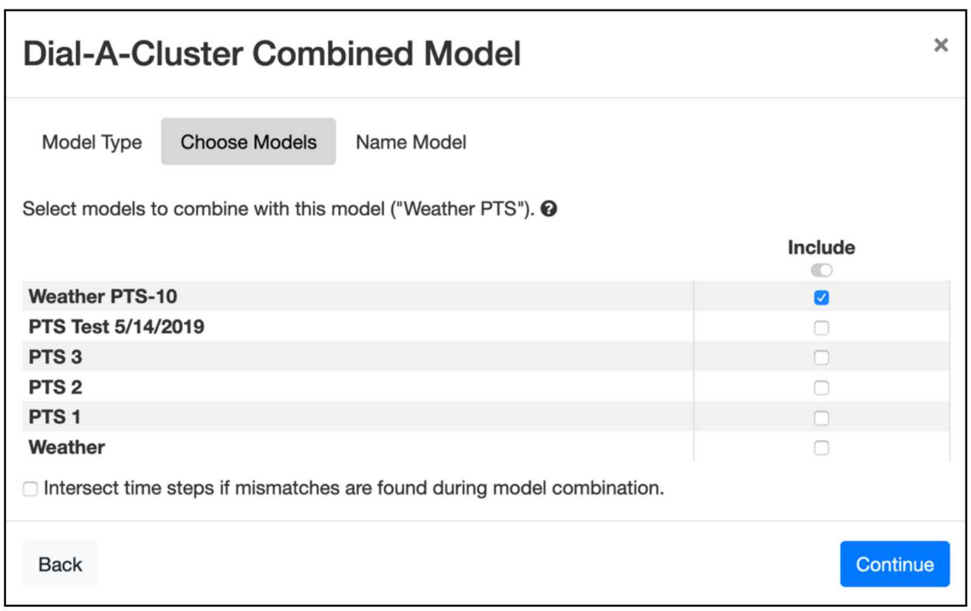


Figure 46: Selecting models to use for creating a combined model.

Once compatible models have been selected, push the *Continue* button to advance the wizard. The final dialog allows you to name and mark the model the combined model, in the same way that new models are named (see **Figure 11**). After naming the model, an indicator and accompanying message window will be presented to show progress and any problems making the combined model (see **Figure 13**).

Time Filtered Models

The second option for altering data in a pre-existing model is the restriction of the time axis in the *Time Series Plots*. Such restrictions, however, also alter between curve similarity values, making it necessary to re-compute the coordinates in the *Scatterplot*. Therefore, restricting the time axis values in the *Time Series Plots* require the creation of a new, derived model. To create a time filtered derived model, select Create > Time Filtered Model, as shown in **Figure 47**. As with combined model creation, this must be done from within an existing model.

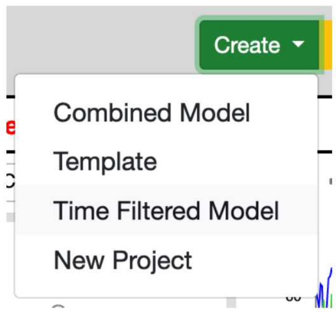
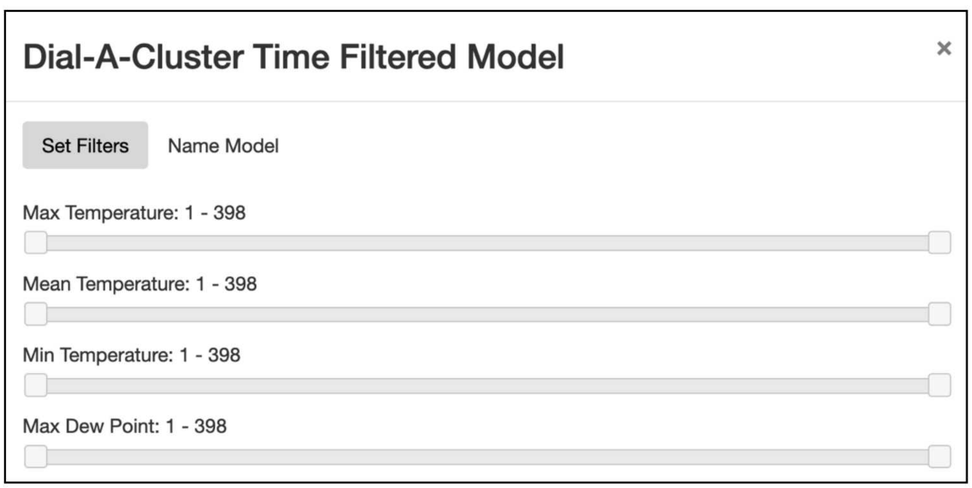


Figure 47: Pulldown menu for creating a time filtered model.

The time filtered model creation wizard consists of two dialogs. The first dialog is shown in **Figure 48**. In this dialog, simply adjust the sliders for a given time series plot to the desired range. After the desired restrictions have been performed, push the *Continue* button. To return the ranges to their full available values, push the *Reset* button.



⋮

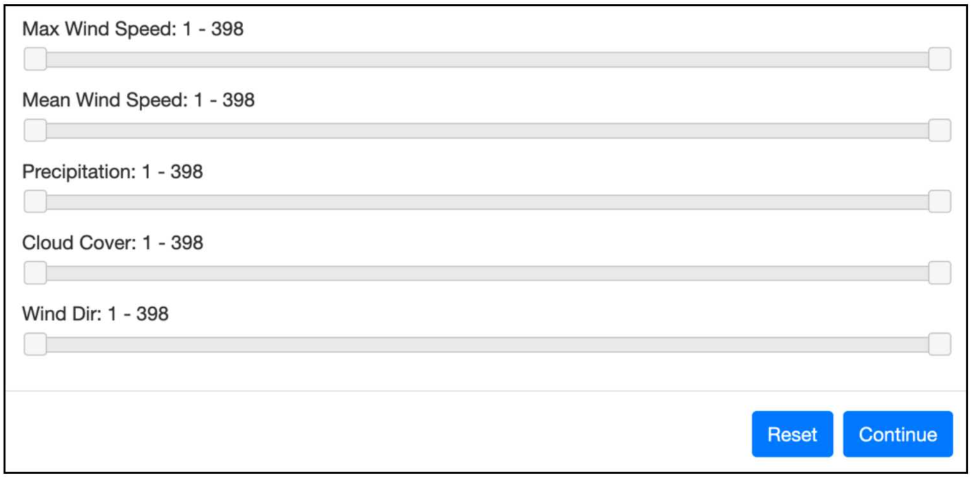


Figure 48: Time filtered creation wizard. Sliders allow the user to restrict any given time axis.

As in the combined model wizard, the final dialog allows you to name and mark the new time filtered model. This is done in the same way that new models are named (see **Figure 11**). After naming the model, an indicator and accompanying message box will be presented to show progress and any problems making the time filtered model (see **Figure 13**).

Model Preferences

There are various parameters that are used to calibrate the DAC user interface experience and performance. These parameters can be adjusted to tune the interface according to preference. The parameters can be changed using the model preferences wizard, accessible under the Edit > Model Preferences pulldown menu, as shown in **Figure 49**. The model preferences are stored separately for each model so are adjusted on a per model basis. In other words, use Edit > Model Preferences from the model you are interested in tuning.

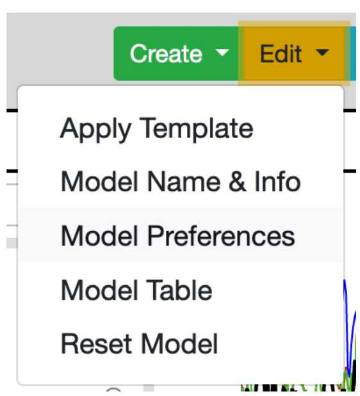


Figure 49: Edit Model Preferences pulldown menu.

Temporal Variables

The first dialog you will see upon launching the Model Preferences wizard is shown in **Figure 50**. This dialog allows the user to display only the variables of interest. Other variables will not be shown in the *Sliders* pane, and the corresponding sliders will be set to zero. Additionally, those variables will not be displayed in the *Time Series Plots* variable selection dropdowns (see **Figure 31**).

The variables are not in fact removed from the model, however, only hidden to provide a less cumbersome interface. By revisiting this dialog, you can re-display variables that were previously hidden. Once the desired variables to display are selected, press the *Continue* button.

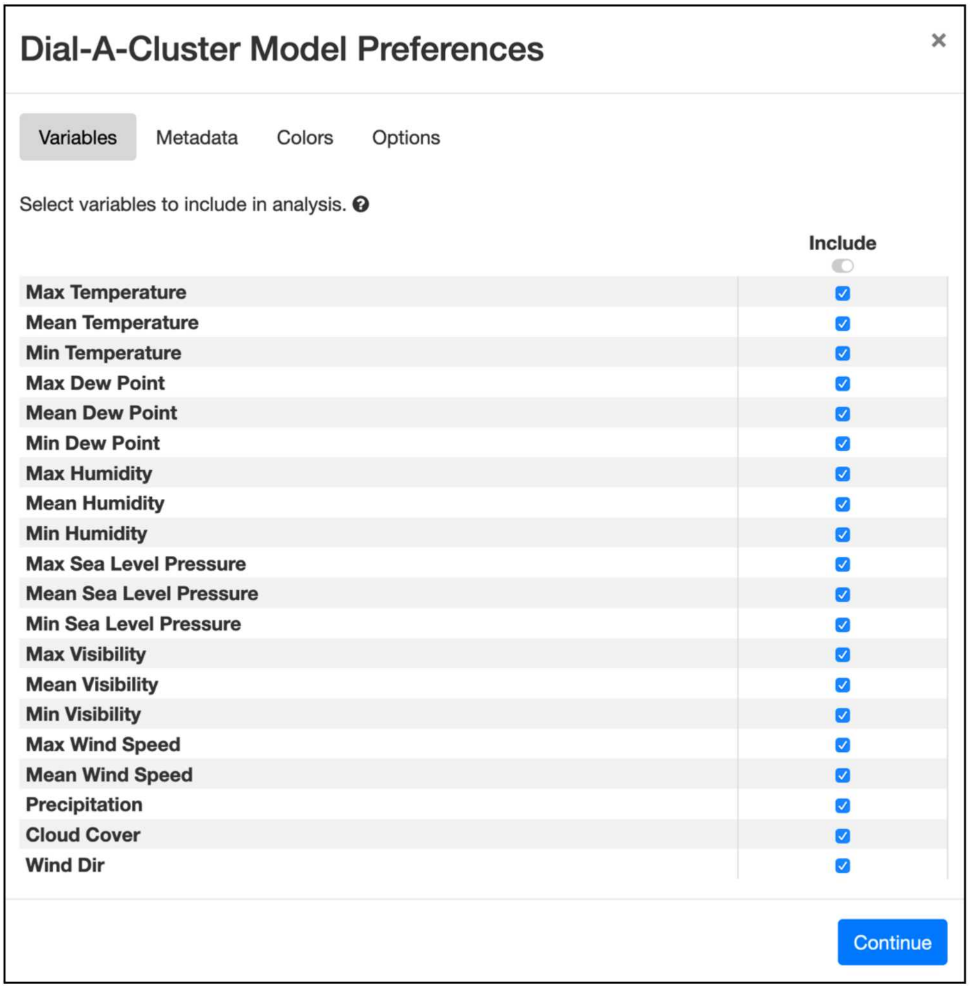


Figure 50: Selecting variables to display in a model.

Table Columns

In addition to selecting variables to display in the *Sliders* pane and *Time Series Plots*, you can also select the columns in the *Metadata Table* that you would like to display. The dialog for selecting columns in the *Metadata Table* is shown in **Figure 51**. By hiding columns, you can create a smaller metadata table. The hidden columns will also be unseen in the color selection pulldown above the *Scatterplot* (see **Figure 20**).

As in the case of hidden variables, columns are not in fact removed from the model. They are only hidden to provide a less cumbersome interface. By revisiting the column selection dialog, you can re-display columns that were previously hidden. Once the desired columns to display are selected, press the *Continue* button.

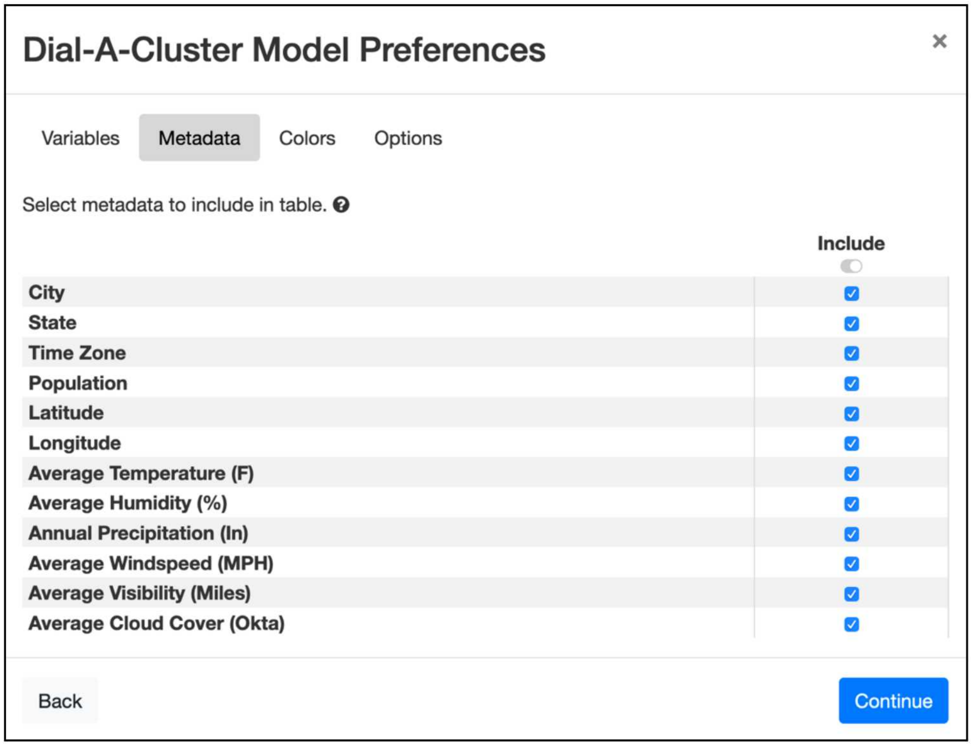


Figure 51: Selecting table columns to display in the Metadata Table.

Color Palettes

The default color scheme for shading points using table columns is gray scale. This color scheme can be overridden using the color palette selection dialog, shown in **Figure 52**. The color selection dialog has three available palette types for use: sequential, diverging, and discrete. These different palette styles can be selected using the radio buttons under "Choose: ● sequential scales, ○ diverging scales, or ○ discrete scales." The diverging scales are shown in **Figure 53**. The discrete scales are shown in **Figure 54**. In each case, click on a scale to make a selection. In **Figure 53**, the red-to-blue scale has been selected. Press the *Continue* button to advance to the final dialog.

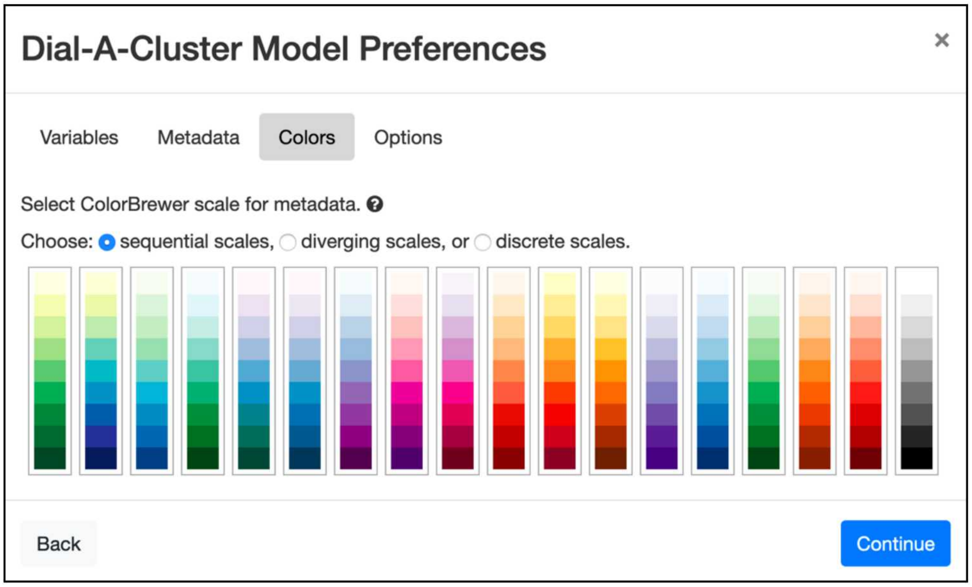


Figure 52: The color selection dialog, showing the sequential palettes.

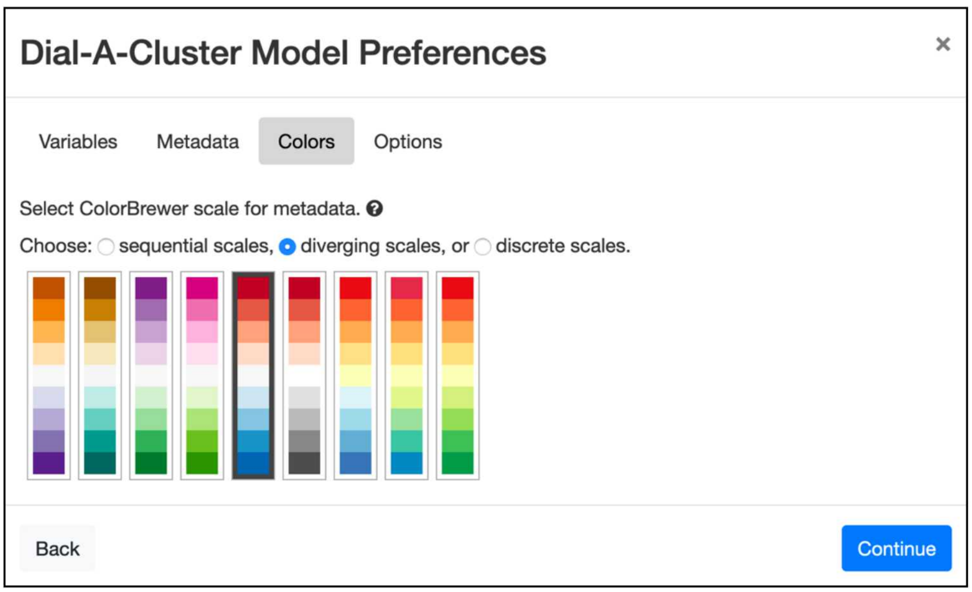


Figure 53: The diverging color scales, with red-to-blue selected.

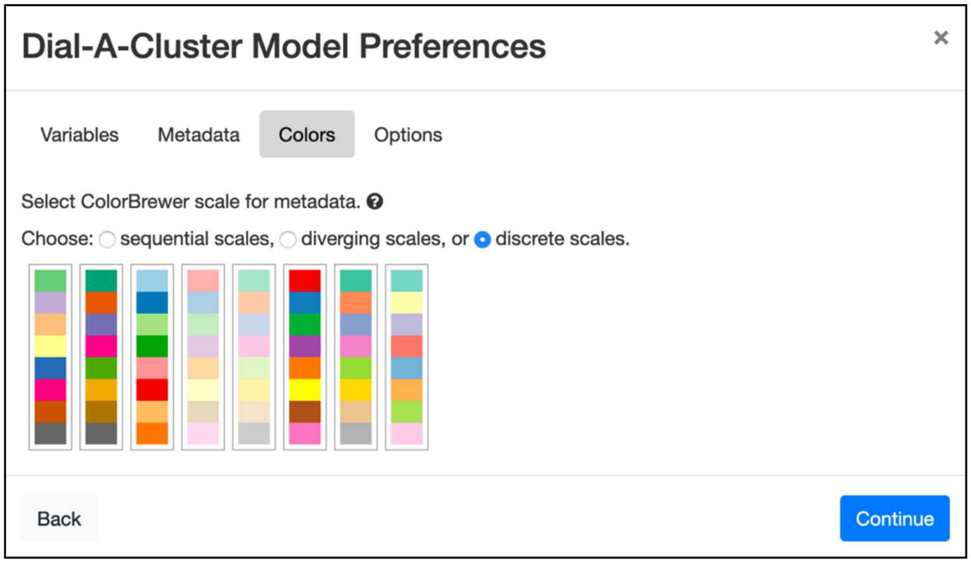


Figure 54: The discrete color scales.

Options

The final dialog in the Model Preferences wizard allows the user to adjust performance-oriented parameters. It is shown in **Figure 55**. The first adjustable parameter is the "Max. label length," an integer which specifies how long variable names can be before they are truncated to prevent pulldown menus from being overly wide. The default value is 20. Names longer than 20 characters are truncated with an appended ellipsis (e.g. "Mean Sea Level Press …" instead of "Mean Sea Level Pressure").

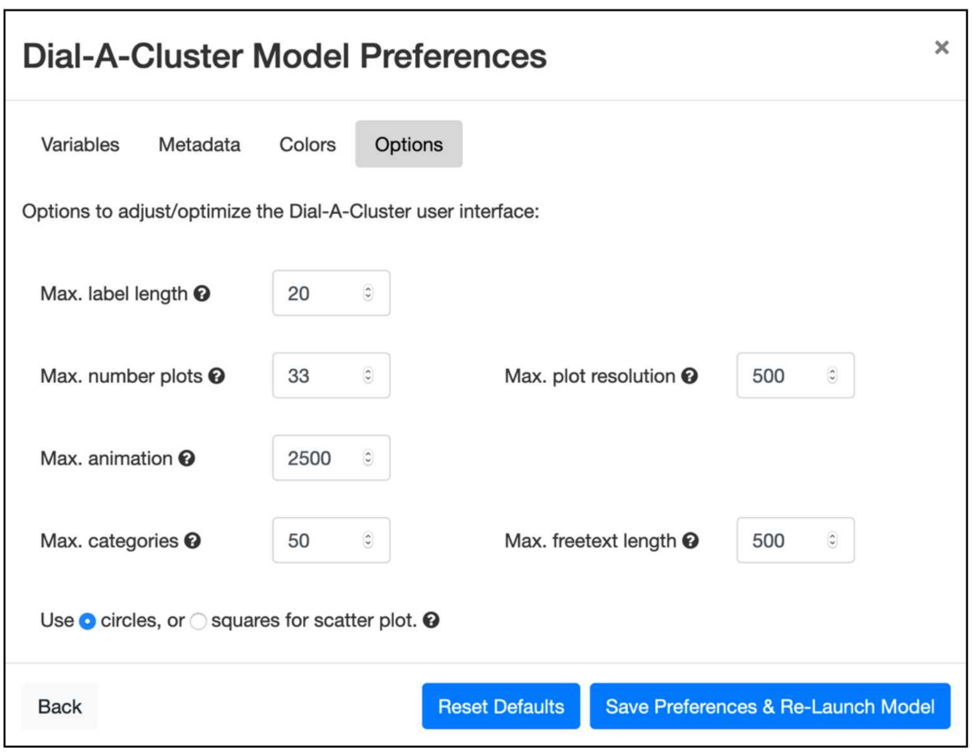


Figure 55: Performance-oriented parameters available to modify in the Model Preferences wizard.

The second adjustable parameter is “Max. number plots,” which specifies the maximum number of *Time Series Plots* to display per variable. The default value for this parameter is 33. This allows the user to make arbitrarily large selections but will only plot a random subset of the corresponding plots. This is necessary to prevent performance delays from rendering too many plots. A smaller number for the maximum number of plots will result in a faster user interface experience and a larger number will result in a slower user interface experience. In case the maximum number of plots is exceeded in the selection, an indicator will be displayed warning that you are not viewing all the plots in the selection (see **Plot Resolution Indicators** for details).

The third adjustable parameter is “Max. plot resolution,” also related to the *Time Series Plots*. The maximum plot resolution specifies how many time points to use when rendering the *Time Series Plots*. The default number is 500. This is again necessary to prevent performance delays due to browser rendering speed. For a faster user interface experience, use a lower maximum plot resolution. There is a corresponding indicator which is displayed when the plot resolution exceeds the maximum specified (see again **Plot Resolution Indicators**) for details.

The fourth adjustable parameter is “Max. animation.” This integer that defaults to 2,500 and determines the maximum number of points in the *Scatterplot* beyond which changes in the coordinates are no longer animated. If animation during zooming, sub-setting, or adjusting the sliders in the *Sliders* pane is jumpy and/or too slow, this parameter can be lowered to avoid animation between the old and new coordinates.

The fifth parameter “Max. categories” specifies the maximum number of categories allowed in a user-added categorical column in the *Metadata Table*. See **Adding Table Columns**. The sixth parameter “Max. freetext length” is also related to the *Metadata Table* user-added columns. The maximum freetext length specifies the maximum allowed length of notes input into the text box of a freetext user-added column.

Finally, the DAC user interface allows the use of squares to designate points in the *Scatterplot*. This option can be specified by toggling the radio buttons in “Use circles, or squares for scatter plot.” The option is included to improve rendering speed for the *Scatterplot* in the browser, as squares are rendered more quickly than circles.

Bookmarks and Templates

In addition to the Dial-A-Cluster specific features that have been so far described, it is worth noting that Bookmarks and Templates are also available to DAC users. These are described in the Overview section of the Slycat™ manual.

Acknowledgements

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. This document is approved for release as xxxxx

References

[1] Martin, S., and Quach, T.-T., *Interactive Visualization of Multivariate Time Series Data.* Human Computer Interaction International (HCII), Foundations of Augmented Cognition: 322-332 (2016). DOI 10.1007/978-3-319-39952-2_31

Sandia National Laboratories

National Nuclear Security Administration