The MobiView graphical user interface

Magnus Dahler Norling

December 11, 2019

Contents

1   Introduction

MobiView is a graphical user interface that can load any model built using the Mobius dll interface. It is created using the Ultimate++ framework and the ScatterCtrl package (`ultimatepp.org`).

This document is a work in progress.

2   The toolbar

The toolbar has 9 buttons.

1. Load (ctrl-O). This asks you to select a model dll, an input file, and a parameter file. (On Linux, you load a shared object .so file instead of a dll). The input and parameter files have to be of the .dat format described in the Mobius file format documentation. You can click load even if you have a model and dataset loaded already. This will delete all data from the earlier dataset from memory, so be sure to save any changes you want to keep first.

2. Save parameters (ctrl-S). This saves any edits you have made to the parameters to the current working parameter file (usually the one you loaded).

3. Save parameters as (alt-S). Saves the parameters to a new file. The new file is now the current working file.

4. Search parameters (ctrl-F). Opens a new window that allows you to search for parameters by name. Any matching parameters are displayed in a list, and clicking an item in the list takes you to the right parameter group in the main view. The search is a case-insensitive substring search.

5. Visualize reach branches (ctrl-R). Only works for models that have an index set called 'Reaches'. Opens a new window that visualizes the connectivity of the indexes in this branched index set (e.g. the connectivity of river sections). No information about actual length or position is used in the visualization, only connectivity.

6. Run model (F7). Runs the model using the parameters that are loaded in the MobiView interface, taking into account any edits. (It does not matter if the edits have been saved to file or not). Results from the latest model run are available for plotting.

7. View model equation batch structure. This displays the equation batch structure, as described in the Mobius model builder documentation. This is mostly interesting if you are a model developer, as it can provide some debug information for your work-in-progress model.

8. Save baseline (ctrl-B). Saves a background copy of the current dataset. This is only used in the "Compare baseline" plot mode.

9. Export to csv (ctrl-E). Allows you to save all selected time series in a .csv format. The separator is ;

3   Parameter viewing and editing

4   The plot

You can use the plot to visualize inputs and results of the model. Time series can be selected in the Equation and Input lists to the right of the plot (results can only be selected if the model has been run at least once). You can select multiple time series at one time by ctrl-clicking (or shift-clicking) them. You can also remove a selected time series by ctrl-clicking it. If a time series varies over one or more index sets, you can select indexes from the lists below the plot. You can do multiselection here too.

The timeseries info box will display info about the selected timeseries. If you are in a residual-type plot mode, it will also display goodness-of-fit statistics (see section 4.5).

The plot will automatically update itself every time you run the model to reflect any changes in the result time series.

4.1   Plot options

You can choose between several different plot options. First, you can choose a few different plot modes. These are described in the next subsection. Depending on the plot mode, you may also have other options available

i Aggregation. You can choose to aggregate the displayed timeseries on monthly or yearly steps. You can also choose between "mean" or "sum" as the aggregation mode. Note that the sum aggregation only works well if all timesteps have a value (which may not be the case for some input series). Note: For models with other than daily timesteps, monthly or yearly aggregation may not always make sense. Other forms of aggregation like min or max could also be added. This is a work in progress.

ii Y axis transformation. You can choose between three Y axis transformations

   (a) You can have no transformation (regular Y axis).

   (b) You can normalize it. In this case, every displayed time series is normalized separately so that it takes values between 0 and 1. This is useful if you want to compare the shape of time series with very different scales.

   (c) Logarithmic Y axis. This changes the scale of the axis to be a base 10 logarithm.

iii Scatter inputs. Determines if input time series should be displayed as scatter plots or line plots. Scatter plots work better if there are a lot of holes in the input data.

4.2   Plot modes

4.2.1   Regular

The Regular plot mode will just plot all selected time series as a function of time. Result time series are plotted as line plots. Input time series can be plotted as scatter plots if the Scatter inputs option is selected. This mode also works together with all other plot options.

4.2.2 Histogram

This option only works if you have exactly one time series selected. it will make a histogram of the time series. The number of bins is determined by Rice's rule

$$k = 2\lceil \sqrt[3]{n} \rceil,$$

where n is the number of data points and k is the number of bins.

4.2.3 Profile

Select one result or input. Moreover, select two or more indexes of exactly one index set that this time series varies over. For a given point in time, a bar plot will be displayed with the selected indexes as the X axis. The point in time can be selected using a slider or a text field. This mode also works together with aggregations, so you can display e.g. yearly mean values (for a selected year).

One use case for this is e.g. to select the yearly sum of the value "Reach nitrate output" in INCA-N to show a bar plot of the yearly output of each reach in a given year.

4.2.4 Compare baseline

This is only available if you have clicked the "Save baseline" button in the toolbar. You must have only one result time series (and optionally one input time series) selected. The plot will display both the current value of the selected time series and the value of the time series at the point you clicked "Save baseline". All plot options are available.

This can be useful for exploring differences in outcome between different parameter sets. For instance, you can see how it affects the stream nitrate concentration if you change the agricultural fertilizer nitrate input in INCA-N.

4.2.5 Residuals

You must have exactly one result time series and one input time series selected. The plots shows the residual time series (observed - modeled). Select "Scatter inputs" to display it as a scatter plot instead of a line plot.

To compare your modeled time series against an observed series, you can load the observed series in as an "additional timeseries", which is explained in the Mobius file format documentation.

A linear regression line of the residuals is also displayed. This shows the trend in the residuals. For instance if this trend goes up, it says that the observed quantity increases over time compared to the modeled one. The regression line is only computed for the GOF interval (see section 4.4).

You can also use aggregation. For instance, the monthly sum of the residuals of something like "Reach flow" versus "Observed discharge" can tell you something about the monthly water balance in a hydrology model.

4.2.6 Residual histogram

You must have exactly one result time series and one input time series selected. The shows a histogram of the residuals. The number of bins are selected using the same rule as for the Histogram option. Moreover, red dots show what the distribution of residuals would look like if it was perfectly normally distributed (with the same mean and standard deviation).

4.2.7 Q-Q

You must have exactly one result time series and one input time series selected. This shows a quantile-quantile plot of the two time series, and can be used to see if your modeled timeseries is roughly similarly distributed to the observed one.

The displayed percentiles are 5, 15, 25, 50, 75, 85, 95. The X axis is the result series, while the Y axis is the input series. The two have the same quantiles if the blue dots are on the red diagonal line.

4.3 Context menu options (edit or save plot)

In addition to what we have implemented in MobiView (described above), the plot has all of the functionality of the ScatterCtrl package from the Ultimate++ framework. You can right click the plot to get a context menu, where you can

   i Select zooming or panning options.

  ii Edit text fields, select colors and plot styles (Properties).

 iii Get a table of the underlying data of the plot (Data). Can e.g. be copied to Excel by selecting multiple cells and ctrl-C.

 iv Copy the plot as an image to the clipboard (Copy image). This only works correctly if you want to paste into certain applications. It works with most image editing software, Microsoft applications (e.g. Word, Outlook, Internet Explorer), but sadly there is a bug right now where it does not work with the Google Chrome browser.

  v Save the image of the plot (Save image). Several formats are available (e.g. png, pdf).

4.4 The GOF interval

4.5 Goodness of fit statistics

The goodness-of-fit statistics are displayed in the time series info box if you have exactly one result series and one input series selected, and have selected one of the Residuals, Residual histogram or Q-Q plot modes.

Most of the goodness-of-fit statistics are implemented following [1]. Further properties of the various statistics are discussed in that paper.

Let $o = \{o_i\}_{i \in I}$ be the observed timeseries, and let $m = \{m_i\}_{i \in I}$ be the modelled timeseries. The set I of comparison points is the set of all timesteps in the GOF interval where both series have a valid value. For instance, the observed timeseries can have missing values, so the timesteps corresponding to the missing values will not be considered when evaluating goodness-of-fit. The GOF interval is the entire model run interval unless something else is specified by the user. Let

$$\overline{m} = \frac{1}{|I|} \sum_{i \in I} m_i$$

denote the mean of a timeseries.

4.5.1 Common data points

The common data points is the size of the set of comparison points I, denoted $|I|$.

4.5.2   Mean error (bias)

The mean error is

$$\overline{o-m} = \bar{o} - \bar{m} = \frac{1}{|I|} \sum_{i \in I} (o_i - m_i)$$

For fluxes or flows, the mean error is related to the discrepancy in mass balance.

4.5.3   MAE

MAE is the mean absolute error

$$\frac{1}{|I|} \sum_{i \in I} |o_i - m_i|,$$

where $|\cdot|$ denotes the absolute value of a number.

4.5.4   RMSE

RMSE is the root mean square error

$$\sqrt{\frac{1}{|I|} \sum_{i \in I} (o_i - m_i)^2}.$$

4.5.5   N-S

N-S is the Nash-Sutcliffe efficiency coefficient [2]

$$1 - \frac{\sum_{i \in I} (o_i - m_i)^2}{\sum_{i \in I} (o_i - \bar{o})^2}.$$

This coefficient takes values in $(-\infty, 1]$, where a value of 1 means a perfect fit, while a value of 0 or less means that the modeled series is a no better fit than a series constantly equal to the mean of the observed series.

4.5.6   log N-S

log N-S is the same as N-S, but where o_i is replaced by $\ln(o_i)$ and m_i replaced by $\ln(m_i)$ for each $i \in I$. Here ln denotes the natural logarithm.

$$1 - \frac{\sum_{i \in I} (\ln(o_i) - \ln(m_i))^2}{\sum_{i \in I} (\ln(o_i) - \overline{\ln(o)})^2}.$$

This coefficient behaves similarly to N-S, but is less sensitive to errors on time steps where both series have large values.

4.5.7   r2

r^2 is the coefficient of determination

$$\left(\frac{\sum_{i \in I} (o_i - \bar{o})(m_i - \bar{m})}{\sqrt{\sum_{i \in I} (o_i - \bar{o})^2} \sqrt{\sum_{i \in I} (m_i - \bar{m})^2}} \right)^2.$$

This coefficient takes values in $[0, 1]$.

4.5.8 Idx. of agr.

The index of agreement is

$$1 - \frac{\sum_{i \in I}(o_i - m_i)^2}{\sum_{i \in I}(|m_i - \overline{o}| + |o_i - \overline{o}|)^2}.$$

4.5.9 Spearman's RCC

Spearman's rank correlation coefficient [3] is computed as follows: For a timeseries $x = x_{i i \in I}$, let $\mathrm{rank}(x_i)$ be the index of x_i (starting from 1) in the list $\mathrm{sort}(x)$, where $\mathrm{sort}(x)$ is x sorted from smallest to largest. The rank correlation coefficient can then be computed as

$$1 - \frac{6 \sum_{i \in I}(\mathrm{rank}(o_i) - \mathrm{rank}(m_i))^2}{|I|(|I|^2 - 1)}.$$

The coefficient takes values in $[-1, 1]$. If the value is 1, the modeled series is a (positively) monotone function of the observed series.

References

[1] P. Krause, D. P. Boyle, and F. Bäse. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5:89–97, 2005.

[2] J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part i - a discussion of principles. *Journal of Hydrology*, 10:282–290, 1970.

[3] C. Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15:72–101, 1904.