BROWSER
A Graphic Tool to Explore Data Relationships

D. McClymont[a], **D.M.Freebairn**[b] and **Y. Abawi**[c]

[a] *National Centre for Engineering in Agriculture, University of Southern Queensland, Toowomba 4350, Queensland, Australia (mcclymon@usq.edu.au)*

[b] *APSRU, Department of Natural Resources and Mines, PO Box 318, Toowoomba 4350, Queensland, Australia (David.Freebairn@dnr.qld.gov.au)*

[c] *Queensland Centre for Climate Applications, Department of Natural Resources and Mines, PO Box 318, Toowoomba 4350, Queensland, Australia (Yahya.Abawi@dnr.qld.gov.au)*

Abstract: Models generate volumes of data, which are typically summarised as tables or static graphs. These summaries give little insight into the interrelationships that may have been built into the models. A simple and flexible graphical means is therefore needed to facilitate efficient exploration of data relationships. BROWSER has been developed as a modelling companion to visually explore time-series data. Data-series can be viewed separately, overlayed or stacked and can be accompanied with simple graphical statistics. The user can "zoom" into particular regions of charts and can "pan" or "scroll" through large data sets using a variety of controls. BROWSER incorporates a database to track model output files that may be stored in many locations on a computer system or network.

Keywords: Modelling; Graphical; Output; Data; Relationships;

1. INTRODUCTION

Models generate volumes of data, which are typically summarised as tables or static graphs. These summaries give little insight into the interrelationships that may have been built into the models. The sheer volume of output means that many users are dependent on specialists to have their "science" correct, with little inspection of relationships and results being generated. Therefore there is a lot of compartmentalisation and trust between specialist disciplines. While this seems a genteel state of affairs, the nature of simulation technology and the way we organise science does not necessarily encourage transparency.

A challenge for all modellers is to sort out what is a new insight and what is a code or conceptual blooper. This is particularly taxing in systems models where many processes are being presented, and there are interactions between processes.

Non-modellers are generally at the mercy of the "magician with the model". We have found that many model users are not aware of quite basic principles of model operation. Credibility is developed through a model providing reasonable explanations of what has been observed in the real world, and occasionally, a new insight is gleaned from model output analyses. This is often also the case for modellers who are not specialists in all disciplines.

A simple and flexible graphical means is therefore needed to support exploration of model outputs. Unfortunately this feature is usually inadequate or missing in most models, many of which are still "DOS-based". It also remains difficult to compare the outputs of multiple models as interfaces seldom contain importing capabilities. This is often achieved by inspecting separate printouts or through importing each output into a spreadsheet or a specialist charting package, which can be a complicated and time consuming process. We decided to create a simple tool to overcome these limitations.

2. DEVELOPMENT OF BROWSER

BROWSER is a tool for visually exploring time-series data (Figure 1) from various model outputs and formats located in different directories on a computer system or network.

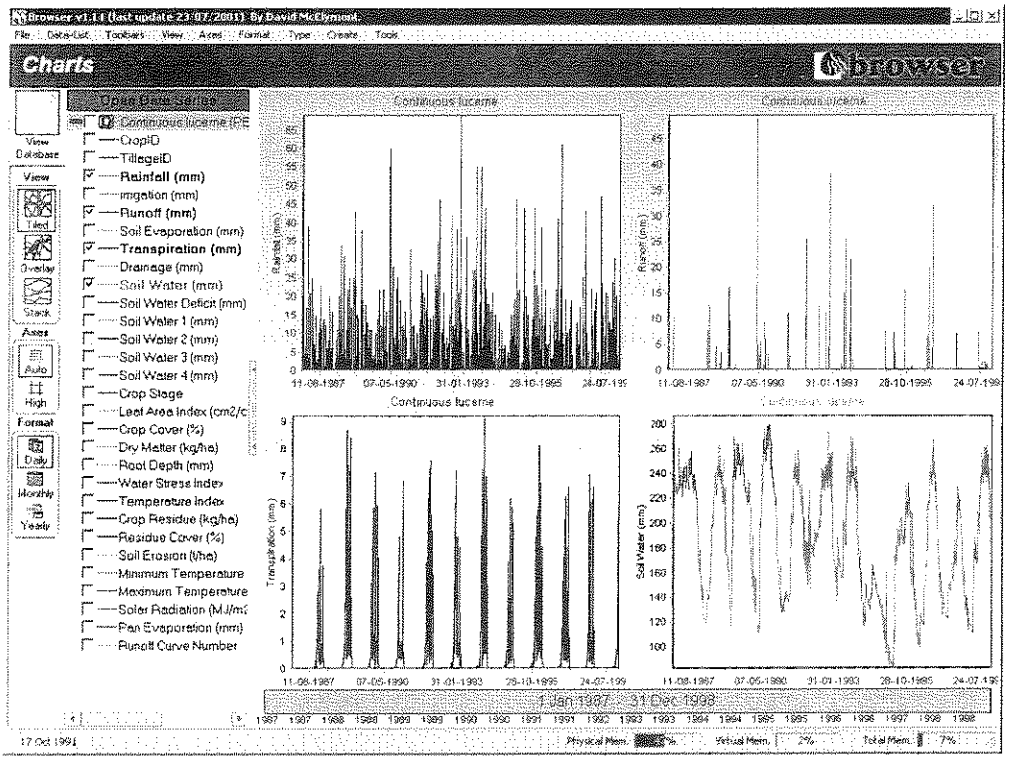

Figure 1. BROWSER interface showing the charting region with "tiled-view" for exploring time-series data.

Development of BROWSER commenced in January 2001 to provide a companion tool for the PERFECT model [Littleboy et al., 1992] whose standard graphical output limited.

The original goals were primarily to interact with the data through scrolling and zooming functions and to overlay data to compare trends. Since then, BROWSER has evolved to incorporate a database, several powerful manipulative tools, summary statistics and limited data transformation capabilities.

BROWSER is currently compatible with fourteen different file types including output from PERFECT, IQQM [DLWC 1998], APSIM [McCown et al. 1996], and SILO [Jeffrey et al., 2001, http://www.dnr.qld.gov.au/silo] models as well as four simple templates. Other formats can easily be accommodated for in future versions.

BROWSER is written in c++ for the Windows operating systems [Win95, 98, NT, 2000 and ME] using Borland's Visual Class Library [http://www.inprise.com/]. Charting is undertaken using Steema's TeeChart component [http://www.teechart.com/].

3. BROWSER INTERFACE

The BROWSER interface is composed of three main components; a file database, an open data-series selector and a charting region. The database and charting components are available as separate views toggled by the user, while the data-series selector is available at all times although it can be hidden.

Commands are available to the user though toolbar buttons, popup menus and through the main application menu. Toolbars are located along the left hand side of the application window and can be hidden.

Output can be printed or exported in "Enhanced Windows Metafile" and "Bitmap" formats.

3.1 Database

BROWSER incorporates a file database to track model output files that may be stored in different locations on a computer system or network.

Creating the database involves defining parent directories, which are automatically searched (including subdirectories) for compatible files.

The user can modify the database search criteria to recognise particular files. These files will be assumed to be of the form of one of the four user templates.

The database can be rebuilt using the "refresh" option to locate any new files or to remove any deleted files in the database directory list. New directories can be added at any stage and particular files or subdirectories can be excluded from the search.

3.2 Data-series Selector

After opening a file from the database, the tabulated data are converted into graphical data-series and are listed in the data-series selector on the left hand side of the application window.

The data are grouped together under the header of their parent file type, or through a common data type (such as rainfall, runoff etc) depending on user preference.

The user is able to show/hide data-series by clicking on its name in the data-series selector. Entire groups of data can be toggled on or off. Editing data-series line properties is achieved by clicking on the line icon next to the series name.

3.3 Chart Output

The power of BROWSER lies in the simple manner in which the user can manipulate the way that the data is presented. Data-series can be viewed separately or overlayed on top or above other series. The user can "zoom" into particular regions of charts and can "pan" or "scroll" through large data-sets using a variety of controls. Time series can be animated under mouse control, thus displaying the dynamic nature of model function. Data-series can automatically be viewed as daily, monthly, and yearly values.

Three main views are presented to the user:

- a "tiled-view" (Figure 1) presents each data-series in its own separate chart;
- an "overlayed-view" (Figure 2) shows several data-series overlayed on a single chart; while
- a "stacked-view" (Figure 3) places each data-series one above another on a single chart.

These views are further enhanced by the ability to locate each data series on its own vertical axis with automatic data range (auto-axes) or to use a single common vertical axis based on the maximum and minimum values of all visible data-series (max-axes).

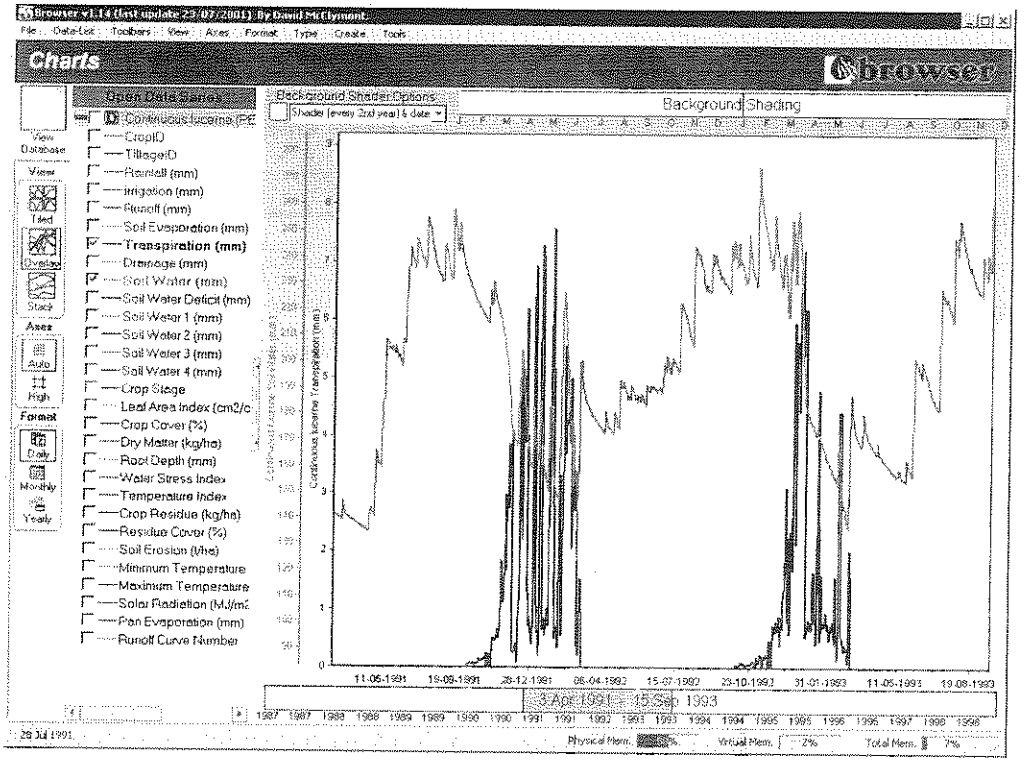

Figure 2. "Overlayed-view" with period shader visible.

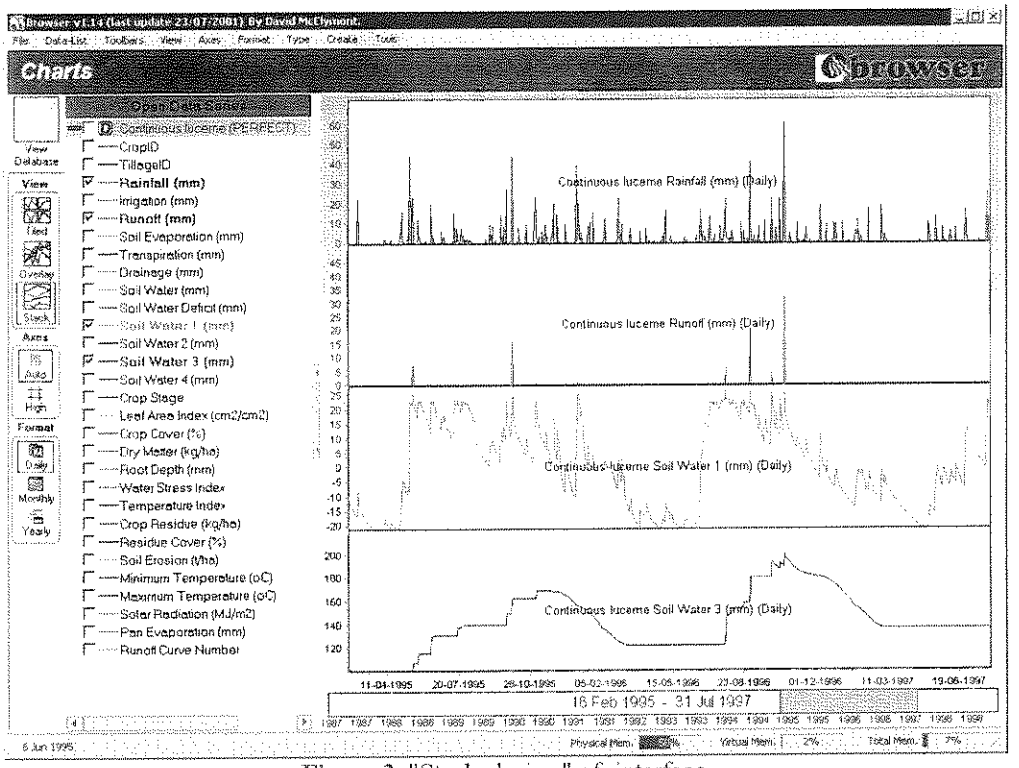

Figure 3. "Stacked-view" of interface.

Zooming into and scrolling (panning) around the charts can be achieved though direct interaction between the mouse cursor and the charts:

- zooming requires tracing a rectangle onto the chart while holding down the left mouse button, and
- scrolling requires moving the mouse cursor around the chart with the right mouse button held down.

The charts are synchronised so that if one chart is "scrolled" or "zoomed", the others are updated accordingly.

A unique feature of BROWSER is the "period-setter" tool (Figure 4) consisting of a timeline-like scrollbar allowing the user to change the viewing period, scroll through the data range or quickly jump to a particular period in the data range. This can represent a "pseudo-animation" of chart data along the timeline. Advanced features of this tool allow precise adjustment of scrolling and zooming parameters.

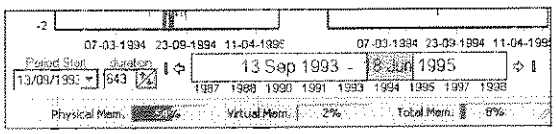

Figure 4. "Period-setter" tool in "advanced-mode"

"Double-clicking" on any chart will temporarily enlarge that chart to the full window area for closer analysis. Selecting any other series from the data-series selector or "double-clicking" the chart again will restore the view back to its previous layout.

Readability of the output is enhanced using a "period-shader" tool (Figure 5) through its ability to highlight areas of the chart background representing annual periods of interest and show the corresponding "year-label". For example, the August to November period could be shown for each year, or a twelve-month period could be highlighted for alternate years.

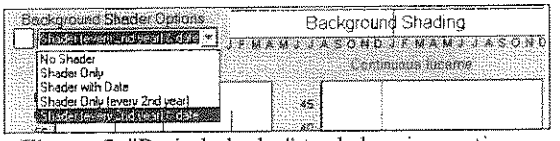

Figure 5. "Period-shader" tool showing options.

Summary statistics of the time-series data including maximum, minimum, median, mean, 25th percentile and 75th percentile can be presented as a line series or a box plot (Figure 6), and can be superimposed behind its corresponding data-series (Figure 8).

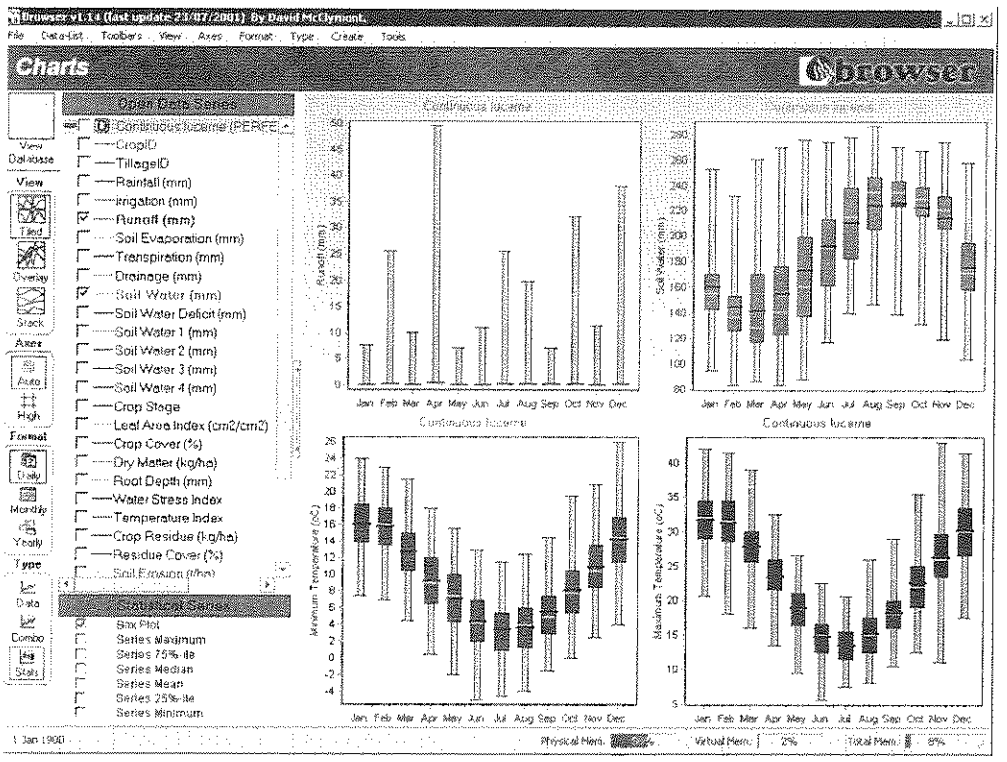

Figure 6. Summary statistics output (box-plot option) in "Tiled-view" layout.

Finally, BROWSER has three options available to create new data-series from existing series:

- the "Combine Series" option allows addition/subtraction of multiple data series,
- the "Scatter Plot" (Figure 7) option creates a separate scatter plot from two data-series, and
- the "Create Cumulative" option converts single data-series into a new cumulative plot.

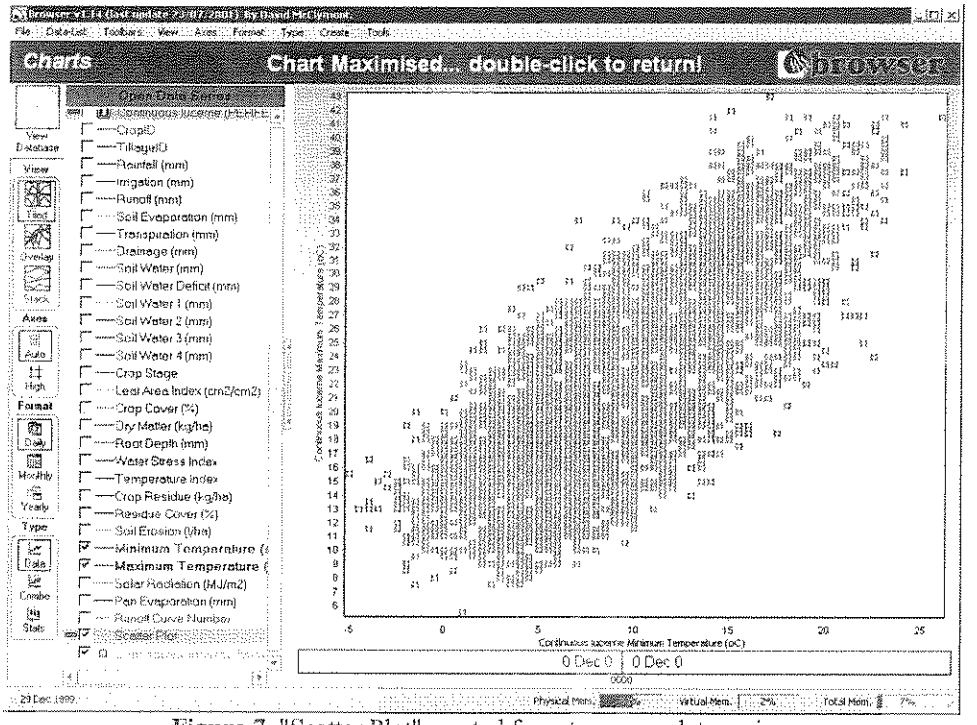

Figure 7. "Scatter Plot" created from two open data-series.

4. USING BROWSER

The range of predefined chart options in BROWSER allows several different tasks to be undertaken.

The "tiled-view" allows quick visualisation of each data series to locate obvious abnormalities or features and aids in selecting relevant data series for further investigation.

The "overlay-view" is useful for comparing both trends and magnitudes between data series. Trends are most easily analysed using the "auto axes" option so that each series has its own separate vertical axis while remaining in the same vertical screen space. Using the "high axes" option allows direct comparison of the magnitude of data using a single vertical axis. However, one disadvantage of this view is that the chart region tends to get a little "cluttered" when too many series are selected, especially when the "auto axes" option is used.

The "stacked-view" alleviates much of the "clutter" of the "overlay-view" while still allowing a clear comparison of trends. Comparing magnitudes is less clear however as each series is now in its own vertical screen space.

The ability to superimpose a data-series over its statistical summaries (Figure 8) provides an efficient method for identifying abnormalities or outliers in an individual data-series. Combined with the advanced navigation features of BROWSER, the user can quickly scan through a time-series investigating how local events compare to the historical measures.

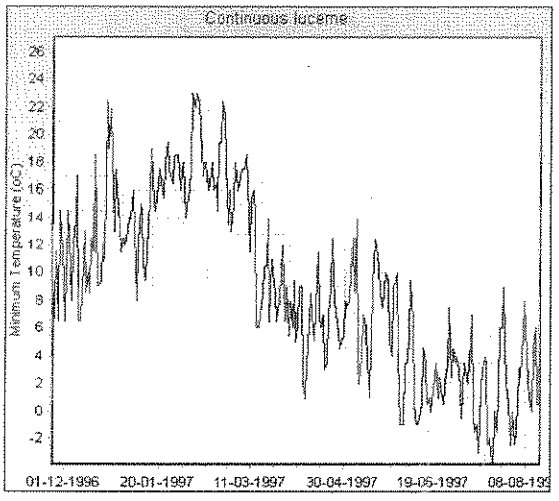

Figure 8. Close-up of a data-series superimposed over a monthly-box plot statistic.

5. CONCLUSIONS

BROWSER is a tool for visually exploring time series data in a simple and efficient manner. While BROWSER does not do anything that can't already be done with an "off-the-shelf" package, the efficiency with which data can be explored, and relationships examined makes the task more attractive. Pseudo animation brings data alive to novices and specialists alike.

It is planned that BROWSER will be readily available as a gesture to encourage enlightenment and transparency in the modelling world.

6. REFERENCES

Littleboy, M., D.M. Silburn, D.M. Freebairn, D.R. Woodruff, G.L. Hammer and J.K Leslie. Impact of soil erosion on production in cropping systems. I. Development and validation of a simulation model. *Australian Journal of Soil Research* 30, 757-74, 1992.

NSW Department of Land and Water Conservation, Integrated Quantity Quality Model Reference Manual, 1998.

McCown, R.L., G.L. Hammer, J.N.G. Hargreaves, D. Holzworth, and D.M. Freebairn, APSIM: A novel software system for model development, model testing, and simulation in agricultural systems research. *Agricultural Systems* 50:255-71, 1996.

Jeffrey, S.J., J.O. Carter, K.B. Moodie, A.R. Beswick, Using spatial interpolation to construct a comprehensive archive of Australian climate data", Vol 16/4, pp 309-330, 2001 *Environmental Modelling and Software.* (in press)