PaleoPlot: A Tool for the Analysis, Integration and Manipulation of Time-series Paleorecords

R. Theron[1], J. A. Flores[2], F. J. Sierro[2], M. Vaquero[1] and F. Barbero[1]
[1]Departamento de Informática y Automática
[2]Departamento de Geología
Universidad de Salamanca, 37008 Salamanca, Spain

***Abstract*—In order to monitor and make predictions, palaeoceanographers need to manipulate, integrate and analyse paleo time-series that are obtained from a number of independent techniques and instruments which, moreover, are usually produced by different researches and/or laboratories. The result is a collection of datasets with unequal age precision. This situation worsens since, at present, there is not an assembly of integrated tools that permits the palaeontologists to perform a study without having to use applications of different natures (spreadsheets, image editors, time-series analysers, etc.) which are of general purpose and, usually, expensive and platform-dependent. Hence, the goal of this project was to develop a platform-independent full featured tool that could simplify the palaeocenographers work, being used during the whole process: from the beginning, merging and correlating different datasets, continuing with interactive analyses, to the end, plotting the results. PaleoPlot is a general tool which could be used for any kind of time series, although it is optimised for the use in the stratigraphy area. Finally, it must be said that PaleoPlot is currently been used as an educational tool on palaeontology courses at the University of Salamanca.**

I. Introduction

As a symptom of evolution, human beings are always trying to better understand current and past events so we are able to foresee future situations and develop new tools or methods. The data registered over millions of years is an impressive source of information that, for instance, help us to model earth dynamics or to make climatic predictions.

We find the largest and oldest record in the oceans. Palaeoceanographers, in order to monitor and make predictions, need to manipulate, integrate and analyse paleo time-series that are obtained from a number of independent techniques and instruments (such as ocean drilling, ocean tracers, AMC 14C datings, astronomic curves, etc.), which, moreover, are usually produced by different researchers and/or laboratories. The result is a collection of datasets with unequal age precision.

In the last years, several tools have been developed to allow best management and analysis of these data. The main available tools are represented by databases (Dbase, Access, Foxpro, Oracle, etc.), spreadsheets (Lotus, Excel, etc.), statistical packages (Axum, Data Desk, S-Plus, SPSS) and graphical and/or image tools (Statgraphics, CorelDraw, and many more).This worsens the situation since, at present, there is not an assembly of integrated tools that allows the palaeontologists to perform a study without having to use applications of different natures, which are of general purpose and, usually, expensive and platform-dependent.

Some freeware tools have appeared in the last few years that were designed for the field of paleoclimatology (AnalySeries [1]) or paleontology (PAST, [2]), which include many commonly used functions, but they are rather limited to a one-by-one analysis, are one-platform-only and the results usually need to be exported to other tools for further analysis or make them more legible.

The goal of this project was to develop a tailor-made, platform-independent, full featured tool that could simplify the palaeocenographers work, being used during the whole process: from the beginning, merging and correlating different datasets, continuing with interactive analyses, to the end, plotting the results.

II. Design Concepts

For the design of a package dedicated to the integration, manipulation and analysis of paleorecords, the following aspects were considered to be most important:

- **Platform-independent software**, since the tool will be used by users with different operating systems.
- **Easily scalable architecture**, providing a way to introduce the feedback of users and to incorporate advanced techniques.
- **Friendly and easy-to-use interface**; some specialists are not proficient using computers and get confused when dealing with file formats or configuration files.
- **Fully-customizable environment**, to prevent the need of additional tools to change number formats or graph colors, for instance.
- **Efficient management of data**, such as missing and incongruent data correction, and realistic data plotting, that is, to avoid misleading representations.

All the previous requirements led to chose Java [3] as programming language.

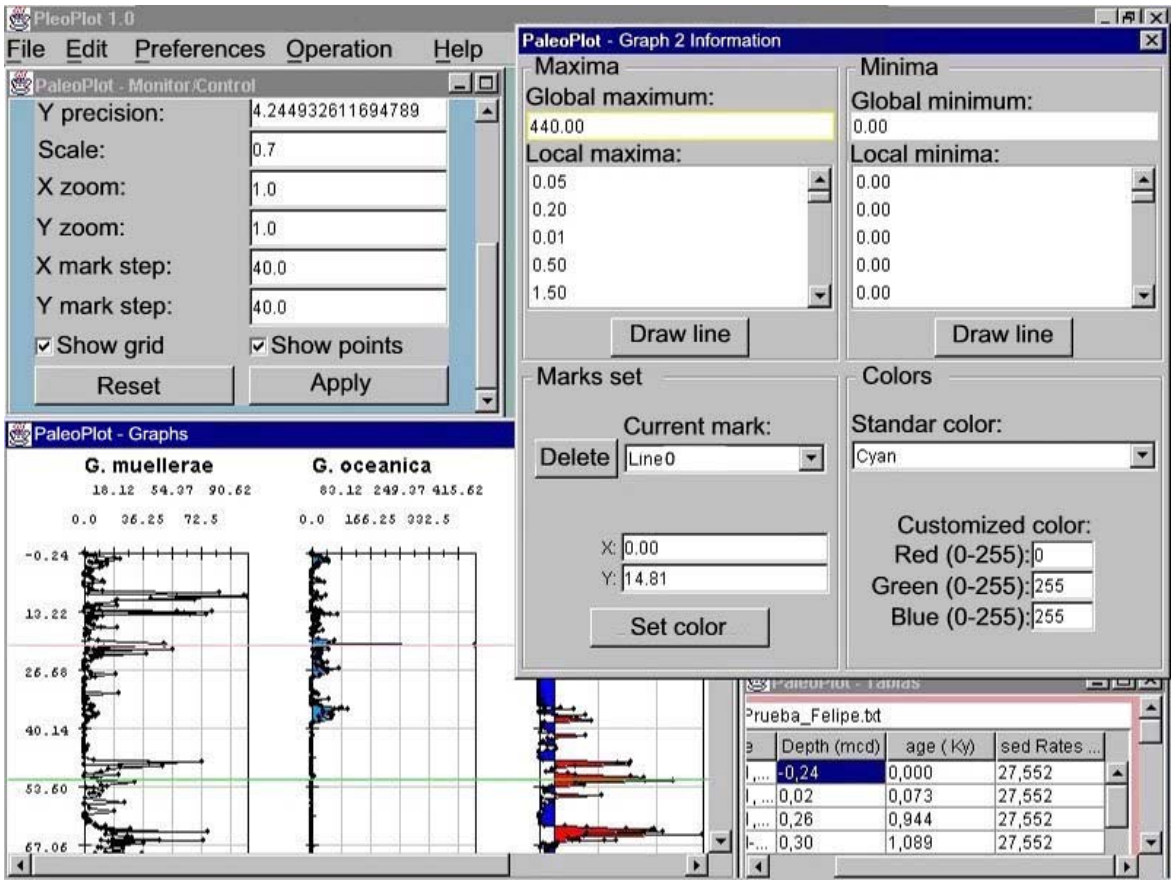

Fig. 1. PaleoPlot user interface

It was established that the tool should include three main parts:

- **Table integration and manipulation**: creation, load, store, edition, data consistency control, merging, interpolation, age-model generation, etc. of tables containing one or several time-series.
- **Customized Graphics Plotting:** features the possibility of plotting any number of time-series in side-by-side tracks in a single chart, including x-y plots, filled polygons, histograms, marked data points, and so forth.
- **Interactive Analysis:** global/local maximum and minimum evaluation of each series, trend analysis, correlation analysis trough the use of *Event marks* (lines marking geological events in one series related to the others with real data values available upon mouse pointing), etc. All the analysis may be performed on the whole dataset or within a selected interval.

In the next section these parts are shown and explained as they are implemented in PaleoPlot 1.0.

III. Implementation and Results

Fig. 1 shows PaleoPlot user interface in a typical analysis stage, where the main three parts stated above are present.

All options are available from the menu bar, at the top of the main window. Inside this area, a window (upper left corner), called *Control/Monitor*, is always present, showing messages of events that may occur and permits the configuration of several options. The rest of the tool is formed by a number of windows and dialogs depending on the stage of operation and the analysis performed.

Among them, two windows (*Graphs*, Fig. 2, and *Tables*, Fig. 3) and two dialogs (*Graph information* and *Merge*), are considered here.

Both windows are created inside the main window and are arranged inside this area, resizing its contents and controlled trough individual sliders. The dialog boxes are independent of the main window and can be placed anywhere on the screen.

Next, the main features of these windows and dialog boxes, and a typical case of time-series analysis are outlined.

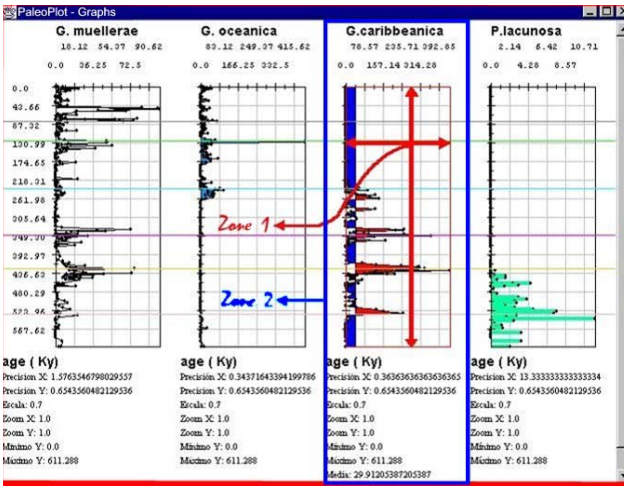

Fig. 2. Graphic time-series analysis window

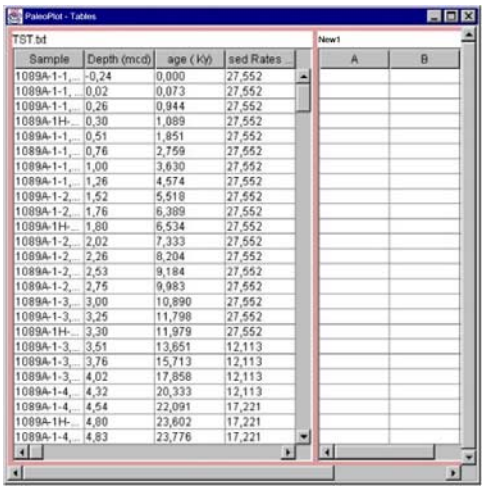

Fig. 3 Time-series tables window

The *tables* window (Fig. 3), placed anywhere inside the main one (Fig. 1, right-bottom corner), contains all opened tables; this may be done opening a file, creating a new table or by means of the *Merging* tool, which generates a new table with columns from two or more existing tables, adapting the precision (of a column present in all source tables, typically age or depth) to one selected by the user. Full editing, copy/paste and drag-and-drop (including external, as from/to MsExcel) operation are allowed.

The *graphs* window (Fig. 2, Fig. 1, left-bottom corner), is the container of all the columns selected for plotting. All features are customizable. This windowmay be used later just for printing or for further study through interactive analysis.

Fig. 2 shows two active zones in the *graphs* window; zone 1, each column graph box, opens the *graphs information* dialog upon mouse clicking (Fig. 1, right-top corner), which permits the selection of important local points to be highlighted with a line in the rest of the graphs; and zone 2, providing statistical information of the graph.

PaleoPlot features the *Event mark;* by clicking anywhere in the *graphs* window, an horizontal line is drawn, across all the graphs, from the nearest real point of the graph. This method allows correlating geological events, without misleading due to graphic interpolation when plotting the data.

A paleo time-series analysis will typically consist of a process of 3 to 5 steps: 1) open one or more files containing time-series tables or create and edit new series; 2) Setting the desired parameters related to each series; 3) Optionally, generate a new table merging several columns of different tables; 4) Plotting the type of graph for each variable of the study; and 5)Optionally, interactive analysis.

PaleoPlot has already been used for integration of time-series originated by different techniques and as a preliminary analyzing tool by our research group, in advanced works [4].

IV. Conclusion and Future Work

This paper presents PaleoPlot, a general tool which could be used for any kind of time series, although it is optimised for the stratigraphy area and specially meant for paleoclimatology. It has been validated being the only tool used to perform typical paleoclimatic studies.

Finally, it must be said that PaleoPlot 1.0 is currently been used as an educational tool on palaeontology courses at the University of Salamanca, while Paleoplot 2.0 is under development incorporating advanced spectral analyses, data mining techniques (decision trees, clustering, etc) and advanced 3D plots.

References

[1] D. Paillard, L. Labeyrie and P. Yiou, "Macintosh program performs time-series analysis," in *Eo, Transactions*, American Geophysical Union, vol. 77, 379, 1996.

[2] Øyvind Hammer, D. A. T. Harper and P. D. Ryan, "PAST: Paleontogical statistics software package for education and data analysis," in *Palaeontologica electronica*, vol. 4, 1, 2001.

[3] James Gosling and Henry McGilton, "The Java language environment: A white paper," Technical report, Sun Microsystems, 1996.

[4] R. Theron, J. A. Flores, F. J. Sierro, C. Pelejero, J. Grimalt and M. Vaquero, "Using data mining and visualization techniques for the reconstruction of ocean paleodynamics," in proceedings of the *IEEE International Geoscience and Remote Sensing Symposium*, 2002, in press.