Research paper

Animated analysis of geoscientific datasets: An interactive graphical application

Peter Morse [a,*], Anya Reading [a], Christopher Lueg [b]

[a] School of Physical Sciences (Earth Sciences), University of Tasmania, Hobart, TAS, 7001, Australia
[b] School of Engineering and ICT, University of Tasmania, Hobart, TAS, 7001, Australia

A B S T R A C T

Geoscientists are required to analyze and draw conclusions from increasingly large volumes of data. There is a need to recognise and characterise features and changing patterns of Earth observables within such large datasets. It is also necessary to identify significant subsets of the data for more detailed analysis.

We present an innovative, interactive software tool and workflow to visualise, characterise, sample and tag large geoscientific datasets from both local and cloud-based repositories. It uses an animated interface and human-computer interaction to utilise the capacity of human expert observers to identify features via enhanced visual analytics. 'Tagger' enables users to analyze datasets that are too large in volume to be drawn legibly on a reasonable number of single static plots. Users interact with the moving graphical display, tagging data ranges of interest for subsequent attention. The tool provides a rapid pre-pass process using fast GPU-based OpenGL graphics and data-handling and is coded in the Quartz Composer visual programing language (VPL) on Mac OSX. It makes use of interoperable data formats, and cloud-based (or local) data storage and compute.

In a case study, Tagger was used to characterise a decade (2000–2009) of data recorded by the Cape Sorell Waverider Buoy, located approximately 10 km off the west coast of Tasmania, Australia. These data serve as a proxy for the understanding of Southern Ocean storminess, which has both local and global implications. This example shows use of the tool to identify and characterise 4 different types of storm and non-storm events during this time. Events characterised in this way are compared with conventional analysis, noting advantages and limitations of data analysis using animation and human interaction. Tagger provides a new ability to make use of humans as feature detectors in computer-based analysis of large-volume geosciences and other data.

1. Introduction

1.1. Geosciences data

The volume of data generated by scientific instruments, sensor systems and computational models is growing at an increasing pace (Hey et al., 2009) and most geosciences disciplines and technologies exhibit this trend (Sellars et al., 2013). Typical datasets might comprise time variant observations at fixed geographic locations such as tide-gauges and oceanographic buoy data, spatially distributed data such as soil geochemistry analyses, or combined spatial/time variant data such as those observed by satellites. Other large datasets include outputs of model simulations and forecasts. Interpretations of geosciences data are commonly carried out using graphs and maps, however, there is a necessary limitation on how much information can be presented in a single representation and how many graphs or maps can be usefully incorporated in one study (Ware, 2012; Munzner, 2014). Statistical analyses afford the ability to summarise in a systematic way but at the expense of exploratory analysis and much pattern characterization.

Graphical and other diagrammatic representations of data can be thought of as visualisations (Tufte, 1990, 1997; Few, 2009). Existing visualisation tools are generally used for the presentation of results rather than being an inherent part of the data inference process (Ware, 2012; Few, 2015; Victor, 2005). However, interactive visual analytics could be used much more widely at earlier stages in the workflow, prospectively generating new insight into the underlying data (Thomas and Cook, 2005; Keim et al., 2008; Ward et al., 2010).

* Corresponding author.
  E-mail address: peter.morse@utas.edu.au (P. Morse).

1.2. Human interaction with data displays

The human comprehension of graphical representations is complex, as it involves the interplay between physiology, visual perception, cognition and experience, together with factors such as information density, dataset size, graphic design and color (Ware, 2012; Healey and Enns, 2012). Some flexibility in the display allows for physiological and experiential differences in scientific data analysts. Expert human analysts can identify structures and patterns that are challenging for statistical classification systems. Both comprehension and pattern detection are influenced by the graphical abstraction of data and the levels of detail present (Carpenter and Shah, 1998; Friel et al., 2001). Taken together, there is a point where visual data density impairs comprehensibility (Walker et al., 2016), as illustrated by the two static plots (Fig. 1).

1.3. Static and animated interactive representations of data

Geoscientists and other data scientists routinely use software packages to visualise and present results as static graphs and charts. This graphing and underpinning 'spreadsheet' document model is user-friendly, robust and has been in use for many decades (Campbell-Kelly, 2003; ch.12) with more recent capabilities that enable dynamic updating of variables. Whilst adequate for many scientific purposes, this approach becomes unwieldy and can lead to missed insights, as datasets increase in size and complexity. As an alternative, many users engaged in scientific research are comfortable creating visual output using programming models and scripting (e.g. Python, R, Matlab) as datasets become larger.

Animated interactive representations of data enable more sophisticated visualisations for interpretation. With the advent of high performance computer graphics, the repertoire of graph types has significantly expanded, including adaptive and network representations (Herman et al., 2000; Liu et al., 2014). Programming languages widely used in science such as Python are also moving beyond the conventional static charts and visualisations produced by MatPlotLib (Hunter, 2007), with an array of tools that leverage programming languages in concert with vector graphics libraries (Khronos Group, 2016), e.g. Bokeh (Continuum Analytics, 2015). Alternatively, using a programming language together with OpenGL enables the processing of significantly greater data volumes (e.g. VTK and Paraview (Kitware, 2016); Vispy (VisPy Developers, 2015)). These softwares combine, in varying degrees, features for interaction and animation, constrained by their underlying architecture (Kloss, 2009). 2D plotting libraries predominantly execute upon the CPU, limiting their ability to handle very large datasets and real-time interactivity. In contrast, OpenGL-based plotting systems (Supplement 1) afford the possibility of handling millions to billions of datapoints as well as handling high-frame-rate animation and interactivity, via the transaction of events and data across CPU and GPU (Rossant and Harris, 2013).

1.4. GPU-enhanced scientific-oriented software

Scientific visualisation has been able to exploit these hardware developments for the analysis and visualisation of massive datasets (Guha et al., 2005; Cuntz et al., 2007; Fogal et al., 2010). Examples of previous studies in the geosciences include the development of sophisticated computer simulations, such as weather forecasting (Rautenhaus et al.,
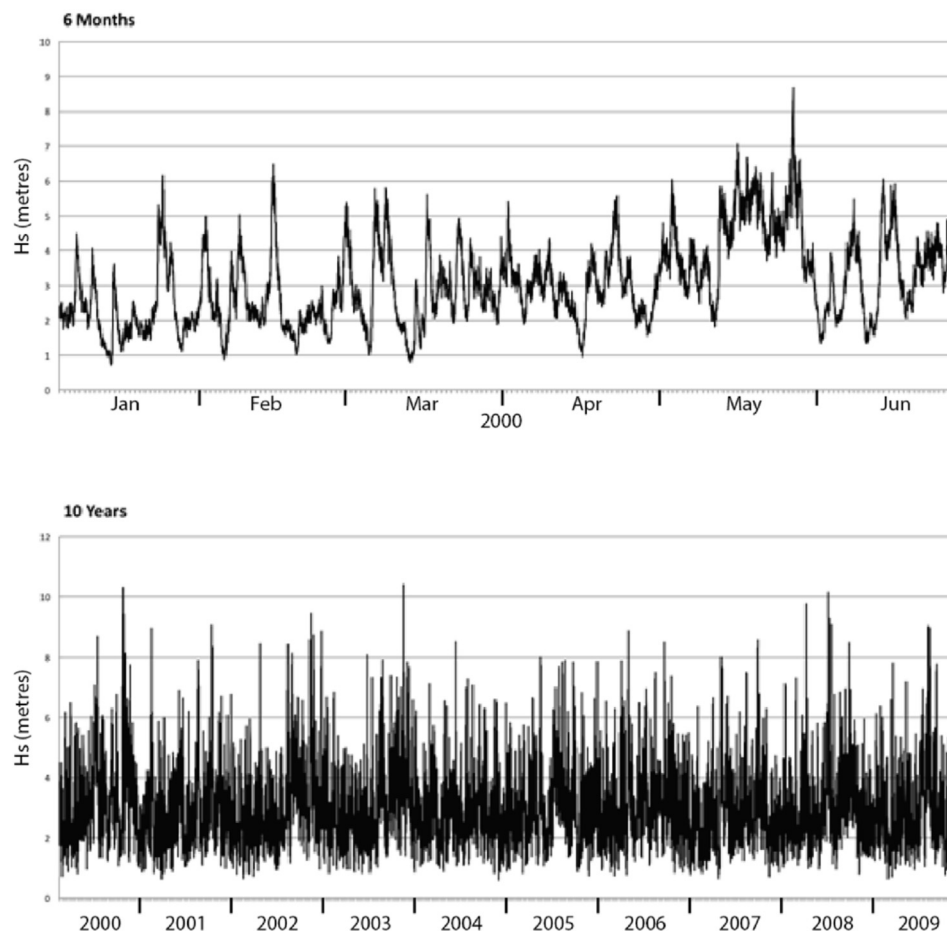

Fig. 1. Significant wave height (Hs) over time, recorded at the Cape Sorell Waverider Buoy, West of Tasmania, Australia: Upper plot: 6 months (1/1/2000-30/6/2000) (8583 Samples), Lower plot 10 years (2000–2009) (165,484 Samples).