ISSN: 2321-8819 (Online) 2348-7186 (Print) Impact Factor: 1.498 Vol. 6, Issue 3, March, 2018

Geospatial Data Pre-processing Packages: Contribution towards an Enhanced Open-source Package

Ahmed Babalaji Ndanusa[1], Zulkhairi B. Md Dahalin[2], Azman Ta'a[3]
[1,2,3] (Dept. of Information Technology, College of Arts and sciences, Universiti Utara Malaysia, Malaysia).
Corresponding Author: Ahmed Babalaji Ndanusa: elahmedn@gmail.com

Abstract: Geographical Information Systems (GIS) packages have been generating immense interest globally. Their relative rate of usability, consistency, functionalities, efficiency and cost-effectiveness have been the determining factors in the selection of packages for any GIS projects. This paper attempts to proffer a comparative assessment of some of the commonly used GIS packages with the aim of proposing the most reliable package based on these factors. By this systematic assessment, both current and potential users will be able to take full advantage of the most efficient GIS package to perform various analytical pre-processing tasks. The outcome of the assessment could be adopted as a guide for selecting an appropriate and reliable open source GIS platform for a timely and efficient pre-preprocessing geospatial data for environmental analysis.

Keywords – *Data Analysis, Geospatial Data Pre-processing, GIS Software*

Introduction

The choice of a Geographic Information System (GIS) tool remains quintessential in the success of any GIS-related projects. Broadly, GIS is a reliable sub-field of Information System with great records of contributions in every aspect of environmental analysis. Currently, GIS packages offer an integrated approach in using spatio-temporal data (Eldrandaly & Naguib, 2013). The emergence of new GIS software packages is largely attributed to the efforts made in providing more GIS software in recent years, which has in turn, led to a corresponding decline in the prices for these packages.

Developing a new GIS project is very vital in carrying out any geospatially-based projects. However, selecting a suitable software package remains a crucial determinant of the success or failure of a project, since any decisions wrongly taken is selecting the package can correspondingly lead to an erroneous result. Additionally, one of the primary functions of GIS is the pre-processing of geospatial data needed for spatial analysis or disaster monitoring/reporting within our environments.

Nevertheless, various authors have displayed their concern on the inherent limitations of analytical capabilities in some GIS packages, such as ArcGIS which the commonly used package. It has been observed that the majority of GIS experts place more emphasis on data input and retrieval than on the analytical capabilities (Neteler, Bowman, Landa, & Metz, 2012). As a result, there is a growing consensus amongst GIS experts on the need to endow GIS packages with efficient spatial analytical functionalities (Fotheringham & Rogerson, 2013). The call for the need to review the existing packages to identify the most suitable one(s) is dependent on the economical and performance standards (Eldrandaly & Naguib, 2013).

Generally, GIS packages proffer an integrated method in utilizing spatio-temporal data sets. The vendors of these packages consider some basic computational specifications, such as privacy, file management system and visualization (Eldrandaly & Naguib, 2013). Therefore, these packages remain an integral aspect of any efficient geographic information system, as the system utilized within a GIS task possesses a determining impact on the type of projects as well as the results that are obtainable.

As discussed thus far, the following four facets will essentially be used in assessing the most efficient GIS package required to proffer enhanced results in GIS projects: Functionality, Usability, Cost and Reliability (Clarke, 1997; Bernhardsen, 2002; Ngai, & Chan, 2005).

Functionality

The functionality factor refers to the features integrated in the GIS packages which are required to meet GIS projects specifications (Keil & Tiwana, 2006). A GIS is defined by its capabilities and not what it is. A systematic review of GIS functionalities is a critical phase in the selection of GIS package, as the inability of GIS package to suitably address environmental issues will mean a failure of the GIS package. The intrinsic challenges in pre-processing geospatial data has made the reliance on GIS tools an indispensable factor in environmental decision-making. Given the voluminous and complex nature of data employed, pre-processing such sets of data requires computation supports.

For a variation of analytical tasks, the required GIS tools do not need to be developed from the scratch but can be chosen amongst the existing ones based on the integrated functionalities (Neteler et al, 2012). A GIS generally integrates a collection of computer-based functionalities that proffers features for inputting, storing and retrieving as well as visualizing geospatial data sets needed for decision-making (Aronoff, 1989). Consequently, functionality is highly considered during the choice of GIS software (Neteler et al, 2012). However, based on the reviewed literatures, functionalities such as operating systems, spatio-temporal data management, GIS data conversion features, spatio-temporal data entry and editing, Attribute data manipulation, Mapping features, Terrain and 3-D features and Raster data processing, have been identified to play a measure role in the selection or choice of GIS package (Maguire, Rhind, Goodchild, & Longley, 2010). Which are the fundamental features that were considered before a solution was proposed in this paper.

Cost

The need and importance of any GIS project determines the likely budget than can be expended on the acquisition of such tools. However, cost can equally influence the choice of GIS tools to be adopted for projects (Maguire, et al, 2010). This factor involves the expenditures related to GIS package, license as well as the maintenance (Ngai & Chan, 2005). It is worth to note that, there are several packages with price difference and also, varied functionalities. However, those with high cost do not necessarily depict their level of efficiency compared to those with low or no cost implications. In fact, this perception induced the need to focus more on an open-source package in this paper.

User-friendliness or usability

The aspects of user-friendliness or usability represents the satisfaction and efficiency which are obtained during the conduct of some tasks within a given environmental point of view (Tiits, 2003). The facet of user-friendliness which is also referred to as usability engineering, is identified as either the ability to understand, to learn and to operate the system conveniently as defined by ISO/IEC 9126-1991 (Tiits, 2003). Fundamentally, GIS packages are intricate to explore. To manipulate them, the knowledge on cartography, databases and statistics, and programming skills are also required. These facets make operating GIS inadvertently tasking. However, the GUI of the software packages that users are forced to use daily are making the manipulation worse. Confusing interfaces, complex operations and extensive processes are all very familiar. But should this be the standard? Can't there be a GIS that is user-friendly to explore and allows the users to be productive? These questions are the topics of usability studies and Usability Engineering, which this paper also considered before the choice of GIS package was proposed well as indicating the real impacts of the lack of a user-friendly GIS package. Most of these inherent complexities are attributed to the inconsistency of the interfaces which are difficult to explore. With virtually every new version with vendors claiming a corresponding enhancement on the interfaces to attract users, even when the enhancements do not actually facilitate its usability. As a result, this paper reviewed the interfaces of ArcGIS before the proposed package (Haklay & Jones, 2008).

By considering the usability features, this paper aims to demonstrate that lack of usability leads to a measurable decrease in productivity and efficiency. Hence, the need to create awareness on usability facets in GIS packages in ensuring the following criteria are met (Haklay & Jones, 2008):

- Effective – enabling the users to accomplish a defined task accurately and entirely.
- Efficient – attaining the specific task correctly with as little work and time as possible.
- Error tolerant – in any system, it is anticipated that the users could make mistakes. The system must recognize these mistakes and allow the users to recover from mistakes they have made (as in the case of "undo" or "Redo").
- Learnable – the system should aid the users to learn its functionalities, as well as learning more powerful options as they develop their knowledge about it. It should also be sufficiently for an infrequent user to work with the system easily without extensive retraining.
- Satisfying – the work with the GIS package should preferably be enjoyable, or at least pleasant and satisfying to explore.

Consistency

The IEEE 610.12-1990 defined consistency as the ability to execute a required task under a standard situation during a defined period of time (Keil & Tiwana, 2006). Generally, consistency of a GIS software in an essential feature which determines the overall performance and reliability of the system used in the geographic information system. Currently, some GIS tools, such as ArcGIS is widely being used by over 81.33 percent of the domain experts (Merry, Bettinger, Grebner, Boston, & Siry, 2016). Evidently, the ArcGIS environment has greatly contributed to GIS-based research. Data scientists as well as environmentalists rely on this application for varied forms of analysis involving the use of spatio-temporal data sets. Despite the continuous reliance on this as well as other similar packages, they equally exude some inherent limitations as presented in the ensuing section.

I COMPARATIVE REVIEW

As earlier mentioned in the previous section, the choice for suitable tools needed for large volume of data remains a challenge. Especially, in the aspect of environmental data analysis (Singh & Fiorentino, 2013). Therefore, the ensuing section will present a concise review of some of the commonly used GIS tools with the aim of identifying their prospects as well as their limitations.

The review of these tools is as follows:

I. ArcGIS

ArcGIS is a licensed tool that allows the creation of maps and other geographical features developed by the world GIS leader, Environmental Systems Research Institute (ESRI). The package has several versions by which data scientists or environmentalists can manipulate maps and spatio-temporal data using its integrated API, JavaScript, Silverlight as well as Sharepoint (Ravi, Dascalu & Harris, 2013).

II. Integrated Data Viewer

IDV is an open source based on Java programming to visualization and analysis of spatio-temporal data (Ravi, Dascalu & Harris, 2013; Ramamurthy & Ho, 2014). The interface is customized to process other functions in domains other than geoscience applications. It also provides labels required for longitude and latitude axes in addition to its ability to display heterogenous data simultaneously.

III GrADS

(GrADS) which is an acronym for Grid Analysis and Display System, is an open source tool used for Geo-science data in a multi-dimensional domain (Ravi, Dascalu & Harris, 2013). It uses command line similar to FORTRAN to process tasks and proffer extensive functions. *Additionally, functions can be added by user in form of PostScripts.*

IV. UV-CDAT

UV-CDAT is an acronym for Ultrascale Visualization Climate Data Analysis Tools. This an open source python set of tools that supports climate data processing (Ravi, Dascalu & Harris, 2013). Several climate data functions such as standard deviations, linear regression and mean are provided.

The following table shows list of various tools commonly used, their strengths as well as their limitations

S/N	Package/Software	Strengths	Limitations
1	ArcGIS	-User-friendly	-Expensive Additional cost required for upgrades - Clipping of images is time consuming and often yields erroneous results [10].
2	Integrated Data Viewer	Scalable with big data -Support heterogenous data sets -Proffers 3D for high quality visualizations. -It can plot remotely	-Its dependence on virtual memory sometimes causes a system to be slow. - It requires a lot of RAM which can make it unsuitable for database systems
3	GrADS	Supports numerous heterogenous data	Graphics have to be drawn by users. This makes it less intuitive and less user-friendly.
5	UV-CDAT	Open source	-Tendency of bugs.

Despite the fact that all the considered criteria cannot be measured unambiguously, they provide a set of principle that can then be translated into specific measurements and expectations to guide in the selection process of the proposed GIS package as presented in the ensuing section.

II PROPOSED SOLUTION

The GIS sector has witnessed an unprecedented growth in the past decade (Neteler et al, 2012). Specifically, Open Source-based GIS packages keep gaining relevance within academia, business, as well as in the administrative domains (Neteler, Bowman, Landa, & Metz, 2012). The idea of Open Source package may be as old as software development itself since code emanating from universities and government laboratories have often been made available in the public domain. In the 1990s, a series of projects on Open Source GIS tools for both desktop and server systems was established in various GIS sectors, including software libraries for map reprojection and data format conversion, desktop GIS, Web mapping/Web GIS, spatial SQL databases, geo-statistics, and metadata catalogues. In the field of environmental analysis, GIS technologies have fundamentally been adopted in an early stage to capture and analyze spatial features. With a considerable level of success recorded. However, even in the face of this success, the packages presented some limitations as highlighted in the previous section. Therefore, to address these limitations, this paper proposes the adoption of QGIS.

QGIS packages have been under development for almost three decades, with a strong link to academia in environmental-related analysis. Its development is generally community-based with global developer network who collaborate with the aid of an online source code repository, mailing lists and a Wiki, users and developers or experts exchange ideas with the view to ameliorating the existing features and functionalities. Presently, researches utilizing geospatial data are continuously being conducted in order to address the effects of climate change to environments as well as individuals within these environments. Even though these sets of data are relatively becoming more accessible, nevertheless, analyzing these complex data is tasking and could be erroneous in the absence of a suitable visualization tool which will yield a graphical representation of data sets to obtain a more comprehensible meaning of the features (Neteler, Bowman, Landa, & Metz, 2012).

Finally, Quantum Geographic Information System (QGIS) is a free and open source GIS package which also a multi-platform that allows visualization, pre-processing as well as analysis of geospatial datasets in addition to composing and exporting the analytical results. QGIS several geospatial data sets such as raster, vector, dx MapInfo and PostGIS. The output can be viewed in 2-D as well as 3-D. QGIS integrates GRASS and SAGA for geospatial data analysis with the aid of plugins.

QGIS is a high-quality cutting edge in GIS with virtually unparalleled depth of proffering pre-processing and visualization features within the main package. The basis of Open Source architecture of QGIS grants long term accessibility as the source code is provided with no financial implications involved. The acceptance of QGIS is for its quality algorithms community contribution and ease of customization**.** In addition, the continuous updates on the plugins makes it suitably efficient in addressing the limitations currently experienced by the existing GIS packages.

III Conclusion

This paper highlighted various limitations of the commonly used GIS tools based on their functionalities, usability, efficiency and cost in order to propose a package devoid of the identified limitations. This paper identified a more GIS package based on open source environment. The advantages of the latter package have been presented within the concept of systematic comparative assessment amongst various commonly used packages. In turn, a novel GIS open source-based tool referred to as Quantum Geographic Information System (QGIS) package was proposed. This package integrates python, Geographic Resources Analysis Support System (GRASS) as well as System for Automated Geoscientific Analyses (SAGA) modules to perform various GIS operations to proffer an efficient and all-encompassing features and plugins for pre-processing spatio-temporal data required for environmental analysis.

References

Aronoff, S. (1989). *Geographic information systems: A management perspective.* Ottawa, Canada: WDL Publications.

Bernhardsen, T. (2002). *Geographic information systems: an introduction*. Milton,Australia: John Wiley & Sons.

Clarke, K. C. (1997). *Getting started with geographic information systems*. Upper Saddle River, NJ: Prentice Hall.

Collier, K., Carey, B., Sautter, D., & Marjaniemi, C. (1999). A methodology for evaluating and selecting data mining software. In p*roceedings of the 32nd Annual Hawaii International Conference on. System Sciences* (pp. 11). Washington, DC: IEEE Computer Society

Eldrandaly, K., & Naguib, S. (2013). A knowledge-based system for GIS software selection. *Arab Journal of Information Technology*, *10*(2), 152–159.

Fotheringham, S., & Rogerson, P. (2013). *Spatial analysis and GIS*. Boca Raton, Florida: CRC Press.

Haklay, M., & Jones, C. E. (2008). Usability and GIS-why your boss should buy you a larger monitor. In *Proceeding of AGI GeoCommunity '08* (pp. 1-11). London, UK: UCL Press.

Keil, M., & Tiwana, A. (2006). Relative importance of evaluation criteria for enterprise systems: A conjoint study. *Information Systems Journal*, *16*(3), 237-262.

Maguire, D. J., Rhind, D. W., Goodchild, M., & Longley, P. A. (2010). Geographic information systems and science. Minneapolis MN: John Wiley & Sons.

Merry, K., Bettinger, P., Grebner, D. L., Boston, K., & Siry, J. (2016). Assessment of geographic information system skills employed by graduates from three forestry programs in the United States. *Forests*, *7*(12), 304-315.

Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling and Software*, *31*(1), 124–130.

Ngai, E. W., & Chan, E. W. C. (2005). Evaluation of knowledge management tools using AHP. *Expert systems with applications*, *29*(4), 889-899.

Ramamurthy, M., & Ho, Y. (2014). Integrated Data Viewer. In *EGU General Assembly Conference Abstracts* (Vol. 16). **Vienna, Austria: EGU General Assembly Online Press.**

Ravi, L., Yan, Q., Dascalu, S. M., & Harris Jr, F. C. (2013, March). A survey of visualization techniques and tools for environmental data. In *Proceedings of the 2013 Intl. Conference on Computers and their Applications (CATA 2013)*. Honolulu, Hawaii: International Society for Computers and Their Applications.

Singh, V., & Fiorentino, M. (2013). *Geographical information systems in hydrology* (Vol. 26). Berlin/Heidelberg, Germany: Springer Science & Business Media.