APPLICATION

PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2

Michael S. Rosenberg* and Corey Devin Anderson

School of Life Sciences and Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, Tempe, AZ, USA

Summary

1. Spatial analysis has become increasingly popular in the biological sciences, particularly in disciplines such as landscape ecology and landscape genetics. However, many statistical functions for performing spatial analysis are not readily available (except in the most limited manner) in common, easy-to-use statistical packages or geographic information systems (GIS) software.

2. Over the last decade, the software package Pattern Analysis, Spatial Statistics and Geographic Exegesis (PASSaGE) has been popular tool for conducting spatial statistics. PASSaGE is completely free and has a user-friendly graphical user interface. A new version of PASSaGE, rewritten from the ground up, has now been released and is available for download.

3. PASSaGE 2 is significantly more user friendly than the original release and provides an excellent platform for both scientific analysis and classroom training. PASSaGE 2 includes a broad array of spatial statistical analyses not commonly found in other software packages or GIS software, all in an easy-to-use framework. It includes support for one-, two- and three-dimensional spatial analysis, including a number of unique and newly developed approaches.

Key-words: geographic analysis, landscape ecology, landscape genetics, software, spatial analysis, spatial statistics

Introduction

Spatial analysis is a fundamental part of scientific inquiry, including ecological, evolutionary, and environmental science, epidemiology, geology, geography and mathematics. Recent technological advances in remote sensing and global positioning systems have led to a rapid expansion of the number and size of spatially explicit data sets available for analysis. Despite the broad use of these methods across the sciences, easy-to-use software for conducting these analyses has been difficult to find.

Many of the computer programs used for conducting spatial analysis in the late 1990s could only be run on local, antiquated mainframe computers. While working in the laboratory of Robert Sokal, MSR began reprogramming many of the methods into a more general spatial analysis package. This new software was designed to combine multiple analyses into a single package, reduce computational redundancy and allow for more flexible analysis (e.g., not forcing

Euclidean distance measures), while also modernizing the code into a user-friendly interface that would run on a desktop computer. As more and more analyses were added, including new and novel statistical approaches (Rosenberg 2000, 2004), the package eventually turned into a free Windows program called: Pattern Analysis, Spatial Statistics and Geographic Exegesis (PASSaGE) (Rosenberg 2001). The original PASSaGE program was somewhat quirky in its interface and methodological approach because it was never designed for broad use, but instead grew out of internal needs of the laboratory group. Despite never being formally advertised, word-of-mouth has led to thousands of downloads from over 60 countries and 170 educational institutions in the United States.

For the last several years, we have been developing a new version of the software, PASSaGE 2, rebuilt from the ground up and designed for a broader audience. The beta releases of the new version (available in various forms for the past 3 years) have seen widespread adoption, having been downloaded at more than twice the pace of the original release. In all, PASSaGE and PASSaGE 2 have seen more than 12 000 downloads from over 76 countries and

*Correspondence author. E-mail: msr@asu.edu
Correspondence site: http://www.respond2articles.com/MEE/

Table 1. Summary of the primary analyses of PASSaGE 2

Distance estimation	Networks & tessellations
Geographic Distances Data Distances (13 measures*) Geodesic/Shortest-Path Distances*	Distance-based, Nearest Neighbours, Minimum Spanning Trees, Relative Neighbourhood Networks, Gabriel graphs, Delaunay/Dirichlet Tessellations, Least Diagonal Networks
Point pattern analyses	**Contiguous data analyses**
Second-Order/Ripley's *K* (1D*, 2D, 3D*·†, bivariate*, anisotropy*) Dispersion indices Join-counts Angular wavelets*·†	Quadrat variance methods (1D, 2D, 3D*·†, covariance*) Wavelets (1D, 2D, 3D*·†, covariance*) Spectral analysis (8 wave forms*) Lacunarity analysis (1D, 2D, 3D*·†)
Scattered data analyses	**Boundary & cluster analyses**
Correlograms (Moran, Geary, Mantel) Variograms* Local statistics* (LISAs, Getis-Ord Stats)	Moving split-window* Agglomerative Clustering* Wombling (continuous & categorical) *
Scattered data anisotropy analyses	**Miscellaneous analyses**
Bearing analysis† Bearing correlograms† Windrose correlograms† Angular correlation†	Modified *t*-test for correlation Mantel tests, including partial mantel with three or more matrices* Point-line relationship analysis*·† Point-polygon relationship analysis*·†

*New in PASSaGE 2 (All older methods are enhanced from PASSaGE 1).
†Unique (as far as we know) to PASSaGE 2.

216 US educational institutions. PASSaGE and PASSaGE 2 have been used in a wide variety of disciplines for scientific research and for training in classrooms around the world. In the following, we describe the major features of PASSaGE 2, as we finalize its official release to the scientific community.

Spatial analysis tools

Spatial analytical tools are available in a variety of forms; they have been created as modules for programming languages, statistical platforms, geographic information systems (GIS) or as Standalone software. For example, PySAL is a
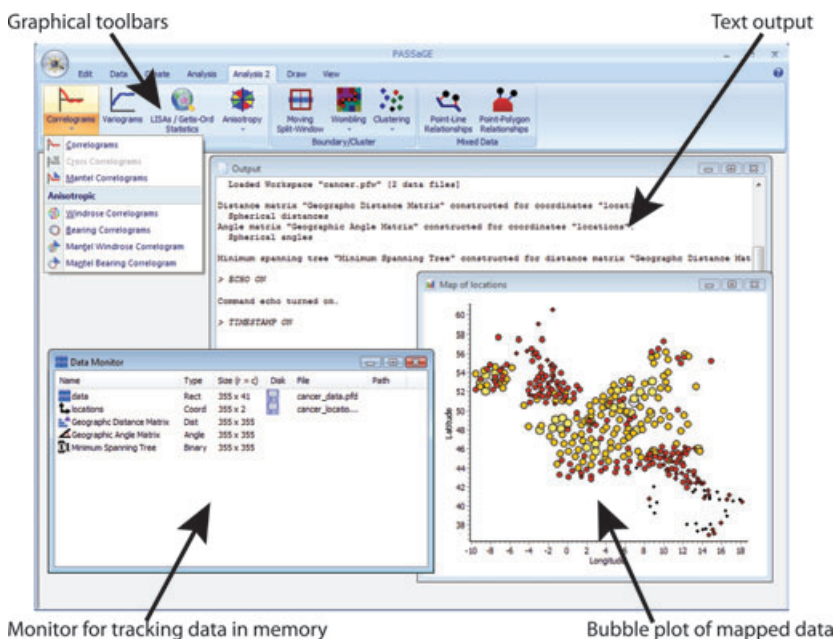

Graphical toolbars

Text output

Monitor for tracking data in memory

Bubble plot of mapped data

Fig. 1. The Graphical User Interface for PASSaGE 2 with some of the main elements highlighted.

library of Python functions intended to aid the development of applications for spatial analysis (Rey & Anselin 2010). Similarly, many spatial analysis packages have been written in statistical languages such as R (R Development Core Team, 2008) and for mathematical platforms such as Matlab (The Mathworks, Natick, MA, USA). R is popular for a variety of reasons, including easy accessibility to other useful statistical functions and graphing tools, while Matlab is good for implementing complex matrix operations. ArcGIS (ESRI, 2009) is popular for spatial data storage and visualization and also offers a handful of commonly used analysis functions in the Spatial Statistics toolbox. Standalone software for spatial analysis has been developed for various fields, including ecology and evolutionary biology (e.g., SAM: Spatial Analysis in Macroecology, Rangel, Diniz Filho, & Bini 2010).

When compared to other spatial statistical tools, PASSaGE 2 includes substantially more options for pattern exploration and spatial data description. PASSaGE 2 also supports data in one, two or three dimensions, with many analytical approaches adapted for three-dimensional data for the first time. While there is some overlapping functionality between PASSaGE 2 and other spatial analysis tools, software such as SAM and spatial analysis packages developed for R tend to have a stronger focus on spatial modelling than PASSaGE 2. Although some effort has been made to allow better data integration between PASSaGE 2 and popular GIS products (such as ArcGIS), PASSaGE 2 is first-and-foremost designed as a statistical tool.

Table 1 summarizes the major classes of methods included in PASSaGE 2. This is not an exhaustive list but highlights the major areas of emphasis as well as the breadth of options within the program. Structurally, PASSaGE 2 divides procedures into those for manipulating data, those for creating new data matrices from existing data matrices and those for analysis. The highlights of the creation component are construction
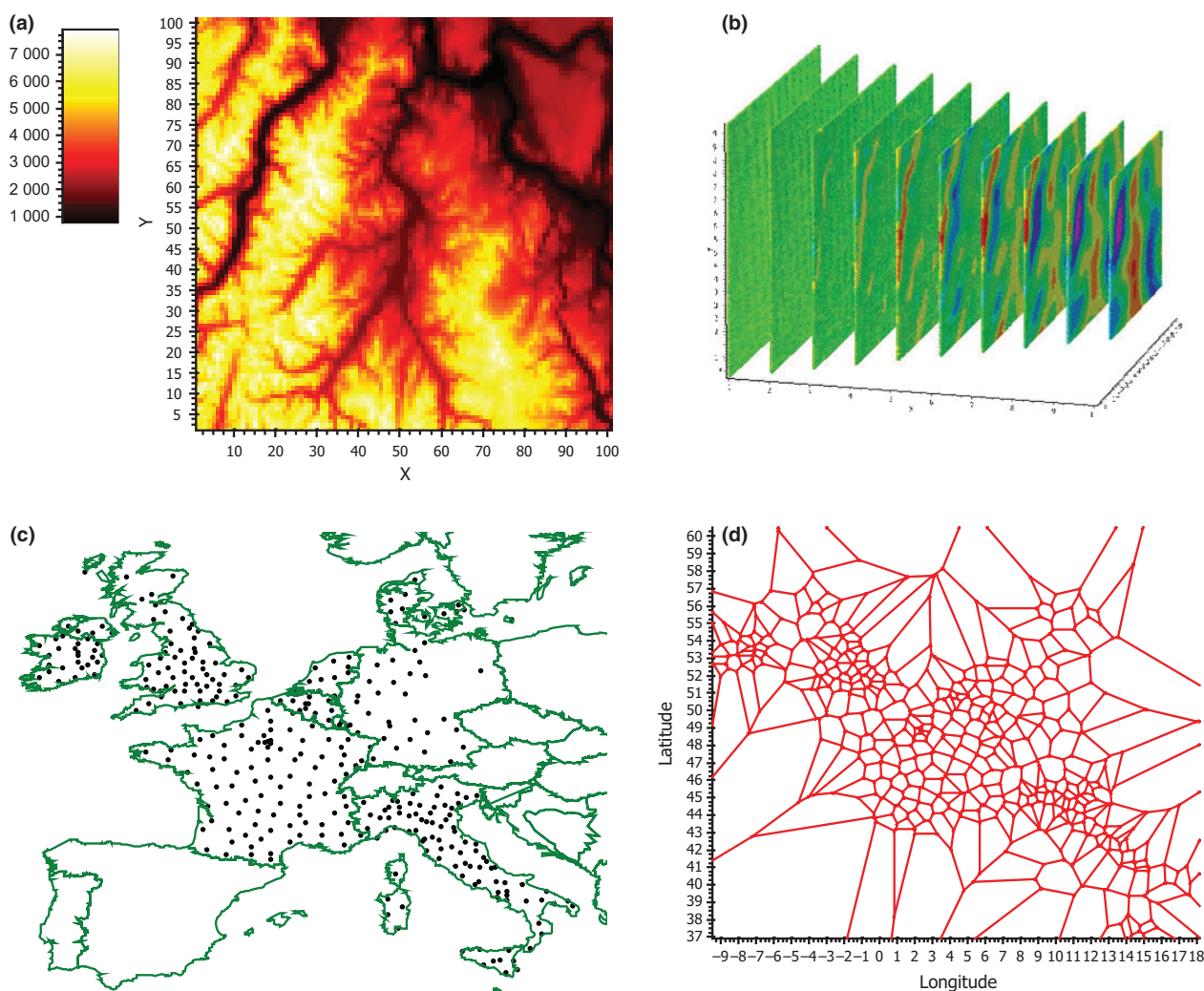

Fig. 2. Examples of some of the graphical output directly produced by PASSaGE 2. (a) A surface map of elevation data with an associated colour gradient scale. (b) The position × scale variance plot from a two-dimensional wavelet analysis of the data in (a), where each plane represents the positions within the plot for a different scale (increasing from left to right), with variance represented by colour. (c) A polygon map of Europe with a point plot of 355 cancer registration locations. (d) The tessellation of the same points represented in (c).

of distance matrices from geographic locations or variate data and the construction of network or connection schemes, such as minimum spanning trees or Delaunay triangulations. Additional creation options include the construction of distance/lag classes and summarizing point location data into overlaying grids.

The analysis section is primarily divided by data type: point pattern analyses, contiguous unit analyses and scattered data analyses, with an additional large focus on anisotropic analysis. Additional methods for boundary and cluster analysis (which can fall across multiple data types), Mantel tests, and other data types, such as line and polygon data, are also included.

Interface

Although designed primarily for use in a Windows environment, PASSaGE 2 comes in two forms: a Graphical User Interface (GUI) version and a command-line version (CMD). The GUI version is designed for 32-bit versions of Windows (all current varieties) and uses a typical mouse-driven point-and-click interface (Fig. 1); the CMD version is designed to compile and run on most operating systems (including Windows, Mac and Linux), and analyses are conducted using a batch language (thoroughly described in the manual). Batch files can also be executed using the GUI version for those desiring to do high-throughput analysis.

Data input/output

Data input is designed to be as flexible as possible. Data can be imported from text files, common spreadsheets and ArcGIS shapefiles, among others. Formats can be varied, with little-to-no restriction on the ordering of columns, the inclusion of column or row labels, etc. Data are stored internally within twelve different matrix types – rectangular matrices, coordinate matrices and distances matrices are among the more common types, allowing for storage optimization as well as specification of allowable analyses based on the type of data currently present. There are many data manipulation functions built into PASSaGE 2 to allow easy merging, splitting and transformation of data from one type or form to another. Data export is also flexible, with many potential file and data formats. In addition, PASSaGE 2 has a native binary format that it can use to store data (both single matrices and entire workspaces) for easy and simple reuse in subsequent runs.

There are no technical hard limits on the amount of data that can be processed by PASSaGE 2. However, from a practical standpoint, very large data sets will become untenable on many systems. For example, a geographic distance matrix for 10 000 locations contains almost 50 million pairwise distances. At double precision (8 bytes per value), this matrix will consume over 390 Meg of memory and disc space. A Mantel test

of two $10\,000 \times 10\,000$ distance matrices would require the simultaneous manipulation of 100 million numbers.

That being said, PASSaGE 2 has successfully performed analyses based on 10 000 locations. They consume a lot of time and resources and will occasionally crash older computers. Analyses of more reasonable numbers of data points should be quite stable. In the end, the practical limitations vary depending on the make-up of the computer system and the specific analyses being performed.

Graphics

The GUI version of PASSaGE 2 includes strong graphical support with a variety of mapping and graphic functions as well as graphical output for many methods (Fig. 2). The new version includes a powerful graphical editor allowing complete customization of all aspects of a graph, as well as the ability to export to numerous graphical file formats.

Documentation and availability

The manual is over 500 pages long, with detailed descriptions of all of the methods. It can be accessed three ways: (i) as a PDF included with the downloaded program, (ii) as an attached help file accessible directly from within the program (GUI version) and (iii) in a searchable and indexed set of web pages on the PASSaGE website. The software is freely available to all users and may be downloaded at http://www.passagesoftware.net.

Acknowledgements

References

ESRI (2009) *ArcGIS*. Environmental Systems Resource Institute, Redlands, CA.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rangel, T.F., Diniz Filho, J.A.F. & Bini, L.M. (2010) SAM: a comprehensive application for spatial analysis in macroecology. *Ecography*, **33**, 46–50.

Rey, S.J. & Anselin, L. (2010) PySAL: a python library of spatial analytical methods. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (eds M. Fischer & A. Getis), pp. 175–193. Springer, Berlin.

Rosenberg, M.S. (2000) The bearing correlogram: a new method of analyzing directional spatial autocorrelation. *Geographical Analysis*, **32**, 267–278.

Rosenberg, M.S. (2001) *PASSAGE: Pattern Analysis, Spatial Statitics and Geographic Exegesis*. Arizona State University, Tempe, AZ.

Rosenberg, M.S. (2004) Wavelet analysis for detecting anisotropy in point patterns. *Journal of Vegetation Science*, **15**, 277–284.