Computers & Geosciences 36 (2010) 539–549

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

New image processing software for analyzing object size-frequency distributions, geometry, orientation, and spatial distribution ☆

Ciarán Beggan [a,*], Christopher W. Hamilton [b]

[a] Grant Institute, School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK
[b] Hawai'i Institute of Geophysics and Planetology, University of Hawai'i, USA

ARTICLE INFO

Article history:
Received 14 November 2008
Received in revised form
21 September 2009
Accepted 23 September 2009

Keywords:
Image analysis
Software
Size-frequency
Spatial distribution
Nearest neighbor
Vesicles
Volcanology

ABSTRACT

Geological Image Analysis Software (GIAS) combines basic tools for calculating object area, abundance, radius, perimeter, eccentricity, orientation, and centroid location, with the first automated method for characterizing the aerial distribution of objects using sample-size-dependent nearest neighbor (NN) statistics. The NN analyses include tests for (1) Poisson, (2) Normalized Poisson, (3) Scavenged $k=1$, and (4) Scavenged $k=2$ NN distributions. GIAS is implemented in MATLAB with a Graphical User Interface (GUI) that is available as pre-parsed pseudocode for use with MATLAB, or as a stand-alone application that runs on Windows and Unix systems. GIAS can process raster data (e.g., satellite imagery, photomicrographs, etc.) and tables of object coordinates to characterize the size, geometry, orientation, and spatial organization of a wide range of geological features. This information expedites quantitative measurements of 2D object properties, provides criteria for validating the use of stereology to transform 2D object sections into 3D models, and establishes a standardized NN methodology that can be used to compare the results of different geospatial studies and identify objects using non-morphological parameters.

1. Introduction

Geological image analysis extracts information from representations of natural objects that may either be captured by an imaging system (e.g., photomicrographs, aerial photographs, and digital satellite imagery) or schematically rendered into visual form (e.g., geological maps). In addition to examining the properties of individual objects, spatial analysis may be used to quantify object distributions and investigate their formation processes.

ImageJ (Rasband, 2005) is a commonly used image processing application that was developed as an open source Java-based program by the National Institutes of Health (NIH). Custom plug-in modules enable ImageJ to solve numerous tasks, including the analysis of vesicle size-frequency distributions within geological thin-sections (e.g., Szramek et al., 2006; Polacci et al., 2007). However, ImageJ and similar programs for Windows (e.g., Scion Image and ImageTool) and Macintosh (e.g., NIH Image) are not specifically designed for geological applications nor do they provide analytical tools for investigating patterns of spatial organization.

To take greater advantage of the information contained within geological images, we have developed Geological Image Analysis Software (GIAS). This program combines (1) an image processing module for calculating and visualizing object areas, abundance, radii, perimeters, eccentricities, orientations, and centroid locations, and (2) a spatial distribution module that automates sample-size-dependent nearest neighbor (NN) analyses. Although other programs can perform the basic functions in the "Image Analysis" module, GIAS is the first program to automate sample-size-dependent analyses of NN distributions.

2. Motivation

Nearest neighbor (NN) analysis is well-suited for investigating patterns of spatial distribution within intrinsically two-dimensional (2D) datasets, such as orthorectified aerial photographs and satellite imagery. Applications of NN analyses to remote sensing imagery include the study of volcanic landforms (Bruno et al., 2004, 2006; Baloga et al., 2007; Bishop, 2008; Hamilton et al., 2010b; Bleacher et al., 2009), sedimentary mud volcanoes (Burr et al., 2010b), periglacial ice-cored mounds (Bruno et al., 2006), glaciofluvial features (Burr et al., 2009), dune fields (Wilkins and Ford, 2007), and impact craters (Bruno et al., 2006). Despite widespread utilization of NN analyses, its value as a remote sensing tool and effectiveness as basis for comparison between

☆ Code available from server at http://www.geoanalysis.org.
* Corresponding author. Tel.: +44 131 650 5916.
E-mail address: ciaran.beggan@ed.ac.uk (C. Beggan).

0098-3004/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2009.09.003

Nomenclature

A	area of a feature field
A_{hull}	area of the convex hull
c	statistic for comparing observed (actual) to expected mean NN distances
k	Poisson index
N	number of features within a sample population
N_i	number of features within the convex hull
N_v	number of features forming the vertices of the convex hull
NN	abbreviation for nearest neighbor
p	probability
R	statistic for comparing actual (r_a) to expected (r_e) mean NN distances
r	radial distance
r_0	threshold distance used to calculate Normalized Poisson NN distributions
r_a	mean NN distance observed in the data
r_e	mean NN distance calculated from a given NN model (e.g., Poisson)
P	probability density
ρ_0	population (spatial) density of the input features ($\rho_0=N/A$)
ρ_i	population (spatial) density of the input features ($\rho_i=N_i/A_{hull}$)
σ	standard deviation
σ_e	standard error

different datasets is limited by the lack of a standardized NN methodology—particularly in terms of defining feature field areas, thresholds of significance, and criteria for apply higher-order NN methods.

In addition to remote sensing applications, NN analyses can be used to study objects in photomicrographs such as crystals (Jerram et al., 1996, 2003; Jerram and Cheadle, 2000) and vesicles. In this study, we emphasize the application of NN analyses to vesicle distributions to demonstrate how GIAS can be used to validate (or refute) the assumption of randomness, which is a prerequisite for effectively applying stereological techniques to derive vesicle volumes from photomicrographs and for selecting appropriate statistical models to characterize those vesicle populations.

Vesicle textures preserve information about the pre-eruptive history of magmas and can be used to investigate the dynamics of explosive and effusive volcanic eruptions (e.g., Mangan et al., 1993; Cashman and Mangan, 1994; Mangan and Cashman, 1996; Cashman and Kauahikaua, 1997; Polacci and Papale, 1997; Blower et al., 2001, 2003; Gaonac'h et al., 2005; Shin et al., 2005; Lautze and Houghton, 2005; Adams et al., 2006; Polacci et al., 2006; Sable et al., 2006; Gurioli et al., 2008). Quantitative vesicle analyses stem from Marsh (1988), who explored the physics of crystal nucleation and growth dynamics to derive an analytical formulation for crystal size distributions. This research was then applied to numerous bubble size-frequency distribution studies (e.g., Sarda and Graham, 1990; Cashman and Mangan, 1994; Blower et al., 2003). Early studies of vesicle distributions (e.g., Cashman and Mangan, 1994) were limited by their inability to characterize the full range of vesicle sizes because their methodology could not resolve the smallest vesicles. Nested photomicrographs solve this problem because photomicrographs captured at multiple magnifications enable the reconstruction of total vesicle size-frequency distributions (e.g., Adams et al., 2006; Gurioli et al., 2008).

In general, vesicle studies are limited by two major factors: transformations of 2D cross-sections into representative vesicle volumes (Mangan et al., 1993; Sahagian and Proussevitch, 1998; Higgins, 2000; Jerram and Cheadle, 2000), and development of reliable statistical characterizations of vesicle populations in terms of distribution functions and their spatial characteristics (Morgan and Jerram, 2006; Proussevitch et al., 2007a). These difficulties have been partially addressed by improved statistical techniques for investigating vesicle populations; however, a single cross-section cannot be used to reconstruct a representative 3D vesicle distribution unless the objects viewed in a 2D section can be characterized using 2D reference textures with known 3D distributions (Jerram et al., 1996, 2003; Jerram and Cheadle, 2000; Proussevitch et al., 2007a). Although synchrotron X-ray tomography is increasingly being used to directly generate 3D vesicle distributions (e.g., Gualda and Rivers, 2006; Polacci et al., 2007; Proussevitch et al., 2007b), nested datasets containing multiple scanning electron microscope (SEM) images remain the most common input for vesicle studies because of their superior spatial resolution relative to X-ray tomography. To facilitate the analysis of vesicles in SEM imagery, GIAS can be used to determine the geometric properties of vesicles within the plane of a given photomicrograph and establish if objects fulfil the criteria of spatial randomness, which is required for transforming 2D sections into 3D models using stereology.

3. Nearest neighbor (NN) analysis

Clark and Evans (1954) proposed a simple test for spatial randomness in which the actual mean NN distance (r_a) in a population of known density is compared with the expected mean NN distance (r_e) within a randomly distributed population of equivalent density. Following Clark and Evans (1954), r_e and expected standard error (σ_e) of the Poisson distribution are as follows:

$$r_e = \frac{1}{2\sqrt{\rho 0}}; \quad \sigma_e = \frac{0.26136}{\sqrt{N\rho_0}}, \tag{1,2}$$

where the input population density, ρ_0, equals the number of objects (N) divided by the area (A) of the feature field ($\rho_0=N/A$). The following two test statistics (termed R and c) are used to determine if r_a follows a Poisson random distribution:

$$R = \frac{r_a}{r_e}; \quad c = \frac{r_a - r_e}{\sigma_e}. \tag{3,4}$$

If a test population exhibits a Poisson random distribution, R should ideally have a value of 1, while c should equal 0. If R is approximately equal to 1, then the test population may have a Poisson random distribution. If $R>1$, then the test population exhibits greater than random NN spacing (i.e., tends towards a maximum packing arrangement), whereas if $R<1$, then the NN distances in the test case are more closely spaced than expected within a random distribution and thus exhibit clustering relative to the Poisson model.

To identify statistically significant departures from randomness at the 0.95 and 0.99 confidence levels, $|c|$ must exceed the critical values of 1.96 and 2.58, respectively (Clark and Evans, 1954); however, these critical values implicitly assume large sample populations ($N>10^4$). Jerram et al. (1996) and Baloga et al. (2007) note that finite-sampling effects introduce biases into the variation of NN statistics. These biases become significant for small populations ($N<300$) and thereby necessitate the use of sample-size-dependent calculations of R and c thresholds.