Journal of Structural Geology 46 (2013) 22–33

Contents lists available at SciVerse ScienceDirect

Journal of Structural Geology

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

Fitting the void: Data boundaries, point distributions and strain analysis

Kieran F. Mulchrone*

Department of Applied Mathematics, University College Cork, Cork, Ireland

ARTICLE INFO

Article history:
Received 14 August 2012
Received in revised form
19 October 2012
Accepted 28 October 2012
Available online 6 November 2012

Keywords:
Strain analysis
Delaunay triangulation
Point data
Objective
Numerical

ABSTRACT

An objective method is developed for identifying the shape of the void generated by analysis of the spatial arrangement of point distributions using the Fry method and in particular Delaunay triangulation nearest neighbour data. The method works by numerically minimising a weighted non-linear least-squares formulation which tends to selectively apply higher weights to points below the boundary and less weight to points above the boundary. Sampling errors are estimated using a bootstrap procedure. A simulation study indicates that the method works best for closely packed object arrangements and tends to underestimate the strain axial ratio at high imposed strains. Overall the method is found to be objective and consistent.

1. Introduction

Geological strain analysis is an indispensable tool for unravelling tectonic histories. Choosing a method of strain analysis is largely guided by the presence of suitable geometrical features in the study area typically known as strain markers. There are three categories of strain markers (Lisle, 2010) commonly in use: 1) individual objects or groups of objects with sufficiently known pre-strain geometric detail, 2) collections of objects with minimal geometric detail but whose shape may be approximated by an ellipse (Mulchrone and Roy Choudhury, 2004), and 3) collections of objects whose mutual spatial arrangement may be used to estimate strain. Category 1 includes various fossils of known unstrained geometry and the basic equations of strain can be applied to one or more markers to determine finite strain (Ramsay, 1967; Ramsay and Huber, 1983). Category 2 usually consists of collections of mineral grains or sedimentary clasts etc. and a variety of methods for analysis have been developed including the R_f/ϕ (Ramsay, 1967), the method of Robin (1977), the mean radial length (Mulchrone et al., 2003), and others (Shimamoto and Ikeda, 1976; Yamaji, 2008). For successful recovery of strain estimates with categories 1 and 2 strain markers, deformation must be homogenous on the scale of measurement and markers must have behaved passively during deformation. Homogeneity of object/fossil morphology is also assumed in category 1 methods and typically isotropy of initial orientation is assumed in category 2 methods. Strain analysis with the third category also assumes homogeneity of deformation at the scale of the sample but importantly does not require markers to have any particular initial orientation or shape within reason and also does not make any make any prescription on how markers ought to deform relative to the enclosing medium i.e. passive behaviour is not required at the marker scale. There is also a cautionary restriction of coaxial strain which is important only if the data initially formed part of regular grid (Lisle, 2010). Methods of strain analysis that work with category three data ought to be very general and widely applicable (Ramsay, 1967, p. 195, Fry, 1979; Hanna and Fry, 1979). There also exist other methods such anisotropy of magnetic susceptibility (AMS) which measure non-visible features and are commonly used as strain proxies (Hrouda and Ježek, 1999; Ježek and Hrouda, 2002; 2007).

Methods which utilise all object–object separations such as Fry's method (Fry, 1979; Hanna and Fry, 1979), the Normalised Fry Method (Erslev, 1988) and the enhanced Normalised method (Erslev and Ge, 1990) are commonly used due to their simplicity, ease of application and perhaps because they can be graphically implemented. These methods have also been extended to work well with irregularly shaped objects (Mcnaught, 1994). Objective interpretation of the resulting data (i.e. fitting an ellipse to a central vacancy) remains problematical (Erslev and Ge, 1990; Waldron and Wallace, 2007; Lisle, 2010). The enhanced normalised method suggests selecting a subset of the data where objects are in close contact. This results in a set of data for which the best fit ellipse may

* Tel.: +353 21 4205822.
E-mail address: k.mulchrone@ucc.ie.

0191-8141/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jsg.2012.10.011

be readily calculated (Mulchrone, 2003). However this tends to make the data converge with category two data because the selected subset of data is sensitive to the shape of objects, thus the benefits of processing category 3 data are lost. In addition, this may not be useful in practice when marker objects to be processed are far apart (Waldron and Wallace, 2007). Another approach is to construct an elliptical annulus and then search by computer for the annulus which maximises the difference between the point count of the inner and outer ellipses (Waldron and Wallace, 2007). Waldron and Wallace, 2007 also suggested a continuous function method which seeks to minimise the number of points on one side of the best fit ellipse. Lisle (2010) developed a method whereby the natural dataset is undeformed for various orientations and axial ratios until the undeformed dataset has the expected characteristics of undeformed point configuration. This technique has the additional benefit of providing information about the precision of the result. Shan and Xiao (2011) present a statistical method of analysis of Fry data by modelling the data with a truncated Poisson process and formulating a maximum likelihood solution which is numerically solved.

The alternative to using all object–object separations is to use nearest neighbours only (Ramsay, 1967, p. 195) which is onerously labour intensive in its initial formulation. However application of techniques from computational geometry (O' Rourke, 1993) lead to the development of a computationally efficient Delaunay Triangulation Nearest Neighbour Method (Mulchrone, 2003). In this paper a method is introduced for determining strain from category 3 strain markers which seeks to apply high weighting to points on the inner boundary of the central vacancy found in a Fry-type plot resulting in a non-linear least squares minimisation problem. The method uses data from the Delaunay Triangulation of the points and thus the method fits into the nearest neighbour suite of methods. Sampling errors associated with the method are readily estimated using a bootstrap procedure. In addition to the method results of a detailed simulation to assess the error characteristics of the method are reported.

2. Point distributions

The raw data required for the proposed method for the analysis are the centroids of the objects (Fry, 1979; Mulchrone, 2003) and for normalisation purposes the average radius ($=\sqrt{r_{\max}r_{\min}}$ for an elliptical object, where $r_{\max}$, $r_{\min}$ are the long and short axes respectively). If average radius data are unavailable then assigning the value of 1/2 to every data point means that the normalisation step is effectively ignored. There are two fundamental assumptions made by all methods of strain estimation using point distributions: 1) an anticlustered initial distribution of objects and 2) the same point set can be identified before and after deformation. Object distributions with these features can be readily generated using the simple sequential inhibition process described in section 4.2 of Mulchrone (2003).

In its original conception (Fry, 1979) and owing to its graphical implementation, the data to be analysed consisted of all object–object separations – this is referred to as the Fry distribution (see Fig. 1(a) and (b)). Another possible distribution for analysis consists of the distribution of nearest neighbours whereby the set of points is made up of the one nearest neighbour to each point in the distribution thus resulting in one data point per analysed object (this is referred to as the Nearest Neighbour distribution, see Fig. 1(c) and (d)). Each of these distributions has difficulties associated with it. The Fry distribution contains excessive data because the central void is defined by a subset of data relating mainly to objects which are close together. Therefore most of the data which relates to far apart objects is redundant and tends to slow down computational approaches. On the other hand the Nearest Neighbour distribution is at the other extreme and tends to be overly sparse. This sparsity tends to be accentuated with increasing strain because the points aligned close to the direction of maximum stretching tend not be the nearest neighbour (see Fig. 1(d)). The preferred subset of data used in this contribution is the set of pairwise points defined by the Delaunay Triangulation (Mulchrone, 2003) and is referred to as the Delaunay Triangulation Nearest Neighbour (DTNN) distribution. This dataset tends to be of a manageable size and the central void continues to be well defined even after considerably large imposed strains (see Fig. 1(e) and (f)).

2.1. Selecting distribution parameters

In order to test the effectiveness of the proposed method of analysis it will be necessary to choose simulation parameters appropriately. There are five main interrelated parameters that may be controlled: packing density (p_d, which is the ratio of the area comprised by the object to that of the total area), number of objects (n), the distribution of radii of those objects from some minimum (a) to some maximum (b) and finally the size of the area enclosing the objects to be analysed. For simplicity it is assumed that the containing area is square with edge size given by l. These factors are geometric in nature and others such as point distribution type, normalisation etc. are separate and considered later.

The packing density is the ratio of the sum of all object areas to that of the containing area. Suppose we generate a distribution of circular objects each with radius r then the total area of the objects is $\sum_{i=1}^{n}\pi r^2$. In practice objects have a variable radius which is modelled here as a random selection from a uniform distribution between a and b. The probability of each radius in the interval is equal and given by $1/(b-a)$. The mean or expected value of the area of such a collection of objects is calculated by evaluating the following integral:

$$\int_a^b \frac{\pi r^2}{b-a}\mathrm{d}r = \frac{\pi}{3}\left(a^2+ab+b^2\right)$$

For simplicity let $a=kb$, $0<k\leq 1$, hence the expected area of a collection of n objects is $n\pi/3(1+k+k^2)b^2$. Notice that the smaller the value of k the more variability in the distribution of radii. Hence

$$p_d = \frac{n\pi}{3l^2}\left(1+k+k^2\right)b^2$$

or more usefully

$$b^2 = \frac{3l^2p_d}{n\pi(1+k+k^2)}$$

Thus by first choosing the required values of n, p_d, l, and k, the appropriate value of b is calculated.

There is a complication associated with this approach to choosing appropriate parameter values because in addition to selection from a uniform distribution there is also a further selection effect due to the simple inhibition process. This tends to modify the uniform distribution of object radii into a right skewed distribution containing more objects of smaller radius than objects with larger radius (see Fig. 2). The principal difficulty is that the number of circular objects generated tends to be greater than that requested. However, simulations indicate that there is a consistency in the number of objects generated for a particular choice of parameters and the effect is pronounced only in the case of low k (i.e. high radii variability) and low p_d. In other words, it is still possible to exert some control the number of objects generated in a simulation.