Variable Star Symbols for Seismicity Plots

Duncan Carr Agnew

Institute of Geophysics and Planetary Physics,

Scripps Institution of Oceanography,

University of California San Diego,

La Jolla, California, USA

April 7, 2014

Supplemental material (plots and scripts)

may be found online at

`http://igppweb.ucsd.edu/~agnew/Miscsoft/starsym.html`

SUPPLEMENT

Introduction

At least since Mallet (1858) seismicity maps have been a way of showing earthquake activity, but, because earthquake size varies greatly, it is difficult to provide an accurate representation of the spatial variation of total seismic energy or moment release. The most common approach, using for each earthquake a single geometric figure of varying size, creates substantial overlap between large symbols. To reduce overlap and provide a more distinctive gradation, I propose a family of symbols in which size and shape vary together, from polygonal to star-shaped as their size increases. Two functions determine how symbol size and shape vary with value, and even a simple parametrization gives considerable flexibility in symbol design. Tests show that, given an appropriate key, the symbol value can be estimated to within better than 5% of the range covered.

Symbolization in seismicity maps is challenging. Using identical symbols plotted at the epicenters (or, in a cross-section, the hypocenters), shows where there are more or fewer earthquakes, but not how (say) moment release is distributed. This can be done by spatially smoothing the amount of energy or moment release and contouring the result, (e.g. Allen *et al.* (1965)), but this in turn removes fine details.

Most seismicity maps use a simple geometrical figure, usually a circle or square, to symbolize for each earthquake, and vary its size with the earthquake magnitude. The interior of each symbol can be filled with a color to denote depth (or, in Web displays, recency of occurrence). One problem with this approach is that one large symbol can easily cover many small ones. Some seismicity maps use use two geometrical figures, one (often a scaled circle) for events less than some magnitude, and another (often a star) for the larger events. In cartography such identical shapes of different size are called scaled (or graduated) point symbols, and, as in seismicity maps, are used to associate geographical locations with some quantity: for example, cities with their population size.

I propose a set of symbols, called variable stars, in which shape and size both depend on an associated value, which might be city population or earthquake magnitude. As the value increases, the shape changes gradually from a scaled polygon (which looks like a circle) to a more and more pointed star. The star shape causes the perceived symbol size to grow much more than the actual symbol area, decreasing the amount of

overlap between symbols. If a key is provided, jointly changing shape and size helps the viewer to better estimate the value associated with a particular symbol.

1 Design

The procedures used for drawing stars with different shapes and sizes are themselves illustrated in Figure 1. To produce a family of variable stars several properties need to be specified: both n, the number of points in the star-shaped form; and two functions of the associated size value z. These functions specify an inner radius $r(z)$, and a scaling factor $a(z)$ which sets the outer radius ar (so $a \geq 1$). To form the shape, points are placed on the outer radius at angles $2\theta = 2\pi/n$, and on the inner radius at the same angular spacing, but offset by $\theta = \pi/n$, with the symbol formed by connect them. If $a = 1$, the symbol is a $2n$-sided polygon. As a increases, the shape changes. For $a = (\cos\theta)^{-1}$, the figure becomes an n-sided regular polygon; as a increases beyond this, the figure becomes star-shaped. When a is equal to $\cos\theta + \sin\theta\tan 2\theta$, the line segments on either side of a point are collinear; in Figure 1 these are (for example) the segments marked AB and CD. I denote this value by s_n; it is defined only for $n \geq 5$; for $n = 5$, as in Figure 1, $s_n = 2.618$. Star shapes with $a < s_n$ would usually be characterized as having stubby points, and those with $a > s_n$ elongated ones, so s_n serves as a boundary between different shapes.

A more formal development starts by defining, For a size parameter z, a nondimensional scaled parameter

$$u = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{1}$$

where z_{min} and z_{max} cover the range of z expected, so $0 \leq u \leq 1$. The size and shape of the symbol as a function of u depends on the functions $r(u)$ and $a(u)$; for the symbol size to increase with u both functions have to be nondecreasing.

A simple choice for these functions that also allows a wide range of behavior is to use powers of u over parts of the range. This choice is partly motivated by the expressions for the area of the symbol, A_s, and the area of the circumscribing polygon, A_c; the latter corresponds more accurately to how "large" the symbol appears

to be. These areas are given by

$$A_s = ar^2 f_n \qquad A_c = a^2 r^2 f_n \tag{2}$$

where $f_n = n \sin(\pi/n)$. For $a = 1$ these are the same, as they should be; otherwise the ratio of areas is $A_s/A_c = a^{-1}$.

The choice made here for $r(u)$ and $a(u)$ is designed to produce polygons that increase in area as u^p for small values of u and stars that do the same (and become more spiky) for larger values:

$$\begin{aligned} r(u) &= bu^{p/2} \qquad a(u) = 1 \quad \text{for} \quad u < u_c \\ r(u) &= bu_c^{p/2} \qquad a(u) = \left(\frac{u}{u_c}\right)^{p/2} \quad \text{for} \quad u \geq u_c \end{aligned} \tag{3}$$

where p is the power-law dependence that makes $A_c \propto u^p$, b is a constant that sets the scale of the symbol, and u_c is the value of u above which the variation is confined to the outer radius. The panels on the right-hand side of Figure 1 show how r and a would vary with u for $b = 1$ and four different values of p.

How the shapes vary depend on what u_c is chosen to be. One choice is to always use a particular value of u, denoted by u_s. The panels in Figure 1 show another choice, which is to make $a(u_s) = s_n$: this associates a particular value of u (in Figure 1 this value is 0.6) with the symbol being a perfect star. Then

$$u_c = \frac{u_s}{s_n^{2/p}} \tag{4}$$

Figure 2 shows what sets of symbols this produces for the four values of p used in Figure 1. Increasing values of p naturally produce more size variation, but they also result in more variations in symbol shape, since the requirement that $a(u_s) = s_n$ means that the value of a for $u = 1$ increases as p increases. This formulation leaves the power dependence of area up to the symbol designer; I evaluate two possible choices in Section 4.

2 Implementation

The actual drawing of star shapes on a map requires some care because the coordinates of the symbol are a mix of geographical ones, describing the center, and coordinates in map units as actually plotted. Once n, p, and u_s are chosen, equation (3) gives r and a for any value of u. The items to be plotted have a latitude ϕ, a longitude λ, and a size z. To plot the symbols in map units (inches or cm), first apply the map projection to convert (ϕ, λ) to map coordinates (x, y). Then use equation (1) to convert z to u, find the values of $r(u)$ and $a(u)$ using equations (4) and (3), and compute the $x - y$ coordinates of the polygon as (x_m, y_m). For m ranging from 0 to $2n$

$$\begin{aligned} x_m &= x + ar\sin(m\theta) \qquad y_m = y + ar\cos(m\theta) \qquad \text{for} \quad m \quad \text{even} \\ x_m &= x + r\sin(m\theta) \qquad y_m = y + r\cos(m\theta) \qquad \text{for} \quad i \quad \text{odd} \end{aligned} \tag{5}$$

which gives the $2n + 1$ locations needed to complete the polygon, and means that one point of the star points vertically up. (The design could include a rotation of the shapes according to some other variable, but this makes overlapping symbols more difficult to distinguish).

For some mapping softwares, it may be possible to use the (x, y) coordinates directly to draw the symbol on the map. With the Generic Mapping Tools (GMT) package (Wessel and Smith, 1991), one would first use the `mapproject` program to project the geographic coordinates (ϕ, λ) of each point to its (x, y) coordinates on the plot; then combine these with the point values to construct stars in (x, y) coordinates; and finally plot these (x, y) coordinates directly, without applying a projection (this is the `-Jx` option in GMT). (The alternative, of inverse projecting (x, y) to (ϕ, λ) and then projecting forward, means that the symbols cannot go beyond the map edges). The electronic supplement SUPPLEMENT contains a sample GMT script, with an associated fortran program to produce the coordinates for the stars.

3 Examples

My first example is a global map of shallow earthquakes (Figure 3) which uses the variable star symbols for different magnitudes. Even in active regions, the shape of the symbols for the largest events make it possible to distinguish them visually; for example, it is possible to see that there are three events with $M_w \geq 8.5$ near the NW end of Sumatra (earthquakes in December 2004, March 2005 and April 2012). In the most active regions (such as Japan) the very high density of symbols is still difficult to resolve, so as a further visual cue the smallest magnitudes are given a less saturated color (gray), with larger events overplotted in two saturated ones: black for $7.5 \leq M_w < 8.5$ and red for $M_w \geq 8.5$. (A larger map with relatively smaller stars is included in the electronic supplement). SUPPLEMENT For symbols that can be clearly distinguished, it is possible to estimate magnitudes to within 0.3 units (see Section 4).

Figure 4 shows a cross-section of seismicity in the Tonga-Fiji seismic zone, in roughly the same region as was used by Brudzinski and Chen (2003) in an earlier discussion of choices for symbol size, though Figure 4 is over a longer timespan and uses the GEM catalog. Two shades are used for different size ranges: the smaller symbols (magnitudes less than 6.4) are filled and gray and the larger ones are unfilled and black.

4 Testing

Of course, the important question for symbols of this type is, how well can people viewing them estimate the value of the attribute that is represented? This requires actual testing (Cleveland and McGill, 1984; Cleveland, 1993), in this case a test in which viewers are presented with a range of symbols, or other graphical elements, and asked to estimate their values. A number of investigators have tested how to scale symbols of the same shape so that viewers will most accurately interpret size as actual value, though how meaningful such psychophysical measurements are has been questioned (Montello, 2002).

As a preliminary test of how accurately the variable star symbols are interpreted, I prepared sheets, each containing 25 symbols corresponding to values distributed randomly from 0 to 1.1, and also including instructions and a symbol key from 0.1 to 1.0 at intervals of 0.1 (a sample sheet is included in the electronic

supplement). SUPPLEMENT Two sets of eight sheets were prepared, one for the progression with $p = 2$ and the other for $p = 4$. Randomly selected graduate students in geophysics were asked to provide, for each symbol, a best estimate of its value and a possible range within which it would fall. Some participants filled out two sheets (one for $p = 2$ and one for $p = 4$, but with different patterns of symbols), and some only one sheet. For each progression this test thus provided 200 estimates and 190 ranges (these were not always filled out). Figure 5 shows the results, plotting the difference between the estimates against the true value. For the estimates of the value, the most applicable statistic is the regression coefficient r^2, and for the intervals, the fraction that cover the true value. For both progressions the estimates are well-correlated with the actual values, but the figure shows that for $p = 4$ the differences have noticeably less scatter; also, several of the participants stated that they found it easier to make an estimate for the $p = 4$ progression. Except for the smallest values, the estimates are slightly smaller than the true value, although the difference is not large: the median value of the differences for $p = 4$ is -0.018, which is less than 2% of the range. This systematic bias means that the estimated ranges only include the true value about 60% of the time.

5 Previous Use of Variable Symbols

A search of the literature on cartography and statistical graphics (for example, Robinson *et al.* (1995), MacEachren (1995), and Brewer (2008)) has not provided any examples of symbols whose size and shape change together. In the literature on statistical graphics variation in symbol shape is discussed in terms of using maximally distinctive shapes to indicate different classes of data on scatterplots (Lewandowsky and Spence, 1989; Tremmel, 1995; Krzywinski and Wong, 2013), but not for indicating associated values. Indeed Bertin (1983/2010) states that symbol shape cannot be used to express an ordered quantity – which is certainly true in general. The preferred practice in cartographic theory is to associate one symbol attribute with a single variable, so that symbol size, symbol shape, and symbol color would each represent different variables. But few point quantities in statistical cartography show the range of size, and amount of clustering, that earthquakes do.

Usually size variations of a particular shape (most often and circle or square) are used to indicate value

(Mersey, 1996), although this is complicated by the fact that, even under ideal conditions, perceived area A_p is a nonlinear function of actual area A, with the best relation being a power law $A_p = A^s$, with s about 0.8 (Williams, 1956; Dent, 1996; Montello, 2002). Such scaling attempts to avoid forcing the viewer to consult a symbol key; but for earthquake magnitude no natural scaling is possible, since there is no zero point, and different plots will include different magnitude ranges. So it is always necessary to have a key relating plotted size to magnitude.

Star-shaped symbols have been used in cartography and statistical graphics, though for a different purpose than proposed here, namely by using the lengths of the arms of the star (or polygon) to represent a multivariate quantity. The first use of this, over 150 years ago, was for representing the distribution of wind directions (Agnew, 2004), but they have been used since for a variety of quantities (Wainer, 1997; Klippel *et al.*, 2009).

There was one type of map in which point symbols were varied systematically in size and shape, namely star maps – though this is no longer true. Early star atlases often used elaborately engraved symbols of different types to indicate different magnitudes of stars (Herlihy, 2007). Over time these patterns became simplified, with a common method being to make fainter stars smaller and also give them fewer points; for example Argelander (1843). But in the nineteenth century this style was replaced by scaled circles (e.g. von_Littrow (1854)), and these are used in all modern star charts (e.g., Tirion *et al.* (2001)).

6 Conclusion

In this note I have proposed a set of symbols, varying systematically in shape and size, for representing items of various sizes located at various locations. A variety of such symbol families are no doubt possible; the particular one offered here, shapes that vary from polygons to more and more pointed stars, would appear to be useful for plotting phenomena, such as seismicity, that are heavily clustered. Because the star shape means that symbol area increases much less rapidly than symbol size, overlap is minimized even for larger symbols close together.

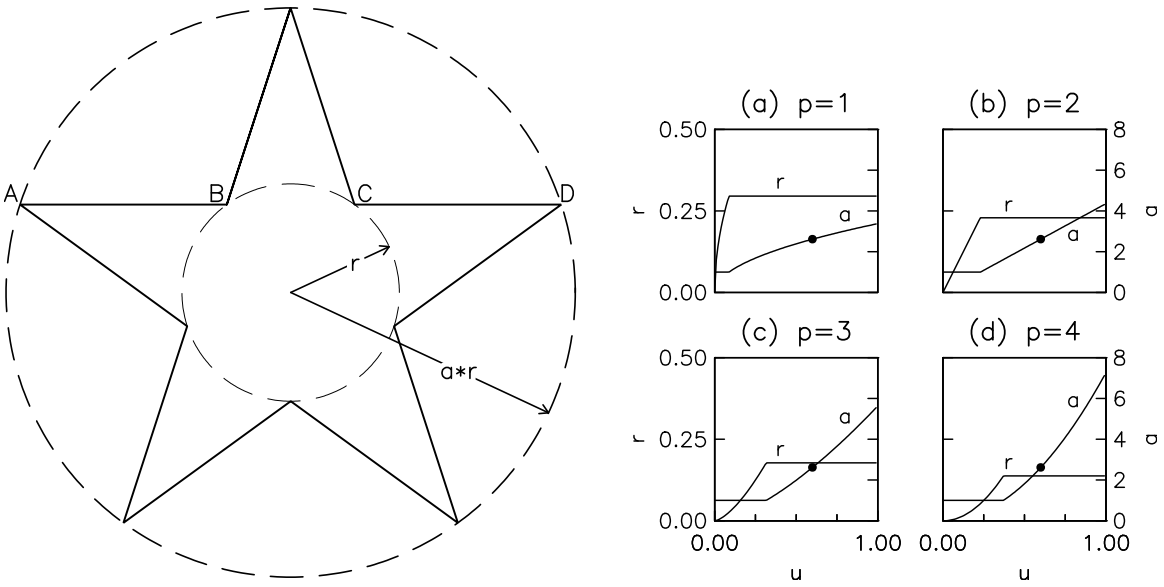

Figure 1: Figure showing (left) the construction of the star symbol, defined by the number of points $n = 5$, an inner circle of radius r and outer circle of radius ar. In this example a is chosen to make the line segments AB and CD collinear, which for $n = 5$ is $a = 2.618$. On the right, the four panels (a) through (d) show the variation in r and a as a function of the size variable u, for the size of the circumscribing polygon varying as u^p. In each plot the left side gives the scale for a and the right side the scale for r. In all four plots, a is required to be 2.618 for $u = 0.6$ (black dot).

A few parameters describe how the shape and size of these symbols depends on the value they represent. For two choices of parameters I have tested how well the values can be estimated for isolated symbols; this test suggests that viewers can estimate the true value to within 5% of the total range covered.

7 Acknowledgments

I thank those who came to the IGPP afternoon tea on October 29, 2013, for being willing to spend their tea-time on the unexpected task of filling the forms whose contents are summarized in Section 4. I also thank Andrew Gelman, Howard Wainer, Daniel Montello, and Sara Fabrikant for comments and references.

This research was supported by the Southern California Earthquake Center. This paper is SCEC contribution 1931.

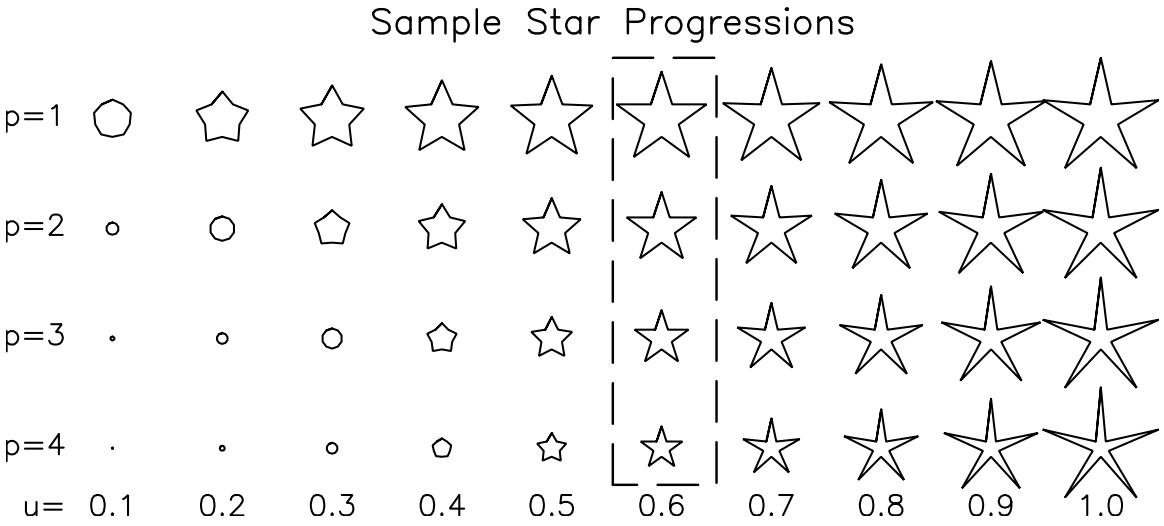

Figure 2: Progressions of star symbols corresponding to the variations of r and a shown in panels (a) through (d) of Figure 1, with the "perfect star" shape always being at $u = 0.6$. Note that because the area of the circumscribing polygon varies as u^p, the radius varies as $u^{p/2}$; so for $p = 1$ the radius varies as $\sqrt{u}$.

References

Agnew, D. C. (2004), Robert FitzRoy and the myth of the 'Marsden square': Transatlantic rivalries in early marine meteorology, *Notes Records Roy. Soc.*, **58**, 21–46.

Allen, C. R., P. S. Amand, C. F. Richter, and J. M. Nordquist (1965), Relationship between seismicity and geologic structure in the southern California region, *Bull. Seismol. Soc. Am.*, **55**, 753–798.

Argelander, F. W. (1843), *Uranometria Nova: Stellae per Mediam Europam Solis Oculis Conspicuae Secundum Veras Lucis Magnitudines e Coelo Ipso Descriptae*, Simon Schropp, Berlin.

Bertin, J. (1983/2010), *Semiology of graphics*, (W. J. Berg, trans.) ESRI Press, Redlands, California.

Brewer, C. A. (2008), *Designed Maps: a Sourcebook for GIS Users*, ESRI Press, Redlands, Calif.

Brudzinski, M. R., and W.-P. Chen (2003), Visualization of seismicity along subduction zones: Toward a physical basis, *Seismol. Res. Lett.*, **74**, 731–738, doi:10.1785/gssrl.74.6.731.

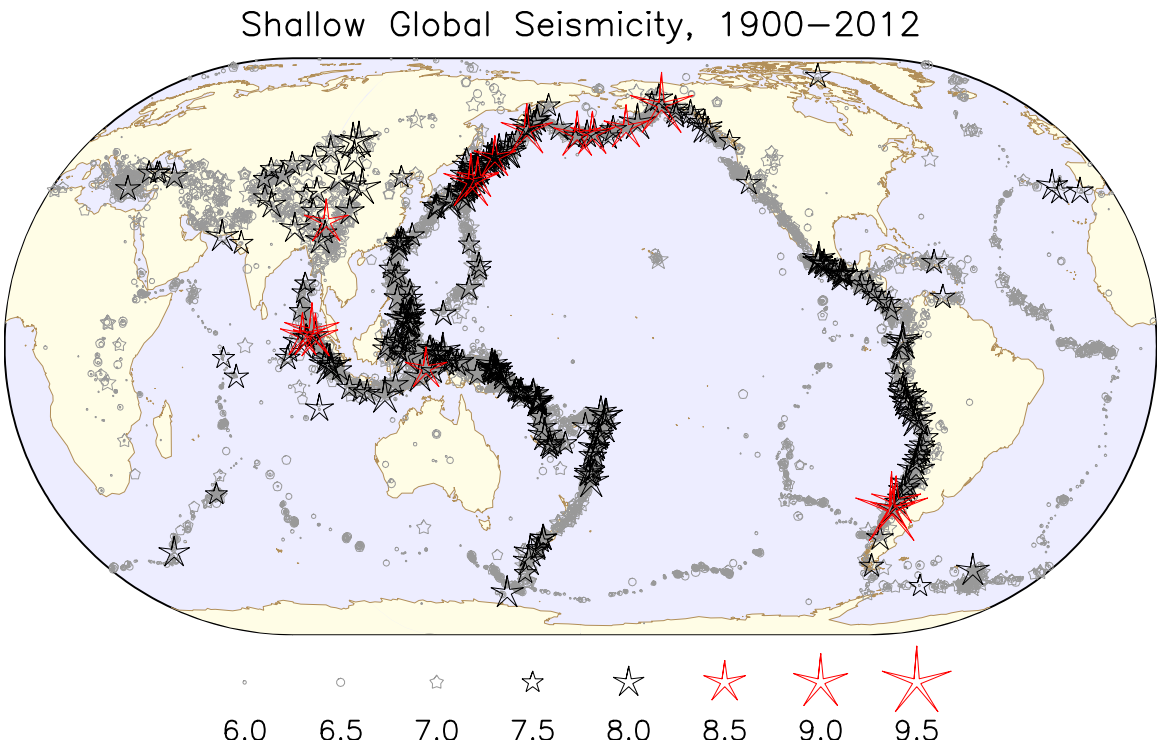

Figure 3: Global seismicity from 1900 through 2012, for shallow earthquakes (depth 70 km or less) with $M_W ge 6.0$ on an equal-area projection (Eckart IV). The data for 1900-2009 inclusive is from the ISC-GEM Global Instrumental Earthquake Catalogue (Storchak *et al.*, 2013), with the years 2010-2012 taken from the NEIC catalog. The symbols and colors used for different magnitudes are shown below the map.

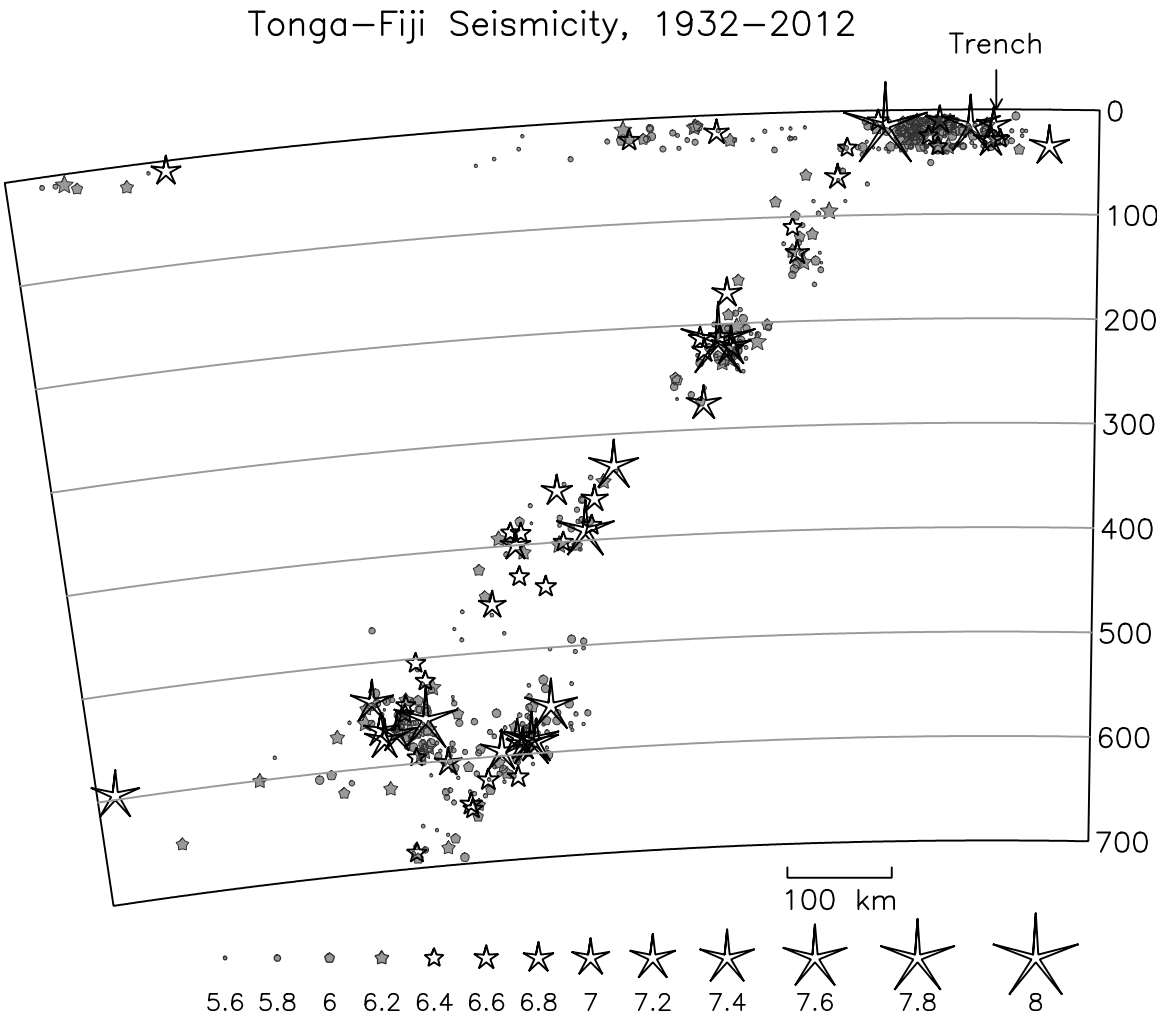

Figure 4: Seismicity of the Tonga-Fiji subduction zone, 1932 through 2012 from the same catalogs used in Figure 3. The earthquakes are projected onto a plane striking perpendicular to the Tonga Trench, and passing through 19.343°S 172.986°W; all events from 3° southerly of this plane, to 1° northerly were included.

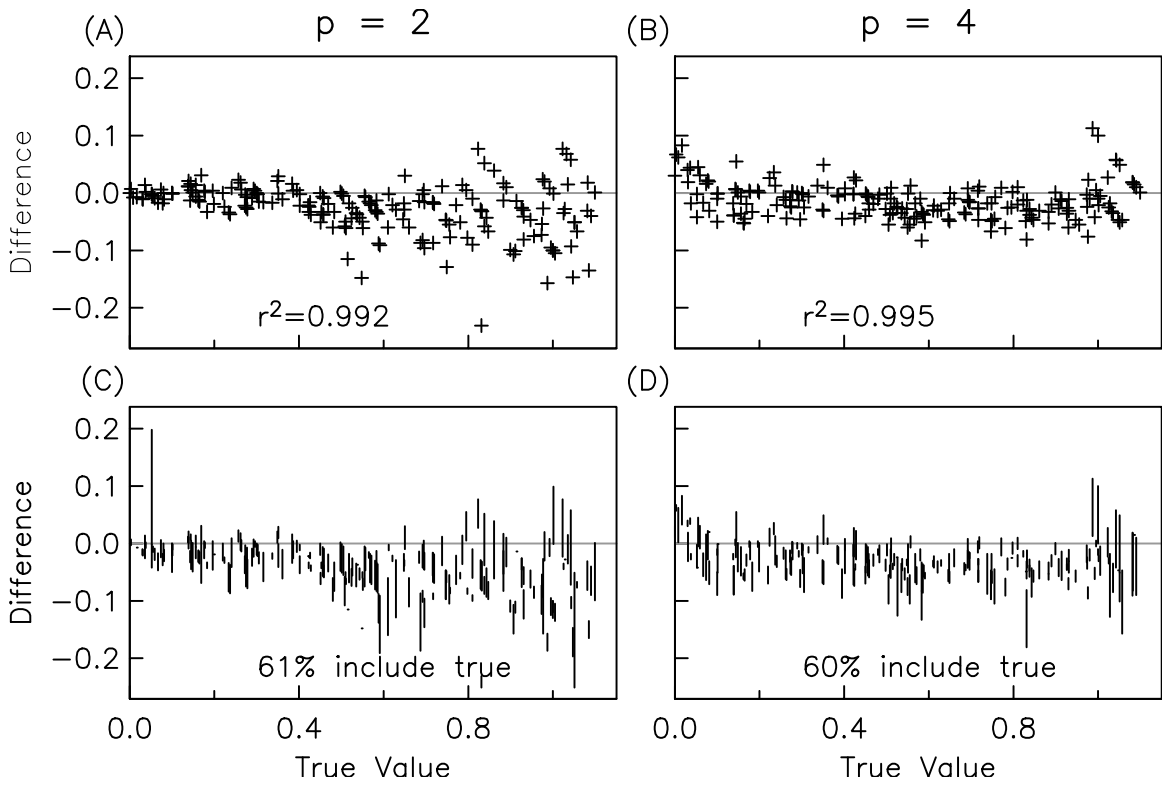

Figure 5: Tests of how accurately the variable stars indicate size. Panels A and B show the ratio between estimated values (uniformly distributed between zero and one) and true values, for two choices of the parameter *p*. Panels C and D show the same ratio for the range of values estimated to be possible, indicated by lines. See text for details of the test.

Cleveland, W. S. (1993), A model for studying display methods of statistical graphics, *J. Comput. Graph. Stat.*, **2**, 323–343, doi:10.2307/1390686.

Cleveland, W. S., and R. McGill (1984), Graphical perception: theory, experimentation, and application to the development of graphical methods, *J. Am. Stat. Assoc.*, **79**, 531–554.

Dent, B. (1996), *Cartography: Thematic Map Design*, W. C. Brown.

Herlihy, A. F. (2007), Renaissance star charts, in *The History of Cartography, Volume Three: Cartography in the European Renaissance*, edited by D. Woodward, pp. 99–122, University of Chicago Press, Chicago.

Klippel, A., F. Hardisty, and C. Weaver (2009), Star plots: How shape characteristics influence classification tasks, *Cartogr. Geog. Infor. Sci.*, **36**, 149–163.

Krzywinski, M., and B. Wong (2013), Points of view: Plotting symbols, *Nature Methods*, **10**, 451–451, doi:10.1038/nmeth.2490.

Lewandowsky, S., and I. Spence (1989), Discriminating strata in scatterplots, *J. Amer. Stat. Assoc.*, **84**, 682–688, doi:10.1080/01621459.1989.10478821.

MacEachren, A. H. (1995), *How Maps Work: Representation, Visualization, and Design*, Guilford Press, New York.

Mallet, R. (1858), Fourth report on the facts of earthquake phenomena, *Brit. Assoc. Adv. Sci. Ann. Rep.*, **28**, 1–136.

Mersey, J. E. (1996), Cartographic symbolization requirements For microcomputer-based geographic information systems, in *Cartographic Design: Theoretical and Practical Perspectives*, edited by C. P. Keller and C. H. Wood, pp. 157–175, John Wiley & Sons, New York.

Montello, D. (2002), Cognitive map-design research in the twentieth century: theoretical and empirical approaches, *Cartogr. Geogr. Inform. Sci.*, **29**, 283–304.

Robinson, A. H., P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill (1995), *Elements of Cartography*, Wiley, New York.

Storchak, D. A., G. DiGiacomo, I. Bondár, E. R. Engdahl, J. Harris, W. H. K. Lee, A. Villaseñor, and P. Bormann (2013), Public release of the ISC-GEM Global Instrumental Earthquake Catalogue (1900-2009), *Seismol. Res. Lett.*, **84**, 810–815, doi:10.1785/0220130034.

Tirion, W., B. Rappaport, and W. Remaklus (2001), *Uranometria 2000.0: Deep Sky Atlas, Vol. 1: The Northern Hemisphere to -6 Degrees*, Willmann-Bell, Richmond, Virginia.

Tremmel, L. (1995), The visual separability of plotting symbols in scatterplots, *J. Comp. Graph. Stat.*, **4**, 101–112.

von_Littrow, J. J. (1854), *Atlas des Gestirnten Himmels für Freunde der Astronomie*, Hoffmann, Stuttgart.

Wainer, H. (1997), Some multivariate displays for NAEP results, *Psychol. Meth.*, **2**, 34–63.

Wessel, P., and W. H. F. Smith (1991), Free software helps map and display data, *EOS Trans. AGU*, **72**, 441.

Williams, R. L. (1956), *Statistical Symbols for Maps: Their Design and Relative Value*, 114 pp., Map Library, Yale University, New Haven.