# Depth Analysis of Directional Data

Claudio Agostinelli and Mario Romanazzi

**Abstract** We illustrate a depth-based approach in directional statistics, concentrating on a prominent depth function, angular simplicial depth. A new local version of this function is proposed and several depth-based summaries, including angular medians, depth regions and dispersion parameters are considered.

**Key words:** Data depth, Angular simplicial depth, Depth regions, Depth-based measures of dispersion, Local depth.

## 1 Local Angular Simplicial Depth

Nonparametric analysis of directional data received an innovative contribution by Liu and Singh (1992), who suggested the use of depth methods. In particular, they proposed new depth functions — angular simplicial depth (ASD) and angular Tukey's depth (ATD)— tailored to fit the geometry of directional data, i. e., points lying on the surface of (hyper)spheres. In this paper we concentrate on ASD, propose a local version for it and illustrate two applications with real data. Broad presentations of data depth are Liu et al (1999) and Zuo and Serfling (2000).

When defining ASD, the standard simplices of $\mathbb{R}^p$ are replaced by *spherical simplices*, whose edges are the shortest arcs connecting the vertices. The shortest arcs belong to the great circles joining all pairs of vertices and having the same center as the (hyper)sphere $\mathbb{C}_{p+1} = \{x \in \mathbb{R}^{p+1} : x^T x = 1\}$, $p \geq 1$. We write $SS_{p+1}(X) = SS_{p+1}(X_1, X_2, \cdots, X_{p+1})$ for a random spherical simplex, whose ver-

Claudio Agostinelli
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, e-mail: claudio@unive.it

Mario Romanazzi
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, e-mail: romanaz@unive.it

tices are $p+1$ independent copies $X_1, X_2, \cdots, X_{p+1}$ of the random variable $X$ with support on the sphere. As in the Euclidean case, ASD is the coverage function of a random spherical simplex (Liu and Singh, 1992), that is

$$ad_X(x) = \mathrm{P}_X(SS_{p+1}(X) : \ x \in SS_{p+1}(X)) \ .$$

Let $x_i, i = 1, \cdots, n$ be a random sample from $X$. The empirical version $ad_n(.)$ is the proportion of all $\binom{n}{p+1}$ sample spherical simplices including $x$. A main requirement for any depth function is ray-monotonicity (Zuo and Serfling, 2000), meaning that it does not increase along the rays from the deepest point. It implies the depth function to have just one (global) maximizer, even when the distribution is multimodal. This is coherent with the aim of finding a global center. However, in certain situations it is desirable to obtain more information. Local depth (Agostinelli and Romanazzi, 2011) is a generalization of the depth function able to recognize partial centers and groups. It is mainly useful when dealing with multimodal and mixture distributions and clustered data. The basic idea is to measure depth conditional on the nearby region of each point. As in the Euclidean case, a local version for ASD is obtained by constraining the size of the spherical simplices not to exceed a positive threshold $\tau$. For directions $x \in \mathbb{C}_{p+1}$, local ASD $lad_X(.;\tau)$ is the coverage probability by a random spherical simplex $SS_{p+1}(X)$ with surface area $\mathrm{t}(SS_{p+1}(X))$ not greater than $\tau$, that is

$$lad_X(x;\tau,F) = \mathrm{P}_X\left(SS_{p+1}(X) : \ x \in SS_{p+1}(X) \cap \mathrm{t}(SS_{p+1}(X)) \leq \tau\right) \ .$$

Other function, in place of the surface area, can be use to measure the size of the spherical simplex. Global ASD is recovered by fixing $\tau$ equal to its maximum value. It can be shown that, in the circular case, local and global ASD behave similarly when $X$ is unimodal.

Together with the values $d_X(x)$ $(d_n(x))$, describing the ordering induced on the space by the depth function, several functionals (sample statistics) associated to $d_X$ are used to investigate the features of the distribution. For example, in the bivariate case, the depth contours help investigating location, spread and shape of the underlying distribution or data. A related functional is the family of depth regions $D_X(\alpha)$, containing the deepest points corresponding to fraction $\alpha$ of cumulative probability. The boundaries of such regions are often interpreted as $\alpha$-level quantile surfaces. Depth-based location and dispersion functionals were also suggested that represent alternatives to the usual moment-based functionals. The location functional is the deepest point and the dispersion functional is the Lebesgue integral of the depth function. The depth-based location parameter (Liu et al, 1999) of $X$ is $\lambda_X = \arg\max_{x \in \mathbb{C}_{p+1}} ad_X(x)$ with corresponding sample statistic $\widehat{\lambda}_n = \arg\max_{x \in \mathbb{C}_{p+1}} ad_n(x)$. Writing $\nu$ for the Lebesgue measure on $\mathbb{C}_{p+1}$, the depth-based dispersion parameter (Romanazzi, 2009) of $X$ is $\gamma_X = \int_{\mathbb{C}_{p+1}} ad_X(x) d\nu(x)$, to be interpreted as the expected measure of the spherical simplex $SS_{p+1}(X)$. The sample statistic is $\widehat{\gamma}_n = \int_{\mathbb{C}_{p+1}} ad_n(x) \ d\nu(x)$. An approximation of $\widehat{\gamma}_n$ is given by $\widetilde{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} ad_n(x_i)$.

## 2 Applications

The first illustration deals with the direction taken by 76 turtles, after treatment (Fisher, 1993, Appendix B3.). The circular mean and circular median coincide at 1.12, circular variance is 0.5 and $\gamma = 1.27$. The raw data clearly suggest some degree of multimodality. ASD correctly identifies the main center of the distribution at 1.10, but fails to show the modes. Instead, local ASD, with $\tau$ equal to 0.2 quantile order, shows (mainly) two groups center at 1.10 and 1.57 (see Figure 1). A further mode is at 4.24. The behaviour is stable for a wide range of $\tau$, from 0.1 to 0.4 quantile orders.

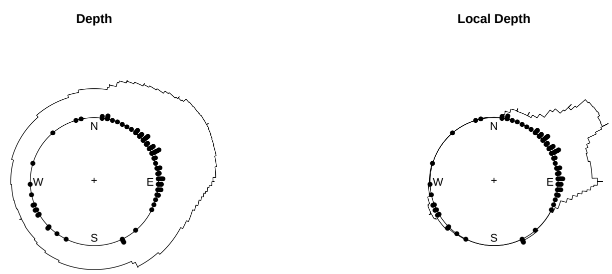


**Fig. 1** Turtle data. Circular plot with ASD (left) and local ASD (right).

The second application deals with locations of earthquakes with modified Mercalli intensity greater than 6, since 1900 (source: `www.ngdc.noaa.gov/`). The number of events is $n = 284$.

The spherical mean is $(72.4, 57.1)$ (latitude, longitude), the spherical median is $(56.7, 67.8)$, and the deepest point is $(44.7, 37.4)$ while the dispersion parameter based on depth is $\widetilde{\gamma} = 0.123$. Figure 2 provides the map of the dataset, where gray scale of points corresponding to the value of ASD (top panel) and local ASD (bottom panel). Darker color means higher depth value. The threshold $\tau$ is equal to 20% quantile of the areas of the spherical simplices. Contours of (normalized) depth are also shown for values 0.95, 0.8 and 0.6. Deepest point, spherical mean and spherical median are represented by square, diamond and triangle symbol, respectively.

The main result is the ordering provided by local ASD, which is remarkably different from ASD ordering. In the bottom map of Figure 2 several clusters are indeed apparent, the two most important corresponding to Greece – Turkey and South-East China – Philippines. Also remarkable the difference between location estimates. While the deepest point, as expected, belongs to the Greece – Turkey cluster, the spherical mean and median are very far from it and their use is questionable in the present application.

## References


Agostinelli C, Romanazzi M (2011) Local depth. Journal of Statistical Planning and Inference 141:817–830

Fisher NI (1993) Statistical Analysis of Circular Data. Cambridge University Press, Cambridge

Liu RY, Singh K (1992) Ordering directional data: concepts of data depth on circles and spheres. The Annals of Statistics 20:1468–1484

Liu RY, Parelius JM, Singh K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. The Annals of Statistics 27:783–858

Romanazzi M (2009) Data depth, random simplices and multivariate dispersion. Statistics & Probability Letters 79:1473–1479

Zuo Y, Serfling R (2000) General notions of statistical depth function. The Annals of Statistics 28:461–482


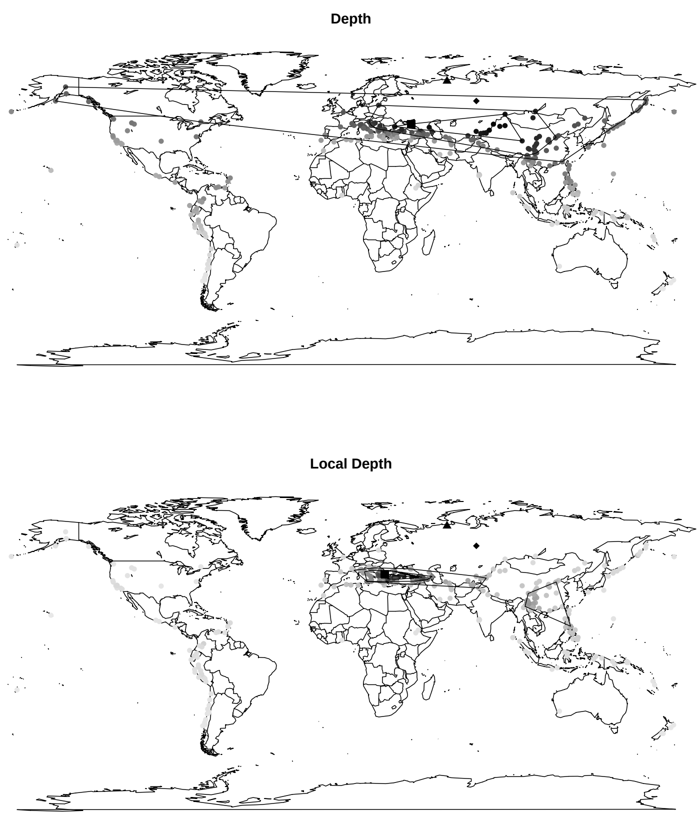


**Fig. 2** Earthquake data. Map with ASD (top panel) and local ASD (bottom panel) values.