A scale space approach for exploring structure in spherical data

Ville Vuollo a,b,c,*, Lasse Holmström c

a Research Unit of Oral Health Sciences, Faculty of Medicine, University of Oulu, Oulu, Finland
b Medical Research Center, Oulu University Hospital, Oulu, Finland
c Research Unit of Mathematical Sciences, Faculty of Science, University of Oulu, Oulu, Finland

HIGHLIGHTS

• A novel scale space technique for analyzing spherical data is proposed.
• Distributions of normal vector directions computed from a 3D head image are analyzed.
• A movie is a convenient way to explore the maps included in a SphereSiZer atlas.

ABSTRACT

A novel scale space approach, SphereSiZer, is proposed for exploring structure in spherical data, that is, directional data on the unit sphere of the three-dimensional Euclidean space. The method finds statistically significant gradients of the smooths of the probability density function underlying the observed data. Bootstrap is used to establish significance and inference is summarized with planar maps of contour plots of smooths of the data, overlaid with arrows that indicate the directions and magnitudes of the significant gradients. An effective way to explore such maps is a movie where each frame corresponds to a fixed level of smoothing, that is, a particular spatial scale on the sphere. The SphereSiZer is demonstrated using simulated data as well as two real-data examples. The first example examines the distribution of infant head normal vector directions. The presence of local maxima in the normal vector distribution may indicate head deformity, such as severe flatness or asymmetry. The second example considers the distribution of earthquakes in the Northern Hemisphere.

1. Introduction

Statistical scale space analysis provides useful tools for finding salient features in the data. The goal in particular is to discover structure underlying the data in different scales. This is typically achieved by applying different levels of smoothing to the observed data. The object of the analysis can be for example a probability density or regression function, a time series, a digital image or a more general random field (Holmström and Pasanen, 2017). The origin of scale space theory can be traced back to the 1950s and it later gained popularity in computer vision research (Lindeberg, 1994). Then, some 20 years ago, Chaudhuri and Marron applied scale space ideas to statistical function estimation (Chaudhuri and Marron, 1999). The underlying idea in their SiZer methodology was, instead of considering one ‘optimal’ smooth, to explore simultaneously a
whole family of different smooths and establish their statistically significant features. This approach side-stepped the need to select an optimal level of smoothing and at the same time avoided the ever-present problem of bias in nonparametric function estimation. Another innovation was the SiZer map which provides an easily interpreted summary of scale space inference. Since the seminal work of Chaudhuri and Marron, statistical scale space techniques have developed into a rich palette of techniques that has found applications in a wide range of fields (Holmström, 2010; Holmström and Pasanen, 2017).

In directional statistics, the sample space is the unit sphere (Mardia, 1972; Fisher et al., 1987). Analysis of such data occurs naturally e.g. in the earth sciences, meteorology, physics and biology. Recently, Oliveira et al. (2014) and Huckemann et al. (2016) proposed versions of SiZer for circular data, that is, directional data on the unit circle. The method described in Oliveira et al. (2014) is called CircSiZer and it applies kernel estimation with the von Mises density as the kernel function. Huckemann et al. (2016) pointed out that smoothing with the von Mises kernel does not necessarily satisfy the causality axiom of scale space smoothing which stipulates that increasing smoothing should not create new local extrema in the analyzed function. To achieve causality, their WiZer (Wrapped SiZer) therefore used the wrapped Gaussian kernel instead of the von Mises kernel.

This article describes SphereSiZer, a scale space technique for the analysis of spherical data, that is, directional data on the unit sphere of the three dimensional Euclidean space. Based on a random sample of such data, the goal is to discover the features of the probability density function of the distribution from which the data arose. The salient features of the density function are discovered by finding the locations on the sphere where the gradient of density surface differs significantly from zero. The role of a SiZer map is played by an area preserving projection onto a planar disk on which a contour plot of a smooth of the data and the significant gradients are displayed. Our approach is therefore similar to the $S^3$ methodology of Godtliebsen et al. (2002) used for analyzing a bivariate density function on a plane. However, since our main interest is only to find the significant modes of the density, unlike in $S^3$, we will not consider local curvature of the density. Like CircSiZer, SphereSiZer uses the von Mises–Fisher kernel so that strict scale space causality is not guaranteed. As noted in Holmström and Pasanen (2017), in many statistical scale space methodologies the useful properties of a particular smoother are often deemed to be more important than strict adherence to the scale space axioms. We also note that the name SphereSiZer is a bit of a misnomer since ‘SiZer’ is an acronym for ‘Significant ZERO crossings of derivatives’ and in the two-dimensional setting of SphereSiZer, instead of the sign of the derivate, magnitude and direction of the gradient are of interest. However, the basic idea in SphereSiZer is analogous to that of the original SiZer, justifying our choice for its name.

Both simulated and real data are used to demonstrate the properties of SphereSiZer. The first real data example considers head normal vector directions of a 3D model of an infant’s head. A local maximum in the density of the normal vector directions can indicate the presence of a flat region or asymmetry in the head. In the second example, the distribution of the spatial locations of earthquakes in the Northern Hemisphere is analyzed with SphereSiZer.

Our scale space approach for spherical data, including the construction of the associated SphereSiZer maps, is described in Section 2. Section 3 gives examples of analyses of directional data and the article concludes with a summary together with some questions for future work in Section 4. Most figures in the printed article are reproduced in black and white; for all color figures, see the online version of the paper. Supplementary material to this article includes movies of the examples discussed in Section 3.

2. SphereSiZer

2.1. Kernel density estimation

Consider the unit sphere $S^2 = \{x \in \mathbb{R}^3 \mid ||x|| = 1 \}$ of the 3-dimensional Euclidean space and a directional random vector $X \in S^2$ with a density $f$. Given a sample $X_1, \ldots, X_n \sim f$, the kernel estimator of $f$ can be defined as

$$
\hat{f}(x; \kappa) = \frac{c(\kappa)}{n} \sum_{i=1}^{n} K(\kappa x^T X_i), \quad x \in S^2, \quad (1)
$$

where $K$ is a suitable kernel function and $\kappa > 0$ is the smoothing parameter. The normalization constant $c(\kappa)$ is chosen so that the estimator integrates to one. Good introductions to ordinary Euclidean and spherical kernel smoothing are Silverman (1986); Wand and Jones (1994); Hall et al. (1987). For thorough general introductions to directional statistics, see Mardia (1972) and Fisher et al. (1987).

According to Hall et al. (1987), the kernel $K$ in (1) should be a rapidly varying function, such as the exponential. Then a small value of $\kappa$ corresponds a smooth estimate and a large value produces a rough, variable result. With $K(t) = e^t$ the formula (1) then becomes

$$
\hat{f}(x; \kappa) = \frac{c(\kappa)}{n} \sum_{i=1}^{n} e^{\kappa x^T X_i} = \frac{\kappa}{4n\pi \sinh(\kappa)} \sum_{i=1}^{n} e^{\kappa x^T X_i}, \quad x \in S^2. \quad (2)
$$

The von Mises-Fisher distribution on $S^2$ can be considered as the spherical counterpart of the circularly symmetric bivariate normal distribution and it has the density

$$
f_{\text{vMF}}(x; \mu, \kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} e^{\kappa x^T \mu}, \quad x \in S^2, \quad (3)
$$
where \( \mu \) is the directional mean and \( \kappa > 0 \) is the concentration parameter. Therefore, the kernel density estimate (2) is in fact an arithmetic mean of von Mises-Fisher densities with different sample units as mean directions,

\[
\hat{f}(x; \kappa) = \frac{1}{n} \sum_{i=1}^{n} f_{\text{vMF}}(x; X_i, \kappa).
\]

(4)

2.2. Feature significance in the scale space

Taking expectation in (2),

\[
f(x; \kappa) = \mathbb{E}[\hat{f}(x; \kappa)] = c(\kappa)\mathbb{E}\left[e^{\kappa x^T y}\right]
\]

\[
= c(\kappa)\int_{S^2} e^{\kappa x^T y} f(y) d\sigma(y) = c(\kappa)(K_{\kappa} * f)(x),
\]

(5)

where \( \sigma \) is the Lebesgue measure on \( S^2 \). The convolution notation \( K_{\kappa} * f \) is justified by its Euclidean analog where, instead of \( x^T y \), the distance between \( x \) and \( y \) is measured by their difference. Clearly, \( f(\cdot; \kappa) \) is a smooth of \( f \) and we regard \( \{ f(\cdot; \kappa) \mid \kappa > 0 \} \) as its scale space representation. By (5), \( \{ \hat{f}(\cdot; \kappa) \mid \kappa > 0 \} \) is an unbiased estimate of this representation.

We want to explore the features of the smooth \( f(\cdot; \kappa) \) using its gradient. This poses a problem since \( f(x; \kappa) \) is defined only for \( x \in S^2 \). Here we follow the approach of Hall et al. (1987) and extend \( f(\cdot; \kappa) \) beyond \( S^2 \) before calculating its gradient. Thus, for \( x \in \mathbb{R}^3 \setminus \{0\} \), let \( g(x) = f(x/\|x\|) \) and

\[
\hat{g}(x; \kappa) = \hat{f}(x/\|x\|; \kappa) = \frac{c(\kappa)}{n} \sum_{i=1}^{n} e^{\kappa(x/\|x\|)^T X_i}.
\]

(6)

Similarly to (5), define

\[
g(x; \kappa) = \mathbb{E}[\hat{g}(x; \kappa)] = c(\kappa)(K_{\kappa} * f)(x/\|x\|)
\]

so that \( g(\cdot; \kappa) \) is the extension of the smooth \( f(\cdot; \kappa) \). We will therefore use \( \nabla \hat{g}(\cdot; \kappa) \) as a test statistic to discover the statistically significant features of \( f(\cdot; \kappa) \). Note that for \( x \in S^2 \),

\[
\mathbb{E} \nabla \hat{g}(x; \kappa) = \nabla \mathbb{E} \hat{g}(x; \kappa) = \nabla g(x; \kappa)
\]

so that \( \nabla \hat{g}(x; \kappa) \) is an unbiased estimator of the quantity of interest, the gradient of the extended version of \( f(x; \kappa) \).

To test the hypothesis

\[ H_0 : \nabla g(x; \kappa) = 0 \]

against

\[ H_1 : \nabla g(x; \kappa) \neq 0 \]

we first estimate the quantile \( q \) for which

\[
P\left\{ \sum_{i=1}^{3} \left[ \frac{D_i \hat{g}(x; \kappa) - D_i \hat{g}(x; \kappa)}{\text{SD}(D_i \hat{g}(x; \kappa))} \right]^2 < q \right\} = 1 - \alpha,
\]

where \( D_i \) is the \( i \)th partial derivative, SD denotes standard deviation and \( 1 - \alpha \) is the confidence level. Here \( D_i \hat{g}(x; \kappa) \) is proportional to a mean of \( n \) independently identically distributed terms so that its standard deviation can be estimated as in Chaudhuri and Marron (1999),

\[
\text{SD}(D_i \hat{g}(x; \kappa)) = \sqrt{\frac{\sum_{i=1}^{n} D_i e^{\kappa(x/\|x\|)^T X_i}}{n}} = \frac{c(\kappa)}{\sqrt{n}} s \left( D_1 e^{\kappa(x/\|x\|)^T X_1}, \ldots, D_3 e^{\kappa(x/\|x\|)^T X_3} \right),
\]

where \( s \) denotes the sample standard deviation. The null hypothesis is then rejected if

\[
\sum_{i=1}^{3} \left[ \frac{D_i \hat{g}(x; \kappa)}{\text{SD}(D_i \hat{g}(x; \kappa))} \right]^2 \geq q.
\]

(7)

2.3. Pointwise and simultaneous inference

For the original SiZer, several methods for obtaining the quantile \( q \) were proposed: a pointwise Gaussian quantile, an approximate simultaneous quantile based on a number of independent blocks of data, and two quantiles based on the
bootstrap (Chaudhuri and Marron, 1999). Hannig and Marron (2006) proposed a better inference method for SiZer in curve estimation that uses extreme value theory for Gaussian processes to approximate the distributions underlying the statistical inference needed in testing the significance of features in multiple locations simultaneously. Their approach has subsequently become the inference tool of choice for SiZer-like methods in curve estimation (Holmström and Pasanen, 2017). However, for simultaneous inference, both the extreme value distribution theory approach and the independent blocks method appear to be difficult to implement for spherical density estimation. We therefore propose to obtain all the required quantiles with a bootstrap-technique (cf. Efron and Tibshirani, 1993). Bootstrap-technique was also used in CircSiZer for pointwise quantile estimation.

The SphereSiZer map that summarizes the inference about significant gradients is constructed on an equispaced Fibonacci grid $G = \{x_1, \ldots, x_l\}$ on $S^2$ (Swinbank and Purser, 2006). Fig. 1 shows an example of a Fibonacci grid with $l = 2001$ points.

Further, as in Chaudhuri and Marron (1999), estimated effective sample size (ESS) is used to exclude from the analysis points around which data are too sparse for meaningful inference. ESS is calculated as

$$ESS(x, \kappa) = \frac{\sum_{i=1}^{n} K(\kappa x^T X_i)}{K(\kappa)}$$

and inference will only be considered in

$$A_\kappa = \{x : ESS(x, \kappa) \geq 5\}.$$  

Therefore, for a fixed $\kappa$, inference is considered for the grid points in the set $H_\kappa = G \cap A_\kappa$.

To apply bootstrap-technique, we generate bootstrap samples $\{X_1^{b}, \ldots, X_n^{b}\}, b = 1, \ldots, B$, by sampling with replacement from the original data set $\{X_1, \ldots, X_n\}$. Denote by $\hat{g}^{b}$ the kernel density estimate (3) based on the bootstrap sample $\{X_1^{b}, \ldots, X_n^{b}\}$ and let

$$Z_b(x; \kappa) = \sum_{i=1}^{3} \left[ \frac{D \hat{g}^{b}(x; \kappa) - D \hat{g}(x; \kappa)}{SD(D \hat{g}^{b}(x; \kappa))} \right]^2, \quad b = 1, \ldots, B. \quad (8)$$

The simplest type of inference is pointwise, where each $x \in H_\kappa$ is considered independently. The quantile $q$ in (7) is then the corresponding empirical quantile of the values $Z_1(x; \kappa), \ldots, Z_B(x; \kappa)$. As pointwise inference is prone to false positives, more useful results can be obtained by using simultaneous inference over all locations $H_\kappa$, or both all locations and all smoothing levels $\kappa$. For a fixed $\kappa$, the appropriate quantile $q$ for inference which is simultaneous over all locations is obtained by the empirical quantile of $\max_{x \in H_\kappa} Z_b(x; \kappa), b = 1, \ldots, B$, and the quantile for inference which is simultaneous over both locations and smoothing levels is obtained from the empirical quantile of $\max_{\kappa \in K} \max_{x \in H_\kappa} Z_b(x; \kappa), b = 1, \ldots, B$. 

Fig. 1. A Fibonacci grid of 2001 points.
Fig. 2. Estimates of the actual significance level of SphereSiZer as a function of the smoothing level, computed as the percentage of rejections of the null hypothesis in 1000 trials. (a) Pointwise inference. (b) Inference simultaneous over x. The red line shows the significance level $\alpha = 0.05$.

2.4. SphereSiZer maps

Visualization of the spherical kernel density estimate is done by using a projection onto a planar contour plot (see Vuollo et al., 2016). The Fibonacci grid is mapped on a planar disk using the Lambert azimuthal equal area projection (Fisher et al., 1987) and the contour plot is computed using the values of the density estimate at the projected grid points. In Lambert projection, the spherical coordinates $\theta$ (azimuthal angle) and $\varphi$ (polar angle) are transformed to polar coordinates $\rho$ and $\psi$ by $\rho = 2 \sin((\pi - \varphi)/2)$ and $\psi = \theta$, where $0 \leq \theta \leq 2\pi$ and $0 \leq \varphi \leq \pi$. Distortion in the projection increases with distance from the center of the projection disk. This means that the shape of the density in the southern hemisphere, and in particular near the southern pole ($\varphi = \pi$), is somewhat difficult to interpret as the projections of these areas lie near the boundary of the disk. We therefore divide the sphere into two hemispheres and project them separately onto two disks for both of which $0 \leq \rho \leq \sqrt{2}$. However, if all the data are located in just one hemisphere, only one projection disk is used. Finally, the significant gradients are plotted on the contour plot of the kernel density estimate, visualized by arrows whose relative lengths correspond to the actual gradient magnitudes. Strictly speaking, the projection of a gradient vector is of course curved but we ignore this in our maps because, especially after scaling the arrow lengths, the visual difference between a straight arrow and a curved one is negligible. An example of a SphereSiZer map is shown in Fig. 3. A collection of several such maps is called a SphereSiZer atlas and an example is shown in Fig. 5. A convenient way to view an atlas that consists of a large number of maps is a movie (see the on-line supplement to this article). Matlab R2016b was used for programming all SphereSiZer analyses.

2.5. Assessment of the actual significance level of SphereSiZer

We examined the performance of the bootstrap based quantiles using simulation. To find out the actual significance level of our testing procedure, we applied SphereSiZer to data generated under the null hypothesis and estimated the probability of rejecting the null hypothesis of a zero gradient. For the both types of simultaneous inference considered, the null hypothesis corresponds to the uniform distribution probability density function $f(x) = 1/4\pi, x \in S^2$. Strictly speaking, this is true when all points of $S^2$ are involved in simultaneous inference but the finite, dense Fibonacci grid used in the actual tests approximates this scenario. In pointwise inference the null hypothesis is composite, as it includes any density with a vanishing gradient at the particular point considered so, in this case, a lower bound for the significance level is obtained.

The nominal significance level used in simulations was $\alpha = 0.05$ so, ideally, in pointwise inference only 5% of the gradients at a fixed location should be significant. Similarly, for inference which simultaneous only over $x$, only 5% of the maps corresponding to a fixed smoothing level $\kappa$ should have at least one significant gradient. For inference which is simultaneous over both $x$ and $\kappa$, only 5% of the atlases should have at least one significant gradient. In simulations, a random sample of size $n = 1000$ was generated from the uniform distribution (Shao and Badler, 1996). This was repeated 1000 times and the significance level of a test was estimated as the fraction of times for which a non-zero gradient was inferred. The grid of smoothing levels had 40 values between 0.1 and 100 and the number of bootstrap samples in each test was $B = 500$.

Theoretically, in pointwise inference, only one particular location $x \in S^2$ needs to be considered since from the point of view of the uniform distribution, all locations are equal. However, due to the finite size of our simulation study, small differences were observed between locations and we therefore estimated the significance level for pointwise inference as an average of estimates obtained at 10 locations. Including more estimates in the average did not change the result displayed in the Fig. 2a. As noted above, pointwise inference is likely to lead to false positives and we therefore do not recommend it.
Fig. 3. An artificial test density and a SphereSiZer map on the complete sphere. (a) A contour plot of the test density and the 200 random sample points used in SphereSiZer analysis. (b) A SphereSiZer map based on this random sample. The smoothing parameter $\kappa = 25$. In the contour plots, whiter shading indicates higher density values. Separate projection disks are used for the northern and southern hemispheres. The azimuthal angle is marked on the boundary of the disks and the polar angle is shown in their interior. Green arrows show statistically significant gradients.

The assessment of its actual significance level also shows that for very small scales (here $\kappa > 53$), the pointwise test breaks down and significant gradients suggested by it should not be trusted. The threshold value of $\kappa$ for which this breaking down occurs decreases with the sample size (results not reported), so the general advice is to be careful when interpreting small scale results in pointwise inference.

The significance level of simultaneous inference was estimated using a Fibonacci grid of 1001 points. Fig. 2b summarizes the results for the inference simultaneous only over $x$. The level is close to its nominal value for small values of the smoothing parameter (larger scale features) but begins to decrease when $\kappa$ increases (smaller scale features), dropping to 0 at $\kappa = 46$. This means in particular that when, in inference simultaneous over both $x$ and $\kappa$, smoothing levels higher than 46 are included, no significant gradients are expected under the null hypothesis. Again, the value $\kappa$ above which the significance level drops to 0 depends on the sample size. Since in the two real data examples discussed in Section 3, $n = 1000$ and the range of $\kappa$ extends beyond 46, we can be highly confident that the features discovered by doubly simultaneous SphereSiZer
inference are ‘really there’ and not just noise artifacts. The situation is analogous to the original SiZer which was reported to be conservative when using bootstrap based inference simultaneous over both location and scale (Chaudhuri and Marron, 1999).

3. Examples

We first illustrate SphereSiZer using simulated data, generating random spherical data from mixtures of von Mises-Fisher densities. In general, let \( f(x) = p_1f_1(x) + \cdots + p_mf_m(x) \) be a mixture of density functions where the non-negative weights \( p_i \) satisfy \( \sum_{i=1}^{m} p_i = 1 \). To produce a random sample point from this mixture, one can first generate an index \( i \) uniformly on \( \{1, \ldots, m\} \) according to the probabilities \( \{p_1, \ldots, p_m\} \), and then generate a point from \( f_i \).

Fig. 3 shows the contour plot of our first test density, \( f(x) = 0.5f_{vMF}(x; \mu_1, \kappa_1) + 0.5f_{vMF}(x; \mu_2, \kappa_2) \), as well as a SphereSiZer map based on 200 random directions sampled from it. Spherical coordinates of the directional means are \( \theta = 7\pi/24, \phi = 4\pi/6 \) and \( \theta = \pi/2, \phi = 0 \) for \( \mu_1 \) and \( \mu_2 \), respectively, while the concentration parameters are \( \kappa_1 = 35 \) and \( \kappa_2 = 20 \). Here the significance level \( \alpha = 0.05 \) and the bootstrap sample size \( B = 500 \), as they are in all examples in this section. The map is part of an atlas simultaneous over both \( x \) and \( \kappa \), and for this map \( \kappa = 25 \). In all our examples, the maximum of (8) was computed using an \( x \)-grid \( G \) of size 1001 and a \( \kappa \)-grid \( K \) of size 40 (cf. Section 2.3). In this first example, the data points are scattered over the whole sphere so the map consists of two disks corresponding to the northern and southern hemispheres. The left side is the projection of the northern hemisphere viewed from the positive \( z \)-axis (\( \phi = 0 \)). The azimuth angle \( \theta = \pi \) is on the left side and \( \theta = 0 \) is located opposite to it where the two disks touch. Drawing the two projection disks in this fashion aids interpretation of the density features that cross the meridian \( \phi = \pi \). The polar angle \( \varphi = 0 \) (north pole) lies at the center and the meridian \( \varphi = \pi/2 \) is represented by the boundary of the projection disk. To avoid clutter, only polar angles \( \pi/6 \) and \( 2\pi/6 \) are marked in the contour plot. The right hand side disk is the projection of the southern hemisphere from the direction of the negative \( z \)-axis (\( \varphi = \pi \)). The azimuth angles \( \pi \) and 0 change places and the polar angle \( \pi \) is now at the center. The SphereSiZer map of Fig. 3 suggests that the underlying density has two local maxima where the other maximum lies on the borderline between the two hemispheres of contour plots near \( \theta = \pi/2 \). As features in such locations can be difficult to interpret, it may be useful to consider projections also from other directions, such as along the \( x \)- (\( \theta = 0, \varphi = \pi/2 \)) or \( y \)-axes (\( \theta = \pi/2, \varphi = \pi/2 \)). Fig. 4 shows the SphereSiZer map of this example when the projection is taken along the \( x \)-axis. Notice that now polar angles are marked outside the disks while the azimuth angle markings lie inside the disks.

The test density function in the second example is

\[
 f(x) = 0.4f_{vMF}(x; \mu_1, \kappa_1) + 0.4f_{vMF}(x; \mu_2, \kappa_2) + 0.2f_{vMF}(x; \mu_3, \kappa_3),
\]

where \( \mu_1 = (8\pi/15, 7\pi/18), \mu_2 = (14\pi/15, 5\pi/24), \mu_3 = (5\pi/3, 7\pi/18), \kappa_1 = 40, \kappa_2 = 40 \) and \( \kappa_3 = 30 \). Now all simulated data lie on the northern hemisphere and therefore only one projection disk is used for each map. The SphereSiZer atlas of

Fig. 4. The SphereSiZer map of the same data as in Fig. 3 but now projected along the \( x \)-axis \( (\theta = 0, \phi = \pi/2) \).
Fig. 5. The second artificial test density and a SphereSiZer atlas based on a random sample of size 200 on the northern hemisphere. (a) A contour plot of the test density with the 200 random sample points and SphereSiZer maps with smoothing level (b) $\kappa = 26$, (c) $\kappa = 20$ and (d) $\kappa = 12$.

Fig. 5 shows how the significant features change with scale. Again, the size of the random sample is 200, the quantile used in the inference is simultaneous over both $x$ and $\kappa$ and the number of bootstrap samples $B = 500$. For $\kappa = 26$, the significant gradients clearly suggest two adjacent peaks. In the larger scale map with $\kappa = 20$, the two adjacent peaks still persist and a few gradients hint at a third peak near $\theta = 5\pi/3$, $\varphi = \pi/2$. Increasing smoothing slightly to $\kappa = 12$ makes the third peak more significant but the two adjacent peaks have merged.

To demonstrate the practical potential of SphereSiZer, we analyze a real data set of surface unit normal vectors of an infant's head. The study of infant head deformations is an active area of research (Hutchison et al., 2004; Laughlin et al., 2011; Robinson and Proctor, 2009). Flat regions or asymmetry of the head are likely to lead to local maxima in the normal direction density function, as shown in the schematic illustration of Fig. 6 (cf. Vuollo et al., 2016).

A 3D head surface model of a three months old infant diagnosed with plagiocephaly, where flattening of one side of the head causes an asymmetrical shape, was recorded using the 3dMDhead™ System manufactured by 3dMD LLC (Atlanta, USA). For the exact description of the head positioning and definition of the head region analyzed, see Vuollo et al. (2016) (position of the head in this example was Pose 1 defined in that article). In Fig. 7, the 3D image of the head is shown from
Fig. 6. (a) A schematic illustration of a generally round surface that includes a flat region, together with a number of normal vector directions. The normal vectors in the flat area are highlighted in red and they all point in the same direction. (b) The normal vector directions interpreted as points on a circle. The red points correspond to red normal vectors of panel (a). They generate a tight cluster of directions that results in a local maximum in the normal direction density function.

Fig. 7. A 3D image of an infant’s head from three different angles.

three different angles. The 3D images of all examples in this section were processed using Rapidform2006 software (INUS Technology, Inc., Seoul, Korea).

The left panel of Fig. 8 shows the contour plot of a kernel density estimate of 6736 surface normal directions placed more or less uniformly on the posterior side of the infant’s head. The smoothing parameter value used was $\kappa = 43.61$, obtained using the rule of thumb formula proposed in García-Portugués (2013). The contour plot suggests two local maxima in the kernel estimate. However, such a large set of normal directions easily makes many features statistically significant and therefore does not offer a very interesting setting for testing the performance of SphereSiZer. Also, being computed at fixed regular grid points, the head normals really cannot be regarded as a true random sample, as required in the bootstrap approach used for inference. To test the potential of SphereSiZer, we therefore regard the kernel estimate based on the full set of 6736 normal directions as the underlying truth, sample from it a much smaller set of 1000 directions and then, as a demonstration of SphereSiZer, use this smaller set for making inferences about the underlying true density. The kernel density estimate (4) can be regarded as a mixture of densities so that sampling can be done as in the case of the two artificial densities analyzed above.

The right panel of Fig. 8 presents a SphereSiZer analysis of the 1000 normal vector directions for $\kappa = 35$. Inference is simultaneous over $x$. The presence of two local maxima in the data gets moderately strong support. Fig. 9 shows an atlas of four SphereSiZer maps for the same 1000 head normal directions as in Fig. 8. From (a) to (d) the values of $\kappa$ are 45, 35, 25 and 15, increasing the level of smoothing progressively. Now the inference is simultaneous over both $x$ and $\kappa$. Looking at the significant gradients, two local maxima are simultaneously present for $\kappa = 25$ but the one on the right, near $\theta = 2\pi/3$, $\varphi = \pi$, appears to be more robust as its presence is suggested through a wider range of scales. The effect of the two levels of simultaneity on inference results can be seen by comparing maps of Fig. 8b and Fig. 9b. Making the inference simultaneous also over the smoothing level $\kappa$ makes the analysis more conservative.
In the final example, we use SphereSiZer to analyze earthquake data collected by the Northern California Earthquake Data Center (NCEDC, 2016). The data consists of the geographic coordinates of the first 1000 earthquakes in 2016 of magnitude 4.8 to 7 on the Richter Scale in the Northern Hemisphere (see Fig. 10a). In geographic coordinates, longitude corresponds to the azimuthal angle and latitude to the polar angle. The range for longitude is $[-180^\circ, 180^\circ]$ and the azimuthal angle 0 corresponds to the longitude of 0°. Similarly, the polar angle 0 corresponds to the latitude of 90° and the latitude range is $[-90^\circ, 90^\circ]$.

The SphereSiZer analysis of the 1000 earthquake locations is summarized in the three SphereSiZer maps of Fig. 10. As before, the number of bootstrap samples used in the analysis was $B = 500$. All three maps show two local maxima, one near $\theta = 3\pi/2, \varphi = \pi/2$ (0°N, 90°W) and the other around $\theta = 3\pi/4, \varphi = \pi/3$ (30°N, 135°E). Most earthquakes are known to occur at the boundaries of tectonic plates (Bird, 2003). Indeed, according to Fig. 10a, the earthquake locations seem to concentrate near these boundaries and the two maxima suggested by SphereSiZer in central America and the Western Pacific region are also located in such boundary areas. Note that for the scales $\kappa = 25$ and $\kappa = 15$, a few significant gradients suggest an elevated earthquake density near $\theta = \pi/2, \varphi = \pi/3$ (30°N, 90°E). This could be explained by proximity to the boundary of the Eurasian and Indian plates (Bird, 2003).

All computations were carried out on a laptop with an Intel Core i7-6600U 2.60 GHz processor. For sample sizes $n = 1000$, $B = 500$, and for inference simultaneous over both $\mathbf{x}$ and $\kappa$, it took about 10 min to compute a SphereSiZer map for a single value of $\kappa$. The computation of a map with the same sample sizes but for inference that is simultaneous only over $\mathbf{x}$ took only about 24 s.

4. Summary and questions for future work

SphereSiZer is a novel method for scale space analysis of spherical data. It can be viewed as a generalization of both CircSiZer that handles circular data as well as the gradient analysis of $S^3$ that deals with the two-dimensional Euclidean case (cf. Oliveira et al., 2014; Godtliebsen et al., 2002). The role of a family plot and feature significance map commonly used in statistical scale space methods is played by an atlas of planar contour plots overlaid with statistically significant gradient vectors. As in $S^3$, a movie is an effective means of exploring the scale-dependent features underlying the data. The test examples demonstrated how SphereSiZer works in practice.

Simulation was used to examine the performance of the bootstrap based quantiles in significance testing. The quantiles for simultaneous inference appeared to work quite well, erring on the conservative side and therefore probably not producing many false positives. In pointwise inference the number of false positives begins to increase rapidly when $\kappa$ exceeds a threshold value that depends on the sample size. On the other hand, in inference simultaneous over $\mathbf{x}$, the significance level drops to 0 for large values of $\kappa$. Because of problems caused by multiple hypothesis testing as well as the observed behavior of the actual significance levels, we strongly recommend using only simultaneous inference in SphereSiZer analyses, even at the risk of potentially being too conservative in inferring statistically significant features of a density.
An interesting topic for future research would be to consider the possibility of replacing the bootstrap with something similar to the approach of Hannig and Marron (2006) based on extreme value theory of Gaussian processes. The goal would be more accurate inference of the salient density function features. A second topic of interest would be to generalize the theory of WiZer to the case of spherical data (Huckemann et al., 2016). While such a generalization would be guaranteed to satisfy the causality property of axiomatic scale space theory, we do believe that when the von Mises–Fisher kernel is used in spherical smoothing, non-causality is likely to occur only in very small scales. Especially with simultaneous inference, SphereSiZer would probably be unlikely to flag such features as significant.

Finally, there is the question of generalizations to higher dimensions. The problem is not the definition of appropriate tests and the associated quantiles, as the mathematical derivations of Section 2 can be readily generalized to higher dimensions, but rather visualization and interpretation of the results. After all, much of the appeal of scale methodology lies with easy-to-interpret visual summaries of inference. While, in the Euclidean case, some proposals have been made for scale space analyses with higher dimensional data (cf. Holmström and Pasanen, 2017), finding useful multi-dimensional analogues of
Fig. 10. SphereSiZer analysis of earthquake location data on the Northern Hemisphere. (a) A map projection with 1000 earthquake locations and SphereSiZer maps with smoothing parameter value (b) $\kappa = 50$, (c) $\kappa = 25$, and (d) $\kappa = 15$. Inference is simultaneous over both $x$ and $\kappa$.

SphereSiZer atlases could be challenging, with the constrained nature of spherical data making the task even more difficult. One interesting possibility would be to visualize the multivariate spherical density using the volume function defined by its level sets. The volume function, considered in the Euclidean context e.g. in Klemelä (2009) and Holmström et al. (2017), is a mode isomorphic, one-dimensional description of the mode structure of a multivariate density and it could perhaps be usefully extended to directed data and visualization of spherical densities.

Acknowledgments

The work of Ville Vuollo was supported by the Finnish Doctoral Program in Oral Sciences. We are grateful to Henri Aarnivala, Virpi Harila, Tuomo Heikkinen, Pertti Pirutiniemi and Arja Marita Valkama for providing us with the 3D infant head data used in the real data example. Earthquake data for this study were accessed through the Northern California Earthquake Data Center (NCEDC), doi: 10.7932/NCEDC.
Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2018.03.014.

References

NCEDC, 2016. Northern California Earthquake Data Center. UC Berkeley Seismological Laboratory. Dataset. URL http://dx.doi.org/10.7932/NCEDC.