Small-sphere distributions for directional data with application to medical imaging

Byungwon Kim1 | Stephan Huckemann2 | Jörn Schulz3 | Sungkyu Jung1,4

1Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania
2Felix Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Göttingen, Germany
3Department of Electrical and Computer Engineering, University of Stavanger, Stavanger, Norway
4Department of Statistics, Seoul National University, Seoul, South Korea

Correspondence
Sungkyu Jung, Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea. Email: sungkyu@snu.ac.kr

Funding information
Niedersachsen Vorab of the Volkswagen Foundation; Deutsche Forschungsgemeinschaft, Grant/Award Number: HU 1575/4

Abstract
We propose novel parametric concentric multi-unimodal small-subsphere families of densities for \( p - 1 \geq 2 \)-dimensional spherical data. Their parameters describe a common axis for \( K \) small hypersubspheres, an array of \( K \) directional modes, one mode for each subsphere, and \( K \) pairs of concentrations parameters, each pair governing horizontal (within the subsphere) and vertical (orthogonal to the subsphere) concentrations. We introduce two kinds of distributions. In its one-subsphere version, the first kind coincides with a special case of the Fisher–Bingham distribution, and the second kind is a novel adaption that models independent horizontal and vertical variations. In its multi-subsphere version, the second kind allows for a correlation of horizontal variation over different subspheres. In medical imaging, the situation of \( p - 1 = 2 \) occurs precisely in modeling the variation of a skeletally represented organ shape due to rotation, twisting, and bending. For both kinds, we provide new computationally feasible algorithms for simulation and estimation and propose several tests. To the best knowledge of the authors, our proposed models are the first to treat the variation of directional data along several concentric small hypersubspheres, concentrated near modes on each subsphere, let alone horizontal dependence. Using several simulations, we show that our methods are more powerful than a recent nonparametric method and ad hoc methods. Using data from medical imaging, we demonstrate the advantage of our method and infer on the
dominating axis of rotation of the human knee joint at different walking phases.

KEYWORDS
Bingham–Mardia distribution, directional data, skeletal representation, small circle, small sphere, von Mises–Fisher distribution

1 | INTRODUCTION

In medical imaging, accurately assessing and correctly diagnosing shape changes of internal organs is a major objective of a substantial challenge. Shape deformations can occur through long-term growth or necrosis as well as by short-term natural deformations. In view of surgery and radiation therapy, it is important to model all possible variations of object deformations by both long- and short-term changes, in order to control the object’s exact status and shape at treatment time. Rotational deformations, such as rotation, bending, and twisting, form a key subcategory of possible shape changes. For instance, shape changes of hippocampi in the human brain have been shown to mainly occur in the way of bending and twisting (Joshi et al., 2002; Pizer et al., 2013).

For the task of modeling three-dimensional (3D) objects, an abundance of approaches have been introduced. Closely related to our work are landmark-based shape models (Cootes, Taylor, Cooper, & Graham, 1992; Dryden & Mardia, 1998; Kurtek, Ding, Klassen, & Srivastava, 2011), where a solid object is modeled by the positions of surface points, chosen either anatomically, mathematically, or randomly. A richer family of models is obtained by attaching directions normal to the sampled surface points. More generally, in skeletal representations (called s-reps, Siddiqi & Pizer, 2008), an object is modeled by the combination of skeletal positions (lying on a medial sheet inside the object) and spoke vectors (connecting the skeletal positions with the boundary of the object). In these models, describing the variation of rotational deformations can be transformed into a problem of exploring the motion of directional vectors on the unit two-sphere. As argued in Schulz et al. (2015), directional vectors representing rotational deformations tend to be concentrated on small circles on the unit sphere; a toy data example in Figure 1 shows a typical pattern of such observations.

For such s-rep data, spread out nonuniformly over several concentric small circles, to the best knowledge of the authors, there are neither parametric models nor inferential methods available.

FIGURE 1 (a) Toy example showing observations (solid green) distributed near a small circle $C(\mu, \nu)$. The heat maps of (b) fitted Bingham–Mardia density and (c) the proposed small-sphere density of the first kind are overlaid. Red: high density; blue: low density [Colour figure can be viewed at wileyonlinelibrary.com]
In order to fill this gap, in particular, to model horizontal (detailed below) dependence across different small subspheres, which are typical for s-rep data, we propose two new families and provide methods for estimation, simulation, and statistical tests. To date, only for the estimation of small circles, not involving horizontal dependence, there is only the nonparametric least squares (LS) method by Schulz et al. (2015) available. For the more simple task of estimating a single small circle, along which data are spread uniformly, there is the parametric family of Bingham–Mardia (BM) distributions by Bingham and Mardia (1978) available. We remove the uniformity constraint by adding a von Mises–Fisher (vMF) term, giving either a special case of a Fisher–Bingham distribution (Kent, 1982) or, more subtly, a new family of distributions. For the former, while simulation and maximum likelihood estimation (MLE) methods are available (cf. Hoff, 2009), for computational feasibility, we adapt the saddle-point approximation of Kume and Wood (2005). For the latter new one, in the case of small circles, we develop even faster numerical methods for simulation and estimation. For application to s-rep data, we propose several multivariate extensions, in particular, in order to model horizontal dependence across different small subspheres. We show the usefulness of our new methods by analyzing s-rep data and comparing to the limited capabilities of LS and BM using methods derived ad hoc. As mentioned, while LS and BM cannot model horizontal dependence, for comparison, we also derive a crude ad hoc method to implement in composite principal nested spheres (CPNS) from Pizer et al. (2013) a test for horizontal dependence.

Let us now provide more details. Throughout this paper, \( S^{p-1} = \{ x \in \mathbb{R}^p | \|x\| = 1 \} \) is the unit sphere in arbitrary dimension \( p \geq 3 \), and \( \|x\| = (x^T x)^{1/2} \) is the usual 2-norm of vector \( x \). To precisely describe the targeted data situation, we define a \((p - 2)\)-dimensional subsphere of \( S^{p-1} \) as the set of all points equidistant from \( \mu \in S^{p-1} \), denoted by

\[
C(\mu, \nu) = \{ x \in S^{p-1} | \delta(\mu, x) = \arccos(\nu) \}, \quad \nu \in (-1, 1).
\]

Here, \( \delta(u, v) = \arccos(u^T v) \) is the geodesic distance between \( u, v \in S^{p-1} \). The subsphere is called a great subsphere if \( \nu = 0 \) and a proper small subsphere if \( \nu \neq 0 \). Note that \( C(\mu, \nu) \subset S^{p-1} \) is well defined for all \( p > 1 \). For the special case of \( p = 3 \), \( C(\mu, \nu) \) is a circle, that is, a great circle if \( \nu = 0 \) and a proper small circle if \( \nu \neq 0 \). To model the data in Figure 1, one may naively use the BM distribution, which is a family of densities on \( S^2 \) with a modal ridge along a small circle. However, typical observations we encountered in applications do not uniformly spread over the full circle, and the BM distribution does not fit well, as shown in Figure 1b. Moreover, when, by a single observation, multiple directional vectors are provided, that is, data are on a polyshpere \((S^2)^K\), to the knowledge of the authors, there is no tool available to date, to model dependencies between directions.

In this paper, we propose two types of new distributional families for random directional vectors on \( S^{p-1} \), which we call small-sphere distributions of the first (S1) and second (S2) kind. If \( p = 3 \), the proposed distributions may be called small-circle distributions. These two distributional families are designed to have higher densities on \( C(\mu, \nu) \) and to have a unique mode on \( C(\mu, \nu) \). An example of a small-sphere density, fitted to the toy data, is shown in Figure 1c. The new densities are natural extensions of the BM distribution with an additional term explaining a decay from a mode. If the additional term is a vMF density on \( S^{p-1} \), we obtain S1, which is a subfamily of the general Fisher–Bingham distribution (Mardia, 1975; Mardia & Jupp, 2000). On the other hand, if the additional term is a vMF density on the subsphere \((\cong S^{p-2})\), we obtain the S2 distribution, in which case the horizontal (inside the small subsphere) and vertical (orthogonal to the small subsphere) components of the directional vectors are independent of each other.
Several multivariate extensions of the new distributions to \((S^{p-1})^K, K \geq 2\), are discussed as well. In particular, we show that a special case, called MS2, of our multivariate extensions is capable of modeling dependent random vectors. It has a straightforward interpretation, and we provide for fast estimation of its parameters. This MS2 distribution is specifically designed with s-rep applications in mind. In particular, s-rep data from rotationally deformed objects have directional vectors that are “rotated together,” share a common axis of rotation, and are “horizontally dependent” (when the axis is considered to be vertically positioned). The component-wise independence of the S2 distributions plays a key role in this simple and interpretable extension. We discuss here likelihood-based parameter estimation and testing procedures of the multivariate distributions.

While the new distributions summarized in Table 1 contribute to the literature of directional distributions (Mardia & Jupp, 2000), the proposed estimation procedures for the S1, S2, and MS2 parameters can be thought of as a method of fitting small subspheres to data, which has been of separate interest. Nonparametric LS-type solutions for such problem date back to the works of Mardia and Gadsden (1977), Gray, Geiser, and Geiser (1980), and Rivest (1999). Jung, Dryden, and Marron (2012) proposed recursively fitting small subspheres in the dimension reduction of directional and shape data. Pizer et al. (2013) proposed to combine separate small-circle fitting results in the analysis of s-rep data. In a similar spirit, Jung, Foskey, and Marron (2011) and Schulz et al. (2015) also considered fitting small circles in applications to s-rep analysis. In a simulation study, we show that our estimators provide smaller mean angular errors in small-circle fits than recent developments listed above.

The rest of this paper is organized as follows. In Section 2, we introduce the proposed densities of the S1 and S2 distributions and discuss their multivariate extensions including the MS2 distribution. Procedures for obtaining random variates from the proposed distributions are also discussed. In Section 3, algorithms to obtain the maximum likelihood estimators of the parameters are proposed and discussed. In Section 4, we introduce several hypotheses of interest and procedures of likelihood-ratio tests. Simulation studies demonstrating the performance of small-circle fitting, estimating dependency, and the power of the proposed test are contained in Section 5. In Sections 6 and 7, we demonstrate applications of the new multivariate distributions to analyze models that represent human organs and knee motions. The Appendix and the Online Supplementary Material contain technical details, additional numerical results, and a table summarizing the notation we used.

### TABLE 1

**Key features of the newly proposed small-sphere distributions (top three rows) and methods (bottom three rows) developed in this paper. Items marked “×” are beyond the scope of this paper**

| Relation among horizontal and vertical components | Dependent | Independent |
|---------------------------------------------------|------------|-------------|
| Univariate                                        | S1         | S2          |
| Multivariate (indep.)                             | iMS1       | iMS2        |
| Multivariate (dep.)                               | ×          | GMS2        |
| Simulation                                        | Gibbs sampling | ×          |
| Estimation                                        | Approximate MLE | ×          |
| Hypothesis testing                                | Likelihood ratio | ×          |
|                                                   |            | MS2 \((p - 1 = 2)\) |

*Note.* S1 = small-sphere distribution of the first kind; S2 = small-sphere distribution of the second kind; iMS1 = independent multivariate S1 distribution; iMS2 = independent multivariate S2 distribution; GMS2 = general multivariate small-sphere distribution of the second kind; MS2 = multivariate S2 distribution; MLE = maximum likelihood estimation.
2 | PARAMETRIC SMALL-SPHERE MODELS

First, we introduce two classical spherical densities, and then, we suitably combine them for our purposes.

2.1 | Two classical distributions on $S^{p-1}$

The vMF distribution (Mardia & Jupp, 2000, p. 168) is a fundamental unimodal and isotropic distribution for directions with density

$$f_{vMF}(x; \mu, \kappa) = \left( \frac{\kappa}{2} \right)^{p/2-1} \frac{1}{\Gamma(p/2)I_{p/2-1}(\kappa)} \exp\{\kappa \mu^\top x\}, \quad x \in S^{p-1}. \quad (1)$$

Here, $\Gamma$ is the gamma function, and $I_\nu$ is the modified Bessel function of the first kind and order $\nu$. The parameter $\mu \in S^{p-1}$ locates the unique mode with $\kappa \geq 0$ representing the degree of concentration.

The BM distribution was introduced by Bingham and Mardia (1978) to fit data in $S^2$ that clusters near a small circle $C(\mu, \nu)$. For an arbitrary dimension $p \geq 3$, the BM density is given by

$$f_{BM}(x; \mu, \kappa, \nu) = \frac{1}{a(\kappa, \nu)} \exp\{-\kappa (\mu^\top x - \nu)^2\}, \quad x \in S^{p-1}, \quad (2)$$

where $a(\kappa, \nu) > 0$ is the normalizing constant.

For our purpose of generalizing these distributions, we represent the variable $x \in S^{p-1}, p \geq 3$, by spherical angles $\phi_1, \ldots, \phi_{p-1}$ satisfying $\cos \phi_1 = \mu^\top x$. Setting $s := \cos \phi_1 \in [-1, 1]$ and $\phi := (\phi_2, \ldots, \phi_{p-1}) \in [0, \pi]^{p-3} \times [0, 2\pi)$, the random vector $(s, \phi)$ following the vMF (1) or the BM (2) distribution has the following respective density values:

$$g_{vMF}(s, \phi; \kappa) = \left( \frac{\kappa}{2} \right)^{p/2-1} \frac{1}{\Gamma(p/2)I_{p/2-1}(\kappa)} \exp\{\kappa s\}, \quad (3)$$

$$g_{BM}(s, \phi; \kappa, \nu) = \frac{1}{a(\kappa, \nu)} \exp\{-\kappa (s - \nu)^2\}. \quad (4)$$

In consequence, for both distributions, $s$ and $\phi$ are independent, and the marginal distribution of $\phi$, which parameterizes a co-dimension one unit sphere $S^{p-2}$, is uniform. In (3), the marginal distribution of $s$ is a shifted exponential distribution truncated to $s \in [-1, 1]$, whereas in (4), the marginal distribution of $s$ is a normal distribution truncated to $s \in [-1, 1]$. Both densities are isotropic, that is, rotationally symmetric with respect to $\mu$. The vMF density is maximal at mode $\mu$ and decreases as the latitude $\phi_1$ increases, whereas the BM density is uniformly maximal on the small sphere $C(\mu, \nu)$ and decreases as $\phi_1$ deviates from $\arccos(\nu)$.

2.2 | Small-sphere distributions of the first and second kind

The proposed small-sphere densities of the first and second kind on $S^{p-1}$, for $x = (x_1, \ldots, x_p) \in S^{p-1}$ with parameters $\mu_0, \mu_1 \in S^{p-1}, \nu = \mu_0^\top \mu_1 \in (-1, 1), \kappa_0 > 0, \kappa_1 > 0$, are given by

$$f_{S1}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, \nu)} \exp\{-\kappa_0 (\mu_0^\top x - \nu)^2 + \kappa_1 \mu_1^\top x\}, \quad (5)$$

$$f_{S2}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, \nu)} \exp\{-\kappa_0 (\mu_0^\top x - \nu)^2 + \kappa_1 \frac{\mu_1^\top P_{\mu_0} x}{\sqrt{\mu_1^\top P_{\mu_0} \mu_1 x^\top P_{\mu_0} x}}\}, \quad (6)$$
respectively, where \( a(\kappa_0, \kappa_1, v) \) and \( b(\kappa_0, \kappa_1, v) \) are normalizing constants. Here, \( P_{\mu_0} \) denotes the matrix of orthogonal projection to the orthogonal complement of \( \mu_0 \); \( P_{\mu_0} = I_p - \mu_0 \mu_0^\top \), where \( I_p \) is the identity matrix. (In (6), we use the convention \( 0/0 = 0 \).)

These distributions are well suited to model observations that are concentrated near the small sphere \( C(\mu_0, v) \) but are not rotationally symmetric. The first kind (5) is a natural combination of the vMF (1) and BM (2) distributions. The parameter \( \mu_1 \) gives the mode of the distribution, which, by the definition of \( v \), is on the small sphere \( C(\mu_0, v) \). These parameters, that is, \( \mu_0, \mu_1, \) and \( v \), are illustrated in Figure 1a for the \( p = 3 \) case. The parameter \( \kappa_0 \) controls the vertical concentration toward the small sphere (with an understanding that \( \mu_0 \) is arranged vertically). In (5), \( \kappa_1 \) controls the isotropic part of the concentration around the mode, forcing the density to decay from \( \mu_1 \).

The rationale for the second kind (6) is better understood with a change of variables. Let us assume for now that \( \mu_0 = (1, 0, \ldots, 0)^\top \). For any \( x = (x_1, \ldots, x_p)^\top \in \mathbb{S}^{p-1} \), write \( s := x_1 = \mu_1^\top x \). If the spherical coordinate system \((\phi_1, \ldots, \phi_{p-1})\) as defined for (4) is used, then \( s = \cos \phi_1 \). The “orthogonal complement” of \( s \) is denoted by

\[
y := (x_2, \ldots, x_p)/\sqrt{1 - s^2} \in \mathbb{S}^{p-2},
\]

where vector \( y \) is obtained from the relation \( P_{\mu_0} x/\|P_{\mu_0} x\| = (0, y) \in \mathbb{S}^{p-1} \). Similarly, define \( \tilde{\mu}_1 \in \mathbb{S}^{p-2} \) as the last \( p - 1 \) coordinates of \( P_{\mu_0} \mu_1/\|P_{\mu_0} \mu_1\| \). Then, the random vector \((s, y) \in [-1, 1] \times \mathbb{S}^{p-2} \) from S1 or S2 has densities

\[
g_{S1}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, v)} \exp \left\{ -\kappa_0(s-v)^2 + \kappa_1 \mu_1^\top \left( s, \sqrt{1-s^2} y \right) \right\},
\]

\[
g_{S2}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, v)} \exp \left\{ -\kappa_0(s-v)^2 + \kappa_1 \tilde{\mu}_1^\top y \right\},
\]

respectively, for \( s \in [-1, 1], y \in \mathbb{S}^{p-2} \). The subtle difference is that for the first kind (8), the “vMF part” (the second term in the exponent) is not statistically independent from the “BM part,” whereas it is true for the second kind (9). That is, \( s \) and \( y \) are independent only in the second kind. Accordingly, in (9), \( \kappa_1 \) controls the horizontal concentration toward mode \( \mu_1 \). The parameters \( \mu_0, \mu_1, \) and \( \kappa_0 \) of the second kind have the same interpretations as those of the first kind.

We use the notation \( X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1) \) and \( Y \sim S2(\mu_0, \mu_1, \kappa_0, \kappa_1) \) for random directions \( X, Y \in \mathbb{S}^{p-1} \) following small-sphere distributions of the first and second kind with parameters \((\mu_0, \mu_1, \kappa_0, \kappa_1)\), respectively. The proposed distributions are quite flexible and can fit a wide range of data. In Figure 2, we illustrate the S1 densities (5) with various values of the concentration parameters \( \kappa_0, \kappa_1 \). In all cases, the density is relatively high near the small circle \( C(\mu_0, v) \) and has the mode at \( \mu_1 \in C(\mu_0, v) \). Despite the difference in their formulations, the S2 densities (6) look similar to S1 densities for each fixed parameter set. We refer to the Online Supplementary Material (Section A2) for several visual examples of the S2 density.

Both distributions are invariant to rotation in the null space of \((\mu_0, \mu_1)\).

**Proposition 1.** Let \( X, Y \in \mathbb{S}^{p-1} \) be random directions with \( X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1) \) and \( Y \sim S2(\mu_0, \mu_1, \kappa_0, \kappa_1) \), and let \( B \) be a \( p \times p \) orthogonal matrix.

(i) \( X \) and \( BX \) (or \( Y \) and \( BY \)) have the same distribution if and only if \( B \mu_0 = \mu_0 \) and \( B \mu_1 = \mu_1 \).

(ii) \( X \sim S1(-\mu_0, \mu_1, \kappa_0, \kappa_1) \) and \( Y \sim S2(-\mu_0, \mu_1, \kappa_0, \kappa_1) \).

An example for matrix \( B \) in Proposition 1(i) is the reflection matrix \( B = I_p - 2UU^\top \), where \( U = [u_1, \ldots, u_p] \) is such that \([u_1, \ldots, u_p] \) is a \( p \times p \) orthogonal matrix whose column vectors \( u_1 \) and \( u_2 \) generate \( \mu_0 \) and \( \mu_1 \).
FIGURE 2 The S1 densities on $S^2$ modeling nonisotropic small-circle distributions. Red: high density; blue: low density. In all figures, $\mu_0$ points to the north pole, and $\mu_1$ satisfies $\mu_0^T \mu_1 = 1/2$. Rows and columns correspond to different choices of concentration parameters $(\kappa_0, \kappa_1)$. (a) $\kappa_0 = 10, \kappa_1 = 4$. (b) $\kappa_0 = 10, \kappa_1 = 1$. (c) $\kappa_0 = 10, \kappa_1 = .5$. (d) $\kappa_0 = 20, \kappa_1 = 4$. (e) $\kappa_0 = 20, \kappa_1 = 1$. (f) $\kappa_0 = 20, \kappa_1 = .5$. (g) $\kappa_0 = 40, \kappa_1 = 4$. (h) $\kappa_0 = 40, \kappa_1 = 1$. (i) $\kappa_0 = 40, \kappa_1 = .5$ [Colour figure can be viewed at wileyonlinelibrary.com]

Remark 1. The S1 distribution is a special case of the Fisher–Bingham distribution (Mardia, 1975). Following the notation of Kent (1982), the S1 distribution may be labeled as an FB$_6$ distribution, in the special case of $p = 3$, emphasizing the six-dimensional parameter space. In terms of the general parameterization of the Fisher–Bingham density (cf. Mardia & Jupp, 2000, p. 174), we write $\gamma = 2\kappa_0 \nu \mu_0 + \kappa_1 \mu_1$ and $A = \kappa_0 \mu_0 \mu_0^T$, so that the S1 density (5) is expressed as

$$f_{S1}(x; \gamma, A) = \frac{1}{a(\gamma, A)} \exp \left\{ \gamma^T x - x^T A x \right\},$$

(10)

where $a(\gamma, A) = a(\kappa_0, \kappa_1, \nu) \exp \{\kappa_0 \nu^2 \}$. This relation to the general Fisher–Bingham distribution facilitates random data generation and MLE, shown later in Sections 2.4 and 3.1.
2.3 | Multivariate extensions

The univariate small-sphere distributions (5) and (6) are now extended to model a tuple of associated random directions, \( \mathbf{X} = (X_1, \ldots, X_K) \in (S^{p-1})^K \). We confine ourselves to a special case where the marginal distributions of \( X_k \) have a common “axis” parameter \( \mu_0 \), but relaxing this condition is straightforward. We begin by introducing multivariate small-sphere distributions for independent random directions, denoted by iMS1 and iMS2.

**Independent extensions.** Suppose that, in the \( K \)-tuple of random directions \( \mathbf{X} \), each \( X_k \in S^{p-1} \) is marginally distributed as \( S1(\mu_0, \mu_k, \kappa_0k, \kappa_k) \). Throughout, we assume that \( \nu_k = \mu_0^T \mu_k \in (-1, 1) \), so that the underlying small spheres do not degenerate. If the components of \( \mathbf{X} \) are mutually independent, then the joint density evaluated at \( \mathbf{x} \in (S^{p-1})^K \) is

\[
f_{\text{iMS1}}(\mathbf{x}) \propto \exp \left\{ \mathbf{\Gamma}^T \mathbf{x} - \mathbf{x}^T \mathbf{A} \mathbf{x} \right\}. \tag{11}
\]

Here, \( \mathbf{\Gamma} = [\gamma_1, \ldots, \gamma_K]^T \), where \( \gamma_k = 2\kappa_0 \kappa_k \mu_0 + \kappa_k \mu_k \), and \( \mathbf{A} = \mathbf{K}_0 \otimes (\mu_0 \mu_0^T) \), where \( \mathbf{K}_0 = \text{diag}(\kappa_01, \ldots, \kappa_{0K}) \). Each marginal density is of the form (10).

If each component is marginally distributed as \( S2(\mu_0, \mu_k, \kappa_0k, \kappa_k) \), then writing the density in terms of \((s, y)\) as done for (9) facilitates our discussion. First, we decompose each \( x_k \) into \( s_k = \mu_0^T x_k \in [-1, 1] \) and \( y_k \in S^{p-2} \) as defined in (7). Furthermore, we denote by \( \tilde{y}_k \) the scaled projection of \( \mu_k \) as done for the univariate case. Then, an independent multivariate extension for the S2 model can be expressed as the joint density of \( \mathbf{s} = (s_1, \ldots, s_K) \) and \( \mathbf{y} = (y_1, \ldots, y_K) \), that is,

\[
g_{\text{iMS2}}(\mathbf{s}, \mathbf{y}) \propto \exp \left\{ H^T \mathbf{s} - \mathbf{s}^T \mathbf{K}_0 \mathbf{s} + \mathbf{M}^T \text{vec}(\mathbf{y}) \right\}, \tag{12}
\]

where \( H = (2\kappa_{01} \nu_1, \ldots, 2\kappa_{0K} \nu_K) \) and \( \mathbf{M} = \text{vec}(\kappa_1 \tilde{\mu}_1, \ldots, \kappa_K \tilde{\mu}_K) \), whereas \( \text{vec}(\cdot) \) denotes the column-wise vectorization of a matrix.

**Vertical and horizontal dependence.** On the basis of (11) and (12), we now contemplate dependent models. Obviously, if we allow in (11) nonzero off-diagonal entries of \( \mathbf{A} \), then we obtain a dependent modification of the S1 model. With our applications in mind, however, we aim at modeling a specific structure of dependence that is natural to the variables \((\mathbf{s}, \mathbf{y})\) in (12).

If \( s_1, \ldots, s_K \) are dependent, we speak of vertical dependence; if \( y_1, \ldots, y_K \) are dependent, we speak of horizontal dependence. In practice, when we deal with small-circle concentrated directional data, association among these vectors usually occurs along small circles with independent vertical errors. For example, when a 3D object is modeled by skeletal representations, as described in more detail in Section 6 and visualized in Figure 6, a deformation of the object is measured by the movements of directional vectors on \( S^2 \). When a single rotational deformation (such as bending, twisting, or rotation) occurs, all the directions move along small circles with a common axis \( \mu_0 \). In this situation, the longitudinal variations along the circles are dependent on each other because nearby spoke vectors are under the effect of similar deformations. (Examples of such longitudinal dependencies can be found in Section 6 as well as in Schulz et al., 2015.) Adding such horizontal (or longitudinal) dependence to a multivariate S1 model requires a careful introduction and parameterization of the off-diagonal entries of \( \mathbf{A} \) in (11). This is not straightforward, and we leave it for future work. On the other hand, it is feasible to extend the S2 model by generalizing the “vMF part” of \( \mathbf{y} \), the last term in the exponent of (12), to a Fisher–Bingham type.
To this end, we introduce a parameter matrix $B$ to model general quadratics in $\text{vec}(y)$. This allows to write the densities for a general multivariate small-sphere distribution of the second kind (GMS2) as follows:

$$
g_{\text{GMS2}}(s, y; H, K_0, M, B) = \frac{1}{T_1(H, K_0) T_2(M, B)} \exp \left\{ H^\top s - s^\top K_0 s + M^\top \text{vec}(y) + \text{vec}(y)^\top B \text{vec}(y) \right\}, \tag{13}
$$

where $H, K_0,$ and $M$ are as defined in (12), and $T_1(H, K_0)$ and $T_2(M, B)$ are normalizing constants. We set $B = (B_{kl})_{k,l=1}^K, B_{kl} = (b_{i,j}^{(k,l)})_{i,j=1}^{p-1},$ as a block matrix with vanishing blocks $B_{kk} = 0$ on the diagonal. The submatrix $B_{kl}$ models the horizontal association between $y_k$ and $y_l$. The fact that $z^\top B z = z^\top (B + B^\top) z$ for any vector $z \in \mathbb{R}^{(p-1)K}$ allows us to assume, without loss of generality, that $B$ is symmetric.

**An MS2 distribution on $(S^2)^K$.** As a viable submodel for the practically important case $p = 3$, we propose to use a special form for the off-diagonal blocks $B_{kl}$ of $B$. In particular, with $\lambda_{kl}$ representing the degrees of association between $y_k$ and $y_l$, we set

$$
B_{kl} = 2 \left( \begin{array}{cc} \tilde{\mu}_K & \tilde{\mu}_K^\perp \\ \tilde{\mu}_K^\perp & \tilde{\lambda}_{kl} \end{array} \right) \left( \begin{array}{c} 0 \\ 0 \end{array} \right) \left( \begin{array}{c} \tilde{\mu}_l \ 	ilde{\mu}_l^\perp \\ \tilde{\mu}_l^\perp \end{array} \right)^\top = 2 \lambda_{kl} \tilde{\mu}_K \tilde{\mu}_K^\perp (\tilde{\mu}_l \ 	ilde{\mu}_l^\perp)^\top, \tag{14}
$$

where $(\tilde{\mu}_K \ 	ilde{\mu}_K^\perp)$ is the rotation matrix given by setting

$$
\tilde{\mu}_K^\perp = \left( \begin{array}{cc} 0 & -1 \\ 1 & 0 \end{array} \right) \tilde{\mu}_K.
$$

The density (13) with the above parameterization of $B$ will be referred to as a multivariate $S2$ (MS2) distribution for data on $(S^2)^K$; its angular representation will be derived in (15) below.

Our choice of the simple parameterization (14) does not restrict the modeling capability of the general model (13) and has some advantages in parameter interpretations and in estimation. To see this, we resort to using an angular representation for $y$ (available to this $p = 3$ case). For each $k$, define $\phi_k$ and $\zeta_k$ such that $y_k = (\cos \phi_k, \sin \phi_k)^\top$ and $\tilde{\mu}_K = (\cos \zeta_k, \sin \zeta_k)^\top$. Accordingly, the inner products appearing in (13) can be expressed as

$$
\tilde{\mu}_K^\top y_k = \cos(\phi_k - \zeta_k), \quad (\tilde{\mu}_K^\perp)^\top y_k = \sin(\phi_k - \zeta_k).
$$

Let $
\phi = (\phi_1, \ldots, \phi_K)^\top, \ \zeta = (\zeta_1, \ldots, \zeta_K)^\top, \ \kappa = (\kappa_1, \ldots, \kappa_K)^\top,$

$$
c(\phi, \zeta) = (\cos(\phi_1 - \zeta_1), \ldots, \cos(\phi_K - \zeta_K))^\top, \quad
s(\phi, \zeta) = (\sin(\phi_1 - \zeta_1), \ldots, \sin(\phi_K - \zeta_K))^\top,
$$

and $\Lambda = (\lambda_{kl})_{k,l=1}^K$, where $\lambda_{kl}(= \lambda_{lk})$ for $k \neq l$ is the association parameter used in (14), and $\lambda_{kk}$ is set to zero. The density of the MS2 distribution, in terms of $(s, \phi)$, is then

$$
g_{\text{MS2}}(s, \phi; H, K_0, \kappa, \zeta, \Lambda) = \frac{1}{T_1(H, K_0) T_3(\kappa, \Lambda)} \exp \left\{ H^\top s - s^\top K_0 s + \kappa^\top c(\phi, \zeta) + \frac{1}{2} s(\phi, \zeta)^\top \Lambda s(\phi, \zeta) \right\}. \tag{15}
$$
From (15), it can be easily seen that the “horizontal angles” $\phi$ follow the multivariate von Mises distribution (Mardia, Hughes, Taylor, & Singh, 2008) and are independent of the vertical component $s$. As we will see later in Section 3.2, this facilitates the estimation for the MS2 distributions. Moreover, since

$$
\kappa^T c(\phi, \zeta) + \frac{1}{2} s(\phi, \zeta)^T \Lambda s(\phi, \zeta)
$$

$$
= \sum_{k=1}^{K} \kappa_k \left(1 - \frac{(\phi_k - \zeta_k)^2}{2}\right) + \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \left(\lambda_{kl}(\phi_k - \zeta_k)(\phi_l - \zeta_l)\right) + \sigma \left(||\phi - \zeta||^2\right),
$$

for large enough concentrations, $\phi$ is approximately multivariate normal with mean $\zeta$ and precision matrix $\Sigma^{-1}$, where $(\Sigma^{-1})_{kk} = \kappa_k$ and $(\Sigma^{-1})_{kl} = -\lambda_{kl}$ for $1 \leq k \neq l \leq K$. These parameters are naturally interpreted as partial variances and correlations. This interpretation of the parameters as entries of a precision matrix is most immediate under MS2 but not under the general case.

### 2.4 Random data generation

Generating pseudorandom samples from the S1 and S2 distributions is important in simulations and development of computer-intensive inference procedures.

For simulation of the S1 (5) and iMS1 (11) distributions, the fact that each marginal distribution of iMS1 is a special case of the Fisher–Bingham distribution is handy. Thereby, one can use the Gibbs sampling procedure developed for generating Fisher–Bingham variate samples (Hoff, 2009).

For simulation of the S2 (6), iMS2 (12), and MS2 (15) distributions, we take advantage of the independence of $s$ and $y$. As we assume vertical independence (i.e., $s_1, \ldots, s_K$ are independent), each $s_k$ can be sampled separately. Therefore, sampling from the MS2 distribution amounts to independently drawing samples from a truncated normal distribution (for $s_k$) and from a multivariate von Mises distribution (for $y$). Specifically, to sample $x = (x_1, \ldots, x_K)$ from $MS2(\mu_0, \mu, \kappa_0, \kappa, \Lambda)$, the following procedure can be used.

**Step 1.** For each $k$, sample $s_k$ from the truncated normal distribution with mean $v_k$ and variance $1/(2\kappa_0 k)$, truncated to the interval $(-1,1)$.

**Step 2.** For the S2 or the iMS2 model, sample each $y_k \in S^{p-2}$ in $y = (y_1, \ldots, y_K)$ independently from the von Mises distribution with mean $(1,0,\ldots,0)$ and concentration $\kappa_k$; for the MS2 distribution (when $p = 3$), sample the $K$-tuple $y \in (S^1)^K$ directly from the multivariate von Mises distribution with mean $(1,0)$ and precision parameters $\kappa$ and $\Lambda$.

**Step 3.** For each $k$, let $E_k$ be a $p \times p$ orthogonal matrix, with $(\mu_0,\mu_0,\mu_0/||\mu_0,\mu_0||)$ being the first two column vectors. Set $x_k = E_k^T(s_k, (1 - s_k^2)^{-1/2}y_k)$.

In our experiments, sampling from the S2 and MS2 distributions is much faster than sampling from the S1 distribution. In particular, when the dimension $p$ or the concentration level is high, the Markov chain simulations for S1 appear to be sluggish. Some examples of random samples from the S2, iMS2, and MS2 distributions are shown in Figure 3. The small circles $C(\mu_0, v_k)$ are also overlaid in the figure. Notably, the MS2 sample in the rightmost panel clearly shows horizontal dependence.
FIGURE 3  Random samples from the S2, iMS2, and MS2 distributions. Same colors represent same simulations. (a) Low concentrations ($\kappa_0 = 10, \kappa = 1$); S2 on $S^2$. (b) Independent directions with high concentrations ($\kappa_{0k} = 100, \kappa_k = 10, k = 1, 2, 3$); iMS2 on ($S^2$)$^3$. (c) Horizontally dependent directions with high concentrations ($\kappa_{0k} = 50, \kappa_k = 30$) and high dependence ($\lambda_{12} = 24$); MS2 on ($S^2$)$^2$ [Colour figure can be viewed at wileyonlinelibrary.com]

3 | MAXIMUM LIKELIHOOD ESTIMATION

The algorithms developed below converge quickly to a local maximum of the likelihood function in all of the data situations of this paper. In the Online Supplementary Material (Section A3), we give an example of our new quick algorithm for the normalizing constant for S1.

3.1 | Estimation for S1 and iMS1 models

The standard way to estimate the parameters of S1 is to use the maximum likelihood estimates (MLE). However, it does not seem possible to obtain explicit expressions of the MLE, partly due to having no closed-form expression of the normalizing constant (5). We propose to approximate the normalizing constant and numerically obtain the MLE. Our procedure naturally extends to estimation for the iMS1 distribution, which will also be discussed.

As a preparation, we first describe an approximation of the normalizing constant, following Kume and Wood (2005). The exact calculation of the normalizing constant of the Fisher–Bingham distribution, including S1, is possible (Kume & Sei, 2018) but computationally very expensive. Thus, to facilitate applicability, we chose to use an approximation of the normalizing constant.

The normalizing constant of S1 has an alternative expression, as shown in the following.

Proposition 2. For any $h > 0$, let $\xi = ((2\kappa_0 + \kappa_1)/2h, \kappa_2\sqrt{1-\kappa_0^2}/2h, 0, \ldots, 0)^T \in \mathbb{R}^p$, and let $\Psi$ be the $p \times p$ diagonal matrix with diagonal elements $(\kappa_0 + h, h, \ldots, h)$. Moreover, let $g(r) (r > 0)$ be the probability density function of $R = Z^T Z$, where $Z \sim N_p(\xi, 1/2\Psi^{-1})$. Then, the normalizing constant $a(\kappa_0, \kappa_1, \nu)$ of (5) is

$$a(\kappa_0, \kappa_1, \nu) = 2\pi^{p/2}|\Psi|^{-1/2}g(1) \exp \left(\xi^T \Psi \xi + h - \kappa_0 \nu^2\right).$$

In Proposition 2, the function $g$ is the density of a linear combination of independent noncentral $\chi^2_1$ random variables. Following Kume and Wood (2005), we use saddle-point density approximations in the numerical computation of $g(1)$. First, note that the derivatives of
the cumulant-generating function, \( K_y(t) = \log \int_0^\infty e^{tr} g(r) dr \), associated with density \( g \) have closed-form expressions. Denoting by \( K_y^{(j)}(t) \) the jth derivative of \( K_y(t) \), for \( j = 1, \ldots, 4 \), we get

\[
K_y^{(j)}(t) = \frac{(j - 1)!}{2} \left( \frac{1}{(\kappa_0 + h - t)^j} + \frac{p - 1}{(h - t)^j} \right) + \frac{j!}{4} \left( \frac{\nu^2(2\kappa_0 + \kappa_1)^2}{(\kappa_0 + h - t)^{j+1}} + \frac{\kappa_1^2(1 - \nu^2)}{(h - t)^{j+1}} \right).
\]

Let \( \hat{t} \) be the unique solution in \((-\infty, h)\) of the saddle-point equation \( K_y^{(1)}(t) = 1 \), which can be easily evaluated by using, for example, a bisection method. Then, a saddle-point density approximation of \( g(1) \) is

\[
\hat{g}(1) = \left(2\pi K_y^{(2)}(\hat{t})\right)^{-1/2} \exp(K_y(\hat{t}) - \hat{t} + T),
\]

where \( T = K_y^{(4)}(\hat{t})/[8(K_y^{(2)}(\hat{t}))^2] - 5(K_y^{(3)}(\hat{t}))^2/[24(K_y^{(2)}(\hat{t}))^3] \). In the following, we approximate the value of \( a(\kappa_0, \kappa_1, \nu) \) by \( \hat{a}(\kappa_0, \kappa_1, \nu) \) obtained by plugging (18) in place of \( g(1) \) in (17).

We are now ready to describe our estimation procedure. Suppose \( x_1, \ldots, x_n \) is a sample from \( S(\mu_0, \mu_1, \kappa_0, \kappa_1) \), and let \( \ell_n(\mu_0, \mu_1, \kappa_0, \kappa_1, \nu) \) be the log likelihood.

Suppose for now that \( \nu \in (0, 1) \) is fixed. Then, the MLE of \( \mu_0 \) and \( \mu_1 \) can be efficiently estimated. In particular, maximizing the likelihood function with respect to \( \mu_0 \) is equivalent to minimizing \( \frac{1}{n} \sum_{i=1}^n (\mu_0^T x_i - \nu)^2 \) subject to the constraint \( \mu_0^T \mu_0 = 1 \). With a Lagrangian multiplier \( \lambda \) using matrix notation, we solve

\[
\min_{\mu_0 \in \mathbb{S}^{p-1}} \left[ \frac{1}{n} \| \mathbb{X}^T \mu_0 - \nu 1_n \|^2 - \lambda (\mu_0^T \mu_0 - 1) \right],
\]

where \( \mathbb{X} \) is the \( p \times n \) matrix whose ith column is \( x_i \), yielding the necessary condition \( S \mu_0 - \nu \mathbb{X} - \lambda \mu_0 = 0 \), where \( S = \mathbb{X} \mathbb{X}^T / n, \mathbb{X} = \frac{1}{n} \sum_{i=1}^n x_i \). For a fixed Lagrangian multiplier \( \lambda \), the solution is \( \hat{\mu}_0 = \nu(S - \lambda I_p)^{-1} \hat{x} \), provided that \( S \) is of full rank. The constraint \( \mu_0^T \mu_0 = 1 \) makes us find a root \( \lambda \) of \( \nu^2 \hat{x}^T (S - \lambda I_p)^{-2} \hat{x} - 1 \). The root \( \hat{\lambda} \) is found by a bisection search in the range \([-\nu^2 \hat{x}^T \hat{x}, \lambda_S]\), where \( \lambda_S > 0 \) is the smallest eigenvalue of \( S \) (Browne, 1967). The solution to (19) is then

\[
\hat{\mu}_0 = \nu(S - \hat{\lambda} I_p)^{-1} \hat{x}.
\]

If \( \nu = 0 \), then \( \hat{\mu}_0 \) is the eigenvector of \( S \) corresponding to the smallest eigenvalue.

Now, with \( \nu \) and \( \hat{\mu}_0 \) (20) given, maximizing the likelihood with respect to \( \mu_1 \) is equivalent to maximizing \( \frac{1}{n} \sum_{i=1}^n \mu_1^T x_i \) subject to the constraints \( \mu_0^T \mu_1 = \nu \) and \( \mu_1^T \mu_1 = 1 \). It can be shown that the MLE of \( \mu_1 \) is a linear combination of \( \mu_0 \) and \( \hat{x} / \| \hat{x} \| \), where \( \hat{x} = \frac{1}{n} \sum_{i=1}^n x_i \). Let \( \hat{x}^* = P_{\mu_0} \hat{x} / \| P_{\mu_0} \hat{x} \| \), then the MLE of \( \mu_1 \) given \( \nu \) and \( \hat{\mu}_0 \) is

\[
\hat{\mu}_1 = \nu \hat{\mu}_0 + \sqrt{1 - \nu^2} \hat{x}^*.
\]

Thus, the MLE of \((\kappa_0, \kappa_1, \nu)\) is

\[
(\hat{k}_0, \hat{k}_1, \hat{\nu}) = \arg \max_{k_0, k_1, \nu} \ell_n(\hat{\mu}_0, \hat{\mu}_1, \nu, \kappa_0, \kappa_1, \nu),
\]

which is solved by a standard optimization package, and the MLE of \((\mu_0, \mu_1)\) is given by (20) and (21) with \( \nu \) replaced by \( \hat{\nu} \). This procedure, beginning with an initial value for \( \hat{\nu} \), is iterated until convergence.

Let us now describe an extension of the above algorithm to the iMS1 model. Suppose \( (x_{i1}, \ldots, x_{iK}) \in (S^{p-1})^K \) for \( i = 1, \ldots, n \) is a sample from an iMS1 model, where each marginal
distribution is $S_1(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$. While the last step above can be applied to estimate $\kappa_{0k}, \kappa_k, v_k$, for $k = 1, \ldots, K$, given $\mu_0$ and $\mu_k$’s, we replace (19) with

$$\min_{\mu_0 \in \mathbb{R}^p} \left[ \frac{1}{n} \sum_{k=1}^{K} (\kappa_{0k} \| \mathbf{x}_k - v_k 1_n \|^2) - \lambda \left( \mu_0 \mu_0 - 1 \right) \right],$$

where the marginal $p \times n$ observation matrix $\mathbf{x}_k$ has the columns $x_{ik}$ ($i = 1, \ldots, n$). This is solved with the obvious analog to (20). For $\mu_k$’s, the above solution (21) can be applied for each $k$ separately.

### 3.2 Estimation for S2, iMS2, and MS2

The S2 model and its extensions have the convenient property that the horizontal components are independent of the vertical ones. To take advantage of this, suppose for now that $\mu_0$ is known. This allows us to decompose an observation $x$ into two independent random variables $(s, y)$, which, in turn, leads to an easy estimation of the remaining parameters $\eta := (\mu, \kappa_0, \kappa_1)$. Thus, the estimation strategy proceeds in two nested steps. Let $\ell_n(\mu_0, \eta)$ be the log-likelihood function given a sample $x_1, \ldots, x_n$ from $S_2(\mu, \eta)$. In the outer step, we update $\mu_0$ to maximize the likelihood, that is,

$$\hat{\mu}_0 = \arg \max_{\mu_0} \ell_n(\mu_0, \hat{\eta}_{\mu_0}),$$

(22)

where evaluating

$$\hat{\eta}_{\mu_0} = \arg \max_{\eta} \ell_n(\mu_0, \eta)$$

(23)

for a fixed $\mu_0$ is the inner step. It is straightforward to see that the MLE of $(\mu_0, \eta)$ is given by $(\hat{\mu}_0, \hat{\eta}_{\mu_0})$.

In the following, we discuss in detail the inner step (23) of maximizing $\ell_n(\mu_0, \eta) := \ell_n(\mu_0, \eta)$ for the iMS2 model (12) and for the MS2 model (15), whereas we resort to a standard optimization package for solving (22).

**Independent multivariate S2 (iMS2).** Suppose $(x_{i1}, \ldots, x_{iK}) \in (S^{p-1})^K$ for $i = 1, \ldots, n$ is a sample from an iMS2, where each marginal distribution is $S_2(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$. For a given $\mu_0$, the joint density can be written in terms of $(s_i, \phi_i)$ as done in (15), but with $\Lambda = 0$. Furthermore, by the definition of $H$ and $K_0$ we used in (15), we can write

$$H^T s_i - s_i^T K_0 s_i = -\sum_{k=1}^{K} \kappa_{0k}(s_{ik} - v_k)^2 + \sum_{k=1}^{K} \kappa_{0k} v_k^2,$$

and hence,

$$\log [T_1(H, K_0)T_2(K, 0)] = \sum_{k=1}^{K} [\log b(\kappa_{0k}, \kappa_k, v_k) + \kappa_{0k} v_k^2].$$

Note that the normalizing constant $b(\kappa_{0}, \kappa_1, \nu)$ satisfies

$$b(\kappa_{0}, \kappa_1, \nu) = \int_{0}^{\pi} e^{\kappa_0 \cos \phi} d\phi \int_{-1}^{1} e^{-\kappa_0 (1-s)^2} ds$$

$$= (2\pi)^{3/2}(2\kappa_0)^{-1/2} I_0(\kappa_1) \left[ \Phi((1 - \nu)\sqrt{2\kappa_0}) - \Phi(-(1 + \nu)\sqrt{2\kappa_0}) \right].$$

where $\Phi(\cdot)$ is the standard normal distribution function. Finally, the log-likelihood function (given $\mu_0$) is, for $\kappa_0 = (\kappa_{01}, \ldots, \kappa_{0K})^T$, $\nu = (v_1, \ldots, v_K)^T$,

$$\ell_{\mu_0}(\nu, \zeta, \kappa_0, \kappa; \{s_i, \phi_i\}_{i=1}^n) = \ell_{\mu_0}^{(1)}(\nu, \kappa_0) + \ell_{\mu_0}^{(2)}(\zeta, \kappa).$$

(24)
where

\[
\ell^{(1)}_{\mu_0}(\nu, \kappa_0) = -\sum_{k=1}^{K} \left[ \kappa_{0k} \sum_{i=1}^{n} (s_{ik} - \nu_k)^2 - \frac{n}{2} \log(2\kappa_{0k}) + \frac{n}{2} \log(2\pi) \right. \\
+ \left. n \log \left( \Phi \left( (1 - \nu_k)\sqrt{2\kappa_{0k}} \right) - \Phi \left( -(1 + \nu_k)\sqrt{2\kappa_{0k}} \right) \right) \right],
\]

\[
\ell^{(2)}_{\mu_0}(\zeta, \kappa) = \sum_{k=1}^{K} \left[ \kappa_k \sum_{i=1}^{n} \cos(\phi_{ik} - \zeta_k) - n \log I_0(\kappa_k) - n \log(2\pi) \right].
\]

Therefore, the optimization for the inner step (23) is equivalent to simultaneously solving 2K subproblems.

Each of the K subproblems of (23) is equivalent to obtaining the MLE of a truncated normal distribution \( \text{trN}(\nu_k, (2\kappa_{0k})^{-1/2}; (-1, 1)) \) based on the observations \( s_{ik} \ (i = 1, \ldots, n) \). Similarly, each of the K subproblems of (26) amounts to obtaining the MLE of a von Mises distribution with mean \( \zeta_k \) and concentration \( \kappa_k \) from the sample \( \phi_{ik} \ (i = 1, \ldots, n) \). The MLEs of the truncated normal are numerically computed, and we use the method of Banerjee, Dhillon, Ghosh, and Sra (2005) to obtain approximations of the MLEs of von Mises.

**MS2.** Under the MS2 model (15) with a dependence structure on \( \phi_i \), a decomposition \( \ell_{\mu_0}(\nu, \zeta, \kappa_0, \kappa, \Lambda) = \ell^{(1)}_{\mu_0}(\nu, \kappa_0) + \ell^{(2)}_{\mu_0}(\zeta, \kappa, \Lambda) \), similar to (24), is valid, where (26) is replaced by

\[
\ell^{(2)}_{\mu_0}(\zeta, \kappa, \Lambda) = -\sum_{i=1}^{n} \left[ \kappa^T c(\phi_i, \zeta) + \frac{1}{2} s(\phi_i, \zeta)^T \Lambda s(\phi_i, \zeta) - \log T_3(\kappa, \Lambda) \right].
\]

Maximizing (27) is equivalent to computing the MLE of the multivariate von Mises distribution (Mardia et al., 2008). We use either the maximum pseudo-likelihood estimate as discussed in Mardia et al. (2008) or moment estimates, yielding

\[
\hat{\zeta}_k = \frac{1}{n} \sum_{i=1}^{n} \phi_{ik}, \quad \hat{\kappa}_k = \frac{1}{n} \sum_{i=1}^{n} \phi_{ik}^2, \quad \hat{\kappa}_{kl} = \frac{1}{n} \sum_{i=1}^{n} \sin(\phi_{ik} - \hat{\zeta}_k) \sin(\phi_{il} - \hat{\zeta}_l) \quad (k \neq l),
\]

where \( \bar{S} = (\bar{S}_{kl}) \) and \( \bar{S}_{kl} = \frac{1}{n} \sum_{i=1}^{n} \sin(\phi_{ik} - \hat{\zeta}_k) \sin(\phi_{il} - \hat{\zeta}_l) \) for \( k, l = 1, \ldots, K \). These estimates coincide with the MLEs when \( K = 2 \). For larger \( K > 3 \), the accuracy of the moment estimates deteriorates, but evaluating MLEs or a maximum pseudo-likelihood estimator becomes computationally highly expensive.

## 4 \ TESTING HYPOTHESES

It is of interest to infer on the parameters of our models. While we adopt the well-known likelihood ratio test, we emphasize that the proposed model enables us to test several important hypotheses, which have been of interest to some researchers, and that our estimation procedure can be easily adapted to compute maximized likelihood under the null parameter space \( \Theta_0 \). Recall that the parameter space for the iMS1 and iMS2 models is given by \( \Theta_\text{ind} = S^{p-1} \times S^{(p-1)K} \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K \) for \( \theta_\text{ind} = (\mu_0, \mu, \kappa_0, \kappa) \). For conciseness, we describe our testing procedure using the MS2 distribution in dimension \( p = 3 \), whose parameter space is \( \Theta = (\Theta_\text{ind} \times (\mathbb{R})^{K(K-1)/2}) \) for \( \theta = (\theta_\text{ind}, \Lambda) \). For some \( \Theta_0 \) that dictates a null hypothesis \( H_0 \) and satisfies \( \Theta_0 \subset \Theta \), we denote the maximized log
likelihood under $\Theta_0$ by $\ell_0$ and the maximized log likelihood under $\Theta$ by $\ell_1$. Our test statistic is $W_n := -2(\ell_0 - \ell_1)$, and $H_0$ is rejected for large enough $W_n$.

We list a few null hypotheses of practical interest, with the alternative being the full MS2 distribution. In all three cases below, the alternative hypothesis is $H_1 : \theta \in \Theta \setminus \Theta_0$.

1. **Test of association.** $H_0 : \Lambda = 0$, that is, there is no horizontal dependence.
2. **Test of axis.** $H_0 : \mu_0 = \mu_0^*$, that is, the common axis of rotation is $\mu_0^*$.
3. **Test of great sphere.** $H_0 : \nu = 0$, that is, the underlying subsphere is indeed a great subsphere.

While the test of association is only available under the MS2 model ($p = 3$), Hypotheses 2 and 3 can also be tested using the S1, S2, iMS1, or iMS2 model in any dimension $p \geq 3$. To validate the use of small-sphere distributions, in any dimension $p \geq 3$, the following hypotheses can be tested.

4. **Test for the vMF distribution.** $H_0 : \kappa_0 = 0$, that is, there is no “small-circle feature.”
5. **Test for the BM distribution.** $H_0 : \kappa = 0$, that is, there is no unique mode.

For each hypothesis, computing the test statistic $W_n$ requires maximizing the likelihood on $\Theta_0$ (or to compute $\ell_0$). This is easily achieved by modifying the iterative algorithms in Section 3. For example, for the test of association, computing $\ell_0$ and $\ell_1$ amounts to obtaining the MLEs under the iMS2 and MS2 models, respectively; for Hypothesis 2 (test of axis), where $\mu_0$ is given, one only needs to solve (23) once. Other cases of restricted MLEs can be easily obtained. In the Online Supplementary Material (Section A4), we confirm that the test statistic $W_n$ using our algorithms under the null hypotheses above is empirically nearly chi-square distributed for sample size $n = 30$. In Section 5.3 and in the Online Supplementary Material, empirical powers of the proposed test procedures are reported for several important alternatives.

## 5 | NUMERICAL STUDIES

We demonstrate the performances of small-circle fitting in Section 5.1, the ability of MS2 in modeling the horizontal dependence in Section 5.2, and a testing procedure to prevent overfitting in Section 5.3.

### 5.1 | Estimation of small circles

The performance of our estimators in fitting the underlying small spheres $C(\mu_0, \nu)$ is numerically compared with those of competing estimators obtained from assuming the BM distribution and the LS methods of Schulz et al. (2015). The BM distribution has originally been defined only for data on $S^2$, but we use a natural extension given by a special case of iMS1. Thus, “BM estimates” refer to the estimates of the iMS1 model with the restriction $\kappa = 0$. The estimates of Schulz et al. are obtained by minimizing the sum-of-squared angular distances from observations to $C(\hat{\mu}_0, \hat{\nu})$, which will be referred to as an LS method.

We first consider four univariate S2 models to simulate data concentrated on a single small circle in order to benchmark against existing BM and LS. The directional parameters $(\mu_0, \mu_1)$ are set to satisfy $\nu = 0.5$. We use $(\kappa_0, \kappa_1) = (10, 1), (100, 1), (100, 10)$ to represent various data situations. Random samples from these three settings are shown in Figure 4a–c. We also consider the BM model as a special case of the S2 distributions (by setting $\kappa_1 = 0$); a sample from the BM distribution is shown in Figure 4d.
FIGURE 4  Random samples of size $n = 50$ from the S2 model on $S^2$ used in our simulations. Small-circle estimation performances are reported in Table 2. (a) $\kappa_0 = 10, \kappa_1 = 1$. (b) $\kappa_0 = 100, \kappa_1 = 1$. (c) $\kappa_0 = 100, \kappa_1 = 10$. (d) $\kappa_0 = 100, \kappa_1 = 0$ [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2  Small-circle estimation performances for univariate data on $S^2$ from Figure 4. Means (standard deviations) of the angular product errors in degrees (28) are shown

| Method | (a) | (b) | (c) | (d) |
|--------|-----|-----|-----|-----|
| S1     | 5.81(3.17) | 1.46(0.72) | 15.56(11.68) | 1.35(0.72) |
| S2     | 5.24(3.14) | 1.44(0.70) | 14.36(11.91) | 1.36(0.72) |
| BM     | 8.37(6.43) | 1.47(0.70) | 17.00(15.68) | 1.35(0.72) |
| LS     | 5.83(3.74) | 1.45(0.69) | 15.27(14.16) | 1.36(0.73) |

Note. S1 = small-sphere distribution of the first kind; S2 = small-sphere distribution of the second kind; BM = Bingham–Mardia distribution; LS = least squares method.

The small-circle estimation performances of the S1, S2, BM, and LS estimates are measured by an angular product error (in degrees), defined as

$$L ((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu})) = \left( \text{Angle}(\hat{\mu}_0, \mu_0)^2 + \frac{180}{\pi} (\arccos \hat{\nu} - \arccos \nu)^2 \right)^{1/2} .$$

Table 2 displays the means and standard deviations of $L((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu}))$ from 100 repetitions for each of the four methods, fitted to random samples of size 50 from each of the settings, labeled (a)–(d). Since horizontal/vertical dependence is not of concern, we expect that S1, S2, and LS perform similarly and that they strongly outperform BM when the small circle features a distinct mode (c). This is indeed the case. Remarkably, in the high-vertical-noise case (a), the standard deviations of S1 and S2 are considerably lower than that of BM and still notably smaller than that of LS.

Next, to show the performance of our multivariate models, we consider six bivariate MS2 models. The directional parameters $(\mu_0, \mu)$ were set to satisfy $\nu = (0.5, -0.3)$, and the concentration parameters were chosen to mimic the concentrations of the univariate models, described above. For cases (a)–(c), we set $(\kappa_{0j}, \kappa_j, \lambda_{12j}) = (10, 1, 0), (100, 1, 0), (100, 10, 0)$, for $j = 1, 2$, so that the models are indeed the iMS2. For the latter three cases (d)–(f), we set $(\kappa_{0j}, \kappa_j, \lambda_{12j}) = (10, 2, 1.5), (100, 2, 1.5), (100, 20, 15), j = 1, 2$, to make their vertical and horizontal dispersions be similar to the iMS2 counterparts. By setting $\lambda_{12j} > 0$, the random bivariate directions are positively associated. (Examples of random samples from these settings can be found in Figure A7 in the Online Supplementary Material.) The small-circle estimation performance of the iMS1, iMS2, MS2, BM, and LS estimates is measured by the canonical multivariate extension of the angular product error (28). Table 3 collects the means and standard deviations of the angular product
TABLE 3  Small-circle estimation performances for bivariate data on $S^2 \times S^2$ from Figure A7 in the Online Supplementary Material. Means (standard deviations) of the angular product errors in degrees (28) are shown

| Method | Independent | Dependent |
|--------|-------------|-----------|
|        | (a)         | (b)       | (c)       | (d)       | (e)       | (f)       |
| iMS1   | 4.84(1.75)  | **1.51(0.56)** | **3.90(3.61)** | 6.21(2.73) | 1.58(0.76) | 4.91(5.82) |
| iMS2   | 4.49(1.80)  | 1.52(0.53) | 4.30(2.46) | 5.90(2.60) | 1.58(0.75) | 4.55(2.56) |
| MS2    | **4.47(1.78)** | 1.52(0.53) | 4.26(2.45) | **5.81(2.67)** | **1.57(0.75)** | **4.50(2.48)** |
| BM     | 5.30(2.21)  | 1.54(0.54) | 8.63(4.66) | 13.87(15.42) | 1.69(0.83) | 9.26(5.25) |
| LS     | 4.76(1.85)  | 1.52(0.55) | 4.30(2.35) | 6.68(2.77) | 1.59(0.71) | 4.53(2.44) |

Note. iMS1 = independent multivariate S1 distribution; iMS2 = independent multivariate S2 distribution; MS2 = multivariate S2 distribution; BM = Bingham–Mardia distribution; LS = least squares method.

errors from 100 repetitions with the sample size $n = 50$. Now, BM clearly performs the worst, also in low-horizontal-concentration scenarios (a), (b), (d), and (e). Since BM is able to model vertical but not horizontal concentration, somewhat unexpected, BM performs considerably worse under additional high vertical noise in (a) and (d), as compared to (b) and (e). The same phenomenon, but more subtly, is also visible for LS. In the case of high vertical concentration (all but (a) and (d)), LS performs comparable to the new parametric models. In particular, in the case of additional horizontal dependence with high concentration in (f), it outperforms the independent parametric models, and it is as good as MS2. In the case of low vertical concentration in (a), however, MS2 is superior; MS2 is considerably superior under additional horizontal dependence in (d).

We check robustness against the model misspecification of the estimators by simulating data from a more general signal-plus-noise model (neither S1 nor S2). The performances of small-circle fitting of the proposed methods are comparable to that of the LS estimator. Relevant simulation results and a detailed discussion can be found in the Online Supplementary Material.

5.2 Estimation of horizontal dependence

The ability of MS2 to model the horizontal dependence is an important feature of the proposed distributions. We emphasize that MS2 is the only method modeling such dependence; hence, in this section, we can only validate it against iMS2. Here, we empirically confirm that the MS2 estimates provide accurate measures of horizontal dependence, using cases (c) and (f) in Section 5.1. For sample sizes $n = 50$ and 200, the concentration and association parameters were estimated under the assumption of MS2 (or iMS2), and Table 4 summarizes the estimation accuracy. In all cases, the MS2 model provides precise estimations of the horizontal dispersion and dependence; as the sample size increases, the mean squared error decreases. For case (c), the underlying model is exactly iMS2; hence, the iMS2 estimates have smaller mean squared errors than the MS2 estimates. However, for case (f), we notice that the iMS2 estimates of $\kappa = (\kappa_1, \kappa_2)$ become inferior. In fact, in the case of existing horizontal dependence, that is, when $\lambda_{12} \neq 0$, the concentration parameters $\kappa$ in the misspecified iMS2 model do not correctly represent the concentrations as correctly represented by the MS2 model. This is so because the marginal distribution of $\varphi_j$, $j = 1, 2$, in (15) is not a von Mises distribution (Singh, Hnizdo, & Demchuk, 2002; Mardia et al., 2008).
TABLE 4  Small-circle estimation performances for bivariate data on $S^2 \times S^2$ from Figure A7 in the Online Supplementary Material. Means (standard deviations) of the angular product errors in degrees (28) are shown.

| n  | Method | $\kappa_1 = 10$ | $\kappa_2 = 10$ | $\lambda_{12} = 0$ | $\kappa_1 = 20$ | $\kappa_2 = 20$ | $\lambda_{12} = 15$ |
|----|--------|----------------|----------------|-----------------|----------------|----------------|----------------|
| 50 | iMS2   | 10.23(2.48)    | 10.54(2.16)    | 11.73(2.19)     | 11.19(2.14)    |
|    | MS2    | 10.48(2.54)    | 10.80(2.27)    | -0.17(1.85)     | 22.63(5.06)    | 21.40(4.32)    | 16.82(4.22)    |
| 200| iMS2   | 10.31(1.08)    | 10.10(1.04)    | 11.00(1.05)     | 11.09(1.04)    |
|    | MS2    | 10.35(1.10)    | 10.14(1.05)    | -0.12(0.73)     | 20.38(2.17)    | 20.54(2.25)    | 15.41(2.06)    |

Note. iMS2 = independent multivariate S2 distribution; MS2 = multivariate S2 distribution.

FIGURE 5  Degrees of the “small-circle feature.” Shown are random samples from an isotropic distribution (case (a)) and the S2 distributions with increasing “small-circle concentrations” (cases (b)–(d)). (a) vMF ($\kappa_1 = 10$). (b) ($\kappa_0, \kappa_1$) = (20, 10). (c) ($\kappa_0, \kappa_1$) = (100, 10). (d) ($\kappa_0, \kappa_1$) = (100, 1) [Colour figure can be viewed at wileyonlinelibrary.com]

5.3 | Detecting overfitting in an isotropic case

When the data do not exhibit a strong tendency of a small-circle feature, the S1 and S2 distributions may overfit the data. For example, to a random sample from an isotropic vMF distribution, as shown in Figure 5a, the S1 or the S2 model fits an unnecessary small circle $C(\hat{\mu}_0, \hat{\nu})$. Indeed, a small-circle fit was observed in 83% of simulations of fitting the S1 model. Using the BM or LS results in a similar overfitting, where very small circles are erroneously fitted for 100% and 68% of the simulations, for the BM and LS, respectively.

This problem of overfitting has been known for a while and discussed in the context of dimension reduction of directional data. In particular, Jung et al. (2011), Jung et al. (2012), and Eltzner, Huckemann, and Mardia (2018) investigated the overfitting phenomenon for the LS estimates and proposed some ad hoc methods for adjustment. To prevent overfitting, we point out that the testing procedure in Section 4 for the detection of isotropic distributions (Hypothesis 4) works well. To confirm this, we evaluated the empirical power of the test at the significance level $\alpha = 0.05$ for several alternatives. The power increases sharply as the distributions become more anisotropic; under the alternative distributions depicted in Figure 5b–d, the empirical powers are, respectively, $\hat{\beta} = 0.435$, 1 and 1, evaluated from 200 repetitions.

6 | ANALYSIS OF S-REP DATA

In this section, an application of the proposed distributions and test procedures to s-rep data is discussed.
6.1 Modeling rotationally deformed ellipsoids via s-reps

Skeletal representations (s-reps) have been useful in the mathematical modeling of human anatomical objects (Siddiqi & Pizer, 2008). Roughly, an s-rep model for a 3D object consists of locations of a skeletal mesh (inside the object) and spoke vectors (directions and lengths), connecting the skeletal mesh with the boundary of the object; examples are shown in the top-left panel of Figure 6. When the object is “rotationally deformed,” Schulz et al. (2015) have shown that the directional vectors of an s-rep model approximately trace a set of concentric small circles on $S^2$, as shown in the top panels of the figure. Such rotational deformations (e.g., rotation, bending, and twisting) of human anatomical objects have been observed in between and within shape variations of hippocampi and prostates (Joshi et al., 2002; Jung et al., 2011; Pizer et al., 2013). We demonstrate a use of the MS2 distribution in modeling (and fitting) a population of such objects via s-reps. Note that the sample space of an s-rep with $K$ spokes is $(S^2)^K \times \mathbb{R}^K_+ \times (\mathbb{R}^3)^K$ (for direction, length, and location). In this work, we choose to analyze the spoke directions in $(S^2)^K$ only, leaving a full-on analysis, accommodating the lengths and locations, to future work.
6.2 Data preparation

For our purpose of validating the use of the proposed distributions, we use an s-rep data set, fitted from \( n = 30 \) deformed ellipsoids; two samples from this data set are shown in Figure 6a. This data set was previously used in Schulz et al. (2015) as a simple experimental representation of real human organs. The data set was generated by “physically bending” a template ellipsoid about an axis \( \mu^* = (0, 1, 0) \) by random angles drawn from a normal distribution with standard deviation 0.4 (radians). Each deformed ellipsoid is then recorded as a 3D binary image. To mimic the procedure of fitting s-reps from, for example, the medical resonance imaging of a real patient, s-reps with 74 spokes were fitted to these binary images. (See Pizer et al., 2013 for details of the s-rep fitting.) As a preprocessing, we chose \( K = 58 \) spoke vectors, excluding the vectors with very small total variation.

6.3 Inference on the bending axis

Fitting the iMS1 distribution, we obtained the axis estimate \( \hat{\mu}_{0}^{(iMS1)} = (0.007, 1.000, -0.008) \) (rounded to three decimal digits). Similarly, from the MS2 fitting, \( \hat{\mu}_{0}^{(MS2)} = (0.006, 1.000, 0.006) \). The LS fit of Schulz et al. (2015) also provides a similar estimate. These estimates are virtually the same, only 0.6 degrees away from the ground truth \( \mu^* \). Estimates of the concentric small circles \( C(\hat{\mu}_{0}^{(MS2)}, \hat{\nu}_j) \) for four choices of \( j \) (the spoke index) are also shown in the top-right panel of Figure 6, in which \( \hat{\mu}_{0}^{(MS2)} \) and \( \mu^* \) are also shown. Although all methods provide virtually the same estimate, only by assuming one of the iMS1, iMS2, and MS2 models are we able to infer on the axis of rotation. For example, under the iMS1 model, we tested \( H_0 : \mu_0 = \mu^* \), and with the \( p \) value of 0.26, we accept that the true axis of rotation is the hypothesized axis \( \mu^* \).

6.4 Inference on horizontal dependence

An advantage of modeling the s-rep spoke directions by the MS2 distribution is the ability of perceiving and modeling the horizontal dependence among directions. As an exploratory step, we have collected the estimated correlation coefficients, computed from the approximate precision matrix \( \hat{\Sigma}^{-1} \), whose elements are \( \kappa \) and \( \hat{\alpha} \) (see (16)). A histogram of \( K(K - 1)/2 \) estimated correlation coefficients is plotted in the bottom-left panel of Figure 6. Notably, pairs of spoke vectors from the same side (e.g., two spoke vectors on the “left side” of the ellipsoids in Figure 6) exhibit strong positive correlations, whereas those from the opposite sides exhibit strong negative correlations. The horizontal dependence is, in fact, apparent from the way data were generated (simultaneously bending all the spoke directions).

We point out that, due to obvious dependence among multivariate directions, to date, MS2 is the only meaningful model for these data. The iMS1 and the multivariate extension of the BM are not capable of modeling such association. We check that when goodness-of-fit tests based on Jupp (2005) and Székely and Rizzo (2013) are applied to the s-rep data, the tests reject all other distributions except MS2. (See the Online Supplementary Material [Section A7] for details.)

For large enough sample sizes, we could use the test of association discussed in Section 4 for testing \( H_0 : \Lambda = \mathbf{0} \). Unfortunately, due to our small sample size, \( n = 30 \), and the large number of parameters tested, 1653 (= \( K(K - 1)/2 \)), this is infeasible. Coping with this high-dimension, low-sample-size situation is beyond the scope of this paper, and we resort to choosing only two spoke directions to test the dependence, but repeating the testing for many different pairs of the total \( K = 58 \) spokes. For each pair of spokes, the likelihood-ratio test produces a \( p \) value for the
pair. Investigating the empirical distribution of these \( p \) values can provide a rough estimate of the power. In Figure 6d, it can be seen that, at the significance level of 0.05, the MS2 test of dependence is indeed powerful, with a rejection rate of 98.5%.

To provide some context to this rate, the MS2 test was compared with other natural choices of tests. We applied two methods that were previously used for s-rep data analysis: the CPNS, discussed in Pizer et al. (2013), and the LS (concentric) small-circle fitting method of Schulz et al. (2015).

The CPNS test is built as follows. First, the LS small circle is fitted to each marginal direction on \( S^2 \). With an understanding that the axis of the fitted small circle points to the north pole, the observations (say, \( x_{ik} \) from the \( i \)th sample, \( k \)th spoke) are represented in spherical coordinates (\( \theta_{ik}, \phi_{ik} \)). For the purpose of testing “horizontal associations,” we only keep the longitudinal coordinates \( \theta_{ik} \). For any given pair \((k, k')\), Fisher’s z-transformation is used to obtain the \( p \) value in testing whether the correlation coefficient between \( \theta_{ik} \) and \( \theta_{ik'} \) is zero. We refer to this test procedure by a CPNS test.

An LS test procedure is defined similarly to the CPNS test, except that the first step of fitting individual small circles is replaced by fitting concentric small circles.

These two tests were also conducted for the same combinations of spoke directions, and the empirical distributions of the respective \( p \) values are also plotted in Figure 6. These alternative tests appear to be too conservative, with rejection rates of 11% for the LS test and 11% for the CPNS test (at level 0.05). Heuristically, the higher power of the MS2 test is due to the superior fitting of the MS2 distribution. In particular, the “horizontal angles” predicted from MS2 tend to be linearly associated, whereas those from the LS fit tend to be arbitrary. We refer to the Online Supplementary Material for more numerical results. All in all, using the MS2 distribution shows a clear advantage in modeling and testing the horizontal dependence of multivariate directions.

7 | HUMAN KNEE GAIT ANALYSIS

In biomechanical gait analysis, accurately modeling human knee motion during normal walking has a potential to differentiate diseased subjects from normal subjects. In particular, the axis of bending (of the lower leg toward the upper leg) is believed to be a key feature in the discrimination among the diseased and normal subjects (Pierrynowski, Costigan, Maly, & Kim, 2010). As a step toward the development of statistical tests for a “two-group axis difference,” in this section, we employ the proposed distributional families in modeling the bending motion of the knee.

The raw data set we use is obtained from a healthy volunteer, and it is a time series of coordinates of markers planted at the volunteer’s leg, recorded for 16 gait cycles. For each time point, the directional vectors on \((S^2)^5\) were computed to be the unit vectors between reference markers, as done in Schulz et al. (2015). As evident from Figure 7, these directional vectors are horizontally dependent of each other, which suggests that we can only fit the MS2 distribution.

The first panel in Figure 7 illustrates the result of MS2 fit to all data points \((n = 1000)\). There, we superimpose the fitted concentric circles to the observed directional vectors, including their estimated axis, together with a hypothesized dominant bending axis \( \mu^*_0 = (0, 1, 0)^\top \), the left–right axis of the subject. The axis estimates from the iMS1 or the LS method also provided similar estimates. The MS2 model seems to fit well with high estimated horizontal correlation
coefficients. We, however, identify seemingly strong evidence against using a single MS2. Specifically, as shown in Figure 7, some directional vectors exhibit higher variations for a subset of time points.

In fact, the data consist of many inhomogeneous periods of the gait cycle. We focus on the “swing” and “stance” periods and separately analyze subsampled data from each period ($n_{sw} = 66$ and $n_{st} = 118$). The MS2 model fits well for the swing period data (See the Online Supplementary Material for a goodness-of-fit analysis.) The axis of swing is estimated at $\hat{\mu}^{(sw)}_0 = (0.013, 1.000, 0.005)^T$, virtually the same as the hypothesized axis $\mu^*_0$. With the $p$ value of 0.16, we accept that the axis of swing is the left–right axis of this healthy person. For the stance period, excluding the highly irregular directions shown as dark-blue points in Figure 7c, the MS2 model also fits well. We confirm that in stance, the axis of major rotation differs from $\mu^*_0$ with the $p$ value less than $10^{-5}$ in testing $H_0 : \mu_0 = \mu^*_0$. The estimated axis for the stance period is $\hat{\mu}^{(st)}_0 = (0.11, 0.994, 0.006)^T$, about 6 degrees away from $\mu^*_0$.

While the MS2 distribution was useful in making inference on the bending axis of partial knee motions, future work for this type of data lies in the development of a two-sample axis difference test.

ACKNOWLEDGEMENTS
Stephan Huckemann gratefully acknowledges funding by the Niedersachsen Vorab of the Volkswagen Foundation and DFG HU 1575/4.

ORCID
Sungkyu Jung https://orcid.org/0000-0002-6023-8956
REFERENCES

Banerjee, A., Dhilon, I. S., Ghosh, J., & Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6, 1345–1382.

Bingham, C., & Mardia, K. V. (1978). A small circle distribution on the sphere. *Biometrika*, 65, 379–389.

Browne, M. W. (1967). On oblique procrustes rotation. *Psychometrika*, 32, 125–132.

Cootes, T. F., Taylor, C., Cooper, D., & Graham, J. (1992). Training models of shape from sets of examples. In D. Hogg & R. Boyle (Eds.), *Proceedings of British Machine Vision Conference 1992* (pp. 9–18). Berlin, Germany: Springer-Verlag.

Dryden, I., & Mardia, K. V. (1998). *Statistical shape analysis*. Chichester, England: Wiley.

Eltzner, B., Huckemann, S., & Mardia, K. V. (2018). Torus principal component analysis with an application to RNA structures. *The Annals of Applied Statistics*, 12, 1332–1359.

Gray, N. H., Geiser, P. A., & Geiser, J. R. (1980). On the least-squares fit of small and great circles to spherically projected orientation data. *Journal of the International Association for Mathematical Geology*, 12, 173–184.

Hoff, P. D. (2009). Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18, 438–456.

Joshi, S., Pizer, S., Fletcher, P. T., Yushkevich, P., Thall, A., & Marron, J. S. (2002). Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Transactions on Medical Imaging*, 21, 538–550.

Jung, S., Dryden, I. L., & Marron, J. S. (2012). Analysis of principal nested spheres. *Biometrika*, 99, 551–568.

Jung, S., Foskey, M., & Marron, J. S. (2011). Principal arc analysis on direct product manifolds. *The Annals of Applied Statistics*, 5, 578–603.

Jupp, P. (2005). Sobolev tests of goodness of fit of distributions on compact Riemannian manifolds. *The Annals of Statistics*, 33, 2957–2966.

Kent, J. T. (1982). The Fisher–Bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B*, 71–80.

Kume, A., & Sei, T. (2018). On the exact maximum likelihood inference of Fisher–Bingham distributions using an adjusted holonomic gradient method. *Statistics and Computing*, 28, 835–847.

Kume, A., & Wood, A. T. A. (2005). Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. *Biometrika*, 92, 465–476.

Kurtek, S., Ding, Z., Klassen, E., & Srivastava, A. (2011). Parameterization-invariant shape statistics and probabilistic classification of anatomical surfaces. In G. Székely & H. K. Hahn (Eds.), *Information Processing in Medical Imaging* (pp. 147–158). Berlin, Germany: Springer.

Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society: Series B*, 37, 349–393.

Mardia, K. V., & Gadsden, R. J. (1977). A small circle of best fit for spherical data and areas of vulcanism. *Journal of the Royal Statistical Society: Series C*, 26, 238–245.

Mardia, K. V., Hughes, G., Taylor, C. C., & Singh, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *The Canadian Journal of Statistics*, 36, 99–109.

Mardia, K. V., & Jupp, P. E. (2000). *Directional Statistics*. Chichester, England: Wiley.

Pierrynowski, M., Costigan, P., Maly, M., & Kim, P. (2010). Patienrs with osteoarthritic knees have shorter orientation and tangent indicatrices during gait. *Clinical biomechanics*, 25, 237–241.

Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., ... Marron, J. S. (2013). Nested sphere statistics of skeletal models. In M. Breus, A. Bruckstein & P. Maragos (Eds.), *Innovations for shape analysis: Models and algorithms* (pp. 93–115). New York, NY: Springer.

Rivest, L.-P. (1999). Some linear model techniques for analyzing small-circle spherical data. *The Canadian Journal of Statistics*, 27, 623–638.

Schulz, J., Jung, S., Huckemann, S., Pierrynowski, M., Marron, J. S., & Pizer, S. M. (2015). Analysis of rotational deformations from directional data. *Journal of Computational and Graphical Statistics*, 24, 539–560.

Siddiqi, K., & Pizer, S. (2008). *Medial representations: Mathematics, algorithms and applications*. Berlin, Germany: Springer.
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kim B, Huckemann S, Schulz J, Jung S. Small-sphere distributions for directional data with application to medical imaging. Scand J Statist. 2019;46:1047–1071. https://doi.org/10.1111/sjos.12381

APPENDIX

We provide a technical lemma, referenced in Section 3.1, and proofs of Propositions 1 and 2.

Lemma 1. If $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$, then $E(X)$ is a linear combination of $\mu_0$ and $\mu_1$.

Proof of Lemma 1. Suppose that for some $a, b, c \in \mathbb{R}$, $v \in S^{p-1}$, which does not lie in the span of $\mu_0$ and $\mu_1$, $E(X) = a\mu_0 + b\mu_1 + cv$. Then, choose a $B \in O(p)$ such that $B\mu_0 = \mu_0$, $B\mu_1 = \mu_1$ but $Bv \neq v$. By Proposition 1(i), $B \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$. Thus, $E(X) = E(BX)$, which, in turn, leads to $a\mu_0 + b\mu_1 + cv = a\mu_0 + b\mu_1 + cBv$ and which is true only if $c = 0$. This gives the result.

Proof of Proposition 1. Note that if $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$, then $BX \sim S1(B\mu_0, B\mu_1, \kappa_0, \kappa_1)$. The argument is true when $S1$ is replaced by $S2$. Claim (ii) is verified by comparing the respective density functions. In (i), verifying the sufficient condition is trivial and, thus, is omitted.

Now, suppose that $X$ and $BX$ both have the S1($\mu_0, \mu_1, \kappa_0, \kappa_1$) distribution. Using the parameterization of (10), in particular, $\gamma = 2\kappa_0v\mu_0 + \kappa_1\mu_1$, we have, for any $x \in S^{p-1}$, $(B\gamma)^{T}x = \gamma^{T}x$ and $((B\mu_0)^{T}x)^2 = (\mu_0^{T}x)^2$, which, in turn, leads to $B\gamma = \gamma$ and $B\mu_0 = \pm \mu_0$. Plugging in $B\mu_0 = -\mu_0$ into the equation $B\gamma = \gamma$, we get

$$||B\mu_1||^2 = \left\|4\frac{\kappa_0}{\kappa_1} \mu_0 + \mu_1 \right\|^2 = 1 + 8\kappa_0/\kappa_1 (1 + 2\kappa_0/\kappa_1) > 1,$$

which contradicts the assumption that $B$ is orthogonal. Thus, $B\mu_0 = \mu_0$, in which case $B\mu_1 = \mu_1$ as well.

Next, suppose that $Y$ and $BY$ both have the S2($\mu_0, \mu_1, \kappa_0, \kappa_1$) distribution. Without loss of generality, suppose $\kappa_0 = \kappa_1 = 1$. From (6), we have, for any $x \in S^{p-1}$,

$$-(\mu_0^{T}x - \nu)^2 + \frac{\mu_1^{T}x - \nu \mu_0^{T}x}{\sqrt{1 - (\mu_0^{T}x)^2}} = -(\mu_0^{T}Bx - \nu)^2 + \frac{\mu_1^{T}Bx - \nu \mu_0^{T}Bx}{\sqrt{1 - (\mu_0^{T}Bx)^2}}. \quad (A1)$$
Plugging in $x = \pm \mu_0$ into (A1) yields
\[
- (1 - \nu)^2 = -(a - \nu)^2 + (b - \nu a)\sqrt{(1 - \nu^2)(1 - a^2)},
\]
\[
- (1 + \nu)^2 = -(a + \nu)^2 - (b - \nu a)\sqrt{(1 - \nu^2)(1 - a^2)},
\]
where $a = \mu_0^T B \mu_0$, $b = \mu_1^T B \mu_0$. Solving the above system of equations, we get $a = \mu_0^T B \mu_0 = \pm 1$; thus, $B \mu_0 = \pm \mu_0$. Suppose $B \mu_0 = -\mu_0$, so that (A1) becomes
\[
((B \mu_1)^T - \mu_1^T) x = c \mu_0^T x, \quad c = \nu^{-1} \left( \sqrt{(1 - \nu^2)} \left( 1 - (\mu_0^T x)^2 \right) - 1 \right),
\]
which implies that $B \mu_1 - \mu_1$ is parallel to $\mu_0$. However, coefficient $c$ is not a constant function of $x$, and there exists an $x$ such that $||\mu_0|| = c^{-1}||B \mu_1 - \mu_1|| \neq 1$, which contradicts the fact that $\mu_0 \in \mathbb{S}^{p-1}$. Thus, $B \mu_0$ must be $\mu_0$, in which case we have $B \mu_1 = \mu_1$ as well. \hfill \Box

**Proof of Proposition 2.** For a given $h > 0$, let $\gamma = 2\kappa_0 \nu \mu_0 + \kappa_1 \mu_1$ and $A_h = \kappa_0 \mu_0 \mu_0^T + h I_p$. Then, the S1 density (5) can be expressed as the Fisher–Bingham form (10), that is,
\[
f_{S1}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \alpha(\gamma, A_h) \exp \left\{ \gamma^T x - x^T A_h x \right\},
\]
where $\alpha(\gamma, A_h)$ satisfies
\[
\alpha(\kappa_0, \kappa_1, \nu) = \alpha(\gamma, A_h) \exp\{ -\kappa_0 \nu^2 + h \}. \tag{A2}
\]
For the purpose of evaluating the value of $\alpha(\kappa_0, \kappa_1, \nu)$ or, equivalently, $\alpha(\gamma, A_h)$ for the given value of $h$, one can assume, without losing generality, that $\mu_0 = (1, 0, \ldots, 0)^T$ and $\mu_1 = (\nu, \sqrt{1 - \nu^2}, 0, \ldots, 0)$, so that $\gamma = (\nu(2\kappa_0 + \kappa_1), \kappa_1 \sqrt{1 - \nu^2}, 0, \ldots, 0)^T$, and the vector of diagonal values of $A_h$ is $\lambda := (2(\kappa_0 + h), h, \ldots, h)$. The $j$th element of $\xi$, in the statement of the proposition, is then given by $\xi_j := (\gamma_j^2) / 2 \lambda_j$. With these notations, proposition 1 of Kume and Wood (2005) gives
\[
\alpha(\gamma, A_h) = 2\pi^{p/2}|A_h|^{-1/2} g(1) \exp\{ \xi^T A_h \xi^\gamma \}.
\]
Hence, by (A2), we have (17). \hfill \Box