Space sciences from a directional statistical point of view

Guendalina PALMIROTTA*[1] and Christophe Ley[1]

[1]University of Luxembourg, Faculty of Science, Technology and Medicine, Department of Mathematics, 6, Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg,
e-mails: guendalina.palmirotta@uni.lu and christophe.ley@uni.lu

Abstract

Already back in the 17-18th centuries, important foundations of modern statistical theory were formulated with the goal to address astronomical problems. This successful interdisciplinary collaboration has been revived since the 1990s, giving rise to the research flow called astrostatistics, which has been particularly active over the past decades. The increasing amount of astronomical data nowadays has posed new challenges and created the need for more innovative modern statistical theories and models. Directional statistics, a branch of statistics involving observations such as directions, axes, rotations, with values on (compact) Riemannian manifolds like our celestial sphere, has proved to be a promising domain to address important space sciences issues such as space weather, cosmology, or even space surveillance.

In this paper, we will instigate directional statistics by providing a review of their old and recent developments stimulated by interesting applications in space sciences.

Keywords: Astronomy, Astrophysics, Cosmology, Directional Statistics, Multivariate analysis, Nonparametric statistics, Space sciences, Space surveillance.

Contents

1 The Big Bang of (directional) astrostatistics **2**

2 Preliminaries on directional statistics **3**
2.1 Distributional models . . . 3
2.1.1 Circular models . . . 3
2.1.2 Models for spherical data . . . 4
2.2 Hypothesis testing (for uniformity) on hyperspheres . . . 5

3 Recent advances in space sciences problems **7**
3.1 Orbits of planets and long-period comets . . . 7
3.1.1 Pluto's origin and small solar system bodies . . . 10
3.2 Craters on planets and moons . . . 11
3.3 Space weather and sunspots . . . 13
3.4 Electromagnetic radiation: the cosmic rays . . . 15
3.5 Galactic environment and stars . . . 20
3.5.1 Stars positions . . . 20
3.5.2 Pulsars . . . 22
3.5.3 Binary stars . . . 24
3.5.4 Supernovae . . . 25
3.5.5 Interstellar medium . . . 26
3.5.6 Star Sgr A* . . . 27
3.5.7 Dark matter: WIMP direct detection . . . 28
3.5.8 Galaxies dynamics and clustering . . . 29
3.5.9 Exoplanets . . . 33

*corresponding author

3.5.10 Hot stars . . . 34
3.6 Tracking satellite orbit phases and space objects . . . 34
3.6.1 Asteroids . . . 35
3.6.2 Space debris and its orbital uncertainty propagation . . . 35
3.6.3 GNSS phase observations . . . 37

4 Conclusion **38**

1 The Big Bang of (directional) astrostatistics

In this golden age of big data, space sciences seems to have found an unlikely hero – statistics or, using the more modern and broader term (including machine learning and computer science), data science. However, this interdisciplinary combination is not utterly new, as a look at the history of astronomy reveals. Indeed, tracing back to ancient Greek, Arabic, and Renaissance astronomers (such as Hipparchus, al-Biruni, Brahe, Galilei, Hubble and Newton, to name but a few), we come across the fact that they were already making extensive use of basic statistical concepts such as the mean and the least squares method for their celestial calculations. As particularly striking example, in 1809 Gauss [Gau09] deduced the law of errors, nowadays known as the Gaussian/normal distribution or "bell curve", as a natural and practical way to model the inevitable errors in astronomical observations[1]. Gauss is a shining example of how astronomical questions have led to major advances in probability and statistics.

It was at the beginning of the 1990s that the connections between the two fields began to sparkle again and they have since then rapidly grown. This was due to the revolution of modern astronomy, characterized mainly by the advent of large astronomical surveys, the launch of space-based observatories like National Aeronautics and Space Administration's (NASA) Hubble or the European Space Agency's (ESA) Giotto, and the dawning of multi-messenger and gravitational-wave astronomy. As a result, the astronomical data became too large to handle by classical tools and simple statistical concepts were no longer sufficient. It was time to adapt, improve, and come up with new statistical methods to extract scientific knowledge from this data and confirm or even review astrophysical theories. In addition, with the emergence of cross-disciplinary interactions between astronomers and statisticians, a small but growing research field known as *astrostatistics* has seen the light under the impetus of the likes of Babu and Feigelson (see e.g. [FB09, Fei09, FDSIB21, Fei16] and references therein for a more detailed history and role of statistics in astronomy). Hence a wide range of statistical methods such as model fitting, hypothesis testing including goodness-of-fit tests of a probability distribution to a dataset, or modern parameter estimation methods have been proposed to assist in the analysis of quantitative astronomical data. Compared to the very early combination of the two fields, characterized mostly by descriptive statistics put to use in astronomy, modern statistics provides astronomers and space scientists a pathway for transitioning from qualitative examination to quantification, in particular quantification of the unavoidable uncertainties associated to any data-driven analysis. For instance, it enables them to assess the quality of predictions, determine their acceptability, and ascertain whether a chosen model possesses the required accuracy and precision to fulfil the scientific objectives.

A pre-requisite for determining the most appropriate modern statistical methods for space sciences and astronomy consists in recognizing that most natural supports for the observed astronomical or space sciences data actually *are not* linear such as the real line or $\mathbb{R}^2$, but are rather of a different topological type such as the celestial sphere or other compact Riemannian manifolds like the unit circle, the torus, and their extensions. For example, a direction observed in the plane $\mathbb{R}^2$, like the direction of a comet, is most naturally represented by an angle θ, typically in $[0, 2\pi)$ or $[-\pi, \pi)$, measured in a specified direction from a specified origin, or by the unit vector $\mathbf{x} = (\cos(\theta), \sin(\theta))'$ for which it holds that $\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = 1$. Such data are referred to as *directional data* in the literature and they typically consist of (unit) vectors with common origin and whose length is of no interest. Consequently, their natural supports are the (circumference of the) unit circle $\mathbb{S}^1$ (circular data), unit (hyper-)sphere $\mathbb{S}^2$ (spherical data), torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ (toroidal data), and cylinder $\mathbb{S}^1 \times \mathbb{R}$ (cylindrical data).

The fact that most astronomical or space sciences data lie on such supports implies that classical statistical techniques and tools cannot be used to analyze directional data. As a simple example, consider two points lying on the surface of a sphere. Their classical average will be a point lying inside the sphere, which is obviously no more a point on the sphere. A concept as simple as the

[1] For the historically inclined reader, we remark that the Gaussian distribution is an illustration of Stigler's law of eponymy [Sti80] as its first introduction dates back to de Moivre in 1738.

average thus needs to be properly defined for such data, and one can imagine that this difficulty goes crescendo for more and more complex statistical quantities and methods. The need to properly deal with directional data has given rise since the 1950s to the research flow called *directional statistics* which has been particularly active over the past two decades. This popularity is due to demands from various fields. Besides astronomy and space sciences, such data occur in fields like meteorology (wind directions), earth and environmental sciences (location of earthquakes on the Earth's surface, modeling of wildfires, sea conditions, paleomagnetism), bioinformatics (angle pairs describing the protein conformation), and social sciences (events occurring on the 24-hour clock or in calendar measurements). For more ample information, we refer the reader to some of the various books written on the topic in recent years [MJ00, PNR13, LV17, LV18].

Paper contribution and organization. The purpose of this review is to bring to the space sciences and in particular astronomy communities some of the concepts and results of the field of modern directional statistics that have proven useful in the analysis of astronomical data. In some sense, our work is an updated and more detailed version of the paper [Jup90] by Jupp from 1990. In our search, we found more than 90 papers that proposed (new) directional statistics methodology or employed existing directional statistics tools for astronomical applications. As we shall see, many of the relevant papers have appeared in the statistics literature, implying that several members of the aforementioned communities likely are not aware of them. We wish to alleviate this situation by a tailor-made review.

To this end, we begin in Section 2 by recalling the main directional statistics tools used to deal with astronomy and space sciences data. Section 3 constitutes the heart of this paper, as it yields, to the best of our knowledge, an as complete as possible survey of papers who analyze astronomy and space sciences (simulated/real) data via directional statistics. It covers a broad range of topics including the orbits of planets and comets, craters on celestial objects, space weather, cosmic rays, our galactic environment, and space safety. Section 4 concludes the paper with final comments.

2 Preliminaries on directional statistics

In order to provide the reader with the necessary technical background to follow the developments in Section 3, we recall the main directional statistics concepts and tools that have been used to model astronomy-driven data. We refer to the books [FLE87, MJ00, LV17, LV18] and references therein, as well as the survey papers [ML18, PGP21] for a primer on directional statistics and their advancement.

2.1 Distributional models

We start by introducing the most common probability distributions for circular (Section 2.1.1) and spherical (Section 2.1.2) data.

2.1.1 Circular models

Consider f_Θ as the probability density function (pdf), also referred to as density, of a continuous circular random variable Θ taking values in $\mathbb{R}$ such that for almost all $\theta \in \mathbb{R}$

- f_Θ is non-negative, i.e. $f_\Theta(\theta) \geq 0$;
- f_Θ is 2π-periodic, i.e. $f_\Theta(\theta + 2\pi) = f_\Theta(\theta)$;
- f_Θ integrates to 1 over any interval of length 2π, i.e. $\int_\theta^{\theta+2\pi} f_\Theta(\omega)\, \mathrm{d}\omega = 1$.

This latter property leads us to define a circular pdf through its values on $[0, 2\pi)$ or $[-\pi, \pi)$. The circular cumulative distribution function (cdf) is defined as the non-periodic function $F_\Theta(\theta) = \int_{\theta_0}^{\theta} f_\Theta(\omega)\, \mathrm{d}\omega$, with θ_0 typically being 0 or $-\pi$. There are usually six general ways used to construct circular pdfs: wrapping, projection, perturbation, conditioning, diffusion, and characterizations such as via maximum likelihood or maximum entropy.

The simplest example of a distribution on the circle is the *circular uniform distribution* with density

$$f_\Theta(\theta) = \frac{1}{2\pi} \tag{2.1}$$

and cdf $F_\Theta(\theta) = \frac{\theta}{2\pi}, \theta \in [0, 2\pi)$. Other classical circular distributions are the cardioid, Wrapped Cauchy and von Mises, and we will expand on the latter.

Von Mises distribution. The von Mises (vM) distribution introduced by von Mises (1918) [vM18] plays a core role among circular distributions and is sometimes referred to as the circular equivalent of the normal distribution. Its pdf is

$$f_\Theta(\theta;\mu,\kappa) = \frac{1}{2\pi\mathcal{I}_0(\kappa)}\exp\{\kappa\cos(\theta-\mu)\}, \tag{2.2}$$

where $\mu \in [0, 2\pi)$ is the mean direction, $\kappa \geq 0$ its concentration parameter and $\mathcal{I}_0$ is the modified Bessel function of the first kind and order zero

$$\mathcal{I}_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi}\exp\{\kappa\cos(\phi-\mu)\}\,\mathrm{d}\phi. \tag{2.3}$$

2.1.2 Models for spherical data

Most of the astronomical phenomena like orbits of comets, craters on planets and moons, position and distribution of stars, cosmic rays, and dark matter are supported by the (hyper-)sphere $\mathbb{S}^{d-1}$, $d > 2$. In the following, we will encounter classical spherical probability models including Fisher-von-Mises-Langevin, Fisher-Bingham, Kent, and Watson models and their variants.
Suppose that $\mathbf{X}$ is a unit random vector on $\mathbb{S}^{d-1}$. Like for the circular model, the best-known and simplest model for spherical data is the *uniform spherical distribution* with pdf $\frac{1}{\omega_d}$ over $\mathbb{S}^{d-1}$, where $\omega_d = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}$ is the surface area of $\mathbb{S}^{d-1}$ and $\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)\,\mathrm{d}t$ is the Gamma function. No other distribution is invariant under both reflection and rotation.

Fisher-von Mises-Langevin distribution. Arguably the most popular non-uniform directional pdf is the Fisher-von Mises-Langevin (FvML or sometimes FvM) pdf

$$f_\mathbf{X}(\mathbf{x};\boldsymbol{\mu},\kappa) = \left(\frac{\kappa}{2}\right)^{d/2-1}\frac{1}{2\pi^{d/2}\mathcal{I}_{d/2-1}(\kappa)}\exp(\kappa\mathbf{x}^T\boldsymbol{\mu}), \tag{2.4}$$

where $\mathbf{x}$ is a point on the unit hyper-sphere $\mathbb{S}^{d-1}$, $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ the mean direction vector, $\kappa \geq 0$ a concentration parameter and $\mathcal{I}_\nu$ the modified Bessel function of the first kind and of order ν. The pdf (2.4) is unimodal about μ and its concentration around the mode is regulated by κ, the limit $\kappa = 0$ yielding the uniform distribution. Observe that we retrieve for $d = 2$ the vM pdf (2.2), while $d = 3$ corresponds to the Fisher (1953) [Fis53] pdf and general d to the Langevin (1905) [Lan05] pdf, hence the terminology.

Fisher-Bingham distribution and its submodels. Another classical spherical distribution is the Fisher-Bingham (FB) distribution with pdf in vector form

$$f_\mathbf{X}(\mathbf{x};\boldsymbol{\mu},\kappa,\mathbf{A}) = \frac{1}{c(\kappa,\mathbf{A})}\exp(\kappa\mathbf{x}^T\boldsymbol{\mu} + \mathbf{x}^T\boldsymbol{A}\mathbf{x}), \tag{2.5}$$

where $\boldsymbol{A}$ is a symmetric $d \times d$ matrix and $\boldsymbol{\mu}, \kappa$ play the same role as in (2.4). The normalizing constant $c(\kappa, \boldsymbol{A})$ is non-trivial to deal with, and we recommend the reader to check [KS18] for an approximation. An alternative formulation for this density is

$$f_\mathbf{X}(\mathbf{x};\boldsymbol{\nu},\mathbf{A}) = \frac{1}{c(\kappa,\mathbf{A})}\exp(\boldsymbol{\nu}^T\mathbf{x} + \mathbf{x}^T\mathbf{A}\mathbf{x}),$$

where $\boldsymbol{\nu} = \kappa\boldsymbol{\nu}_0$ is a vector with $\boldsymbol{\nu}_0$ a unit d-vector. Without loss of generality, one can assume that $\mathrm{tr}(\mathbf{A}) = 0$ since $\mathbf{x}^T\mathbf{x} = 1$. The full FB distribution has eight parameters, thus preferably denoted as FB8, and contains some special cases:

- *Fisher distribution.* When $\mathbf{A} = 0$ is the zero matrix, then (2.5) reduces to the FvML pdf (2.4), also known as the Fisher distribution for $d = 3$. The Fisher pdf is invariant under rotations of $\mathbb{S}^2$ about the $\boldsymbol{\nu}_0$ axis and thus it can be described as "isotropic" bivariate normal distribution in the tangent plane to $\mathbb{S}^2$ at the modal direction.
- *Bingham distribution.* If $\kappa = 0$ (or $\boldsymbol{\nu} = 0$), we obtain the Bingham pdf, which was introduced by Bingham (1964) [Bin64]. It is a suitable model for axial data, meaning that data are axes and each direction is considered equivalent to the opposite direction. For $d = 3$, we consider $\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$ as the spectral decomposition of $\mathbf{A}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_{\max}, \lambda_{\mathrm{mid}}, \lambda_{\min})$ contains the eigenvalues in decreasing order, and the columns of the 3×3 orthogonal matrix $\boldsymbol{\Gamma} =$

$[\boldsymbol{\gamma}_{(\max)}, \boldsymbol{\gamma}_{(\mathrm{mid})}, \boldsymbol{\gamma}_{(\min)}]$ contain the corresponding eigenvectors. There are two common special cases, both characterized by a concentration parameter $\beta > 0$, the *bimodal case* ($\lambda_{\max} = \beta > \lambda_{\mathrm{mid}} = \lambda_{\min} = 0$) with the mode along the $\pm\boldsymbol{\gamma}_{(\max)}$ axis, and the *girdle case* ($\lambda_{\max} = \lambda_{\mathrm{mid}} > \lambda_{\min} = -\beta$) with the mode at all points on the great circle perpendicular to the $\pm\boldsymbol{\gamma}_{(\min)}$ axis.

- *Kent or Fisher-Bingham-Kent distribution and its special cases.* A sub-family of the FB8 is the "aligned" Fisher-Bingham distribution (FB6) with six parameters. It is obtained when $\boldsymbol{\nu}_0$ equals one of the eigenvectors of $\mathbf{A}$. A special case of FB6 and thus of FB8 is the Kent or Fisher-Bingham-Kent distribution with 5 parameters, denoted as FB5, proposed by Kent (1982) [Ken82] under the additional condition that $\mathbf{A}^T\boldsymbol{\mu} = 0$. It is an oval density with contours around $\boldsymbol{\mu}$ and the matrix $\mathbf{A}$ is a shape parameter. It is ideal to simulate and track the space debris floating uncontrollably around our Earth. According to [KHJ16], there are two variants, the "balanced" (FB5b) ($\boldsymbol{\nu} = \boldsymbol{\gamma}_{(\mathrm{mid})}$, $\lambda_{\max} = \beta, \lambda_{\mathrm{mid}} = 0$, $\lambda_{\min} = -\beta$, for some $\beta \geq 0$) and the "extreme" (FB5e) ($\boldsymbol{\nu} = \boldsymbol{\gamma}_{(\max)}$, $\lambda_{\max} = \lambda_{\mathrm{mid}} = 0, \lambda_{\min} = -\delta$, for some $\delta \geq 0$). The FB5b density has a mode in the $\boldsymbol{\nu}$ direction provided that $\kappa \geq 2\beta$, and the contours of constant probability are oval-shaped with the major axis pointing along the $\boldsymbol{\gamma}_{(\mathrm{mid})}$ axis and the minor axis pointing along the $\boldsymbol{\gamma}_{(\min)}$ axis. In addition, the FB5b distribution is approximately bivariate normal and concentrated near the modal direction $\boldsymbol{\nu}$ under higher concentration. The FB5e distribution can be viewed as a combination of a Fisher and a girdle Bingham, where the mode of the Fisher lies on the great circle mode of the Bingham. To describe unimodal data concentrated near a great circle such as space objects orbiting the Earth, the FB5e is better suited than the FB5b (see Sect. 3.6.2).

(Dimroth-)Watson distribution. Dimroth (1962) [Dim62] and Watson (1965) [Wat65] introduced independently a density that is invariant under reflections and hence suited for modeling axial data. Their density has the form

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2} M(\frac{1}{2}, \frac{d}{2}, \kappa)} \exp\{\kappa(\mathbf{x}^T\boldsymbol{\mu})^2\}, \tag{2.6}$$

where $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ is the location parameter, $\kappa \in \mathbb{R}$ the concentration, and

$$M(a, b, z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_{-1}^{1} e^{t^2 z} t^{2a-1}(1-t^2)^{b-a-1}\, \mathrm{d}t$$

denotes the Kummer function with $a, b, z \in \mathbb{R}$. For $\kappa < 0$ the Watson density is concentrated around the great circle like girdle densities, while for $\kappa > 0$ it is bipolar with maxima at $\pm\boldsymbol{\mu}$. Note that this distribution is a submodel of the FB exponential family of distributions (2.5).

Rotationally symmetric distributions. The FvML (2.4) and Watson (2.6) distributions share the common feature that they are rotationally symmetric about their location $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$. This means that their pdfs depend on $\mathbf{x}$ only through $\mathbf{x}^T\boldsymbol{\mu}$ and, as a consequence, have contours that are circular when $\mathbf{x} \in \mathbb{S}^2$.

Definition 1 (Rotational symmetry)**.** *The distribution of a random $\mathbf{X} \in \mathbb{S}^{d-1}$ is said to be rotationally symmetric about $\boldsymbol{\mu}$ if $\mathbf{OX}$ is equal in distribution to $\mathbf{X}$ for any orthogonal $d \times d$ matrix $\mathbf{O}$ satisfying $\mathbf{O}\boldsymbol{\mu} = \boldsymbol{\mu}$.*

Rotational symmetry leads to densities of the form $\mathbf{x} \mapsto c_{f_a,d} f_a(\mathbf{x}^T\boldsymbol{\mu})$, where $c_{f_a,d}$ is a normalizing constant and $f_a : [-1, 1] \longrightarrow \mathbb{R}^+$ an absolutely continuous function. Since the distribution of $\mathbf{X}$ only depends on the angle (colatitude angle when $d = 3$) between $\mathbf{X}$ and $\boldsymbol{\mu}$, the function f_a is thus referred to as angular function. The respective angular functions of the FvML and Watson distribution are $t \mapsto \exp(\kappa t)$ and $t \mapsto \exp(\kappa t^2), t \in [-1, 1]$.

2.2 Hypothesis testing (for uniformity) on hyperspheres

Hypothesis tests for uniformity, symmetry, location, concentration, goodness-of-fit, and other testing scenarios are indispensable in astronomy problems. Such tests provide crucial information about the direction and distribution of spatial and astronomical observables, e.g. the direction of cosmic rays, cometary orbits, and the position of stellar objects, craters, sunspots, galaxies, and satellites. Source detection and sample comparison problems can be formulated as "yes or no" questions to be addressed with statistical hypotheses testing. In the subsequent sections, we will mention several times these tests. For general information, we refer the reader to [LV17, GPPV20]. One type of tests however plays a particular role. Assessing the presence of uniformity (or isotropy) is one of the first and most natural modeling questions that practitioners like astronomers focus on when dealing with directional data, hence its importance. It is also one of the oldest problems in directional statistics. We will therefore provide some background in the rest of this section.

Given an independent and identically distributed (i.i.d.) random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from the unit random vector $\mathbf{X}$ on $\mathbb{S}^{d-1}$, the problem of testing the presence of uniformity is formalized as the testing of the null hypothesis

$$\mathcal{H}_0^{\text{unif}} : P = \text{Unif}(\mathbb{S}^{d-1}) \text{ against } \mathcal{H}_1 : P \neq \text{Unif}(\mathbb{S}^{d-1}),$$

where P stands for the probability distribution of $\mathbf{X}$ (short $\mathbf{X} \sim P$) and $\text{Unif}(\mathbb{S}^{d-1})$ denotes the uniform distribution on $\mathbb{S}^{d-1}$. For $d = 2$, the polar coordinates of $\mathbb{S}^1$ yield $\mathbf{X}_i = (\cos(\Theta_i), \sin(\Theta_i))^T$ with random angles $\Theta_i \in [0, 2\pi), i = 1, \ldots, n$.

We distinguish between *parametric* and *nonparametric* tests. Parametric tests nest $\mathcal{H}_0^{\text{unif}}$ within a parametric distribution P_θ (with $\theta \in \mathbb{R}^q$ $(q \geq 1)$ a vector of parameters) and aim to perform well against alternatives within P_θ, but may fail against alternatives outside P_θ. Nonparametric tests aim for uniform consistency against a wide range of non-uniform alternatives, though they may not be powerful against specific deviations. García-Portugués and Verdebout [GPV18] provided an extensive review of uniformity tests.

Historically, Rayleigh (1880) [Ray80] was the first to study the resultant length of bivariate uniform unit vectors, proposing the initial uniformity test, which drew significant attention. Classical tests like Rayleigh, Ajne, Kuiper, and Watson are typical for circular data, with Rayleigh and Ajne tests generalizable to $\mathbb{S}^{d-1}$ through the Sobolev class of tests. In the following, in historical order, we will inspect uniformity tests that were applied for astronomy-driven data:

- The **Rayleigh test**, proposed in 1919 by Rayleigh [Ray19], was the first test of uniformity for circular data and Jupp (2001) [Jup01] extended it later to the hypersphere. It is based on the simple fact that when $\mathbf{X} \sim \text{Unif}(\mathbb{S}^{d-1})$, then $\mathbb{E}[\mathbf{X}] = \mathbf{0}$ or, equivalently, $\|\mathbb{E}[\mathbf{X}]\|^2 = 0$ for the Euclidean norm $\|\cdot\|$. This provides a simple way of testing $\mathcal{H}_0^{\text{unif}}$ by using Rayleigh's statistic
$$R_{n,d} = \frac{d}{n} \sum_{i,j=1}^{n} \mathbf{X}_i^T \mathbf{X}_j = dn\|\bar{\mathbf{X}}\|^2, \tag{2.7}$$
with $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i$. The Rayleigh test rejects $\mathcal{H}_0^{\text{unif}}$ for large values of (2.7). Since the $\mathbf{X}_i$'s have mean zero and covariance $d^{-1} I_d$ with I_d the diagonal matrix with only 1s, under $\mathcal{H}_0^{\text{unif}}$ the multivariate Central Limit Theorem implies that the asymptotic distribution of $R_{n,d}$ is χ^2_d, a chi-squared distribution with d degrees of freedom. Therefore Rayleigh's test rejects $\mathcal{H}_0^{\text{unif}}$ at asymptotic level α whenever $R_{n,d} > \chi^2_{d;1-\alpha}$. In particular, the Rayleigh test is the score test or the likelihood ratio test for testing uniformity against FvML distributions (2.4).
- The **Kuiper test** for circular data is a rotation-invariant version of the classical (one-sample) *Kolmogorov-Smirnov test* for uniformity studied by Kuiper (1960) [Kui60] and based on
$$V_n = D_n^+ + D_n^-, \tag{2.8}$$
with statistics
$$D_n^+ = \sqrt{n} \sup_{\theta \in [0,2\pi)} \{F_n(\theta) - F(\theta)\} = \sqrt{n} \max_{1 \leq i \leq n} \left\{\frac{i}{n} - \mathbf{X}_i\right\}$$
$$D_n^- = \sqrt{n} \sup_{\theta \in [0,2\pi)} \{F(\theta) - F_n(\theta)\} = \sqrt{n} \max_{1 \leq i \leq n} \left\{\mathbf{X}_i - \frac{i-1}{n}\right\},$$
where $F(\theta)$ is the cdf of the uniform distribution (see Sect. 2.1.1), $F_n(\theta) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{\Theta_i \leq \theta\}}$ the empirical cdf (ecdf) by assuming that the origin is set as the angle 0, and $\mathbf{X}_i = (2\pi)^{-1}\Theta_i, i = 1, \ldots, n$. Hence testing uniformity is achieved through the examination of the discrepancy between the postulated F and the data-driven F_n. The tail probability of V_n, under $\mathcal{H}_0^{\text{unif}}$, is given by series expansion
$$V_n^* = V_n\Big(n^{1/2} + 0.155 + \frac{0.24}{n^{1/2}}\Big).$$
- The **Watson test**, introduced by Watson (1961) [Wat61], also evaluates the deviation between F_n and F but here through the quadratic norm associated with *Cramér-von Mises type tests*. Similar to the Kuiper test, it is a rotation-invariant test. In fact, in contrast to the classical Cramér-von Mises (CvM) test, Watson considered the variance of $F_n(\theta) - F(\theta)$ instead of the

second moment. The Watson statistics on $\mathbb{S}^1$, sometimes referred as the U^2 *statistics*, is given by

$$\begin{aligned} U_n^2 &= n \int_0^{2\pi} \left[F_n(\theta) - F(\theta) - \int_0^{2\pi} (F_n(\theta) - F(\theta)) \, \mathrm{d}F(\theta) \right]^2 \mathrm{d}F(\theta) \\ &= \sum_{i=1}^{n} \left[\left(\mathbf{X}_{(i)} - \frac{i - \frac{1}{2}}{n} \right) - \left(\bar{\mathbf{X}} - \frac{1}{2} \right) \right]^2 + \frac{1}{12n}. \end{aligned} \tag{2.9}$$

It can be shown that the asymptotic distributions of U_n^2 and $(\frac{V_n}{\pi})^2$ are the same and that the asymptotic tail probability of U_n^2 is given by the Kolmogorov distribution function K, i.e. $\lim_{n\to\infty} \mathbb{P}[U_n^2 > u] = 1 - K(\sqrt{u}\pi)$.

- **Sobolev test** types of tests of uniformity on $\mathbb{S}^{d-1}$ have been proposed by Beran [Ber68, Ber69] and Giné [Gin75]. The motivation behind this lies in the Sobolev procedures to root on the eigenfunctions of the Laplace-Beltrami operator Δ acting on $\mathbb{S}^{d-1}$. For a given sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on $\mathbb{S}^{d-1}$, the Sobolev statistics, specified by a sequence of coefficients $\{v_k\}$ for $k \in \mathbb{N}_*$, take the form

$$S_n(\{v_k^2\}) = \frac{1}{n} \sum_{i,j=1}^{n} \sum_{k=1}^{\infty} v_k^2 h_k(\mathbf{X}_i, \mathbf{X}_j), \tag{2.10}$$

where

$$h_k(\mathbf{u}, \mathbf{v}) = \begin{cases} 2\cos(k \cos^{-1}(\mathbf{u}^\mathbf{T}\mathbf{v})), & d = 1 \\ (1 + \frac{2k}{d}) C_k^{(d-1)/2}(\mathbf{u}^\mathbf{T}\mathbf{v})), & d > 1, \end{cases}$$

and $C_k^{(d-1)/2}$ is the k-th Gegenbauer polynomial of index $(d-1)/2$. Note that particular choices for $\{v_k^2\}$ give different local optimality properties, consistencies, and power against specific kinds of alternatives. However v_k^2 should decay fast enough to ensure convergence in (2.10). For example, we obtain the test statistics of

  - Rayleigh with $v_k = \delta_{k1}$, where $\delta_{k\ell} = \begin{cases} 0, & k \neq \ell \\ 1, & k = \ell, \end{cases}$
  - Bingham with $v_k = \delta_{k2}$,
  - Watson with $v_k = (\pi k)^{-1}$ for $k \geq 1$,
  - Ajne with $v_k = \begin{cases} 0, & k \text{ even} \\ (\pi k)^{-1}, & k \text{ odd}, \end{cases}$
  - Rotham (Rt) with $v_k = \frac{\sin(k\pi t)}{2k\pi t}$ for $k \geq 1$.

To test uniformity on compact Riemannian manifolds and on $\mathbb{S}^1$, a "data-driven" Sobolev test is more adapted by using an information criterion to truncate the infinite series in (2.10). Moreover, the Sobolev tests can be extended to a class of tests based on the projected ecdf that yields extensions for data on $\mathbb{S}^{d-1}$ of the Watson tests for circular uniformity and a novel Anderson-Darling-like (AD) test for uniformity on $\mathbb{S}^{d-1}$ [GPNECA23].

3 Recent advances in space sciences problems

Equipped with the necessary statistical baggage, we can now delve into the review per se. This section is split into six subsections, starting from the orbits of planets and comets in Subsection 3.1, continuing with craters on celestial objects in Subsection 3.2, space weather in Subsection 3.3, cosmic rays in Subsection 3.4, our galactic environment in Subsection 3.5 going from the different stars (Subsections 3.5.1-3.5.6 and 3.5.10), dark matter (Subsection 3.5.7) to galaxies dynamics (Subsection 3.5.8) and exoplanets (Subsection 3.5.9), and ending with space safety in Subsection 3.6 containing the space debris (Subsection 3.6.2) dilemma and other space objects (Subsections 3.6.1-3.6.3). Each subsection follows the same scheme and is consequently divided into three parts: "Motivation", "Database", and "Statistical tools and data analysis".

3.1 Orbits of planets and long-period comets

Motivation. It is a well-known fact that the planes defined by the orbits of the planets in the Solar System are all nearly coincident with the ecliptic, the only partial exception being Pluto which is no longer considered a true planet according to the decision of the International Astronomical

Union (IUA). Such an "almost coincidence" even caught Bernoulli's (in the 1730s) attention, who wondered if this fact could happen "by chance". Nowadays the clustering of the planets' orbits about the ecliptic is explained by their origin in the protoplanetary disk. In addition, Mardia and Jupp (2000) [MJ00, p. 209] showed that there is strong evidence against the uniformity hypothesis of the planets' orbits.

The study of comet orbits has been more intricate. It is believed that most of the comets are produced in the Kuiper Belt or the Oort Cloud following an unpredictable mechanism. Hence, their orbits could be seen as multiple realizations of an experiment whose output is not deterministic. There are two main classes of comets:

- the short-period comets, also called ecliptic-clustered comets, emerging either in the flattened Kuiper belt or the scatter disc, with relatively small orbits following the ecliptic plane, and
- the long-period comets, also referred to as nearly isotropic comets, originating directly from the outer Oort cloud, with very large orbits (larger than 200 years) appearing from every direction in the sky. The Oort cloud is a widely accepted model posing the existence of a roughly spherical reservoir of icy planetesimals in the limits of our Solar system that were ejected from protoplanetary disks by giant planets. It is believed that these icy planetesimals became long-period comets when randomly captured in heliocentric orbits due to the effect of several gravitational forces as well as perturbations of passing stars and the galactic tide (see, e.g. Sects. 5 and 7.2. in [DBKR15] and references therein). This conjectured past of the Oort cloud explains the very characteristic nearly isotropic distributions of the long-period comets' orbits (evidenced, e.g. in [WT99]).

Since the comet orbits are always conical sections, they are either elliptical (with our Sun at one focus) leading to periodic comets or parabolic or hyperbolic, leading to non-periodic comets. Their orbit plane is commonly determined by two angles (i, Ω), where the angle $i \in [0, \pi]$ is the inclination between the normal vector of the orbit and the normal vector of the ecliptic, and $\Omega \in [0, 2\pi)$ is the longitude of the ascending node. The directed normal vector to the orbit plane is given by

$$v = (\sin(i)\sin(\Omega), -\sin(i)\cos(\Omega), \cos(i))^T. \tag{3.1}$$

The sign of the vector reflects if the orbit is prograde (with positive $\cos(i)$ in (3.1), i.e. revolves in the same direction as the Earth about the ecliptic polar axis, so that the directed normal points to the north of the ecliptic plane) or retrograde (with negative $\cos(i)$ in (3.1), i.e. revolves in the opposite direction, so that the directed normal points to the south of the ecliptic plane). In particular, the periodic cometary orbit is featured by the perihelion direction u (the direction from the Sun to the point of closest approach of the comet) and the directional normal v to the orbit (with direction specified by the sense of rotation). In addition, putting $w = u \wedge v$ yields a third unit vector such that the matrix $\mathbf{X} = (u, v, w)$ is a rotation matrix, i.e., $\mathbf{X}^T\mathbf{X} = I_3$ and $\det(\mathbf{X}) = 1$. Thus the directional aspects of the orbit of a periodic comet can be represented by a rotation matrix, according to [Jup95, Sect. 2.1].

There are astronomical reasons for believing that observations of comet orbits are subject to considerable selection effects. This can be translated into three natural questions of astronomical interest:

(i) Are the comet orbits uniformly distributed?
Since the comets' orbits are defined by the corresponding normal vectors, they reduce to a sample of $n \in \mathbb{N}$ points in the unit sphere $\mathbb{S}^2$. Strictly speaking, the comets' orbit can be seen as an observation from the $\mathbb{S}^2$-valued random variable v. Testing the null hypothesis that v is uniformly distributed on $\mathbb{S}^2$ from a random sample $v_1, \dots, v_n$ of this variable can be performed.

(ii) Are the perihelion directions uniformly distributed?

(iii) Are the orbital planes of comets with a given perihelion direction distributed uniformly about that axis?

Database. The dataset for long-period cometary orbits can be found in the catalog of Marsden and Williams (1993) [MW93]. It is composed of 658 single-apparition long-period cometary orbits, 315 of these are prograde and 343 are retrograde. Another different and more updated (since 2007) database for orbits of long-period comets (Chiron-type comet `CTc`, Encke-type comet `ETc`, Halley comet `HTC`, Jupiter-family comets (`JFc, JFC`), hyperbolic cometary orbits `HYP` and parabolic cometary orbits `PAR`) is available through the `comets` object of the `spunif` R package in [GPVN21]. It is based on the records of NASA's JPL Small-Body Database Search Engine (`https://ssd.`

jpl.nasa.gov/sdbd_query.cgi) and sorted through the JPL's database ID, which is assigned chronologically based on the discovery of new comets.

Datasets for planet orbits are provided by different observatories or space missions, for example, the Planetary Data System (PDS), which archives and distributes digital data from past and present NASA planetary missions, astronomical observations, and laboratory measurements. The R package `planets` in `spunif` (see [GPVN21]) contains planet orbits data from the JPL Keplerian Elements for Approximate Positions of the Major Planets.

Concerning the database for solar system small bodies, the JPL Solar System Dynamics Group's Small-Body Database Query (Group 2023a: https://ssd.jpl.nasa.gov/tools/sdbd_query.html and Group 2023b: https://ssd-api.jpl.nasa.gov/doc/sdbd.htm) and MPC Orbit (MPCORB) database from the Minor Planet Center (MPC) 2022 (https://minorplanetcenter.net/iau/MPCORB/MPCORB.DAT) contain updated databases of Plutinos.

Statistical tools and data analysis.

Back to the roots - the first tests. To address (i)-(iii), a natural way to proceed consists of using a uniformity test (Sect. 2.2) on the sphere. Uniformity of the orbits (resp., perihelion directions) means that the underlying random rotation matrix $\mathbf{X}$ (resp., u) has the same distribution as $\mathbf{RX}$ (resp., $\mathbf{R}u$), for any fixed rotation matrix $\mathbf{R}$. In particular, Mardia (1975) [Mar75] tackled (ii) by employing the classical Rayleigh test; Watson (1983) [Wat83, pp. 28-32] did the same but for directed normals; Fisher et al. (1987) [FLE87, p. 161] studied (iii) by a uniformity test against bipolar alternatives; Jupp and Mardia (1979) [JM79] applied their likelihood ratio test for uniformity to analyze the joint distribution of the directed normals and the perihelion directions; and Mardia and Jupp (2000) [MJ00, pp. 283, 288-289] used several classical uniformity tests such as the Rayleigh (2.7) and Sobolev (2.10) tests to study the joint distribution of both the directed normals and the perihelion directions. The latter as well as Jupp and Spurr (1983) [JS83] also tested rotational symmetry of orbital planes about perihelion directions. Furthermore, Hall et al. (1987) [HWC87] employed a completely different approach by proposing a kernel density estimation for the directed comet orbit's normals.

In all these studies, strong statistical evidence against the uniformity assumption for the distribution of comet orbits has been obtained. Thus the non-uniformity of the orbits comes largely from the non-uniformity of the perihelion distributions. Such a conclusion is somewhat surprising as it contradicts the isotropy properties which could be expected a priori. This motivated Jupp et al. (2003) [JKKW03] to do a careful data analysis (such as QQ-plots) by investigating the intrinsic distribution of the directed normals, that is, the distribution prior to selection by any observational biases, and in particular to test it for uniformity. The previous studies by statisticians have not taken selection effects into account and have tended to reject uniformity. [JKKW03] believed that the lack of uniformity is largely due to an observational bias that favors the observation of comet orbits near the ecliptic. Therefore they proposed a new probabilistic model on the celestial sphere $\mathbb{S}^2$ which incorporated an observational window defined by the inequality $|\sin(i)| \leq \epsilon$. The parameter ϵ can be estimated by maximum likelihood. Their proposed model shows satisfactory goodness of fit of the data set at hand.

New approaches. The paper [JKKW03] inspired other statisticians to propose new approaches to tackle the uniformity assumption of the comet orbits. Cuesta-Albertos et al. (2009) [CACF09, Sect. 5] presented a new class of nonparametric uniformity tests based on random projection including uniformity for spherical (directional) as well as compositional data, sphericality of the underlying distribution, and homogeneity in two-sample problems on the sphere or the simplex. Their proposed procedures have a number of benefits: flexibility, computational simplicity, and ease of application even in high-dimensional cases. They applied their two nonparametric random projection tests termed RP50 and RP100 (in comparison with uniformity tests by Giné and Rayleigh (2.7)) to test if the comet orbits are uniformly distributed with NASA's JPL Small-Body database by December 2017, consisting of 208 observations. As one can observe in Table 1, [CACF09] found no statistical evidence against the uniformity hypothesis.

Inspired by the work of [CACF09], Garcia-Portugués et al. (2023) [GPNECA23, Sect. 5.2] proposed also a projected-based class of uniformity tests on the hypersphere (see Sect. 2.2) but using the weighted quadratic discrepancy between the empirical cumulative distribution function of the projected data and the projected uniform distribution. Their proposed class, despite its different origin, can be related to the well-studied Sobolev class of uniformity tests and allows deriving new tests for hyperspherical data that neatly extend the circular tests by Watson (2.9) and Rothman (2.10). In fact, their tests constitute the first instance of an Anderson-Darling-like test for such data. They applied their new tests on three astronomical investigations including the analysis of 208 long-period elliptical-type single apparition comets as in [CACF09] and of 438 comets as of May 7 2020 by restricting to comets with distinct (i, Ω) up to second digit. In conclusion, they obtained a noticeable difference between the two datasets, for the first with 208 data (see Table 1) uniformity is not rejected at significance level 5% (findings similar to those in [CACF09]), contrarily to the second dataset of 438 comets, where a significant non-uniformity in the orbits of long-period comets with updated records has been detected. As described in [JKKW03], the observational bias of long-period comets may explain the leading rejection cause.

Comets	Orbits	RP50	RP100	Giné	Rayleigh	CvM	AD	Rt
Long-periods	208	0.248	0.250	0.060	0.210	0.1011	0.0744	0.1207
	842	/	/	/	/	0.0041	0.0023	0.0052

Table 1: p-values of different uniformity tests for the NASA comet data set (208 data) from [CACF09] and [GPNECA23].

Park (2012) [Par12] proposed a geometric kernel density estimator on the tangent space of $\mathbb{S}^2$ that he applied on the tangent vectors corresponding to the observations on normals to the orbital planes of long-period comets. His numerical study has been done using a very old long-period comet with a sample size of 240 from [FLE87, pp. 160-161].

Garcia-Portugués et al. (2022) [GPdMMV24, Sect. 5.1] showed that there is a possible manifestation of observational bias, both in long- and short-period comets, by using their nonparametric tests of independence based on trigonometric moments. For their study, they made use of the dataset of 455 long-period comets and 842 short-period comets. Similarly as in previous works, they found no evidence against autocorrelation in long-period comets contrary to short-period comets.

3.1.1 Pluto's origin and small solar system bodies

Pluto's origin. As mentioned at the beginning of this section, the planet Pluto is no longer a true planet in our Solar System. Since 2006, IAU downgraded the status of Pluto to that of a dwarf planet, because it did not meet the third of three criteria the IAU uses to define a full-sized planet, i.e., it has not "cleared its neighboring region of other objects". However, for 76 years, the planet Pluto was the beloved ninth planet and was the runt of our Solar system. At that time, astronomers questioned whether Pluto is a major planet or not, if it would be better classified as a large asteroid or comet rather than as a planet, or if it was rather an entrapment by the Sun of an extrasolar object. Due to its highly eccentric orbits, some scientists consider it to be the largest of the Kuiper Belt objects (also known as Trans-Neptunian Objects). In fact, Pluto rotates in the opposite direction from most of the other planets and its orbital inclination is also much higher than that of the other planets. In addition, Pluto does neither fit in the group of Jovian planets, which are large, gaseous, low-density worlds, nor into the Terrestrial group because it lies far away. Therefore, at that time, there was (and still is today!) a controversy over Pluto's status.
In 2002, Patrangenaru and Mardia (2002) [PM02] tested the solar nebula theory for Pluto using both parametric and non-parametric bootstrap methods. According to the solar nebula theory, originally proposed by Kant, Laplace, and others, the solar system is presumed to be formed from a nebula that evolved into a disk and thus the planets are formed from an accretion disk of material that, over time, condenses into dust, small planetesimals. Hence the planets should have, on average, coplanar, nearly circular orbits. However, Pluto's orbit is more eccentric and more tippy. This led [PM02] to propose a new non-parametric bootstrap approach based on a large sample theory for extrinsic means of asymptotic distributions on a manifold. They regarded the major planets as a sample and determined their mean axis of rotation. Similar as for the cometary orbits, they determined each planetary orbit by the directed unit vector v (3.1). For their analysis, they used the University of Uppsala data [MJ00, Table 10.2, p. 209] giving measurements of (i, Ω) for the nine planets. Their parametric and non-parametric analyses provided very strong evidence that the solar nebula theory does not hold for Pluto. Both approaches showed that Pluto is too strong an "outlier" within the population of the major planets of the solar system.

Small solar system bodies. Recently, Matheson et al. (2023) [MMK23] applied the FvML distribution (2.4) to the orbit pole distribution of the observed Plutinos, which are Kuiper Belt objects having orbital periods very similar to that of Pluto. In fact, small solar system bodies such as Plutinos have widely dispersed orbital poles, posing challenges to dynamical models of solar system origins and evolution. With the help of a functional form for a model of the distribution function, it is possible to characterize the orbit pole distribution of dynamical groups of such small bodies. Moreover, since the orbital pole is a directional variable, its distribution can be more appropriately modeled with directional statistics. That is exactly what Matheson, Malhotra, and Kean have done in [MMK23] by using the FvML distribution which naturally accommodates all physical inclinations (and no others), whereas Gaussian models must be truncated to the physical inclination range $0° - 180°$. They obtained an exact value of the Plutinos' FvML mean poles and concentration parameters. For their study, they used a current observational sample of Plutinos of the JPL Solar System Dynamics Group's Small-Body Database Query and MPCORB.DAT database.

Inspired by [MMK23], Malhotra and Roy (2023) [MR23] used the FvML distribution (2.4) to model the free inclinations of classical Kuiper belt objects (CKBOs). The CKBOs are known to have two components of orbital planes: the "cold" component concentrated near the forced plane and the more widely dispersed "hot" component. Similar to the comet's orbit, the orientations of orbital planes of minor planets are directional random variables and their free inclination is the deviation of the orbit plane from the plane forced by the major planets. To model the CKBOs' free inclination distribution, [MR23] adopted a linear combination of two FvML functions f_{FvML} for an inclination $i \in [0, \pi]$ as

$$f(i; \kappa_1, \kappa_2, \omega) = \omega f_{\text{FvML}}(i; \kappa_1) + (1 - \omega) f_{\text{FvML}})(i; \kappa_2),$$

where ω is a weight factor and $f_{\text{FvML}}(i; \kappa) = \frac{\kappa}{\epsilon^{\kappa} - \epsilon^{-\kappa}} \exp(\kappa \cos(i)) \sin(i)$ is the functional form of the FvML distribution with concentration parameter κ (not too small). They found that the cold and hot components respectively account for 57% and 43% and that the mixture of two FvML functions is a good fit for the cold CKBOs but not very good for the hot CKBOs.

3.2 Craters on planets and moons

Motivation. In astrogeology, counting the craters on planets or moons is the primary method for determining remotely the relative age of a planetary surface, plus it also gives valuable insights into the past geologic processes, resurfacing history, and the planetary subsurface structure, according to Fassett [Fas16] who gave a review on crater statistics and their applications. In fact, craters are roughly circular depressions resulting from impact or volcanic activity. Among the main generators of the non-isotropic impact cratering process are short-period comets, in particular, those in the outer solar system (see, e.g., [ZSLD03] and references therein). Thus, crater locations and diameters D provide important information on the history of a planet or moon. To put this in a statistical context, consider the observed craters on $\mathbb{S}^2$ containing the planetocentric coordinates of the crater's centers. Then one might test if:

(i) the planets' and moons' craters are uniformly distributed, in particular, the distribution of Venusian craters, and for four Saturnian moons including Rhea;

(ii) there is a form of uniformity for small craters (diameter D between 15 km and 20 km, i.e., $15 < D < 20$), for large craters ($D > 15$) or all reliable-detected craters ($D > 20$).

In Table 2, we observe that among the irregularity of uniform crater distributions in the solar system, the apparent uniform distribution of Venusian and Saturnian moons' craters, especially Rhea, is truly remarkable. Venus is the closest planet to Earth and the most Earth-like planet in the solar system in terms of size and composition. The filtering of small meteoroids by the dense Venusian atmosphere may be one of the causes explaining the uniform distribution of craters.

Rhea orbits Saturn synchronously, it has a leading hemisphere that always faces forward into the orbit motion and a trailing hemisphere that faces backward. Hirata (2016) [Hir16] pointed out that planetocentric impactors weakly favor the centers of the leading and trailing hemispheres, referred to as apex and antapex, respectively, whereas the cratering on the leading hemisphere is expected from heliocentric impactors. This may therefore lead to a non-uniform crater distribution by both populations of impactors. Furthermore, [Hir16] found no apparent apex-antapex asymmetry in small craters as opposed to larger craters.

Database. The Gazetteer of Planetary Nomenclature (GPN) database (`https://planetarynames.wr.usgs.gov/AdvancedSearch`) of the IUA is a publicly available search engine containing the data of planetary surface feature names, i.e., provides maps of planets and locations of planetary craters. It is maintained by the Planetary Geomatics Groups of the United States Geological Survey (USGS) Astrogeology Science Center (`https://www.usgs.gov/centers/astrogeology-science-center/data`). Note that the position of craters is recorded in areocentric coordinates so that the areocentric longitudes range from $0°$ to $360°$ and areocentric latitudes range from $-90°$ to $90°$.

In the R package `spunif` of [GPVN21], the object `craters` contains the named craters of the GPN database, and the objects `venus, rhea` provide a separate dataset of the coordinates of the Venusian craters (from the USGS Astrogeology Science Cente) and Rhea's (from [Hir16]) craters, respectively. According to the database of named craters of the IUA, Venus has the largest number of observed craters.

Statistical tools and data analysis.

Martian craters. Zhang and Chen (2021) [ZYC21] applied their proposed kernel density estimator together with the directional mean shift algorithm to further estimate the density of craters and determined crater clusters on the surface of Mars. They made use of Martian crater data provided by the GPN database. Before applying their tests, they trimmed the data to 1653 craters by removing the small craters whose diameters are less than 5 kilometers, as their presence may provide spurious information. Moreover, since the craters with areocentric longitudes greater than $180°$ are on the western hemisphere of Mars, they transformed their longitudes back to the interval $(-180°, 0°)$. Their finding aligns with prior research on the Martian crater distribution, stating that Mars can be divided into two general classes of terrain [SCW+74]. More precisely, their mode clustering succeeded in capturing two major crater clusters (or basin of attraction) on Mars, in which one cluster is densely cratered while the other is lightly cratered.

Crater distributions in the Solar system with a focus on Rhea's craters. Garcia-Portugués et al. (2023) [GPNECA23, Sect. 5.3] evaluated the rareness of uniform crater distributions in the solar system by analyzing the GPN database of May 31 2020 containing 4818 observations on $\mathbb{S}^2$ for 14 bodies after filtration of non-asteroid bodies (the full database contains 5236 craters for 44 bodies). They conducted their three different uniformity tests (CvM, Rt, and AD-like test, see Sect. 2.2) for this dataset and rejection the hypothesis of uniformity at significance level 5% in all bodies except Venus and four Saturnian moons (see Tab. 2).

Class	Name	Craters	CvM	AD	Rt
Planets	Mars	1127	0	0	$1 \cdot 10^{-8}$
	Venus	881	0.2726	0.2749	0.2806
	Mercury	409	0	0	0
Dwarf	Ceres	115	0.0133	0.0127	0.0159
Moons	Moon	1578	0	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$
	Callisto	141	0	0	$3 \cdot 10^{-8}$
	Ganymede	129	0.0132	0.0087	0.0184
	Europa	41	0.0010	0.009	0.0010
Saturn's	Rhea	128	0.2793	0.2954	0.2705
moons	Dione	73	0.5195	0.4989	0.5418
	Iapetus	58	0.0034	0.0037	0.0032
	Enceladus	53	$1 \cdot 10^{-7}$	$2 \cdot 10^{-8}$	$5 \cdot 10^{-7}$
	Tethys	50	0.7910	0.8425	0.7199
	Mimas	35	0.1701	0.1704	0.1754

Table 2: Asymptotic p-values of the CvM, AD and Rt tests when applied to Hirata's crater database [Hir16]. Outcomes from the study of [GPNECA23].

According to [GPNECA23], these few non-rejections are suspected to be driven by a uniformity bias in the data. Well-separated craters that cover the body are likely more probable to be named than those that cluster. Thus a detailed crater database (for the moment only available for certain bodies such as Venus and Rhea) is required to bypass this source limitation. For this reason, [GPNECA23] investigated in detail the crater distribution of Rhea, the second most cratered body in Table 2 by performing their three tests on 867 craters with $15 < D < 20$, 1373 craters with $D > 20$, and 2440 craters with $D < 15$ using Hirata's database [Hir16]. In Table 3, their tests revealed that uniformity is not rejected for small craters ($15 < D < 20$) at significance level 5% but is emphatically rejected for large and all reliable-detected craters ($D > 20$). The non-rejection of uniformity for small craters may be explained by Rhea's "crater saturation" [SHLL97] or by the dominance of planetocentric impactors, as the largest crater generated by the debris ejected from large crater impact has $D \approx 20$ [AZDH05]. The rejection of uniformity for large craters may be attributed to the predominantly heliocentric origins of the impactors associated with large craters [Hir16].

Diameter	Craters	CvM	AD	Rt
$15 < D < 20$	867	0.1176	0.0721	0.1856
$D > 20$	1373	$2 \cdot 10^{-9}$	0	$3 \cdot 10^{-9}$
$D < 15$	2240	$2 \cdot 10^{-8}$	0	$3 \cdot 10^{-8}$

Table 3: Asymptotic p-values of the CvM, AD, and Rt tests when applied to Hirata's Rhea crater database [Hir16]. Outcomes from the study of [GPNECA23].

Venusian craters. Garcia-Portugués et al. (2021) [GPNECA21, Sect. 5] tested the uniformity of 967 Venusian craters by performing the Rayleigh (2.7) and Giné F_n tests, the [CACF09] test (using $k = 50$), and their novel projected-based CvM-based test on $\mathbb{S}^2$. They found no statistical evidence to reject the uniform hypothesis at usual significance levels.

Recently, Hallin et al. (2022) [HLV22] applied their new quantile concepts and goodness-of-fit tests to the Venus craters dataset. Concretely, they divided the data into two groups: 453 craters with $D > 15$ and 514 craters with $D < 15$, and then they computed the p-values of the n uniformity measure transportation-based goodness-of-fit tests. Their results revealed that, while the null hypothesis of uniformity is essentially not rejected at traditional levels, the p-values associated with their test are uniformly larger for the large craters than for the small craters. On the contrary, the other tests (projected CvM, AD and Rt procedures by [CACF09] and [GPNECA21]) have similar p-values for both types.

3.3 Space weather and sunspots

Motivation. Predicting our Sun's activity is a crucial task in the heliophysics and space weather community. It permits us to protect our technology on Earth as well as satellites and astronauts in space which might be impacted by geomagnetic storms when they are heading towards our Earth. They are initially triggered by the nestling of groups of sunspots when they explode to solar flares or coronal mass ejections. Over the last decade, some attempts have been made to study these darker and cooler regions – sunspots – which vary throughout solar cycles on the Sun's photosphere, particularly when they are heading to peak levels. In particular, the focus was to:

(i) investigate the significance of (non-)rotationally symmetric patterns of sunspots, and

(ii) detect temporary longitudinal non-uniformity in sunspots.

Database. Sunspots data containing daily positions and areas of sunspots can be found in the Debrecen Photoheliographic Data (DPD) sunspot catalog (`http://fenyi.solarobs.csfk.mta.hu/DPD/`, [BGL16, GBL13] and [GLB17] are good references on how to employ the DPD data correctly). It contains observations of sunspot locations since 1974 and is a continuation of the Greenwich Photoheliographic Results catalog, which spanned 1872-1976. Note that the DPD catalog delimits the central position of a group of sunspots through an area-weighted average of the sunspots within the group. This means that within this setting both the temporal (observations belong to a single cycle that aggregates the different latitude-appearance regimes) and spatial (clusters of related sunspots are treated as a single observation) dependency of the data is mitigated. To better accommodate this averaging, Garcia-Portugués et al. (2020) [GPPV20] advised considering an i.i.d. framework when working with those data. Moreover for R users, [GPPV22] created in their available R package `rotasym` the dataset `sunspots_births`, which accounts just for the first-ever observation (referred to as "birth" henceforth) of a group of sunspots. For example, the 21st solar cycle (March 1976 – September 1986) includes 6195 observations, the 22nd solar cycle (September 1986 – August 1996) has 4551 observations, and the 23rd solar cycle (August 1996 – December 2008) contains 5373 observations.

Statistical tools and data analysis.

Rotational symmetry and uniformity tests. Motivated by the central role played by rotationally symmetric distributions (see Sect. 2.1.2) in directional statistics and its presence in the formation of sunspots that follows a nearly rotationally symmetric pattern [Bab61], Garcia-Portugués et al. (2020) [GPPV20] considered the problem of testing rotational symmetry on the hypersphere where the location of symmetry axis is either specified or unspecified. By developing new distributional-based tests for rotational symmetry that inspect the circular uniformity of the longitudes of sunspots with respect to a specified or unspecified location, they investigated a quantification of the significance of non-rotationally symmetric patterns in the emergence of sunspots, even though the main driving force generating sunspots is of a rotationally symmetric nature (intense magnetic field concentrations at equal latitudes due to the Sun's differential rotation). In fact, Babcock (1961) [Bab61, pp. 574, 581] pointed out that sunspots tend to emerge in "preferred zones of occurrence" associated with longitudes where there has been prior activity. [GPPV20] analyzed the rotational symmetry about the north pole $\boldsymbol{\theta}_0 = (0,0,1)^T$ (or, equivalently, about the south pole $-\boldsymbol{\theta}_0$) of the appeared central position groups of sunspots. They obtained a significant non-rotationally symmetric emergence pattern during that cycle for both unspecified-$\boldsymbol{\theta}$ and specified-$\boldsymbol{\theta}$. However, due to the non-omnibus nature of the tests employed in their analysis, non-rotationally symmetric deviations for which the tests are not consistent may have been undetected. This explains the need for a more complex modeling.

A further investigation of the rotational symmetry of sunspots has been done by [GPNECA23]. They applied the CvM, AD and Rt tests (see Sect. 2.2) to the longitudes about $\boldsymbol{\theta}_0 = (0,0,1)^T$ for the same observed sunspots in cycle 23 as in [GPPV20] but also for the cycle 22. Their outcomes are coherent with those in [GPPV20], see Table 4, and thus highlight the varying behavior of different cycles.

Cycle	Sunspots	CvM	AD	Rt	$\phi_\theta^{\mathrm{sc}}$	$\phi_\theta^{\mathrm{loc}}$	$\phi_\theta^{\mathrm{hyb}}$
23	5373	0.3959	0.8393	0.3285	0.1656	0.4571	0.2711
22	4551	0.0067	0.0139	0.0091	0.1077	0.0125	0.0103

Table 4: Asymptotic p-values of the CvM, AD, Rt tests from [GPNECA23] and rotational symmetry tests $\phi_\theta^{\mathrm{sc}}$ (s-sc), $\phi_\theta^{\mathrm{loc}}$ (s-loc), $\phi_\theta^{\mathrm{hyp}}$ (s-hyp) from [GPPV20].

Another interesting approach is given by Jammalamadaka and Terdik (2019) [JT19, Sect. 8]. They analyzed the longitudinal sunspot activity using their model-free inference for spherical distributions based on *harmonic analysis*. More precisely, they estimated the shift α yearly for the DPD sunspots data set (including 187223 sunspots positions) between 1976-2014 by making use of the Dimroth-Watson distribution (2.6) for their model. Their plotted estimated shifts $\hat{\alpha}$ and $-\hat{\alpha}$ showed a nearly "Butterfly" pattern, which refers to the movement of girds by the years.

Fernandez-de-Marcos and Garcia-Portugués (2023) [FdMGP23, Sect. 4] investigated further the longitudinal non-uniformity patterns in sunspots births for the 21st-23rd solar cycles within a rolling window conformed by 10-Carrington rotations (approximately, nine months). Their analysis consisted of three tests, (1) their projected AD-based uniformity test of sunspot births longitudes, (2) circular-linear kernel density estimation [GPCGM13] level sets of sunspot births through time and longitude, and (3) scatter plot of sunspots births with the circular Nadaraya-Watson regression [MPT12] for northern, southern and both hemispheres. They obtained interesting facts, namely (a) the flip-flop effect between 180° active longitudes is not obvious throughout all the cycles, (b) the length and quantity of uniformity periods differ between solar cycles, (c) non-uniformity periods are scarce to claim the existence of active longitudes, and (d) in general, both hemispheres seem to have different behavioral patterns, both in terms of longitudinal non-uniformities and sunspots activity level, along solar cycles.

Recently, Jeon and Van Keilegom (2023) [JvK23] proposed novel nonparametric density and regression estimators to test the rotational symmetry of sunspots by considering the dataset "sunspots_births" in the R package of [GPPV22]. Since it is well known that sunspots area observation may contain measurement errors [BGL16], they considered the Laplace distribution on the special orthogonal group SO(3) for the measurement error distribution and estimated the density of the birth locations of sunspots based on their deconvolution density estimator. They obtained contour plots of the estimated density based on their proposed method, and of the 95% pointwise confidence intervals for the true density based on asymptotic normality (see Figs. 1 and 2 in [JvK23]). Moreover, they observed that the true density is likely to have a higher mass as the latitude approaches zero and that the density levels are horizontal. This can be explained by considering the sunspots as a consequence of the twisted solar magnetic field caused by the fast rotating speed.

A toroidal point of view. An interesting exploration of the dependence structure of sunspots births longitudes for cycle 23 is given by Zoubouloglou et al. (2021) [ZGPM21, Sect. 5.1]. Since the

Earth rotates around the Sun, the visibility of sunspots is limited by which side the Sun is facing the Earth and the sunspots can thus be considered as toroidal data. More precisely, [ZGPM21] considered the series of longitudes and its lagged versions of order one and two

$$\mathbf{\Theta}_i := (\theta_i, \theta_{i+1}, \theta_{i+2})^T \in \mathbb{T}^3,\ i = 1, \ldots, 5371,$$

where $\theta_i \in [-\pi, \pi)$ are the longitudes of sunspots births. Hence, if there existed significantly preferred longitudes, a significant deviation from a linear dependence structure in the series of sunspots longitudes should be expected. To investigate this serial structure, they employed their new Scaled Torus Principal Component Analysis (ST-PCA), a dimension reduction method for toroidal data, to show graphically the main driver behind that selection as the output on $\mathbb{T}^3$. The term scaled refers to the construction of a map via multidimensional scaling to minimize the discrepancy between pairwise geodesic distances in both spaces $\mathbb{S}^d$ and $\mathbb{T}^d$. Roughly speaking, the ST-PCA seeks a data-driven map from a torus to a sphere with the same size and radius as the torus. The analyses were done on spheres by obtaining first the nested sequence of subspheres using spherical multidimensional scaling, and after finding the best fits of data, it can be inverted back to the torus. It turns out that the apparent linearity of the innovations in the series of sunspots births longitudes points towards the non-existence of (at least) major preferred longitudes during the 23rd solar cycle. This is coherent with the non-rejection of rotational symmetry reported in [GPPV20] and [GPNECA23].

3.4 Electromagnetic radiation: the cosmic rays

Motivation. Since the unexpected discovery in 1912 from the Austrian physicist Victor Hess that our Earth was receiving cosmic radiation from our Universe, cosmic ray physicists have been interested in the problem of discovering the sources of cosmic rays, now known to be the nuclei of hydrogen and the centers of heavier elements. Hence nowadays observed cosmic rays are seen as ordinary particles (electrons, protons, and nuclei), propagating in empty space, and deflected by various galactic magnetic fields through which they will have passed.

Since then a number of experiments have been carried out in a number of fieldsm providing a boosting amount of cosmic ray data. In the beginning, the data was far from the level needed to produce precise estimates of parameters and conclusive tests. This changed completely in the last couple of years by improving the size and precision of existing observations by several orders of magnitude. Let us name some important experiments:

- **Cosmic microwave background (CMB)** is the cooled leftover radiation, the first and thus oldest light that could ever travel freely throughout our Universe from the time when the cosmos began, what is universally known as the Big Bang (see for instance [Dod15] and [Bal07] for a more detailed description). Hence, this radiation provides an image of our Universe as it was approximately 13.7 billion years ago. Although the existence of CMB radiation dates back to 1965 in a celebrated experiment by Penzias and Wilson, it was only in the current century with the help of several ballon-borne experiments, ground-based observatories, and in particular of sophisticated satellite missions such as *COBE* (NASA satellite mission in 1992), *WMAP* (NASA satellite mission in 2003), or *Planck* (ESA satellite mission in 2009), that one could manage to produce high-resolution low-noise full-sky maps of the Last Scattering Surface. These photons, now observed in microwave frequencies and in every direction on the sky, constitute the CMB.
- **Ultrahigh-energy cosmic rays (UHECRs)** are particles of unknown origin that arrive at the Earth from apparently random directions of the sky. When these particles interact with atoms of the upper atmosphere, they generate a huge cascade of billions of secondary particles. Observing these secondary particles (with appropriate detectors on the ground) allows us to measure the direction of arrival as well as the energy of the original cosmic ray (see [KO11] and [B$^+$19] for a recent review and open questions on the astrophysics of UHECRs). The Pierre Auger Observatory (PAO) is one of the huge international collaborations that investigate UHECRs.
- **Gamma ray (or γ-ray).** Gamma-ray photons are quanta of light in the highest energy range and are usually generated from accelerated charged particles, such as electrons or protons, boosted by extreme celestial objects such as supermassive black holes, pulsars, active galactic nuclei (AGN) or supernova remnants. They provided the basis for a large number of astronomical discoveries. In particular, the γ-rays emitting sources improve our understanding of high-energy astrophysical phenomena, and might even resolve the mystery of the fundamental

nature of dark matter [MBM23]. Gamma rays sources have been probed by satellite missions such as *Fermi-LAT* and *AGILE*, as well as from several observatories on the ground.

Gamma-ray bursts (GRBs) are emissions of gamma rays linked with extremely energetic explosions that have been recorded for distant galaxies. Bursts vary in duration from ten milliseconds to several minutes. GRBs may originate from different processes, e.g., a supernova event such as the birth of a neutron star, a quark star, or a black hole, as well as of the death of a massive star. GRBs have been classified mostly as "long" for the ones having a duration of greater than two seconds and "short" otherwise. The former class is of greater interest since almost all of its members could be linked to a galaxy with rapid star formation or the deaths of massive stars.

Database. Old catalogues of several hundred high energy (at least 10^{19} eV) cosmic rays exist which give the direction of their arrival from space and the associated estimate of their energy. The data have been recorded at four arrays over a number of years: Haverah Park, UK (1°38′ W, 53°58′ N) with 224 showers [RW80]; Volcano Ranch, USA (106°47′ W, 35°10′ N) with 44 showers [Lin80]; SUGAR, near Narrabri, NSW, Australia (149°36′ E, 30°32′ S) with 423 showers [WUP+86a, WUP+86b]; and Yakutsk, Russia (129° E, 62°N) with 233 showers [EEK+80].

Fisher et al. (1987) [FLE87, data set B3] have digitized the arrival directions of 148 low mu showers of cosmic rays (Bolivian Air Shower Joint Experiment) from [T+65]. This cosmic ray data set has been used by [PV17] and [LSV14] for their study.

A more updated catalog is provided by the PAO (`https://www.auger.org/collaboration/`) which has a huge public catalog of UHECRs [AAA+10]. For example, [FDKP13, LPN14] and [KKPN13] considered 69 arrival directions of cosmic rays with energy above 55 EeV which were detected by the PAO between 1 January 2004 and 31 December 2009. In addition, the Pierre Auger Collaboration [AAA+08, AAA+09, AAA+10] also performed a catalog-free test for anisotropy with no reference to any catalog, using the TwoPC procedure.

The dataset on GRBs in celestial coordinates is available on `http://swift.gsfc.nasa.gov/docs/swift/archive/grb_table`. This data has been recorded by the Burst Alert Telescope (BAT), which is one of the three instruments on the Swift MIDEX spacecraft by NASA, launched into orbit on 20 November 2004.

The Fermi Large Area Telescope (LAT) Collaboration (`https://glast.sites.stanford.edu/`) provides a γ-ray database of over 1 billion photons in the energy range from about 20 MeV to more than 300 GeV collected in over a decade of operations. The dataset typically consists of an event list that gives the direction in the sky of each detected photon together with additional information like its energy content and the so-called event type which expresses the quality of the measurement. The recorded data by the LAT can be represented in polar coordinates or in Galactic coordinates on the unit sphere. This dataset is particularly used for the search of yet undetected γ-ray sources as done in [MBM23].

Concerning CMB data, LAMBDA (Legacy Archive for Microwave Background Data Analysis) (`https://lambda.gsfc.nasa.gov/`), which is part of NASA's High Energy Astrophysics Science Archive Research Center (HEASARC), provides free public CMB data. The same holds for the ESA satellite mission Planck (`https://www.cosmos.esa.int/web/planck` and `https://pla.esac.esa.int/#home`).

For radio survey catalogs, NRAO VLA Sky Survey (NVSS) (`https://heasarc.gsfc.nasa.gov/W3Browse/all/nvss.html`), Parkes-MIT-NRAO (PMN) Surveys (`https://www.parkes.atnf.csiro.au/observing/databases/pmn/pmn.html`) and the Green Bank 1987 (897GB) survey (`https://www.parkes.atnf.csiro.au/observing/databases/aitoff.html`) contain a database of ~ 40000 sources brighter than 50 mJy at 4.85 GHz, over about 70 percent of the sky, and represent the largest and least biased sample of X-ray and radio-loud AGN available at present.

Other relevant cosmic rays data are Cosmic Neutrinos investigated by IceCube (`https://icecube.wisc.edu/`), Gravitational Waves data by the LIGO-Virgo Collaboration (`https://www.ligo.org/`), Radio sources data by huge collaborations such as the Square Kilometre Array Observatory (SKAO) (`https://www.skao.int/`), Weak gravitational lensing data found in [B+09] and [K+11b], and Large-scale structure of the Universe as part of the upcoming mission *Euclid*, surveys can be found in [Cra09, Sect. 4].

Statistical tools and data analysis.

A conditional and experimental approach. Mardia and Edwards (1982) [ME82] observed that in the analysis of astrophysical data (e.g., the arrival of cosmic rays and asteroids), when the monitoring station is on the surface of the Earth, the field of vision is often an off-center cap on the sphere. The coordinate system usually employed is to form the celestial sphere, which arises from

the ancient interpretation of the universe with the Earth at the center. However, it may happen that the observer does not sit at this center and records the directions from another point. For instance, in determining the arrival directions of high-energy cosmic rays, the cosmic radiation is not observed directly. Standard directional data techniques are not applicable since measurements are not possible over the whole sphere at the same instant, but are confined to the off-center cap. It is further complicated by the rotation of the Earth causing rotation of this cap about the Earth's axis along a line of constant latitude. This has led to assume that the data arise from a rotationally symmetric distribution.
Therefore [ME82] proposed the use of weighted distributions assuming uniform rotation of an off-center cap along a line of constant colatitude and tested whether the resulting distribution shows concentration along the galactic plane or perpendicular to it. They applied their conditional model [ME82, Eq. (3.2)] to the data set of the arrival directions of UHECRs provided by Pollock [ME82, Fig. 3 and Table 1.-2.] and obtained a likelihood conditional equation that numerically yields two distinct sets of maximum likelihood estimators. One of them corresponds to a girdle distribution, while the other is bipolar, which is almost perpendicular to the first one. They did not reject their null hypothesis of a "weighted" uniform distribution, meaning that the original longitudes are uniformly distributed.

An alternative approach for modeling the arrival directions of cosmic rays has been proposed by Boulerice and Ducharme (1994) [BD94]. They suggest to generate families of models for directional data that can arise from the physical context of an experiment. This means that experimental conditions can prevent the coincidence described above from happening. By applying their modified Directional Decentered Uniform model (MDCU) [BD94, Eq. (6.1)], they tested whether the cosmic rays data from [T$^+$65] stem from a uniform distribution over a given subset of the sphere. They did not reject the hypothesis at the 5% level, which agrees with the empirical finding of [T$^+$65] and [ME82].

Classical uniformity tests. Marchant (1990) [Mar90] investigated the arrival directions of cosmic rays recorded at four different latitudes on the Earth's surface [RW80, Lin80, WUP$^+$86a, WUP$^+$86b, EEK$^+$80]. By using two classical uniformity tests on the circle (Kuiper's (2.8) and Rayleigh's (2.7) test), he tested if there is some evidence of uniformity of these observations in right ascension. In his analysis, he found out that Kuiper's test is more robust to such kind of data than Rayleigh's test and that the results, for both tests, indicated that the data are consistent with the hypothesis that the cosmic rays arrive uniformly from all directions. By examining the data more closely, Marchant is convinced that there may be a tendency for the highest energy rays to arrive more frequently from the direction of the center of the local supercluster and there is possibly evidence of anisotropy in the northern hemisphere. However, more data and further analysis are needed to verify or reject both hypotheses. More precisely, he suggested constructing a regression model in which the parameters of a circular distribution (e.g., vM distribution (2.2)) are functions of declination and energy, and some knowledge of any intergalactic magnetic fields which would affect cosmic ray nuclei on their long journey to Earth would also be an asset.

Uniformity tests for noisy and weak data. Astrophysicists have at their disposal directional data which are measurements of the incoming directions of the UHECRs on Earth. However, their trajectories are deflected by galactic and intergalactic fields. As this deflection is inevitable in the measurements, it is essential yet challenging to take into account this uncertainty in statistical modeling. Mathematically, these *noisy* directional observations can be regarded as

$$\mathbf{Z}_i = \epsilon_i \mathbf{X}_i, \quad i = 1, \ldots, n, \tag{3.2}$$

where the ϵ_i's are i.i.d. random variables of SO(3), the rotation group in $\mathbb{R}^3$, and the $\mathbf{X}_i$'s are randomly sampled from a distribution over $\mathbb{S}^2$. The rotational Laplace distribution (ordinary smooth noise) or Gaussian distribution (supersmooth noise) are typical choices of noises. Suppose that $\mathbf{X}_i$ and ϵ_i are independent and that the distributions of $\mathbf{Z}_i$, $\boldsymbol{X}_i$ and ϵ_i are absolutely continuous with respect to the uniform measure on $\mathbb{S}^2$ and the Haar measure on SO(3), respectively. Lacour and Pham Ngoc (2014) [LPN14] as well as Kim et al. (2015) [KKPN13] considered the problem of testing uniformity expressed as

$$\mathcal{H}_0^{\text{unif}} : f = f_0 = \text{Unif}(\mathbb{S}^2) \quad \text{against} \quad \mathcal{H}_1 : f \in \mathbb{H}(\mathcal{F}, \delta, M)$$

in a setup in which the noise density $f_{\boldsymbol{\epsilon}}$ of the $\boldsymbol{\epsilon}$'s is assumed to be known, f stands for the density of $\mathbf{X}_i$, and the alternative consists in the set of densities $\mathbb{H}(\mathcal{F}, \delta, M) = \{f \in \mathcal{F} \mid \|f - f_0\|_{L^2} \geq M\delta\}$, which is a Hilbert space, where M is a constant, δ is referred to as the separation rate, and $\mathcal{F}$ can be considered as a Sobolev class on $\mathbb{S}^2$ with a known smoothness s. In [LPN14, KKPN13], this set $\mathcal{F}$ is explicitly defined and consists of Sobolev densities satisfying some norm-inequality. As Hilbert space, one can consider the space of square-integrable functions $L^2(\mathbb{S}^2)$ with its associated inner product. Observe that the model (3.2) is the spherical convolution model, since for $f_\epsilon \in L^2(\mathrm{SO}(3))$ and $f \in L^2(\mathbb{S}^2)$, we have

$$f_{\mathbf{Z}}(\omega) = f_\epsilon * f(\omega) = \int_{\mathrm{SO}(3)} f_\epsilon(u) f(u^{-1}\omega)\,\mathrm{d}u,$$

where $f_{\mathbf{Z}}$ denotes the density of $\mathbf{Z}_i$ and $*$ the convolution product. By using the Fourier analysis on SO(3) to deal with the noise and on $\mathbb{S}^2$ for the unknown density, [LPN14, KKPN13] provided a non-parametric minimax adaptive testing procedure on f from the noisy observations $\mathbf{Z}_i$.

Another issue is that the signal of the incoming directions of cosmic rays is *weak*. Traditional tests, like the Rayleigh test (2.7), reject the null hypothesis of uniformity for this kind of spherical data that looks uniformly distributed. More precisely, the Rayleigh test rejects the null at asymptotic level 5%, while visual inspection of the left panel of Fig. 4 in [PV17] suggests that concentration is quite moderate, implying that inference on the modal location $\boldsymbol{\theta}$ is likely delicate. To tackle this challenge, Paindaveine and Verdebout (2017) [PV17] considered the problem under asymptotic scenarios for which the signal strength λ_n goes to zero at an arbitrary rate $\boldsymbol{\eta}_n$ as $n \to \infty$. This is equivalent to allowing the underlying distribution P_n to converge to the uniform distribution $P_0 = \mathrm{Unif}(\mathbb{S}^{d-1})$ at an arbitrary rate. In this context, it can be considered as a location testing problem for the null hypothesis

$$\mathcal{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{against} \quad \mathcal{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

in the vicinity of the uniform distribution, also known as (asymptotic) score test problem with $\boldsymbol{\mu}_0 \in \mathbb{S}^{d-1}$ fixed and spherical mean $\boldsymbol{\mu} = \boldsymbol{\mu}(P)$. Performing inference on $\boldsymbol{\mu}$ is a semiparametric problem whose difficulty depends on the underlying distribution P. If P is much concentrated about $\boldsymbol{\mu}$, then it is in principle easy to, for example, identify small confidence zones for $\boldsymbol{\mu}$. On the contrary, if P is close to the uniform distribution P_0 over the unit sphere, then performing inference on $\boldsymbol{\mu}$ is much more delicate and the corresponding confidence zones will be very broad. The latter corresponds to the situation of cosmic rays data. Under this setting, the known asymptotic distributions of certain statistics of interest may fail, even if n is large, to provide satisfactory approximations of the corresponding fixed-n distribution. To cope with this difficulty, [PV17] revisited, in an original and challenging perspective, the location testing problem. To fully characterize how challenging the problem is as a function of η_n, they adopted a Le Cam, convergence-of-statistical-experiments, point of view and showed that the resulting limiting experiments crucially depend on η_n. In the light of their study, the classical Watson score test based on

$$W_n = \frac{n(d-1)\bar{\mathbf{X}}^T(I_d - \boldsymbol{\mu}_0\boldsymbol{\mu}_0^T)\bar{\mathbf{X}}}{1 - \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i^T\boldsymbol{\mu}_0)^2},$$

where $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$, unlike the Wald test based on the spherical mean, is shown to be robust to asymptotic scenarios, *adaptively* rate-consistent and essentially adaptively Le Cam optimal. In conclusion, the Watson score test remains an extremely competitive test in semi-parametric rotationally symmetric models when one has to deal with distributions that are close to uniformity as the case, for instance, for the cosmic rays data set.

Quantile-based goodness-of-fit test. Ley et al. (2014) [LSV14] envisaged a study of primary cosmic rays in certain energy regions by proposing a new concept of quantiles for directional data and the angular Mahalanobis depth. The notion of *quantile* of a probability distribution is vastly popular in the statistics world to make a visual inspection of the (univariate) data. In recent years, the success of univariate quantiles has stimulated several researchers to try to extend this fundamental one-dimensional concept to higher dimensions. *Depth functions*, such as the half-space, the simplicial depth, the Mahalanobis depth or the zonoid depth, to cite but these, also enjoy a strong popularity among statisticians. Their main appeal is the fact that they provide a center-outward ordering for a given (typically multivariate in $\mathbb{R}^d$) dataset by affecting each point $\boldsymbol{x} \in \mathbb{R}^d$ with a value, the depth of $\boldsymbol{x}$, determining its centrality within the data cloud. Thus depth-based procedures are applied to location estimation, regression, classification (e.g., via DD-plots), trimming, testing procedures (e.g., testing for distinct notions of symmetry), functional data analysis, and several other problems. [LSV14] developed the notion of Mahalanobis depth based on a new (the first proper) concept of quantiles for spherical data, as well as a quantile-based goodness-of-test to tackle the problem

$$\mathcal{H}_0 : P = P_0 \quad \text{against} \quad \mathcal{H}_1 : P \neq P_0$$

for some P_0 in the class $\mathcal{F}$ of probability laws on $\mathbb{S}^{d-1}$ with bounded density and which admits a unique median direction $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. They applied their goodness-of-test to analyze the cosmic rays data set and found that the FvML distribution (2.4) with concentration 0.7 provides the best fit to the data.

A spherical needlets approach. In the last few years, a rather extensive literature has been developed on the construction of *wavelets systems on the sphere* and its new generation, the so-called *spherical needlets* that possess several attractive properties. Some of these attempts have been explicitly motivated by applications in cosmology, in particular, theoretical predictions involving CMB (e.g., [MPB$^+$07, BKMP09, CM09], but also partially [MP11] and references therein), where quantities such as the angular power spectra or the bispectra play an important role. In fact, the statistical procedures for the estimation of angular power spectra and bispectra are based upon spherical harmonic transforms and manipulations of the random coefficients. However, the evaluation of these random coefficients requires that the *spherical random field* is fully observed, which is not the case in practice. More precisely, in the context of CMB data analysis, the presence of foreground emissions from our Milky Way and other astrophysical sources prevents the exact evaluation of spherical harmonic transforms on approximately 20% of the entire sphere. This means that these random coefficients derived from such incomplete maps lose many of their standard properties and thus are no longer uncorrelated.

To address these difficulties, a promising approach is to consider spherical needlets. Their construction is based on two main ideas, the discretization of the sphere by an exact quadrature formula (i.e., discretize the sphere into cubature points and cubature weights) and the Littlewood-Paley decomposition. They possess peculiar asymptotic properties and their random needlets coefficients enjoy a capital uncorrelation property in the sense that, for any fixed angular distance, random needlets coefficients are asymptotically uncorrelated as the frequency parameters grow larger and larger [BKMP09]. This makes it possible for instance to test for cross-correlation among CMB and Large Scale Structure data (see Chap. 10 in [MP11] and the references therein).

Faÿ et al. (2013) [FDKP13] focused on the important question of the isotropy of the UHECRs on the unit sphere. More precisely, they proposed two methods using the decomposition of directional data onto a frame of spherical needlets.

Finally, Duque and Marinucci (2021) [DM23] have given a survey of recent developments in spherical needlets with a view to applications in Cosmology. In their last section, they discuss more recent and challenging issues such as the analysis of polarization data, which can be viewed as realizations of random fields taking values in spin fiber bundles.

Analysis of γ-rays and other radio sources. Protheroe (1985) [Pro85] proposed a new statistic for the analysis of circular data. The statistic is designed specifically for situations where one requires a test of uniformity which is powerful against alternatives in which a small fraction of the observations are grouped in a small range of directions, or phases. In particular, it is sensitive to (i) the distance between pairs of points on the circle, (ii) the grouping of several points in the same region, and (iii) the sensitivity to a very small distance between individual pairs of points should not be so great that an "accidentally" close pair dominates. It has been used in a search for ultra-high energy γ-ray from neutron star binary X-ray sources. The Rayleigh test (2.7) and tests using Kuiper's (2.8) and Watson's (2.9) statistics have very similar powers for broad pulses but Protheroe's new statistic (Eq. (2) in [Pro85]) is considerably more powerful than the others for narrow pulses.

Uesu et al. (2015) [USS15] proposed an asymmetric multivariate Möbius distribution on the unit hyper-disc, which has support in the interior of the circle, sphere, or more generally, hyper-sphere. They applied their distribution to study GRBs (for 247 observations from 17 December 2004 - 16 January 2012) and obtained a (first) statistical modeling of GRBs based on probability distributions that can lead to further insights into the formation of clusters within galaxies, as well as to the discovery, i.e., classification, of possibly the origin of a new galaxy or of a stars-forming region as hypothesized or postulated by astrophysics.

Crawford (2009) [Cra09] described a method for detecting a velocity dipole anisotropy in the sky distribution of sources in large-scale radio surveys. This detection can be used to constrain the magnitude and direction of our local motion with respect to an isotropically distributed extragalactic radio source population. He used their three one-dimensional Gaussian distributions in

Cartesian coordinates to determine the probability distribution of the magnitude of the final radial displacement. In his analysis, he considered the feasibility of detecting this dipole with statistical significance in existing and proposed future large-scale surveys. Note that his study corresponds to the most optimistic detection case possible. He found out that existing large-scale radio survey catalogs such as 87GB/PMN or NVSS do not have a sufficient number of sources to detect the expected velocity dipole with statistical significance.

Montin et al. (2023) [MBM23] advocated the use of a nonparametric approach, also known as modal clustering on the unit sphere, to detect the source of γ-ray photons, which are usually generated from accelerated charged particles (e.g., electrons or protons), boosted by extreme celestial objects like supermassive black holes or pulsars. More precisely, since the distance to the emitting source of γ-rays is not given, the sources will be identified using a von Mises-Fisher kernel estimate of the photon count density on the unit sphere via an adjustment of the mean-shift algorithm to account for the directional nature of data. Thus, using directional statistics tools, [MBM23] proposed an automatic way to skim off sound candidate sources from the γ-ray emitting diffuse background and to quantify their significance. As explained in their paper, this approach entails a number of desirable benefits: (i) it allows by-passing the difficulties on the border of any 2-dimensional projection of the photon directions, and (ii) it guarantees high flexibility and adaptability. In order to calibrate their algorithm, they analyzed a 5-year observation period of the simulated Fermi-LAT γ-ray photons database.

3.5 Galactic environment and stars

From our blue planet (on a clear night and void of light pollution), we can catch a glimpse of the bright lights of the galactic city on the night sky. Our window into the universe, this milky white band of stars (estimated about 100 thousand million stars), dust, and gas is where our galaxy, the Milky Way, gets its name from. It is a barred spiral galaxy (our Earth is located in one of its spiral arms, the Orion Arm) and includes our Solar system (containing the Sun, 8 planets, and a vast collection of rocks, ice, and dust). 27 000 lightyears from our Earth, at the core of our galaxy, lies the glutton of a supermassive black hole – Sagittarius A* (about 4 million times the mass of our Sun) – which we captured for the very first time in 2022 with help of the Event Horizon Telescope (EHT). Outside our Solar System are other planets, which we refer to as exoplanets or extrasolar planets (according to NASA, 5470 confirmed exoplanets in 4069 planetary systems have been discovered so far) and which orbit around other stars than our Sun; there are also other spiral, even elliptical and irregular, galaxies (estimated at roughly 200 billion) distributed not uniformly, but rather tending to congregate in groups or clusters, throughout our Universe. This naturally spurs the investigation of the distribution of the galaxies, exoplanets, and stars through our Universe by means of directional statistics tools, more precisely the rotation of the galaxies, the position and orientation of the stars in the celestial sphere, or the dispersions in the visibility phase of the black hole Sagittarius A* on a circle, to cite but these.

A possible large-scale asymmetry of the Universe. In 1983, Kendall and Young [KY84] verified Birch's hypothesis [Bir82] that there is a statistical asymmetry of our Universe, which could indicate that the universe is rotating. For this, they proposed an "indirectional" statistical approach since the position angle of an axis of plane symmetry is only defined up to multiples of π, rather than 2π. The elongation and the polarization of double radio source, that Birch examined, have both this character. More precisely, the discrepancy angle is the acute angle embraced by these two indirections in the tangent plane to the celestial sphere, and the unit vector locating the line of sight to the object is a direction in the ordinary sense. Their proposed test showed that the reported effect is strongly supported by the observation. Moreover, their tests estimated the pole of the asymmetry and the angular error.

3.5.1 Stars positions

The location on our Earth's surface that arises as the intersection with the imaginary line that joints the center of our Earth with a given star is referred to as the position of the star in the celestial sphere $\mathbb{S}^2$. In short, we are dealing with spherical data. According to Garcia-Portugués (2013) [GP13], for many years histograms that were adapted to the spherical case were the usual statistical tools to describe this kind of data. Nowadays more advanced and adapted statistical tools can be used to manipulate a star database as has been done by [GP13, BL15], and [Sei13].

Data of stars locations. The Bright Star Catalog, distributed from the Astronomical Data Center, consists of the locations of stars of a certain magnitude.

The Hipparcos mission carried out by the ESA in 1989-1993 contains a massive enumeration of near stars [P$^+$97]. An improved version of the original dataset is available from [VL07] which contains a corrected collection of the position of the stars on the celestial sphere as well as other star variables.

For the study of [BCDM22], they considered the available database HYG (Hipparcos-Yale-Gliese) (`https://www.astronexus.com/hyg`; the HYG 2.0 database contains around 11500 observations), the data introduced by [AL14], and the much bigger database GAIA [DDD20] (including 1.3 billion parallaxes (second release)), all three consisting of star-distance data from Earth.

Statistical tools and data analysis. Garcia-Portugués (2013) [GP13] illustrated his new bandwidth selectors for kernel density estimation with real spherical star observations (117955 star positions from the dataset of [VL07]) from the Hipparcos catalog. Such density estimators are a clear step beyond the abovementioned histograms. [GP13] estimated the spherical density by using the optimal smoothing of the bandwidth selector based on asymptotic and exact error expressions for the kernel density estimator combined with mixtures of vM distributions (2.2). For such a large dataset, [GP13] has shown that his technique is faster (256.01 seconds running time for the bandwidth selector, 247.34 seconds for the mixtures fitting, and 8.67 seconds for bandwidth optimization) than usual cross-validation techniques that demand an enormous amount of computing time and memory resources. [GP13] observed that the higher concentrations of stars are located around two spots, Orion's arm (left) and the Gould's Belt (right) of our Milky Way galaxy.

Budavári and Loredo (2015) [BL15] investigated an old astronomical problem of source identification in separate observations. In fact, stars and galaxies look different through the eyes of different instruments, and their independent measurements have to be carefully combined to provide a complete, sound picture of a multicolor and eventful universe. The association of an object's independent detections form multiple sources is a difficult scientific, computational as well as statistical problem. According to [BL15] the aim is to find records in survey databases with directions that match within the direction uncertainties by using cross-matching, cross-identification, and directional, positional, or spatio-temporal coincidence assessment. [BL15] surveyed recent developments (most of them based on directional statistics) and highlighted important open areas for future research in their paper (e.g., in Sect. 3.4 of [BL15]). They showed that a closed-form Bayes factor for directional data can be applied to stars with unknown motion [BL15, Sect. 3.3]. It turns out that stars pose a significant challenge for cross-identification methods due to their movement through space at different speeds to alter their apparent positions measurably with time.

Sei (2013) [Sei13] analyzed the Bright Star Catalog (5th Revised Ed.) data of the locations of 188 stars of magnitude than or equal to 3.0 by using a new quadratic model on the sphere defined in terms of cost-convex functions to estimate the parameters. He compared the quadratic model and the null model (uniform distribution) by using Akaike's Information Criterion.

Barabesi et al. (2022) [BCDM22] described the statement that a random sample of n observations is made of realizations of an absolutely continuous random variable X which is distributed according to Benford's law as *Benford hypothesis*. In fact, Benford's law was discovered by the astronomer Simon Newcomb in 1881 and independently later by the physicist Frank Benford. Since then it has become a much-studied probability distribution for significant digits and it has been applied in many real-world situations. Given a random sample $(X_1, \ldots, X_n)$ of n observations from X, the most straightforward translation of the Benford hypothesis into a null hypothesis to be tested is then

$$\mathcal{H}_0 : s(X) \stackrel{\mathcal{L}}{=} U,$$

where $s(X)$ refers to the (base-10) log-significant function defined as $s(x) = \langle \log_{10} |x| \rangle$ with assumption that $s(0) = 0$, and U is a uniform random variable on $[0, 1)$. Statistical assessment of the Benford hypothesis is important in many fields. For instance, in astrophysics, it has been recently shown that the Benford hypothesis has been used to provide empirical evidence that the distributions of the first digits of many physical quantities of celestial bodies are in good agreement with Benford's law (see Sect. 6 of [BCDM22] and references therein). In fact, by considering distances from Earth to stars, [HF16] explained theoretically (using astrophysical arguments and the mathematical properties of Benford's law) that empirical observation of such an argument may be viewed as new independent evidence of the validity of Hubble's law. Although according to [BCDM22], most of the applied astrophysical literature either relies on diagnostic checks of the data or performs a statistical evaluation of first-digit compliance, an accurate assessment of the Benford hypothesis would then be required. Therefore, [BCDM22] provided a principled framework for the statistical evaluation of the Benford hypothesis by first studying the probabilistic structure of many standard univariate models when they are framed in the space of the significant. Establishing thus the closeness of a particular model to the Benford hypothesis through a suitable Kolmogorov distance.

3.5.2 Pulsars

Although the name pulsar blends "pulse" and "star", a pulsar is not a pulsating star but is (from a pulsating radio source) a highly magnetized rotating neutron star that emits beams of electromagnetic radiation (analogous to a lighthouse) out of its magnetic poles. The radiation can only be observed when the beam of emission is pointing towards the distant observer every time a beam sweeps past them. For instance, its radiation and photons can be detected by either low-Earth-orbit or ground-based telescopes. The effect described before is known as the lighthouse effect and this gives rise to the pulsed nature that ultimately is responsible for the name of these stars – pulsars [Sch14].

A simplified pulsar model can be portrayed such that the magnetic field consists of open and closed field lines, with the outermost closed field lines touching the light cylinder and defining the polar cap region in the surface of the pulsar. The electromagnetic radiation from pulsars is observed in the form of two conical beams directed along the magnetic axis, believed to be due to the pulsar's spin and the magnetic axis being misaligned. [LK05] and [LGS06] have written excellent reference books about pulsars, and in Schutte's doctoral thesis [Sch14] one can find a good overview of the statistical developments of pulsars.

With the launch of several telescopes and satellites, it is possible to detect pulsating radio sources (e.g., by the Mullard Radio Astronomy Observatory), spot the Vela[2] and Crab[3] pulsar by historical evolution of supernovae and supernova remnants, identify pulsar wind nebulae from the γ-rays data, observe γ-ray pulsars (e.g., with the Energetic Gamma Ray Experiment Telescope), define important characteristics of pulsars such as its pulse (Crab pulsar) and off-pulse window (nebula), pulsar timing, etc. This gives rise to even more questions about the source, formation, and evolution of pulsars in our Universe. According to [Sch14] certain analyses used to be performed on histogram-type light curves with the so-called "eye-ball" technique or visual inspection of the data in order to identify the off-peak phase interval for example. However, some authors have recently shown that directional statistics can be used to investigate pulsars data and obtained interesting outcomes.

Pulsars data. The launch of NASA's Fermi Gamma-ray Space Telescope (formerly GLAST) in 2008 with its Large Area Telescope (LAT) on board contributed to the gathering of excellent data to study the high energy behaviour of many pulsars in great detail such as the Crab and Vela pulsars. It provides a database of γ-ray pulsars, which gives the possibility to study the properties of the off-pulsed emission of each pulsar, estimate light curves of pulsars to perform subsequent analyses on these curves in order to understand the pulsar magnetosphere and generate times of arrival of the photons, attempt to detect the potential emission associated with its pulsar wind nebulae, and much more (see [Sch14] and references therein). A pulsar timing package *Tempo2* is also available on the Fermi Science Support Center website (`https://fermi.gsfc.nasa.gov/ssc/data/access/lat/ephems`) which fits the time of arrivals to a timing model. The Fermi pulsar database has been used by [Sch14] and [SS16] for their studies.

The National Radio Astronomy Observatory (`https://public.nrao.edu/radio-astronomy/pulsars/`), which is a facility of the National Science Foundation operated under a cooperative agreement by Associated Universities, Inc., provides polarization observations data of single pulses from radio pulsars. These data were used for the analysis of [McK09] and [McK10].

Statistical tools and data analysis. Pulsar radio emission is renowned for its complicated polarization behavior, which has motivated McKinnon (2009) [McK09] to study polarization patterns in pulsar radio emission by presenting a generalized empirical model of pulsar polarization to accommodate a wide variety of polarization fluctuation geometries. When polarization observations of single pulses from radio pulsars are displayed in a two-dimensional projection of the Poincaré sphere, a variety of intriguing polarization patterns are created. Although in many pulsars, two clusters of data points that reside at antipodal points on $\mathbb{S}^2$ are formed by fluctuations in polarization amplitude that are parallel to the unit vectors, in other pulsars, the formation of the patterns resembling

[2] The Vela pulsar is the brightest non-transient source of γ-rays in the sky and due its brightness it is traditionally the first target object for any new γ-ray observatory.

[3] The Crab pulsar is associated with the supernovae explosion.

annuli and bow ties is more complex and largely unexplored. McKinnon's model [McK09] showed that these patterns arise from polarization fluctuations that are perpendicular to the mode vectors and that the modulation index of the polarization amplitude is an indicator of polarization pattern complexity.

One year later, McKinnon (2010) [McK10] presented a simple model for the polarization of pulsar radio emission used to derive the three-dimensional statistics of radio polarimetry. His model is based on the fact that the observed polarization is due to the incoherent superposition of two, highly polarized, orthogonal modes. Orthogonal modes are thought to be the natural modes of radio wave propagation in the pulsar magnetosphere and their intensities become statistically independent when the generalized Faraday rotation in the magnetosphere causes the difference in their phases to be large. The directional statistics derived from its model follow the Fisher family of distributions (2.5) as well as the Bingham-Mardia distribution with density

$$f_{\mathbf{X}}(\theta;\kappa,\rho) = \frac{1}{c(\kappa,\rho)}\frac{\sin(\theta)}{2\pi}\exp\{\kappa(\cos(\theta-\rho)^2)\}, \quad \theta\in[-\pi,\pi),$$

with rotational symmetry about the z-axis as mean direction and where κ and $\rho > 0$ are two shape parameters, and $c(\kappa,\rho)$ is the normalizing constant.

Lombard and Maxwell (2012) [LM12] proposed a non-parametric sequential cumulative sum, short cusum, procedure based on the Rayleigh test (2.7) to detect deviations from a uniform distribution on the circle and applied it to some data consisting of arrival times of cosmic rays from the vicinity of a pulsar. The objective is to detect periods of sustained high-energy radiation. Following a standard procedure in astrophysics, the data are wrapped around a circle of circumference equal to the period of the pulsar. If no high-energy radiation is present, the wrapped data should be more or less uniformly distributed on the circumference of the circle, while a non-uniform distribution should manifest itself during periods of high-energy radiation. The idea behind their method is to approximate the distribution of the partial sum of a sequence of standard normal random variables by the distribution of the partial sums of a sequence of transformations of the Rayleigh statistic (2.7), leading to a standard normal cusum procedure that can be applied to the transformed data. By applying their method to real pulsar data, the Rayleigh test based on (2.7) assigned a p-value of 0.96 when applied to the first 192 observations and a p-value of 0.05 when applied to observations 193 through 303. This suggests that the data from 193 onwards are non-uniform and that the data of the first 192 observations confirm that the signals from the lower cusum are indeed a false positive (when the offset is 0, see [LM12, Eq. (7)]). Moreover, for these pulsar data, they obtained a kernel density estimate

$$\hat{f}(\theta) = \frac{1}{n\mathcal{I}_0(\rho)}\sum_{j=193}^{304}\exp(\rho\cos(\theta-\theta_j)),$$

where $\theta\in(\pi/2, 5\pi/2]$, $n\in\mathbb{N}$ is the number of samples SIZE and ρ the concentration parameter. This kernel density estimate has been constructed using a kernel vM density (2.2) with concentration parameter $\rho = 3.5$ and its apparent bimodality suggests that the pulsar in question may be a binary object, that is, two bodies in relatively close proximity rotating about a common axis. [LM12] also observed that the appropriate decision constant, determined by simulations, is somewhat smaller than the value suggested by the normal approximation theory and that the multiple cusum (where the cusum procedures is directly applied to two sequences of partial sums of cosines and sines of the observed data) also signals at $n = 303$. For pulsar data that are rotated through an angle of $\pi/4$, i.e., there is no signal at all , their non-parametric cusum is unaffected by any such rotation. Hence their cusum procedure is rotation invariant in the sense that the outcome of the analysis should not depend upon which point on the circle is chosen as the origin of angular measurement. However, as pointed out by [LM12], a limitation of their proposed procedure, shared with the fixed sample Rayleigh test (2.7), is that it is not useful when the alternative to uniformity is a distribution that produces a zero expected resultant length.

Later, Lombard et al. (2020) [LHP20] developed a non-parametric rotation invariant circular cusum for detecting changes in the mean direction and concentration of a circular distribution for pulsar data. Their methods were designed for situations in which the initial mean direction and concentration are unspecified. Hence the objective is to detect a change from the initial values, whatever the latter may be. They applied their concentration cusums to observation 191 through 1250 pulsar data and obtained a uniform distribution in the segment $[523, 1250]$ and non-uniformity in the segment $[191, 522]$.

Schutte and Swanepoel (2014) [Sch14, SS16] developed a non-parametric sequential estimation technique, which they referred to as SOPIE (Sequential Off-Pulse Interval Estimation), for the off-pulse interval(s) of a source function originating from a pulsar. Their estimation technique SOPIE is available on R [Sch22]. In astrophysics, in order to attempt to detect potential emissions from the associated pulsar wind nebula, it is important to study the properties of the off-pulse emissions by identifying the off-pulse interval of each pulsar accurately. In the literature, the common identification technique is based on the visual inspection of the histogram estimate of the pulsar light curve. The "eye-ball" (visual) technique and Bayesian Block method are the most used ones. SOPIE is a new statistical procedure to estimate the off-pulse interval in the presence of noise. It is based on a sequential application of p-values obtained from goodness-of-fit tests for uniformity such as the Kolmogorov-Smirnov, CvM, AD and Rayleigh tests (see Sect. 2.2). They applied their model to two real-life pulsar datasets, the PSR J1709-4429 pulsar (energy > 0.1 GeV, containing $n = 21153$ photon arrival times) and the well-known Crab pulsar PSR J0534+220 (energy > 0.1 GeV, $n = 21145$), and obtained comparability estimated interval(s) compared to the "eye-ball" (visual) technique and Bayesian Block method. In [Sch14], the SOPIE model was applied to other pulsars datasets, namely PSR J0835-4510 (Vela), PSR J2021+3651, PSR J1057-5226, and PSR J0034-0534 [Sch14, Sect. 5].

3.5.3 Binary stars

Binary stars or binary star systems consist of a pair of stars that orbit about their common center of mass. So far only 65 thousand binaries are known and it is thought that at least 50% of stars are binaries. Our Sun might once have had a binary companion star [SL20]. There exist different classes of binary stars, the important one is the visual binaries in which the two stars can be distinguished from one another when seen through an optical telescope. According to [Jup95], the directional aspects of the observations on visual binary stars are its position on the celestial sphere and the direction $\boldsymbol{z}$ of the axis about which the two stars rotate (with a sign specified by the sense of rotation). One may consequently address the following questions:

(i) Are the directions $\boldsymbol{z}$ uniformly distributed on the celestial sphere? (see [Gil88] and [Jup95])

(ii) What is the preferred orbital direction for binary stars? (see [Gil88])

(iii) Do current theories of binary star formation predict alignment of orbital poles for all binaries resulting from the same event of star formation? (see [ABJ+15])

(iv) Assuming that the answer to (iii) is positive: Is this coherence in the pole orientation of young binary systems preserved by the Galactic kinematics in a volume-limited sample (the solar neighborhood) consisting of much older systems? (see [ABJ+15])

Database. There is an old dataset by [WH83] of orbits of visual binary stars. More precisely, the orbits included several categories of binary stars and all orbits have a reliability index of 4 or better. The first category contains all stars for which the ascending node has been determined from radial velocity (amounts to 61 pairs). Since a pole to a fully characterized orbit is defined uniquely by using a right-hand rule, the dataset consists of unit vectors. This dataset has been used for the study of [Gil88] and [Jup95] (the latter used a suitable data set of 170 orbits of grade 1 or 2).

A more updated database of a visual or an astronomic orbit is available in the 6th Catalog of Orbit of Visual Binary Stars (COVBS) (`http://ad.usno.navy.mil/wds/orb6.html`) at USNO (2001) [HMW01] and of HIPPARCOS parallax (2007) [PLK+97, VL07]. Note that [ABJ+15] (see their Table 1 and Table 2) selected 95 systems in the master sample based on these two datasets. In addition, they made use also of other catalogs such as the Simbad database at the Centre de Données Astronomiques de Strasbourg [G+00] for the spectral type, the 9th Catalogue of Spectroscopic Binary Orbits (`http://sb9.astro.ulb.ac.be/intro.html`) [Pou00], the CORAVEL spectro velocimeter [BMP79] and the HERMES spectrograph installed on the Mercator telescope [R+11] for the radial-velocity measurements. For more information see [ABJ+15, Sect. 3].

As pointed out by [ABJ+15], [R+10, Table 11] collected orbits for visual binaries with HIPPARCOS parallaxes larger than 40 mas and with primary spectral type F6-K3, and [K+12] compiled an exhaustive list of stars nearer than 8 ps from the Sun.

Statistical tools and data analysis. Gillett (1988) [Gil88] gave a statistical demonstration (a more rigorous analysis than [Ber83]) that the orbital planes of visual binary stars are randomly oriented. For this, he used directional statistics. More precisely, Giné's nonparametric test statistics (see Sect. 2.2) yielded that the axes to the orbits of 102 visual binary stars, for which the ambiguity of inclination can be resolved to $\pm 180°$, are randomly distributed at $> 95\%$ confidence and that their distribution is also random against the alternative of Bingham distribution (see special case under the Eq. (2.5)). Moreover, by using Beran's test (see Sect. 2.2) and the alternatives of a Bingham and a Fisher distribution (2.5), binaries for which the true ascending node can be determined are also randomly oriented.

Jupp (1995) [Jup95, Sect. 3] studied if the directions of the axis about which the two stars rotate (with a sign specified by the sense of rotation) are uniformly distributed on the celestial sphere. He observed that testing the hypothesis of uniformity is complicated by (i) the *mirror-image ambiguity* of many visual stars binaries, i.e., for these binaries, it is not possible to distinguish between the direction $\boldsymbol{z}$ and its reflection $(2uu^T - I_3)\boldsymbol{z}$ in the line of sight along its position u; and (ii) possible selection effects, with orbits having $\boldsymbol{z}$ close to $\pm u$ being more likely to be observed than other orbits. Therefore since $\boldsymbol{z}$ is subject to mirror-image ambiguity, the Fisher distribution (2.5) cannot be used directly. However, according to [Jup95], it is appropriate to consider the conditional distribution given u of the observable parts $U\boldsymbol{z}$ and $\pm(I_3 - U)\boldsymbol{z}$ of $\boldsymbol{z}$, where $U = uu^T$ denotes the orthogonal projection onto the line of sight. In fact, the distribution of observable u is influenced by the positions of observatories on the Earth's surface, and both $U\boldsymbol{z}$ and $\pm(I_3 - U)\boldsymbol{z}$ depend on u. In order to allow for possible selection effects, [Jup95] considered the Watson distribution (2.6) with β as a parameter measuring the strength of the selection effect, i.e. for $\beta = 0$ there is no selection, for $\beta > 0$ we tend to observe those $\boldsymbol{z}$ which are parallel to $\pm u$, and for $\beta < 0$ we tend to observe those $\boldsymbol{z}$ which are normal to u. Combining the conditional Fisher distribution (2.5) and the Watson densities (2.6) yields a model that can be regarded as a directional regression model in the spirit of that used by [KY84] to analyze a problem about a possible large-scale asymmetry of the Universe (for further details, see the introduction of Sect. 3.5). [Jup95] obtained no evidence of non-uniformity, which is in agreement with the analysis of [Ber83] and [Gil88] when restricted to orbits not subject to mirror-image ambiguity. Concerning possible selection effects, he concluded that there is a higher probability of selection for orbits with $\boldsymbol{z}$ parallel to $\pm u$, this means with orbits having inclination nearer to 0° or 180° are more likely to be selected. The latter agrees with the results obtained informally by [PL67]. Moreover, [Jup95] found out that the Watson model fits the data well by means of the Kolmogorov-Smirnov test, which is not significant at the 5% level.

Agati et al. (2015) [ABJ+15] tested if the orbital poles of binary stars in the solar neighbourhood are isotropically distributed on the celestial sphere by using spherical statistics, more precisely the Rayleigh and Beran tests. The problem is plagued by ambiguity on the position of the ascending node, which led them to first analyze the binary stars data and filter among the 95 systems with an orbit in the 6th Catalogue of Orbits of Visual Binaries 51 systems thanks to the radial-velocity data (obtained from the CORAVEL database). After ordering the binary systems by increasing distance from the Sun, the false-alarm probability of rejecting the null hypothesis of isotropy while it is true was computed for subsamples of increasing size, from $n = 1$ up to the full sample of 51 systems. By using a jack-knife approach to remove one system at a time from the full sample, Rayleigh's and Beran's tests delivered a false-alarm probability between 1.5% and 0.1% (depending on which system is removed from the sample). This means that no uncertain conclusion shall be taken. [ABJ+15] fostered to do further studies on this problem, especially in the Gaia era.

3.5.4 Supernovae

A supernova is the set of phenomena resulting from the implosion of a star at the end of its life, in particular a gigantic explosion that is accompanied by a brief but fantastically large increase in its luminosity. Seen from Earth, a supernova therefore often appears as a new star, whereas in reality, it corresponds to the disappearance of a star. Observing a galactic supernova is very difficult, but locating one in the sky by the neutrinos alone is more reasonable. One of the main reasons is that the MeV-neutrino burst precedes the optical explosion by several hours so that an early warning can be issued to the astronomical community and hence they can specify the direction to look for the explosion. Therefore the optical signal from a supernova, if observed, can give the most accurate determination of its position in the sky. These observations are detected by optical telescopes, or by ground-by-air Cherenkov telescopes (permitting to measure multi-GeV to TeV photons associated with the accelerated protons in the shock after the supernova environment becomes transparent to high energy photons), or a suitable x- or γ-ray satellite (like ESA's INTEGRAL (International Gamma-Ray Astrophysics Laboratory) satellite that could resolve the supernova with an angular resolution of 12 ar-minutes, that means that the supernova would be visible in these wavebands starting from the optical explosion). However, it is not always possible that the supernova is seen in the entire electromagnetic spectrum (e.g., due to not suitable x- or γ-ray satellite operating, the air Cherenkov telescopes are blinded by daylight, or the satellites and telescopes simply do not look in the right direction at the right time). Therefore the best way to locate a supernova by its core-collapse neutrinos is through the directionality of elastic scattering in a water Cherenkov detector such as Super-Kamiokande [BV99, AS02]. This naturally calls to use directional statistics tools to extract information from these "directional data".

Database. The data for the supernova SN1987A in the Large Magellanic Cloud [H+88] can be found in the Kamiokande-II experiment group (https://www-sk.icrr.u-tokyo.ac.jp/en/sk/about/detector/).

There are some catalogs and databases where supernovae datasets can be found: the Lyon-Meudon Extragalactic Database (http://www-obs.univ-lyon1.fr/hypercat/), the NASA/IPAC Extragalactic Database (http://nedwww.ipac.caltech.edu), the Sloan Digital Sky Survey (http://www.sdss.org/), the IAU Central Bureau for Astronomical Telegrams (http://cfa-www.harvard.edu/iau/cbat.html), the CfA List of Supernovae (http://cfa-www.harvard.edu/cfa/ps/lists/Supernovae.html), the Asiago Supernova Catalogue (https://heasarc.gsfc.nasa.gov/W3Browse/all/asiagosn.html), the Sternberg Astronomical Institute Supernova Catalogue (http://www.sai.msu.su/sn/sncat/), and the Astronomy Section Rochester Academy of Science (https://www.rochesterastronomy.org/supernova.html).

Concerning the detection of high energy neutrinos, the IceCube at the South Pole [K+03] or nothern projects like the Gigation Water Detector at Baikal [D+01] and Nemo in the Mediterranean [C+02] provided such data.

Statistical tools and data analysis. Konishi et al. (1993) [KKC+93] tested if there is a possible spherical distribution of the electrons from neutrinos emitted from the supernova SN 1987A, which is in the large magellanic cloud. Many investigators have claimed, using various statistical methods on the basis of the polar angle, that the IMB data seem to have a nonisotropic distribution, whereas the Kamiokande data show an isotropic distribution. However, they do not seem to have succeeded in testing spherical distributions using both the polar and azimuthal angles. In most of the cases, they only consider the polar angle while keeping the azimuthal angle uniform. [KKC+93] used a test for the mean direction of the spherical data and concluded that both angular distributions observed by the IMB and Kamiokande groups do not contradict the assumption that all events are due to a neutrinos capture process in accordance with the standard model of a supernova explosion.

Another holy grail of low-energy neutrino astronomy is the observation of a galactic supernova. An important question is how well can one locate the galactic supernova in the sky by the neutrinos alone. Tomàs et al. (2003) [TSR+03] studied this question by analyzing the performance of different methods for extracting the galactic supernova direction and identified a simple approach that is nearly optimal, yet independent of the exact galactic supernova neutrino spectra. They used a two-dimensional Gaussian distribution on a sphere as a toy model to extract information from directional data in order to explore some parameter-free methods that only use the data and exploit the symmetries inherent in the physical situation.

3.5.5 Interstellar medium

The interstellar medium (ISM) contains primordial leftovers from the formation of galaxies, detritus from stars, and the raw ingredients for future planets and stars. Studying ISM is primordial to better understand the structure of the galaxy and the life cycle of stars. Astronomers discovered that magnetic fields are a major agent in the ISM of spiral, barred, irregular, and dwarf galaxies (see, e.g., [Cru12] and [HH12]). Recently in astrophysics, a particular interest has been given to examining the relationship between the orientations of the magnetic field projected in the plane of the sky and the morphology of the structures in the total column density of the ISM. This can be translated by the interest of characterizing the relative orientations between two sets of overlapping directional data, meaning to determine whether or not the two sets of directions are preferentially parallel or perpendicular to each other. This leads to the need for robust statistical modeling of orientation effects.

For instance, in 2012, the ESA's space satellite *Herschel* viewed and collected important data of the Vela-C molecular cloud, which is a giant molecular cloud and one of four subregions of the Vela Molecular Ridge – a vast star-forming complex in the plane of our Milky Way galaxy. In this cloud, stars are being born and it is considered to be a rare example of a nearby and massive cloud in the early evolutionary stage [BNN+04, N+09, S+17]. Coming back to our problem, studying the overlap of the Balloon-borne Large Aperture Submillimeter Telescope for Polarimetry (BLASTPol) and the columns density maps density in the regions of the Vela C cloud is of great interest.

Database. For the Vela C molecular cloud data set, the BLASTPol provided, during its flight in 2012, the linearly polarized emission (stokes I, Q, and U) observations [G+14] and the ESA's

Herschel satellite dust-continuum observations its total column density maps [P+10].

Other public column density maps of molecular clouds are available in Soler's GitHub (`http://github.com/solerjuan/magnetar`) (see also [Sol19] and references therein).

Statistical tools and data analysis. Motivated by the astrophysical problem of characterizing the relative alignment of the magnetic field projected on the plane of the sky and the orientations of structures in the ISM, Jow et al. (2018) [JHS+18] presented a robust and efficient method based on the projected Rayleigh statistics for determining whether two pseudo-vector fields tend to align, either preferentially perpendicular or preferentially parallel. The projected Rayleigh statistic (PRS) is a modification of the classical Rayleigh statistic (2.7) and thus is a test for non-uniform and unimodal clustering of the angle distribution. For a set of n relative orientation vectors θ_i, the weighted PRS is defined as

$$V := \frac{2}{\sum_{i=1}^{n} w_i} \left[\sum_{i=1}^{n} \cos(2(\theta_i - \theta_0)) \right]^2, \tag{3.3}$$

where the w_i correspond to the statistical weights assigned to each angle θ_i. [JHS+18] compared the efficiency of the PRS against histogram binning methods such as the histogram of relative orientations (HRO) method by [SHM+13, S+17] which has been used for characterizing the relative orientations of gas column density structures with the magnetic field projected in the plane of the sky. The HRO method is a technique for analyzing the relative orientations of interstellar magnetic fields and emission morphologies in molecular clouds. It is an obvious way to investigate the underlying distribution of a sample of angles by plotting a histogram of that sample. For this study, [JHS+18] examined the data for the Vela C molecular cloud from Herschel submillimetre observations and the magnetic field from observations by the BLASTPol. They found out that (i) the PRS is statistically more powerful than and loses sensitivity slower than the HRO as the distribution of angles approaches uniformity, (ii) the sensitivity of the PRS also improves faster with larger sample sizes as the HRO, (iii) for various subregions of the Vela C molecular complex, the PRS reports a higher statistical significance for the correlations. [JHS+18] showed that their method has the potential for a broad range of applications in astronomy.

By taking into consideration the outcomes of [JHS+18], Soler (2019) [Sol19] presented an update of their HRO method by making use of PRS for circular data. They studied the relative orientation between the magnetic field projected on the plane of sky $B_\perp$ inferred from the polarization thermal emission of Galactic dust (observed by Planck) and the distribution of gas column density N_H (derived from the observations by Herschel). In (3.3), if $|V|$ is small, there is no evidence that N_H and $B_\perp$ are oriented towards the reference angles 0 or 90°, while for $|V|$ relatively large, there must be some concentration around the reference angles. By testing their improved HRO to dark (molecular) stars clouds and nebulae such as Taurus, Ophiuchus, Lupus, Chamaeleon-Musca, Corona Australis, Aquila Rift, Perseus, IC5146, Cepheus, and Orion, they found that (i) the mean relative orientation between N_H and $B_\perp$ towards these regions increases progressively from 0°, where the N_H structures lie mostly parallel or perpendicular to $B_\perp$, (ii) the slopes of the N_H pdf tail are steepest in the regions where N_H and $B_\perp$ are close to perpendicular, (iii) no evident correlation exists between the star formation rates, estimated from the counts of young stellar objects, and the relative orientation between N_H and $B_\perp$ in these regions.

3.5.6 Star Sgr A*

The EHT imaged for the first time black holes at event horizon scales such as the Sagittarius A* (Sgr A*) which was one of its primary observing targets. Sgr A* is a spinning black hole inside our Milky Way's heart. In addition, closure phases along different baseline triangles carry a large amount of information regarding the structures of the images of black holes in interferometric observations with EHT. From a directional statistics point of view, closure phases are an example of circular statistics, as they can be described as points on the unit circle characterized by an angle.

Database. General relativistic magnetohydrodynamic (GRMHD) simulations have successfully recovered observational parameters of black holes such as the Sgr A* and M87. Medeiros et al. (2017) [MCÖ+17] as well as Roelofs et al. (2017) [RJS+17] used these GRMHD simulations (e.g., b0-high from [Shi13]) as basis for their study and adapted them for Sgr A*.

Statistical tools and data analysis. [MCÖ+17] computed the dispersions in the visibility phase of Sgr A* by taking into account the fact that angles are directional quantities. More precisely, they used five different GRMHD models (denoted by Model A-E), which permitted them to simulate the Sgr A*, to (i) present the structure of the complex visibilities (where the average visibility phases are shown in contours and the visibility amplitudes in color maps), and (ii) explore the structure of the variability in the visibility phases (where the dispersion in visibility phase is shown in color and the average visibility phase in white contours for comparison) of the black hole [MCÖ+17, Fig. 3-4]. They obtained these averages by finding the phases and amplitudes of each snapshot and subsequently averaging them. Since the angles are directional (periodic) quantities, they made use of circular statistics to compute (i) the mean of the unit vectors that correspond to the distribution of angles, and (ii) the dispersion of a distribution of angles.

[RJS+17] did a similar study as [MCÖ+17] by quantifying intrinsic closure-phase variability in a measured closure-phase track containing variations from image and observational variability. The constructed metric is based on circular statistics and represents the fraction of variability in a given closure-phase track that is not due to thermal noise. In fact, closure phases are an example of circular data. This brought [RJS+17] to employ classical circular techniques [MJ00] to estimate the spread on a circular data set.

3.5.7 Dark matter: WIMP direct detection

It turns out that roughly 95% of our Universe is made of a mysterious, invisible substance that no one has ever seen. From this 95%, 68% is dark energy, since we know how it affects the universe's expansion, the other 27% makes up the dark matter. The remaining 5% of the universe is made of normal matter that has been observed with our instruments. Detecting dark matter is very difficult since it does not appear to interact with electromagnetic fields, i.e., it does not absorb, reflect, or emit electromagnetic radiation. That is also why it is referred to as "dark" matter.

Weakly Interacting Massive Particles (WIMPs) are hypothetical particles that are one of the proposed candidates for dark matter. Via the elastic scattering of WIMPs on detector nuclei it is possible to directly detect non-baryonic dark matter. In fact, these direct detection experiments are designed to measure the collision of WIMPS with nuclei in the laboratory. In addition, they are presently reaching the sensitivity required to detect neutralinos, which are the lightest supersymmetric particles and an excellent WIMP candidate. In the search for WIMP with a directional detector, one can consider three simple hypotheses to test:

(i) Is the recoil direction distribution uniform? In case of rejection, we would detect a WIMP signal.

(ii) Is the recoil distribution rotationally symmetric about the direction of solar motion? In case of rejection, we would detect the flattening of the Milky Way halo.

(iii) Does the mean direction deviate from the direction of solar motion due to a tidal stream?

Investigating these questions can be done for a range of observationally motivated halo models (taking into account the detector response) by using directional statistical tests.

Database. Data from direct detection experiments can be found in [MG05b, MGS05] and [OG14]. For instance, the data in [OG14] consists of 304 signal events in the halo-only case and 316 in the halo+stream case. O'Hara and Green (2014) [OG14] have binned the data on a sphere using a HEALPix [GHB+05] equal angular area discretization.

Statistical tools and data analysis. Morgan and Green (2005) [MG05b] investigated the detection of a WIMP signal from possible backgrounds due to, for instance, neutrons from cosmic-ray induced muons or natural radioactivity by using a variety of (non-parametric) spherical statistical tests. In fact, WIMPs are hypothetical, new elementary particles that have been postulated to solve the cosmological problem of dark matter in space. Due to the very small expected event rates, it is crucial to distinguish a putative WIMP signal from an isotropic background when the uncertainty in the reconstruction of the nuclear recoil direction is included in the calculation of the expected signal. These recoil directions constitute vectors, or, if the senses are unknown, undirected lines or axes, and so can equivalently be represented as points on a sphere. [MG05b] studied the number of events required (i) to reject isotropy and hence detect a WIMP signal by using the modified Rayleigh, modified Bingham, Beran, and Giné statistics (see Sect. 2.2), (ii) to reject rotational symmetry (see Definition 1) and detect flattening of the Milky Way halo (which can be modeled as an isotropic sphere with a certain distribution) for a range of observationally motivated halo models, taking into account the detector response, by considering the modified Kuiper statistics, and (iii) to detect a deviation in the mean direction from the direction of solar motion due to a tidal stream. The advantage of these spherical tests, unlike likelihood analysis, is that they do not require any assumptions about the form of the local WIMP velocity distribution (even if the properties such as the shape, velocity anisotropy, and density profile of the Milky Way halo at the solar radius were accurately determined there would likely be a wide range of local velocity distributions consistent with these properties). Furthermore, [MG05b] also proposed a WIMP search strategy with a directional detector that can be divided into three regimes: (1) the simple search phase (aiming to detect a non-zero signal, i.e. an anomalous recoil signal above what is expected from backgrounds), (2) the confirmation stage (at this point the experiment would collect more data and aim to confirm the recoil signal as Galactic in origin by searching for the expected anisotropy in the recoil directions; in statistical terms one might ask if the distribution of observed recoil directions is isotropic), and (3) the exploitation phase (extraction of information about the form of the recoil anisotropy and hence the location WIMP velocity distribution, flattening of the halo or the presence of a tidal stream). At the end of their analysis, they found out that if the senses (the signs) of the recoils are known then between 10 and 30 events, depending on the read-out plane, will be sufficient to distinguish a WIMP signal from an isotropic background for all of the considered halo models, with the uncertainties in reconstructing the recoil direction only mildly increasing the required number of events. On the contrary, if the senses are unknown then these numbers are increased by roughly an order of magnitude.

Morgan et al. (2005) [MGS05] is a continuation of [MG05b]. They repeated the same analysis as in [MG05b] but for a detector with a less complex 2-d read-out by using non-parametric circular statistics. More precisely, the recoil directions projected onto the read-out plane constitute 2-d vectors, or, if the senses are not known, undirected lines or axes, and so can equivalently be represented as points on a circle, parametrized by their angle relative to some fixed point/direction. For their analysis, they applied the Rayleigh (2.7), Kuiper (2.8), and Watson (2.9) statistics to test isotropy, the Wilcoxon signed-rank statistic to test reflective symmetry, and the Watson mean direction test. They observed that the potential for detecting a WIMP signal (via its anisotropy) with a 2-d read-out detector is similar to that for a detector with a full 3-d read-out, provided that the reduced angle distribution is utilized and that the sense of the recoils can be measured.

Some time later, O'Hara and Green (2014) [OG14] presented non-parametric directional statistics tests based on the modified Kuiper test and a profile likelihood test to determine the number of events required to distinguish a WIMP signal from isotropic backgrounds and thus detect a dark matter stream. In particular, they tested if there is the presence of a distribution without a stream against when a stream is present in the data. They considered the direction of the nuclear recoils, which consist of 304 signal events in the halo-only case and 316 in the halo+stream case, and binned them on a sphere using a HEALPix equal angular discretization. As a concrete example, they performed their tests by using Bayesian parameter estimation on data containing recoils from a Sagittarius-like tidal stream with varying density fractions and noted that the observed distributions of the test statistics become further separated from the null distributions when the density fraction increases. They showed that (i) the detectability of a stream is highly dependent on its speed and position in relation to the lab velocity and (ii) if a moderately high-density WIMP stream is present in the Solar neighbourhood then there are good prospects for its detection using directional dark matter detectors.

3.5.8 Galaxies dynamics and clustering

One of the pillars of cosmology is that our Universe is isotropic, meaning the same in all directions. However, some discoveries showed that there may be cracks in that pillar. For instance, since 1980, possible signatures of a comic anisotropy began to appear. At that time, in 1982, Birch [Bir82] found an apparent anisotropy in the relation between the polarization vectors of the magnetic field of the radiation received from galaxies and their respective position angles. Birch reinforced the hypothesis from Gamow that the whole universe is rotating as a possible explanation for the galaxies' rotation. Later other scientists obtained similar results to Birch's and found further evidence of anisotropy in observations of supernovae, fine structure of quasars, and the CMB dipole, quadrupole and octopole to name a few (for more information see, e.g., [MPC17] and references therein). All this potential evidence motivates the search of new observational anisotropy signatures, analyses of the consequences and to conquer the galaxy dynamics in general (in the sense of distribution, rotations, clusters, etc.). For instance, it is of high interest to

(i) look for anisotropies in galaxies' distributions, more precisely in the distributions of their position angles. The position angles distribution will not be uniform if there is a cosmic anistropy related to a cosmic relation.

(ii) check if there is any anisotropy in the propagation of radio polarizations from cosmologically distant galaxies.

(ii) extract rotation measures from spectral polarization data in order to provide useful information about the magnetic fields in our Galaxy and in the host galaxy.

(iv) study the scale and dispersion of galactic alignments by modeling the random tidal torques of spiral rotating galaxies. The angular moment is conserved as quantity (so long as only gravitational forces act) and its distribution amongst the galaxies should give interesting information about their origin and history.

(v) test if there is any statistically significant alignment observed between the brightest galaxy and its parent cluster or with some other relevant position angle (such as that of the first brightest galaxy or with the line of separations joining the two). Detection of alignment may offer support for either the pancake scenario of galaxy formation or the tidal interaction of member galaxies with the cluster's potential in the hierarchical scenario.

Database. Coutts (1996) [Cou96] already remarked in 1995 that the size of the samples of spiral galaxies has varied from 20 [HS82] to 30760 [MD79], and the total number of galaxies examined in published studies is now in excess of millions.

For the brightest binary galaxies, the data is obtained from Palomar Observatory Sky Survey (POSS) (`https://skyserver.sdss.org/dr5/en/proj/advanced/skysurveys/poss.asp`) prints on original Kodak paper (e.g., [Str88]). There are two different measurements: the red prints (POSS E) corresponding to binary components and cluster galaxies which appear brightest on them, and blue prints (POSS O) representing pairs whose members appear close together or nearly overlapping on E prints. Among the POSS O prints, intrinsically red (early-type) galaxy images appear smaller and thus render them more easily resolvable for position-angle determination. In the study of [TCF92], a sample of 55 clusters of galaxies obtained by machine scans of this Palomar 48-inch Oschin Schmidt plates (a telescope at Mount Palomar in southern California carried out in the 1905's the first POSS) were used as angular data. In [Str88, Str90] one can find the position-angle data of 405 Abell clusters (Table I in [Str88]) and of 68 clusters of the Rood-Sastry B type (Table I in [Str90]). These data consist of position-angles of the major axis of each member of the brightest binary members, the separation vector joining their centers, and the long axis of the galaxy distribution of the parent cluster.

On NASA's HEASARC portal, the user can find the Morphological catalog (`https://heasarc.gsfc.nasa.gov/W3Browse/galaxy-catalog/mcg.html`) and the ABELL database (`https://heasarc.gsfc.nasa.gov/W3Browse/all/abell.html`). The Morphological catalog is a database of 30642 galaxies (of magnitude 15 or brighter) assembled in the 1960s-1970s, from inspection of the plates of the POSS, carried out in the 1950s. The ABELL database contains information from catalogs of clusters of galaxies. Each cluster of galaxies has at least 30 members within the magnitude M3 (which is the third brightest cluster member (BCM)) to M3+2 and each with a nominal redshift less than 0.2.

The publicly available HyperLEDA catalog (`https://leda.univ-lyon1.fr/intro.html`) contains a database and a collection of tools to study the physics of galaxies and cosmology. For instance, it provides a series of bright galaxy catalogs (RC1, RC2, RC3). For the study of [MPC17], the HyperLEDA catalog [P$^+$03] with 833844 galaxies was used for their simulations.

For the polarization database, the catalog of [TI80] contains position angles of polarization and was used in the study of [SJ01]. For the study of [JR99], they collected different galaxies databases from [CFJ90], [ER79, ER80] (which are available on the NASA-ADC archives), [SNKB] containing polarization angles, [Bie84, Bie84], and [Bir82].

The Nearby Galaxies catalog from [Tul88] provides a collection of observational data on galaxies (with systemic velocities less than 3000 km/s) from Tully and Fisher [FT81] and Shapley-Ames samples.

Concerning the dataset for atomic hydrogen emission, there are publicly available observations of the emission by atomic hydrogen at 21 cm wavelength in the HI 4π (HI4PI) survey [B$^+$16]. This survey is based on data from the Effelberg-Bonn Hi survey [K$^+$11a] and the Galactic All-Sky Survey [MG$^+$05a, K$^+$05]. This data has been used for the study of [S$^+$22].

Statistical tools and data analysis. Struble (1988) [Str88] analyzed the position-angle data of the brightest binary galaxies for 405 Abell clusters (obtained from POSS prints). He examined the alignment of the second brightest galaxy with a parent cluster for the specific case of the components of the brightest binary galaxies in Abell clusters. In particular, he applied the Rayleigh test (2.7) to test the uniformity of axial directions for the position-angle of the major axis of the first (ϕ_1)

and second (ϕ_2) brightest component. In fact, the orientation of each component of a binary galaxy, the separation vector, and the long axis of the parent cluster are unrelated to Earth's rotation axis, so the distributions of position-angles on the sky are useful in detecting systematic errors in measurements. The Rayleigh test (2.7) showed that the ϕ_1 distribution deviated from uniformity at the 1% significance level and the ϕ_2 distribution at the 2% significance level. This led [Str88] to expect that these two position-angles show some correlation. [Str88] evaluated the statistical significance of circular correlation coefficients. In contrast to the results for isolated field binaries, he found significant correlations between position-angles of both components of the binary and their line of separation. These strong alignments were attributed to both radial and circumferential orientations of the major axis of the fainter component with respect to the brighter present among the binaries. Compared to previous studies, the position-angles of the brightest component showed a weaker alignment with the long axis of the galaxy distribution of the parent cluster, but the position-angles of the second brightest component showed no correlations with the cluster position-angle.

Later in 1990, Strublle [Str90] did a similar statistical study as in [Str88] by considering a sample of 68 clusters of the Rood-Sastry B type (like Coma, containing two luminous galaxies brighter than the remaining members). He found that (i) the major axis of the brightest galaxy is aligned not only with the long axis of the parent cluster (as found in the previous studies), but also with the position-angle of the line of separations between the first and second brightest galaxy, and that (ii) the second brightest galaxy is aligned with the long axis of the parent cluster and weakly aligned with the first.

Tucker (1988) [TP88] investigated the alignment of clusters with brightest member galaxies, i.e., if the major axis of a BCM galaxy tends to be aligned with the major axis of the parent cluster. In their study, they used the Kuiper test (2.8) for angular data to test nonuniformity in the distribution of φ_{CG}, the acute angle made by the galaxy and cluster. For their dataset of 32 BCM galaxies in Abell clusters, they found that (i) the alignment of BCM galaxies with clusters is not significant, (ii) there is no significant trend of φ_{CG} with isophote size, and (iii) significant isophote rotation in individual galaxies is uncommon.

Trèvese et al. (1992) [TCF92] studied the orientation of clusters of each galaxy, and the direction of the line joining each galaxy with the cluster center or with the brightest galaxy, as well as all possible relative angles between these directions. For this, they did a statistical analysis by applying various statistical tests for angular data of a sample of 55 clusters of galaxies obtained by machine scans of Palomar 48-inch Schmidt plates, as part of a program that has been undertaken to analyze in a uniform manner the properties of a larger sample of galaxy clusters. Among the various statistical tests, they applied the Rayleigh test (2.7) (i) to detect possible deviations from uniformity of the distributions of the position angle of the cluster θ_C, the position angle of the first-ranked galaxy θ_1, the position angle of the second-ranked galaxy θ_2, and the position angle of the separation vector between the first and the second-ranked galaxy θ_R, and (ii) to the distribution of a subsample of galaxies in each cluster defined by 10-20 brightest galaxies within 3 core radii from the cluster center, and having ellipticity $\epsilon \geq 0.2$. Furthermore, they computed (iii) the correlation between all the different pairs formed with the angles $\theta_C, \theta_1, \theta_2$, and θ_R by using the statistical parameters appropriate for circular data

$$C_{ab} = \mathrm{Cov}(\theta_a, \theta_b)(V_a, V_b)^{-1/2}, \quad a, b, = C, R, 1, 2,$$

where $V_a = \mathbb{E}[1 - \cos(\theta_a - \overline{\theta}_a)]$ is the sample variance of θ_a and $\mathrm{Cov}(\theta_a, \theta_b) = \mathbb{E}[2\sin(\frac{1}{2}(\theta_a - \overline{\theta}_a)\sin(\frac{1}{2}(\theta_b - \overline{\theta}_b)]$ is the covariance of θ_a and θ_b. They obtained that for (i) none of the four distributions showed significant deviations from uniformity, indicating that the data acquisition and reduction procedures do not introduce important biases, and (ii) there is no preferred orientation for 48 cases, for the five clusters (A2088, A2093, A2100, A2122, and A2124) belonging to the same plate, there is a significantly preferred orientation but due to guiding problems during the exposure (as shown by the alignment of the stellar images), and for the two remaining clusters (A376 and A505) there is a preferred orientation with 92% and 95% significant confidence levels, respectively. Hence for (ii), they concluded that there is, in general, no effect present, except in some peculiar cases. For (iii), they found a correlation of θ_1 with θ_C (significant at the 99.8% level) and a correlation of θ_R with θ_C (significant at the 99.9% level).

In order to determine a possible alignment of the galaxy, one of the most favored theories for the origin of angular momentum is the tidal torque theory of Peebles (1969) [Pee69]. Coutts (1896) [Cou96] used circular statistics to model random tidal torques exerted by galaxies for a dataset of 1000 clusters of various sizes. More precisely he used the Dimroth-Watson distribution (2.6) to model the dispersion of a cluster of galaxies initially aligned in the plane of the shock produced by a pancake collapse. He showed that the Dimroth-Watson distribution is a good model for the results of numerous random small rotations on a sphere in a way analogous to that in which the Gaussian is the result of numerous small movements on the line. Hence, he concluded that an initial alignment is not generally dispersed by subsequent torques.

Jain and Ralston (1999) [JR99] tested a new and independent data set of 361 points under the null proposal of statistical independence of linear polarization alignments relative to galaxy axes versus their angular positions. The null hypothesis was tested via maximum likelihood (ML) analysis of best fits among numerous independent types of factors distributions. More precisely they used Jupp and Mardia's invariant correlation test statistics for this study. They found that the null proposal is not supported at the level of less than 5% to less than 0.1%.

Sarala and Jain (2001) [SJ01] proposed a new method for the extraction of rotation measures from spectral polarization data. The polarizations of radio waves from cosmologically distant sources undergo Faraday rotation upon propagation through galactic magnetic fields. This effect provides very useful information about the magnetic fields in our Galaxy as well as in the host galaxy [ZRS83, Val97]. Their method is based on the maximum likelihood analysis of the Mac-Donald-Bunimovitch (MB) and vM distribution (2.2) of the polarization angle. The MB distribution has the pdf

$$f_{\mathrm{MB}}(\theta) := \frac{1-\xi^2}{\pi(1-\mu^2)^{3/2}}\left[\mu \sin^{-1}(\mu) + \frac{\pi}{2}\mu + (1-\mu^2)^{1/2}\right],$$

where $\mu = \xi\cos(2\theta - 2\overline{\theta})$ and $0 \leq \xi \leq 1, \overline{\theta} \in [0, 2\pi)$ are respectively concentration and location parameters. It turns out that their method is statistically more efficient as well as computationally more convenient than the standard χ^2 procedure if the number of data points is very large.

Menezes et al. (2017) [MPC17] studied the distribution of position angles around 1 million galaxies belonging to the HyperLEDA catalog by applying diverse statistical methods to this directional data sample. Each direction in the sample can be represented by a unit vector that forms an angle with a reference direction. By using classical circular statistical tools, they tested (i) the (local) medium alignment (this test consisted in determining the value of a parameter that quantifies how near the uniformity a directional data sample is) and (ii) the directional alignment. They even proposed a statistical test S_D designed for sets of directional data that measures the level of angles concentration for objects relatively near in space. It is defined as the sum of all the dispersion of a set of directional angular data $\theta_1, \theta_2, \ldots, \theta_n$ around a given angle (also known as circular average standard deviation) D divided by the total number n of galaxy data

$$S_D = \frac{1}{n}\sum_{j=1}^{n} D_j.$$

They showed that S_D tends to be smaller in distributions of directional data with some alignment, as compared to those where the angles are randomly distributed. This means that the value of S_D can be used as an indicator of the presence of an alignment in the distribution. In conclusion, they obtained that, with a high probability, these galaxies do not seem to be randomly positioned in the sky (anisotropy appeared in the form of a non-homogeneous position angle distribution in the celestial sphere, as well as in the presence of local alignments), and that there are strong alignments in the directions $[120°, 180°]$, $[\pm 30°, \pm 75°]$, and $[210°, 270°]$. However, this could be due to the size of the used catalogue or that possible alignments in different redshifts may be hidden by the effect of superposition of data in the celestial sphere. Nevertheless, this shows evidence of a possible existence of a cosmological anisotropy in the Universe or an observational bias due to diverse local effects, so further studies are needed to favor the existence of a global anisotropy.

Soler et al. (2022) [S+22] used circular statistics to identify the filamentary structures in the neutral atomic hydrogen (Hi) emission toward the Galactic plane and quantified their orientations. More precisely, they applied the projected modified Rayleigh statistic proposed by Durand and Greenwood (1958) [DG58] and given by

$$V = \frac{\sum_{ij} w_{ij}\cos(2\theta_{ij})}{\sqrt{\sum_{ij} w_{ij}/2}}, \tag{3.4}$$

for specific directions of interest $\theta = 0°$ and $90°$ such that $V > 0$ or $V < 0$ correspond to preferential orientations parallel or perpendicular to the Galactic plane, respectively, to determine whether the distribution of angles is not uniform and peaked at a particular angle. Here the indices i and j run over the pixel locations in the two spatial dimensions for a given velocity channel and w_{ij} is the statistical weight of each angle θ_{ij}. They found that the orientation of the Hi filaments with respect to the Galactic plane changes progressively with Galactocentric distance, from mostly perpendicular or having no preferred orientation in the inner Galaxy to mostly parallel. This means that the regions of the Milky Way's disk display Hi filamentary structures predominantly parallel to the Galactic plane and that for regions at lower Galactocentric radii, the Hi filamentary structures are mostly perpendicular or do not have a preferred orientation with respect to the Galactic plane. The observations indicate that the Hi filamentary structures provide insight into the dynamical processes shaping the Galactic disk, leading to interpret their results as the imprint of the supernova feedback in the inner Galaxy and Galactic rotation and shear in the outer Milky Way.

Recently, Godlowski and Mryzygłód (2023) [GM23] have proposed a new method to analyze the orientation of galaxies by taking into account the information about the morphological types of galaxies. Studies have shown that to obtain correct analysis results it is preferable to consider galaxies as oblate spheroids, with the real axis ratio depending on the morphological type, which is unfortunately not contained in most of the astronomical data available today (e.g., the Tully Nearby Galaxies catalog). For the development of their new method of investigation, they employed the classical directional test statistics such as the Kolmogorov-Smirnov, CvM and Watson tests (see Sect. 2.2). Their method shows that in small groups and clusters of galaxies, such as Tully clusters, there is no galaxy alignment. Moreover, the authors believe that their method allows the study of catalogs of galaxy clusters in the absence of knowledge of the morphological types of member galaxies.

3.5.9 Exoplanets

In 1995, Mayor and Queloz [MQ95] detected the first exoplanet around a Sun-like star, and since then astronomers have found more than 1000 exoplanets. More precisely, as of 27 June 2024, we count 5678 confirmed exoplanets in a total of 4231 planetary systems with 952 systems being composed of more than one exoplanet. This has opened a recent page in astronomy, which could eventually answer questions regarding the formation and evolution of planetary systems, including our Solar system. With cluster analysis, it is possible to gain insight into the information provided by exoplanets data. For instance, it permits to find clusters of a data set with the most similarity in the same cluster and the most dissimilarity between different clusters, thus leading to a possible better understanding of the formation and evolution of planetary systems.

Database. The Extrasolar Planets Catalog compiled by [SDL⁺11] provides data on exoplanets as well as [FLE87, Appendix B20]. A more updated exoplanets dataset can also be downloaded from The Extrasolar Planets Encyclopedia (`http://exoplanet.eu/catalog/`), which contains for each registered exoplanet several relevant astronomical variables such as projected mass, orbital period, semimature axis, orbital eccentricity, stellar metallicity, and stellar mass.

Statistical tools and data analysis. Many authors [JIY03, PR02, ZM02] have indicated that there is a possible correlation between the planetary mass and the orbit period in the study on the exoplanets. This motivated Hung et al. (2015) [HCCY15] to propose a simple and intuitive clustering algorithm for spherical data based on an intuitive concept and examine where groups for exoplanets on these two important features are. Their algorithm is based on the d-dimensional FvML distribution (2.4). By considering the data of 731 exoplanets from the Extrasolar Planets Catalog (data as of December 14 2012, where incomplete data are excluded), where each of them has the values of projected mass and orbital period, [HCCY15]'s clustering results suggest two main implications, namely that there are three major clusters which might associate with the disc, ongoing tidal and tidal interactions, and that stellar metallicity does not play a key role in exoplanet migration.

Yang et al. (2016) [YCCH16] presented a similarity-based clustering method for spherical data based on a mixture of Fisher distributions (2.5) that identify characteristics without initialization. They applied their unsupervised clustering algorithm on data from the Extrasolar Planets Catalog of May 2005 with 782 exoplanets as well as on data from [FLE87] and obtained similar clustering results as [HCCY15]. Later [YCCH17] proposed a learning-based EM algorithm with FvML distribution (2.4) to cluster spherical exoplanet data in extrasolar planets (Extrasolar Planets Catalog of December 2012 with 731 exoplanets and December 2015 with 870 exoplanets). The obtained clustering results allow an interpretation of exoplanet migration.

Recently, Saavedra-Nieves and Fernández-Péerez (2023) [SNFP23] have extended the directional density-based clustering methodology for the unit hypersphere by solving the computational problems associated to high dimensional spaces and applied it to analyze an exoplanets dataset. For their study, they considered the exoplanets dataset from *The Extrasolar Planets Encyclopedia* and checked, in the same spirit as [HCCY15], where exoplanet groups on the two features, orbital period P and projected mass M_p, are located, and checked the existence of correlation within the astronomical variables M_p, P, semimature axis a, orbital eccentricity e, stellar metallicity $[Fe/H]$, and stellar mass M_s. More precisely, they employed their density-based clustering algorithm to (i) the data $(\ln(M_p), \ln(P))$ on $\mathbb{S}^1$ and (ii) the dataset $(M_p, a, e, [Fe/H], M_s)$ on $\mathbb{S}^4$. They obtained interesting results for different cases of exoplanets and cluster cores, which are more or less similar to the results obtained by [HCCY15].

3.5.10 Hot stars

In astronomy, hot stars are not well represented in the Kepler sampler of planet-hosting stars, and hot Jupiters (which are in close proximity to their stars and have high surface-atmosphere temperatures) are a class of gas giant exoplanets that are inferred to be physically similar to Jupiter but that have very short orbital periods. While it has been known that hot stars with hot Jupiter tend to have high obliquities, less is known about the degree of spin-orbit alignment for hot stars with other kinds of planets.

Database. The California Kepler Survey (CKS) [P+17, J+17] provides data for the planet host with precise determinations of the effective temperature, surface gravity, iron metallicity, and projected rotation velocity. The data include the main-sequence stars late-G and late-F, which have been used for the study of [LWP+21].

Statistical tools and data analysis. Louden et al. (2021) [LWP+21] reassessed the obliquities of the hot Kepler stars with transiting planets smaller than Neptune. By employing model-independent tests (such as the two-dimensional Kolmogorov-Smirnov test for differences between the projected rotation velocities distributions of the two samples, planet hosts and the control stars) and the simple FvML distribution model (2.4) (to characterize the obliquity distribution of the planet hosts), [LWP+21] described possible implications for theories of obliquity or inclination excitation. They found that (i) planet hosts have systematically higher values of rotation velocities distributions than the control stars, but not by a large enough amount to be compatible with perfect spin-orbit alignment, and (ii) there is evidence that the hottest stars have a broader obliquity distribution. Based on these results, they conclude that obliquity excitation for early-G and late-F stars appears to be a general outcome of stars and planet formation instead of being exclusively linked to hot Jupiter formation.

3.6 Tracking satellite orbit phases and space objects

Since the launch of the first Earth's artificial satellite Sputnik 1 in 1957 more than a thousand satellites have been launched and there is no end in sight. Many are still there, and we face an ever-increasing risk of collision as we launch more. In fact, humanity has created a bit of a mess around our Earth along with bits of debris from the rockets we have launched over the years, too. This mess is referred to as space debris or space junk, and includes any noncontrollable defunct human-made object (such as rocket bodies, solar pannels, unused thermal blankets of astronaut) or even pieces and debris which no longer serve a useful function.

Aside from these floating space debris, there are other flying space objects that might be dangerous for us, for example, asteroids. For instance, the Chicxulub asteroid - the most legendary space rock in the world − impacted our Earth 66 million years ago causing a mass extinction event that saw most non-flying dinosaurs and many other species wiped out. Fortunately, the impacts from such large asteroids are immensely rare, however small and medium-sized rocks are far more common in our Solar System and these can still do serious damage by creating explosive airbursts, with resulting shockwaves that may shatter glass, damage buildings and injure anyone who happens to be nearby.

Therefore, tracking space objects or "near-Earth objects" (NEOs) (the term refers to any natural object, like an asteroid, whose orbit brings it close to Earth) is important to protect our Earth and ourselves, e.g. by developing statistical models that permit us to estimate the distribution of space objects orbiting around our Earth. This can be split into two major challenges:

- the ability to accurately associate space object detections with unique space objects;
- the ability to accurately predict where any space object will be as a function of time.

3.6.1 Asteroids

Asteroids are rocky and airless remnants left over from the early formation of our Solar system about 4.6 billion years ago. They are a belt of several thousand minor planets orbiting around our Sun in orbits between the planets Mars and Jupiter. According to NASA (`https://solarsystem.nasa.gov/asteroids-comets-and-meteors/asteroids/in-depth/#many_shapes_and_sizes_otp`), nowadays 1.282.499 asteroids are known. Many asteroids have an axis around which they are observed to spin. [Jup95] mentioned that a hypothesis of astronomical interest is that the directed spin axes $\boldsymbol{z}$ are uniformly distributed. However, for numerous asteroids the sense of spin is unknown, so that we cannot distinguish between $\boldsymbol{z}$ and $-\boldsymbol{z}$. This means that the classical directional distributions on $\mathbb{S}^2$ such as the Fisher distribution (2.5) or the Watson distribution (2.6) cannot be employed, since only the axis $\pm\boldsymbol{z}$ is observed instead of the direction $\boldsymbol{z}$.

Database. The catalog of Magnusson (1988) [Mag88] is a suitable data set that contains 54 asteroids with known spin axis. A more updated database of asteroids is available on NASA's JPL Solar System Dynamics Small-Body Database Lookup (`https://ssd.jpl.nasa.gov/tools/sbdb_lookup.html#/`), Asteroid Data Sets from Planetary Science Institute (`https://sbn.psi.edu/pds/archive/asteroids.html`), Astertank - Asteroid Database and Mining Rankings (`https://www.asterank.com/`), Minor Planet Physical Properties Catalogue – Asteroid Database (`https://mp3c.oca.eu/`,) and ESA's Near-Earth Objects Coordination Centre (`https://neo.ssa.esa.int/`), to name only a few.

In the R package `spunif` [GPVN21], we can select in the `craters` the target type of celestial body "`Asteroid`", which contains the asteroids database of IUA. Note that these updated databases have not been used in the listed survey papers.

Statistical tools and data analysis. Jupp (1995) [Jup95] proposed to replace the Fisher distribution (2.5) for $\boldsymbol{z}$ (prior to selection) by a Bingham distribution (see special cases under Eq. (2.5)) to test the uniform distribution of the directed spin axes $\pm\boldsymbol{z}$ of an asteroid. More precisely, consider the standard model for axial data with probability density function

$$g_0(\boldsymbol{z};\mathbf{A}) = \exp\{\boldsymbol{z}^T\mathbf{A}\boldsymbol{z} - c(\mathbf{A})\}, \tag{3.5}$$

where the parameter $\mathbf{A}$ is a symmetric 3×3 matrix with trace 0 (since $\boldsymbol{z}^T\boldsymbol{z} = 1$) and $c(\mathbf{A})$ is the logarithm of the normalizing constant. Then, if $\boldsymbol{z}$ has the Bingham distribution (3.5), the conditional density of the observable parts $(\pm\mathbf{U}\boldsymbol{z}, \pm(I_3 - \mathbf{U})\boldsymbol{z})$ of $\pm\boldsymbol{z}$ due to the mirror-image ambiguity (where $\mathbf{U} = \boldsymbol{u}^T\boldsymbol{u}$ is the orthogonal projection onto the line of sight) for given $\boldsymbol{u}$ is

$$g_M(\boldsymbol{z};\mathbf{A}|\boldsymbol{u}) = \exp\{\mathrm{tr}(\mathbf{A}[\mathbf{U}\boldsymbol{z}^T\boldsymbol{z}\mathbf{U} + (I_2 - \mathbf{U})\boldsymbol{z}\boldsymbol{z}^T(I_3 - \mathbf{U})]) - c(\mathbf{A},\boldsymbol{u})\}.$$

Now if this model is combined with the selection of the Watson distribution $f_W(\boldsymbol{z};\boldsymbol{u},\kappa)$ (2.6) (here $\boldsymbol{z}^T = \boldsymbol{x}, \boldsymbol{u} = \boldsymbol{\mu}$ and κ is the parameter measuring the strength of the selection effect), [Jup95] obtained as the conditional probability density function for a given $\boldsymbol{u}$ of the observable parts $(\pm\mathbf{U}\boldsymbol{z}, \pm(I_3 - \mathbf{U})\boldsymbol{z})$ of the selected axes $\pm\boldsymbol{z}$

$$g(\boldsymbol{z};\mathbf{A},\kappa|\boldsymbol{u}) = g_M(\boldsymbol{z};\mathbf{A}|\boldsymbol{u}) f_W(\boldsymbol{z};\boldsymbol{u},\kappa). \tag{3.6}$$

Note that $\mathbf{A} = \mathbf{0}$ corresponds to uniformity (prior to selection) of $\pm\boldsymbol{z}$. For an asteroid database of 54 asteroids (where for each asteroid [Jup95] made an arbitrary choice for the determination of its spin axis $\pm\boldsymbol{z}$), [Jup95] obtained the score test $w_{\mathbf{A}} = 4.59$ ($\mathbf{A} = 0$ in (3.6)) and $w_\kappa = 0.6$ ($\kappa = 0$ in (2.6)). He concluded that the uniform distribution may be fitted to this data set.
Note that Jupp had already published a paper about directional statistics in astronomy in 1990 [Jup90], this work [Jup95] is an updated version.

3.6.2 Space debris and its orbital uncertainty propagation

Satellites in orbit underpin our modern lives, however, due to the aforementioned increasing amount of satellites in our space, we are faced with space debris. Based on a report published by ESA (`https://www.esa.int/Space_Safety/Space_Debris/Space_debris_by_the_numbers`) on June 18, 2024, more than 130 million space debris (most of them of size less than 1 mm) are orbiting our Earth and only around 34.000 objects are traceable due to the size limitation. The biggest danger of space debris is posed to our spacecraft, astronauts, and satellite operations. Due to the high population and large relative velocities of space debris, it is extremely difficult to track them accurately using for example optical observations and to associate them with past observations. Hence a main challenge is to represent the uncertainty in predicted location, orbit, and velocity of space debris and space objects in general, especially for tracking and association purposes. On top of that, in orbital uncertainty propagation and orbital space object tracking one needs to deal with the nonlinearity of the system equation when expressed in Cartesian coordinates. According to [BKH+17] there are two basic strategies to deal with nonlinearity:

(i) transform the coordinate system to remove the nonlinearity, or

(ii) develop sophisticated methods or use a higher order polynomial to accommodate it.

Database. For the moment, it is difficult to obtain a real dataset of space debris. However, the North American Aerospace Defense Command (NORAD) started compiling a database of all known debris. The Department of Defense maintains a highly accurate satellite catalog on objects in Earth's orbit that are larger than a softball. Moreover, ESA's developed softwares (like DRAMA (Debris Risk Assessment and Mitigation Analysis), MASTER (Meteoroid and Space Debris Terrestrial Environment Reference), and DISCOS (Database and Information System Characterising Objects in Space)) serve to generate a simulated space debris database.

Statistical tools and data analysis.

Orbit determination by the Gauss-vM cylindrical distribution. Horwood and Poore (2012) [HP12] proposed a new class of multivariate probability density functions on the $(n+1)$-dimensional cylindrical manifold $\mathbb{R}^n \times \mathbb{S}^1$, that they named Gauss von Mises (Gauss-vM) distribution. According to them, this distribution determines the orbit of satellites and provides a more statistically rigorous treatment of uncertainty in the space surveillance tracking environment. In several papers [HP12, HP14, HASP14], the same authors (plus some coauthors) have demonstrated the usefulness of their Gauss-vM distribution to characterize the uncertainty in a space object's orbital state. Their Gauss-vM distribution has nice properties, it reduces to the normal $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for a suitable limit, and it is related to the Mardia-Sutton density from [MS78] (see [HP14], Sect. 5). However, they mentioned that their Gauss-vM distribution is preferred for space surveillance problems since the parameter set provides more control over the magnitude of the higher-order cumulants and the bending of the banana- or boomerang-shaped level sets. Moreover, from their analysis, it is demonstrated that uncertainty propagation with their Gauss-vM distribution can be achieved at the same cost as the traditional unscented Kalman filter and that it can maintain "uncertainty realism" for up to eight times as long.

Space debris. Kent et al. (2016) [KHJ16] proposed a special directional distribution to track space debris. Their new distribution is a special case of the Fisher-Bingham distribution on the unit sphere, which they called the "extreme" 5-parameter Fisher-Bingham distribution (FB5e), sometimes also referred to as the Fisher-Bingham-Kent (FBK) distribution (see special cases under Eq. (2.5)). The adjective "extreme" is used to contrast it with another 5-parameter subfamily, the "balanced" or Kent distribution FB5b. It turns out that the FB5e is more appropriate in the semi-concentrated setting. In particular, in contrast to the FB5b, the FB5e density is better to describe data that lie very near a great circle (i.e., where the mode lies on the equator and the data are tightly clustered near the equator), but whose projection onto the great circle has a unimodal distribution. The range of the data along the equator roughly fills a semicircle, and this pattern cannot be created with the FB5b distribution. Since space debris data is not available, they demonstrated the potential usefulness of their new distribution along with two realistic space debris problems. The aim was to map the uncertainty in an object's trajectory down to the surface of a sphere and model point clouds concentrated near a great circle on $\mathbb{S}^2$.They showed that their FB5e distribution is promising but an extended analysis is required with more realistic orbital models.

In 2017, Kent et al. [KBHJ17] showed that it is possible to describe succinctly the angular part of the position vector at later times of a space object by using their FB5e distribution (see special cases under Eq. (2.5)). In fact, it was already observed in [KHJ16] that the propagated distribution of a space object under noisy initial conditions tends to follow a banana-shaped distribution in Euclidean coordinates. The projection of this distribution onto $\mathbb{S}^2$ is well-approximated under a variety of conditions by their FB5e. They showed that by simplifying the initial noise to a small set of sigma points, it is possible to gain qualitative insight into the way in which the five parameters of the FB5e distribution vary with time. Thus their FB5e can also be applied for Space Situational Awareness (SSA) applications.

In the same year, Bhattacharjee et al. (2017) [BKH+17] stressed that the FB5e on $\mathbb{S}^2$ can be used to approximate the propagated uncertainty and that the 2-dimensional representation of their distribution is typically approximately bivariate normal. Moreover, they demonstrated the usefulness of their distribution in various data association and manoeuver analysis problems and discussed about a newly defined coordinate system, which they referred to as the "Adapted STructural (AST) coordinate system" to represent the orbital structure, normal direction, and location of an object along its orbit. The purpose behind this AST coordinate system is to preserve normality and show the usefulness and efficiency of such a system in debris tracking problems. Note that other works of the same authors have been published [KBFH19, KBFH20], in particular Bhattacharjee's Ph.D. thesis [Bha20], where they developed more in detail their AST coordinate system and gave various uses for suitable examples.

In 2018, Kent et al. [KBH+18] investigated how to fit the FB5e model using the EM algorithm and discussed a variety of SSA applications. In fact, although their FB5e distribution is a suitable model to describe the angular position of a propagated space object when the initial position and velocity are known up to small errors, in settings where the initial conditions are less constrained, e.g. when the manoeuvering capabilities of the object are uncertain, it is better to fit a mixture of FB5e distributions. They showed that the FB5e distribution on $\mathbb{S}^2$ provided a powerful and tractable model to summarize uncertainty and, moreover, it is straightforward to implement the classic methods of discriminant analysis for this classification problem. Hence in supervised learning, the mixture models can be used to allocate observations to one of a set of known classes or populations. Regarding the unsupervised learning setting, they proposed an EM algorithm, where under high concentration their FB5e reduces to the bivariate normal distribution, and much of the methodology reduces to the corresponding methodology for Gaussian mixture models. In contrast to Gaussian mixture models, the FB5e distribution remains tractable even when the normal approximation is too crude (in the case of a "semi-concentrated" behavior on the sphere).

3.6.3 GNSS phase observations

Global Navigation Satellite System (GNSS) refers to a constellation of Earth-orbiting satellites (Europe's Galileo, USA's NAVSTAR Global Positioning System (GPS), Russia's Global'naya Navigatsionnaya Sputnikovaya Sistema (GLONASS), and China's BeiDou Navigation Satellite System) that broadcast their locations in space and time, of networks of ground control stations, and of receivers that calculate ground positions by trilateration. They are used in all forms of transportation (e.g., space stations, aviation, maritime) and also to control computer networks, air traffic, power grids, and more. Besides GNSS, the use of a reflected GNSS signal for Earth observation, referred to as GNSS reflectometry (GNSS-R), has become popular, too, and is used for research applications in altimetry, oceanography and soil moisture monitoring, to name but a few.

The GNSS carrier phase measurements form the basis of high-precision satellite positioning, as they measure the apparent distance between the satellite and the receiver. However, there are phase biases due to effects at a satellite and a receiver and thus these biases have to be carefully modeled and substracted from GNSS measurements. One approach to do this is using circular statistics. In addition, according to [KRS+17] the GNSS-R phase delay observations are angular and affected by a noise assumed to follow the vM distributions (2.2).

Database. GPS data is collected from 45 stations of the global International GNSS Service (IGS) tracking network. The global distribution of the stations provides almost full coverage of GPS satellites orbits without large tracking gaps. Depending on the study, the upper limit for the root-mean-square of the L1 and L2 multipath observables (MP1 and MP2) is usually set to 0.5m in order to mitigate the multipath effect, and a 2-month time span is chosen to assess a day-to-day behavior of the satellite-satellite single-difference wide lane (SSWL) phases biases (this was the case for [Kes04]'s study).

Concerning [CGH07]'s study, they used the GPS observation data collected in 2005 from four short baselines (2-3 km), which are measured in two two-hour observation periods with a 20-second sampling rate. They computed the baseline length using observations above 10°. In total, they considered 7198 L1 double difference phase observables with 0.0097 cycles.

One can also work with synthetic data corresponding to a realistic GPS constellation scenario, which was the case for [KRS+17]'s study. In particular, Kucwaj et al. (2017) [KRS+17] made use of the observed GPS signals collected in June 2016 which were reflected on an artificial basin in Calais, France (see [KRS+17, Table 2 and Table 3]).

Statistical tools and data analysis. Keshin (2004) [Kes04] investigated the problem of single-receiver real-time carrier phase ambiguity resolution by using circular statistics. More precisely, he considered the simplified situation where one needs to deal with satellite-satellite single difference ambiguities instead of undifferenced phase ambiguities which are fraction phase biases due to effects at a satellite and a receiver (the initialization constants, the clock errors). Hence, in this way, any receiver-based effects and the need of SSWL biases calibration are completely canceled. Keshin's idea is to analyze SSWL biases using goodness-of-fit tests performed for the basic circular distributions (see Sect. 2.1.1). For this study, he considered a 2-month GPS data set from 45 stations of the global IGS tracking network from November to December 2003. It turns out that from his results, the vM (2.2) and Wrapped Normal (obtained by wrapping the classical normal distribution around the circle) distributions are most likely to be a reasonable model for the SSWL phase biases and can be used as a base model in further investigations related to statistical analysis of the SSWL phase biases. Moreover, the vM model makes it possible to establish strong criteria for detecting outliers and to investigate the impact of systematic effects on the SSWL phases biases.

Cai et al. (2007) [CGH07] also observed that the vM distribution (2.2) on the circle is more suitable than the Gauss-Laplace normal distribution to investigate the distribution of the GPS carrier phase observable as well as double difference GPS phase observations. From the GPS observation data collected in 2005, they showed that the statistical properties of the vM distribution and their related hypothesis tests improve the theoretical background and application reality of GPS geodesy and navigation and can play an essential role for the forthcoming Galileo, Europe's civilian-managed GNSS.

Kucwaj et al. (2017) [KRS+17] recognized as well the usefulness of the vM distribution (2.2) to derive a linear-circular regression estimator for application to ground-based GNSS-R altimetry. Their approach is to assume that the GNSS-R phase delay observations are processed as angles, defined in the circular domain, and the measurement noise follows a vM distribution. The aim of their linear-circular regression estimator is to estimate the phase delay slope in the maximum likelihood sense. By testing on various scenarios, meaning experimentation on synthetic and real data (collected in 2016 over the city of Calais), they showed that the proposed estimator is able to fuse phase observations obtained from several satellite signals and allows the sea-surface height to be estimated from datasets with large data gaps.

4 Conclusion

In this paper we have provided a broad and, to the best of our knowledge, as complete as possible review of the importance of directional statistics for space sciences and astronomy problems. We hope to have showcased how suitable the field of directional statistics is for this type of data and problems, and hereby generate further research in this direction. Historically, the domains of astronomy and statistics have mutually benefited each other, and future interdisciplinary collaborations involving directional statistics appear very promising to us. On the one hand space scientists and astronomers will be able to analyze their data with tailor-made cutting-edge tools, and on the other hand statisticians will be able to develop new methodological approaches and test them on incredibly rich and challenging data.

Acknowledgement. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n° 811017 and also by the National Research Fund, Luxembourg under grant project reference AFR/17917651/BIL_PDR_ASTROSTATS.

References

[AAA+08] J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, D. Allard, I. Allekotte, J. Allen, P. Allison, J. Alvarez-Muñiz, M. Ambrosio, et al., *Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei.*, APh **29** (2008), 188–204.

[AAA+09] J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, E.J. Ahn, D. Allard, I. Allekotte, J. Allen, P. Allison, J. Alvarez-Muniz, et al., *Upper limit on the cosmic-ray photon fraction at eev energies from the Pierre Auger observatory.*, APh **31** (2009), 399–406.

[AAA+10] P. Abreu, M. Aglietta, E.J. Ahn, D. Allard, I. Allekotte, J. Allen, J. Castillo, J. Alvarez-Muñiz, M. Ambrosio, A. Aminaei, et al., *Update on the correlation of the highest energy cosmic rays with the positions of nearby extragalactic matter.*, APh **34** (2010), 314–326.

[ABJ+15] J.-L. Agati, D. Bonneau, A. Jorissen, E. Soulié, S. Udry, P. Verhas, and J. Dommanget, *Are the orbital poles of binary stars in the solar neighborhood anisotropically distributed?*, A&A **574** (2015).

[AL14] T. Alexopoulos and S. Leontsinis, *Benford's Law in Astronomy.*, A&A **35** (2014), 639–648.

[AS02] S. Ando and K. Sato, *Determining the supernove direction by its neutrinos.*, Prog. Theor. Phys. **957** (2002).

[AZDH05] J. L. Alvarellos, K. J. Zahnle, A. R. Dobrovolskis, and P. Hamill, *Fates of satellite ejecta in the Saturn system.*, Icarus **178** (2005), 104–123.

[B+09] S. Bridles et al., *Handbook for the GREAT08 Challenge: An image analysis competition for cosmological lensing.*, Ann. Appl. Stat. **3** (2009), 6–37.

[B+16] N. B. Bekhti et al., *HI4PI: A full-sky Hi survey based on EBHIS and GASS.*, A&A **594** (2016).

[B+19] R. A. Batista et al., *Open Questions in Cosmic-Ray Research at Ultrahigh Energies.*, Front. Astron. Space Sci., Sec. - High-Energy and Astroparticle Physics **6** (2019).

[Bab61] H. W. Babcock, *The topology of the sun's magnetic field and the 22-year cycle.*, ApJ **133** (1961), 572–587.

[Bal07] A. Balbi, *The Music of the Big Bang.*, Springer-Verlag, 2007.

[BCDM22] L. Barabesi, A. Cerioli, and M. Di Marzio, *Statistical models and the Benford hypothesis: A unifying framework.*, Stat. Sci., 2022.

[BD94] B. Boulerice and G. R. Ducharme, *Decentered directional data.*, AISM **46** (1994), 573–586.

[Ber68] R.J. Beran, *Testing for uniformity on a compact homogeneous space.*, J. Appl. Probab. **5** (1968), no. 1, 177–195.

[Ber69] ———, *Asymptotic theory of a class of tests for uniformity of a circular distribution.*, Ann. Math. Statist. **40** (1969), no. 4, 1196–1206.

[Ber83] F. Berglund, *The Orientation of the Orbit Planes of Visual Binary Stars.*, Lunds Universitets Arsskrift **34** (1983), 2–12.

[BGL16] T. Baranyi, L. Győri, and A. Ludmány, *On-line tools for solar data compiled at the Debrecen observatory and their extensions with the Greenwich sunspot data.*, Sol. Phys. **291** (2016), 3081–3102.

[Bha20] S. Bhattacharjee, *Tracking space debris using directional statistics.*, Ph.D. thesis, University of Leeds, 2020.

[Bie84] M. F. Bietenholz, *An investigation of the polarization properties of extragalactic radio sources of different morphologies, and their relation to claimed large scale anisotropies in the universe.*, Ph.D. thesis, University of Toronto, 1984.

[Bin64] C. Bingham, *Distributions on the sphere and on the projective plane.*, Yale University, 1964.

[Bir82] P. Birch, *Is the Universe rotating?*, Nature **298** (1982), 451–454.

[BKH+17] S. Bhattacharjee, J. T. Kent, I. I Hussein, W. R Faber, and M. K. Jah, *Application of Directional Statistics to Problems in SSA.*, Acta Astronautica, 1st IAA Conference on Space Situational Awareness (ICSSA), 2017.

[BKMP09] P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, *Adaptive density estimation for directional data using needlets.*, Ann. Stat. **37** (2009), 3362–3395.

[BL15] T. Budavári and T. J. Loredo, *Probabilistic Record Linkage in Astronomy: Directional Cross-Identification and Beyond.*, Annu. Rev. Stat. Appl. **2** (2015), 113–139.

[BMP79] A. Baranne, M. Mayor, and J. L. Poncet, *Coravel - A new tool for radial velocity measurements.*, Vistas Astron. **23** (1979), 279–316.

[BNN+04] D. Baba, T. Nagata, T. Nagayama, C. Nagashima, D. Kato, M. Kurita, S. Sato, Y. Nakajima, M. Tamura, H. Nakaya, et al., *Deep Near-Infrared Imaging toward the Vela Molecular Ridge C. I. A Remarkable Embedded Cluster in RCW 36.*, ApJ **614** (2004).

[BV99] J. F. Beacom and P. Vogel, *Can a supernova be located by its neutrinos?*, Phys. Rev. D. **60** (1999).

[C+02] A. Capone et al., *Measurements of light transmission in deep sea with the AC9 transmissometer.*, Nucl. Instrum. Meth. **487** (2002).

[CACF09] J. A. Cuesta-Albertos, A. Cuevas, and R. Fraiman, *On projection-based tests for spherical and compositional data.*, Stat. Comput. **19** (2009), 367–380.

[CFJ90] S. M. Carroll, G. B. Field, and R. Jackiw, *Limits on a Lorentz- and parity-violating modification of electrodynamics.*, Phys. Rev. D. **41** (1990).

[CGH07] J. Cai, E. W. Grafarend, and C. Hu, *The Statistical Property of the GNSS Carrier Phase Observations and its Effects on the Hypothesis Testing of the Related Estimators.*, Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (TX) (Ion GNSS, ed.), Fort Worth, 2007, pp. 331–338.

[CM09] P. Cabella and D. Marinucci, *Statistical challenges in the analysis of cosmic background radiation.*, Ann. Appl. Stat. **3** (2009), 61–69.

[Cou96] A. Coutts, *The scale and dispersion of galactic alignments.*, MNRAS **278** (1996), 87–94.

[Cra09] F. Crawford, *Detecting the cosmic dipole anisotropy in large-scale radio surveys.*, ApJ **692** (2009), 887–893.

[Cru12] R. M. Crutcher, *Magnetic fields in molecular clouds.*, ARA&A **50** (2012), 29–63.

[D+01] G. V. Domogatskii et al., *Status of the BAIKAL neutrino project.*, Nuc. Ph. B. (PS) **91** (2001), 438–444.

[DBKR15] L. Dones, R. Brasser, N. Kaib, and H. Rickman, *Origin and evolution of the cometary reservoirs.*, Space Sci. Rev. **197** (2015), 191–269.

[DDD20] J. DeJong, J. DeBruijne, and J. DeRidder, *Benford's law in the Gaia universe.*, A&A **642** (2020).

[DG58] D. Durand and J. A. Greenwood, *Modifications of the Rayleigh test for uniformity in analysis of two-dimensional orientation data.*, JG **66** (1958), no. 3, 229–238.

[Dim62] E. Dimroth, *Untersuchungen zum Mechanismus von Blastesis und Syntexis in Phylliten und Hornfelsen des südwestlichen Fichtelgebirges: I. Die statistische Auswertung einfacher Gürteldiagramme.*, TMPM **8** (1962), 248–274.

[DM23] J. C. Duque and D. Marinucci, *Geometric methods for spherical data, with applications to cosmology.*, Preprint, 2023.

[Dod15] S. Dodelson, *Modern Cosmology.*, Academic Press, 2015.

[EEK+80] N. N. Efimov, T. A. Egorov, D. D. Krasilnikov, M. I. Pradvin, and I. Y. Slepstov, *Catalogue of highest energy cosmic rays.*, WDC-C for Cosmic Rays **2** (1980).

[ER79] W. Eichendorf and M. Reinhardt, *Polarization properties of extragalactic radio sources.*, Astro. Sp. Sci. **61** (1979), 153–167.

[ER80] ———, *A catalogue of polarization measurements and related data of extragalactic radio sources.*, Ac. Cos. Zesc. **9** (1980), 7–48.

[Fas16] C. I. Fassett, *Analysis of impact crater populations and the geochronology of planetary surfaces in the inner solar system.*, J. Geophys. Res. Planets **121** (2016), 1900–1926.

[FB09] E. D. Feigelson and G. J. Babu, *Modern statistical methods for astronomy: With R applications.*, Cambridge University Press, 2009.

[FDKP13] G. Fay, J. Delabrouille, G. Kerkyacharian, and D. Picard, *Testing the isotropy of high energy cosmic rays using spherical needlets.*, Ann. Appl. Stat. **7** (2013), 1040–1073.

[FdMGP23] A. Fernandez-de Marcos and E. García-Portugués, *Data-driven stabilizations of goodness-of-fit tests.*, Comput. Stat. Data Anal. **179** (2023), 107653.

[FDSIB21] E. D. Feigelson, R. S. De Souza, E. E. O. Ishida, and G. J. Babu, *Twenty-First-Century Statistical and Computational Challenges in Astrophysics.*, Annu. Rev. Stat. Appl. **8** (2021), 493–517.

[Fei09] E. D. Feigelson, *Statistics in astronomy.*, Preprint, 2009.

[Fei16] ———, *The changing landscape of astrostatistics and astroinformatics.*, Proc. Int. Astron. **12** (2016), 3–9.

[Fis53] R.A. Fisher, *Dispersion on a sphere.*, Proc. R. Soc. (London) A - Mathematical and Physical Sciences **217** (1953), no. 1130, 295–305.

[FLE87] N. I. Fisher, T. Lewis, and B. J. Embleton, *Statistical Analysis of Spherical Data.*, Cambridge University Press, Cambridge, 1987.

[FT81] J. R. Fisher and R. B. Tully, *Neutral hydrogen observations of a large sample of galaxies.*, ApJS **47** (1981), 139–200.

[G+00] F. Genova et al., *The CDS information hub.*, A&AS **143** (2000), 1–7.

[G+14] N. Galitzki et al., *The Balloon-borne Large Aperture Submillimeter Telescope for Polarimetry-BLASTPol: Performance and results from the 2012 Antarctic flight.*, Preprint, 2014.

[Gau09] C.F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss.*, sumtibus Frid. Perthes et IH Besser, 1809.

[GBL13] N. Gyenge, T. Baranyi, and A. Ludmány, *Variations of solar non-axisymmetric activity.*, Cent Eur. Astrophys. Bull **1** (2013), 1–16.

[GHB+05] K. M. Gorski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, and M. Bartelmann, *HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere.*, ApJ **622** (2005), no. 2, 759.

[Gil88] S. L. Gillett, *Orbital planes of visual binary stars are randomly oriented: A statistical demonstration.*, AJ **96** (1988), 1967–1970.

[Gin75] E. Giné, *Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms*, Ann. Stat. **3** (1975), no. 6, 1243–1266.

[GLB17] L. Győri, A. Ludmány, and T. Baranyi, *Comparative analysis of debrecen sunspot catalogs.*, MNRAS **465** (2017), 1259–1273.

[GM23] W. Godłowski and B. Mrzygłód, *A new method of investigation of the orientation of galaxies in clusters in the absence of information on their morphological types.*, Appl. Sci. **13** (2023).

[GP13] E. García-Portugués, *Exact risk improvement of bandwidth selectors for kernel density estimation with directional data.*, Electron. J. Statist. **7** (2013), 1655–1685.

[GPCGM13] E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga, *Kernel density estimation for directional-linear data.*, J. Multivar. Anal. **121** (2013), 152–175.

[GPdMMV24] E. García-Portugués, P. L. de Micheaux, S.G. Meintanis, and T. Verdebout, *Nonparametric tests of independence for circular data based on trigonometric moments.*, Stat. Sin. **34** (2024), no. 2, 567–588.

[GPNECA21] E. García-Portugués, P. Navarro-Esteban, and J. A. Cuesta-Albertis, *A Cramér-von Mises test of uniformity on the hypersphere.*, Statistical Learning and Modeling in Data Analysis, Part of the Studies in Classification, Data Analysis, and Knowledge Organization, Springer, 107-116, 2021.

[GPNECA23] E. García-Portugués, P. Navarro-Esteban, and J. A. Cuesta-Albertos, *On a projection-based class of uniformity tests on the hypersphere.*, Bernoulli **29** (2023), 181–204.

[GPPV20] E. García-Portugués, D. Paindaveine, and T. Verdebout, *On optimal tests for rotational symmetry against new classes of hyperspherical distributions.*, JASA **115** (2020), 1873–1887.

[GPPV22] ———, *`rotasym`: Tests for rotational symmetry on the hypersphere.*, CRAN, 2022.

[GPV18] E. García-Portugués and T. Verdebout, *An overview of uniformity tests on the hypersphere.*, Preprint, 2018.

[GPVN21] E. García-Portugués, T. Verdebout, and P. Navarro, *`sphunif`: Uniformity tests on the circle, sphere, and hypersphere.*, CRAN, 2021.

[H+88] K. S. Hirata et al., *Observation in the Kamiokande-II Detector of the Neutrino Burst from Supernova SN 1987a.*, Phys. Rev. D. **38** (1988), 448–458.

[HASP14] J. T. Horwood, J. M. Aristoff, N. Singh, and A. B. Poore, *A comparative study of new non-linear uncertainty propagation methods for space surveillance.*, In Proceedings of the SPIE, Signal, and Data Processing of Small Targets **9092**, 90920H, 2014.

[HCCY15] W.-L. Hung, S.-J. Chang-Chien, and M.-S. Yang, *An intuitive clustering algorithm for spherical data with application to extrasolar planets.*, J. Appl. Stat. **42** (2015).

[HF16] T. P. Hill and R. F. Fox, *Hubbles' law implies Benford's law for distances to galaxies.*, A&A **37** (2016).

[HH12] C. Heiles and M. Haverkorn, *Magnetic fields in the multiphase interstellar medium.*, A&A **556** (2012), 293–305.

[Hir16] N. Hirata, *Differential impact cratering of saturn's satellites by heliocentric impactors.*, J. Geophys. Res. Planets **121** (2016), 111–117.

[HLV22] M. Hallin, H. Liu, and T. Verdebout, *Nonparametric measure-transportation-based methods for directional data.*, Tech. report, Working Papers ECARES 2022-18, ULB, 2022.

[HMW01] W. I. Hartkop, B. D. Mason, and C. E. Worley, *The 2001 US Naval Observatory Double Star CD-ROM. II. The Fifth Catalog of Orbits of Visual Binary Stars.*, AJ **122** (2001).

[HP12] J. T. Horwood and A. B. Poore, *Orbital state uncertainty realism.*, Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, Ryan S., The Maui Economic Development Board, Id.48, September, 2012, pp. 11–14.

[HP14] ______, *Gauss von mises distribution for improved unvertainity realism in space situational awareness.*, JUQ **2** (2014), 276–304.

[HS82] G. Helou and E. E. Salpeter, *Galaxy spins in the virgo cluster.*, ApJ **252** (1982), 75–80.

[HWC87] P. Hall, G. S. Watson, and J. Cabrera, *Kernel density estimation with spherical data.*, Biometrika **74** (1987), 751–762.

[J^+17] J. A. Johnson et al., *The California-Kepler survey. II. Precise physical properties of 2025 Kepler planets and their host stars.*, AJ **154** (2017).

[JHS^+18] D. L. Jow, R. Hill, D. Scott, J. D. Soler, P. G. Martin, M. J. Devlin, L. M. Fissel, and F. Poidevin, *An application of an optimal statistic for characterizing relative orientations.*, MNRAS **474** (2018), 1018–1027.

[JIY03] I.-G. Jiang, W.-H. Ip, and L.-C. Yeh, *On the fate of close-in extrasolar planets.*, AJ **582** (2003), 449–454.

[JKKW03] P. E. Jupp, P. T. Kim, J.-Y. Koo, and P. Wiegert, *The intrinsic distribution and selection bias of long-period cometary orbits.*, JASA **98** (2003), 515–521.

[JM79] P. E. Jupp and J. Mardia, *Maximum Likelihood Estimation for the Matrix von Mises-Fisher and Bingham distributions.*, Ann. Stat. **7** (1979), 599–606.

[JR99] P. Jain and J. Ralston, *Anisotropy in the propagation of radio polarizations from cosmologically distant galaxies.*, Mod. Phys. Lett. A **14** (1999), 417–432.

[JS83] P. E. Jupp and B. D. Spurr, *Sobolev test for symmetry of directional data.*, Ann. Stat. **11** (1983), 1225–1231.

[JT19] S. R. Jammalamadaka and G. Terdik, *Harmonic analysis and distribution-free inference for spherical distributions.*, J. Multivar. Anal. **171** (2019), 436–451.

[Jup90] P. E. Jupp, *Directional statistics in astronomy.*, Cambridge (Bias Errors, Uncertainties in Astronomy Uncertainties in Astronomy Jaschek, and Murtagh, eds.), University Press, 1990, pp. 229–244.

[Jup95] ______, *Some applications of directional statistics to astronomy.*, New Trends in Probability and Statistics, VSP, 1995, pp. 123–133.

[Jup01] P.E. Jupp, *Modifications of the Rayleigh and Bingham tests for uniformity of direction.*, J. Multivar. Anal. **77** (2001), no. 1, 1–20.

[JvK23] J. M. Jeon and I. van Keilegom, *Density estimation and regression analysis on $\mathbb{S}^d$ in the presence of measurement error.*, Preprint, 2023.

[K^+03] A. Karle et al., *IceCube: The next generation neutrino telescope at the South Pole.*, NPB(PS) **188** (2003), 388–395.

[K^+05] P. M. W. Kalberla et al., *GASS: the Parkes Galactic All-Sky Survey.*, A&A **440** (2005), 775–782.

[K+11a] J. Kerp et al., *The Effelsberg Bonn H I Survey (EBHIS).*, Astron. Nachrichten **332** (2011).

[K+11b] T. Kitching et al., *Gravitional lensing accuracy testing 2010 (GREAT10) challenge handbook.*, Ann. Appl. Stat. **5** (2011), 2231–2263.

[K+12] J. D. Kirkpatrick et al., *Further defining spectral type "y" and exploring the low-mass end of the field brown dwarf mass function.*, ApJ **753** (2012).

[KBFH19] J. T. Kent, S. Bhattacharjee, W. R. Faber, and I. I. Hussein, *Revisiting the orbital tracking problem.*, Preprint, 2019.

[KBFH20] ______, *A unified approach to the orbital tracking problem.*, 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2020, pp. 14–16.

[KBH+18] J. T. Kent, S. Bhattacharjee, H. Hussein, W. R. Raber, and M. K. Jah, *Fisher-Bingham-Kent Mixture models for angles-only observation processing.*, AIAA Session: Space Situational Awareness, Orbital Debris, and Conjunction Analysis I (2018), 2018–1972.

[KBHJ17] J. T. Kent, S. Bhattacharjee, I. I. Hussein, and M. K. Jah, *Orbital error propagation analysis using directional statistics for space objects.*, McMahon (Leve and Guo, ed.), Proceedings of the 27th AAS/AIAA Space Flight Mechanics Meeting, Adv. Astronaut. Sci. (160), 2017.

[Ken82] J.T. Kent, *The Fisher-Bingham distribution on the sphere.*, J. R. Stat Soc. Ser. B (Methodological) **44** (1982), no. 1, 71–80.

[Kes04] M. Keshin, *Directional statistics of satellite-satellite single-difference widelane phase biases.*, J. Geod. **39** (2004), 303–324.

[KHJ16] J. T. Kent, I. I. Hussein, and M. K. Jah, *Directional distributions in tracking of space debris.*, 19th International Conference on Information Fusion, 2016, pp. 2081–2086.

[KKC+93] T. Konishi, T. Kitamura, M. Chikawa, K. Tsuji, T. Nakatsuka, and S. Konishi, *Spherically statistical tests on the angular distributions of electrons of the IMB events associated with SN 1987A.*, Phys. Rev. D. **47** (1993), 5228.

[KKPN13] P. T. Kim, J.-Y. Koo, and T. M. Pham Ngoc, *Supersmooth testing on the sphere over analytic classes.*, JNPS **28** (2013), 84–115.

[KO11] K. Kotera and A. V. Olinto, *The astrophysics of ultrahigh-energy cosmic rays.*, ARA&A **49** (2011), 119–153.

[KRS+17] J.-C. Kucwaj, S. Reboul, G. Stienne, J.-B. Choquel, and M. Benjelloun, *Circular regression applied to GNSS-R phase altimetry.*, Remote Sens. **9** (2017).

[KS18] A. Kume and T. Sei, *On the exact maximum likelihood inference of Fisher–Bingham distributions using an adjusted holonomic gradient method.*, Statist. Comput. **28** (2018), 835–847.

[Kui60] N.H. Kuiper, *Tests concerning random points on a circle.*, Nederl. Akad. Wetensch. Proc. Ser. A, vol. 63, 1960, pp. 38–47.

[KY84] D. G. Kendall and G. A. Young, *Indirectional statistics and the significance of an asymmetry discovered by Birch.*, MNRAS **207** (1984), 637–647.

[Lan05] P. Langevin, *Magnétisme et théorie des électrons.*, Ann. chim. et phys. (1905), no. 8, 203.

[LGS06] A. G. Lyne and F. Graham-Smith, *Pulsar astronomy.*, Cambridge University Press, Cambridge, UK, 2006.

[LHP20] F. Lombard, D. M. Hawkins, and C. Potgieter, *Nonparametric cusum charts for circular data with applications in health science and astrophysics.*, Revstat. Stat. **18** (2020), 461–481.

[Lin80] J. Linsely, *Catalogue of highest energy cosmic rays.*, WDC-C for Cosmic Rays **2** (1980).

[LK05] D. R. Lorimer and M. Kramer, *Handbook of pulsar astronomy.*, Cambridge University Press, Cambridge, UK, 2005.

[LM12] F. Lombard and R. K. Maxwell, *A cusum procedure to detect deviations from uniformity in angular data.*, J. Appl. Stat. **39** (2012), 1871–1880.

[LPN14] C. Lacour and T. M. Pham Ngoc, *Goodness-of-fit test for noisy directional data.*, Bernoulli **20** (2014), 2131–2168.

[LSV14] C. Ley, C. Sabbah, and T. Verdebout, *A new concept of quantiles for directional data and the angular Mahalanobis depth.*, Electron J. Statist. **8** (2014), 795–816.

[LV17] C. Ley and T. Verdebout, *Modern directional statistics.*, Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, 2017.

[LV18] C. Ley and T. Verdebout (eds.), *Applied directional statistics.*, Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Boca Ration, 2018.

[LWP+21] E. M. Louden, J. N. Winn, E. A. Petigura, H. Isaacson, A. W. Howard, K. Masuda, S. Albrecht, and M. R. Kosiarek, *Hot stars with Kepler planets have high obliquities.*, AJ **161** (2021).

[Mag88] P. Magnusson, *Pole determinations of asteroids.*, Asteroids II (Binzel and others Binzel al, eds.), University of Arizona Press, Tucson, 1988.

[Mar75] K. V. Mardia, *Statistics of directional data.*, J. R. Stat. Soc. Ser. B **37** (1975), 349–393.

[Mar90] P. R. Marchant, *An analysis of the arrival directions of the highest energy cosmic rays.*, J. Phys. G: Nucl. Part. Phys. **16** (1990), 1385–1398.

[MBM23] A. Montin, A. R. Brazzale, and G. Menardi, *Locating γ-ray sources on the celestial sphere via modal clustering.*, Preprint, 2023.

[McK09] M. M. McKinnon, *Polarization patterns in pulsars radio emission.*, ApJ **692** (2009).

[McK10] ———, *Directional statistics for polarization observations of individual pulses from radio pulsars.*, AIP Conf Proc. **1283** (2010).

[MCÖ+17] L. Medeiros, C.-K Chan, F. Özel, D. Psaltis, J. Kim, D.P. Marrone, and A. Sadowski, *Variability in GRMHD simulations of Sgr: Implications for EHT closure phase observations.*, ApJ **844** (2017), no. 1, 35.

[MD79] H. T. MacGillivray and R. J. Dodd, *Properties of faint galaxies in two fields near the South Galactic Pole.*, MNRAS **186** (1979), 69–84.

[ME82] K. V. Mardia and R. Edwars, *Weighted distributions and rotating caps.*, Biometrika **69** (1982), 323–330.

[MG+05a] N. M. McClure-Griffiths et al., *The Southern Galactic Plane survey: H I Observations and Analysis.*, ApJ **158** (2005).

[MG05b] B. Morgan and A. M. Green, *Directional statistics for WIMP direct detection II: 2-d read-out.*, Phys. Rev. D. **72** (2005).

[MGS05] B. Morgan, A. M. Green, and N. J. C. Spooner, *Directional statistics for WIMP direct detection experiments.*, Phys. Rev. D. **71** (2005).

[MJ00] K. V. Mardia and P. E. Jupp, *Directional statistics.*, Wiley Series in Probability and Statistics, Wiley, Chichester, 2000.

[ML18] K.V. Mardia and C. Ley, *Directional distributions.*, Wiley StatsRef: Statistics Reference Online (2018), 1–13.

[MMK23] I. C. Matheson, R. Malhotra, and J. T. Keane, *A von Mises–Fisher distribution for the orbital poles of the plutinos.*, MNRAS **522** (2023), 3298–3307.

[MP11] D. Marinucci and G. Peccati, *Random fields on the sphere.*, Mathematical Society Lecture Note Series 389, London, 2011.

[MPB+07] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio, *Spherical needlets for cosmic microwave background data analysis.*, MNRAS **383** (2007), 539–545.

[MPC17] Jr. RS. Menezes, C. Pigozzo, and S. Carneiro, *Testing cosmic isotropy with galaxies position angles distribution.*, Preprint, 2017.

[MPT12] Di Marizo, A. Panzera, and C. C. Taylor, *Smooth estimation of circular cumulative distribution functions and quantiles.*, J. Nonparametr. Stat. **24** (2012), 935–949.

[MQ95] M. Mayor and D. Queloz, *A Jupiter-mass companion to a solar-type star.*, Nature **378** (1995), 355–359.

[MR23] R. Malhotra and S. Roy, *Modeling the free inclinations of the classical Kuiper belt with the von Mises–Fisher distribution.*, Res. Notes AAS **7** (2023).

[MS78] K.V. Mardia and T.W. Sutton, *A model for cylindrical variables with applications.*, J. R. Stat Soc. Ser. B (Methodology) **40** (1978), no. 2, 229–233.

[MW93] B. G. Marsden and G. V. Williams, *Catalogue of cometary orbits (8th ed.).*, Minor Planet Center, Smithsonian Astrophysical Observatory, Cambridge, MA, 1993.

[N+09] C. B. Netterfield et al., *Blast: The mass function, lifetimes, and properties of intermediate mass cores from a 50 deg submillimeter galactic survey in Vela ($l \approx 265°$).*, ApJ **707** (2009).

[OG14] C. A. J. O'Hare and A. M. Green, *Directional detection of dark matter streams.*, Phys. Rev. D. **90** (2014).

[P+97] M. A. C. Perryman et al., *The Hipparcos and Tycho catalogues.*, European Space Agency, 1997.

[P+03] G. Paturel et al., *HYPERLEDA II. The homogenized HI data.*, A&A **412** (2003).

[P+10] G. L. Pilbratt et al., *Herschel space observatory.*, A&A **518** (2010).

[P+17] E. A. Petigura et al., *The California-Kepler Survey. I. High-resolution spectroscopy of 1305 stars hosting kepler transiting planets.*, AJ **154** (2017).

[Par12] H. S. Park, *Asymptotic behavior of the kernel density estimator from a geometric viewpoint.*, Commun. Stat. - Theory Methods **41** (2012), 3479–3496.

[Pee69] P.J.E. Peebles, *Cosmology.*, J. R. Astron. Soc. Can. **63** (1969), 4.

[PGP21] A. Pewsey and E. Garcia-Portugués, *Recent advances in directional statistics.*, TEST **30** (2021), 1–58.

[PL67] J. Pensado and J. F. Lahulla, *Orientactión de los planos orbitales y distribución de las órbitas de las estrellas dobles visuales.*, Bol. Astr. Obs **9** (1967), 21–35.

[PLK+97] M.A.C. Perryman, L. Lindegren, J. Kovalevsky, E. Hoeg, U. Bastian, P.L. Bernacca, M. Crézé, F. Donati, M. Grenon, M. Grewing, et al., *The HIPPARCOS catalogue.*, A&A **323** (1997), L49–L52.

[PM02] V. Patrangenaru and K. V. Mardia, *A bootstrap approach to Pluto's origin.*, J. Appl. Stat **29** (2002), 935–943.

[PNR13] A. Pewsey, M. Neuhäuser, and G. D. Ruxton, *Circular statistics in R.*, Oxford University Press, Oxford, 2013.

[Pou00] D. Pourbaix, *Resolved double-lined spectroscopic binaries: A neglected source of hypothesis-free parallaxes and stellar masses.*, A&AS **145** (2000), 215–222.

[PR02] M. Pätzold and H. Rauer, *Where are the massive close-in extrasolar planets?*, ApJ **568** (2002), L117–L120.

[Pro85] R. J. Protheroe, *A new statistic for the analysis of circular data in gamma-ray astronomy.*, In NASA Goddard Space Flight Center 19th Intern. Cosmic Ray Conf **3** (1985), 485–488.

[PV17] D. Paindaveine and T. Verdebout, *Inference on the mode of weak directional signals: Le Cam perspective on hypothesis testing near singularities.*, Ann. Statist. **45** (2017), 800–832.

[R+10] D. Raghavan et al., *A survey of stellar families: Multiplicity of solar-type stars.*, ApJS **190** (2010).

[R+11] G. Raskin et al., *HERMES: a high-resolution fibre-fed spectrograph for the Mercator telescope.*, A&A **526** (2011).

[Ray80] L. Rayleigh, *On the resultant of a large number of vibrations of the same pitch and of arbitrary phase.*, The London, Edinburgh, and Dublin Philosophical Magazine and J. Sci. **10** (1880), no. 60, 73–78.

[Ray19] ——, *On the problem of random vibrations, and of random flights in one, two, or three dimensions.*, The London, Edinburgh, and Dublin Philosophical Magazine and J. Sci. **37** (1919), no. 220, 321–347.

[RJS+17] F. Roelofs, M. D. Johnson, H. Shiokawa, S. S. Doeleman, and H. Falcke, *Quantifying intrinsic variability of Sagittarius A* using closure phase measurements of the event horizon telescope.*, ApJ **847** (2017).

[RW80] R. J. O. Reid and A. A. Watson, *Catalogue of highest energy cosmic rays.*, WDC-C **2** (1980).

[S+17] J. D. Soler et al., *The relation between the column density structures and the magnetic field orientation in the vela c molecular complex.*, A&A **603** (2017).

[S+22] ______, *The Galactic dynamics revealed by the filamentary structure in atomic hydrogen emission.*, A&A **662** (2022).

[Sch14] W. D. Schutte, *Non-parametric estimation of the off-pulse interval(s) of a pulsar light curve.*, Ph.D. thesis, North-West University, 2014.

[Sch22] ______, *SOPIE: Non-parametric estimation of the off-pulse interval of a pulsar.*, CRAN, 2022.

[SCW+74] L. A. Soderblom, C. D. Condit, R. A. West, B. M. Herman, and T. J. Kreidler, *Martian planetwide crater distributions: Implications for geologic history and surface processes.*, Icarus **22** (1974), 239–263.

[SDL+11] J. Schneider, C. Dedieu, P. LeSidaner, R. Savalle, and I. Zolotukhin, *Defining and cataloging exoplanets: the exoplanet.eu database.*, A&A **532** (2011).

[Sei13] T. Sei, *A Jacobian inequality for gradient maps on the sphere and its application to directional statistics.*, Commun. Stat. - Theory Methods **42** (2013), 2525–2542.

[Shi13] H. Shiokawa, *General - relativistic magnetohydrodynamics-simulations of black hole accretion disks: Dynamics and radiative properties.*, Ph.D. thesis, Univ. Illinois, 2013.

[SHLL97] S. W. Squyres, C. Howell, M. C. Liu, and J. J. Lissauer, *Investigation of crater "saturation" using spatial statistics.*, Icarus **125** (1997), 67–82.

[SHM+13] J. D. Soler, P. Hennebelle, P. G. Martin, Miville-Deschênes M. A., C. B. Netterfield, and L. M. Fissel, *An imprint of molecular cloud magnetization in the morphology of the dust polarized emission.*, ApJ **744** (2013).

[SJ01] S. Sarala and P. Jain, *A circular statistical method for extracting rotation measures.*, A&A **23** (2001), 137–141.

[SL20] A. Siraj and A. Loeb, *The case for an early solar binary companion.*, ApJL **899** (2020).

[SNFP23] P. Saavedra-Nieves and M. Fernández-Pérez, *Directional density-based clustering.*, Preprint, 2023.

[SNKB] M. Simard-Normandine, P. P. Kronberg, and S. Button, *Linear polarization of extragalactic radio sources at 3.71 and 11.1-centimeters.*, ApJSupp **46**.

[Sol19] J. D. Soler, *Using Herschel and Planck observations to delineate the role of magnetic fields in molecular cloud structure.*, A&A **629** (2019).

[SS16] W. D. Schutte and J. W. H. Swanepoel, *SOPIE: an R package for the non-parametric estimation of the off-pulse interval of a pulsar light curve.*, Mon. Not. R. Astron. Soc. **461** (2016), 627–640.

[Sti80] S.M. Stigler, *Stigler's law of eponymy.*, Transactions of the New York academy of sciences **39** (1980), no. 1 Series II, 147–157.

[Str88] M. F. Struble, *Position-Angle statistics of the brightest binary galaxies in Abell clusters.*, AJ **96** (1988).

[Str90] ______, *Position angle statistics of the first and second brightest galaxies in a sample of Coma-like Abell clusters.*, AJ **99** (1990).

[T+65] Y. Toyoda et al., *Studies of primiary cosmic rays in the energy position region* 10^{14} *eV to* 10^{17} *eV (Bolivian Air Shower Joint Experiment).*, PICRC **2** (1965), 708–711.

[TCF92] D. Trèvese, G. Cirimele, and P. Flin, *The orientation of galaxies in clusters.*, AJ **104** (1992).

[TI80] H. Trabara and M. Inoue, *A catalogue of linear polarization of radio sources.*, A&AS **39** (1980), 379–393.

[TP88] G. S. Tucker and J. B. Peterson, *The alignment of clusters with brightest member galaxies.*, AJ **95** (1988).

[TSR+03] R. Tomàs, D. Semikoz, G. G. Raffelt, M. Kachelrieß, and A. S. Dighe, *Supernova pointing with low- and high-energy neutrino detectors.*, Phys. Rev. D. **68** (2003).

[Tul88] R. B. Tully, *Nearby galaxies catalog.*, Cambridge, University Press, 1988.

[USS15] K. Uesu, K. Shimizu, and A. Sengupta, *A possibly asymmetric multivariate generalization of the Möbius distribution for directional data.*, J. Multivar. Anal. **134** (2015), 146–162.

[Val97] J. P. Vallée, *Observations of the magnetic fields inside and outside the Milky Way, starting with globules (~ 1 parsec), filaments, clouds, superbubbles, spiral arms, galaxies, superclusters, and ending with the cosmological universe's background surface (at ~ 8 Teraparsecs).*, Fundam. Cosmic Phys. **19** (1997), 1–89.

[VL07] F. Van Leeuwen, *Hipparcos, the new reduction of the raw data.*, Springer, 2007.

[vM18] R. von Mises, *Über die "Ganzzahligkeit" der Atomgewichte und verwandete Fragen.*, Phys. Z. **19** (1918), 490.

[Wat61] George S Watson, *Goodness-of-fit tests on a circle*, Biometrika **48** (1961), no. 1/2, 109–114.

[Wat65] G.S. Watson, *Equatorial distribution on a sphere.*, Biometrika (1965), no. 52, 193–201.

[Wat83] G. S. Watson, *Statistics on spheres.*, New York, Wiley, 1983.

[WH83] C. E. Worely and W. D. Heintz, *Fourth catalog of orbits of visual binary stars.*, Publ. Us Naval Obs XXIV Part VII (1983), 1–53.

[WT99] P. Wiegert and S. Tremaine, *The evolution of long-period comets.*, Icarus **137** (1999), 84–121.

[WUP+86a] M. M. Winn, J. Ulrichs, L. S. Peak, C. B. McCusker, and L. Horton, *Catalogue of highest energy cosmic rays.*, WDC-C for Cosmic Rays **2** (1986).

[WUP+86b] ———, *Catalogue of highest energy cosmic rays.*, J. Phys. G: Nucl. Phys. **2** (1986), 675.

[YCCH16] M.-S. Yang, S.-J. Chang-Chien, and W.-L. Hung, *An unsupervised clustering algorithm for data on the unit hypersphere.*, Appl. Soft. Comput. **42** (2016), 290–313.

[YCCH17] ———, *Learning-based EM clustering for data on the unit hypersphere with application to exoplanet data.*, Appl. Soft Comput. **60** (2017), 101–114.

[ZGPM21] P. Zoubouloglou, E. Garcia-Portugués, and J. S. Marron, *Scaled torus principal component analysis.*, J. Comput. Graph. Stat., 2021.

[ZM02] S. Zucker and T. Mazeh, *On the mass-period correlation of the extrasolar planets.*, ApJ **568** (2002), L113–L116.

[ZRS83] Y. B. Zeldovich, A. A. Ruzmaikin, and D. D. Sokoloff, *Magnetic fields in astrophysics.*, Gordon and Breach Science Publishers, New York, 1983.

[ZSLD03] K. Zahnle, P. Schenk, H. Levison, and L. Dones, *Cratering rates in the outer Solar System.*, Icarus **163** (2003), 263–289.

[ZYC21] Y. Zhang and Chen Y.-C., *Kernel smoothing, mean shift, and their learning theory with directional data.*, JMLR **22** (2021), 1–92.