

Statistica Sinica Preprint No: SS-2024-0106	
Title	On Runs Tests for Directional Data and Their Local and Asymptotic Optimality Properties
Manuscript ID	SS-2024-0106
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0106
Complete List of Authors	Maxime Boucher, Christian Francq, Yuichi Goto and Thomas Verdebout
Corresponding Authors	Thomas Verdebout
E-mails	tverdebout@gmail.com

On runs tests for directional data and their local and asymptotic optimality properties

Maxime Boucher[†], Christian Francq^{*}, Yuichi Goto^{**} and Thomas Verdebout[†]

*Kyushu University^{**}, Université de Lille and CREST (Paris)^{*}*

and Université libre de Bruxelles (ULB)[†]

Abstract: In the present paper, we tackle the problem of detecting serial correlation in directional data. We introduce a concept of runs properly adapted to the directional context. We then show that tests based on the latter runs enjoy some local and asymptotic optimality properties against local alternatives with serial dependence. We evaluate the finite-sample performances of our tests using Monte Carlo simulations and show their usefulness on a real data illustration that involves the analysis of sunspots locations for various solar cycles.

Key words and phrases: directional data, runs tests, randomness, hypothesis testing

1. Introduction

The problem of testing randomness of a series of observations is one of the most important problems in time series analysis. To tackle the problem,

runs tests are nowadays very classical tools in statistics. In particular the runs tests for randomness of Wald and Wolfowitz (1940) is very popular. More recently, runs have been used for instance in Henze and Penrose (1999) and Biswas et al. (2014) to compare samples, in McWilliams (1990), Corzo and Babativa (2013) and Dyckerhoff et al. (2015) to test for symmetry, in Dufour et al. (1998) to test for nonhomogeneous white noise and in Paindaveine (2009), Cho and White (2011) and Hentati-Kaffel and De Peretti (2015) to test for randomness.

Classical runs tests typically reject the null hypothesis of randomness when the number of *runs* in the sequence is too large or too small. In a sequence of observations, a *run* is defined as a consecutive series of observations with the same sign. From a univariate sample $X_1, \dots, X_n$, the number of runs can be computed from the quantity $\sum_{t=2}^n U_t(\theta)U_{t-1}(\theta)$, where $U_t(\theta) := \text{sign}(X_t - \theta)$ is the sign of a centered version of X_t ; θ plays the role of a location parameter here. More precisely if $X_1, \dots, X_n$ are mutually independent random variables with median θ , we have that

$$\frac{1}{\sqrt{n-1}} \sum_{t=2}^n U_t(\theta)U_{t-1}(\theta) = \frac{N_n(\theta) - \mathbb{E}[N_n(\theta)]}{\sqrt{n-1}}, \quad (1.1)$$

where $N_n(\theta) := 1 + \sum_{t=2}^n \mathbb{I}[U_t(\theta) \neq U_{t-1}(\theta)]$ is the number of runs associated with $X_1, \dots, X_n$. Based on a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ of p -dimensional random vectors, a multivariate extension of the notion of runs has been provided

in Marden (1999) who proposed to replace the univariate signs in (1.1) by spatial signs $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{U}_{\boldsymbol{\theta}}(\mathbf{X}_n)$, where $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{X}) := \mathbf{X} - \boldsymbol{\theta} / \|\mathbf{X} - \boldsymbol{\theta}\|$ to consider a test statistic of the form

$$R_{1m}^{(n)}(\boldsymbol{\theta}) := \frac{1}{\sqrt{n-1}} \sum_{t=2}^n \mathbf{U}_{\boldsymbol{\theta}}'(\mathbf{X}_t) \mathbf{U}_{\boldsymbol{\theta}}(\mathbf{X}_{t-1}), \quad (1.2)$$

that is a measure of closeness of the successive spatial signs of the observations. Elliptical extensions of the Marden (1999) runs have been studied more recently by Paindaveine (2009), where signs of sphericized observations are used.

Directional data consist in observations that are directions/unit vectors. In most cases, these observations therefore lie on the circumference of the unit circle of $\mathbb{R}^2$ (one then speaks of circular data) or on the surface of the unit hypersphere $\mathcal{S}^{p-1} := \{\mathbf{s} \in \mathbb{R}^p, \mathbf{s}'\mathbf{s} = 1\}$ of $\mathbb{R}^p$. Although we will focus in this work on data on hyperspheres (such as the sunspots data analyzed below), directional data also include data on the torus (product of two circles or spheres) and the cylinder (product of $\mathbb{R}^p$ with a circle or sphere), but also on other manifolds such as the Stiefel or the Grassmann manifolds. Directional datasets are encountered in various fields, such as meteorology (wind direction), biology (animal migration patterns) and many more. Analyzing and interpreting directional data requires specialized techniques that account for the curved nature of the observation space.

Nonparametric methods recently designed specifically for directional data include the independence tests of García-Portugués et al. (2024), optimal transport-based methods in Hallin et al. (2024) and methods for regression in García-Portugués et al. (2016); Di Marzio et al. (2017); Meilán-Vila et al. (2020); Alonso-Pena et al. (2021) and Alonso-Pena et al. (2024) to cite only a few. For a general overview of the topic, we refer the reader to Mardia and Jupp (1999), Rao and SenGupta (2001) and Ley and Verdebout (2017). To the best of our knowledge, runs for directional data have never been considered before. In the present work, we define a concept of directional runs. We show that tests based on our concept of directional runs enjoy some local and asymptotic optimality properties to test for iidness (the observations $\mathbf{X}_1, \dots, \mathbf{X}_t$ are i.i.d.) against alternatives under which some serial dependence is introduced. More precisely, we show that our runs provide tests that are locally and asymptotically optimal against alternatives called *Tangent Markov* local alternatives.

The paper is organized as follows: in Section 2, we define precisely our runs tests. In Section 3, we provide some asymptotic properties of tests based on runs; in particular we show that they enjoy some local and asymptotic optimality properties against Tangent Markov alternatives we define. Generalized versions of our tests are presented in Section 4. Some numeri-

cal illustrations of the performances of our tests are proposed in Section 5. Finally, we close the paper with some conclusions we can draw on sunspots data for various solar cycles in Section 6. A supplementary material collects the proofs of the various results and complementary informations on the real data analysis of sunspot locations.

2. Directional runs tests

Assume that we have a sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ of directions taking values on the unit hypersphere $\mathcal{S}^{p-1}$ of $\mathbb{R}^p$. For some $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$, the classical tangent-normal decomposition of $\mathbf{X}_t$ around $\boldsymbol{\theta}$ reads (below $\|\mathbf{v}\|$ is the Euclidean norm of $\mathbf{v}$)

$$\begin{aligned}\mathbf{X}_t &= (\mathbf{X}_t' \boldsymbol{\theta}) \boldsymbol{\theta} + (\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t \\ &= (\mathbf{X}_t' \boldsymbol{\theta}) \boldsymbol{\theta} + \|(\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t\| \frac{(\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t}{\|(\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t\|},\end{aligned}\quad (2.1)$$

where letting $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ be a $p \times (p-1)$ semi-orthogonal matrix such that $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} \boldsymbol{\Gamma}_{\boldsymbol{\theta}}' = \mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}'$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}' \boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \mathbf{I}_{p-1}$, a notion of multivariate sign $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$ for directions can be naturally defined via

$$\frac{(\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t}{\|(\mathbf{I}_p - \boldsymbol{\theta} \boldsymbol{\theta}') \mathbf{X}_t\|} = \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \frac{\boldsymbol{\Gamma}_{\boldsymbol{\theta}}' \mathbf{X}_t}{\|\boldsymbol{\Gamma}_{\boldsymbol{\theta}}' \mathbf{X}_t\|} =: \boldsymbol{\Gamma}_{\boldsymbol{\theta}} \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t); \quad (2.2)$$

see Figure 1 for an illustration. Note that $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$ is a random vector taking values on a unit sphere which is orthogonal to $\boldsymbol{\theta}$. We tacitly assume

throughout that the distribution of the $\mathbf{X}_t$'s does not charge $\boldsymbol{\theta}$ so that the $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$'s are well-defined almost surely. In the circular case $p = 2$ with $\mathbf{X}_t = (\cos(U_t), \sin(U_t))$ for some random angle U_t , $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$ is the (univariate) sign of $\sin(U_t - \theta)$, where $\boldsymbol{\theta} = (\cos(\theta), \sin(\theta))'$ for some angle θ . Tests for location precisely based on this concept of signs have been proposed in Schach (1969) for the circular case while tests based on the signs in (2.2) for the (hyper)spherical case have been studied for instance in Paindaveine and Verdebout (2016) and García-Portugués et al. (2020). Using such directional signs, it is very natural (following (1.2)) to consider a runs test statistic of the form

$$R_{1d}^{(n)}(\boldsymbol{\theta}) := \frac{1}{\sqrt{n-1}} \sum_{t=2}^n \mathbf{S}_{\boldsymbol{\theta}}'(\mathbf{X}_t) \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_{t-1}). \quad (2.3)$$

It follows from the discussion above that this concept of runs reduces to classical univariate concept of runs for random angles in the ($p = 2$) circular case (see (1.1)). Under the hypothesis of iidness of $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ with $E[\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1)] = \mathbf{0}$ ($\boldsymbol{\theta}$ can be seen as a directional median), the central limit theorem for 2-dependent stationary processes directly entails that the standardized version $s_n^{-1/2} R_{1d}^{(n)}(\boldsymbol{\theta})$ of $R_{1d}^{(n)}(\boldsymbol{\theta})$ in (2.3), where

$$s_n(\boldsymbol{\theta}) := \text{tr} \left[\left(n^{-1} \sum_{t=1}^n \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t) \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)' \right)^2 \right], \quad (2.4)$$

converges weakly to a standard Gaussian variable. As a result, nonparametric tests can be obtained using $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$. For instance the test $\phi_1^{(n)}$ that rejects the null hypothesis of iidness at the asymptotic level α when

$$s_n^{-1}(\boldsymbol{\theta})(R_{1d}^{(n)}(\boldsymbol{\theta}))^2 > \chi_{1,1-\alpha}^2,$$

where $\chi_{\nu,\beta}^2$ is the quantile of order β of the chi-square distribution with ν degrees of freedom, is an asymptotically valid test for the problem. Obviously tests based on $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$ will be able to detect serial correlation of order 1 only; generalized runs tests that can detect serial correlation of larger order are discussed in Section 4. In the next Section, we study some asymptotic properties of tests based on $R_{1d}^{(n)}(\boldsymbol{\theta})$.

3. Theoretical guarantees

In the present section, our objective is to show that tests based $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$ enjoy nice asymptotic properties. We show below that this is particularly the case in the vicinity of rotational symmetry. We say that $\mathbf{X}_t$ is rotationally symmetric around $\boldsymbol{\theta}$ if for any rotation $\mathbf{O}$ such that $\mathbf{O}\boldsymbol{\theta} = \boldsymbol{\theta}$, $\mathbf{O}\mathbf{X}_t$ has the same distribution as $\mathbf{X}_t$. If $\mathbf{X}_t$ is absolutely continuous with respect to

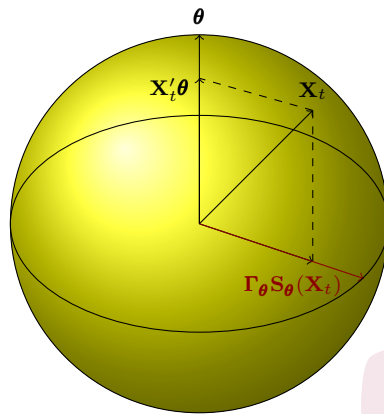


Figure 1: The *tangent-normal* decomposition of $\mathbf{X}_t$ with respect to $\boldsymbol{\theta}$. The 3-dimensional unit vector $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$ belongs to the orthogonal complement to $\text{span}(\boldsymbol{\theta})$ ($\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$ is bivariate).

the surface area measure on $\mathcal{S}^{p-1}$, then its density is of the form

$$\mathbf{x} \mapsto f_{\boldsymbol{\theta},g}(\mathbf{x}) = c_{p,g} g(\mathbf{x}'\boldsymbol{\theta}), \quad (3.1)$$

where $c_{p,g}(> 0)$ is a normalizing constant and $g : [-1, 1] \rightarrow [0, \infty)$ is referred to as an *angular function*; below we write $P_{\boldsymbol{\theta},g}^{(n)}$ for the joint distribution of the $\mathbf{X}_t$'s. Note that the very classical von Mises distribution is obtained by taking $g(u) = \exp(\kappa u)$ for some positive concentration parameter κ . We also have that if $\mathbf{X}_1, \dots, \mathbf{X}_n$ are i.i.d. with a rotationally symmetric distribution with parameters g and $\boldsymbol{\theta}$, $v_{\boldsymbol{\theta}}(\mathbf{X}_1) := \mathbf{X}_1'\boldsymbol{\theta}, \dots, v_{\boldsymbol{\theta}}(\mathbf{X}_n) := \mathbf{X}_n'\boldsymbol{\theta}$ are i.i.d. with density

$$v \rightarrow \tilde{g}_p(v) := \omega_{p-1} c_{p,g} (1 - v^2)^{(p-3)/2} g(v), \quad (3.2)$$

where $\omega_{p-1} := 2\pi^{\frac{p-1}{2}}/\Gamma(\frac{p-1}{2})$ is the surface area of $\mathcal{S}^{p-2}$. They are moreover independent of the signs $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ defined above. Finally, the signs $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ are i.i.d. uniformly distributed over $\mathcal{S}^{p-2}$ under $P_{\boldsymbol{\theta},g}^{(n)}$. The following result confirms the comment just below (2.4) under rotational symmetry around $\boldsymbol{\theta}$.

Proposition 1. *Under $P_{\boldsymbol{\theta},g}^{(n)}$, $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta}) = (p-1)^{1/2}R_{1d}^{(n)}(\boldsymbol{\theta}) + o_P(1)$ converges weakly to a standard normal random variable.*

See the supplement for a proof. As we will see below, tests based on $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$ enjoy some local and asymptotic optimality properties against alternatives under which the signs $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ are serially correlated. We also show that the location parameter $\boldsymbol{\theta}$ in $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$ can be replaced by root- n consistent estimators without any asymptotic cost under $P_{\boldsymbol{\theta},g}^{(n)}$ (and therefore under contiguous alternatives). Consider as in (2.1) the tangent normal decomposition

$$\mathbf{X}_t = v_{\boldsymbol{\theta}}(\mathbf{X}_t)\boldsymbol{\theta} + \sqrt{1 - v_{\boldsymbol{\theta}}^2(\mathbf{X}_t)}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t),$$

$t = 1, \dots, n$. As mentioned already, under rotational symmetry (with parameters g and $\boldsymbol{\theta}$), the $v_{\boldsymbol{\theta}}(\mathbf{X}_t)$'s are i.i.d. with density (3.2) over $[-1, 1]$ and are independent of the $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)$'s. In the sequel we need to assume some regularity conditions on angular functions g . More precisely, we will

restrict to the collection $\mathcal{G}_a$ of positive monotone nondecreasing angular functions $g : [-1, 1] \rightarrow (0, \infty)$ that are absolutely continuous (with almost everywhere derivative $\dot{g}$) and for which

$$\mathcal{J}_p(g) := \int_{-1}^1 \varphi_g^2(t)(1-t^2)\tilde{g}_p(t) dt \quad (3.3)$$

is finite, where $\varphi_g := \dot{g}/g$. These regularity assumptions ensure a proper identifiability of the parameter $\boldsymbol{\theta}$ together with the fact that the corresponding sequence of models is local and asymptotic normal (LAN) (see for instance Ley et al. (2013) or Paindaveine and Verdebout (2017)).

Assume now that the joint distribution of $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ is not a product of uniform distributions over $\mathcal{S}^{p-2}$ but is given by a density of the form

$$(\mathbf{s}_1, \dots, \mathbf{s}_n) \rightarrow c_{\lambda}^n \exp\left(\lambda \left(\sum_{t=2}^n \mathbf{s}'_t \mathbf{s}_{t-1}\right)\right) \quad (3.4)$$

with respect to the surface area measure over $(\mathcal{S}^{p-2})^n$ for some normalizing constant c_{λ}^n obtained through

$$c_{\lambda}^{-1} := \int_{-1}^1 \exp(\lambda u)(1-u^2)^{(p-3)/3} du.$$

Following Accardi et al. (1987), the marginal distributions of $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ with density (3.4) are uniform over $\mathcal{S}^{p-2}$ but the conditional density $f_{\mathbf{S}_t|\mathbf{S}_{t-1}=\mathbf{s}_{t-1}}$ of $\mathbf{S}_t|\mathbf{S}_{t-1}=\mathbf{s}_{t-1}$ is von Mises with location parameter $\mathbf{s}_{t-1}$ and concentration parameter λ . In the von Mises model, λ plays the role of a concen-

tration parameter that drives the probability mass in the vicinity of the location parameter while below it will influence the dependence between observations; recall that a von Mises distribution is obtained by taking $g(u) = \exp(\kappa u)$ as an angular function in (3.1) for some positive concentration κ . In the following result we obtain the shape of the joint density of random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ when their signs have a joint density (3.4).

Theorem 1. *Assume that $v_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, v_{\boldsymbol{\theta}}(\mathbf{X}_n)$ are (i) i.i.d. with density*

$$v \rightarrow \omega_{p-1} c_{p,g} (1 - v^2)^{(p-3)/2} g(v)$$

with respect to the Lebesgue measure on $[-1, 1]$ and (ii) independent of $\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ jointly distributed with density (3.4). Then $\text{vec}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ has density

$$\text{vec}(\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto c_{p,g}^n c_{\lambda}^n \exp\left(\lambda \left(\sum_{t=2}^n \mathbf{S}'_{\boldsymbol{\theta}}(\mathbf{x}_t) \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{x}_{t-1})\right)\right) \prod_{t=1}^n g(v_{\boldsymbol{\theta}}(\mathbf{x}_t)) \quad (3.5)$$

with respect to the surface area measure over $\mathcal{S}^{p-1}$.

See the supplement for a proof. Note that when $\lambda = 0$, the joint distribution of $\text{vec}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is simply $P_{\boldsymbol{\theta},g}^{(n)}$ defined below (3.2). In the sequel we write $(\mathbf{X}_1, \dots, \mathbf{X}_n) \sim P_{\boldsymbol{\theta},\lambda,g}^{(n)}$ when $\text{vec}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ has density (3.5) and call the corresponding distribution the Tangent Markov distribution with location

$\boldsymbol{\theta}$, angular function g and dependence parameter $\lambda \geq 0$. Obviously $P_{\boldsymbol{\theta},0,g}^{(n)}$ and $P_{\boldsymbol{\theta},g}^{(n)}$ coincide and both notations are used in the rest of the paper. We now show that the runs tests defined in Section 2 enjoy some local and asymptotic optimality properties for testing $\mathcal{H}_0 : \lambda = 0$ against $\mathcal{H}_1 : \lambda > 0$. Consider a local perturbation $(n^{-1/2}\ell_n, \boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}_n)$ of a null value of the parameter $(0, \boldsymbol{\theta})$, where the sequence $\boldsymbol{\tau}_n$ in $\mathbb{R}^p$ converges to $\boldsymbol{\tau} \in \mathbb{R}^p$ and ℓ_n is a positive real bounded sequence. Of course, it is assumed that $\boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}_n$ belongs to $\mathcal{S}^{p-1}$ for any n , which imposes that

$$1 = (\boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}_n)'(\boldsymbol{\theta} + n^{-1/2}\boldsymbol{\tau}_n) = 1 + 2n^{-1/2}\boldsymbol{\theta}'\boldsymbol{\tau}_n + n^{-1}\|\boldsymbol{\tau}_n\|^2,$$

or equivalently that $\boldsymbol{\theta}'\boldsymbol{\tau}_n = -\frac{1}{2}n^{-1/2}\|\boldsymbol{\tau}_n\|^2$. In the next result, we study the local log-likelihood ratio

$$\Lambda^{(n)} = \log \frac{dP_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}_n, n^{-1/2}\ell_n, g}^{(n)}}{dP_{\boldsymbol{\theta},0,g}^{(n)}}$$

of a perturbed distribution $P_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}_n, n^{-1/2}\ell_n, g}^{(n)}$ with respect to a distribution that belongs to the null hypothesis $P_{\boldsymbol{\theta},0,g}^{(n)}$. We have the following result.

Theorem 2. *Letting $\mathbf{v}_n := (\ell_n, \boldsymbol{\tau}_n')'$ as described above, we have that*

$$\Lambda^{(n)} = \mathbf{v}_n' \boldsymbol{\Delta}^{(n)} - \frac{1}{2}\mathbf{v}_n' \boldsymbol{\Gamma} \mathbf{v}_n + o_P(1)$$

as $n \rightarrow \infty$ under $P_{\boldsymbol{\theta},0,g}^{(n)}$, where letting

$$\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} := n^{-1/2} \sum_{t=1}^n \varphi_g(v_{\boldsymbol{\theta}}(\mathbf{X}_t))(1 - v_{\boldsymbol{\theta}}^2(\mathbf{X}_t))^{1/2} \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t),$$

and

$$\Delta_{\lambda}^{(n)} := n^{-1/2} \sum_{t=2}^n \mathbf{S}'_{\boldsymbol{\theta}}(\mathbf{X}_t) \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_{t-1}),$$

the central sequence $\boldsymbol{\Delta}^{(n)} := (\Delta_{\lambda}^{(n)}, (\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)})')'$ converges weakly (still under $P_{\boldsymbol{\theta},0,g}^{(n)}$) to a Gaussian vector with mean zero and covariance matrix (see (3.3))

$$\boldsymbol{\Gamma} := \text{diag} \left(\frac{1}{(p-1)}, \mathcal{J}_p(g) \mathbf{I}_{p-1} \right).$$

See the supplement for a proof. The Local Asymptotic Normality (LAN) theorem above is very important. Indeed, we readily see that the dependence (λ -)part of central sequence $\Delta_{\lambda}^{(n)}$ is proportional to $R_{1d}^{(n)}(\boldsymbol{\theta})$. As described below, this has important consequences in terms of local and asymptotic optimality for tests based on $R_{1d}^{(n)}(\boldsymbol{\theta})$. It directly follows from Theorem 2 above (and in particular from the block-diagonal structure of the Fisher information matrix of the LAN property) that

- (i) $R_{1d}^{(n)}(\hat{\boldsymbol{\theta}}) - R_{1d}^{(n)}(\boldsymbol{\theta}) = o_P(1)$ as $n \rightarrow \infty$ provided that the estimator $\hat{\boldsymbol{\theta}}$ (with values in $\mathcal{S}^{p-1}$) is part of a sequence that is: (i) *root-n consistent* under any $g \in \mathcal{G}'$ for some $\mathcal{G}' \subset \mathcal{G}_a$, i.e., $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_P(1)$ under $\bigcup_{g \in \mathcal{G}'} \{P_{\boldsymbol{\theta},g}^{(n)}\}$; (ii) *locally and asymptotically discrete*, i.e., for all $\boldsymbol{\theta}$ and for all $C > 0$, there exists a positive integer $M = M(C)$ such that the number of possible values of $\hat{\boldsymbol{\theta}}$ in $\{\mathbf{t} \in \mathcal{S}^{p-1} : \sqrt{n} \|\mathbf{t} - \boldsymbol{\theta}\| \leq C\}$ is bounded by M , uniformly as $n \rightarrow \infty$.

(ii) for testing $\mathcal{H}_0 : \lambda = 0$ against $\mathcal{H}_1 : \lambda > 0$, the locally and asymptotically most powerful test $\phi_{\text{opt}}^{(n)}$ rejects the null hypothesis at the asymptotic level α when

$$(p-1)^{1/2} \Delta_{\lambda}^{(n)} > z_{1-\alpha},$$

where z_{β} denotes the quantile of order β of the standard Gaussian distribution.

Point (i) above directly follows from the block-diagonal structure of the Fisher information matrix in Theorem 2. Indeed, the latter block-diagonal structure implies that the non-specification of the location parameter does not have any asymptotic cost when inference on λ is considered. Since, we also have that $s_n(\hat{\theta}) - s_n(\theta)$ is $o_P(1)$ under $P_{\theta,0,g}^{(n)}$ as $n \rightarrow \infty$, the Slutsky Lemma and Proposition 1 directly entail that

$$s_n^{-1/2}(\hat{\theta}) R_{1d}^{(n)}(\hat{\theta}) = s_n^{-1/2}(\theta) R_{1d}^{(n)}(\theta) + o_P(1) = (p-1)^{1/2} R_{1d}^{(n)}(\theta) + o_P(1) \quad (3.6)$$

as $n \rightarrow \infty$ under $P_{\theta,0,g}^{(n)}$. Note that the local discreteness property is a purely technical requirement in point (i) with little practical implications since for fixed n ; any estimate can be considered part of a locally and asymptotically discrete sequence of estimators. Point (ii) above together with (3.6) entail that the locally and asymptotically optimal test for testing $\mathcal{H}_0 : \lambda = 0$ against $\mathcal{H}_1 : \lambda > 0$ is $\phi_{\text{opt}}^{(n)}$ and not $\phi_1^{(n)}$. The testing problem considered here

is one-sided so that this is not really surprising. The test $\phi_1^{(n)}$ is actually the “two-sided version” of $\phi_{\text{opt}}^{(n)}$. The Tangent Markov model defined above is such that $\mathbf{S}_\theta(\mathbf{X}_t)|\mathbf{S}_\theta(\mathbf{X}_{t-1}) = \mathbf{s}_{t-1}$ is von Mises with location parameter $\mathbf{s}_{t-1}$ and concentration λ and therefore conditionally on $\mathbf{S}_\theta(\mathbf{X}_{t-1})$, the unit vector $\mathbf{S}_t$ belongs to the hemisphere $\{\mathbf{u} \in \mathcal{S}^{p-2}, \mathbf{u}'\mathbf{S}_\theta(\mathbf{X}_{t-1}) \geq 0\}$ with more probability than to the complementary hemisphere $\{\mathbf{u} \in \mathcal{S}^{p-2}, \mathbf{u}'\mathbf{S}_\theta(\mathbf{X}_{t-1}) < 0\}$ and can be seen as positively correlated with $\mathbf{S}_\theta(\mathbf{X}_{t-1})$ in that sense. The test $\phi_1^{(n)}$ (two-sided test) will clearly perform better than $\phi_{\text{opt}}^{(n)}$ (one-sided test) under alternatives allowing for both positive and negative dependence. A Tangent Markov model with $\lambda \in \mathbb{R}$ rather than $\lambda \in \mathbb{R}^+$ is also perfectly valid and would allow for both positive and negative dependence in the sense described just above. In such a model, it is easy to show that $\phi_1^{(n)}$ is locally and asymptotically maximin for testing $\mathcal{H}_0 : \lambda = 0$ against $\mathcal{H}_1 : \lambda \neq 0$. Note that a test ϕ^* is called maximin in the class $\mathcal{C}_\alpha$ of level- α tests for some null hypothesis $\mathcal{H}_0$ against the alternative $\mathcal{H}_1$ if (i) ϕ^* has level α and (ii) the power of ϕ^* is such that

$$\inf_{P \in \mathcal{H}_1} E_P[\phi^*] \geq \sup_{\phi \in \mathcal{C}_\alpha} \inf_{P \in \mathcal{H}_1} E_P[\phi].$$

For a definition of locally and asymptotically maximin tests, see for instance Chapter 5 of Ley and Verdebout (2017).

4. Generalized directional runs tests

In the present section, we discuss extensions of the tests proposed in the previous sections. Following Dufour et al. (1998), a generalized version of the test $\phi_1^{(n)}$ can be obtained by considering generalized runs of the form

$$R_{hd}^{(n)}(\boldsymbol{\theta}) := \frac{1}{\sqrt{n-h}} \sum_{t=h+1}^n \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t)' \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_{t-h}).$$

Such generalized runs are able to detect serial dependence of order h . Based on $R_{hd}^{(n)}(\boldsymbol{\theta})$ above, a test that can detect serial dependence until lag H say can be based on the test statistic

$$s_n^{-1}(\boldsymbol{\theta}) \sum_{h=1}^H (R_{hd}^{(n)}(\boldsymbol{\theta}))^2.$$

The following result summarizes some of its asymptotic properties.

Proposition 2. *The test statistic $s_n^{-1}(\boldsymbol{\theta}) \sum_{h=1}^H (R_{hd}^{(n)}(\boldsymbol{\theta}))^2$ converges weakly to*

(i) a chi-square random variable with H degrees of freedom under $P_{\boldsymbol{\theta},0,g}^{(n)}$

and

(ii) a chi-square random variable with H degrees of freedom and non-

centrality parameter $(p-1)^{-1}\ell^2$ under $P_{\boldsymbol{\theta},n^{-1/2}\ell_n,g}^{(n)}$, where $\ell := \lim_{n \rightarrow \infty} \ell^{(n)}$.

See the supplement for a proof. Following Proposition 2 above, a natural generalized runs test $\phi_H^{(n)}$ rejects the null hypothesis at the nominal level α when

$$s_n^{-1}(\boldsymbol{\theta}) \sum_{h=1}^H (R_{hd}^{(n)}(\boldsymbol{\theta}))^2 > \chi_{H,1-\alpha}^2.$$

Note that as in $s_n^{-1/2}(\boldsymbol{\theta})R_{1d}^{(n)}(\boldsymbol{\theta})$, replacing $\boldsymbol{\theta}$ in $s_n^{-1/2}(\boldsymbol{\theta})R_{hd}^{(n)}(\boldsymbol{\theta})$ by an estimator $\hat{\boldsymbol{\theta}}$ satisfying the same assumptions as the ones described below Theorem 2 has no asymptotic cost under $P_{\boldsymbol{\theta},0,g}^{(n)}$ as well as under contiguous alternatives. In the next Section, we compare the various tests proposed in this work through Monte Carlo simulations.

5. Monte Carlo study

In this section, we investigate the finite-sample performances of the proposed tests through two Monte Carlo exercises. In the first exercise, we generated $N = 2,500$ mutually independent random samples of the form

$$\mathbf{X}_{i;\ell}, \quad i = 1, \dots, n, \quad \ell = 0, 0.5, 1, \dots, 5,$$

with values in $\mathcal{S}^2$. The $\mathbf{X}_{i;\ell}$'s follow a Tangent Markov model (defined in (3.5)) with location $\boldsymbol{\theta} := (1, 0, 0)'$, angular function $t \mapsto g_1(t) := \exp(5t)$, and dependence parameter $\lambda_\ell := \ell/\sqrt{n}$. The value $\ell = 0$ corresponds to the null hypothesis of iidness, whereas $\ell = 0.5, \dots, 5$ provide increasingly severe

alternatives with serial dependence. For each replication, we performed, at asymptotic level $\alpha = .05$, the tests $\phi_1^{(n)}$, $\phi_2^{(n)}$, $\phi_3^{(n)}$ and $\phi_{\text{opt}}^{(n)}$ (respectively based on $R_{1d}^{(n)}(\boldsymbol{\theta})$, $R_{2d}^{(n)}(\boldsymbol{\theta})$, $R_{3d}^{(n)}(\boldsymbol{\theta})$ and $(p-1)^{1/2}\Delta_\lambda^{(n)}$) computed with (i) the correctly specified $\boldsymbol{\theta}$ and (ii) the classical spherical mean estimator (the arithmetic mean of the observations divided by its norm) of $\boldsymbol{\theta}$. In Figure 2, we provide the rejection frequencies of the various tests for the different values of the sample size $n \in \{250, 500, 750, 1000\}$. Inspection of Figure 2 reveals that, as expected, the larger ℓ , the larger the empirical power. We also observe that as n increases, the empirical rejection frequencies of the tests with a fixed $\boldsymbol{\theta}$ or with an estimated version of $\boldsymbol{\theta}$ share the same values. This could also have been expected since the estimation of $\boldsymbol{\theta}$ has no asymptotic cost here as discussed below Theorem 2. In Figure 3, we provide the same rejection frequencies of the tests $\phi_1^{(n)}$, $\phi_2^{(n)}$, $\phi_3^{(n)}$ and $\phi_{\text{opt}}^{(n)}$ computed with the correctly specified $\boldsymbol{\theta}$ together with the theoretical asymptotic powers of the tests obtained with Theorem 2. Inspection of Figure 3 clearly reveals that the empirical rejection frequencies almost coincide with the theoretical asymptotic power curves; our asymptotic results are therefore corroborated.

In a second simulation exercise, we generated $N = 2500$ mutually random samples of random vectors

$$\mathbf{X}_{i;\ell}, \quad i = 1, \dots, n, \quad \ell = 0, 0.5, 1, \dots, 5,$$

with values in $\mathcal{S}^2$ where $\mathbf{X}_{1;\ell}$ is uniformly distributed over $\mathcal{S}^2$ and for $i = 2, \dots, n$, $\mathbf{X}_{i;\ell} | \mathbf{X}_{i-1;\ell} \sim \text{vMF}(\mathbf{X}_{i-1;\ell}, \ell/\sqrt{n})$ where $\text{vMF}(\boldsymbol{\mu}, \kappa)$ denotes the von Mises Fisher distribution with location parameter $\boldsymbol{\mu}$ and concentration parameter κ . At each replication, we performed at the nominal level $\alpha = .05$ the same tests as in the first simulation exercise; that is we performed two versions of $\phi_1^{(n)}$, $\phi_2^{(n)}$, $\phi_3^{(n)}$ and $\phi_{\text{opt}}^{(n)}$ (one with $\boldsymbol{\theta} = (1, 0, 0)$ and one with an estimated version of $\boldsymbol{\theta}$) at the nominal level $\alpha = .05$. In Figure 4, we provide the rejection frequencies of the various tests for the different values of the sample size $n \in \{250, 500, 750, 1000\}$. Inspection of Figure 4 reveals that our runs tests are clearly able to detect such type of serially correlated observations. Using an estimated version of $\boldsymbol{\theta}$ or a fixed $\boldsymbol{\theta}$ does not result in a significant gap between the empirical powers.

6. Real data illustration

Several phenomena can be observed on top of the sun underlying structure. Sunspots are among the most important aspects of the global solar activity whose effects, among others, may affect Earth's long-term climate (see,

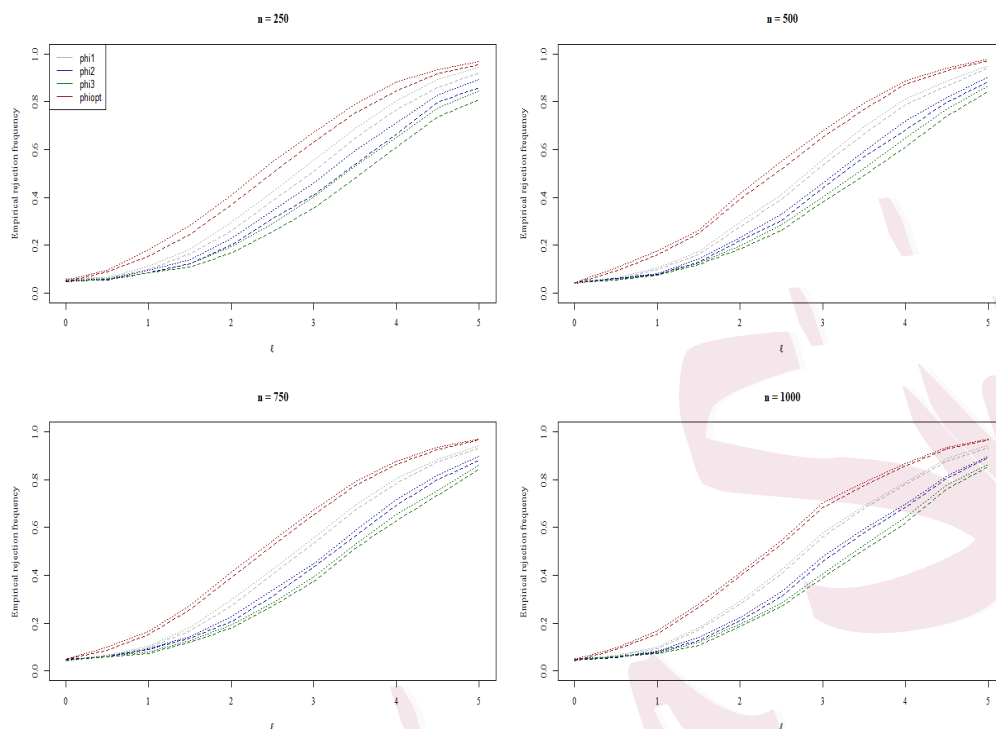


Figure 2: Empirical rejection frequencies of several directional runs tests: $\phi_1^{(n)}$, $\phi_2^{(n)}$, $\phi_3^{(n)}$ and $\phi_{\text{opt}}^{(n)}$. In dotted lines, the curves are the rejection frequencies of the tests performed with $\boldsymbol{\theta} = (1, 0, 0)$ while the curves in dashed lines coincide with the tests performed with an estimated version of $\boldsymbol{\theta}$. All the tests have been performed at the nominal level $\alpha = .05$.

e.g., Haigh, 2007). Sunspots are dark regions on the photosphere (region where visible photons are emitted) associated with strong magnetic field structures. The number of sunspots visible on the sun is not constant; it varies in each *solar cycle* (over about 11 years). During a solar cycle,

sunspots are frequently found in groups and tend to appear in bands on the sun situated just in the north and the south of the equator. Further details on sunspots and their origin can be consulted in Babcock (1961) and Solanki et al. (2006). The dataset displayed in Figure 5 is based on the Debrecen Photoheliographic Data (DPD) sunspot catalogue (Baranyi

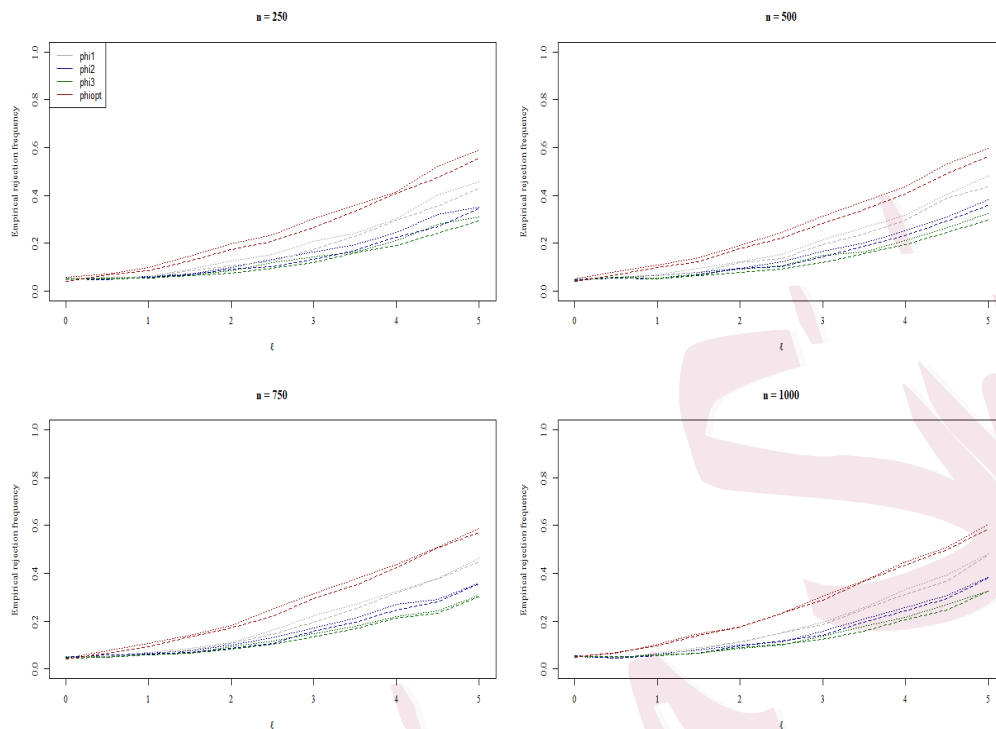


Figure 4: Empirical rejection frequencies of several directional runs tests: $\phi_1^{(n)}$, $\phi_2^{(n)}$, $\phi_3^{(n)}$ and $\phi_{\text{opt}}^{(n)}$. In dotted lines, the curves are the rejection frequencies of the tests performed with $\boldsymbol{\theta} = (1, 0, 0)$ while the curves in dashed lines coincide with the tests performed with an estimated version of $\boldsymbol{\theta}$. All the tests have been performed at the nominal level $\alpha = .05$.

et al., 2016; Györi et al., 2016). It consists in $n = 5,373$ central positions of groups of sunspots during the 23rd solar cycle, understood as the first-ever observations of each group.

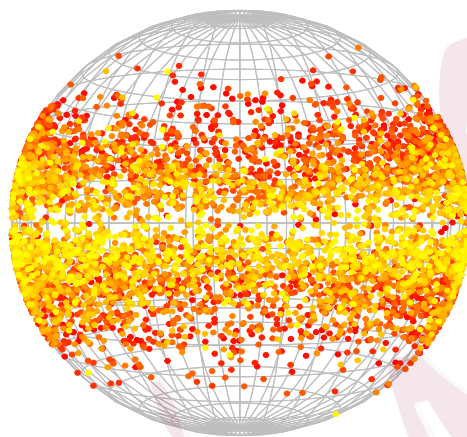


Figure 5: Emerging locations of sunspot groups during the 23rd solar cycle. The locations are colored with a red-yellow gradient according to the relative position of the sunspot appearance date within the solar cycle in order to visualize the Spörer's law.

Figure 5 strongly suggests that the distribution of the $(p=)3$ -dimensional random positions $\mathbf{X}_1, \dots, \mathbf{X}_n$ of sunspots is symmetric around the north pole $\boldsymbol{\theta} = (0, 0, 1)'$ in the sense that $\mathbf{O}\mathbf{X}_t$ has the same distribution as $\mathbf{X}_t$, $t = 1, \dots, n$ for any rotation $\mathbf{O}$ such that $\mathbf{O}\boldsymbol{\theta} = \boldsymbol{\theta}$. An explanation for this phenomenon is given by the Babcock (1961), see also García-Portugués et al. (2020). It also strongly suggests that the sequence of (the absolute values) of the latitudes $|\mathbf{X}'_1\boldsymbol{\theta}|, \dots, |\mathbf{X}'_n\boldsymbol{\theta}|$ contains some serial correlation. This is known as the Spörer's law which is illustrated by the *butterfly diagram* (available on <https://solarscience.msfc.nasa.gov/SunspotCycle.shtml>) in Figure 6. The Spörer's law mentions a drift of the average latitude of sunspots towards the Sun's equator during a solar cycle.

While serial correlation in the latitudes is well-known, little is known about the potential serial correlation that may be present within the longitudes $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ of the sunspots locations. The objective here is therefore to learn about a potential presence of serial correlation within sunspots longitudes $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_1), \dots, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_n)$ for various solar cycles: cycles 11 to 24. In Figure 5 we provide a plot of the dataset for the cycle 23 while in Figure 10 of the supplementary material we provide plots of the datasets for the cycles 16, 17, $\dots$, 24. The 11th solar cycle started in 1867 while the 24th solar cycle ended in 2019. We are currently in the 25th solar cycle.

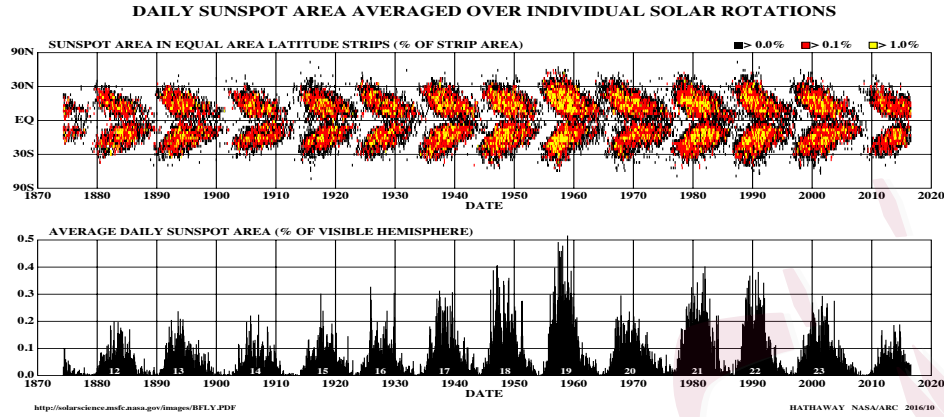


Figure 6: Top: butterfly diagram of the latitudes of sunspots (of various sizes) through time (for several solar cycles). Bottom: the average daily sunspot area through time.

As mentioned above, Figure 5 strongly suggests that the distribution of sunspots central positions is rotationally symmetric distribution around the north pole $\theta = (0, 0, 1)'$. Since sunspots are usually clustered in groups that evolve with time, serial correlation both in the latitude and in the longitude may definitely be present in the data. For each solar cycle, we performed various runs tests. Letting $\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i}$ stands for the observed locations of cycle i ($i = 11, \dots, 24$), we performed

- (i) classical Wald and Wolfowitz (1940) univariate runs tests of randomness on the absolute values $|\mathbf{X}'_{i1}\theta|, \dots, |\mathbf{X}'_{in_i}\theta|$ of the latitudes for each solar cycle;

-
- (ii) classical Wald and Wolfowitz (1940) univariate runs tests of randomness on the latitudes $\mathbf{X}'_{i1}\boldsymbol{\theta}, \dots, \mathbf{X}'_{in_i}\boldsymbol{\theta}$ for each solar cycle and;
 - (iii) our runs tests that are able to detect a potential serial correlation on the longitudes for each solar cycle.

In Figures 7, 8 and 9, we provide boxplots of (generalized) runs tests p -values for various choices of $H \in \{1, \dots, 4\}$ (number of lags in the runs statistic) and for each solar cycle. Each boxplot is obtained as follows: we built 200 different subsamples obtained by randomly keeping 75% of the original observations in each sample (for each solar cycle). Note that the conclusions drawn below are not strongly influenced by the percentage of the original observations we decide to keep in the subsamples. Boxplots for other choices are available on request. Then, on each such subsample, we performed the three runs tests described above; for each runs test we therefore obtained boxplots computed from 200 p -values. Inspection of Figures 7, 8 and 9 clearly reveals that there is frequently some serial correlation in the absolute values of the latitudes (as expected thanks to the Spörer's law) and in the longitudes. We cannot highlight the presence of autocorrelation in the absolute values of the latitudes in Cycle 11. We are not able to explain why cycle 11 is different but we nevertheless mention here that for cycle 11 we only have 251 measures of sunspots location (which is quite

less than for the other cycles). The signs of the latitudes seem more often random. The null hypothesis of iidness is often rejected in solar cycles 14, 15, 20 and 24 only. Therefore the fact that the location of a sunspot belongs to the northern or southern hemisphere seems to be often random.

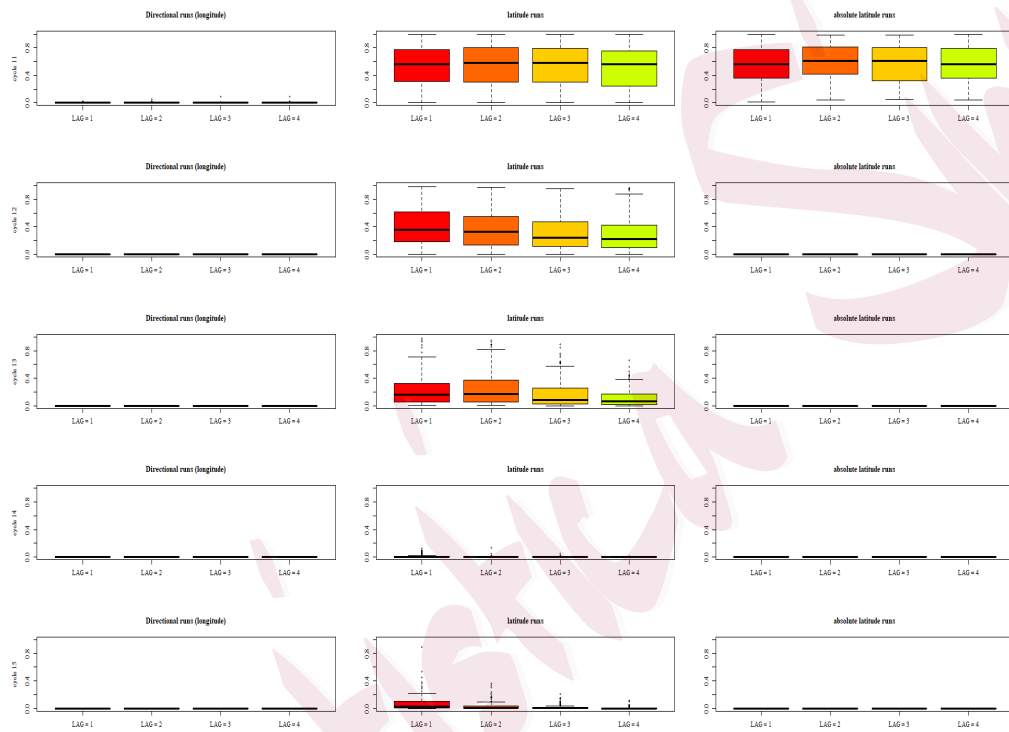


Figure 7: Boxplots of the various runs tests p -values for solar cycles 11 to 15.

To conclude, we provide in Figures 11 and 12 of the supplementary material the partial autocorrelation functions of (i) the absolute values $|\mathbf{X}'_{i1}\boldsymbol{\theta}|, \dots, |\mathbf{X}'_{in_i}\boldsymbol{\theta}|$ of the latitudes for various solar cycles; (ii) the latitudes

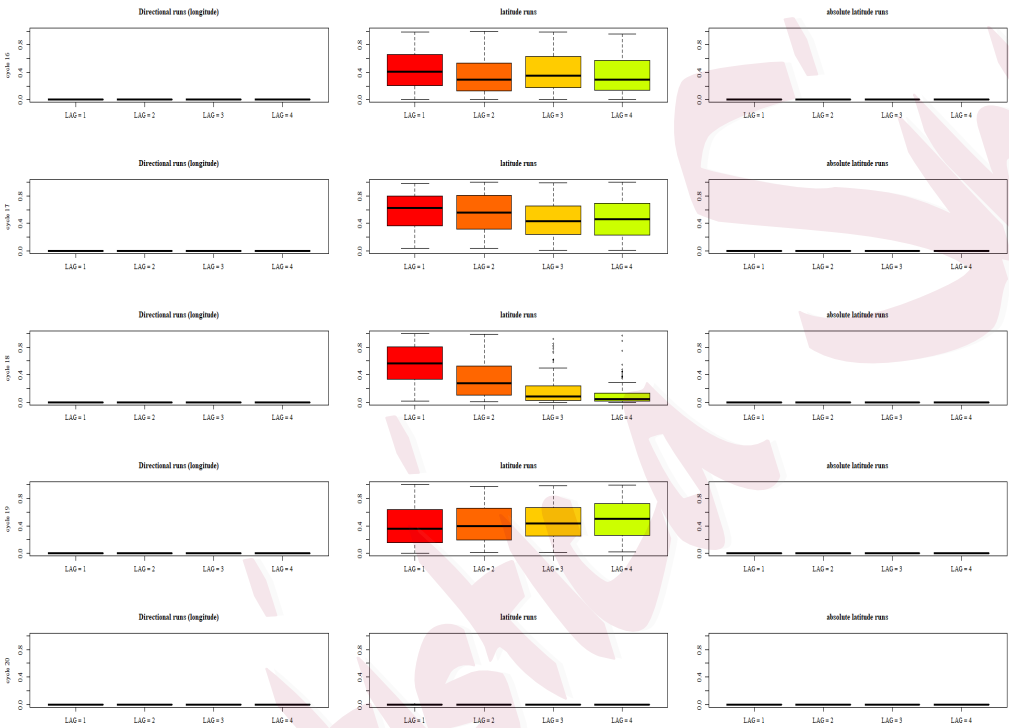


Figure 8: Boxplots of the various runs tests p -values for solar cycles 16 to 20.

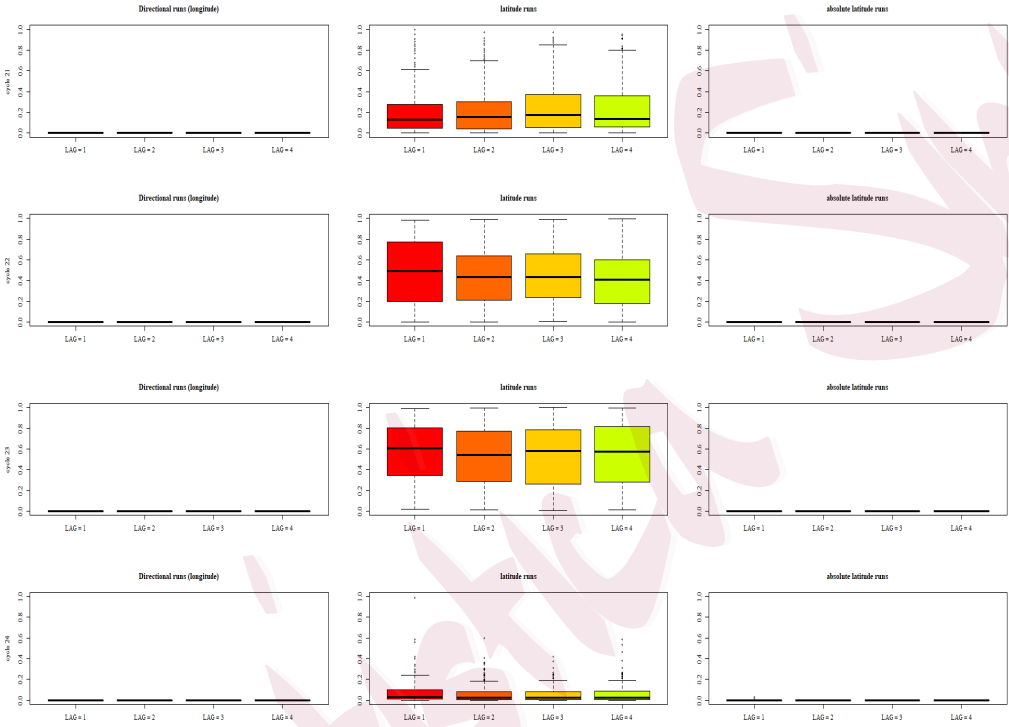


Figure 9: Boxplots of the various runs tests p -values for solar cycles 21 to 24.

$\mathbf{X}'_{i1}\boldsymbol{\theta}, \dots, \mathbf{X}'_{in_i}\boldsymbol{\theta}$ for various solar cycles and (iii) the angle associated with the longitudes for various solar cycles (each longitude or meridian is a bivariate unit vector and is therefore characterized by an angle). Figures 11 and 12 of the supplementary material essentially confirm the results obtained with the runs tests.

7. Perspectives for future research

In the present paper, we introduce runs tests for directional data. We show that our first-order runs tests $\phi_{\text{opt}}^{(n)}$ (or its two-sided versions) enjoy local and asymptotic optimality properties to test for iidness against alternatives under which some serial dependence is introduced. Monte-Carlo simulations strongly support our results while we show through the use of our tests that the longitudes of sunspots emerging locations are correlated. Optimality properties under local alternatives are obtained for first-order runs tests only. This is partly due to the fact that the Tangent Markov model involves “first order dependencies” only. Higher-order runs tests involving larger degrees of freedom are therefore less efficient against such Tangent Markov alternatives. Models with “higher-order dependencies” can be obtained for instance with observations such that

$$\mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_t) | \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_{t-1}) = \mathbf{s}_{t-1}, \dots, \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_{t-h}) = \mathbf{s}_{t-h}$$

is a mixture of h von Mises distributions with location parameters $\mathbf{s}_{t-1}, \dots, \mathbf{s}_{t-h}$ respectively. The study of the corresponding experiments together with the study of the asymptotic behavior of high-order runs tests or full-rank matrix-valued runs tests as in Paindaveine (2009) against such alternatives are left for future research.

Supplementary Material

The online supplementary material contains all the technical proofs and complements to the real data illustration.

Acknowledgments

Maxime Boucher gratefully acknowledges the PDR grant 40008027 from the Fonds National de la Recherche Scientifique (FNRS), Belgium, Christian Francq gratefully acknowledges the support of the Agence Nationale de la Recherche (ANR) through the Project MLEforRisk (ANR-21-CE26-0007), Yuichi Goto gratefully acknowledges the JSPS Grant-in-Aid for Early-Career Scientists JP23K16851 (Y.G.) while Thomas Verdebout gratefully acknowledges the PDR grant 40008027 from the Fonds National de la Recherche Scientifique (FNRS), Belgium and an advanced ARC grand from the Communauté Française de Belgique .

REFERENCES

References

- Accardi, L., J. Cabrera, G. Watson, et al. (1987). Some stationary markov processes in discrete time for unit vectors. *Metron* 45(1-2), 115–133.
- Alonso-Pena, M., J. Ameijeiras-Alonso, and R. M. Crujeiras (2021). Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation* 91(3), 477–500.
- Alonso-Pena, M., I. Gijbels, and R. M. Crujeiras (2024). A general framework for circular local likelihood regression. *Journal of the American Statistical Association* 119 (548), 2709–2721.
- Babcock, H. W. (1961). The topology of the Sun’s magnetic field and the 22-year cycle. *Astrophysical Journal* 133(2), 572–587.
- Baranyi, T., L. Györi, and A. Ludmány (2016). On-line tools for solar data compiled at the Debrecen observatory and their extensions with the Greenwich sunspot data. *Solar Physics* 291(9), 3081–3102.
- Biswas, M., M. Mukhopadhyay, and A. K. Ghosh (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* 101(4), 913–926.
- Cho, J. S. and H. White (2011). Generalized runs tests for the iid hypothesis. *Journal of Econometrics* 162(2), 326–344.
- Corzo, J. and G. Babativa (2013). A modified runs test for symmetry. *Journal of Statistical Computation and Simulation* 83(5), 984–991.

REFERENCES

- Di Marzio, M., S. Fensore, A. Panzera, and C. C. Taylor (2017). Nonparametric estimating equations for circular probability density functions and their derivatives. *Electronic Journal of Statistics* 11(2), 4323–4346.
- Dufour, J.-M., M. Hallin, and I. Mizera (1998). Generalized runs tests for heteroscedastic time series. *Journal of Nonparametric Statistics* 9(1), 39–86.
- Dyckerhoff, R., C. Ley, and D. Paindaveine (2015). Depth-based runs tests for bivariate central symmetry. *Annals of the Institute of Statistical Mathematics* 67, 917–941.
- García-Portugués, E., P. L. de Micheaux, S. G. Meintanis, and T. Verdebout (2024). Nonparametric tests of independence for circular data based on trigonometric moments. *Statistica Sinica*, 567–588.
- García-Portugués, E., D. Paindaveine, and T. Verdebout (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *Journal of the American Statistical Association* 115(532), 1873–1887.
- García-Portugués, E., I. Van Keilegom, R. M. Crujeiras, and W. González-Manteiga (2016). Testing parametric models in linear-directional regression. *Scandinavian Journal of Statistics* 43(4), 1178–1191.
- Györi, L., T. Baranyi, and A. Ludámny (2016). Comparative analysis of Debrecen sunspot catalogues. *Monthly Notices of the Royal Astronomical Society* 465(2), 1259–1273.
- Haigh, J. D. (2007). The Sun and the Earth’s climate. *Living Reviews in Solar Physics* 4, 2.

REFERENCES

- Hallin, M., H. Liu, and T. Verdebout (2024). Nonparametric measure-transportation-based methods for directional data. *Journal of the Royal Statistical Society, Series B* 86 (5), 1172–1196.
- Hentati-Kaffel, R. and P. De Peretti (2015). Generalized runs tests to detect randomness in hedge funds returns. *Journal of Banking and Finance* 50, 608–615.
- Henze, N. and M. D. Penrose (1999). On the multivariate runs test. *Annals of Statistics* 27(1), 290–298.
- Ley, C., Y. Swan, B. Thiam, and T. Verdebout (2013). Optimal R-estimation of a spherical location. *Statistica Sinica* 23(1), 305–332.
- Ley, C. and T. Verdebout (2017). *Modern Directional Statistics*. Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton: CRC Press.
- Marden, J. (1999). Multivariate rank tests. *Statistics textbooks and monographs* 159, 401–432.
- Mardia, K. V. and P. E. Jupp (1999). *Directional Statistics*. Wiley Series in Probability and Statistics. Chichester: Wiley.
- McWilliams, T. P. (1990). A distribution-free test for symmetry based on a runs statistic. *Journal of the American Statistical Association* 85(412), 1130–1133.
- Meilán-Vila, A., M. Francisco-Fernández, R. M. Crujeiras, and A. Panzera (2020). Nonparametric multiple regression estimation for circular responses. *Test* 30(3)(650–672), to appear.
- Paindaveine, D. (2009). On multivariate runs tests for randomness. *Journal of the American*

REFERENCES

- Statistical Association* 104(488), 1525–1538.
- Paindaveine, D. and T. Verdebout (2016). On high-dimensional sign tests. *Bernoulli* 22(3), 1745–1769.
- Paindaveine, D. and T. Verdebout (2017). Inference on the mode of weak directional signals: a Le Cam perspective on hypothesis testing near singularities. *Annals of Statistics* 45(2), 800–832.
- Rao, J. and A. SenGupta (2001). *Topics in Circular Statistics*, Volume 5 of *Series on Multivariate Analysis*. Singapore: World Scientific.
- Schach, S. (1969). Nonparametric symmetry tests for circular distributions. *Biometrika* 56(3), 571–577.
- Solanki, S. K., B. Inhester, and M. Schüssler (2006). The solar magnetic field. *Reports on Progress in Physics* 69(3), 563–668.
- Wald, A. and J. Wolfowitz (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics* 11(2), 147–162.

Maxime Boucher

Département de Mathématique and ECARES

Université libre de Bruxelles (ULB)

Campus Plaine, Boulevard du Triomphe, CP210

B-1050 Brussels, Belgium

REFERENCES

E-mail: maxime.boucher@ulb.be

Christian Francq CREST and University of Lille, France

Laboratoire Finance-Assurance

5 avenue Henry Le Chatelier

91764 Cedex, France

E-mail: christian.francq@ensae.fr

Yuichi Goto Faculty of Mathematics, Kyushu University

744, Motooka, Nishiku

Fukuoka, Japan 819-0395

E-mail: goto.yuichi.436@m.kyushu-u.ac.jp

Thomas Verdebout

Département de Mathématique and ECARES

Université libre de Bruxelles (ULB)

Campus Plaine, Boulevard du Triomphe, CP210

B-1050 Brussels, Belgium

E-mail: thomas.verdebout@ulb.be