Bernoulli **22**(4), 2016, 2113–2142
DOI: 10.3150/15-BEJ722

The circular SiZer, inferred persistence of shape parameters and application to early stem cell differentiation

STEPHAN HUCKEMANN[1,*], KWANG-RAE KIM[2,**], AXEL MUNK[3,†], FLORIAN REHFELDT[4,‡], MAX SOMMERFELD[1,§], JOACHIM WEICKERT[5,¶] and CARINA WOLLNIK[4,‖]

[1]*Felix Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen. E-mail:* [*]*huckeman@math.uni-goettingen.de;* [§]*max.sommerfeld@math.uni-goettingen.de*
[2]*School of Mathematical Sciences, University of Nottingham. E-mail:* [**]*Kwang-rae.Kim@nottingham.ac.uk*
[3]*Max Planck Institute for Biophysical Chemistry, Göttingen and Felix Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen. E-mail:* [†]*munk@math.uni-goettingen.de*
[4]*3rd Institute of Physics – Biophysics, University of Göttingen. E-mail:* [‡]*rehfeldt@physik3.gwdg.de;* [‖]*carina.wollnik@phys.uni-goettingen.de*
[5]*Faculty of Mathematics and Computer Science, Saarland University. E-mail:* [¶]*weickert@mia.uni-saarland.de*

We generalize the SiZer of Chaudhuri and Marron (*J. Amer. Statist. Assoc.* **94** (1999) 807–823; *Ann. Statist.* **28** (2000) 408–428) for the detection of shape parameters of densities on the real line to the case of circular data. It turns out that only the wrapped Gaussian kernel gives a symmetric, strongly Lipschitz semi-group satisfying "circular" causality, that is, not introducing possibly artificial modes with increasing levels of smoothing. Some notable differences between Euclidean and circular scale space theory are highlighted. Based on this, we provide an asymptotic theory to make inference about the persistence of shape features. The resulting circular mode persistence diagram is applied to the analysis of early mechanically-induced differentiation in adult human stem cells from their actin-myosin filament structure. As a consequence, the circular SiZer based on the wrapped Gaussian kernel (WiZer) allows the verification at a controlled error level of the observation reported by Zemel *et al.* (*Nat. Phys.* **6** (2010) 468–473): Within early stem cell differentiation, polarizations of stem cells exhibit preferred directions in three different micro-environments.

Keywords: circular data; circular scale spaces; mode hunting; multiscale process; persistence inference; stem cell differentiation; variation diminishing; wrapped Gaussian kernel estimator

1. Introduction

Mode (maxima and minima) and bump (maxima of the derivative) hunting of a density has a long history in statistical research and has been tackled from various perspectives. Good and Gaskins [24] argued that actually this would be a problem of significance testing rather than estimation, and, to some extent, we agree. We stress, however, that for practical purposes, it seems attractive to accompany any testing decision on the number of modes, say, with an estimator and corresponding visualization tools which are in agreement as much as possible with such a test decision. Indeed, many tests which have been developed (mainly in the context of detecting modes of a density) implicitly or explicitly offer this additional information or parts of it. Most

 © 2016 ISI/BS

of them are based on smoothing techniques with variable bandwidth which provides a reconstruction of modes and other shape features at a range of scales. Prominent methods include the critical bandwidth test of Silverman [48] (see also Ahmed and Walther [1] for a generalization to multivariate data), the dip test by Hartigan and Hartigan [26], the excess mass approach by Müller and Sawitzki [36], see also Polonik [42], the test of Cheng and Hall [11], the SiZer (SIgnificant ZERo crossings of the derivatives) by Chaudhuri and Marron [9,10], or the mode tree of Minnotte and Scott [35]; see also Minnotte [34], Ooi [39] and Klemelä [29] for extensions and related ideas. More recently, more sophisticated multiscale methods which do not rely on variable bandwidth kernel estimators have been developed for this purpose as well, for example, Dümbgen and Spokoiny [15], Davies and Kovac [14], Dümbgen and Walther [16], Schmidt-Hieber *et al.* [45].

In this paper, we are concerned with circular data for which rigorous inferential methods on the number and location of modes have not been established yet to the best of our knowledge. Recently, Oliveira *et al.* [38] suggested a circular version of the SiZer, and argued that this is of particular use for their problem at hand, the analysis of Atlantic wind speeds and directions, however, without providing a circular scale space theory or methods assessing the statistical significance of empirically found modes. Also in the application of the present paper, there are both biological and practical reasons that let us also favor the relatively simple circular SiZer's methodology although we are aware of some potential drawbacks, for example, a loss in power and asymptotic accuracy for small scales. In the below study of early differentiation of human mesenchymal stem cells, the structure characteristic for a specific cell type appears to be of the relative size captured by the first few largest modes while modes on much smaller scales most likely feature individual cell effects, not of immediate interest. The circular SiZer will allow to investigate the coarser mode structure and relative importance of modes in terms of bandwidth, which renders it a relatively simple tool for visualization of dynamics of modes through circular scales.

Causality. For data on the real line, it has been shown by Chaudhuri and Marron [10] that the SiZer controls the estimated modes at a large range of scales (represented by a bandwidth h) as long as these are above a smallest scale $h_0 > 0$, say. Crucial for a valid interpretation of these modes is "causality" which prevents the estimator from creating artifacts with increased smoothing. More specifically, causality of a family of kernels $\{L_h : h > 0\}$ is the requirement that for any integrable function f,

$$h \mapsto \#\,\mathrm{Modes}(L_h * f) \qquad \text{is decreasing.} \tag{1}$$

Note that on the circle, the number of modes of a differentiable function is simply half the number of sign changes of its derivatives or less (cf. Definition 2.5).

On the real line, it is well known that the Gaussian kernel yields the only causal family under suitable assumptions (Lindeberg [30]). Nevertheless, violation of causality will usually only matter for relatively small "critical" scales (Hall *et al.* [25]), and other (compactly supported) kernels may be used without too much concern as long as the true number of modes is small. However, for the SiZer it appears to be difficult to decide for given data whether and when such a critical scale is achieved, and hence causality violation may be a concern.

The above quoted circular SiZer method of Oliveira *et al.* [38] is a numerical extension of the SiZer methodology from the line to the circle based on the von-Mises kernel $g(e^{it}) \sim$

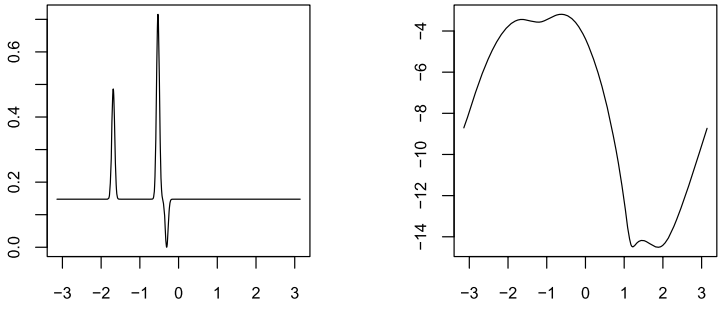

Figure 1. The von Mises kernel is not causal (mode reducing) on the circle; displaying on the right the tri-modal circular density (its nonuniform part in log-scale) of a convolution with a von Mises kernel ($\kappa = 6$) of the bi-modal circular distribution displayed on the left (its density is the combination of two spikes and a cleft with a uniform distribution).

$\exp(\kappa \cos(t - \mu))$ with variable concentration parameter κ playing the role of the (inverse) bandwidth. The von Mises density is often considered as the natural counterpart of the Gaussian on the circle, for example, maximizing entropy with given mean (Mardia and Jupp [33], Section 3.5.4). For circular nonparametric smoothing, this kernel has also been used, for example, in Fisher and Marron [22], Taylor [49]. However, Lemma 1 of Munk [37] states that the von Mises kernel is not variation diminishing (i.e., the number of sign changes is not nonincreasing under smoothing, cf. Definition 2.1) for $\kappa > 1/2$ (the precise statement is that "order 2 variation diminishing" – which implies variation diminishing – holds if and only if $0 \le \kappa \le 1/2$). From our Theorem 2.6, it follows now that a convolution with a von Mises distribution with concentration parameter $\kappa > 1/2$ violates circular causality (1). Figure 1 shows such an example where the convolution with a von Mises kernel increases the number of modes (maxima) from two to three. As a consequence, the von Mises kernel seems questionable in the context of the SiZer since a lower bound for the number of modes of the smoothed density does not imply a lower bound for those of the true underlying density. Therefore, the question remains whether a circular causal kernel exists admitting a circular scale space approach.

Circular scale spaces – The uniqueness of the wrapped Gaussian. In this paper, we give an affirmative answer to this question. Let

$$K_h(e^{it}) = \frac{1}{\sqrt{h}} \sum_{k \in \mathbb{Z}} \phi\left(\frac{t + 2\pi k}{\sqrt{h}}\right) \tag{2}$$

be the wrapped Gaussian kernel (on the circle) with bandwidth $h > 0$ where ϕ denotes the standard normal density. We will show that the wrapped Gaussian is the only circular causal kernel. Therefore, from this perspective it can be viewed as the most natural analog to the normal density on the real line. This result requires some preparation. We will:

(1) propose circular scale space axiomatics and discuss its relationships to circular variation and mode reducing properties,

(2) show that under reasonable assumptions the wrapped Gaussian kernel gives the one and only semigroup guaranteeing causality.

This will then allow us to:

(3) assess asymptotically the statistical significance of shape features obtained from the WiZer,
(4) and to infer persistence over smoothing scales of shape features.

The analogs of (1) and (2) are well investigated in the linear case (cf. Lindeberg [30], Weickert *et al.* [52]). This result also complements numerous linear scale-space axiomatics for images that have attempted to establish uniqueness of the Gaussian kernel; see, for example, Weickert *et al.* [52] for an overview. However, it should be noted that a justification of Gaussian smoothing via a variation diminishing axiom is only possible for 1-D signals (cf. Babaud *et al.* [3]). It is well known that for 2-D images, Gaussian convolution can create new extrema (see Lindeberg [30]). To avoid this, one has to consider scale-spaces based on nonlinear partial differential equations such as mean curvature motion or the affine morphological scale-space. For more details, we refer to Alvarez *et al.* [2]. In contrast to the Euclidean case, circular scale spaces have hardly been considered so far. Notable exceptions are a periodic scale-space filtering by Wada *et al.* [50] and a feature extraction study by Briggs *et al.* [5]. However, there is no axiomatic foundation of circular scale-spaces in terms of a variation diminishing property. We will show in the following that on the circle the situation is indeed peculiar and, for example, causality is fully equivalent to the circular diminishing property. In summary, as a byproduct of this paper, the circular case fills a gap of classical scale space axiomatics.

Statistical guarantees. Concerning (3) above, we find for the SiZer based on wrapped Gaussians that the probability of overestimating the true number of modes on *any* scale can be bounded by some small number α, say, while the probability of underestimating the true number goes to zero for any fixed bandwidth as the sample size goes to infinity. Causality as in (1) ensures that, if we find significant evidence for, say, k circular modes in the smoothed density, we automatically have significant evidence that there are also at least k modes in the true density.

Inferred mode persistence diagrams. Of particular importance are those scales where the lower bounds on the number of inferred modes change and the intervals of scales in which these bounds are constant. More general, under causality, the latter give a notion of *significant scale space persistences of shape features* over smoothing scales. In fact, these expand the general scheme of topological persistence introduced by Edelsbrunner *et al.* [17], cf. also Ghrist [23], Carlsson [8] among many others. This originally deterministic scheme is currently gaining high momentum in statistics and medical imaging (e.g., Chung *et al.* [12], Heo *et al.* [27]). Illustrated by the application at hand, we propose *inferred persistence diagrams* simultaneously depicting significant bandwiths for *births* and *splits* of modes. In addition to statistically estimating (as done for persistent homologies by Bubenik and Kim [7]), we also provide for confidence statements for the estimate being a lower bound. This is related to recent work by Balakrishnan *et al.* [4] who provide for inference on persistence diagrams and by Schwartzman *et al.* [46,47] for inferential mode detection of linear densities within specific classes. Complementing this, however, we are firstly "truly" circular. Secondly, we do not infer on the support of the data but on the shape of its very density, and thirdly, for this density we require no assumptions.

A measure for early stem cell differentiation. Utilizing inferred persistence, we give a proof of concept to use persistence diagrams of shape parameters – namely modes – to elucidate precisely how the elasticity of the micro-environment directs early human mesenchymal stem cell

(hMSC) differentiation; hMSCs from bone marrow are considered highly promising for regenerative medicine and tissue engineering. However, for successful therapeutic applications understanding and controlling of differentiation mechanisms is of paramount importance. The seminal study of Engler *et al.* [20] demonstrated that the mechanical properties (Young's modulus E) of the micro-environment can direct stem cell differentiation. In Section 4, we re-analyze the fluorescence microscopy images of nonmuscle myosin IIa in stem cells from Zemel *et al.* [53]. Here, hMSCs have been cultured for 24 hours on substrates of varying Young's modulus E mimicking the physiological mechanical properties of different tissue in vivo: 1 kPa corresponding to neural cells and brain tissue, 11 kPa corresponding to muscle fibres and 34 kPa corresponding to pre-mineralized collagen densifications in bone, cf. Rehfeldt *et al.* [43]. From these images, we extract filament-process valued data[1] using the filament sensor of Eltzner *et al.* [18], in particular total fibre length over angular directions, which live on the circle. We do not only reproduce the observed nonmonotonicity over matrix rigidity, now for mode persistences, but also, more subtly, the respective mode-persistence diagrams indeed reflect different matrix rigidities elucidating the impact of micro-environment rigidities on early stem cell differentiation.

Plan of the paper. In the following Section 2, we begin by establishing circular scale space axiomatics, while the details on circular sign changes have been deferred to Appendix A. All of the proofs of the theorems subsequently developed in this and in Section 3 have been deferred to Appendix B. For the application to human stem cell differentiation in Section 4, in Appendix C we give a numerical foundation for the choice of the number of wrappings required to compute the p-values underlying the tests performed by the WiZer with a desired accuracy. For the bandwidths considered in our application, an error less than 10^{-4} can be achieved with six wrappings.

2. Circular scale space theory

2.1. Notation

Let $\mathbb{S} = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle in the complex plane $\mathbb{C}$ equipped with the measure $d\mu(z) = \frac{dt}{2\pi}$ for $z = e^{it}$, with $t \in [0, 2\pi)$ of the uniform distribution on $\mathbb{S}$. This is the normalized Haar measure on the compact Abelian group $\mathbb{S}$ with the ordinary multiplication of complex numbers. Then

$$\hat{f}_k := \int_{\mathbb{S}} f(z) z^{-k} \, d\mu(z)$$

is the kth Fourier coefficient, $k \in \mathbb{Z}$, of a function $f \in L^1(\mathbb{S})$. Moreover, for $f, g \in L^1(\mathbb{S})$ we have the convolution

$$(g * f)(w) = \int_{\mathbb{S}} g\big(wz^{-1}\big) f(z) \, d\mu(z) = \frac{1}{2\pi} \int_0^{2\pi} g\big(e^{i(s-t)}\big) f\big(e^{it}\big) \, dt, \qquad w = e^{is} \in \mathbb{S},$$

which is well defined for $w \in \mathbb{S}$ μ-a.e. and in $L^1(\mathbb{S})$.

[1] Which are provided for along with the WiZer R-package including inferred persistences on the website of the second author, http://rayerk.weebly.com/files.html.

In particular, every function $g \in L^1(\mathbb{S})$ with $\int_{\mathbb{S}} g(z)\,d\mu(z) = 1$ is a *circular kernel* as it conveys a bounded operator $g : L^1(\mathbb{S}) \to L^1(\mathbb{S})$, $f \mapsto g * f$. A family of circular kernels $\{L_h : h > 0\}$ generates from every $f \in L^1(\mathbb{S})$ a *scale space tube*

$$\big\{L_h * f(z) : z \in \mathbb{S}, h > 0\big\}.$$

We write $\hat{L}_{h,k}$ for the kth Fourier coefficients of L_h, $k \in \mathbb{Z}$.

Finally, we say that a function $f : \mathbb{S} \to \mathbb{R}$ is differentiable if the *counter-clockwise derivatives*

$$(Df)(z) = \frac{d}{dt}\Big|_{t=0}\big(t \mapsto f\big(e^{it}z\big)\big)$$

exist for all $z \in \mathbb{S}$, it is m-times differentiable if the mth counter-clockwise derivatives exist which are defined iteratively as $D^m f = D(D^{m-1} f)$.

2.2. Regular scale space tube axiomatics

In the following, $S_c(f)$ denotes the number of *cyclic sign changes* of a function $f : \mathbb{S} \to \mathbb{R}$ as introduced by Mairhuber *et al.* [32] and detailed in the Appendix.

Definition 2.1. *A family of circular kernels $\{L_h : h > 0\}$ is*

(SG) *a* semigroup *if $L_h * L_{h'} = L_{h+h'}$ for all $h, h' > 0$,*
(VD) variation diminishing *if*

$$S_c(L_h * f) \le S_c(f)$$

for every $f \in L^1(\mathbb{S})$ and $h > 0$,
(SM) symmetric *if $L_h(z) = L_h(z^{-1})$ for all $z \in \mathbb{S}$ and $h > 0$,*
(SL) strongly Lipschitz *if there exists $r > 0$ such that the limit*

$$\lim_{h\to 0}\left(\frac{\hat{L}_{h,k} - 1}{h|k|^r}\right)_{k\in\mathbb{Z}}$$

exists in the space $\ell^\infty(\mathbb{Z})$ of bounded sequences on $\mathbb{Z}$.

A scale space tube generated by a strongly Lipschitz, symmetric and variation diminishing semigroup will be called a regular scale space tube.

Straightforward computation gives for the wrapped Gaussian kernel that $\hat{K}_{h,k} = e^{-k^2 h/2}$, and hence the following.

Remark 2.2. The family of wrapped Gaussian kernels is a symmetric semigroup that is strongly Lipschitz with $r = 2$.

Of special interest are those families $\{L_h : h > 0\}$ of kernels that are *differentiable* in the sense that

$$\partial_h (L_h * f) = \lim_{t \to 0} \frac{L_{h+t} * f - L_h * f}{t}$$

exists for all $h > 0$ and $f \in C^\infty(\mathbb{S})$. The following is a straightforward consequence of the definition.

Remark 2.3. A semigroup of circular, strongly Lipschitz kernels is differentiable. To see this, consider the Fourier coefficients

$$\big((L_{h+t} * f \hat{-} L_h * f)/t\big)_k = \hat{L}_{h,k}(\hat{L}_{t,k}\hat{f}_k - \hat{f}_k)/t = \hat{L}_{h,k}|k|^r \hat{f}_k(\hat{L}_{t,k} - 1)/\big(t|k|^r\big).$$

The latter has a limit in $\ell^2(\mathbb{Z})$ for $t \to 0$ since smoothness of f guarantees that $(\hat{L}_{h,k}|k|^r \hat{f}_k)_k$ is in $\ell^2(\mathbb{Z})$ and $\lim_{h\to 0}(\hat{L}_{t,k} - 1)/(t|k|^r)$ exists in $\ell^\infty(\mathbb{Z})$.

At this point let us relate the heat equation

$$\tfrac{1}{2}\Delta u = \partial_h u \tag{3}$$

on a Riemannian manifold M with Laplace–Beltrami operator Δ for functions $u : M \times [0, \infty) \to \mathbb{R}$ with some initial condition $u(\cdot, 0) = f$ (e.g., Sakai [44], Section IV.3) to scale space axiomatics. Using the Fourier transform, the heat equation can be solved in the spectral domain such that the unique solution of (3) is given by the convolution of the initial condition with the usual Gaussian $x \mapsto \phi(x/\sqrt{h})/\sqrt{h}$ (in case of $M = \mathbb{R}$) and the wrapped Gaussian $z \mapsto K_h(z)$ (in case of $M = \mathbb{S}$), if the convolution exists; where in the circular case $\Delta = D^2$. Requiring that equality (3) holds only for the signs of both sides at proper local maxima or minima, that is, that maxima and minima are not enhanced under smoothing, is another axiom frequently found in the scale space literature. This axiom has been used by Babaud *et al.* [3], for deriving the Gaussian as the unique scale space kernel in 1D. Later on, Lindeberg [31] emphasised that nonenhancement of local extrema distinguishes the Gaussian from other scale space kernels such as the ones corresponding to the so-called α scale spaces. However, if one goes beyond scale space representations that can be expressed in terms of convolutions with kernels, this property can also be fulfilled. For instance, nonenhancement of local extrema has been established in Weickert [51], Section 2.4.2, for a family of nonlinear diffusion evolutions. Similar reasonings can be used in the circular setting as well.

Definition 2.4. *A family of differentiable circular kernels* $\{L_h : h > 0\}$ *is*

(NE) not enhancing local extrema *if for every smooth* $f : \mathbb{S} \to \mathbb{R}$ *and* $h > 0$

$$\operatorname{sign}\big(D^2(L_h * f)\big) \operatorname{sign}\big(\partial_h(L_h * f)\big) \geq 0$$

at every nondegenerate critical point (*i.e.*, *points at which* $D(L_h * f) = 0 \neq D^2(L_h * f)$) *of* $L_h * f$ *on* $\mathbb{S}$.

Finally, we formalize (1). From this, we obtain one more axiom.

Definition 2.5. *Let $k \in \mathbb{N}_0$. A differentiable function $f \in L^1(\mathbb{S})$ is k-*modal (*i.e., it has k modes or less*) *if*

$$S_c(Df) \leq 2k.$$

A circular kernel L_h, $h > 0$, is

(MR) mode reducing (*i.e., reducing the number of modes*) *if for any function f that is differentiable except for finitely many points and that satisfies $\lim_{t\downarrow 0} Df(e^{it}z) = \lim_{t\uparrow 0} Df(e^{it}z)$ for all $z \in \mathbb{S}$, we have*

$$S_c(L_h * Df) \leq S_c(Df),$$

where Df is continued to all of $\mathbb{S}$ via its left and right limits.

Theorem 2.6. *A family of circular kernels is variation diminishing if and only if it is mode reducing. Moreover, a differentiable semigroup of variation diminishing circular kernels is not enhancing local extrema.*

Here is the complete picture.

$$\left.\begin{array}{rcl} \text{(MR)} & \Leftrightarrow & \text{(VD)} \\ & & \text{and} \\ \text{(SL), (SG)} & \Rightarrow & \text{differentiable (SG)} \end{array}\right\} \Rightarrow \text{(NE)}.$$

On the line, an analog definition of mode reducing is more simple, as no periodicity of the antiderivative is required. For functions $f, g \in L^1(\mathbb{R})$, denote by $(f *_{\mathbb{R}} g)(y) = \int_{-\infty}^{\infty} f(x)g(y-x)\,dx$ the usual convolution on the line, if existent and by $S^-(f)$ the number of sign changes of f. (cf. Karlin [28], Brown *et al.* [6] and Appendix A.) Then call a kernel K on the line *mode reducing* if for any differentiable function $f : \mathbb{R} \to \mathbb{R}$ we have $S^-((K *_{\mathbb{R}} f)') \leq S^-(f')$.

Remark 2.7. Inspection of the proofs of Lemmata B.1 and B.2 in the Appendix shows that on the line $\mathbb{R}$, with the analog axioms, Theorem 2.6 also holds true.

2.3. The uniqueness of the wrapped Gaussian kernel

The analogs of the following two characterizations of the wrapped Gaussian kernel as generating regular scale space surfaces and being up to scaling the only variation diminishing, symmetric and strongly Lipschitz semigroup are well known on the line and, under suitable adaptations on Euclidean spaces. For an overview over various scale space axiomatics for Euclidean spaces, see Weickert *et al.* [52]. Suitably adapted arguments, sometimes even simpler lead to the following two circular versions, which, to the knowledge of the authors, have not been established before.

Theorem 2.8. *The family of wrapped Gaussian kernels $\{K_h : h > 0\}$ generates a regular scale space tube.*

Theorem 2.9. *Let $\{L_h : h > 0\}$ be a semigroup generating a regular scale space tube. Then there is a constant $\alpha > 0$ such that for all $h > 0$ the function $L_{\alpha h}$ is the wrapped Gaussian kernel K_h with bandwidth h.*

Remark 2.10. Often in scale space literature, the axiom of *strong continuity*

(SC) $\|(L_h - \mathrm{I})f\|_{L^2} \to 0$ as $h \downarrow 0$ for all $f \in L^2(\mathbb{S})$,

is introduced which is weaker than the strong Lipschitz property (SL). Here, I denotes the identity. It seems that this property is not sufficient to ensure the critical fact (detailed in the Appendix) that the smooth functions are in the domain of definition of the infinitesimal operator

$$\mathcal{A} = \lim_{h \downarrow 0} \frac{L_h - \mathrm{I}}{h},$$

which is required if $\mathcal{A}$ is to turn out to be a multiple of the Laplacian.

More precisely, we first show that $\mathcal{A}$ is a local operator, that is, if a smooth function vanishes in an neighborhood of a point, then its image under $\mathcal{A}$ also vanishes at this point. This implies that $\mathcal{A}$ is a differential operator. In a second step, we argue that, due to causality, $\mathcal{A}$ must be a multiple of the Laplacian.

This strain of arguments has been employed by Lindeberg [31], page 41, for the Euclidean case, who has suggested to require a property different from the strong Lipschitz property (SL) which we require for the circle.

3. Inference on shape parameters

In the previous Section 2, we showed that the wrapped Gaussian generates a regular scale space tube and that it is only the wrapped Gaussian that has this property. For this reason, in the following, we only consider the wrapped Gaussian kernel K_h defined in (2) although empirical scale space tubes (Definition 3.2) can of course be defined for arbitrary families, Remark 3.5 holds for any semigroup and the proofs of Theorem 3.6 and Corollary 3.7 require only that second moments be finite. In the literature, the properties variation diminishing, nonenhancement of modes, etc. are often referred to as preserving *causality*. Thus, the wrapped Gaussians are the one and only (under reasonable assumptions) kernels preserving circular causality.

3.1. Circular causality

Assumption 3.1. *From now on assume that X is a random variable (with or without a density with respect to the Haar measure μ) taking values on the circle $\mathbb{S}$ and we observe $X_1, \ldots, X_n \overset{i.i.d.}{\sim} X$.*

Definition 3.2. *We call the two-parameter stochastic process indexed in $\mathbb{S} \times \mathbb{R}^+$*

$$\left\{ f_h^{(n)}(z) := \frac{1}{n} \sum_{j=1}^{n} K_h\left(z X_j^{-1}\right) : z \in \mathbb{S}, h > 0 \right\}$$

the empirical circular scale space tube *and*

$$\{f_h(z) := Ef_h^{(n)}(z) : z \in \mathbb{S}, h > 0\}$$

the population circular scale space tube.

Remark 3.3. Note that if X has density f with respect to the Haar measure μ on $\mathbb{S}$ then under Assumption 3.1

$$E\big(f_h^{(n)}(z)\big) = E\big(K_h\big(zX^{-1}\big)\big) = \int_{\mathbb{S}} K_h\big(zw^{-1}\big) f(w)\mu(dw) = (K_h * f)(z).$$

Now we can state the causality theorem.

Theorem 3.4 (The circular causality theorem). *Let K_h be the wrapped Gaussian. Under Assumption 3.1, for any $m = 0, 1, \ldots$ the following holds*:

(i) *The mappings $h \mapsto S_c(D^m f_h)$ and $h \mapsto S_c(D^m f_h^{(n)})$ are decreasing and right continuous functions on $(0, \infty)$.*

(ii) *If X has a density f with respect to μ which is m-times differentiable and we set $f_0 = f$ then $h \mapsto S_c(D^m f_h)$ is decreasing and right continuous on $[0, \infty)$.*

3.2. Weak convergence of the scale space tube

Remark 3.5. All information of the empirical and the population scale space tube for any bandwidth is already contained in that of every smaller bandwidth. More precisely, for $h_0 > 0$, $\{f_h^{(n)}(z) : z \in \mathbb{S}, h \geq h_0\}$ and $\{f_h(z) : z \in \mathbb{S}, h \geq h_0\}$ can be reconstructed from $\{f_{h_0}^{(n)}(z) : z \in \mathbb{S}\}$ or $\{f_{h_0}(z) : z \in \mathbb{S}\}$, respectively.

The reason for this is of course the identity $f_{h_0+h} = K_h * f_{h_0}$ for any $h > 0$ (and similarly for the empirical counterpart), which also holds for $h = 0$ if we set $K_0 * f := f$.

Moreover, $(z, h) \mapsto K_h(z)$ is a solution to the heat equation (3), and hence $f_h^{(n)}(z)$ and $f_h(z)$ are solutions as well. By the well known maximum principle for solutions of the heat equation, maxima are attained at the boundary, that is, we have

$$\sup_{z\in\mathbb{S}, h\geq h_0} \big|f_h^{(n)}(z) - f_h(z)\big| = \sup_{z\in\mathbb{S}} \big|f_{h_0}^{(n)}(z) - f_{h_0}(z)\big| \qquad \text{for all } h_0 > 0. \tag{4}$$

Theorem 3.6. *Under Assumption 3.1, let $m \geq 0$ be an integer and define*

$$\mathfrak{c}(z_1, z_2; h) = \operatorname{cov}\big(D^m K_h\big(z_1 X^{-1}\big), D^m K_h\big(z_2 X^{-1}\big)\big) \tag{5}$$

for $z_1, z_2 \in \mathbb{S}$ and $h > 0$. Then, for any fixed $h > 0$,

$$n^{1/2}\big(D^m f_h^{(n)}(z) - D^m f_h(z)\big) \to G_h, \tag{6}$$

weakly in $C(\mathbb{S})$, *where* G_h *is a Gaussian process on* $\mathbb{S}$ *with mean zero and covariance structure defined by* $\mathfrak{c}(z_1, z_2; h)$.

Moreover, G_h *has continuous sample paths with probability one. In particular,* $P(\sup_{z\in\mathbb{S}} |G_h(z)| < \infty) = 1$.

This gives at once the following.

Corollary 3.7. *Under Assumption 3.1 with* $h_0 > 0$,

$$\sup_{z\in\mathbb{S}, h\geq h_0} n^{1/2} \left| D^m f_h^{(n)}(z) - D^m f_h(z) \right|$$

converges weakly to $\sup_{z\in\mathbb{S}} |G_{h_0}(z)|$ *as* $n \to \infty$, *where* G_{h_0} *is a Gaussian process with zero mean and covariance structure* $\mathfrak{c}(z_1, z_2, h_0)$ *given by* (5).

In the following denote by $q_{(1-\alpha)}$ the $(1-\alpha)$ quantile of the random variable $\sup_{z\in\mathbb{S}} |G_{h_0}(z)|$, for $0 \leq \alpha \leq 1$.

The convergence results can be used to test whether the derivative $D^m f_h(z)$ at some point $z \in \mathbb{S}$ and for some bandwidth $h \geq h_0 > 0$ is different from zero. We perform the tests for the hypotheses

$$H_0^{(h,z)} : D^m f_h(z) = 0, \qquad z \in \mathbb{S}, h \geq h_0$$

as follows:

$$\text{If } \begin{cases} D^m f_h^{(n)}(z) > n^{-1/2} q_{(1-\alpha)}, & \text{reject } H_0^{(h,z)} \text{ and conclude } D^m f_h^{(n)}(z) > 0, \\ \left| D^m f_h^{(n)}(z) \right| \leq n^{-1/2} q_{(1-\alpha)}, & \text{accept } H_0^{(h,z)}, \\ D^m f_h^{(n)}(z) < -n^{-1/2} q_{(1-\alpha)}, & \text{reject } H_0^{(h,z)} \text{ and conclude } D^m f_h^{(n)}(z) < 0. \end{cases} \tag{7}$$

The following theorem states two key properties of this test:

Firstly, by using the quantile $q_{(1-\alpha)}$ of the supremum $\sup_{z\in\mathbb{S}} |G_{h_0}|$ we control the *family-wise error rate* (FWER) of the tests. More precisely, we can assert that the asymptotic probability that one or more of the hypotheses $\{H_0^{(h,z)} : z \in \mathbb{S}, h \geq h_0\}$ is falsely rejected is at most α.

Secondly, we see that for a given bandwidth h all sign changes of $D^m f_h$ are detected by this test asymptotically with probability one. Note that this does not require the test to detect all points of positive/negative derivative. Indeed, without prior information on the smallest scales this is not possible for any test since the absolute value of the derivative can be arbitrarily small.

Theorem 3.8. *If under Assumption 3.1 with* $h_0 > 0$, $0 \leq \alpha \leq 1$, *the test* (7) *is performed for each of the hypotheses* $\{H_0^{(h,z)} : z \in \mathbb{S}, h \geq h_0\}$ *then the probability that one or more of them are falsely rejected is asymptotically at most* α.

Moreover, if for any fixed bandwidth $h \geq h_0$ *the function* $D^m f_h(x)$ *has* $2k \geq 1$ *sign changes then this test will detect them with asymptotic probability one.*

3.3. Inferred persistence

A scale space satisfying causality gives a notion of *persistence* of features of a (density-) function f which are given by zero-crossings of derivatives. The number of such features is a decreasing integer valued function in the bandwidth and, therefore, is constant except for finitely many jumps. We can consider the bandwidths associated with these jumps as the amount of smoothing that is necessary to remove the features, one by one. Indeed, with the wrapped Gaussian K_h as kernel, we can define a sequence decreasing in k,

$$\operatorname{ssp}_k^{(m)} := \operatorname{ssp}_k^{(m)}(f) = \inf_{h>0}\big\{h : S_c\big(K_h * D^m f\big) < 2k\big\},$$

and call $\operatorname{ssp}_k^{(m)}$ the *scale space persistence* of $D^{m-1}f$.

We will now use the family of tests introduced in (7) to define an empirical counterpart of $\operatorname{ssp}^{(m)}$, which can be obtained from the data $X_1, \ldots, X_n$. To this end, let

$$W_h^{(m)}(z) = \begin{cases} +1, & \text{if } D^m f_h^{(n)}(z) > n^{-1/2} q_{(1-\alpha)}, \\ 0, & \text{if } \big|D^m f_h^{(n)}(z)\big| \le n^{-1/2} q_{(1-\alpha)}, \\ -1, & \text{if } D^m f_h^{(n)}(z) < n^{-1/2} q_{(1-\alpha)} \end{cases} \quad \text{for } h \ge h_0 \text{ and } W_h = W_{h_0} \text{ for } h < h_0, \tag{8}$$

which we call the *WiZer signature function*, and define

$$\widehat{\operatorname{ssp}}_k^{(m)} := \widehat{\operatorname{ssp}}_k^{(m)}(f) = \inf_{h>0}\big\{h : S_c\big(W_h^{(m)}\big) < 2k\big\} \tag{9}$$

as the *significant empirical scale space persistence* of $D^m f_h$ or just *inferred persistence*. Note that we are *not* simply defining the empirical scale space persistence as the persistence of the kernel density estimator $\inf_{h>0}\{h : S_c(D^m f_h^{(n)}) < 2k\}$. The reason is, that we want to eliminate statistically insignificant features.

The following is an immediate consequence of Theorem 3.8.

Corollary 3.9. *Under the assumptions of Theorem 3.8, the following holds for the scale space persistence of $D^{m-1}f$*

$$\liminf_{n\to\infty} P\big(\widehat{\operatorname{ssp}}_k^{(m)} \le \operatorname{ssp}_k^{(m)} \text{ for all } k \in \mathbb{Z}\big) \ge 1 - \alpha.$$

Remark 3.10. Note that by definition the sequence $\operatorname{ssp}_1^{(m)}, \operatorname{ssp}_2^{(m)}, \operatorname{ssp}_3^{(m)}, \ldots$ is decreasing. For $k \ge 1$, $\operatorname{ssp}_k^{(m)}$ can be considered as the *birth* bandwidth of the kth mode of $D^{m-1}f_h$ and for $k \ge 2$ it is also the *split* bandwidth where $k-1$ modes of $D^{m-1}f_h$ *split* into k modes. As birth and splits (or births) occur with decreasing bandwidths, our scenario is twofold opposite to that of usual persistence diagrams (e.g., Edelsbrunner *et al.* [17], Cohen-Steiner *et al.* [13], Chung *et al.* [12]) where births and mergers (or deaths) occur with increasing bandwidth.

The application in the following Section 4 illustrates the case $m = 1$ of scale space persistence of modes of f_h.

4. Application to early stem cell differentiation

Datasets – acto-myosin cytoskeleton. We re-analyze a data set from Zemel *et al.* [53] hMSCs (human mesenchymal stem cells) cultured for 24 hrs. on elastic substrates of different Young's moduli E (1, 11 and 34 kPa), subsequently chemically fixed and immuno-stained for NMM IIa, the motor proteins making up small filaments that are actually responsible for cytoskeletal tension. Fluorescence images were recorded for 30 cells on each of the three conditions:

1. $E = 1$ kPa which form the data set *Black*,
2. $E = 11$ kPa which form the data set *Red* and
3. $E = 34$ kPa which form the data set *Blue*.

From these micrographs, the filament structure is reconstructed using the filament sensor of Eltzner *et al.* [18]. Figure 2 shows typical cell images with their traced skeleton from each of the three datasets.

The statistical model. Each traced image gives a realization $(z_j, w_j)_{j=1}^J$ of a bounded filament process where $J \in \mathbb{N}$ denotes the number of filaments, $z_j \in \mathbb{C}$ the jth filament's center in complex notation, $w_j = \lambda_j e^{i\phi_j}$ encodes its length $\lambda_j > 0$ and orientation angle $\phi_j \in [-\pi/2, \pi/2)$ with the positive real axis ($j = 1, \ldots, J$). For a given binning number $N \in \mathbb{N}$ we have the empirical histogram

$$\left(\sum_{j:\phi_j \in [-\pi/2+\pi k/N, -\pi/2+\pi(k+1)/N)} \lambda_j \Big/ \sum_{j=1}^{J} \lambda_j \right)_{k=0}^{N-1}. \tag{10}$$

We assume that there is a true underlying filament process such that

$$\mathbb{P}\{z \in A, \lambda \in B, \phi \in C\} = \int_{A \times B \times C} g(z, \lambda, \phi)\, dz\, d\lambda\, d\phi$$

for Borel sets $A \subset \mathbb{C}$, $B \subset (0, \infty)$ and $C \subset [-\pi/2, \pi/2)$ with a density g w.r.t. the corresponding Lebesgue measures. Then the histogram (10) is an estimator for the true conditional density

$$f(\phi) = \frac{\mathbb{E}[\lambda|\phi]}{\mathbb{E}[\lambda]}, \qquad \phi \in [-\pi/2, \pi/2). \tag{11}$$

This statistical model relates to the previous theoretical analysis as follows. For every observation of the filament process, there are n pixels carrying orientations and we denote by $X_l \in [-\pi/2, \pi/2)$ the orientation of the lth pixel, $1 \le l \le n$, where a pixel is multiple counted, each with the orientation of the corresponding filament, if two or more filaments intersect at this pixel. The binned histogram of $X_1, \ldots, X_n$ is given by (10). Similarly, the filament process is a point process carrying weighted (by relative filament length) orientation marks, where the distribution of weighted orientations $X \in [-\pi/2, \pi/2)$ has the density $\phi \mapsto f(\phi)$ given by (11).

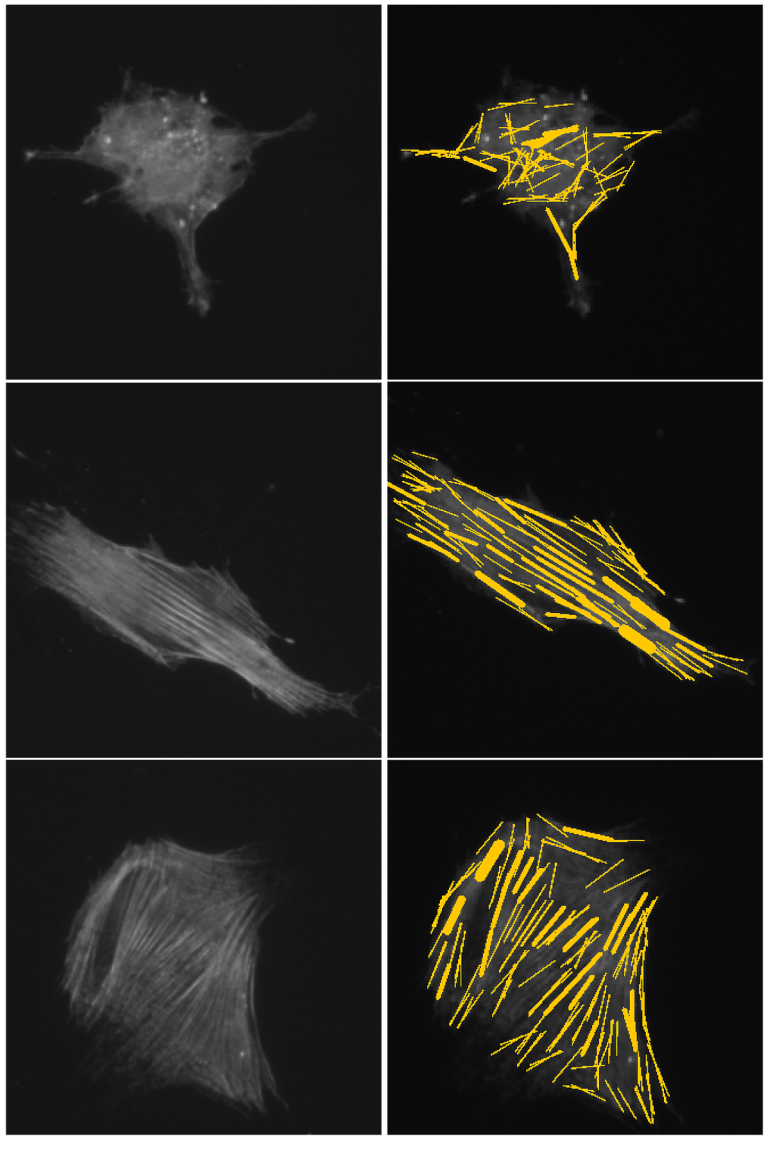

Figure 2. Fluorescence microscopy images of human mesenchymal stem cells, showing immuno-labeled myosin structures, cultured for 24 hrs. on matrices of varying young moduli $E_m = 1$ kPa (top row), $E_m = 11$ kPa (middle row) and $E_m = 34$ kPa (bottom row). Left column: raw fluorescence images, right column: superimposed traced filament structure. (Colors are visible in the online version of the article.)

Taking into account pixel and machine discretization, $X_1, \ldots, X_n$ can be viewed as a discretized sample of the weighted orientation mark X. If we double the orientations, we are in the situation of Assumption 3.1.

Empirical histograms and WiZer functions. From the empirical histograms with total number n of observed pixels (filament crossings result in pixels occurring in more than one filament; such pixels are accordingly multiply counted), as well as from the true density f we obtain densities $f_h^{(n)}$ and f_h, respectively, smoothed with the wrapped Gaussian kernel (2) of bandwidth $h > 0$. For the former, the WiZer signature functions W_h are computed as in (8). For the hMSCs from Figures 2, 3 depicts empirical histograms, corresponding $f_h^{(n)}$ for some values of h and the WiZer signature functions.

Note that not all of the modes of the smoothed empirical densities may be statistically significant at the specified level, cf. Figure 4.

Inferred mode-persistence diagrams. In the next step, the inferred kth persistence bandwidths $\widehat{\mathrm{ssp}}_k^{(1)}$ for modes of $f_h^{(n)}$ have been computed from the WiZer signature functions. Recall that $\widehat{\mathrm{ssp}}_k^{(1)}$ gives the inferred bandwidth of birth of the kth mode which is for $k > 1$ the split bandwidth of the $(k-1)$st mode. For the three hMSCs of the previous Figures 3 and 4, Table 1 depicts the first few inferred mode birth bandwidths in degrees.

Finally, we depict inferred mode persistence bandwidths graphically. Typically, in logarithmic scale, inferred persistences of several modes can be included in a single diagram such that the verticals measure inferred persistence of odd order modes while the horizontals measure inferred persistence of even order modes, cf. Figure 5. With confidence level $1 - \alpha = 0.95$, the black cell has least persistent first modes and most persistent second and higher modes. This is reversed for the red cell: most persistent first mode and least persistent second and third mode. The blue cell's mode persistences are intermediate. For all cells, the fourth mode comes to lie on the first diagonal which accounts for the fact that fourth modes have not been detected for any cell at the given resolution level (200 steps between the bandwidth reported in the WiZer signature function), cf. Figure 1.

Data analysis. We have now applied the above analysis to the data set of a total of 179 cells: 60, 58 and 61 on each elasticity (1 kPa, black, 11 kPa, red, 34 kPa, blue). Each of the three datasets comes in two files reflecting two experimental batches. For each of these six files, common parameters (reflecting average intensity, blur and noise of the specific experimental setup) for the filament sensor (cf. Eltzner *et al.* [18]) have been determined by an expert. Figure 6 depicts the boxplots for the first 9 inferred mode persistences and Figure 7 the persistence diagram for the means of the first seven, all in logarithms of radians. Recalling that in the persistence diagram, odd order modes are more persistent if they have a higher vertical component, even order are so if their horizontal component is higher, the trend observed in Figure 5 is consolidated. In mean and median,

1. cells on 1 kPa (black) have least persistent first modes and most persistent next higher modes,
2. cells on 11 kPa (red) have most persistent first modes and least persistent next higher modes,
3. cells on 34 kPa (blue) range close to red cells with a clear tendency (except for the first mode) toward black cells.

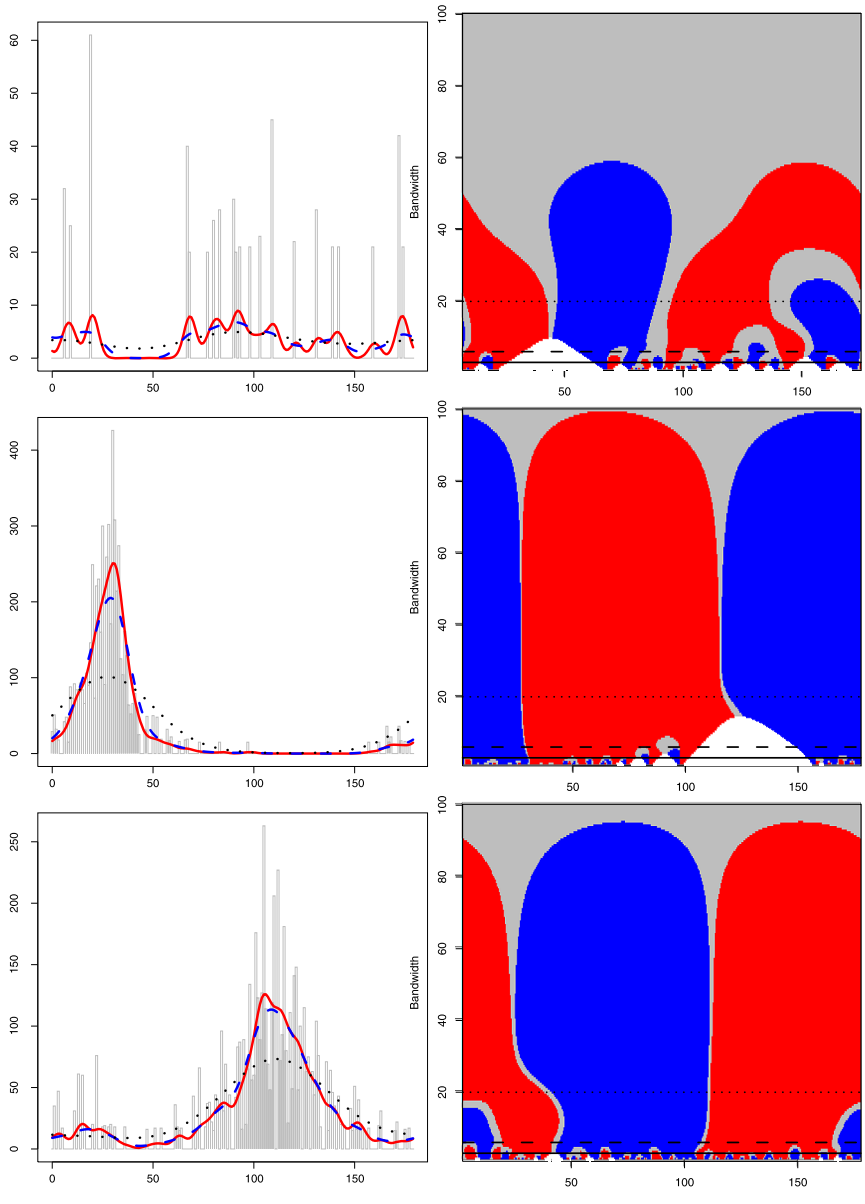

Figure 3. Processed traced filament structures from Figure 2 (rows correspond). Left column: histograms of total filament length × width (vertical) per angle orientation (horizontal) with smoothed densities via wrapped Gaussian kernels of bandwidths 3 (solid red), 6 (dashed blue) and 20 (dotted black). Right column: the respective WiZer signature functions to the significance level $\alpha = 0.05$ (horizontal: angle orientation, vertical: bandwidth). Here, blue depicts significant increase ($W_h^{(1)}(z) = 1$), red significant decrease ($W_h^{(1)}(z) = -1$), grey regions where neither increase nor decrease is significant ($W_h^{(1)}(z) = 0$) and white areas with too few data (effective sample size $\sum_{i=1}^{n} K_h(x - x_i)/K_h(0) \leq 5$, Chaudhuri and Marron [9], page 812). The three bandwidths from the left column are horizontal lines in the right. (Colors are visible in the online version of the article.)

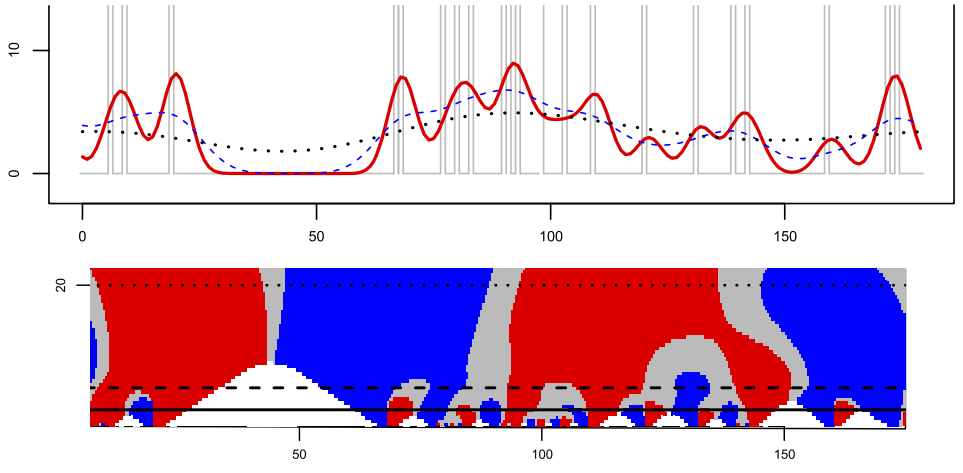

Figure 4. Detail of the first row of Figure 3. Careful investigation shows that only one of the two first modes of the kernel smoothed empirical density with bandwidth 3 (red, solid) is statistically significant. Similarly, the first mode of the kernel smoothed empirical density with bandwidth 6 (blue, dashed) is statistically insignificant. (Colors are visible in the online version of the article.)

5. Discussion

In this research, we have proposed a nonparametric methodology that can be applied to circular data to detect differences in shape features of underlying unknown circular densities. This method builds on a wrapped SiZer (WiZer) because we found that the wrapped Gaussian kernel is the one and only kernel (under reasonable assumptions) that preserves circular causality. In a natural way, with the W/SiZer's methodology comes a notion of shape feature persistences. We have used this to propose a measure assessing early differentiation of human mesenchymal stem cells (hMSCs). Our results warrant larger studies involving larger sample sizes. In such studies, we expect that inferred mode persistence diagrams (IMPDs) and combinations of IMPDs with other cell parameters will give a useful tool to measure early differentiation of hMSCs also over time.

The key benefit of the W/SiZer methodology, providing for IMPDs, is also a key limitation: statistical inference is possible only above a specific minimal bandwidth h_0, which is of course free to choose, although it may affect the asymptotic approximation, cf. Theorem 3.6 and Corol-

Table 1. Inferred bandwidths $\widehat{\mathrm{ssp}}_k^{(1)}$ for $k = 1, \ldots, 6$ and $k = 16$ in degrees of births of modes with confidence $1 - \alpha = 0.95$ for the three cells from Figure 2. The first four are depicted in log-scale in Figure 5

Mode no.	1	2	3	4	5	6	⋯	16	⋯
Cells on 1 kPa (black)	58.420	26.245	8.425	4.960	4.960	4.465		0.573	
Cells on 11 kPa (red)	99.010	5.455	3.970	2.485	2.485	2.485		1.495	
Cells on 34 kPa (blue)	95.050	15.355	4.960	3.970	3.970	3.475		1.990	

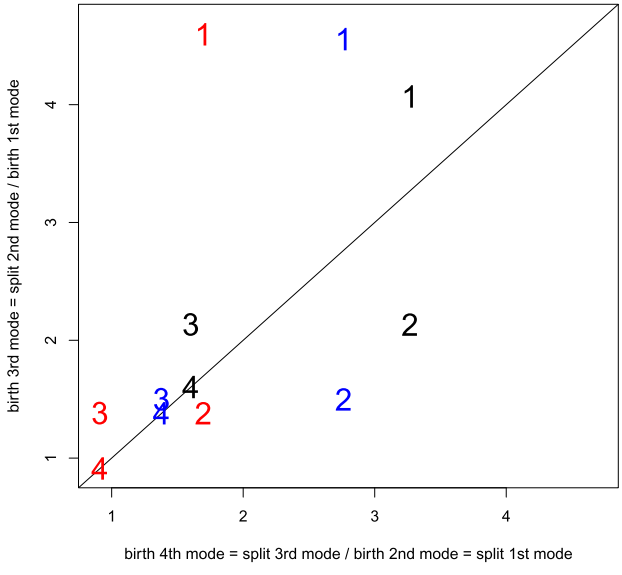

Figure 5. Inferred mode-persistence diagram for the first four modes (cf. Table 1) of the three cells of Figure 2 in log-scale. Colors indicate cell number (cell 1: black, cell 2: red, cell 3: blue) and coordinates of numbers indicate (birth, split) for even number of modes (1 and 3) and (split, birth) for odd number of modes (2 and 4). Odd order modes are more persistent if they have a higher vertical component, even order are so if their horizontal component is higher. (Colors are visible in the online version of the article.)

lary 3.7. This is due to the fact that the corresponding asymptotic distributions are not tight as $h_0 \to 0$ and a suitable calibration as $h_0 \to 0$ and $n \to \infty$ would have to be found that allows to compensate for nontightness (cf. Chaudhuri and Marron [10], page 420). For the line, this problem has been recently solved by Schmidt-Hieber *et al.* [45], now allowing for inference on shape parameters simultaneously over all scales (see also the discussion in Panaretos *et al.* [40] for a similar reasoning in the context of flow estimation). We expect that these arguments can be extended to our setup.

Appendix A: Sign changes

Let S^- denote the usual sign change functional as defined in, for example, Brown *et al.* [6], Definition 2.1, Karlin [28], that is, for real numbers $x_1, \ldots, x_n$, the expression

$$S^-(x_1, \ldots, x_n)$$

simply counts the number of changes from positive to negative values in the sequence $x_1, \ldots, x_n$, or vice versa, ignoring zeros. We now follow Mairhuber *et al.* [32].

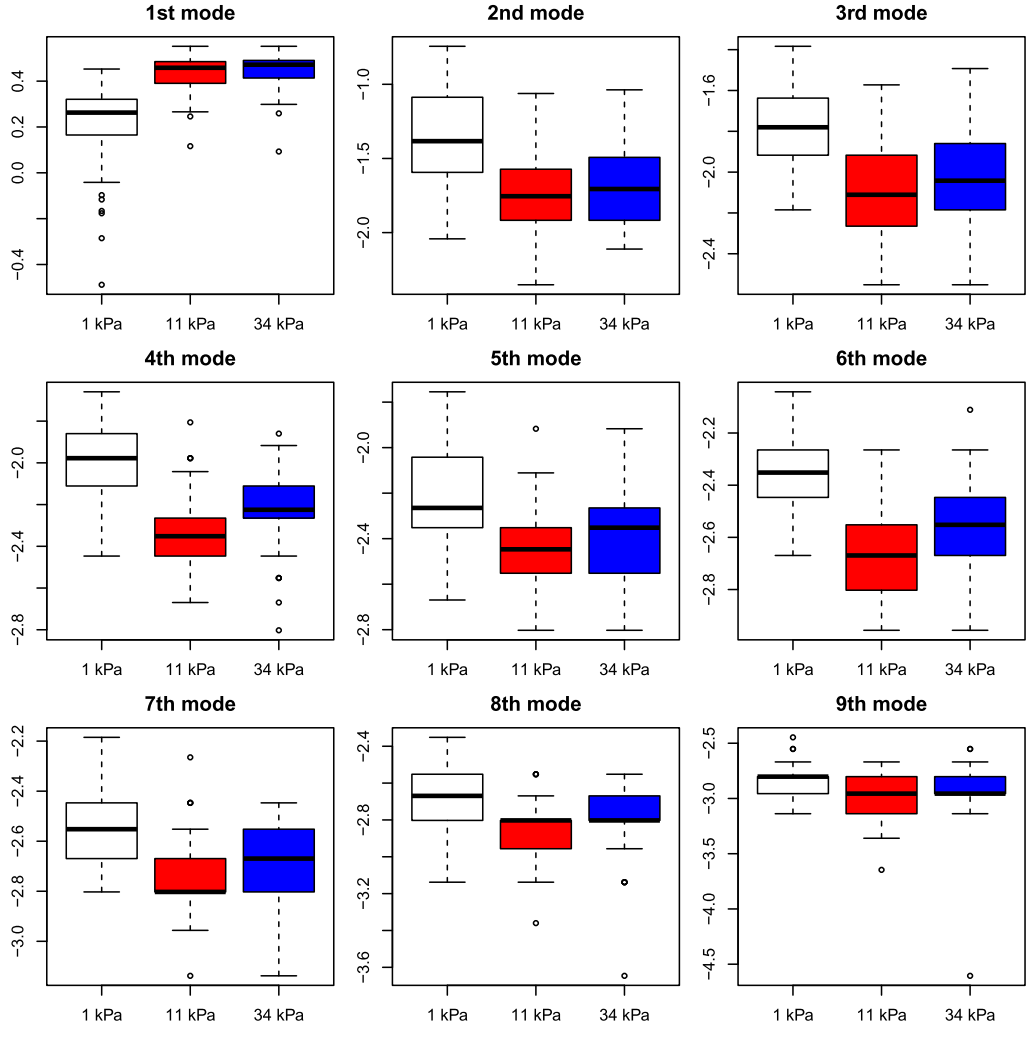

Figure 6. Boxplots of inferred mode-persistences in log-scale radians for the first nine modes for a sample of 179 cells in total exposed for 24 hours to matrix rigidities of 1 kPa (black, 60 cells), 11 kPa (red, 58 cells) and 34 kPa (blue, 61 cells). (Colors are visible in the online version of the article.)

Definition A.1. *The number of* cyclic sign changes *for a vector* $\mathbf{x} = (x_1, \ldots, x_n)$ *is defined as follows*: *If* $\mathbf{x} = 0$ *set* $S_c(\mathbf{x}) = 0$. *If there is a nonzero entry, say* $x_j \neq 0$, *then set*

$$S_c(\mathbf{x}) = S^-(x_j, x_{j+1}, \ldots, x_n, x_1, \ldots, x_{j-1}, x_j).$$

Remark A.2. Obviously, this definition does not depend on the choice of the nonzero entry x_j. Moreover, as can be easily seen by induction, the number of cyclic sign changes is always even.

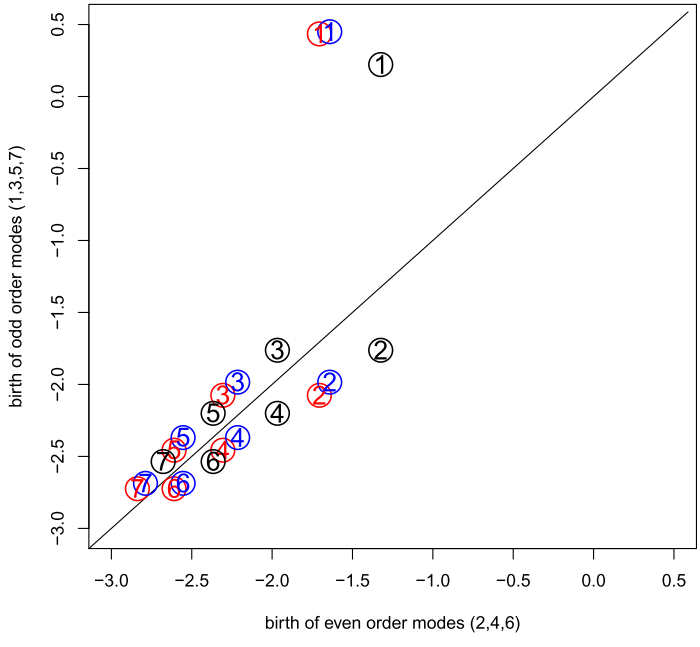

Figure 7. Inferred mode-persistence diagram in log-scale radians for the means of the first seven modes of the data set from Figure 6 with same coloring scheme. (Colors are visible in the online version of the article.)

We say that a vector $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{S}^n$ with distinct entries is in *cyclic order* if there exist real numbers

$$t_1 < t_2 < \cdots < t_n < t_1 + 2\pi$$

such that $z_l = e^{it_l}$ for $l = 1, \dots, n$.

Definition A.3. *For a function* $f : \mathbb{S} \to \mathbb{R}$ *and points* $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{S}^n$ *in cyclic order we set*

$$f(\mathbf{z}) = \big(f(z_1), \dots, f(z_n)\big).$$

Then we define the number of cyclic sign changes of f *as*

$$S_c(f) = \sup_{\substack{n\in\mathbb{N},\mathbf{z}\in\mathbb{S}^n \\ \text{in cyclic order}}} S_c\big(f(\mathbf{z})\big).$$

Appendix B: Proofs

Proof of Theorem 2.6.

Lemma B.1. (VD) $\Leftrightarrow$ (MR).

Proof. The implication "⇒" is obvious. For "⇐", we show the assertion first for a continuous $f \in L^1(\mathbb{S})$. For this, set $F(e^{it}) = \int_0^t f(e^{is})\,ds$. Then $\lim_{t\downarrow 0}(F(e^{it}) - F(1))/t = f(1) = \lim_{t\uparrow 0}(F(e^{i(2\pi+t)}) - F(1))/t$ and $DF = f$, which gives as desired,

$$S_c(L * f) = S_c(L * DF) \le S_c(DF) = S_c(f).$$

Next, assume that $f \in L^1(\mathbb{S})$ is arbitrary with $S_c(f) = k < \infty$ (the assertion is trivial for $k = \infty$). Due to continuity of L, there is a sequence of continuous functions f_n such that $f_n \to f$ and $L * f_n \to L * f$ in $L^1(\mathbb{S})$ as $n \to \infty$. In fact, because $\mathbb{S}$ decomposes into k disjoint intervals (some of which may be points only) in which either sign $f \ge 0$ or ≤ 0, we may choose the functions f_n such that $S_c(f_n) \to S_c(f)$, that is, $S_c(f_n) = k$ for all $n \ge n_0$ for some $n_0 \in \mathbb{N}$. Now let $m \in \mathbb{N}$ and assume that $S_c(L * f) \ge m$. As $\mathbb{S}$ is compact, $L * f$ is uniformly continuous. In consequence, we have that $S_c(L * f_n) \ge m$ for n sufficiently large. Hence, as all f_n are continuous, $m \le k$. In particular, this shows that $S_c(L * f) \le k$, concluding the proof. □

To complete the proof of Theorem 2.6, we show the following.

Lemma B.2. *Suppose that* $\{L_h : h > 0\}$ *is a differentiable semigroup of circular variation diminishing kernels. Then it is not enhancing local extrema.*

Proof. Assume that f is some smooth function and z_0 is an isolated maximum of the function $z \mapsto L_{h_0} * f(z)$ for some $h_0 > 0$. Let $C = L_{h_0} * f(z_0)$ and $\mathbf{z} = (z_1, \dots, z_k)$ be such that

$$S_c\big(L_{h_0} * f(\mathbf{z}) - C\big) = S_c(L_{h_0} * f - C).$$

We find $w_1, w_2 \in \mathbb{S}$ such that

$$\tilde{\mathbf{z}} = (z_1, \dots, w_1, z_0, w_2, \dots, z_k)$$

is in cyclic order and $L_{h_0} * f(w_j) - C < 0$ for $j = 1, 2$.

Assume now that $\partial_h L(z_0, h_0) > 0$ were true then we would find $h' > 0$ small enough such that

$$L_{h_0+h'} * f(z_0) - C > 0$$

and also, because $h \mapsto L_h * f(z)$ is continuous for fixed z,

$$\operatorname{sign}\big(L_{h_0+h'} * f(z) - C\big) = \operatorname{sign}\big(L_{h_0} * f(z) - C\big) \qquad \text{for } z = w_1, w_2, z_1, \dots, z_n.$$

This would imply

$$\begin{aligned} S_c(L_{h_0+h'} * f - C) \ge S_c\big(L_{h_0+h'} * f(\tilde{\mathbf{z}}) - C\big) &= S_c\big(L_{h_0} * f(\mathbf{z}) - C\big) + 2 \\ &= S_c(L_{h_0} * f - C) + 2, \end{aligned}$$

a contradiction, since $L_{h_0+h'} * f - C = L_{h'} * (L_{h_0} * f - C)$. The analog argument works for local minima. □

Proof of Theorem 2.8. With Remark 2.2, we are left to show that the wrapped Gaussian is variation diminishing. We proceed as follows. First, we note that the Gaussian kernels give an exponential family on the line and that exponential families give rise to variation diminishing kernels on the line (cf. Brown *et al.* [6], Example 3.1). Here, we use *variation diminishing* on $\mathbb{R}$ as introduced by Karlin [28] and Brown *et al.* [6]. Secondly, note that kernels made from exponential families on the line are *mean continuous* in the following sense.

A function $K : \mathbb{R} \to [0, \infty)$ with $\int_{\mathbb{R}} K(t)\,dt = 1$ is a *mean continuous kernel* if

$$2K(t) = \lim_{h \downarrow 0} \big(K(t+h) + K(t-h)\big) \qquad \text{for all } t \in \mathbb{R}.$$

Then the assertion of Theorem 2.8 follows from the following.

Theorem B.3. *Let K be a mean continuous kernel that is variation diminishing on the line. Then its* wrapped kernel

$$\widetilde{K}\big(e^{it}\big) = \sum_{k \in \mathbb{Z}} K(t + 2\pi)$$

is variation diminishing on the circle.

Proof. This is a consequence of the following two theorems of Mairhuber *et al.* [32]. To this end recall the Laguerre–Pólya class of entire functions $\Psi : \mathbb{C} \to \mathbb{C}$ of form

$$\Psi(s) = e^{-\alpha s^2 + \beta_0 s} \prod_{k=1}^{\infty} (1 + \beta_k s) e^{-\beta_k s},$$

where $\alpha \geq 0$ and β_k are real such that

$$0 < \alpha + \sum_k \beta_k^2 < \infty.$$

The first theorem connects functions in the Laguerre–Pólya class to mean continuous variation diminishing kernels on the line.

Theorem B.4 (Mairhuber *et al.* [32], Theorem B). *If Ψ belongs to the Laguerre–Pólya class then there exists an $\varepsilon > 0$ such that $\frac{1}{\Psi}$ is holomorphic in the open strip*

$$S = \big\{ z \in \mathbb{C} : -\varepsilon < \Re(z) < \varepsilon \big\}. \tag{12}$$

Further, there exists a mean continuous variation diminishing kernel $\Lambda : \mathbb{R} \to \mathbb{R}$ such that

$$\frac{1}{\Psi(z)} = \int_{-\infty}^{\infty} e^{-zt} \Lambda(t)\,dt \tag{13}$$

holds for all $z \in S$. Conversely, if $\Lambda : \mathbb{R} \to \mathbb{R}$ is a mean continuous variation diminishing kernel then there exists a function Ψ in the Laguerre–Pólya class such that equation (13) holds in an open strip of the form (12).

The second theorem completes the proof of Theorem B.3 as it links Laguerre–Pólya functions to circular variation diminishing kernels via wrapping.

Theorem B.5 (Mairhuber *et al.* [32], Theorem 3). *Every element Ψ of the Laguerre–Pólya class gives rise to a mean continuous cyclic variation diminishing kernel via*

$$\Omega\big(e^{it}\big) = \sum_{k\in\mathbb{Z}} \frac{1}{\Psi(ik)} e^{ikt}.$$

If $\Lambda(t)$ is a function for which (13) holds then we have the identity

$$\Omega\big(e^{it}\big) = 2\pi \sum_{k\in\mathbb{Z}} \Lambda(t + 2\pi k).$$

□

Proof of Theorem 2.9. In this proof, we follow the argument of Lindeberg [31] starting with the following lemma which is a special case of Engel and Nagel [19], Lemma 1.3. and Theorem 1.4. To this end, recall the property of strong continuity (SC) introduced in Remark 2.10.

Lemma B.6. *For a strongly continuous semigroup of kernels $L_h : \mathbb{S} \to \mathbb{R}$, $h > 0$ there is a (not necessarily bounded) closed linear operator $\mathcal{A} : L^2(\mathbb{S}) \to L^2(\mathbb{S})$ defined on a dense domain $\mathcal{D}(\mathcal{A}) \subset L^2(\mathbb{S})$, such that for $f \in \mathcal{D}(\mathcal{A})$ we have*

$$\mathcal{A}f = \lim_{h\downarrow 0} \frac{L_h * f - f}{h},$$

as well as for all $h > 0$ that

$$\partial_h (L_h * f) = \mathcal{A}(L_h * f). \tag{14}$$

The operator $\mathcal{A}$ is called the *infinitesimal generator* of the semigroup $\{L_h : h > 0\}$. By an argument analogous to that of Remark 2.3, we see $C^\infty(\mathbb{S}) \subset \mathcal{D}(\mathcal{A})$.

For $w \in \mathbb{S}$ let $\Delta_w : C^\infty(\mathbb{S}) \to C^\infty(\mathbb{S})$ be the shift operator defined by $\Delta_w f(z) = f(w^{-1}z)$. Then it is easy to see that $\mathcal{A}$ commutes with Δ_w since

$$\Delta_w \left(\frac{L_h * f - f}{h} \right) = \frac{L_h * (\Delta_w f) - \Delta_w f}{h}.$$

Moreover, the identity $R(L_h * f - f) = L_h * (Rf) - Rf$ (by symmetry of L_h) implies that $\mathcal{A}$ also commutes with the reflection operator R, defined via $Rf(z) = f(z^{-1})$.

In the next step, we show that $\mathcal{A}$ is a differential operator. To this end, we exploit Peetre [41], Théorème, stating that $\mathcal{A}$ is a differential operator if and only if it has the following property (of a *local* operator)

> $\mathcal{A}f(z_0) = 0$ for any $z_0 \in \mathbb{S}$ and smooth function $f : \mathbb{S} \to \mathbb{R}$ vanishing in a neighborhood U of z_0.

Indeed, assuming that the smooth function $f : \mathbb{S} \to \mathbb{R}$ vanishes in a neighborhood U of z_0, let $g : \mathbb{S} \to \mathbb{R}$ be smooth, having a nondegenerate maximum at z_0 with support in U. In consequence, for every $\beta \in \mathbb{R}$ the function $r = \beta f + g$ has a nondegenerate maximum at z_0, and hence $\partial_h|_{h=0}(L_h * r)(z_0) \leq 0$ by Lemma B.2 which implies by Lemma B.6 that $\mathcal{A}(r)(z_0) \leq 0$. By linearity,

$$\beta \mathcal{A}(f)(z_0) + \mathcal{A}(g)(z_0) \leq 0,$$

where the left-hand side can be given any sign (as $\beta \in \mathbb{R}$) unless $\mathcal{A}(f)(z_0) = 0$. In consequence, we have shown

$$\mathcal{A} = \sum_{0 \leq k \in \mathbb{Z}} a_k D^k \tag{15}$$

with suitable $a_k \in \mathbb{R}$, $(k = 0, 1, \ldots)$ at least in each chart. Using $\mathcal{A}\Delta_w = \Delta_w \mathcal{A}$ we can conclude that (15) holds globally.

In the penultimate step, we show that $a_k = 0$ in (15) for $k > 2$. Let f be a smooth function on $\mathbb{S}$ such that

$$f\big(\exp(it)\big) = t^2 + \beta t^n$$

for t in some neighborhood of zero and for some integer $n > 2$. Since f has an isolated local minimum at 1 for all β we have by Lemmata B.2 and B.6 that $\partial_h|_{h=0}(L_h * f)(1) \geq 0$ (arguing as above), and hence $\mathcal{A}f(1) \geq 0$. On the other hand, we have $\mathcal{A}f(1) = a_n \beta$ and hence $a_n \neq 0$ would give a contradiction.

If we set $n = 0$ in the above argument, we conclude that $a_0 = 0$.

Finally, the identity $R\mathcal{A} = \mathcal{A}R$ yields $a_1 = 0$ and, therefore, $\mathcal{A} = a_2 D^2$. In consequence, for $f \in C^\infty(\mathbb{S})$, (14) is a multiple of the heat equation (3) with initial condition f, which proves the theorem. □

Proof of Theorem 3.4. In order to see monotonicity, let $g_h(z)$ be either $D^m f_h(z)$ or $D^m f_h^{(n)}(z)$ for $h > 0$ corresponding to case (i) and for case (ii) set $g_0(z) = D^m f(z)$. Now let $h_2 > h_1 > 0$ in case (i) or $h_2 > h_1 \geq 0$ in case (ii). Then $g_{h_2} = K_{h_2 - h_1} * g_{h_1}$ and hence

$$S_c(g_{h_2}) \leq S_c(g_{h_1}) \tag{16}$$

since K_h is cyclic variation diminishing for all $h > 0$ by Theorem 2.9.

To see right continuity, suppose that $S_c(g_h) = 2k$ for integer $k \geq 0$ and some $h > 0$ in case (i) or $h \geq 0$ in case (ii). If $k > 0$, there exists $\mathbf{z} = (z_1, \ldots, z_{2k}) \in \mathbb{S}^{2k}$ such that $\operatorname{sgn} g_h(z_j) = (-1)^j$. Since $h \mapsto g_h(z)$ is continuous for all $z \in \mathbb{S}$ there is an $\varepsilon > 0$ such that $\operatorname{sgn} g_{h'}(z_j) = (-1)^j$ for all $h' \in [h, h + \varepsilon)$. Therefore,

$$S_c(g_{h'}) \geq S_c\big(g_{h'}(\mathbf{z})\big) = 2k.$$

Together with the monotonicity (16), this proves $S_c(g_{h'}) = 2k$, and thus right continuity. If $k = 0$, we have at once $S_c(g_{h'}) \geq 0 = S_c(g_h)$ for all $h' \geq h$ again yielding right continuity. □

Proof of Theorem 3.6. We first show convergence of the finite dimensional distributions (I), then weak convergence (II), and finally a.s. continuity of sample paths.

I: Let $z_1, \dots, z_k \in \mathbb{S}$ and $t_1, \dots, t_k \in \mathbb{R}$ and set

$$Y_l = \sum_{i=1}^{k} t_i \big(D^m K_h\big(z_i X_l^{-1}\big) - D^m f_h(z_i)\big), \qquad l = 1, \dots, n.$$

Then $EY_l = 0$ and

$$\operatorname{var} Y_l = \frac{1}{n} \sum_{i,j=1}^{k} t_i t_j \mathfrak{c}(z_i, z_j; h).$$

In consequence, we have that

$$Z_n := n^{1/2} \sum_{i=1}^{k} t_i \big(D^m f_h^{(n)}(z_i) - D^m f_h(z_i)\big) = n^{-1/2} \sum_{l=1}^{n} Y_l$$

is asymptotically normal with zero mean and covariance $\sum_{i,j=1}^{k} t_i t_j \mathfrak{c}(z_i, z_j; h)$. Since $t_1, \dots, t_k \in \mathbb{R}$ are arbitrary, by the Cramér–Wold device we conclude that

$$U_n(z_1, \dots, z_k) := \big(n^{1/2} \big(D^m f_h^{(n)}(z_i) - D^m f_h(z_i)\big)\big)_{i=1}^{k}$$

is asymptotically normal with zero mean and covariance matrix $(\mathfrak{c}(z_i, z_j; h))_{i,j=1}^{k}$.

II: To establish weak convergence, it now suffices to show that

$$n^{1/2} \big(D^m f_h^{(n)}(z) - D^m f_h(z)\big)_{z \in \mathbb{S}}$$

is tight. To this end, we give a bound for the second moments of the increments. Observe that for $z_1, z_2 \in \mathbb{S}$,

$$\begin{aligned} E\big(U_n(z_1) - U_n(z_2)\big)^2 &= \operatorname{var}\big\{ D^m K_h\big(z_1 X^{-1}\big) - D^m\big(K_h\big(z_2 X^{-1}\big)\big)\big\} \\ &\le E\big\{ D^m K_h\big(z_1 X^{-1}\big) - D^m\big(K_h\big(z_2 X^{-1}\big)\big)\big\}^2 \\ &\le \big\| D^{m+1} K_h \big\|_\infty^2 \, d(z_1, z_2)^2 \end{aligned} \tag{17}$$

using the mean value theorem for the last inequality where d denotes the intrinsic (geodesic) distance in $\mathbb{S}$. Now we argue with Theorem 8.6 (page 138) and Theorem 8.8 (page 139) of Ethier and Kurtz [21]. First note that (17) yields condition (8.36) of Theorem 8.8. which asserts condition (8.29) of Theorem 8.6 as well as condition (8.39). Taking the conditional expectation of the latter gives condition (8.28) of Theorem 8.6. Directly from (17), we infer the third condition (8.30) of (b) of Theorem 8.6. From the fact that in particular for every fixed $z \in \mathbb{S}$, $U_n(z)$ is asymptotically normal, we obtain condition (a) of Theorem 7.2 (Ethier and Kurtz [21], page 128). Thus, we conclude with Theorem 8.6. that the sequence $U_n(z)$, $n \in \mathbb{N}$ of processes $z \in \mathbb{S}$ is relatively compact, yielding the desired convergence result.

III: Since G_h is the weak limit of a sequence of a.s. continuous processes, there is a version of G_h that also has continuous sample paths with probability one.

Proof of Theorem 3.8. Let $\mathcal{S}_0 \subset \{(h,z) : h \geq h_0, z \in \mathbb{S}\}$ for which $H_0^{(h,z)}$ holds. Then by Corollary 3.7, the probability that one or more of the null hypotheses are falsely rejected is

$$\begin{aligned} &P\big(\exists (h,z) \in \mathcal{S}_0 \text{ s.t. } \big|D^m f_h^{(n)}(z)\big| \geq n^{-1/2} q_{(1-\alpha)}\big) \\ &\leq P\Big(\sup_{h\geq h_0, z\in\mathbb{S}} n^{1/2}\big|D^m f_h^{(n)}(z) - D^m f_h(z)\big| \geq q_{(1-\alpha)}\Big) \\ &\to P\Big(\sup_{z\in\mathbb{S}} |G_{h_0}| \geq q_{(1-\alpha)}\Big) = \alpha \qquad \text{as } n \to \infty. \end{aligned}$$

For the last assertion, note that if $D^m f_h(z)$ has $2k \geq 2$ sign changes then we can find a vector $(z_1, \ldots, z_{2k})$ in cyclic order such that $\operatorname{sgn} D^m f_h(z_j) = (-1)^j$ $(j = 1, \ldots, 2k)$. Because of $q_{(1-\alpha)} n^{-1/2} \to 0$ and

$$\max_{1\leq j\leq 2k} \big|D^m f_h^{(n)}(z_j) - D^m f_h(z_j)\big| \leq \sup_{z\in\mathbb{S}} \big|D^m f_h^{(n)}(z) - D^m f_h(z)\big| = \mathcal{O}_P\big(n^{-1/2}\big)$$

the above test procedure will correctly conclude the sign of $D^m f_h(z_j)$ for all j asymptotically with probability one. □

Appendix C: Numerical considerations

Let $K : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a kernel on the real axis and $\tilde{K} : [-\pi, \pi] \to \mathbb{R}_{\geq 0}$ the corresponding wrapped kernel defined by $\tilde{K}(x) = \sum_{j\in\mathbb{Z}} K(x + 2\pi j)$ and for $C \in \mathbb{Z}$ the cut off kernel $\tilde{K}^{(C)}(x) = \sum_{|j|\leq C} K(x + 2\pi j)$.

For points $x, x_1, \ldots, x_n \in [-\pi, \pi]$, we define

$$f(x) = \frac{1}{n}\sum_{i=1}^n \tilde{K}(x - x_i), \qquad f^{(C)}(x) = \frac{1}{n}\sum_{i=1}^n \tilde{K}^{(C)}(x - x_i).$$

We estimate

$$\begin{aligned} &\big|f(x) - f^{(C)}(x)\big| \\ &= \left|\frac{1}{n}\sum_{i=1}^n \sum_{|j|>C} K(x - x_i + 2\pi j)\right| \leq \frac{1}{n}\sum_{i=1}^n \sum_{|j|>C} \big|K(x - x_i + 2\pi j)\big| \\ &\leq \sum_{j>C}\Big(\sup_{x\in[2\pi(j-1),2\pi(j+1)]} \big|K(x)\big| + \sup_{x\in[-2\pi(j+1),-2\pi(j-1)]} \big|K(x)\big|\Big) =: \delta_K(C). \end{aligned}$$

Now we want to estimate $\delta_K(C)$ for the derivative of the Gaussian kernel

$$\phi_h(x) = \frac{1}{\sqrt{2\pi h}} e^{-x^2/(2h)}.$$

Note that $\partial_x \phi_h$ has two extremal points at $\pm\sqrt{h}$ and is monotonically increasing on $(-\infty, -\sqrt{h})$ and monotonically decreasing on $(\sqrt{h}, \infty)$. Hence, if $2\pi(C-1) \geq \sqrt{h}$ we have

$$\sup_{x \in [-2\pi(j+1), -2\pi(j-1)] \cup [2\pi(j-1), 2\pi(j+1)]} \left|\partial_x \phi_h(x)\right| \leq -\frac{1}{2\pi} \int_{2\pi(j-2)}^{2\pi(j-1)} \partial_x \phi_h(y)\, dy$$

for all $j \geq C+1$ and, therefore,

$$\delta_{\partial_x \phi_h}(C) \leq -\frac{1}{\pi} \sum_{j \geq C+1} \int_{2\pi(j-2)}^{2\pi(j-1)} \partial_x \phi_h(y)\, dy = \frac{1}{\pi} \phi_h\big(2\pi(C-2)\big).$$

Theorem C.1. *For any empirical measure $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ with points $x_1, \ldots, x_n \in [-\pi, \pi]$, we have*

$$\left\| \frac{\partial(\tilde{\phi}_h * \nu)}{\partial x} - \frac{\partial(\tilde{\phi}_h^{(C)} * \nu)}{\partial x} \right\|_\infty \leq \frac{1}{\pi} \phi_h\big(2\pi(C-1)\big)$$

provided that $2\pi(C-1) \geq \sqrt{h}$.

In consequence, for the bandwidths $0.01 \leq h \leq 3.5$ considered in our applications in Sections 3.3 and 4, an error less than

$$\frac{e^{-10}}{\pi\sqrt{2\pi h}} \leq 10^{-4}$$

can be achieved for $2\pi(C-1) \geq \sqrt{20h}$, for example, $C = 3$.

Acknowledgments

All authors gratefully acknowledge support by DFG SFB 755 "Nanonscale Photonic Imaging" Projects B8, A4 and A6. Furthermore, M. Sommerfeld is grateful for support by the "Studienstiftung des Deutschen Volkes". A. Munk also gratefully acknowledges support by DFG FOR 916. F. Rehfeldt gratefully acknowledges funding through a Feodor–Lynen fellowship from the Alexander-von-Humboldt foundation. S. Huckemann and A. Munk are further indebted to the Volkswagen Stiftung. Finally, M. Sommerfeld and S. Huckemann are thankful for support from the SAMSI 2013–2014 program on Low-Dimensional Structure in High-Dimensional Systems.

References

[1] Ahmed, M.O. and Walther, G. (2012). Investigating the multimodality of multivariate data with principal curves. *Comput. Statist. Data Anal.* **56** 4462–4469. MR2957886

[2] Alvarez, L., Guichard, F., Lions, P.-L. and Morel, J.-M. (1993). Axioms and fundamental equations of image processing. *Arch. Ration. Mech. Anal.* **123** 199–257. MR1225209

[3] Babaud, J., Witkin, A.P., Baudin, M. and Duda, R.O. (1986). Uniqueness of the Gaussian kernel for scale space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8** 26–33.
[4] Balakrishnan, S., Fasy, B., Lecci, F., Rinaldo, A., Singh, A. and Wasserman, L. (2013). Statistical inference for persistent homology. Preprint. Available at arXiv:1303.7117.
[5] Briggs, A.J., Detweiler, C., Mullen, P.C. and Scharstein, D. (2004). May. scale-space features in 1D omnidirectional images. In *Proc. Fifth Workshop on Omnidirectional Vision, Camera Networks and Nonclassical Cameras* (P. Sturm, T. Svoboda and S. Teller, eds.) 115–126. Prague, Czech Republic.
[6] Brown, L.D., Johnstone, I.M. and MacGibbon, K.B. (1981). Variation diminishing transformations: A direct approach to total positivity and its statistical applications. *J. Amer. Statist. Assoc.* **76** 824–832. MR0650893
[7] Bubenik, P. and Kim, P.T. (2007). A statistical approach to persistent homology. *Homology, Homotopy Appl.* **9** 337–362. MR2366953
[8] Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 255–308. MR2476414
[9] Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94** 807–823. MR1723347
[10] Chaudhuri, P. and Marron, J.S. (2000). Scale space view of curve estimation. *Ann. Statist.* **28** 408–428. MR1790003
[11] Cheng, M.-Y. and Hall, P. (1999). Mode testing in difficult cases. *Ann. Statist.* **27** 1294–1315. MR1740110
[12] Chung, M.K., Bubenik, P. and Kim, P.T. (2009). Persistence diagrams of cortical surface data. In *Information Processing in Medical Imaging* 386–397. Berlin-Heidelberg: Springer.
[13] Cohen-Steiner, D., Edelsbrunner, H. and Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* **37** 103–120. MR2279866
[14] Davies, P.L. and Kovac, A. (2004). Densities, spectral densities and modality. *Ann. Statist.* **32** 1093–1136. MR2065199
[15] Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961
[16] Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455
[17] Edelsbrunner, H., Letscher, D. and Zomorodian, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.* **28** 511–533. MR1949898
[18] Eltzner, B., Gottschlich, C., Wollnik, C., Huckemann, S. and Rehfeldt, F. (2015). The filament sensor for near real-time detection of cytoskeletal fiber structures. *PLoS ONE* **10** e0126346.
[19] Engel, K.-J. and Nagel, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations. Graduate Texts in Mathematics* **194**. New York: Springer. MR1721989
[20] Engler, A.J., Sen, S., Sweeney, H.L. and Discher, D.E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell* **126** 677–689.
[21] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838085
[22] Fisher, N.I. and Marron, J.S. (2001). Mode testing via the excess mass estimate. *Biometrika* **88** 499–517. MR1844848
[23] Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.* (*N.S.*) **45** 61–75. MR2358377
[24] Good, I.J. and Gaskins, R.A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. *J. Amer. Statist. Assoc.* **75** 42–73. MR0568579
[25] Hall, P., Minnotte, M.C. and Zhang, C. (2004). Bump hunting with non-Gaussian kernels. *Ann. Statist.* **32** 2124–2141. MR2102505

[26] Hartigan, J.A. and Hartigan, P.M. (1985). The dip test of unimodality. *Ann. Statist.* **13** 70–84. MR0773153
[27] Heo, G., Gamble, J. and Kim, P.T. (2012). Topological analysis of variance and the maxillary complex. *J. Amer. Statist. Assoc.* **107** 477–492. MR2980059
[28] Karlin, S. (1968). *Total Positivity. Vol. I*. Stanford, CA: Stanford Univ. Press. MR0230102
[29] Klemelä, J. (2006). Visualization of multivariate density estimates with shape trees. *J. Comput. Graph. Statist.* **15** 372–397. MR2256150
[30] Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Boston: Kluwer.
[31] Lindeberg, T. (2011). Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. *J. Math. Imaging Vision* **40** 36–81. MR2782119
[32] Mairhuber, J.C., Schoenberg, I.J. and Williamson, R.E. (1959). On variation diminishing transformations of the circle. *Rend. Circ. Mat. Palermo* (2) **8** 241–270. MR0113109
[33] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1828667
[34] Minnotte, M.C. (1997). Nonparametric testing of the existence of modes. *Ann. Statist.* **25** 1646–1660. MR1463568
[35] Minnotte, M.C. and Scott, D.W. (1993). The mode tree: A tool for visualization of nonparametric density features. *J. Comput. Graph. Statist.* **2** 51–68.
[36] Müller, D.W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. MR1147099
[37] Munk, A. (1999). Optimal inference for circular variation diminishing experiments with applications to the von-Mises distribution and the Fisher–Efron parabola model. *Metrika* **50** 1–17. MR1749579
[38] Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2013). CircSiZer: An exploratory tool for circular data. *Environmental and Ecological Statistics* **21** 143–159.
[39] Ooi, H. (2002). Density visualization and mode hunting using trees. *J. Comput. Graph. Statist.* **11** 328–347. MR1938139
[40] Panaretos, V.M., Pham, T. and Yao, Z. (2014). Principal flows. *J. Amer. Statist. Assoc.* **109** 424–436. MR3180574
[41] Peetre, J. (1959). Une caractérisation abstraite des opérateurs différentiels. *Math. Scand.* **7** 211–218. MR0112146
[42] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters – An excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204
[43] Rehfeldt, F., Engler, A.J., Eckhardt, A., Ahmed, F. and Discher, D.E. (2007). Cell responses to the mechanochemical microenvironment – implications for regenerative medicine and drug delivery. *Advanced Drug Delivery Reviews* **59** 1329–1339.
[44] Sakai, T. (1996). *Riemannian Geometry*. *Translations of Mathematical Monographs* **149**. Providence, RI: Amer. Math. Soc. MR1390760
[45] Schmidt-Hieber, J., Munk, A. and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812
[46] Schwartzman, A., Gavrilov, Y. and Adler, R.J. (2011). Multiple testing of local maxima for detection of peaks in 1D. *Ann. Statist.* **39** 3290–3319. MR3012409
[47] Schwartzman, A., Jaffe, A., Gavrilov, Y. and Meyer, C.A. (2013). Multiple testing of local maxima for detection of peaks in ChIP-Seq data. *Ann. Appl. Stat.* **7** 471–494. MR3086427
[48] Silverman, B.W. (1981). Using kernel density estimates to investigate multimodality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **43** 97–99. MR0610384
[49] Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.* **52** 3493–3500. MR2427365

[50] Wada, T., Gu, Y.H. and Sato, M. (1991). Scale-space filtering for periodic waveforms. *Systems and Computers in Japan* **22** 45–54.
[51] Weickert, J. (1998). *Anisotropic Diffusion in Image Processing*. *European Consortium for Mathematics in Industry*. Stuttgart: B. G. Teubner. MR1666943
[52] Weickert, J., Ishikawa, S. and Imiya, A. (1999). Linear scale-space has first been proposed in Japan. *J. Math. Imaging Vision* **10** 237–252. MR1695946
[53] Zemel, A., Rehfeldt, F., Brown, A.E.X., Discher, D.E. and Safran, S.A. (2010). Optimal matrix rigidity for stress-fibre polarization in stem cells. *Nat. Phys.* **6** 468–473.

Received April 2014 and revised November 2014