Working Papers in Statistics
No 2016:5

Department of Statistics
School of Economics and Management
Lund University

Distributions at random events

KRZYSZTOF PODGÓRSKI, LUND UNIVERSITY

IGOR RYCHLIK, CHALMERS UNIVERSITY OF TECHNOLOGY

Distributions at random events

Krzysztof Podgórski
Department of Statistics
Lund University
Email: Krzysztof.Podgorski@stat.lu.se

Igor Rychlik
Mathematical Sciences
Chalmers University of Technology
Email: rychlik@chalmers.se

***Abstract*—We review the generalized Rice formula approach to deriving long-run distributions of a variety of characteristics defined at random events defined on a stochastic process. While the approach stems from the same principle originally introduced by Rice for the level crossing intensity in a random signal, we show how it generalizes to more general contexts. Firstly, we discuss events defined on random surfaces through crossing levels of possibly multivariate valued stochastic fields. Secondly, the dynamics is introduced by adding time argument and introducing the concept of velocity measured at moving surface. Thirdly, extensions beyond the Gaussian model are shown by presenting effective models for sampling from the distribution of a non-Gaussian noise observed at instances of level crossing by a process driven by this noise. The importance of these generalizations for engineering applications is illustrated through examples.**

I. Introduction

In the physical world, a random function can be often conveniently described as a sequence of local maxima or minima, constituting a series of random waves. In fact, this is not only the matter of visual representation of the process but also many technologically important implications in such fields as metal fatigue caused by random vibrations, failure caused by excess load on a construction, etc., depend on the character of the process represented in such a random wave form. Similarly, random fields can be efficiently summarized by level crossing contours and some local maxima inside of these contours. One methodological approach to the so defined summaries of the fields is through event based statistical distributions of the stochastic process or field in hand. The basic (random) events in this theory are level crossings and local extremes and the central theoretical tool generalizations of the Rice formula. The latter allows to obtain the distribution of the process at the instants of level crossings or local extrema (which occur at the zero-crossing of the derivatives). Generalizations of the approach can go in several directions: multidimensional extensions, dynamical evolutions for models in space and time, more general than Gaussian models. All of them are of great importance for engineering applications.

In particular, in ocean engineering, it is important to know the joint distributions of characteristics for the apparent waves in dynamical models of the sea surface elevation. One can evaluate such distributions for wave height, length and period and study statistical properties of velocities both for the sea surface and for the envelope field based on this surface, see [1], [2]. In Figure 1, we see the definition of the wave characteristics based on level crossing events for one dimensional records together with the joint distribution of two characteristics computed by the means of Rice formula.

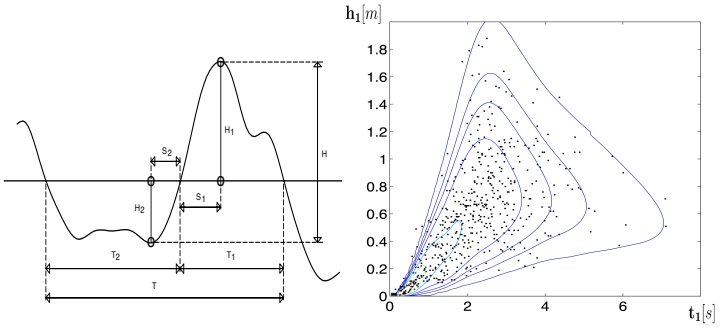


Fig. 1. Events defining wave characteristics for temporal record of the sea surface *(left)* and the joint distributions *(right)* of the crest height H_1 and half-period T_1 computed by means of the generalized Rice formula. The distributions are obtained for a Gaussian sea modeled by a spectrum obtained from an actual sea elevation record. Dots represent empirical values of T_1 and H_1 observed in the record.

The events can be defined by some non-linear functionals of a process. For example, the envelope process that wraps smoothly around the original process is convenient to study extremes or wave groups. The method was applied to the envelope, yielding the upper bound for the distribution of the maximum of a process and summarizing differences between individual waves and wave groups, see [3].

The most common applications are based on the Gaussian or closely related models. However, a variety of asymmetries is frequently observed in stochastic records. These asymmetries among other things can indicate that the underlying process is no longer Gaussian for which there is distributional symmetry both in time and in space. These asymmetries can be summarized by measures that are motivated by Rice's formula for crossing level distributions of the slope, see [4]. In Figure 2, we present a trajectory of a Gaussian (left) and non-Gaussian moving average process together with the zero crossing distribution and the crossing level distributions of the slope as a function of the level. It is well known that for the Gaussian process, the slope at a crossing level has the Rayleigh distribution and it does not depend on the level of the crossing, as seen in the figure. This is non longer true for non-Gaussian processes as clearly seen in the left hand side graphs. Notably, the covariance functions of both the processes are the same, as they are all moving averages with the shown kernel. Through this approach one can demonstrate not only distributional skewness but also more complex geometrical asymmetries in sample paths such as tilting, front-back slope asymmetry and time irreversibility. Non-Gaussian processes that exhibit such asymmetries find applications to model road topography and, in particular, the road surface roughness, [5]. The Slepian model for the non-Gaussian road profile can be

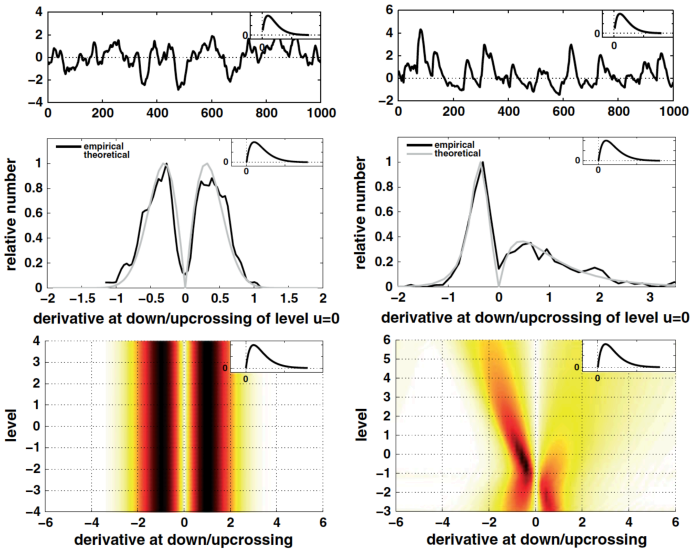


Fig. 2. Crossing distributions of the slope at the level crossings of Gaussian and non-Gaussian moving averages with the kernel shown in the upper-left corner of the graphs. The second process is driven by an asymmetric Laplace noise. Trajectories are shown in the top graphs. In the middle ones the empirical zero crossing distribution is compared with the one computed using the Rice's formula. The bottom-right graph shows asymmetries at different crossings level that are not present in the Gaussian case (left).

derived which accurately describes the response of a vehicle encountering extreme transients in the road profile, [6].

In this paper, an account of the recent advances in computing level crossing distributions is presented and illustrated by a number of applications to engineering sciences. The paper is starting with an introductory section presenting the generalized Rice formula. It is followed by two sections which demonstrates its applicability to different contexts. In the first section, the multivariate events for randomly evolving spatial surfaces are discussed with the emphasize on velocity distributions. The method is illustrated by velocity distribution of waves and group waves observed at the sea surface. The second section presents an approach to event based distributions for models involving non-Gaussian moving average. An application to models of the road surface and vehicle responses is presented to illustrate the effect of non-gaussianity.

II. Rice's formula and its generalization

It is a somewhat surprising and often confusing fact that if one samples observations of a stationary stochastic process $W(\mathbf{p})$ only at points $\mathbf{p}$ at which $W(\mathbf{p})$ equals zero, the obtained distribution is not equal to the conditional distribution of W given that $W(\mathbf{0}) = 0$. This is due to the fact that sampling at random events affects the distribution.

To discuss such a biased sampling distribution, first we should find how many points $\mathbf{p}$ satisfying $W(\mathbf{p}) = 0$ reside in an area (volume) of the unit size. The answer is given by the celebrated Rice formula, [7], [8]. Originally it was formulated as a one dimensional version of the problem, i.e. when W depended only on one variable, say, x.

Here and in the rest of the paper process W is assumed to be ergodic. Let $N(X)$ be the number of times W takes the value zero in $[0, X]$. Then if X increases without bound, the proportion of $N(X)$ to the length X converges with probability one to

$$\int |y| f_{\dot{W},W}(y,0)dy = \mathrm{E}\left[|\dot{W}|\Big|W=0\right]\cdot f_W(0), \qquad (1)$$

where $f_{\dot{W},W}$, f_W are the density functions of $(\dot{W}, W)$ and W, respectively. The above formulation is a combination of the ergodic theorem and the original Rice formula, which states that the average number of crossing $N(1)$, i.e. $\mathrm{E}(N(1))$ equals the right hand side of (1).

More generally, for the number $N(X, A)$ of times the process W takes value zero in $[0, X]$ and at the same time has a property A:

$$\lim_{X\to\infty}\frac{N(X,A)}{N(X)} = \frac{\mathrm{E}\left[\{W\in A\}|\dot{W}|\Big|W=0\right]}{\mathrm{E}\left[|\dot{W}|\Big|W=0\right]}, \qquad (2)$$

where the set $\{W \in A\}$ is identified with its indicator function. Consequently, the right hand side represents the biased sampling distribution when sampling is made over the 0-level contour $\mathcal{C}_0$. We denote this distribution by $\mathbf{P}(W \in A|\mathcal{C}_0) = \mathbf{P}(W(x+\cdot) \in A|x \in \mathcal{C}_0)$ and refer to it as the distribution of W on the contour $\mathcal{C}_0$.

These concepts extend to the vector valued and vector argument setup. Namely, the biased distribution of a vector valued field $\mathbf{V}(\mathbf{p})$, with the biased introduced by the vector field $\mathbf{W}(\mathbf{p})$ is denoted by $\mathbf{P}(\mathbf{V}(\mathbf{p}) \in A|\mathbf{p} \in \mathcal{C}_{\mathbf{w}}\}$ on the contour $\mathcal{C}_{\mathbf{w}} = \{\mathbf{p} : \mathbf{W}(\mathbf{p}) = \mathbf{w}\}$.

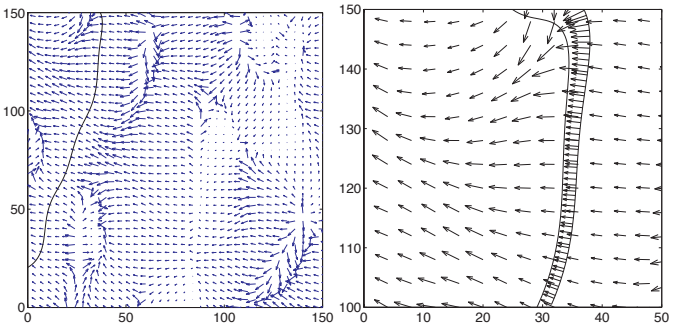


Fig. 3. *(Left)* Unbiased sample of velocities on the sea surface. *(Right)* Biased sample of velocities along the level crossing contours vs. unbiased sample. Here the level crossing contour is presented at the initial time 0 and then at the time dt. (The scale of axes on both the pictures is in meters, velocities are rescaled for clarity.)

The significance of this distribution follows from its interpretation, which is along the same argument as in the one dimensional case. The distribution represents the average size of the part of $\mathcal{C}_{\mathbf{w}}$ on which $\mathbf{V} \in A$ divided by the average size of the entire $\mathcal{C}_{\mathbf{w}}$. By the ergodic theorem, the so-defined distribution coincides with the limiting statistical distribution of $\mathbf{V}$ sampled on the contour $\mathcal{C}_{\mathbf{w}}$ over a large region.

The sampling interpretation in the multivariate case can be illustrated using the Gaussian model of sea surface $W(\mathbf{p}, t)$. The time variable t is considered to be fixed and thus by stationarity it can be set to zero. For this reason and for simplicity, it is dropped from the notation. The sea surface

$W(\mathbf{p})$ (at time zero) is taken to be the process used in defining contour $\mathcal{C}_w$. The distributions of the properly defined velocity $\mathbf{V}(\mathbf{p})$ in the gradient direction are of interest, see [2] for the velocity definition.

In Figure 3*(Left)*, an unbiased sample of velocities recorded on the entire field is presented, that could be used to estimate $\mathrm{P}(\mathbf{V}(\mathbf{p}) \in A)$. In Figure 3*(Right)*, an example of a biased sample is presented by velocities sampled along the contour $\mathcal{C}_0 = \{\mathbf{p} : W(\mathbf{p}) = 0\}$ of the fixed (zero) sea level. The sample distribution of the velocity vectors obtained along this contour represents (approximately if the area is large enough) the biased sampling distribution that could be used to estimate $\mathrm{P}(\mathbf{V}(\mathbf{p}) \in A | \mathbf{p} \in \mathcal{C}_0)$. As pointed in the above discussion, the two distributions of $\mathbf{V}$ are different.

Our aim is to present techniques for expressing sampling distributions in terms of the theoretical distribution of the involved processes that are based on a generalized multivariate Rice formula to which we turn next.

A. Generalized Rice formula

There exists vast literature on various generalizations of the Rice formula, see [9], [10], [11], [12] and references therein. Here we present only a generic formulation in the most general form and we refer to the literature for proofs and technical assumptions.

Consider a pair of jointly stationary stochastic processes $\mathbf{V}(\mathbf{p})$, $\mathbf{W}(\mathbf{p})$, $\mathbf{p} \in \mathbb{R}^k$, taking values in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Assume that $n \leq k$ and from now on treat them as fixed. Further let $\mathcal{V}$ stand for the relative volume in $\mathbb{R}^k$ of the dimension $k-n$ ($\mathcal{V}$ represents the length of a set if $k=3$ and the area if $n=2$, while for $n=3$ it simply counts points in a set). The distribution of $\mathbf{V}(\mathbf{p})$ on the contour $\mathcal{C}_{\mathbf{w}} = \{\mathbf{p} \in [0,1]^2 : \mathbf{W}(\mathbf{p}) = \mathbf{w}\}$ is given by

$$\mathrm{P}(\mathbf{V}(\mathbf{p}) \in A | \mathbf{p} \in \mathcal{C}_{\mathbf{w}}) \stackrel{def}{=} \frac{\mathrm{E}\left[\mathcal{V}\left\{\mathbf{p} \in \mathcal{C}_{\mathbf{w}} : \mathbf{V}(\mathbf{p}) \in A\right\}\right]}{\mathrm{E}\left[\mathcal{V}\left(\mathcal{C}_{\mathbf{w}}\right)\right]}$$

A generalized Rice formula is utilized to compute this distribution. Let $\mathbf{W}(\mathbf{p})$ have continuous finite dimensional distributions and let $f_{\mathbf{W}(\mathbf{0})}(\mathbf{w})$ be the density of $\mathbf{W}(\mathbf{0})$. Let $\dot{\mathbf{W}}(\mathbf{p})$ be the matrix of partial derivatives of $\mathbf{W}(\mathbf{p})$ (which are assumed to exist). The generalized determinant of this matrix is denoted by $\left|\dot{\mathbf{W}}(\mathbf{p})\right|$. This allows us to write a generic form of the Rice formula as

$$\begin{aligned}&\mathrm{E}\left[\mathcal{V}\left\{\mathbf{p} \in \mathcal{C}_{\mathbf{w}} : \mathbf{V}(\mathbf{p}) \in A\right\}\right] = \\ &\quad = \mathrm{E}\left(\left\{\mathbf{V}(\mathbf{0}) \in A\right\} \cdot \left|\dot{\mathbf{W}}(\mathbf{0})\right| \,\middle|\, \mathbf{W}(\mathbf{0}) = \mathbf{w}\right) f_{\mathbf{W}(\mathbf{0})}(\mathbf{w}).\end{aligned}$$

Notice that if the joint density of $\mathbf{V}$, $\mathbf{W}$, $\dot{\mathbf{W}}$ is available, which is always the case in this paper, the right hand side can be written simply in the form of an integral and the biased sampling distribution can be written as

$$\begin{aligned}&\mathrm{P}(\mathbf{V}(\mathbf{p}) \in A | \mathbf{p} \in \mathcal{C}_{\mathbf{w}}) = \\ &\qquad = \frac{\int_A \int f_{\mathbf{V},\dot{\mathbf{W}},\mathbf{W}}(\mathbf{v}, \dot{\mathbf{w}}, \mathbf{w}) \cdot \det\, \dot{\mathbf{w}}\; d\dot{\mathbf{w}} d\mathbf{v}}{\int f_{\dot{\mathbf{W}},\mathbf{W}}(\dot{\mathbf{w}}, \mathbf{w}) \cdot \det\, \dot{\mathbf{w}}\; d\dot{\mathbf{w}}}. \quad (3)\end{aligned}$$

Several aspect of this general formulation should be pointed out. Firstly, the process $\mathbf{V}(\mathbf{p})$ can be vector valued or scalar, for example $m=2$ or $m=1$. Secondly, the crossing contours can be also considered multidimensional, unidimensional or even zero-dimensional (points). Thus, in the case of sampling on a level crossing contour of two dimensional surface $W(\mathbf{p})$, $n=1$ and $k=2$. The determinant is then equal to $\left|\dot{\mathbf{W}}\right| = \sqrt{W_x^2 + W_y^2}$ and $\mathcal{V}$ measures the length of the contour on the plane. The sampling at level crossing points of, say, $W(x,0)$, i.e. along the line $y=0$, corresponds to $n=1$ and $k=1$. The determinant is then simply equal to $|W_x(x,0)|$ and $\mathcal{V}$ counts the number of level crossing points along the line $y=0$. Sampling is at the so-called specular points, i.e. points for which $(W_x(\mathbf{p}), W_y(\mathbf{p}))$ takes a specified value, implies that $n=2$ and $k=2$ and the generalized determinant of the 2×2 matrix of partial derivatives equals to $|W_{xx}W_{yy} - W_{xy}^2|$ with $\mathcal{V}$ counting points on the plane. Thirdly, the distribution of the underlying process does not need to be Gaussian and the formula can be utilized as long as we can evaluate the right hand side of (3).

These three types of biased sampling distributions can be of special interest for certain applications, which can be dictated by the nature of the problem in hand. *However here they were chosen mainly to illustrate techniques of deriving theoretical forms of distributions sampled at various cases of contours. The derivation should help to approach problems of finding sampling distributions on contours in many other multivariate situations of practical interest.*

III. Events for dynamically varying spatial fields

Sea surface elevation, atmospheric pressure, air pollution are examples from a variety of phenomena which can be modeled as a random two-dimensional field evolving in time. For such models it is of interest to describe statistical properties of the motions observed on the surface. This can be best achieved by studying appropriately defined velocities.

A. Velocity in the gradient direction

While there are many different ways of observing motion of a surface, here we focus on the velocity defined for crossing contours moving in the direction of the gradient of the surface, see [2] and the references therein. In this section, $W(\mathbf{p}, t)$ represents moving surface at point $\mathbf{p}$ and time t.

Definition 1: The velocity in the direction of gradient is denoted by $\mathbf{V}_{gr} = V_{gr}\mathbf{n}_\beta$, and is given by

$$\begin{bmatrix} W_x & W_y \\ -W_y & W_x \end{bmatrix} \mathbf{V}_{gr} = -\begin{bmatrix} W_t \\ 0 \end{bmatrix}. \quad (4)$$

This is the velocity in the variable direction between the gradient and the x-axis. We discuss distributions obtained by sampling over the entire field $W(\mathbf{p}, t)$, i.e. unbiased sampling distributions as well as distributions obtained by sampling at $\mathcal{C}_w = \{\mathbf{p} : W(\mathbf{p}, t) = w\}$.

The following vector, called *principal velocity*, enters as an important parameter

$$\mathbf{v}_{\max} = \begin{bmatrix} -\lambda_{101}/\lambda_{200} \\ -\lambda_{011}/\lambda_{020} \end{bmatrix}, \quad (5)$$

where *spectral moments* λ_{ijk} are defined as

$$\lambda_{ijk} = 2\int_{\Lambda^+} \lambda_1^i \lambda_2^j \lambda_3^k \, S(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 \, d\lambda_2 \, d\lambda_3,$$

with $\Lambda_+ = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 \geq 0\right\}$.

Next, the parameter $\gamma = \sqrt{\lambda_{020}/\lambda_{200}}$ which could be referred to as short-crestedness, equals the square root of the ratio of the intensities of zero-crossings along the y-axis and along the x-axis. We have $\gamma = 0$ for long-crested sea and $\gamma = 1$ for the most irregular, sometimes called short-crested or "confused" sea. The distributions of $\mathbf{V}_{gr}$ are given in the next result.

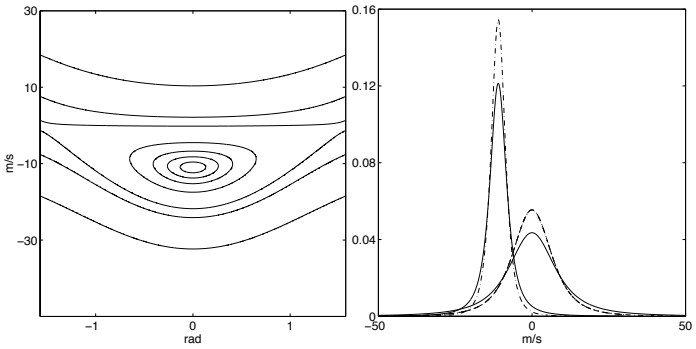


Fig. 4. *(Left)* The biased joint density of β and V_{gr} for the directional Gaussian sea. The isolines are drawn at the levels: 0.025, 0.01, 0.005, 0.001, 0.0005.. *(Right)* The conditional densities of V_{gr} given $\beta = 0, \pi/2$ for the biased (dashdot) and unbiased (solid) case.

Proposition 1: The speed V_{gr} in the direction of the gradient has distribution

$$\mathbf{v}_{\max}^T \mathbf{n}_\beta + \\ + T\sqrt{\frac{\lambda_{002}}{\lambda_{020}} - \frac{\lambda_{101}^2}{\lambda_{020}\lambda_{200}} - \frac{\lambda_{011}^2}{\lambda_{020}^2}}\sqrt{\gamma^2\cos^2\beta + \sin^2\beta},$$

where $\sqrt{n}T$ is t-distributed with n degrees of freedom and is independent of β. In the case of unbiased sampling distribution, $n = 2$ and β has the density

$$\frac{1}{2\pi}\frac{\gamma}{\gamma^2\cos^2\beta + \sin^2\beta},$$

For the case of biased sampling on $\mathcal{C}_w$, $n = 3$ and β has density

$$\frac{\gamma^2}{4\mathcal{E}(\sqrt{1-\gamma^2})}\frac{1}{\left(\gamma^2\cos^2\beta + \sin^2\beta\right)^{3/2}},$$

where $-\pi < \beta \leq \pi$ and the Legendre elliptic integral $\mathcal{E}(k) = \int_0^{\pi/2}\sqrt{1-k^2\sin^2\beta}\,d\beta$.

B. Envelope velocity

It should be noted that in the previous section we considered the velocity along a line with random azimuth. One can also consider a specified fixed direction and ask about velocity of moving in this direction. We apply this approach to illustrate difference in dynamics of a random field and its envelope defined as $E = \sqrt{W^2 + W_H^2}$, where W_H is the Hilbert transform of W, see Figure 5 for illustration and [3] for further details.

We are interested in the statistical distribution of V when measured at an arbitrarily selected point on the sea as well as the so-called biased distribution obtained by measuring this velocity on the fixed level contour. It is well known that these two distributions are essentially different, the first one is simply the distribution of random variable V while the second one has to be computed with a use of generalized Rice's formula. Relatively straightforward although tedious calculations [see Baxevani et al. (2002)] lead to the following form of the distributions of velocity V in the direction of the line $y = 0$:

$$-a \cdot \left(b + \sqrt{c \cdot d - e^2}\frac{X}{Y}\right),$$

$$\begin{aligned} a &= \frac{1}{\lambda_{200} - \lambda_{100}^2/\lambda_{000}}, \quad b = \lambda_{101} - \lambda_{100}\lambda_{001}/\lambda_{000}, \\ c &= \lambda_{200} - \lambda_{100}^2/\lambda_{000}, \quad d = \lambda_{002} - \lambda_{001}^2/\lambda_{000}, \\ e &= \lambda_{101} - \lambda_{100}\lambda_{001}/\lambda_{000}, \end{aligned}$$

variables X and Y are independent, X having the standard normal distribution while the distribution of Y is

a) the standard normal if we deal with the unbiased sampling,

b) the Rayleigh distribution if we deal with the biased sampling distribution of velocity V sampled at points $(\mathbf{p}, t)$ such that $E(\mathbf{p}, t) = u$.

For comparison, the analogous velocity of the sea surface has the same form but with the constants $a = 1/\lambda_{200}$, $b = \lambda_{101}$, $c = \lambda_{200}$, $d = \lambda_{002}$, and $e = \lambda_{101}$.

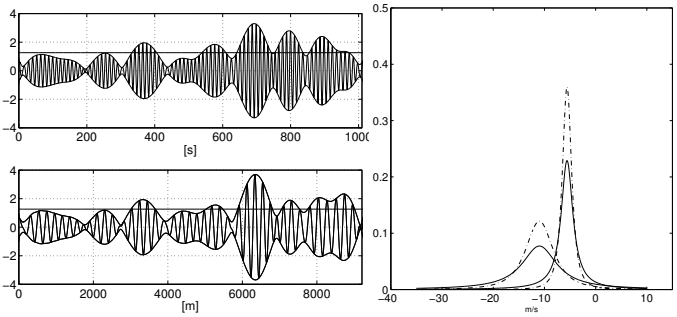


Fig. 5. *(Left)* Envelope process in space an in time. *(Right)* Velocities in along the principal wave direction for waves and wave groups.

C. Illustration

In this section, we illustrate the distributions derived in the previous section. We consider directional Gaussian sea given by the directional spectrum $S(\omega, \theta) = S(\omega)D(\omega, \theta)$ with frequency spectrum $S(\omega)$ being the JONSWAP spectrum with parameters: significant wave height $H_{1/3} = 7$ [m], peak period 11 [s] (peak frequency $\omega_p = 0.57$ [rad/s]) and shape parameter $\rho = 2.3853$, see [1] for details.

In Figure 4*(Left)*, the joint density of β and V_{gr}, for the biased sampling case, is illustrated. The biased density measures that part of the contour in which both the direction,

described by β, and the V_{gr} assume specified values. It is evident from this Figure that the biggest part of the contour is almost perpendicular to the main direction of wave propagation, while only a small part of the wave front is parallel to the main direction. Futhermore, it is also of interest to notice that the wave front moves with speed close to $\mathbf{v}_{max} = (-11, 0)$ (area included in the third isoline in Figure 4 *(Left)*. In Figure 4 *(Right)*, the densities of V_{gr} conditionally on $\beta = 0$ and $\beta = \pi/2$ are compared. The conditional densities $V_{gr} \Big| \beta = 0$ are well concentrated about $\mathbf{v}_{max}$. This is a consequence of the symmetry of the spreading function ($\lambda_{011} = 0$).

The case $\beta = \pi/2$, corresponds to the part of the contour with gradient parallel to the y-axis. Such a part, although small, still exists. The velocity V_{gr} in this case, is due to the vertical motion of the surface and hence has median equal to zero.

In Figure 5, we present the unbiased and biased sampling distributions of velocities both for the envelope and for the sea surface. The solid lines represent the unbiased densities and the dashed-dotted ones corresponds to the biased sampling densities. We see that the biased sampling distribution which are more important for applications, are more concentrated around its center. The group velocity is smaller than that of individual waves as it is observed in the real life records. The peaks are at $-5.58[m/s]$ and $-10.98[m/s]$, i.e. the waves are about twice as fast as the wave groups.

IV. Non-Gaussian models

For non-Gaussian models the problem in using the Rice formula is in computational challenges – the joint distributions rarely are available in explicit forms. To circumvent these difficulties an alternative approach was developed [6] that obtains effectively crossing distribution of the noise that is driving the considered models. The novelity of the approach is its focus on the noise distribution at the crossings and then obtain other crossing level distributions by simply replacing the underlying noise by the one that observed at the instants of crossings. One advantage of the presented approach is a possibility of simultaneous studies of various random functionals of such a noise without necessity of separate calculations for each of the functionals, or from the joint distribution of the functionals what would be even more challenging. For illustration, we start with a recap of the approach for the Gaussian case.

A. Noise at Gaussian moving average crossing

The Gaussian moving average model is given by

$$X(t) = \int_{-\infty}^{\infty} g(s-t)\, dB(s)$$

and its derivative $\dot{X}$ is given as the moving average with $-\dot{g}$ as the kernel. The question one can ask is how the noise B behaves at the instants of crossings by X of a level u. It can be argued that the biased sampling distribution of $dB(x)$ is represented by the distribution of the following stochastic process $B_u(t)$, $t \in \mathbb{R}$:

$$B_u(t) = \\ = u \int_0^t g - R \int_0^t \dot{g} - \int_0^t g \cdot \int g\, dB - \int_0^t \dot{g} \cdot \int \dot{g}\, dB + B(t),$$

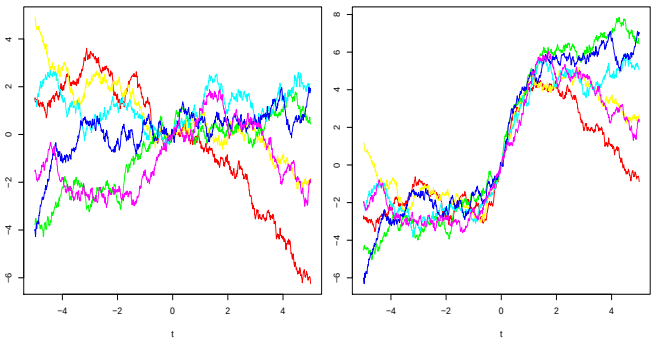


Fig. 6. *Left:* Six BM samples used in computing samples from the Slepian model $B_u(t)$. *Right:* Six samples of B_u corresponding to the BM samples and $u = 5$. A single value for the Rayleigh variable is used for all these samples.

where random variable R has the Rayleigh distribution and is independent of $dB(t)$, while $B_u(t)$ is understood as a random measure of $[0, t]$, with the convention that for $t < 0$, the measure is understood as minus the measure of $[t, 0]$. The models that produce distributions observed at the crossings are often referred to as Slepian models introduced first in [13], see also [14] and [15] for a survey on this topic.

Example 1: Let us consider (normalized) kernel $g(t) = (2/\pi)^{1/4}\, e^{-t^2}$, $t \in \mathbb{R}$. Direct calculations lead to the following crossing distribution model

$$B_u(t) = F_{u,g}(t) + G_g(t) + B(t),$$

with

$$F_{u,g}(t) = \sqrt[4]{2\pi}\; u\; \Phi_0\left(\sqrt{2}t\right),$$

$$G_g(t) = \left(\sqrt[4]{\frac{2}{\pi}}\; R - \sqrt{\frac{8}{\pi}} \int s e^{-s^2}\; dB(u)\right) \cdot \left(1 - e^{-t^2}\right) \\ - \sqrt{2} \int e^{-s^2}\; dB(s) \cdot \Phi_0\left(\sqrt{2}t\right),$$

where $\Phi_0(s) = (2\pi)^{-1/2} \int_0^s e^{-u^2/2}\; du$.

In Figure 6, we show simulations of this noise for a high level u and compare them with corresponding samples from a regular BM. For a high level u the main contribution comes from the deterministic part $F_{u,g}$.

To illustrate the convenience of the approach, we consider a pair of linear functionals of dB, $\mathbf{Y} = (Y_1, Y_2)$. The first component $Y_1(t)$, $t \in \mathbb{R}$, is a filtered original process $X(t)$ by means of a filter $h(t)$, i.e. the output from a linear system that is described by $h(t)$ when the input is $X(t)\; dt$. Thus we have

$$Y_1(t) = \int h(s-t) X(s)\; ds = \int h * g(s-t)\; dB(s).$$

The second component, $Y_2(t)$, $t \in \mathbb{R}$ is a far more complex functional of B as it arises from a linear scheme that alter Gaussian distribution of the moving average process. Namely, we consider the moving average driven by a Lévy motion build upon the Laplace distribution – the Laplace motion, [16]. The Laplace motion is obtained through subordination of the original BM to a gamma motion. For a kernel f and the Lévy process Γ such that $\Gamma(1)$ has the gamma distribution

with shape τ and scale $1/\tau$ (for negative t, the process $-\Gamma(t)$ is an independent copy of $\Gamma(t), t \geq 0$), we define the Laplace moving average

$$Y_2(t) = \int f(s-t)\ dB \circ \Gamma(s),$$

see also [17] and [18], for more details.

The simplicity of the approach that in the models we simply replace B by B_u while keeping all other independent components unchanged. For illustration, we take X as in Example 1 and consider $Y_1 = X$, while $Y_2 = Y$, where

$$Y(t) = \int g(s-t)\ dB \circ \Gamma(s), \tag{6}$$

which could be viewed as a modified X obtained by random distortion of time represented by gamma process Γ. We have the following formulas

$$\begin{aligned} X_u(t) =& u \cdot e^{-\frac{t^2}{2}} + \left(R - 2\sqrt[4]{\frac{2}{\pi}} \int s e^{-s^2} dB(s)\right) t e^{-\frac{t^2}{2}} \\ & - \sqrt[4]{\frac{2}{\pi}} \int e^{-s^2} dB(s) \cdot e^{-\frac{t^2}{2}} + X(t), \\ Y_u(t) =& \sqrt[4]{2\pi}\ u \cdot \int e^{-(s-t)^2}\ d\Phi_0\left(\sqrt{2}\Gamma(s)\right) \\ & - \sqrt{\frac{2}{\pi}} \left(R - 2\sqrt[4]{\frac{2}{\pi}} \int s e^{-s^2} dB(s)\right) \int e^{-(s-t)^2}\ de^{-\Gamma^2(s)} \\ & - \sqrt[4]{\frac{2^3}{\pi}} \int e^{-s^2} dB(s) \int e^{-(s-t)^2}\ d\Phi_0\left(\sqrt{2}\Gamma(s)\right) \\ & + Y(t), \end{aligned}$$

Using the above relation, we sampled from the joint upcrossing distribution of (Y_1, Y_2). We have chosen $\tau = 0.5$ for the shape parameter of the gamma process. The samples of underlying Brownian motion are the same as those in Figure 6.

In Figure 7 *(top)*, we observe samples simulated from bivariate process $(X(t), Y(t))$ (to facilitate better visual comparison we have used the same sample of the underlying gamma process for all six samples of the Laplace moving average). They reveal complex leptikurtic behavior of Y, which shows much larger extreme values than X. In the middle and right columns we see a sample from the Slepian model at level $u = 0.5$ and $u = 5$, respectively. The level crossing occurs at $t = 0$ as seen at the top middle/right plots. We observe in the bottom graphs that the random time change introduced by the gamma motion is adding to variability of Y at the crossing instants of X. For large level u the variability relatively to the level is reduced however the process Y still significantly overshoots the crossing value $u = 5$.

B. Noise at Laplace moving average crossing

The biased sampling distribution of Laplace noise at crossings by a moving average driven by this noise is more complicated. However, it can be effectively implemented although a Gibbs sampler is needed to obtain samples from the noise. We consider crossings of

$$X(t) = \int g(s-t)\ dL(s) = \int g(s-t)\ dB \circ \Gamma(s), \tag{7}$$

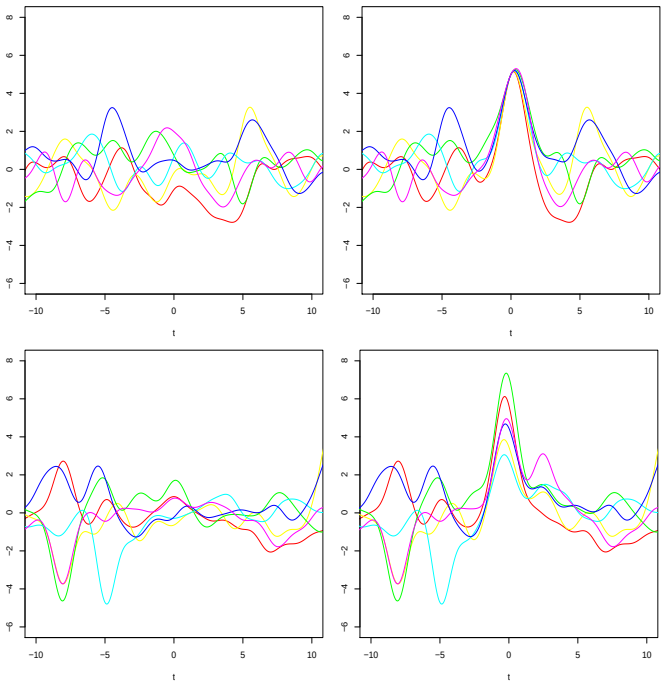


Fig. 7. *Left:* Six samples from the Gaussian moving average X – *(top)* and corresponding samples from the Laplace moving average Y – *(bottom)*. Samples are based on six samples of the Brownian motion and a single sample of the gamma process that is used for process Y. *Right:* Analogous samples at the crossings by X level $u = 5$.

where, as before, $\Gamma(t)$ is a gamma process with shape τ and scale $1/\tau$.

Let us consider an arbitrary process Y and a process $Y(\cdot|\gamma, z, u)$ with the distribution equal to conditional distribution of Y given $\Gamma = \gamma$, ($\gamma = \gamma(\cdot)$ is a trajectory of Γ), $\dot{X}(0) = z$, and $X(0) = u$. Then if t$(\Gamma_u, \dot{X}_u)$ is he biased sampling model for $(\Gamma, \dot{X})$ at crossings of u by X, then

$$Y_u(t) = Y(t|\Gamma_u, \dot{X}_u, u),$$

where for shortness $\dot{X}_u = \dot{X}_u(0)$. This approach splits the problem of finding Y_u into two separate tasks: firstly, finding $Y(\cdot|\gamma, z, u)$, then, secondly finding a model for $(\Gamma_u, \dot{X}_u)$. While finding $Y(\cdot|\gamma, z, u)$ is specific to a given process Y and need to be addressed in each case of Y individually, obtaining $(\Gamma_u, \dot{X}_u)$ is universal in the sense that it has the same structure dependent only on the elements defining the moving average X but independent of the choice of Y.

In fact, it is easier to consider an extended model $(L_u, \Gamma_u, \dot{X}_u)$ and express the crossing level distribution model by a convenient Gibbs sampler. Namely, the model will based on alternate samples from Γ_u conditionally on $L_u, \dot{X}_u$ and $L_u, \dot{X}_u$ conditionally on Γ_u. These two conditional distributions, given through the Bayes relation

$$\begin{aligned} f_{\Gamma_u|L_u,\dot{X}_u}(\gamma|l,z) &\sim f_{L_u|\Gamma_u,\dot{X}_u}(l|\gamma,z) f_{\dot{X}_u|\Gamma_u}(z|\gamma) \\ f_{L_u,\dot{X}_u|\Gamma_u}(l,z|\gamma) &\sim f_{L_u|\Gamma_u,\dot{X}_u}(l|\gamma,z) f_{\dot{X}_u|\Gamma_u}(z|\gamma), \end{aligned}$$

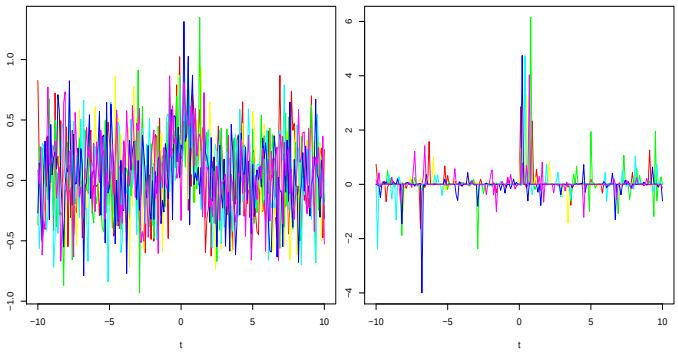


Fig. 8. *Left:* The samples from the Slepian noise for the Gaussian (left) and the Laplace (right), both for the crossing level $u = 5$.

can be simulated in a straightforward fashion. For further details we refer to [6].

Example 2: To illustrate the approach, we consider the u-level crossings of the Laplace moving average defined by (6) with the Gaussian kernel. We compare how the noise at the crossings differ from the case when the crossings were taken by the Gaussian moving average discussed in Example 1. As before, the shape parameter for the Laplace noise is $\tau = 0.5$.

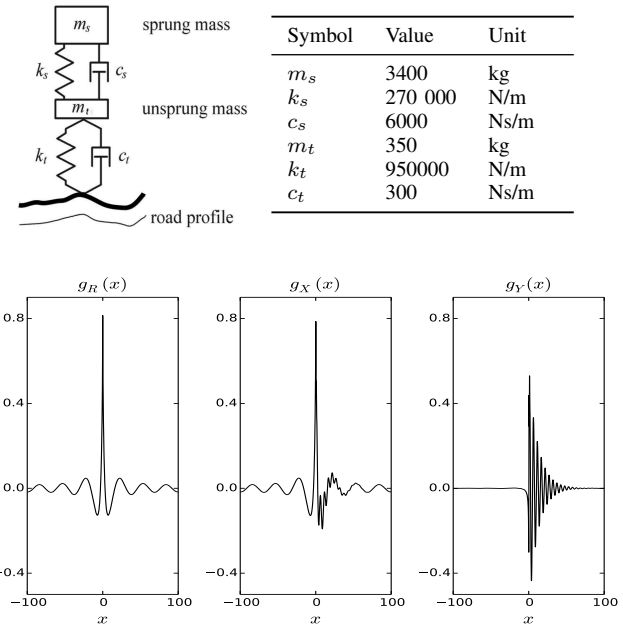
Using the Gibbs sampler, six samples $(\dot{X}_u, \Gamma_u, L_u)$ are obtained for level $u = 5$. For a large value of the crossing level $u = 5$, the Laplace motion is having large jumps at the crossing level and thus these jumps convoluted with the kernel are responsible for the shape of the process at the crossing. The jumps for the noise at level $u = 5$ are shown in Figure 8 in the right hand side graph are accumulating near the crossing instant. This is in contrast to the Gaussian case where irregularities of the noise around the crossing instant are spread over many values as presented at the left hand side graph of that figure.

C. Application

The road profile roughness is often quantified by means of the response of a quarter-vehicle model traveling at a constant velocity on road profiles, see Figure 9 *(left)*. Such a simplification of a physical vehicle cannot be expected to predict loads exactly, but it will highlight the most important road characteristics as far as durability is concerned. Often one choses the force acting on the sprung mass m_s as the response $Y(x)$ which then is used to compute suitable indexes to classify severity of road roughness. The parameters in the model are set to mimic heavy vehicle dynamics developed in SCANIA. Here the parameters have the following physical interpretation. Properties of the tire are described by k_t, c_t, which relate to vertical stiffness and damping of the tire, while properties of the suspension are given by vertical stiffness and damping k_s, c_s, respectively.

The road profile $R(x)$ will be modeled as moving average having a symmetrical kernel $g_R(x)$ which is introduced here through its Fourier transform

$$G_R(\omega) = \mathcal{F}g(\omega) = \sqrt{2\pi S(\omega)}. \tag{8}$$



Symbol	Value	Unit
m_s	3400	kg
k_s	270 000	N/m
c_s	6000	Ns/m
m_t	350	kg
k_t	950000	N/m
c_t	300	Ns/m

Fig. 9. *Top:* Quarter vehicle model. *Bottom plots:* Kernels $g_R(x)$ given in (8) and the pair $g_X(x), g_Y(x)$ given in (9).

The responses $X(x)$, $Y(x)$, defined in Figure 9 by means of a mechanical system, are obtained by linearly filtering road profile $R(x)$. The filters transfer functions will be given next.

First, by writing the equations of motion for the two masses, we obtain the transfer function for X:

$$H_X(\omega) = \frac{i\,\omega_v\,c_t + k_t}{H_t(\omega_v) - m_s\omega_v^2(i\,\omega_v\,c_t + k_t)/H_t(\omega_v)},$$

where $\omega_v = \omega \cdot v$ and ω is a wave number and v is vehicle speed in $[m/s]$.

Now, the response $Y(t)$ is filtered $X(t)$ by a filter having the following transfer function

$$H_Y(\omega) = \frac{-m_s\omega_v^2(i\,\omega_v\,c_s + k_t s)}{H_s(\omega_v)},$$

where

$$H_t(\omega) = -m_t\,\omega^2 + i\,\omega\,c_t + k_t,$$
$$H_s(\omega) = -m_s\,\omega^2 + i\,\omega\,c_s + k_s.$$

Consequently, the processes $X(x)$ and $Y(x)$ are moving averages with kernels g_X and g_Y defined through theirs Fourier transforms

$$\begin{aligned} G_X(\omega) &= H_X(\omega)G_R(\omega), \\ G_Y(\omega) &= 4\cdot 10^{-6}H_Y(\omega)G_X(\omega). \end{aligned} \tag{9}$$

Here, to ease comparisons, we scaled $Y(x)$ by a factor $4\cdot10^{-6}$. The kernels g_R, g_X and g_Y are shown in Figure 9.

The Gaussian moving average is commonly used to model the road profile variability. Although it is well known that Gaussian processes do not describe the road profiles well, see [19] and references therein, they are used because many tools are available for fast computations of probabilities of interest for durability evaluations. In [20], the Laplace moving average

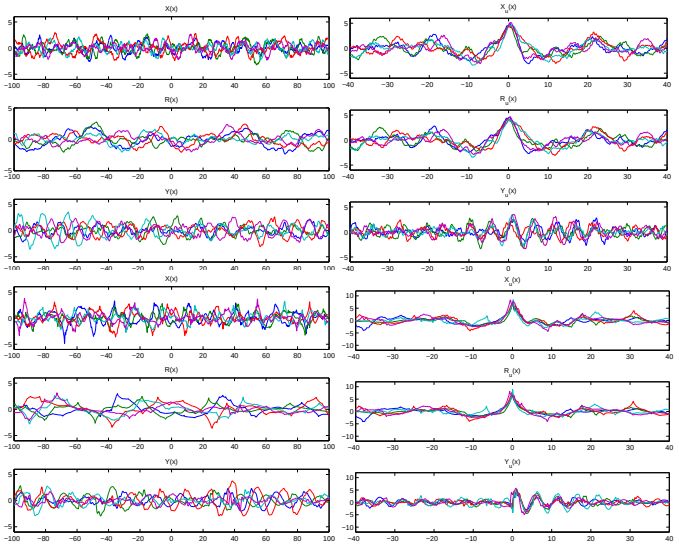


Fig. 10. The Laplace and Gaussian cases shown in the bottom and the top three graphs of each column, respectively. *Left:* Road profile $R(x)$ (middle) and responses $X(x)$ (top), $Y(x)$ (bottom). *Right:* Models $X_u(x)$, $R_u(x)$ and $Y_u(x)$ around the $u = 7$ upcrossing of $X(x)$ in the Laplace case and around the $u = 4.5$ upcrossing of $X(x)$ for the Gaussian case.

road profile model was proposed and it was demonstrated that it gives much more accurate than the Gaussian counterpart predictions of fatigue damage accumulations in vehicle components. We shall illustrate some further properties that can be useful for a design of components. Our application is kept simple for transparency of the example but it can be easily developed further to address more realistic situations by changing kernels g_X, g_Y and to include additional responses, linear or even nonlinear functionals of X. The purpose of the example is to illustrate quantitative differences between Gaussian and Laplace modeling.

To choose a level u which upcrossed by X defines the center of an extreme episode, we consider the level crossed about once per 600 km. In Figure 10 *(Left)*, some sample paths of the Gaussian and Laplace moving average models for road profile R and responses X, Y are shown. One can see that the Laplace model reaches more extreme values than the Gaussian model does. For example it can be evaluated using the Rice formula that the frequency of upcrossings of level $u = 4.5$ (measured in standard deviation) by the Gaussian process is about the same at the frequency of upcrossings of $u = 7$ (also in the standard deviation unit) by the Laplace process. This happens rarely but still frequently enough to be of importance in durability analysis. Note that many components are designed to hold 200 thousands km with high probability. Thus the levels $u = 4.5, 7$ have been chosen for the Gaussian, Laplace models of X, respectively.

The difference induced by jumps (transients) in the Laplace model is actually very influential on processes X, R, Y around u level upcrossings by X, i.e. X_u, R_u and Y_u, which can be observed in Figure 10 (the plots in the right column). First by studying kernels g_X and g_R, given in Figure 9, we expect that the paths of X_u and R_u should be similar. There will be some extra vibrations in the tire after passing a large bump but those are relatively small. In contrast, the kernel g_Y is oscillatory and asymmetric. These oscillations are characteristic for the shape of the Y_u in the Laplace model which, as oppose to the Gaussian case, is not a time reversible process. Basically, the shape of extreme episodes Y_u resembles the (asymmetric) kernel while for Gaussian model the shape is given by the correlation function of Y which is symmetric in time.

ACKNOWLEDGMENT

The first author was supported from Riksbankens Jubileumsfond Grant Dnr: P13-1024 and Swedish Research Council Grant Dnr: 2013-5180, while the second was partial support by Swedish Research Council Grant 340-2012-6004 and by Knut and Alice Wallenberg stiftelse.

REFERENCES

[1] K. Podgórski, R. I., and U. E. B. Machado, "Exact distributions for apparent waves in irregular seas," *Ocean Eng.*, vol. 27, pp. 979–1016, 2000.

[2] A. Baxevani, K. Podgórski, and I. Rychlik, "Velocities for moving random surfaces," *Prob. Eng. Mech.*, vol. 18, pp. 251–271, 2003.

[3] K. Podgórski and I. Rychlik, "Envelope crossing distributions for Gaussian fields," *Probabilistic Engineering Mechanics*, vol. 23, pp. 364–377, 2008.

[4] A. Baxevani, K. Podgórski, and J. Wegener, "Sample path asymmetries in non-gaussian random processes," *Scandinavian Journal of Statistics*, vol. 41, pp. 1102–1123, 2014.

[5] M. Kvarnström, K. Podgórski, and I. Rychlik, "Laplace moving average model for multi-axial responses applied to fatigue analysis of a cultivator," *Probabilistic Eng. Mechanics*, vol. 34, pp. 12–25, 2013.

[6] K. Podgórski, I. Rychlik, and J. Wallin, "Slepian models for moving averages driven by a non-Gaussian noise," *Extremes*, vol. 18, pp. 665–695, 2015.

[7] S. O. Rice, "The mathematical analysis of random noise," *Bell Syst Tech J*, vol. 23, no. 3, pp. 282–332, 1944.

[8] ——, "The mathematical analysis of random noise," *Bell Syst Tech J*, vol. 24, no. 1, pp. 46–156, 1945.

[9] R. Adler, G. Samorodnitsky, and J. Taylor, "High level excursion set geometry for non-gaussian infinitely divisible random fields," *The Annals of Probability*, vol. 41, pp. 134–169, 2013.

[10] M. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and related properties of random sequences and processes*. Springer-Verlag, 1983.

[11] U. Zähle, "A general Rice formula, Palm measures, and horizontal-window conditioning for random fields," *Stochastic Process. Appl.*, vol. 17, pp. 265–283, 1984.

[12] J.-M. Azaïs and M. Wschebor, *Level Sets and Extremes of Random Processes and Fields*. Wiley & Sons, 2009.

[13] D. Slepian, "On the zeros of Gaussian noise," in *Time Series Analysis*, M. Rosenblatt, Ed. Wiley, New York, 1963, pp. 104–115.

[14] M. Kac and D. Slepian, "Large excursions of Gaussian processes," *Ann. Math. Statist.*, vol. 30, pp. 1215–1228, 1959.

[15] G. Lindgren and I. Rychlik, "Slepian models and regression approximations in crossing and extreme value theory," *International Statistical Review*, vol. 59, pp. 195–225, 1991.

[16] S. Kotz, T. Kozubowski, and K. Podgórski, *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering and finance*. Birkhäuser, Boston, 2001.

[17] K. Podgórski and J. Wegener, "Estimation for stochastic models driven by laplace motion," *Communications in Statistics: Theory and Methods*, vol. 40, pp. 3281–3302, 2011.

[18] S. Åberg and K. Podgórski, "A Class of Non-Gaussian second order Spatio-Temporal Models," *Extremes*, vol. 14, pp. 187–222, 2010.

[19] K. Bogsjö, "Road profile statistics relevant for vehicle fatigue." Ph.D. dissertation, Mathematical Statistics, Lund University, Lund, Sweden, 2007.

[20] K. Bogsjö, K. Podgórski, and I. Rychlik, "Models for road surface roughness," *Vehicle System Dynamics*, vol. 50, pp. 725–747, 2012.

http://journals.lub.lu.se/stat

LUND UNIVERSITY
School of Economics and Management

Working Papers in Statistics 2016
LUND UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT
Department of Statistics
Box 743
220 07 Lund, Sweden