January 30, 2013

Unified characterization of surface topography for vehicle dynamics applications

Paul Pukite, Steve Bankes, Dan Challou : *BAE Systems*
Carl Becker : *CSIR/University of Pretoria*
Richard Bradford : *Rockwell Collins*

Abstract: Environmental models by their nature contain a great deal of uncertainty. Since the underlying behavior of the model is rarely controlled by an ordered process, any model characteristics will carry along with it a level of aleatory uncertainty governed by the natural disorder. This paper applies novel uncertainty characterization approaches to classes of topographic models to demonstrate how to quantify the natural order and distinguish from artifical (man-made) order.

Preface

Comprehensive vehicle testing requires *environmental context* models to evaluate a design's robustness and fitness for use. For a ground vehicle designed for both on-road and off-road use, interactions with the terrain provide an important testing context. For amphibious vehicles, the environmental context extends to sea states and aquatic currents. As an overriding factor, being able to traverse and navigate successfully various surface topographical features plays a large role in verifying that a vehicle's requirement specifications can be met.

To streamline the verification process, terrains need to be first well characterized and then modeled comprehensively enough that probability of correctness approaches can be applied. Without the ability to quantify for the probability of occurrence for a specific environmental effect, any requirements based on topography, such as traverse grade, can't be propagated to a virtual test of the vehicle's design ─ and thus we cannot substantiate with any degree of certainty or accuracy whether the vehicle meets the customer's needs.

This paper applies general methods for stochastic modeling of environment contexts described in a companion paper (volume 1 of this series) "Stochastic Analysis For Context Modeling".

Primer on Characterization and Modeling

To develop some breadth, the models described here will cover both land and aquatic topographies. The landform aspects we refer to as *terrain modeling* [1] and the aquatic as *wave* (or *wave energy*) *modeling*[2].

For terrains, we consider a rigid model of altitudes as a function of lateral dimension. Often, a one-dimensional path, with provision for left and right tracks, is sufficient to characterize a terrain. The interaction of the vehicle with this type of terrain path provides enough of a context to virtually verify requirements such as absorbed power at speed and fuel efficiency on inclined grades.

Stochastic models[3] work very effectively to describe natural as well as artificial, man-made terrain. Enough disorder and randomness exists in the natural world, and even within an artificial world of cobblestones, chatter bumps, rumble strips, and wave machines, that a synthetic model using probability concepts appears indistinguishable from an empirically measured set of real-world data[4].

The stochastic tools of the trade include probability distribution functions, autocorrelation functions, power spectral densities, and moment distributions [5].

Probability density functions (PDF) model sample spaces. A single PDF graph facilitates the characterization of natural phenomena that are prevalent in context modeling, including terrain slope distributions as well as wind, rainfall, lakes, cloud, etc.[6] Many of these models reduce to fairly simple analytical forms. For disordered systems and data containing aleatory uncertainty[7], we use techniques such as the *maximum entropy principle*[8] and *super-statistics*[9], which have a more formal basis than heuristic models can provide[1]. For narrow distributions, we can apply Normal or Gaussian statistics.

Autocorrelations (ACF) and *power spectral densities* (PSD) are tied closely together. In basic terms, an autocorrelation describes the amount of similarity between two points separated spatially or temporally. For static terrain this is a purely spatial consideration, but at a constant vehicle speed, an autocorrelation turns into a temporal characteristic, as time equates to motion across a terrain. A PSD contains the same information as an ACF but works in the inverse space, either the wavenumber (k) space in the spatial domain, or frequency (ω) space in the time domain. The decision to select an autocorrelation function or power spectral density will depend on the application. A spectral representation can easily provide reconstructed instances for either spatial or temporal domains in a general system context. The power of the real space autocorrelation is that it can recover a synthetic representation of the local terrain as a random walk model, while the PSD can approximate a varied terrain as a superposition of sine waves.

In addition, certain *multiscale* metrics[10] are useful to further characterize and in particular disambiguate potentially degenerate model descriptions. In particular, the multiscale entropy measure can be applied to temporal and spatial scales covering a wide dynamic range, and multiscale variances are sensitive to asymptotic behavior. What this tells us is the amount of disorder and uncertainty in the data, which is important as a concise supporting characterization metric.

Table 1: Stochastic Model Categories

Stochastic Model		Includes and Related	
Probability Density Function	PDF	Cumulative Distribution Function, Histogram	Position Independent
Autocorrelation Function	ACF	Correlation, Autocovariance	Relative
Power Spectral Density	PSD	Power Spectrum, Periodogram, Fourier Spectrum	Relative
Multiscale Analysis	MSA	Multiscale Variance, Multiscale Entropy	Relative

Table 1 above categorizes the stochastic models used to characterize terrain and waves. These can work together to model some sophisticated profiles. For example, a probability density function can model a length distribution, which then gets applied to a semi-Markov formulation for an autocorrelation analysis. By then applying a *Fourier transform* to the autocorrelation, we can arrive at a very concise representation. We will describe this in the analysis section.

Data Characterization

The rules for construction of probability density functions follow from some elementary principles. Samples from experimental data are ranked, and then normalized to the largest (i.e. scarcest) sample. This establishes a cumulative probability of unity when integrated over the sampled data space. The rank histogram is then converted to a cumulative distribution function by interpolating across a continuum, and then a PDF derives from the first derivative with respect to the random variate. Strictly speaking, a PDF is a discrete (binned) form while a probability mass function (PMF) is the continuum, but we will uniformly use the term PDF, even though we generally assume the continuum.

---

[1] The sample size needed to reduce the variance of the statistics can be reduced by efficient applications of random sampling such as importance sampling.

Given that *X* is some variate of interest, then a PDF described by *f* (*X*) can be used to generate moments of the distribution, such as the expected value **E** [*X*].

$$\mathbf{E}\,[X] = \bar{X} = \int_{-\infty}^{\infty} X\,f(X)\,dX$$

Autocorrelation functions consider the pair-wise expected value of all samples separated by a distance measure. This describes the affinity for localized interactions that a PDF lacks sensitivity to. The computational complexity of calculating an autocorrelation can go as N^2, but a full computation is not always necessary for determining only closely separated distance correlations. Due to the *Wiener-Khinchin* theorem, the Fourier Transform of the data itself, calculated as a power spectrum of magnitude squared, is equal to the Fourier Transform of the autocorrelation, so that an inverse Fourier Transform of the directly calculated power spectrum will recover an autocorrelation. Due to the efficiency of a FFT, the computational time is order N×log(N), so this is often used to produce an ACF and a PSD with only frequency spectrum tools.

The power spectrum and autocorrelation are related by the following equation:

$$S_{xx}(f) = \int_{-\infty}^{\infty} r_{xx}(\tau)e^{-2\pi i f \tau} d\tau$$

Where the autocorrelation is defined as

$$r_{xx}(\tau) = \mathbf{E}\,[\,x(t)\cdot x^{*}(t-\tau)\,]$$

The double *xx* indicates that the expected value is between pairs of points along the terrain, with the τ serving as the sample-to-sample distance measure. (The asterisk indicates the complex conjugate, which does not apply for these signals.)

Foundation of Stochastic Model Analysis

The analysis of real-world PDF's is aided by the fact that independence is assumed in the sample space. Each draw from a PDF is by definition independent and *stationary* with respect to the sampling point. We don't always have to understand and assert independence but enough empirical studies have been done to understand when the premise will work and when it doesn't.

For distributions based on energetic processes, the *principle of maximum entropy* often results in parsimonious fits to collected data (later will show extended examples for ocean waves and terrain). The selection of the variate and how to apply the prior is important, as the shape of the distribution can be thin-tailed (exponential or Normal distributed) or fat-tailed (a power-law) depending on its modeled derivation. For example, a timing distribution may be fat-tailed because the analyzed variate is velocity (which would ostensibly give a thin-tail) but since time appears in the denominator with respect to velocity, the eventual PDF generates much more weight in the tail (i.e. a ratio distribution [3]).

For an ACF, the typical realization is via a random-walk model, often described by a *Markov* or *semi-Markov* process [11]. Higher order localized interactions beyond that, such as the near-neighbor memory-less Markovian random walk, can generate smoothed/filtered or ordered/periodic profiles, depending on the signs and strengths of the interactions. This has significance for decoding both an ACF and a PSD. An important consideration for autocorrelations is the concept of *coherence* length. A coherence length is that distance at which long-range correlations cease to factor in. Beyond that point, the state of the system could just as easily be determined by a draw from a PDF. This has significant implications on whether to apply a Fourier Transform on a model to match a power spectrum or to simply treat the PSD as an inverted probability density function in the frequency spectrum.

The interplay between the use of PDF and ACF profiles for characterization is abetted by some very practical aspects of working in the spatial frequency domain. One novel way of looking at the problem derives from the world of diffraction spectroscopy, where because of the micro-scales involved, the only way possible to get insight is to immerse oneself in the inverse of the spatial domain, into the spatial frequency domain, in what is known as *reciprocal space.* [11]

We will work in reciprocal space here, not because we can't detect the spatial features with measurements, but because the convenient mating of stochastic processes combined with powerful spectral algorithms, allows us to work out an analysis with greater rigor and statistical accuracy. This has huge implications for generating synthetic terrains based on limited real-world data (for example, where the actual course is classified but the PSD is not).

Unified Autocorrelation Analysis

Our analysis based on characterizing stationary profiles. This approach is fully documented elsewhere[11] in which we derive the *pair correlation* (autocorrelation) functions of arbitrary surfaces. The derivation assumes a path in essentially a 2-dimensional slice. One dimension is the distance along the path of traversal and the other dimension is the elevation of the point along that path.

A surface is constructed with N_c rigid cells arranged in the x- and z-directions. In a discrete approximation, the cell spacing in the x-direction is *a,* and in the z-direction is *d.* This surface is described by a function *f(x,* z) which is equal to 1 if there is a surface cell at coordinates *(x, z)* and 0 otherwise. The Fourier power spectrum is defined as:

$$I(\vec{S}) = \left| \sum_{\vec{r}} f(\vec{r}) e^{-i\vec{S}\cdot\vec{r}} \right|$$

Where $\vec{S}$ is the vector wavenumber defined in the reciprocal space of *(x, z)* as *(S_x, S_z).* This can be rewritten as the equivalent Weiner-Khinchin relation:

$$I(\vec{S}) = N_c \left| \sum_{\vec{u}} C(\vec{u}) e^{-i\vec{S}\cdot\vec{u}} \right|, \qquad N_c \to \infty$$

Where $\vec{u}$ is a real-space vector and

$$C(\vec{u}) = \frac{1}{N_c} \sum_{\vec{u}} f(\vec{r}) \cdot f(\vec{r} + \vec{u})$$

is the pair correlation function on the surface. It is the probability of finding two cells on the surface separated by a vector $\vec{u}$. By including the array of cells as a sum of delta functions, this equation can be changed to the integral form:

$$I(\vec{S}) = N_c \int_{-\infty}^{\infty} dx \, e^{-iS_x x} \sum_{l} e^{-iS_z ld} \left[\sum_{n} \delta(x - na) \, C(x, ld) \right]$$

Where $C(x, ld)$ is the continuum portion of $C(\vec{u})$ written expressed along the two dimensions of interest. This can be rewritten making use of the *convolution theorem* as:

$$I(\vec{S}) = N_c \left[\int_{-\infty}^{\infty} dx \, e^{-iS_x x} \, \delta(x - na) \right] \otimes \left[\sum_l e^{-iS_z ld} \int_{-\infty}^{\infty} dx \, e^{-iS_x x} C(x, ld) \right]$$

Where $\otimes$ is the convolution operator, which expands the embedded summations. By taking the Fourier transform of the first term:

$$I(\vec{S}) = \left[\frac{2\pi N_c}{a} \sum_n \delta \left(S_x - \frac{2\pi n}{a} \right) \right] \otimes \left[\sum_l e^{-iS_z ld} \int_{-\infty}^{\infty} dx \, e^{-iS_x x} C(x, ld) \right]$$

The second term repeats along the periodicity of the cell spacing, which is the result of a discrete representation, and is what we are interested in for a continuous system. Expanding the pair correlation term $C(x, ld)$, we first consider a surface in which the surface cells are allowed to be on any of an infinite number of levels as illustrated below. (Note that infinite specifically applies to the number of levels; for finite level systems, only the x-distance is infinite.)
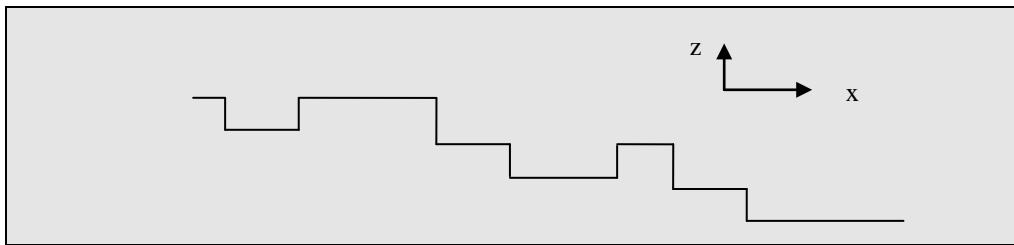

Figure 1: A surface composed of terraces separated by levels

The probability of finding a cell at the origin and at a position (x, ld) away is the sum of the probabilities of all possible configurations of steps that separate these two points. For example, assume for the moment that all steps are same height and are always in the same direction as in the classic descending stair. And suppose that a sequence of levels with lengths $(L_1, L_2, ..., L_n,)$ separate a cell at an origin that is L_0 away from the first step from a cell at (x, ld) which is L_n away from the last step. Then the probability of this configuration is

$$P_0(L_0) \, P_1(L_1) \cdots P_{n-1}(L_{n-1}) \, P_n(L_n) \, \Delta L_0 \, \Delta L_1 \, \cdots \, \Delta L_{n-1} \Delta L_n$$

Where the $P_i(L_i)$ with i = 1, 2,· . . . , n -1 are the probabilities per unit length of a terrace of length L_i occurring on the ith level of the surface. The function $P_0(L_0)$ is the probability per unit length that there is a cell at the origin that is L_0 before the first step. The function $P_i(L_i)$ is the probability that there is a cell L_n away from the last step. The middle $n - 1$ $P_i(L_i)$ are identical and equal to $P(L_i)$ for the infinite surface. This equation implicitly assumes that the "terrace" lengths are statistically independent from one another. The length of one does not determine the length of subsequent terraces except in so far as the sum of the L_i equals the total x-coordinate and that the number of steps is l, which is equal to n for the strictly descending staircase. To determine the pair correlation function, one must integrate over all possible configurations of steps that satisfy these constraints.

If in addition, we allow the step heights to be any multiple of the level separation, either positive or negative, with probability distribution $H(h_i)$ for a step of height $h_i d$, then the correlation function $C(x, ld)$ is

$$C^+(x, ld) = N_l \left\{ P_{0f}(x)\delta_{l,0} \right.$$

$$+ \sum_{n=1}^{\infty}\sum_{h_i=}^{\infty} \cdots \sum_{-\infty}^{\infty} \int_{L_i=}^{\infty} \cdots \int_{0}^{\infty} P_0(L_0) \times H(h_1)\, P(L_1) \cdots P(L_{n-1})H(h_n)P_f(L_n)$$

$$\left. \times \left[\delta\left(x - \sum_{i=0}^{n} L_i\right) \delta\left(x - \sum_{i=1}^{n} h_i\right) dL_0 \cdots dL_n \right] \right\}$$

This formulation is equivalent to a convolution integral of all the possible configurations of steps. The right hand side of this equation integrates over all possible configurations of terrace lengths, L_i, and sums over all possible step heights, $h_i d$, which will add to give a displacement *(x, ld)* from the origin. The plus sign in the superscript indicates the correlation function for the positive x and z-directions; to obtain the negative directions, simply replace x by $-x$ and z by $-z$. N_l is the total number of levels involved and is required to approach infinity. The function $P_{0f}(x)$ is the probability that there is a cell at the origin and one x away on the same terrace with no jumps in between.

The two distribution functions $H(h)$ and $P(L)$ are normalized according to

$$1 = \sum_{h=-\infty}^{\infty} H(h). \qquad 1 = \int_{\tau=L}^{\infty} P(\tau)d\tau$$

The functions $P_0(L)$ and $P_f(L)$ can be written in terms of the terrace length distribution function $P(L)$. $P_0(L)$ is the probability that there is a cell at the origin and the first step is L away. It is given by the product of the probability that there is a cell at the origin, which is given by the coverage θ, and the number of configurations that allow the first step to be at L. For the infinite staircase, the coverage is $1/N_l$ and

$$P(L) = \frac{\theta}{\langle L\rangle}\int_{\tau=L}^{\infty} P(\tau)d\tau$$

Where

$$\langle L\rangle = \int_{\tau=L}^{\infty} \tau\, P(\tau)d\tau$$

To find $P_f(L)$ it is important to realize the asymmetry of the method. Since the probability of being at step n is 1 by construction, $P_f(L)$ is simply given by the probability that the step is followed by a terrace of length $\tau > L$. Thus it is given as

$$P_f(L) = \int_{\tau=L}^{\infty} P(\tau)d\tau$$

For the important case of no level changes, or the $n = 0$ portion of the summation, $P_0(L_0)$ and $P_f(L_n)$ must be replaced by

$$P_{0f}(L) = \frac{\theta}{\langle L\rangle}\int_{\tau=x}^{\infty} (\tau - x)P(\tau)d\tau$$

as this is the correlation function of a single level. Similar derivations of the preceding probabilities applied to a microscopic domain can be found in [11].

To calculate the PSD, we want the Fourier transform of the correlation function for $C^+(x, ld)$ For this, first substitute the delta function representations

$$\delta\left(x - \sum_{i=0}^{n} L_i\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dS_x \, e^{-iS_x(x - \Sigma_{i=0}^{n} L_i)}$$

$$\delta\left(l - \sum_{i=0}^{n} h_i\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dS_z \, e^{-iS_z(l - \Sigma_{i=0}^{n} h_i)d}$$

into the long equation and define $P(S_x)$ and $H(S_z)$ by

$$P(S_x) = \int_{0}^{\infty} P(L)e^{-iS_x L}$$

$$H(S_z) = \sum_{-\infty}^{\infty} H(h)e^{-iS_z h d}$$

to be the Fourier transforms of the individual probability density functions. Also, the Fourier transforms of the other probability functions can be written in terms of $P(S_x)$ according to

$$P_0(S_x) = \frac{\theta}{iS_x\langle L\rangle}[1 - P(S_x)]$$

$$P_f(S_x) = \frac{1}{iS_x}[1 - P(S_x)]$$

$$P_{0f}(S_x) = \frac{\theta}{S_x^2\langle L\rangle}[1 - P(S_x)] + \frac{\theta}{iS_x}$$

Then the correlation function becomes

$$C^+(x, ld) = N_l\left\{P_{0f}(x)\delta_{l,0} + \frac{1}{4\pi^2}\sum_{n=1}^{\infty}\left[\int_{-\infty}^{\infty} dS_z e^{-iS_z ld}[H(S)]^n\right] \times \left[\int_{-\infty}^{\infty} dS_x \, e^{-iS_x x}P(S_x)\big(P(S_x)\big)^{n-1}P(S_x)\right]\right\}$$

Taking the transform to generate a PSD, we obtain:

$$I_0(S_x, S_z) = 2 N_l \, \text{Re}\left\{P_{0f}(S_x) + \sum_{n=1}^{\infty}[H(S_z)]^n P_0(S_x)[P(S_x)]^{n-1}P_f(S_x)\right\}$$

Here, the subscript 0 on the power spectrum intensity indicates the Fourier transform of the continuum part of the correlation function. Calculating the geometric sum, the result is

$$I_0(S_x, S_z) = 2 N_l \, \text{Re}\left\{P_{0f}(S_x) + H(S_z) \, P_0(S_x)P_f(S_x)\sum_{n=0}^{\infty}[H(S_z)P(S_x)]^n\right\}$$

$$= 2 N_l \, \text{Re}\left\{P_{0f}(S_x) + \frac{H(S_z) \, P_0(S_x)P_f(S_x)}{1 - H(S_z)P(S_x)}\right\}$$

$$= \frac{2\,N_l}{S_x{}^2 \langle L \rangle}\; \mathrm{Re} \left\{ 1 - P(S_x) - \frac{H(S_z)\,(1 - P(S_x))^2}{1 - H(S_z)P(S_x)} \right\} \tag{1}$$

One simple realization of a surface forms a staircase of steps. Each step descends or ascends in one direction. For this case, the transform of the step height distribution function, $H(S_z)$, becomes $e^{-iS_z d}$. We also assume that the staircase is not perfectly regular so that the sum in the equation always converges.

$$I_0(S_x, S_z) = \frac{2}{S_x{}^2 \langle L \rangle}\; \frac{1 - |P(S_x)|^2}{|1 - P(S_x)e^{-iS_z d}|^2} \tag{2}$$

If we then flatten the staircase, this emulates a sawtooth pattern of topography, either a set of breaking waves or a graded terrain with that structure. To level the reciprocal space, all we need to do is follow a path along the equality $x + z/\theta$, as shown in the figure below. This maps the average terrain slope onto an affine transformed coordinate system that is flat from the perspective of the viewer. In reciprocal coordinates, the transformed path is $S_x + S_z\,\theta$
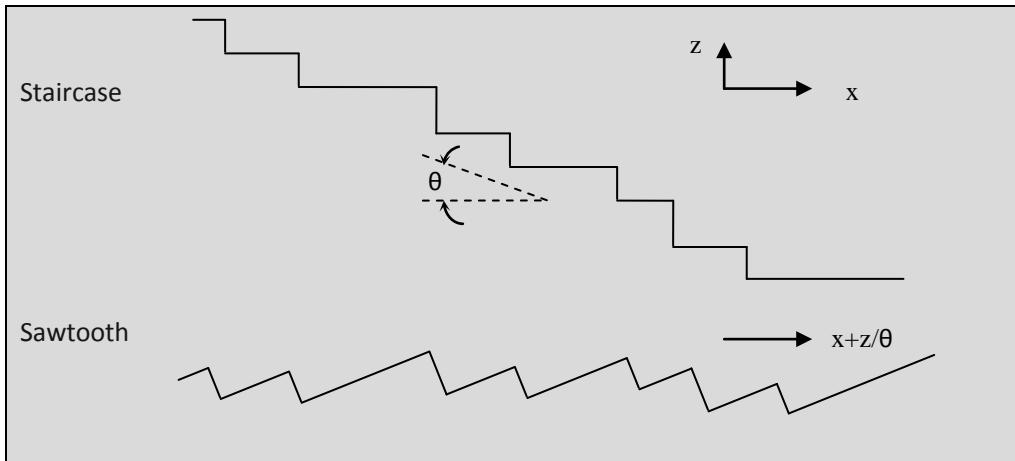
Figure 2: A descending staircase is used to derive an autocorrelation and then the coordinate system is rotated to model a sawtooth terrain profile.

The general approach falls under the categorization of a *semi-Markov* analysis. We allow the distribution for $P(S_x)$ and $H(S_z)$ to take on any form. To recover the elementary Markov spectrum, we apply a maximally disordered distribution to $P(L)$.

$$P(L) = \frac{1}{\langle L \rangle}\; e^{-L/\langle L \rangle}$$

This selection results in the Fourier transform for $P(S_x)$

$$P(S_x) = \frac{1}{1 + iS_x \langle L \rangle} \tag{3}$$

Applying the same transform to the step height, the resulting PSD is the Cauchy or Lorentzian profile:

$$I_0(S_x) = \frac{1}{4 + \langle L \rangle S_x{}^2} \tag{4}$$

This has the classic PSD second-order power-law fall-off of a random terrain, shown below. The deviations from this fall-off are due either to further order in the terrain relief or filtering of the data set. This particular data was extracted from an unclassified military test course.
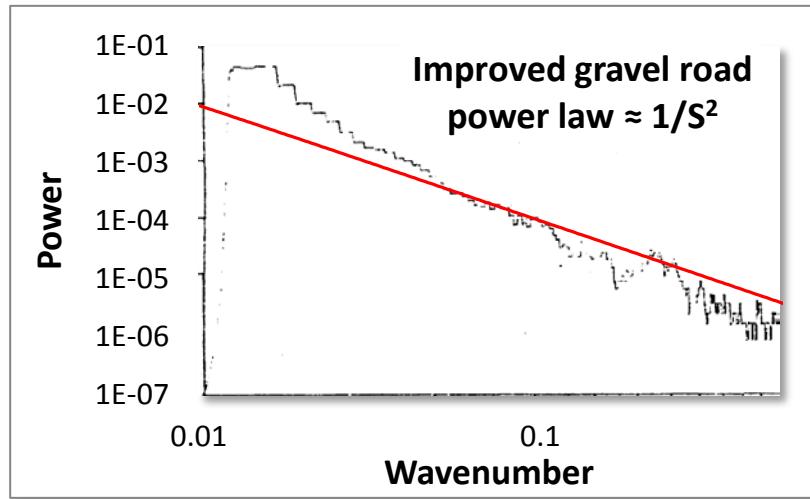
*Figure 3: **Gerotek instead?** Power Spectral Density (PSD) plot from a declassified military test course, with a random walk terrain profile overlaid as 1/S²*

The data and spectra from Figure 3 are taken from a historical unclassified test document. Figure 4 presents data from the present day "rough road" test track data from a Mercedes-Benz course[12], note the reduced fluctuation noise and wide dynamic range in the data.
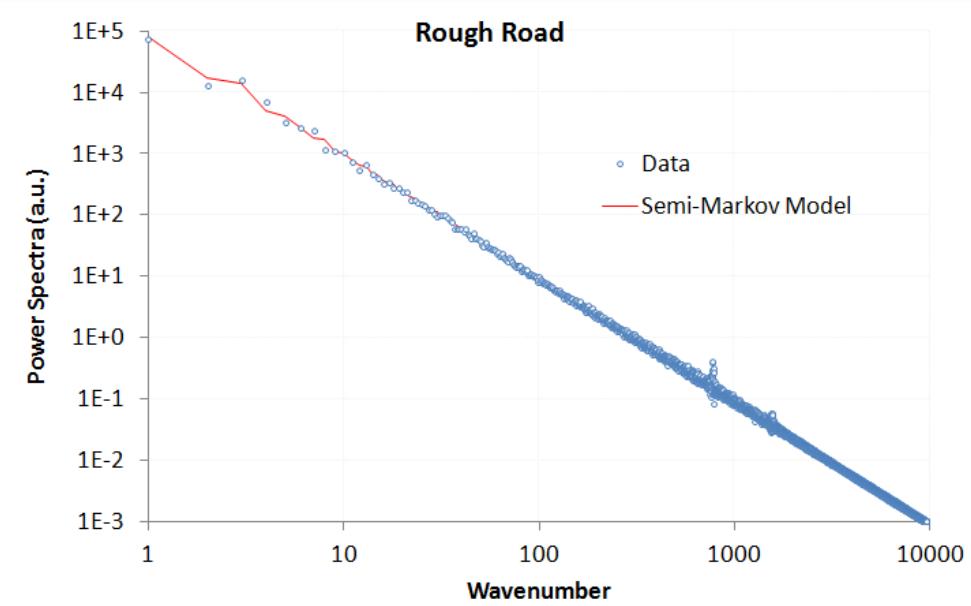

Figure 4: PSD of data collected from a higher resolution and more extensive terrain data set from a Mercedes-Benz test track. Note the weak high frequency spike with a definite even harmonic signal. The overall tendency of the terrain profile follows a random semi-Markov model very well.
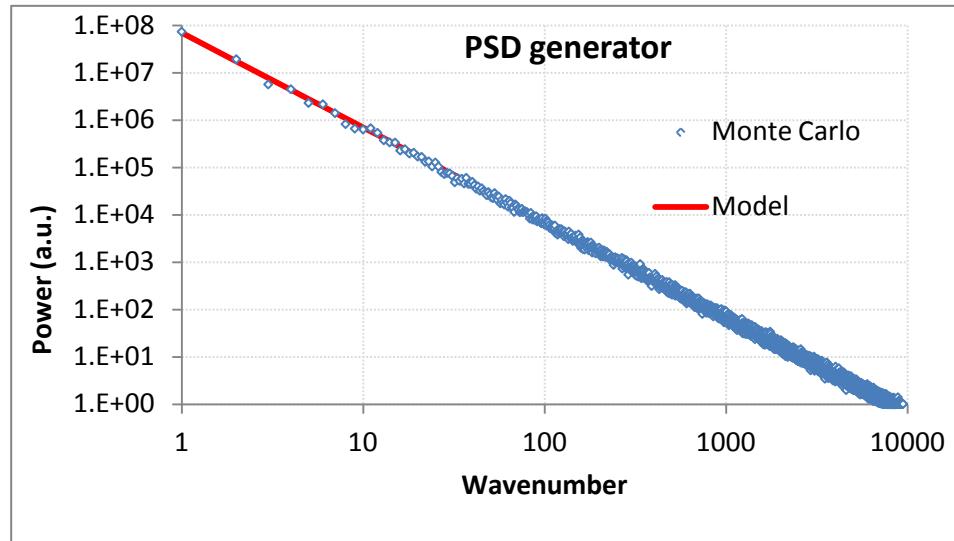
Figure 5: Monte Carlo generated course based on random walk model

The power of the analysis is that it can detect asymmetries in the surface relief. For example, a terrain that shows sawtooth waves will contain even harmonics (note the weak harmonic in Figure 4), while a terrain that contains symmetric square or rolling pseudo-sinusoidal waves will contain odd harmonics.

Additionally, for the surface shown above, much detail may be obscured by a long-wavelength trend in the underlying terrain. The slope for the "rough road" example of Figure 4 is fairly large and this contributes to a strong $1/S^2$ tail above a certain wave number, i.e. a linear slope in real space generates an inverse squared response in reciprocal space according to Fourier analysis. This can also be inferred by the reduction of noise at high wave numbers, since a static slope is deterministic and dominates over the stochastic fluctuations. This also explains why a course profile needs to be "detrended" for it to be practical for determining actual roughness. A non-detrended course will generate a $1/S^2$ and provide less value in extracting the smaller scale roughness.

Regular features. If on occasion, we come across a terrain structure with extremely regular or periodic features (such as a highway rumble strip or a lengthy grated section), the PSD calculation simplifies to a Fourier series.
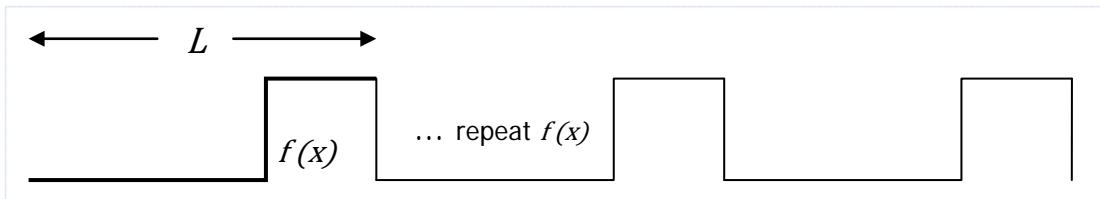

Figure 7: A perfectly ordered sequence of steps of periodicity L

This reduces to the following intensity, with the PSD described as a series of harmonic peaks of very narrow width modulated by an envelope determined by the Fourier transform of one cycle

$$I(S_x) = |F(S_x)|^2 \sum_{i=0}^{\infty} \delta\left(S_x - \frac{2\pi}{L}i\right) \xrightarrow{where} F(S_x) = \int_0^L f(x)\, e^{-iS_x x}$$
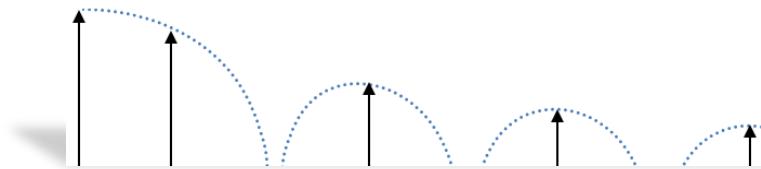
Figure 8: The PSD of ordered steps has delta-function harmonic peaks, modulated by the Fourier transform envelope of a single period.

These insights derive from a the Fourier series decomposition of periodic waveforms, but our goal is to apply this technique to stochastically varying waveforms – as stochastic profiles are very prevalent in artificial (see the following figure) and natural terrain. In the following sections, we will look more closely at the spectral features of disordered aquatic and land terrain.

Figure 9: Examples of roads with random and pseudo-periodic cracks and patches

Application to Aquatic Waves

We first consider aquatic wave spectra because these have a more consistent character than land terrain. Gravity plays a critical role as it automatically detrends the data to maintain a level profile over the distances of interest. Sea waves also are very sensitive to long-range *coherence*, that is, the ability to maintain phase relationships over many wavelengths such as seen with capillary waves[13][14] or over a significant distance as with cnoidal swells[15]. The term coherence is defined as correlation when applied specifically to wave-like properties [16]. This is noteworthy also when one considers that a dispersion relation holds between the spatial frequencies and temporal frequencies of aquatic waves.

Spatially Incoherent Wave Spectra

In the wild, waves usually display little coherence over long spatial scales. In other words, the knowledge that one has about one wave has limited implications regarding another wave separated by several undulations.

We make a maximum entropy estimation of the energy of a one-dimensional propagating wave driven by a prevailing wind direction. The mean energy of the wave is a function of the square of the wave height, **H**.[17] This makes sense because a taller wave needs a broader base to support that height, leading to a scaled pseudo-triangular shape, as shown in the figure below.
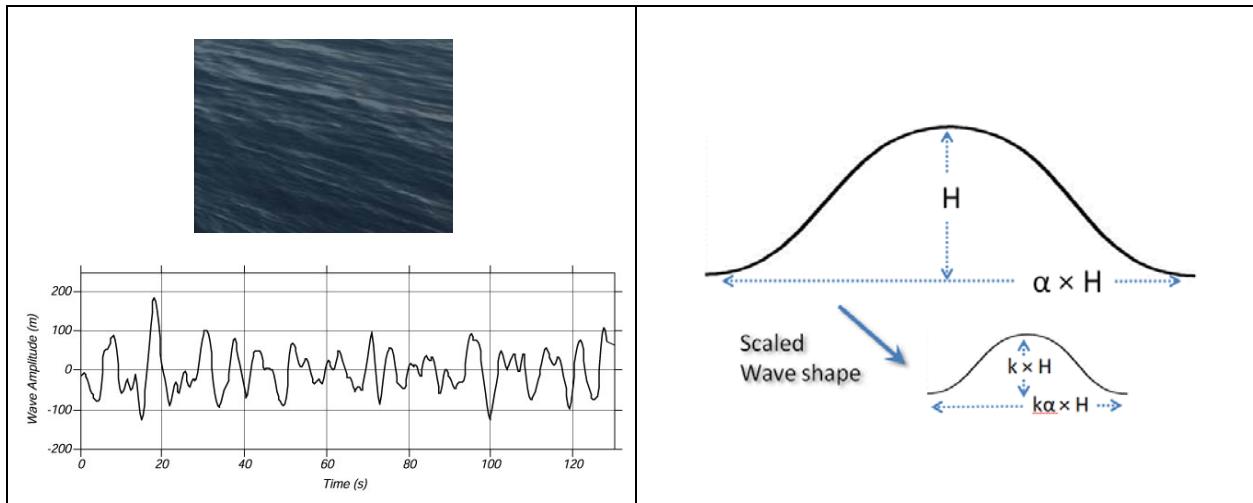

Figure 10: (left) The goal is to model the spectral density of waves. Enough disorder exists in open water that periodic coherence between spatially separate waves is not maintained, measurement from wave buoy in north Atlantic.(from [2]).   (right) The initial assumption is that waves contain energy proportional to their width and height. This proportionality scales independent of volume. Total energy in a directed wave goes as the square of the height, and the macroscopic fluid properties suggest that it scales to size. This leads to a dispersive form for the wave size distribution.

Since the area of such a scaled triangle goes as **H**^{2}, the MaxEnt cumulative probability is:

$$P(H) = e^{-aH^2}$$

where *a* is related to the mean energy of an ensemble of waves. This relationship is empirically observed from measurements of ocean wave heights over a sufficient time period. However, we can proceed further and try to derive the dispersion results of wave frequency, which is the typical oceanography measure. So we consider — based on the energy stored in a specific wave — the time, *t*, it will take to drop a height, *H*, by the Newton's law relation:

$$t^2 \sim H$$

and since *t* goes as 1/*f*, then we can create a new PDF from the height cumulative as follows:

$$p(f)df = \frac{dP(H)}{dH}\frac{dH}{df}df$$

where[2]

$$H \sim \frac{1}{f^2}$$

$$\frac{dH}{df} \sim -\frac{1}{f^3}$$

then

$$p(f) \sim \frac{1}{f^5} e^{-c/f^4} \qquad\qquad \textbf{(5)}$$

which is just the Pierson-Moskowitz wave spectra that oceanographers have observed for years (developed first in 1964, variations of this include the JONSWAP, Bretschneider and ITTC wave spectra[18]).

This concise derivation works well despite not applying the correct path of calculating an auto-correlation from *p(f)* and then deriving a power spectrum from the Fourier Transform of *p(f)*. As smooth ocean waves approach a sinusoidal ideal, a single wave will contribute a Fourier component at that frequency alone and the spectrum can be evaluated almost by inspection. This convenient shortcut remains useful in understanding the simple physics and probabilities involved.

The utility of this derived form can be demonstrated using data[19] obtained from public-access stations. The following data is extracted from a pair of measuring stations located off the coast of San Diego [20].
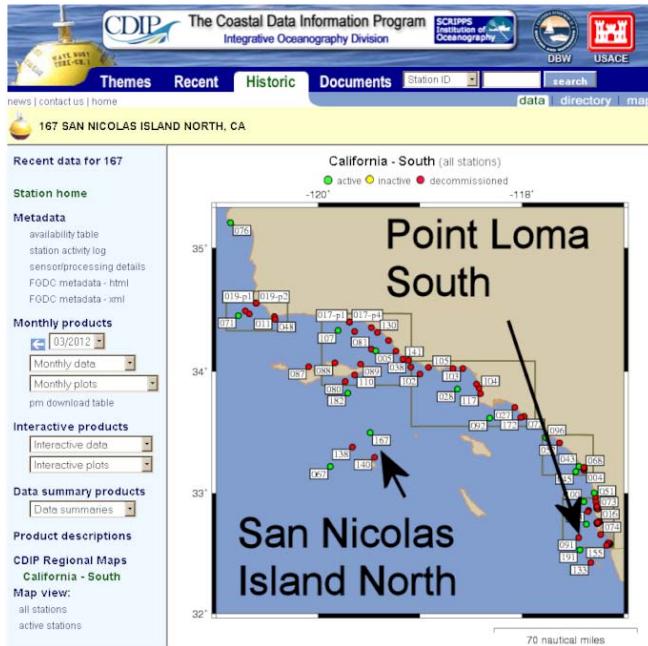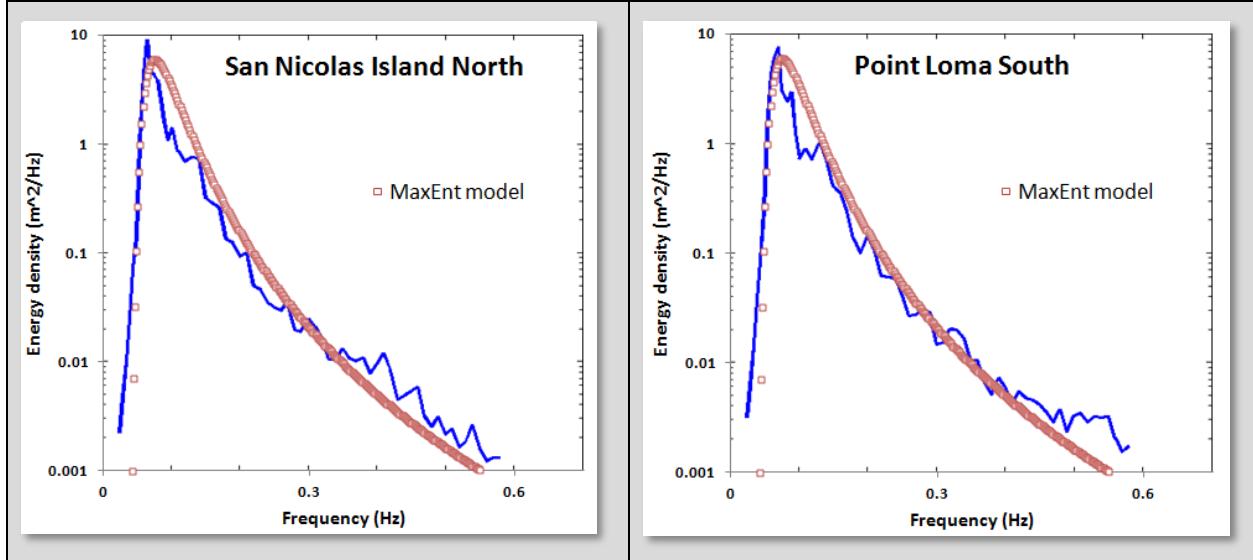
Figure 11: The CDIP data source provided archival wave energy statistics for assessing models.

The data shown here is averaged wave spectra for the entire day 1/1/2012. The red points in Figure 8 correspond to best fits from the derived MaxEnt algorithm applied to the blue data set. [2]


Figure 12: Wave energy spectra from two sites off of the San Diego coastal region.
The Maximum Entropy estimate is in red.

Like many similar spectra such as wind and EMI, the wave spectrum derives simply from maximum entropy conditions.

Note that we did not need to invoke the full spectral decomposition model presented earlier in this paper. The correlations are short over the sinusoidal shape of an individual wave and so those frequency components show up strongly in the energy density PSD. The implications are that the driving forcing function needed to provide order in the case of natural waves is missing, and this makes sense as the main stimulus, wind energy, on its own is highly disordered. The wind simply stimulates the wave to maximize its entropy subject to the kinetic and potential energy that are provided.

On the other hand, under controlled or reinforcing (positive interference) conditions, more order can be supplied to make the waves appear more coherent over a spatial distance.

Spatially Coherent Wave Spectra
Coherent waves are easier to generate in the laboratory than in nature, apart from the occasional cnoidal swells peculiar to a region. A large wave tank, with waves generated by a consistent wind will clearly expose any ordered features in the PSD [21].

In Figure 9 below, taken from [22], the measured PSD clearly shows clear harmonic peaks strongly suggestive of longer-range order in the wave periodicities. Superimposed on the data profile (shown in black) is a model adapted from the derivation of the previous section of this paper. Note that from the spatial/temporal dispersion relation, the time- domain frequency scale with the spatial-domain frequencies, as short, choppy waves have much smaller periods or cycle-times than the large, rolling

---

[2] The dataset is available from :
http://cdip.ucsd.edu/?nav=historic&sub=data&units=metric&tz=UTC&pub=public&map_stati=1,2,3&stn=167&stream=p1&xyrmo=201201&xitem=product25

waves. This means that we can apply a spatial analysis to infer the temporal characteristics with some generality (and vice versa).
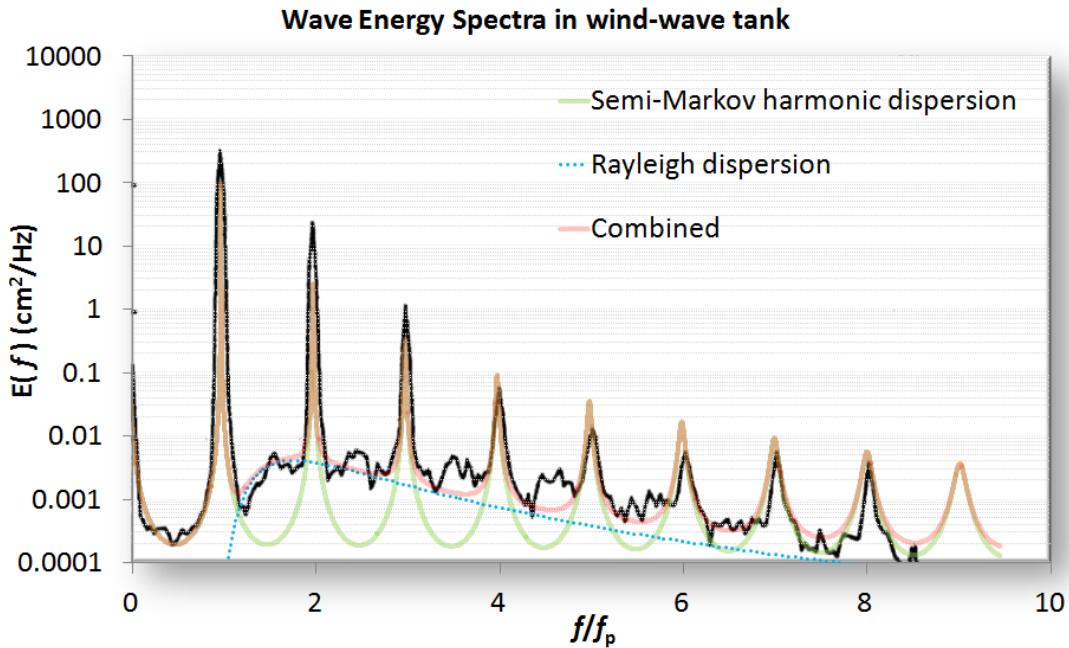

Figure 13: Highly ordered waves with all harmonics generated in a wave tank. This is in terms of temporal frequency and the spatial frequency can be recovered via the dispersion relation.

The basic model is to choose a *P(L)* wave density such that it reproduces the spacing and envelope of the harmonics. One of the simplest density functions is known as the shifted exponential. This maintains a minimum spacing with the possibility of an occasional longer wavelength:

$$P(L) = \begin{cases} 0, & L < L_0 \\ e^{-\alpha(L-L_0)}, & L > L_0 \end{cases} \tag{6}$$

This generates a Fourier transform

$$P(S_x) = \frac{e^{-iS_x L_0}}{1 + iS_x/\alpha}$$

And then the full power spectrum assuming a coherent relationship exists (see **Equation 1**):

$$I(S_x) = \frac{1}{(S_x - \alpha\sin(S_x L_0))^2 + \alpha^2(1 + \cos(S_x L_0))^2} \tag{7}$$

$$I(S_x) = \frac{1}{(S_x + \alpha\sin(2S_x L_0))^2 + \alpha^2(1 - \cos(2S_x L_0))^2} - \frac{1}{S_x^2(1 + \alpha L_0)^2} \tag{8}$$

These two formulations, scaled accordingly, generate the semi-Markov harmonic dispersion model shown in Figure 9. The first equation (7) generates an odd-harmonic waveform due to a symmetric view of up and down steps. The second equation (8) assumes an all-harmonic composition derived by assuming a surface with a saw-tooth set of steps, which is likely nearer the real situation for stimulated waves.

Note that as L_0 approaches zero, then the Markov random walk is recovered (see equation **4**). Thus a simple parametric model with the two coefficients related by a *characteristic period, $< L >$*, can emulate the continuum of order to disorder:

$$\langle L \rangle = L_0 + 1/\alpha$$

As L_0 approaches $< L >$ the spectrum will show strong harmonics. As L_0 approaches 0, the random aspect will suppress the harmonics. As an example of this continuum consider Figure 14 which shows very weak harmonics and higher disorder than the waves of Figure 13: [23]
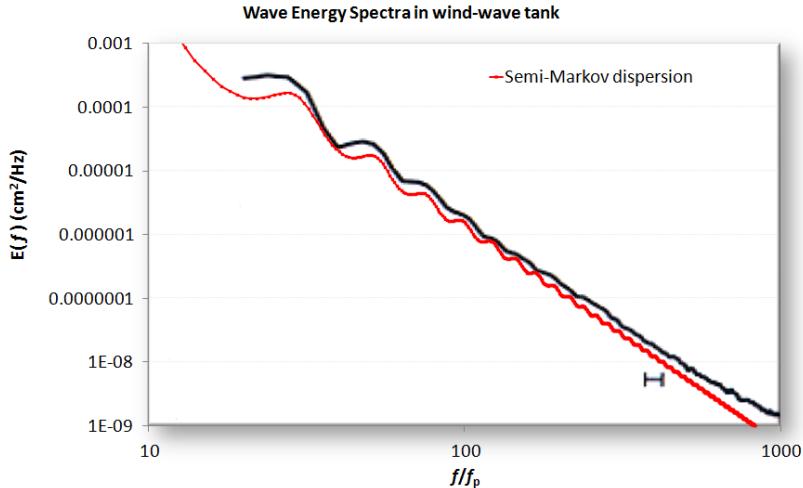

Figure 14: A PSD from a wave tank showing ripples

Note that the dispersion model does, by definition, generate square or sharp edges. Unless there are breakers in the waves, the actual profiles will get rounded. By applying a low-pass filter to the model, the actual waveform will reveal reduced high frequency components and give better agreement with the spectral data. Another example, fit from a simulated wave[24] is shown in figure 11:
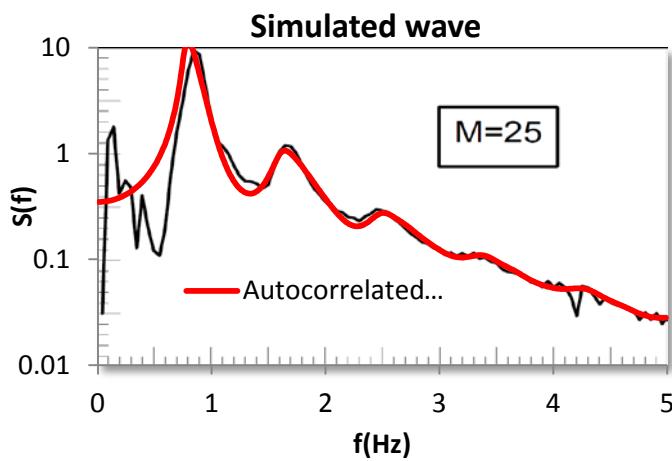

Figure 15: A simulated wave which emulates crest-focused dynamics shows intermediate order

In general, aquatic wave coherence showing strong harmonics occurs under more controlled conditions than normally exist in the wild.

Application to Landforms and Terrain

For terrains we use the same approach as for wave spectra. The significant geometric scale distinction between terrain profiles and aquatic wave profiles is that the former can have very long range variations, stretching to the scale of mountain ranges.

Incoherent Terrain Correlations

Statistics over an extensive large-scale terrain database provides the best example of incoherent variations in surface topography. Figure 12 derives from a detailed terrain slope case study reported in a previous volume [6]. The derivation assumed that maximally sloped regions had larger potential energy and a maximum entropy (MaxEnt) statistical distribution would follow. The chart in Figure 12 is from a single region and that of Figure 13 from the entire country. The former data follows an exponential distribution and the latter super-statistical data follows a Bessel distribution.
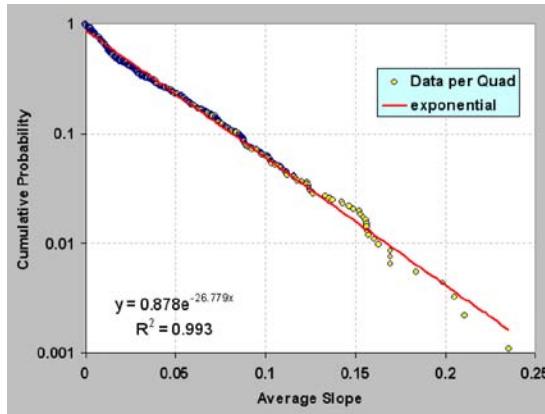

Figure 16 : Probability density function of terrain slopes for an isolated region
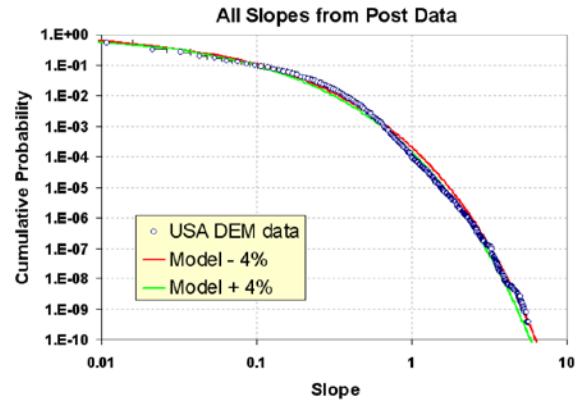

Figure 17 : Probability density function for a wide area

In this case, the term incoherent implies that the slopes at two points separated by a distance can lose correlation at wide spatial separations. In other words, the slope at one terrain point is independent of the value at another point. This is certainly an approximation that does not hold as the two points decrease in spatial separation and merge onto a single location.

Markov Terrain Correlations

We can use a random walk model as a very useful first-order approximation to characterize terrain correlations. Random walk applied in this way falls into the category of a Markov model, where the correlations occur between near-neighbor steps. In general, a pure random walk model featuring stochastic up-and-down steps will show an unbounded variance. This means that the probability of two points separated by any elevation is uniformly distributed between zero and infinity — *given a large enough surface separation*.

A pure random walk defined in this way is impractical and unphysical, as elevation differences are in fact finitely bounded. To allow for the boundedness of real terrains, a random walk process that has *reversion to the mean* properties is required. The classic reversion-to-the-mean variation to the random walk model is known as the Ornstein-Uhlenbeck process.

If one assumes the mean is the average terrain elevation for a localized region, the Ornstein-Uhlenbeck is the simplest stochastic behavior that will retain both Markov properties and a Gaussian spread in

elevation changes. It also generates the commonly observed damped exponential pair-correlation (or autocorrelation) function.

Given a stationary run of random walk transitions, there are known solutions to the Ornstein-Uhlenbeck process, calculated from solving the Fokker-Planck equation; this solution turns out fairly concise in terms of representation.

For a transition probability of elevation change (z) given a surface translation (x)

$$p(z|x) = \sqrt{\frac{\theta}{2\pi D(1 - e^{-2\theta x})}} \cdot e^{\left\{-\frac{\theta}{2D}\left[\frac{(z-\xi e^{-\theta x})^2}{1-e^{-2\theta x}}\right]\right\}}$$

Here, θ represents a drag term for reversion to the mean, and D is the random walk diffusivity or a relative hopping rate in a discrete simulation. The ξ term is an initial condition necessary for representing a starting point away from the mean. If we assume a stationary run, the ξ term disappears and the expression reduces to:

$$p(z|x) = \sqrt{\frac{\theta}{2\pi D(1 - e^{-2\theta x})}} \cdot e^{\left\{-\frac{\theta}{2D}\left[\frac{z^2}{1-e^{-2\theta x}}\right]\right\}}$$

The factor $(1 - e^{-2\theta x})/\theta$ serves as an asymptotic limit which prevents the elevation (z) excursions from getting too large, as a non-linear scaling factor gets applied to create an *effective* surface translation.

A normalized view of the marginal translation probability is shown in Figure 14 below, where y takes the place of z. Note that the contour profile of shows a roughly parabolic shape and this shape begins to asymptotically approach a horizontal along the x-axis. In contrast, for a classic random walk which does not have Ornstein-Uhlenbeck reversion-to-the-mean tendencies, this profile would continue to rise with increasing x, in keeping with a Fickian square-root growth law characteristic of a pure diffusional process. It is this asymptotic behavior which keeps the variance of z-excursions bounded.

$$\sqrt{\frac{\frac{1}{2}}{1 - \exp(-x)}} \; \exp\left(\frac{-\frac{y^2}{2}}{1 - \exp(-x)}\right)$$

Contour plot:

3D plot:
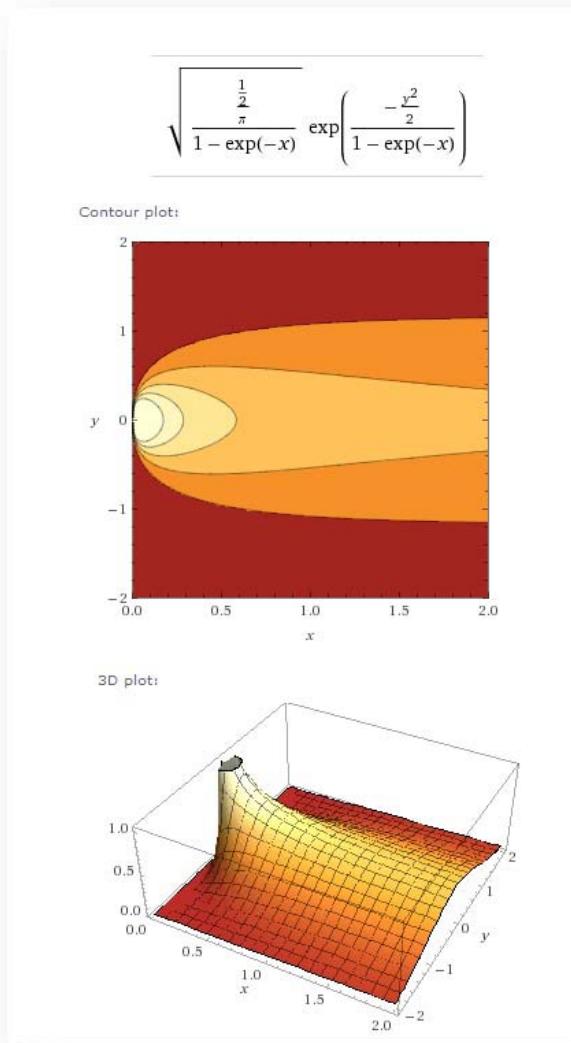
Figure 18: Contour plots of transition probability for the Ornstein-Uhlenbeck process

For an alternative view of what is happening, consider that the marginal transition probability for a pure random walk is given by:

$$p(x, z) = \frac{1}{\sqrt{2\pi D x}} e^{-\frac{z^2}{2Dx}}$$

An Ornstein-Uhlenbeck random walk assumes the transformation

$$x = \frac{(1 - e^{-2\theta\tilde{x}})}{2\theta}$$

This provides a reversion-to-the-mean property, which prevents large or infinite excursions from occurring in a pure unconstrained random walk.

For small x relative to $1/\theta$, this reverts to pure random walk, where

$$\tilde{x} = x$$

For the following derivation, we leave this in the pure random walk formulation and mix in a probability density function reflecting uncertainty in D. We initially assume maximum entropy uncertainty, where we characterize the value of D by its mean value $\langle D \rangle$.

$$p(D) = 1/\langle D \rangle \cdot e^{-D/\langle D \rangle}$$

Then by integration over marginal probabilities, we can estimate a "fuzzy" Ornstein-Uhlenbeck random walk PDF.

$$p(x,z) = \int_{D=0}^{D=\infty} p(x,z|D) \cdot p(D) \, dD$$

$$p(x,z) = \int_{D=0}^{D=\infty} \frac{1}{\sqrt{2\pi Dx}} e^{-\frac{z^2}{2Dx}} \cdot 1/\langle D \rangle \cdot e^{-D/\langle D \rangle} dD$$

This looks like a complicated integral, not amenable to closed form evaluation since the variate D shows up in both the numerator and denominator of an exponential, as well as within a square root. Yet, this does indeed reduce to a very manageable expression thanks to a key integral identity, available from any comprehensive integration table reference. The resultant expression is

$$p(x,z) = \frac{1}{2\sqrt{\langle D \rangle x}} e^{-\frac{|z|}{\sqrt{\langle D \rangle x}}}$$

where $\langle D \rangle$ replaces the original D in the original random walk formulation, assuming the role of the mean diffusivity. No new parameters are added to the original, as we have simply swapped out certainty in D with uncertainty characterized by $\langle D \rangle$. To get to the Ornstein-Uhlenbeck model we simply apply the x-transformation.

$$x = \frac{(1 - e^{-2\theta \tilde{x}})}{2\theta}$$

To mix-in any other probabilities, we simply need to factor in marginal prior probabilities for different values of $\langle D \rangle$. So for instance, assume that very low diffusion, D_0, areas are mixed in by a fraction f, where $0 < f < 1$. Then

$$p'(x,z) = f \cdot p(x,z) + (1-f) \cdot p_0(x,z)$$

$$p'(x,z) = \frac{f}{2\sqrt{\langle D \rangle x}} e^{-\frac{|z|}{\sqrt{\langle D \rangle x}}} + \frac{(1-f)}{2\sqrt{D_0 x}} e^{-\frac{|z|}{\sqrt{D_0 x}}}$$

In general, a model fit would require parametric values for $\langle D \rangle$ and θ, and potentially f and D_0, if an extra inhomogeneity is suggested from the empirical data. The prime example of this would be flat regions that occur due to absorbing boundary conditions via a natural random walk, or due to non-random features such as bodies of water, graded roads, or flatter agricultural regions.

This foundational analysis gives us some room to work with when we try to characterize actually topographies. The premise is that the Ornstein-Uhlenbeck formulation with a single diffusivity D may map well to terrains with a homogeneous random character, while the maximum entropy "fuzzy" Ornstein-Uhlenbeck formulation works better with terrain regions that have mixed heterogeneous character, in the sense that the diffusivity $\langle D \rangle$ has a (prior) range in values which match the variability of the region.

We tested out the two formulations with respect to a set of digital elevation models (DEM) representing the lower 48 states. Each DEM file corresponded to a section that was 1° in latitude by 1° in longitude. This corresponded to approximately 90 meter post spacing in the north-south direction and a variant amount below that value in the east-west direction for increasing latitudes. Each "post" contains an elevation value with respect to sea-level reported to the nearest meter (there are 2401×2401 posts in total per DEM section).

What we wanted to demonstrate was how well an Ornstein-Uhlenbeck model works to describe correlated elevation transitions for relatively small surface translations (< 40 post spacings, or < 4 kilometers). The data from a DEM section was sampled uniformly to capture good counting statistics.

In Figure 15 below, raw histogram counts are shown for an area around Yuma, CA (the El_Centro DEM section). What is readily apparent is the trend toward a contour profile such as that shown in Figure 14.
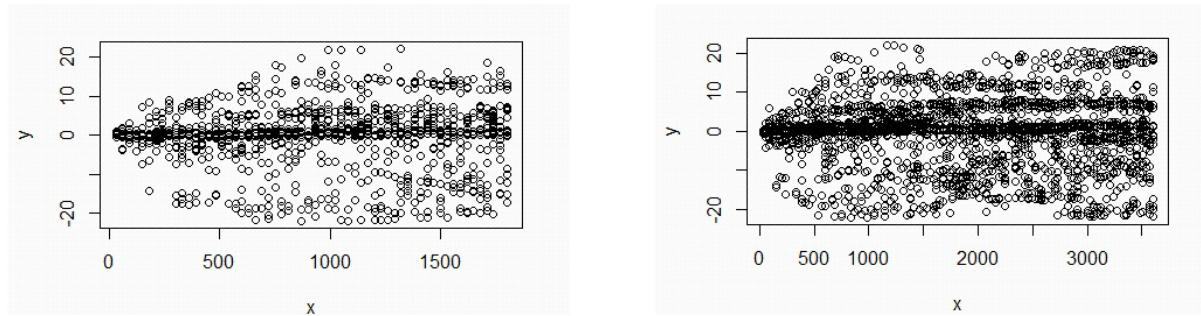

Figure 19: Raw histogram counts for a DEM section around Yuma, California.

A relief plot is shown below, along with a log-scaled contour plot:
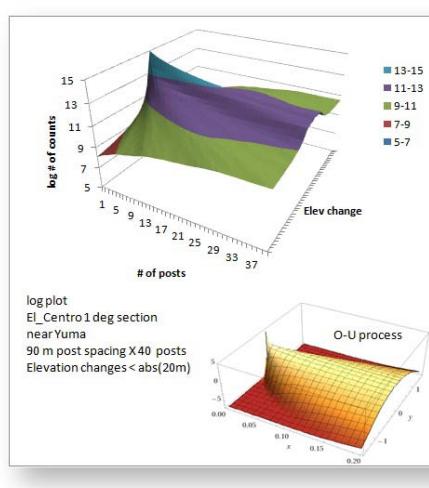

Figure 20: Relief plot
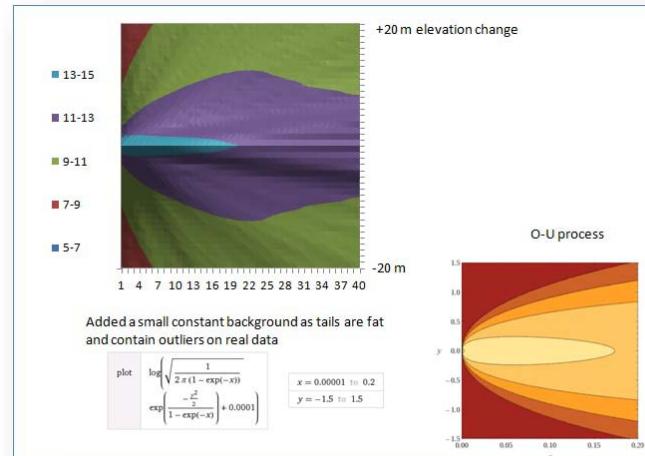
Figure 21: Contour plot of marginal probability

To demonstrate the relative fitting abilities of the conventional Ornstein-Uhlenbeck model against the maximum entropy O-U model, we searched the DEM database for examples that gave good qualitative fits to each model.

Figure 22

The *Andalusia-E* DEM section is located in mid-south Alabama, with terrain that is in between the flatness of the Gulf lowlands and the hilly terrain of the start of the southern Appalachians.

The *Eau Claire-W* DEM section is located in southwestern Wisconsin, with terrain that features hills and valleys that are remnants of one of the few un-glaciated areas of the upper Midwest. To get an appreciation for the topography, the following image is converted from a portion of this DEM located in southeastern Minnesota.
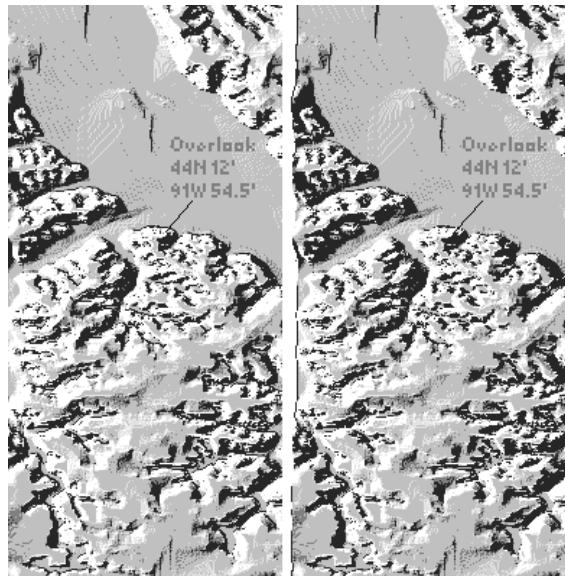

Figure 23 : A stereoscopic rendering of the southwest portion of the Eau Claire-W section. The GPS reading is next to the Upper Mississippi River National Wildlife and Fish Refuge with a view northeast across to Wisconsin.

The fitting of the appropriate O-U model to each data set was accomplished by minimizing the log of the error between the model and the data for each separation and elevation. For the pure Ornstein-Uhlenbeck process, Figure 20 illustrates the salient features, which includes a rounded profile in the relief plot, indicating a single diffusivity. Note that the plots are one-sided as the negative elevation excursion is removed due to symmetry.
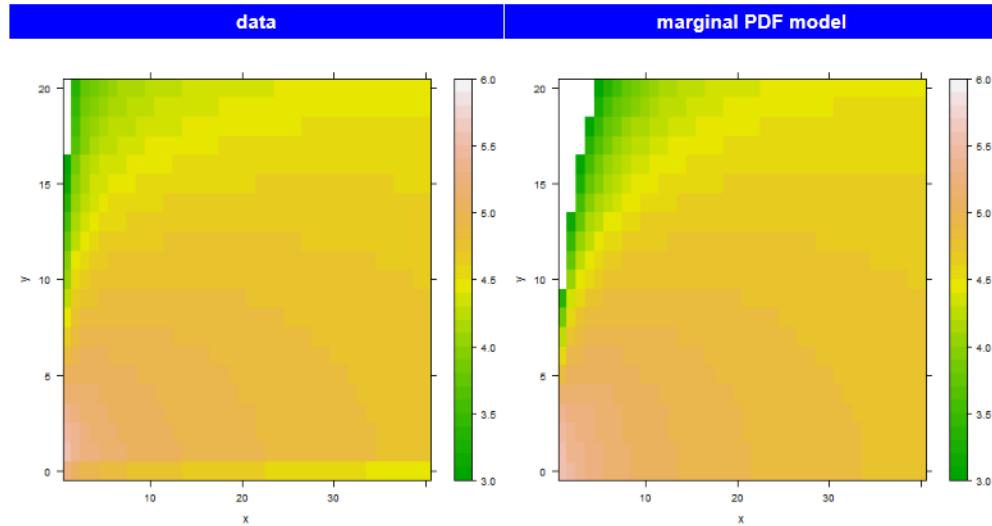
Figure 24: Andalusia-E contour plot

For the maximum entropy Ornstein-Uhlenbeck process, the profile shown in Figure 21 has more of a tail, indicating a spread in diffusivities. This indicates that the unglaciated region of the Eau Claire section has greater variability in its terrain makeup.
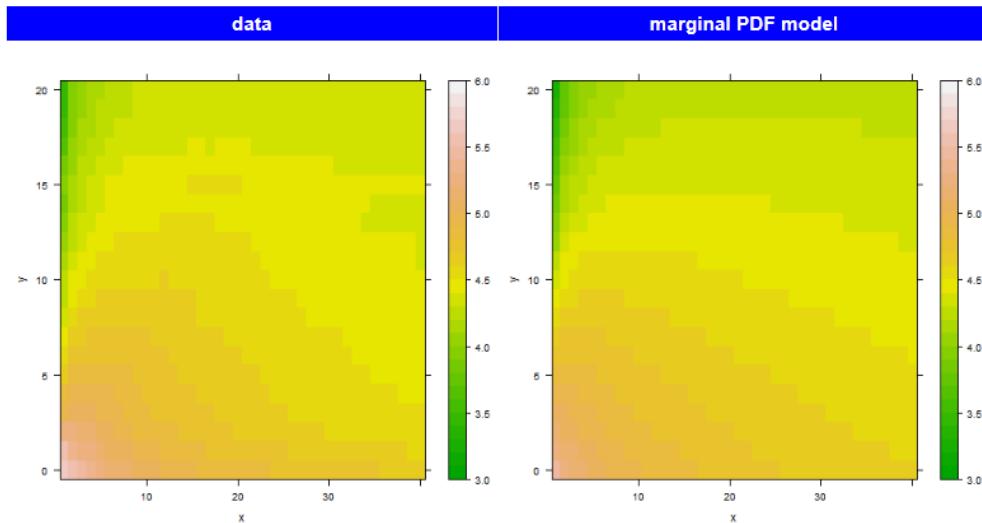

Figure 25: Eau Claire-W contour plot

In Figure 22, we place the data plots along the center to indicate which way the profiles tend toward. For the upper Andalusia contour, the color contours indicate closer agreement to the Ornstein-Uhlenbeck model, while for the lower Eau Claire data, the contour lines map more closely to the Max Entropy prior of the Ornstein-Uhlenbeck model.
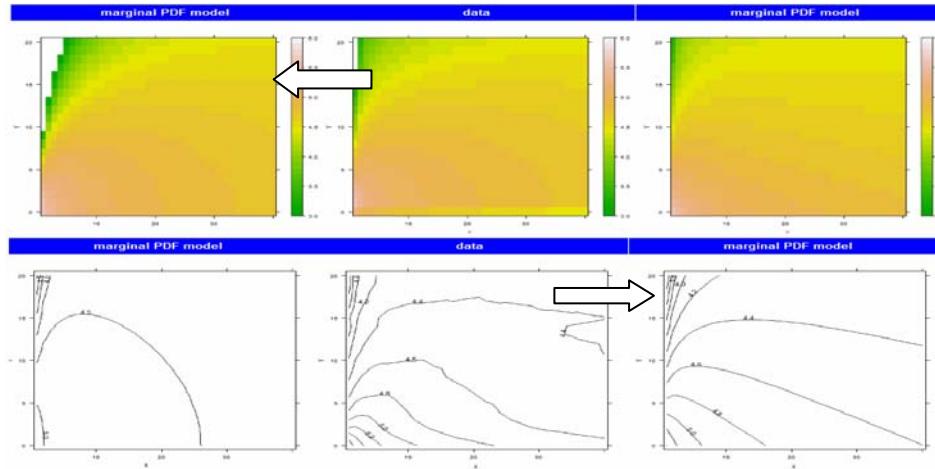
Figure 26: Regions trend either toward an Ornstein-Uhlenbeck profile (left column) or a Maximum Entropy prior of the O-U profile (right column).

A very strong case can be made for the effectiveness of the maximum entropy model in remote topography regions that contain little human terrain modification.
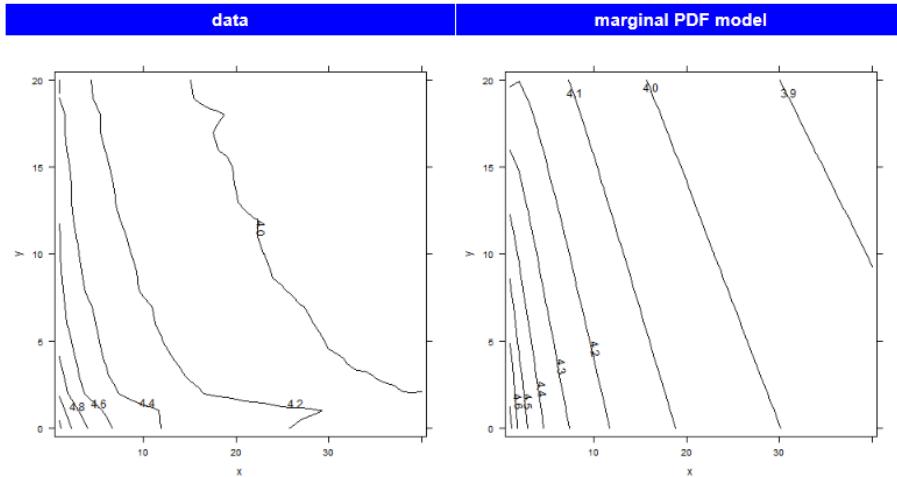
Figure 27: Knoxville-W contour plot of elevation difference probabilities. This is in the middle of the Great Smoky Mountains region, and has very low integrated error against the Maximum Entropy prior of the Ornstein-Uhlenbeck model.

We can use the maximum entropy formulation directly to calculate the censored (or truncated) variance and mean for the data sets and compare that to the theoretical values. A censored variance means that only the collected data points go into the calculation and those that contribute to the true variance ignored. Figure 24 shows the RMS deviation in elevation and the mean elevation as a function of surface translation for the *Knoxville-W* data set.
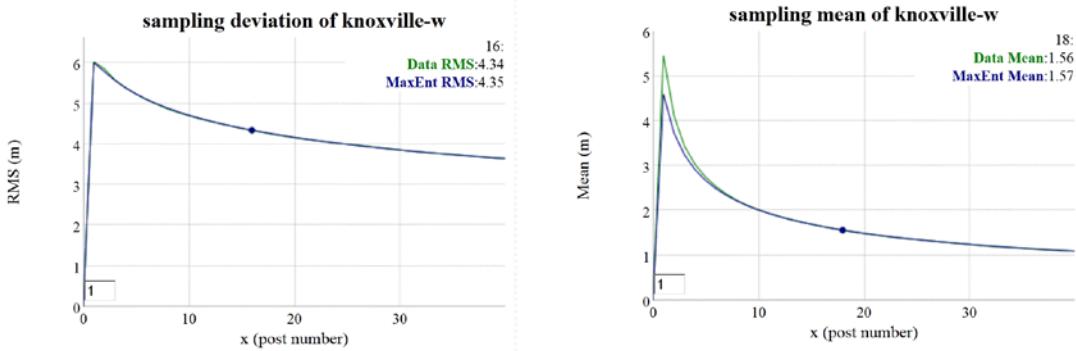
Figure 28: Sampling RMS deviation (left) and mean (right) of Knoxville section data against the Maximum Entropy prior of the Ornstein-Uhlenbeck model. The RMS deviation for both model and data align precisely enough that the curves cannot be distinguished.

The alignment between data and model of this measure mirrors that of the contour plots. This supports the idea that the terrain in these locations is well suited to a MaxEnt-varied random walk model. The simulation of the disordered random walk involves drawing a sample of a diffusion coefficient and then applying that to an Ornstein-Uhlenbeck random walk (see next section).
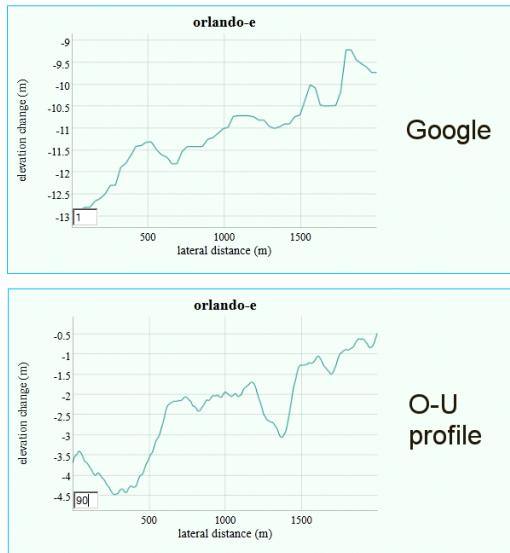
Figure 29: Simulated random walk applied to the Orlando DEM section. Above is an extracted Google elevation profile from a random location within that section. The scale of the excursions match the diffusivity of the random walk.

In addition, the effects of systemic errors and non-stochastic anomalies are readily apparent from the marginal probability density. If artificial correlations exist, they will appear as high density regions in the contour plot (see the strong horizontal bar at elevation changes of ~15m shown in the right inset of Figure 26). The correlations in question come about from a propensity of surveying crews to report elevations to the nearest 50 foot round-off. For example, one can see that red dotted lines in Figure 26 passing through 1800', 1850', 1900', and 1950' contour levels map to elevation areas that tend to flatten out over short intervals. This can only be ascribed to poor data collection or data translation and is systemically observed in DEM data sets that were taken over the relatively flat terrain of the Midwest USA .
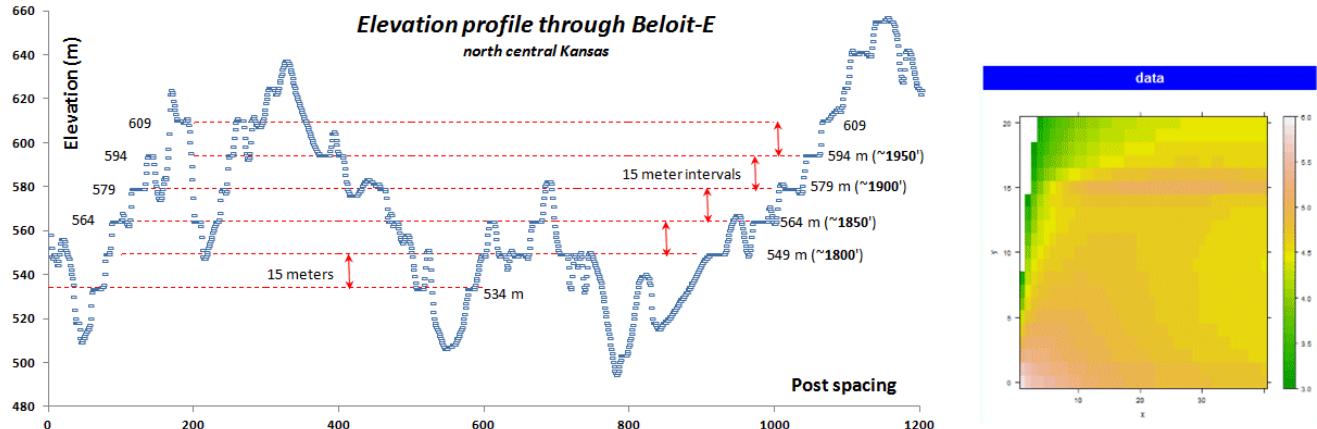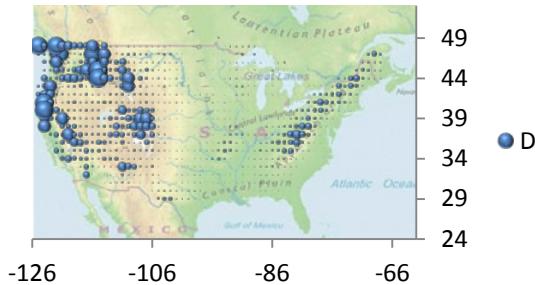
Figure 30: Determining anomalies in terrain elevation modeling. Notice the systemic correlations between elevations separated by rounded 50 foot elevation contours.

The utility of this approach is dependent on the quality and the availability of the data. Mapping out the lower 48 states into 1° sections, we can clearly discern the high diffusivity regions of the country's terrain, bounded by the western and eastern mountains.

We have used this approach to infer models which we can use for formal verification and model checking, as further detailed in Appendix ?.

Semi-Markov Terrain Correlation

Terrain correlations also exist on a much shorter scale, which can also show subtle periodic features. The objective is to model the fine terrain, both random roughness as described in Figure 4, and more periodic features such as cobblestones[25] and washboard[26] as shown in Figure 27 below: [27]
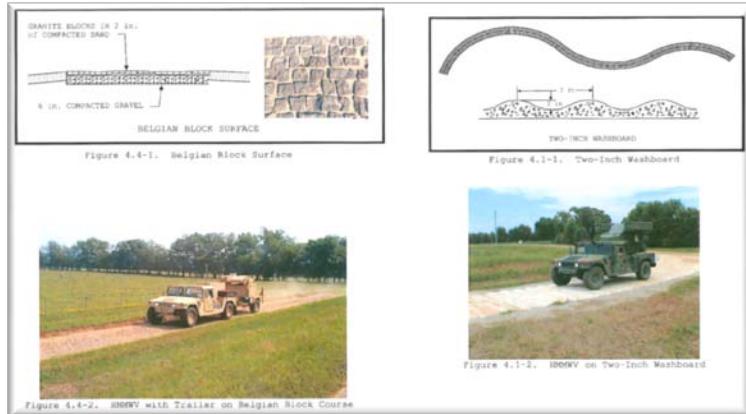
Figure 31: (left) Cobblestone test track and (right) Washboard test track

The Mercedes-Benz test track data used in Figure 4 was also applied to a particular style of road cobblestone paving known as *Belgian Block*. This data was supplied in *OpenCRG* format[12], and covered the track shown in the Google Earth snapshot shown below in Figure 15.

Figure 32: The OpenCRG terrain format features a header which indicates the geospatial location of the data set. This winding Belgian Block course was located underneath an overpass on the Mercedes-Benz campus in Stuttgart, with the start of the path indicated by the green arrow.

A typical one-dimensional profile trace is shown below in Figure 29. Clearly visible are two levels of terrain variability. One variation on the order of 10 cm height contributes to swales on the test track and the other, a fine-grained roughness on the order of 1 cm, are caused by the cobblestone blocks themselves.
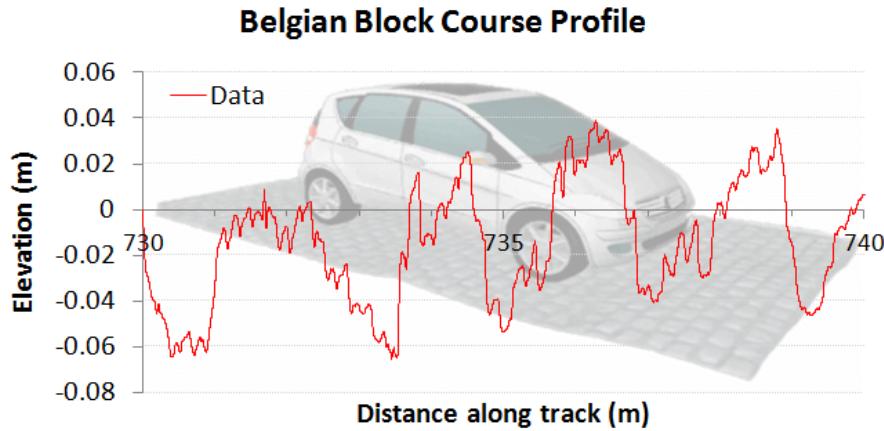
Figure 33: A profile trace along a short path of the test track indicates the roughness.

The fine resolution of the data is shown in Figure 30, which is on the order of 1 cm in the lateral dimension.
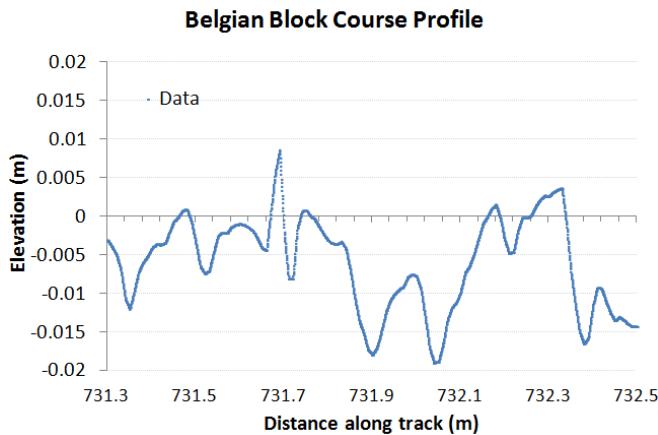
Figure 34 : A very short profile along the test track showing the Belgian Block variability.

The PSD calculated for this course is shown in Figure 31. Although the harmonics are not as striking as that shown for the aquatic wave PSD of Figure 9, their positions correlate well with a model of the terrain consisting of varying degrees of order (shown by the blue curve). The low-frequency odd harmonics correspond to washboard-like swales on the order of 1.5 meters. These are odd-harmonics because they show a tendency to an asymmetric slant or tilt. The higher frequency even harmonics are caused by the Belgian block perturbations which are about 10 cm in lateral spacing.

The overall envelope suggests that a strong random walk component also exists, with a low density filter cutting off the high frequency sharpness in the terrain roughness. Over time, the cobblestones will wear down their sharp edges, and this is observed in the filtered trace.

The idea of using a superposition of various profiles from **Equation 7** makes intuitive sense since the various periodicities have very little mutual coherence, and the pair correlations are additive in contributing to the power spectrum.
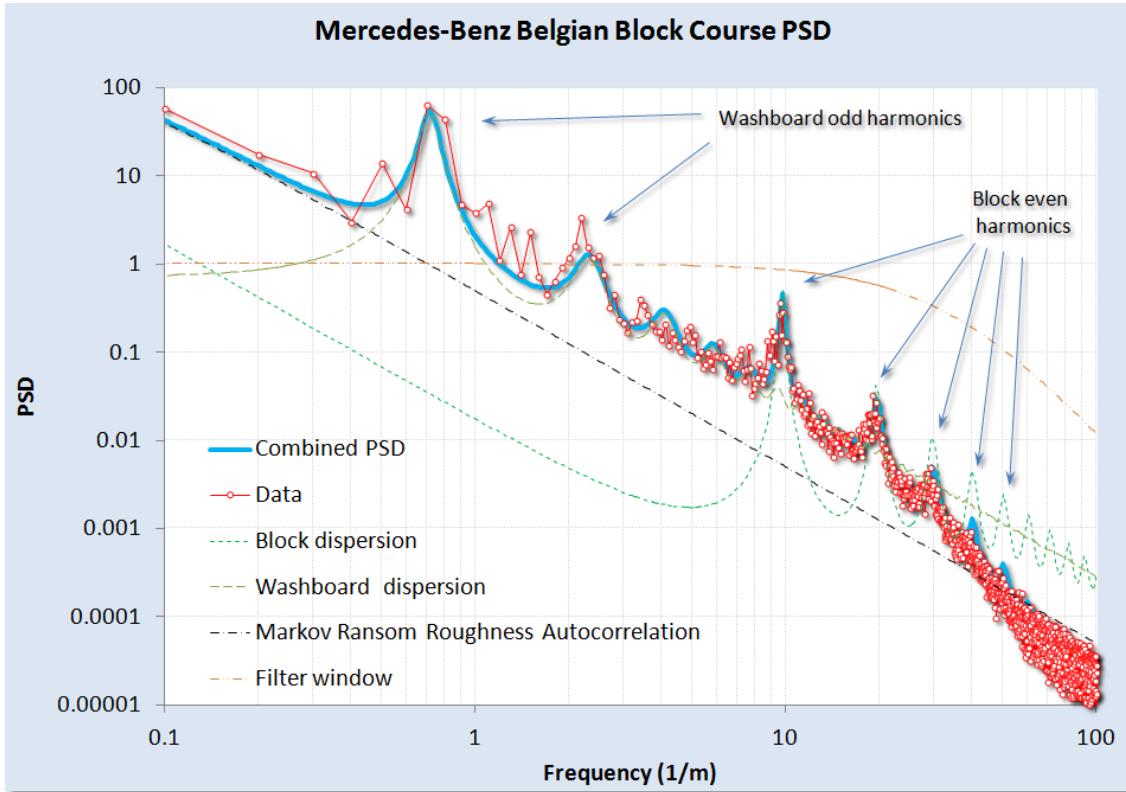
Figure 35: A model fit to the Belgian Block course PSD. A rich variety of surface textures can be gleaned from the harmonics. At high wavenumbers, a change in slope is detected, likely due to a smoothing spatial filter applied to the fine terrain relief and partly due to interpolation for the OpenCRG data set

For the Belgian block course, the data was not detrended prior to computing a spectrum. A slight tilt does exist in the data as can be gleaned from Figure 32, where several sections of 2D Fourier Transform PSD's for simulated (left and right) and real (center) terrains. The general approach of treating both the Z and X dimensions in a full power spectrum allows us to sensitively test for stepped or staircase surfaces.
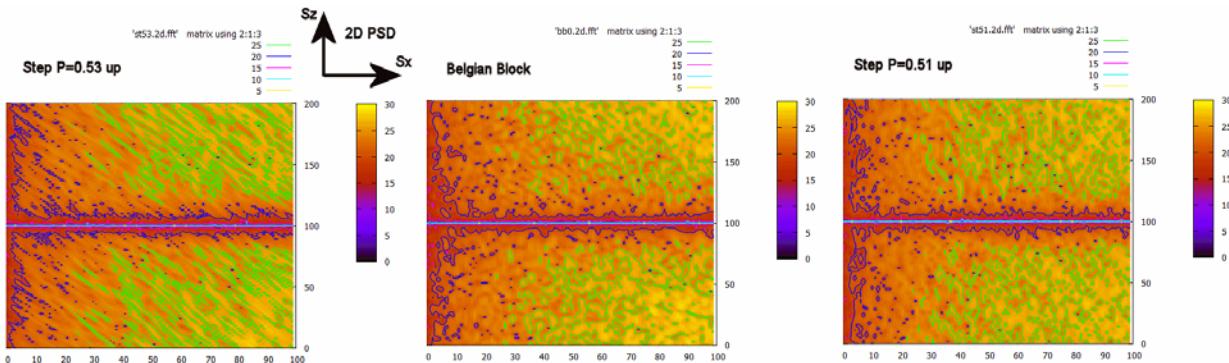
Figure 36: Using the full 2D power spectrum, we can discern subtle changes in slope from any staircase pattern by looking at slanted texture. The center is the real data, and the two side maps show a slightly higher slope (left) and a slightly lower slope (right). The logarithm was taken of the PSD after the transform was multiplied by S_x^2 and S_z^2

As a control study, we also collected PSD data from a military test course and compared the results in Figure 33. This data was detrended prior to use, and was the only unclassified data available for analysis. Similar features are observed as with the Mercedes-Benz Belgian Block course, with the same washboard and block harmonics. These are slightly shifted in period, an effect likely due to different road construction and maintenance procedures.
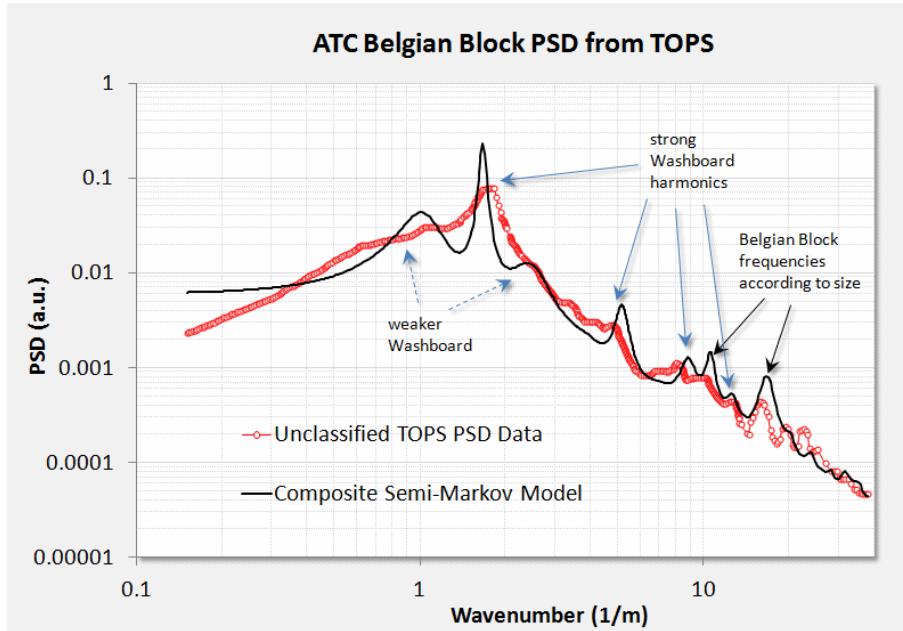
Figure 37: Belgian Block course from a military test operational procedure (TOP) document.

A few other profiles are worth considering. For Figure 34, from [28], we aligned an even-harmonic semi-Markov PSD with the empirically calculated PSD. The strong harmonics of a washboard road are expected as the repeated travel of vehicles over the course reinforces the washboard resonant frequency.
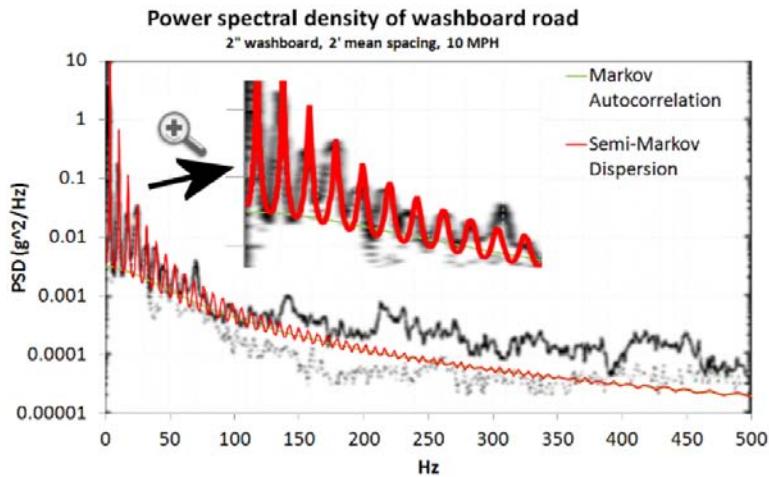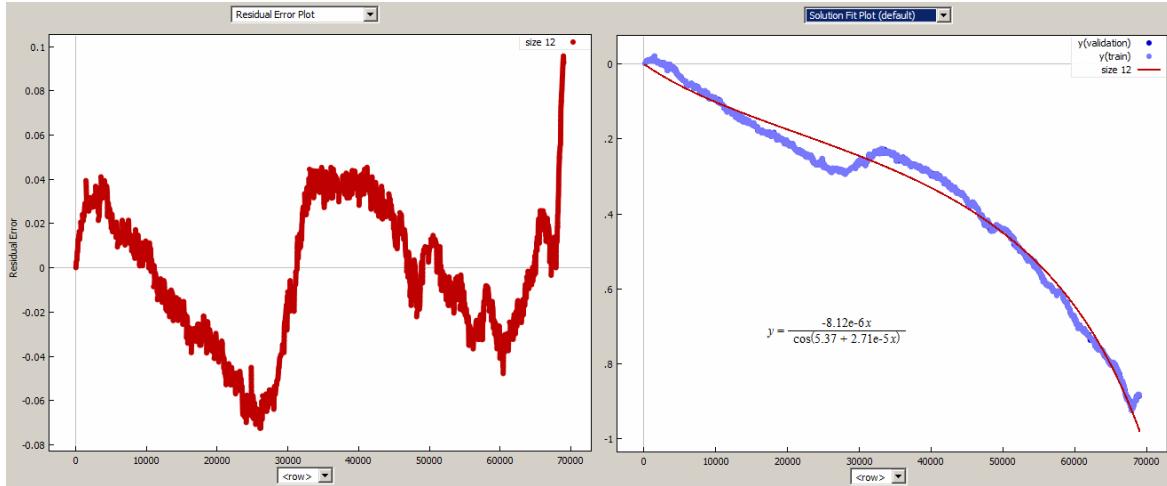
Figure 38: A model fit to a pure washboard. The x-axis is expressed in Hz to indicate the test vehicle was maintaining travel at a constant speed.

We can also return to the rough road power spectrum of Figure 4. We noted that this course likely contained a long-wavelength (deterministic) slope. The data was detrended as shown in Figure 35 below.

*Figure 39: The elevation trend in the rough road data set (right) was removed by a curve fitting program (**Eureqa**). The detrended residuals are shown to the left.*

After detrending the terrain profile, the rough-road PSD was calculated and fit to a semi-Markov model with one low-frequency component and a high-frequency periodic component as shown in Figure 36. The harmonics on the high-frequency spikes suggested that the underlying roughness was shaped similarly to a cobblestone surface.. The first and second harmonics are strong, but the third roughly cancels out. By creating a two-level surface where the troughs were half as long as the tops, we could duplicate the missing third harmonic as shown by the solid red line in the figure.

A sinc function filter was added to reproduce the reduced strength of wavenumbers above 10/m.
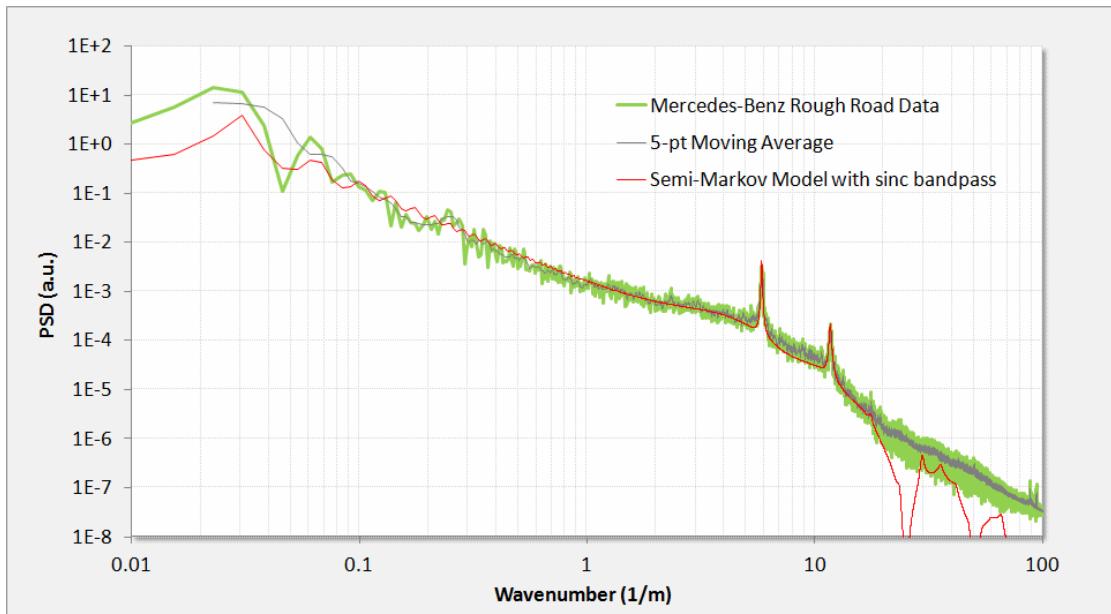
Figure 40: Rough road PSD. From the location and spacing of the harmonics, we can infer a periodic structure very close to a Belgian block structure.

The two-level spectra is a generalization of the result shown in **Equation 8**.

$$I_0(S_x, S_z) = \mathbb{Re}\left\{\frac{1}{S_x{}^2\langle L\rangle}\frac{(1-P(S_x))(1-Q(S_x))}{1-P(S_x)Q(S_x)}\right\} \tag{9}$$

Here each level has its own probability density of lengths, which allows for a surface with wider plateaus and narrower grooves.

$$P(S_x) = \frac{e^{-iS_xL_\alpha}}{1+{iS_x}/{\alpha}}, \qquad Q(S_x) = \frac{e^{-iS_xL_\beta}}{1+{iS_x}/{\beta}}$$

A simulated random walk for this terrain relief is shown in Figure 37 alongside an arbitrary path length from the actual terrain. To mimic the sinc filter, a rectangular window moving average was applied to the random walk profile. Note that after calibration of the relative step heights, very good qualitative agreement exists with the actual surface.

The simplified expression from **Equation (9)** is complicated only because of the number of cross terms needed to pair correlate the two levels

$$I = \frac{(\alpha S)^2 + (\beta S)^2 + (\alpha\beta)^2 S^4 - (\beta\,S)^2\cos(SL_\alpha) - (\alpha\,S)^2\cos(SL_\beta) + \alpha\beta^2 S^3\sin(SL_\alpha) + \alpha^2\beta S^3\sin(SL_\beta)}{\left(1-\alpha\beta S^2 - \cos(S(L_\alpha+L_\beta))\right)^2 + (\alpha S + \beta S + \sin(S(L_\alpha+L_\beta)))^2}$$

To generate a synthetic terrain, each level is treated as a sampling distribution drawn from the distributions from each level (see next section).
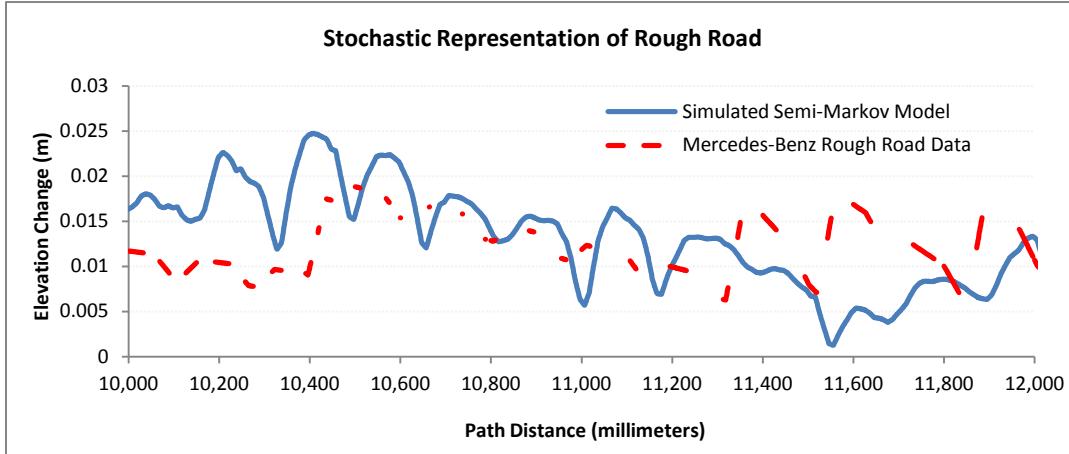
Figure 41: Simulated random walk for the rough road terrain.

As a check to determine whether we can recover the PSD from the simulated semi-Markov random walk, we applied a Fourier transform to 10 sets of simulation traces and squared the amplitude. Note the close alignment with the peaks and with the sinc filter poles. This demonstrates the general utility of switching between a stochastic analytical representation and a Monte Carlo generated simulation.
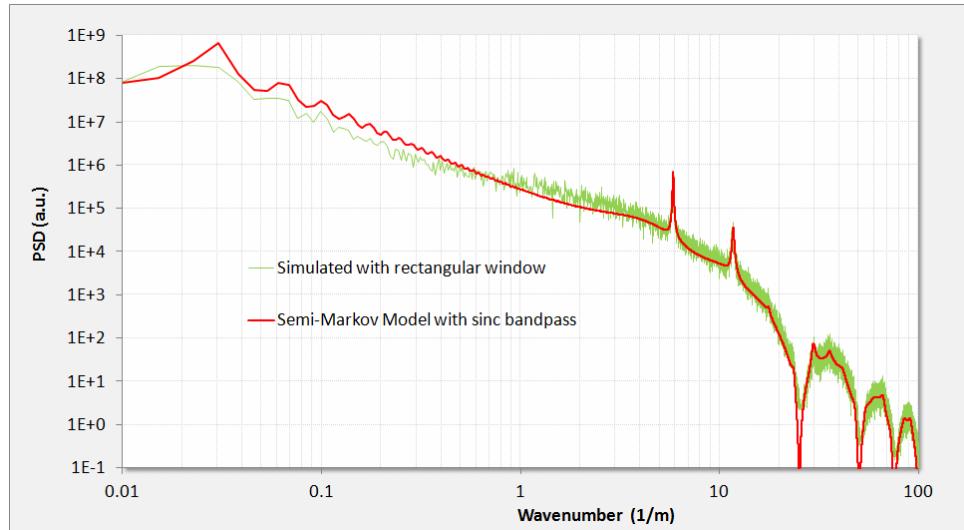
Figure 42: A Monte Carlo simulation of the rough road semi-Markov behavior reveals nearly the same feature amplitudes for the assumed pair-correlation weightings. Contrast this with Figure 36

High Resolution Data

To verify the utility of the semi-Markov approach, we evaluated a fitting procedure against several high-resolution profiles of real vehicle test courses; specifically we chose the Gerotek site in South Africa as this was being measured with advanced profiling tools [29].

To calibrate the data format and determine the limits of resolution, we started with a simulated pothole course. We experimented with the size of the data set so as to clearly reveal features in the PSD, which are shown in Figure 43.
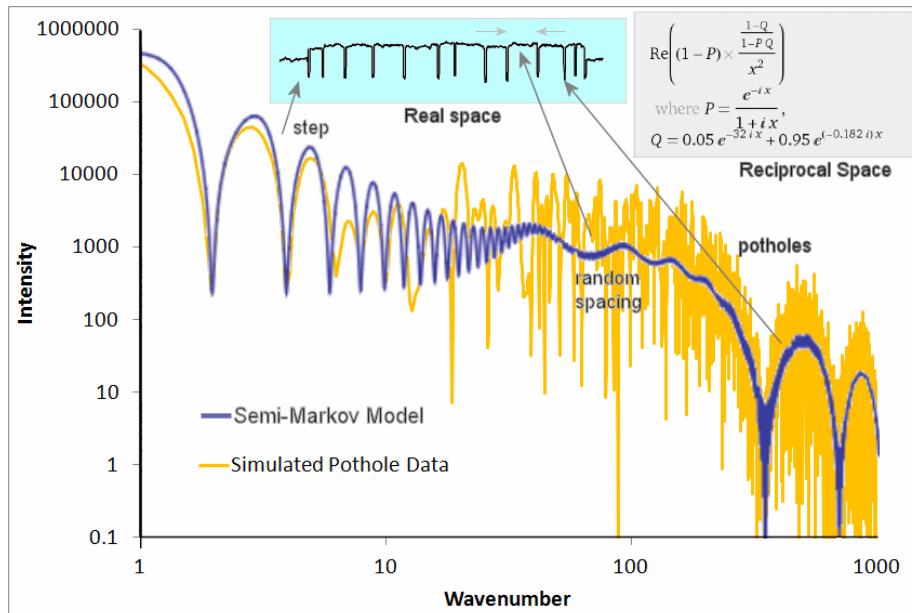
Figure 43: Simulated pothole course PSD. The typical fitting procedure uses knowledge of the underlying terrain profile (top middle inset, showing a pot-hole course with steps) and a

stochastic representation of the PSD (upper right inset) to match the PSD calculated from the data.

For another course at the Gerotek site, we chose the "corrugated" track, consisting of approximately 20 cm high bumps, separated at rather regular intervals. This also provided a very accurate fit to a semi-Markov model (see Figure 44), with the PSD showing a suppressed high harmonic at the spatial frequency predicted by the ratio of the length of the bump to the average spacing.
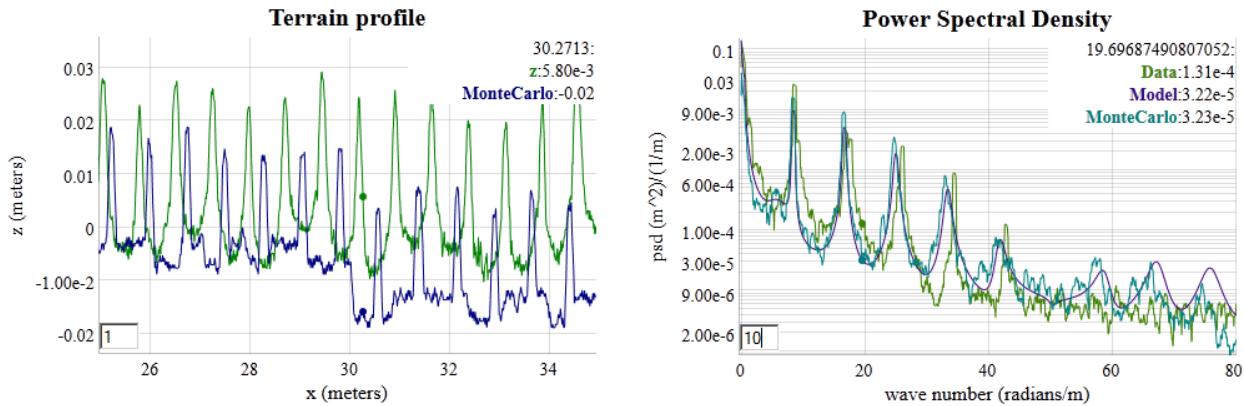
Figure 44: Gerotek Corrugation course. (left) track profile and simulated profile (right) PSD data, simulated PSD, and analytical data

The semi-Markov model with very few parameters is thus able to reproduce a data set containing potentially gigabytes of data.

Generating Synthetic Terrains and Waves and Monte Carlo Sampling

Having a stochastic model of the terrain allows one to generate instances of synthetic terrain that have the same spectral content as the real terrain. This becomes useful for mapping out an *ergodic* representation of possible terrain states suitable for *sampling* (i.e. *Monte Carlo*) simulations.

Several mechanisms exist to generate a random walk, which includes the classical Brownian motion based on a Markov model, the semi-Markov step transition, and the Ornstein-Uhlenbeck random walk process.

Classical Random Walk Model

The simplest *random walk* models will generate a PSD with a $1/S^2$ fall-off, and exemplified by the empirical data in Figure 4. The problem with the simple random walk profile is that it will generate an infinitely high spike for $S=0$, as a random walk is unbounded and will show undulations of infinite length and height. For an example of a typical algorithm for generating a Markov random walk in discrete steps, assume R is a sample drawn as a uniform variate: $R \in [0..1]$.

```
-- classical random walk = rw
rw(Diffusion,  Z)
    random(R)
    if R < 0.5 then
        Z = Z + Diffusion
    else
        Z = Z - Diffusion
    end
```

Ornstein-Uhlenbeck Process

The issue with a pure random walk model is that the absolute excursions become unbounded, whereas real data shows bounds in altitude in the terrain or waves. One way around this, which does not impact the spectral fall-off, is to attach a *reversion-to-the-mean* correction term to the random walk algorithm. This procedure is known as the *Ornstein-Uhlenbeck* random walk process[4], and derives from a physical model of an attractor or potential well which "tugs" on the random walker to bring it back to the mean state (see Figure 45 below).
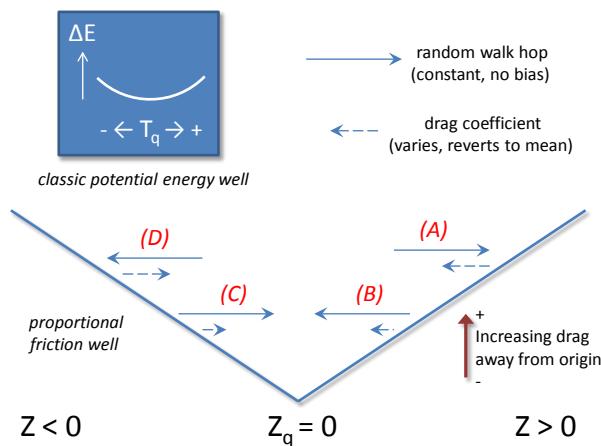
Figure 45: Representation of an Ornstein-Uhlenbeck random walk process for terrain elevation changes. The hopping rate works similarly to a potential well, with a greater resistance to hopping the further excursion ways from a quiescent elevation (Z_q) changes.

The following pseudo-code snippet sets up an Ornstein-Uhlenbeck random walk model with a reversion-to-the-mean term. The *diffusion* term is the classical Markovian random walk transition rate. The *drag* term places an attractor which opposes large excursions in the terrain elevation term, *Z*.

```
-- Ornstein-Uhlenbeck random walk = ou
ou(X1, X2, Drag,  Z)
    random(R)
    if R < 0.5 then
        Z = Z*X1 + X2
    else
        Z = Z*X1 - X2
    end

--  This is how it gets parameterized
ou_random_walker (dX, Diffusion, Drag, Z)
    X1 = exp(-2*Drag*dX)
    X2 =  sqrt(Diffusion*(1-exp(-2*Drag*dX))/2/Drag)
    ou(X1, X2, Drag, Z)
```

To determine whether an Ornstein-Uhlenbeck process is apparent on a set of data, one can apply a simple multiscale variance (or a multiscale entropy measure [10]) to the result of a *Z* array of length *N* :

```
variance(Z,N) {
   L = N/2
   while(L > 1) {
     Sum = 0.0
     for(i=1; i<N/2; i++) {
        Val = Z[i] - Z[i+L]
        Sum += Val * Val
     }
     print L " " sqrt(Sum/(N/2))
     L = 0.95*L;
   }
}
```

So that for a given random walk simulation, the asymptotic variance will tend to saturate at longer correlation length scales. A typical multiscale variance plot will look like Figure 42.

The autocorrelation of the Ornstein-Uhlenbeck process is, where $\theta=Drag$:

$$R(x) = \frac{D}{\theta}\, e^{-\theta|x|}$$

Even though this shows a saturation level, the power spectrum still obeys a $1/S^2$ fall-off.[30]

$$I(S_x) = \frac{D}{\pi} \cdot \frac{1}{S_x^{\,2} + \theta^2}$$
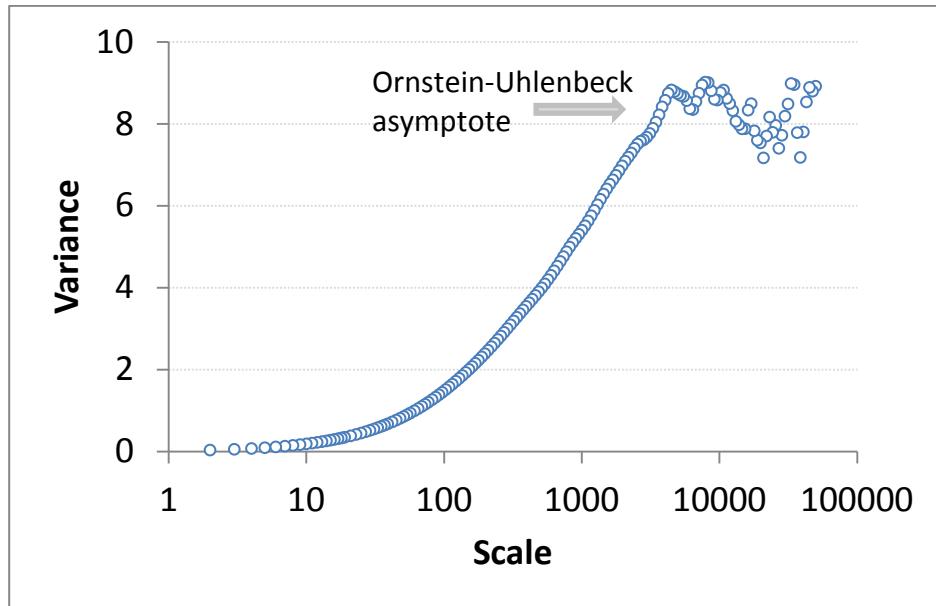
Figure 46: Ornstein-Uhlenbeck process saturates on variance

The Orenstein-Uhlenbeck process is often referred to as *red noise* and the two parameters of Diffusion and Drag can be determined either from the autocorrelation function or from the PSD. For the PSD, on a log-log plot, this involves reading the peak near *S*=0 and then determining the shoulder of the power-law roll-off. Between these two measures, one can infer both parameter values.

Semi-Markov Process
The key to mapping to semi-Markov distributions is to draw from the appropriate PDF. For example, the delayed exponential (see **Equation 6**) is simply a draw from an exponential distribution with a fixed offset (i.e. the delay) added to the variate.

The step deltas and terrace deltas can both draw from a semi-Markov distribution; taking the PSD of the resulting step sequence will generate a noisy **Equation 7**, depending on the amount of samples taken.

The algorithm for generating this semi-Markov model is as follows, where **P** is a probability of stepping up or down after each length. This has no offset in length. If **P** is not 0.5 precisely, then the two-dimensional PSD will show an asymmetry reflecting the asymmetry of the terrain profile.

```
BEGIN {
   N = 100000
   h = 0
   srand()
   i = N
   while(i>0) {
     n = - 10.0 * log(rand())
     for(j=0; j<n; j++) {
        i--
        print h
     }
     if(rand()>P) {
        h += +0.0008 * log(rand())
     } else {
        h += -0.0008 * log(rand())
     }
   }
}
```

The "rough road" data model shown previously was generated according to this recipe.

As for the analytical analog semi-Markov model which reproduces the spectral characteristics, consider Figure 47 below and compare it to the Monte Carlo simulation which generates Figure 36. This again is the simplest semi-Markov model, a damped exponential size for both step length and step height.

| log plot	$\mathrm{Re}\left(\left(1 - P - H \times \dfrac{(1-P)^2}{1-HP}\right) z^2\right)$ where $P = \dfrac{1}{1+ix}, H = \dfrac{1}{1+iz}$
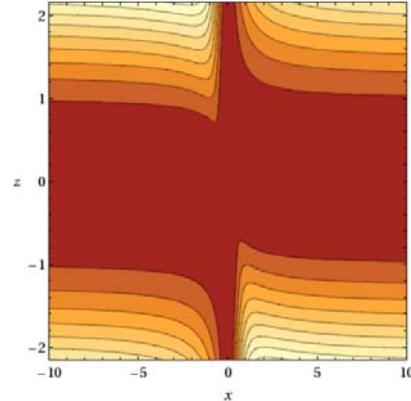
Figure 47 : 2D contour plot of power spectrum for a randomly ascending staircase. To illustrate a 2D power spectrum in terms of the orthogonal wavenumbers S_x and S_z, we multiply by the wavenumber squared and then take the logarithm to give the greatest dynamic range.

The Belgian Block course suggests an asymmetry 0.51 ± 0.01 as shown in Figure 36, indicating sensitivity to incline.

Process of Fitting to Synthetic Terrain
If we have a terrain that indicates some periodicity, these steps will synthesize a semi-Markov profile:

1. Get terrain data [*x*, *z*] as distance/displacement pairs
2. Run FFT on data assuming *x* equally spaced, and take magnitude squared
3. Fit the FFT curve to $w \cdot \mathrm{I}(s \mid \alpha, L)$
   w = scaling
   α = local order
   L = quasi-periodicity
   s = wavenumber

$$\mathrm{I}(s \mid \alpha, L) = \frac{1}{(s - \alpha \cdot \sin(sL))^2 + \alpha^2(1 + \cos(sL))^2}$$

4. Draw a number of Monte Carlo samples from the PDF => $P(x \mid \alpha, L)$.
   For each sample, use the cookbook inversion of a PDF to generate a step length, *x*

$$x = L - \frac{1}{\alpha} \cdot \ln(\mathrm{rand}())$$

5. Move the step up and down alternately with amount corresponding to the RMS weighting of *w*

$$z = c \cdot \sqrt{L}$$

   (if we need to detrend the data against large excursions, apply Ornstein-Uhlenbeck to *z*)
6. Save the step lengths and step heights as a compact array of [*x*, *z*] pairs

Verifying the synthetic PSD

1. Get [*x, z*] data from last terrain synthesis and discretize so that *x* is on equal intervals (for the FFT)
2. Run FFT and the magnitude squared
3. Compare the curves. Adjust the step size if magnitude is not to scale

The intensity of a rough or cobbled terrain with differing lengths of troughs and tops is more complicated than that shown here. For the "rough road" model, we need to modify the step lengths to alternate depending on whether they lie on a trough or a top. The approach is straightforward but the pair correlation increases the number of terms in the intensity profile (*I(s)*) quadratically (see Equation 10).

Process Filters

If we wish to add a second-order Markovian feedback, we can filter the data with beyond near-neighbor interactions, i.e. not only **Z[i]** and **Z[i-1]** but **Z[i-2]** and beyond. This adds an additional $1/S^2$ factor as it acts as a low-pass smoothing filter. In the case of an autocorrelation for a second-order step change:

$$c(x) = \int_{y=0}^{\infty} [\alpha^2 y\, e^{-\alpha y}] \cdot [\alpha^2 (y + x)\, e^{-\alpha(y+x)}] \; dy$$

This will generate a power spectrum of order 4:

$$I(S_x) = \frac{1}{\sqrt{2\pi}} \frac{\alpha^4}{(\alpha^2 + S_x{}^2)^2}$$

Ordinarily, we can apply the class of filters known as autoregressive models - AR(*p*), where *p* indicates the interaction order of the model, or in terms of digital processing, finite impulse response (FIR) filters. In general, these will all tend to smooth the simulation step changes, rounding out the edges as a low-pass filter is designed to work.

Superposition of Waves

A non-stochastic method for generating a synthetic terrain involves extracting the Fourier coefficients from the PSD and superposing sine waves to recreate the original real-space terrain profile. This kind of inversion will generate a repeat sequence with multiples of the longest time period in the spectrum. No phase information is available from the power spectrum, so any long range coherence is artificially introduced, unlike that from a pair correlation. Although potentially useful and practical[31], the superposition approach will definitely not traverse a complete state space and so is not ergodic. It will also not capture skewness and kurtosis in the terrain profile[32], of which the stochastic model can more easily (see Figure 37).

A good application of the spectral superposition approach is for generating aquatic waves for various sea-states in the incoherent regime. Since aquatic waves are already close to sinusoidal, they do not need extra harmonics with phase relationships to match angular shapes (as a terrain might). And the randomness in superposing incoherent waves is visually familiar. This is why most graphics applications use the superposition approach for rendering waves [33].

Calibration of Spectra

The calibration of simulated terrains and their corresponding PSD curves to PSD calculations of actual terrains is aided by the application of Parseval's theorem [34]:

$$\langle z^2 \rangle = \int I(S)dS$$

This states that the statistical variance in the terrain elevation displacement is equivalent to the integrated power spectrum. In practice, as long as the variance in the real space sampling is stored, then the reciprocal space curve can be shifted by a constant scaling factor to match the variance.

Discussion

Stochastic terrain modeling as described within this study finds practical application for vehicle test and verification applications [35], with the terrain and energy models providing an environmental context for such activities as safety testing. We are essentially trying to map out the *Importance × Likelihood* space for the users of the context models. The users must know the impact or significance of the model effects as well as it likelihood of occurring.

The relevance for a particular topographical model depends on its intended usage. For example, one needs to ask the basic questions:

- What is the importance or impact of the event?
- What is the likelihood of the event?

These occupy largely orthogonal roles as shown in figure 30 below:

Figure 48: Likelihood versus importance

In the case of terrain, a vehicle operating over a rough profile would experience nuisance effects (vibration and absorbed power) that persistently occur. It could also potentially experience a critical effect (extreme slope) that rarely if ever occurs.

Table 2: Orthogonal spaces

	Description	**Examples**
High Likelihood	Environmental stimulus that have high likelihood, but only a cumulative impact	Persistent and perpetual nuisance effects ▪ Terrain roughness and vibration ▪ Wear and tear ▪ Abrasion ▪ Wind friction and rolling resistance
High Impact	Environmental stimulus that have high impact but often low likelihoods	A critical effect that rarely if ever occurs ▪ A large shock

	▪ Accidents
	▪ 100-year climate events
	▪ … lots of other possible cases
	▪ … or some other pathological or unknown cases

The overall goal is to provide models of environmental stimulus for design verification, using the probabilities inferred from the stochastic models to estimate correctness of the designs.

Note that assessing the impact for a specific design requires composing the context model with a performance model for the design being tested. The context model serves as a stimulus, and the simulation for the design being tested then generates a response. Assessment of a design in terms of probability of failure or probability of satisfying a requirement, involves a likelihood weighted sampling of impact, in general done with Monte Carlo techniques.[3] Note that as failure conditions will typically be rare, and it is desirable to estimate probably of failure accurately, importance sampling techniques are frequently needed. This is true even though highly parallel computer hardware now allow many thousands of cases to be executed in parallel.

Conclusion

A unified approach for characterizing and modeling topographies has definite advantages for environmental context applications. We have demonstrated how often simple probability density functions (PDF) can characterize the impact/likelihood factors for various environmental behaviors. A stochastic basis for the models provides an explanation for empirically observed behavior that heuristics often miss. Further, unifying an established set of scientific and ontological criteria helps us to derive and classify the models. This creates an opportunity to establish a collection of context models as a community resource, with supporting technology facilitating its correct utilization for a given purpose. Such a resource can substantially improve the utility of available data for supporting Model Based Engineering, and also create opportunities for improving existing design methods by facilitating the discovery of highly resilient designs.

---

[3] techniques such as importance sampling[36] may be required

References

[1]   M. Waechter, F. Riess, H. Kantz, and J. Peinke, "Stochastic analysis of surface roughness -," *EPL (Europhysics Letters)*, vol. 64, no. 5, p. 579, 2003.

[2]   R. H. Stewart, *Introduction to physical oceanography*. A & M University, 2003.

[3]   A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*. Prentice Hall, 1994.

[4]   D. Mumford and A. Desolneux, *Pattern Theory: The Stochastic Analysis Of Real-World Signals*. A K Peters, Ltd., 2010.

[5]   P. R. Pukite and J. Pukite, "Digital signal processors for computation intensive statistics and simulation," *SIGSIM Simul. Dig.*, vol. 21, no. 2, pp. 20–29, Dec. 1990.

[6]   P. R. Pukite, *The Oil Conundrum: Vol. 1 Decline, Vol. 2 Renewal*, vol. 1,2, 2 vols. Daina, 2011.

[7]   K. Chowdary, "Distinguising and Integrating Aleatoric and Epistemic Variation in UQ," Brown University, 2011.

[8]   E. T. Jaynes and G. L. Bretthorst, *Probability theory: the logic of science*. Cambridge Univ Pr, 2003.

[9]   C. Beck, "Generalized statistical mechanics for superstatistical systems," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1935, pp. 453–465, Dec. 2010.

[10]   P. Pukite and S. Bankes, "Entropic Complexity Measured in Context Switching," in *Applications of Digital Siignal Processing*, vol. 17, InTech, 2011.

[11]   P. R. Pukite, C. S. Lent, and P. I. Cohen, "Diffraction from stepped surfaces:: II. Arbitrary terrace distributions," *Surface science*, vol. 161, no. 1, pp. 39–68, 1985.

[12]   OpenCRG, "OpenCRG Homepage." [Online]. Available: http://www.opencrg.org/. [Accessed: 26-Mar-2012].

[13]   F. Behroozi and A. Perkins, "Direct measurement of the dispersion relation of capillary waves by laser interferometry," *American journal of physics*, vol. 74, p. 957, 2006.

[14]   E. Falcon and C. Laroche, "Observation of depth-induced properties in wave turbulence on the surface of a fluid," *EPL (Europhysics Letters)*, vol. 95, p. 34003, 2011.

[15]   "Cnoidal wave - Wikipedia, the free encyclopedia." [Online]. Available: http://en.wikipedia.org/wiki/Cnoidal_wave. [Accessed: 31-Jan-2013].

[16]   P. R. Pukite, *Reflection High Energy Electron Diffraction Studies of Interface Formation*. University of Minnesota, 1988.

[17]   PATRICK HOLMES, PhD, "A COURSE IN COASTAL DEFENSE SYSTEMS I CHAPTER 5 COASTAL PROCESSES: WAVES," *CDCM Professional Training Programme, 2001*. [Online]. Available: http://www.oas.org/cdcm_train/courses/course21/chap_05.pdf. [Accessed: 24-Apr-2012].

[18]   FormSys, "Wave Spectra." [Online]. Available: http://www.formsys.com/extras/FDS/webhelp/seakeeper/wave_spectra1.htm. [Accessed: 24-Apr-2012].

[19]   Huang, T.; Alarcon, C.; Bingham, A.; Henderson, M. L.; Kessling, M.; Takagi, A.; Thompson, C. K, "NASA ADS: Data-Driven Oceanographic Web Portal." [Online]. Available: http://adsabs.harvard.edu/abs/2010AGUFMIN23A1350H. [Accessed: 12-Apr-2012].

[20]   CDIP, "CDIP Homepage." [Online]. Available: http://cdip.ucsd.edu/?units=metric&tz=UTC&pub=public&map_stati=1,2,3. [Accessed: 24-Apr-2012].

[21]   T. S. Rhee, P. D. Nightingale, D. K. Woolf, G. Caulliez, P. Bowyer, and M. O. Andreae, "Influence of energetic wind and waves on gas transfer in a large wind–wave tunnel facility," *J. Geophys. Res.*, vol. 112, no. C5, p. C05027, May 2007.

[22]   F. Feddersen and F. Veron, "Wind effects on shoaling wave shape," *Journal of physical oceanography*, vol. 35, no. 7, pp. 1223–1228, 2005.

[23]   W. B. Wright, R. Budakian, D. J. Pine, and S. J. Putterman, "Imaging of Intermittency in Ripple-Wave Turbulence," *Science*, vol. 278, no. 5343, pp. 1609 –1612, Nov. 1997.

[24]   D. Ning, J. Zang, S. Liu, R. Eatock Taylor, B. Teng, and P. Taylor, "Free-surface evolution and wave kinematics for nonlinear uni-directional focused wave groups," *Ocean Engineering*, vol. 36, no. 15, pp. 1226–1243, 2009.

[25]   BTRC, "BTRC Track Features." [Online]. Available: http://www.vss.psu.edu/BTRC/btrc_track_features.htm. [Accessed: 24-Apr-2012].

[26]   E. E. Fitch, "Durability Analysis Method for Nontationary Random Vibration," MIT, 1996.

[27]   Automotive Directorate, "Test Operations Procedure (TOP) 1-1-010 Vehicle Test Course Severity (Surface Roughness)," ATEC, Aberdeen Proving Ground, MD, Final TOP 1-1-010, 2006.

[28]   G. F. Sievert, "Effects of stabilizer bars on road vehicle ride quality," Rice, 1994.

[29]   C. Becker and P. S. Els, "GEROTEK data from CSIR for C2M2L program." .

[30]   B. Lindner, "A BRIEF INTRODUCTION TO SOME SIMPLE STOCHASTIC PROCESSES," *Stochastic Methods in Neuroscience*, p. 1, 2009.

[31]   J. E. Alexander, "Synthesis of a PSD Compatible Acceleration Time-History," *Submitted to Shock & Vibration Symposium (November 5-9, 2012)*.

[32]   A. Steinwoff and W. H. Connon III, "Limitations of the Fourier transform for describing test course profiles," *Sound and Vibration*, vol. 39, no. 2, pp. 12–17, 2005.

[33]   J. Tessendorf, "Simulating ocean water," *Simulating Nature: Realistic and Interactive Techniques. SIGGRAPH*, 2001.

[34]   L. Romero and W. K. Melville, "Spatial Statistics of the Sea Surface in Fetch-Limited Conditions," *Journal of Physical Oceanography*, 2011.

[35]   Jonathan Cameron, Abhi Jain, Terry Huntsberger, Garett Sohl, and Rudranarayan Mukherjee, "Vehicle-Terrain Interaction Modeling and Validation for Planetary Rovers," NASA JPL, Dartslab, JPL Publication 10-15, 2009.

[36]   M. Denny, "Introduction to importance sampling in rare-event simulations," *European Journal of Physics*, vol. 22, p. 403, 2001.