# Stochastic Analysis for Context Modeling

**Abstract:** Many common behaviors in the natural science contain similar patterns. These are often related by fundamental principles of thermodynamics and statistical physics. This series of papers sets forth a set of approaches for the basic probability categories that several natural processes belong to.

## Preface

Whenever we discover new patterns in our natural environment, we can often take technical advantage of the gained knowledge. In many cases, we can characterize and then potentially apply the information to everyday situations and *contexts*, such as transportation and energy use. More and more frequently we find that applying contextual models based on environmental patterns has become a vital task in engineering product development. Contextual variables are defined as *"a set of characteristics that influences the organization's design processes".* A significant cause of design failure often relates to providing an insufficient safety margin to deleterious environmental factors and contextual variables, such as an exceedingly rough terrain or corrosive atmospheric elements.

As many of the environmental patterns stem from fundamental processes deriving from statistical mechanics and thermodynamics, we can gain significant insight from approaching the patterns from first-principles physics. When we put the patterns and a fundamental understanding together, we can often extrapolate and infer behavior outside of the data set that the observations were drawn from. This makes them powerful for prediction and verification tasks.

This paper introduces a set of research results relating to stochastic modeling of environmental contexts. We first provide some background to probability theory and for formulating the building blocks for stochastic analysis — in particular, that of *probability density functions* and Markov and semi-Markov processes. The building blocks comprise the core fundamental aspects of the stochastic modeling, whereby we can reason about probabilities, sampling, and uncertainty. The more sophisticated techniques of autocorrelation and power spectrum densities are elaborated in a report on terrain characterization, and then models of dispersive growth and uncertainty propagation are described within a diffusion context.

Since these models taken together are comprehensive in their scope, we will need to tie them together and classify them according to a contextual ontology. In practice, the reader or end-user would get steered to the correct probability model based on the working knowledge available and the environmental context required.

## Background: The role of the environment in cyber-physical systems

For practicality's sake, we consider two criteria for applying probability to some observed behavior: (1) the importance or impact of the event and (2) the likelihood of the event. This allows us to concentrate on contexts that play a significant role either through their frequency of occurrence or the severity of their impact. To be useful, the metric of interest has to be measurable or countable. For a select set of environmental behaviors, the measures could involve extensive variables such as volumes or of measures consisting of ratios such as rate (i.e. volume over time).

*Table 1 : The use of probabilities can describe high likelihood and high impact events.*

|  | High Likelihood | High Impact |
|---|---|---|
| Terrain | RMS roughness | Steep slopes |
| Rainfall | Humid climates | Heavy downpours |
| Wind | Prevailing winds | Gusts |
| Waves | Chop and swell | Rogue waves |

| Particulates | Aerosols | High density volcanic dust |
|---|---|---|

The fundamental building block is the use of probability density functions (PDF) to model sample spaces. These facilitate the characterization and modeling of natural phenomena that are prevalent in human environmental contexts, including slope distribution, wind, rainfall, lakes, etc. These can both model the high likelihood events through sampling of the meat of the distribution curves, but also provide for the rare cases through the concept of exceedance probabilities. To make them more useful, the goal is to find if any of these models reduce to fairly simple analytical forms. For disordered systems and data containing uncertainty, techniques such as the *maximum entropy principle* (MaxEnt) and *superstatistics* will be used; these often have a more formal basis than the heuristic fractal models often employed.

To discover patterns in data, we can make heavy use of data analysis techniques such as rank histogram plots. A set of data, binned according to frequency of occurrence for the parameter of interest reveals most of the structure of the probability density. These views can be manipulated or marginalized against conditional or joint probabilities.

One of the significant observations that one can make about typical environmental parameters is in the extent of their randomness. On occasion an environmental parameter, such as temperature, can exist within a narrow range of values and thus become well-suited for a normal Gaussian distribution model, but more commonly, the values have enough spread that skewed (i.e. asymmetric) and fat-tail distributions are much more applicable. The data along with some physical reasoning will direct the model away from a normal distribution toward a higher variance distribution.

Environmental context modeling relies on knowledge of exogenous behavior – that behavior that exists outside the confines of the vehicle we are designing. In fact, any behavior that we have little control over fits under the umbrella of a context. When nature gets involved, the behavior often tends toward disorder, as the thermodynamic laws describing entropy are immutable (and entropy describes the movement toward disorder).

As part of the motivation for thinking in terms of substantial randomness, consider the paper by D. Mumford, *"The Dawning of the Age of Stochasticity"*.[1]

> *"stochastic models and statistical reasoning are more relevant (i) to the world,(ii) to science and many parts of mathematics and iii) particularly to understanding the computations in our own minds, than exact models and logical reasoning"*

This paper served as the motivation for one of the several mathematical challenges[2] posed by DARPA's Defense Sciences Office -- the third challenge, **"Capture and Harness Stochasticity in Nature"**, has the potential for providing the DoD with significant and immediate benefit. Mumford concludes with this:

> *"stochastic methods will transform pure and applied mathematics in the beginning of the third millenium. Probability and statistics will come to be viewed as the natural tools to use in mathematical as well as scientific modeling. The intellectual world as a whole will come to view logic as a beautiful elegant idealization but to view statistics as the standard way in which we reason and think."*

We assert that this benefit can be immediate, as recent advances in our understanding of stochastic phenomena have benefited greatly from the availability of data from a variety of sources. In the past, modeling of physical behavior has often been hampered by the lack of sufficient statistics to substantiate the original formulation. In combining stochastics and information elements for modeling, we can incorporate probability and information theory (Jaynes, Shannon), pattern theory (Mumford, Grenander), fat-tail statistics (Mandelbrot, Taleb, Sornette), and superstatistics and complexity theory (Beck, Gell-Mann) and then apply these contemporary ideas to the characterization of environmental contexts.

Probability theory as advanced by E.T. Jaynes suggests using probability as an extended logic, and we should consider inference and plausible reasoning under various levels of uncertainty. The key idea of Jaynes is to meld Shannon's information theory concept of entropy together with the statistical mechanics definition of entropy. Many important and non-trivial applications exist where Jaynes' maximum entropy principle is the only tool we may need, as it describes the minimal application of prior knowledge when appropriate — often a mean value is all that is required.[3]

Pattern theory as advanced by Mumford[4] and Grenander seeks to identify the hidden variables of a data set, characterizing real world observations as patterns. The approach uses the observed patterns to infer information about the unobservable factors, formulating prior distributions for those too complex or difficult to observe. If we can determine efficient inferential models for the observed variables by modeling observations as partly stochastic and partly deterministic, and apply the randomness and variability of these distributions along with considering their natural structures and constraints (i.e. symmetries, independences of parts, and marginals on key statistics) we can create classes of stochastic models by applying transformations to patterns. We can then synthesize (sample) from the models, and the stochastic factors affecting an observation exhibit strong conditional independence, making it easily decomposable. We will see this approach demonstrated when we consider terrain characterization.

The analysis of fat-tail statistics as advanced by Sornette[5] and Taleb[6] has shown promise for the prediction of crises and extreme events in complex systems and risk management, both for social[7] and natural systems. The general theory encompasses scale-free properties, fractals, and power-laws, and provides an alternative to normal or Gaussian statistics. This introduces black and gray swan terminology and the idea of rare dragon-kings which relates to extreme value analysis (EVA), and also via 100-year events.

Considerations of complexity theory as advanced by Gell-Mann[8] leads to the idea that :*"when defining complexity, it is always necessary to specify a level of detail up to which the system is described, with finer details being ignored"*. Seemingly complex representations can often be represented by rather concise descriptions and we can apply concepts such as dispersion and coarse graining to simplify the complexity. One such idea is that of superstatistics[9], which ties in closely to the ideas of maximum entropy [10]. The essential approach here is to admit that randomness can exist on different scales and by combining these scales, the underlying real-world statistical distributions are revealed.

The awareness that we can indeed use probability to characterize larger scale phenomena has often been fought tooth-and-nail by opposing schools of thought. For example, Mumford describes how classical statisticians opposed contextual Bayesian modeling when they claimed that *"this approach denies that statistical inference can have anything to do with real thought because real-life situations are always buried in contextual variables and cannot be repeated."* In this case, the contextual variables appear to get in the way of our understanding of the desired effect, whereas they should become part of the understanding of the *system*: i.e. the context of the vehicle within the environment. In reality, we can and have created very useful models by incorporating prior contextual knowledge to infer possible and potential behaviors. The key is that many of the contextual variables are governed by properties of nature that repeatedly occur under conditions of the thermodynamic arrow of time, which always leads to greater amounts of entropy. In this sense, nature can perhaps be more predictable than we think, and at worst, we can expect that it remain predictable in its unpredictability, and thus we can make progress by applying a stochastic characterization to the empirical observations.

# Maximum Entropy Principle Modeling

To apply context modeling effectively one has to approach the environmental domains pragmatically. We want to find solutions with the minimal amount of complexity that conversely generates the most general benefit. Consider the situation of spending an inordinate amount of time generating a model with a very detailed representation. Chances are that no one would use the model and, if used, they may apply it only in specific cases. So we first seek the simplest possible approach and find out if that has general applicability.

Fortunately, nature helps us out with the bookkeeping, through its implicit use of information theory, or as it is known in the physical sciences world – statistical mechanics. Thermodynamics and the essential notion of entropy both derive from statistical physics. The essential idea behind information theory is to try to describe an observed phenomenon with the least amount of words as possible. Gell-Mann uses a connectivity pattern to drive home this point. Consider Figure 1 below, a series of networks showing increasing levels of seeming complexity. Graph **A** seems the least complex as it has no connections, while Graph **F** seemingly shows the most amount of complexity as it has all nodes interconnected.
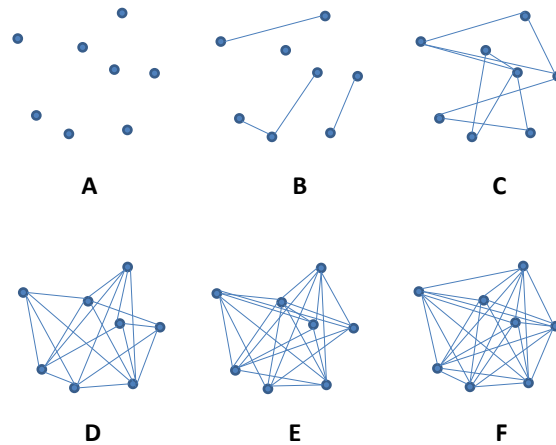


*Figure 1: Gell-Mann's connectivity patterns. Increasing levels of complexity shown as an interconnected graph.*

Using information theory, we can estimate the actual complexity by algorithmically describing each of the connectivity arrangements. Graph **A** can be described as *"no nodes connected"*, while Graph **F** can be described as *"all nodes connected"*. These therefore have very close to the same information content, because they can be reduced to an equally concise description. The graphs in between these end-points have higher complexity because they will require a more elaborate description. That seems a particularly non-intuitive contradiction, but that is the way that scientists have harnessed the ideas of entropy – seemingly complex representations can often be represented by surprisingly concise descriptions.

In physics, this approach has long been applied to describe topics such as the ideal gas law, which stems from seminal ideas of statistical mechanics. Ensembles of particles, though complex at a micro-scale, can be described more concisely at the macro-scale, with the connecting thread elegantly applied from entropy and information theory [Jaynes 2001].

The question is whether we can apply these ideas to other natural phenomena. The key idea that we propose to apply is that of the Maximum Entropy Principle (MaxEnt). The principle of maximum entropy

states that those configurations described by the least amount of constraints and moments (such as mean, variance, and higher-order moments) will tend toward probability density functions that maximize entropy. The MaxEnt solution is essentially to maximize **S** = entropy subject to constraints of probability, $f(p_i)$.

$$\max{}_{f(p_i)} S = -\sum p_i \cdot \ln p_i$$

(1)

In practice, MaxEnt gets routinely applied, even without the knowledge of the practitioners. For example, with the assumption of only a mean and a standard deviation in a set of empirical observations, the conventional approach is to apply the standard Gaussian or normal statistical distribution. That indeed coincides with the MaxEnt derived probability distribution for that particular set of moments. The canonical example of this is a manufactured component with a tight tolerance; this would routinely show a Normal distribution in the measurements of the produced parts. (see the RLC example of ratio distributions from the AVM META report[11])

Another obvious example is that of a dye mixing in a glass of water. The eventual spatial distribution of the dye in the volume will approach a uniform level over time. This is straightforwardly modeled as a MaxEnt calculation given the boundary conditions of the volume. There is no preference for the dye occupying one subvolume over another so that we can only apply the known physical constraints. The barometric pressure law with altitude is one of the simplest examples of adding an energy constraint to Jaynes formulation; and this gives the characteristic damped exponential decrease of pressure with altitude.

In addition, countless other phenomena, especially those showing great amounts of disorder, do not follow either a Normal distribution, or the uniform distribution or exponential just described. This includes many of the so-called "fat-tail" distributions, popularized by recent events [Taleb 2007]. Further, some of the fat-tails come about not from any intrinsic physical property, but from derived property of the measure space.

This leads into consideration of epistemic versus aleatory uncertainty.

## Uncertainty Quantification

In addition to the natural random variation in the values of a particular environmental observable (**Type 1**: aleatoric uncertainties), our limited abilities to quantify the numbers can lead to a further uncertainty spread (**Type 2**: epistemic or systemic uncertainties). Whether the epistemic uncertainties allow us to capture the underlying randomness depends on the strength of the characterization and modeling.

For example, we can have uncertainty in the actual measurements due to calibration or precision errors, uncertainty due to counting statistics, and uncertainty in the applicability of a model.

If the observable has a wide dynamic range and is readily quantified, the uncertainty in measurement is often absorbed in the natural randomness and has minimal impact on the data characterization. For counting statistics, uncertainty is minimized by drawing from a sufficient sample size that represents the complete dynamic range of the observable. By accumulating data over time, using approaches such as Bayes rule via additional prior knowledge, the additional evidence will reduce this uncertainty.

The applicability of the model for describing a natural phenomenon is referred to as the *likelihood* of the model and of its parameters. The certainty or confidence we have in a particular model is ultimately best gauged by comparison to an alternative model in which we can apply standard inference metrics such as

conditional entropy, maximum likelihood, log-likelihood, and information entropy criteria such as Akaike Information Criterion (AIC) or Bayes Information Criterion (BIC). The information criteria techniques are valuable because they penalize models that contain too many fitting parameters. Models based on first-principles with minimal parameterization (such as those derived from the Maximum Entropy Principle) will always score higher than, for example, a naïve high-order polynomial fit that contains many adjustable parameters.

The goal of context modeling is to converge to only **Type 1** uncertainties, where we can apply what we consider the most likely model and use its probability distribution function to provide Monte Carlo or importance sampling for design verification.

This is a natural follow-on to our previous META metrics work where we applied propagation of uncertainty in combination with various physical and information-entropy-based models to arrive at estimates of a probabilistic certificate of correctness (PCC) for a given design. This included exogenous artificial effects and sets of metrics dealing specifically with exogenous variables, those variables whose value is determined outside the model in which it is used.

- Manufacturing variance (the RLC example)
- Semantic network links
- Travel dispersion
- Wireless signal latency
- Human reaction times

Our innovation comes from further application of the maximum entropy formulation and a careful consideration of the measure space[1]. In terms of applicability to engineering, we assert that by applying minimal information to stochastic models of various behaviors, that we can infer the essential probability distributions, and therefore concisely model contextual behaviors suitable for conversion (in reverse) to uncertainty bounds — necessary for tasks such as verifying vehicle or system environmental suitability. Following this process, we have several innovative claims regarding stochastic modeling and thus provide a deep knowledge about physical processes that others can build on and use in interesting ways. For one, our description of modeling approaches for disordered systems and data containing uncertainty, using techniques such as the maximum entropy principle and superstatistics, has a more formal basis than the heuristic or fractal models conventionally used.

The basics described in the following will lead to more sophisticated analyses, including:

➔ Growth curves

➔ Generalized correlation functions in the real-space domain

➔ Combining correlation functions with spectral representations, both in the spatial and temporal domains of system context.
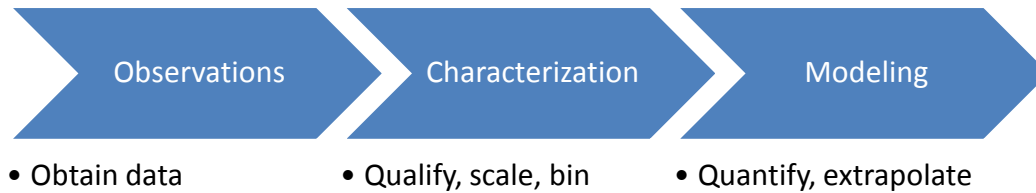
## Characterization to Modeling

In some sense, context modeling is really about applying the scientific method to describe the physical world. Figure 2 below shows the path going from (1) initial observations, to (2) characterization, and then to (3) modeling.

---

[1] This approach is not as common as one would think, considering the amount of detailed analysis undertaken with conventional statistical techniques (by non-domain experts who have a knowledge of statistics). We want to occupy the pragmatic middle ground and borrow insight from both physics and statistical camps.

*Figure 2: Process of Modeling*

Stage 1 represents gathering the initial data from observations and measurements.

Stage 2 transforms the raw data to usable information — referring to it as characterization imparts deeper categorical qualities and meaning to the data. By characterizing the initial data, we may not know understand the deeper theory but we have the informational available, at least in a graphical form or as a lookup table. Such characterization of categorized data will allow the differentiation from other sets of data. For example, by charting the frequency of slopes in an environmental terrain data set, we can tell if one region is on average steeper than another region.

Stage 3 transforms the information to real knowledge via scientific modeling. By modeling we can bring a much deeper understanding to the information at hand than someone who only has lookup tables. To provide one benefit, we can extrapolate to regions outside the scope of our data ranges. Then, we can not only say how much steeper one area is than another but we can explain the reason and thus have additional knowledge to infer from.[2]

We will treat these models in as concise a formulation as possible to avoid undue complexity, and we apply physical first principles to justify their description

## Context Modeling of Environmental Domains

Let's start by applying maximum entropy principles and superstatistics to several practical examples: (1) wind speed, (2) terrain slopes, (3) rainfall intensity, (4) wave crest heights, and (5) electro-magnetic signal clutter. These all share a common foundation, as each draws from energy from the environment to give a range in intensity of some measure. In each case, the intensity is a readily measured observable, with the general trend of higher intensity values occurring less frequently than the low intensity values. The zero energy state is indeed the most common in all these measures.

An initial set of questions to consider from a context modeling perspective is what general characterization can we make from these observables, and then what universal modeling approaches can we apply?

Drawing from the perspective of maximum entropy, we first consider that the probability of a high energy configuration is typically treated as a variation of a Boltzmann factor:

$$e^{-E/E_A}$$

where $E_A$ is the activation energy for occurrence. In this approach, the likelihood of higher energies become exponentially damped, scaled by the activation energy constant. The connection to maximum

---

[2]  The path we want to follow is citizen-based science and citizen participation, which is one of the 'democratization' angles that we may want to stress. This has turned into one of the best avenues to pick up real knowledge on our environment, as the amount of data and information available on the internet is widespread.

entropy is that we can treat this factor as a probability and then apply it as a probability density function of the measure of interest:

$$p(E)dE = \frac{1}{E_A} \; e^{-E/E_A} \, dE$$

The choice of the exponential in terms of MaxEnt is that it is the least biased estimator considering that $E_A$ is the **average** energy of the configuration. If we happened to know the variance of the ensemble configuration, this would lead to a normal, Gaussian PDF. Yet, since we in general lack this knowledge, we need to rely on the least amount of information available, and this is the exponential, or Boltzmann factor.

Next we consider the application of the activation energy for the individual cases. For wind distributions, the kinetic energy is related to the square of the wind speed, $v$. This turns into the Rayleigh distribution.

$$E \sim kv^2$$
$$dE = 2 \, k \, dv$$
$$p(v) \sim 2kv \, e^{-kv^2}$$

Electromagnetic signal clutter follows a similar derivation as the energy is the square of the amplitude of the electric signal.

For terrain distributions, we can to first order suggest that a potential energy is directly proportional to the terrain slope.
$$p(s) \sim e^{-ks}$$

For rainfall intensity, we consider the potential energy associated with a volume of water under gravitational forces. The larger the volume, the greater the encapsulated energy, which gets released scaled to the rate intensity of the rainfall.

Finally, for aquatic wave crests, the energy of the waves is proportional to square of the crest height. This works only to some level, as shoaling and non-linear fluid mechanics can prevent or attenuate taller waves.

Now consider that in each of these cases, the value of the activation energy can vary depending on regional or environmental conditions. For the case of rainfall, the intensity of the rainfall can be predicated on other conditions besides the volume of the water vapor alone. This leads to the idea of a super-statistical distribution. Here we not only apply the exponential PDF to the measures of interest, but we grant that probability an extra layer of uncertainty. That uncertainty would commonly apply to the value of $E_A$ or to some other constant of proportionality.

Depending on whether we look at an activation energy proportionality or some other variant measure, we can come up with different kinds of statistical distributions. For example, if we consider variations in capacity and growth time, we can model such phenomena as cloud sizes, lake sizes, and particulate sizes. These become fat-tail distributions due to the weighting of the rate calculation, as a strong variant situated in the denominator of a stochastic ratio turns into a heavy weighting in the tail of a distribution.

These become exemplars of the general superstatistical approach we will apply to natural context domains. The table below describes several of the practical examples. For many of these natural phenomena, we can use empirical data sets and draw samples from this observational space, or we can

work at a more fundamental level and try to infer information from a model that may more accurately represent the underlying behavior. The latter has a number of appealing qualities, including conciseness of representation and potentially better estimation of rare events.

*Table 2: List of stochastic models*

| Stochastic Metric | Elements | Description | Data |
|---|---|---|---|
| Wind | PDF, ME, SS | Model of wind speeds | Bonneville Power Authority[12] |
| Rainfall | PDF, ME, SS | Model of rainfall amount | Hydrometeorology Laboratory University of Iowa[13] |
| Clutter | PDF, ME, SS | Model of EMI | **(maybe Rockwell Collins or a JPL site)** |
| Clouds | PDF, ME | Model of cloud sizes | NASA Goddard[14] |
| Lakes | PDF, ME | Model of lake sizes | Global Lakes and Wetlands Database[15] |
| Particles | PDF, ME | Model of particle sizes | NASA JPL[16] |
| Waves | PDF, ME | Model of crest heights | CDIP[17] and US Army Corps of Engineers[18] |
| Terrain slopes | PDF, ME, SS | Model of inclination | USGS DEM [19] |

## Examples of predictable unpredictability

Examples of almost predictable unpredictability can arise in many natural environmental contexts, such as variability in terrain slopes, as well as in artificially man-made situations, such as a large highly-interconnected network. The key is to apply these models in either analytic form or in a form suitable for Monte Carlo-type simulations.

 *Slope modeling and Topography*

First is an example of the use of a longitudinal slope context distribution. The macroscopic slope model of USA terrain follows closely that described by the Maximum Entropy Principle (MaxEnt).

How do we model and thus characterize disorder in the earth's terrain? Can we actually understand the extreme variability we see? If we consider that immense forces cause upheaval in the crust then we can reason that the energy can also vary slope topography widely.

The process that transfers potential energy into kinetic energy to first order has to contain elements of randomness. To the huge internal forces within the earth, generating relief textures equates to a kind of Brownian motion in relative terms — over geological time, the terrain amounts to nothing more than inconsequential particles to the earth's powerful internal engine.
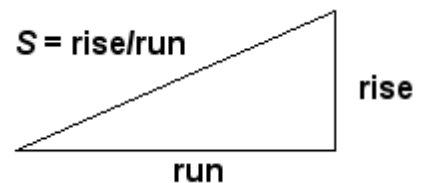

$S = rise/run$

We take the terrain slope *S* as our random variable (defined as rise/run).

The initial premise is the higher the slope, the more energetic the terrain. Applying the Maximum Entropy Principle to a section of terrain, we can approximate the local variations as a MaxEnt conditional probability density function, where E is the local mean energy and c is a constant of proportionality. But we also assume that the mean E varies over a larger area that we are interested in, as in the superstatistical sense of applying a prior distribution, where k is another MaxEnt measure of our uncertainty in the energy spread over a larger area.

The final probability is an integral over the marginal distribution consisting of the conditional multiplied by the prior. This integrates as a modified `BesselK` function of the zero order, $K_0$.

$$p(S) = \frac{2}{S_0} \cdot K_0(2\sqrt{S/S_0})$$

The average value of the terrain slope for this distribution is simply the value $S_0$.

Now we can try it on a large set of data. The digital elevation model (DEM) data for the 1 degree quadrangles (aka blocks/tiles) in the USA from the USGS web site was characterized. This consists of post data at approximately 92 meter intervals (i.e. a fixed value of run) at 1:250,000 scale for the lower 48 USA and some spillover into Canada. From individual DEM files, we calculate the slopes between adjacent posts and came up with an average slope (rise/run) of 0.039, approximately a 4% grade or 2.2 degrees pitch. The characterization takes the absolute values of all slopes so that the average is not zero.

The cumulative plot of terrain slopes for all 5 billion calculated slope points appears on the following chart (see ref). The cumulative probability distribution of the `BesselK` model is plotted with the calculated average slope as the single adjustable parameter.
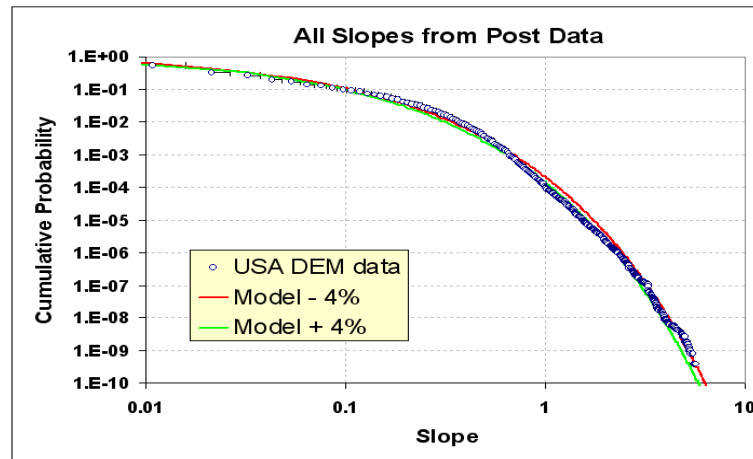


*Figure 3: Longitudinal slope cumulative distribution function for the USA. The BesselK model with a small variation in S0 (+/-4% about the average 0.037 rise/run) demonstrates sensitivity to the fit.*

The good agreement occurs because random forces contribute to maximizing the entropy of the topography. Enough variability exists for the terrain to reach an ergodic limit in filling the energy-constrained state space. As supporting evidence, it turns out that we can generate a distribution that maps well to the prior by estimating the average slope from the conditional PDF of each of the 922 quadrangle blocks and then plotting this aggregate data set as another histogram.

For context modeling, a library function can generate Monte Carlo sample draws for the BesselK model without requiring a probability inversion. The algorithm turns out surprisingly simple. First, draw two independent random samples from a uniform [0.0 .. 1.0] interval, then apply the natural log to each, multiply them together, and then multiply by the BesselK $S_0$ scaling constant.

This random draw algorithm will give the following cumulative if done 5 billion times, which is the same sample size as the real USA DEM data sample (see Figure 4). The only statistical noise is at the $10^{-9}$ level, which is roughly the same as in the DEM data set.
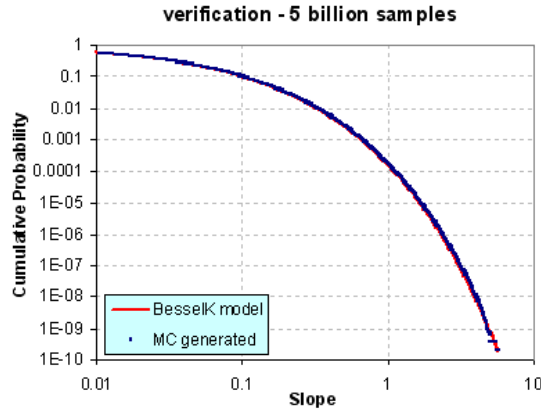


*Figure 4: Drawing random samples to approximate slope distribution*

Practically speaking, we see the variability in slopes expressed at the two different levels: the entire USA at the integrated (BesselK model) level and the aggregated regions at the localized (exponential prior) level. These remain consistent as they agree on the single adjustable parameter $S_0$.
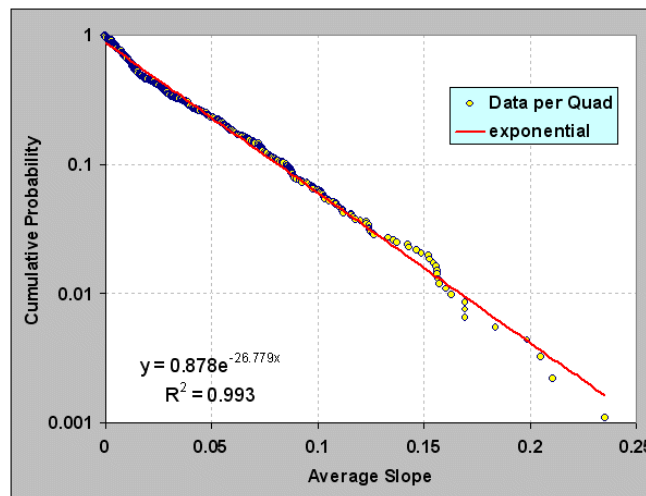


*Figure 5: Generation of the prior distribution by taking the average slope of each of nearly 1000 quadrangles. The best fit generates a value of $S_0$ (1/ 27=0.037) close to that used in the previous figure.*

The modeled distribution has many practical uses for analysis, including mobility studies and fuel efficiency planning. Obviously, vehicles traveling up slopes use a significant amount of energy and a context modeler would need a model to base an analysis on without having to rely on the raw data by itself. Further, spatial correlations also exist and will prove useful as well. A survey of recent research gives no indication that others have discovered this rather simple model (See further Vico[20], Gagnon[21], Gonçalves[22], Guarnieri[23])

### *Wind Distribution*

Wind velocities demonstrate a wide dynamic variability, ranging from calm to gale force. With minimal knowledge as to the origin of the wind variance, we can apply the maximum entropy principle to its energy content. If we assume only a temporally-averaged mean wind energy and then relate the energy, **$E$**, as the square of the wind speed, **$v^2$**, the resultant maximum entropy probability distribution matches the Rayleigh distribution.

$$p(v) = p(E) \cdot \frac{dE}{dv} = 2cv \cdot e^{-cv^2}$$

However intuitive the concept of "windiness", we may often miss the underlying mathematical simplicity behind wind speed variability. The complexity of the earth's climate and environment actually contributes to this simplicity as it generates more states for the system to exist within, which increase the likelihood of variability.

The derivation of wind dispersion follows a few easy steps. We start with the premise that every location on Earth has a mean or average wind speed. This speed has a prevailing direction but assume that it can blow in any direction. Next we safely assume that the kinetic energy contained in the aggregate speed goes as the square of its velocity.

This comes about from the Newtonian kinetic energy law ½$mv^2$ and it shows up empirically as the aeronautical drag law (i.e. wind resistance) which also goes as the square of the speed. (Note that we can consider $E$ as an energy or modified slightly as a power, since the energy is sustained over time) Then apply the principle of maximum entropy to the possible states of energy that exist and come up with a probability density function that has no constraints other than a mean value (with negative speeds forbidden). In the equation above $c$ is a constant and $1/c$ defines the mean energy (i.e. essentially acting as the Boltzmann activation energy). This describes a declining probability profile, with low energies much more probable than high energies. To convert to a wind dispersion PDF we substitute velocity for energy and simplify.

Figure 6 shows an empirically observed wind speed distribution, showing a peak away from zero wind speeds and a rapid decline of frequency at higher velocity. Heuristically, many scientists refer to it as a Rayleigh or Weibull distribution. The Rayleigh comes out as the simpler model because it derives from first principles and any deviation from the quadratic exponent works as a refinement. The first data set shown consisted of about 36,000 sequential hourly measurements in terms of energy (kilowatt-hours) for Ontario.
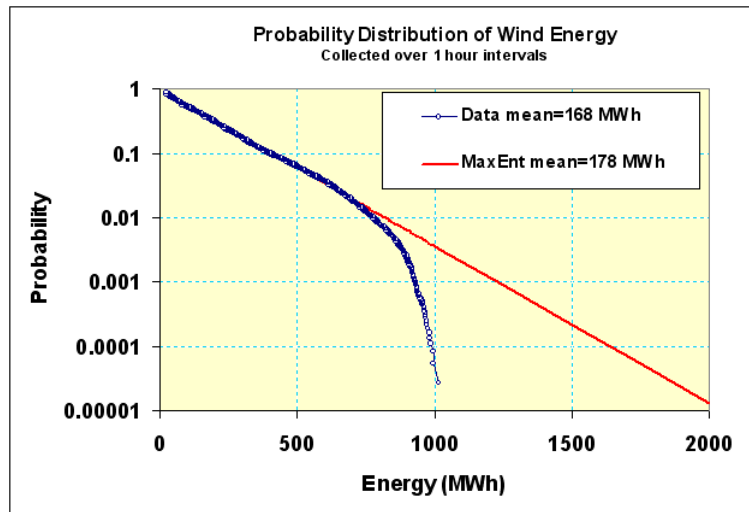
*Figure 6: Wind distribution follows Rayleigh distribution closely. This data was collected from an archive of wind energy statistics collected in Ontario. The upper level cut-out is due to turbine governor regulation at high wind speeds.*

By adding more data to our knowledge on wind dispersion, we can observe how dispersion in wind speeds has a universal character. The second data set (Figure 7) comes from northwest Germany and consists of wind power collected at 15 minute intervals.
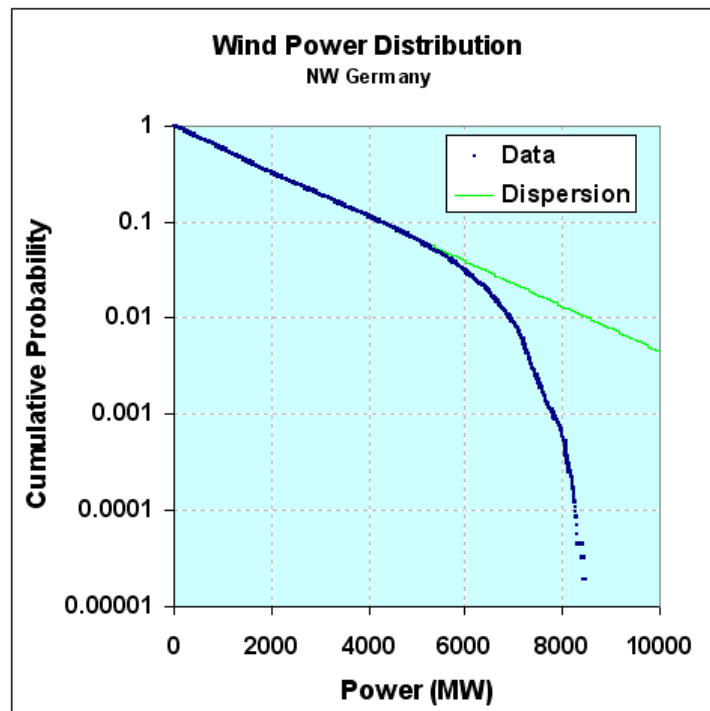


*Figure 7: Wind variations for Germany. The curve has all the same characteristics as that for Ontario, demonstrating the universal behavior in wind speed variability.*

Note that the same wind dispersion holds as for Ontario. Both curves display the same damped exponential probability distribution function for frequency of wind power (derived from wind speed). We also see the same qualitative cut-out above a certain power or wind energy level.

But we know that the $E_A$ can vary from region to region, so we leave it as a conditional, and then set that as a maximal entropy estimator as well

$$p(E_i) = \alpha \cdot e^{-\alpha E_i}$$

then we integrate over the conditional's range according to standard practice and arrive at a cumulative.

$$P(E) = \int_0^\infty P(E|E_i)p(E_i)dE_i$$

This results in a simple lookup in your favorite comprehensive table of cataloged integration formulas, which leads to the following solution:

$$P(E) = 2\sqrt{\frac{E}{\overline{E}}}\, K_1\left(2\sqrt{\frac{E}{\overline{E}}}\right)$$

where $K_1$ is the modified BesselK function of the second kind, in this case of order 1, which is found in any spreadsheet program (such as Excel). Note that this is the same function that we used for the distribution of terrain slopes. The order 1 is the variant used for the *cumulative* distribution function

We tested this formulation against wind data from Bonneville Power Administration,which has over 20 meteorological stations set up around northern Oregon. The download consisted of over 2.5 million data points collected at 5 minute intervals, archived over the span of a little less than 2 years.
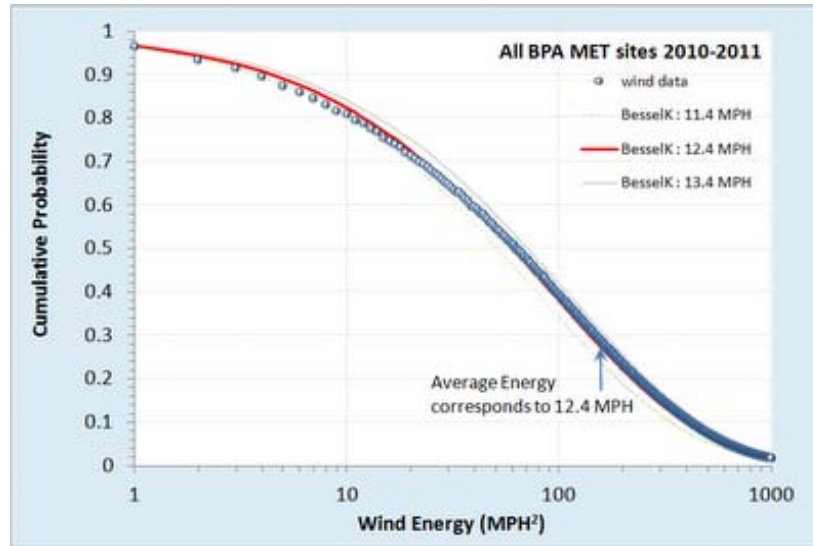


*Figure 8: Cumulative distribution function of wind energies from Bonneville with model fit.*

For the fit of this curve, the average energy was derived directly by computing the mean over the entire set of data separately. This corresponded to a value of 12.4 MPH, and placed a pair of positive and negative tolerances to give an idea of the sensitivity of the fit.

As this is a single parameter model, the only leeway we have is in shifting the curve horizontally along the energy axis, and since this is locked by an average, the fit becomes essentially automatic with no room for tweaking and little for argument. The probabilities are automatically normalized.

Figure 9 shows the log-log plot, which reveals a departure at high wind speeds. This shows that excessive gale force winds (greater than 60 MPH) did not occur over the extended region during the span of two years data collection.

Wind dispersion analysis has obvious applications for context modeling. Fuel efficiency is impacted by aerodynamics and drag goes up as the square of the wind speed. Vehicle cooling also is impacted by convection due to local winds.. Applying this approach for context modeling has the benefit of allowing simple sampled data generation for verification and PCC bounding, similar to that applied for terrain slopes.
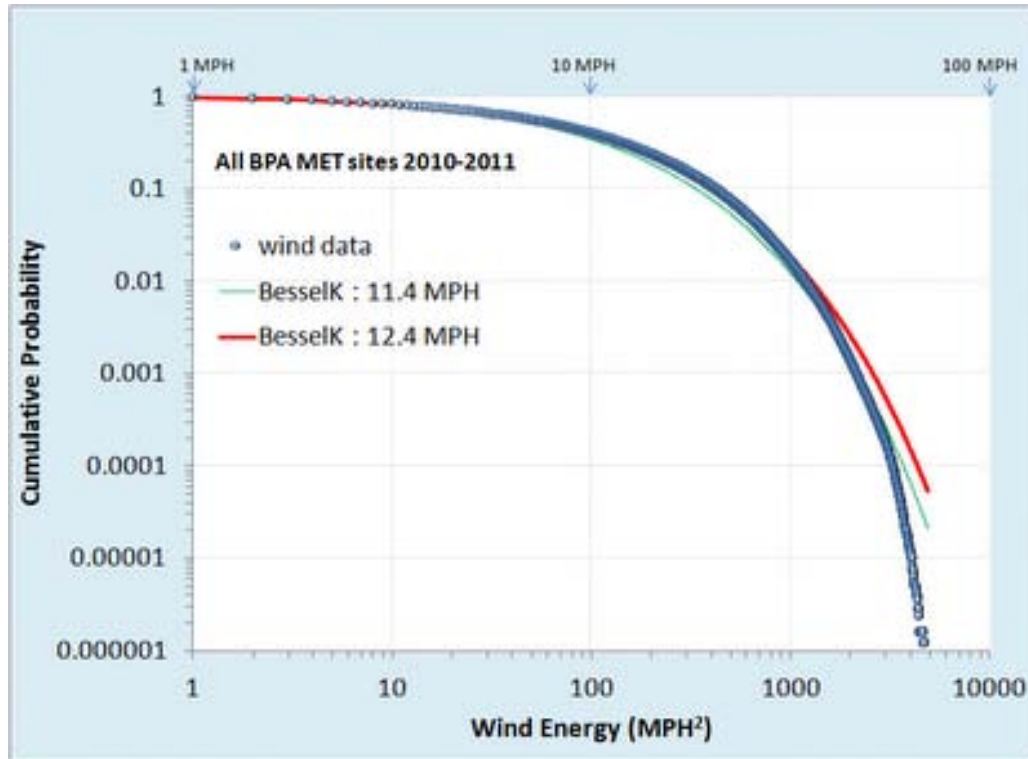


*Figure 9: Cumulative distribution function of wind energies on a log-log plot.*

The previous maximum entropy derivation assumed only a known mean of wind energy levels (measured as power integrated over a fixed time period). From this simple formulation, one can get a wind speed probability graph. Knowing the probability of wind speed, one can also perform all kinds of interesting extrapolations — for example, we can project the likelihood of how long it would take to accumulate a certain level of energy (see reference to energy modeling in Afghanistan).

*Aquatic Waves*

(this section needs a few more derivations for the data) Ocean waves exist in as disordered state as the wind. We may not notice this as the scale of waves is smaller and more focused. In practice, the wind

energy distribution relates to an open water wave energy distribution via similar maximum entropy disorder considerations. The following derivation assumes a deep enough water such that the wave troughs do not touch bottom

First, 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 related to the wave height by the square of the height, **H**. 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 Figure 10 below.
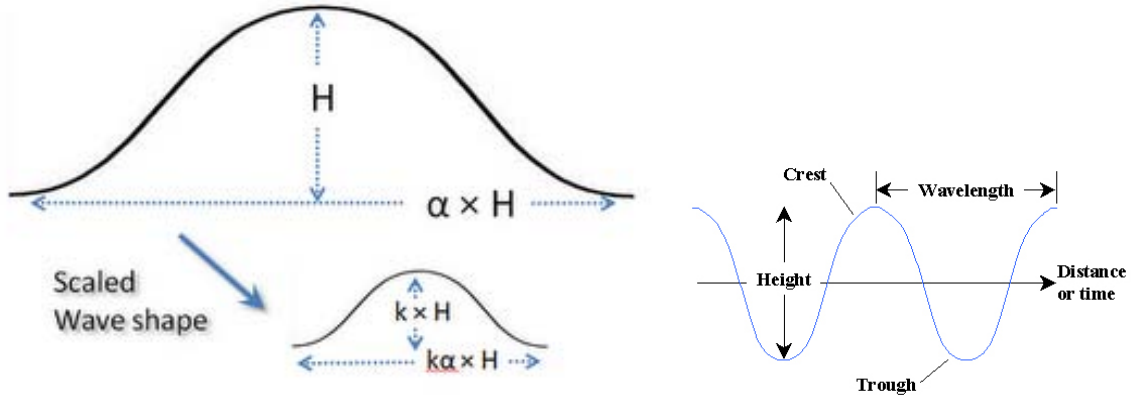


*Figure 10: Total energy in a directed wave goes as the square of the height, and the macroscopic fluid properties suggest that it scales to size. This leads to a dispersive form for the wave size distribution*

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

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

where $\alpha$ 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. This looks at height alone.
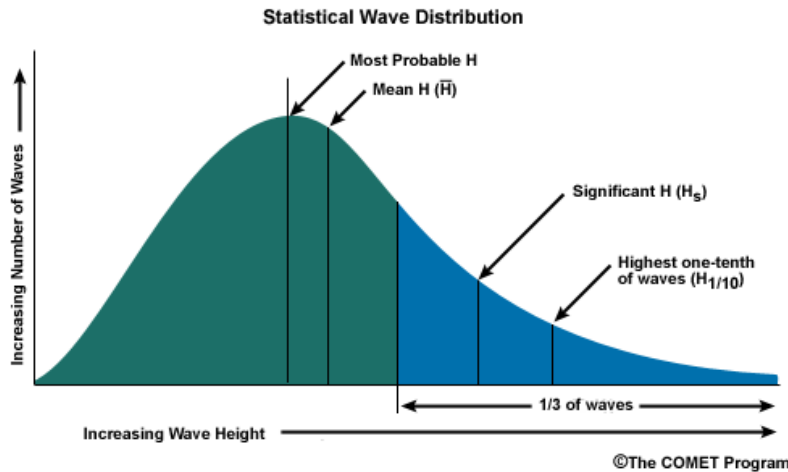


*Figure 11: Statistical Wave Distribution*

Long-lived ocean and lake measuring stations have recorded historical records of wave crest data over the span of decades. From the US Army Corps of Engineer's Wave Information Studies project[24], the following figure collects chop and swell data from over several hundred million data points on Lake Michigan:
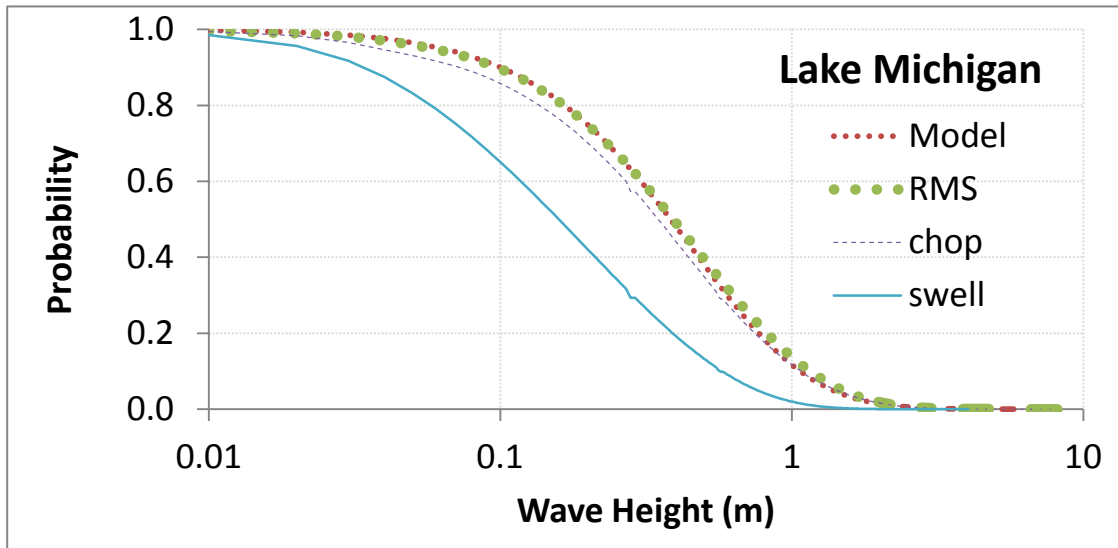


*Figure 12: Lake Michigan log-linear*

The model again assumes maximum disorder in a wave height parameter in the BesselK distribution. The full log-log plot is shown below.
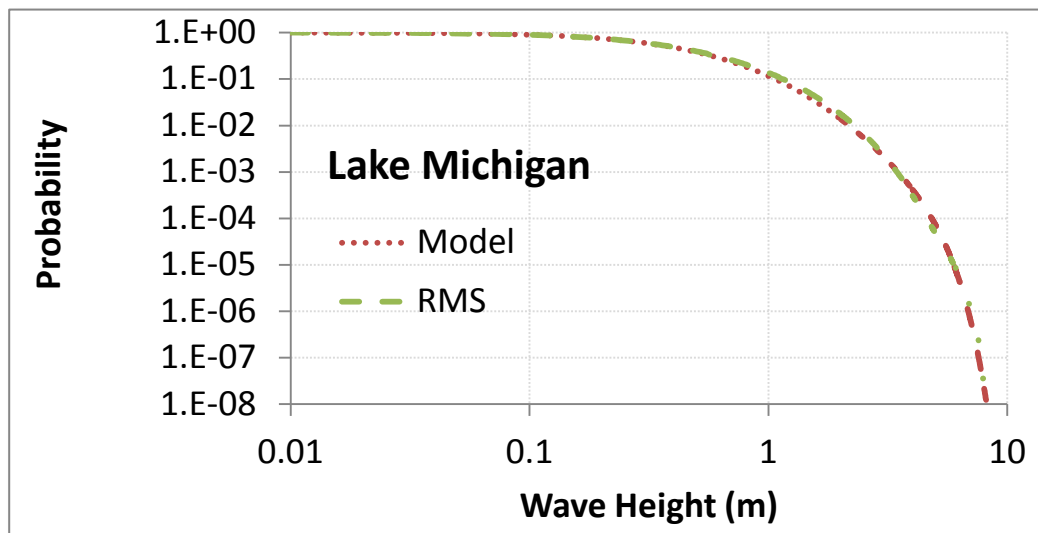


*Figure 13: Lake Michigan log-log*

The difference between the maximum entropy estimator for wind than for waves is due to nonlinear effects at large wave-height. A correction factor caused by cusping generates larger heights than the triangular base wave predicts. The correction from a variation the Rayleigh-Stokes (Tayfun[25] model[26])

P(x)= exp(-(s*x+1-sqrt(2*s*x+1))) (narrow-banded random process)[27] And probability of consecutive waves.[3]

Lake Superior measurements (Figure 14and Figure 15) were more sparse yet the same profile is observed.
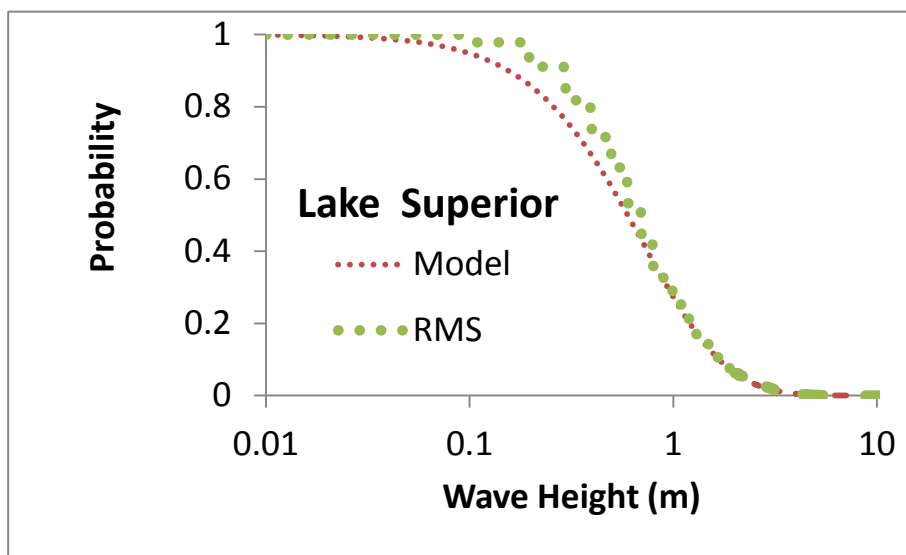


*Figure 14: Lake Superior log-linear (limited data set)*
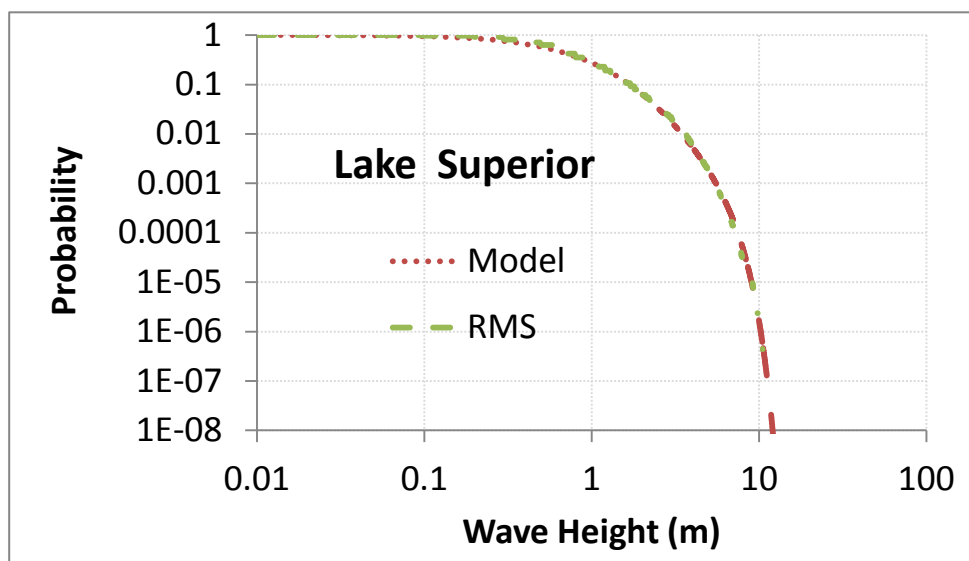


*Figure 15: Lake Superior log-log*

---

[3]  Sharpened crests and flattened troughs:  The crests are steeper and troughs are flatter; Phase velocity increases with the increase in wave steepness.   Non-closed trajectories of particles movement.  Nonlinear wave characteristics (up to 2nd order).3 (chapter 16.2)

It can be shown that the probability distribution of normalized nonlinear crests $\zeta_n$  is of the form

$$P(\zeta_n \leq s) = 1 - \exp\left( -\frac{\left\{-1+\left[1+4\delta(s-\gamma)\right]^{1/2}\right\}^2}{8R\delta^2} \right)$$

Wave measurements from the Atlantic Coast along the length of the USA were less widely dispersed. It is much more unlikely to find very calm waters in the data set. In this case, a higher order BesselK function was used to model the wave height distribution.
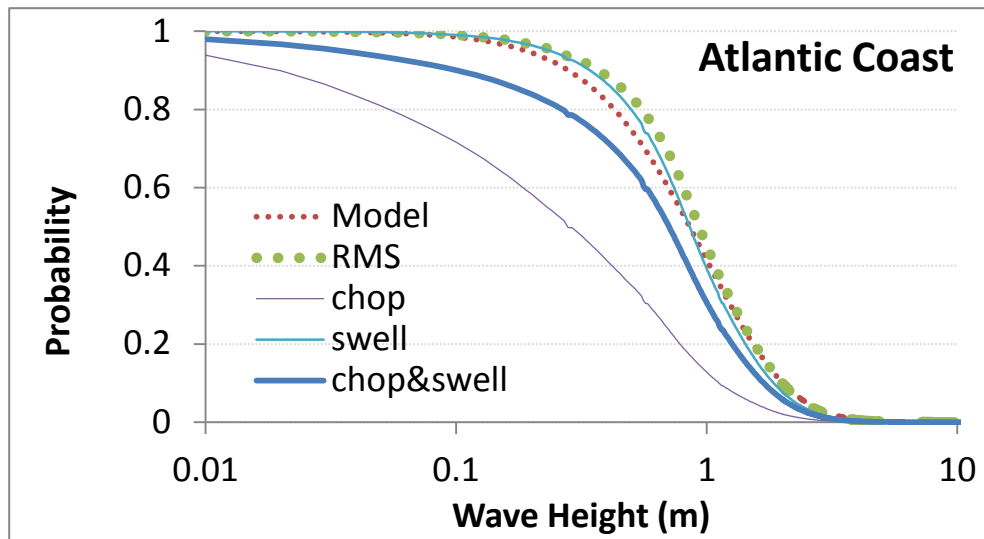


*Figure 16: Atlantic Coast log-linear*
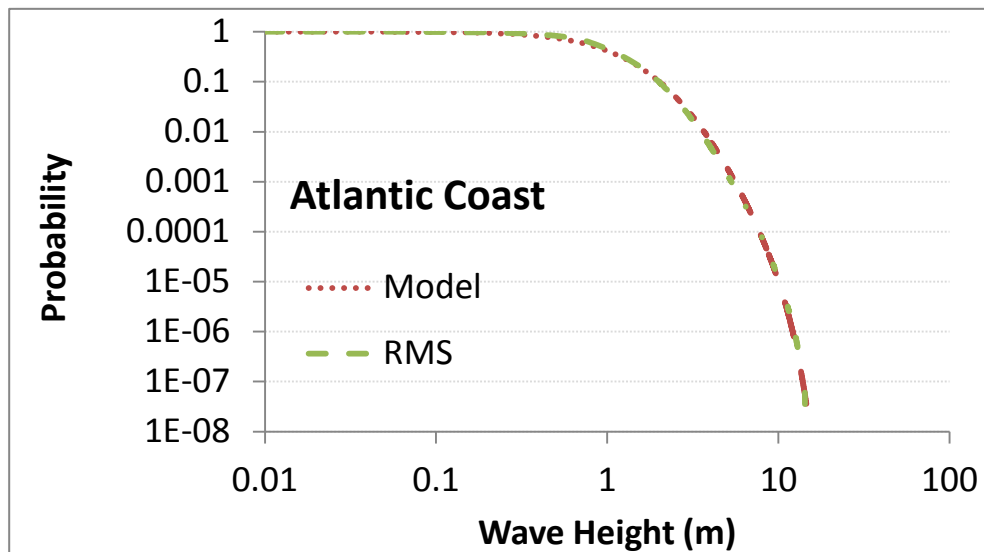


*Figure 17: Atlantic Coast log-log*

*Rainfall Intensity*

As a premise, we want to consider whether a simple stochastic model can generate statistical patterns of rainfall events. We postulate that a critical point exists for rain to fall. The volume and density at which nature decides it reaches this critical point has much to do with the rate at which a cloud develops in its intensity and payload.

We assume a maximum entropy probability distribution function which assumes an average energy of rainfall rate for a storm i:

$$P(E|E_i) = e^{-E/E_i}$$

The rationale for this is that the rainfall's energy is proportional to the rate of the rainfall, since that amount of moisture had to be held aloft by gravity.

$$Rate_i \sim E_i$$

Yet we know that the $E_i$ can vary from storm to storm, so we leave it as a conditional, and then set that as a maximal entropy estimator as well

$$P(E_i) = \alpha\, e^{-\alpha E_i}$$

then we integrate over the conditional's range.

$$P(E) = \int_0^\infty P(E|E_i) p(E_i) dE_i$$

This results again in an integration formula lookup, which leads to the following solution:

$$P(E) = 2 \sqrt{\frac{E}{\bar{E}}}\, K_1\left( 2 \sqrt{\frac{E}{\bar{E}}} \right)$$

where $K_1$ is the modified BesselK function of the second kind, in this case of order 1. This is the same general derivation as we performed for wind speed.

So let us see how it works. The analysis was compared against this recent paper by Papelexio: "Can a simple stochastic model generate rich patterns of rainfall events?"[28] and graphed as shown below in Figure 18. The green points constitute the BesselK fit which lies right on top of the blue empirical data set.
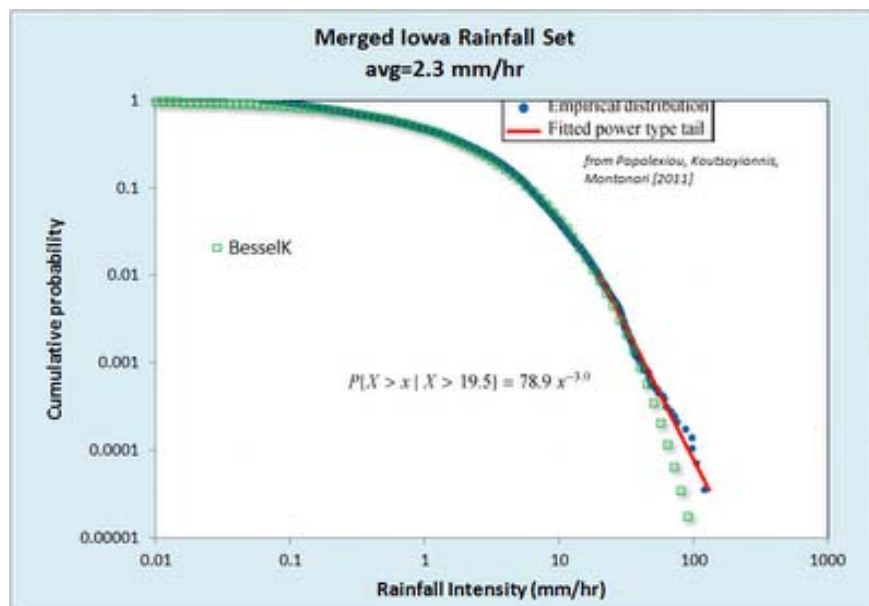
*Figure 18: Cumulative rainfall distribution statistically gathered from several storm events.*

From the table below reproduced from the paper, we can see that the mean value used in the BesselK distribution was *exactly the same* as that from the standard statistical moment calculation

*Table 3: Rainfall moments*

| Event# | 1 | 2 | 3 | 4 | 5 | 6 | 7 | All |
|---|---|---|---|---|---|---|---|---|
| SampleSize | 9697 | 4379 | 4211 | 3539 | 3345 | 3331 | 1034 | 29536 |
| Mean(mm/h) | 3.89 | 0.5 | 0.38 | 1.14 | 3.03 | 2.74 | 2.7 | **2.29** |
| StandardDeviation(mm/h) | 6.16 | 0.97 | 0.55 | 1.19 | 3.39 | 2.2 | 2 | 4.11 |
| Skewness | 4.84 | 9.23 | 5.01 | 2.07 | 3.95 | 1.47 | 0.52 | 6.54 |
| Kurtosis | 47.12 | 110.24 | 37.38 | 5.52 | 27.34 | 2.91 | -0.59 | 91 |
| HurstExponent | 0.94 | 0.79 | 0.89 | 0.94 | 0.89 | 0.87 | 0.97 | 0.89 |

In contrast, Papalexio try to apply Hurst-Kolmogorov statistics to the problem in search of a power law solution. They claim that the power law tail of -3 is somehow significant. Instead we suggest that the slight deviation in the tail region is likely caused by insufficient sampling in the data set. A slight divergence starts to occur at the 1 part in 5,000 level and there are only 30,000 points in the merged data set, indicating that statistical fluctuations could account for the difference. See Figure 19 below which synthesizes a BesselK distribution from the same sample size, and can clearly duplicate the deviation in the tail.
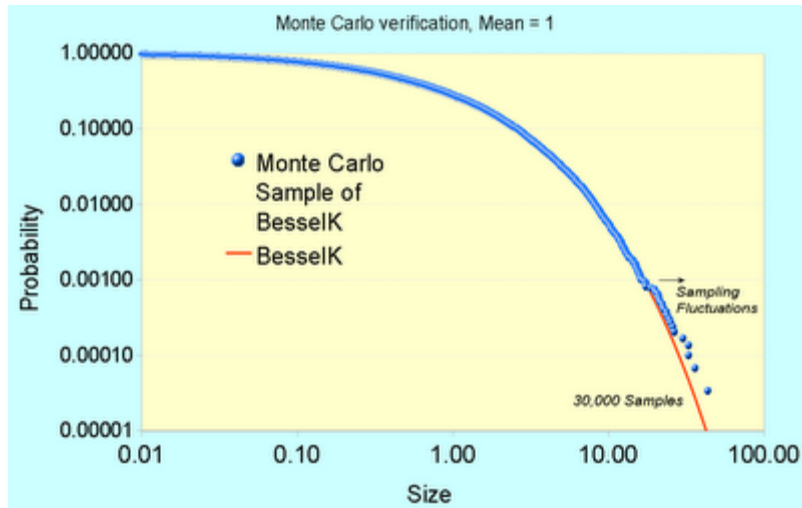
*Figure 19: Monte Carlo sampling from a synthetic BesselK distribution reveals the same departure at low probability events.*

Fluctuations of 1 in 30,000 are clearly seen at this level, so care must be taken to not read fat-tail behavior that may not exist.

The bottom-line is that this simple stochastic model outdoes Papalexio's simple stochastic model, which doesn't seem as simple anymore once we apply an elementary uncertainty analysis to the data. The only physical premise needed is based on observations and intuition that rainfall buildup is a potential energy argument. This translates to a Boltzmann/Gibbs probability with a large spread in the activation energy. The large spread is modeled by maximum entropy -- we don't know what the activation energy is so we assume a mean and let the fluctuations about that mean vary to the maximum amount possible. That is the maximum entropy activation which is proportional to a mean rainfall rate -- the stronger the rate, the higher the energy.

That essentially becomes the premise for predicting the probability of a given rainfall rate within the larger ensemble defined by a mean. It seems as if Papalexio[29] was on a similar track but decided to stop short of making the thesis quite this simple. They seem to imply that the PDFs can go beyond the BesselK and into the MiejerG territory, and beyond that in variability. That is certainly acceptable as it will simply make it into a fatter tail distribution, and if a fatter tail matches the empirical observations on a grander ensemble scale, then that might be a better characterization of the natural process.

Cf. Koutsoyiannis paper
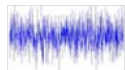
### *EMI Clutter*

The intermittent nature of wind power has a fundamental explanation based on entropy arguments. It turns out that this same entropy-based approach explains some other related noisy and intermittent phenomena that modelers have to deal with. The obvious cases involve the use of mobile wireless gadgets such as WiFi devices, cell phones, and global positioning system (GPS) navigation aids in an imperfect (i.e. noisy) environment.

First of all, consider that we often use these wireless devices in cluttered environments where the supposedly constant transmitted power results in frustrating fadeouts that we have all learned to live with. An example of Rayleigh fading appears in Figure 20 below. You can find some signal interference-based explanations for why this happens, originating via the same intentional phase cancellations that occur in noise-cancelling headphones. For the headphones, the electronics flip the phase so all interferences turn destructive, but for wireless devices, the interferences turn random, some positive and some negative, so the result gives the random signal shown. In the limit of a highly interfering environment the amplitude distribution of the signal shows a Rayleigh distribution, the same observed for wind speed.
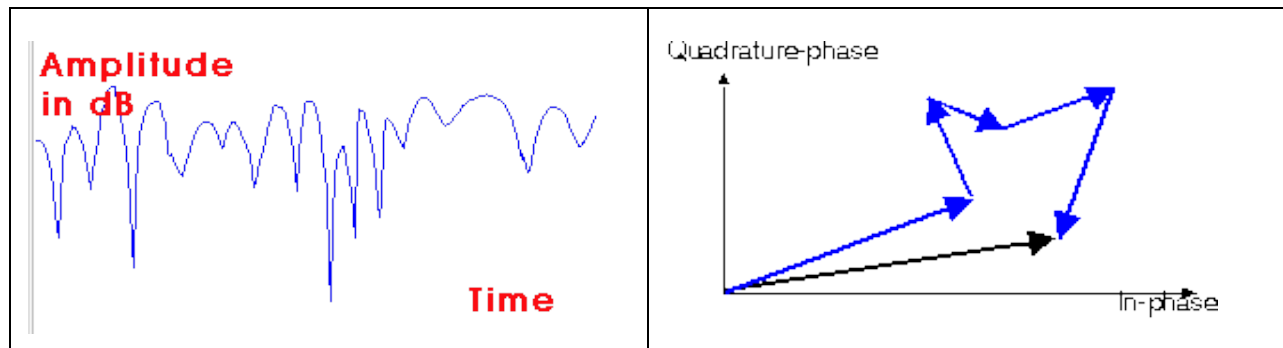


*Figure 20: Noisy variation of signal amplitudes (left) is typically explained as a duadrature summation of randomly directed vector amplitudes (right). The variation in noise amplitude levels with time and the effect this has on a signal is known as Rayleigh fading.*

Our knowledge of the situation reduces to that of knowing only the average power level of the signal. In that case, we can use Maximum Entropy Principles to estimate the amplitude from the energy stored in the signal, just like one can derive it for wind speed. So, as a starting premise, if we know the average power alone, then we can derive the Rayleigh distribution. The following figure shows the probability density function of the correlated power measured from a GPS signal. Since power in an electromagnetic signal relates to energy as a flow of constant energy per unit time, then we would expect the energy or power distribution to look like a damped exponential, in line with the maximum entropy interpretation. And sure enough, it does exactly match a damped exponential.

$$p(E) = k \cdot e^{-kE}$$

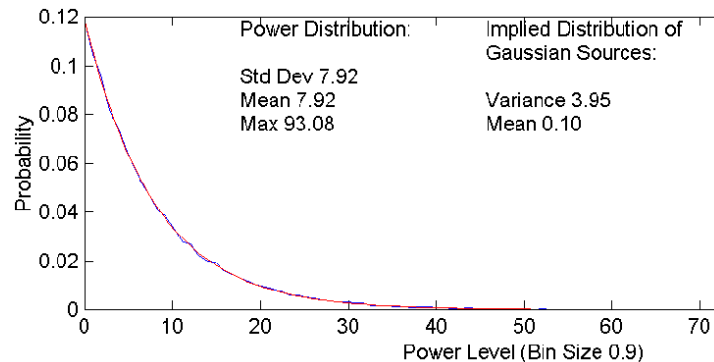This matches the observation of noise power level as shown in Figure 21.



*Figure 21: Note that the standard deviation equals the mean, a clear indication of a damped exponential. (from Watson [30])*

Yet since power (*E*) is proportional to Amplitude squared, we can derive the probability density function by invoking the chain rule.

$$p(A) = p(E) \cdot \frac{dE}{dA} = e^{-kA^2} \frac{d}{dA}(A^2) = 2kr \cdot e^{-kA^2}$$

This precisely matches the Rayleigh distribution, implying that Rayleigh fits a Maximum Entropy (MaxEnt) distribution. So too does the uniformly random phase in the destructive interference process qualify as a MaxEnt distribution, which will range from 0 to 360 degrees (which gives an alternative derivation of Rayleigh). So all three of these distributions, Exponential, Rayleigh, and Uniform all act together to give a rather parsimonious application of the maximum entropy principle.
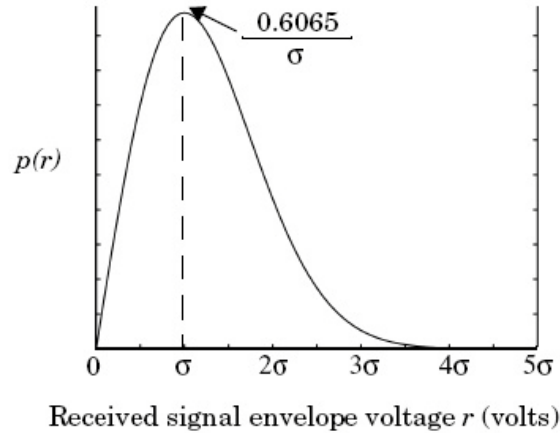


*Figure 22: The Rayleigh Distribution of signal strength.*

The most interesting implication of an entropic signal strength environment relates to how we deal with this power variation in our electronic devices. In a GPS, a vehicle's navigation system will experience this when trying to acquire a GPS signal from a cold-start. The amount of time it takes to acquire GPS satellites can range from seconds to minutes, and sometimes we don't get a signal at all, especially if we have tree cover with branches swaying in the wind.

Explaining the variable delay in GPS comes out quite cleanly as a fat-tail statistic if we consider how the GPS locks into the set of satellite signals. The solution assumes the entropy variations of the signal strength and integrating this against the search space that the receiver needs to lock-in to the GPS satellites. Since the search space involves time on one axis and frequency in the other, it takes in the limit $\sim N^2$ steps to decode a solution that identifies a particular satellite signal sequence for a particular unknown starting position. This gets reduced because of the mean number of steps needed on average in the search space. We can use some dynamic programming matrix methods and parallel processing (perhaps using an FFT) to get this to order $N$, so the speed-up for a given rate is $t^2$. So this will take a stochastic amount of time according to the MaxEnt criteria.

However because of the Rayleigh fading problem we don't know how long it will take to integrate our signal with regard to the rate $R$. This rate has a density function proportional to the power level distribution, then according to the rules for marginal distributions the conditionals line up to give the probability of acquiring a signal within time $t$.

$$P(t < T) = \frac{1}{1 + \left(\frac{T}{a}\right)^2}$$

This leads to the dispersion result where $a$ is an empirically determined number derived from $k$ and $c$. This is not considered an extremely fat tail because the acceleration of the search by quadrature tends to mitigate very long times.

**Data Analysis.** Data was collected from a GPS project that has a goal to speed up wild-fire response times by cleverly using remote transponders[31]. They published data for cold-start times as shown in the histogram below. Note that the data shows many times that approach 1000 seconds. The single parameter entropic dispersion fit (a=62 seconds) appears as the blue curve, and it fits the data quite well.
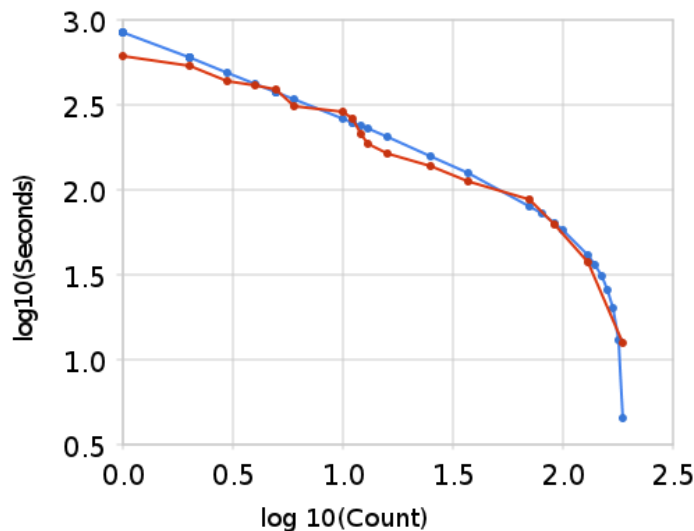


*Figure 23: Fit of GPS acquisition times to a dispersive model*

Beyond the Rayleigh, we will also find the BesselK PDF, as that is what describes the K-distributed clutter which appears in large regions collected by radar stations against random terrestrial and oceanographic features. The heterogeneous nature of that environment will generate clutter with longer tails than the Rayleigh. See Grenander[32].  (I don't have a figure of this yet).

 *Clouds*

If we assume that clouds develop by some sort of preferential attachment, then the uncertainties at which the preferential attachment process increases with time balanced against the uncertainties in the critical point contribute to the dispersion.  This is an analysis based on work by Yuan at NASA Goddard, "Cloud macroscopic organization: order emerging from randomness"

The cloud size distribution fits the same process as lake size distribution.

$$P(Size) = \frac{1}{1 + \frac{Median}{Size}}$$

We first assume that water vapor disperses through the atmosphere freely. The rate $r$ at which it does this we treat as a stochastic variable with a probability density

$$p(r|g) = \frac{1}{g} \cdot e^{-\frac{r}{g}}$$

This introduces two concepts at once: the idea that we do not assume a single rate (i.e. assume instead *dispersion*) together with the idea that we can only assume at best a mean (as the growth rate $g$) *and* treat the standard deviation as equivalent to the mean. This type of assumption makes the least presuppositions as to what has happened — we know we have a mean value but beyond that, the rate can vary to the maximum entropy limit.

If we next assume that a collection of these rates can act to sweep out a selected uniform unit of volume, then over time we can imagine that a cloud will capture this moisture.
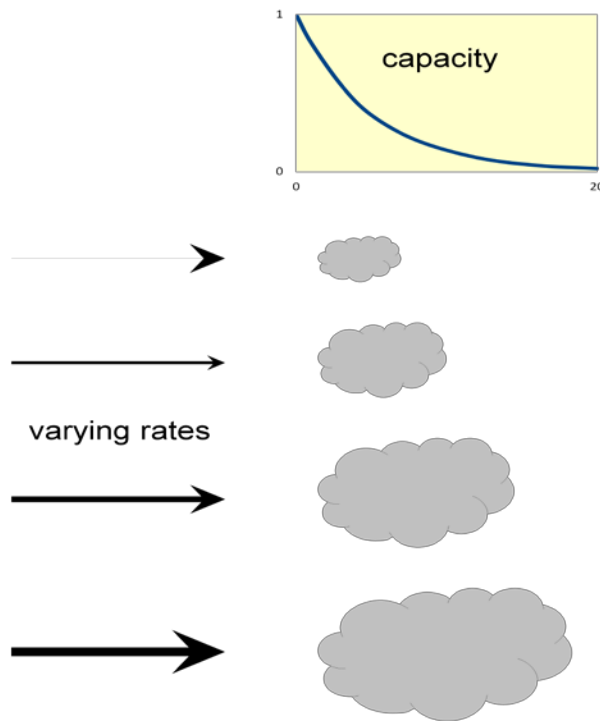


*Figure 24: Cloud formation*

Let's say that the water vapor diffuses outward, so for a given time period $t$, the moisture will diffuse over a distance $x = r\,t$, a simple variable change gives

$$p(r|g) = \frac{1}{gt} \cdot e^{-\frac{x}{gt}}$$

Over time, the probability that some moisture will migrate at least a length $x_0$ distance is:

$$p(x > x_0|g, t) = \int_0^\infty p(x|g, t)\, dx = e^{-\frac{x_0}{gt}}$$

Alternatively, the following relation tells us the cumulative probability of the distance covered by material after time $t$. This again assumes a distance travelled $x = r\,t$.

$$P(x_0|g) = \int_{r=x_0/t}^{\infty} p(r)\, dr = e^{-\frac{x_0}{gt}}$$

This relation also crops up in terms of the *population balance equation*. It basically relates a conservation of particles law, in that we do not lose track of any material due to a flow.

So next we have to accumulate this over a volume or depth at which a cloud develops. The simplest approximation assumes that the water droplets get distributed to a mean height (*H*) with a similar exponential distribution — this is like a capacity for the cloud formation (see Figure 24) :

$$f(x|H) = \frac{1}{H} e^{-\frac{x}{H}}$$

Combining the two relations turns into an *a priori* probability for the expected cumulative transfer after time *t* through the volume. Integrating over the atmospheric layer over which clouds can form, gives the average water volume accumulated:

$$C(t|H) = \int_0^{\infty} f(x|H) \cdot P(x|g)\,dx = \int_0^{\infty} f(x|H) \cdot e^{-\frac{x}{gt}}\, dx$$

$$C(t|H) = \frac{1}{1 + \dfrac{H}{gt}}$$

For the last assumption, we note that if *t* gets evenly spread from the over time, then the value *gt* becomes the effective collected thickness *W* of a distribution of clouds, where we add a factor *k* to indicate collection efficiency. Alternatively, we can interpret the stochastic variable *W* as the maximum net cloud thickness that would develop over a diffusion time *t*. The term *kH* sets the potential maximum net thickness achievable turning it into a hyperbolic discounting probability distribution.

$$C(W|H) = \frac{1}{1 + \dfrac{k\,H}{W}}$$

Or as we first surmised:

$$P(Size|Median) = \frac{1}{1 + \dfrac{Median}{Size}}$$

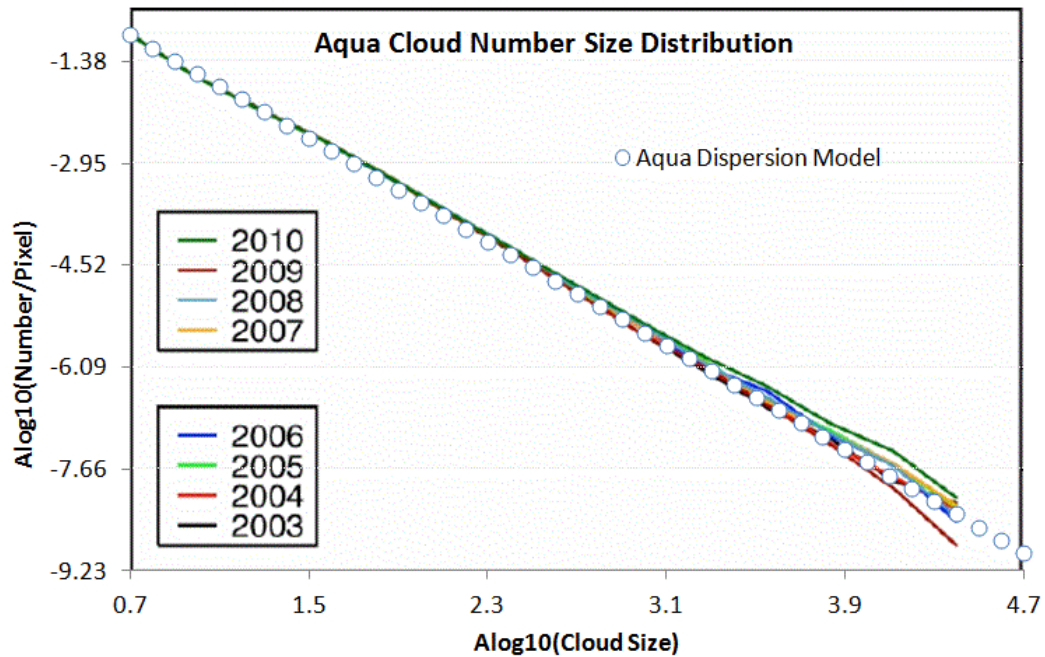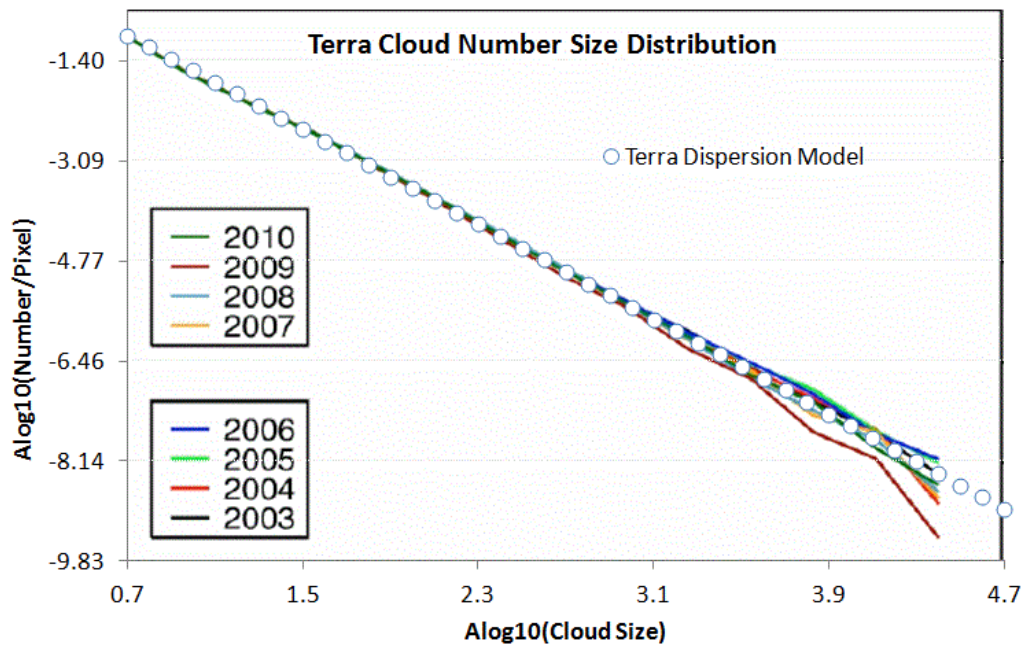The agreement to the collected data is remarkable over several orders of magnitude..

*Figure 25: Cloud statistics over water[33]*



*Figure 26: Cloud statistics over land*

Cloud size information may have utility for weather (durations of shade and potential for lightning, the latter has some direct implications).

## *Lake sizes*

Our environment shows great diversity in the size and abundance in natural structures. Freshwater lakes accumulate their volume in a dispersive manner. Over geologic time, water drifts into a basin at various rates and over a range in collecting regions. As lakes capture most of their volume through water drainage, one can imagine that the rate of influx plays a factor in how large a lake can become. The Maximum Entropy prediction of the size distribution leads to the following expression, exactly the same as for cloud formation:

$$P(Size) = \frac{1}{1 + \frac{Median}{Size}}$$

Surveys of lake size show the same reciprocal power law dependence, with the exponent usually appearing arbitrarily close to one. In the Figure 27 below, the data plotted on a ranked plot clearly shows this dependence over several orders of magnitude.
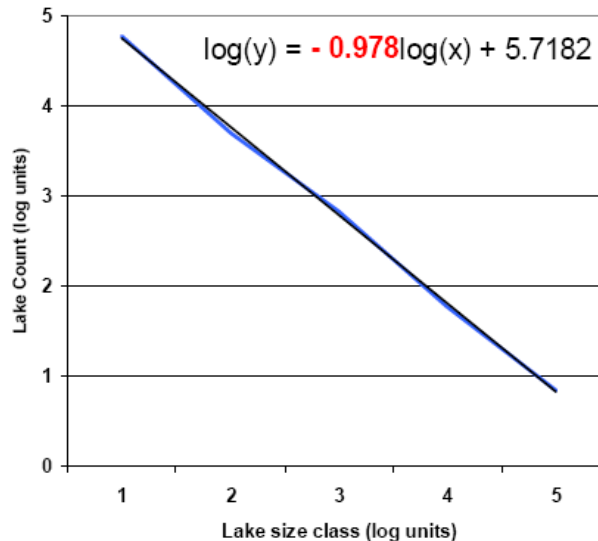


*Figure 27: Northern Quebec lakes map and size distribution adapted from Telmer [34]*

More revealing is that in Figure 28 below, we can observe the bend in the curve that limits the number of small lakes in exact accordance to the equation shown above. The agreement with such a simple model suggests that a universal behavior.
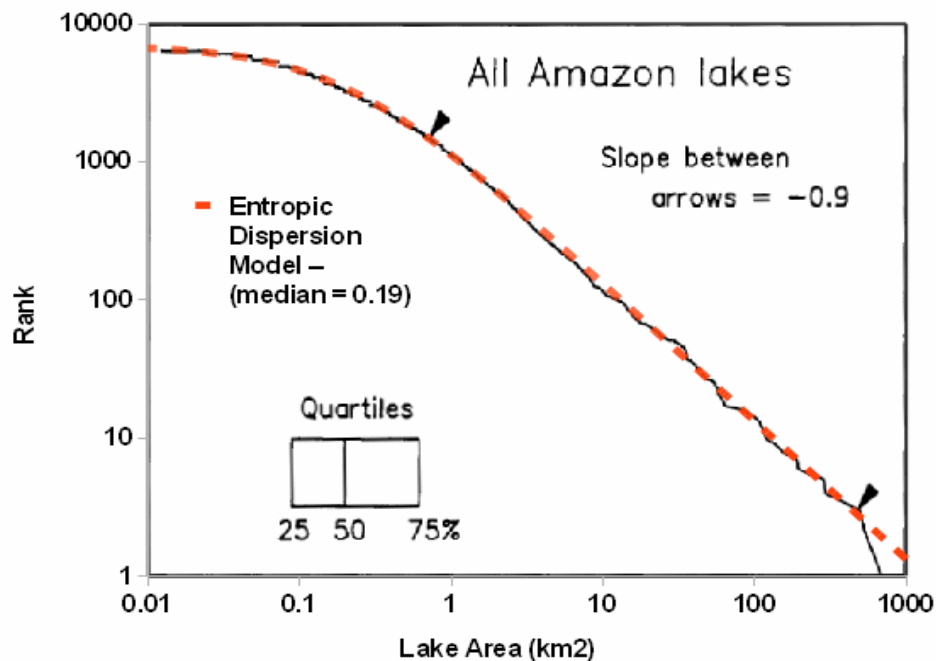
*Figure 28: Amazon lake size distribution from "Estimation of the fractal dimension of terrain from Lake Size Distributions"[35]*

This general trend is repeated for lakes around the world. All that is required is to have a median value for lake area and the rest of the distribution will rougly scale to this factor.
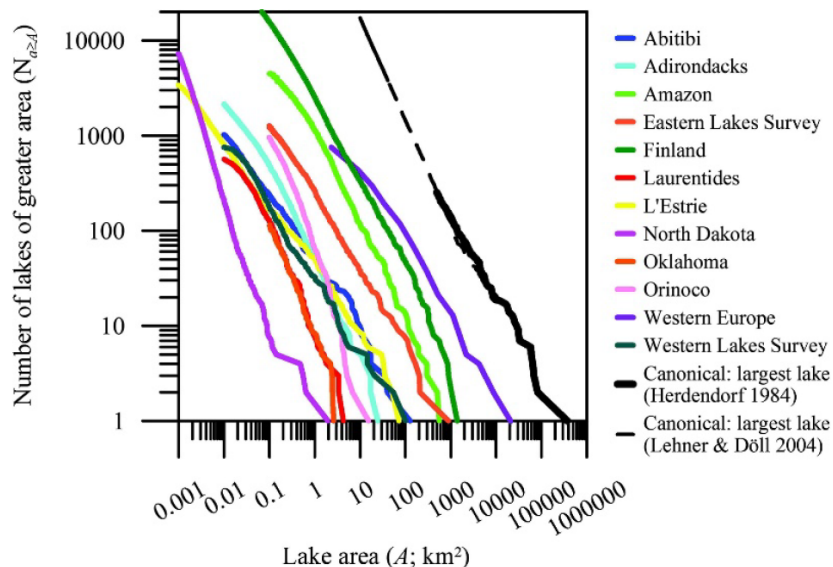


*Figure 29: Near universal scaling of lake size distributions across the world[36]*

Models of lake size distribution may have utility for vehicle fording and amphibious vehicle simulations.

*Particulates*

The role of entropy overrides other factors enough so that some simple dispersion arguments can explain the size distribution completely. Take as an example the formation of ice crystals in a cirrus cloud. Depending on the surrounding temperature, a crystal nucleates on some foreign particle and then starts growing. The atmospheric conditions have enough variety that the growth rate will disperse to the maximum entropy amount given a mean rate value. The end state for volumetric growth will also show the same amount of variation, where x is the size variate and S is the mean size.

The following particle size distribution (PSD) graph shows measurements taken from high altitude cloud experiments[37],[38],[39]. The size gets measured along a single length dimension and the density of the particles takes the place of a probability. This assumes that they have plotted the density as a mass function along the x-axis. Converting this to a volumetric size growth problem, integrate the growth for a time duration corresponding to the cloud formation period, and plot the entropic dispersion model gives the result below. The value for S is 2 and the integration scale is 16.
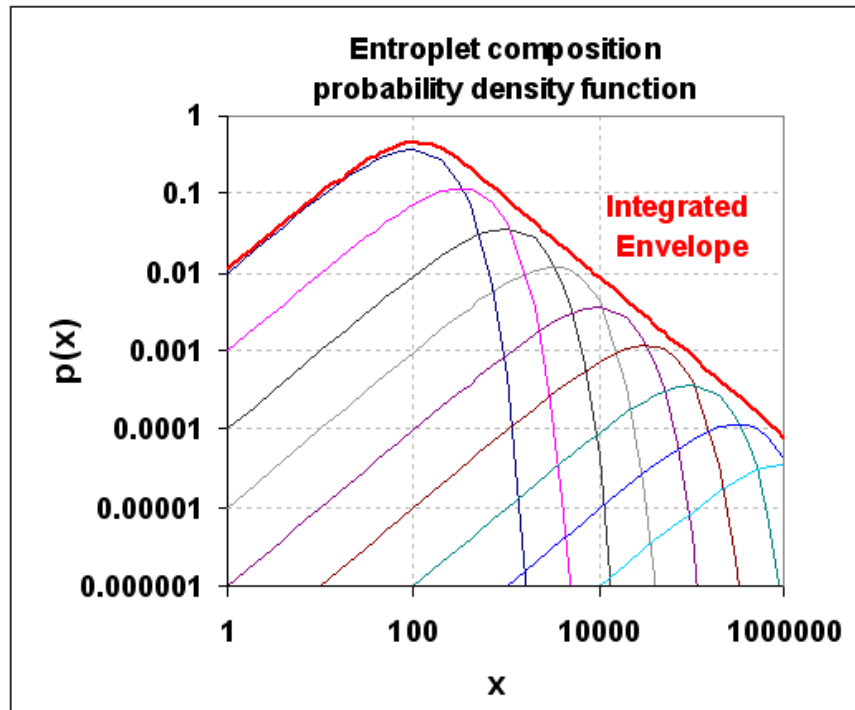


*Figure 30: Superposition of exponential density profiles leads to a entropic dispersion envelope.*

By smearing over start time we can essentially broadens the peak but retain the power law tail. Convert the linear growth to a time varying parameter (x=kt) and then treat the growth as an evolving pattern which starts from multiple points in time as the cloud develops.

This is a cumulative so take the derivative with respect to x to get the PDF.

Crystal sizes get reported as a length and we have to convert that to a volume. This means the derivative has to include a chain rule to convert the volume x to a length parameter L, $x \sim L^3$ generates $dx/dL \sim L^2$.

The data fits the dispersion model nicely (green line), but notice at low density that an extra mode shows up as the blue line. This clearly has a sharp exponential drop so likely has a non-dispersive origin. In terms of the higher density entropic model, this stands out as an ordered nucleation regime in the midst of a sea of disordered ice crystal growth modes. Compare to the first figure again and one can see how a distinct peak could occur.

Why or how physically can this bump occur? Some unknown nucleation process has provided an optimal growth environment for these crystals to deviate from the entropic distribution. Hypothesizing, this could take the form of a catalyst or an accommodating growth substrate. With a power tail of -3/2 this might well have a strong diffusive growth component. However, the nuclei occur rarely enough so they do not drown out the much more common random or spontaneously occurring growth centers. It thus shows up as a clear non-dispersive growth mode in a sea of non-uniformity.

On a micro-level, we do have a population of reproducible structured shapes to bind against — as the airborne particulate world shows some uniformity in its density. This original analysis may prove of some help to those looking at particle size distributions of volcanic ash (see the figure below). Researchers routinely apply a log-normal fit to the data — yet one that uses an MaxEnt dispersion formulation with the appropriate volume/diameter exponent often can work just as well. Below, we use a root 1/2 dispersive growth rate on volume which may indicate a diffusion-controlled rate.

To account for wear/tear/corrosion modeling, one could pull in particulate and ash models that feed into models for physical breakdown. Again, these models will often reduce to simple analytical forms.
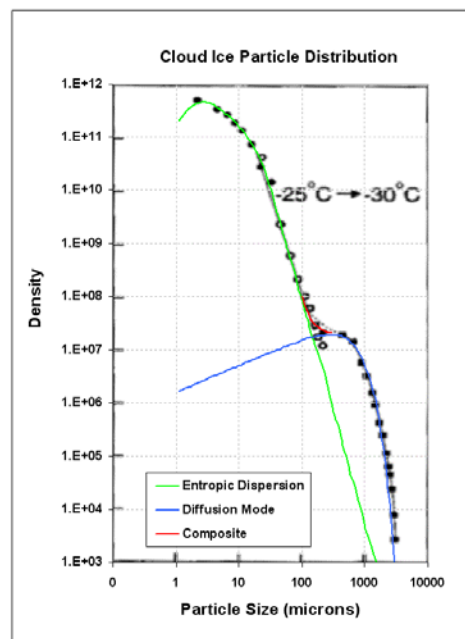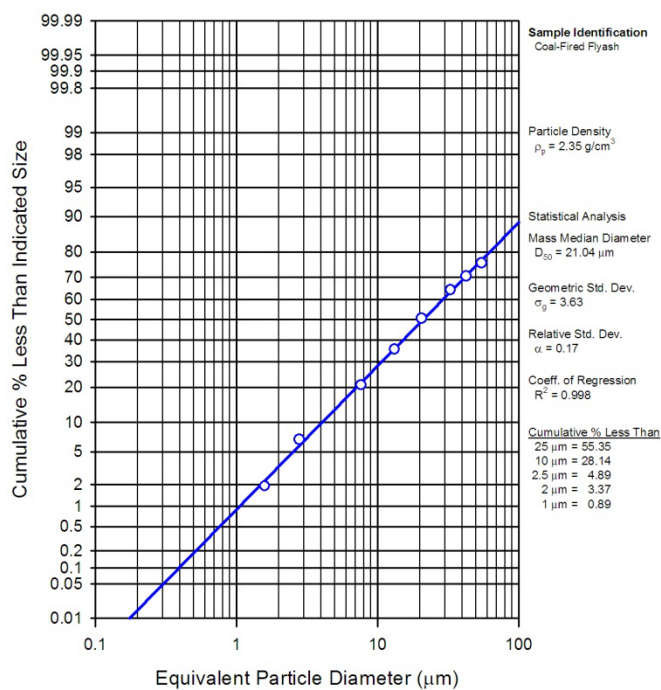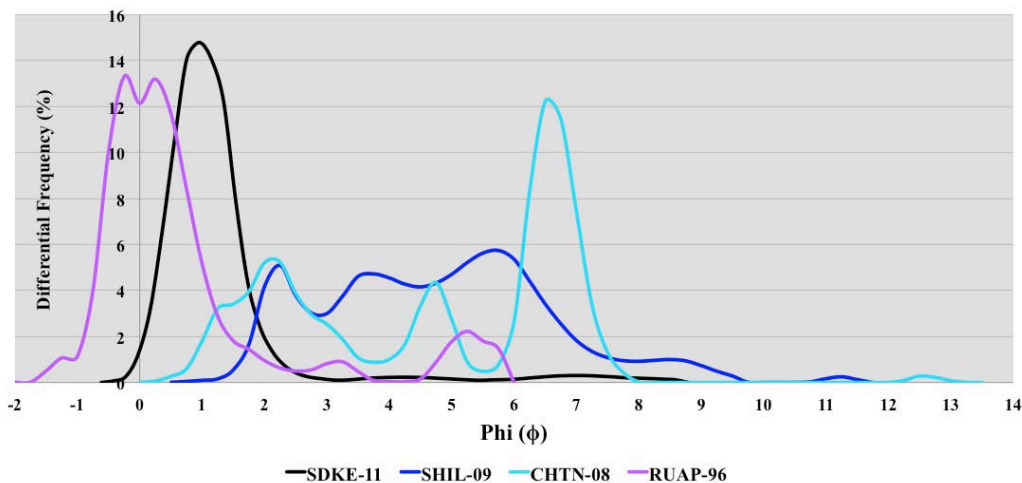


*Figure 31. Cloud ice particle number density n(D) vs. the long dimension of particles as observed at the temperature range of - 25× C and -30× C [from Platt[40]]. It shows a bimodal structure in the ice crystal distribution with the second peak at ~500 microns"*

*Figure 32: Wiki data for fly ash size distribution.[41]*



*Figure 33: Volcanic Ash[42]. This is on the Krumbein phi (φ) particle size scale[43] which is negative of the log base 2 of the particle diameter. That means smaller diameters are to the right side of the scale.*

## Discussion

This study focused on the characterization and modeling of natural phenomena in a form suitable for contextual use. We worked through examples of slope distribution, wind speed, rainfall rates, lake sizes, etc. Most of the models reduce to fairly simple analytical forms.

The probability measures derived from theory and applied to the data are often snapshots in time of a continuous growth process. Physical processes can accumulate or discard material following rates of material flow. While this may seem overly simplistic to consider, it does set the stage for revealing patterns in behavior.

For example, a statistically measured volume could be the result of various rates operating over varying time spans. Along with this variation, the concept of an average or mean value plays a role into modeling the behavior.

Two categories of models exist:

- *Models of variations around a mean energy*: Measures are bound by uncertainty in the mean energy value of a measured value, such as wind, rainfall rate, clutter, terrain slopes, wave crest height; these all generate the same universal curve, the ubiquitous modified Bessel of the second kind.
- *Models of variations around a rate and enclosed volume*: Several measures are bound by considerations of the rate of accumulation and the uncertainty of the eventual volume: such as clouds, particles, and lake sizes give the same entropic power law.[4]

We also described how the statistical spreads come about, via the concept of dispersion and max entropy. One of the outcomes of modeling is that a significant amount of data, 5 billion in the case of USA terrain and several hundred million for just the Lake Michigan wave data, can be expressed in a much more concise form.[5]

Sampling approaches for simulation are often very straightforward for these distributions. We described several efficient applications of random sampling for transcendental functions such as the BesselK. We can further apply techniques such as importance sampling to reduce the sample size needed or to reduce the variance of the statistics (see Meta demos for good examples).

In certain cases it may be possible to invert the stochastic models for assume-guarantee models and PCC applications. The classical case is for normal distributions, but for fat-tail distributions such as the hyperbolic power law, this is rather straightforward.

The fact that many of these models follow similar patterns makes it them useful for model automation. We can generate abstract interfaces for common distributions such as the BesselK and hyperbolic power law.

Other characterization techniques such as the multiscale entropy measure can be applied to temporal and spatial scales covering a wide dynamic range. What this tells us is the amount of disorder and uncertainty in the data, which is important as a quick characterization metric.

**Other application areas**

---

[4]  Plus oil reservoirs.

[5]  Archived data often gets stale, and eventually gets deleted via some bureaucratic measure. Possessing a model that can potentially condense 4 billion data points into a single function containing a single parameter has clear benefits. It amounts to a 4 Billion to 1 reduction in storage size and encapsulated domain knowledge.

Frequency characteristics will be described in Volume 2.

The use of some universal principles allows us to apply the same pattern to characterize other disparate natural phenomenon, similar to what we did for terrain slopes, such as corrosion and thermal dispersion in Volume 3.

In general, support for context modeling and the general notion of Internet-based data collection allows us to leverage a vast amount available data and reduce its dimensionality and scope by careful characterization and modeling. Further, ontology-based approaches for organizing models and techniques sets the stage for discovery and formulating standard access protocols for context modeling.

# References

[1] D. Mumford, "The Dawning of the Age of Stochasticity," *Mathematics: Frontiers and Perspectives* (2000): 197–218.

[2] DARPA, "23 Mathematical Challenges", n.d., http://www.darpa.mil/Our_Work/DSO/Programs/23_Mathematical_Challenges.aspx.

[3] E.T. Jaynes and G.L. Bretthorst, *Probability Theory: The Logic of Science* (Cambridge Univ Pr, 2003).

[4] D. Mumford and A. Desolneux, *Pattern Theory: The Stochastic Analysis Of Real-World Signals* (A K Peters, Ltd., 2010), http://www.dam.brown.edu/ptg/MDbook/index.html.

[5] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization, and Disorder: Concepts and Tools* (Springer Verlag, 2004).

[6] N.N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (Random House Inc, 2010).

[7] N. Taleb, *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets* (Random House Inc, 2005).

[8] M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex* (St. Martin's Griffin, 1995).

[9] C. Beck and EGD Cohen, "Superstatistics," *Physica A: Statistical Mechanics and Its Applications* 322 (2003): 267–275.

[10] Rudolf Hanel, Thurner, and Murray Gell-Mann, "Generalized Entropies and the Transformation Group of Superstatistics", n.d., http://www.pnas.org/content/108/16/6390.full.

[11] S. Bankes et al., *META Adaptive, Reflective, Robust Workflow (ARRoW)*, Final Report, Phase 1B (BAE Systems, 2011), http://www.darpa.mil/Our_Work/TTO/Programs/AVM/AVM_Meta_Final_Deliverables.aspx.

[12] "BPA Meteorological Information", n.d., http://transmission.bpa.gov/business/operations/wind/MetData.aspx.

[13] S. M Papalexiou, D. Koutsoyiannis, and A. Montanari, "Can a Simple Stochastic Model Generate Rich Patterns of Rainfall Events?," *Journal of Hydrology* (2011).

[14] T. Yuan, "Cloud Macroscopic Organization: Order Emerging from Randomness," *Atmos. Chem. Phys* 11 (2011): 7483–7490.

[15] J. A. Downing et al., "The Global Abundance and Size Distribution of Lakes, Ponds, and Impoundments," *Limnology and Oceanography* (2006): 2388–2397.

[16] Dong L. Wu, "MLS Cloud Ice Measurements", n.d., http://mls.jpl.nasa.gov/dwu/cloud/index.html.

[17] "CDIP Homepage", n.d., http://cdip.ucsd.edu/?units=metric&tz=UTC&pub=public&map_stati=1,2,3.

[18] Corps of Engineers US Army, "Wave Information Studies", n.d., http://wis.usace.army.mil/hindcasts.shtml.

[19] "USGS/EROS Find Data/Products and Data Available/DEMs", n.d., http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/DEMs.

[20] G. Vico and A. Porporato, "Probabilistic Description of Topographic Slope and Aspect," *J. Geophys. Res* 114 (2009): F01011.

[21] J.S. Gagnon et al., "Multifractal Earth Topography," *Nonlinear Processes in Geophysics* 13, no. 5 (2006): 541–570.

[22] G. Gonçalves and J. Santos, "Propagation of Dem Uncertainty: An Interval Arithmetic Approach.," in *XXII International Cartographic Conference, Spain*, 2005.

[23] A.M. Guarnieri, "SAR Interferometry and Statistical Topography," *Geoscience and Remote Sensing, IEEE Transactions On* 40, no. 12 (2002): 2567–2581.

[24] US Army, "Wave Information Studies."

[25] M.A. Tayfun, "Narrow-band Nonlinear Sea Waves," *Journal of Geophysical Research* 85, no. C3 (1980): 1548–1552.

[26] Hervé Socquet-Juglard et al., "Probability Distributions of Surface Gravity Waves During Spectral Changes," *Journal of Fluid Mechanics* 542, no. -1 (October 25, 2005): 195.

[27] Izadparast, "Empirical Moment-Based Estimation of Rayleigh-Stokes Distribution Parameters", n.d., https://docs.google.com/viewer?a=v&q=cache:1zZ2W7kUE-QJ:www.isope.org/publications/proceedings/isope/isope%202011/data/papers/11DA-03Izadpara.pdf+&hl=en&gl=us&pid=bl&srcid=ADGEEShQwoT6ngDp3fqRhJvNVTb-IyCgI0k1VQnuwWhsDAdpUypm8_QXd9zGLIGwcUzs6suWAvsvccOS0FmxTIn_Je5qO83JfuQz4pkK9KN10vSFyG1QKV3Jhs4giw3Ws8-5_74EHIfI&sig=AHIEtbR2IUAiO_UDczN12Z1vQVOBYLYiDQ.

[28] Papalexiou, Koutsoyiannis, and Montanari, "Can a Simple Stochastic Model Generate Rich Patterns of Rainfall Events?".

[29] Ibid.

[30] J.R.A. Watson, *High-sensitivity GPS L1 Signal Analysis for Indoor Channel Modelling* (University of Calgary, Department of Geomatics Engineering, 2005).

[31] M. Chen et al., "Design and Construction of a Wildfire Instrumentation System Using Networked Sensors," *Network Embedded Systems Technology (NEST) Retreat, Oakland California. Retrieved April* 5 (2004).

[32] U. Grenander and A. Srivastava, "Probability Models for Clutter in Natural Images," *Pattern Analysis and Machine Intelligence, IEEE Transactions On* 23, no. 4 (2001): 424–429.

[33] Yuan, "Cloud Macroscopic Organization."

[34] K. Telmer, M. Costa, and J. MacGregor, "K&C Science Report–Phase 1 Global Lake Census" (n.d.).

[35] S.K. HAMILTON et al., "Estimation of the Fractal Dimension of Terrain from Lake Size Distributions," *Lowland Floodplain Rivers: Geomorphological Perspectives. Wiley* (1992): 145–163.

[36] Downing et al., "The Global Abundance and Size Distribution of Lakes, Ponds, and Impounds."

[37] C.M.R. Platt, "A Parameterization of the Visible Extinction Coefficient of Ice Clouds in Terms of the Ice/water Content," *Journal of the Atmospheric Sciences* 54, no. 16 (1997): 2083–2098.

[38] Wu, "MLS Cloud Ice Measurements."

[39] D.L. Wu, J.H. Jiang, and C.P. Davis, "EOS MLS Cloud Ice Measurements and Cloudy-sky Radiative Transfer Model," *Geoscience and Remote Sensing, IEEE Transactions On* 44, no. 5 (2006): 1156–1165.

[40] Ibid.

[41] "Particle-size Distribution - Wikipedia, the Free Encyclopedia", n.d., http://en.wikipedia.org/wiki/Particle-size_distribution.

[42] "Volcanic Ash - Wikipedia, the Free Encyclopedia", n.d., http://en.wikipedia.org/wiki/Volcanic_ash.

[43] "Particle Size (grain Size) - Wikipedia, the Free Encyclopedia", n.d., http://en.wikipedia.org/wiki/Particle_size_(grain_size).