HPD REVOLUTIONIZES STOCHASTIC MODELING, ANALYSIS, AND OPTIMIZATION; MAJOR IMPLICATIONS

Jean M. Parks[1] and Mohammad Noori[2]

[1]Founder, Variability Institute
10103 E Graythorn Dr., Scottsdale, AZ
jparks20@gmail.com; https://variability.institute/

[2]Emeritus Professor, Mechanical Engineering
California Polytechnic Institute, San Luis Obispo, CA 93401
mnoori52@yahoo.com; https://www.linkedin.com/in/mnoori/

Key Words: Reliability, Risk, Probability Distributions, Optimization, Complex Systems, Stochastic Modeling

Abstract. This brief paper gives an overview of our breakthrough Holistic Probabilistic Design (HPD) Software and Methodology that revolutionizes Stochastics with fundamentally new techniques and tools for Stochastic Modeling, Analysis, and Optimization for simple or complex systems. Even more fundamentally, a major implication of two of HPD's key advancements is that it "liberates" probability distributions. One aspect of liberation is the following: Currently Stochastic Analysis utilizes existing, or known, types of distributions, but these are just a small discrete set within the continuum of distribution space consisting of an infinitude of types, or no types at all. We will also discuss why most of the known distribution types are actually not reasonable for real-world applications. The second aspect of liberation is that distributions no longer need to be given as functions; they are digitized and exist as "Data-Type" distribution files and plots. Handling data-type distributions is included in HPD which has taken many intermittent years for the authors to develop. The methodology prescribes how to structure the work process for, and stochastically model, a complex problem with numerous random variables that should be addressed. The software (i) is monumental in size, (ii) has an extremely user-friendly graphical user interface (GUI), and (iii) will be imminently made available open-source. The fundamental breakthrough of the HPD software is being able to compute the "exact" output distribution for any relationship, $\mathbf{z} = g(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ for "any" set of input distributions, for that led to liberating distributions. HPD is based on our SMB (≡ Stochastic Modeling Based) technique which has overcome the well-known "curse of dimensionality" issue. The authors are currently completing a book on revolutionizing Stochastics that includes rigorous mathematical concepts and computational techniques for the HPD capabilities. Such documentation is necessary for leaving to posterity. The authors originally intended to market the software, but have now concluded that making it freely available to the Stochastics community is a far more satisfying outcome. Liberating distributions will have a consequential impact on existing literature, techniques, and tools in Applied Probability and Stochastics. And HPD's availability would enable revolutionizing Stochastics to take hold.

INTRODUCTION

Due to the length limitation for this paper, and since material on HPD is extensive in scope, the authors have intentionally included a long abstract to provide a more substantive preview of HPD and its important implications and thus, shorten the paper's length.

This paper will only provide a concise overview of HPD; much more information in this regard is available on our website. The main focus of this short paper is on explaining (i) how HPD liberates distributions and what that means, (ii) why most existing, or known, types of distributions are unreasonable for real-world situations, and (iii) why this would dictate that current Applied Probability should be *redefined* and completely separated from Theoretical Probability. Therefore, herein, aside from the main focus stated above, we will mostly provide information that supplements or clarifies those in the abstract and on the website in the following sections:

1. A concise history of HPD and its goal
2. A glimpse of the HPD Software and Methodology
3. HPD's breakthroughs and advancements
4. Some important implications
5. On liberating distributions
6. How is adequate Stochastic Modeling conducted; an example
7. Our forthcoming book on revolutionizing Stochastics
8. Why did HPD take so long to reach "product-readiness"

Each section will be quite brief. Please note that portions of Sections 3 and 4 are highlighted to emphasize the breakthroughs/advancements and significant implications.

1. A CONCISE HISTORY OF HPD AND ITS GOAL

Our effort to seriously address probability of failure started in Xerox' golden age of engineering years ago. Its systems were rife with variability, with only a few "analysis-oriented" engineers being slightly familiar with the Monte Carlo (MC) technique, but were inappropriately applying it. Our background in Mathematics and Engineering led to the beginnings of new techniques, not MC-based, that ultimately became HPD.

We were able to apply our techniques and tools to many technology and product development projects, some of which were very complex. Our ambitious *goal*, however, was to enable *every* engineer, i.e., including those who were not analysis-oriented, but were "experimentally- or hardware-oriented," to use them, not realizing that such a goal would take a tremendous amount of effort – as described in Section 8. We finally reached "product-readiness" for HPD and froze the development effort in 2020.

2. A GLIMPSE OF HPD SOFTWARE AND METHODOLOGY

There are mainly three categories of stochastic needs that require techniques & tools (T&Ts): Stochastic Modeling (SM), Stochastic Analysis (SA), and Stochastic Optimization (SO). T&Ts for SA and SO have received all the main attention, but stochastic models are what SA and SO act upon. Adequately conducting SM is just as important; thus, HPD addresses all three, where SM is part of its methodology as shown in Figure 1.

In the figure, The HPD_VA suite is for SA which is described as the "Umbrella Area of Variability Analysis." It includes seven tools for:

a) Failure Analysis
b) Contribution/Sensitivity Analysis
c) Computing output distributions

The most important one is for the fundamental breakthrough – see next section – which includes (c) and the basic technique for (a). The SO tool, HPD_Opt, has 36 paths and sub-paths. Section 3 will give an overview of some of these breakthroughs and advancements.

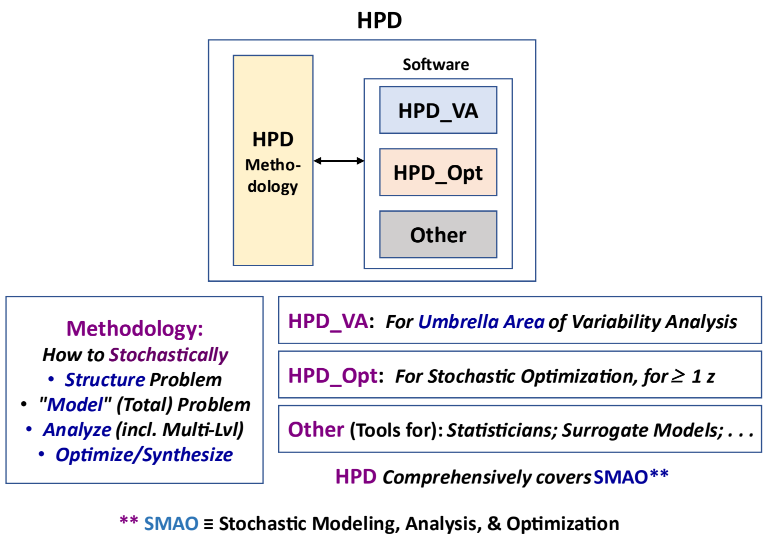

Figure 1: Glimpse of HPD's total set of capabilities

3. HPD'S BREAKTHROUGHS AND ADVANCEMENTS

Only five breakthroughs/advancements are presented herein, although there are more; and only very brief statements are provided for these. Where appropriate, an essential clarification on how the breakthrough or advancement was attained is given.

Let us introduce a few notations beforehand since these would be too cumbersome to define when they are encountered. Some of the notations below are not used in this paper, but they rigorize those that are, such as $\mathbf{n}$ in $\mathbf{x_n}$ instead of using n; also, they are to introduce the readers to rigorous notations that are used in the mathematical concepts and techniques behind HPD where both deterministic and stochastic notations are needed:

-- Random variables (RVs) are in boldface lowercase letters. E.g., $\mathbf{z}$, $\mathbf{x_n}$

-- Deterministic variables or other items are in regular font style. E.g., z, x_n, the index n in, say, x_n, and the g as in $\mathbf{z} = g(\mathbf{x}, \mathbf{y})$. Exception: The $\mathbf{n}$ in $\mathbf{x_n}$ serves double duty, as part of the name, $\mathbf{x_n}$, and as the index in place of n. Reason for the latter is that $\mathbf{x}_n$, where n is not in boldface, has an entirely different meaning.

-- Capital letters indicate a set of lowercase items: E.g., $\mathbf{X}$ is the set $\mathbf{x_1}$, $\mathbf{x_2}$, …, and $\mathbf{x_N}$.
-- $f_\mathbf{x}(x)$ (or just f(x)), $f_n(x_n)$, and $f_\mathbf{z}(z)$ (or just f(z)) are density distributions in the *ranges of variability* of $\mathbf{x}$, $\mathbf{x_n}$, and $\mathbf{z}$. Likewise, $F_\mathbf{x}(x)$, . . . for the cumulative distributions.
-- For failure analysis: 1 FC means 1 model plus a failure criterion on $\mathbf{z}$; 2 FCs mean models $\mathbf{z_1} = g_1(\mathbf{X_1})$ and $\mathbf{z_2} = g_2(\mathbf{X_2})$, each with its failure criterion; and so on.
-- RVs are labeled as parameters, as in design, noise, or performance parameters

3.1 The fundamental breakthrough

The fundamental breakthrough [1] is the capability to compute the "exact" probability distribution of $\mathbf{z}$ for "any" stochastic model,

$$\mathbf{z} = g(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N});$$

thus, also the "exact" probability of failure. This completely supersedes Monte Carlo.

Discussions on the innovation:

- The innovation is our three-part SMB technique, one part of which is the SG&E (≡ Sparse Grid & Explosion) technique which overcomes the well-known "curse of dimensionality" issue for multiple integration for the failure integral. This will be discussed in our forthcoming book.
- "Exact," of course, takes into account that computational techniques carry minor errors.
- "Any" does have some limitations on N; limit in the current version of the software is 10. For N ≤ 5, computation is executed in a blink of time with high accuracy where accuracy is measured against using sums of Normal distributions to compare. For larger Ns, HPD includes lesser tools or work processes.

One significance of the innovation: Being able to compute the 'exact" output distribution is *analogous* to an *existence theorem* in Mathematics. That the distribution of $\mathbf{z}$ exists as a legitimate distribution, and that it is dependent on input distributions *and* the function $\mathbf{z} = g(\mathbf{X})$, means that $f_\mathbf{z}(z)$ almost certainly would not be of any known types of distribution, i.e., it most likely would be of "no type" at all, but be any in the continuum of distribution space.

3.2 Novel Stochastic Modeling technique

Stochastic modeling must address *all* contributors to variability for the problem. HPD's methodology [1, 2] prescribes how to appropriately model by (i) addressing a Flow of variabilities (FoV) from lowest, or lower, to performance levels as necessary, (ii) defining a "total model," and (iii) structuring the work process.

This enables addressing two complex types of problems: a) A problem that has a multitude of RVs that must be included – Section 6 gives an example; b) a problem of sequential processes.

This, along with the fundamental breakthrough, has impact on liberating distributions – see Subsection 4.1.

3.3 Advanced Contribution/Sensitivity Analysis techniques

HPD's Contribution/Sensitivity Analysis techniques [1] supersede those for existing Sensitivity Analysis by removing the latter's omissions and errors through addressing cross-term effects (CTEs) and "NU-ness," respectively.

Herein we will just state that the limit for this HPD tool is 25 RVs and that the tool internally uses the fundamental breakthrough when computing the first and second order terms in the Taylor series expansion of $\mathbf{z} = g(\mathbf{X})$. Defining NU-ness will be discussed in the book.

3.4 Advanced techniques for Failure Analysis

The fundamental breakthrough includes computing the "exact" probability of failure for one FC. Extension of the technique allows computing exact probability of failure for two or more FCs under certain conditions on $\mathbf{X_2}$, $\mathbf{X_3}$, . . . for $\mathbf{z_1} = g(\mathbf{X_1})$, $\mathbf{z_2} = g(\mathbf{X_2})$, $\mathbf{z_3} = g(\mathbf{X_3})$, . . . Three other application techniques for Failure Analysis are also included.

3.5 Advanced Stochastic Optimization technique

HPD's Stochastic Optimization [3, 4] addresses "latitude," robustness, meeting targets, and both internal and external noise. The software optimizes for the operating setpoint ($\equiv$ the combination of the means of the design parameter being optimized) when an entity has one *or more* performance parameters.

This goes far, far beyond Taguchi methods (TM); also, TM has major limitations and a basic fallacy.

4. SIGNIFICANT IMPLICATIONS

The following four implications from the breakthroughs/advancements (B/As) stand out. Given in parenthesis are the B/As, numbered as the subsections, e.g., #3.1- #3.5, that enabled each implication.

4.1 Liberates probability distributions [5] (Due to #3.1 and #3.2)

This is the *most striking* implication. The currently "known" types of distributions, i.e., the "named" types, are but a *miniscule discrete set* in the continuum of "distribution space."

Distributions are implemented in HPD as follows: Distributions are *digitized*; they exist as *datafiles* and plots. All output distributions are in digitized form. Inputs that are of known types are digitized before being used. An added advantage is that the digitized format contributes *immensely* to computational *efficiency*; this will be clarified in our book.

Section 5 will further discuss distributions.

4.2 Enables computing Flows of variabilities (FoV), distributions, and models, as in "total model" (Due to #3.1 and #3.2 being symbiotic)

That output distributions are accessible as input distributions for any next level computation is extremely important since it enables modeling and analyzing with input distributions based on causal variabilities from any lower levels as needed.

This essentially makes the term epistemic uncertainty moot since one can model and analyze down to the root cause of any level of input variability.

4.3 Enables *correctly* identifying high contributors to variability (Due to #3.1and #3.3)

Correctly doing so is so important to predicting probability of failure and for tolerancing. Also, #3.3 and #3.1 form a dynamic duo in capabilities.

4.4 Enables Transcendence to the Stochastic Realm (Due to elements #3.1 and #3.2)

HPD's comprehensiveness and extreme user-friendliness are intended to enable anyone to easily think stochastically as one could deterministically. This is what is meant by transcendence; it would be a paradigm shift.

Note that the stochastic "realm" is far, far larger than the deterministic realm. In this larger "space," so much is unifiable – a deep philosophical comment here. And so much can be simplified.

5. ON LIBERATING DISTRIBUTIONS [5]

In this section we will discuss (i) real world variabilities and currently known types of distributions and (ii) the reason that most known types are unreasonable for real-world situations. Then we will give two examples of output distributions and an important point.

5.1 On current known types of distributions

As we all know, currently, input RVs for any model are assumed to be any of the known types, or "named types," of distributions. Except for the Normal distribution, which is the continuous extension of the Binomial distribution and which can be assumed to have validity for an extremely large sample if the dispersion is binomially distributed, almost all other existing continuous distributions do not have any justification for their representing real-world situations – we say "almost all" because the Uniform distribution, for example, is a named type that does represent some real-world situations, That currently there are ~98 continuous distributions types being options for practitioners to choose from is purely due to those being available because probabilists (≡ probability specialists) derived most of those from t*ransformations* of the Normal distribution. I.e., those are just mathematically derivable distribution types, *not* typically what one would encounter in real-world situations. There are other approximate distribution types people use, e.g., the Pearson System or the Johnson System for fitting data; these types are just as limited and do not merit further discussions here.

5.2 For real-world situations, distributions do not extend to infinity

It is important to understand that in contrast to Theoretical Probability, for real-world situations/problems, i.e., in Applied Probability and Stochastics, distributions do not extend to infinity because for applications, the ranges of variability of an RV are bounded. Of course, one can truncate any named distribution that does have an infinite range of variability and then normalize it, and that is done in HPD.

5.3 Why are most known types of distributions not reasonable for real-world situations

Aside from the issue of variability for real-world situations/problems pointed out in Subsection 5.2, consider the *flaw* of, e.g., the Lognormal distribution for design engineering**:** Almost none of the *natural* engineering relationships in, say, a Flow of variabilities with any set of input distributions would have resulted in a Lognormal distribution – this also applies to the distribution of any *part's* feature since the *part* would have been manufactured from *its* design process. And that is true for *most* of the named-types of distributions because of their origins as explained in Subsection 5.1.

5.4 Examples of liberated distributions

Two examples of computed outputs are given here. Although the input distributions were known types of distributions, the output distributions are definitely of no type at all, i.e., they are liberated.

The first example shows the output density for the cantilever beam equation **z** = g(**X**) where all inputs were (truncated) Normal distributions.

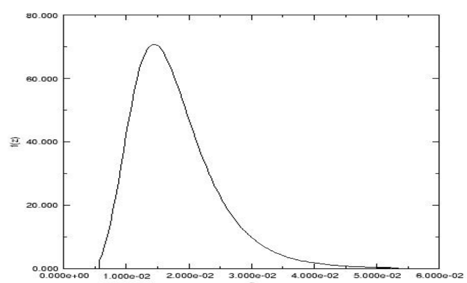

Figure 2: Density for deflection of cantilever beam

The density shown in Figure 3 is for the simple equation **z** = $\mathbf{x}^2$ – **y** where **x:** U(– 3.0, 3.0) and **y:** N(0, 1.0). The strange shape gave us pause at first, but later we were able to "prove" that it was correct.

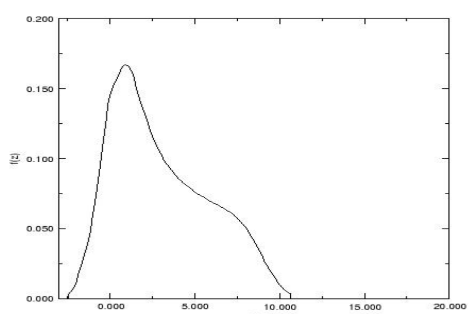

Figure 3: "Strange-Shaped" distribution

These output distributions are, of course, automatically accessible as inputs for some models that include them as input RVs.

Very simple instructions are given for turning experimental variability information into data-type distributions. The HPD software infrastructure enables all this with ease.

An important point: There is no need for distributions to be given as *functions*, thus there is no need for the terms cdf and pdf; simply use the terms density and cumulative distributions, or those can be just densities and "cumes," for short.

6 HOW IS ADEQUATE STOCHASTIC MODELING CONDUCTED [5]

There is quite a logical process for this; however, it would be too detailed to describe for this short paper, but we can give a brief essence of it. Here we will use the copier's Fuser as an example. Among its several performance parameters, let's consider just one, say, Fuser Roll Life. This was an actual Xerox application.

Simplistically, the process proceeds as follows: Define an engineering metric for Fuser Roll Life, say, Coating Effectivity, Y. Next, specify all parameters with variability that affect Y; there were 14 in this case. Then give each a mathematically efficient symbol (which is something that experimentally-oriented engineers are not accustomed to doing). After this, methodically define what *next* level (i.e., downwards) parameters Y *depends* upon *and* specify how one is to obtain that relationship (i.e., model), analytically or by design of experiments. Repeat with each of the next level parameters, and so on. These steps are similar to *how* engineers define a FAST diagram in industry, but the major difference is that HPD's process is to enable a *quantifiable* total model whereas the FAST diagram is *not* meant to be quantifiable.

Figure 4 is the Flow of *Dependencies* only, i.e., those become a flow of piecewise models (between symbols that point to the next level symbol) when the work process for obtaining the models have been carried out.

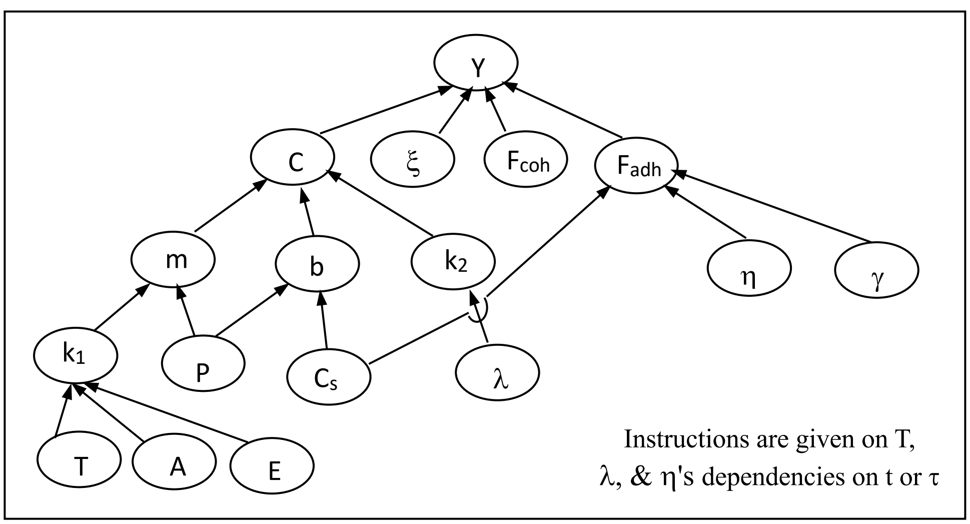

Figure 4: Flow of Dependencies for Fuser Roll's Coating Effectivity

This process of modeling is just as comfortable for the experimentally-oriented engineer as for the analysis-oriented engineer, and its comprehensiveness is novel for *both* types. That is one part of what we meant regarding our goal of enabling *any* engineer to be stochastically literate. Note: The HPD methodology also addresses non-independence of RVs that arise in the Flows.

When we guided the Fuser technologist – a PhD engineer quite experienced in developing Fusers – on this, he was ecstatic that together we had *structured* his work process for the next two to three years. Also, in an HPD-based course we taught at the University of Rochester, some students thought this was extremely useful and used it in their individual course projects.

7 OUR FORTHCOMING BOOK ON REVOLUTIONIZING STOCHASTICS

The book is 95% completed. It will present Part A of Revolutionizing Stochastics; Part B, which will include most of Stochastic Optimization and some advanced topics and applications of Stochastic Analysis, will appear in a later volume. The book will include rigorously discussing the mathematical concepts and computational techniques (MCCTs) of HPD's breakthroughs and advancements. Details of such rigor also include precise notations and discussing (i) when is one operating in the deterministic or stochastic realm, (ii) limitations, (iii) minor needed debugging, (iv) possible extensions, (v) future R&D, and (vi) why HPD simplifies and unifies. The book should interest all in the Stochastics community and is necessary for passing the MCCTs of the breakthroughs/advancements down to posterity.

8 WHY DID HPD TAKE SO LONG TO REACH "PRODUCT-READINESS"

Briefly, the following categories of reasons caused HPD's lengthy development time: Intermittency of effort, retirements and serious health issues, overestimation of what *any* practitioner could understand, and making its set of the tools extremely user-friendly and comprehensive.

8.1 Intermittency of effort:

Intermittency differs for the two authors: Below, (a) refers to the first author and (b) refers to the second author:

(a) After we started late in life at Xerox and developed a new technique for Failure Analysis, we diverted into management for a number of years, but that enabled our understanding details of technology and product development, and the dire need for addressing variability. We left management to focus on developing what became HPD with the goal stated in Section 1. Other causes of intermittency: (i) First retirement followed by pre-planned five years in academia to deploy HPD; (ii) second retirement followed by a necessary 10-year break; (iii) difficult resuscitation of effort with new team (of mostly retirees); (iv) team problems due to old age. Note: Resuscitating the effort was because we realized that HPD still constituted a breakthrough.

(b) Due to our common expertise in Probabilistic Methods, we interacted a great deal with the first author and supported each other in many ways. We then diverted to academic administration for many years and gained significant understanding of industrial needs.

Because of this and our past collaboration, we joined forces after the HPD effort was resuscitated.

8.2 Overestimation of what *any* practitioner could understand; need to develop HPD to a product-like readiness state.

As HPD evolved, we started to train Xerox engineers in short sessions and saw how difficult it was for especially the experimentally-oriented engineers to grasp probabilistic concepts beyond the Normal distribution, standard deviation, and RSS (root sum squares). While HPD had already developed a first stage user-interface by then, we realized that extreme user-friendliness and comprehensiveness would be necessary for them to learn quickly which meant needing to develop HPD to be product-like. And as people in industry know, product development is extremely tedious and time-consuming. That was what we undertook after resuscitating the HPD effort mentioned in (iii) of Subsection 8.1. That our university students – mentioned in (i) of Subsection 8.1 – were able to conduct their HPD-based course projects with only the rudimentary user-interface (at that time) was because they had time to learn, a luxury that practitioners do not have.

SUMMARY; CONCLUSIONS

- HPD has revolutionized Stochastics: Its fundamental breakthrough and its Stochastic Modeling methodology that can yield total sets of dependencies/models enable completely liberating distributions from the very limited discrete set of known types in the continuum of distribution space.
- The focus of this paper, a complement to the conference presentation, is to give an overview of the HPD capabilities and fully clarify what liberating distributions means.
- HPD's user-friendliness will enable both analysis- and experimentally-oriented engineers to be stochastically literate, thus enabling significantly improving engineering efficiency.
- Making HPD available open-source will also lead to much research in the Stochastic community towards future enhancements.
- HPD will impact many fields as all revolutions do. Foremost, it has redefined what Applied Probability should include; thus, at the least, it will affect all existing material, tools, and literature on Applied Probability.
- We anticipate that engineering software suppliers, such as those into CAE, MBSE, PLM tools and include Monte Carlo based capabilities – example corporations are Ansys, Siemens, and Dassault – might:

  i) First replace their current tools with HPD with its infrastructure that includes the liberated distribution capability, and

  ii) Later join forces to create a "home" for the open-source HPD *analogous* to that for Ubuntu – the open-source Linux-based operating system – so that future advancements from researchers in the Stochastics community would contribute to HPD's evolution.

REFERENCES[1]

[1] Parks, JM., 1997, "The two enablers for Holistic Probabilistic Design: Methodology and a complete suite of tools for stochastic analysis." *Proceedings of the Third international Conference on Stochastic Structural Dynamics,* Puerto Rico, pp. 9.1-9.29

[2] Parks, JM., 1996, "Holistic Approach and Advanced Techniques and Tools for Tolerancing and Tools for Tolerance Analysis and Synthesis." *Computer-Aided Tolerancing* (Kimura, F., editor), Chapman & Hall, London, pp. 283-300

[3] Parks, Jean M. and Chun Li, 1998, "Advanced Methodology and Software for Tolerancing and Stochastic Optimization." *Geometric design tolerancing: theories, standards, and applications*, (ElMaraghy, HA., editor), Chapman and Hall, London, pp. 325-336

[4] Parks, JM., 2001, "On stochastic optimization**:** Taguchi Methods™ demystified; its limitations and fallacy clarified," *Probabilistic Engineering Mechanics*, ***16***(1), pp. 87-101

[5] Parks, Jean, Li, Chun, and Noori, Mohammad, 2025, *Fundamental Advancements in Modeling and Analyzing for Failure/Risk/Reliability, Revolutionizing Stochastics, Part A,* in preparation.

[1] Since HPD is our own work, and the material for this paper is being newly shared, the few listed are the key references. Note that those in the "D" side of R&D in Industry, as we mostly have been in, do not typically publish; they keep innovations proprietary.