Would you jump? or the uncertainty of living...

R. Muñoz-Carpena, Ph.D., Professor Agricultural & Biological Engineering

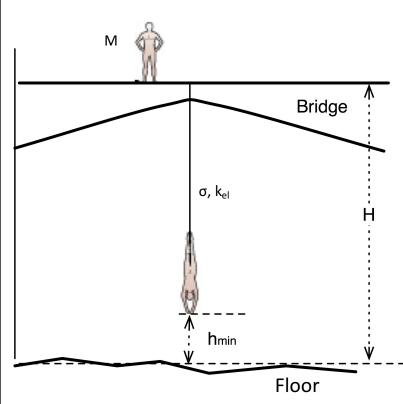


Would you jump?

Class discussion (1)...



Bungee Jumping Model



Simplified solution to linear oscillator equation (string length=0):

$$h_{\min} = H - \frac{2Mg}{k_{el}\sigma}$$

h_{min}= minimum distance to surface [m]
H=distance from platform to surface[m]
M=mass of jumper [kg]
σ= number of strands in cord [-]
k_{el}=elastic constant of material [N/m]≈1.5
g= gravity constant (9.81 m/s²)

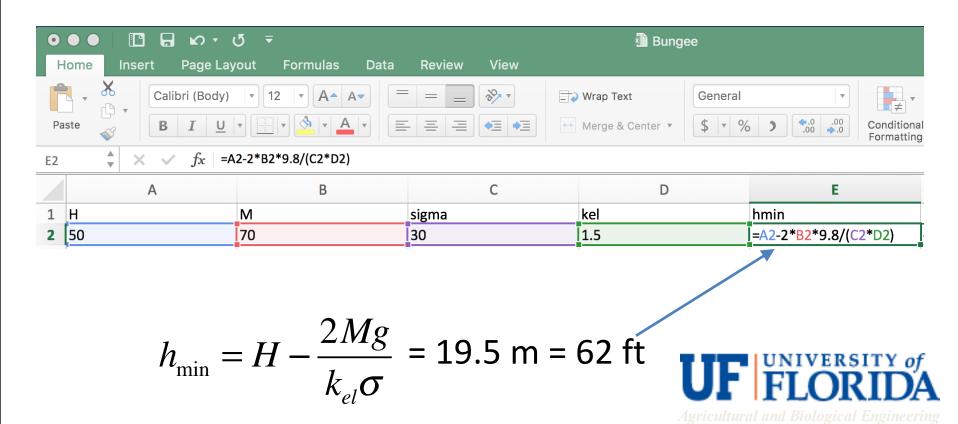




Bungee Jumping Model

Example: a 70 Kg (M) person jumping 50 m (H) with a cord made of 30 strands (σ =30) and elasticity k_{el} = 1.5 N/m

 $\rightarrow h_{min}$ = ? (Let's calculate in Excel!!)



And now... Would you jump?

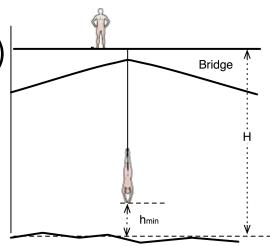
Class discussion redux (2) ...



On the uncertainty of life... Monte-Carlo Uncertainty Analysis

Uncertain **input factors*** when jumping?

- H: U(40,60) (m, bottom variation, topo survey)
- M: U(67,74) (kg, ±5%, physiological)
- σ: U(20,40) ([-], based on <u>vendor survey</u>)
- k_{el}: U(1.475-1.525) (N/m, <u>5% manufacturer</u>)

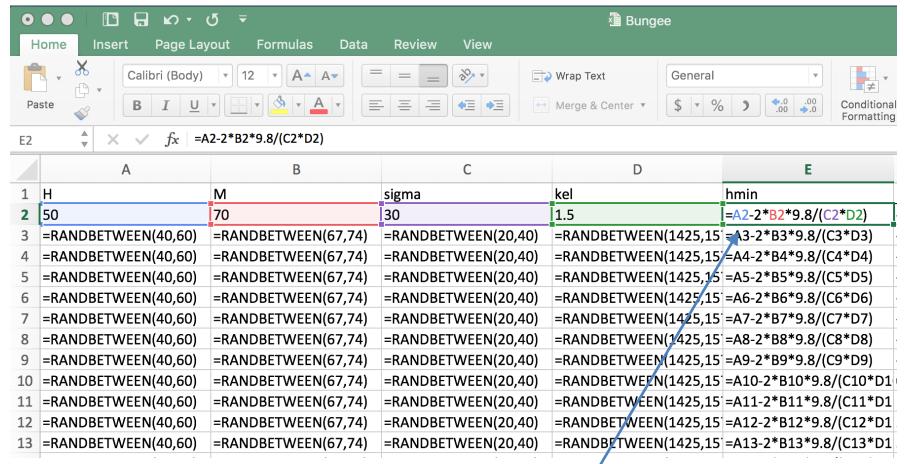


$$h_{\min} = H - \frac{2Mg}{k_{el}\sigma}$$

(*) Input factors: anything that would change the model/system outputs (parameters, initial and boundary conditions)



Let's calculate probable values for each factor and propagate those into the h_{min} model –> Monte Carlo UA



... x 3000 times

$$h_{\min} = H - \frac{2Mg}{k_{el}\sigma}$$



Monte-Carlo Uncertainty Analysis (UA)

-3.6614583

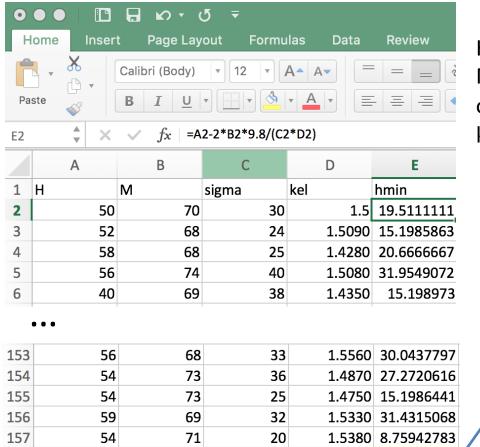
17.7314138

18.951361

1.5360

1.4320

1.4940



20

39

30

H: U(40,60) (m, bottom variation, survey)

M: U(67,74) (kg, ±5%, physiological)

σ: U(20,40) (no. based on vendor survey)

 k_{el} :U(1.475-1.525) (5% manufacturer)

Negative value =





41

43

50

70

72

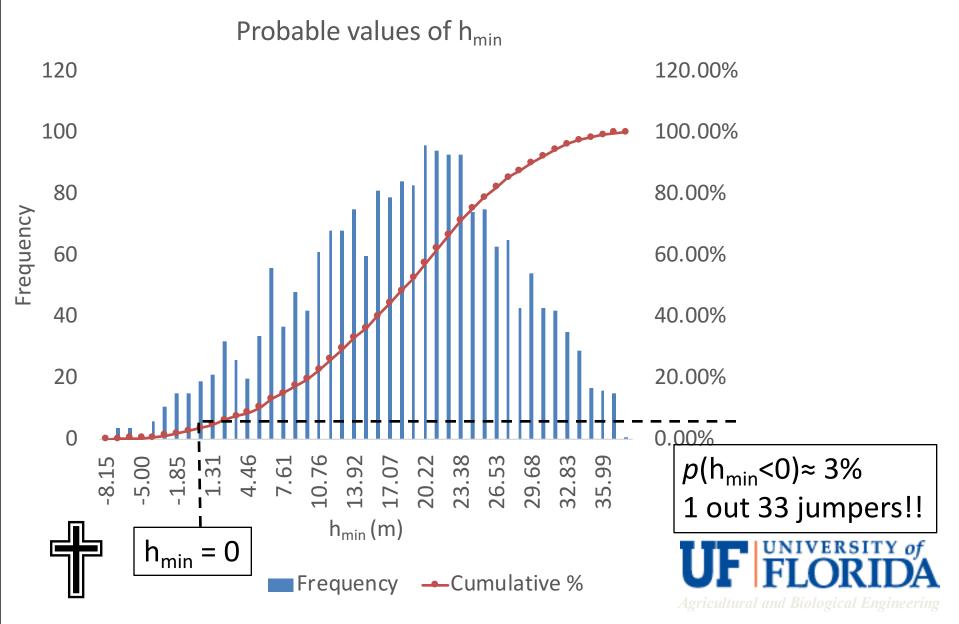
71

158

159

160

Monte-Carlo Uncertainty Analysis (UA)



And now... Would you jump?

Class discussion redux (3)...



Why so much uncertainty?

- What are the important factors controlling it?
- Can we reduce/control uncertainty through management/policy actions?
- Can we control/reduce uncertainty in **complex environmental systems**, with many inputs where the response is more that the sum of the individual components (interactions)?



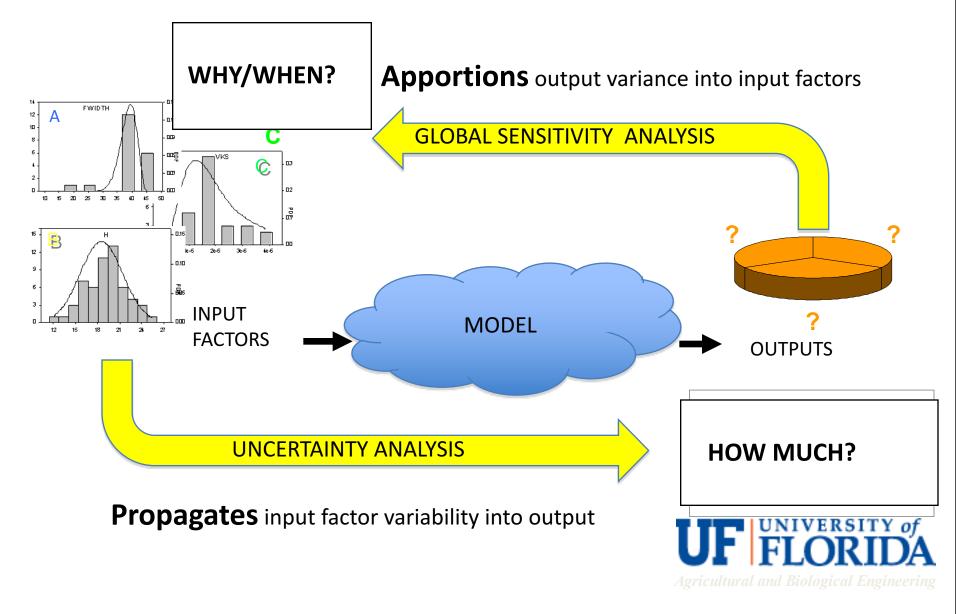
Global Sensitivity & Uncertainty Analysis (GSUA)

Some definitions:

- Mathematical models are built in the presence of uncertainties of various types (input variability, model algorithms, model calibration data, and scale).
- Uncertainty analysis is used to propagate all these uncertainties, using the model, onto the model output of interest
- <u>Sensitivity analysis</u> is used to determine the strength of the relation between a given uncertain input and the output



Global Sensitivity/Uncertainty Analysis



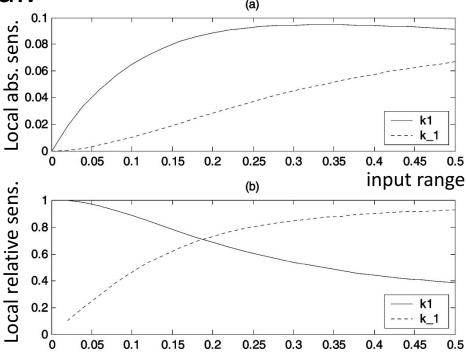
Classical SA approach - derivative local OAT

- Often model sensitivity is expressed as a simple derivative of model output (y) with respect to the variation of a single input factor (x_i) :
 - $-S = \partial y/\partial x_i$ (absolute sensitivity)
 - $-S^{r}=(x_{m}/y_{m})(\partial y/\partial x_{i})$ (relative sensitivity w/means)
 - S^σ = (σ_{xi}/σ_y)(∂y/∂x_i)(relative sensitivity w/uncertainty)
- While the last two measures are unit independent, σ_{ν} requires an uncertainty analysis of the model output.
- These derivative measures can be efficiently computed (direct differentiation of model equations, automatic differentiation algorithms, etc.).
- Local, One-factor-At-a-Time (OAT).



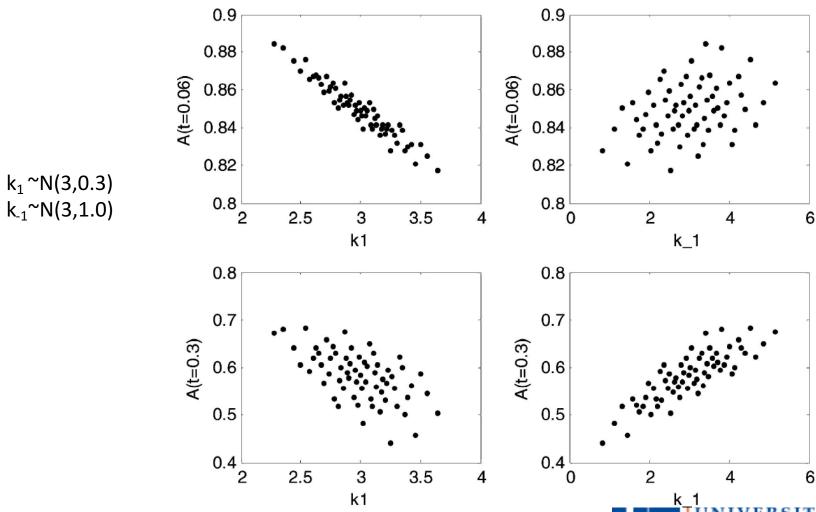
However, these measures only provide *local*information of one-factor-at-a-time (OAT). Often the
derivative is non-linear.

Simple 2-parameter reversible chemical reaction. Values of local sensitivity indexes: (a) absolute; (b) bottom pseudoglobal or normalized by $\sigma_{ki}/\sigma_{[A]}$



 What happens to the more common problem of a model driven by more than one factor with varying effects across the input range?

Dependency of sensitivity with time



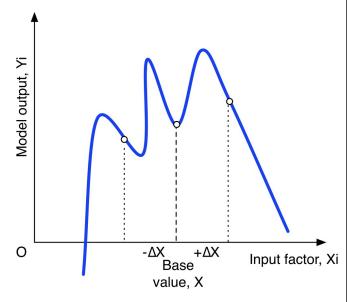
Scatter plots of [A] versus k_1 , k_1 at time t = 0.3 and t = 0.06. k_1 appears mo k_1 at t = 0.06, while the reverse is true at t = 0.3

Agricultural and Biological Engineering

Global Sensitivity Analysis

Local vs. global sensitivity analysis (SA)

	Local SA (classic)	Global SA
Model	Linear Monotonic additive	No assumptions
No. of factors	O-A-T	All together
Factor range	Local (derivative)	Whole PDF
Interactions	No	Yes





Global Sensitivity Analysis

HOW MUCH, WHY, WHEN...



- surprise the analyst,
- find technical errors in the model,
- gauge model adequacy and relevance,
- identify critical regions in the space of the inputs (including interactions),
- establish priorities for research,
- simplify models,
- verify if policy options make a difference or can be distinguished.
- anticipate (prepare against) falsifications of the analysis

• ...



Potentially large input set.

How to handle the model evaluation with many inputs at the time?

 Effect of model structure complexity on sensitivity of inputs.

Often we can have several choices of model formulation, each adding, removing or simply changing the conceptual basis of existing components. What is the effect the model output and other inputs in the model of this changes in the model structure?

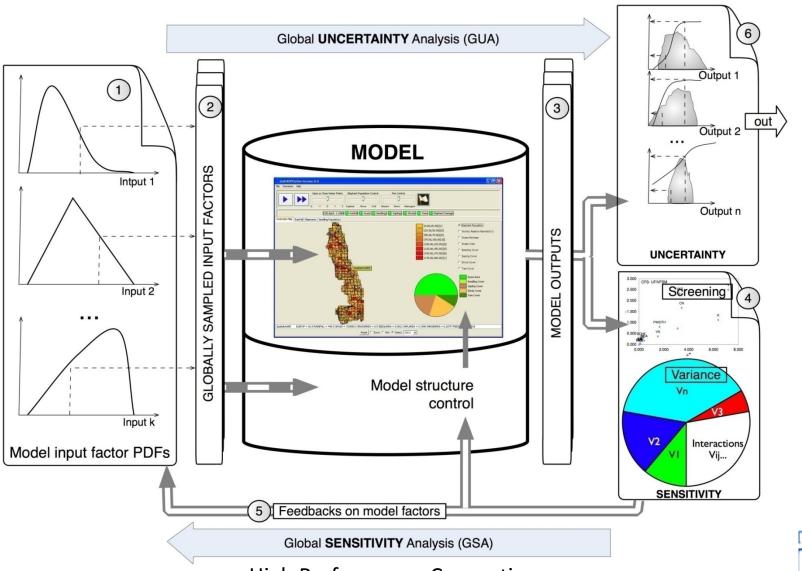
GSUA evaluation framework

For model with large number of factors a two-step global process is recommended:

- 1. Qualitative Screening with limited number of simulation (p.ej. Morris Method)
 Ranking and selection of important factors (μ^*); Presence of interactions (σ)
- **2. Quantitative Variance-based** method: (i.e. Extended FAST, Sobol, etc.) First order indexes (S_i , direct effects); Total order (S_{Ti} , interactions) + Uncertainty analysis (!)



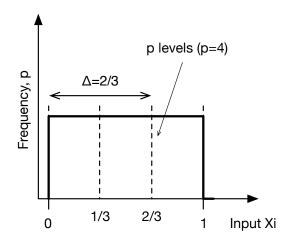
GSUA Evaluation Framework

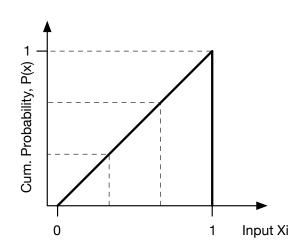


High Performance Computing

- Morris (1991) proposed conducting individually randomized experiments that evaluate the effect of changing one input at a time globally.
- Each input assumes a discrete number of values, called levels, that are selected from within an allocated probability density function (PDF) for the inputs.
- Uses few simulations to map relative sensitivity
- It's Qualitative, not quantitative (low sampling). Provides an early indication of the importance (ranking) of first order effects vs. interactions
- Identifies a subset of more important inputs (could be followed by quantitative analysis)

The key to Morris sampling is that it is based on the "unit hypercube", i.e. every input factor x_i is always sampled from a uniform distribution U[0,1], regardless of its actual distribution (like normal, lognormal, beta, etc.) assigned to that factor (in the ".fac" file). The reason for this is that the unit uniform distribution has the very nice property that the value of the factor is equal to its cumulative probability value:





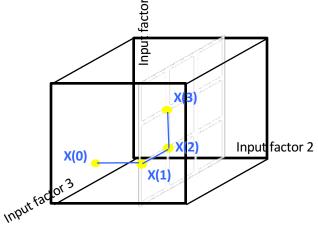
Notice that to limit sampling, the U[0,1] is only sampled at a few places ("levels", where p=number of levels), and the sampling jump Δ across levels is $\Delta = p/(2 p - 2)$ (i.e. two levels in each jump).



The "unit uniform" is a **nice trick**! It means that we can sample all the x_i factors (i=1,k no. factors) at once from a Cartesian k-dimensional hyperspace, each between $\{0,1\}$, and then back-transform them into their actual value based on the inverse probability function of the actual distribution assigned to that value. This is,

Sample U(0,1) =
$$P(x_i)$$
 -> $x_i = P^{-1}(x_i)$

where $P(x_i)$ is the actual cumulative probability assigned to the factor (normal, lognormal, etc.) and $P^{-1}(x_i)$ is inverse probability function for that distribution.



The sampling strategy on the k-dimensional unit-hypercube space is to follow a 'trajectory" with as many "turns" as dimensions (k), each turn modifying by a jump Δ in *only one of the coordinates*. The input space is then evenly sampled with a number of trajectories (r) with as much separation from each other as possible.

Number of model simulations N = r (k + 1)

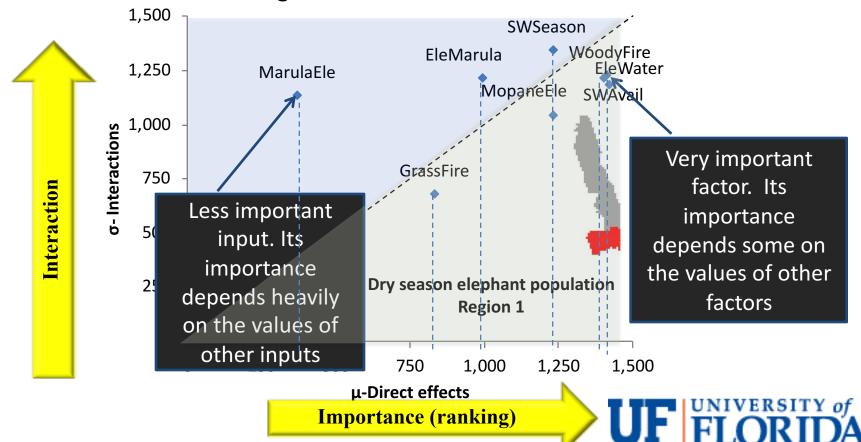
Example: k=20, *r*=10, N = 210

• The elementary effects (EE_i) are calculated for each factor x_i as derivatives for each jump Δ in the trajectories,

$$EE_i = [y(x_1, x_2...x_i, x_{i+1}, ...x_k) - y(x_1, x_2...x_i + \Delta, x_{i+1}, ...x_k)]/\Delta$$

- For each input x_i , two sensitivity measures are proposed by Morris (1991):
 - 1. μ : the mean of EE_i, which estimates the overall *direct effect* of the input on a given output
 - 2. σ : the standard deviation of EE_i, which estimates the higher-order characteristics of the input (such as curvatures and *interactions*).
- Since the model output could be non-monotonic, Campolongo and others (2003) suggested considering the absolute values of the elementary effects, μ^* , to avoid the effects of canceling EE_i of opposite signs.

- A very nice feature of Morris is the graphical representation of the factor importance in the "Morris plane" where input factors are plotted on (μ_i^*, σ_i) axes. μ^* = average of the $|EE_i|$
- The factors closest the origin are less influential!!



HDMR*: Variance decomposition Fourier Amplitude Sensitivity Analysis Test (FAST)

• FAST, in a nutshell, decomposes the output (Y) variance $V = s_Y^2$ using spectral analysis, so that $V = V_1 + V_2 + ... + V_k + R$, where V_i is that part of the variance of Y that can be attributed to x_i alone, k is the number of uncertain factors, and R is a residual. Thus, $S_i = V_i/V$ can be taken as a measure of the sensitivity of Y with respect to x_i .

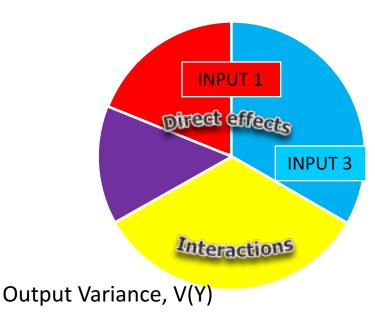
•
$$V(Y) = \sum V_i + \sum V_{ij} + \sum V_{ijl} + ... + V_{123...k}$$

 FAST is a GSA method which works irrespective of the degree of linearity or additivity of the model.



HDRM*: Variance decomposition Fourier Amplitude Sensitivity Analysis Test (FAST)





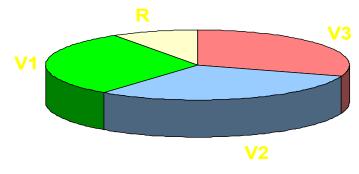


(*) HDMR: High-Dimension Model Representation

HDRM: Variance decomposition Fourier Amplitude Sensitivity Analysis Test (FAST)

$$V(Y) = V_1 + V_2 + ... + V_k + R$$

V(Y) – variance of output, V_i – variance due input factor X_i , k – nr of uncertain factors, R - residual



Number of FAST model simulations $N = N \approx M k$; where $M = 2^b$, b = 9 to 10, M = 512 to 1024 Example: k = 20, N = 10240 to 20480



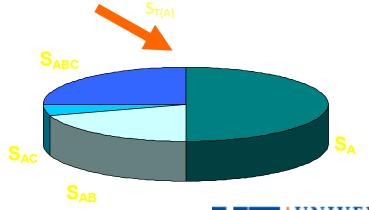
HDRM: Variance decomposition Fourier Amplitude Sensitivity Analysis Test (FAST)

- 1. S_i first-order sensitivity index: $S_i = V_i / V(Y)$
- 2. S_{T(i)} total sensitivity index

For model with 3 inputs: A, B, and C, for input A:

$$\mathbf{S}_{\mathsf{T}(\mathsf{A})} = \mathsf{S}_{\mathsf{A}} + \mathsf{S}_{\mathsf{AB}} + \mathsf{S}_{\mathsf{AC}} + \mathsf{S}_{\mathsf{ABC}}$$

$$S_{Ti} - S_i = higher order effects$$





- Although FAST is a sound approach to the problem, it has seen little used in the scientific community at large.
- If specific decompositions of interactions is needed, Sobol further proposed:

$$Y = f(X_1, X_2, ..., X_k) = f_0 + \sum_i f_i(X_i) + \sum_i f_{ij}(X_i, X_j) ... + f_{12...k}$$

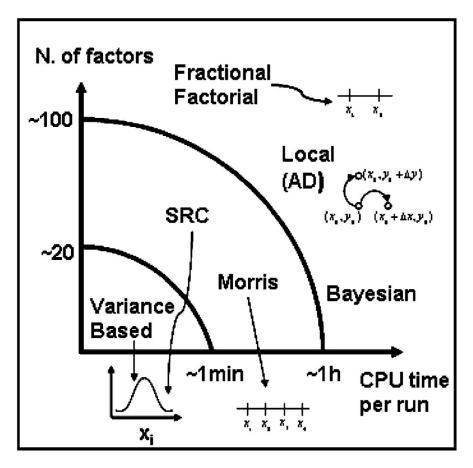
Number of Sobol model simulations $N = N \approx M (2k + 2)$; where $M = 2^b$, b=8-12, M=256-4096) Example: k=20, N=10,752 to 172,032 (typical 21504)

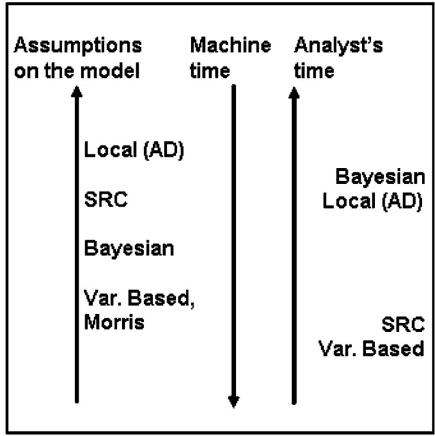


- Using variance-based techniques in numerical experiments is the same as applying ANOVA (analysis of variance) in experimental design, as the same variance decomposition scheme holds in the two cases.
- One could hence say that modelers are converging with experimentalists treating Y, the outcome of a numerical experiment, as an experimental outcome whose relation to the control variables, the input factors, can be assessed on the basis of statistical inference.



Choice of SA methods: Tradeoffs



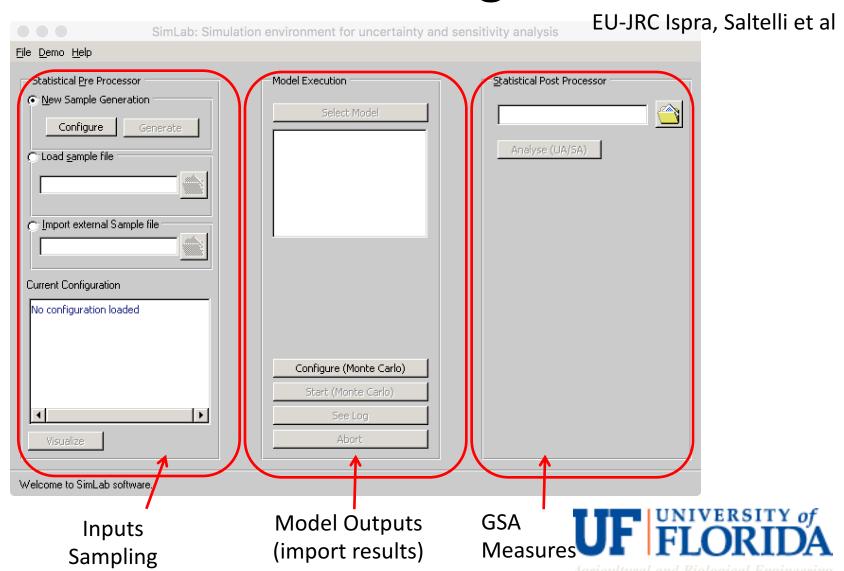




Let's find out the most important input factors controlling the bungee jumping risk ... Morris GSA



Simlab GSA teaching software



GSA bungee jumping

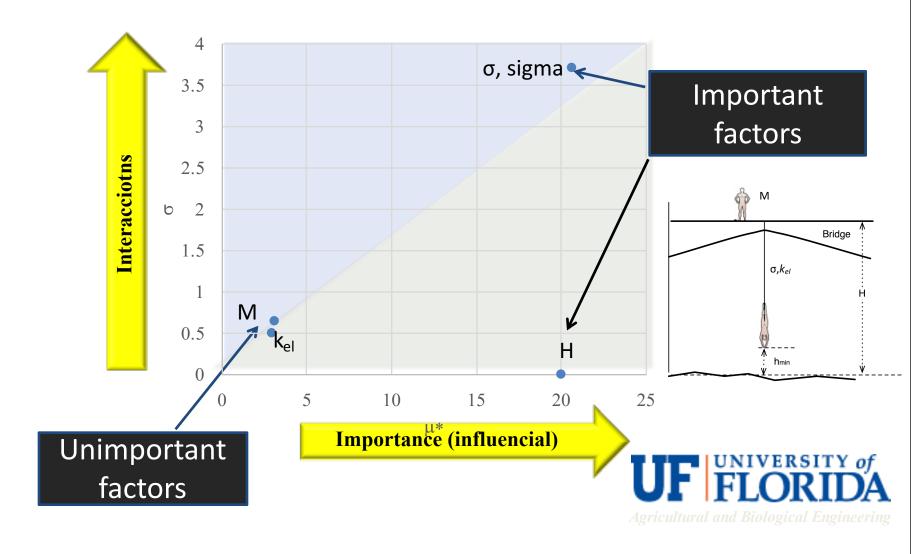
Apply GSA methods of Morris (and FAST). Follow these steps:

- 1. In SimLab (EU-JRC) teaching software define the 4 input factors (H, M, σ , k_{el})
- 2. Copy the inputs sample matrix into the Bungee Excel sheet and calculate the model output vector (h_{min}) .
- 3. Bring back the input matrix and output vector and process the results in Simlab



GSA Morris Results: Bungee jumping

Selection of important factors (μ^*); Presence of interactions (σ)



Risk control: Monte Carlo Filtering

Discussion:

- How could we reduce risk when jumping?
- What factor(s) should we design our risk control management strategy around?
- How?

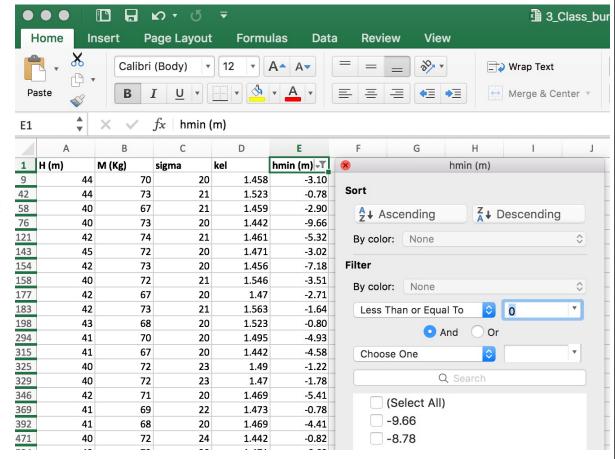




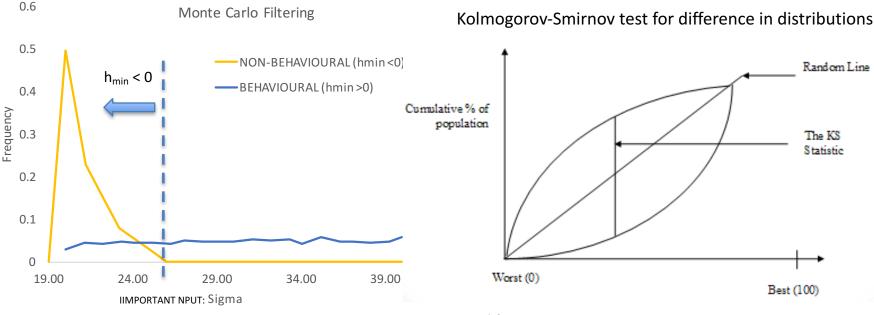
Risk control: Monte Carlo Filtering

Let's MCF the UA outputs as "non-behavioral" ($h_{min} \le 0$) and "behavioral" ($h_{min} > 0$) and "map" the most important (and manageable) input factor (σ).

- Copy the h_{min} values from the Uncertainty Analysis sheet
- 2. Paste "as values" into a new sheet
- Filter the data in two subsets (h_{min} ≤ 0 and h_{min} < 0)]
- Copy each of the subsets into another spreadsheet and create histograms for σ from each subset.

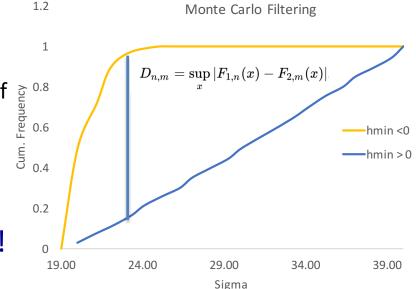


Risk control: Monte Carlo Filtering



The Kolmogorov-Smirnov test (see Wikipedia for calculation) is used to test if behavioral and non-behavioural subsets are statistically different and if so, MCF identified a candidate for an intervention (i.e. reduce the initial variability through some action). In our case sigma: U[25,40]

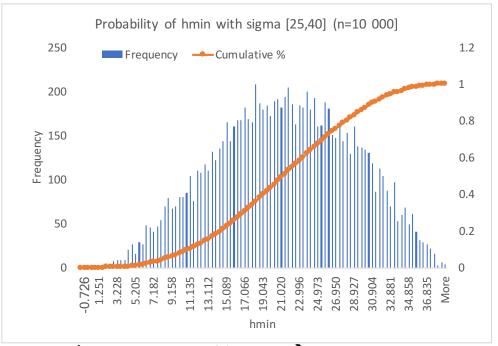
So, $h_{min} < 0$ only when $\sigma < 25$ strands!!



Risk control: increase σ !

Through policy, regulation and enforcement the mininum number of bungee cord strands allowed is now

increased from 20 to 25.

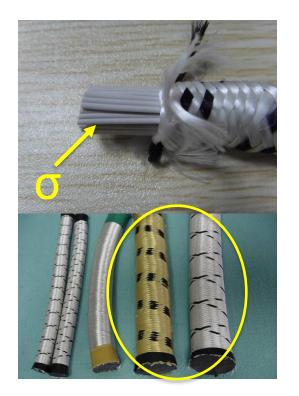


 $p(h_{min}<0) < 0.001\%!! \rightarrow 1:100,000$

Traffic accident deaths \rightarrow 1:32,000

"Normal" cancer rate \rightarrow 1:30,000

Death by lightning strike → 1:1,000,000





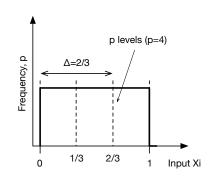
And now... Would you jump?

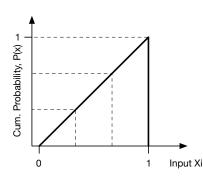
Class discussion redux (4)...



- Researchers after Morris found that the original trajectory-based sampling does not guarantee a uniform sampling (like on the left of the Figure).
- Instead, one gets somewhat irregular sampling across the levels. Although this irregularity does not affect most of the problems, for difficult outputs (highly non-linear or non-continuous across the range), this implies inaccuracies in the ranking of factors.
- We propose the SU (Sampling for Uniformity) (Khare et al., 2015) (and its enhanced version eSU that requires no oversampling, Chitale et al., 2017) that guarantees uniformity.
- This translates into accurate input ranking even for the most demanding test functions, like those tested on Khare et al., (2015) and at a much lower computational cost than other improved methods proposed.

Khare, Y.P., Muñoz-Carpena, R., Rooney, R.W., Martinez, C.J. A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects. *Environmental Modelling & Software* 64:230-239. doi:10.1016/j.envsoft.2014.11.013.





- We currently only use Khare et al. (2015) eSU sampling as the best and most efficient method for sampling with the Matlab package (http://abe.ufl.edu/carpena/software/SUMorris.shtml).
- We included a user interface 'sampler.m' to run a single sampling or in mixed scripts (like unix) after compilation of the Matlab program. Typical settings (recommended) for eSU are no oversampling (1 sample), 8 levels to get a more spread sampling across the range, 16 trajectories, and write the outputs as a text file for later processing, i.e.

```
>> sampler 'bungee4.fac' 'eSU' 1 8 16 'Text'
```



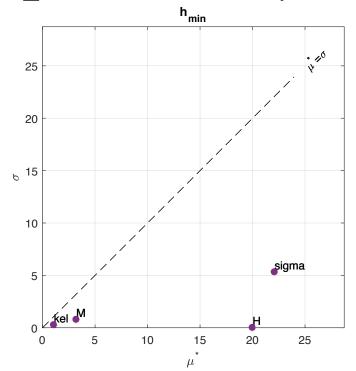
Let's repeat the bungee analysis with eSU in Matlab.

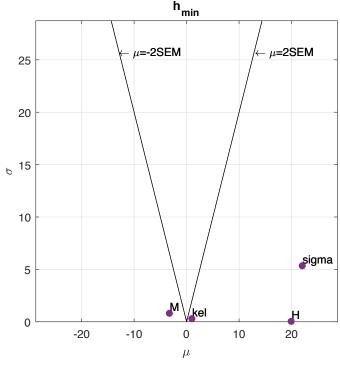
- 1. Download and uncompress the EE Sampling and EE Measures packages from http://abe.ufl.edu/carpena/software/SUMorris.shtml
- 2. Copy the 'bungee4.fac' file in the directory EE_Sampler.
- 4. This produces a CSV text sample file 'bungee4_FacSample.sam'. Open the file in Excel and paste the samples in the bungee spreadsheet.
- 5. Create a new Excel output file ('bungee4_outputs.xlsx') with outputs in a single column, and the output label in the first row (hmin).
- 6. Copy the files bungee4_FacSample.sam, bungee4_outputs.xlsx and bungee4_FacSamChar.txt (also created in the EE_sampler directory after the run) and paste them into the EE_Measures directory.
- 7. In Matlab change to the EE_Measures directory and run,

```
>>EE_SenMea_Calc('bungee4_FacSample.sam','bungee4_FacSamChar.txt','bungee4_outputs.xlsx')

UNIVERSITY of FLORIDA
```

8. The Morris plot (modified and original) are presented on the screen during the run, the graphical (bungee4_Fig1.pdf) and text (bungee4_EE_Measures.txt) outputs are written in the EE_Measures directory,





The comparison of these graphs offers useful information on dominance of interactions and monotonicity of effects.

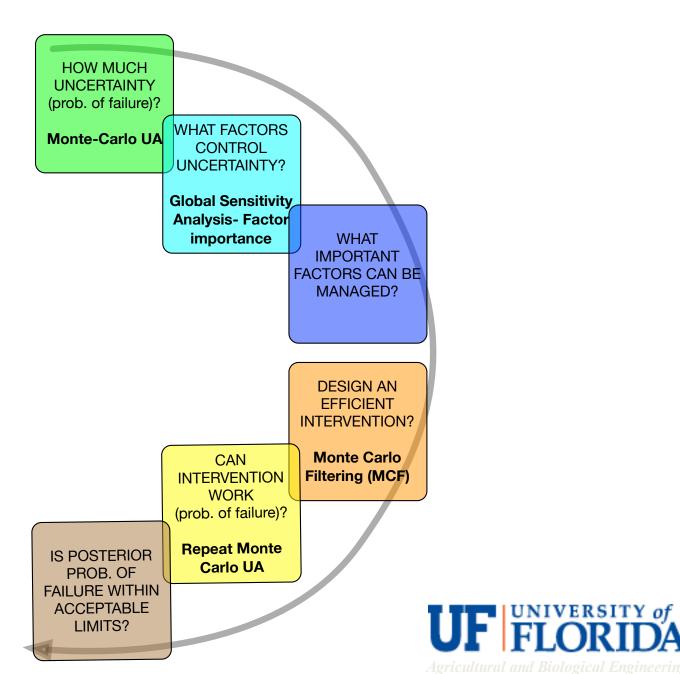


Wrap up... What did we learn?



Wrap up- what did we learn?

- Uncertainty –intrinsic & conditions management decisions
- Models allow objective assessment of system responses to inform decisions (only as good as the model!).
- Uncertainty Analysis (UA) helps to quantify how much uncertainty and risk associated.
- Global Sensitivity Analysis (GSA) serves to identify important risk factors.
- Monte Carlo Filtering (MCF) allows to identify ranges
 of important factors for effective policy/intervention
 strategies.

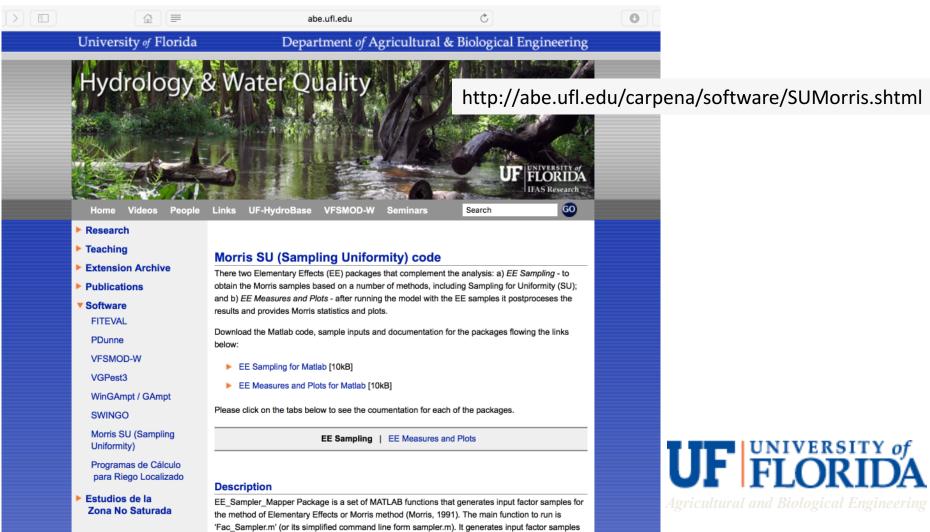


Further reading:

- Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo (2005), Sensitivity analysis for chemical models, *Chemical Reviews*, 105, 2811-2827.
- Saltelli., A., S. Tarantola, F. Campolongo, M. Ratto. 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Wiley: London.
- Beven, K (2006), On undermining the science? *Hydrological Processes*, 20:1-6.
- Chu-Agor, M.L., R. Muñoz-Carpena, G. Kiker, A. Emanuelsson and I. Linkov. 2011. Exploring sea level rise vulnerability of coastal habitats through global sensitivity and uncertainty analysis. *Env. Model. & Software* 26(5):593-604. doi:10.1016/j.envsoft.2010.12.003.
- Khare*, Y.P., R. Muñoz-Carpena,. R.W. Rooney. and C.J. Martinez. 2015. A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects. Environmental Modelling & Software 64:230-239. <a href="https://doi.org/d
- Muñoz-Carpena, R., Z. Zajac and Yi-Ming Kuo. 2007. <u>Global Sensitivity and uncertainty Analyses of the Water Quality Model VFSMOD-W[268KB]</u>. <u>Trans. of ASABE 50(5):1719-1732</u>
- Perz, S., R. Muñoz-Carpena, G.A. Kiker, R.D Holt. 2013. Evaluating ecological of resilience with global sensitivity and uncertainty analysis. Ecological Engineering Modelling 263:174-186. doi:10.1016/j.ecolmodel.2013.04.024 Biological Engineering

Morris SU sampling: Matlab tool

Khare, Y.P, Muñoz-Carpena, R., Rooney, R.W., Martinez, C.J. A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects. *Environmental Modelling & Software* 64:230-239. doi:10.1016/j.envsoft.2014.11.013



in a unit hyperspace and then transforms them according to the specified input probability