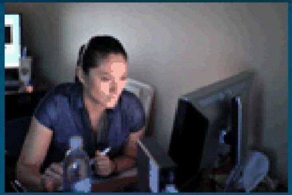
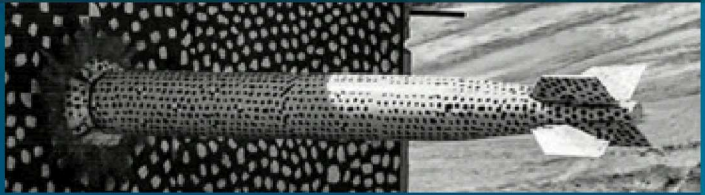
SAND2019-10414PE

Introduction to the Basics of Uncertainty Quantification





PRESENTED BY

Erin C.S. Acquesta

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

1. Motivating Uncertainty Quantification
   - "All Models are Wrong"
   - UQ in the Real World: Dorian
   - Predictive Science and UQ

2. Fundamentals of Probability
   - Frequentist vs Bayesian
   - Probability Space and Random Variables

3. Coin Toss Experiment

All Models are Wrong, But Some are Useful

- George Box

All Models are Wrong, But Some are Useful

- George Box

“Modelling in science remains, partly at least, an art. Some principles do exit, however, to guide the modeler. The first is that *all models are wrong;* some, though are better than others and we can search for the better ones. **At the same time we must recognize that eternal truth is not within our grasp.”**

- McCullagh and Nedler (1989)

“… it does not seem helpful just to say that all models are wrong. The very word model implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd.”

- Cox (1995)

“A model is a simplification or approximation of reality and hence will not reflect all of reality.”

- Burnham and Anderson (2002)

“In general, when building statistical models, we must not forget that the aim is to understand something about the real world. Or predict, choose an action, make a decision, summarize evidence, and so on, but always about the real world, no an abstract mathematical world: our models are not the reality”

- Hand (2014)

* All above quotes were pulled from the Wikipedia page for *“All models are wrong” on 2019/09/01*

All Models are Wrong, But Some are Useful

- George Box

"Modelling in science remains, partly at least, an art. Some principles do exit, however, to guide the modeler. The first is that *all models are wrong;* some, though are better than others and we can search for the better ones. **At the same time we must recognize that eternal truth is not within our grasp."**

- McCullagh and Nedler (1989)

"… it does not seem helpful just to say that all models are wrong. The very word model implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd."

- Cox (1995)

"A model is a simplification or approximation of reality and hence will not reflect all of reality."

- Burnham and Anderson (2002)

"In general, when building statistical models, we must not forget that the aim is to understand something about the real world. Or predict, choose an action, make a decision, summarize evidence, and so on, but always about the real world, no an abstract mathematical world: our models are not the reality"

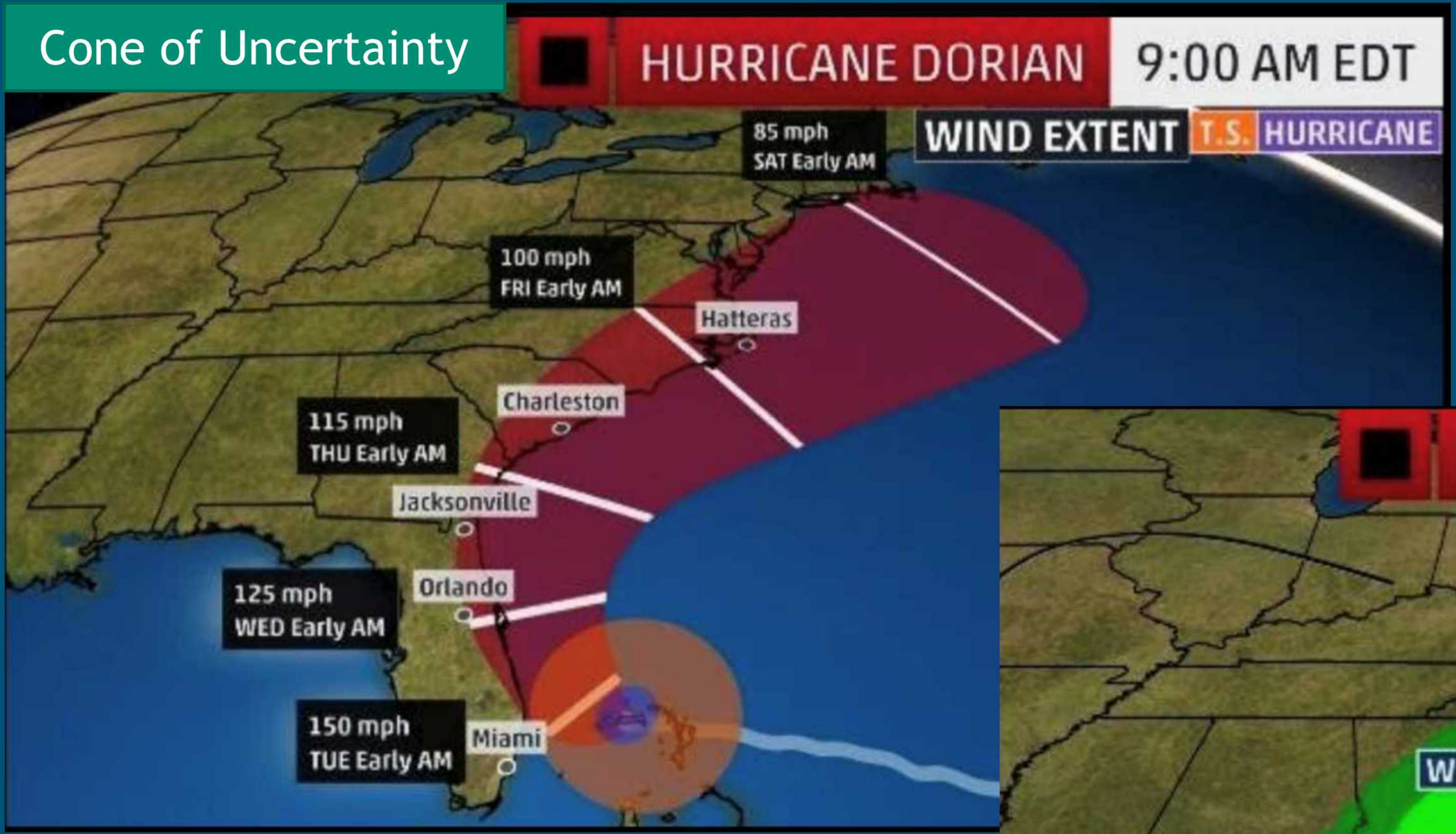
- Hand (2014)

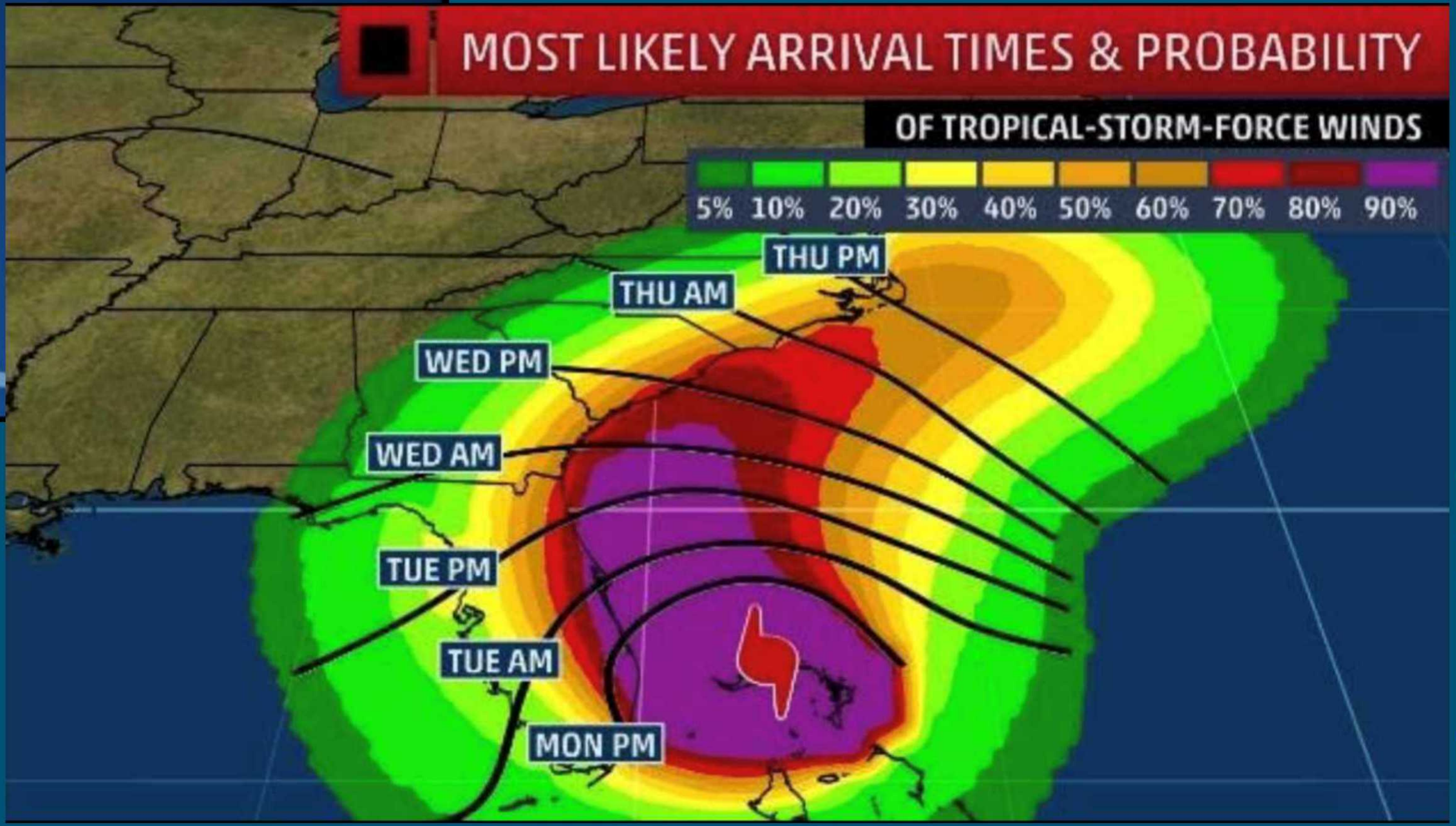
Why do we even bother then?

* All above quotes were pulled from the Wikipedia page for *"All models are wrong" on 2019/09/01*

Dorian: Hurricane Trajectory Projections
as of 9:00AM EDT on Monday 09/02/2019

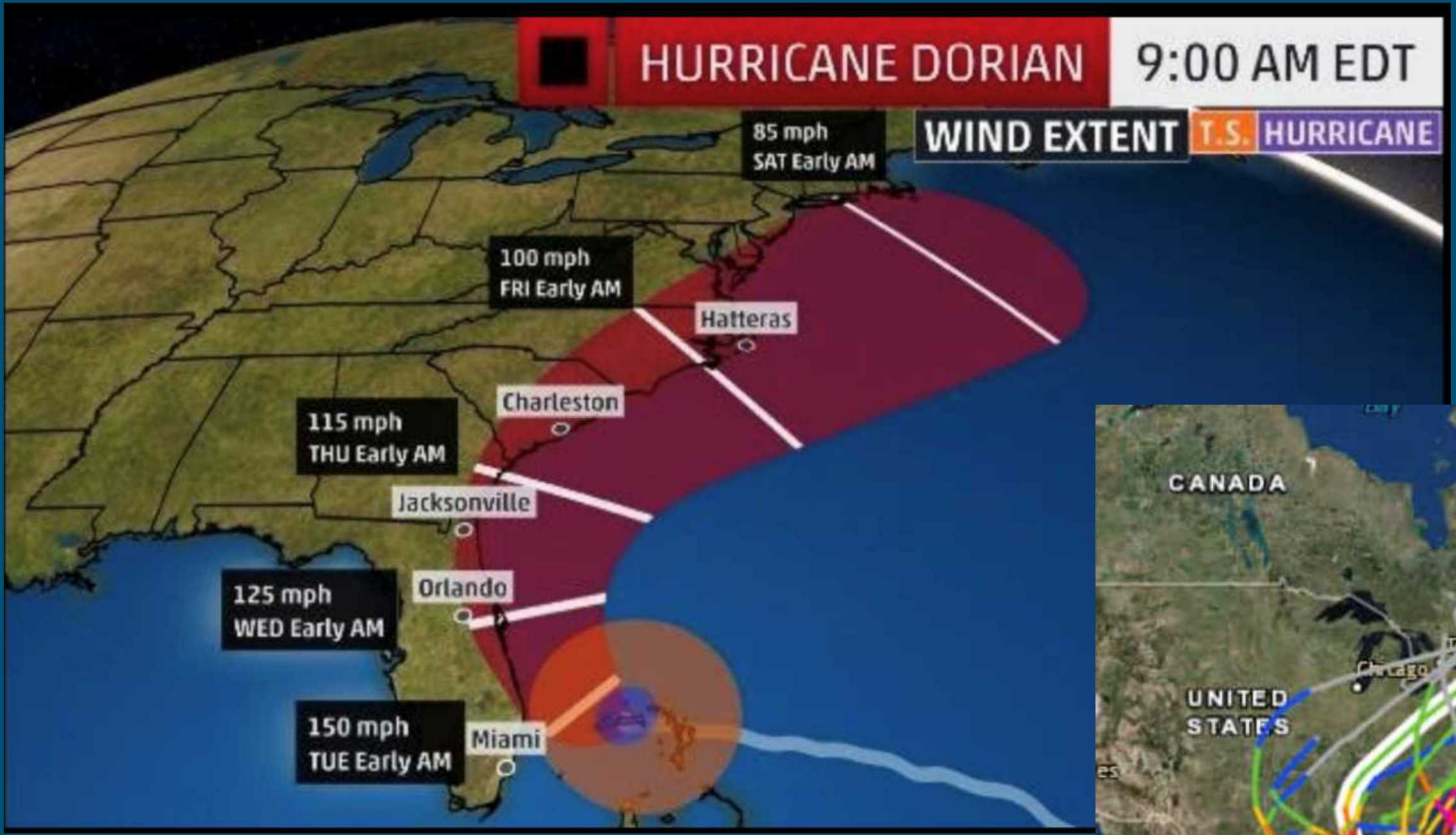
Cone of Uncertainty





* https://www. weather.com/amp/storms/hurricane/news/2019-09-02-hurricane-dorian-labor-day-bahamas-florida-georgia-carolinas.html

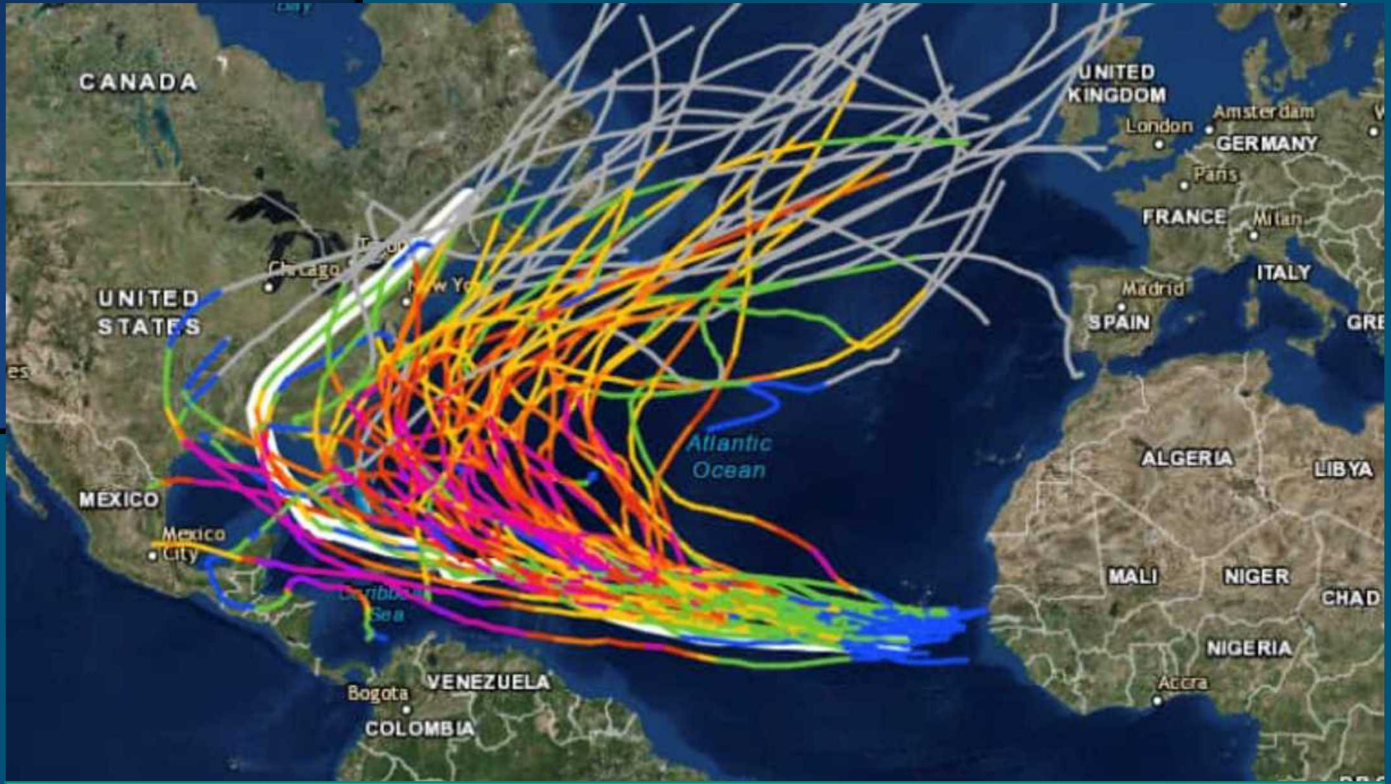
How do the weather forecasters make these projections?



* https://www. weather.com/amp/storms/hurricane/news/2019-09-02-hurricane-dorian-labor-day-bahamas-florida-georgia-carolinas.html

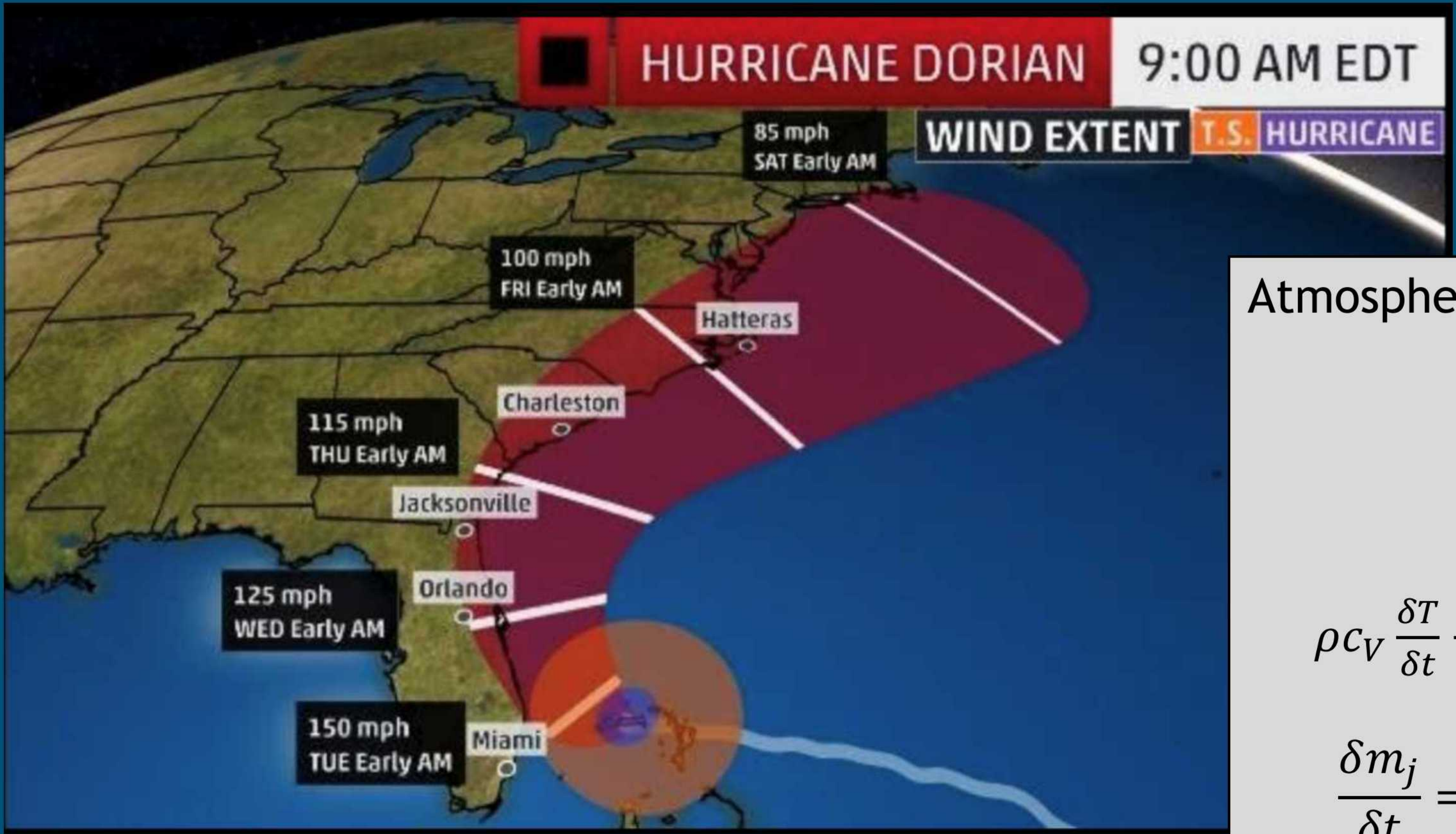
Historical Data

* https://oceanservice.noaa.gov/news/historical-hurricanes/



>150 years of category 4 and 5 hurricane tracking data

How do the weather forecasters make these projections?



* https://www. weather.com/amp/storms/hurricane/news/2019-09-02-hurricane-dorian-labor-day-bahamas-florida-georgia-carolinas.html

Mathematical Models

Atmospheric Physics Equations:

$$\frac{\delta\rho}{\delta t} + \nabla\cdot(\rho v) = 0,$$

$$\frac{\delta v}{\delta t} = -v\cdot\nabla v - \frac{1}{\rho}\nabla\rho - g\hat{k} - 2\Omega\times v,$$

$$\rho c_V\frac{\delta T}{\delta t} + \rho\nabla\cdot v = -\nabla\cdot F + \nabla\cdot(k\nabla T) + \rho\dot{q}(T,p,\rho),$$

$$p = \rho RT,$$

$$\frac{\delta m_j}{\delta t} = -v\cdot\nabla m_j + S_{m_j}(T,m_j,\chi_j,\rho), \qquad j = 1,2,3,$$

$$\frac{\delta\chi_j}{\delta t} = -v\cdot\nabla\chi_j + S_{\chi_j}(T,\chi_j,\rho), \qquad j = 1,\dots,J$$

Where $\rho, v, T, p, k,$ and c_V respectively denote the density, velocity, temperature, pressure, thermal conductivity, and specific heat of air

* Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014. page: 2

…Because "Some are Useful"

…but…How do we know a model is "Useful"?

…Because "Some are Useful"

…but…How do we know a model is "Useful"?

We consider a model "useful" if:

- It provides enhanced insight into a problem.
- It provides a means to test theory/hypothesis

…Because "Some are Useful"

…but…How do we know a model is "Useful"?

We consider a model "useful" if:

- It provides enhanced insight into a problem.
- It provides a means to test theory/hypothesis

How do we determine if a model is "useful"?

- Traditional Methods: Validation and Verification

…Because "Some are Useful"

…but…How do we know a model is "Useful"?

We consider a model "useful" if:

- It provides enhanced insight into a problem.
- It provides a means to test theory/hypothesis

How do we determine if a model is "useful"?

- Traditional Methods: Validation and Verification

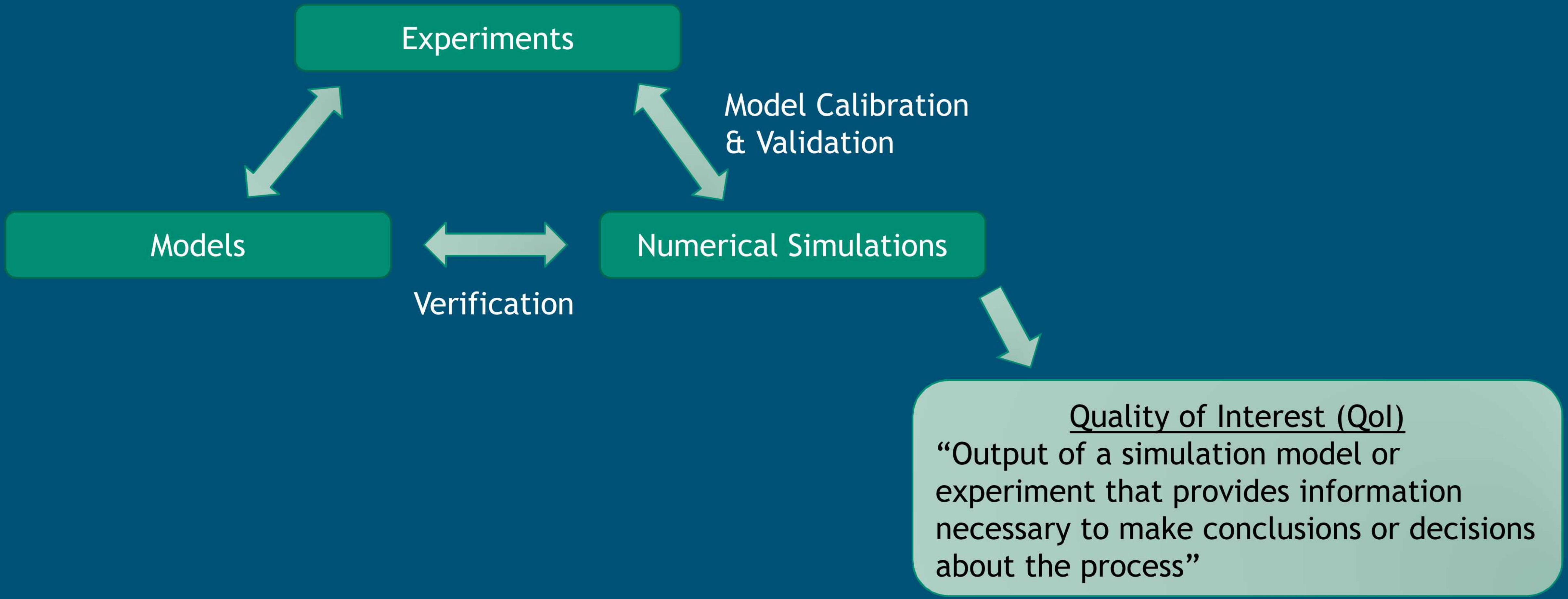
Validation: "describes the process of determining the accuracy with which mathematical models quantify the physical process of interest"[1]

Verification: "refers to the process of quantifying the accuracy of simulation codes used to implement mathematical models" [1]

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Predictive Science

The following introduction to Uncertainty Quantification (UQ) is a summary of what is provided by Dr. Ralph Smith in his 2014 SIAM publication, *Uncertainty Quantification: Theory, Implementation, and Applications. [1]*



[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

What is Uncertainty Quantification?

"Uncertainty quantification is both a new field and one that is as old as the disciplines of probability and statistics."
-Smith (2014)

Uncertainty Quantification: "[in the context of predictive science] uncertainty quantification is the science identifying, quantifying, and reducing uncertainties associated with models, numerical algorithms, experiments, and predicted outcomes or quantities of interest."

Multi-disciplinary Field:

- Probability
- Statistics
- Analysis
- Numerical Analysis

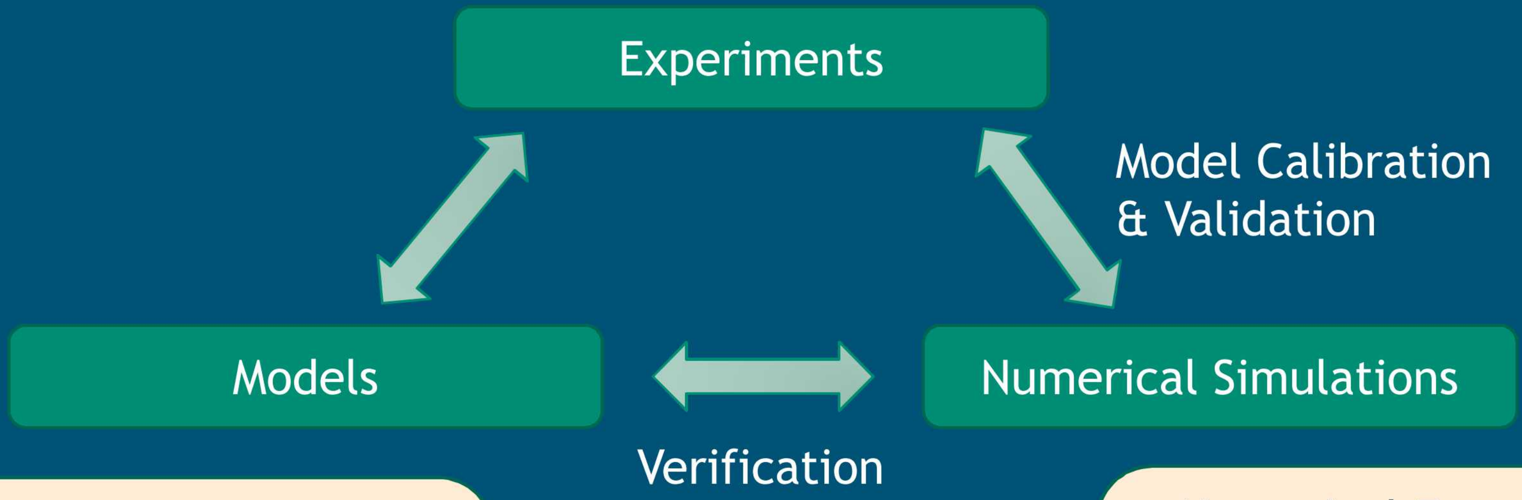
Areas of Research Interest:

- Parameter Selection
- Surrogate Model Construction
- Local & Global Sensitivity Analysis
- Quantification of Model Discrepancies

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Uncertainty in Predictive Science

<u>Experimental Uncertainties and Limitations</u>
- Limited or incomplete data
- Limited accuracy or resolution of sensors



<u>Model and Input Uncertainties</u>
- Model errors or discrepancies
- Input uncertainties due to uncertain parameters, forcing functions, and initial & boundary conditions

<u>Numerical Errors and Uncertainties</u>
- Round off, discretization, or approximation errors
- Bugs or coding errors
- Bit-flip and hardware failures

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Uncertainty in Predictive Science

Experimental Uncertainties and Limitations
- Limited or incomplete data
- Limited accuracy or resolution of sensors

Experiments

Model Calibration & Validation

"Essentially, all models are wrong, but some are useful"
-Box

Models

Numerical Simulations

Verification

Model and Input Uncertainties
- Model errors or discrepancies
- Input uncertainties due to uncertain parameters, forcing functions, and initial & boundary conditions

Numerical Errors and Uncertainties
- Round off, discretization, or approximation errors
- Bugs or coding errors
- Bit-flip and hardware failures

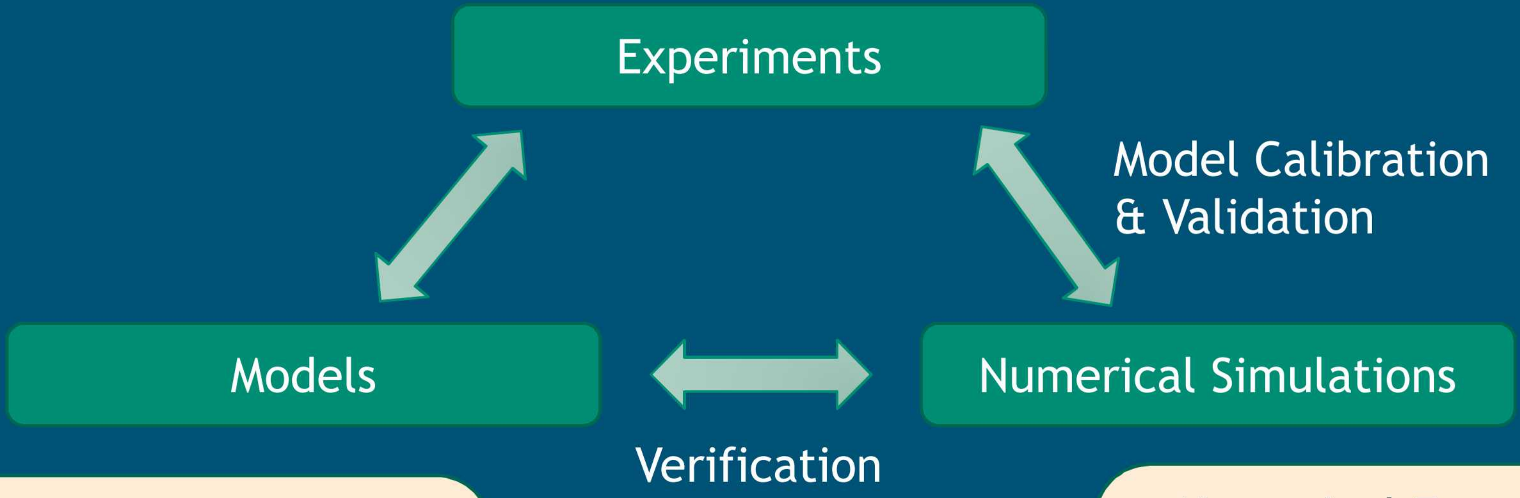
[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Uncertainty in Predictive Science

<u>Experimental Uncertainties and Limitations</u>

- Limited or incomplete data
- Limited accuracy or resolution of sensors

"Experimental results are believed by everyone, except for the person who ran the experiment"
-Gunzburger



<u>Model and Input Uncertainties</u>

- Model errors or discrepancies
- Input uncertainties due to uncertain parameters, forcing functions, and initial & boundary conditions

<u>Numerical Errors and Uncertainties</u>

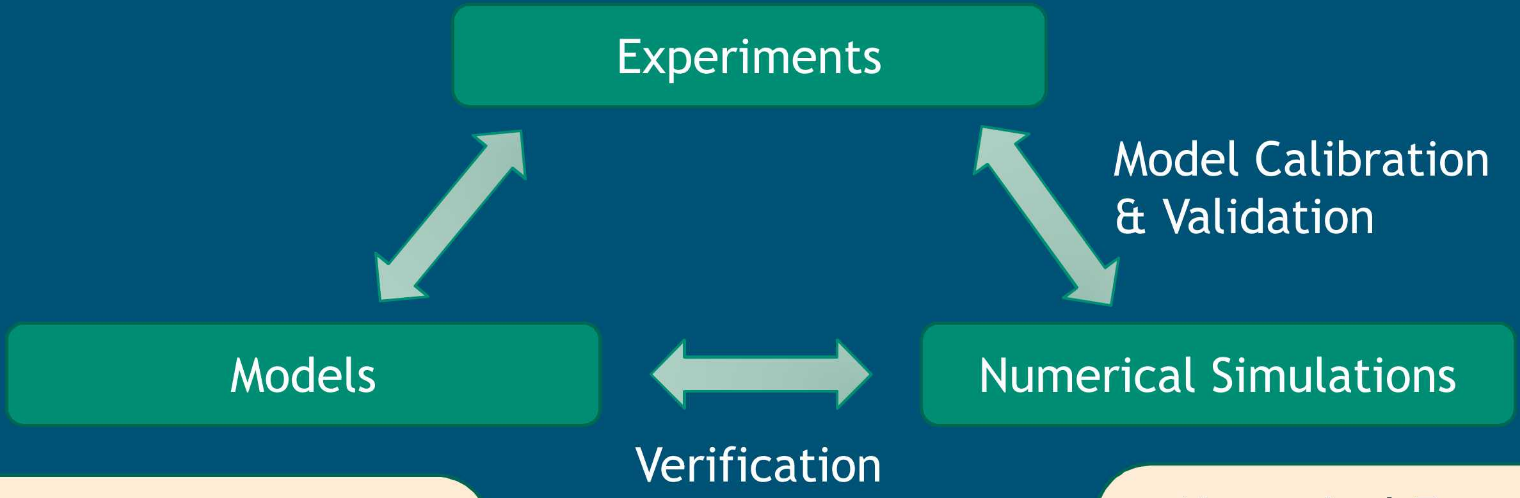
- Round off, discretization, or approximation errors
- Bugs or coding errors
- Bit-flip and hardware failures

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Uncertainty in Predictive Science

<u>Experimental Uncertainties and Limitations</u>

- Limited or incomplete data
- Limited accuracy or resolution of sensors



"Computational results are believed by no one, except the person who wrote the code"
-Gunzburger

<u>Model and Input Uncertainties</u>

- Model errors or discrepancies
- Input uncertainties due to uncertain parameters, forcing functions, and initial & boundary conditions

<u>Numerical Errors and Uncertainties</u>

- Round off, discretization, or approximation errors
- Bugs or coding errors
- Bit-flip and hardware failures

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Types of Uncertainty

Aleatoric (Statistical, Stochastic, or Irreducible Uncertainty):

"Uncertainty inherent to a problem or experiment that in principle cannot be reduced by additional physical or experimental knowledge."

Epistemic (Systemic Uncertainty):

"Uncertainty due to simplifying model assumptions, missing physics, or basic lack on knowledge."

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Statistics of Probability Theory

"Consider parameters and measurement errors to be random variables whose statistical properties or distributions we wish to infer using measured data"

Frequentist:

> "Probabilities are defined as the frequency with which an event occurs if the experiment is repeated a large number of times"

Bayesian:

> "treats probabilities as a distribution of subjective values, rather than a single frequency, that are constructed or updated as data is observed"

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Probability Space and Univariate Random Variable

Probability Space: A probability space $(\Omega, \mathcal{F}, P)$ is comprised of three components:

- Ω: sample space is the set of all possible outcomes from an experiment
- $\mathcal{F}$: σ- field of subsets of Ω that contains all events of interest
- $P: \mathcal{F} \rightarrow [0,1]$: probability or measure that satisfies the postulates
  1. $P(\emptyset) = 0$
  2. $P(\Omega) = 1$
  3. If $A_i \in \mathcal{F}$ and $A_i \cap A_j = \emptyset$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Random Variable: a univariate random variable is a function $X: \Omega \rightarrow \mathbb{R}$ with the property that

$\{\omega \in \Omega \,|X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$; i.e. it is measurable

Realization: the value

$$x = X(\omega)$$

of a random variable X for an event $\omega \in \Omega$ is termed a realization of X.

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Some Properties of a Univariate Random Variable

Cumulative Distribution Function (cdf):

$$F_X\colon \mathbb{R} \to [0,1]$$
$$F_X(x) = P\{\omega \in \Omega \mid X(\omega) \leq x\}$$
$$F_X(x) = P\{X \leq x\}$$

Probability Density Function (pdf):

For a continuous random variable X the cdf can be expressed as,

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds, \qquad x \in \mathbb{R}$$

where the derivative $f_X = \frac{dF_X}{dx}$ is the pdf.

Probability Mass Function (pmf):

The pmf of a discrete random variable X is given by $f_X(x) = P(X = x)$

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Moments of a Univariate Random Variable

$$\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n f_X(x)dx$$

Mean (Expected Value): Density central location

$$\mu = \int x f_x(x)dx$$

Variance: Density's variability or width

$$\sigma^2 = \int (x-\mu)^2 f_x(x)dx$$

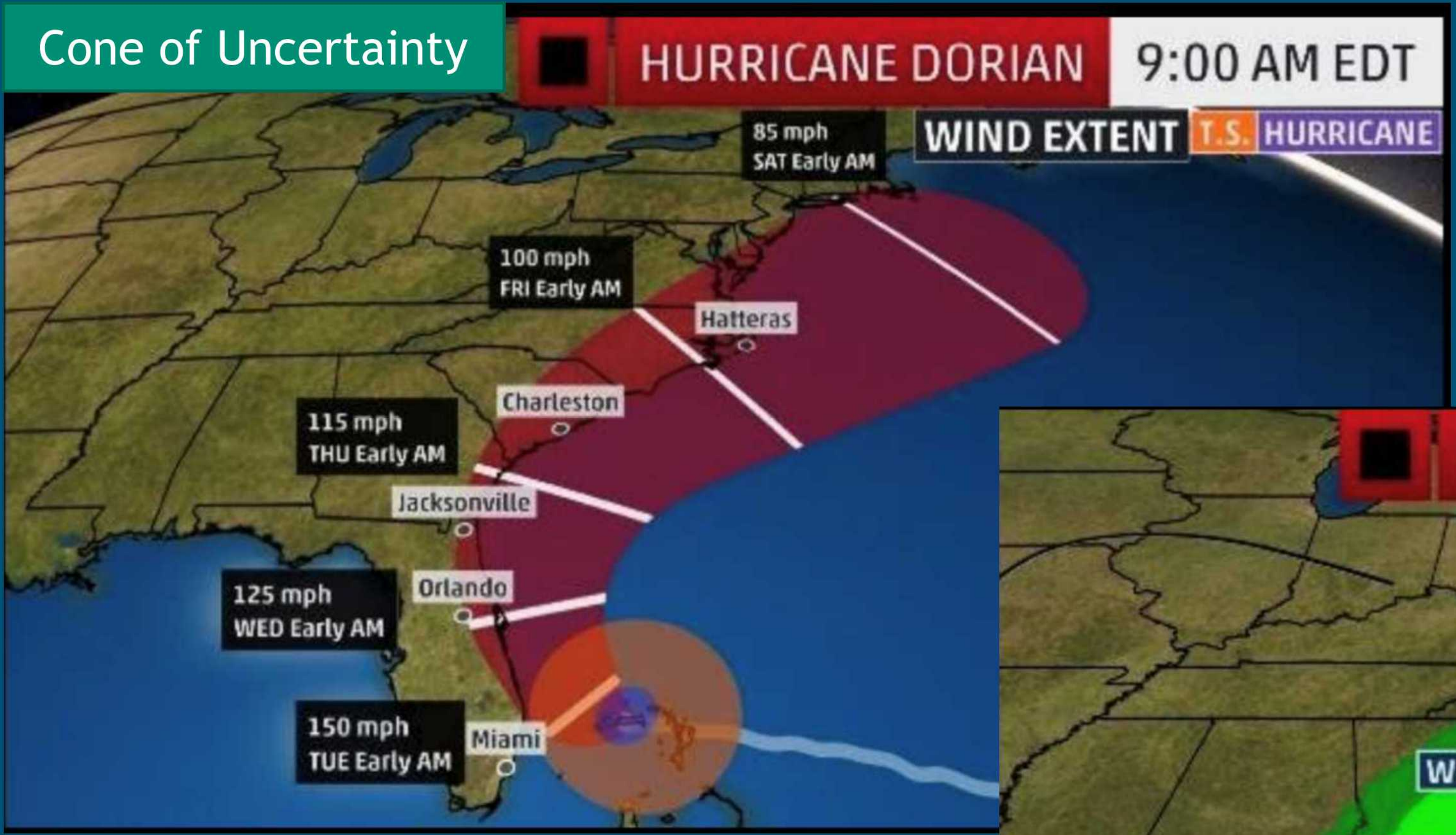
Skewness: Density's symmetry about μ

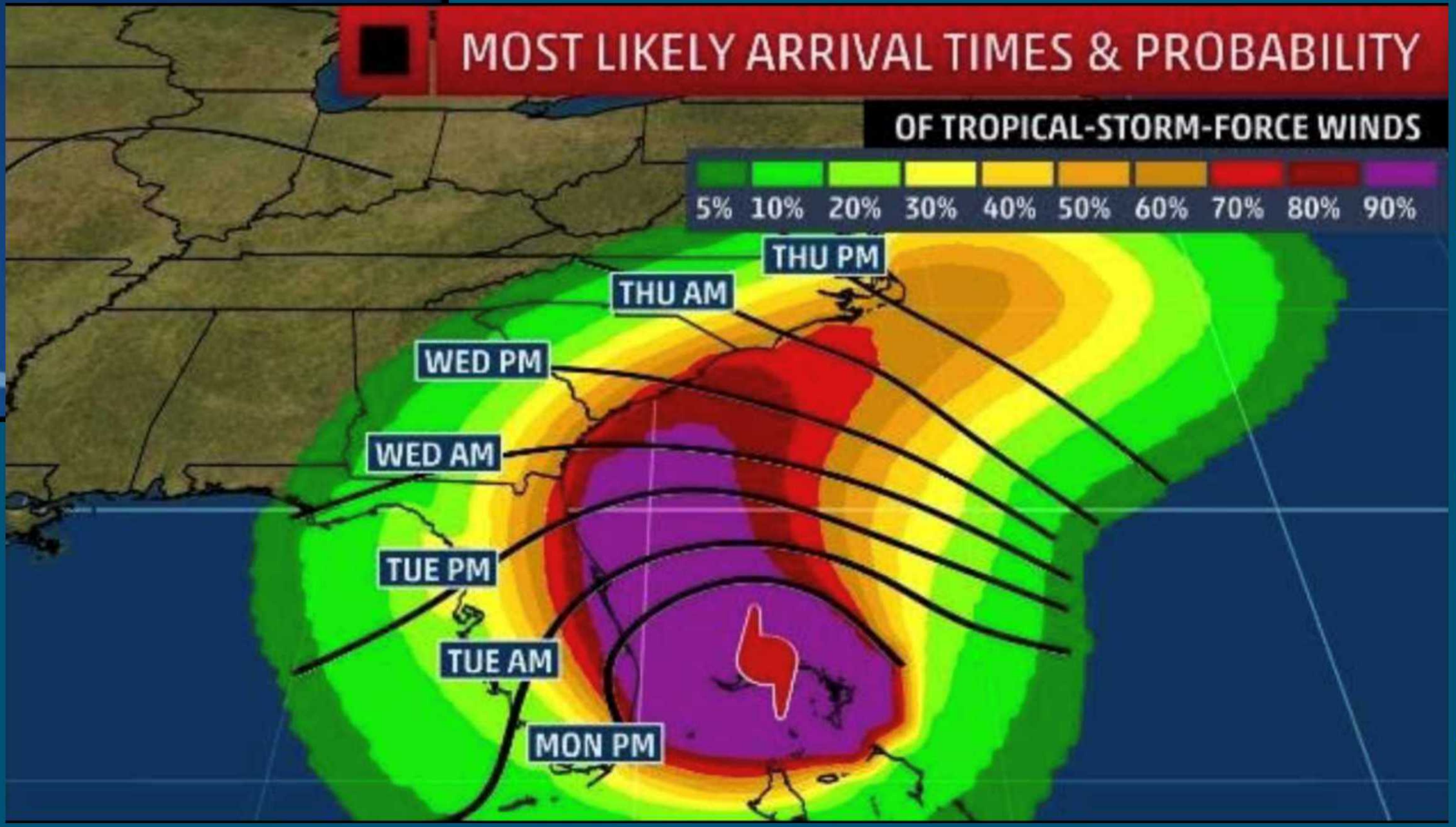
Kurtosis: Magnitude of tail contributions

[1] Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, 2014.

Why is the cone of uncertainty ”tighter” than the probability distribution?

Cone of Uncertainty





* https://www. weather.com/amp/storms/hurricane/news/2019-09-02-hurricane-dorian-labor-day-bahamas-florida-georgia-carolinas.html

Coin Toss Experiment