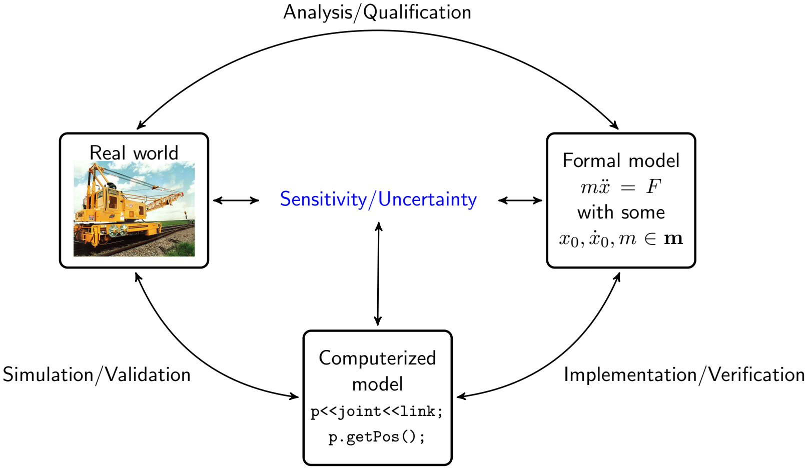
Kinds of Uncertainty and Possibilities for Their Treatment during Modeling and Simulation in Engineering

Ekaterina Auer

University of Technology, Business and Design Wismar

July 14, 2015

Modeling and Simulation Cycle



Uncertainty and sensitivity analyses are needed at each stage!

Imperfect Information: Philosophical Questions

Open question: Is there imprecision and uncertainty in the real world?

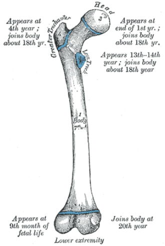
Fact: Data/information as available to an engineer are always imperfect

Modeling imperfect data: Probability theory only (until $\approx$ 1960s)

Currently: Many different possibilities for modeling, which are not equally suitable for a given situation

Classification possibility: Aspects of imperfect information*

- Imprecision
  → The length of Mr. X's femur is either 52.6cm or 53.2cm
- Uncertainty
  → The length of Mr. X's femur is probably 52.6cm
- Inconsistency, vagueness, ambiguity, error
  → The length of the femur is on average 26.74% of the height, measured 70cm on Mr. X (whose height is 190cm)



*Ph. Smets, *Imperfect information: Imprecision – Uncertainty*, 1999

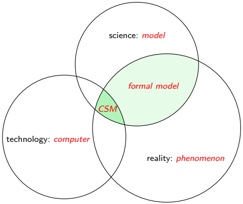
Basic Notions

Model (traditional): A set of mathematical equations along with the *computational expression* that describe a physical phenomenon

Computerized model (CM): *Code*

Accuracy: The agreement between estimated values and their true values

Credibility: The degree of trust that the CM answers a specific research question



Impreciseness: Characterized by the absence of an error component

Uncertainty: Arises from e.g. a gap in knowledge about the real system or its inherent variability

Sensitivity: A measure of the effect of a change in a particular variable on the simulation outputs

Hicks et al., Is My Model Good Enough? DOI: 10.1115/1.4029304

Errors in an Engineering Application

P — real system
V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$

Reality

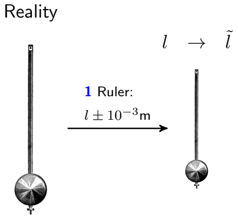
1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Errors in an Engineering Application

P — real system

V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$



1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Errors in an Engineering Application

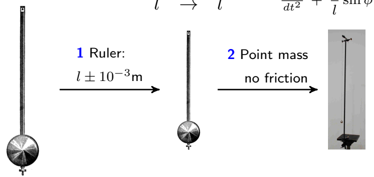
P — real system
V — representation of the CM solution

$$\max\{|P-M|, |M-D|, |D-L|, |L-V|\} \leq |P-V|$$
$$\leq |P-M| + |M-D| + |D-L| + |L-V|$$

Reality

$l \rightarrow \tilde{l}$

$$\frac{d^2\phi}{dt^2} + \frac{g}{l}\sin\phi = 0$$



1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Errors in an Engineering Application

P — real system
V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$

1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Reality

1 Ruler: $l \pm 10^{-3}$m $\longrightarrow$

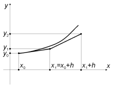
$l \rightarrow \tilde{l}$

2 Point mass, no friction $\longrightarrow$

$$\frac{d^2\phi}{dt^2} + \frac{g}{l}\sin\phi = 0$$

3 Grid: $h = t_i - t_{i-1}$ $\longrightarrow$

$$y_{i+1} = y_i + hf(t_i, y_i) + \Delta_i$$



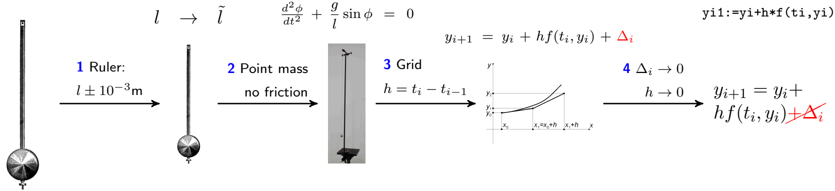
Errors in an Engineering Application

P — real system
V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$

1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Reality



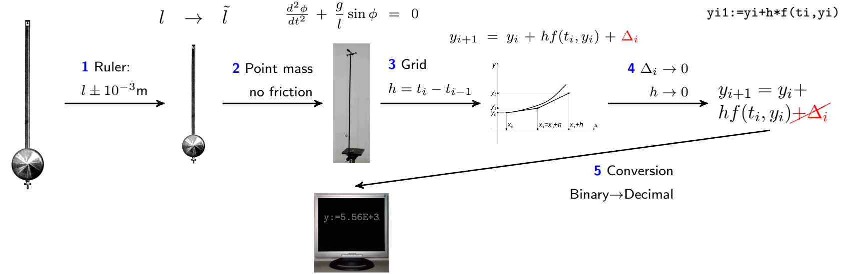
Errors in an Engineering Application

P — real system
V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$

1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$

Reality

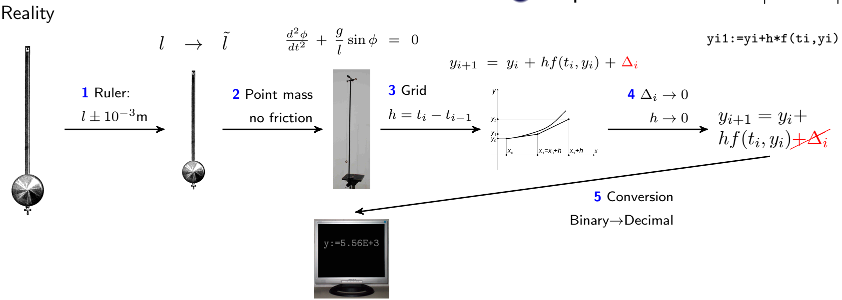


Errors in an Engineering Application

P — real system
V — representation of the CM solution

$$\max\{|P-M|,|M-D|,|D-L|,|L-V|\} \leq |P-V|$$
$$\leq |P-M|+|M-D|+|D-L|+|L-V|$$

1. Errors in the input data
2. Modeling error $|P-M|$
3. Discretisation error $|M-D|$
4. Truncation error $|D-L|$
5. Representation error $|L-V|$



These 'errors' can be seen as *sources* of uncertainty

Outline

Uncertainty versus Sensitivity Analyses

Uncertainty analysis

Quantify the uncertainty in the the model output from the uncertainty in the input or vice versa

Direct:

- Probabilistic (e.g. Monte-Carlo, polynomial chaos expansion)
- Non-probabilistic (e.g. interval, fuzzy)

Inverse:

- Frequentist
- Bayesian

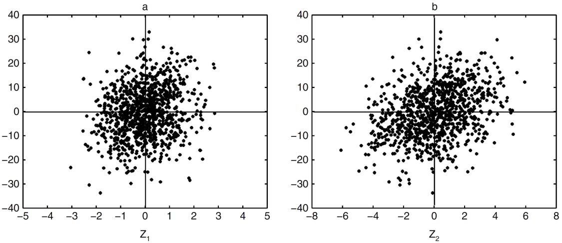
Sensitivity analysis

Apportion the uncertainty in the model output to different sources of uncertainty in the model input

- Scatter plots
- First derivative $\mathbf{s}_i = \dfrac{\partial x(\mathbf{p}_1, \ldots, \mathbf{p}_n)}{\partial p_i}$ for $p_i \in \mathbf{p}_i$

Uncertainty versus Sensitivity Analyses

Example:* $Y = \Omega(Z_1 + Z_2)$, $Z_i \sim \mathcal{N}(0, \sigma_i)$, Z_1 less uncertain $(\sigma_1 < \sigma_2)$



$$S_i = \frac{\partial Y}{\partial Z_i} = \Omega$$

$\rightsquigarrow$ Z_1, Z_2 are equal

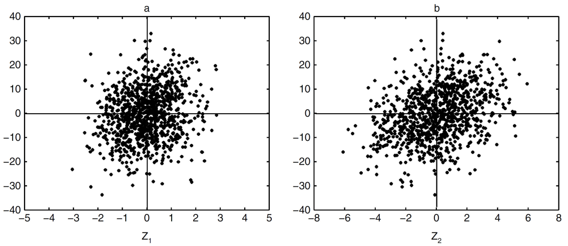
$\rightsquigarrow$ Z_1 is less influential

- Point derivatives can lead to wrong conclusions $\rightsquigarrow$ $S^\sigma = \dfrac{\sigma_i \partial Y}{\sigma_Y \partial Z_i}$
- Another view: S^σ combines uncertainty and sensitivity!

* A. Saltelli et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, 2008

Uncertainty versus Sensitivity Analyses

Example:* $Y = \Omega(Z_1 + Z_2)$, $Z_i \sim \mathcal{N}(0, \sigma_i)$, Z_1 less uncertain $(\sigma_1 < \sigma_2)$



$\rightsquigarrow$ scatter plots propagate the uncertainty

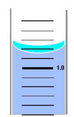
$$S_i = \frac{\partial Y}{\partial Z_i} = \Omega$$

$\rightsquigarrow$ the model response is indeed equal

- Sensitivity is *the response of the model to the changes in parameters*
- Uncertainty is quantified by *propagating* it from input to output
- S^σ or similar notions *combine* both in one indicator

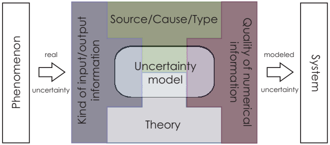
* A. Saltelli et al., *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, 2008

Uncertainty Modeling: Choice of the Method

Uncertainty 

→ Uncertainty model
Interval $[1.2, 1.6]$ → Observer 

1. Kind/source/cause of uncertainty
   lack of info/complexity/conflict/belief/ambiguity
2. Type of input information
   numerical/interval/linguistic/symbolic
3. Quality of numerical data
   nominal/ordered/metric/precise/interval/absolute
4. Required output information
   numerical/interval/linguistic/symbolic



Uncertainty models require a certain scale level of numerical information

Scale of method's operations $\leq$ Scale of provided information

Example: Frequentist Kolmogorov probability theory: $(LoI, Num, Cardinal, Num)$

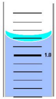
Sources of Uncertainty

Two major kinds of uncertainty

Aleatory (irreducible)

environmental stochasticity
(as in games of chance)

Epistemic (reducible)

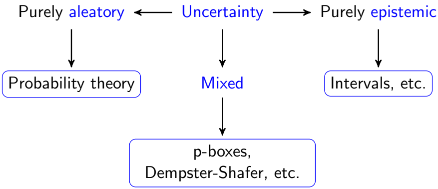


lack of knowledge:
measurement uncertainty,
unobservability, censoring

Sources of uncertainty

1. Uncertainty in the model itself (e.g., due to simplifications or parameter/dimension reduction)
2. Possible numerical discretization
3. Uncertainty in parameters (e.g., due to physical reasons or measurement errors)
4. Errors due to the finite nature of floating-point arithmetics

Main Types of Methods



Fuzzy methods handle impreciseness

1. A.C. Cullen, H.C. Frey, *Probabilistic techniques in exposure assessment*, 1999
2. H.M. Wadsworth, *Handbook of statistical methods for engineers and scientists*, 1998
3. S. Ferson et al, *Constructing Probability Boxes and Dempster-Shafer Structures*, 2003
4. V. Kreinovich et al, *Monte-Carlo-type techniques for processing interval uncertainty...*, 2004
5. H.-J. Zimmermann, *Fuzzy Set Theory and its Applications*, 2003

	Model	Discretization	Parameters	Arithmetic
PT	1		2	✓
IP	3		4	
F	5		5	
VR	✓	✓	✓	✓

PT=probability theory, IP=p-boxes or Dempster-Shafer, VR= methods with result verification, F=fuzzy

Types of Algorithms with Uncertain Numbers

Rigor-preserving (≈ with result verification): the result is guaranteed to enclose the uncertainty completely, if inputs enclose it completely

Best possible (≈ inner enclosure): the result cannot get any tighter without more information

Statistical confidence: guarantee of the type "in x percent of the trials, the result is sure to enclose the uncertainty completely"

We will focus (mostly) on parametric uncertainty, direct case

Probabilistic methods →	Monte-Carlo, polynomial chaos exp.
Set-based methods →	Result verification
Mixed methods →	p-boxes, Dempster-Shafer

Monte-Carlo Simulation

Applications in connection with uncertainty

→ Properties of random variables with unknown distributions

→ Uncertainty propagation, failure analysis

Advantage: simple to implement **Disadvantage:** no proof of correctness

Random numbers

$z_1, z_2, \ldots, z_N$

↓

Model with N variables

(metric $\eta(z) = y$)

↓

System response

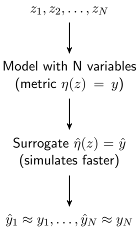
$y_1, y_2, \ldots, y_N$

1. Use the existing data to create a CDF for each input
2. Create an empty Frequency Distribution Histogram
3. The ith iteration step ($i = 1 \ldots 50000$):
   1. Loop over each input variable; calculate a weighted random number for inputs
   2. Use the weighted value of all input variables in the metric to calculate a representative answer
   3. Adjust the FDH appropriately
4. Repeat Step 3 if the final FDH is not complete
5. Normalize the FDH into DPDF; interpret the results

`www.drjfwright.com/c/montecarlosimulation.html`

Polynomial Chaos Expansion

Monte-Carlo can become computationally prohibitive for complex metrics



Goal: Quantify uncertainty (e.g. in differential equations)

Method: Represent stochastic quantities as spectral expansions of orthogonal polynomials; $X = f(\Xi)$, Ξ with a given distribution, $f \approx$ polynomial expansion

Example: $X \sim \chi^2_1$, $\Xi \sim \mathcal{N}(0,1)$, then $X = \Xi^2$, Hermite polynomials $\phi_0(\xi) = 1$,$\phi_1(\xi) = \xi$,$\phi_2(\xi) = \xi^2 - 1$,...

Generally: F_x, F_ξ the CDFs of X, Ξ, then $X = F_x^{-1}(F_\xi(\Xi)) = f(\Xi)$

Polynomial chaos expansion 1D: $X \approx \sum_{j=0}^{p} x_j \phi_j(\Xi)$, $x_j = \frac{<f,\phi_j>}{<\phi_j,\phi_j>}$ (truncated)

Propagation: $Y \approx \sum_{j=0}^{p} y_j \phi_j(\Xi) \approx \eta(\sum_{j=0}^{p} x_j \phi_j(\Xi))$ (e.g. by Galerkin projection)

Non-intrusive: Solve $y_k = \frac{<\eta(\sum_{j=0}^{p} x_j \phi_j(\Xi)),\phi_k>}{<\phi_k,\phi_k>}$, $k = 0 \ldots p$

A. O'Hagan, *Polynomial Chaos: A Tutorial and Critique*, 2013, www.tonyohagan.co.uk/academic/pdf/Polynomial-chaos.pdf

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Methods with Result Verification

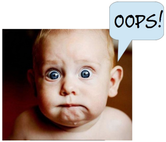
Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools

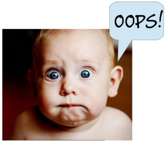


Not that easy....

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools



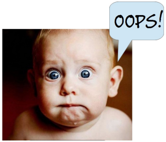
Not that easy....

Major problem: Many higher-level techniques with result verification (ODE solver etc.) need exact derivatives

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools



Not that easy....

Major problem: Many higher-level techniques with result verification (ODE solver etc.) need exact derivatives

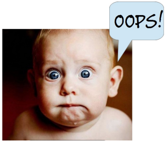
Derivatives can be easily (if not really quickly) obtained by algorithmic differentiation, if the expressions for the model are known

Some CA System ⟶ Expressions ⟶ Verification block ⟶ Verified results

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools



Not that easy....

Major problem: Many higher-level techniques with result verification (ODE solver etc.) need exact derivatives

Derivatives can be easily (if not really quickly) obtained by algorithmic differentiation, if the expressions for the model are known

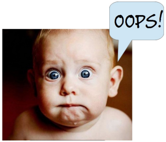
Some CA System ⟶ Expressions ⟶ Verification block ⟶ Verified results

Many simulation tools do not produce models as expressions!

Methods with Result Verification

Idea: Use set-based methods, if uncertainty can be bounded

Task: Interface result verification and existing simulation tools



Not that easy....

Major problem: Many higher-level techniques with result verification (ODE solver etc.) need exact derivatives

Derivatives can be easily (if not really quickly) obtained by algorithmic differentiation, if the expressions for the model are known

Some CA System → Expressions → Verification block → Verified results

Many simulation tools do not produce models as expressions!

Nonetheless possible, if simulation tools are open-source

Tools: SmartMOBILE for (Bio)Mechanics

1. Verified kinematics/dynamics + uncertainty management
2. Free choice of the underlying arithmetic: templates + solvers

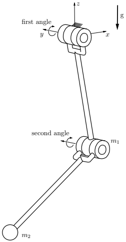
Type	Integrator	Purpose
`MoReal`	`MoAdams,...`	nonverified dynamics

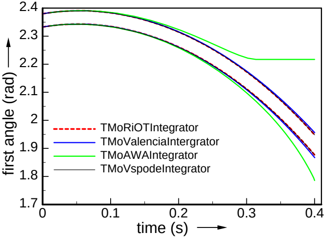
`TMoInterval`
`TMoFInterval`
`TMoTaylorModel`
`TMoTaylorModel`
	`TMoAWA`
	`TMoValencia`
	`TMoRiOT`
	`TMoVSPODE`
		verified dynamics of ODE based systems

`RDAInterval`	`---`	Taylor model based kinematics
`MoFInterval`	`MoIGradient`	verified equilibria kinematics with constraints
`MoSInterval`	`TMoValenciaS`	verified sensitivity
...	...	...

3. Converters MOBILE ⟶ SmartMOBILE

Example: Dynamics of a Double Pendulum





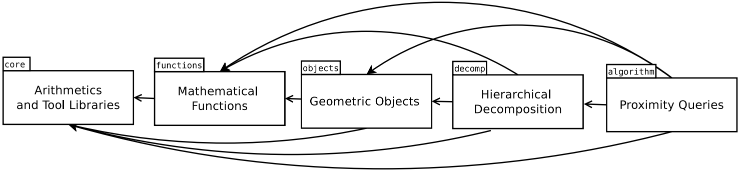

```
# define TMoInterval t;
TMoFrame<t> K0, K1, K2, K3, K4;
TMoAngularVariable<t> psi1, psi2;
// transmission elements
TMoVector<t> l1(0,0,-1), l2(0,0,-1);
TMoElementaryJoint<t> R1(K0,K1,psi1,xAxis);
TMoElementaryJoint<t> R2(K2,K3,psi2,xAxis);
TMoRigidLink<t> rod1(K1,K2,l1),rod2(K3,K4,l2);
t m1(1),m2(1);
TMoMassElement<t> Tip1(K2,m1),Tip2(K4,m2);
// the complete system
TMoMapChain<t> Pend;
Pend << R1<<rod1<<Tip1<<R2<<rod2<<Tip2;
// dynamics
TMoVariableList<t> q; q << psi1<<psi2;
TMoMechanicalSystem<t> S(q,Pend,K0,zAxis);
TMoAWAIntegrator I(S,0.0001,ITS_QR,15);
I.doMotion();
```

Strategy	TMoAWA (variable h)	TMoRiOT ($0.0002 \leq h \leq 0.2$)	TMoValencia ($h = 10^{-4}$)	TMoVSPODE (variable h)
Break–down	0.420	0.801	0.531	0.656
CPU Time*	5	285	22	10

* computed on 8 × Intel Xeon CPU 2.00GHz under Linux 2.6.25.14-69.fc8

UniVerMeC Instead of Templates

Unified Framework for Verified GeoMetric Computations



Relaxed layered structure:

core	*Adapter for underlying arithmetic libraries*
functions	*Uniform representation for functions*
objects	Implicit surfaces, CSG models, polyhedrons
decomp	Spatial decomposition, Multisection schemes
algorithms	Distance computation, Global optimization, ...

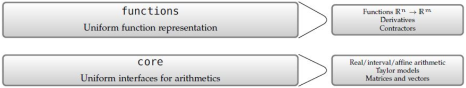
S. Kiel, UniVerMeC – *A Framework for Development, Assessment and Interoperable Use of Verified Techniques*, 2014

UniVerMeC: Function Specification

Important: Interoperability

The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units

Necessary: Formalizations for arithmetics, types of enclosures, etc.



$f : \mathbb{R}^n \mapsto \mathbb{R}^m$, tools for user-defined functions (inductive), analytical expressions or C++ code blocks

Function extensions: Evaluated with all arithmetics supported by `core`

$\rightarrow$ e.g. natural interval extension (replace everything with interval versions)

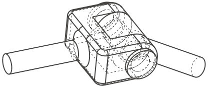
Features: set of functionalities associated with f (e.g., differentiability)

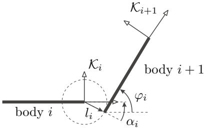
FR object: Tuple $F_{f,n,m} = (\mathcal{I}, \mathcal{F})$ where $\mathcal{I}$ is the set of inclusion functions, $\mathcal{F}$ is a choice out of r supported features

If $\mathcal{I}$ and $\mathcal{F}$ are defined appropriately, e.g. probabilistic arithmetics can be used!

Bounded Uncertainty at the Analysis Stage

`TMoSloppyJoint:`

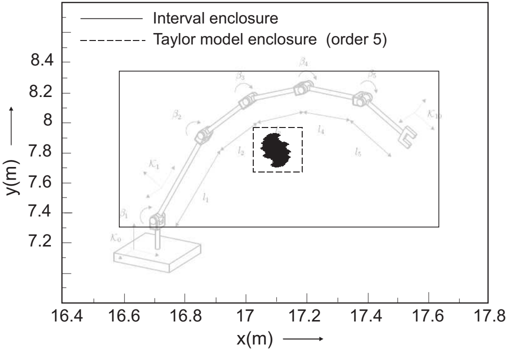




Parameter:

Lengths	$\pm 1\%$
Slackness	± 2mm
Angle	$\pm 0.1°$

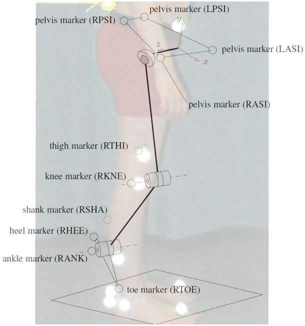
Results:



	x	y	Time (s)
`TMoInterval`	1.047	1.041	0.02
`TMoTaylorModel`	0.163	0.290	0.14

Purely Parametric Bounded Uncertainty

Body segment motion



Parameters (mm):

knee width	120 ± 10
ankle width	80 ± 10
displacements	tangential/soft tissue ± 10
	normal ± 5

Femur length (mm):

	TMoRDA	INTERVAL
Knee, ankle	[377.6; 396.7]	$[0;\infty]$
Skin displacement	[0.000; 621.4]	no answer

Point sensitivity of femur wrt.

Knee	Ankle	Tangential	Normal	Soft
0.4	-0.3	-2	0.7	1.4

Knee, Ankle: ±7mm; Tangential, Normal, Soft: ±37.5mm

Reference $\mathbf{r} = \sum_{i=1}^{n} s_i \cdot \mathbf{p}_i$:

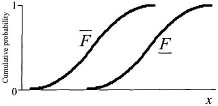
Probability Boxes

Given: $\overline{F}$ and $\underline{F} : \mathbb{R} \mapsto [0,1]$ nondecreasing, $\underline{F}(x) \leq \overline{F}(x)$, $\forall x \in \mathbb{R}$

$[\overline{F}, \underline{F}]$: Set of all nondecreasing functions $F : \mathbb{R} \mapsto [0,1]$ with $\underline{F}(x) \leq F(x) \leq \overline{F}(x)$

Definition: $[\overline{F}, \underline{F}]$ is called a p-box (probability box) when $\underline{F}$ and $\overline{F}$ circumscribe an imprecisely known probability distribution

Meaning: If X is a random variable with the unknown distribution $F \in [\overline{F}, \underline{F}]$ then $\underline{F}(x)$ is a lower bound on $F(x) = P(X \leq x)$



$$\overline{F}(x) = 1 - \underline{P}(X > x),\ \underline{F}(x) = \underline{P}(X \leq x)$$

$\underline{P}$ is a lower probability for an event A (the maximum rate one would be willing to pay for the gamble that pays 1 unit of utility if A occurs)

Bounds on the result of $+, -, \cdot, /$ of random variables defined using only bounds on their input distributions can be given

S. Ferson et al, *Constructing Probability Boxes and Dempster-Shafer Structures*, 2003

(Finite) Dempster-Shafer Structures

Discrete distributions

Probability density

$$f(x) = \begin{cases} \underbrace{P(X = x_i)}_{\approx\text{probability mass}} & \text{for } x = x_i \\ 0 & \text{for } x \neq x_i \end{cases}$$

$i = 1, 2, \ldots$

Idea Dempster-Shafer

$$\begin{cases} \underbrace{P(X = x_i)}_{\text{probability mass}} & \text{for } x \in \text{Set}_i \\ 0 & \text{otherwise} \end{cases}$$

$i = 1, 2, \ldots, n$

Interpretation: Classical probability theory in a topologically coarser space (where each focal element is identified as a point)

Focal elements $a_i \subseteq \mathbb{R}$: Sets associated with nonzero mass; may overlap one another

Basic probability assignment: Correspondence of probability masses associated with the focal elements; $m : 2^{\mathbb{R}} \mapsto [0, 1]$, $m(\emptyset) = 0$, $m(a_i) = p_i$, $i = 1, 2, \ldots, n$, $p_i > 0$, $\sum_i p_i = 1$

Dempster-Shafer Structures: Plausibility and Belief

Assumption: Let a_i be closed intervals, $b \subseteq \mathbb{R}$

Plausibility function: $Pl : 2^{\mathbb{R}} \mapsto [0, 1]$, $Pl(b) = \sum_{a_i \cap b \neq \emptyset} m(a_i)$

Belief function: $Bel : 2^{\mathbb{R}} \mapsto [0, 1]$, $Pl(b) = \sum_{a_i \subseteq b} m(a_i)$

Property: $Bel(b) \leq Pl(b)$

Arithmetic operations generalize the notion of convolution between distribution functions

The upper bound for the distribution function: $\sum_{\inf(a_i) \leq z} p_i$, $z \in \mathbb{R}$

(step function with n discontinuities)

The lower bound: $\sum_{\sup(a_i) \leq z} p_i$

P-boxes Versus Dempster-Shafer Structures

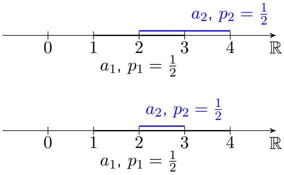
Dempster-Shafer: Uncertainty in the x-value, certainty in the p-value

P-box: Uncertainty about probabilities, certainty about events

Nonetheless: Dual, can be converted into each other

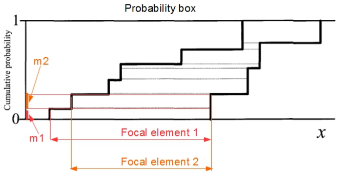
DS → p-box: $([x_i, y_i], p_i) \rightarrow \overline{F}(z) = \sum_{x_i \leq z} p_i,\ \underline{F}(z) = \sum_{y_i \leq z} p_i$

Relationship: Many Dempster-Shafer structures, a single p-box



$$\underline{F}(z) = \begin{cases} 0, & z < 3 \\ \frac{1}{2}, & z \in [3,4) \\ 1, & z \geq 4 \end{cases} \qquad \overline{F}(z) = \begin{cases} 0, & z < 1 \\ \frac{1}{2}, & z \in [1,2) \\ 1, & z \geq 2 \end{cases}$$

P-box → DS:



When to Apply?

1. Imprecisely specified distributions
2. Poorly known or even unknown dependencies
3. Non-negligible measurement uncertainty
4. Non-detects or other censoring in measurements
5. Small sample size
6. Inconsistency in the quality of input data
7. Model uncertainty
8. Non-stationarity (non-constant distributions)

Tools: Codes, References, Links for DS-Structures and P-Boxes

P-boxes: RiskCalc (commercial)

- → http://www.ramas.com/riskcalc.htm
- → probability bounds analysis, standard fuzzy arithmetic, and classical interval analysis

DS-structures: IPPToolbox (in MATLAB or R, open-source)

- → www.uni-due.de/informationslogistik/ipptoolbox.php

DS-structures with verified intervals: DSI Toolbox (MATLAB, open-source)

- → www.scg.inf.uni-due.de/forschung/software/dsi-toolbox.php

Applications, comparisons: www.lix.polytechnique.fr/~bouissou/

Stochastic arithmetic and the CADNA software

Stochastic arithmetic: Model for exact computation on imprecise data $\sim \mathcal{N}(\mu, \sigma)$

Origin: Cestac method by J. Vignes and J. M. Chesneaux, 1992

Idea: Interpret imprecise data as stochastic numbers
($\mathbb{S}$ is the set of Gaussian random variables)

Therefore: $X \in \mathbb{S}$ is characterized by μ and σ

Property: $\exists \lambda_\nu$: $P(X \in [\mu - \lambda_\nu \sigma, \mu + \lambda_\nu \sigma]) = 1 - \nu$

Confidence interval of μ with the probability $1 - \nu$: $[\mu - \lambda_\nu \sigma, \mu + \lambda_\nu \sigma]$,
(e.g. $\lambda_\nu = 1.96$ for $\nu = 0.05$)

Significant decimal digits on μ: $\log_{10}\left(\frac{|m|}{\lambda_\nu \sigma}\right)$ if $\frac{|m|}{\lambda_\nu \sigma} \geq 10$ otherwise 0

Operations: Follow from the properties of the Gaussian distribution

Software: CADNA, `www-pequan.lip6.fr/cadna/`, estimates round-off errors

Example: $P(x, y) = 9x^4 - y^4 + 2y^2$, $P(10864, 18817) = 2.0$ (wrong, exact=1.0), $P(1/3, 2/3) = 0.802...$ (correct). Is there any way to distinguish the quality?
In CADNA, the number of significant digits is computed to be zero in the first case, normal (15) in the second!

Fuzzy: Short Overview

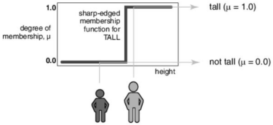
Goal: Model impreciseness/vaugness (*a little*, *very*, etc.)

Definition: Pair (U, μ) is a fuzzy set, where U is a set and $\mu : U \mapsto [0, 1]$ a membership function, $\mu(x)$ the grade of membership of x in U

Set vs. fuzzy set

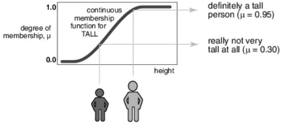
Traditional

$x \in U$ or not



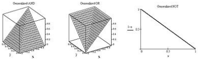
Fuzzy

μ decides if $x \in U$



Fuzzy number Convex, normalized fuzzy set $A \subseteq \mathbb{R}$ with continuous $\mu_A(x) = 1$ at precisely one element

Operations: xAND$y = \min(x, y)$, xOR$y = \max(x, y)$, NOT$x = 1 - x$



Fuzzy decisions: Rules for credibility, aggregation etc.

www.calvin.edu/~pribeiro/othrlnks/Fuzzy/fuzzysets.htm

Mixed Methods: Some References

A set-membership method to characterize a probabilistic set:
L. Jaulin et al., *Inner and outer approximations of probabilistic sets*, 2014
`www.ensta-bretagne.fr/jaulin/teaching.html`

Fuzzy probability theory: M. Beer, *Fuzzy Probability Theory*, In: Meyers, R. (ed.), Encyclopedia of Complexity and Systems Science, 2009

Comparisons: M Beer, V. Kreinovich, *Interval or Moments: Which Carry More Information?*, 2012

Imprecise probabilities: `http://www.sipta.org/`

Intervals and probabilities:
`http://ualr.edu/jdberleant/intprob/`

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

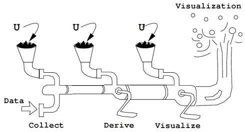
⚠ Most works deal with random aleatory uncertainty

L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, `graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/`

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

⚠ Most works deal with random aleatory uncertainty



→ Depends on uncertainty *sources*,

→ figure does not consider *model* uncertainty

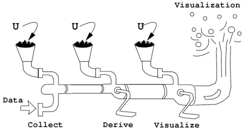
Figure from A. Pang et al., *Approaches to uncertainty visualization*, 1997

L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

⚠ Most works deal with random aleatory uncertainty



→ Depends on uncertainty *sources*,

→ figure does not consider *model* uncertainty

Figure from A. Pang et al., *Approaches to uncertainty visualization*, 1997

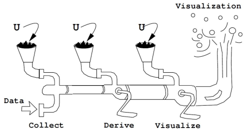
Taxonomies: Dimensionality of data vs that of uncertainty

L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, `graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/`

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

⚠ Most works deal with random aleatory uncertainty



→ Depends on uncertainty *sources*,

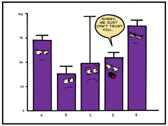
→ figure does not consider *model* uncertainty

Figure from A. Pang et al., *Approaches to uncertainty visualization*, 1997

Taxonomies: Dimensionality of data vs that of uncertainty

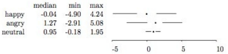
Methods: For example, for scalar/multivariate/vector discrete data

Error bars

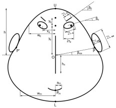


by Velica

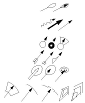
Tufte plots



Chernoff



Glyphs

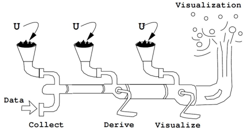


L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, `graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/`

Approaches to Uncertainty Visualization

Goal: Present data with auxiliary uncertainty information

Most works deal with random aleatory uncertainty



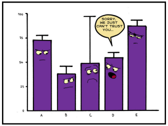
→ Depends on uncertainty *sources*,
→ figure does not consider *model* uncertainty

Figure from A. Pang et al., *Approaches to uncertainty visualization*, 1997

Taxonomies: Dimensionality of data vs that of uncertainty

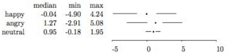
Methods: For example, for scalar/multivariate/vector discrete data

Error bars

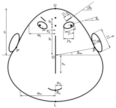


by Velica

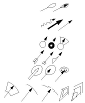
Tufte plots



Chernoff



Glyphs



Evaluation: e.g. K. Potter et al., *From Quantification to Visualization...*, 2011

L. Gosink et al., *Characterizing and Visualizing Predictive Uncertainty*, 2013, graphics.cs.ucdavis.edu/~joy/NSF-IIS-1018097/

Summary

Learned in this workshop:

Uncertainty: Definition, types, sources, models

Model uncertainty: Mostly not considered here

Parametric uncertainty: At modeling and simulation level

- → Classical uncertainty: Monte-Carlo, polynomial chaos expansions
- → *Bounded uncertainty:* intervals, Taylor models etc.
- → Imprecise probabilities: *p-boxes, DS-structures*
- → Impreciseness: fuzzy sets

Interesting research topics:

- → Uncertainty visualization
- → Interoperability/implementation issues
- → Parallelization/templatization (e.g. on the GPU)