A Verification and Validation (V&V) framework for Computational Science

R. Tempone

SCS, Florida State University

Joint work with: Ivo Babuška and Fabio Nobile

SANDIA CSRI Workshop on Mathematical Methods for Validation and Verification. August 14, 2007.

Support from SANDIA project # 523695 is gratefully acknowledged.

Outline

1. V&V framework
   - Introduction
   - Basic notions
   - A framework for V&V
   - Design of validation/accreditation experiments
2. SANDIA Static Frame Validation Problem
   - Definition of the problem
3. Solution procedure
   - Definition of the mathematical model
   - Calibration
   - Validation
   - Accreditation
   - Prediction
4. Conclusions
   - Conclusions

Outline

1. V&V framework
   - Introduction
   - Basic notions
   - A framework for V&V
   - Design of validation/accreditation experiments
2. SANDIA Static Frame Validation Problem
   - Definition of the problem
3. Solution procedure
   - Definition of the mathematical model
   - Calibration
   - Validation
   - Accreditation
   - Prediction
4. Conclusions
   - Conclusions

Introduction

- Verification and Validation concerns the reliability of computational models (mathematical models + numerical simulations).

- It aims to answer the question: *can we base an important or critical decision only on the outcome of a mathematical models?*

Goal (from ASME Guide to V&V, 2006)

To develop standards for assessing the correctness and credibility of modeling and simulation in computational science.

Introduction

- Verification and Validation concerns the reliability of computational models (mathematical models + numerical simulations).

- It aims to answer the question: *can we base an important or critical decision only on the outcome of a mathematical models?*

Goal (from ASME Guide to V&V, 2006)

To develop standards for assessing the correctness and credibility of modeling and simulation in computational science.

Introduction

- Verification and Validation concerns the reliability of computational models (mathematical models + numerical simulations).

- It aims to answer the question: *can we base an important or critical decision only on the outcome of a mathematical models?*

Goal (from ASME Guide to V&V, 2006)

To develop standards for assessing the correctness and credibility of modeling and simulation in computational science.

Where V&V is relevant

- To reduce the amount of experiments in regulatory requirements. Ideally, one wishes to prove that a new industrial design satisfies given regulatory requirements only upon a computational analysis. Can we trust the results?

- Whenever experiments on the physical system of interest are too expensive or cannot be done.

Where V&V is relevant

- To reduce the amount of experiments in regulatory requirements. Ideally, one wishes to prove that a new industrial design satisfies given regulatory requirements only upon a computational analysis. Can we trust the results?
- Whenever experiments on the physical system of interest are too expensive or cannot be done.

Definition of V&V (from ASME Guide)

- Verification. The process of determining that a computational model accurately represents the underlying mathematical model and its solution.

Verification: are we solving the equations right?

- Validation. The process of determining the degree to which a model is an accurate representation of the real world for the perspective of the intended uses of the model.

Validation: are we solving the right equations?

Definition of V&V (from ASME Guide)

- Verification. The process of determining that a computational model accurately represents the underlying mathematical model and its solution.

Verification: are we solving the equations right?

- Validation. The process of determining the degree to which a model is an accurate representation of the real world for the perspective of the intended uses of the model.

Validation: are we solving the right equations?

Definition of V&V (from ASME Guide)

- Verification. The process of determining that a computational model accurately represents the underlying mathematical model and its solution.

Verification: are we solving the equations right?

- Validation. The process of determining the degree to which a model is an accurate representation of the real world for the perspective of the intended uses of the model.

Validation: are we solving the right equations?

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Ingredients for V&V

- Verification includes:
  - Code Verification: correctness of code implementation; check with exact solutions; etc...
  - Calculation Verification: analysis of the numerical method, convergence, a-posteriori error estimation for the desired output, etc...
- Validation implies comparison with experiments
- Verification should come before Validation
- Experimental measurements should be verified, too.

Basic notions and definitions

Prediction problem

The purpose of computation is to provide the quantitative data of interest (sometimes called quantities of interest) on which a decision is made.

Experimental analysis of the prediction problem is too costly or unfeasible. The prediction is based only on the mathematical model.

Mathematical model.

- Structure: type of equations and functional relations between input and output
- Input Data: parameters values or ranges (coeffs., boundary conds., geometry, etc.) in the math. model

Basic notions and definitions

Prediction problem

The purpose of computation is to provide the quantitative data of interest (sometimes called quantities of interest) on which a decision is made.

Experimental analysis of the prediction problem is too costly or unfeasible. The prediction is based only on the mathematical model.

Mathematical model.

- Structure: type of equations and functional relations between input and output
- Input Data: parameters values or ranges (coeffs., boundary conds., geometry, etc.) in the math. model

Basic notions and definitions

Prediction problem

The purpose of computation is to provide the quantitative data of interest (sometimes called quantities of interest) on which a decision is made.

Experimental analysis of the prediction problem is too costly or unfeasible. The prediction is based only on the mathematical model.

Mathematical model.

- Structure: type of equations and functional relations between input and output
- Input Data: parameters values or ranges (coeffs., boundary conds., geometry, etc.) in the math. model

Basic notions and definitions

The mathematical model only transforms the available information into the prediction of the quantity of interest.

$$Q = M(\Theta),$$

Q : quantity of interest

M : structure of the mathematical model

Θ : input data

It can also map uncertainty ranges (sets) for input data into corresponding uncertainty sets for the output quantity.

Basic notions and definitions

The mathematical model only transforms the available information into the prediction of the quantity of interest.

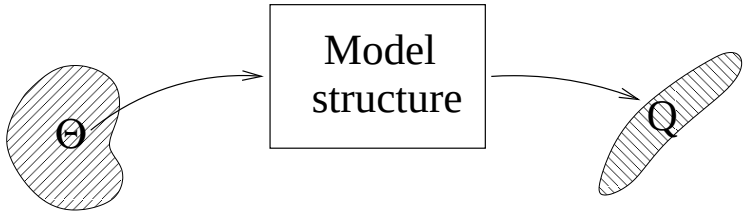
$$Q = M(\Theta),$$

Q : quantity of interest

M : structure of the mathematical model

Θ : input data

It can also map uncertainty ranges (sets) for input data into corresponding uncertainty sets for the output quantity.



Basic notions and definitions

Calibration experiments

- Simple experiments to identify (if necessary) some of the input data of the mathematical model.
- They are much simpler than the prediction problem and can be described, eventually, by simpler math. models.

Validation/Accreditation experiments

- Are easier than the prediction problem and experimental measurements can be obtained.
- The comparison between computational model and experiments is based on a suitable validation metric.
- If the computational model fails to meet tolerances set a priori by the analyst, the model is rejected.
- The validation experiments, metrics and tolerances have to be related (and relevant) to the prediction.

Basic notions and definitions

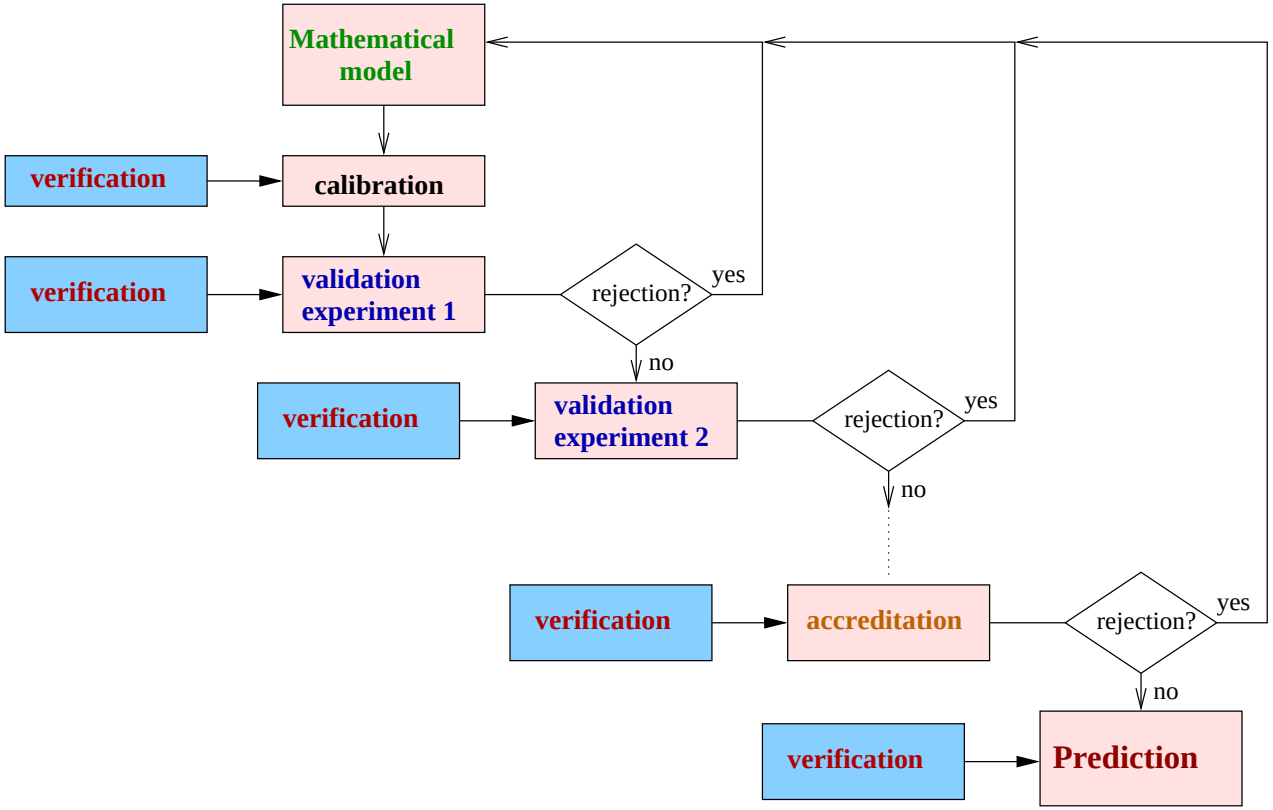
Calibration experiments

- Simple experiments to identify (if necessary) some of the input data of the mathematical model.
- They are much simpler than the prediction problem and can be described, eventually, by simpler math. models.

Validation/Accreditation experiments

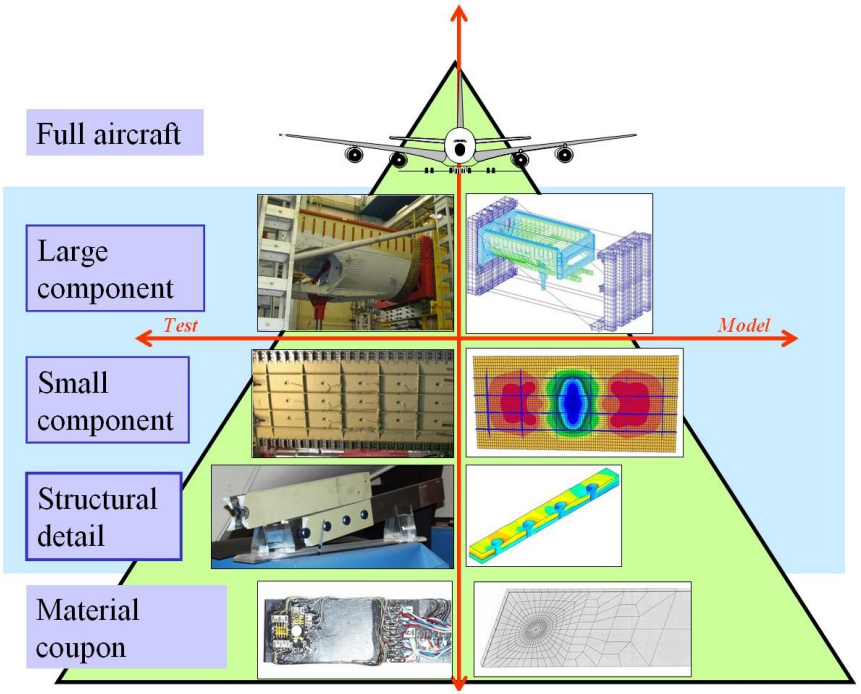
- Are easier than the prediction problem and experimental measurements can be obtained.
- The comparison between computational model and experiments is based on a suitable validation metric.
- If the computational model fails to meet tolerances set a priori by the analyst, the model is rejected.
- The validation experiments, metrics and tolerances have to be related (and relevant) to the prediction.

A framework for V&V



The validation pyramid

Ref: Prof. Babuska's morning discussion.



[Courtesy of Mr. S. Guinard, EADS Corporate Research]

Typically, number of experimental measurements decreases with increasing complexity of the validation/accreditation experiments

On the validation/accreditation steps

- Choice and amount of validation/accr. experiments
- Choice of validation/accr. metric and rejection criterion
- Validation/accreditation experiments show discrepancy wrt computations. Can we exploit this info. in prediction?

Proposed approach to V&V

- Produce model (parameter) updates using validation/accr. experiments.
- Then, estimate the implied change in the prediction from calibrated & updated models (val. metric). Decide on rejection.

The approach is general: Applications to stochastic (see Webster's talk), worst scenario (see Nobile's talk).

On the validation/accreditation steps

- Choice and amount of validation/accr. experiments
- Choice of validation/accr. metric and rejection criterion
- Validation/accreditation experiments show discrepancy wrt computations. Can we exploit this info. in prediction?

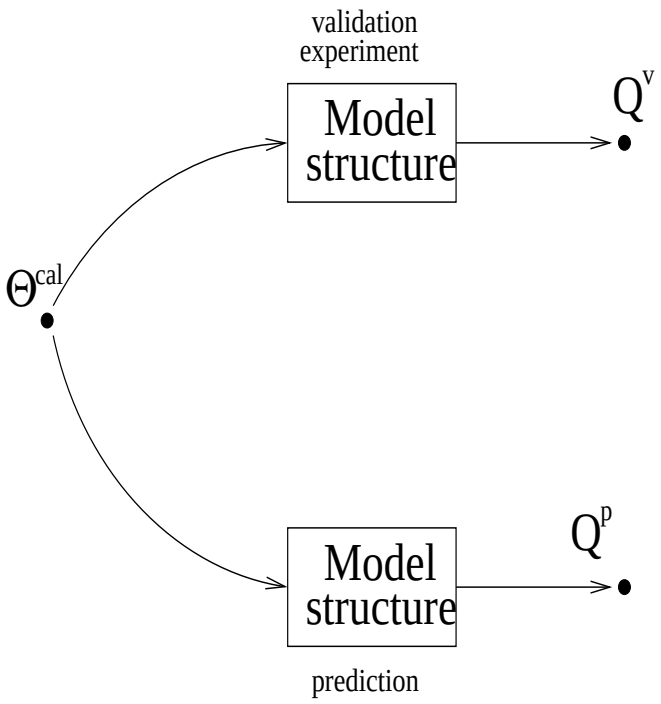
Proposed approach to V&V

•Produce model (parameter) updates using validation/accr. experiments.

•Then, estimate the implied change in the prediction from calibrated & updated models (val. metric). Decide on rejection.

The approach is general: Applications to stochastic (see Webster's talk), worst scenario (see Nobile's talk).

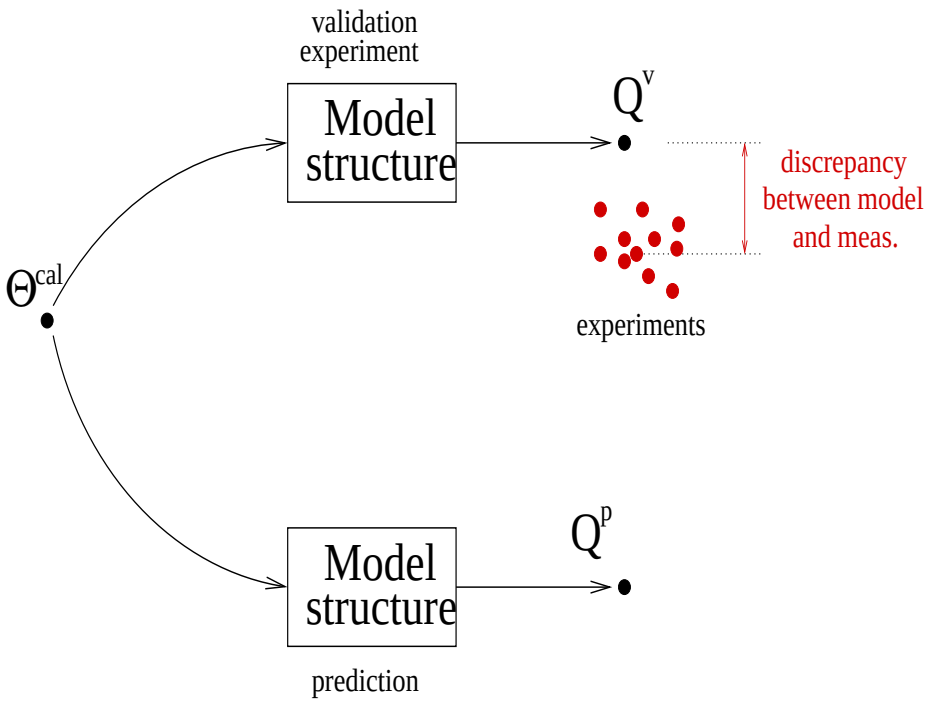
From validation to prediction



Rejection criterion: $\Delta Q^p \leq tol$

Obs: ΔQ^p estimates model's predictive capability

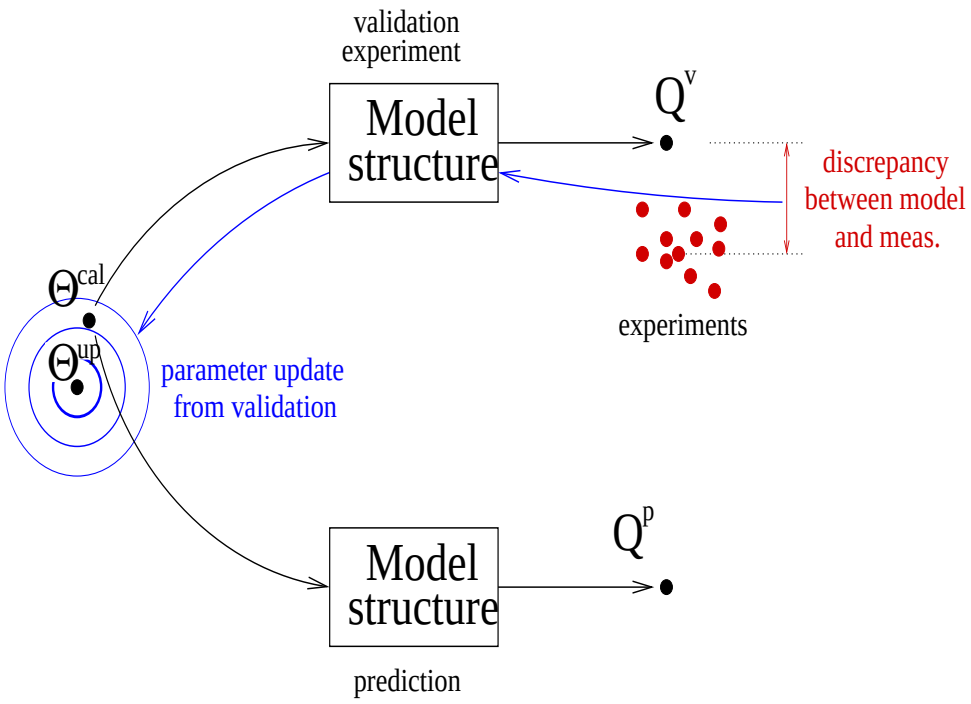
From validation to prediction



Rejection criterion: $\Delta Q^p \leq tol$

Obs: ΔQ^p estimates model's predictive capability

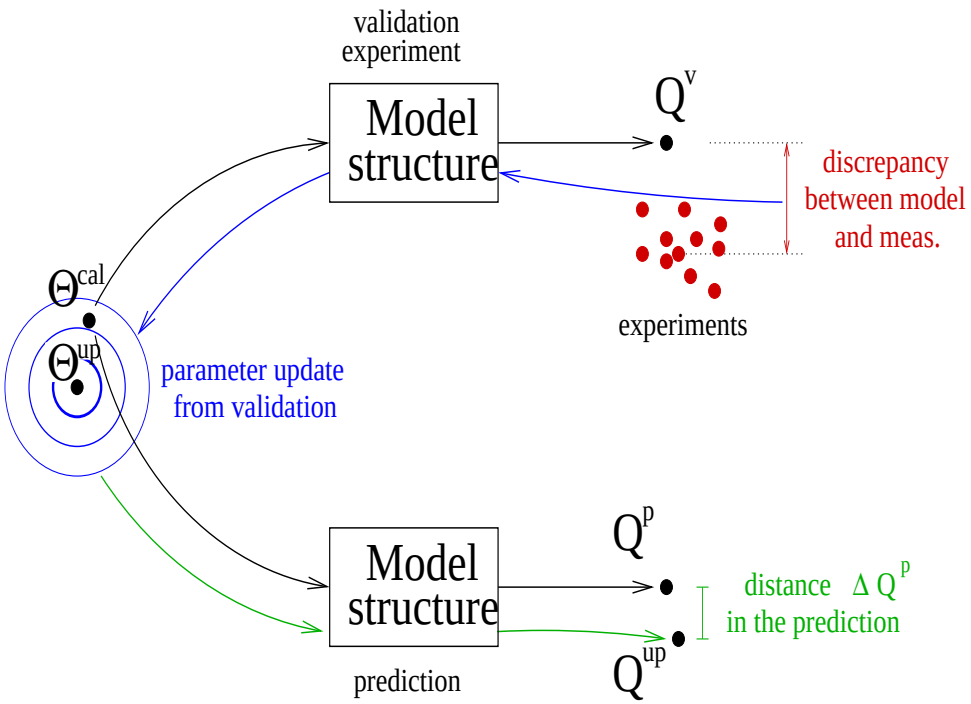
From validation to prediction



Rejection criterion: $\Delta Q^p \leq tol$

Obs: ΔQ^p estimates model's predictive capability

From validation to prediction



Rejection criterion: $\boxed{\Delta Q^p \leq tol}$

Obs: ΔQ^p estimates model's predictive capability

On the choice of validation experiments

Let $\Theta = [\theta_1, \ldots, \theta_n]$. Sensitivity analysis should be used in the prediction problem to identify the parameters θ_i with largest influence in the prediction. Where to start?

$\left\| \frac{\partial Q^p}{\partial \theta_i} \right\| \ll 1$ Q^p is not sensitive to θ_i; no need for validation

$\left\| \frac{\partial Q^p}{\partial \theta_i} \right\| \gg 1$ The prediction problem is too sensitive!! computation not reliable

$\left\| \frac{\partial Q^p}{\partial \theta_i} \right\| \approx 1$ Need to check with a validation experiment the "uncertainty" in θ_i.

Recall talks: Helton (Sensitivity), Santner (Experiment design)

On the choice of validation experiments

Let $\Theta = [\theta_1, \ldots, \theta_n]$. Sensitivity analysis should be used in the prediction problem to identify the parameters θ_i with largest influence in the prediction. Where to start?

$\left\lVert \frac{\partial Q^p}{\partial \theta_i} \right\rVert \ll 1$	Q^p is not sensitive to θ_i; no need for validation
$\left\lVert \frac{\partial Q^p}{\partial \theta_i} \right\rVert \gg 1$	The prediction problem is too sensitive!! computation not reliable
$\left\lVert \frac{\partial Q^p}{\partial \theta_i} \right\rVert \approx 1$	Need to check with a validation experiment the "uncertainty" in θ_i.

Recall talks: Helton (Sensitivity), Santner (Experiment design)

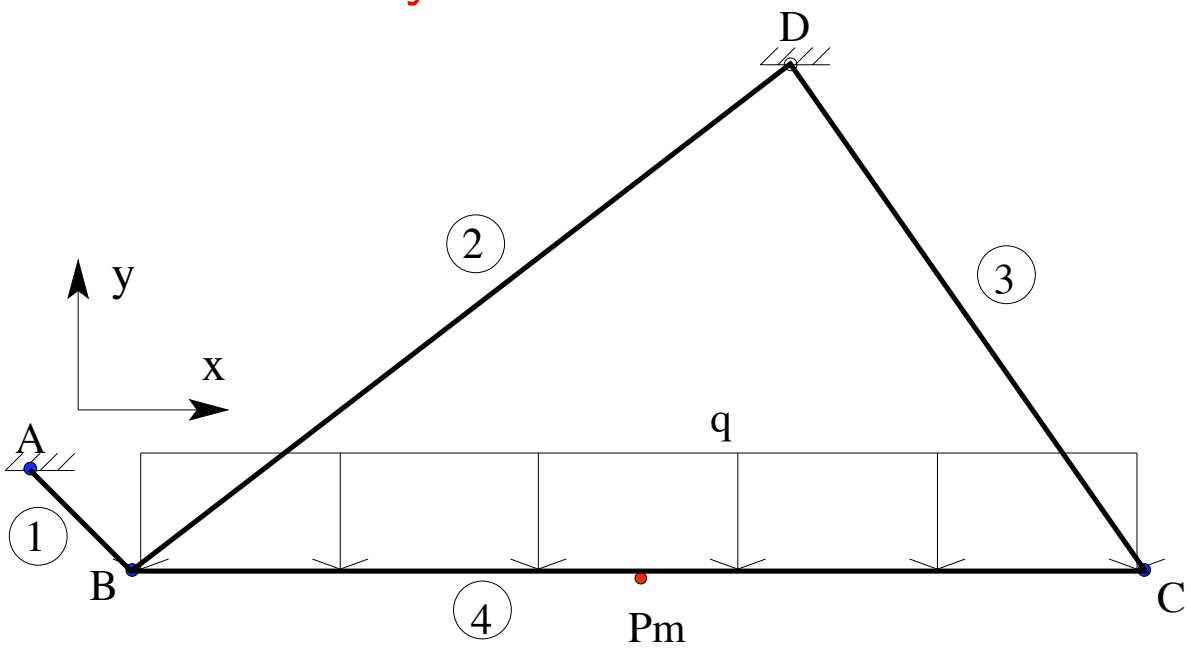
Outline

1. V&V framework
   - Introduction
   - Basic notions
   - A framework for V&V
   - Design of validation/accreditation experiments
2. SANDIA Static Frame Validation Problem
   - Definition of the problem
3. Solution procedure
   - Definition of the mathematical model
   - Calibration
   - Validation
   - Accreditation
   - Prediction
4. Conclusions
   - Conclusions

SANDIA Static Frame Validation Problem

Reference: SANDIA Validation Challenge Workshop, 05/2006.

Frame reliability



Point	x(cm)	y(cm)
A	0	20
B	20	0
C	220	0
D	150	100

Bar #	$A(cm^2)$	$I(cm^4)$
1	16	
2	16	
3	16	
4	80	5333

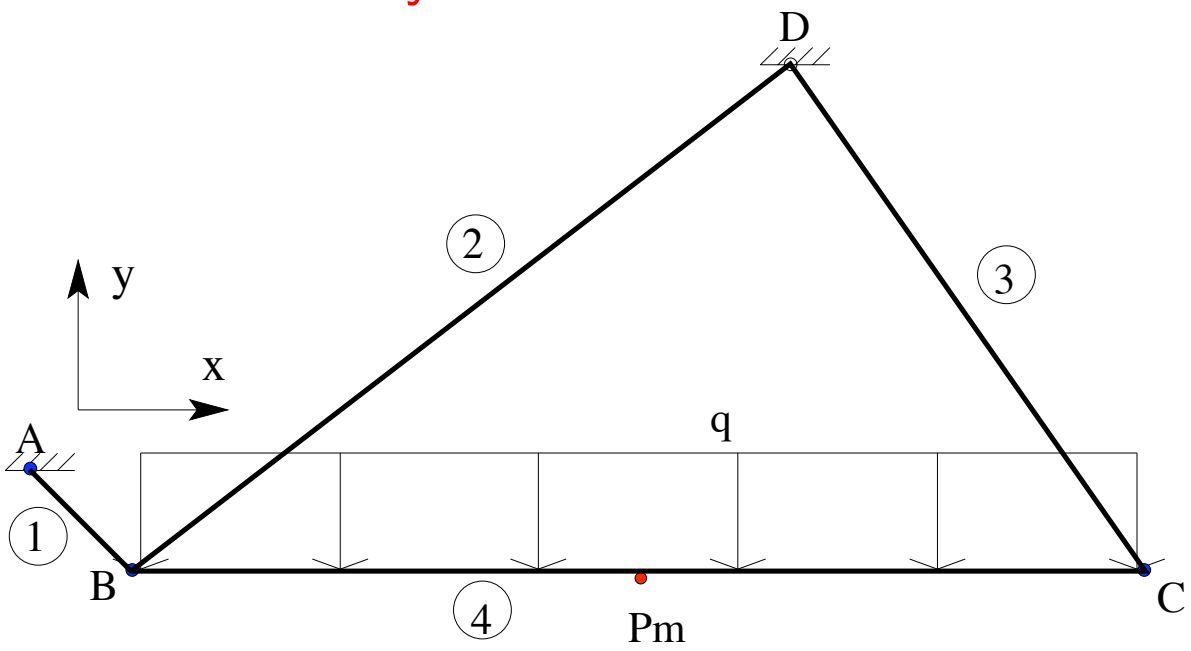
Quantity of interest: vertical displacement $w(P_m)$ of the point P_m.

In particular $P(|w(P_m)| \leq 3mm)$

SANDIA Static Frame Validation Problem

Reference: SANDIA Validation Challenge Workshop, 05/2006.

Frame reliability



Point	x(cm)	y(cm)
A	0	20
B	20	0
C	220	0
D	150	100

Bar #	$A(cm^2)$	$I(cm^4)$
1	16	
2	16	
3	16	
4	80	5333

Quantity of interest: vertical displacement $w(P_m)$ of the point P_m.

In particular $P(|w(P_m)| \leq 3mm)$

Assumptions on the prediction problem

- Joints (hinges) are perfect.
- Supports in points A and D are perfectly rigid.
- Geometric data are completely accurate
- The load q is given by regulation
- Linear elasticity theory acceptable for the bars' deformation. In particular Kirchhoff bending theory can be used fo bar #4.
- Material properties (Young's modulus) significantly influence the target displacement (hence the reliability of the frame)
- Also, the material might be inhomogeneous.

Model structure: linear elasticity
Input data: elasticity coefficient

Assumptions on the prediction problem

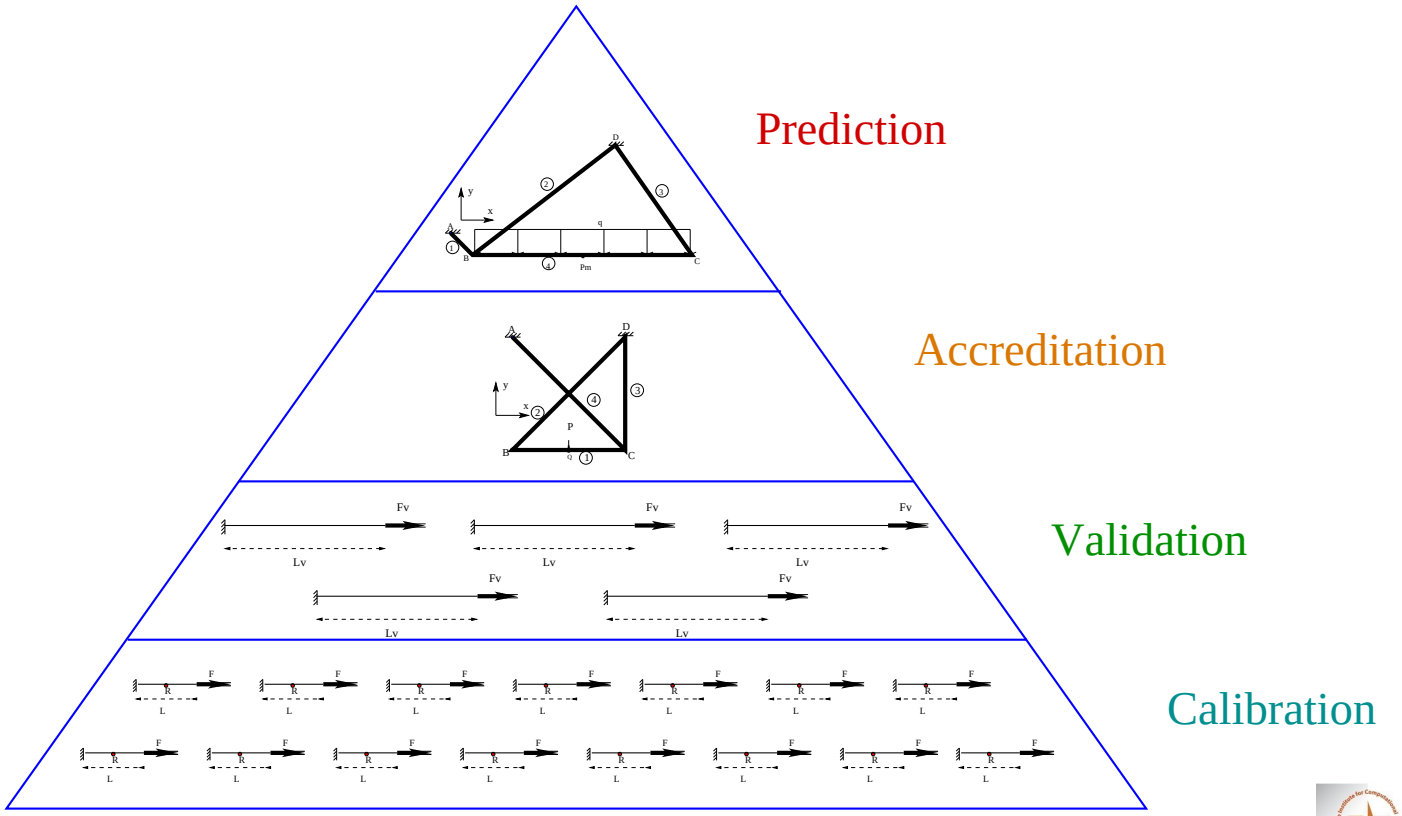
- Joints (hinges) are perfect.
- Supports in points A and D are perfectly rigid.
- Geometric data are completely accurate
- The load q is given by regulation
- Linear elasticity theory acceptable for the bars' deformation. In particular Kirchhoff bending theory can be used fo bar #4.
- Material properties (Young's modulus) significantly influence the target displacement (hence the reliability of the frame)
- Also, the material might be inhomogeneous.

Model structure: linear elasticity
Input data: elasticity coefficient

The validation pyramid

Challenge problem provided calibration / validation / accreditation experiments + experimental data (synthetic experiments).

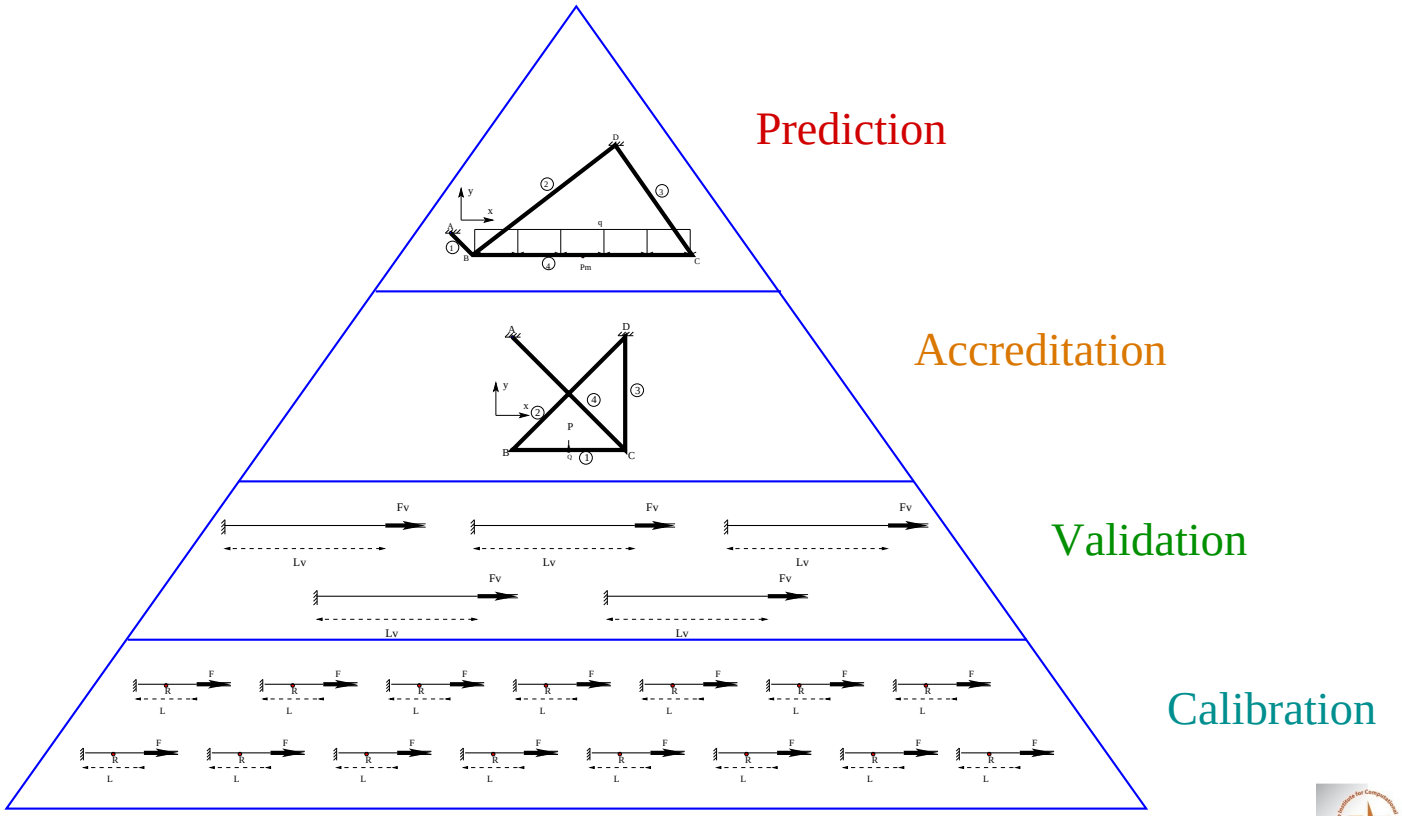
Three cases were proposed, with increasing number of data.



The validation pyramid

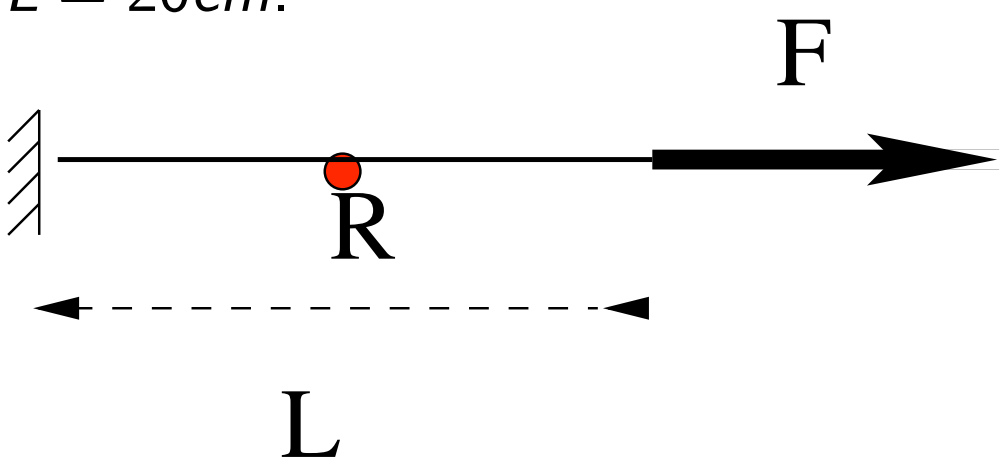
Challenge problem provided calibration / validation / accreditation experiments + experimental data (synthetic experiments).

Three cases were proposed, with increasing number of data.



Calibration

simple tensile test on a bar of length $L = 20cm$.



Quantities measured:

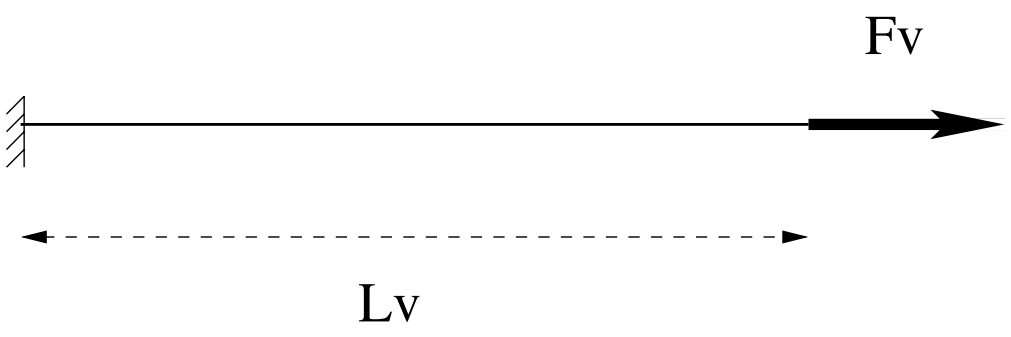
- Young's modulus $E(L/2)$ on the middle of the bar (by strain gage)
- Elongation δL of the bar

Sample #	δL(mm)	$E(L_c/2)$ (GPa)
1	5.15e-02	13.26
2	5.35e-02	10.86
3	5.24e-02	14.77
4	5.51e-02	10.94
5	5.14e-02	11.05
6	5.38e-02	11.06
7	4.97e-02	11.97
8	5.41e-02	11.66
9	4.95e-02	12.09
10	5.42e-02	11.30
11	5.47e-02	10.98
12	5.74e-02	11.92
13	5.36e-02	11.12
14	5.42e-02	12.00
15	5.34e-02	10.98
16	5.60e-02	10.71
17	5.06e-02	10.91
18	4.99e-02	11.89
19	5.22e-02	11.43
20	5.57e-02	10.87
21	5.28e-02	11.75
22	5.10e-02	13.47
23	5.48e-02	11.44
24	5.35e-02	12.44
25	4.92e-02	12.13
26	5.51e-02	11.38
27	5.27e-02	10.75
28	5.14e-02	11.98
29	5.61e-02	10.82
30	5.56e-02	11.04

Validation

Same as calibration on a longer bar of length $L_v = 80cm$.

Goal: assess the effect of inhomogeneities in the material



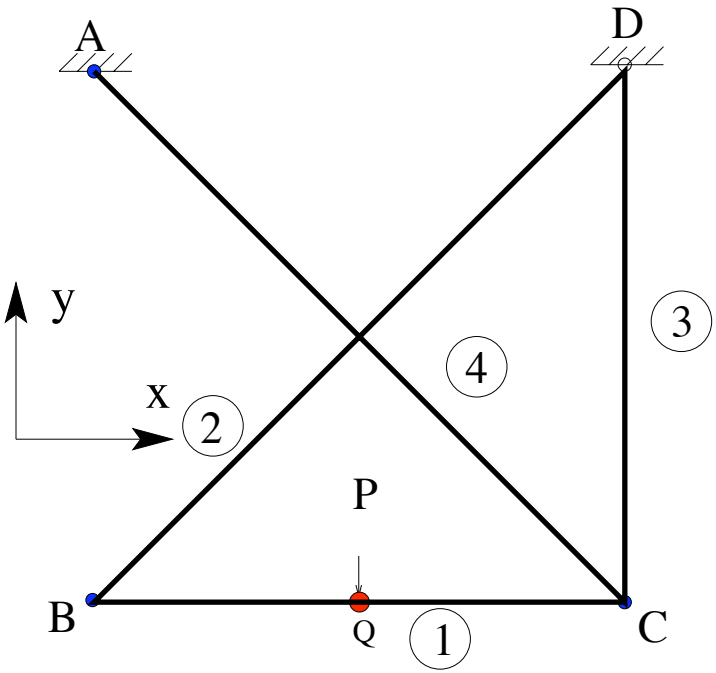
Quantity measured:

- Elongation δL_v of the bar

Sample #	δL(mm)
1	2.01e-01
2	2.06e-01
3	2.01e-01
4	2.08e-01
5	2.04e-01
6	2.01e-01
7	2.06e-01
8	2.11e-01
9	1.98e-01
10	2.08e-01

Accreditation

More complex frame



Sample #	$w(P)$(mm)
1	-6.50e-01
2	-6.73e-01

Case#1 and #2 have only one datum

Case #3 has two measurements

Quantity measured:

- Vertical displacement of the point Q

Outline

1. V&V framework
   - Introduction
   - Basic notions
   - A framework for V&V
   - Design of validation/accreditation experiments
2. SANDIA Static Frame Validation Problem
   - Definition of the problem
3. Solution procedure
   - Definition of the mathematical model
   - Calibration
   - Validation
   - Accreditation
   - Prediction
4. Conclusions
   - Conclusions

Models for elasticity coefficient

Let $G(x,\omega)$ be a Gaussian random field with zero mean, unit variance and (stationary) covariance function

$$Cov[G](x-y) = e^{-\left(\frac{|x-y|}{L_c}\right)^{\alpha}} \qquad (*)$$

Model for the material compliance $\mathcal{C} = 1/E$

$$\mathcal{C}(x,\omega) = F^{-1} \circ \Phi\left(G(x,\omega)\right) \qquad (**)$$

where Φ is the cumulative distribution of standard normal.

The function F determines the marginal distribution of $\mathcal{C}$ while the parameter L_c defines the correlation length.

The model structure is given in (*) and (**).

The input data are the function F and the parameters L_c and α

Models for elasticity coefficient

Let $G(x,\omega)$ be a Gaussian random field with zero mean, unit variance and (stationary) covariance function

$$Cov[G](x-y) = e^{-\left(\frac{|x-y|}{L_c}\right)^{\alpha}} \qquad (*)$$

Model for the material compliance $\mathcal{C} = 1/E$

$$\mathcal{C}(x,\omega) = F^{-1} \circ \Phi\left(G(x,\omega)\right) \qquad (**)$$

where Φ is the cumulative distribution of standard normal.

The function F determines the marginal distribution of $\mathcal{C}$ while the parameter L_c defines the correlation length.

The model structure is given in (*) and (**).

The input data are the function F and the parameters L_c and α

Models for elasticity coefficient

Let $G(x,\omega)$ be a Gaussian random field with zero mean, unit variance and (stationary) covariance function

$$Cov[G](x-y) = e^{-\left(\frac{|x-y|}{L_c}\right)^{\alpha}} \qquad (*)$$

Model for the material compliance $\mathcal{C} = 1/E$

$$\mathcal{C}(x,\omega) = F^{-1} \circ \Phi\left(G(x,\omega)\right) \qquad (**)$$

where Φ is the cumulative distribution of standard normal.

The function F determines the marginal distribution of $\mathcal{C}$ while the parameter L_c defines the correlation length.

The model structure is given in (*) and (**).

The input data are the function F and the parameters L_c and α

Models for elasticity coefficient

Let $G(x, \omega)$ be a Gaussian random field with zero mean, unit variance and (stationary) covariance function

$$Cov[G](x-y) = e^{-\left(\frac{|x-y|}{L_c}\right)^{\alpha}} \qquad (*)$$

Model for the material compliance $\mathcal{C} = 1/E$

$$\mathcal{C}(x, \omega) = F^{-1} \circ \Phi\left(G(x, \omega)\right) \qquad (**)$$

where Φ is the cumulative distribution of standard normal.

The function F determines the marginal distribution of $\mathcal{C}$ while the parameter L_c defines the correlation length.

The model structure is given in (*) and (**).

The input data are the function F and the parameters L_c and α.

Models for elasticity coefficient

We focused on the following models:

Marginal distribution: the function F is chosen to obtain 5 different distributions

- 4 parametric distributions: Uniform, Normal, Log-Normal, Inverse Log-Normal
- 1 non parametric distribution: KDE (kernel density function) with Gaussian kernel.

Covariance structure

- fully correlated: $\alpha = 0, L_c > 0$
- partially correlated: $\alpha = 2, L_c > 0$
- perfectly uncorrelated: $\alpha > 0, L_c = 0$

In total we have 15 models to calibrate and validate.

Models for elasticity coefficient

We focused on the following models:

Marginal distribution: the function F is chosen to obtain 5 different distributions

- 4 parametric distributions: Uniform, Normal, Log-Normal, Inverse Log-Normal
- 1 non parametric distribution: KDE (kernel density function) with Gaussian kernel.

Covariance structure

- fully correlated: $\alpha = 0, L_c > 0$
- partially correlated: $\alpha = 2, L_c > 0$
- perfectly uncorrelated: $\alpha > 0, L_c = 0$

In total we have 15 models to calibrate and validate.

Models for elasticity coefficient

We focused on the following models:

Marginal distribution: the function F is chosen to obtain 5 different distributions

- 4 parametric distributions: Uniform, Normal, Log-Normal, Inverse Log-Normal
- 1 non parametric distribution: KDE (kernel density function) with Gaussian kernel.

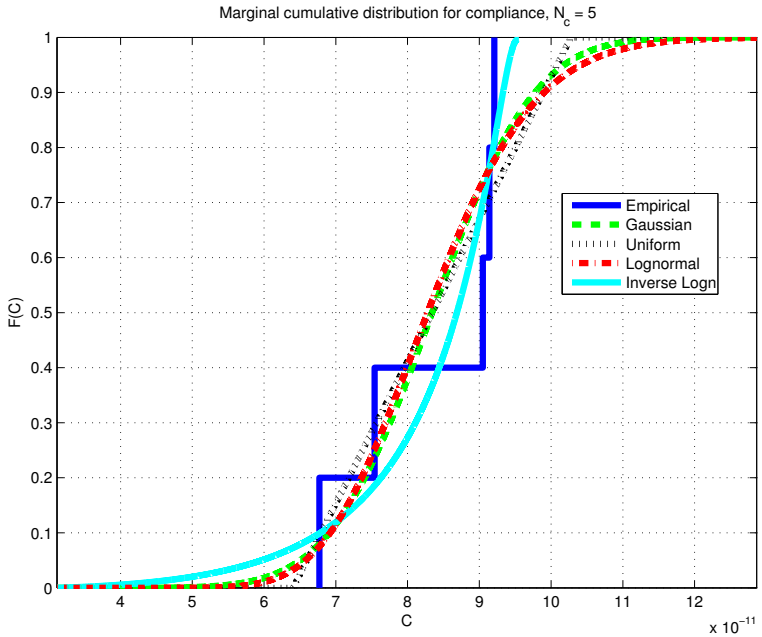
Covariance structure

- fully correlated: $\alpha = 0, L_c > 0$
- partially correlated: $\alpha = 2, L_c > 0$
- perfectly uncorrelated: $\alpha > 0, L_c = 0$

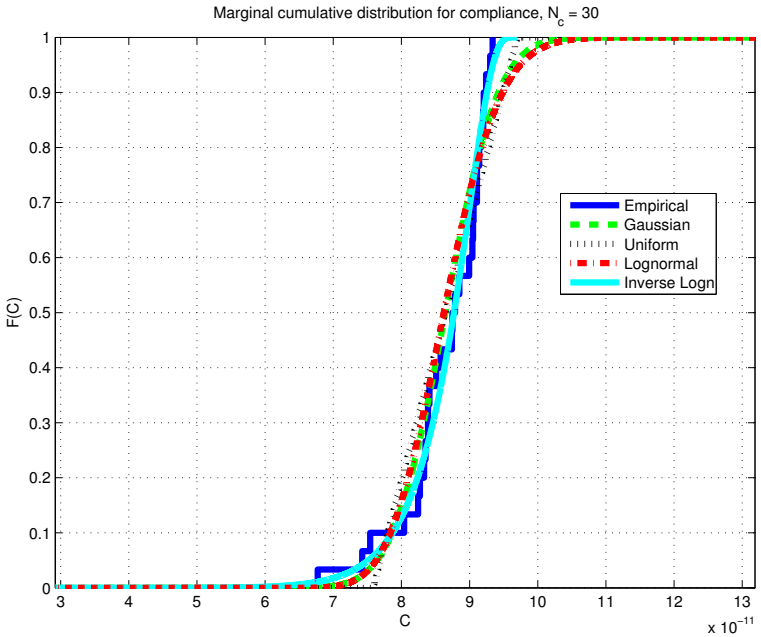
In total we have 15 models to calibrate and validate.

Calibrating the marginal distribution

For the 4 parametric distributions the parameters are chosen to fit the first two moments of the compliance data



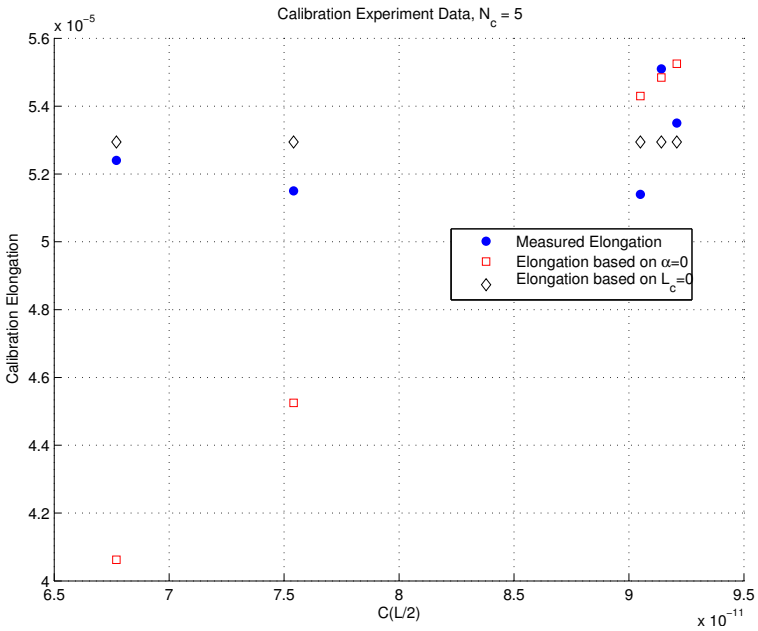
Case 1 ($N_c = 5$)



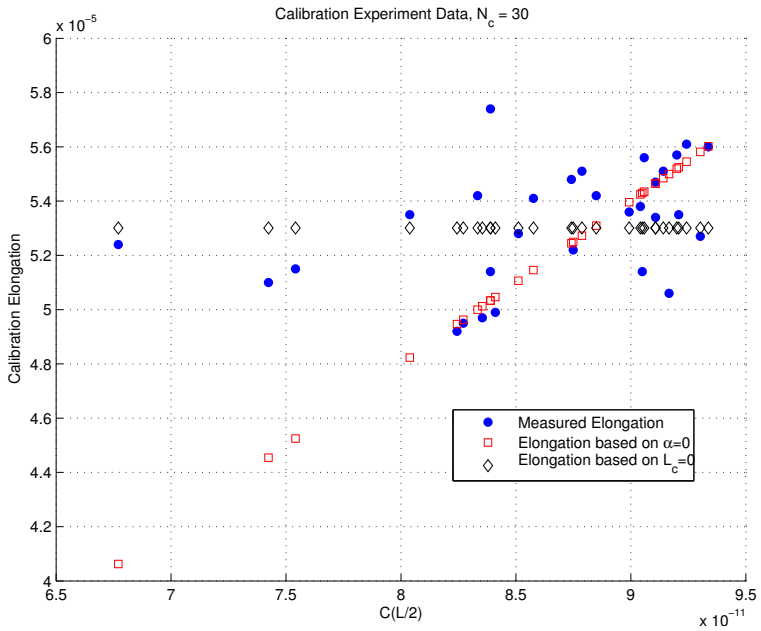
Case 3 ($N_c = 30$)

Calibrating the correlation length

To fit the correlation length we use both sets of calibration data: pointwise compliance $\mathcal{C}(L/2)$ and elongation δL.



Case 1 ($N_c = 5$)



Case 3 ($N_c = 30$)

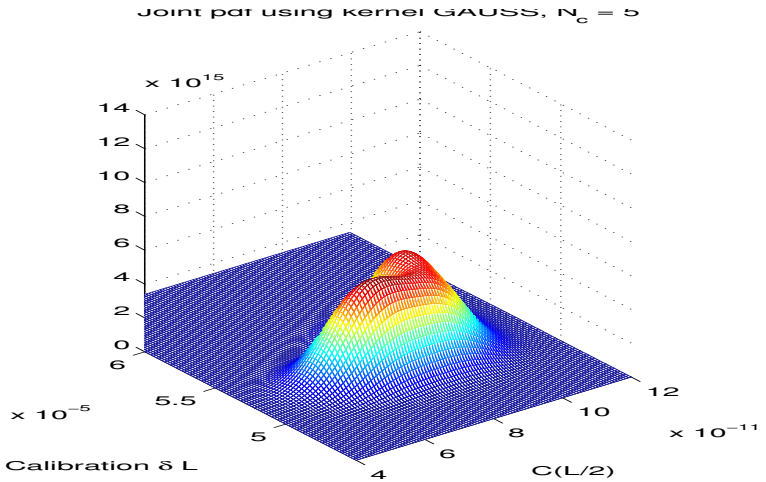
- The parameter L_c has to be fitted only for the partially correlated model.

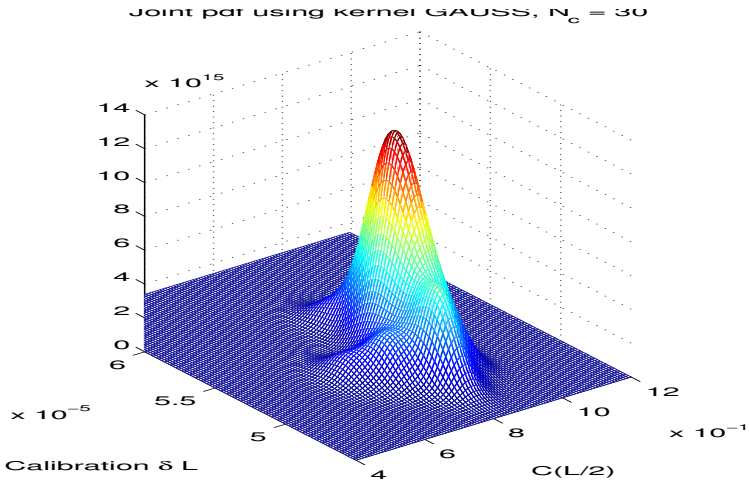
Case	L_c (m)
1, $N_c = 5$	0.009
2, $N_c = 20$	0.030
3, $N_c = 30$	0.035

Variability in the fitted data

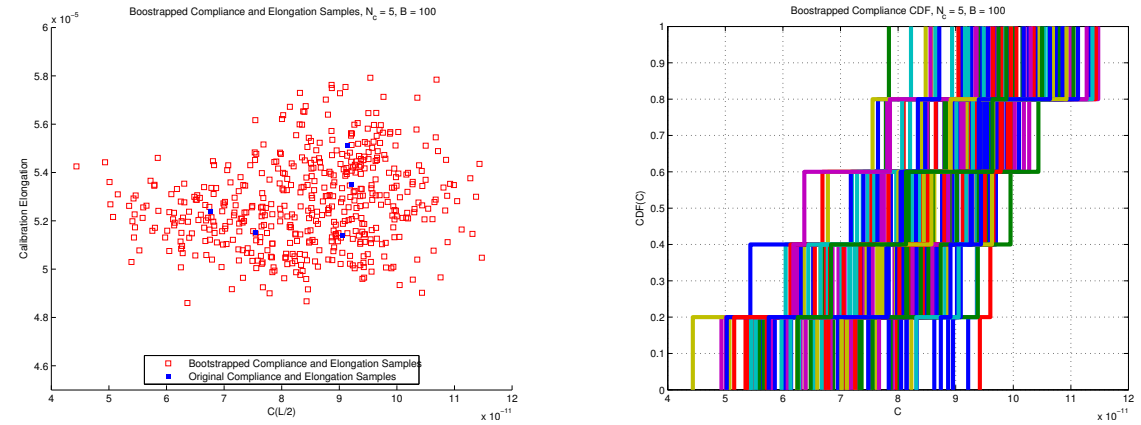
We estimated the variability in the fitted data by a bootstrapping procedure:

- Reconstruct the joint pdf of $\mathcal{C}(L/2)$ and δL by a kernel density estimate starting from the N_c calibration measurements.
- Draw samples of size N_c from the reconstructed pdf and, for each drawing, fit the parameters of the compliance model.
- Repeat this procedure $B = 1000$ times and get distributions for the fitted parameters.

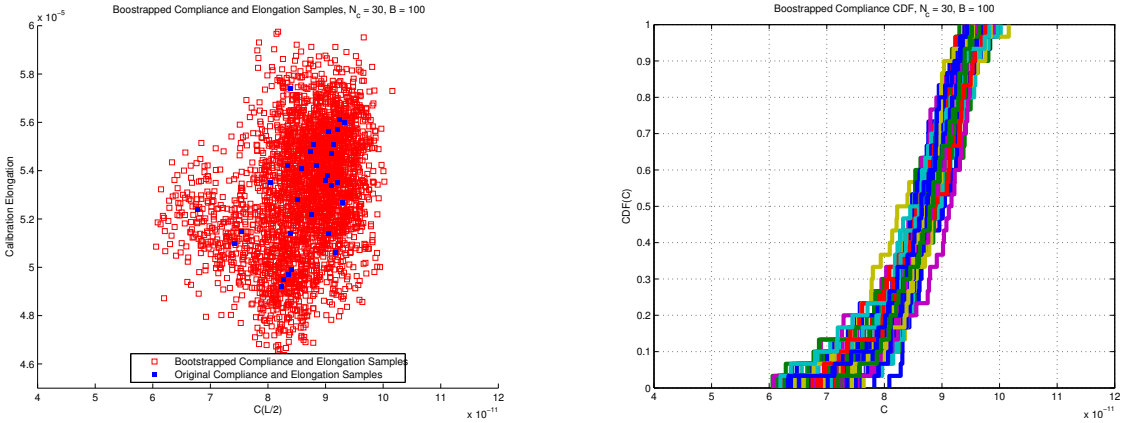




Variability in the fitted data



Case 1 ($N_c = 5$)



Case 3 ($N_c = 30$)

On the left: 100 bootstrapped samples from the reconstructed joint pdf of $C(L/2)$ and δL. On the right: bootstrapped empirical cumulative distribution of the compliance.

The validation process

Due to the small amount of validation data we decided to adopt a Bayesian approach. The validation process is as follows

a) Using the calibrated model, the probability field of the validation elongation $Y(\Theta^{cal})$ is constructed, typically by the Monte Carlo method;

b) Using the available validation measurements a Bayesian update of the compliance model parameters Θ^{up} is computed.

c) Using both the calibrated (Θ^{cal}) and updated (Θ^{up}) model, the probability distribution of the quantity of interest in the prediction problem is constructed. The distance (validation metric) is based on a comparison between the two cumulative distributions.

Validation metric – notion of distance

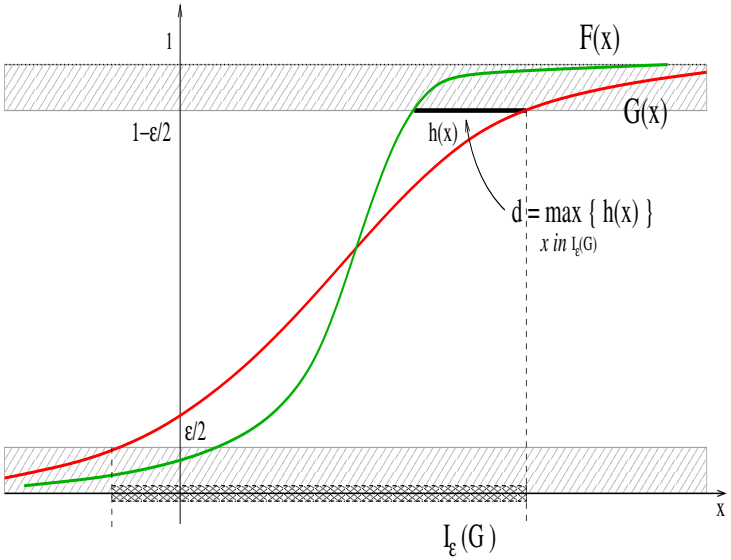
We want to compute a failure probability $F_Y(\bar{y}) = P(Y < \bar{y})$. We compare (at the prediction level) the two cumulative distributions $F_{Y^{cal}}$ and $F_{Y^{up}}$.

Definition (Horizontal distance)

Let F and G be two cumulative distributions. Define

$$d_\varepsilon(F, G) := \max_{x \in \mathcal{I}_\varepsilon(G)} |F^{-1} \circ G(x) - x|$$

with $\mathcal{I}_\varepsilon(G) \equiv \{x \in \mathbb{R} \mid \frac{\varepsilon}{2} \leq G(x) \leq 1 - \frac{\varepsilon}{2}\}$.



Definition (Rejection criterion)

Given a tolerance level *tol* we reject the model Y^{cal} if

$$d_\varepsilon(F_{Y^{up}}, F_{Y^{cal}}) \geq tol \times 3mm$$

Validation metric – notion of distance

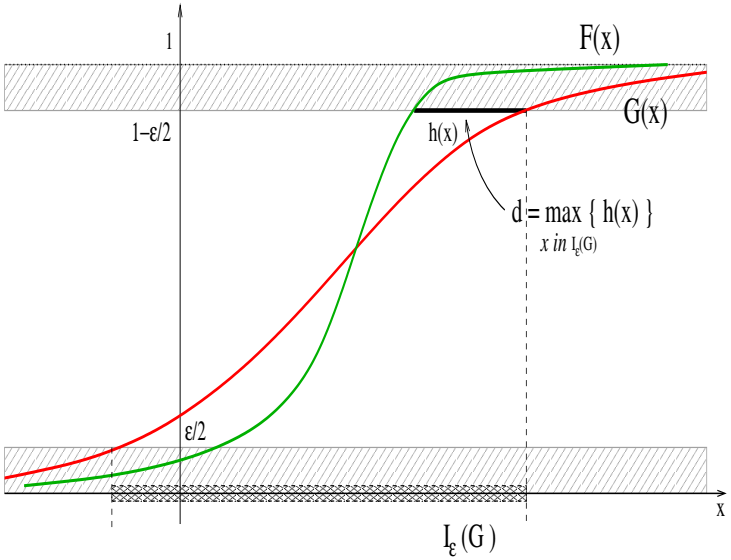
We want to compute a failure probability $F_Y(\bar{y}) = P(Y < \bar{y})$. We compare (at the prediction level) the two cumulative distributions $F_{Y^{cal}}$ and $F_{Y^{up}}$.

Definition (Horizontal distance)

Let F and G be two cumulative distributions. Define

$$d_\varepsilon(F, G) := \max_{x \in \mathcal{I}_\varepsilon(G)} |F^{-1} \circ G(x) - x|$$

with $\mathcal{I}_\varepsilon(G) \equiv \{x \in \mathbb{R} \mid \frac{\varepsilon}{2} \leq G(x) \leq 1 - \frac{\varepsilon}{2}\}$.



Definition (Rejection criterion)

Given a tolerance level *tol* we reject the model Y^{cal} if

$$d_\varepsilon(F_{Y^{up}}, F_{Y^{cal}}) \geq tol \times 3mm$$

Bayesian update of the calibrated model

The posterior distribution is defined as

$$p(\Theta|\mathbf{Y}) = cL(\Theta|\mathbf{Y})p(\Theta).$$

We have considered as an updated value Θ^{up} the maximum likelihood estimator

$$\Theta^{up} = \underset{\Theta}{\operatorname{argmax}}\ p(\Theta|\mathbf{Y}).$$

- In principle Θ^{up} can be computed numerically by solving an optimization problem.
- We have taken a shortcut: approximate $p(Y|\Theta)$ with a Gaussian (which is a sound approximation). Then all calculations can by done by hand.

Bayesian update of the calibrated model

The posterior distribution is defined as

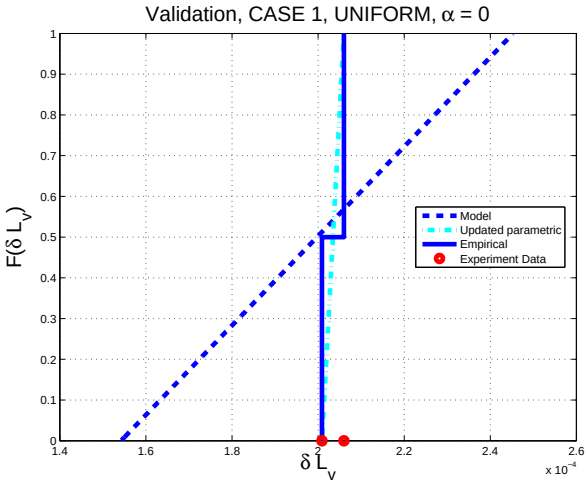
$$p(\Theta|\mathbf{Y}) = cL(\Theta|\mathbf{Y})p(\Theta).$$

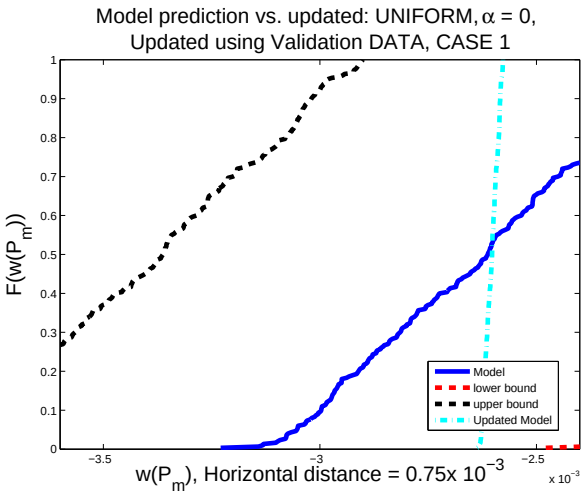
We have considered as an updated value Θ^{up} the maximum likelihood estimator

$$\Theta^{up} = \underset{\Theta}{\operatorname{argmax}}\ p(\Theta|\mathbf{Y}).$$

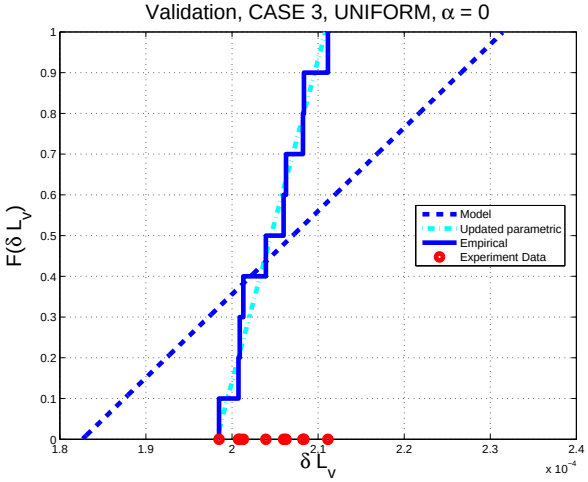
- In principle Θ^{up} can be computed numerically by solving an optimization problem.
- We have taken a shortcut: approximate $p(Y|\Theta)$ with a Gaussian (which is a sound approximation). Then all calculations can by done by hand.

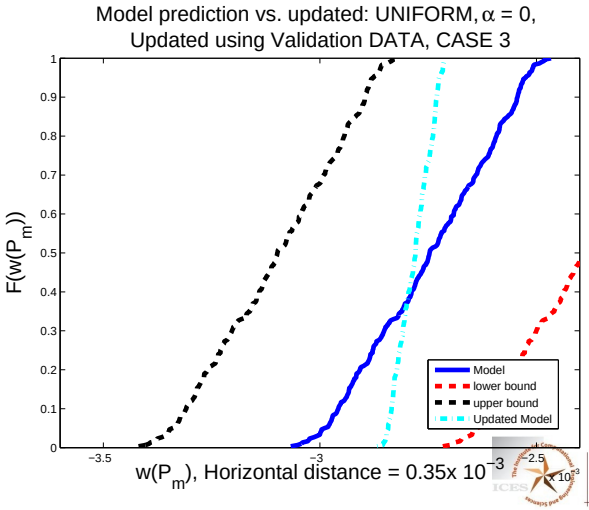
Validation results – fully correlated model





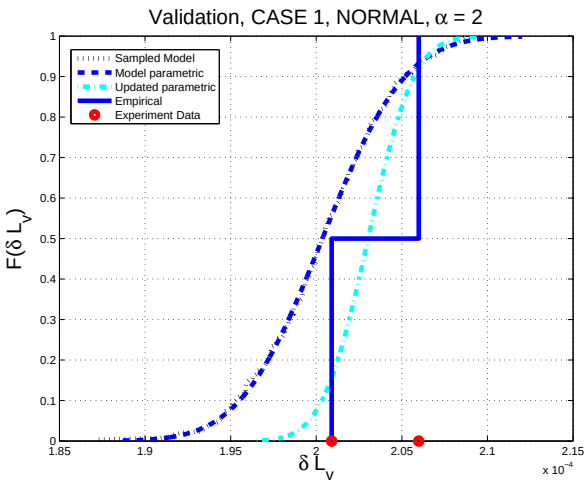
Case 1: $N_c = 5,\ N_v = 2$

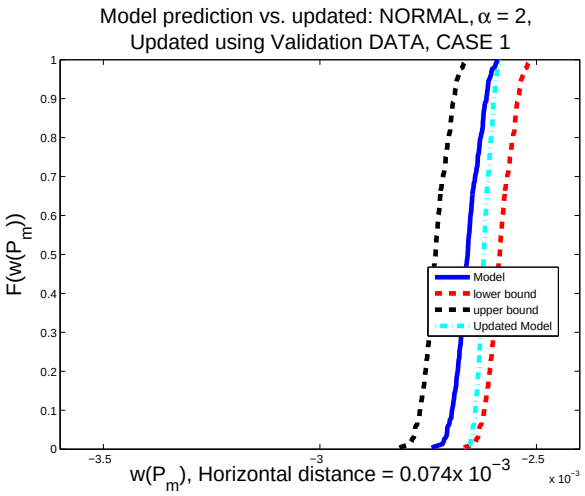




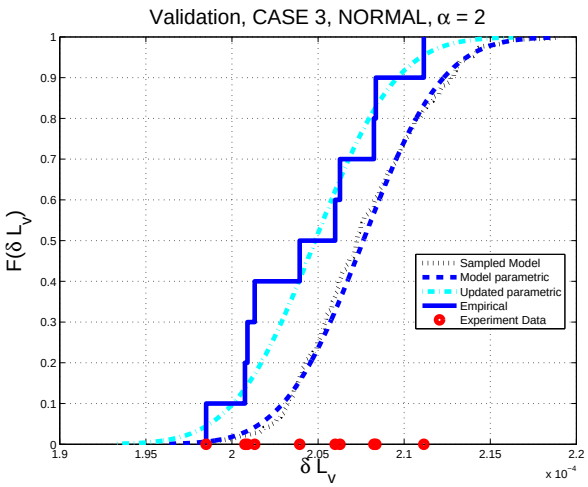
Case 3: $N_c = 30,\ N_v = 10$

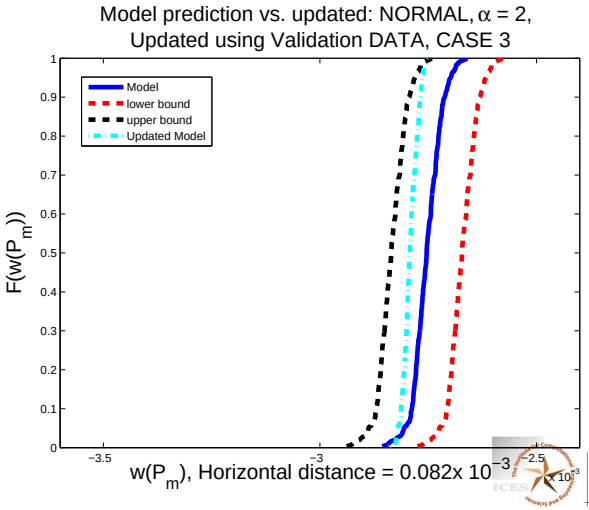
Validation results – partially correlated model





Case 1: $N_c = 5,\ N_v = 2$





Case 3: $N_c = 30,\ N_v = 10$

Validation results – computed distances

$(N_c, N_v) =$	Case 1 $(5, 2)$	Case 2 $(20, 4)$	Case 3 $(30, 10)$
	(a) *Fully correlated model* $\alpha = 0$		
Uniform	**0.25**	**0.14**	**0.12**
Normal	**0.25**	**0.13**	**0.11**
Inverse LN	**0.25**	**0.17**	**0.12**
	(b) *Partially correlated model* $\alpha = 2$		
Uniform	0.02	0.04	0.025
Normal	0.03	0.035	0.025
Inverse LN	0.025	0.035	0.03
	(c) *Perfectly uncorrelated model* $L_c = 0$		
Effective model	0.07	0.08	0.09

Ratio of the horizontal distances and the critical prediction displacement.
Models are rejected is the normalized distance is larger than $tol = 0.1$.

Accreditation

Similar to Validation.

- We adopted also in this case a Bayesian approach.
- Since the accreditation problem is more complex than the validation one, the likelihood function is more difficult to construct. Also in this case we have employed a Gaussian approximation.

$(N_c, N_a) =$	Case 1 $(5,1)$	Case 2 $(20,1)$	Case 3 $(30,2)$
	(a) *Partially correlated model* $\alpha = 2$		
Uniform	0.015	0.028	0.037
Normal	0.015	0.028	0.037
Inverse LN	0.015	0.026	0.033
	(b) *Perfectly uncorrelated model* $L_c = 0$		
Effective model	0.072	0.072	0.1

Accreditation

Similar to Validation.

- We adopted also in this case a Bayesian approach.
- Since the accreditation problem is more complex than the validation one, the likelihood function is more difficult to construct. Also in this case we have employed a Gaussian approximation.

$(N_c, N_a) =$	Case 1 $(5,1)$	Case 2 $(20,1)$	Case 3 $(30,2)$
	(a) *Partially correlated model* $\alpha = 2$		
Uniform	0.015	0.028	0.037
Normal	0.015	0.028	0.037
Inverse LN	0.015	0.026	0.033
	(b) *Perfectly uncorrelated model* $L_c = 0$		
Effective model	0.072	0.072	0.1

Prediction

- The fully correlated model has been rejected in the validation process.
- Both the partially correlated and the perfectly uncorrelated models have been validated, for any choice of marginal distribution. They can then be used for prediction.
- The final prediction should include
  - the "distance" information coming from validation and accreditation. It is a measure of the modeling error,
  - the variability on the input data measured by bootstrapping.

How to include bootstrapping information: we want to compute the quantity

$$P(|w(P_m, \Theta)| \geq 3mm) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m, \Theta(\tilde{\omega}), \omega)| \geq 3(mm)\}} dQ(\omega) \, d\tilde{P}(\tilde{\omega})$$

We repeat the prediction computation B times, every time generating new calibration data by bootstrapping and corresponding calibrated parameter Θ^{cal}.

Prediction

- The fully correlated model has been rejected in the validation process.
- Both the partially correlated and the perfectly uncorrelated models have been validated, for any choice of marginal distribution. They can then be used for prediction.

- The final prediction should include
  - the "distance" information coming from validation and accreditation. It is a measure of the modeling error,
  - the variability on the input data measured by bootstrapping.

How to include bootstrapping information: we want to compute the quantity

$$P(|w(P_m, \Theta)| \geq 3mm) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m, \Theta(\tilde{\omega}), \omega)| \geq 3(mm)\}} dQ(\omega) \, d\tilde{P}(\tilde{\omega})$$

We repeat the prediction computation B times, every time generating new calibration data by bootstrapping and corresponding calibrated parameter Θ^{cal}.

Prediction

- The fully correlated model has been rejected in the validation process.
- Both the partially correlated and the perfectly uncorrelated models have been validated, for any choice of marginal distribution. They can then be used for prediction.
- The final prediction should include
  - the "distance" information coming from validation and accreditation. It is a measure of the modeling error,
  - the variability on the input data measured by bootstrapping.

How to include bootstrapping information: we want to compute the quantity

$$P(|w(P_m, \Theta)| \geq 3mm) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m, \Theta(\tilde{\omega}), \omega)| \geq 3(mm)\}} dQ(\omega) \, d\tilde{P}(\tilde{\omega})$$

We repeat the prediction computation B times, every time generating new calibration data by bootstrapping and corresponding calibrated parameter Θ^{cal}.

Prediction

- The fully correlated model has been rejected in the validation process.
- Both the partially correlated and the perfectly uncorrelated models have been validated, for any choice of marginal distribution. They can then be used for prediction.
- The final prediction should include
  - the "distance" information coming from validation and accreditation. It is a measure of the modeling error,
  - the variability on the input data measured by bootstrapping.

How to include bootstrapping information: we want to compute the quantity

$$P(|w(P_m, \Theta)| \geq 3mm) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m, \Theta(\tilde{\omega}), \omega)| \geq 3(mm)\}} dQ(\omega) d\tilde{P}(\tilde{\omega})$$

We repeat the prediction computation B times, every time generating new calibration data by bootstrapping and corresponding calibrated parameter Θ^{cal}.

Prediction

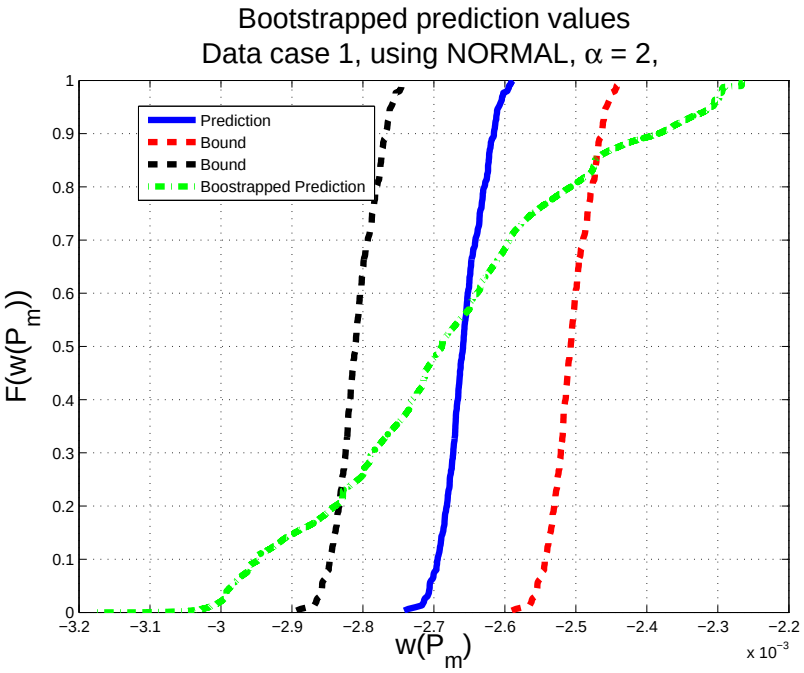
- The fully correlated model has been rejected in the validation process.
- Both the partially correlated and the perfectly uncorrelated models have been validated, for any choice of marginal distribution. They can then be used for prediction.
- The final prediction should include
  - the "distance" information coming from validation and accreditation. It is a measure of the modeling error,
  - the variability on the input data measured by bootstrapping.

How to include bootstrapping information: we want to compute the quantity

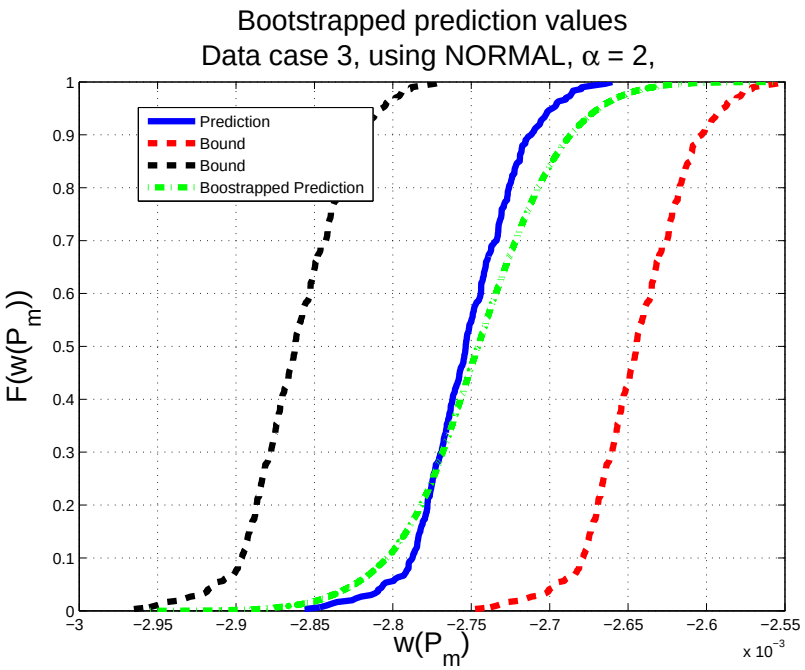
$$P(|w(P_m,\Theta)| \geq 3mm) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m,\Theta(\tilde{\omega}),\omega)| \geq 3(mm)\}} dQ(\omega) d\tilde{P}(\tilde{\omega})$$

We repeat the prediction computation B times, every time generating new calibration data by bootstrapping and corresponding calibrated parameter Θ^{cal}.

Prediction results – partially correlated model



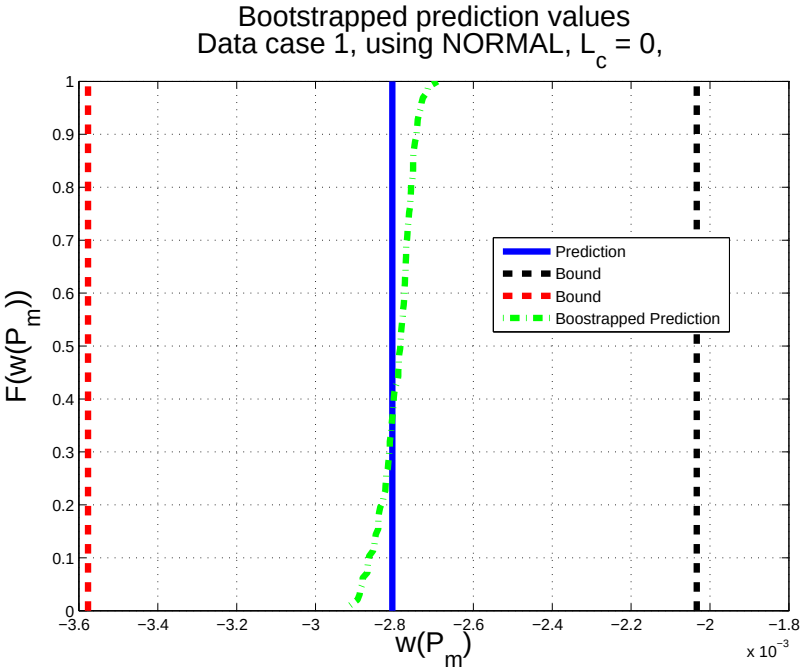
Case 1: $(N_c, N_v, N_a) = (5, 2, 1)$



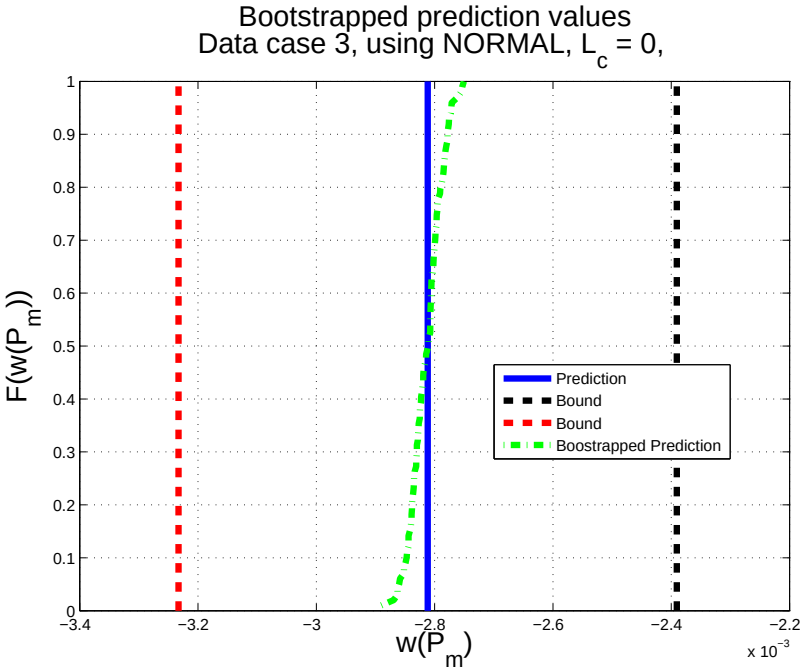
Case 3: $(N_c, N_v, N_a) = (30, 10, 2)$

The cumulative distribution of the predicted quantity of interest obtained by the calibrated model (blue), the bootstrapped model (green) and the bounds. Partially correlated model with Normal marginal distribution.

Prediction results – perfectly uncorrelated model



Case 1: $(N_c, N_v, N_a) = (5, 2, 1)$



Case 3: $(N_c, N_v, N_a) = (30, 10, 2)$

The cumulative distribution of the predicted quantity of interest by the calibrated (blue) and the bootstrapped model (green) for the perfectly uncorrelated model.

Prediction results – failure probability

Failure probability computed with the partially correlated model including:

bootstrapping

$(N_c, N_v, N_a) =$	Case 1 $(5, 2, 1)$	Case 2 $(20, 4, 1)$	Case 3 $(30, 10, 2)$
Uniform	2.0×10^{-2}	4.7×10^{-5}	5.6×10^{-8}
Normal	2.1×10^{-2}	5.8×10^{-5}	8.6×10^{-8}
Inverse LN	1.4×10^{-2}	4.6×10^{-5}	5.6×10^{-8}

bootstrapping + validation and accreditation distances

$(N_c, N_v, N_a) =$	Case 1 $(5, 2, 1)$	Case 2 $(20, 4, 1)$	Case 3 $(30, 10, 2)$
Uniform	0.18	1.8×10^{-2}	3.0×10^{-3}
Normal	0.19	2.1×10^{-2}	2.1×10^{-3}
Inverse LN	0.19	1.7×10^{-2}	1.4×10^{-3}

Outline

1. V&V framework
   - Introduction
   - Basic notions
   - A framework for V&V
   - Design of validation/accreditation experiments
2. SANDIA Static Frame Validation Problem
   - Definition of the problem
3. Solution procedure
   - Definition of the mathematical model
   - Calibration
   - Validation
   - Accreditation
   - Prediction
4. Conclusions
   - Conclusions

Conclusions

- We have presented a general framework for Verification and Validation (V& V) in Computational Science
- The main idea is to update the model using the validation data and to measure the distance between calibrated and updated prediction. The rejection criterion is defined on the prediction level directly.
- The most crucial step is how to design the validation pyramid: calibration/validation/accreditation experiments, metrics and tolerances for rejection.
- We have illustrated and implemented on the SANDIA Challenge Validation Problem the steps for model validation. The link between validation experiments and prediction has been achieved via Bayesian updating.
- The result are encouraging; the bad models have been rejected, as expected.
- There are still a lot of interesting mathematical questions to be investigated. Space for research .....