Bayesian Optimal Experimental Design Framework for Data-Driven Uncertainty Quantification (UQ) in Dynamical Systems

Costas Papadimitriou

University of Thessaly, Greece

Acknowledgements:

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 764547.

Outline

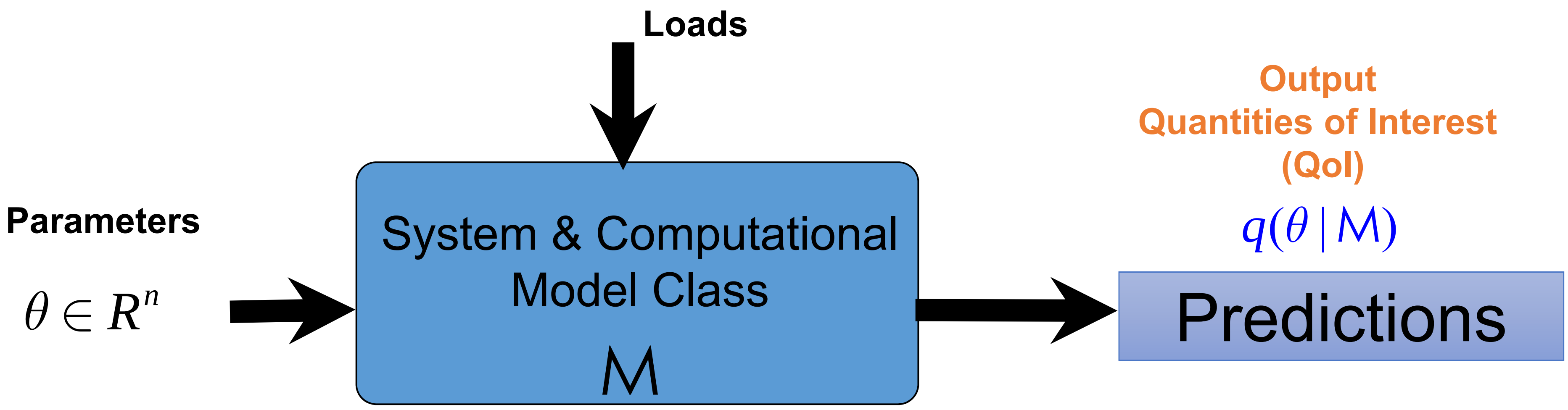
Bayesian Learning Framework for Data-Driven UQ

Bayesian Optimal Experimental Design

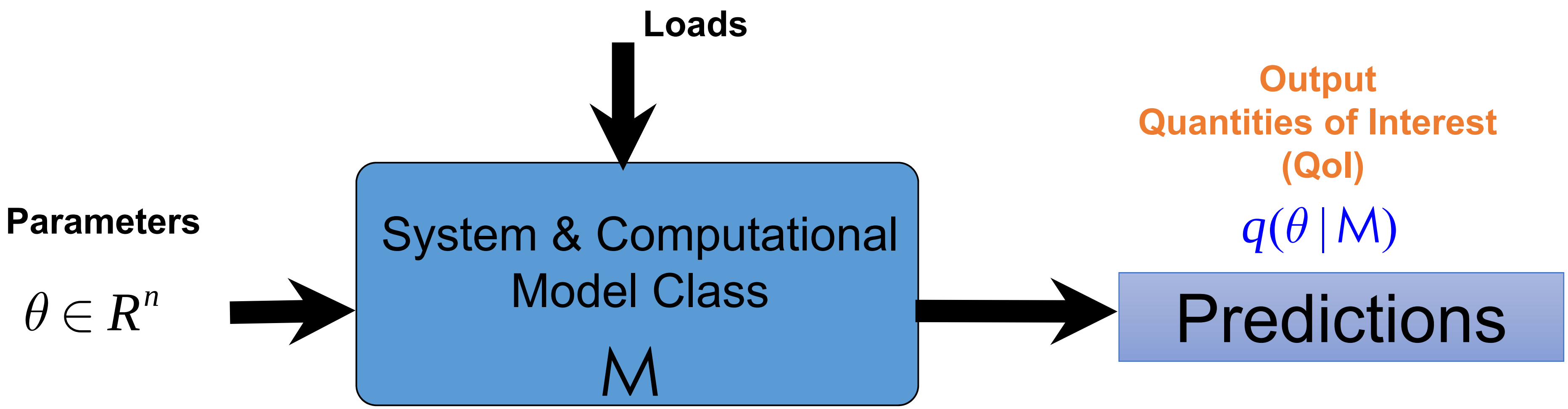
Applications in Structural Dynamics

Conclusions

Models in Structural Dynamics



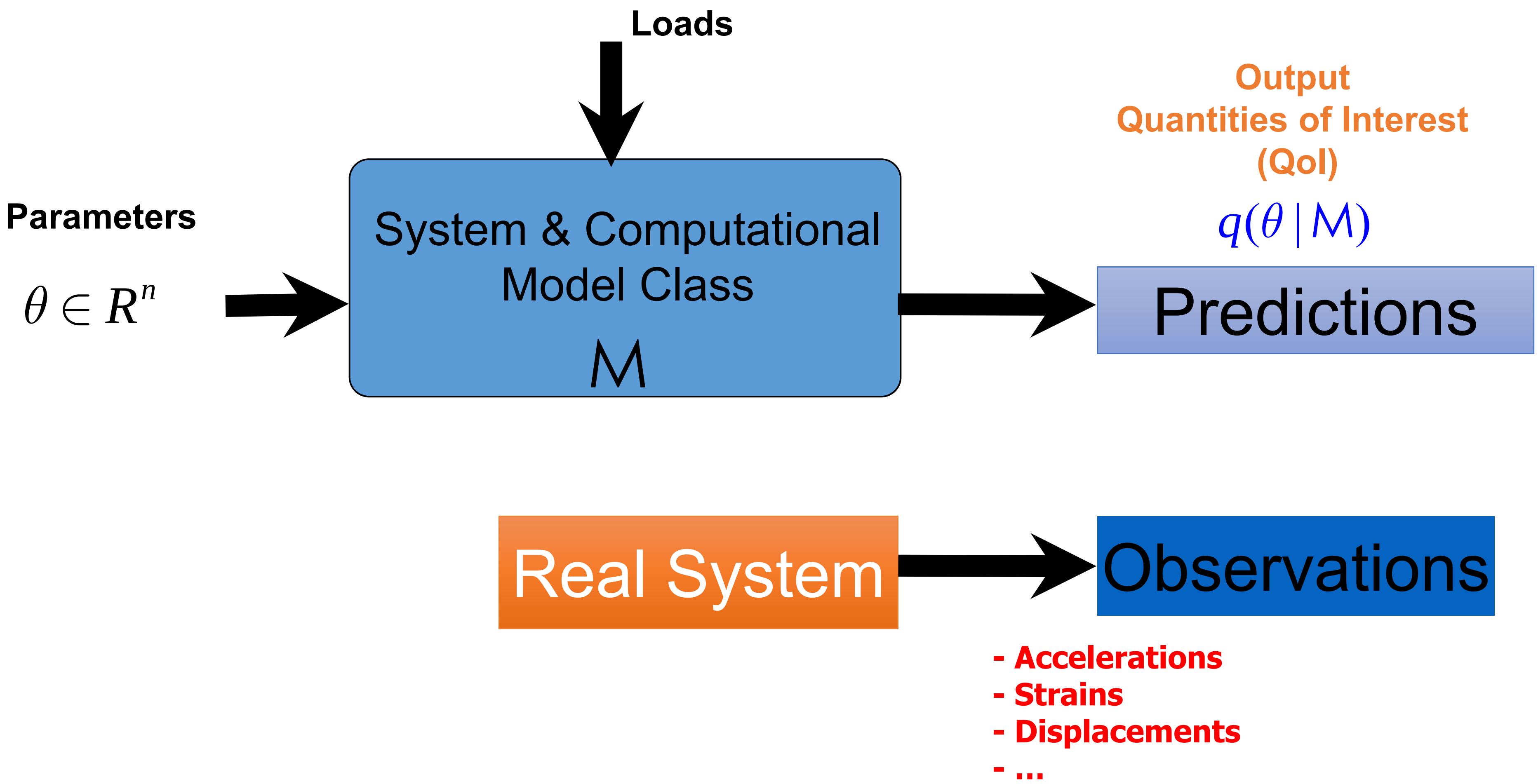
Models in Structural Dynamics



- **Equations of motion in structural dynamics**
  - **Linear or nonlinear (elastic, plastic, sliding, impact, etc)**
  - **Loads known or unknown**

- **PARAMETERS**
  - **Stiffness, restoring force models, mass, boundary conditions, damping, friction)**
  - **Spectral characteristics of unknown loads**

Data-Driven Modeling and UQ



Optimal Experimental Design Goal

Optimal Experimental Design is used to select the "best" experimental setup out of many alternative ones.

"Best" is the one that yields most informative measurements for

- **Selecting models and inferring model parameters**
- **Estimating output QoI (e.g. response reconstruction, stresses)**

<u>Design Variables</u>

- **Type, location and number (density) of sensors and/or actuators**
- **Excitation characteristics (e.g. frequency content, amplitude)**

Bayesian Learning Framework for Data-Driven UQ

A. PARAMETER ESTIMATION

Posterior Likelihood Prior

$$p(\theta \mid y) = \frac{p(y \mid \theta) \;\; \pi(\theta)}{p(y)}$$

Evidence

Prediction Error Model

$$y = q(\theta) + e$$

$$e = e^e + e^m \sim N(0, \Sigma)$$

Likelihood

$$p(y \mid \theta) = \frac{|\Sigma(\theta)|^{-1/2}}{(2\pi)^n} \exp\left[-\frac{1}{2}[y - q(\theta)]^T \Sigma^{-1}(\theta)[y - q(\theta)]\right]$$

B. PREDICTIONS

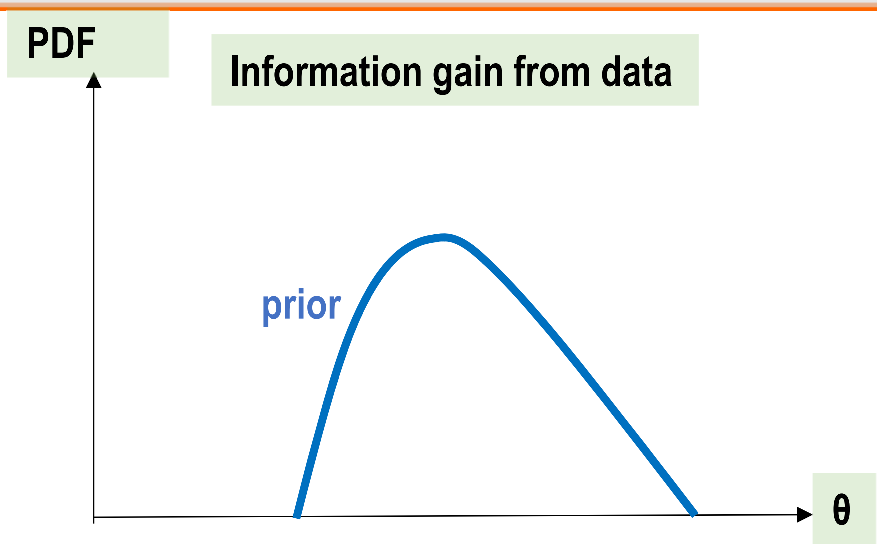
$$p(z \mid y) = \int p(z \mid \theta) \, p(\theta \mid y) \, d\theta$$

Bayesian Learning Framework for Data-Driven UQ

A. PARAMETER ESTIMATION

Posterior, Likelihood, Prior, Evidence

$$p(\theta \mid y) = \frac{p(y \mid \theta)\ \pi(\theta)}{p(y)}$$



Prediction Error Model

$$y = q(\theta) + e$$

$$e = e^e + e^m \sim N(0, \Sigma)$$

Likelihood

$$p(y \mid \theta) = \frac{|\Sigma(\theta)|^{-1/2}}{(2\pi)^n} \exp\left[-\frac{1}{2}[y - q(\theta)]^T \Sigma^{-1}(\theta)[y - q(\theta)]\right]$$

B. PREDICTIONS

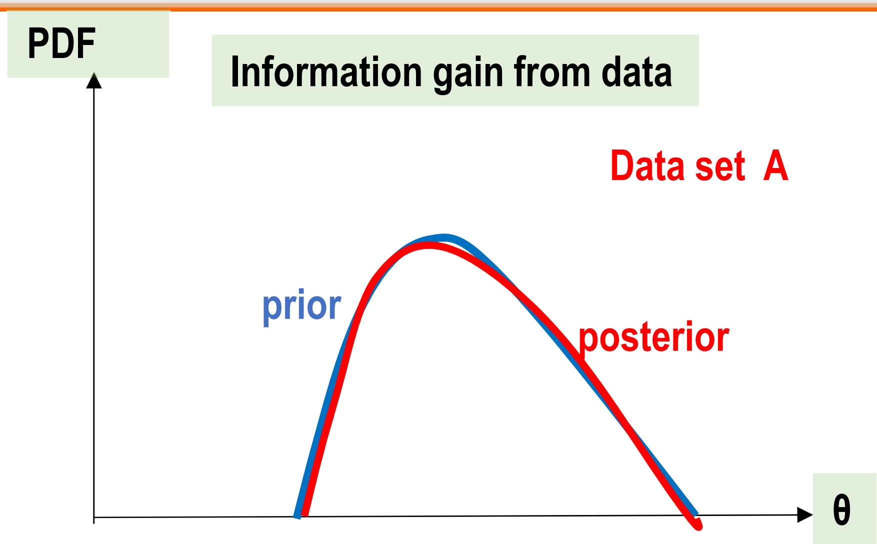
$$p(z \mid y) = \int p(z \mid \theta)\, p(\theta \mid y)\, d\theta$$

Bayesian Learning Framework for Data-Driven UQ

A. PARAMETER ESTIMATION

Posterior, Likelihood, Prior, Evidence

$$p(\theta \mid y) = \frac{p(y \mid \theta)\ \pi(\theta)}{p(y)}$$



Prediction Error Model

$$y = q(\theta) + e$$

$$e = e^e + e^m \sim N(0, \Sigma)$$

Likelihood

$$p(y \mid \theta) = \frac{|\Sigma(\theta)|^{-1/2}}{(2\pi)^n} \exp\left[-\frac{1}{2} [y - q(\theta)]^T \Sigma^{-1}(\theta) [y - q(\theta)] \right]$$

B. PREDICTIONS

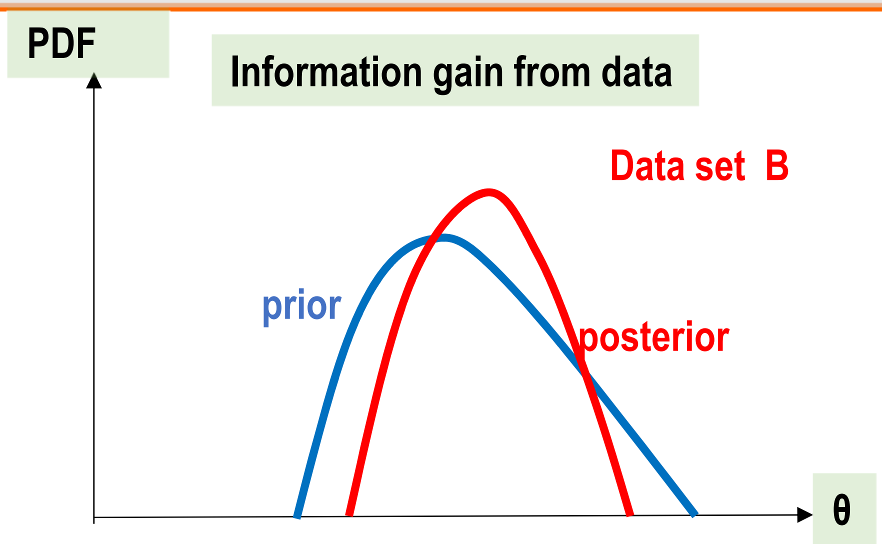
$$p(z \mid y) = \int p(z \mid \theta)\ p(\theta \mid y)\ d\theta$$

Bayesian Learning Framework for Data-Driven UQ

A. PARAMETER ESTIMATION

Posterior, Likelihood, Prior, Evidence

$$p(\theta \mid y) = \frac{p(y \mid \theta) \;\; \pi(\theta)}{p(y)}$$



Prediction Error Model

$$y = q(\theta) + e$$

$$e = e^e + e^m \sim N(0, \Sigma)$$

Likelihood

$$p(y \mid \theta) = \frac{|\Sigma(\theta)|^{-1/2}}{(2\pi)^n} \exp\left[-\frac{1}{2} [y - q(\theta)]^T \Sigma^{-1}(\theta) [y - q(\theta)] \right]$$

B. PREDICTIONS

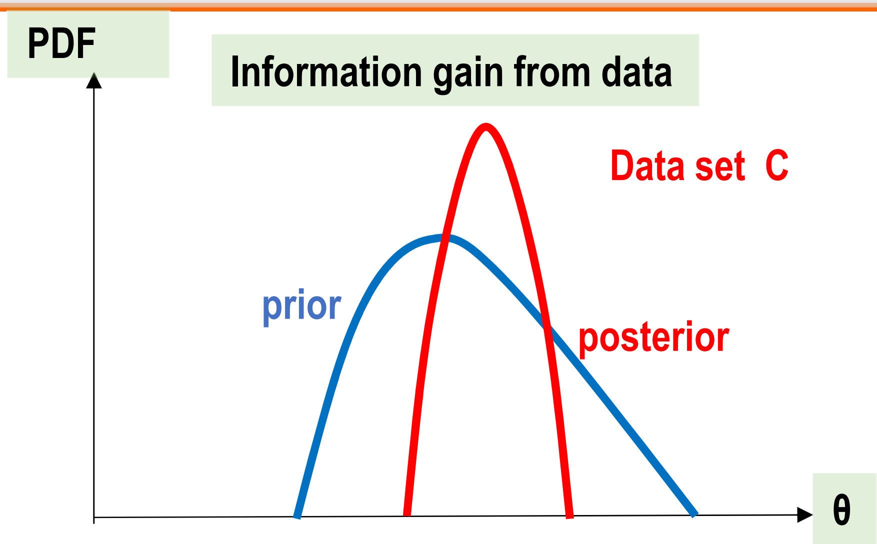
$$p(z \mid y) = \int p(z \mid \theta) \, p(\theta \mid y) \, d\theta$$

Bayesian Learning Framework for Data-Driven UQ

A. PARAMETER ESTIMATION

Posterior, Likelihood, Prior, Evidence

$$p(\theta \mid y) = \frac{p(y \mid \theta)\ \pi(\theta)}{p(y)}$$



Prediction Error Model

$$y = q(\theta) + e$$

$$e = e^{e} + e^{m} \sim N(0, \Sigma)$$

Likelihood

$$p(y \mid \theta) = \frac{|\Sigma(\theta)|^{-1/2}}{(2\pi)^{n}} \exp\left[-\frac{1}{2}[y - q(\theta)]^{T} \Sigma^{-1}(\theta)[y - q(\theta)]\right]$$

B. PREDICTIONS

$$p(z \mid y) = \int p(z \mid \theta)\ p(\theta \mid y)\ d\theta$$

Bayesian Optimal Experimental Design (OED)

- Bayesian Learning: Formulate posterior PDF of model parameters and/or posterior PDF of QoI given the data
- The Kullback-Leibler Divergence (KL-div), a scalar measure of the "distance" between the posterior and prior, is used to represent the information gain due to data.
- The objective of the experiment is to maximize the KL-div with respect to the design variables of the experiment.

- **KL-div**

$$U_y(\delta, y, \varphi) = \int_{\Theta} p(\theta \mid y, \delta, \varphi) \ \ln\left[\frac{p(\theta \mid y, \delta, \varphi)}{p(\theta)}\right] d\theta$$

Bayesian Optimal Experimental Design (OED)

- Bayesian Learning: Formulate posterior PDF of model parameters and/or posterior PDF of QoI given the data
- The Kullback-Leibler Divergence (KL-div), a scalar measure of the "distance" between the posterior and prior, is used to represent the information gain due to data.
- The objective of the experiment is to maximize the KL-div with respect to the design variables of the experiment.
- Need the data in order to compute KL-div. Data are not available before the experiment.
- Average the KL-div over all possible data, based on the likelihood & prior PDF of the model.

$$y = q(\theta \mid \mathsf{M}) + e$$

- **Expected KL-div**

$$e \sim N(0, \Sigma)$$

$$U(\delta, \varphi) = \int_Y \int_\Theta p(\theta \mid y, \delta, \varphi) \ \ln\left[\frac{p(\theta \mid y, \delta, \varphi)}{p(\theta)}\right] d\theta \ p(y \mid \delta) \ dy$$

Bayesian Optimal Experimental Design (OED)

- Bayesian Learning: Formulate posterior PDF of model parameters and/or posterior PDF of QoI given the data
- The Kullback-Leibler Divergence (KL-div), a scalar measure of the "distance" between the posterior and prior, is used to represent the information gain due to data.
- The objective of the experiment is to maximize the KL-div with respect to the design variables of the experiment.
- Need the data in order to compute KL-div. Data are not available before the experiment.
- Average the KL-div over all possible data, based on the likelihood & prior PDF of the model.

- **Expected KL-div**

$$U(\delta,\varphi)=\int_{Y}\int_{\Theta} p(\theta \mid y,\delta,\varphi)\ \ln\left[\frac{p(\theta \mid y,\delta,\varphi)}{p(\theta)}\right] d\theta\ p(y \mid \delta)\ dy$$

- **Computation of Multi-Dimensional Integrals**
  - **Sampling algorithms (nested sampling)**

    Huan & Marzouk, J. of Computational Physics, 2013

  - **Asymptotic approximation**

Asymptotic Approximation

- **REFORMULATE utility as a function of the design variables:**

$$-U(\delta,\varphi) \sim \int_{\Theta} \bar{H}(\theta,\varphi,\delta)\ \pi(\theta)\ d\theta$$

Expected posterior entropy

$$\bar{H}(\theta,\varphi,\delta) = \int_{\Upsilon} H(y,\varphi,\delta)\ p(y\,|\,\theta,\varphi,\delta)\ dy$$

Posterior entropy averaged over data for specific parameters

$$H(y,\varphi,\delta) = -\int_{\Theta} p(\theta\,|\,y,\varphi,\delta)\ \ln p(\theta\,|\,y,\varphi,\delta)\ d\theta$$

Posterior entropy for specific data set

- **Asymptotic approximation**
  - **Large number of data ~ Normal posterior distribution**
  - **Small prediction error ~ Laplace type approximation of integral on Y**

- **Integration**
  - **Sparse grid**
  - **Monte Carlo (use same samples for different values of the design variables)**

Asymptotic Approximation

$$-U(\delta,\varphi) \approx \sum_{j=1}^{n} w_j \; H(q(\theta^{(j)},\varphi,\delta))$$

- **Information entropy**

Inverse of Prior Covariance

$$H(q(\theta,\varphi,\delta)) = \frac{1}{2} N_\theta \ln(2\pi) - \frac{1}{2} \ln \det[\hat{Q}(\theta,\varphi,\delta) + Q_\pi(\theta)]$$

Fisher Information Matrix

$$\hat{Q}(\theta,\varphi,\delta) = \nabla_\theta q^T(\theta;\varphi,\delta)\; \Sigma^{-1}(\sigma)\; \nabla_\theta^T q(\theta;\varphi,\delta)$$

- **Result is same as one in [Papadimitriou, Beck & Au 2000], based on the concept of robust information entropy for optimal experimental design**

Papadimitriou, Beck & Au, Journal of Vibration and Control, 2000
Papadimitriou & Papadimitriou, International Journal for Uncertainty Quantification, 2015
Papadimitriou, Lombaert, Mechanical Systems and Signal Processing, 28, 105-127, 2012
Argyris, Ph.D. Thesis, University of Thessaly, 2017

Asymptotic Approximation

$$-U(\delta,\varphi) \approx \sum_{j=1}^{n} w_j \; H(q(\theta^{(j)},\varphi,\delta))$$

Inverse of Prior Covariance

- **Information entropy**

$$H(q(\theta,\varphi,\delta)) = \frac{1}{2} N_\theta \ln(2\pi) - \frac{1}{2} \ln\det[\hat{Q}(\theta,\varphi,\delta) + Q_\pi(\theta)]$$

Fisher Information Matrix

$$\hat{Q}(\theta,\varphi,\delta) = \nabla_\theta q^T(\theta;\varphi,\delta)\; \Sigma^{-1}(\sigma)\; \nabla_\theta^T q(\theta;\varphi,\delta)$$

- **Result is same as one in [Papadimitriou, Beck & Au 2000], based on the concept of robust information entropy for optimal experimental design**

Papadimitriou, Beck & Au, Journal of Vibration and Control, 2000
Papadimitriou & Papadimitriou, International Journal for Uncertainty Quantification, 2015
Papadimitriou, Lombaert, Mechanical Systems and Signal Processing, 28, 105-127, 2012
Argyris, Ph.D. Thesis, University of Thessaly, 2017

- **Spatial correlation avoids sensor clustering**

$$\Sigma_{ij} = \sigma_1^2 \delta_{ij} + \sigma_2^2 q_i q_j R(\Delta_{ij}) \qquad R(\Delta_{ij}) = \exp\left[-\frac{|x_i - x_j|}{\lambda}\right]$$

Papadimitriou, Lombaert, Mechanical Systems and Signal Processing, 28, 105-127, 2012

Robust Bayesian OED

- Bayesian Learning: Formulate posterior PDF of model parameters and/or posterior PDF of QoI given the data
- The Kullback-Leibler Divergence (KL-div), a scalar measure of the "distance" between the posterior and prior, is used to represent the information gain due to data.
- The objective of the experiment is to maximize the KL-div with respect to the design variables of the experiment.
- Need the data in order to compute KL-div. Data are not available before the experiment.
- Average the KL-div over all possible data, based on the likelihood & prior PDF of the model.

- **Robust measure of Expected KL-div**:

$$U(\delta) = \int_{\Phi}\int_{Y}\int_{Z} p(\theta \mid y, \delta, \varphi) \ \ln\left[\frac{p(\theta \mid y, \delta, \varphi)}{p(\theta)}\right] d\theta \ p(y \mid \delta) \ dy \ \pi(\varphi) \ d\varphi$$

Asymptotic Approximation

$$-U(\delta) \approx \sum_{j=1}^{n} w_j \; H(q(\theta^{(j)}, \varphi^{(j)}, \delta))$$

- **Information entropy**

Inverse of Prior Covariance

$$H(q(\theta,\varphi,\delta)) = \frac{1}{2} N_\theta \ln(2\pi) - \frac{1}{2} \ln\det[\hat{Q}(\theta,\varphi,\delta) + Q_\pi(\theta)]$$

Fisher Information Matrix

$$\hat{Q}(\theta,\varphi,\delta) = \nabla_\theta q^T(\theta;\varphi,\delta)\; \Sigma^{-1}(\sigma)\; \nabla_\theta^T q(\theta;\varphi,\delta)$$

- **Optimization (Multiple local/global optima)**
  - **CMA-ES**
  - **Forward Sequential Sensor Placement (FSSP)**
  - **Backward Sequential Sensor Placement (BSSP)**

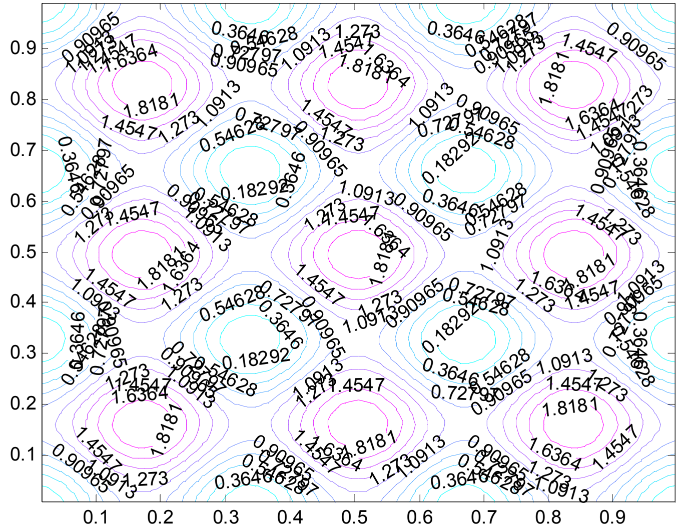
Papadimitriou 2005
Papadimitriou & Papadimitriou, International Journal for Uncertainty Quantification, 2015
Argyris, Ph.D. Thesis, University of Thessaly, 2017

Effect of Spatial Correlation

Optimal placement of two (2) sensors on simply supported beam (Length 1m)

Only the third mode contributes (Characteristic length ~0.3 m)

Uncorrelated – Length 0.0



Spatially uncorrelated prediction error results in sensor clustering

For spatially correlated prediction error model, the optimal location of a new sensor is sufficiently away from the locations of existing sensors. Spatial correlation avoids sensor clustering.

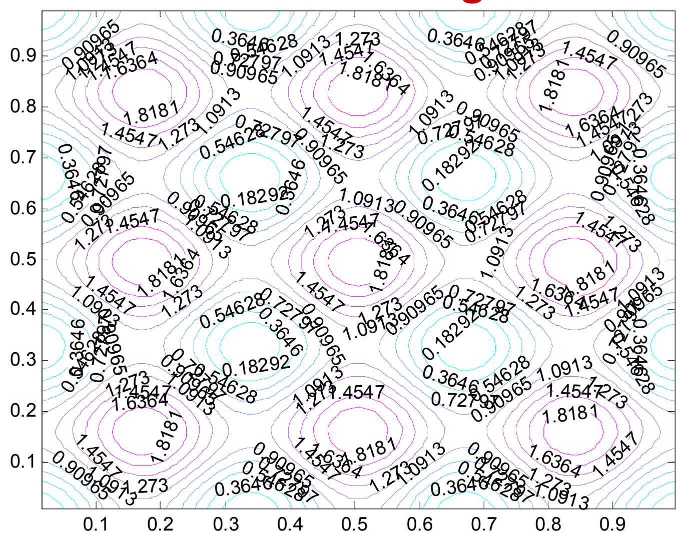
Papadimitriou, Lombaert, Mechanical Systems and Signal Processing, 28, 105-127, 2012

Effect of Spatial Correlation

Optimal placement of two (2) sensors on simply supported beam (Length 1m)

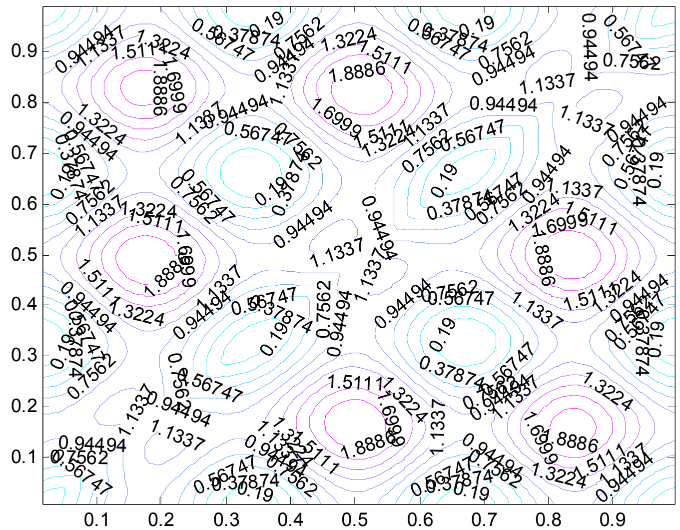
Only the third mode contributes (Characteristic length ~0.3 m)

Uncorrelated – Length 0.0



Spatially uncorrelated prediction error results in sensor clustering

Correlated – Length 0.1

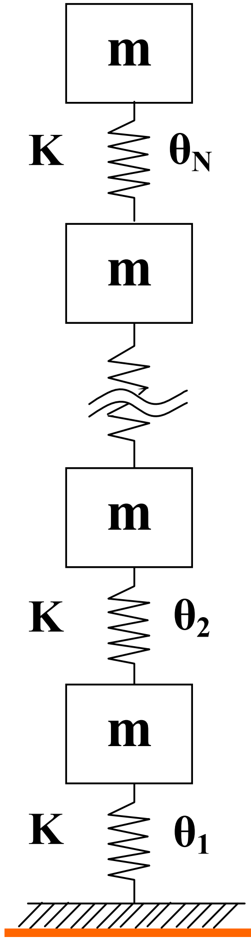


Spatially correlated prediction error avoids sensor clustering

For spatially correlated prediction error model, the optimal location of a new sensor is sufficiently away from the locations of existing sensors. Spatial correlation avoids sensor clustering.

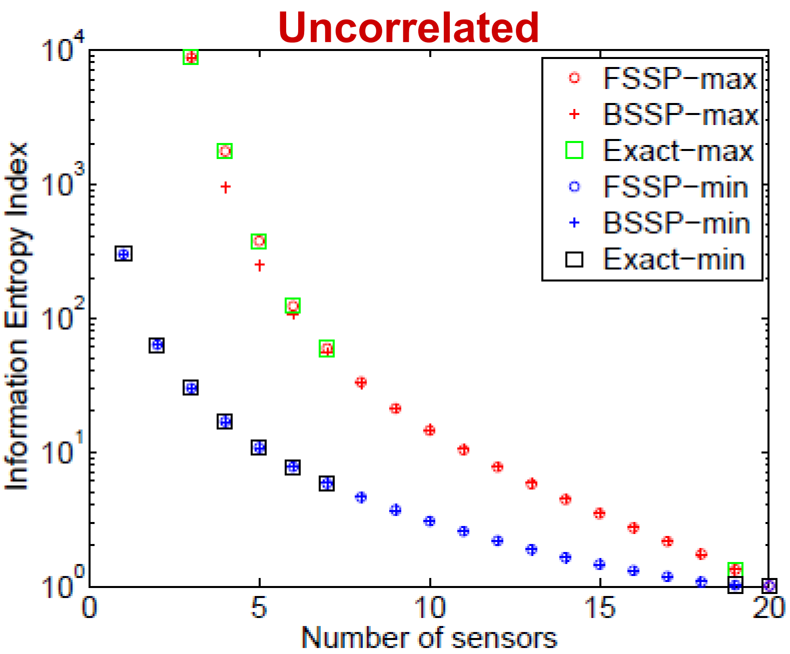
Papadimitriou, Lombaert, Mechanical Systems and Signal Processing, 28, 105-127, 2012

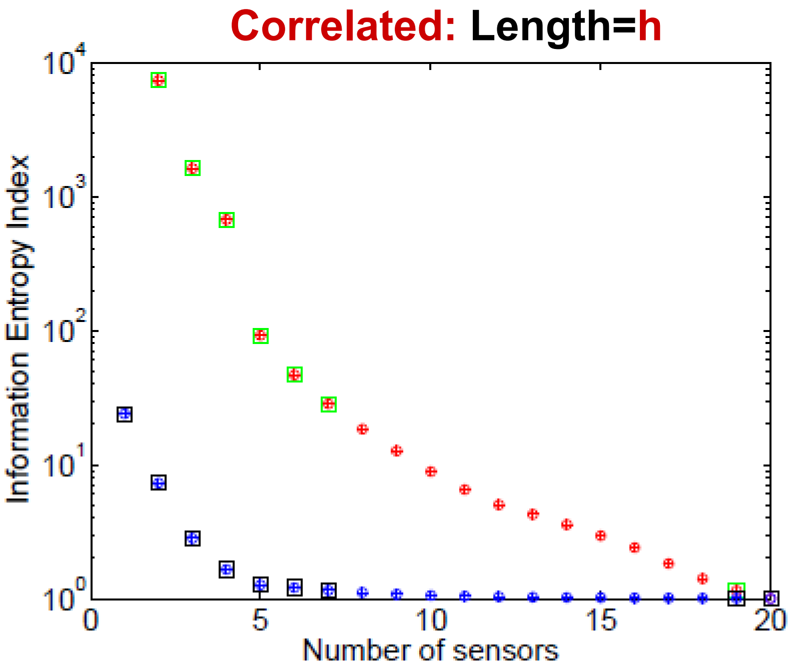
Effect of Spatial Correlation on Information Gain



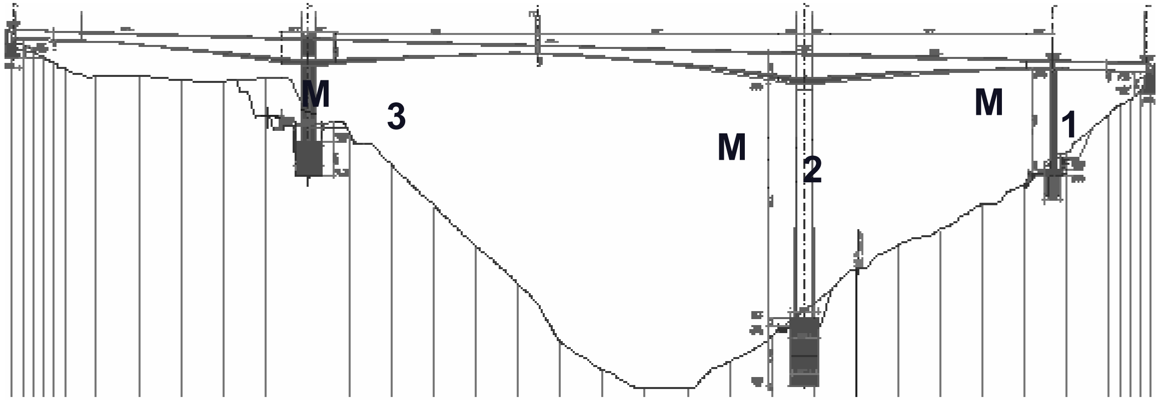
20 - DOF Spring-Mass Model

5-parameters
3 contributing modes: Characteristic length is 7h
Design for impulse excitation at top mass





Demonstration: Metsovo Bridge

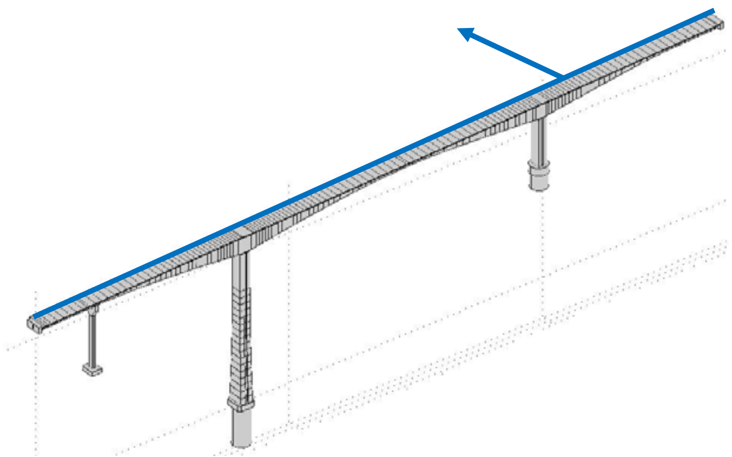


R/C BRIDGE

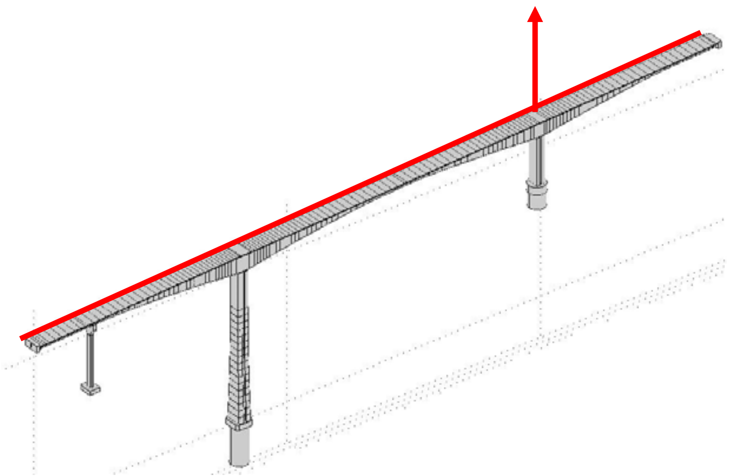
- **Total length: 537m**
- Deck width: 14m
- M1 height: 45m tall
- **M2 height:110m tall**
- M3 height: 35m tall
- **Central span length: 235m**

Monitoring Equipment

Sensor Locations



Optimization for the transverse modes.
d in [0 1]



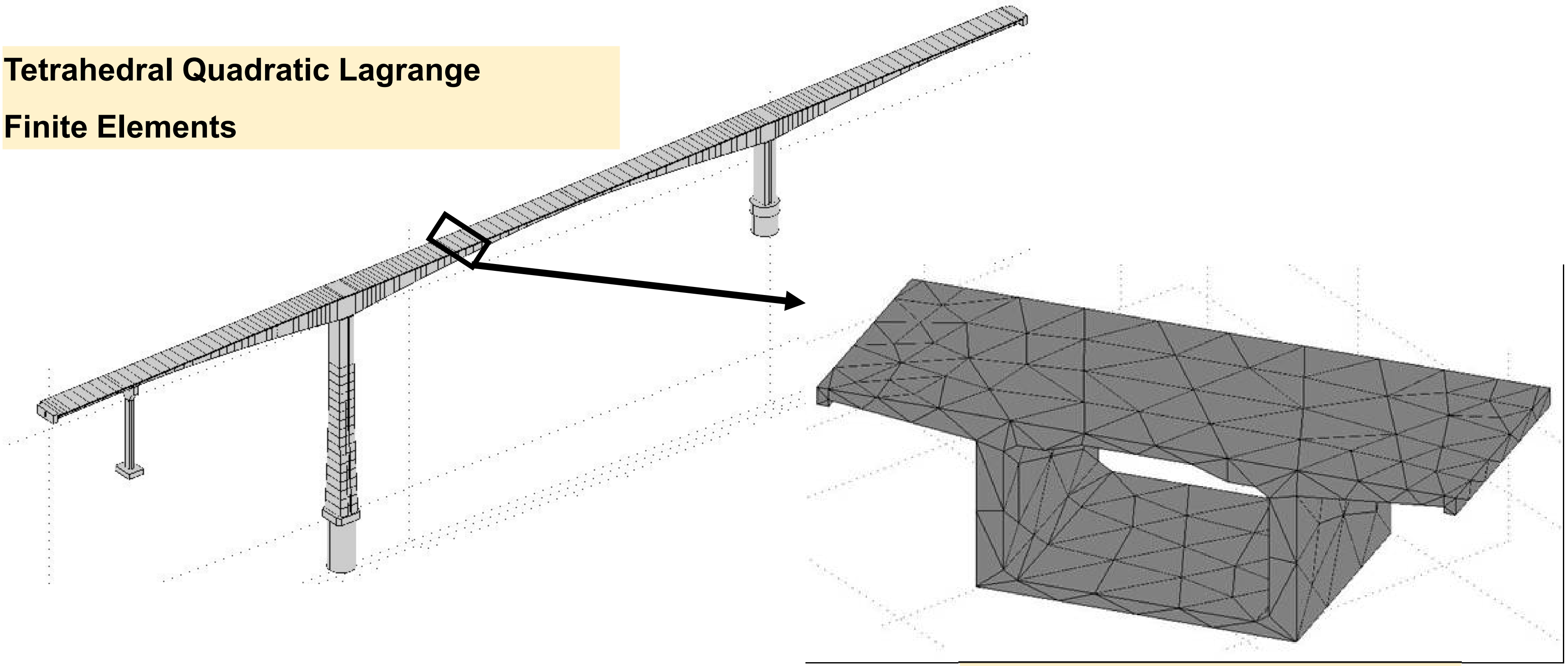
Optimization for the vertical modes.
d in [1 2]

- Optimization with continuous design variables
- Implementation: Map continuous design domain in the bridge on a parent domain
- d in [0 2] cover the entire length for both type of sensors

Finite Element Model

Tetrahedral Quadratic Lagrange

Finite Elements



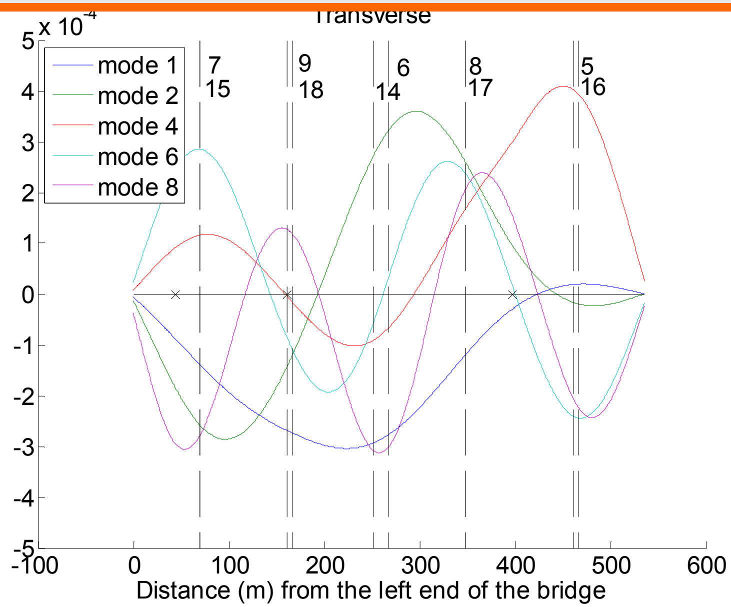
FINITE ELEMENT mesh:

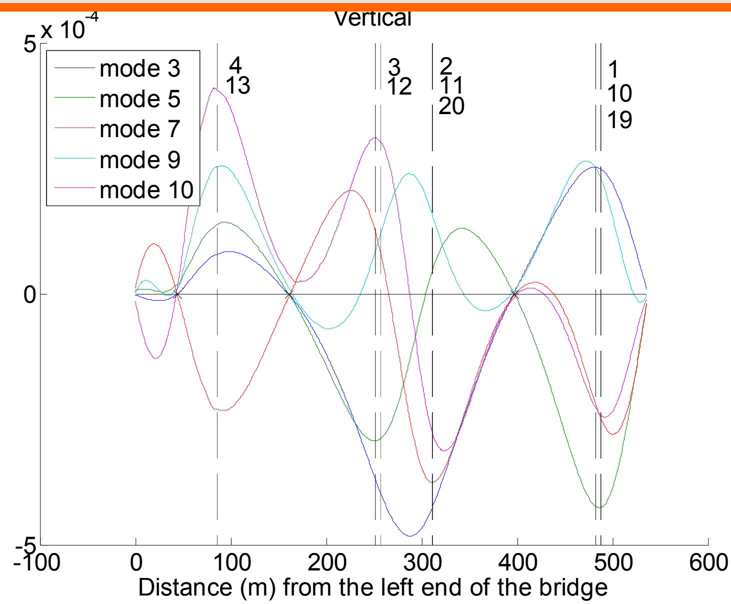
Largest FE size is limited by

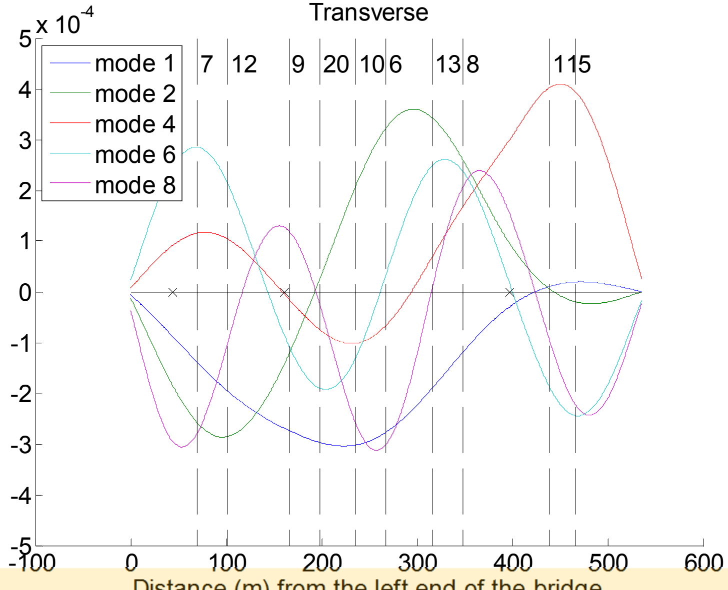
box cross-section thickness.

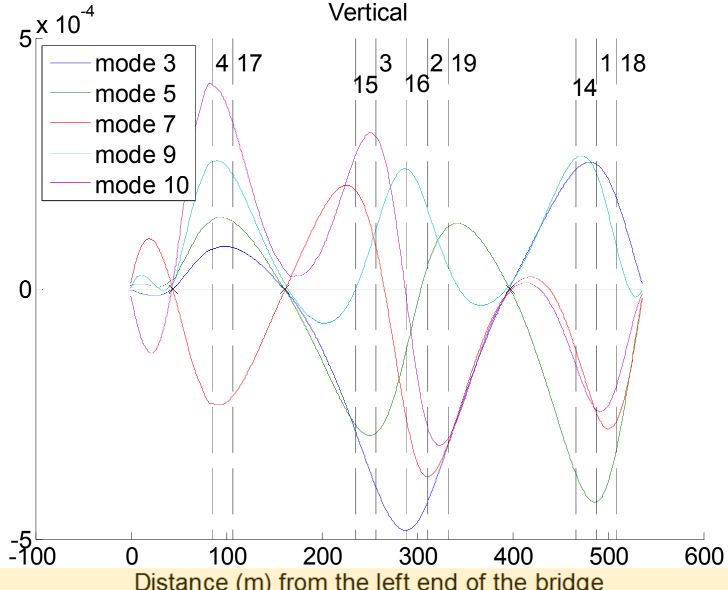
FE size < 0.4m

Metsovo Bridge: Optimal Sensor Placement





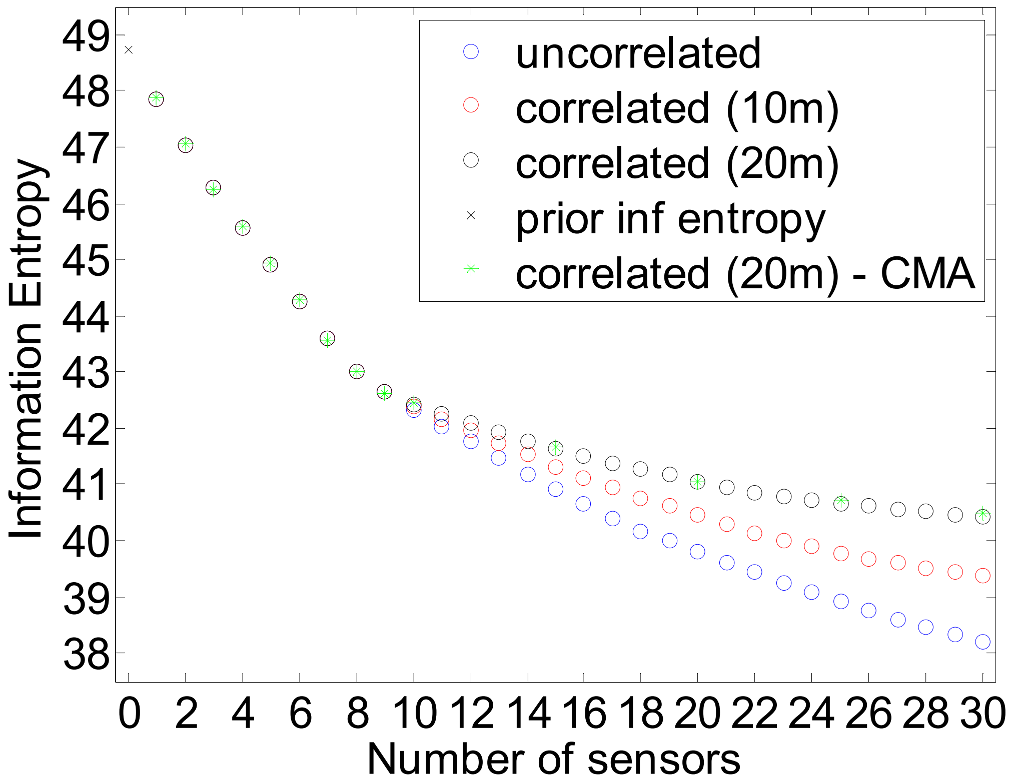




Up = uncorrelated error. Down = exponentially correlated error with λ=10m

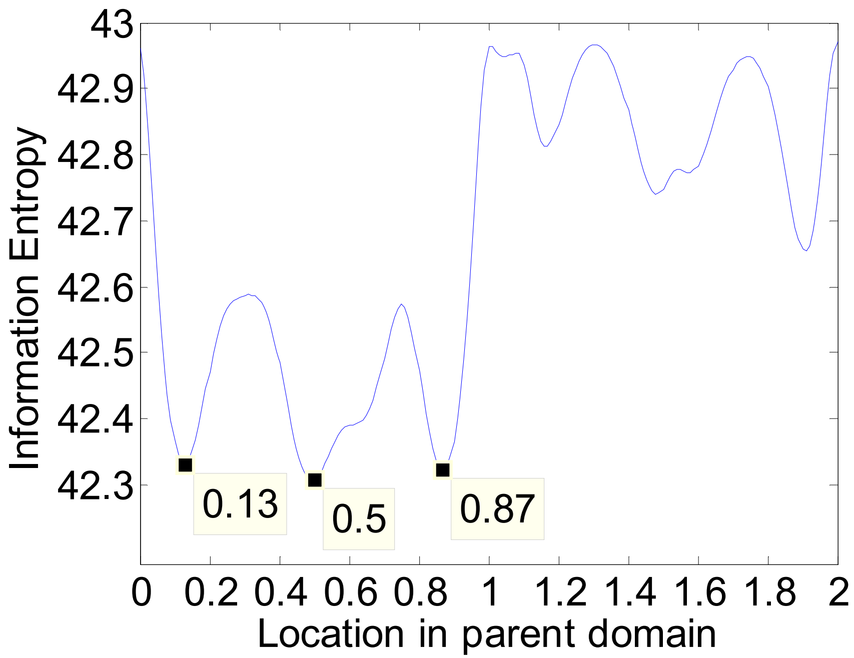
Argyris & Papadimitriou, J. of Smart Cities 2017

Optimal Information Entropy vs Number of Sensors

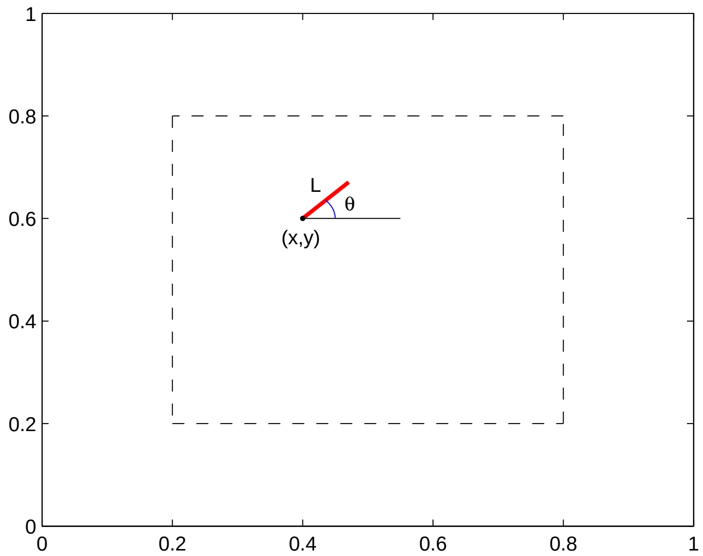


- **Uncorrelated prediction error: Information is gained as new sensors are added**
- **Correlated prediction error: Information gain reduces as number of sensors increases.**
- **Sequential sensor placement strategy is very accurate**

Optimization: Local/Global Optima

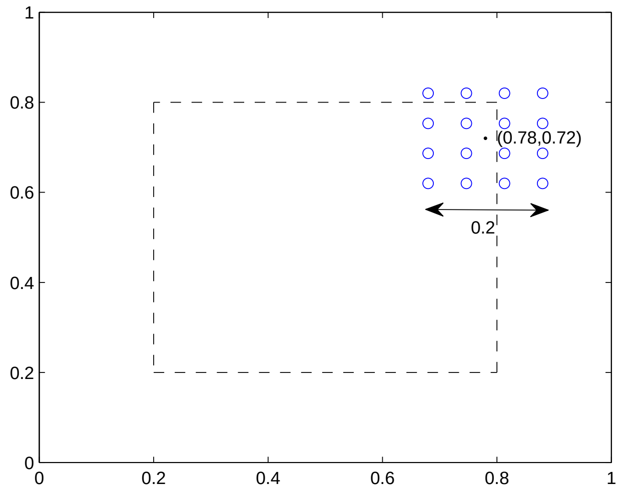


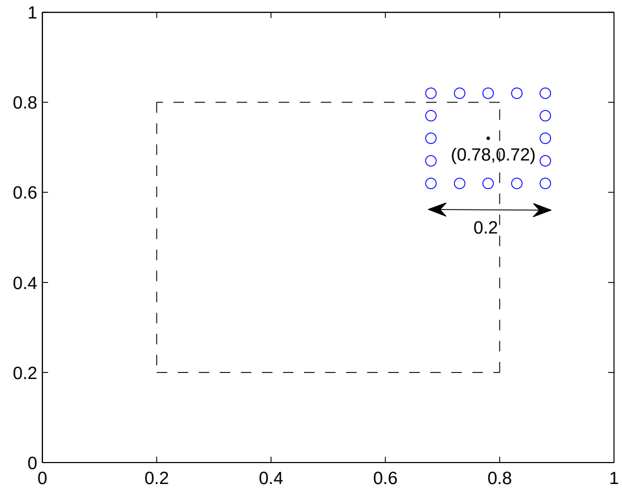
Optimal Sensor Placement for Crack Identification

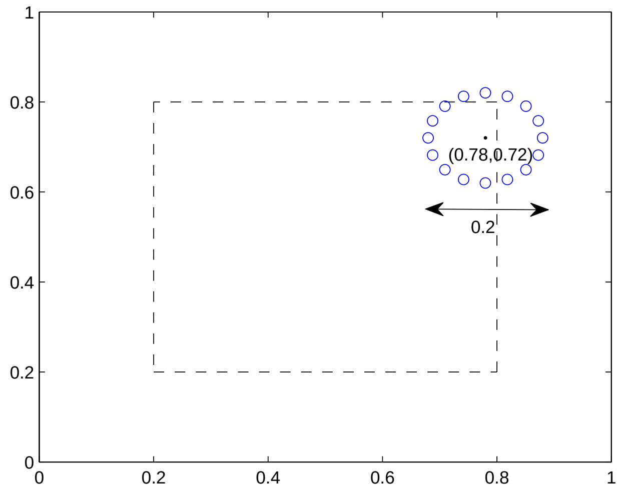


Argyris, Choudhury, Zabel & Papadimitriou, Structural Control and Health Monitoring, 2018

Sensor Configurations



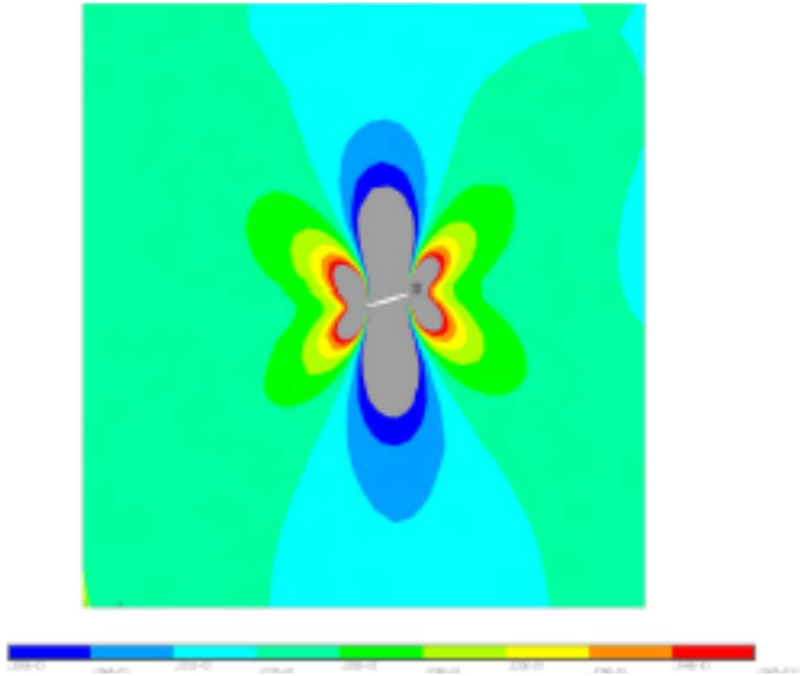




ANSYS FE Model of a Plate with Crack

Upward distributed load

fixed



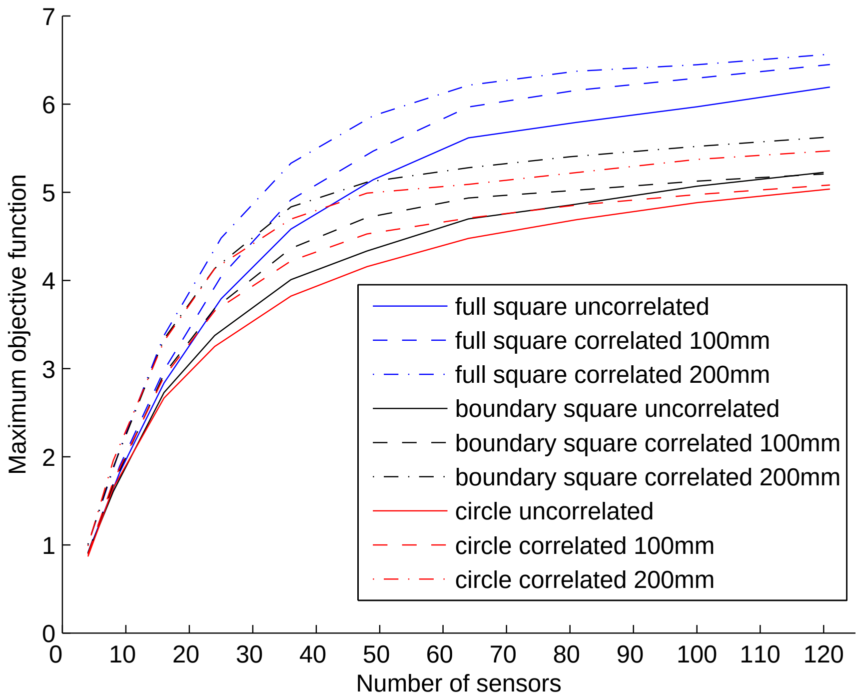
fixed

Variation of strain field (y-direction) around the crack

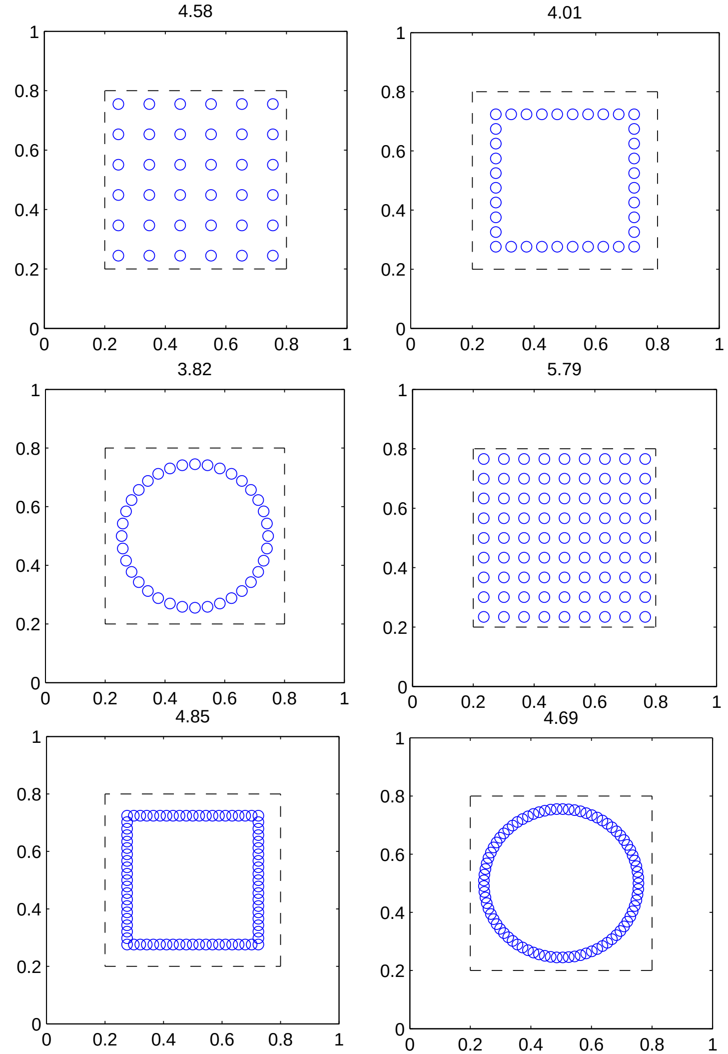
The model has about 6000 DOFs and is parametrized by the crack location, length and angle

Utility and Optimal Designs

Utility as a function of the number of sensors

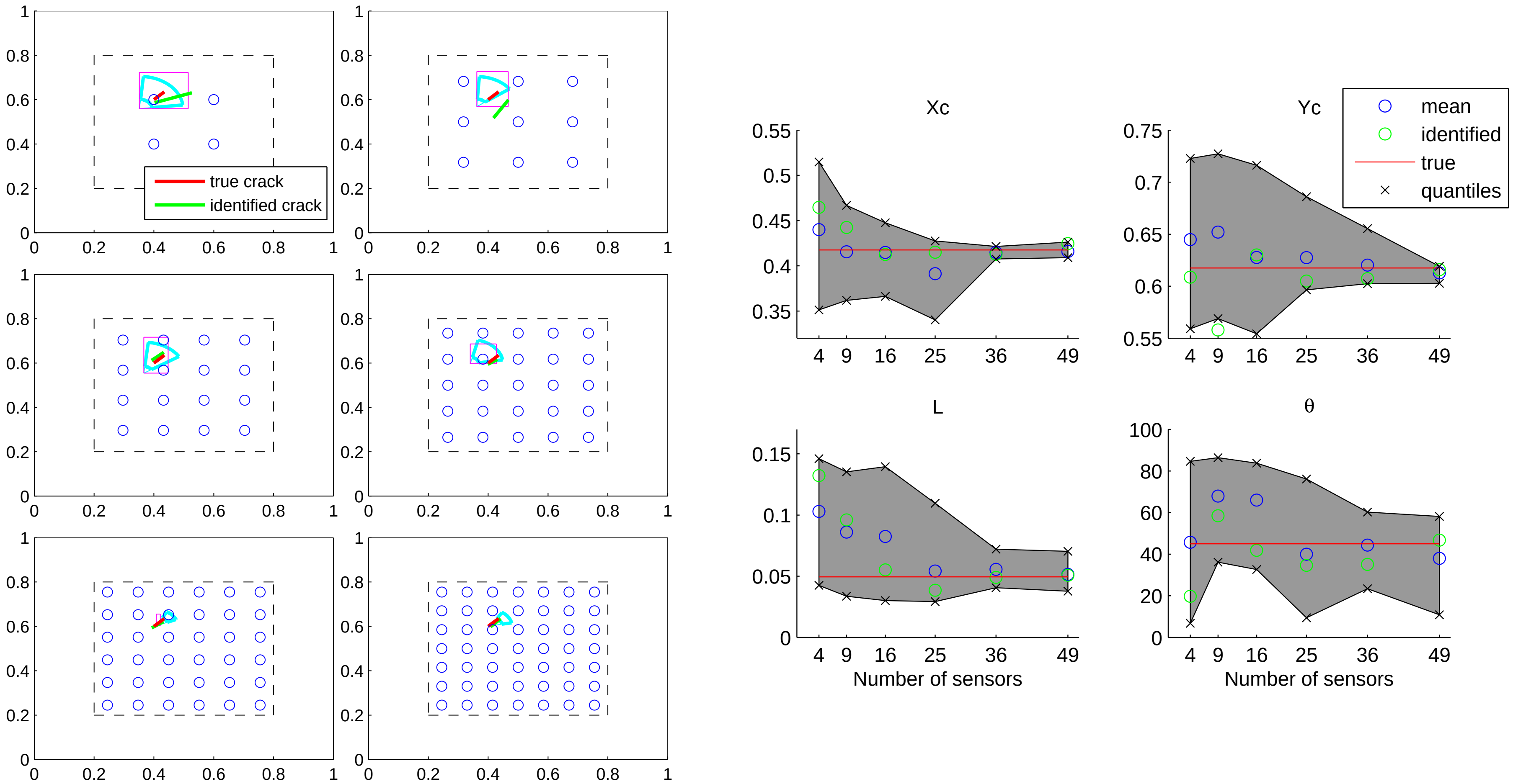


Selected optimal grids for 36 and 81 sensors



Effectiveness of Optimal Sensor Placement

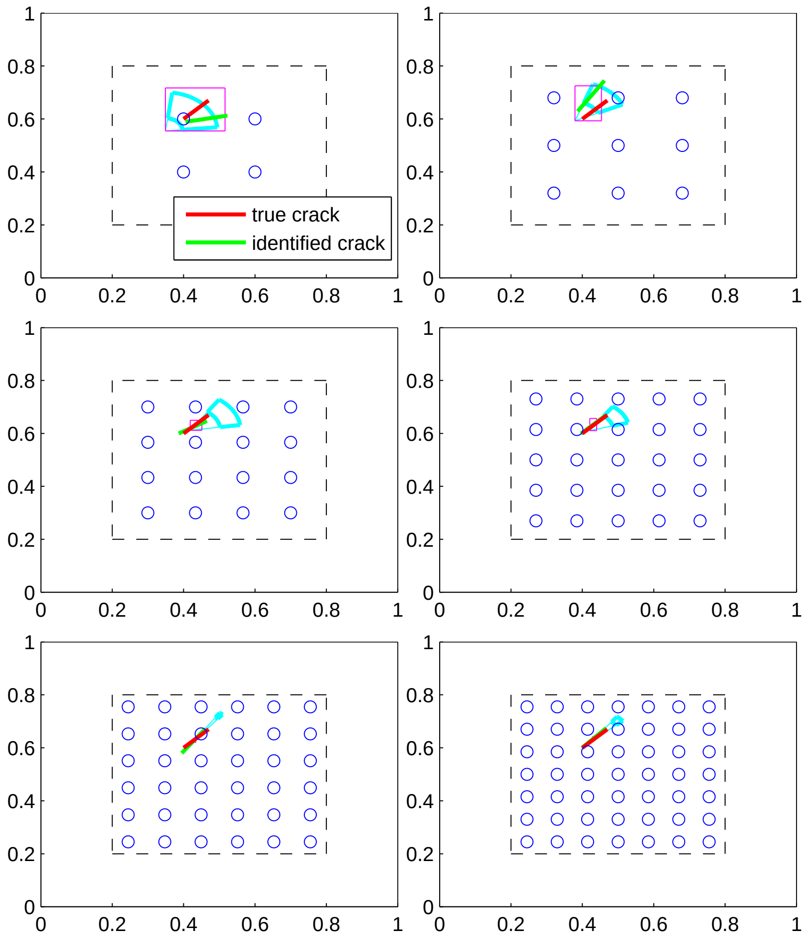
50mm crack full grid identification

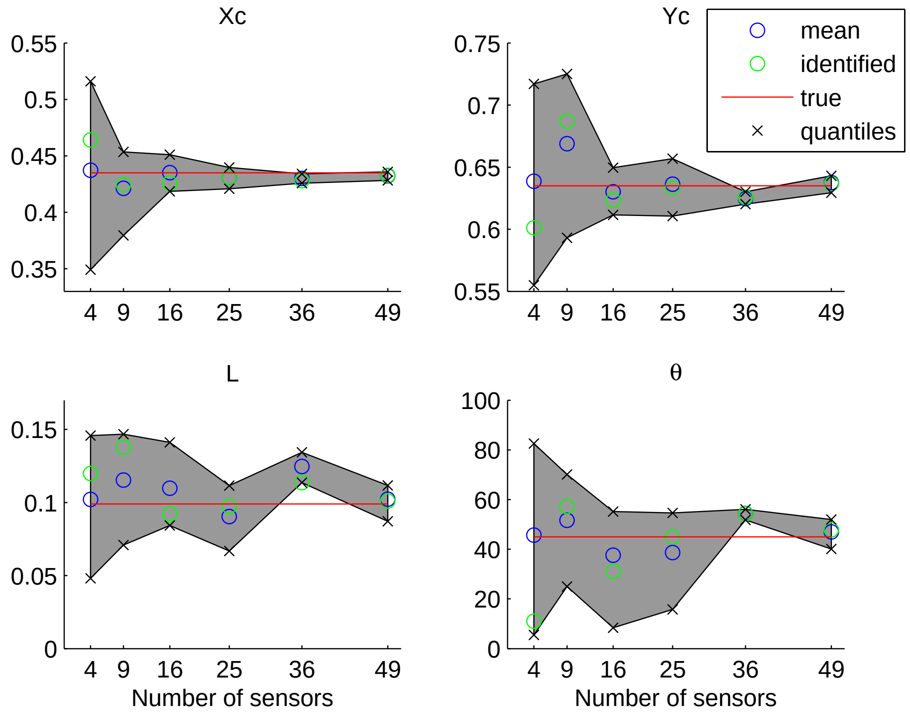


Argyris, Choudhury, Zabel & Papadimitriou, Structural Control and Health Monitoring, 2018

Effectiveness of Optimal Sensor Placement

100mm crack full grid identification





Bayesian OED for Response Predictions

OBJECTIVES:

- Response reconstruction (e.g. accelerations, displacements, strains & stresses)
- Fatigue estimation at hotspot locations using limited number of sensors

APPROACH: Integrate/Combine

1. Validated FE models

2a. Modal Expansion Tehcnique

2b. Bayesian filtering for input-state-parameter estimation

SIGNIFICANCE: DIAGNOSIS - PROGNOSIS

- Prediction of remaining fatigue lifetime;
- Decision making for cost-effective inspection, maintenance, repair

EXACT FORMULATION FOR LINEAR SYSTEMS

Posterior PDF: $z \mid D \sim N(\hat{z}, P_z(\delta, \varphi))$

Prior

$$-U(\delta) = \frac{1}{2} \int_{\Phi} \left[\ln \det P^z(\delta, \varphi) \right] \pi(\varphi)\, d\varphi + b$$

Posterior covariance

parameters

OSP for Predictions Using Modal Expansion

Modal Expansion

$$y = \Phi(\delta, \varphi)\ \xi + e, \qquad\qquad e \sim N(0, Q_e)$$

Prediction Equation (for response reconstruction)

$$z = \Psi(\varphi)\ \xi + \varepsilon, \qquad\qquad \varepsilon \sim N(0, Q_\varepsilon)$$

Given observations, use Bayesian inference to estimate modal coordinates and then propagate for making predictions.

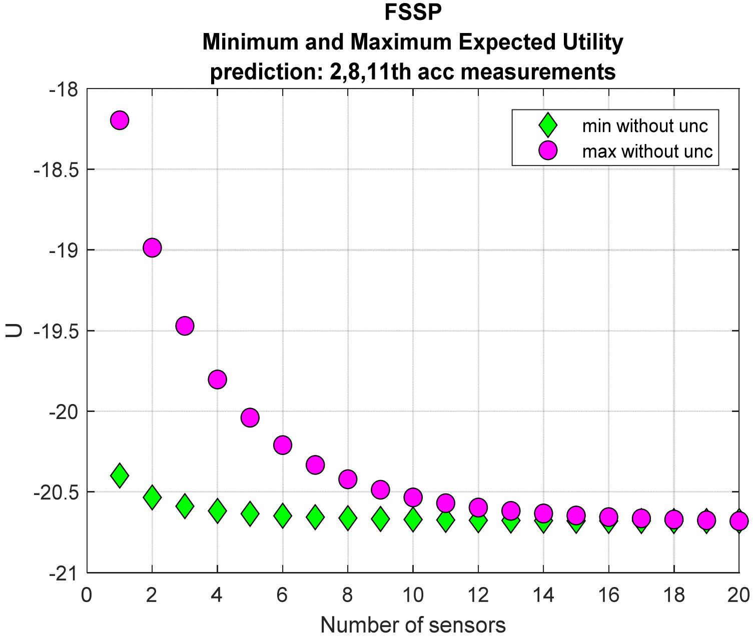
Predictions of unmeasured QoI follow a Normal Distribution.

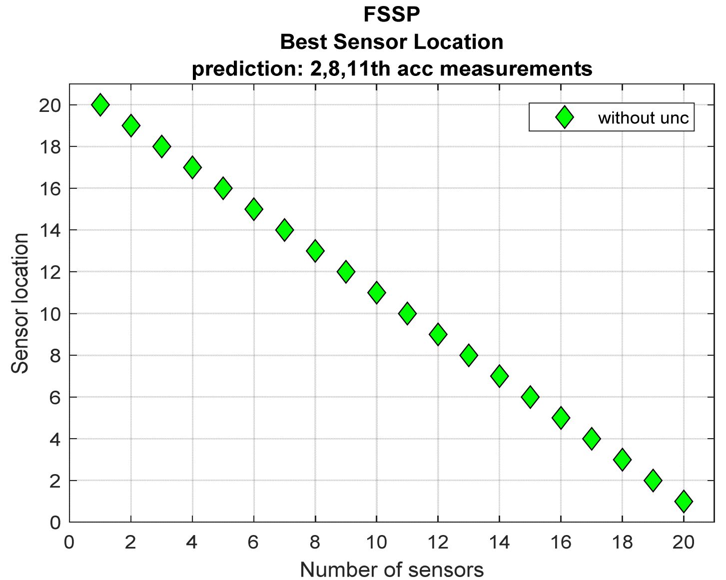
$$z \mid D \sim N(\hat{z}, P_z(\delta, \varphi))$$

Covariance is independent of the measurements

$$P_z(\delta, \varphi) = \Psi(\varphi) \left[\Phi^T(\delta, \varphi)\ Q_e^{-1}(\delta, \varphi)\ \Phi(\delta, \varphi) + \underbrace{S^{-1}}_{\text{Prior covariance}} \right]^{-1} \Psi^T(\varphi) + Q_\varepsilon(\varphi)$$

Example: 20-DOF Spring-Mass Chain Model





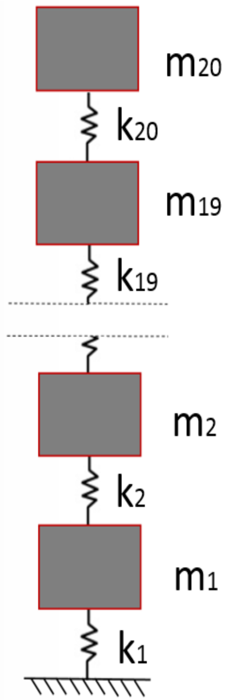
Method: Modal expansion

Modes: 1

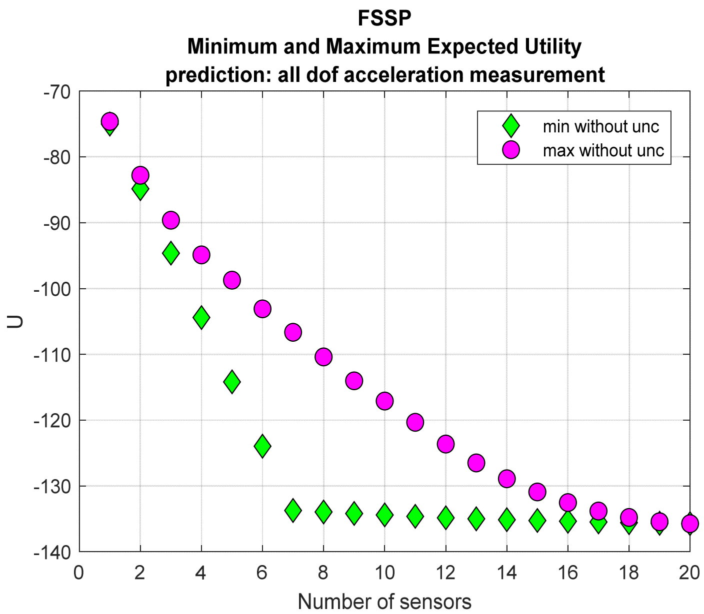
Prediction: Accelerations

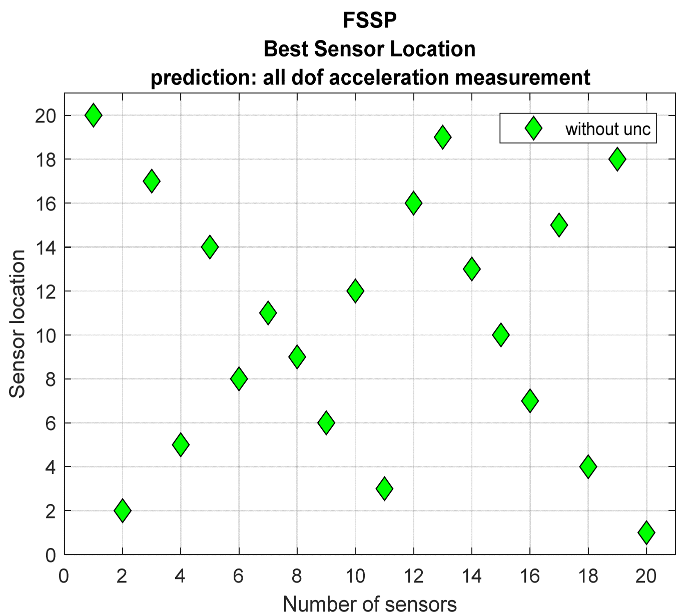
Prediction DOF: 2, 8, 11

$Q_e=Q_\varepsilon$: very small



Example: 20-DOF Spring-Mass Chain Model





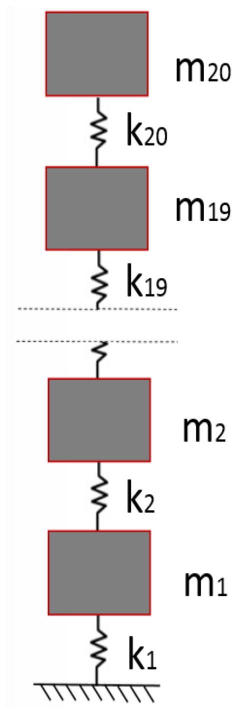
Method: Modal expansion

Modes: 7

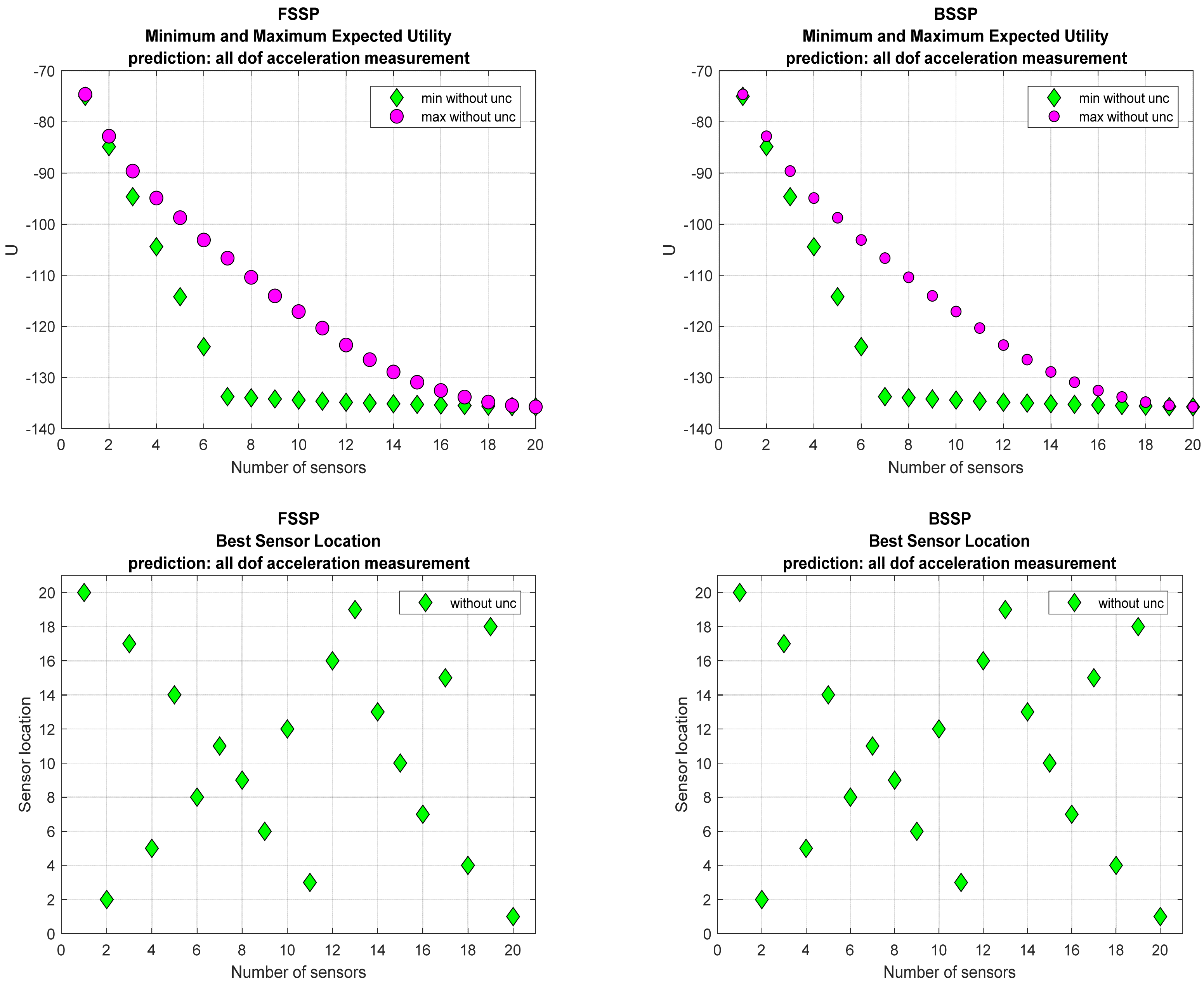
Prediction: Accelerations

Prediction DOF: 1:20

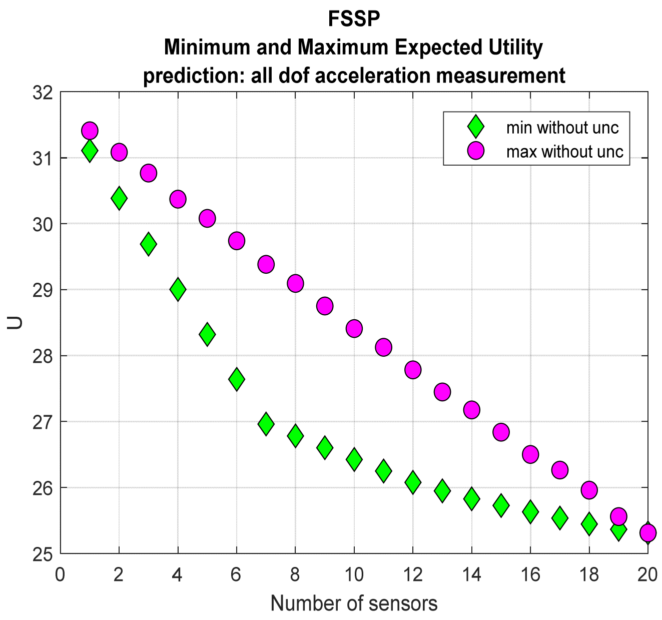
Q_e=Q_ε: very small

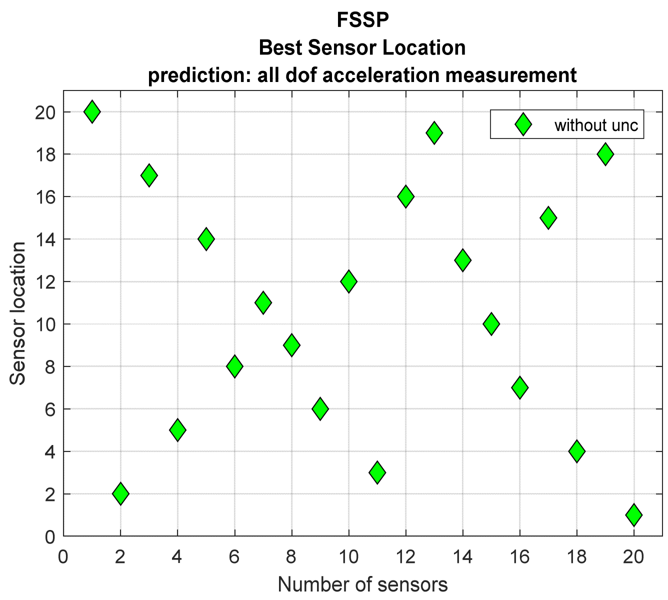


Example: 20-DOF Spring-Mass Chain Model



Example: 20-DOF Spring-Mass Chain Model





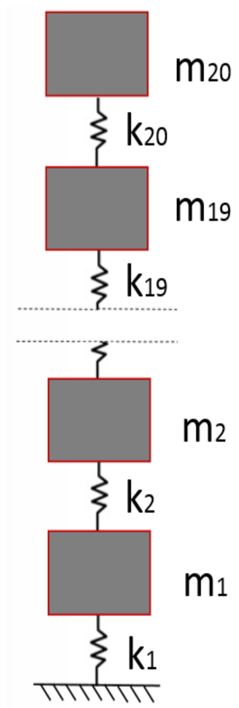
Method: **Modal expansion**

Modes: **7**

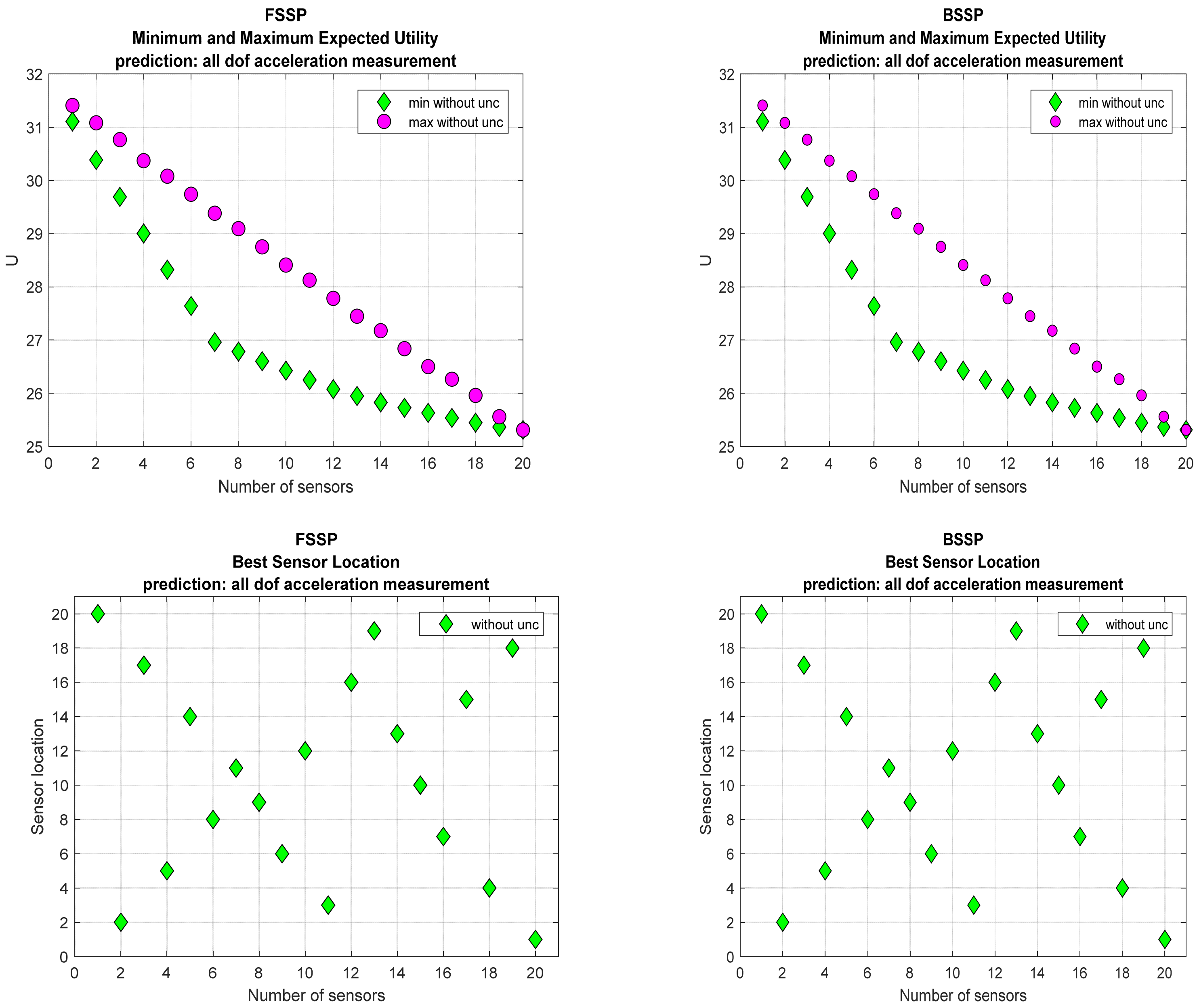
Prediction: **Accelerations**

Prediction DOF: **1:20**

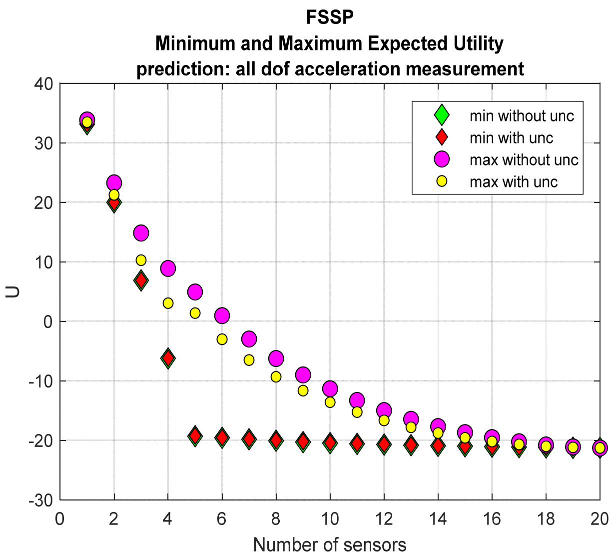
Q_e=Q_ε: **large**

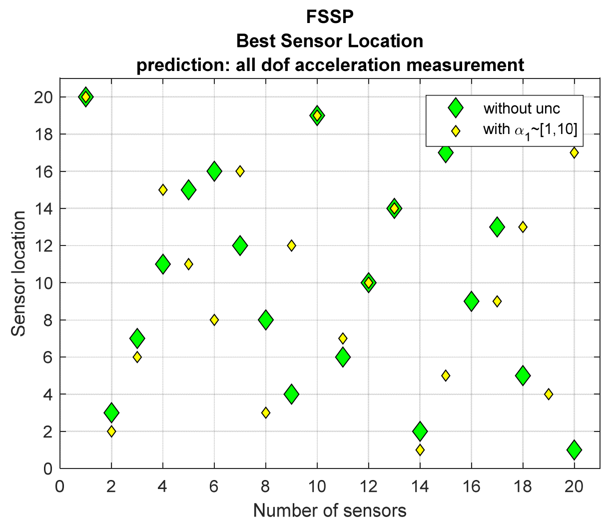


Example: 20-DOF Spring-Mass Chain Model



Example: 20-DOF Spring-Mass Chain Model





Robust Design

Uncertain degree of fixity

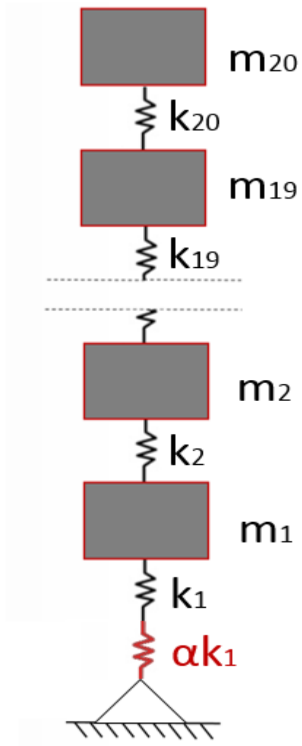
Method: Modal expansion

Modes: 5

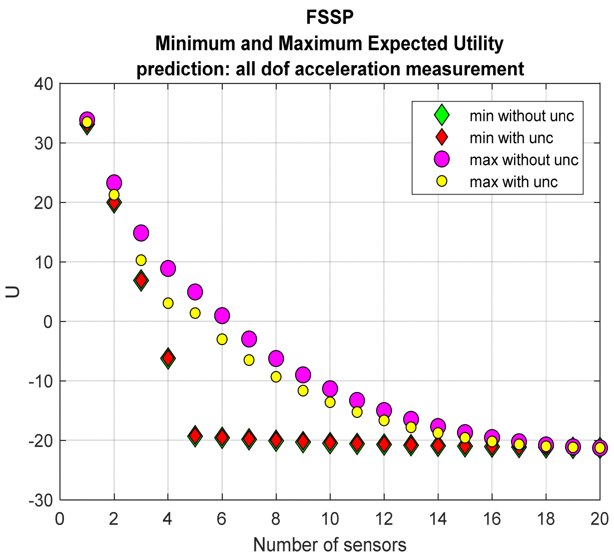
Prediction: Accelerations

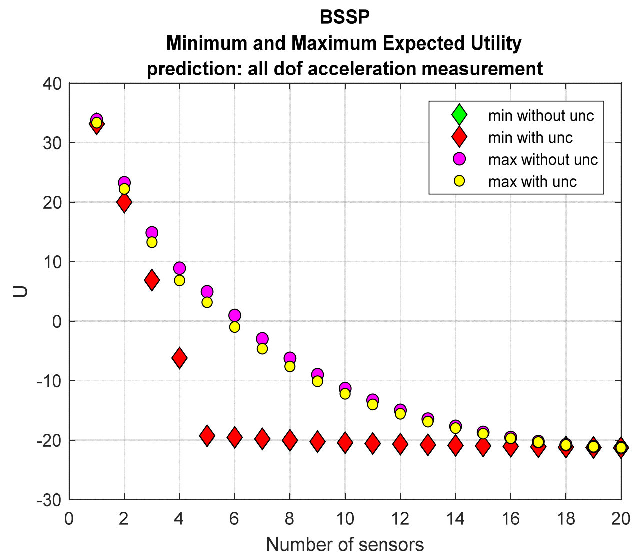
Prediction DOF: 1:20

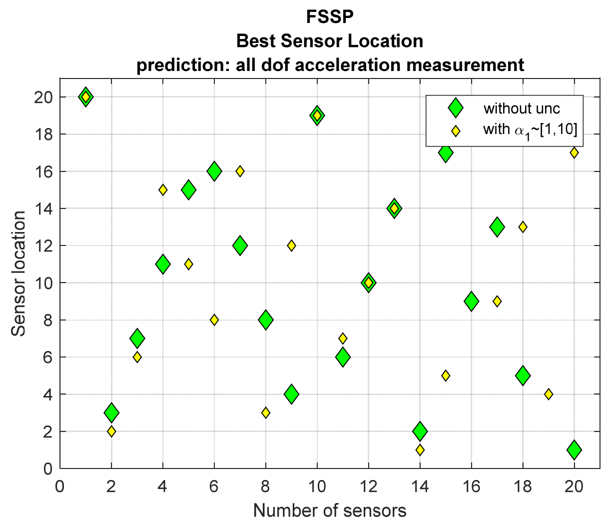
Q_e=Q_ε: small

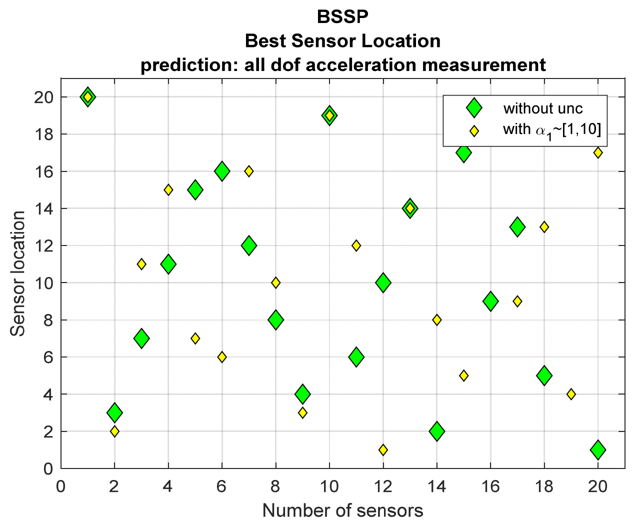


Example: 20-DOF Spring-Mass Chain Model









OSP for Joint Input-State Estimation (Steady-State)

Bayesian Filtering Techniques

- Lourens, Reynders, De Roeck, Degrande, Lombaert, MSSP, 2012

- Lourens, Papadimitriou, Gillijns, Reynders, De Roeck, Lombaert, MSSP, 2012

- Eftekhar Azam, Papadimitriou, Chatzi, MSSP, 2015

- Naets, Cuadrado, Desmet, 2015

- Maes, Smyth, De Roeck, Lombeert, MSSP, 2016, 2019

- Eftekhar Azam, Chatzi, Papadimitriou, Smyth, J. Vibration & Control, 2017

- Sedehi, Papadimitriou, Teymouri, Katafygiotis, MSSP, 2019

Given observations, use filtering techniques to estimate the input and the state and then propagate for making predictions

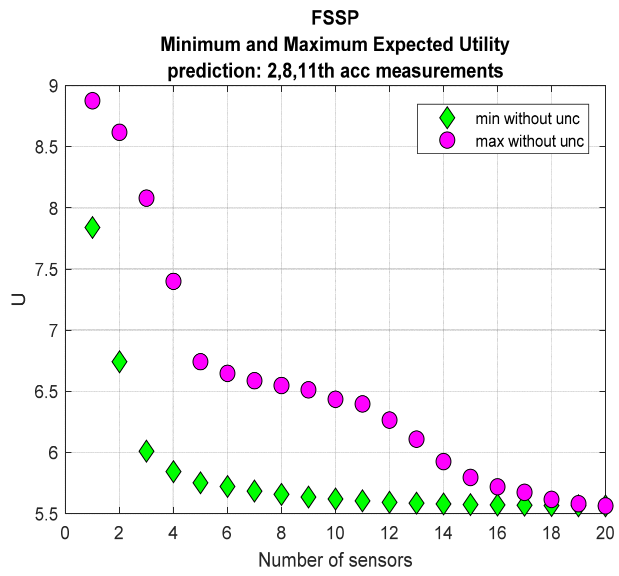
Predictions of QoI follow a Normal Distribution

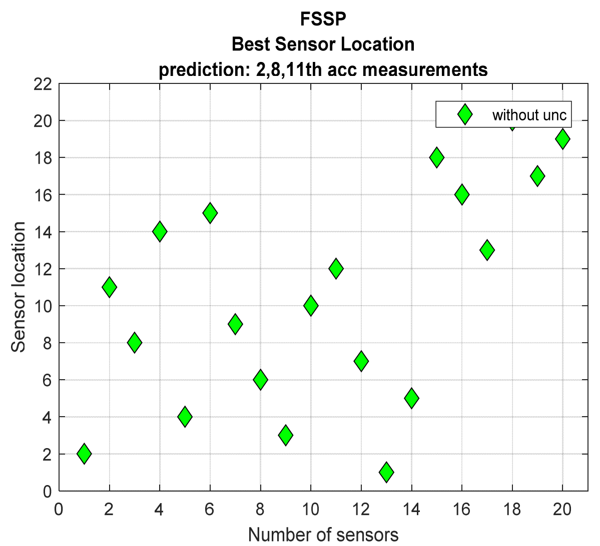
Covariance is independent of the measurements

$$\underline{z}_k \mid D \sim N(\underline{\hat{z}}_k, P^z(\underline{\delta}, \underline{\varphi}))$$

Posterior PDF

Example: 20-DOF Spring-Mass Chain Model





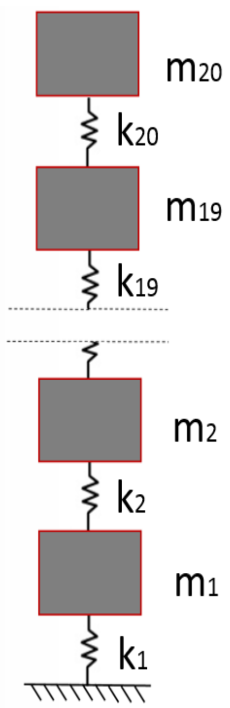
Method: Seq Bayesian Filter

Modes: 5

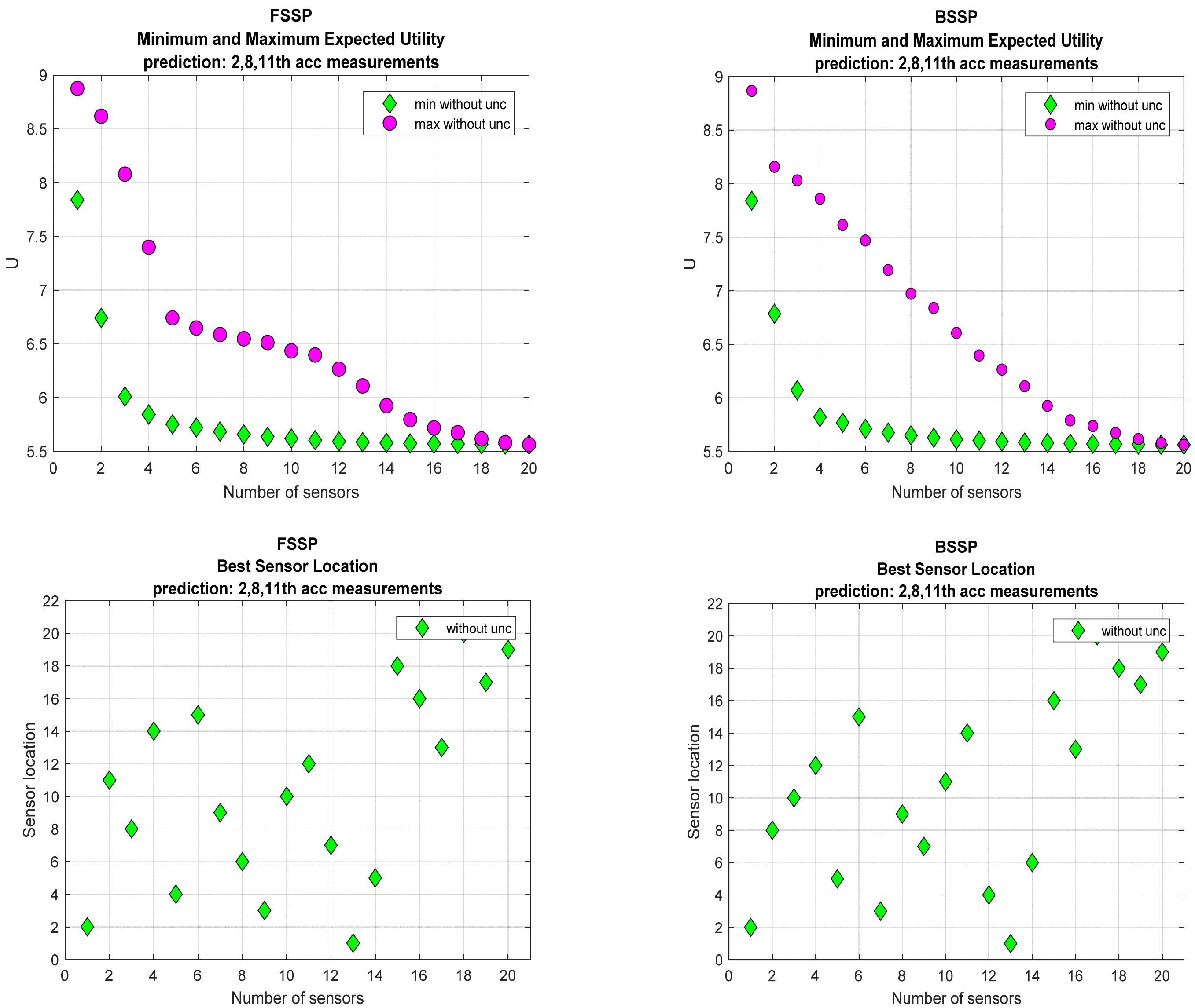
Prediction: Accelerations

Prediction DOF: 2, 8, 11

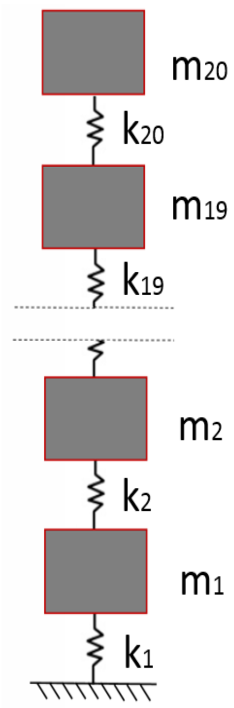
Q_e=Q_ε: small

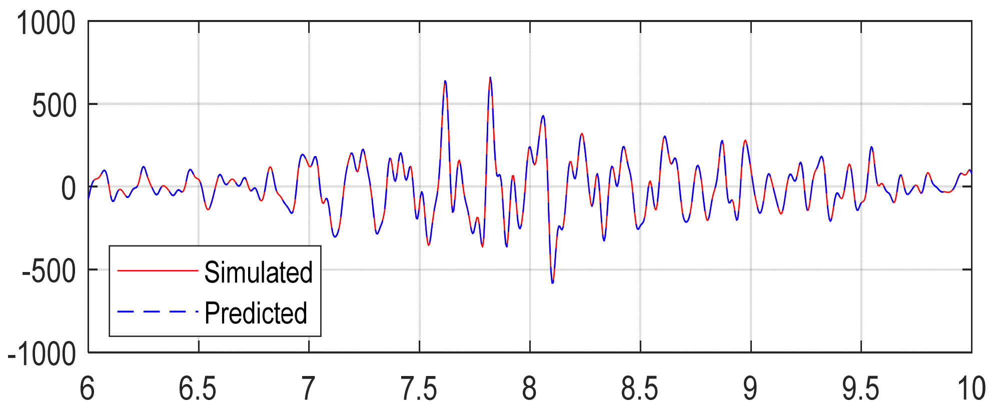


Example: 20-DOF Spring-Mass Chain Model



Validation of Optimal Sensor Placement

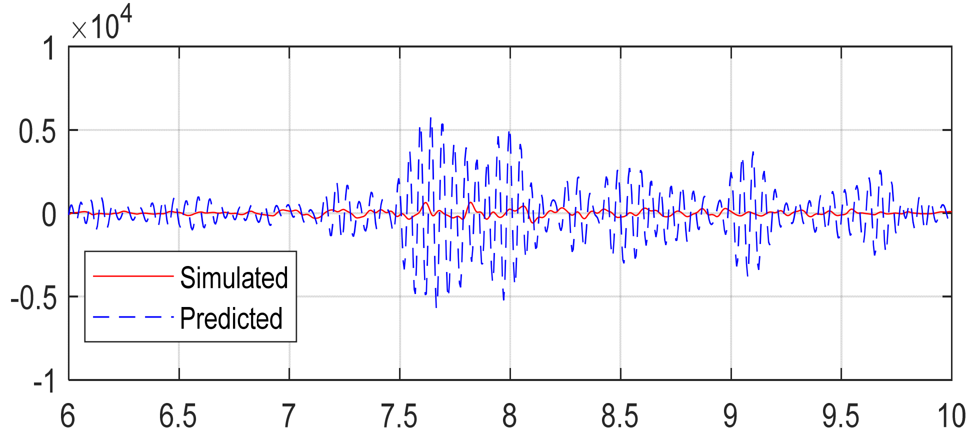




OPTIMAL

Locations of

7 Sensors



WORST

Locations of

7 Sensors

Concluding Remarks

- **OED for parameter estimation can handle nonlinear structural dynamics models. Also can optimize the excitation characteristics (Metallidis, Verros, Natsiavas & Papadimitriou 2003)**

- **OED for response predictions can be used to design effective sensor arrays for reconstructing/monitoring stresses and fatigue damage accumulation at hot-spot locations using output only vibration measurements.**

- **Sequential sensor placement technique is effective for solving the optimization problem and accounting for multiple local/global optima**

- **Optimization problem is best formulated in the continuous space**

- **Bayesian OED formulation is applicable to any system in engineering and applied sciences**

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764547.

Tulay Ercan, PhD student, University of Thessaly