**Principles of Data and Error Analysis in Engineering Measurements   
(Topics in Signal Processing)**

*Instructor*: Miodrag Bolic, [www.site.uottawa.ca/~mbolic](http://www.site.uottawa.ca/~mbolic)

*Time and place:* Monday 10:00 - 13:00, MNO E218

*Course code:* ELG 7172B (EACJ 5600)

Summary

Our goal in this course is to quantify uncertainty in our measurements, data analysis and machine learning algorithms. Uncertainty quantification will be done mainly through Bayesian approach and will rely on computational statistics (Monte Carlo). Uncertainty means: getting systems to estimate how much they do not know.

The course will be focused more on practical than theoretical aspects of uncertainty quantification and therefore we will study new probabilistic programming methods (Python PyMC3) and tools for modeling uncertainties in complex systems (Python OpenTurns).

Topics that will be covered are related to:

* Design of experiments
* Prediction
* Decision making and risk analysis
* Estimation
* Data fusion
* Quality assurance

*Key words:* Bayesian analysis, Uncertainty quantification, Probabilistic programming, Data analysis, Modeling, Monte Carlo analysis, Bayesian machine learning, Measurement, Errors, Time series analysis

## Different perspectives of this course

*Probabilistic Machine learning:* Over the last several years, deep neural networks advanced many applications including vision, language understanding, speech understanding, robotics and so on. But a major challenge still remain and that is: how to modeling uncertainty. Good models of uncertainty are crucial whenever decision needs to be made or an algorithm needs to decide how and when to acquire new information.

*Uncertainty quantification* is related to combining computational models, physical observations, and possibly expert judgment to make inferences about a physical system. Types of uncertainties include:

* Experimental uncertainty (measurement errors)
* Model uncertainty/discrepancy.
* Input/parameter uncertainty.
* Prediction uncertainty.

Why uncertainty:

* Uncertainty quantification is a fundamental component of model validation
* The objective is to replace the subjective notion of confidence with a mathematical rigorous measure
* Uncertainties relate to the physics of the problem of interest and not to the errors in the mathematical description/solution.

Comparison between different approaches that will be covered in this course is presented in the table below. The approaches include uncertainty in measurements, uncertainty quantification in engineering and science and probabilistic machine learning.

Comparison of different approaches that will be partially covered in this course:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Uncertainty in measurements (traditional approach GUM) [3]** | **Uncertainty quantification in engineering and science [1]** | **Probabilistic machine learning [7, 15]** |
| Model | Sensor model usually given as a function or regression model | Physical model given using set of equations or differential equations | Mixture models / k-means,  Hidden Markov models,  State space models, … |
| Quantifying sources of uncertainties, Input data representation | Commonly assume that data and parameters follow normal distribution | Fitting data to the model, MLE, Kernel density estimates | Fitting data to the model, MLE, Kernel density estimates |
| Learning, Parameter optimization, Model calibration | Commonly linear regression | MAP estimates, Regression,  Inverse problems, Bayesian approaches |  |
| Computational tools | Frequentist statistical methods | Monte Carlo sampling, MCMC, polynomial chaos | MCMC, Hamiltonian Monte Carlo |
| Uncertainty propagation in models | Approximation using first order Taylor expansion, Monte Carlo methods, Bootstrap | Sampling methods, Perturbation methods, Providing confidence intervals | Analyzing posterior distribution |
| Sensitivity analysis and model reduction | - | Global sensitivity analysis -Sobol | Through Bayesian modeling |
| Prediction | - | Bayesian or frequentist methods |  |
| Programming | Statistical computations of confidence intervals: | Uncertainty quantification tools, e.g. OpenTurns | Probabilistic programming, e.g. Python PyMC3 |

## Syllabus:

### Topics

Part I **Statistical concepts required to understand uncertainty**

* Intro: Definitions and motivation
  + Uncertainty, data quality, data analysis
  + Calibration, Precision, Accuracy, traceability, reproducibility, error
  + Measurement model
* Monte Carlo methods
  + Random variable generation, Importance sampling, Metropolis-Hastings Algorithm, MCMC
* Bootstrap principles
  + Principles or resampling, pivoting, bootstrap for time series
* Statistical intervals
  + Confidence intervals for a Normal distribution, Bootstrap based statistical intervals
* Bayesian analysis and inference

Part II **Uncertainty when no data and/or with historical data is available**

* Uncertainty in Metrology, GUM
  + Terminology, uncertainty quantification based on GUM
* Uncertainty propagation
  + Example of uncertainty propagation for temperature, pressure and other sensors
* Sensitivity analysis
  + Global sensitivity analysis, variance based method, Monte Carlo approaches, application to exploring sensitivity to parameters in the models in biomedical instrumentation
* Regression analysis
  + Linear and non-linear fitting, Confidence intervals of the estimates
* Model calibration and parameter estimation
  + Adjusting model parameters in order to improve the agreement between the model output and collected data, Regression analysis for calibration

**Part III Uncertainty for real-time analysis**

* Bayesian inference
  + Bayesian theorem, importance of prior, implementation using Markov Chain Monte Carlo, Prediction and Credible intervals
  + Model checking
  + Hierarchical Bayesian models
* Time series and HMM
* Particle filtering
  + State-space model, Bayesian filtering and Monte Carlo simulations, From complex probabilistic formulas to implementation
* Data and Sensor fusion
* Bayesian neural networks

Programming Topics

* PyMC3 – probabilistic programming
* OpenTurns – modeling complex systems
* …

### Applications topics

* Uncertainty in Metrology, GUM
* Physical models, models of sensors and electrical circuits
* Time series – biomedical data, bearings only tracking
* Sensor fusion

## Expected Learning Outcomes

After successful completion of this course, you will be able to:

* Understand basic concepts behind probabilistic machine learning and uncertainty quantification
* Write programs in Python to do quick analysis
* Apply your knowledge in metrology, data analytics, analysis of models in engineering and science
* Analyse problems using Bayesian approaches
* Perform sampling and MCMC to solve a variety of problems such as integrals and inference

## References:

### Uncertainty quantification

Books

1. R. C. Smith, Uncertainty Classification: Theory, Implementation and Applications, SIAM, 2014.
2. W. Q. Meeker, G. J. Hahn, L. A. Escobar, Statistical Intervals: A Guide for Practitioners and Researchers, 2nd Edition, Wiley, 2017.

Courses:

### Metrology

Books

1. F. Pavese, A. B. Forbes, Data Modeling for Metrology and Testing in Measurement Science, Springer, 2009.
2. S. V. Gupta, Measurement Uncertainties, Physical Parameters and Calibration of Instruments, Springer-Verlag Berlin Heidelberg 2012.
3. I. Lira, Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance, IoP 2002.
4. H. W. Coleman, W. G. Steele, Experimentation, validation and uncertainty analysis for Engineers, Wiley, 2009.

### Bayesian analysis

Books

1. R. McElreath, Statistical Re-thinking: A Bayesian Course with Examples in R and Stan, Chapman and Hall/CRC, 2015, http://xcelab.net/rm/statistical-rethinking/
2. A. Gelman, et al, Bayesian Data Analysis, 3rd edition, Chapman & Hall/CRC Texts in Statistical Science, 2013.
3. C. Davidson-Pilon, Bayesian Methods for Hackers, Addison-Wesley Data & Analytics, 2015, https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers
4. C. Bailer-Jones, Practical Bayesian Inference: A primer for Physical Scientists, Cambridge University Press, 2017.
5. J. Winn, C. Bishop, Model Based Machine Learning, http://www.mbmlbook.com/toc.html

Computational statistics

Books

1. W.M. Bolstad, Understanding Computational Bayesian Statistics, Wiley, 2010
2. Allen B. Downey, Think Stats 2e, online at <http://greenteapress.com/wp/think-stats-2e/>

Probability and statistics

1. José Unpingco, Python for Probability, Statistics, and Machine Learning, Springer 2016.
2. Cosma Rohilla Shalizi, Advanced Data Analysis from an Elementary Point of View, online book, <http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/>

Other references:

1. Z. Ghahramani, Probabilistic Machine Learning and AI, talk, 2017, <https://www.youtube.com/watch?v=-47G_ULKAHk>

Courses

* Rahul Dave, MIT, AM207, Computational statistics … https://am207.github.io/2017/
* STA-663-2017, Computational Statistics in Python, <http://people.duke.edu/~ccc14/sta-663-2017/>
* CSC2541: Scalable and Flexible Models of Uncertainty, University of Toronto, Fall 2017, <https://csc2541-f17.github.io/#/contents/01-info>
* CSE 515T: Bayesian Methods in Machine Learning – Spring 2017 by Roman Garnett, <http://www.cse.wustl.edu/~garnett/cse515t/spring_2017/>
* Cosma Shalizi, 36-402, Undergraduate Advanced Data Analysis, Spring 2017, <http://www.stat.cmu.edu/~cshalizi/uADA/17/>
* Ralph Smith, MA 540 Uncertainty Quantification for Physical and Biological Models, http://www4.ncsu.edu/~rsmith/MA540\_s17.html

Useful links

## Detailed plan

Part I **Statistical concepts required to understand uncertainty**

### Lecture 1: Intro: Definitions and motivation

Topics:

* Uncertainty, data quality, data analysis
* Calibration, Precision, Accuracy, traceability, reproducibility, error
* Measurement model

Slides:

Reference and links:

1. Definitions related to the measurement system and its performance:
   * Chapter 3 from T. A. Reddy, Applied Data Analysis and Modeling for Energy Engineers and Scientists, Springer 2011
   * A.F. Mills and B.H. Chang, ERROR ANALYSIS OF EXPERIMENTS, A Manual for Engineering Students, University of California, 2004
2. STA-663-2017 Excellent course for statistics and data analysis in Python

[http://people.duke.edu/~ccc14/sta-663-2017/#](http://people.duke.edu/~ccc14/sta-663-2017/)

1. M. D. Bloice, A. HolzingeA, Tutorial on Machine Learning and Data Science Tools with Python, pp. 435–480, Machine Learning for Health Informatics, Volume 9605, Lecture Notes in Computer Science, Springer, 2016. ISBN: 978-3-319-50477-3.

<https://www.researchgate.net/publication/311555646_A_Tutorial_on_Machine_Learning_and_Data_Science_Tools_with_Python>

Github link: <https://github.com/mdbloice/MLDS>

* Video: Probabilistic Machine Learning and AI, <https://www.youtube.com/watch?v=-47G_ULKAHk>

Z. Ghahramani, “Probabilistic machine learning and artificial intelligence,” Nature, 2015, Excellent intro on uncertainty in neural networks and need for Bayesian approach in quantifying uncertainty.

Notebooks:

Homework:

Go through the following material:

Writing and organizing code for data analysis

* Scientific Software Engineering

<https://link.springer.com/content/pdf/10.1007/978-3-319-29439-1_5.pdf>

Jupiter

* Notes on using Jupyter: <http://people.duke.edu/~ccc14/sta-663-2017/00_Jupyter.html>

Python

* Read reference 2 and run notebooks NumPy.ipynb, Pandas Pandas.ipynb and Plotting.ipynb from <https://github.com/mdbloice/MLDS>.

Latex

* Mat Leonard, A Primer on Using LaTeX in Jupyter Notebooks, October 2016, http://data-blog.udacity.com/posts/2016/10/latex-primer/

Lecture scribing template

* <https://github.com/mbolic2/Uncertainty_Course/blob/master/Introduction/Lecture%20scribing%20template.ipynb>

Markdown

* Markdown basics: <http://markdown-guide.readthedocs.io/en/latest/basics.html>

### Lecture 2: Probability review – probability distributions

Topics:

Slides:

Reference and links:

Introduction to Probability Theory, Lecture 2,3 from the course MA598 Machine Learning and Uncertainty Quantification for Data Science <https://github.com/PredictiveModelingMachineLearningLab/MA598/blob/master/lectures/lec_02.ipynb>

<https://github.com/PredictiveModelingMachineLearningLab/MA598/blob/master/lectures/lec_03.ipynb>

Homework:

### Lecture 3: Statistical intervals

Topics:

* Confidence intervals for a Normal distribution
* Bayesian analysis and inference, Prediction and Credible intervals

Slides:

* Frequentism and Bayesianism: What's the Big Deal? (SciPy 2014)  
  <https://speakerdeck.com/jakevdp/frequentism-and-bayesianism-whats-the-big-deal-scipy-2014>

Reference and links:

* Confidence intervals, from J. Unpingco, Python for Signal Processing, Springer 2014.

<http://nbviewer.jupyter.org/github/unpingco/Python-for-Signal-Processing/blob/master/Confidence_Intervals.ipynb>

* **Advanced Python based course on confidence intervals**: Nonparametric Inference, Auditing, and Litigation, <https://github.com/pbstark/MX14/blob/master/index.ipynb>

Notebooks:

* Confidence Intervals, Chapter 26 from Philip B. Stark’s [SticiGui interactive statistics textbook](https://www.stat.berkeley.edu/~stark/SticiGui/index.htm) book, <https://www.stat.berkeley.edu/~stark/SticiGui/Text/confidenceIntervals.htm>
* Confidence Intervals and Gaussian Distribution
* Point Estimates and Confidence Intervals, <http://hamelg.blogspot.ca/2015/11/python-for-data-analysis-part-23-point.html> from <http://hamelg.blogspot.ca/2015/>
* Confidence Interval and Hypothesis Testing: <https://www.quantconnect.com/tutorials/introduction-python-confidence-interval-hypothesis-testing/>

Homework:

### Lecture 4: Bootstrap principles

Topics:

* Principles or resampling, pivoting, bootstrap for time series

Slides:

* Statistics for Hackers by Jake VanderPlas  
  https://speakerdeck.com/jakevdp/statistics-for-hackers

Reference and links:

* D. Shasha, M. Wilson, Statistics is Easy!, Second Edition, Morgan & Claypool Publishers 2011
* Resampling: <https://people.duke.edu/~ccc14/sta663/ResamplingAndMonteCarloSimulations.html>
* Notebook on resampling from STA-663-2017: <http://people.duke.edu/~ccc14/sta-663-2017/15B_ResamplingAndSimulation.html>

Further reading:

* Implementation of bagging and boosting: https://github.com/fonnesbeck/ngcm\_sklearn\_2017

Notebooks:

* Python code from “Statistics is Easy”
* Bootstrapped - confidence intervals made easy, <https://github.com/facebookincubator/bootstrapped#bootstrapped---confidence-intervals-made-easy>

Homework:

### Lecture 5: Monte Carlo methods

Topics:

* Random variable generation, Importance sampling, Metropolis-Hastings Algorithm, MCMC

Slides:

Reference and links:

* Chapters 5 and 6 from the book: W.M. Bolstad, Understanding Computational Bayesian Statistics, Wiley, 2010
* Interactive MCMC demos <https://chi-feng.github.io/mcmc-demo/>

Notebooks:

* Monte Carlo methods, STA-663-2017: <https://people.duke.edu/~ccc14/sta-663/MonteCarlo.html>
* Markov Chain Monte Carlo (MCMC), STA-663-2017: <https://people.duke.edu/~ccc14/sta-663/MCMC.html>
* Importance and rejection sampling:   
  Ch 10 Gelman - Python demos: https://github.com/avehtari/BDA\_py\_demos   
  Gibbs sampling Ch 11: https://github.com/avehtari/BDA\_py\_demos
* PyMC3 examples: https://people.duke.edu/~ccc14/sta-663/PyMC3.html

Homework:

### Lecture 6: Mathematical models

Topics:

Slides:

Reference and links:

* D. M. Blei, “Build, compute, critique, repeat: Data analysis with latent variable models.” Annual Review of Statistics and Its Application 1 (2014): 203-232
* Allen B. Downey, Modeling and Simulation in Python, Version 1.1, Green Tea Press, 2017, <http://greenteapress.com/wp/modsimpy/>
* Model specification, accuracy and reliability: Statistics and Data Analysis in MATLAB, <http://www.cmrr.umn.edu/~kendrick/statsmatlab/>

Python Notebooks: http://nbviewer.jupyter.org/github/mwaskom/Psych216/tree/master/

Notebooks:

* Modelling Dynamical Systems, from the course Computational Modelling in Neuroscience by Paul Gribble, https://www.gribblelab.org/compneuro/2\_Modelling\_Dynamical\_Systems.html
* Building a model: Ch. 2 - "From Words to Models: Building a Toolkit" of the book: Computational Modeling in Cognition: Principles and Practice by Lewandowsky and Farrell <http://smash.psych.nyu.edu/courses/spring12/modeling/>

Homework:

Part II **Uncertainty when no data and/or with historical/test data is available**

### Lecture 7: Uncertainty propagation

Topics:

* GUM
* Example of uncertainty propagation for temperature, pressure and other sensors

Slides:

* R. Bettencourt da Silva et. al. , Analytical measurement: measurement uncertainty

and statistics, Training in Metrology in Chemistry, <http://www.chem-soc.si/dokumenti/analytical-measurement-measurement-uncertainty-and-statistics/view>

* Tutorials on chaosPy at https://github.com/jonathf/chaospy/tree/development/tutorial

ChaosPy: <http://chaospy.readthedocs.io/en/master/index.html>

Reference and links:

* Numerical Methods using Python (scipy) – integration, differential equations, root finding, interpolation, curve fitting, Chapter 16 of Python for Computational Science and-Engineering, <https://www.southampton.ac.uk/~fangohr/teaching/python/book/html/16-scipy.html> . Also Symboli computation might be of interest: <http://localhost:8888/notebooks/Code/Books/introduction-to-python-for-computational-science-and-engineering-master/12-symbolic-computation.ipynb>
* A.F. Mills, B.H. Chang, Uncertainty propagation and error analysis, 46-147 Engineering IV, University of California, 2004. <https://pdfs.semanticscholar.org/0c35/cbd410af71673f0bcfd3107ccf6b30c8f2e7.pdf>

Notebooks:

* Further reading: https://www.ices.utexas.edu/media/reports/2017/1701.pdf Foundations of Predictive Computational Science CSE 397 / EM 397: Special Topics in Computational Science - OPAL method
* Python package for scientific data analysis: <https://github.com/Queens-Physics/qexpy/tree/garrow>

Homework:

### Lecture 8: Sensitivity analysis

Topics:

* Global sensitivity analysis, variance based method, Monte Carlo approaches, application to exploring sensitivity to parameters in the models in biomedical instrumentation

Slides:

Reference and links:

Notebooks:

* Vinzenz Eck, Jacob T. Sturdy, An introductory notebook on uncertainty quantification and sensitivity analysis, <http://nbviewer.jupyter.org/github/lrhgit/uqsa_tutorials/blob/master/index.ipynb>

<https://github.com/lrhgit/uqsa_tutorials>

V. G. Eck, W. P. Donders, J. Sturdy, J. Feinberg, T. Delhaas, L. R. Hellevik, and W. Huberts. A guide to uncertainty quantification and sensitivity analysis for cardiovascular

applications. Int J Numer Method Biomed Eng, 32(8):e02755, 2016.

* SALib - Sensitivity Analysis Library in Python, http://salib.readthedocs.io/en/latest/index.html

Homework:

* Modify the Notebook Uncertainty quantification and sensitivity analysis for arterial wall models

### Lecture 9: Parameter optimization - Regression analysis

Topics:

* Linear and non-linear fitting, Confidence intervals of the estimates

Slides:

Reference and links:

* Analysing regression results using Monte Carlo sampling: Emcee Example: Fitting a Model to Data <http://dfm.io/emcee/current/user/line/>
* Linear Regression and Its Cousins, This is linear regression without uncertainty with sklearn and statsmodels, <http://nbviewer.jupyter.org/github/leig/Applied-Predictive-Modeling-with-Python/blob/master/notebooks/Chapter%206.ipynb#6.-Linear-Regression-and-Its-Cousins>

Notebooks:

Homework:

### Lecture 10: Measuring agreement and performance

Topics:

Slides:

Reference and links:

* Measuring performance in Classification Models, <http://nbviewer.jupyter.org/github/leig/Applied-Predictive-Modeling-with-Python/blob/master/notebooks/Chapter%2011.ipynb>

Notebooks:

Homework:

Part III **Uncertainty through Bayesian analysis**

### Lecture 11: Probabilistic programming and Bayesian inference

Topics:

Slides:

Reference and links:

* Probabilistic programming PyMC3 examples Lab 7 - Bayesian inference with PyMC3 from the course AM207: https://am207.github.io/2017/lectures/lab7.html
* Bayesian Statistical Analysis with PyMC3:

https://github.com/ericmjl/bayesian-stats-talk/blob/master/slides.ipynb

Notebooks:

Homework:

### Lecture 12: Bayesian learning and parameter estimation

Topics:

Slides:

Reference and links:

Advanced:

* Bayesian and classical approach Errors and uncertainty in variables – When to worry and when to Bayes? <http://www.biometrische-gesellschaft.de/fileadmin/AG_Daten/BayesMethodik/workshops_etc/2016-12_Mainz/Muff2016-slides.pdf>
* A Primer on Bayesian Methods for Multilevel Modeling: https://github.com/fonnesbeck/multilevel\_modeling/blob/master/multilevel\_modeling.ipynb

Notebooks:

* A mixure model - Fitting straight line with outliers

Homework:

### Lecture 13: Classifying outcomes with logistic regression

Topics:

Slides:

Reference and links:

Notebooks:

Homework:

### Lecture 14: Surrogate models and Gaussian processes

Topics:

Slides:

Reference and links:

* Lectures from the course MA598 Machine Learning and Uncertainty Quantification for Data Science Priors on Function Spaces: Gaussian Processes
* Excellent video: Richard E. Turner, Keynote Lecture: Gaussian Processes for Signal Processing, MLSP, 2016 <http://rc.signalprocessingsociety.org/sps/product/conference-videos-and-slides/SPSVID00121?source=IBP>
* Gaussian processes for regression, classification and reduction: Gaussian Processes for Regression: A Quick Introduction, M. Ebden, August 2008 https://arxiv.org/pdf/1505.02965.pdfSlides: <https://duvenaud.github.io/sta414/lectures/Lecture11_GPs_2.pdf>

Notebooks:

* GPFlow library [http://gpflow.readthedocs.io/en/latest/#](http://gpflow.readthedocs.io/en/latest/)
* Propagating uncertainty using Gaussian processes: scikit-GPUPPY: Gaussian Process Uncertainty Propagation with Python: <https://github.com/snphbaum/scikit-gpuppy>
* New library based on Torch: GPyTorch (Pre-release, under development), <https://github.com/jrg365/gpytorch>
* magpie, Model Analysis with Gaussian Processes in Python, https://github.com/samcoveney/maGPy

Homework:

### Lecture 15: Bayesian optimization

Topics:

Slides:

Reference and links:

* Lecture notes, Lecture 13, CSE 515T: Bayesian Methods in Machine Learning – Spring 2017 by Roman Garnett, http://www.cse.wustl.edu/~garnett/cse515t/spring\_2017/

Notebooks:

* GPyOpt, http://sheffieldml.github.io/GPyOpt/

Homework:

### Lecture 16: Time series analysis

Topics:

Slides:

Reference and links:

* Modeling time series: Chapter 8 from PETER CONGDON, Bayesian Statistical Modelling, Second Edition

Notebooks:

* Analysis of An AR(1)AR(1) Model in pyMC3, https://hub.mybinder.org/user/pymc-devs-pymc3-f0eg9n8b/notebooks/docs/source/notebooks/AR.ipynb

Homework:

### Lecture 17: Particle filters

Topics:

* State-space model, Bayesian filtering and Monte Carlo simulations, From complex probabilistic formulas to implementation

Slides:

Reference and links:

* Good course: ELEC-E8105 - Non-linear filtering and parameter estimation

Notebooks:

Homework:

### Lecture 18: Sensor fusion

Topics:

Slides:

Reference and links:

Notebooks:

Homework: