Uncertainty Evaluation in Engineering Measurements and Machine Learning

Instructor: *Miodrag Bolic, University of Ottawa, Ottawa, Canada*

January 13, 2026

Course Information

Time and place: Winter 2026, Wednesday 2:30 pm – 5:30 pm
Course code: ELG 5218, CSI 5218, EACJ 5600
Github page: https://github.com/Health-Devices/Course-Uncertainty-Machine-Learning-2026
Contact: Miodrag Bolic's email, Xuanyu Su's email

Calendar Style Description: Uncertainty, uncertainty propagation, Bayesian inference, Bayesian filtering, data fusion, metrology, measurement science, error analysis, measures of agreement, data quality, and data quality index. Case studies will be drawn from various fields, including biomedical instrumentation, sensors and signal processing.

Prerequisites

We expect participating students to bring basic knowledge and experience in:

- Elementary probability
- Elementary statistics
- Machine learning

Grading

For collecting the credits the students are expected to complete:

- Assignments or projects (30% of the grade)
- Midterms (30% of the grade), on February 11 and March 18.
- Final exam (40% of the grade)

Course organizing principle

This course begins with *measurement uncertainty* and progressively builds modern uncertainty-aware ML:

- **Bayesian modeling** (modeling simple problems),

- **Classical probabilistic inference** (IID and sequential),
- **Latent-variable models** (state-space models, Kalman filters)
- **Uncertainty scores** (proper scores, calibration, conformal prediction),
- **Modern architectures** (multimodal fusion, diffusion models, SciML),
- **Sequential decision making** (intro to reinforcement learning).

Course materials will be provided as weekly slides and Jupyter notebooks. Each week, the notebooks will be posted in advance of (or shortly after) the lecture and will include the key concepts, derivations, examples, and implementation exercises covered that week. Students are expected to review the notebooks regularly, work through the examples, and ensure they understand the underlying concepts. The notebooks are an essential component of the course and will be used to support assignments, discussions, and exam preparation.

Lecture plan

Each lecture is one 3-hour block with (A) concepts/math, (B) implementation blocks, (C) evaluation. The 12 lectures are shown in Table 1. Please note that the lecture plan is preliminary and will likely change during the semester.

Why Uncertainty Quantification?

Over the last several years, deep neural networks have advanced many applications, including vision, language understanding, speech understanding, robotics, and more. But a major challenge still remains: how to model uncertainty. Good models of uncertainty are crucial whenever decisions need to be made or an algorithm must decide when and how to acquire new information. Uncertainty quantification involves combining computational models, physical observations, and, if applicable, expert judgment to infer about a physical system. Types of uncertainties include aleatoric and epistemic uncertainties as shown in Figure 1.

Motivation in Engineering

Uncertainty quantification (UQ) is the practice of (i) identifying important sources of uncertainty in a physical system, (ii) representing them in a model, and (iii) propagating them to understand how they affect quantities of interest such as stress, temperature, efficiency, or time-to-failure. Instead of reporting a single number (a point estimate), UQ reports uncertainty-aware statements such as ranges, confidence levels, or failure likelihoods under realistic variability in loads, materials, manufacturing tolerances, and operating conditions.

- **Safety and reliability (risk-aware design):** Engineering decisions are often constrained by rare but high-consequence events. UQ supports safety by translating variability in inputs (e.g., wind gusts, traffic loads, temperature cycles, sensor errors) into quantified reliability margins and failure risk. *Example:* a bridge design can be evaluated under variability in vehicle loads and material strength to determine how often stresses may exceed allowable limits, which is more informative than a single worst-case scenario chosen ad hoc.

Table 1: Preliminary lecture plan.

L#	Theme	Detailed topics
L1	Bayesian modeling and likelihoods	Measurement models; MLE/MAP; identify aleatoric vs epistemic uncertainty
L2	Bayesian regression and classification	Compute predictive distributions; intervals; detection/classification
L3	Monte Carlo, bootstrap, propagation	Propagate uncertainty through pipelines; bootstrap uncertainty; sensitivity basics
L4	State-space models	Kalman/HMM fundamentals; interpret covariance evolution; Particle filters;
L5	Deep learning + latent representations	VAE/ELBO; connect state-space and deep learning
L6	Proper scores and calibration	NLL/Brier/CRPS; reliability diagrams; ECE; temperature scaling
L7	Conformal prediction	Build prediction intervals/sets with finite-sample coverage; efficiency analysis
L8	Multimodal fusion with uncertainty	Encoders + fusion; UQ wrappers
L9	Diffusion models I (DDPM)	Implement forward noising + denoising loss; reverse sampling intuition
L10	Diffusion models II (signals/time series, latent diffusion)	Conditional diffusion; latent diffusion; fast sampling; sample-based evaluation
L11	Scientific ML + out-of-distribution (OOD)	PINNs/operators/inverse problems; attach uncertainty; OOD detection and drift monitoring
L12	Sequential decision making	Uncertainty in sequential decision making, POMDP models

- **Cost–performance–robustness trade-offs (avoiding over- and under-design):** Deterministic "worst-case" margins can be overly conservative (high cost, high weight, low efficiency) or accidentally insufficient (unexpected failures). UQ enables selecting safety margins that are consistent with measured variability and required reliability targets. *Example:* in aerospace structures, explicit uncertainty modeling can reduce unnecessary mass while still meeting strict reliability requirements; in energy systems, UQ can prevent oversizing components while maintaining reliability across operating conditions.

- **Sensitivity analysis (what drives uncertainty?):** Sensitivity analysis identifies which uncertain inputs (e.g., a calibration constant, a friction coefficient, a material parameter) dominate the variability in outcomes, guiding where to invest in better measurements, tighter tolerances, or improved models. *Example:* if output variability is driven mainly by sensor calibration drift rather than by environmental variability, improving calibration procedures yields a larger reliability gain than redesigning the mechanical structure.

Motivation in Machine Learning

In machine learning, uncertainty quantification addresses the question: *how much should we trust this prediction, and what should we do when we do not trust it?* This is critical in deployment, where inputs may differ from training conditions and errors can have real-world cost.

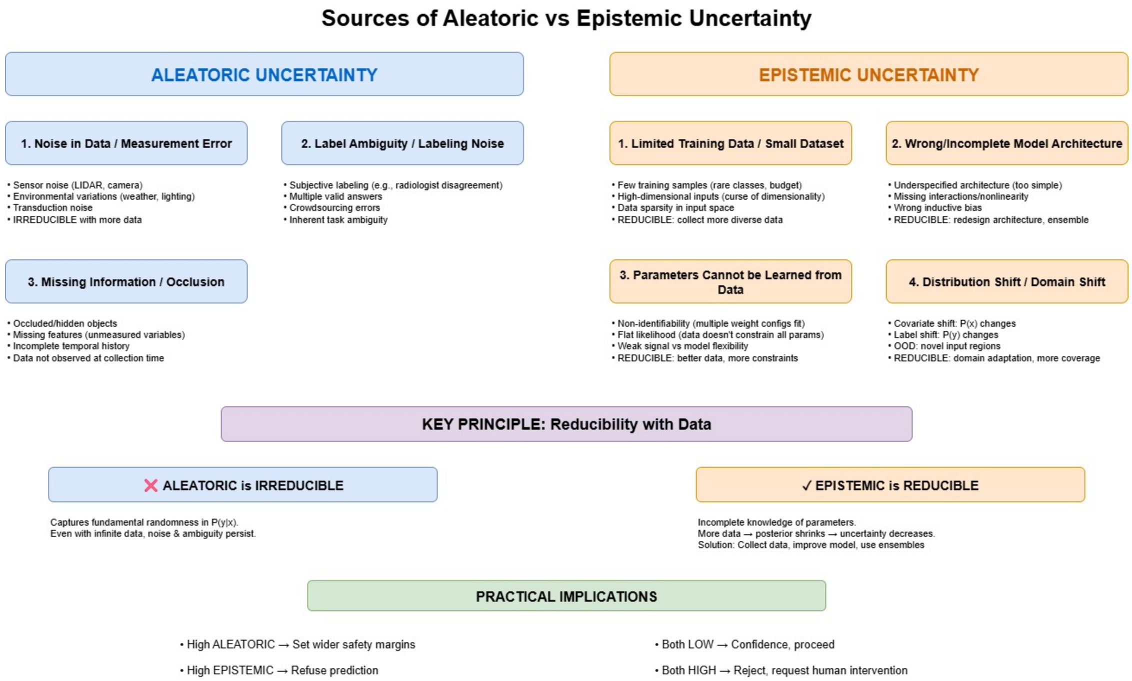

Figure 1: Sources of Aleatoric vs Epistemic Uncertainty

- **Safe deployment and autonomy (detecting low-confidence and unfamiliar inputs):** Deployed models will encounter cases that are under-represented in training (distribution shift) or fundamentally new (out-of-distribution). Reliable uncertainty estimates enable systems to trigger safer behavior such as abstaining, requesting human review, switching to conservative rules, or collecting more information. *Example:* a medical imaging system can flag uncertain cases for clinician review; an autonomous robot can slow down or stop when perception uncertainty is high; an industrial inspection system can request re-scans when data quality is poor.

- **Decision-making and downstream optimization (risk-aware use of predictions):** Many pipelines use ML outputs inside optimizers, controllers, or decision-support tools. UQ allows these downstream components to account for uncertainty, not only average performance, thereby reducing brittle behavior. *Example:* in reinforcement learning, uncertainty can guide exploration toward informative actions while avoiding risky ones.

- **Model selection, debugging, and trust (calibration and interpretability):** UQ helps determine whether errors come from irreducible noise (aleatoric) or lack of knowledge (epistemic), guiding whether to collect more data, redesign features, or change the model class. Calibration is essential: uncertainty outputs should be consistent with observed error frequencies, so that statements like "high confidence" or "90% interval" carry their intended meaning. *Example:* if a classifier repeatedly assigns high confidence to wrong predictions on a new device or new hospital site, UQ diagnostics can reveal miscalibration and motivate recalibration, domain adaptation, or additional data collection before deployment.

Use of Generative AI Tools

(This section is modified from UToronto CSC311 - Introduction to Machine Learning (Fall 2024))

You may use generative AI tools (e.g., ChatGPT, Copilot) as learning aids, but they are not required for this course. Treat them as supplementary resources for exploration—not as substitutes for your own work. You are fully responsible for your learning and for all submitted work.

Important Rules

- No AI use on exams (midterm or final).
- Assignments are designed to be completed without AI tools using course concepts.
- Overreliance on AI can harm your understanding and lead to poor performance later.

Risks and Warnings

- AI tools may produce incorrect or misleading content, false citations, or violate privacy and security standards.
- Submitting false citations or uncredited AI-generated content is an academic offense.

Permitted Uses

- Answer general conceptual questions.
- Provide examples of library/API usage.
- Summarize information.
- Generate test cases.
- Assist with debugging.

Prohibited Uses

- Copying AI-generated text or code into your submission without acknowledgment.
- Using AI beyond the permitted purposes.

Documentation Requirements

- Include a statement in your submission describing which tool you used and how.

Links

Relevant Courses

- Our course in 2021
- Bayesian Machine Learning and Information Processing (5SSD0) by Prof. dr. ir. Bert de Vries
- Advanced Bayesian Learning by Mattias Villani

- Probabilistic Machine Learning (Summer 2020)
- Prince, Simon J.D. (2025). Understanding Deep Learning

Books

- Murphy, Kevin P. (2021). *Probabilistic Machine Learning: An Introduction*. MIT Press.
- Murphy, Kevin P. (2022). *Probabilistic Machine Learning: Advanced Topics*.
- Izbicki, Rafael (2025). *Machine Learning Beyond Point Predictions: Uncertainty Quantification,* `https://github.com/rizbicki/UQ4ML`.
- Villani, Mattias, (2025). *Bayesian Learning.*
- Prince, Simon J.D. (2025). Understanding Deep Learning
- Theodoridis, Sergios (2025) Machine Learning: From the Classics to Deep Networks, Transformers, and Diffusion Models
- Barber, David. *Bayesian Reasoning and Machine Learning*.
- Davidson-Pilon, Cameron. *Bayesian Methods for Hackers*.
- Goodman, N. D., Tenenbaum, J. B., and The ProbMods Contributors (2016). *Probabilistic Models of Cognition* (2nd ed.).
- Winn, John. *Model-Based Machine Learning*: `https://www.mbmlbook.com/`
- Cinelli, L. P. et al. (2021). *Variational Methods for Machine Learning with Applications to Deep Networks*.
- Thuerey, N. et al. (Dec 2021). *Physics-based Deep Learning*: `http://physicsbaseddeeplearning.org`
- Martin, Osvaldo et al. (2021). *Bayesian Modeling and Computation in Python*.
- Yadav, Neha; Yadav, Anupam (2015). *An Introduction to Neural Network Methods for Differential Equations*.
- Goulet, James-A. (2020). *Probabilistic Machine Learning for Civil Engineers*: `http://profs.polymtl.ca/jagoulet/Site/Goulet_web_page_BOOK.html`
- ME 539 - Introduction to Scientific Machine Learning: `https://github.com/PredictiveScienceLab/data-analytics-se?_ga=2.66952878.1678226653.1765915278-1039759646.1765915278`

Acknowledgements/ Document preparation

This document was converted into LaTeX and formatted with assistance from ChatGPT (OpenAI). The instructor provided the source text and verified the final structure and wording.