

PX914-15 Predictive Modelling and Uncertainty Quantification

20/21

Department

Physics

Level

Taught Postgraduate Level

Module leader

James Kermode

Credit value

15

Module duration

10 weeks

Assessment

60% coursework, 40% exam

Study location

University of Warwick main campus, Coventry

Description

Introductory description

N/A.

[Module web page](#)

Module aims

This module covers predictive modelling techniques including probability theory, machine learning, data analytics and data mining. These methods are essential for solving problems in the interdisciplinary area of predictive modelling. The module aims to equip students with a knowledge of random processes, statistical learning theory, Bayesian inference, Monte Carlo methods, model selection, and supervised and unsupervised machine learning techniques. This will enable students to solve complex predictive modelling problems using advanced, cutting edge techniques, as well as adapt the techniques or develop new techniques for data analysis and predictive modelling.

Links will be made to simulations of molecular dynamics with classical force field models, electronic structure ab initio approaches such as Density Functional Theory, Monte Carlo sampling techniques, as applied to diverse materials systems. Particular emphasis will be given to scalable

approaches for uncertainty quantification and propagation in multiscale materials models (from ab initio to continuum), description of random microstructures, information theoretic approaches to coarse graining, and statistical learning approaches for exploring high-dimensional structure/property/process relations.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

1. Probability theory (4 lectures - optional for students with strong Maths background)
 - a. Basic concepts such as probability space, events, expectation, moments, densities, generating functions, conditioning, marginalization, independence
 - b. Joint/conditional probability densities, (conditional) expectations
 - c. Random variables/vectors, covariance, correlation, random processes, mean and covariance functions, cross-covariance, classification of processes, ergodicity
 - d. Markov chains, Poisson processes, continuous time Markov chains, Brownian motion, martingales
 - e. Classical Monte Carlo (rejection, importance sampling), convergence property, Markov chain Monte Carlo (Gibbs, Metropolis-Hastings) Monte Carlo, Random number generation, univariate and multivariate distributions
 - f. Bayes formula, illustration with binary variables (events), Bayes formula for continuous variables: likelihood, conjugate priors
2. Machine learning essentials (4 lectures)
 - a. Statistical learning introduction and background
 - i. Decision theory; Bayes risk
 - ii. Probabilistic models
 - iii. Complexity, regularization, bias vs. variance
 - iv. Resampling, cross-validation
 - b. Unsupervised techniques
 - i. Linear dimensionality reduction: PCA and SVD, MDS
 - ii. Nonlinear dimensionality reduction: LLE, Isomap, kPCA, diffusion maps
 - iii. Clustering methods, K-means; hierarchical algorithms; probabilistic model-based clustering; graph-based/spectral clustering
 - iv. Density estimation
 - v. Gaussian mixture models
 - vi. Expectation-maximization
 - c. Supervised techniques for regression and classification
 - i. Linear methods: linear, logistic, Bayesian regression and generalized linear models, naive Bayes, LDA, SVM
 - ii. Nonlinear methods: kernel methods, nearest neighbor, decision trees, neural networks, Gaussian process regression
 - d. Semi-supervised techniques
 - e. Ensemble methods (bagging, boosting, random forests)
3. Uncertainty propagation through surrogate-model construction (4 lectures)
 - a. Statistical emulators
 - b. Deterministic vs Bayesian training and cross-validation

- c. Gaussian processes and limitations,
 - d. Multivariate RVM
 - e. Mixtures/products of models (mixtures/products of experts)
 - f. Spectral Stochastic Methods (generalized polynomial Chaos, gPC)
 - i. Intrusive vs non-intrusive, collocation, sparse-grid, tensor products
 - ii. Sparse Polynomial Chaos
 - g. Illustration for ODE with various input dimensions
4. Predictive materials modelling (4 lectures)
- a. Statistical thermodynamics and Monte Carlo methods (stochastic exploration of potential energy surfaces, ab-initio thermodynamics, structure prediction).
 - b. Random microstructures (effective properties, property variability, sampling of microstructures, microstructure and materials failure). 4 lectures.
 - c. Model errors (constitutive model errors, limits of density functional theory, transferability of exchange-correlation functionals and pseudopotentials, uncertainty quantification of effective potentials, sampling of thermodynamic quantities).
 - d. High dimensionality in materials modelling (challenges in simulation and uncertainty quantification, dimensionality reduction, coarse graining and microscopic model reconstruction, model selection).
 - e. Machine learning and information (statistical learning approaches, materials genome, materials informatics)

Learning outcomes

By the end of the module, students should be able to:

- Demonstrate knowledge of statistical and mathematical methods for predictive modelling.
- Perform detailed, advanced analyses of complex data sets, extracting information and developing relationships using linear and nonlinear regression and classification techniques.
- Systematically develop models for predictive purposes using advanced techniques of model selection and evaluation.
- Understand and apply cutting-edge methods of machine learning.
- Demonstrate an understanding of complex modelling transferability issues arising from, e.g. choices of exchange-correlation functionals and pseudo-potentials in electronic structure, or the choice of force fields in atomistic and molecular models.
- Demonstrate a detailed knowledge of, and be able to apply models, for quantifying uncertainties arising in material structure and properties, constitutive models, from limited data scenarios and through coarse graining.

Indicative reading list

Probability Random Variables and Stochastic Processes, by A. Papoulis

Pattern Recognition and Machine Learning, by Christopher Bishop

Bayesian Data Analysis, 3rd Edit., by A. Gelman et al.

The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, by Trevor Hastie, Robert Tibshirani, Jerome Friedman

Monte Carlo Strategies in Scientific Computing, by J.S. Liu

Information Theory, Inference, and Learning Algorithms, by D. MacKay

Subject specific skills

Demonstrate knowledge of statistical and mathematical methods for predictive modelling
Perform detailed, advanced analyses of complex data sets, extracting information and developing relationships using linear and nonlinear regression and classification techniques
Systematically develop models for predictive purposes using advanced techniques of model selection and evaluation
Understand and apply cutting-edge methods of machine learning
Demonstrate an understanding of complex modelling transferability issues arising from, e.g. choices of exchange-correlation functionals and pseudo-potentials in electronic structure, or the choice of force fields in atomistic and molecular models.
Demonstrate a detailed knowledge of, and be able to apply models, for quantifying uncertainties arising in material structure and properties, constitutive models, from limited data scenarios and through coarse graining.

Transferable skills

Mathematical analysis, statistics, coding, writing

Study

Study time

Type	Required
Lectures	8 sessions of 2 hours (11%)
Practical classes	4 sessions of 2 hours (5%)
Private study	86 hours (57%)
Assessment	40 hours (27%)
Total	150 hours

Private study description

Further reading, project preparation

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Assessment group D

	Weighting	Study time
Assessed work	60%	30 hours
1. Based on the machine learning workshop exercises.		
2. Based on the uncertainty propagation workshop.		
3. Based on predictive multiscale modelling.		
Viva voce Exam	40%	10 hours
On the core material. 30 minutes.		

Feedback on assessment

-\tWritten annotations to submitted computational notebooks\r\n-\tVerbal discussion during viva voce exam\r\n-\tWritten summary of viva performance

[Past exam papers for PX914](#)

Availability

Courses

This module is Core for:

- Year 1 of TPXA-F344 Postgraduate Taught Modelling of Heterogeneous Systems
- Year 1 of TPXA-F345 Postgraduate Taught Modelling of Heterogeneous Systems (PGDip)