Padova, October 30, 2023

COMPLEX AND DATA DRIVEN CHEMISTRY

GENERAL OVERVIEW

I° Semester	CFU	SSD
Statistical Learning for Chemistry (with coding lab)	10	02
Characterization of complex chemical systems	6	02/03
Analysis of Complex Chemical Systems	6	01
Catalysis	6	03
II° Semester		
Multiscale Chemical Modeling	6	02/03
Chemometrics	6	01
Organic Synthesis Design	6	06
Systems Chemistry	6	06
Optional Course	6	
III° Semester		
Applied Chemistry	6	
Optional Course	6	
Internship Project	10	
IV° Semester		
Final research Project	40	

DETAILED PROGRAMS

I° SEMESTER

	CFU	SSD
Statistical Learning for Chemistry (with coding lab)	**10**	**02**
Characterization of complex chemical systems	**6**	**02/03**
Analysis of Complex Chemical Systems	**6**	**01**
Catalysis	**6**	**03**

Statistical Learning for Chemistry

Lectures (6CFU):

A few notions of statistics and probability. Introduction: machine learning general terminology and machine learning in chemical sciences: classification, regression, supervised, unsupervised and reinforcement learning. Data sets and scaling: generalization and statistical learning theory, regularization; model selection and validation. Linear models: multiple linear regression; bias, normalization, condition, regularization; nonlinear regression and the kernel trick; trees and random forests. Representations of atomistic systems: feature selection, linear filtering and other selection methods; overview of software employed for chemical representation. Dimension reduction techniques: principal component analysis. Neural networks and learned representations: the multilayer perceptron, optimization and training, regularization and hyperparameter selection. Convolutional neural networks; neural network potentials; generative models.

Advanced topics: unsupervised learning, deep learning, and transfer learning in chemical applications; genetic algorithms and other statistical learning methods

Practice (4CFU):

1. Introduction to Python in chemistry: Python programming, relevant libraries for chemical applications (e.g. Numpy, scikit-learn, PyTorch, DeepChem); Google Colab overview
2. Visualization and dimension reduction; molecular representations and format interconversions;
3. Classification via K-nearest neighbours algorithm; regression analysis
4. Setting up and training a neural network
5. Applications and case-studies in chemistry and materials sciences.

Characterization of complex chemical systems

Part A: Optical spectroscopies and big data analysis

Theory about matter/EMR (electro-magnetic radiation) interactions: absorption, transmission, diffusion, and scattering. Molecular vibrations: harmonic oscillator model and vibrational normal modes.

Vibrational spectroscopy #1 (IR & ATR): basic theory and instrumental implementation. Vibrational spectroscopy #2 (Raman): basic theory and instrumental implementation.

Applications of IR & Raman spectroscopies: big data acquisition for calibration; micro-spectroscopic techniques; chemical imaging (concepts and purposes); big data acquisition for chemical imaging.

Good practices in big data acquisition: sample nature (solution, surface, layers, colloid, wet/dry, solvents, …); figures of merit and how to deal with them (replicates and/or mapping, choose of optics, choose of containers/substrates, signal-to-noise ratio, spectral and spatial resolution); typical post-acquisition procedures (background removal, baseline, artifacts, normalizations); examples of mono- and multivariate analysis for the interpretation of IR and Raman data (e.g.: PCA, Cluster Analysis, Pearson's Correlation, Regression models, PLSR).

Part B: Synchrotron radiation and x-rays spectroscopies

Synchrotron radiation, irradiated power, bending magnets and wigglers, beam-lines and experimental stations. X-ray photoelectron spectroscopy (xps), the photoelectric effect, the binding energy scale, photoelectric cross-section, spin orbit coupling and doublet terms, Auger peaks, multiplet splitting, final state effects, inelastic mean-free path and surface sensitivity, practical measurement (survey scan and high resolution spectra),

conventional sources and synchrotron radiation, electron analyser and UHV chambers, elemental composition of a surface, oxidation state assignment, angle- resolved xps and thickness measurement of ultrathin films. UPS spectroscopy, valence band edge and work function measurement. X-ray absorption spectroscopy, pre-edge XANES. EXAFS. Short-range order, coordination numbers. Spectra simulation and practical applications. In “operando” measurements for the characterization of catalysts and photo-electrodes."

Analysis of Complex Chemical Systems

Complex systems and omics. Problem definition and data acquisition. Targeted and untargeted analysis. Metabolomics and proteomics. Multiomics and data fusion. Development/optimization of an analytical workflow (from samples selection and preparation to data acquisition and analysis).
Sample preparation and pre-treatment method. Analyte extraction and concentration. Solid-phase extraction (SPE), QuEchERS, solid-phase microextraction (SPME), thermal desorption. On-line methods. Examples for environmental, clinical/biochemical and food&beverages applications.
Advanced mass spectrometry and hyphenated techniques. Gas (GC) and liquid chromatography (HPLC, UPLC). Column selection and method development. Electrophoresis and capillary electrophoresis. Bidimensional applications. Advanced inductively coupled plasma (ICP) methods for elemental analysis. Ionization techniques and mass analyzers. Accurate mass and high-resolution mass spectrometry. Tandem mass spectrometry. Method development.
Big Data elaboration: from instrumental raw data to compound identification. Data pre-processing. Examples of statistical analysis. Mass spectra analysis and structure elucidation. Identification levels. Databases and software (freeware, open-access and proprietary tools). Validation procedures. Examples on real datasets.
Examples of workflows (open discussion with students) related to environmental and biochemical applications.
NMR spectroscopy: physical principles. The NMR spectrometer. Practical recommendations for recording spectra. One- and two-dimensional NMR spectroscopy: practical strategical approaches to characterize the chemical structure of metabolites. Interpretation of NMR spectra of organic molecules. Quantitative NMR: different approaches to quantify the concentration of metabolites in complex matrices. Multivariate analysis of NMR data. Targeted and untargeted approach. Typical metabolomic analysis workflows. Identification of metabolites. Metabolites databases.

Catalysis

Part 1 (4 CFU)

Catalyst definition; the concept of kinetic control; the concept of active site; catalytic cycle; turn-over number; turnover frequency; catalytic activity and selectivity; functionality; atom efficiency. Homogeneous and heterogeneous catalytic reactions: ideal mechanisms; degree of coverage in heterogeneous catalysis; molecular and dissociative adsorption and coadsorption; Langmuir isotherms; derivation of rate laws; Michaelis-Menten kinetics; variability of the reaction order in complex reactions and its relationship with the apparent activation energy. Diffusion limitations on the rate of heterogeneous catalytic processes: catalytic efficiency; Thiele module. Homogeneous acid-base catalysis: general and specific acid and base catalysis; hydrolysis, condensation and alkylation reactions. Principles of applied organometallic chemistry of transition metals: σ-donation and π-backdonation; metal-CO and metal-alkene bond; the rule of 16-18 electrons; typical binders. Coordination and dissociation of ligands; oxidative addition, reductive elimination; migratory insertion, α- and β-elimination. Industrial homogeneous catalytic processes: hydroformylation of alkenes; carbonylation of methanol to acetic acid: Monsanto and Cativa processes; oligomerization stereospecific polymerization of alkenes with “single-site” and Ziegler-Natta catalysts; olefin metathesis; cross-coupling reactions. Unsupported and supported industrial heterogeneous catalysts: typical components; role of support; preparation methods. Measurement of the surface area of solids using equilibrium physisorption methods; real isotherms; BET isotherm and BET method; determination of sites on the active surface through chemisorption and “temperature programmed desorption” methods. Porosity, tortuosity and roughness in solids; micro-,

meso-, and macro-porous solids; measurement of pore size (diameter, volume) and their relationship with the specific surface area.
{Synthesis, structure and properties of zeolites, mesoporous silica and MOF (possibly move to part 2)}. Surface energy of solids and catalytic activity; fraction exposure of the active component and its relationship with catalytic activity; structure of the crystal lattice and the metal surface; the "broken-bond" model for surface energy; Wulff construction; surface defects; surface sites. Heterogeneous catalysts lifetime and deactivation. Ammonia industrial synthesis mechanism; "structure sensitivity"; electronic promoters. Band theory e chemisorption on metals surface; Sabatier principle; volcano diagrams. Catalytic converters: general concepts; "three-way" catalytic converters; mechanism of CO oxidation and NOx reduction; reconstruction of the active surface in the oxidation of CO. Oxidation of ethylene to ethylene oxide. Chemisorption on semiconducting and insulating oxides. Solid acids: isoelectric point; Hammett acidity function; superacidity; structure of acid sites and their strength in typical solids; difference between Brønsted-Lawry acid sites and Lewis acid sites; activation of alkanes and alkenes; carbonium and carbenium ions. Bifunctional catalysis: catalytic reforming; "selective catalytic reduction". New directions in catalysis: single atom catalysis, photocatalysis, electrocatalysis etc. Possible inclusion of elements of biocatalysis.

Part 2 (2 CFU)

Synthesis of catalytic materials and supports for heterogeneous catalysis. Review of inorganic materials description and structure: crystalline, amorphous, polymeric and nanodimensional, relationship between structure and chemical bond type, defects in solids (with particular reference to point defects). Solid state synthesis methods (mechanosynthesis, carbothermal reduction, combustion synthesis, ceramic synthesis). Formation of solids from gas phase (aerosol process, spray pyrolysis, CVD). Formation of solids from liquid phase: nucleation and growth, La Mer model, classical and non-classical theories of crystallization, Ostwald phase rule; sol-gel techniques; dispersion methods and nucleation/growth methods. Nucleation from solution and seeded growth. Hydro- and solvothermal synthesis (subcritical and supercritical). Polyol-assisted synthesis. Microwave-assisted synthesis. Laser-assisted synthesis, and sonochemical and radiochemical methods. Flow synthesis and microfluidics. Confined space synthesis: microemulsions, miniemulsions. Biogenic syntheses. Approaches to morphological and dimensional control: anisotropy.

II° SEMESTER

	CFU	SSD
Multiscale Chemical Modeling	**6**	**02/03**
Chemometrics	**6**	**01**
Organic Synthesis Design	**6**	**06**
Systems Chemistry	**6**	**06**
Optional Course	**6**	

Multiscale Chemical Modelling

The hierarchy of the physical models of matter: from quantum mechanics to the classical motion of nuclei on a potential energy surface, to continuum models. These models will be apply both to isolated molecular systems and to periodic crystal structures.
Basics of statistical mechanics and thermodynamics, time and ensemble averages, molecular distribution functions, partition function, canonical and isothermal-isobaric ensembles, fluctuations, equilibrium distribution functions, calculation of free energies.
Sampling of the classical phase space: Montecarlo methods, enhanced sampling, biased dynamics. Coarse-graining in time and in space. Network models of conformational dynamics: partition of the phase space into a discrete set of metastable states, Markov models, master equation.
Case-study: charge transfer/transport in (bio)molecular systems. Marcus theory of electron transfer, reorganization energy, multiscale computational approach to the problem.

Boundary between quantum and classical descriptions, hybrid methods: models of solvation, embedding schemes, QM/MM and QM/MM/MD. For the periodic systems, ab initio quantum embedding techniques such as: density embedding, density-matrix and Green's function (GF) embedding, QM/MM embedding.
Introduction to density functional theory and time dependent density functional theory, theoretical modeling of excited states and excited state processes.
Geometrical description of solids. Direct and Reciprocal lattices, band theory of crystals, electronic properties and treatment of dopants. Scattering experiments.
Case study: description of the metal centre in soft matter (QM/MM enzyme), application to spectroscopy (absorption/emission spectra by using TDDFT and solvent effect), heterogeneous catalysis (Machine-learning guided search of active sites).

Chemometrics

Design of experiment in complex systems. Interactions. Modeling. Factorial designs. Face Centered Design. Mixture design.
Univariate statistical analysis. Types of error. Elements of classical statistics. Normal distribution. Student's t-distribution. Fisher's F-distribution. Sample distribution of the mean. Central limit theorem. Statistical inference (Parametrical Tests). Non-parametric statistics (non-parametric tests). Applicative examples using MATLAB or R.
ANOVA for comparing various confidence intervals. Calibration. Data elaboration and presentation. Result interpretation. List of possible sources of uncertainty. Applicative examples.
Quality control and norms. Performance parameters of an analytical method. Estimating Trueness. Estimating Precision. Different approaches to the estimation of uncertainty. Detection limit. Quantification limit. Linearity range. Youden's factorial plan. Evaluation of the Robustness of the Experimental Method. Internal Controls. External controls. Internal audit for maintaining laboratory accreditation according to UNI CEI EN ISO/IEC 17025. Control charts. Accreditation and certification. Applicative examples.
Multivariate data exploration. Introduction to multivariate data. Data pretreatment. Missing data. Variable transformation. Centering. Scaling. Variable transformation. Covariance. Correlation.
Principal component analysis (PCA). Loading plots. Score plots. Selecting principal components. Scree plot.
Clusters analysis. Distance matrix. Similarity matrix. Agglomerative hierarchical methods. Dendrograms. Partitional methods. Applicative examples.
Multivariare modeling. Model linearity. Model order. Control parameters. Model validation.
Classification models. Confusion matrix. Loss matrix. Evaluation parameters of the classification. Misclassification risk (MR%). K-NN. Discriminant analysis (DA). The SIMCA classification method.
Calibration models. Multiple Linear Regression (MLR). Leverages. Regression coefficients. Evaluation parameters for a regression model. Correlation coefficient. Prediction coefficient. Standard error of the estimate. Diagnostic methods for regression models. Principal Component Regression (PCR). Partial Least Squares Method (PLS). Applicative examples.

Organic Synthesis Design

Carbanions in organic synthesis. Metal-halogen exchange, ortho-lithiation, Olefination reactions and related methods for the construction of double bonds. Other carbanion reactions (sulphonium and sulphoxonium ylides, Julia reaction, Darzens reaction, and related transformations).
Chemo-, regio and diastereoselectivity in organic transformations. Selective 1,2-, 1,4- and 1,6-addition reactions and organocuprates chemistry. Diastereoselectivity of the aldol and related reactions (Mannich, Michael, Mukaiyama). Principal stereodeterminging models. Stabilised enolates reactivity (Mukaiyama Silyl enol ethers, Boron-enolates, Tin-enolates). The vinylogy principle in organic synthesis.
Functional groups interconversion. Strategic application of protecting groups. Oxidations (hypervalent iodine reagents, Stahl oxidation, Shi epoxidation, Oppenauer oxidation). Reductions (Birch reduction, Corey-Shibata reduction, Meerwein reduction, modern reagents). Hydroborations.

Basic reactivity of the heterocyclic systems. Construction and reactivity of the principal heterocycles: indole, pyrrole, furan, benzofuran, pyridine and derivatives.
The reaction will be studied and applied to the synthesis of complex organic targets.

Systems Chemistry

Equilibrium self-assembly (non-covalent interactions, multivalency and cooperativity, design principles, surfactants aggregates, self-sorting, supramolecular polymerization, applications in molecular recognition and catalysis, dynamic combinatorial chemistry)
Non-equilibrium self-assembly (introduction to non-equilibrium thermodynamics, characterization of non-equilibrium steady-states: currents and entropy production rate, biological and artificial molecular machines: ratcheting mechanisms, energy consumption, efficiency, thermodynamics and kinetics)
Biological and chemical reaction networks (motifs of enzyme catalysis, signal transduction pathways, feedback mechanisms, oscillating reactions, bistability, self-replication)
DNA nanotechnology (structural DNA nanotechnology: thermodynamics, secondary structure formation, strand displacement reaction, functional DNA nanotechnology: DNAzymes, aptamers, DNA-based devices and materials)
Responsive soft matter (classes: supramolecular polymers, surfactant-based assemblies, hydrogels, polymers, reaction-diffusion systems, concept of physical intelligence, light-responsive materials, actuators, soft-robotics)
Chemical computing (molecular logic gates, network computing, neuromorphic computing, reservoir computing)

III° SEMESTER

	CFU	SSD
Applied Chemistry	**6**	
Optional Course	**6**	
Internship Project	**10**	

IV° SEMESTER

	CFU	SSD
Final research Project	**40**	

Optional Courses

From Chemistry, Industrial Chemistry, Materials Science and Sustainable Chemistry and Technologies for Circular Economy

Internship project

Multidisciplinary individual project co-supervised by two supervisors from different areas. The students will be offered a list of projects pre-proposed by the research groups.