

Padova, September 27, 2023

COMPLEX AND DATA DRIVEN CHEMISTRY

Note introduttive: Motivazioni e Contesto

Il gruppo di lavoro per la didattica del Progetto di Eccellenza 2023-27 del DiSC, composto da Elisabetta Collini (coordinatore), Valerio Di Marco, Camilla Ferrante, Barbara Fresch, Fabrizio Mancin, Antonino Polimeno, e Alessandro Soncini, propone la creazione di un curriculum in inglese per la LM in Chimica.

La creazione del curriculum è indicata nel Progetto di Eccellenza ai punti OD1 (*sviluppo nell'offerta didattica delle tematiche complex e data-driven chemistry*) ed SD1 (*creazione di un percorso formativo (curriculum) su Complex and Data-driven Chemistry in inglese all'interno della LM in Chimica, dedicato allo studio della reattività, sintesi, modellazione di sistemi chimici complessi*), e dettagliato nel quadro D.7 (... *Il percorso formativo conterrà insegnamenti, in inglese, su base fortemente interdisciplinare relativi alla sintesi, caratterizzazione e modellazione di sistemi e materiali complessi. Il chimico ... acquisirà expertise sul trattamento di grandi dati e l'esplorazione dello spazio chimico, e ottimizzazione di processi; progettazione di sintesi organiche e inorganiche mirate; caratterizzazione qualitativa e quantitativa di sistemi chimici; design molecolare di materiali. L'avvio ... è previsto per l'AA 2024/25*).

Il curriculum si pone i seguenti obiettivi didattici/strategici.

- 1) Proporre agli studenti un percorso formativo che li renda in grado di gestire ed utilizzare i nuovi strumenti emergenti in campo scientifico (data science, machine learning, artificial intelligence).
- 2) Incrementare l'attrattività della LM in Chimica erogando contenuti non presenti (o solo marginalmente presenti) nelle attuali LM afferenti al DiSC.
- 3) Attrarre studenti stranieri per aumentare l'internazionalizzazione del DiSC e quindi dell'Ateneo. (Sarà in ogni caso necessario un background chimico per l'iscrizione. Per di più, per tener conto della variegata preparazione di base degli studenti internazionali, saranno attivati anche dei MOOC (*massive open online courses*) di supporto, che gli studenti interessati potranno seguire prima dell'inizio delle lezioni.
- 4) Sperimentare una modalità didattica innovativa, più simile a quelle utilizzate nei paesi occidentali (research master).

Il percorso formativo individuato si propone quindi di introdurre i concetti fondamentali nel campo della "data-driven chemistry" (statistical learning, chemiometry, design of experiments) contestualizzandoli nell'ambito di una formazione magistrale in chimica. A tal fine sono stati individuati alcuni contenuti fondamentali in cui gli approcci "data-driven" trovano maggiore applicazione. Questi riguardano la raccolta di dati chimici, la sintesi e progettazione di materiali e catalizzatori, la progettazione di strategie sintetiche, la modellizzazione e studio di sistemi chimici complessi.

Ciascun ambito "contenutistico" individuato è coperto da due corsi da 6 CFU:

- analisi ed elaborazione di grandi basi di dati chimici (digital learning e DoE/Chemiometry)
- metodi per l'acquisizione di grandi basi di dati chimici e per la caratterizzazione di sistemi complessi (analysis of complex chemical systems and characterization of complex chemical systems)
- modellizzazione razionale e proprietà emergenti di sistemi complessi (multiscale modeling e systems chemistry).

- catalizzatori, materiali catalitici e sintesi multistadio come aree di applicazione dei metodi AI-based (catalysis and synthetic design)

viene inoltre previsto un laboratorio di coding di 4 CFU per fornire le indispensabili competenze fondamentali nell'ambito della programmazione.

Il percorso teorico, che si conclude essenzialmente nel primo anno, viene completato da un percorso "pratico", attraverso l'internship lab e la tesi che coprono quasi tutto il secondo anno.

Il curriculum si differenzia quindi dalla LM in Chimica attualmente esistente non solo per la lingua di erogazione e per i contenuti, ma anche per i seguenti aspetti:

- a) È predisposto all'avvio delle nuove lauree abilitanti, per cui prevede dei contenuti flessibili che possano essere facilmente modificabili quando la riforma abilitante dovrà essere attuata. A tale scopo si attiva un corso da 6 CFU, Applied Chemistry, che nella laurea non abilitante conterrà competenze di soft skills (etica, presentazione e comunicazione della scienza).
- b) Limita la scelta degli esami opzionali a due, anziché a sei come nella LM esistente (dal momento che tali opzionali ricadono nella categoria "a scelta dello studente", non è necessaria un'offerta in inglese, sono comunque presenti alcuni corsi in inglese nelle lauree magistrali in chimica, chimica industriale e scienza dei materiali e si auspica che tale offerta possa aumentare in futuro).
- c) Limita le ore di laboratorio didattico, per gravare solo in minima misura sui laboratori didattici del DiSC attualmente quasi saturi e permettere l'introduzione dell' "internship project" (punto d).
- d) Introduce un laboratorio "Internship project", cioè un piccolo progetto (10 CFU) da svolgere in un laboratorio di ricerca oppure come stage esterno. Idealmente, tale progetto prenderà la collaborazione tra più gruppi di ricerca per garantire un'ampia esposizione a diverse pratiche di laboratorio e metodi di caratterizzazione.
- e) tutti i corsi sono da 6 CFU in modo da permettere lo "scambio" bidirezionale di esami opzionali tra il curriculum e le attuali LM in Chimica, Chimica Industriale, Scienza dei Materiali e Science and Sustainable Chemistry and Technologies for Circular Economy.

I laureati del curriculum dovrebbero trovare i seguenti sbocchi lavorativi:

- sviluppo di catalizzatori (e materiali) innovative
- progettazione di strategie sintetiche complesse
- analisi di sistemi complessi (omiche, ambiente, alimenti, ecc.)
- sviluppo di formulazioni e ottimizzazione di processi
- sviluppo di sistemi chimici e biochimici complessi
- sviluppo di sistemi di "predizione chimica" (drug design, structural biochemistry).

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

Parte 1 (4 CFU)

Definizione di catalizzatore; il concetto di controllo cinetico; il concetto di sito attivo; il ciclo catalitico; turn-over number; turn-over frequency; attività e selettività catalitiche; funzionalità; efficienza atomica. Le reazioni catalitiche omogenee ed eterogenee: meccanismi ideali; grado di ricoprimento in catalisi eterogenea; adsorbimento e coadsorbimento, molecolari e dissociativi; isoterme di Langmuir; derivazione delle leggi cinetiche; cinetica di Michaelis-Menten; variabilità dell'ordine di reazione nelle reazioni complesse e relazione con l'energia di attivazione apparente. Limitazioni diffusive alla velocità dei processi catalitici eterogenei: efficienza catalitica; modulo di Thiele. Catalisi omogenea acido-base: catalisi acida e basica specifiche e generali; reazioni di idrolisi, condensazione e alchilazione. Principi di chimica metallorganica applicata dei metalli di transizione: donazione- σ e retrodonazione- π ; legame metallo-CO e metallo-alchene; la regola dei 16-18 elettroni; leganti tipici. Coordinazione e dissociazione di leganti; addizione ossidativa, eliminazione riduttiva; inserzione migratoria, α - e β - eliminazione. Processi catalitici omogenei industriali: idroformilazione degli alcheni; carbonilazione del metanolo ad acido acetico: processi Monsanto e Cativa; oligomerizzazione e polimerizzazione stereospecifica di alcheni con catalizzatori “single-site” e Ziegler-Natta; metatesi di olefine; reazioni di cross-coupling. Catalizzatori eterogenei industriali non supportati e supportati: componenti tipici; ruolo del supporto; metodi di preparazione. Misura dell'area superficiale dei solidi mediante metodi di fisisorbimento all'equilibrio; isoterme reali; isoterma e metodo BET; determinazione dei siti sulla superficie attiva mediante metodi di chemisorbimento e “temperature programmed desorption”. Porosità, tortuosità e

rugosità nei solidi; solidi micro-, meso-, e macro-porosi; misura delle dimensioni dei pori (diametro, volume) e loro relazione con l'area superficiale specifica. {Sintesi, struttura e proprietà di zeoliti, silice mesoporosa e MOF (*possibile spostamento nella parte 2*)}. Energia superficiale dei solidi e attività catalitica; frazione esposta del componente attivo e sua relazione con l'attività catalitica; struttura del reticolo cristallino e della superficie dei metalli; il modello "broken-bond" per l'energia superficiale; costruzione di Wulff; difetti superficiali; siti superficiali. Vita utile e disattivazione dei catalizzatori eterogenei. Meccanismo della sintesi industriale dell'ammoniaca; "structure sensitivity"; promotori elettronici. Teoria delle bande e chemisorbimento sulla superficie dei metalli; principio di Sabatier; diagrammi a vulcano. Marmite catalitiche: generalità; catalizzatori a "tre vie"; meccanismo dell'ossidazione di CO e della riduzione di NO_x; ricostruzione della superficie attiva nella ossidazione di CO. Ossidazione dell'etilene a ossido di etilene. Chemisorbimento su ossidi semiconduttori e isolanti. Solidi acidi: punto isoelettrico; funzione di acidità di Hammett; superacidità; struttura dei siti acidi e loro forza in solidi tipici; distinzione fra siti acidi di Brønsted-Lowry e di Lewis; attivazione di alcani e alcheni; ioni carbonio e carbenio. Catalisi bifunzionale: reforming catalitico; "selective catalytic reduction". Nuovi orientamenti in catalisi: single atom catalysis, fotocatalisi, elettrocatalisi etc. *Possibile inserimento di cenni di biocatalisi*

Parte 2 (2 CFU, proposta sulla base di contenuti di chimica inorganica 3)

Sintesi di materiali catalitici e supporti per catalisi eterogenea. Richiami sulla descrizione e struttura dei materiali inorganici: cristallini, amorfi, polimerici e nanodimensionali, relazione tra tipo di legame chimico e struttura, difettualità nei solidi (con particolare riferimento ai difetti puntuali). Metodi di sintesi da stato solido (meccanosintesi, riduzione carbotermica, sintesi da combustione, metodo ceramico). Formazione di solidi da fase gas (processo aerosol, spray pyrolysis, CVD). Formazione di solidi da fase liquida: nucleazione e crescita, modello di La Mer, teorie classiche e non classiche di cristallizzazione, regola delle fasi di Ostwald; tecniche sol-gel; metodi di dispersione e metodi di nucleazione/crescita. Nucleazione da soluzione e seeded growth. Sintesi idro- e solvotermale (subcritiche e supercritiche). Sintesi assistita da polioili. Sintesi assistita da microonde. Sintesi assistita da laser e metodi sonochimici e radiochimici. Sintesi in flusso e microfluidica. Sintesi in spazio confinato: microemulsioni, miniemulsioni. Sintesi biogeniche. Approcci al controllo morfologico e dimensionale: anisotropia.

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.

Multivariate 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 sulfoxonium 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 stereodetermining 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.