

“Talktorials” in computational chemistry and data science

S. Christodoulou, R. Poteau



Laboratoire
de Physique & Chimie
des Nano-Objets



Motivation

integration of verbal explanations with numerical demonstrations or computer algebra system-based demonstrations proves to be an influential pedagogical tool

let's call them "talktorials"

(portmanteau of "talks" and "tutorials")



such mixing is not new, but until recently it was restricted to rather simple applications that required a great deal of development effort

What is new is:

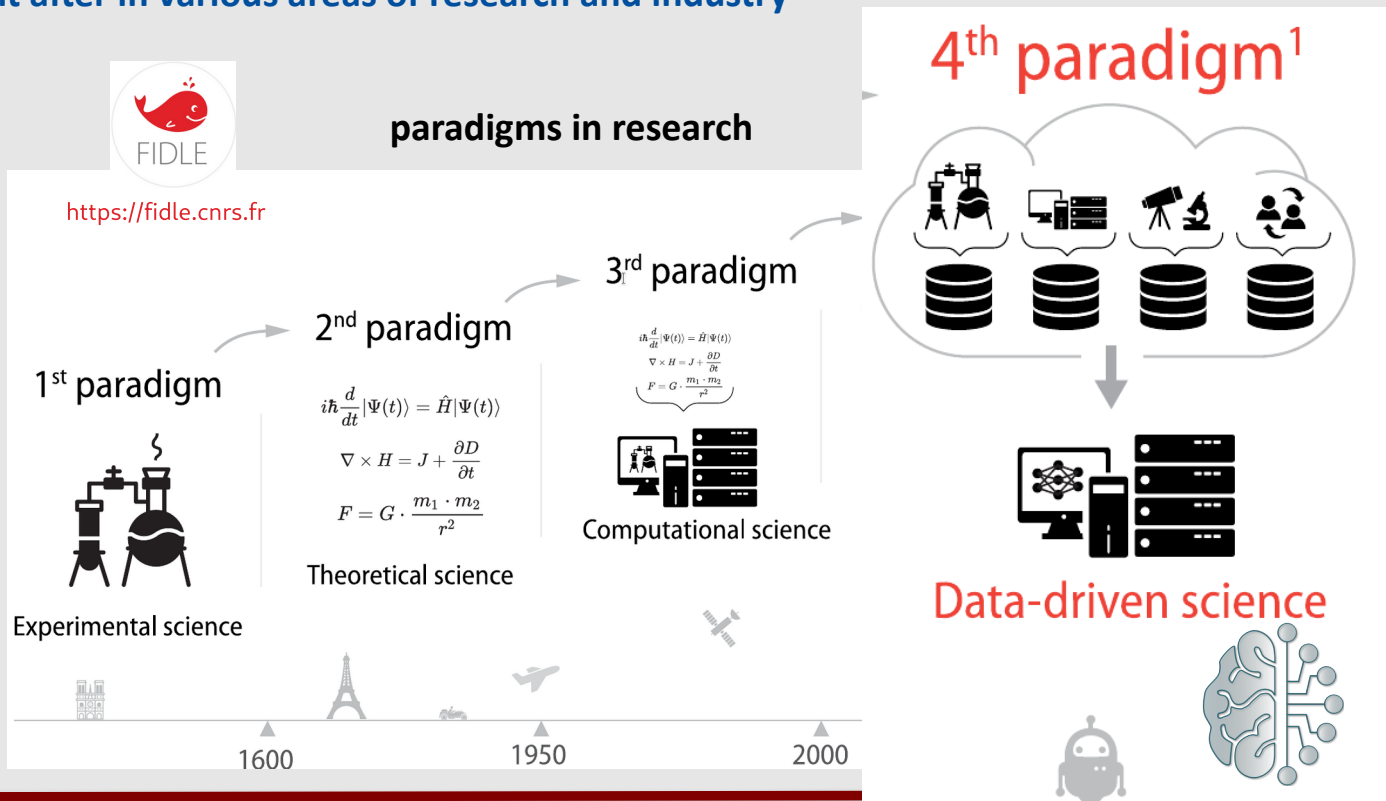
- the combination of Python's popularity and libraries
- the interactive nature of Jupyter - jupyterlab - notebooks
- personal computers performance
- the possibility to easily adapt real-world applications thanks to Python libraries available in a lot of domains
- the strong community support
- the ease of reproducibility that makes tutorials more effective, as learners can directly use the code provided to experiment and build upon it

Motivation

The emergence of such innovative approaches in the realm of computational chemistry is truly encouraging

It not only enables learners to grasp theoretical concepts but also offers a practical perspective on their application

For students specialising in computational chemistry who develop their own scripts, they acquire a **dual skill set that could be sought after in various areas of research and industry**



Github repository

Python in the Physical Chemistry lab (pyPhysChem) github repository, release v. 1.8.0 (2023)

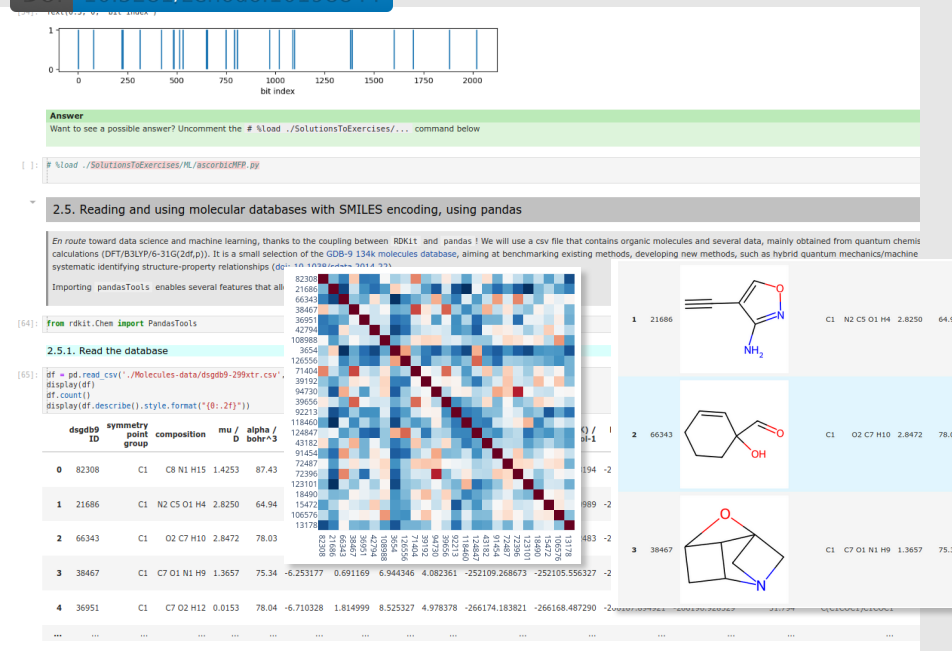
Jupyterlab
notebooks/talktorials-ready

specifically tailored for
computational chemistry
and data science/machine
learning



- interactive python
- images / videos
- mathematical equations
- enriched text (markdown)
- links > data
- embed pdf documents

DOI 10.5281/zenodo.10198844



Which students (so far...)?

bachelor degree

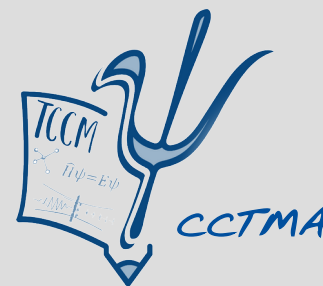
all chemistry students, 2nd year (120 students)



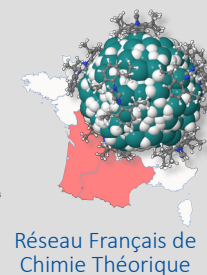
masters' (graduate) degrees



International students of the Nanoscale Science and Engineering master's degree, @Toulouse, France



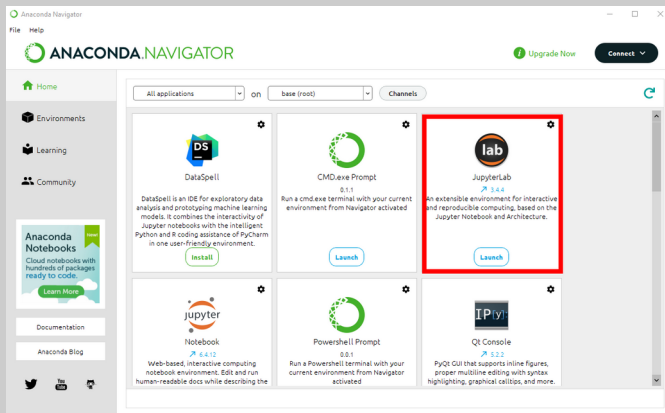
European and French students in Computational Chemistry



Educational scenario: introduction to data science and ANNs in chemistry

Technical prerequisite

Students can install on their computer a python distribution, such as anaconda



educational activity

or

They must install a ready-to-use Linux virtual machine



VirtualBox

less educational...



Remedial help

an online install-party is organized afterwards

Educational scenario: introduction to data science and ANNs in chemistry

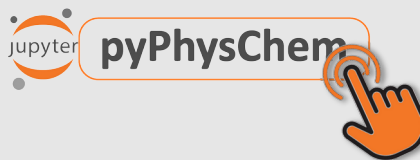
Prerequisite: a basic knowledge in python

Most of the students have already been exposed to python

A python in the physical chemistry lab in a nutshell notebook is part of pyPhysChem

→ revising python within a Phys. Chem. context

→ students become familiar with jupyterlab notebooks and with the spirit in which pyPhysChem notebooks are developed and can be used



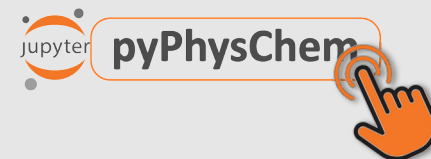
Educational scenario: introduction to data science and ANNs in chemistry

Talktorial

Basics of Artificial Neural Networks (ANNs) for supervised learning



1 jupyterlab notebook



Self-learning with short questions



4 jupyterlab notebooks

Reading and analysis of the "iris" database with the pandas library

Statistics and regression

Supervised Machine Learning applied to classification

eXplainable AI (XAI)



Remedial help

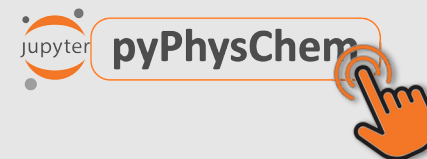
Educational scenario: introduction to data science and ANNs in chemistry

Talktorial

Molecular representations



1 jupyterlab notebook



Self-learning and exercises



2 jupyterlab notebooks

Self-learning: prediction by an artificial neural network of the solubility of CO₂ in ionic liquids

after “Z. Song, H. Shi, X. Zhang & T. Zhou (2020), Prediction of CO₂ solubility in ionic liquids using machine learning methods, *Chem. Eng. Sci.* **223**: 115752”

Exercise: Assessment of the stability of the original ML algorithm of Song *et al.* by K-fold cross validation

solution is available on pyPhysChem → self-assessment and critical discussion about this study



Remedial help

Project-based learning and assessment

TCCM Homework: Data-Driven and ML Color Analysis of Portuguese Wines



Skills tested

- data analysis and curation
- building and training of a DNN
- holdout method
- hyperparameters optimization
- K-fold cross validation
- Shapley Additive exPlanations
- critical thinking

} = ML workflow

students must submit their python code and report in the form of a jupyterlab notebook

Summary & Outlook

<https://github.com/rpoteau/pyPhysChem>

DOI: [10.5281/zenodo.10198844](https://doi.org/10.5281/zenodo.10198844)



Python in the
Physical Chemistry Lab



[pyPhysChem]



- python for physicists and chemists in a nutshell
- Computer Algebra System
- Physical chemistry, including quantum chemistry
- coding and use of representations of molecular structures and related data
- Data science and ML (supervised and unsupervised learning, mainly NN)

live talktorials

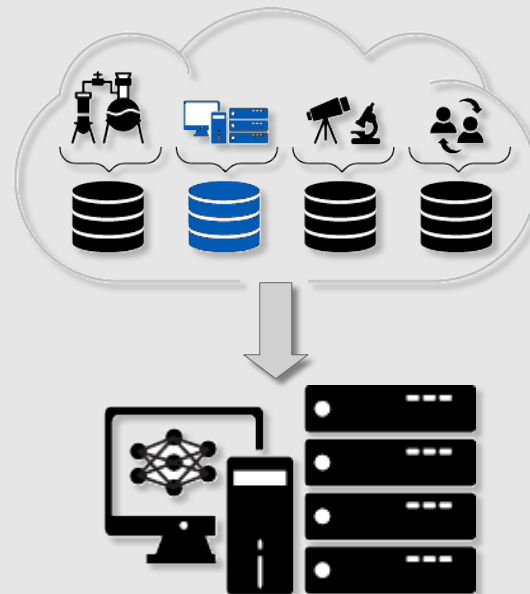
self-learning

project-based learning

[online]remediation help

Very important to give **dual skill set** to Comput. Chem. Students

→ strong convergence between **quantum and computational chemistry**, data science and machine learning



data-driven science

→ could be sought after in various areas of research and industry

AI = black boxes?



Yes and **NO!**... better explainability of AI models than usually supposed

