Data Visualization 201

Advanced concepts using OpenSource software

A. Tavant

LPP- April 3, 2018

Introduction

A picture speaks a thousand words

Napoléon 1er

Outline

- Introduction : Why plotting ?
- Guidelines to plot **WELL**
- Tutorial: using matplotlib

Introduction : Why plotting ?

As scientists, we plot in order to:

- Visualize raw data (from experiment or simulation)
- Present results to our supervisor/colleges
- Present results at conference/ for thesis/ paper

Introduction : Why plotting ?

As scientists, we plot in order to:

- Visualize raw data (from experiment or simulation)
  - **systematic** (once/twice a day)
  - need to be **efficient**
- Present results to our supervisor/colleges
- Present results at conference/ for thesis/ paper

Introduction : Why plotting ?

As scientists, we plot in order to:

- Visualize raw data (from experiment or simulation)
  - **systematic** (once/twice a day)
  - need to be **efficient**
- Present results to our supervisor/colleges
  - less systematic ($\sim$ once a week)
  - clear and consistent (easy to understand)
- Present results at conference/ for thesis/ paper

Introduction : Why plotting ?

As scientists, we plot in order to:

- Visualize raw data (from experiment or simulation)
  - **systematic** (once/twice a day)
  - need to be **efficient**
- Present results to our supervisor/colleges
  - less systematic ($\sim$ once a week)
  - clear and consistent (easy to understand)
- Present results at conference/ for thesis/ paper
  - very well done -> **many corrections**
  - Easy to understand by a general public

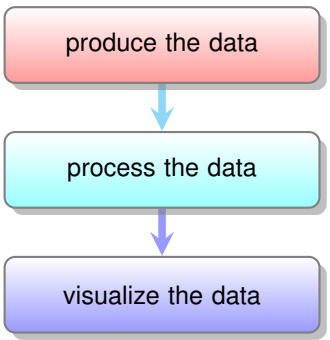
Introduction : Why plotting ?

As scientists, we plot in order to:

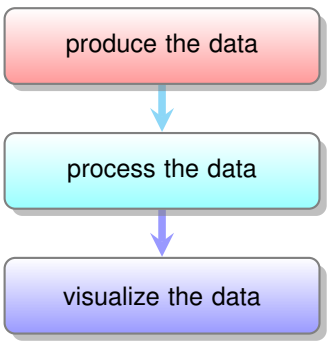
- Visualize raw data (from experiment or simulation)
  - **systematic** (once/twice a day)
  - need to be **efficient**
- Present results to our supervisor/colleges
  - less systematic ($\sim$ once a week)
  - clear and consistent (easy to understand)
- Present results at conference/ for thesis/ paper
  - very well done -> **many corrections**
  - Easy to understand by a general public

Hence: having good tools and work-flow is mandatory to be *efficient*

Introduction : Why plotting ?



Introduction : Why plotting ?



- Produce the data
  - Experiment/ Simulmation
  - Takes a lot of time
- process the data
  - use whatever tool you prefer (python, matlab, excel...)
  - stock and keep both the raw and the processed data
- Visualize
  - Use whatever tool you prefer
  - not necessary the same tool as for processing

Introduction : Why plotting ?

My own tools

There is My own tools and work-flow:

- Produce data via simulation : takes **2-5 days** each
- store data in hierarchic format (**HDF5**, for management of extremely large and complex data collections).
- **Python** for processing
- **Matplotlib** (python library) for plotting
  - **homemade** library for systematic plots
  - generates $\sim$**25 plots and** $\sim$**10 movies** automatically for each plots
  - dedicated **scripts** for presentation/paper/reports

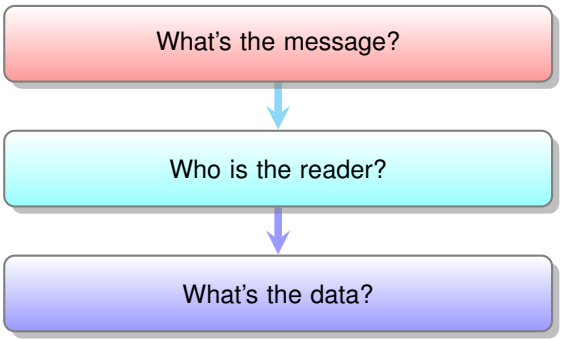
Guidelines to plot WELL

Plotting nice figures **is not magic**.

There is **some recipes** to help you.

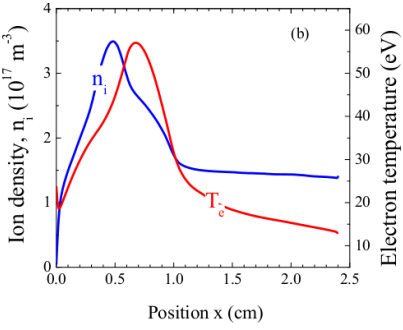
Guidelines to plot WELL

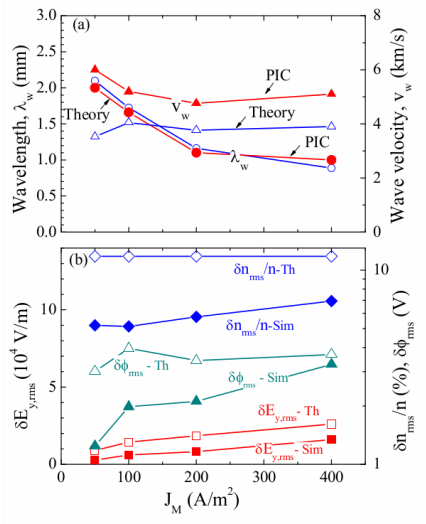
Questions to ask yourself before even starting:



Quiz time

Quiz time : Which figure is the best ?


(b)
n_i
T_e
Ion density, n_i (10^{17} m^{-3})
Electron temperature (eV)
Position x (cm)


(a)
PIC
Theory
V_w
Theory
PIC
λ_w
Wavelength, λ_w (mm)
Wave velocity, v_w (km/s)
(b)
$\delta n_{rms}/n$-Th
$\delta n_{rms}/n$-Sim
$\delta\phi_{rms}$ - Th
$\delta\phi_{rms}$ - Sim
$\delta E_{y,rms}$ - Th
$\delta E_{y,rms}$ - Sim
$\delta E_{y,rms}$ (10^4 V/m)
$\delta n_{rms}/n$ (%), $\delta\phi_{rms}$ (V)
J_M (A/m^2)

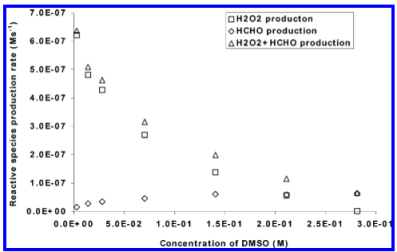


Figure 8. Production rate of stable molecular products from hydroxyl radical reactions as a function of DMSO concentration.

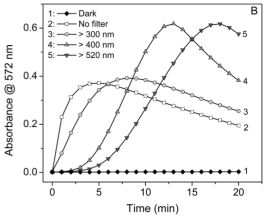
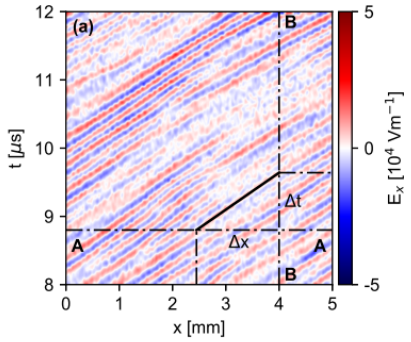
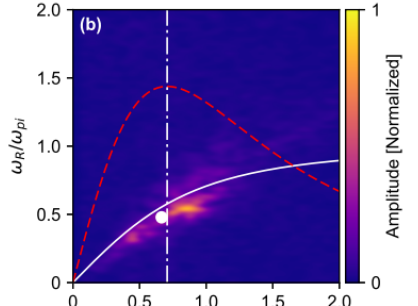
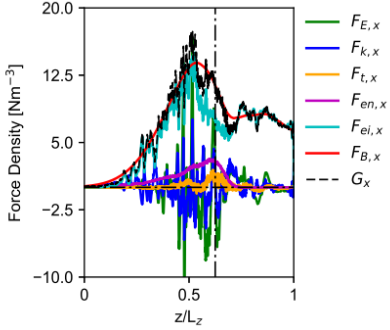
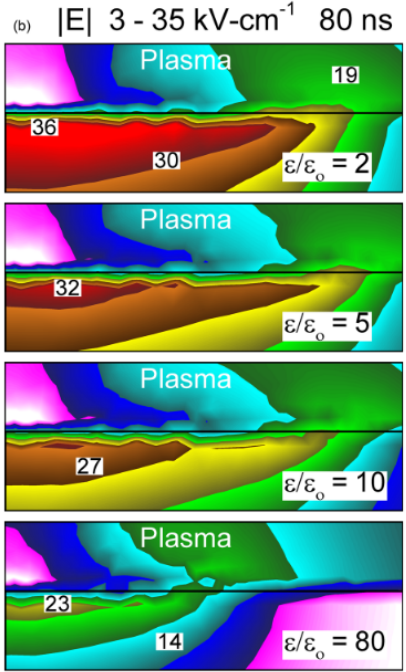


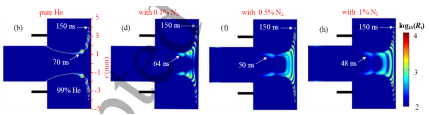
Figure 1.
Photo-induced oxidation of Amplex Red with production of resorufin during irradiation as measured by UV-vis absorption. (A) Absorption spectra of 10 μM AR solution at pH 7.4 after irradiation (> 520 nm) for every minute from 0 to 18 min. The arrow indicates direction of the changes. (B) Plot of peak absorbance of RSF at 572 nm against irradiation time (1, dark; 2, no filter; 3, > 300 nm; 4, > 400 nm, and 5, > 520 nm). The fluence rate (from 250 to 850 nm) was 0, 0.392, 0.329, 0.322, and 0.271 W/cm^2 as measured with a SPR-4001 Spectroradiometer (Luzchem Research Inc., Ottawa Ontario, Canada).


(a)
B
A
A
B
Δt
Δx
t [μs]
x [mm]
E_x [10^4 Vm^-1]

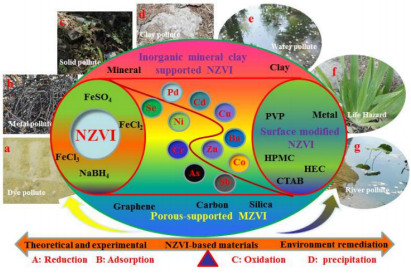

(b)
ω_R/ω_pi
Amplitude [Normalized]


Force Density [Nm^-3]
z/L_z
F_E,x
F_k,x
F_t,x
F_en,x
F_ei,x
F_B,x
G_x

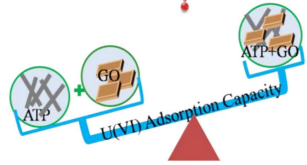

(b) |E| 3 - 35 kV-cm-1 80 ns
Plasma
19
36
30
ε/ε_o = 2
Plasma
32
ε/ε_o = 5
Plasma
27
ε/ε_o = 10
Plasma
23
14
ε/ε_o = 80


pure He
with 0.5% N2
with 1% N2
(b)
150 ns
70 ns
99% He
(d)
150 ns
64 ns
(f)
150 ns
50 ns
(h)
150 ns
48 ns
r (mm)
log10(Re)
4
3
2

TOC



Two kinds of single adsorbents ? Heterogeneous aggregate



Guidelines to plot WELL

What is a good figure; a good plot?

Simple to understand

Enjoyable to look at

Guidelines to plot WELL

What is a good figure; a good plot?

Simple to understand

- simple, not too much information
- readable: font size and family, quality
- Good choice of markers, color, etc.

Enjoyable to look at

Guidelines to plot WELL

What is a good figure; a good plot?

Simple to understand

- simple, not too much information
- readable: font size and family, quality
- Good choice of markers, color, etc.

Enjoyable to look at

- simple, not too much information
- readable: font size and family, quality
- Good choice of markers, color, etc.

Guidelines to plot WELL

What is a good figure; a good plot?

Simple to understand

- simple, not too much information
- readable: font size and family, quality
- Good choice of markers, color, etc.

Enjoyable to look at

- simple, not too much information
- readable: font size and family, quality
- Good choice of markers, color, etc.

an **Enjoyable** *figure will convey* ***more information*** *!*

Some Rules, not absolute and not exhaustive:[1][2][3]

- 2-3 lines max if complex graph
- Think of the black-and-white prints
- Consistent representation
- Adapt to the Support Medium
- Be concise (save ink)
- Be efficient (script, optimize)
- Check of typos and consistency
- Do Not Trust the Defaults
- large font size
- Uses a good the font
- Use vectoriel format (*.eps, *.pdf, *svg)
- Use Color Effectively
- Label axis and use units

[1] Ten Simple Rules for Better Figures, journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

[2] Graphical Excellence in Scientific Presentations and Papers, www3.nd.edu/ pkamat/pdf/graphs.pdf

[3] personal opinion and experience

Live example of matplotlib tips and tricks

Never do a live example

everyone