

Exercice reproductibilité

Installation

Première partie: les logiciels courants

- Python (2 ou 3)
- NumPy
- h5py
- matplotlib

La façon la plus simple est [Anaconda](#) qui contient déjà tout ça. Pour ceux qui sont sous Debian/Ubuntu, faites

```
apt-get install python-h5py
```

Deuxième partie: les logiciels plus exotiques

```
pip install tempdir ActivePapers.Py
```

Télécharger le code et les données

Un seul [fichier HDF5](#) contient tout.

Mettez-le dans un repertoire fraîchement créé:

```
mkdir temp
cd temp
cp $HOME/Downloads/lysozyme_diffusion.ap ,
```

Extraire les figures d'origine et en garder une copie

```
aptool checkout documentation
mv documentation figures_origine
```

Recalculer les résultats

C'est particulièrement facile parce que dans ce fichier, des données intermediaires ont été supprimées pour reduire le poids du fichier publié. Il suffit de les reconstituer, ce qui déclenche un recalcul complet:

```
aptool update -v
```

Le `-v` affiche chaque étape dans le terminal. Sur un MacBookPro de 2016, le calcul prend 5 minutes.

Extraire les nouvelles figures et comparer

```
aptool checkout documentation
```

Puis comparer visuellement les PDFs dans `documentation` et dans `figures_origine`.

Comparer les données binaires

```
h5diff $HOME/Downloads/lysozyme_diffusion.ap lysozyme_diffusion.ap | more
```

La majorité des différences affichées concernent les métadonnées du type `ACTIVE_PAPER_TIMESTAMP`. Ce sont les autres différences qu'il faut réparer. Par exemple:

```
dataset: </data/correlation_functions/mean_square_displacement_molecular_frame_1> and </data/correlation_functions/mean_square_displacement_molecular_frame_1>  
23 differences found
```

Pour regarder les différences numériques:

```
h5diff -v $HOME/Downloads/lysozyme_diffusion.ap lysozyme_diffusion.ap /data/correlation_functions/mean_square_displacement_molecular_frame_1
```

Je vous laisse juger de la pertinence de ces différences. Est-ce simplement les différences d'arrondis cumulés?

Ou y a-t-il de quoi s'inquiéter? Pour voir la totalité de ce jeu de données:

```
h5dump -d /data/correlation_functions/mean_square_displacement_molecular_frame_1 lysozyme_diffusion.ap
```

Pour en savoir plus

Ce [tutoriel](#) vous propose une exploration plus approfondie des données que vous venez d'utiliser.