A python library to manage ADCP big data

Starting joint work with Intechmer

François-Xavier Jollois

May 6th, 2021

diNo – LIPADE – Université de Paris

Résumé

Acoustic Doppler Current Profiler (ADCP) are devices initially developed to measure vertical water currents. The measurement rates still very recently below 2 Hz now reach 16 Hz, or even 64 Hz for the next generations. In addition, these devices now allow simultaneous measurements at several points of the water level. These technological advances involve, on the one hand, increasingly large volumes of digital data. On the other hand, the range of conditions of use of the devices is widened. The software tools available for processing and analyzing ADCP data, for the most part (whether proprietary or not), have not taken into account having to manage large volumes of data. To allow management and analysis of the large volumes of data generated by the most recent ADCP type measuring devices, we are proposing to develop a new python calculation library.

Outline

What is an ADCP ?

For which purpose ?

What is complicated to handle these data ?

What we want to propose ?

Our work (in progress)

Conclusion

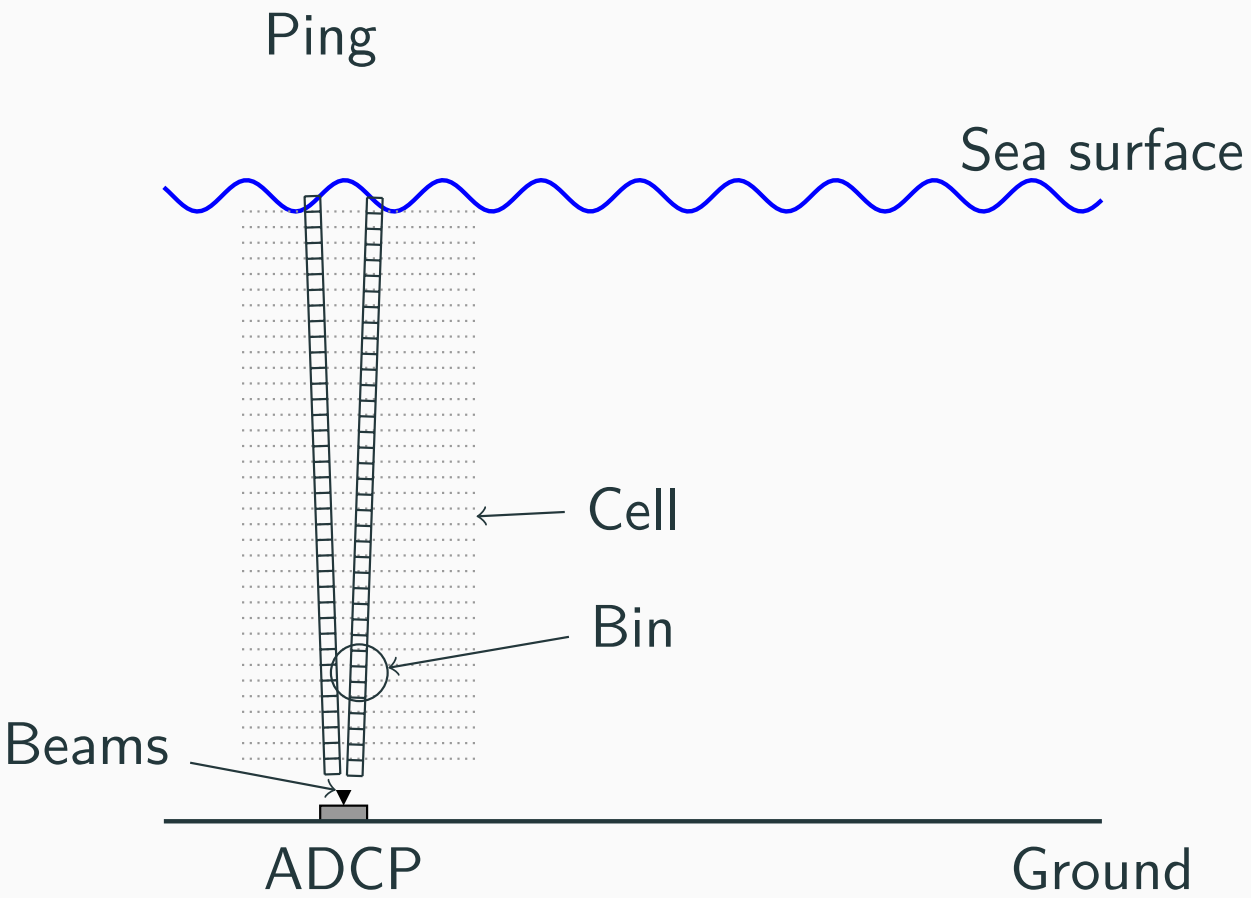
What is an ADCP ?

What is an ADCP ?

Definition

Acoustic Doppler Current Profiler are devices initially developed to measure vertical water currents velocities, based on *Doppler effect*

How ADCP works ?



What are the data collected ?

Global data

ADCP type (especially number of beams), Position (on the ground or under a boat), Number of cells (and size), Blancking (device near blind area)

Ping specific data

Timestamp, ADCP rotation (on 3 axes), ENU position (if moving), Temperature, Salinity, Speed of sound, Depth

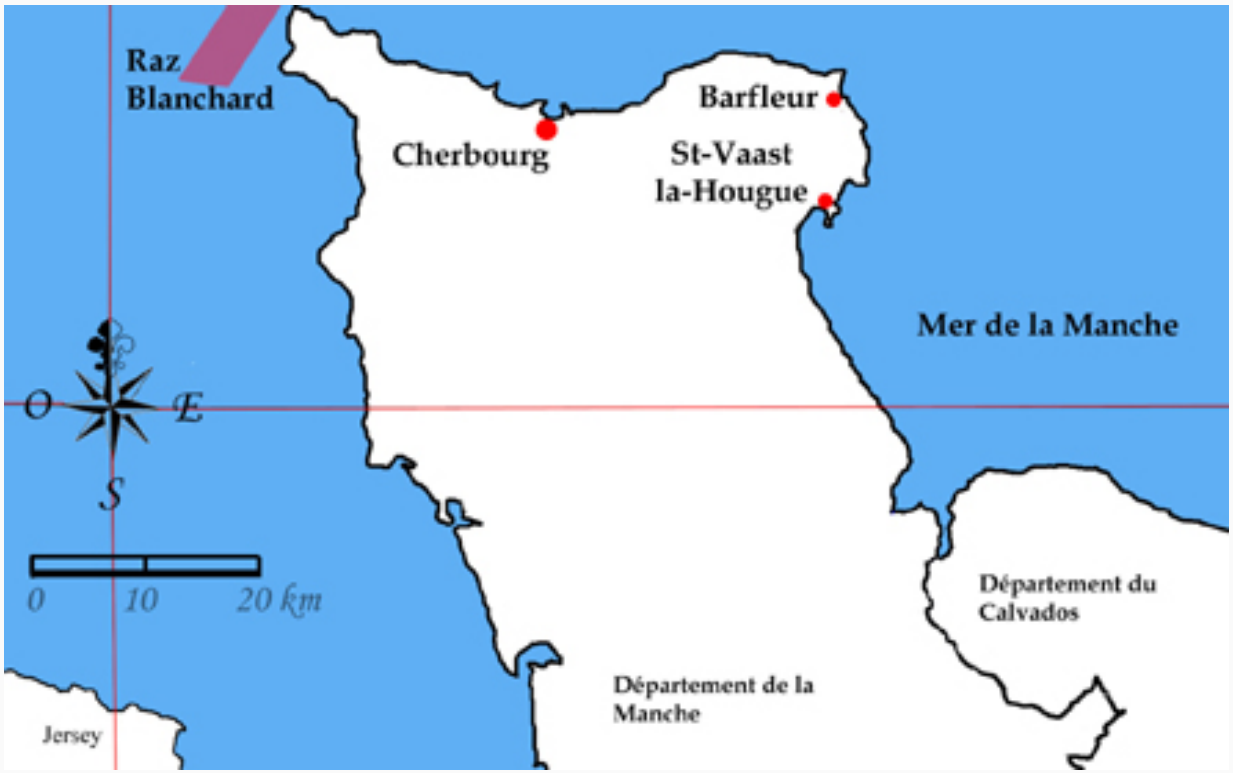
Bin specifica data

Velocity, correlation (quality of measure), intensity (of measure)

For which purpose ?

For which purpose ?

- Estimation of the water current
- Development of Marine Renewable Energy (i.e. tidal turbine)
- Alderney Race (*Raz Blanchard*) [1]
  - One of the most powerful current in Europe
  - Current velocity (arrows) up to 5 m/s (12 knots)



1. http://www.wikimanche.fr/Raz_Blanchard

What is complicated to handle these data ?

What is complicated to handle these data ?

- Raw binary files
- Data quality should be estimated
- Possible huge amount of data
- Data (after transformations) are vectors : velocity and direction

Raw binary files

Raw binary files

- Files are pure binary that we need to decode
- Each ping is defined by a specific starting and ending binary sequence
- Each manufacturer has its own binary format

Data quality

- Measurements are subject to defects
- Shoaling fish, water infiltration, lost device...

How huge are the data collected ?

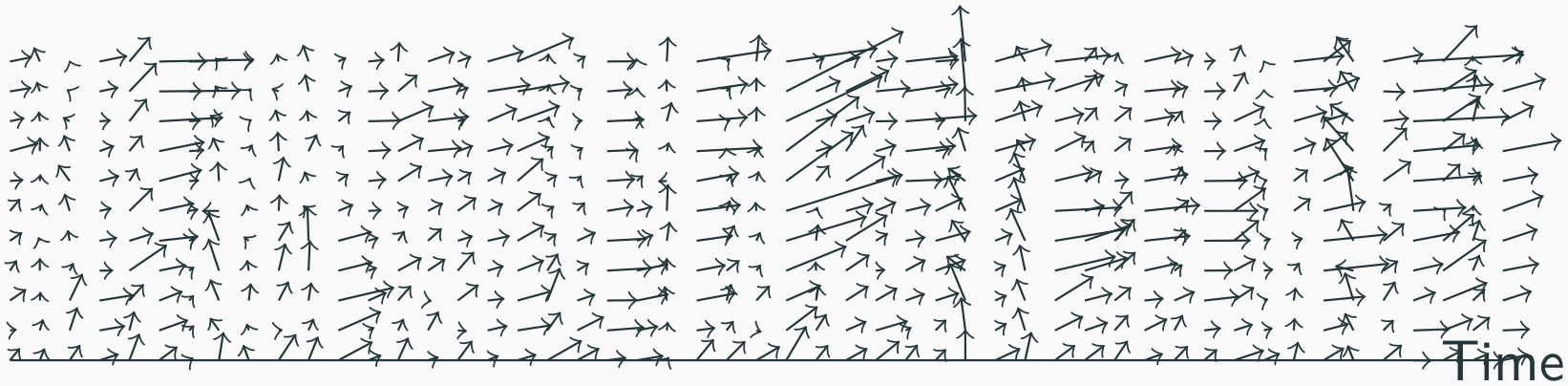
- Time series : data are collected during a given period (from some hours to several weeks or months)
- Frequency of ping : 2Hz to 16Hz, even more in the next years
- Multiple beams : 3 or 4, even more in the next years
- Multiple cells : usual size of cells is 1 meter (even smaller in the next years), so it can be 30 to 40 cells in our case

For instance

2 weeks at 8Hz with 5 beams on 40 cells $\rightarrow$ 1.9 G bins

Data are vectors

- For each ping, for each cell,



Data need to be reconstructed

- Velocity in a bin : numerical value indicating if the current is going to the beam cell (positive) or not (negative) $\rightarrow$ vertically measure
- Combination of values in the different beams to get the current measure in each cell
- Device could rotate due to the current

What we want to propose ?

What we want to propose ?

A python library to manage ADCP files, using parallel processing

- Binary files transformation into NetCDF files
- Coordinates transformation (if needed)
- Compute classical values on NetCDF files
- Visualisation of NetCDF files
- Analysis such as
  - Detection of optimal sliding window for current average
  - Detection of current turbulences

Our work (in progress)

Our work (in progress)

Conclusion

Conclusion

content