

# Internship report on “Automation of interpolating large data using PyDIVA” at INCOIS, Hyd

Rohan Bera  
IMtech, CSE  
University Of Hyderabad

June 18, 2019

# Contents

<b>1</b>	<b>Acknowledgement</b>	<b>3</b>
<b>2</b>	<b>Executive Summary</b>	<b>3</b>
<b>3</b>	<b>Overview of internship activities</b>	<b>4</b>
3.1	Real time to Delayed mode . . . . .	4
3.2	Data Intepolation . . . . .	4
3.3	DIVA . . . . .	5
3.3.1	Installation . . . . .	5
3.3.2	Programs . . . . .	5
<b>4</b>	<b>Conclusion</b>	<b>7</b>
4.1	Learning . . . . .	7
4.2	Challenges Faced . . . . .	7

# 1 Acknowledgement

I, Rohan Bera (17MCME21), pursuing Integrated MTech from University Of Hyderabad, would like to convey my gratitude to Dr. Uday Bhaskar for giving me the opportunity to do a 6 week summer internship at INCOIS, Hyderabad and giving me the platform to interact with the industry professionals.

I would also like to thank Pavan Kumar Sir for guiding me throughout the process.

I extend my warm gratitude and regards to everyone who helped me during my internship.

# 2 Executive Summary

The 6 week Summer Internship Program at Indian National Centre for Ocean Information Services (INCOIS, Hyderabad) is a vital part of the 5-year Integrated MTech Course. Since majority of the students come without any prior work experience, the Summer Internship adds worth to their CVs by giving each student immense learning. At INCOIS, a student can bag an internship through various means since the institute gives us ample opportunities to interact with industry experts. I got my internship by mailing Nagaraja Kumar sir, who then forwarded my request to Dr. Francis Pavanathara. I then worked under the guidance of Dr. Uday Bhaskar and Pavan Kumar sir on the topic “Automation of interpolating large data using PyDIVA”.

## 3 Overview of internship activities

### 3.1 Real time to Delayed mode

1. I was given 2 files (ARGO\_Observations and ARGO\_Global). I had to check for lines in the file which changed modes from 'R' to 'D' and return them in an output file. Furthermore, if there were unique ID's in the ARGO\_Global file, but not in the ARGO\_Observations file, then those were to be returned too.
2. I initially began to code in Java but soon shifted to Python.
3. The Program (update.py) takes in 3 filenames as command line arguments (ARGO\_Observations, ARGO\_Global and the output filename). The unique ID's which met the conditions are returned in the output file, with no repetitions.

### 3.2 Data Intepolation

1. I was given a file containing the unique float ID's (ID\_date) and data (pressure, temperature, salinity) over a year. The temperature and salinity values were distributed over a range of pressure values. I had to interpolate these temperature and pressure values over this range, and extract the temperature and pressure values for a given finite set of pressure values.
2. These interpolated values then were to be stored in files based on the date, i.e. the main directory (year) should contain 12 directories (months). Each month should be divided into 3 directories (1-10, 11-20 & 21-last). The files then, were to be stored in these folders.
3. The Program (interpolFinal.py) takes in 3 command line arguments (ARGO\_Core, ARGO\_Observations and the output path). The output files contains 4 coloumns, they being longitude, latitude, temperature and salinity.

## 3.3 DIVA

### 3.3.1 Installation

- DivaPythonTools : <https://github.com/gher-ulg/DivaPythonTools>
- DIVA : <https://github.com/gher-ulg/diva>

Additional requirements :

1. Gfortran
2. NetCDF library for fortran
3. gnuplot (optional)
4. geos
5. geojson
6. basemap

### 3.3.2 Programs

#### 1. **data1.py**

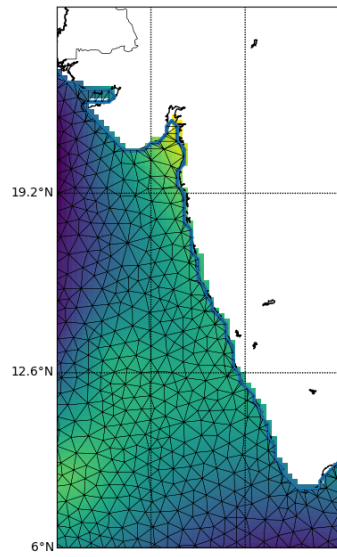
Takes an MLD file, coast.cont file and param.par file as command line arguments and gives the plots for temperature and salinity values (shown in Fig (1a) and Fig (1b)) while storing the .nc (temp.nc and sal.nc) files in an output directory. The directory is created if it didn't exist before.

**Example :** `python3 data1.py example/dp10mts.dat example/coast.cont example/param.par`

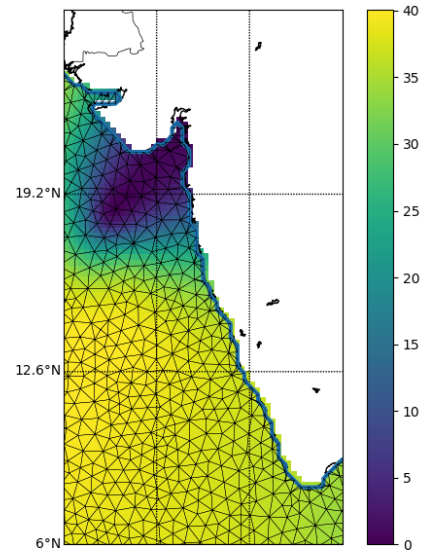
#### 2. **fastdiva.py**

Takes an MLD file, coast.cont file and param.par file as command line arguments and stores .nc (temp.nc and sal.nc) files in an output directory. The directory is created if it didn't exist before. This Program saves time as additional python plots are not created.

**Example :** `python3 fastdiva.py example/dp10mts.dat example/coast.cont example/param.par`



(a) temperature analysis at 10m depth



(b) Salinity analysis at 10m depth

Figure 1: Temperature and Salinity analysis using data1.py

### 3. analyse.py

Takes the following as command line arguments : <dir path containing years> <yyyymmdd> <invalid str> <path to input file>.

The input file should have coast.cont.\* files, coast.depth file and param.par file.

This program creates and stores the .nc (temp.nc and sal.nc) files for all the depths over a 10 days period in an output directory. The output directory is created in the directory in which the input dir exists. This program runs on fastdiva.py.

**Example :** python3 analyse.py example/. 20181211 9999.000 example.py

#### NOTE:

The programs above are dependant on the path to DIVA-4.7.2 directory. Modify the path (divadir) in the source code accordingly.

## 4 Conclusion

### 4.1 Learning

The 6 week Summer Internship exposed me to various aspects of Data Management. I spent the initial week learning about my branch of work and making myself familiar with programming languages that could help me with my work. For the first few days I worked using Java and then shifted to Python.

In my third week, I had to install and work using PyDIVA, an interface to run DIVA software tool for spacial interpolation.

I also learned about .nc files, how to view the netCDF datasets using ncdump and graphically visualize them using ncview.

### 4.2 Challenges Faced

- Since I was only familiar with C, coding took a lot of time and hard-work. Soon I became familiar with Python, thus making my codes simpler and much faster.

- Installing DIVA was a challenge which alone took around 5 days to solve. I managed to solve this problem with the help of Pavan Kumar Sir.
- Understanding how to make use of Diva Python Tools was another issue. After going through the source codes and understanding the underlying concepts, I overcame this challenge too.
- Forming the coast.cont files was the final challenge that I faced, which remained unsolved. As I could not find a solution to this problem, I used the coast.cont files given to me by Dr. Uday Bhaskar as an alternative to continue my work.