Statistics calculation

Year #3 2020

Björn, Carlos, Caroline, Martin and Rez

Stockholm University

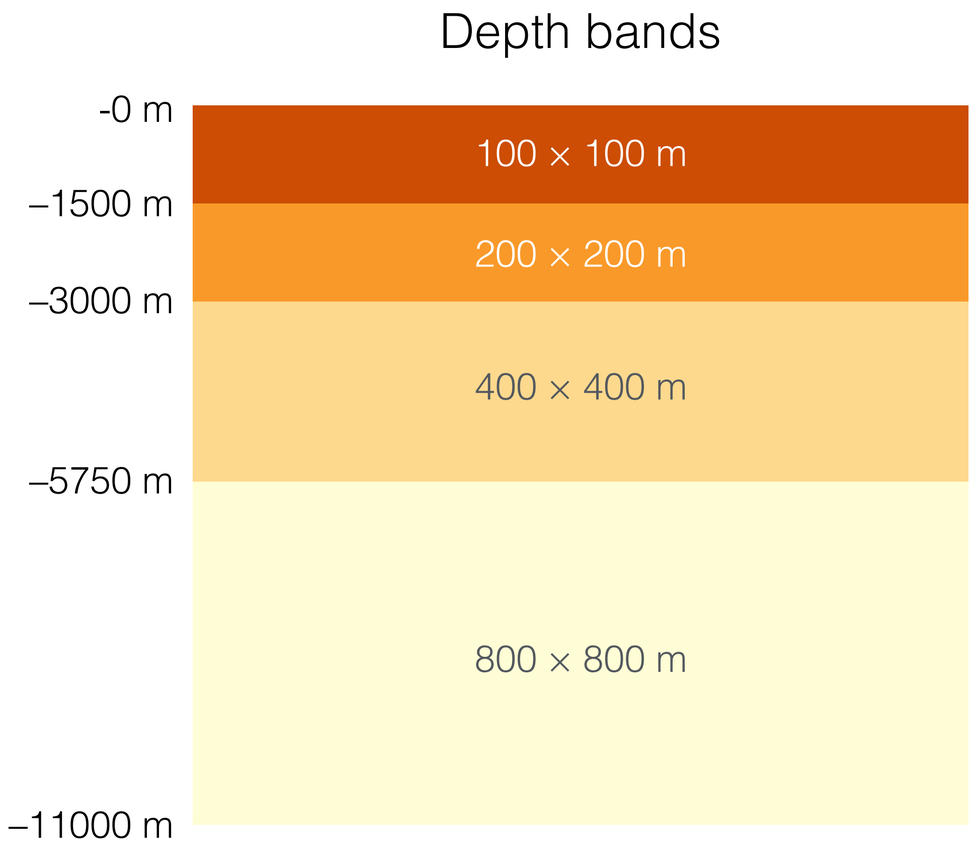
rezze@geo.su.se

Presentation overview

1. Preliminary coverage results
2. Seabed depth bands
3. Under the hood
    - python
    - pandas, numpy
4. Setup
5. Calculation steps
6. Possibilities

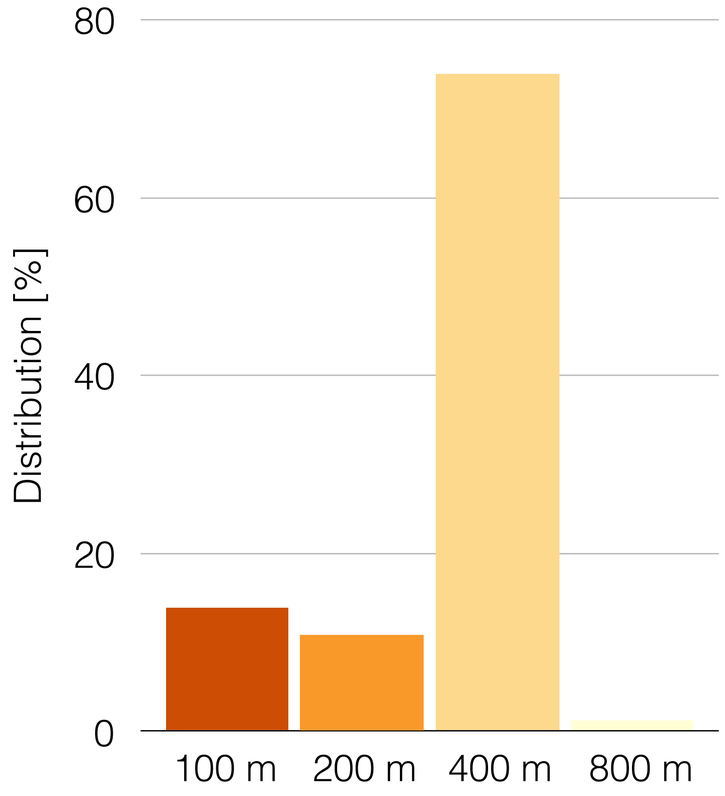
Seabed depth bands

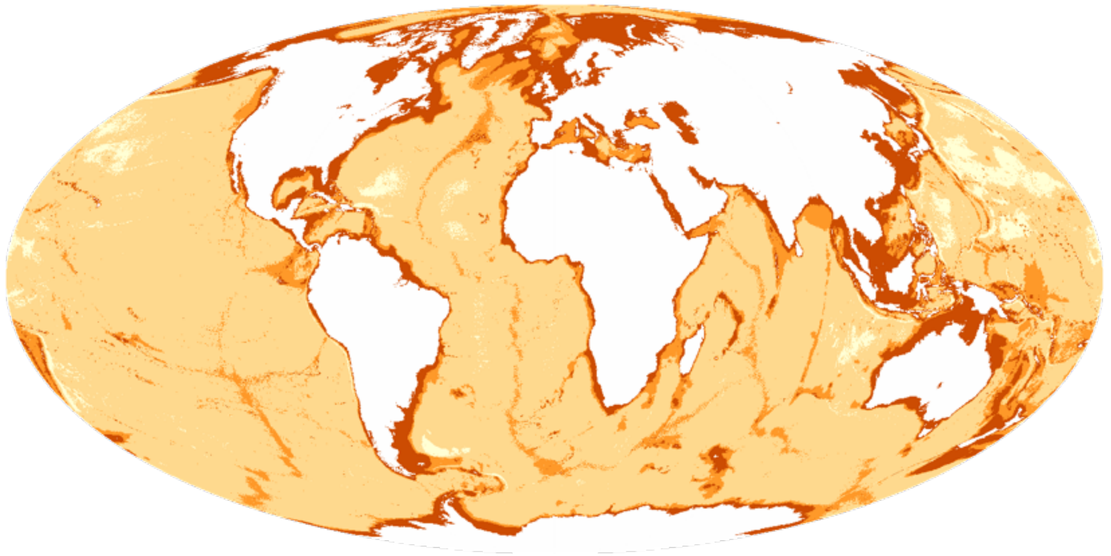
Decreasing resolution with depth



Seabed depth bands

Ocean surface distribution

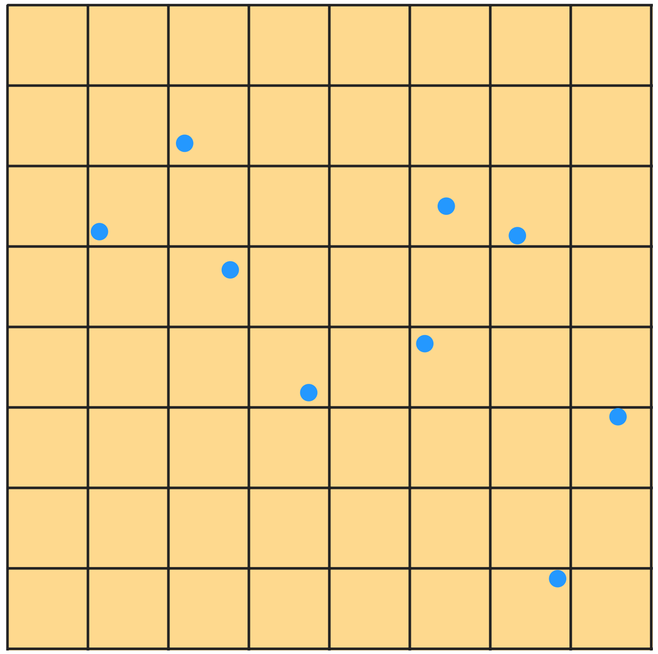




Seabed depth bands

Impact on data coverage

9 / 64 ≈ 14,0 %



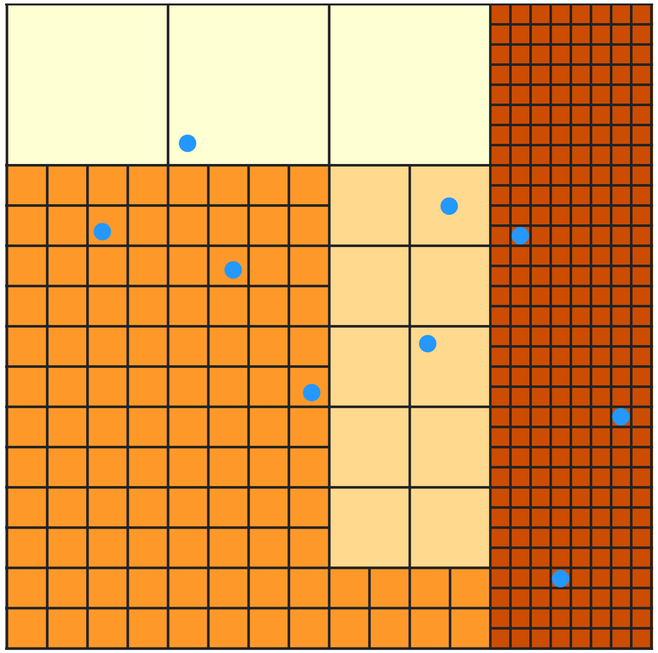
800 × 800 m

400 × 400 m

200 × 200 m

100 × 100 m

9 / 369 ≈ 2,4 %



Under the hood

Python

1. High-level programming language
2. Popular for rapid application development
3. Relatively simple
4. Easy to learn
5. Flexible
6. Open source

Under the hood

Python – Example code

```
# Python 3: Simple arithmetic
>>> 1 / 2
0.5
>>> 2 ** 3
8
>>> 17 / 3  # classic division returns a
float
5.666666666666667
>>> 17 // 3  # floor division
5
```

```
# Python 3: Simple output (with Unicode)
>>> print("Hello, I'm Python!")
Hello, I'm Python!

# Input, assignment
>>> name = input('What is your name?\n')
>>> print('Hi, %s.' % name)
What is your name?
Python
Hi, Python.
```

Under the hood

Python packages (libraries)

1. pandas
   Tabular data

2. numpy
   Grid data (multi-dimensional arrays)

3. xarray
   netCDF storage (labeled multi-dimensional arrays)

Setup

Installation

1. Download python
   E.g. Anaconda (2 min)

2. Download python packages
   `conda install xarray, netcdf4` (4 min)

3. Download python source code
   https://git.bolin.su.se/rez/seabed2030-grid.git (2 min)

4. Prepare input XYZ data in Cylindrical Equal Area 45°
   Five tab separated columns: **x**, **y**, **z**, **priority**, **sid**

Input data example

XYZ ascii data

```
                  x                    y          z   p s
-435409.748999895 8061048.12752250 -3102.0 500 1
-432343.479952218 8061048.12979623 -3105.3 500 1
-435409.738462229 8061202.45983174 -3102.1 500 1
-432781.510083075 8061202.45760693 -3117.0 500 1
-437161.896444168 8061356.77457441 -3088.8 500 1
-436285.817898122 8061356.77142196 -3090.8 500 1
-435409.741859238 8061356.77476238 -3103.8 500 1
-434533.665307618 8061356.77262414 -3089.2 500 1
-433219.558589795 8061356.77239710 -3110.4 500 1
-431029.369722176 8061356.77513360 -3107.7 500 1
```

Calculation steps

From input XYZ data to statistics summary

1. Tile / partition
2. Bin
3. Pyramid
4. Summary
5. Render

Preliminary coverage results

Disclaimer

1. Centre of Greenland below sea level needs to be masked out
2. Some few files are missing or have no data
3. The adjustments are hopefully minor

Preliminary coverage results

Data volume

1. 395 GB ascii input data

2. Total earth area: 510,142,833 km^2

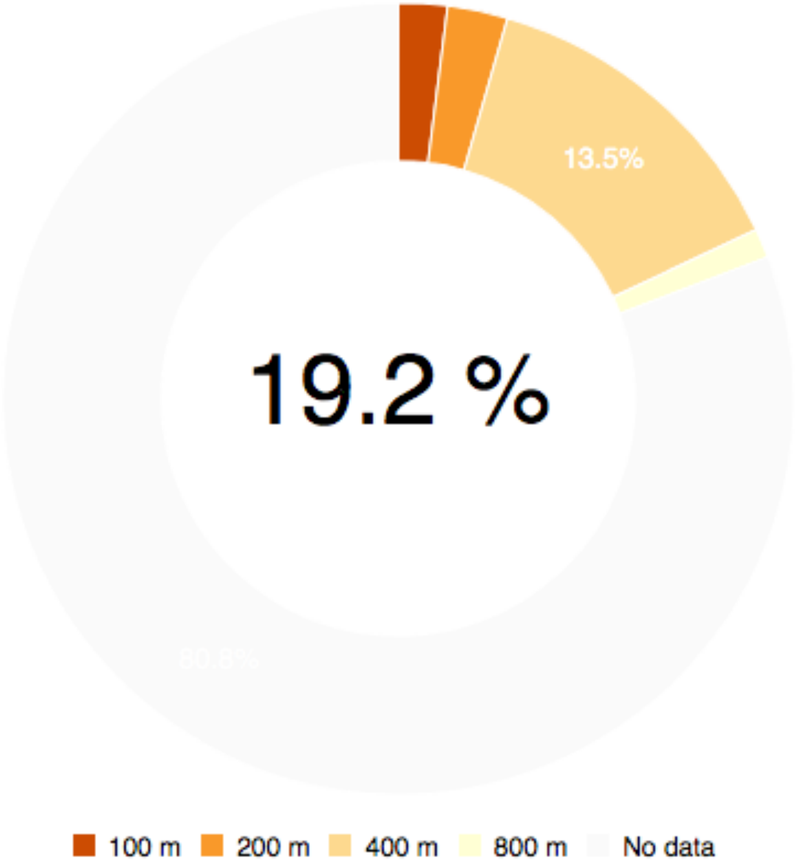
3. Total ocean area: 361,181,126 km^2
   70.8 % of earth's area

4. Total number of ocean grid cells: 36,118,112,551

5. 14 hours of calculation time
   Power work station 225 GB memory (one processor)

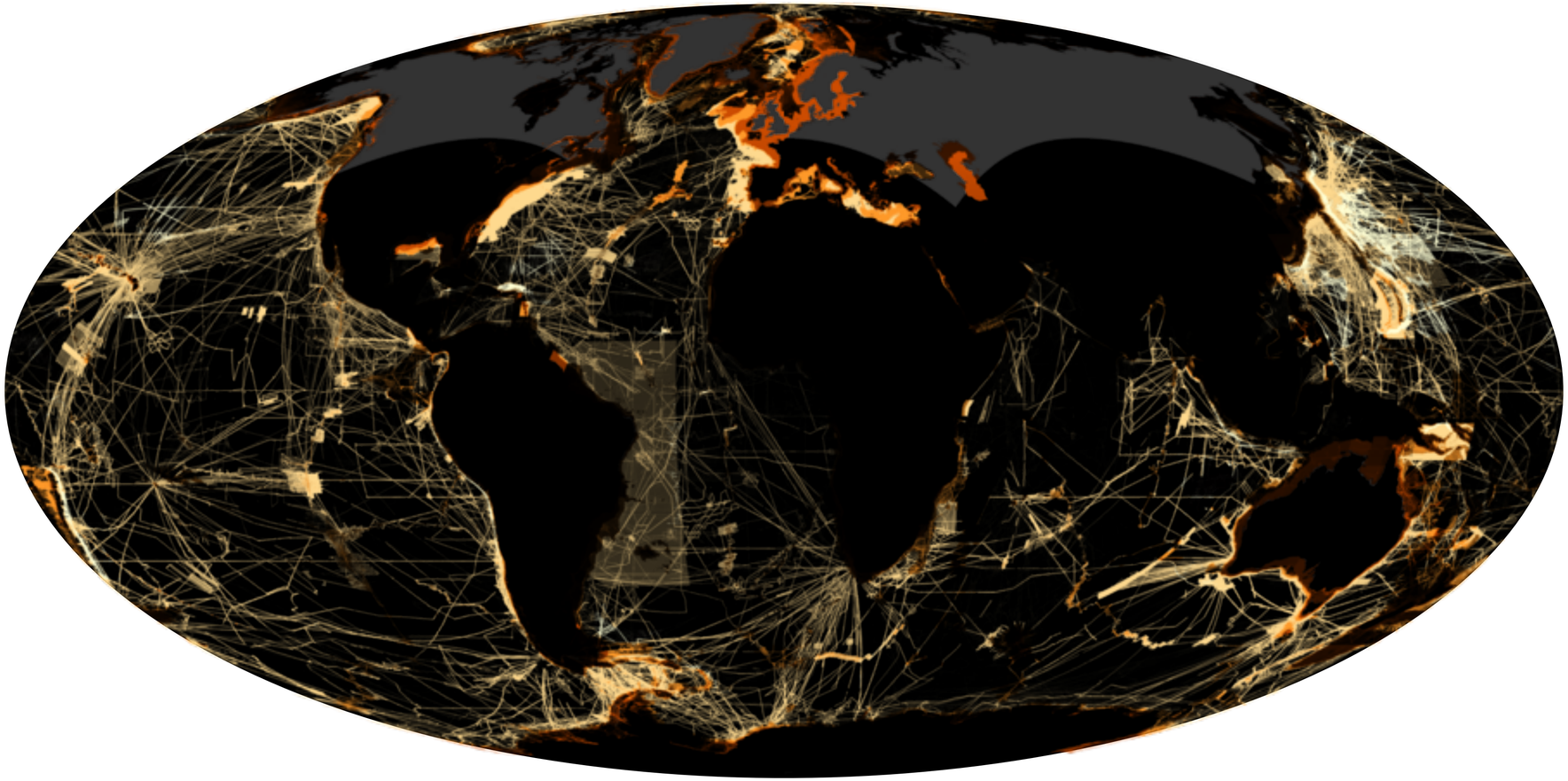
Preliminary coverage results

Seabed depth band ocean coverage chart



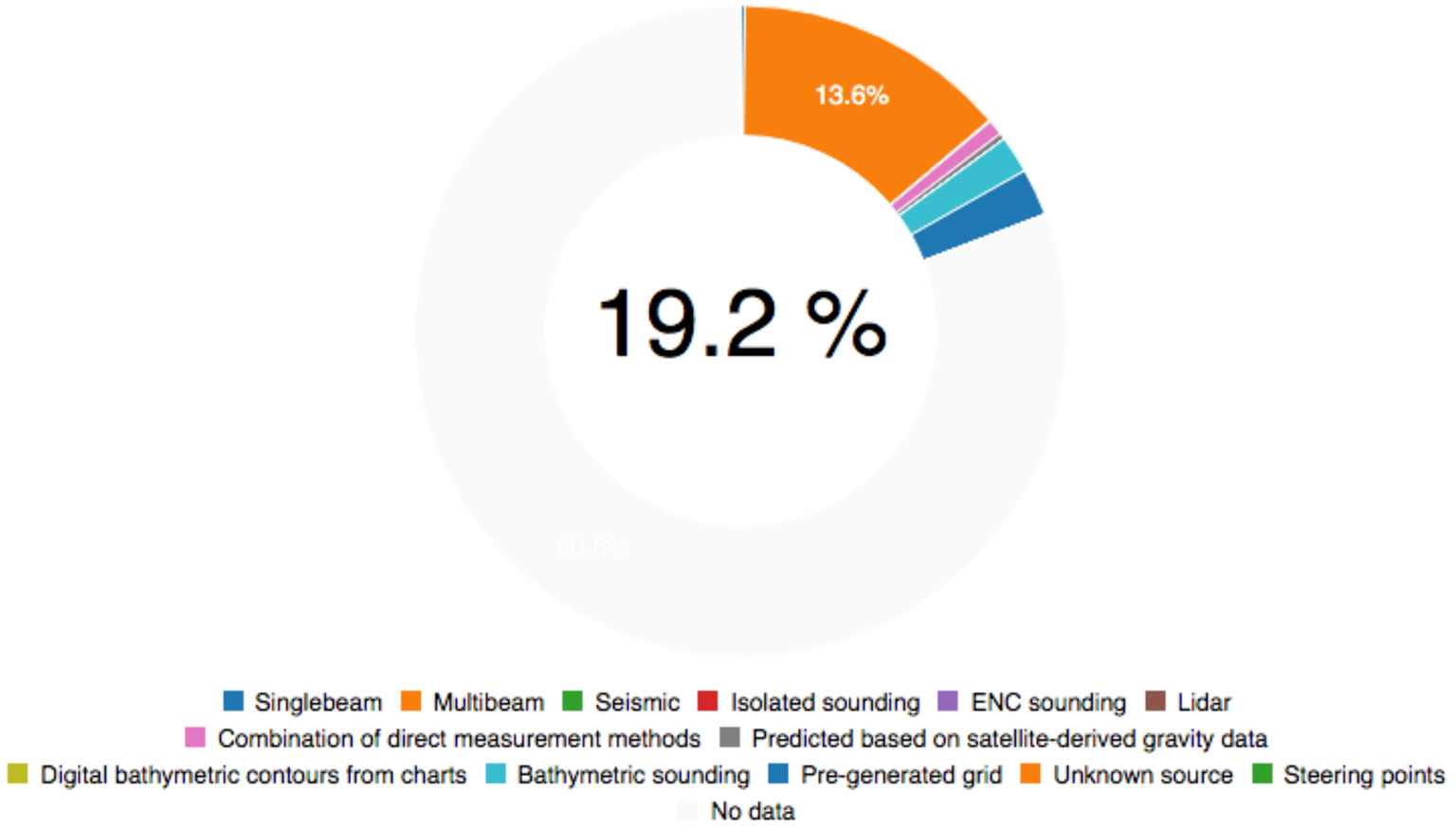
Preliminary coverage results

Seabed depth band coverage map



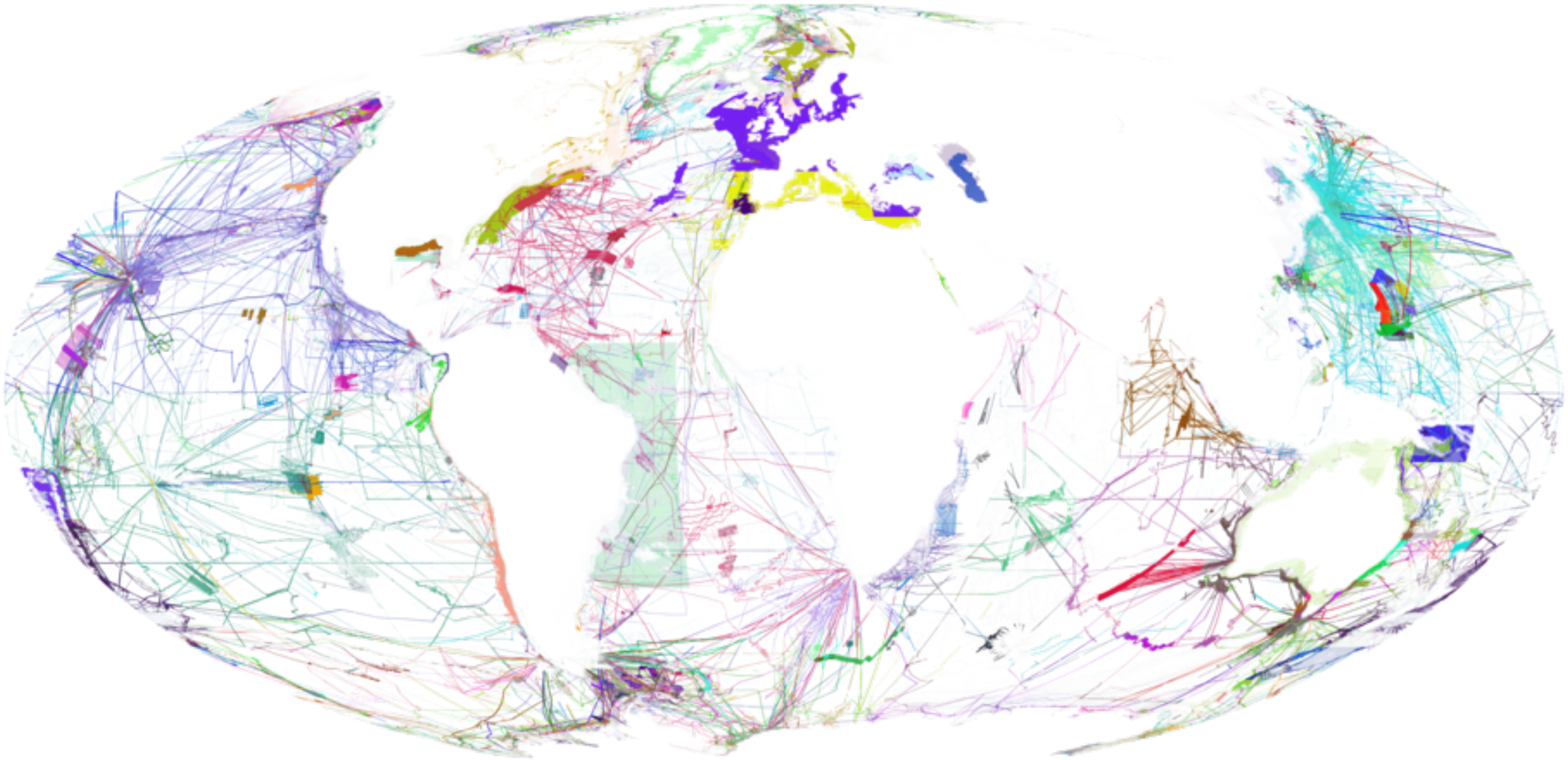
Preliminary coverage results

Seabed data type coverage chart



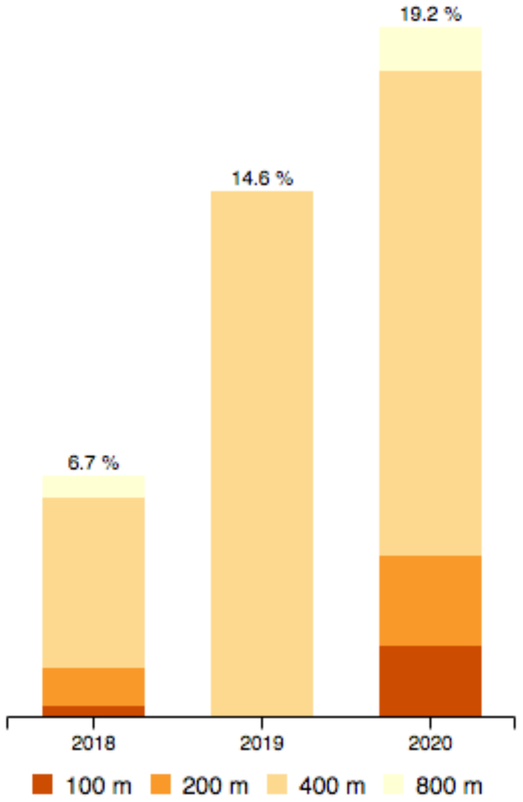
Preliminary coverage results

Seabed data source coverage map



Preliminary coverage results

Coverage increase during the project



Preliminary coverage results

Charts, maps and table

`seabed.geo.su.se/gebco`

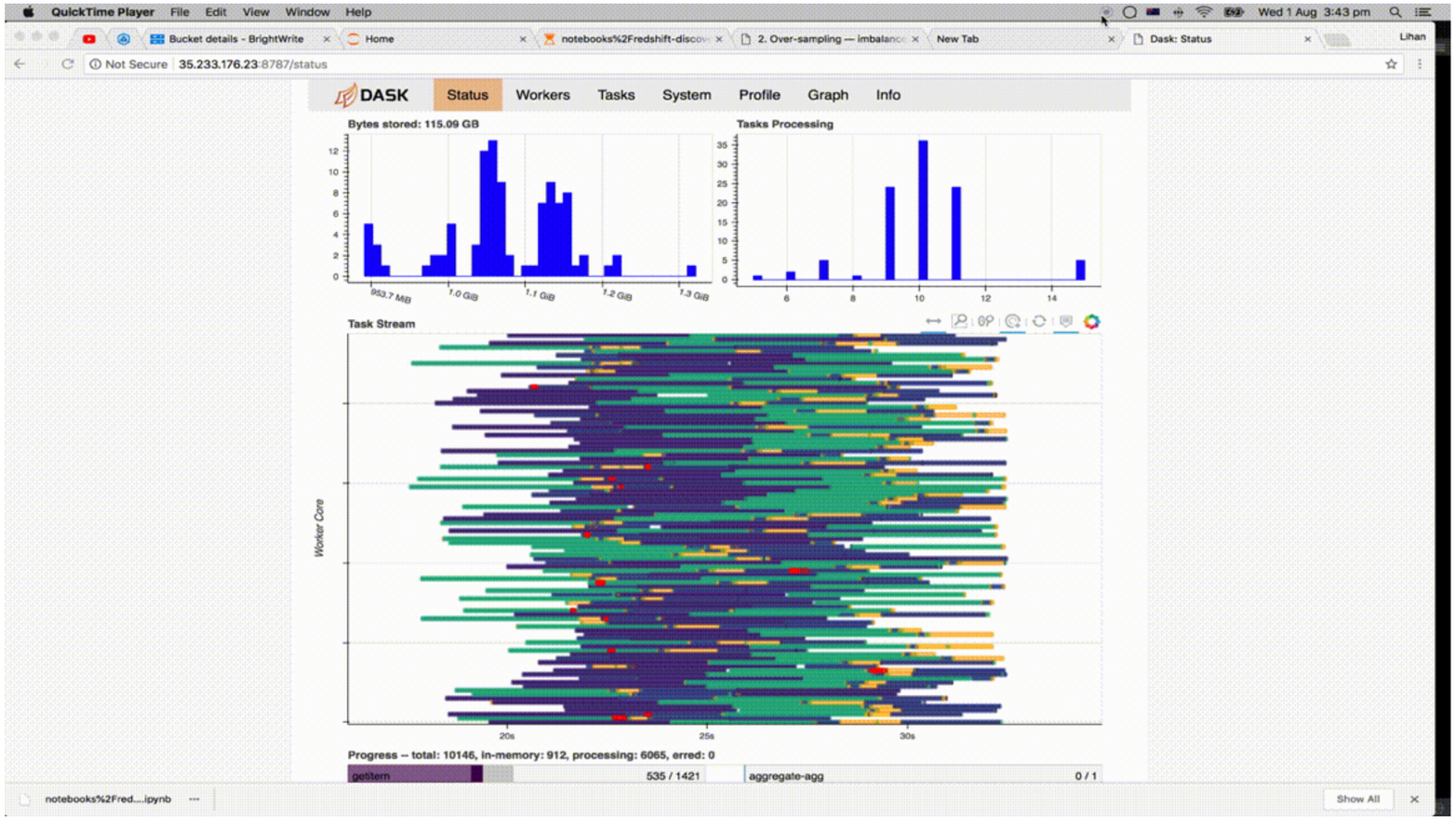
Preliminary coverage results

More stats coming…

1. Data type (TID, i.e. multibeam, singlebeam, etc.)
2. Centre
3. IHO – Regional Hydrographic Commissions (RHC)
4. EEZ

Parallelization

Partition and chunk



Possibilities

For the future

- Simple setup
- Upscaling possibilities in HPC / cloud environments
- Streamlined and harmonized data upload from regional centres
- Iterative and automatic process
- Possible to report statistics quarterly