Project Advancement:
CLAIM 2020 Project

FerryBox database

Managing and administrating the database in order to get faster and more accurate statistic and scientific results

Analyse

Describe

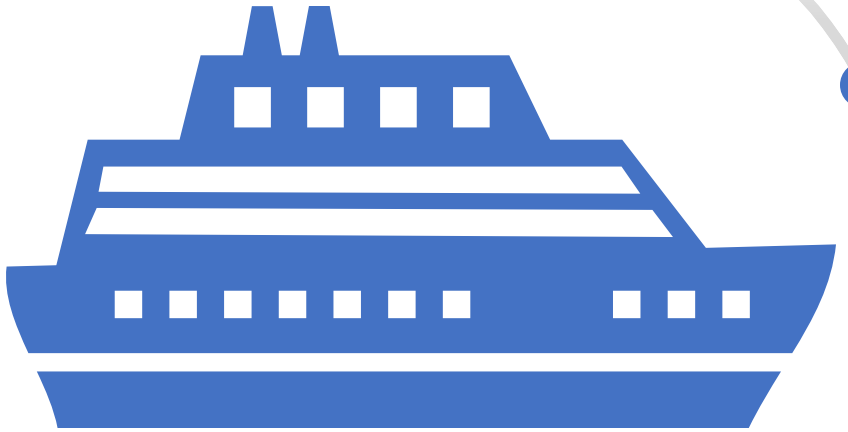
Visualize

Control

Juin 2019 – Mai 2020

Aouachri Sondos

Presentation plan



Introduction
What is the FB database administrator mission ? Where does this mission fit in the big scheme ?

The Tunisian FerryBox project
Why did INSTM team opt for FerryBox technologies ? What has already been done ?

Local management
What data pre-treatments are required to make it scientifically exploitable ? What process to use ?

Web application
What forms of visualization is a user interested in rendering ? How to introduce worldwide FB community to the Tunisian team works ?

Conclusion
What conclusions can we draw at this step ? What should come next ?

FerryBox : Water masses properties and dynamics

- ✓ Device at 5 meter depth
- ✓ Sampling frequency of 1 minute
- ✓ Measured parameters :
  - Temperature
  - Salinity
  - Dissolved oxygen
  - Turbidity
  - pH..

- ✓ The Tunisian FerryBox is currently involved in the **CLAIM EU 2020** project

- ✓ The first launch of FerryBox data collecting campaigns was on 2016

- ✓ The growing database offers several interesting scientific possibilities :
  - Statistical studies
  - Comparison with satellite data
  - Insight into the Mediterranean marine dynamics..

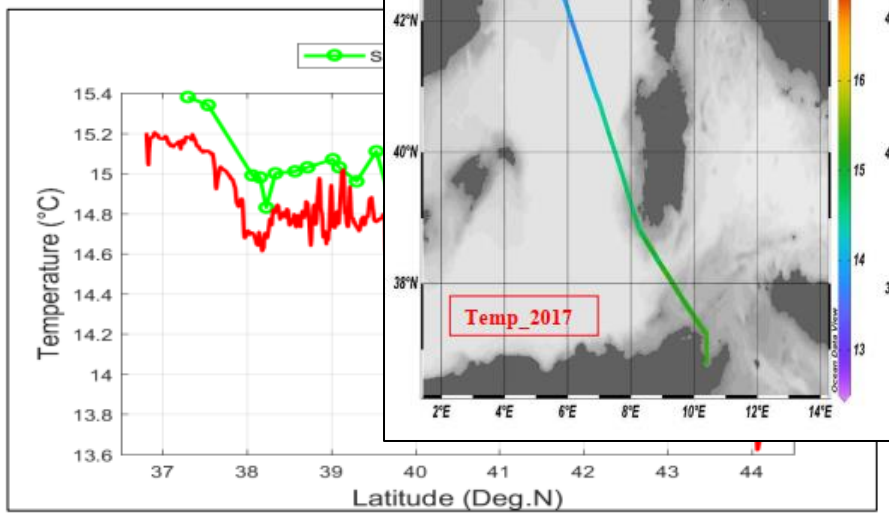
Inventory

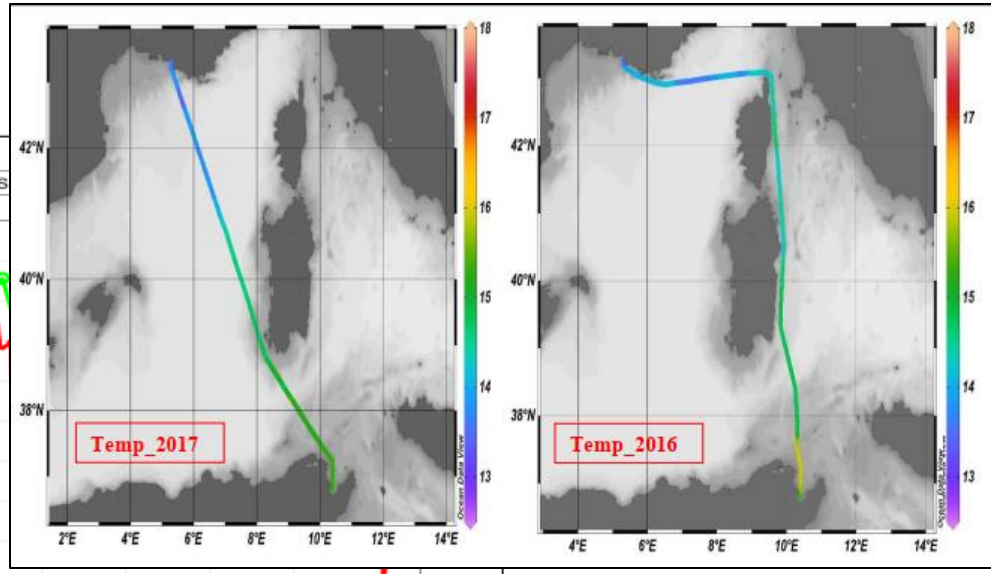
Involvement within the Seadatanet 1 and 2 and SeaDataCloud projects has been successfully completed

INSTM oceanographic data plays a central role in Euro-Mediterranean and African projects

The time series of the Ferry Box data as well as the few missing CTD stations along the Tunisian coast are among the futur ameliorations

Currently, more than 500 FerryBox transects have been processed. Only 18 examples were used to test the next steps regarding Download Manager (DM) and Request Management System (RMS)

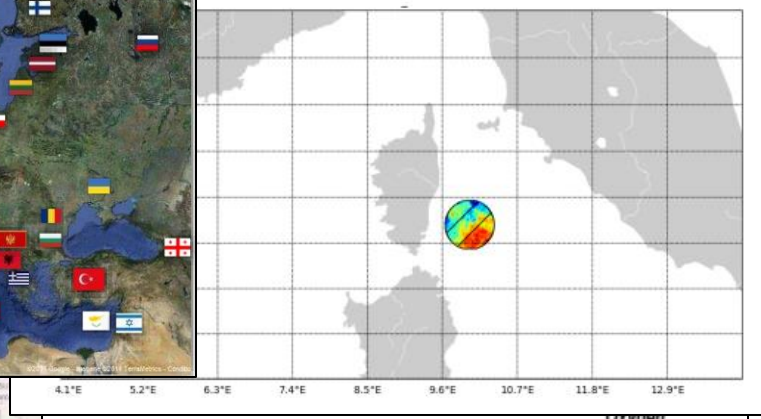


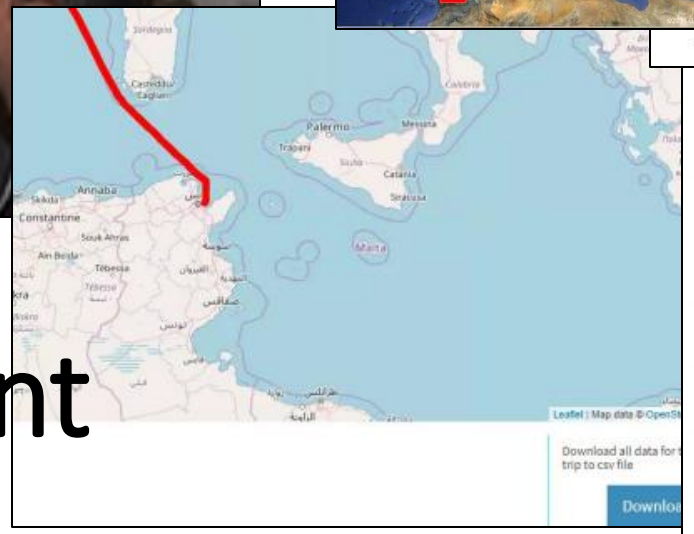


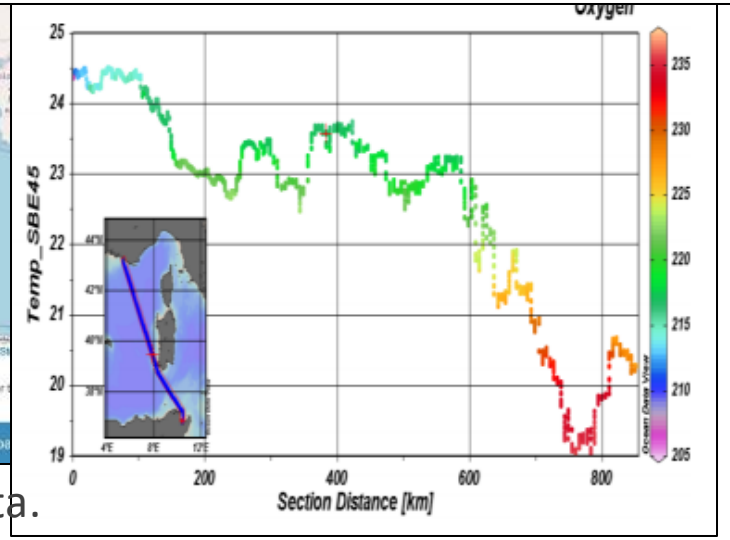
FerryBox system installed on board C/f Carthage of the Tunisian Navigation Company (CTN)

Professional data management is required with agreements on standardization, quality control protocols, archiving and access..

Collect once, Use many times !







4 years of advancement

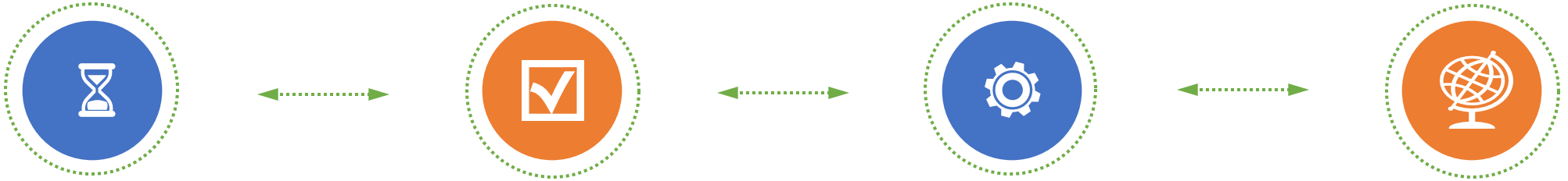
It has been carried out throughout the last 4 years several attempts to manage FerryBox data.
Each contributor used different technologies to tackle one part or more of the project:

- Eliminating the files heading, data plots, quality control.. **(Matlab)**
- Insertion of data in the web app, transect visualization on a map, relational database creation.. **(Php/ Symphony/PostgreSQL)**
- Plotting data, creation of time series, comparing with satellite data.. **(Excel/Matlab)**
- Data pre processing, classification, quality control.. **(Manual)**

Etc..


Php / Symphony
Matlab
Quality systems
Matching data
Satellite data
Files
Scientific articles
Quality control
Visualization
E-mail download
Statistics
Web application
Relevant parameters
Files classification
German database
Python
SeaDataNet
Relational database
Seasonality

Different aspects of the mission



A Files management

Gathering, downloading, automating the classification and the protection of the raw data

B Data management

Indexing files, transforming data into new forms (plots, general stats, time series), pre treatments and quality control

C Database administration

Creation of relational database on Sqlite3 SGBD, automating data insertion

D Web application

Displaying scientific data in an interactive application, different forms of charts, a blog and user management

Quality system design

ISO 9001 Clauses - PLAN

- 1 Scope
- 2 Normative references
- 3 Terms and Definitions
- 4 Context of the organization
  - 4.1 Understanding the organization and its context
  - 4.2 Understanding the needs and expectations of interested parties
  - 4.3 Determining the scope of the quality management system
  - 4.4 Quality management system and its processes
- 5 Leadership
  - 5.1 Leadership and commitment
    - 5.1.1 Leadership And Commitment For The Quality Management System
    - 5.1.2 Customer Focus
  - 5.2 Policy
    - 5.2.1 Establishing the quality policy
    - 5.2.2 Communicating the quality policy
  - 5.3 Organizational roles, responsibilities and authorities
- 6 Planning
  - 6.1 Actions to address risks and opportunities
  - How to address risk in ISO 90001
  - 6.2 Quality objectives and planning to achieve them
  - 6.3 Planning of changes

Key processes are steps that you go through to give the customer what they want, e.g. from order acceptance to design through to delivery.

A good way to do this is to think about how work flows through your organization. Consider how the inputs and outputs to the key processes flow from one process to the next, what sub-processes might exist within it and how the support processes link in.

We have to check that process inputs and outputs are defined and review how each of the processes are sequenced and how they interact.

FerryBox database

A very abundant information about the Mediterranean water surface

Local management

Managing files and data

Web application

Data interactive visualization

Python main libraries

Pandas

Manipulation and analysis of data: structuring of data and operations of manipulation of numerical tables and time series.

Numpy

Manipulation of multidimensional arrays and arrays, as well as mathematical functions operating on these arrays.

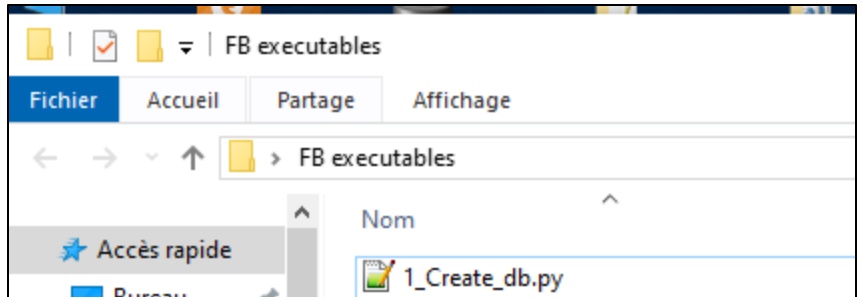
Matplotlib

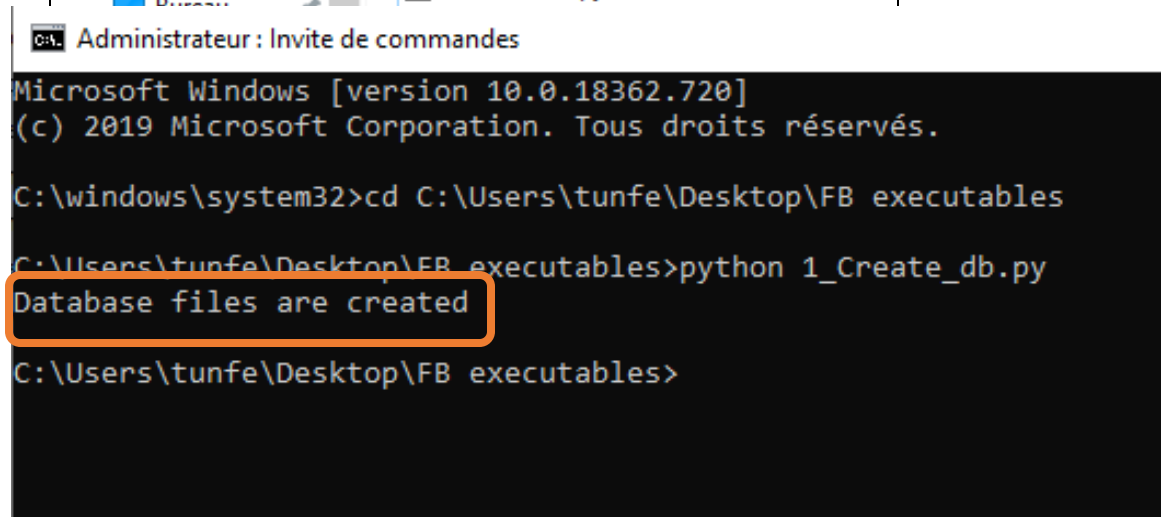
Plotting and visualization of data in the form of graphs

Seaborn

It allows you to create statistical graphs in Python. It is built on matplotlib and is tightly integrated with pandas data structures (hence the choice of this library).

Daily refreshment process





Download

1_Create_db.py
2_E-mail_download.py

Classification

3_Files_classification.py

Indexation

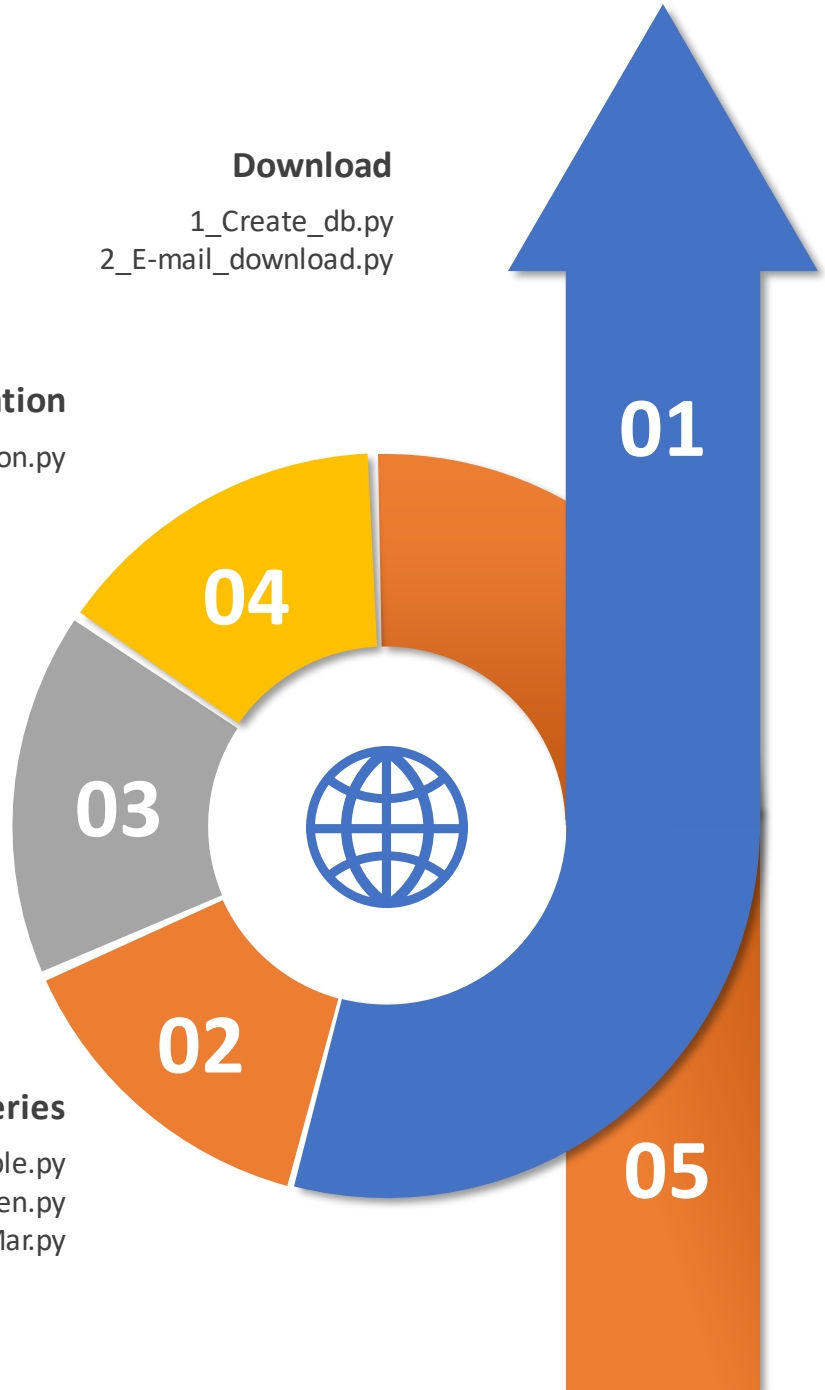
4_Files_pretreatment.py

Visualization

5_plots_generation.py

Metadata / time series

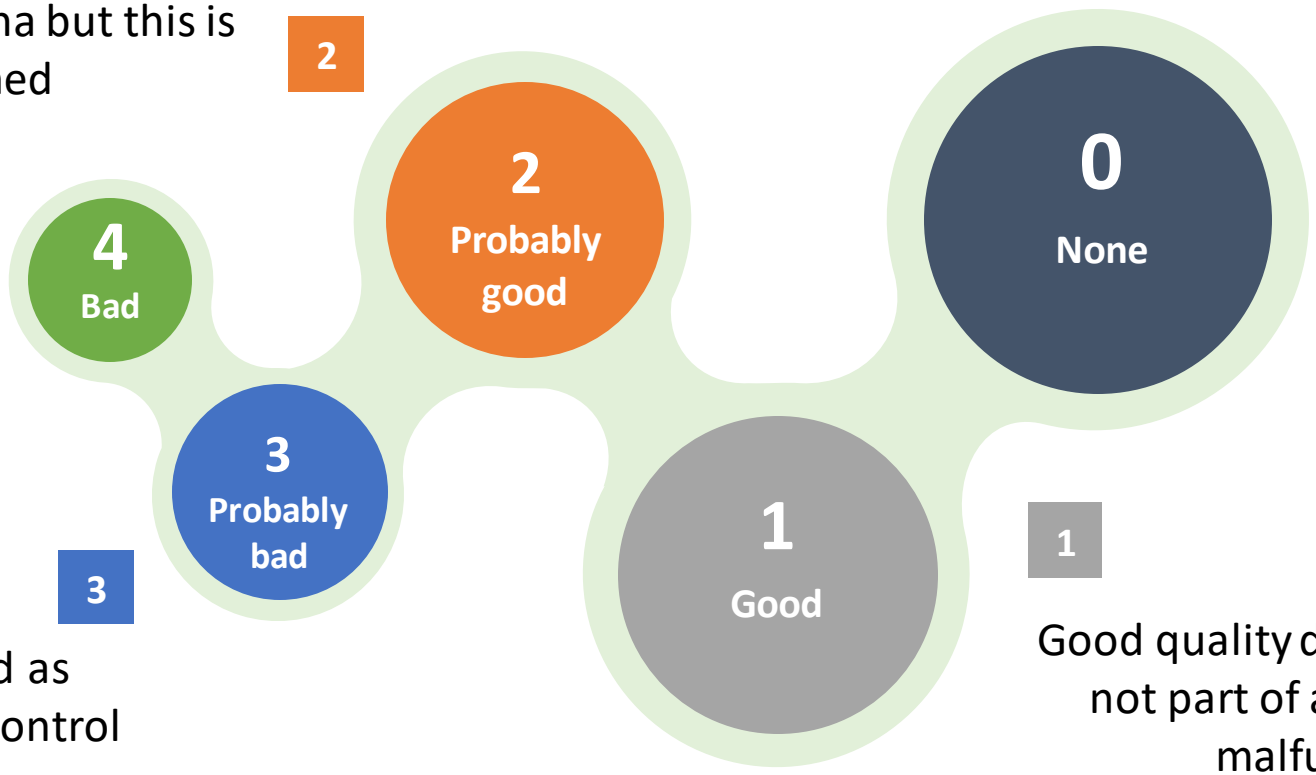
9_Metadata_table.py
7_Time_series_Gen.py
8_Time_series_Mar.py



Quality control flags

0 No quality control procedures have been applied to the data value

Data value that is probably consistent with real phenomena but this is unconfirmed 2



4 An obviously erroneous data value.

3 Data value recognized as unusual during quality control

1 Good quality data value that is not part of any identified malfunction

Quality control flags

- A quality flag is assigned to each data value.
- Quality flags are used to describe the data value, no changes are made to the data values.

Spike test

Differences between sequential measurements, where one measurement is quite different than adjacent ones, is a spike in both size and gradient.

$$Test\_value = \left| \frac{V_2 - (V_3 + V_1)}{2} \right| - \left| \frac{(V_3 - V_1)}{2} \right|$$

where V2 is the measurement being tested as a spike, and V1 and V3 are the values previous and next.

- **Temperature**: The V2 value is flagged when the test value exceeds 6.0 degree C.
- **Salinity**: The V2 value is flagged when the test value exceeds 0.9 PSU

Values that fail the spike test should be flagged as wrong and should not be distributed.

Gradient

This test is failed when the difference between adjacent measurements is too steep.

$$Test\_value = \left| \frac{V_2 - (V_3 + V_1)}{2} \right|$$

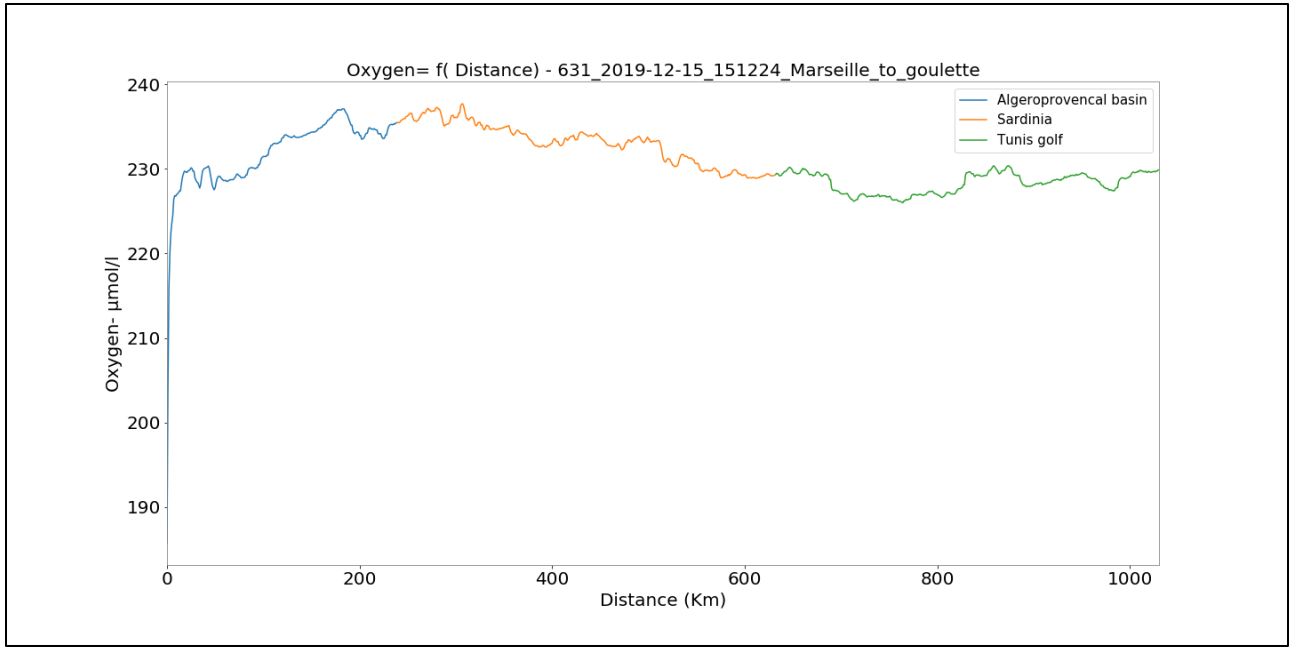
where V2 is the measurement being tested as a spike, and V1 and V3 are the previous and next values.

- **Temperature**: The V2 value is flagged when the test value exceeds 9.0 degree C.
- **Salinity**: The V2 value is flagged when the test value exceeds 1.5 PSU

Values that fail the test (i.e. value V2) should be flagged as wrong.

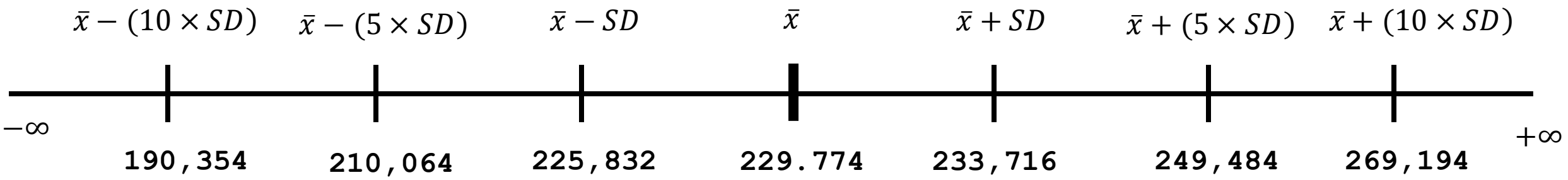
Example

File : 631_2019-12-15_151224_Marseille_to_goulette
Parameter : Oxygen



Mean value : 229.774

Standard deviation: 3.942



Time series creation process

Calculate the corresponding average value

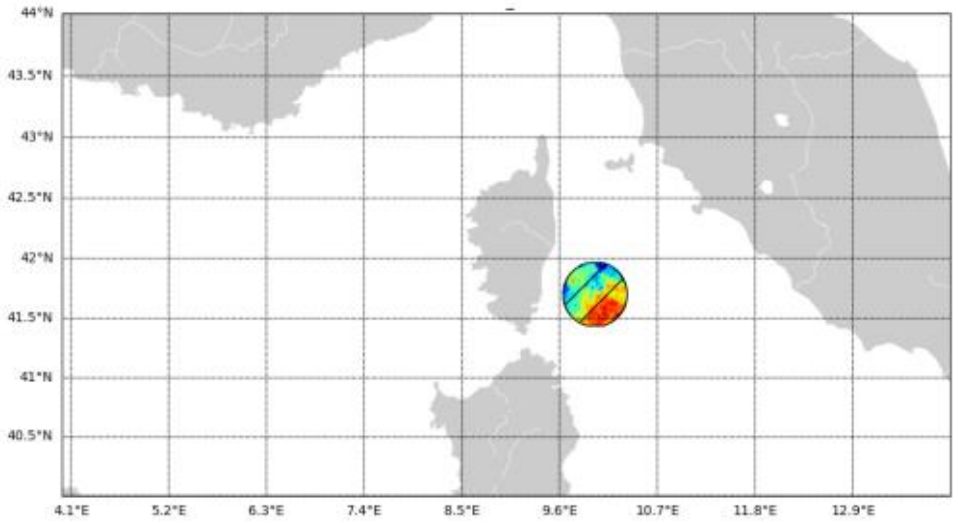
The set of values that were captured in the corresponding set are represented by their mean value

Fix the position - cell

The number of cells is : total distance / size of the cell

Detect the date

The date will be inserted in the x-axis



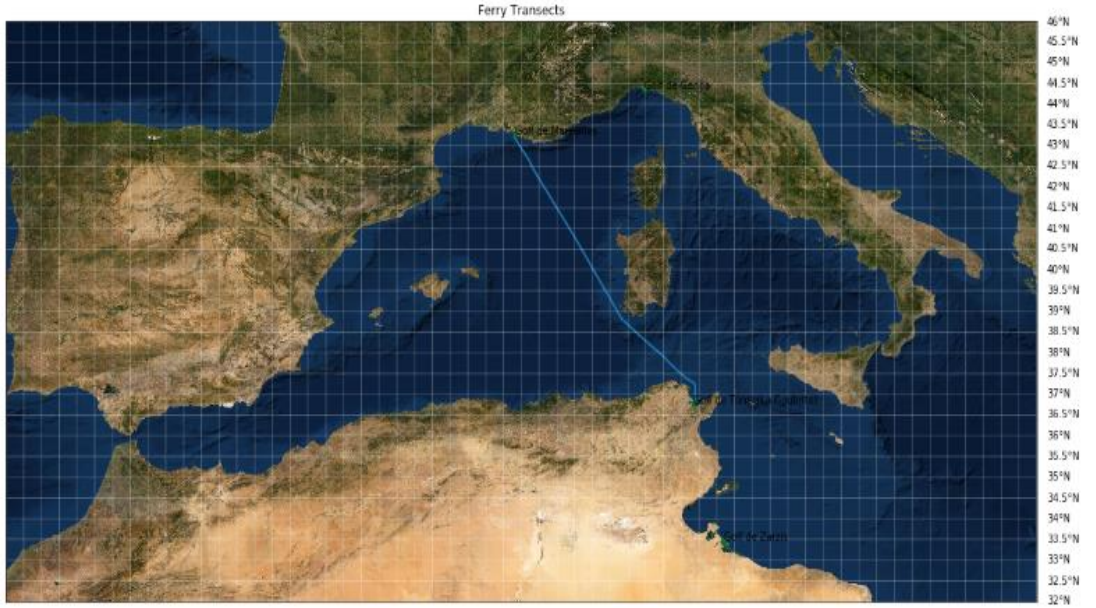
Detect the first path cell (5 km)

Original FerryBox file

Date	Latitude	Longitude	Salinity_SBE45	Distance	Cumul_Distance
09/08/2016	44.325697	8.986357	38.149008	0.601666	850.2433024
09/08/2016	44.330328	8.984447	38.132975	0.536668	850.7799709
09/08/2016	44.334445	8.982676	38.106117	0.478786	851.2587565
09/08/2016	44.338354	8.981048	38.065742	0.453353	851.7121094
09/08/2016	44.342096	8.979485	38.024975	0.434087	852.1461963
09/08/2016	44.345642	8.978024	37.997525	0.41089	852.5570864
09/08/2016	44.349429	8.976265	37.997525	0.443566	853.0006528
09/08/2016	44.353245	8.973875	37.994292	0.464881	853.4655336
09/08/2016	44.356914	8.971277	38.004733	0.457295	853.9228285
09/08/2016	44.360167	8.968976	38.004733	0.405353	854.3281818
09/08/2016	44.363083	8.966903	38.025183	0.363731	854.691913
09/08/2016	44.365833	8.964949	38.0367	0.342986	855.034899

Insert the parameter's mean value

Detect the date



Generated time serie

	A	B	C	D	E	F	G
1	Date	C1	C2	C3	C4	C5	C6
2	09/08/2016	38.0457098	38.201788	38.2851334	38.223326	38.2056789	38.190334
3	08/08/2016	37.3216933	37.2855288	37.2991949	37.2811453	37.2750834	37.250528
4	07/08/2016	38.253919	38.1939946	38.2428334	38.304363	38.1905784	38.22508
5	04/08/2016	37.3654954	37.3230844	37.2925143	37.2979819	37.2904074	37.279241
6	03/08/2016	38.2876265	38.2378459	38.2547643	38.2769547	38.261706	38.218794
7	02/08/2016	37.2999584	37.2895023	37.3070143	37.2898843	37.2484334	37.24317
8	01/08/2016	38.2484234	38.2917585	38.2591799	38.2518169	38.2321976	38.251579
9	29/07/2016	38.2422392	38.2089283	38.228669	38.2632441	38.2394476	38.234801
10	27/07/2016	37.2655537	37.226259	37.219541	37.2319145	37.2163466	37.205981
11	26/07/2016	38.201912	38.1671115	38.1781219	38.2041857	38.206432	38.182373
12	25/07/2016	37.2025665	37.1650584	37.1696249	37.1829989	37.1735787	37.169695
13	24/07/2016	38.1804583	38.1697621	38.1830429	38.2352071	38.2063904	38.180780
14	23/07/2016	37.3310463	37.1769214	37.1608585	37.1464381	37.1464071	37.158096
15	22/07/2016	38.201077	38.1691484	38.1487763	38.1172179	38.1209119	38.13203
16	20/07/2016	37.2608233	37.139301	37.1615947	37.1609644	37.1340291	37.113691
17	19/07/2016	38.1484135	38.1885822	38.1710561	38.1476013	38.1278679	38.129236

File : 99_2016-08-09_185756_Goulette_to_genova_731501_pre.csv

FerryBox database

A very abundant information about the Mediterranean water surface

Local management

Managing files and data

Web application

Data interactive visualization

Web application components

Administrator panel

This part of the dashboard requires administrator login/password access, It enables the import of data in the database.

Community blog

This part of the application is where updates about the project, the advancement and encountered problems are discussed with the international community.

Data description

In this part, the projects current advancement, collected measurements and general statistics are displayed,

Data access

The international users of the Tunisian FerryBox Dashboard can, using this part of the application, visualize filtered data dynamically, in real time.

Data download

Users are offered the possibility to download data by requesting it via an online form.

Motivations

It is important to mention that a web application has already been created , in the previous years. This new web application is inspired from that previous result, while:

- Using a new database structure, based on the new files contents
- Providing a more rich user experience (user interface, data description, blog and articles ..)
- Displaying data in more than one charts type and form
- Filtering data on real-time basis
- Showing the ferry transect while coloring the path based on the measured parameter
- Using a Python-based developing platform, to ensure the continuity and coherence of the used technologies ..

Tunisian FerryBox dashboard

Helmholtz-Zentrum Geesthacht

European FerryBox Database

Transect Plot I

[*Transect Plot I* | Transect Plot II | Time-Series Plot | Scatter Plot | Map Plot | Home]

[Help | Forward | Back]

Areas: ☑ Arctic-ROOS ☑ BOOS ☑ IBI-ROOS ☑ MONGOOS ☑ NOOS [Area Map]

• Route: Troms-Longyearb (NIVA, NO)

• Transect: 24.11.2019 12:12 Bare-Long

• Parameter Selection: ○ Single ◉ Multiple

• Parameters: FLU2, PSAL, TEMP, TUR4

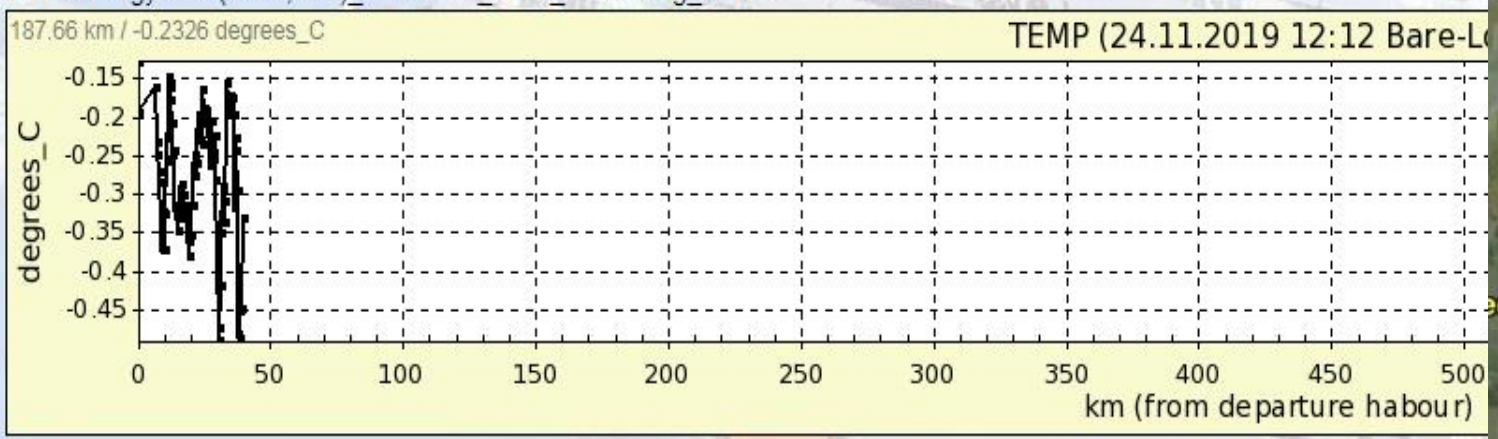
Quality: ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4

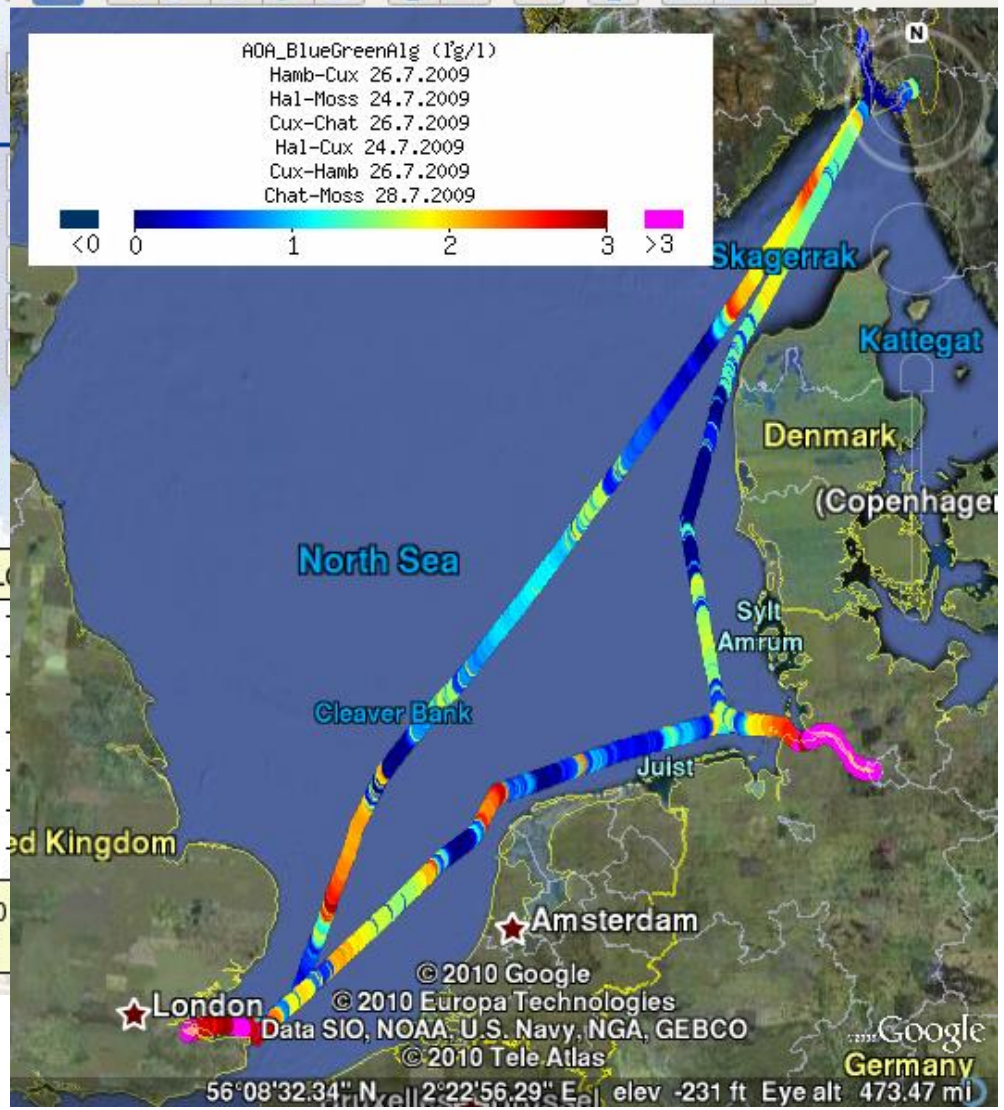
• Show Quality: ◉ no ○ yes

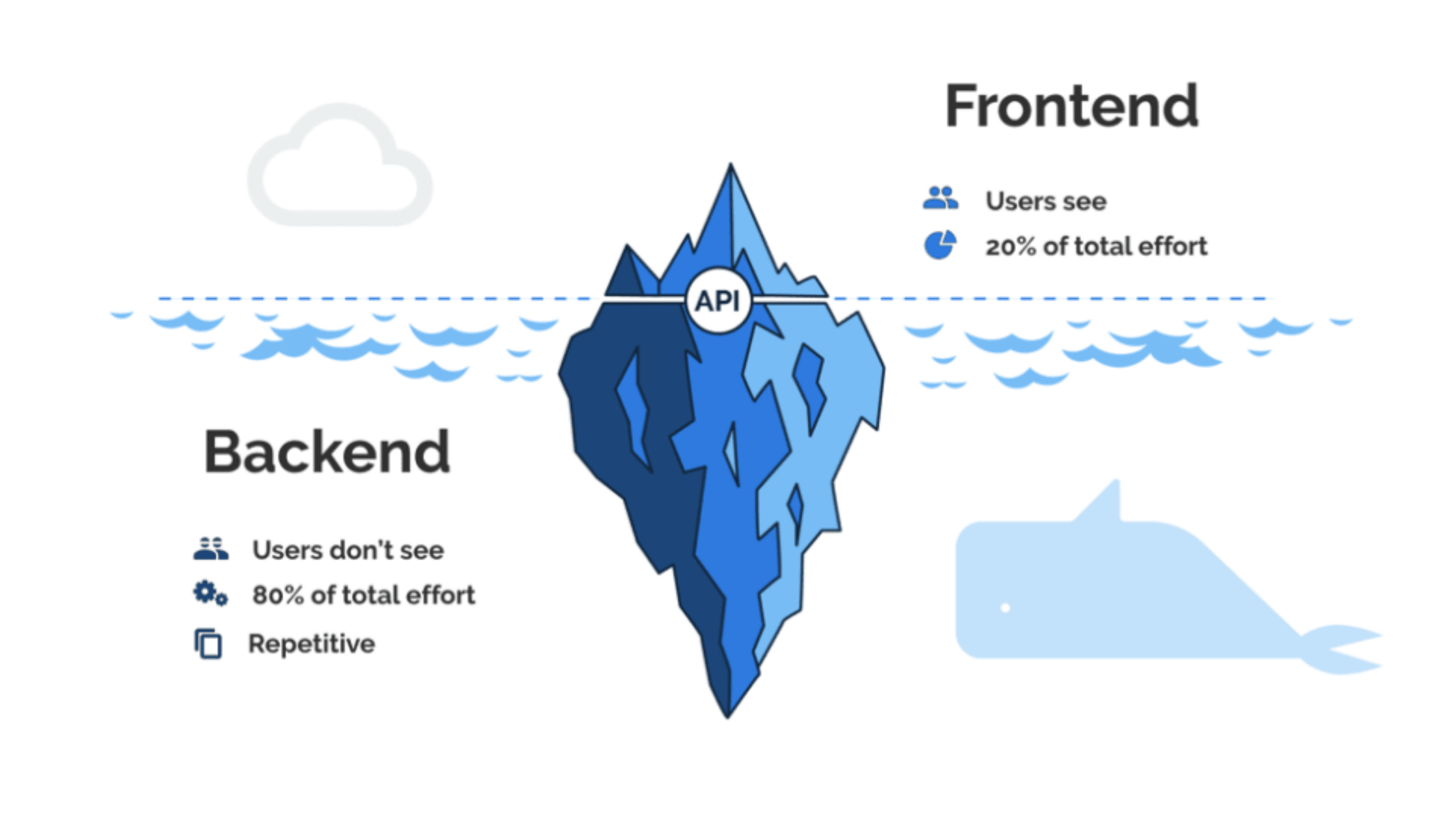
Plot | Cancel

Max. Deviation:
X-Axis Dimension:
X-Axis Range:
Y-Axis Scaling:
Plot Geometry:
Plot Colors:
Font Type/Size:
Line Thickness:
Marker Size:
Grid:
Info Text:
Plot Title:

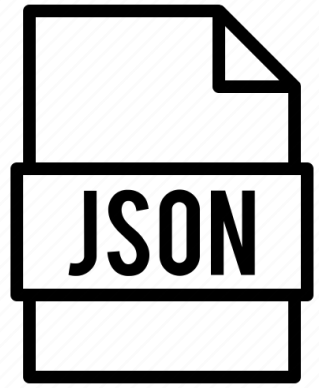
Troms-Longyearb (NIVA, NO)_20191124_1212_Bare-Long_TEMP






Frontend
Users see
20% of total effort
API
Backend
Users don't see
80% of total effort
Repetitive

Bootstrap



JavaScript

FERRYBOX DASHBOARD

Hello admin

Data Overview

Data Access

Data Download

Logout

Graphs: TRANSECT MAP TIME SERIES SCATTER

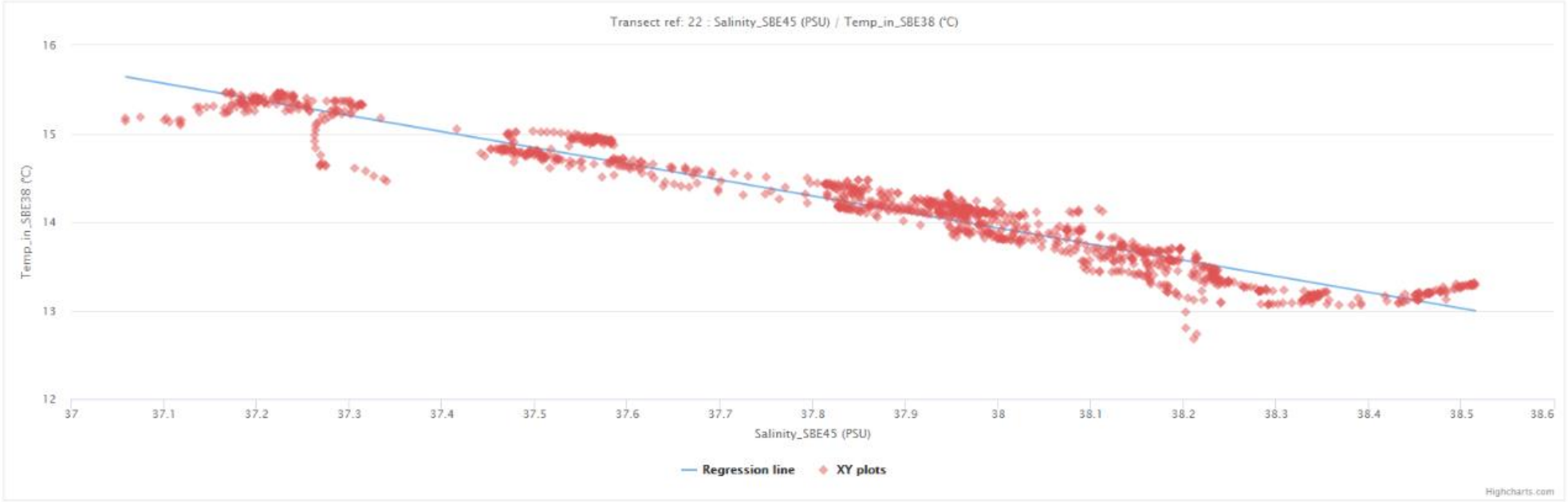
Transect reference: 22_2016-03-10_114236_Marseille_to_Goulette_731501

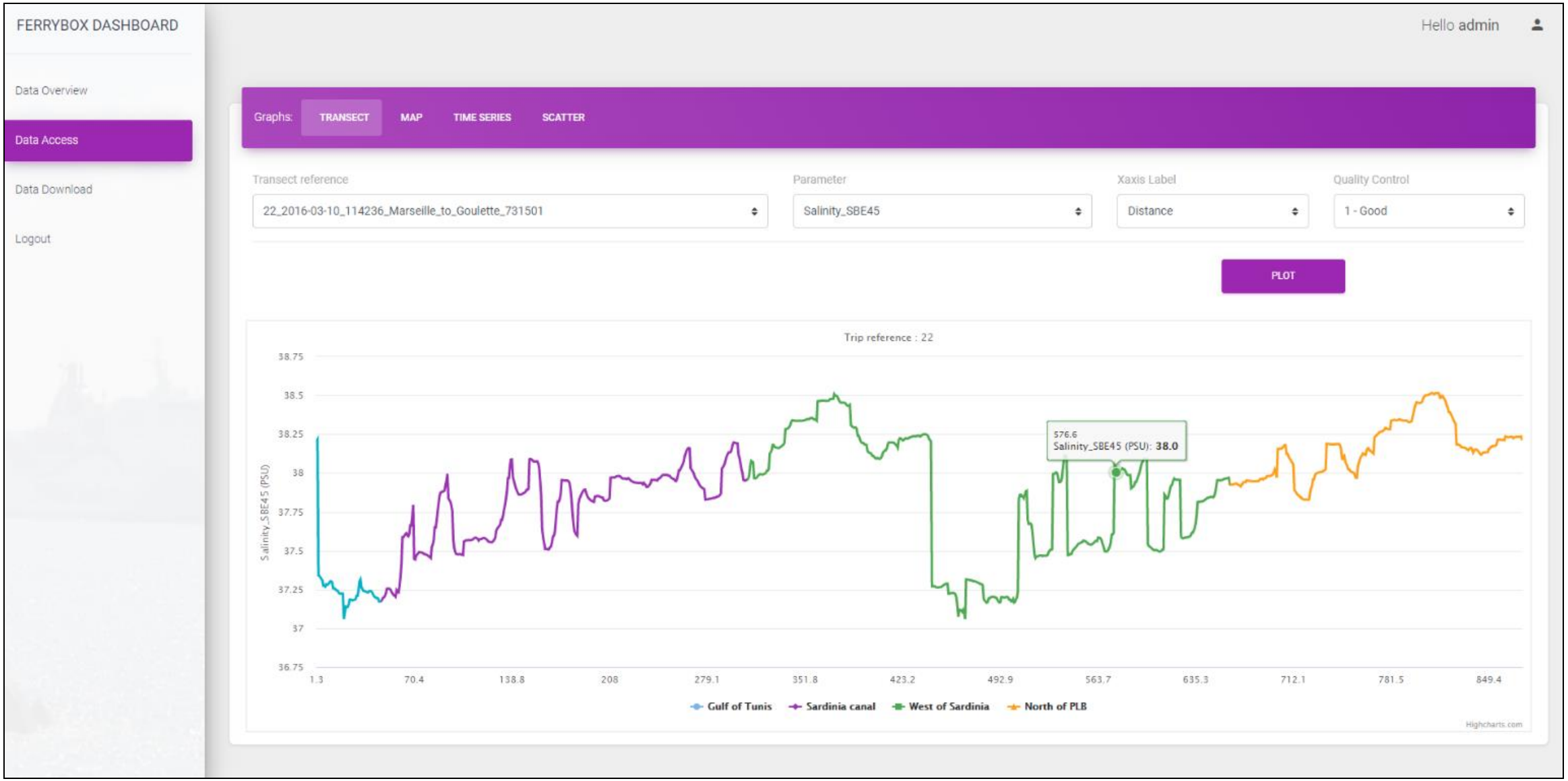
Parameter 1: Salinity_SBE45

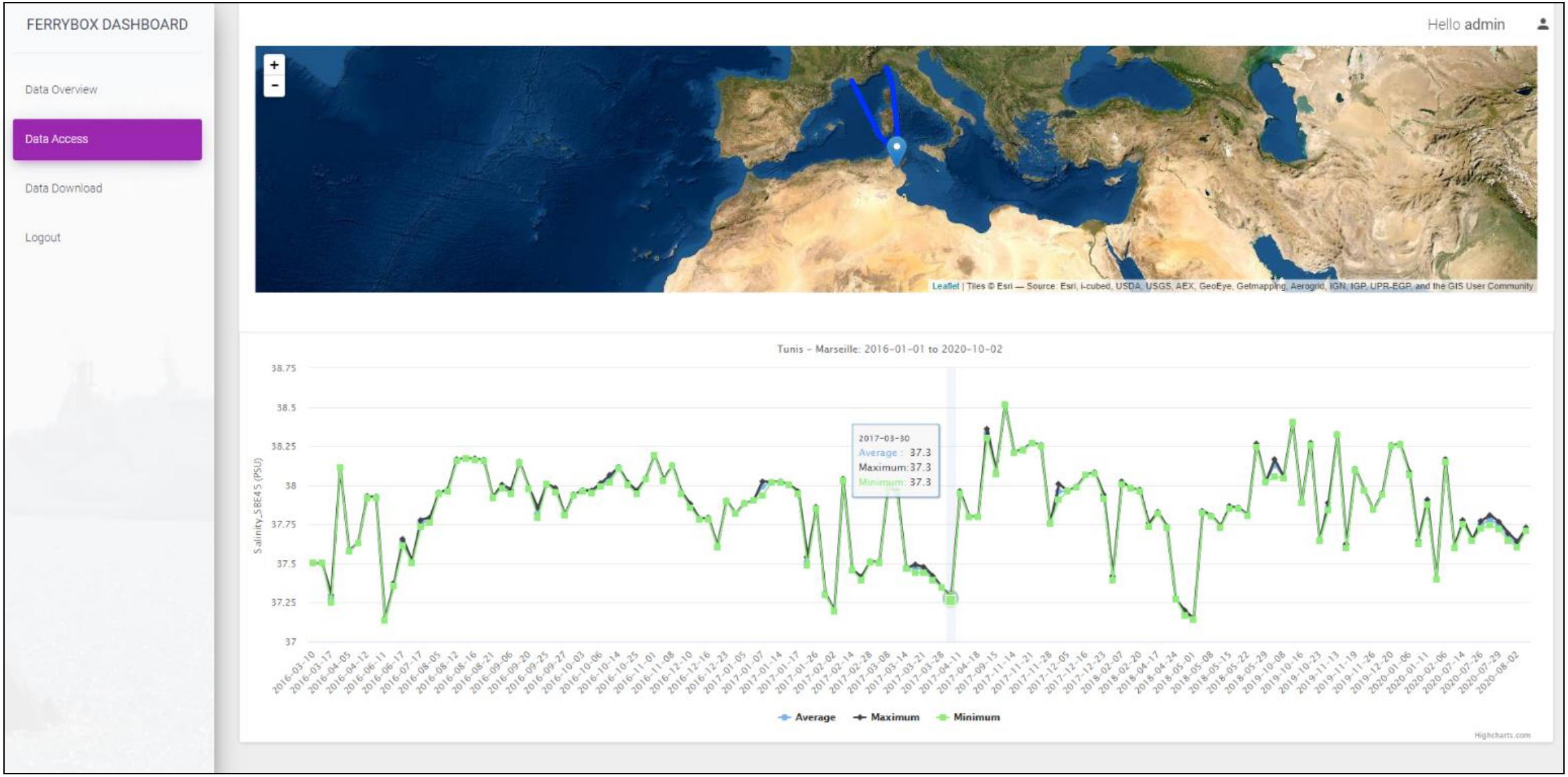
Parameter 2: Temp_in_SBE38

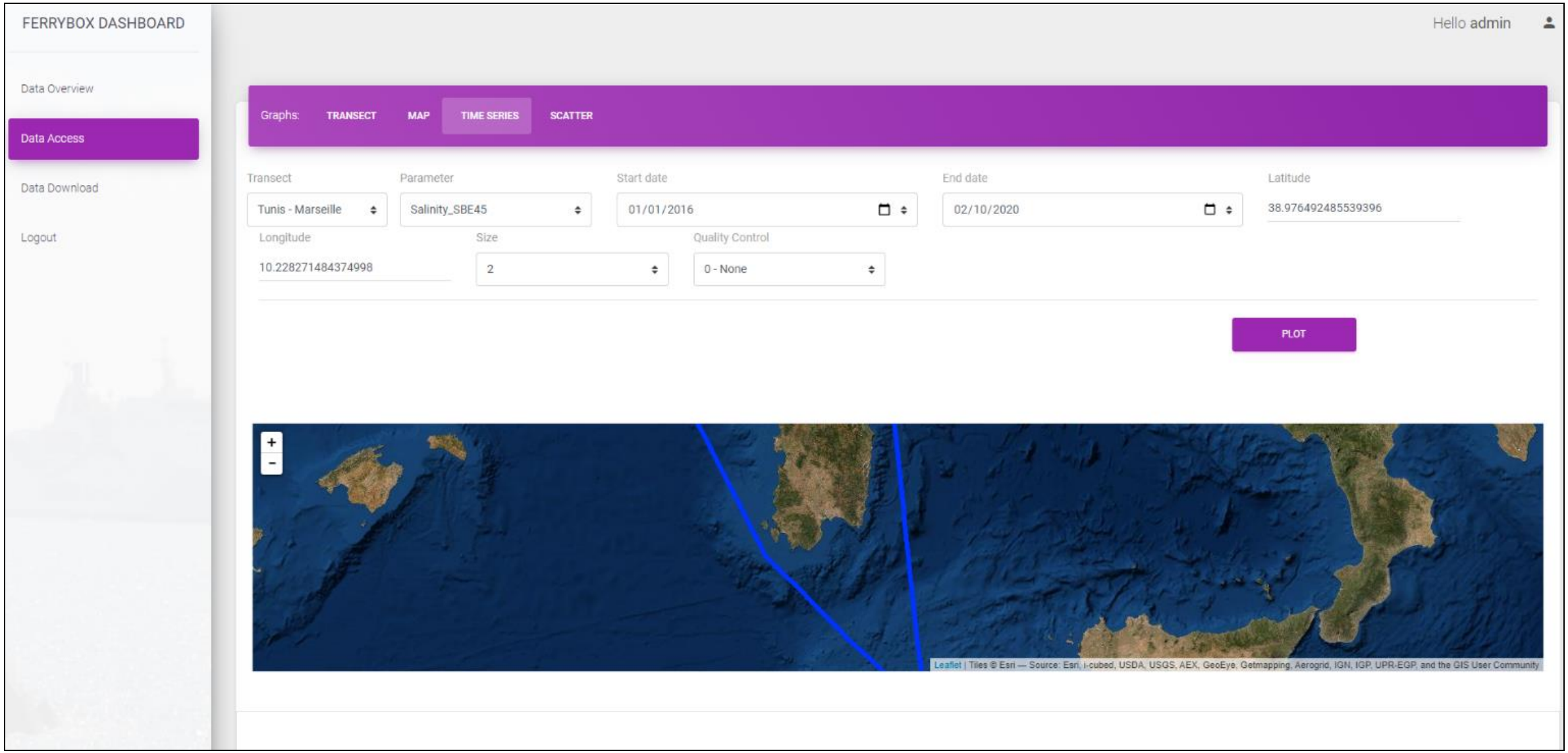
Quality Control: 1 - Good

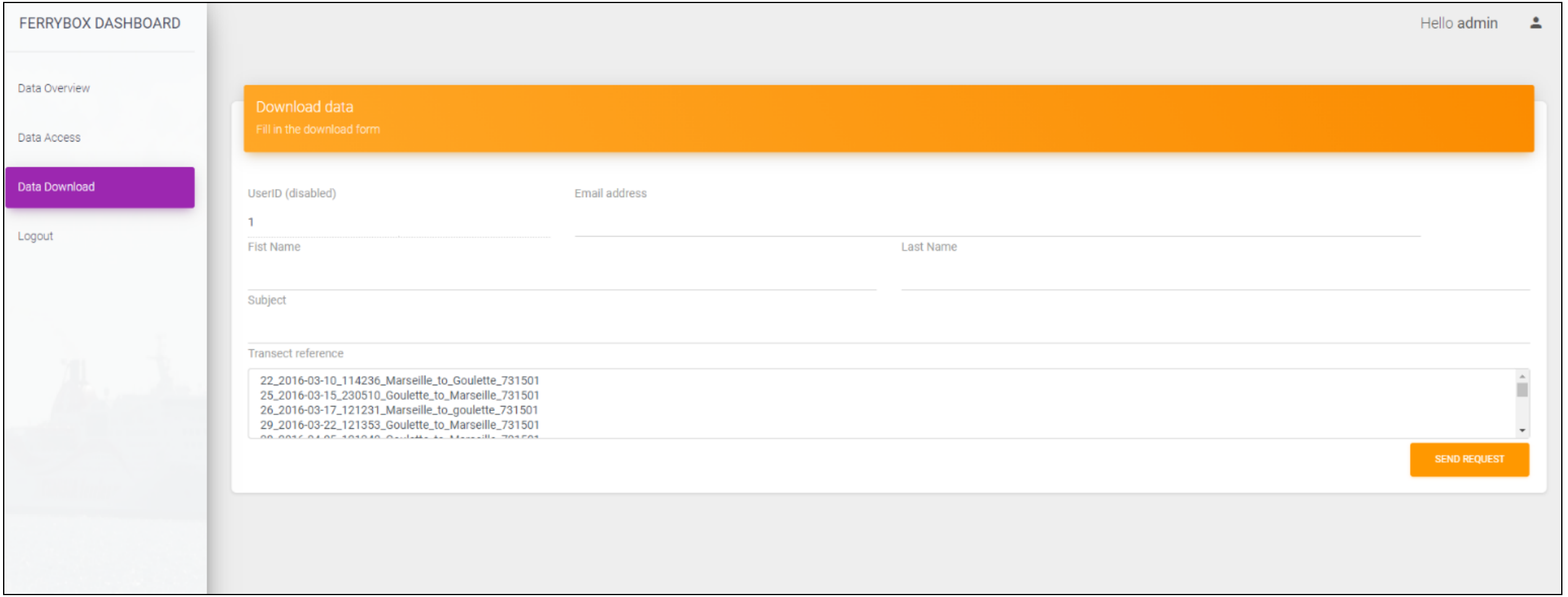
PLOT



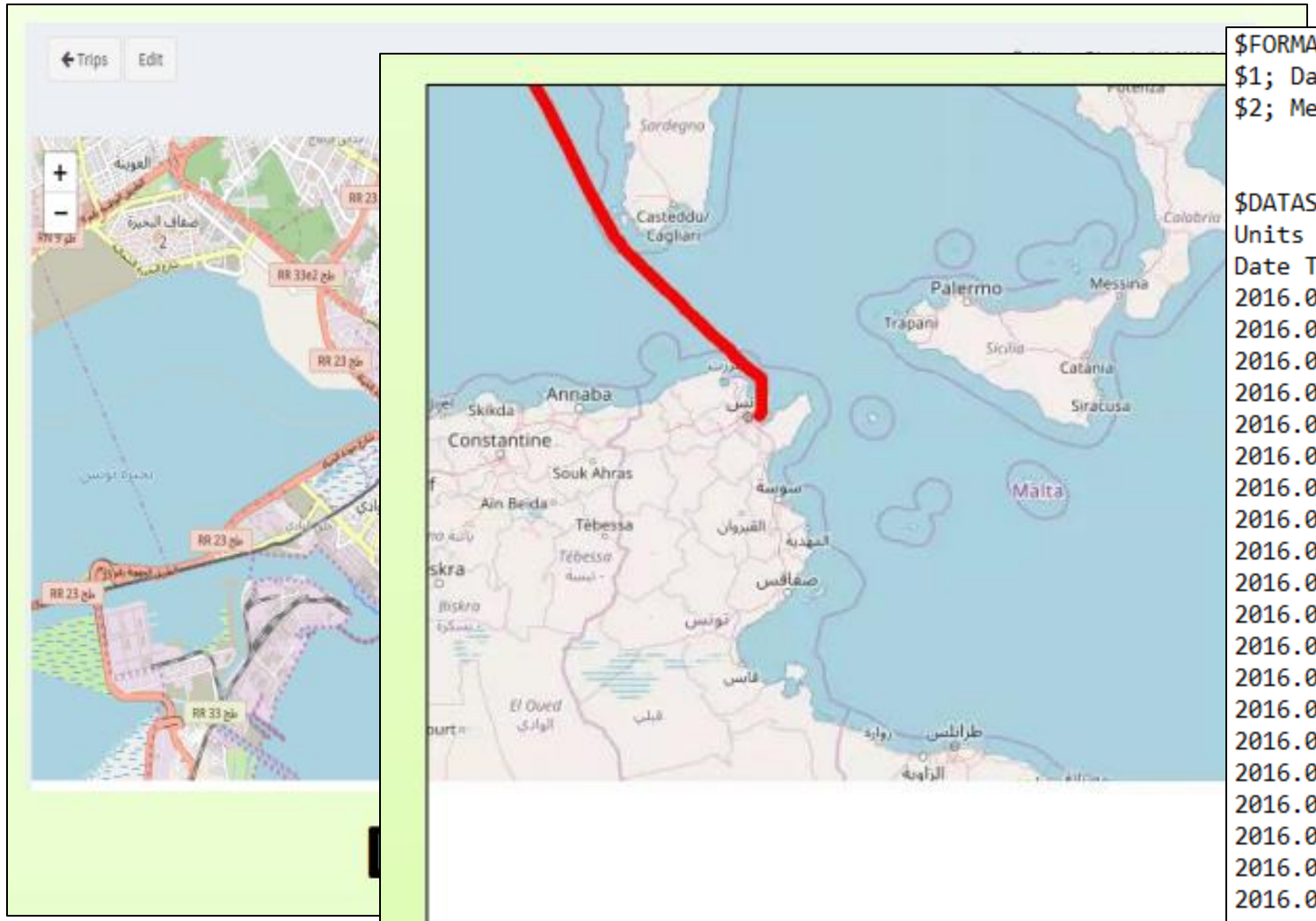

FERRYBOX DASHBOARD
Hello admin
Data Overview
Data Access
Data Download
Logout
Graphs:
TRANSECT
MAP
TIME SERIES
SCATTER
Transect reference
22_2016-03-10_114236_Marseille_to_Goulette_731501
Parameter
Salinity_SBE45
Xaxis Label
Distance
Quality Control
1 - Good
PLOT
Trip reference : 22
Salinity_SBE45 (PSU)
38.75
38.5
38.25
38
37.75
37.5
37.25
37
36.75
576.6
Salinity_SBE45 (PSU): 38.0
1.3
70.4
138.8
208
279.1
351.8
423.2
492.9
563.7
635.3
712.1
781.5
849.4
Gulf of Tunis
Sardinia canal
West of Sardinia
North of PLB
Highcharts.com


FERRYBOX DASHBOARD
Data Overview
Data Access
Data Download
Logout
Hello admin
Leaflet | Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community
Tunis - Marseille: 2016-01-01 to 2020-10-02
Salinity_SBE45 (PSU)
2017-03-30
Average : 37.3
Maximum: 37.3
Minimum: 37.3
Average
Maximum
Minimum
Highcharts.com


FERRYBOX DASHBOARD
Hello admin
Data Overview
Data Access
Data Download
Logout
Graphs:
TRANSECT
MAP
TIME SERIES
SCATTER
Transect
Tunis - Marseille
Parameter
Salinity_SBE45
Start date
01/01/2016
End date
02/10/2020
Latitude
38.976492485539396
Longitude
10.228271484374998
Size
2
Quality Control
0 - None
PLOT
Leaflet | Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community


FERRYBOX DASHBOARD
Hello admin
Data Overview
Data Access
Data Download
Logout
Download data
Fill in the download form
UserID (disabled)
1
Email address
Fist Name
Last Name
Subject
Transect reference
22_2016-03-10_114236_Marseille_to_Goulette_731501
25_2016-03-15_230510_Goulette_to_Marseille_731501
26_2016-03-17_121231_Marseille_to_goulette_731501
29_2016-03-22_121353_Goulette_to_Marseille_731501
SEND REQUEST

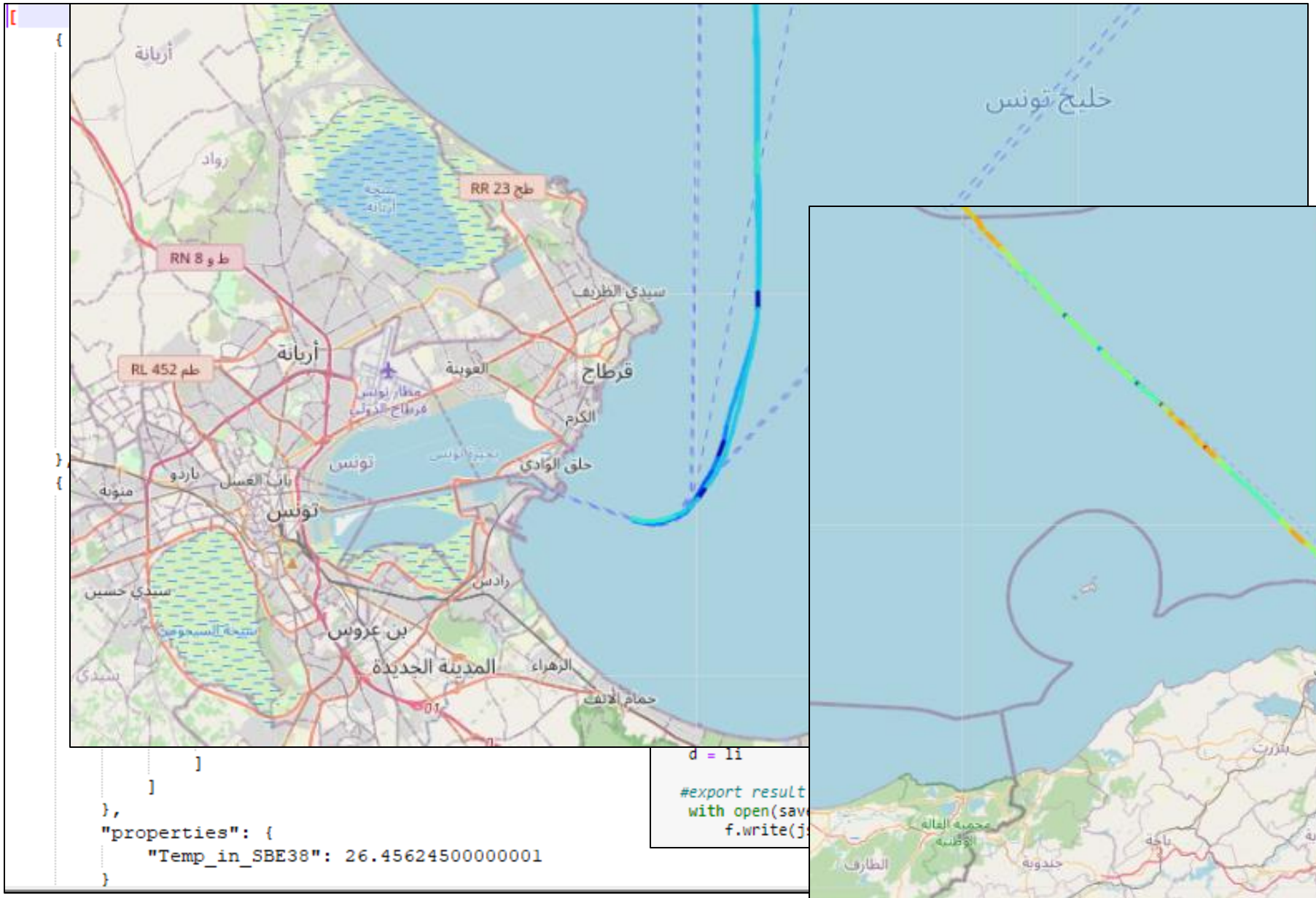
The reference map




```
$FORMATS
$1; DateTime; YYYY.MM.DD hh:mm:ss
$2; Measdata; Float


$DATASETS
Units    °          °          °                 Nm                °C
Date Time           Latitude   Longitude         Course            Variance
2016.03.15 23:08:00 39.244732  8.001719          330.742558
2016.03.15 23:09:00 39.250258  7.997810          331.341218
2016.03.15 23:10:00 39.255713  7.993902          330.775581
2016.03.15 23:11:00 39.261162  7.990056          331.871562
2016.03.15 23:12:00 39.266631  7.986307          331.469930
2016.03.15 23:13:00 39.272143  7.982461          331.521911
2016.03.15 23:14:00 39.277615  7.978658          332.332251
2016.03.15 23:15:00 39.283139  7.974973          332.384149
2016.03.15 23:16:00 39.288631  7.971216          332.026682
2016.03.15 23:17:00 39.294140  7.967473          332.261717
2016.03.15 23:18:00 39.299617  7.963729          331.514747
2016.03.15 23:19:00 39.305093  7.959935          332.306713
2016.03.15 23:20:00 39.310534  7.956166          331.271163
2016.03.15 23:21:00 39.315982  7.952355          331.770833
2016.03.15 23:22:00 39.321441  7.948591          331.116856
2016.03.15 23:23:00 39.326871  7.944659          330.821233
2016.03.15 23:24:00 39.332283  7.940842          331.269841
2016.03.15 23:25:00 39.337650  7.937012          330.551364
2016.03.15 23:26:00 39.342993  7.933140          331.170295
2016.03.15 23:27:00 39.348352  7.929306          330.492308
2016.03.15 23:28:00 39.353693  7.925423          331.136697
2016.03.15 23:29:00 39.359071  7.921671          331.345805
2016.03.15 23:30:00 39.364431  7.917864          331.361625
2016.03.15 23:31:00 39.369862  7.914098          331.429345
```

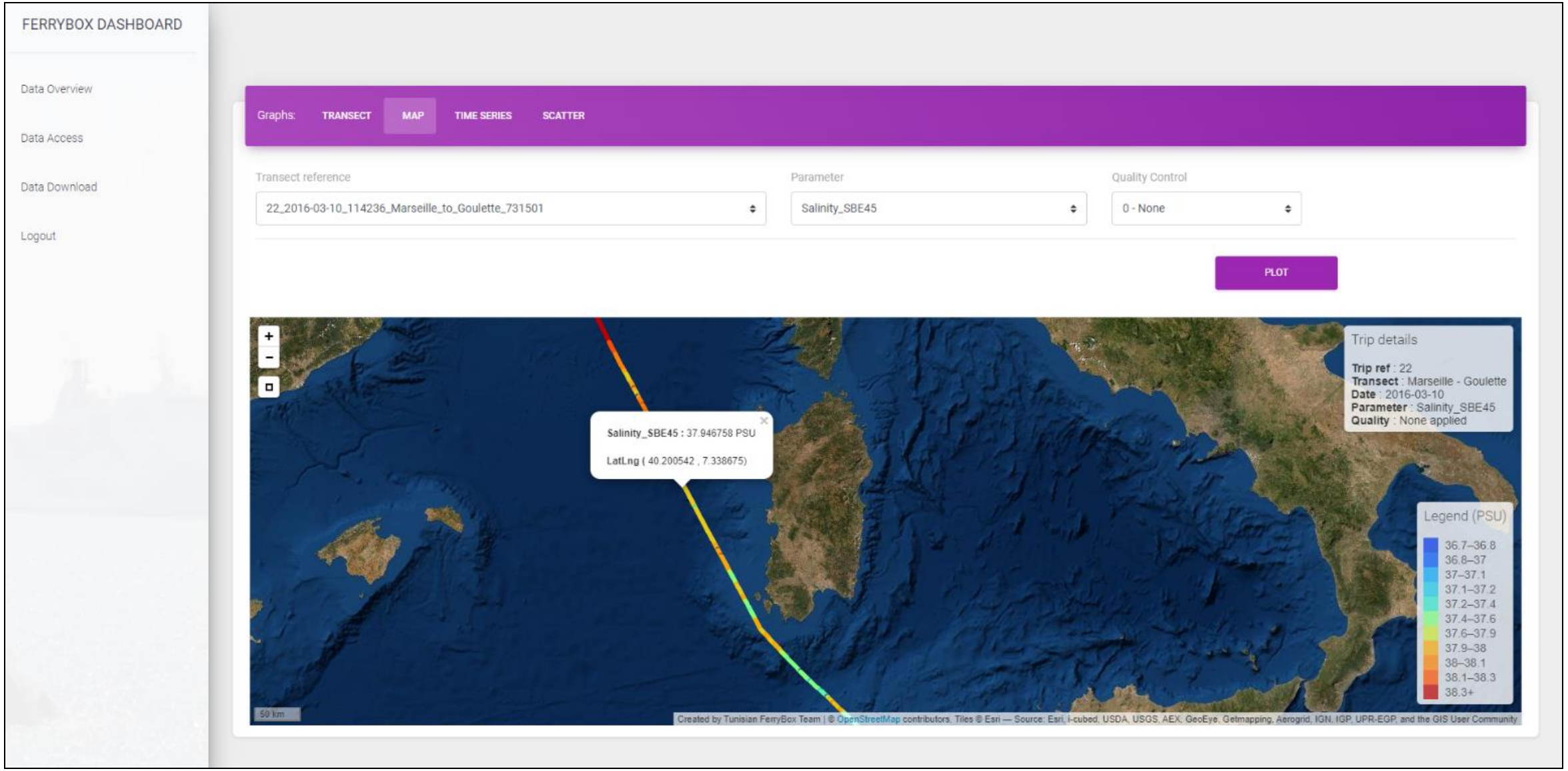
The colored map




```
        ]
      ]
    },
    "properties": {
        "Temp_in_SBE38": 26.45624500000001
    }
```

```
d = li

#export result
with open(save
    f.write(j
```


FERRYBOX DASHBOARD
Data Overview
Data Access
Data Download
Logout
Graphs:
TRANSECT
MAP
TIME SERIES
SCATTER
Transect reference
22_2016-03-10_114236_Marseille_to_Goulette_731501
Parameter
Salinity_SBE45
Quality Control
0 - None
PLOT
Salinity_SBE45 : 37.946758 PSU
LatLng (40.200542 , 7.338675)
Trip details
Trip ref : 22
Transect : Marseille - Goulette
Date : 2016-03-10
Parameter : Salinity_SBE45
Quality : None applied
Legend (PSU)
36.7–36.8
36.8–37
37–37.1
37.1–37.2
37.2–37.4
37.4–37.6
37.6–37.9
37.9–38
38–38.1
38.1–38.3
38.3+
50 km
Created by Tunisian FerryBox Team | © OpenStreetMap contributors. Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community

What’s next ?

December

Scientific article redaction

The FerryBox data should have, at this point of my mission, achieved a level of maturity that allows it to be inserted in some valuable scientific content.

Aout

Further data exploitation / Satellite data matching

The project coordinator may assign further data quality controls, based on the web application behavior and the local refreshment process efficiency. Preparing data for the matching process could be tackled too.

Mai

Finalizing the web application and deployment

The web application, after team meetings and discussions, articles and web redaction, and unit tests validation, can be deployed to international use

Mission’s overview

- ✓ FerryBox data structuring and preprocessing
- ✓ Data analysis related to the project’s tasks and goals
- ✓ Writing reports and scientific articles
- ✓ Participation in the various scientific events of the project

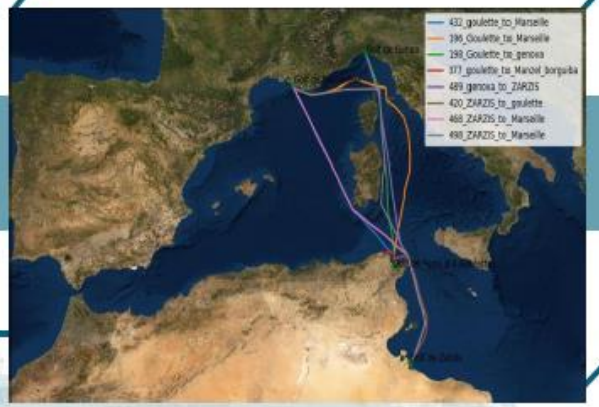
Brochure

Tunisian FerryBox Dashboard

Very abundant information about the Mediterranean water surface

FerryBox

The measuring device for marine water characteristics is fixed at the water depth of 5 meters, collecting data each 1 minute of the ferry trip



Main transects

The two mainly visited transects are :

- **Tunis - Genova**
- **Tunis - Marseille**

Other golfs are randomly visited such as Zarzis, Lyon...

FERRYBOX TECHNOLOGY

The Tunisian FerryBox device is part of an initiative that's targeting water masse tracking and measuring. It is materialized by a set of sensors that are implemented in Carthage ferry, at 5 meters depth. It is measuring, for each minute of the ferry's trip, several parameters (Temperature, Salinity, Dissolved oxygen, Turbidity, pH...)

AUTO-SAMPLER

Auto-sampler is a new device that was joined to FerryBox equipment on 2020, it's designed to automate marine water sampling .

AADI Oxygen op-

Seapoint Turbidity

Seapoint Chloro-

SBE45 MicroTSG Thermosalinograph

SBE38 Digital Oceanographic Thermometer

The first launch of FerryBox data collecting campaigns was on 2016. The growing database offers several interesting scientific possibilities:

- Statistical studies
- Comparison with satellite data
- Insight into the Mediterranean marine dynamics…

The Tunisian FerryBox dashboard renders the required graphics and maps, based on a user's instant demand. The web application is totally dynamic and is exporting the required data to visualize it, in real time.

The application's user is offered the possibility to visualize:

- **Transects plot**, for each parameter
- **Map plot**, with a colored transect
- **Scatter plot**, with a dynamic quality control
- **Time series plot**, on each position of the ferry trip
- **Data description** ...

CARTHAGE FERRY

The FerryBox sensors are being hosted in the Tunisian ferry Carthage, from CTN, since February 2016.

INSTM

DATA REFRESHMENT PROCESS

1 Files management

Gathering, downloading, automating the classification and the protection of raw data files

2 Data analysis

Indexing files, pretreatments, quality control and transforming data into new more exploitable forms (general statistics, time series…),

3 Database creation

Creation of PostgreSQL relational database and automating the daily data insertion in the database

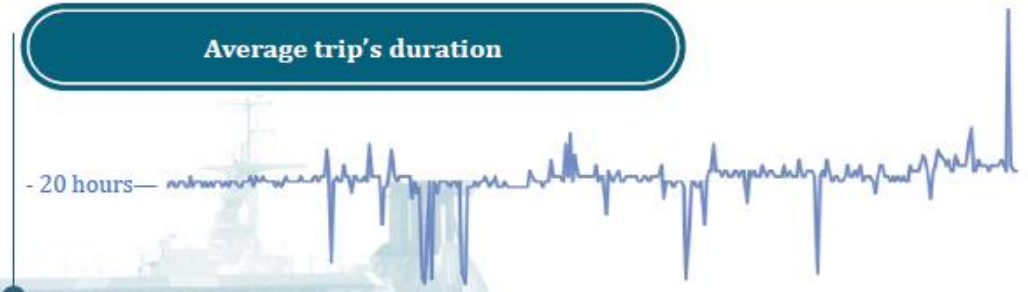
4 Web visualization

Displaying scientific data in an interactive application based on : **Django framework, Python and Javascript :**

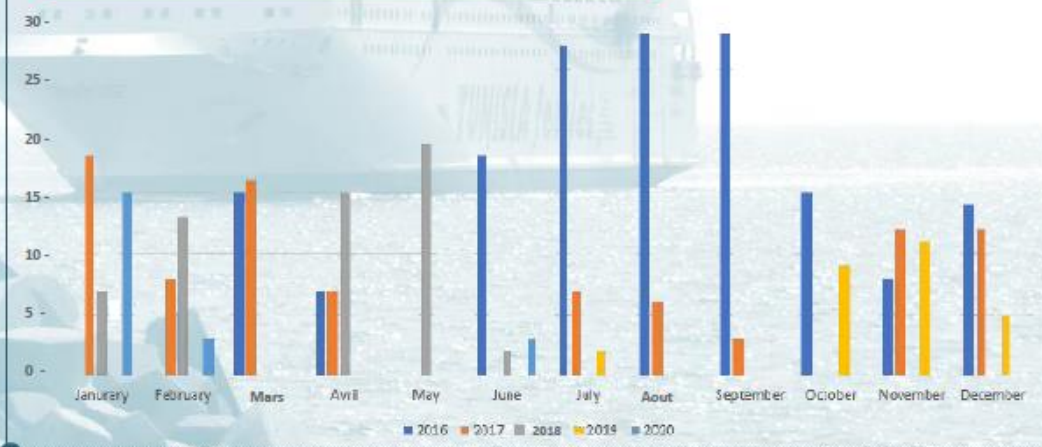
Tunisian FerryBox Dashboard

DATABASE STATISTICS

Average trip's duration



Data's repartition per month / per year

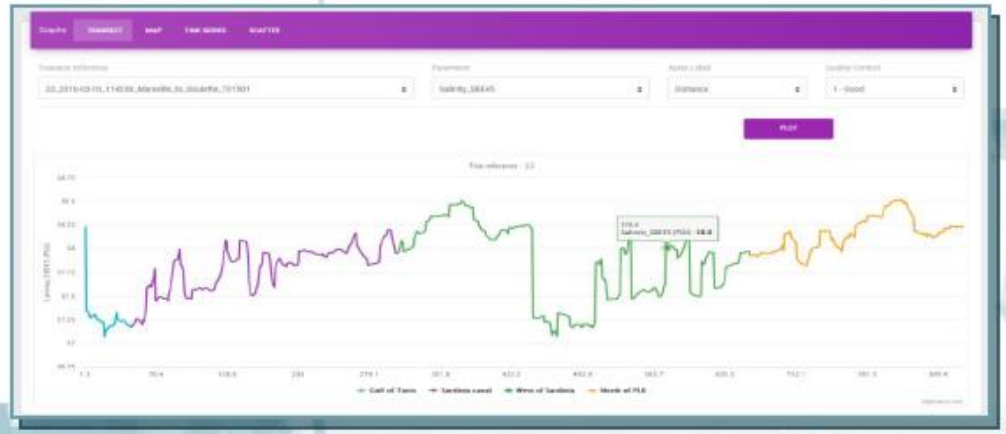


Data's repartition per transect

DYNAMIC WEB APPLICATION : DATA ACCESS

TRANSECT PLOT

Plot several parameter's distribution along the ferry trip regions (Temperature, Salinity, Chlorophyll, Turbidity…), with several quality control options

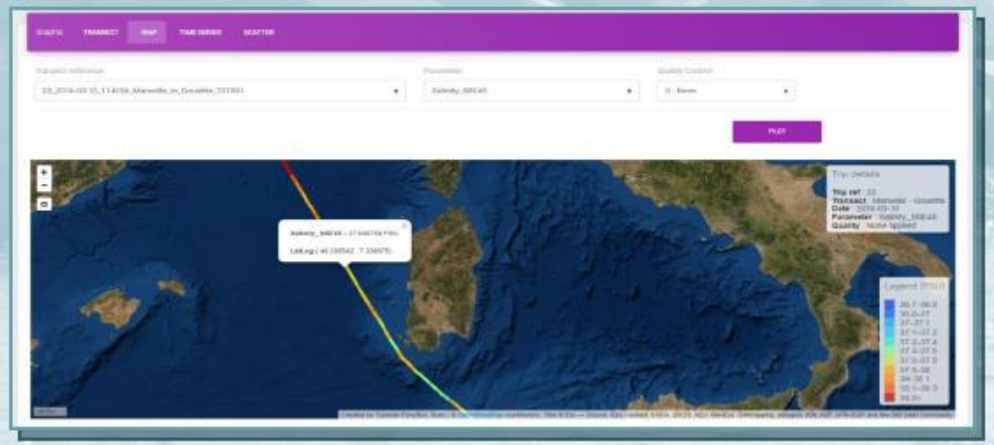


TIME SERIES

Visualize time series fluctuations on any given point of FerryBox trips, and any time range since 2016 (max, min and average values)

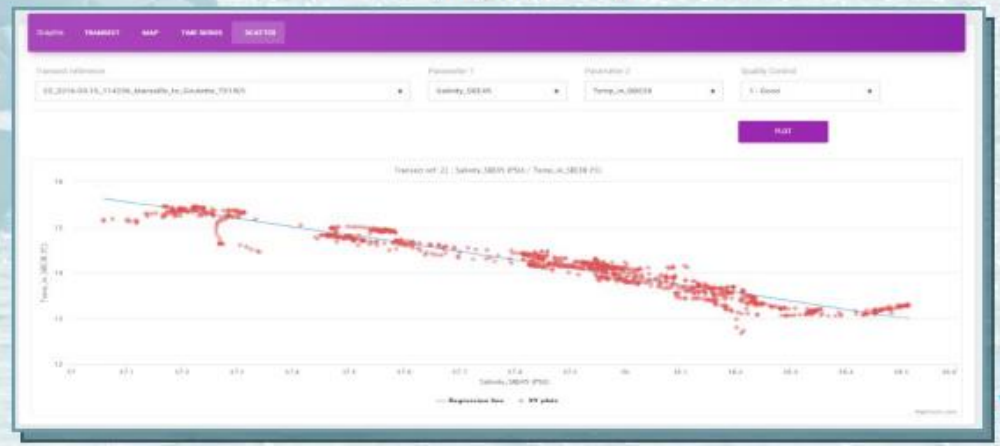
MAP PLOT

Draw the ferry's transect on a georeferenced map, with colored information about scientific data repartition (example below: Salinity



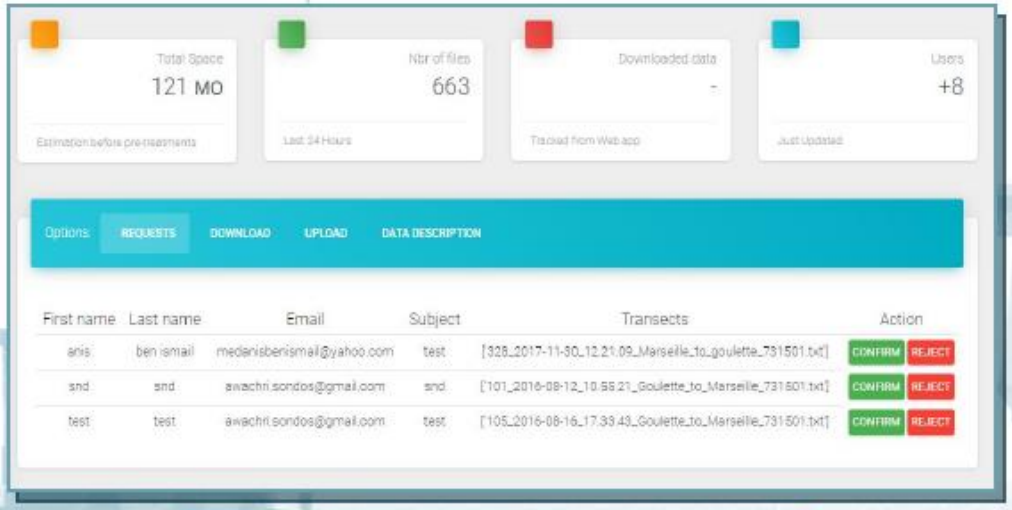
SCATTER PLOT

Draw all possible combinations of FerryBox parameters' Correlations (XY plots) and regression line

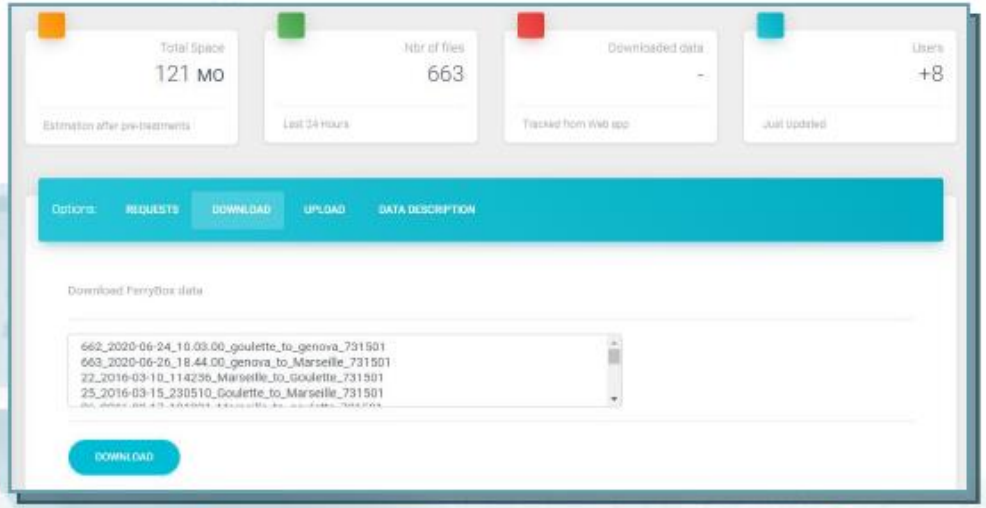


DYNAMIC WEB APPLICATION : ADMIN INTERFACE

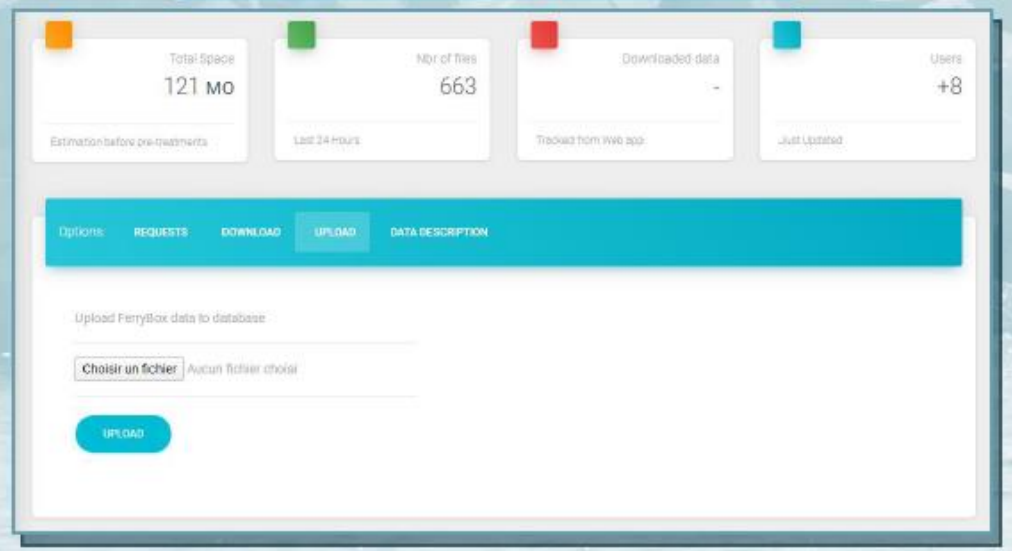
REQUESTS MANAGEMENT



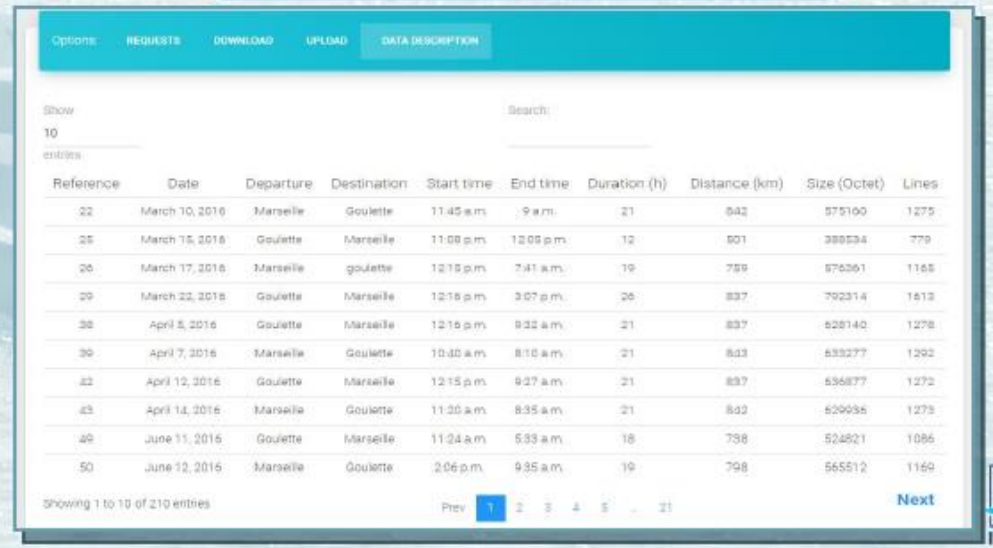
DOWNLOAD DATA



UPLOAD DATA



DATA STATISTICS



INSTM