Predictive analytics system documentation

Version 4

17/12/2017

Acciona agua www.acciona-agua.com	Alejandro Beivide (alejandro.beivide.garcia@acciona.com) Frank Zamora (frank.zamora.avendano@acciona.com) Clara Sanroma (clara.sanroma.flores@acciona.com) Elena Crespo (elena.crespo.olazabal@acciona.com) Cosmin Koch (ckoch@acciona.com)
Artelnics www.artelnics.com	Roberto Lopez (robertolopez@artelnics.com) Fernando Gomez (fernandogomez@artelnics.com) Emilio Lucas (emiliolucas@artelnics.com)

Abstract

This document collects a list of the variables used in this first study and the functions and interfaces of the API.

We will use two satellites for the acquisition of the data: the MODIS Aqua, from the mission MODIS of the NASA, and the Sentinel-3, from the Copernicus mission of the European Commission.

The API will be made of three modules.

The Data module is going to send requests to the satellite servers for retrieving the data.

The Model module trains the neural network and stores the model for later evaluations.

The Query module returns the data of the satellite data for some day or the evaluation of the model for a given day.

Index

1. Data acquisition 6
1.1. Satellites 6
1.1.1. MODIS Aqua 6
1.1.2. Sentinel 9
1.1.3. CSFR 11
1.1.4. ECMWF 11
1.2. XML format 12
1.2.1. Data acquisition fields 12
1.3. Data acquisition example file 13
1.4. API documentation 14
1.5. Data acquisition API 14
1.6. Predictive model API 15
1.7. Output calculation API 15
2. Data preparation 17
2.1 Process description 17
2.2 Data sets 18
2.3 Missing values 19
2.3.1 Plant data 19
2.3.2 Satellite data 20
2.4 Outliers 27
2.4.1 Plant data 27
2.4.2 Satellite data 30
2.5 Data imputation 31
2.5.1 Strong presence 31
2.5.2 Weak presence 32
2.6 Variables selection 33
2.6.1 Modis Aqua 33
2.6.2 Sentinel 34
3. Time series 35
3.1 Plant data 35
3.2 Satellite data (entry point) 41
3.3 Satellite data (maximum depth) 45
4. Descriptive analytics 46
4.1 Plant data statistics 46

4.2 Satellite data statistics (entry point) 47
4.3 Satellite data statistics (maximum depth) 49
4.4 Satellite data distribution (entry point) 50
4.5 Satellite data distribution (maximum depth) 53
5. Diagnostic analytics 54
5.1 Plant scatter charts 54
5.1.1 Conductivity – Temperature 54
5.1.2 Turbidity – Temperature 55
5.1.3 Turbidity - Conductivity 56
5.2 Plant correlations 57
5.2.1 Conductivity 58
5.1.2 Turbidity 59
5.2 Satellite scatter charts (entry point) 60
5.2.1 Northward velocity – Eastward velocity 60
5.2.2 Photosynthetically Available Radiation - Temperature 61
5.2.3 Salinity - Temperature 62
5.2.4 Salinity – Sea Surface Height 63
5.3 Satellite correlations (entry point) 64
5.4 Satellite scatter charts (maximum depth) 65
5.5 Satellite correlations (maximum depth) 65
5.6 Plant-satellite correlations (entry point) 66
5.6.1 Temperature 66
5.6.2 Conductivity 66
5.6.3 Turbidity 66
5.7 Plant-satellite correlations (defined regions) 67
1.7.1. Next steps 68
1.7.2. Turbidity problem 70
6. Predictive analytics ..... **¡Error! Marcador no definido.**
7. Predictive analytics 74
2.7 Temperature model 74
8. Front-end prototype 75
7.1 Design concepts 75
Objeto 75
General issues 75
End user wireframe 78
Administrator wireframe 81

9. Tools and libraries......87

1. Data acquisition

The variables used for the prediction come from two sources:

1- Satellite data
2- Internal data

The satellite data will be downloaded from the data source of each one.

The internal data will come from the plant as a comma separated values file with the data obtained with their measurement elements.

We will use several satellites with different data in each one.

The data coming from the satellite will represent the images as physical, biological or atmospheric variables.

1.1.Satellites

1.1.1. MODIS Aqua

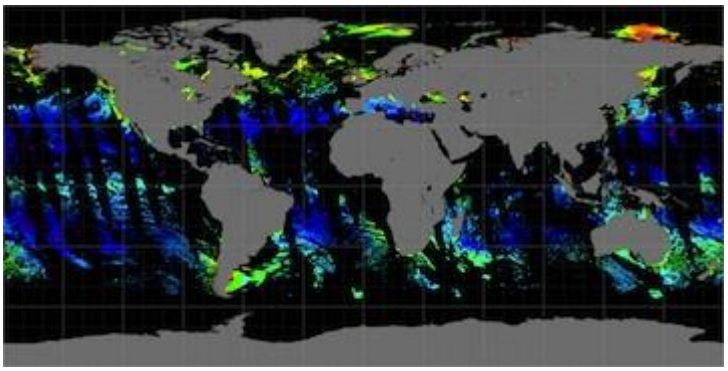
The MODIS Aqua is a satellite property of the NASA. It collects data from the ocean several biological and physical variables.

https://oceancolor.gsfc.nasa.gov/

The acquisition of data from MODIS aqua is performed with the application wget.

We are using the level 3 mapped data of the satellite. This data consists in separated NETCDF data files for each day and variable.

The satellite retrieve data daily. It gets the data while orbiting the globe. This cause that the data could have zones without valid data.



The next table shows the possible variables that the satellite offers.

The yellow variables are the same information with different algorithms. A correlation study must be done to determine which one is the best.

Variable	Description	Units
chl_oc3	Chlorophyll Concentration, OC3 algorithm	mg/m^3
chlor_a	Chlorophyll Concentration, OCI algorithm	mg/m^3
Ipar	Instantaneous Photosynthetically Available Radiation	$einstein/m^2\ s$
Nflh	Normalized fluorescence line height	$W/m^2\ um^1\ sr^1$
adg_443_gsm	Absorption due to gelbstoff and detrital material (443nm), GSM algorithm	1/m
bbp_443_gsm	Particulate backscattering (443nm), GSM algorithm	1/m
chl_gsm	Chlorophyll Concentration, GSM model	mg/m^3
a_412_giop	Total absorption (412nm), GIOP model	1/m
a_443_giop	Total absorption (443nm), GIOP model	1/m
a_469_giop	Total absorption (469nm), GIOP model	1/m
a_488_giop	Total absorption (488nm), GIOP model	1/m
a_531_giop	Total absorption (531nm), GIOP model	1/m
a_547_giop	Total absorption (547nm), GIOP model	1/m
a_555_giop	Total absorption (555nm), GIOP model	1/m
a_645_giop	Total absorption (645nm), GIOP model	1/m

a_667_giop	Total absorption (667nm), GIOP model	1/m
a_678_giop	Total absorption (678nm), GIOP model	1/m
adg_443_giop	Absorption due to gelbstoff and detrital material (443nm), GIOP model	1/m
adg_s_giop	Detrital and gelbstoff absorption spectral parameter for GIOP model	$1/m^1$ nm^1
adg_unc_443_giop	Uncertainty in absorption due to gelbstoff and detrital material (443nm), GIOP model	1/m
aph_443_giop	Absorption due to phytoplankton (443nm), GIOP model	1/m
aph_unc_443_giop	Uncertainty in absorption due to phytoplankton (443nm), GIOP model	1/m
bb_412_giop	Total backscattering (412nm), GIOP model	1/m
bb_443_giop	Total backscattering (443nm), GIOP model	1/m
bb_469_giop	Total backscattering (469nm), GIOP model	1/m
bb_488_giop	Total backscattering (488nm), GIOP model	1/m
bb_531_giop	Total backscattering (531nm), GIOP model	1/m
bb_547_giop	Total backscattering (547nm), GIOP model	1/m
bb_555_giop	Total backscattering (555nm), GIOP model	1/m
bb_645_giop	Total backscattering (645nm), GIOP model	1/m
bb_667_giop	Total backscattering (667nm), GIOP model	1/m
bb_678_giop	Total backscattering (678nm), GIOP model	1/m
bbp_443_giop	Particulate backscattering (443nm), GIOP model	1/m
bbp_s_giop	Backscattering spectral parameter for GIOP model	$1/m^1$ nm^1
bbp_unc_443_giop	Uncertainty in particulate backscatter (443nm), GIOP model	1/m
Kd_490	Diffuse attenuation coefficient (490nm), KD2 algorithm	1/m
Ndvi	Normalized difference vegetation index	---
Sst	Sea surface temperature	ºC
Par	Photosynthetically available radiation	$einstein/m^2$
Pic	Calcite concentration	mol/m^3
Poc	Particle organic carbon	mg/m^3
a_443_qaa	Total absorption (443nm), QAA algorithm	1/m
adg_443_qaa	Absorption due to gelbstoff and detrital material (443nm), QAA algorithm	1/m
aph_443_qaa	Absorption due to phytoplankton (443nm), QAA algorithm	1/m
bbp_443_qaa	Particulate backscattering (443nm), QAA algorithm	1/m
angstrom	Aerosol angstrom exponent	---
aot_869	Aerosol optical thickness (869nm)	---
Rrs_412	Remote sensing reflectance (412nm)	1/sr
Rrs_443	Remote sensing reflectance (443nm)	1/sr
Rrs_469	Remote sensing reflectance (469nm)	1/sr
Rrs_488	Remote sensing reflectance (488nm)	1/sr
Rrs_531	Remote sensing reflectance (531nm)	1/sr
Rrs_547	Remote sensing reflectance (547nm)	1/sr
Rrs_555	Remote sensing reflectance (555nm)	1/sr
Rrs_645	Remote sensing reflectance (645nm)	1/sr
Rrs_667	Remote sensing reflectance (667nm)	1/sr

Rrs_678	Remote sensing reflectance (678nm)	1/sr
sst4	4um sea surface temperature	ºC
Zeu_lee	Euphotic depth, Lee algorithm	M

1.1.2. Sentinel

The Sentinel-3 satellite belongs to the European programme Copernicus. It provides data of concentration of several components of the water and information of the sea and the currents.

http://marine.copernicus.eu/

The acquisition of data from Sentinel aqua is performed using a.dedicated Python API.

The access to the data must be done via a Python script in the server. Also, this satellite requires an account to retrieve the data.

This variables of this satellite are separated in physical and biological variables.

Physical variables

The physical variables come from the status of the sea itself.

Variable	Description	Units
usi	Sea ice eastward velocity	m/s
so	Salinity	practical salinity unit
Sithick	Sea ice thickness	M
Thetao	Temperature	ºC
Siconc	Ice concentration	fraction
Uo	Eastward velocity	m/s
Depth	Depth	m
Vo	Northward velocity	m/s
Mlotst	Density ocean mixed layer thickness	m
Zos	Sea surface height	M
Vsi	Sea ice northward velocity	m/s
bottomT	Sea floor potential temperature	ºC

Biological variables

The biological variables represent the concentration of different elements in the sea.

Weekly variables

Variable	Description	Units
Fe	Mole concentration of dissolved iron	$mmol/m^3$
PP	Net primary productivity of carbon	$g/m^3/day$
Si	Mole concentration of silicate	umol/L
NO3	Mole concentration of nitrate	$mmol/m^3$
CHL	Mass concentration of chlorophyll	mg/m^3
PHYC	Mole concentration of phytoplankton expressed as carbon	$mmol/m^3$
PO4	Mole concentration of phosphate	$mmol/m^3$
O2	Mole concentration of dissolved oxygen	$mmol/m^3$

Daily variables

Variable	Description	Units
CHL	Mass concentration of chlorophyll	mg/m^3
SPM*	Inorganic suspended particle matter	g/m^3

(*) Data from 1997-09-04 to 2016-08-31 and from 2017-07-06 to present.

1.1.3. CSFR

The acquisition of data from MODIS aqua is performed with the application wget and an authentication file.

1.1.4. ECMWF

The acquisition of data from MODIS aqua is performed with a Python script.

It needs an annual subscription.

1.2.XML format

1.2.1. Data acquisition fields

Python path: Path to the Python folder.

MODIS Aqua folder path: Folder to store the data from the MODIS Aqua satellite

Sentinel folder path: Folder to store the data from the Sentinel satellite

CSFR folder path: Folder to store the data from the CSFR satellite

Ecmwf folder path: Folder to store the data from the Ecmwf satellite

User Sentinel: User required for the retrieving of the data from Sentinel.

Password Sentinel: Password of the Sentinel user.

Longitudes: Minimum and maximum longitudes of the zone where to acquire the data.

Latitudes: Minimum and maximum latitudes of the zone where to acquire the data.

Date: Last date for the data required.

Days: Number of days for the acquisition of the data.

MODIS variables: List of the names of the variables required for the MODIS Aqua satellite.

Sentinel variables: List of the names of the variables required for the Sentinel satellite.

Ecmwf variables: List of the names of the variables required for the Ecmwf satellite.

1.3. Data acquisition example file

@todo Insert new format

1.4. API documentation

The API for the server will be separated in three different modules:

1- Data module
2- Model module
3- Query module

1.5. Data acquisition API

This module will be in charge of getting the data from the satellite and the plant. I will also prepare the data file for the creation of the model.

bool satelliteDataCapture (string plantName, string XMLFilePath) Downloads the satellite data in the given region. **Parameters:** string plantName : name of the plant where the data will be acquired. string XMLFilePath: XML file with the name of the desired variables, region and dates. **Return:** true if successful, false otherwise.
bool documentDataCapture (string plantName, string dataFileName) Gets the internal data of the plant. **Parameters:** string plantName : name of the plant where the data come from. string dataFileName: path of the data file obtained in the plant. **Return:** true if successful, false otherwise.
bool prepareData (string plantName) Prepare the data to be readable by the model module. **Parameters:** string plantName : name of the plant where the data come from. **Return:** true if successful, false otherwise.

1.6. Predictive model API

This module will create and train the model of the prediction.

bool trainModel (string plantName, string variables[], int daysAhead)
Starts the training of the model with the given values.

Parameters:
string plantName : name of the plant of the model to be trained.
string variables : vector of variables used for the training. If it only contains the value “all”, the model will be trained with all the available data.
int daysAhead: number of days ahead that the model will predict.

Return: true if successful, false otherwise.

1.7. Output calculation API

This module will be used to the exchange of information of the satellite and the model between the client and the server.

string heatMap (string plantName, string variable)
Gets a variable data for a plant to be plotted as heat map.

Parameters:
string plantName : name of the plant where the map will be plotted.
string variable : variable to be plotted.

Return: JSON string with the coordinates and values of the variable in the points. If the function fails, it will return the error cause.

string getGraphicData (string plantName, string variable, string date)
Returns the previous data and prediction of a variable in a date.

Parameters:
string plantName : name of the plant where the prediction is done.
string variable : name of the variable to be predicted.
string date : date of prediction.

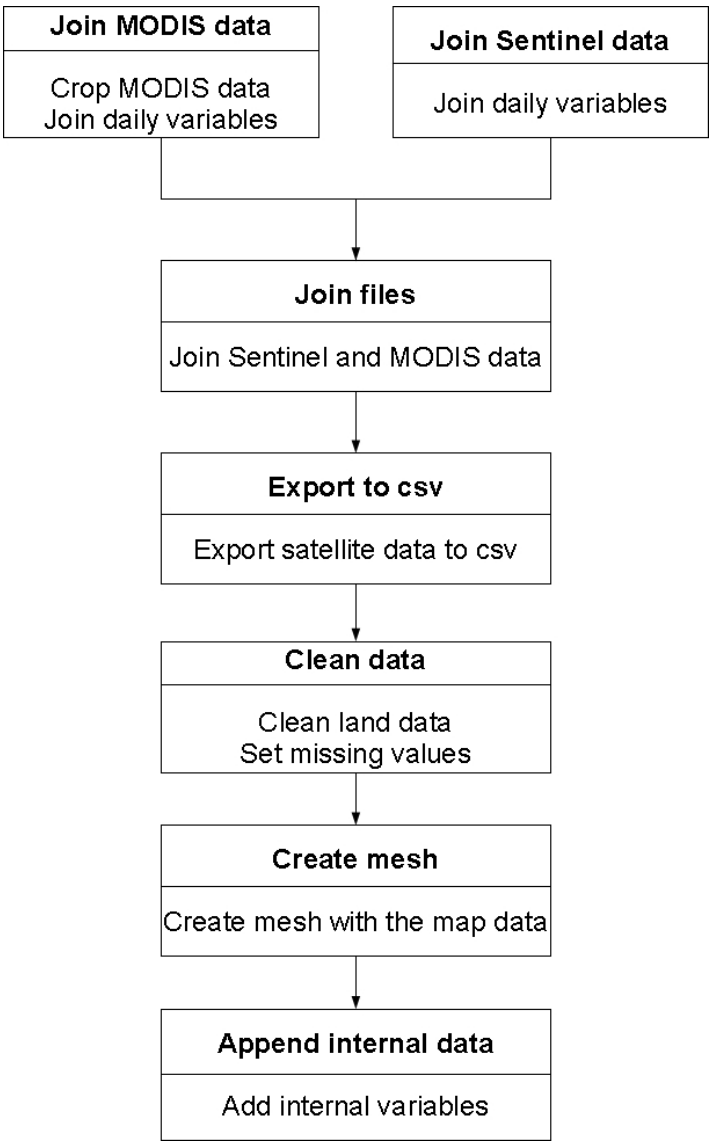
Return: JSON string with the values of the precious days and the prediction, it also returns the average error for that prediction. If the function fails, it will return the error cause.

2. Data preparation

2.1 Process description

The data preparation module joins satellites data, cleans the data and appends the internal data with the desired lag days.

This is the diagram flow of the module.



2.2 Data sets

Plant data	
Total data:	435.665
Number of days:	304
Number of variables	5

Satellite data	
Total data:	54.777.456
Number of days:	304
Number of variables	70
Coordinate points	180.189

The satellite data available is enormous. A clean of the data and a meshing of the map must be done.

2.3 Missing values

2.3.1 Plant data

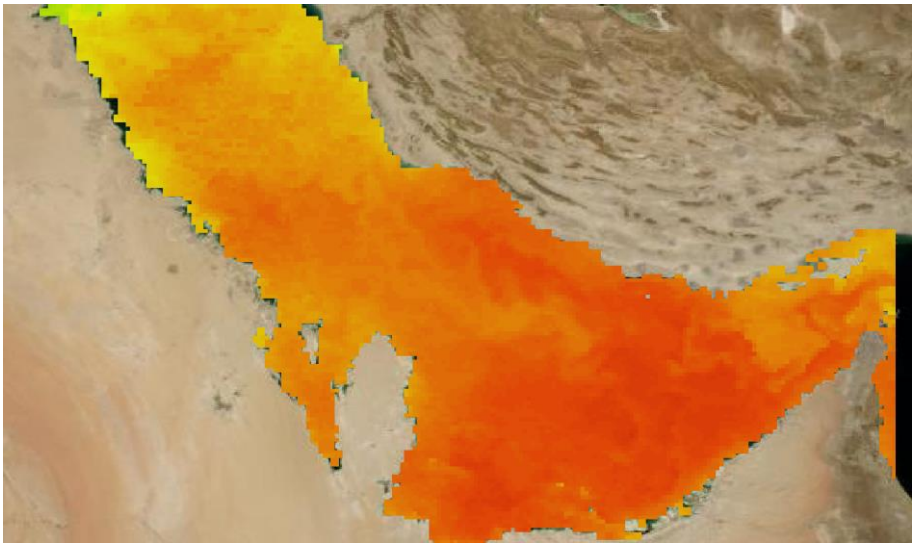
Variable	Missing	Percentage
Temperature 1	54	0.01%
Temperature 2	54	0.01%
Conductivity 1	54	0.01%
Conductivity 2	54	0.01%
Turbidity	62	0.01%

The internal data contains very little missing values.

2.3.2 Satellite data

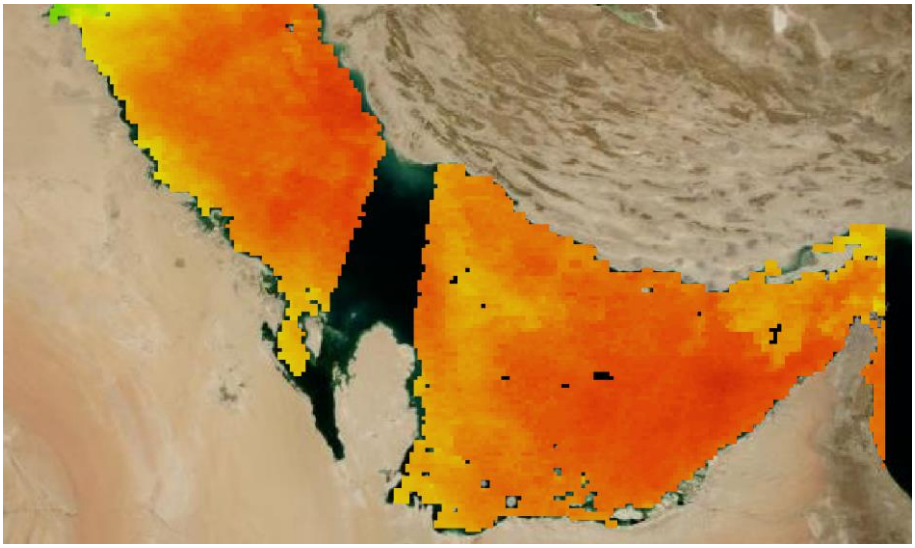
The satellite data have several problems with the missing values.

The perfect case is when we have all the data in the map.

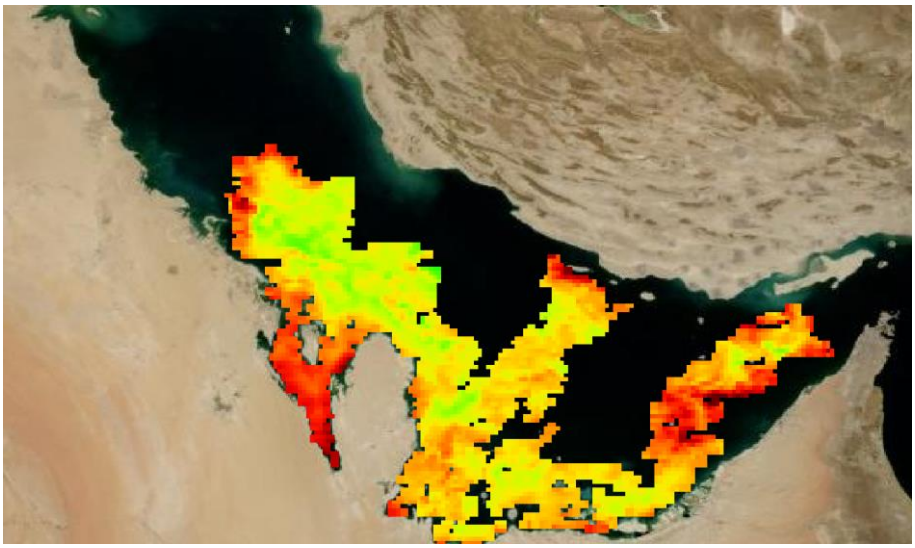


But we can have various cases where the satellite does not measure all the map perfectly.

The first one is caused because of the passes of the satellite around the globe.



The cloudy weather is also an issue for the measure of some variables.



Finally, the satellite can give us an empty map full of missing values. This is the worst case.

Rank	Variable	Description	Satellite	Missing	Percentage
1	chl_gsm	Chlorophyll (GSM model)	MODIS	53762955	98.1%
2	bbp_s_giop	Backscattering spectral parameter	MODIS	53509098	97.7%
3	adg_s_giop	Detrital and gelbstoff absorption	MODIS	53509097	97.7%
4	CHL	Chlorophyll	Sentinel	50777810	92.7%
5	chlor_a	Chlorophyll (OCI algorithm)	MODIS	48683235	88.9%
6	chl_oc3	Chlorophyll (OC3 algorithm)	MODIS	48683235	88.9%
7	sst	Temperature	MODIS	22777518	41.6%
8	bottomT	Floor temperature	Sentinel	8946838	16.3%
9	thetao	Temperature	Sentinel	8946838	16.3%
10	so	Salinity	Sentinel	8946838	16.3%
11	vsi	Sea ice northward velocity	Sentinel	593	0.001%
12	zos	Sea surface height	Sentinel	592	0.001%
13	mlotst	Density ocean mixed layer thickness	Sentinel	589	0.001%
14	vo	Northward velocity	Sentinel	585	0.001%
15	uo	Eastward velocity	Sentinel	581	0.001%
16	siconc	Ice concentration	Sentinel	576	0.001%
17	sithick	Sea ice thickness	Sentinel	568	0.001%
18	usi	Sea ice eastward velocity	Sentinel	565	0.001%
19	Zeu_lee	Euphotic depth	MODIS	562	0.001%
20	sst4	4um sea surface temperature	MODIS	557	0.001%
21	Rrs_678	Remote sensing reflectance (678nm)	MODIS	555	0.001%
22	Rrs_667	Remote sensing reflectance (667nm)	MODIS	551	0.001%
23	Rrs_645	Remote sensing reflectance (645nm)	MODIS	548	0.001%
24	Rrs_547	Remote sensing reflectance (547nm)	MODIS	543	0.001%
25	Rrs_555	Remote sensing reflectance (555nm)	MODIS	543	0.001%
26	Rrs_531	Remote sensing reflectance (531nm)	MODIS	539	0.001%

27	Rrs_488	Remote sensing reflectance (488nm)	MODIS	536	0.001%
28	Rrs_469	Remote sensing reflectance (469nm)	MODIS	532	0.001%
29	Rrs_443	Remote sensing reflectance (443nm)	MODIS	528	0.001%
30	Rrs_412	Remote sensing reflectance (412nm)	MODIS	522	0.001%
31	aot_869	Aerosol optical thickness (869nm)	MODIS	519	0.001%
32	angstrom	Aerosol angstrom exponent	MODIS	515	0.001%
33	bbp_443_qaa	Particulate backscattering (443nm), QAA algorithm	MODIS	510	0.001%
34	aph_443_qaa	Absorption due to phytoplankton (443nm), QAA algorithm	MODIS	507	0.001%
35	adg_443_qaa	Absorption due to gelbstoff and detrital material (443nm), QAA algorithm	MODIS	504	0.001%
36	a_443_qaa	Total absorption (443nm), QAA algorithm	MODIS	503	0.001%
37	poc	Particle organic carbon	MODIS	500	0.001%
38	pic	Calcite concentration	MODIS	496	0.001%
39	par	Photosynthetically available radiation	MODIS	495	0.001%
40	Kd_490	Diffuse attenuation coefficient (490nm), KD2 algorithm	MODIS	481	0.001%
41	bbp_unc_443_giop	Uncertainty in particulate backscatter (443nm), GIOP model	MODIS	478	0.001%
42	bbp_443_giop	Particulate backscattering (443nm), GIOP model	MODIS	469	0.001%
43	bb_678_giop	Total backscattering (678nm), GIOP model	MODIS	466	0.001%
44	bb_667_giop	Total backscattering (667nm), GIOP model	MODIS	464	0.001%

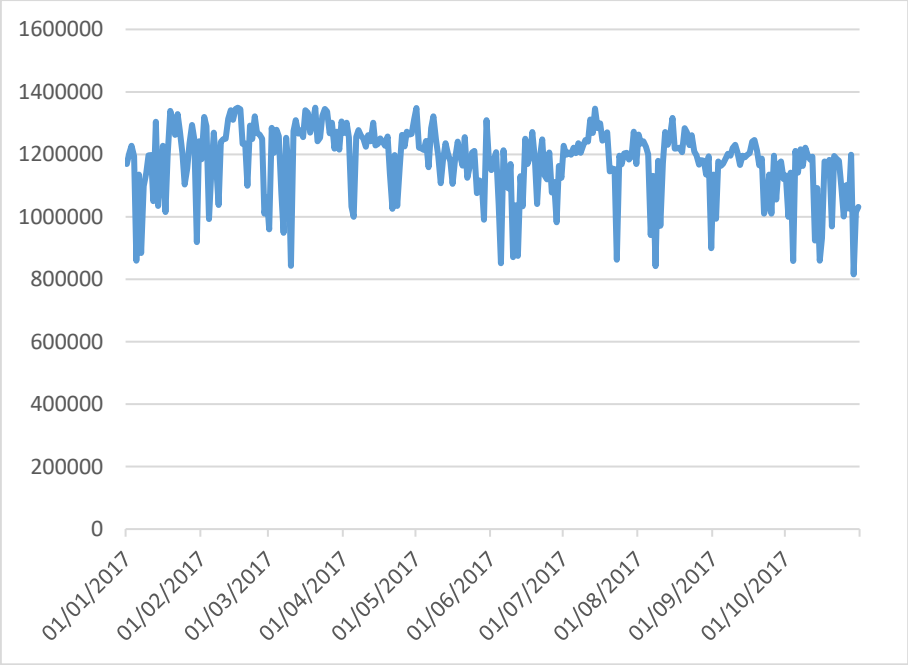
45	bb_645_giop	Total backscattering (645nm), GIOP model	MODIS	458	0.001%
46	bb_555_giop	Total backscattering (555nm), GIOP model	MODIS	449	0.001%
47	bb_547_giop	Total backscattering (547nm), GIOP model	MODIS	442	0.001%
48	bb_531_giop	Total backscattering (531nm), GIOP model	MODIS	439	0.001%
49	bb_488_giop	Total backscattering (488nm), GIOP model	MODIS	436	0.001%
50	bb_469_giop	Total backscattering (469nm), GIOP model	MODIS	430	0.001%
51	bb_412_giop	Total backscattering (412nm), GIOP model	MODIS	423	0.001%
52	bb_443_giop	Total backscattering (443nm), GIOP model	MODIS	423	0.001%
53	aph_unc_443_giop	Uncertainty in absorption due to phytoplankton (443nm), GIOP model	MODIS	416	0.001%
54	aph_443_giop	Absorption due to phytoplankton (443nm), GIOP model	MODIS	412	0.001%
55	adg_unc_443_giop	Uncertainty in absorption due to gelbstoff and detrital material (443nm), GIOP model	MODIS	408	0.001%
56	adg_443_giop	Absorption due to gelbstoff and detrital material (443nm), GIOP model	MODIS	401	0.001%
57	a_678_giop	Total absorption (678nm), GIOP model	MODIS	400	0.001%
58	a_667_giop	Total absorption (667nm), GIOP model	MODIS	395	0.001%
59	a_645_giop	Total absorption (645nm), GIOP model	MODIS	389	0.001%
60	a_555_giop	Total absorption (555nm), GIOP model	MODIS	385	0.001%
61	a_547_giop	Total absorption (547nm), GIOP model	MODIS	382	0.001%
62	a_531_giop	Total absorption (531nm), GIOP model	MODIS	378	0.001%
63	a_488_giop	Total absorption (488nm), GIOP model	MODIS	373	0.001%
64	a_469_giop	Total absorption (469nm), GIOP model	MODIS	370	0.001%

65	a_443_giop	Total absorption (443nm), GIOP model	MODIS	362	0.001%
66	a_412_giop	Total absorption (412nm), GIOP model	MODIS	358	0.001%
67	bbp_443_gsm	Particulate backscattering (443nm), GSM algorithm	MODIS	353	0.001%
68	adg_443_gsm	Absorption due to gelbstoff and detrital material (443nm), GSM algorithm	MODIS	349	0.001%
69	nflh	Normalized fluorescence line height	MODIS	346	0.001%
70	ipar	Instantaneous Photosynthetically Available Radiation	MODIS	340	0.001%

The variable with more missing values is **Chlorophyll**.
The satellite with more missing values is **MODIS**.

The variable with less missing values is **Instantaneous Photosynthetically Available Radiation**.
The satellite with less missing values is **Sentinel**.

Missing values per day

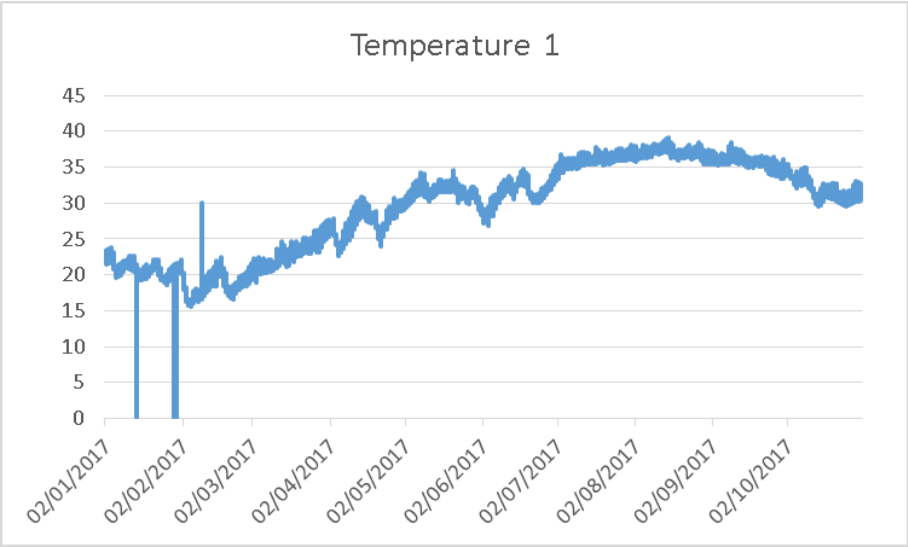


The number of missing values is regular along the time.

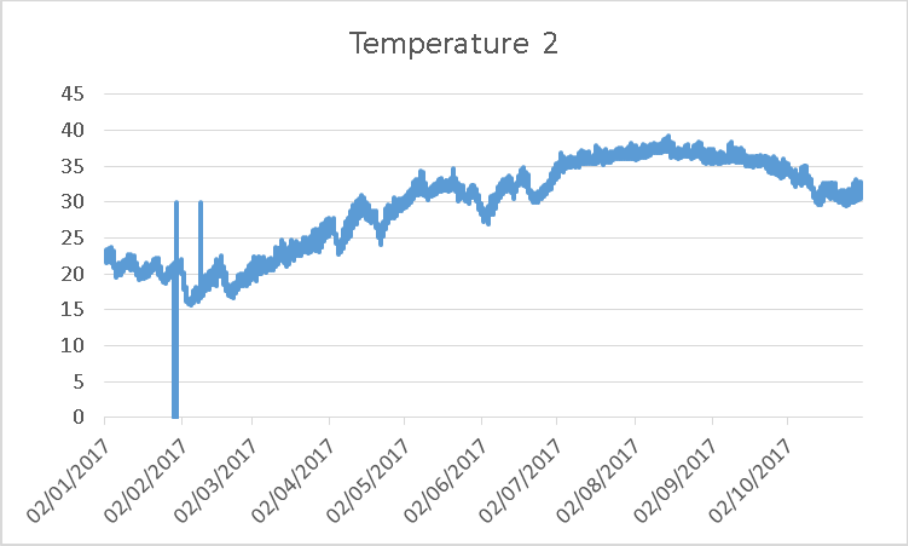
2.4 Outliers

2.4.1 Plant data

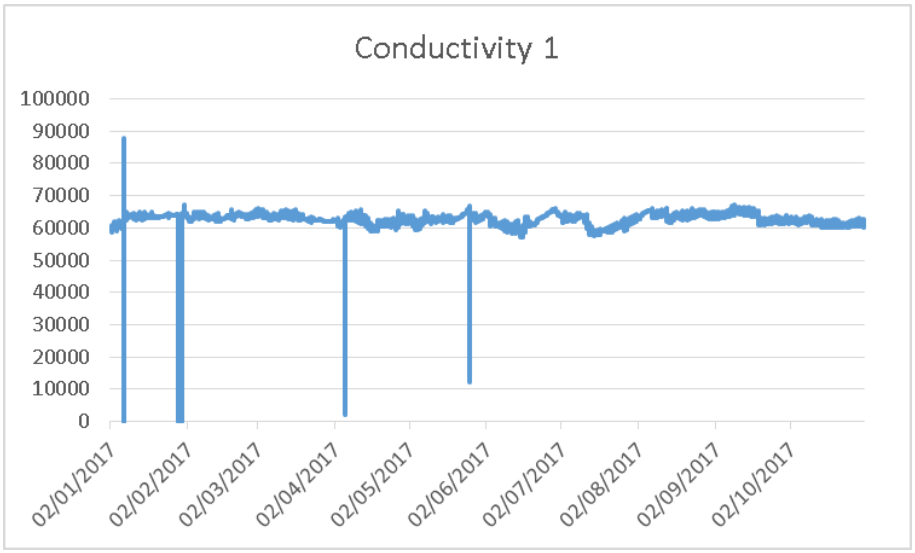
Temperature 1 outliers



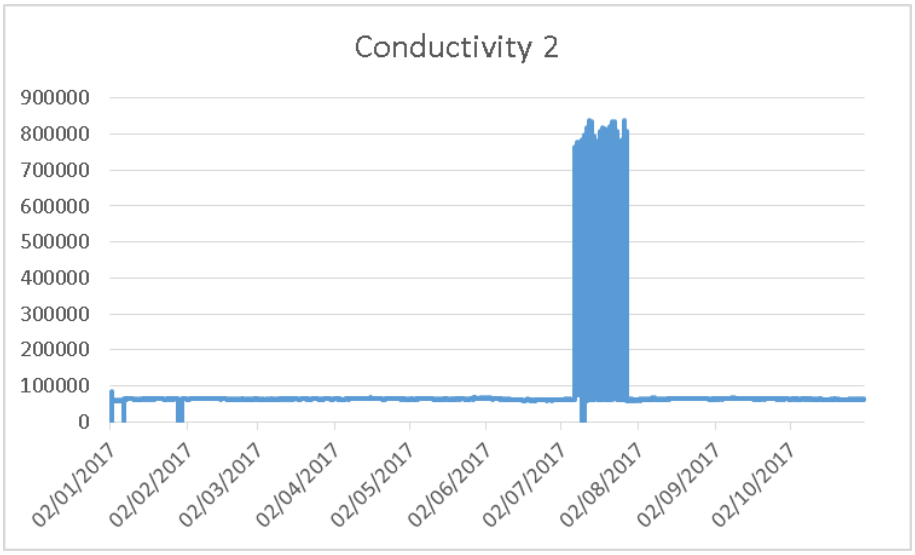
Temperature 2 outliers



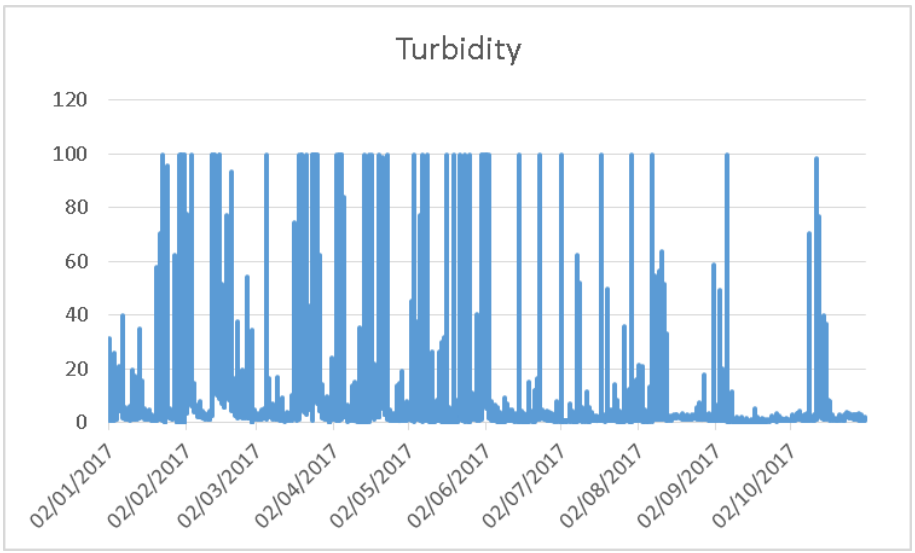
Conductivity 1 outliers



Conductivity 2 outliers



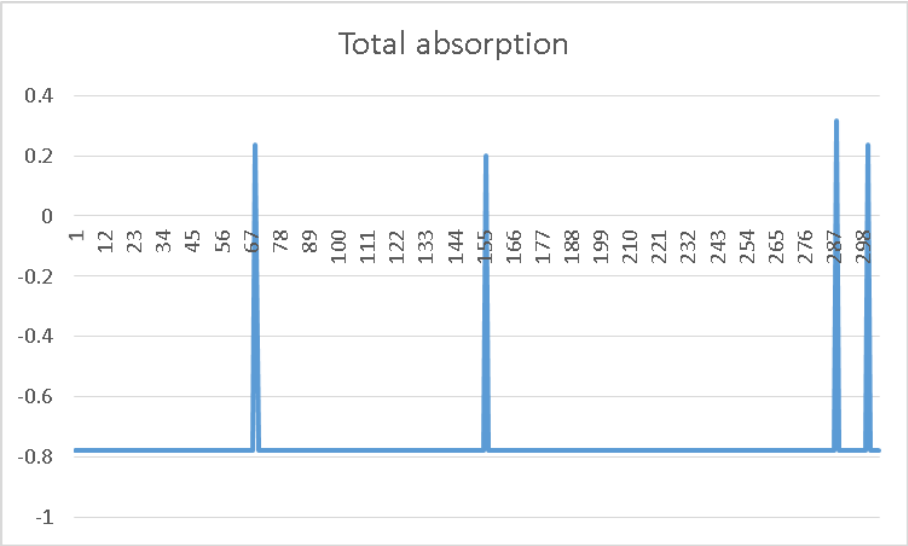
Turbidity outliers

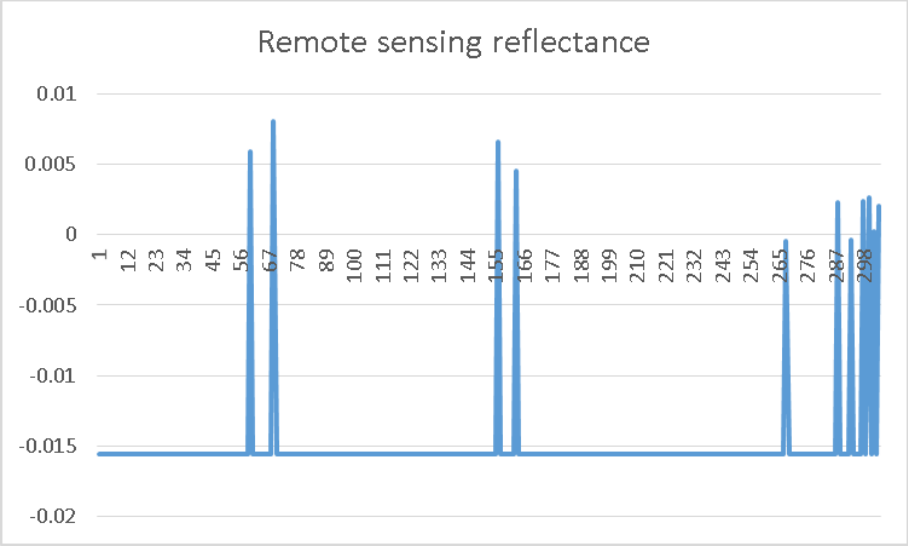


All plant variables have outliers.

Outliers will be set as missing values.

2.4.2 Satellite data





In this case, the missing values seems to be the constant line with negative values.

2.5 Data imputation

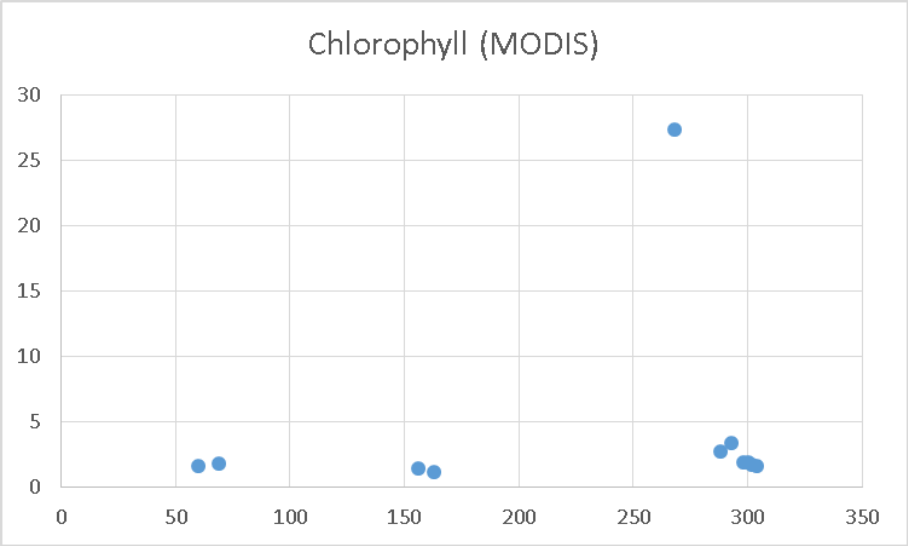
2.5.1 Strong presence

Missing and outlier data are imputed using data interpolation along the time series.

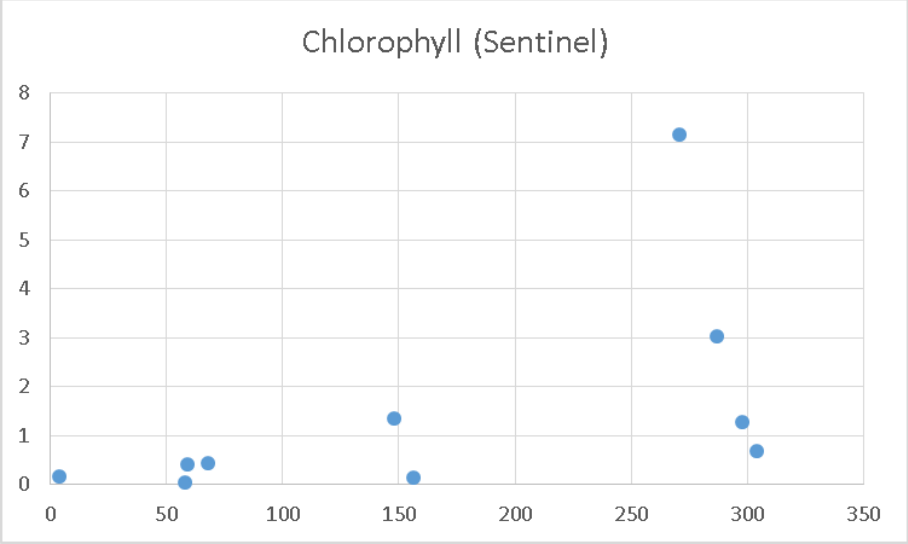
2.5.2 Weak presence

If the percentage of missing values of a variable is greater than a certain threshold, they cannot be imputed, and that variable is discarded.

Time series of MODIS chlorophyll (days) [mg/m3]



Time series of Sentinel chlorophyll (days) [mg/m3]



The daily data of the chlorophylls at a given point has more than 90% of missing values, and it cannot be imputed.

2.6 Variables selection

Some Satellite variables might not be representative for this case study.

Specially those concerning absortion of radiation at different wavelengths.

That variables are removed from the dataset.

2.6.1 Modis Aqua

chl_oc3	Chlorophyll Concentration, OC3 algorithm
chlor_a	Chlorophyll Concentration, OCI algorithm
Ipar	Intantaneous Photosynthetically Avaliable Radiation
Nflh	Normalized fliorescence line height
chl_gsm	Chlorophyll Concentration, GSM model
Kd_490	Diffuse attenuation coefficient (490nm), KD2 algorithm
ndvi	Normalized difference vegetation index
sst	Sea surface temperature
par	Photosynthetically available radiation
Pic	Calcite concentration
Poc	Particle organic carbon
sst4	4um sea surface temperature
Zeu_lee	Euphotic depth, Lee algorithm
adg_443_giop	Absorption due to gelbstoff and detrital material (443nm), GIOP model
adg_s_giop	Detritial and gelbstoff absorption spectral parameter for GIOP model
aph_443_giop	Absorption due to phytoplankton (443nm), GIOP model

2.6.2 Sentinel

Usi	Sea ice eastward velocity
So	Salinity
Sithick	Sea ice thickness
Thetao	Temperature
Siconc	Ice concentration
Uo	Eastward velocity
Vo	Northward velocity
Mlotst	Density ocean mixed layer thickness
Zos	Sea surface height
Vsi	Sea ice northward velocity
bottom	Sea floor potential temperature
Fe	Mole concentration of dissolved iron
PP	Net primary productivity of carbon
Si	Mole concentration of silicate
NO3	Mole concentration of nitrate
CHL	Mass concentration of chlorophyll
PHYC	Mole concentration of phytoplankton expressed as carbon
PO4	Mole concentration of phosphate
O2	Mole concentration of dissolved oxygen
CHL	Mass concentration of chlorophyll
SPM*	Inorganic suspended particle matter

3. Time series

3.1 Plant data

We have two entry points in the plant.

The coordinates of the Point 1 are (25.2124585, 51.640231).

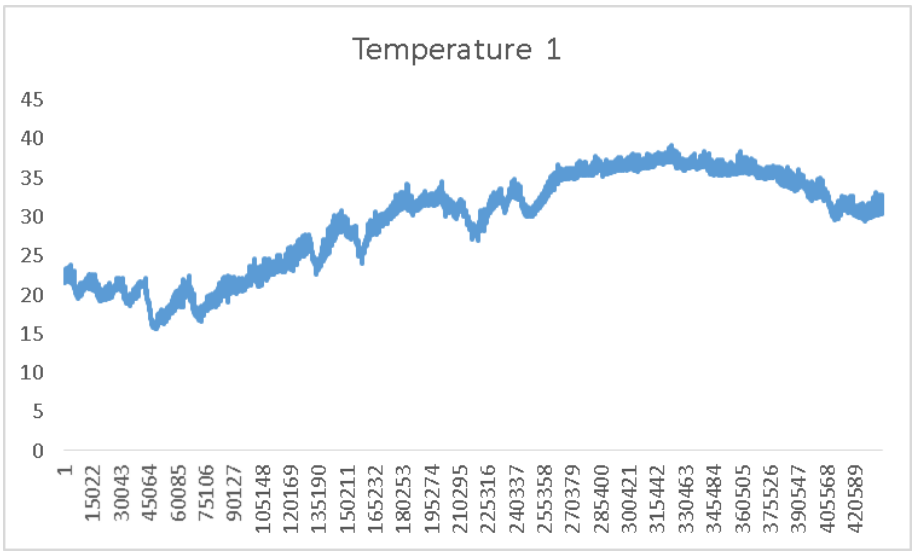
The coordinates of Point 2 are (25.210405, 51.637468). T

The distance between both points is 0.36km.

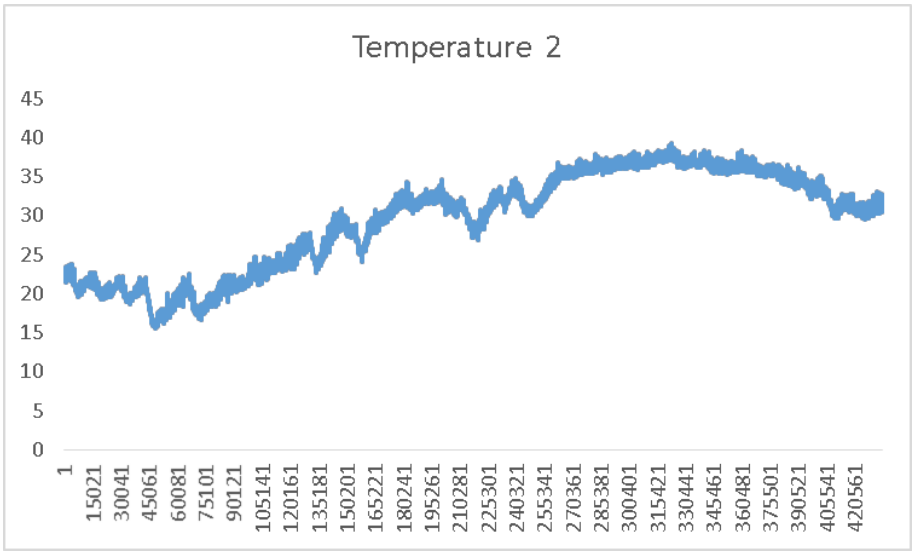
We have data of 304 days with a measure each minute.

The data starts at 02/01/2017 and ends at 31/10/2017.

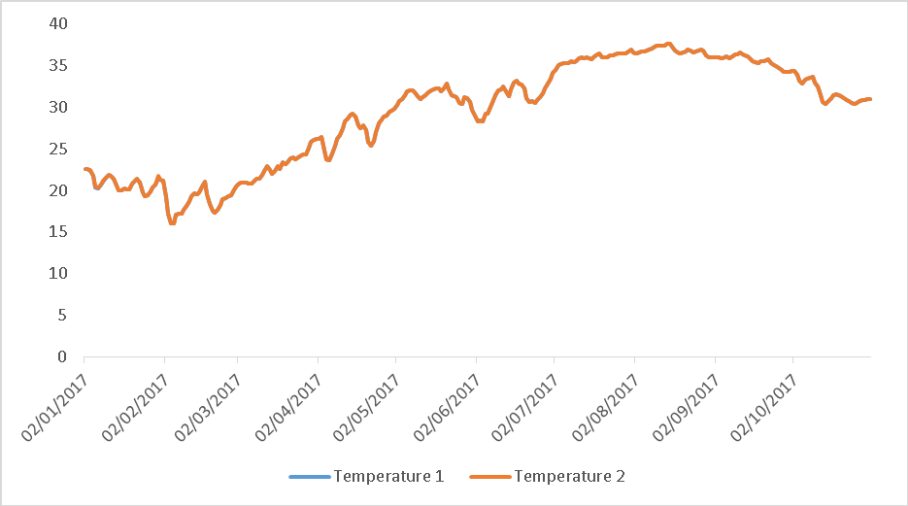
Time series of temperature at Point 1 (minutes)



Time series of temperature at Point 2 (minutes)



Time series of temperatures at Points 1 and 2 (days)

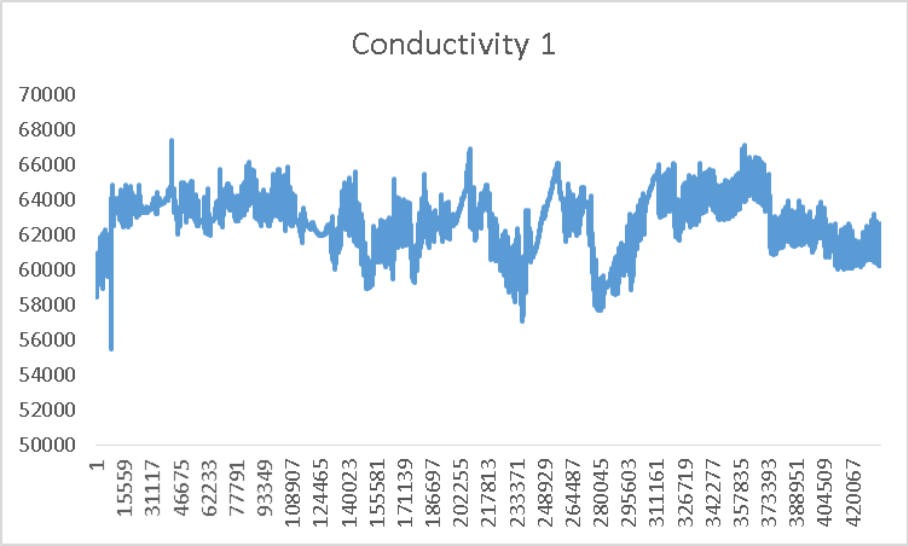


We can see the seasonal trend of the temperature very clearly.

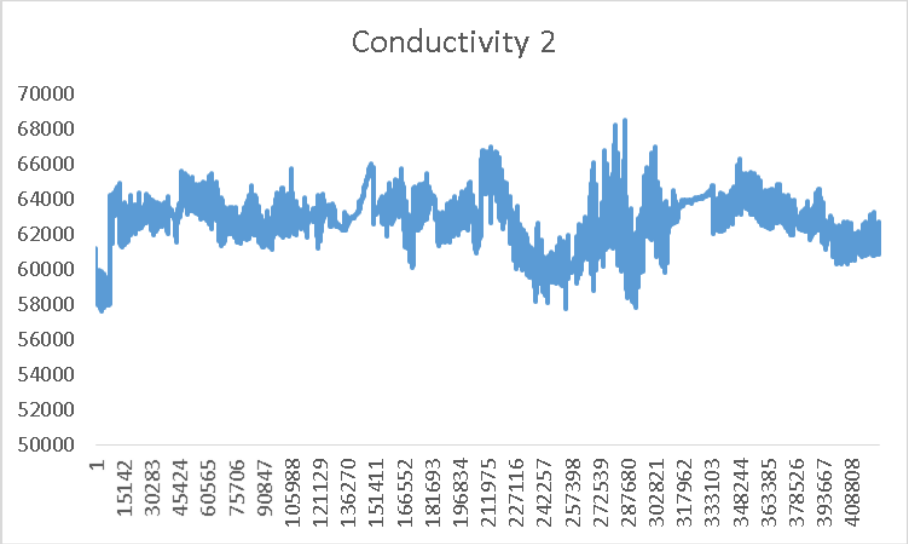
The temperatures at points 1 and 2 are identical.

From now on, we could use a unique value for the temperature at the entry point.

Time series of conductivity at Point 1 (minutes)



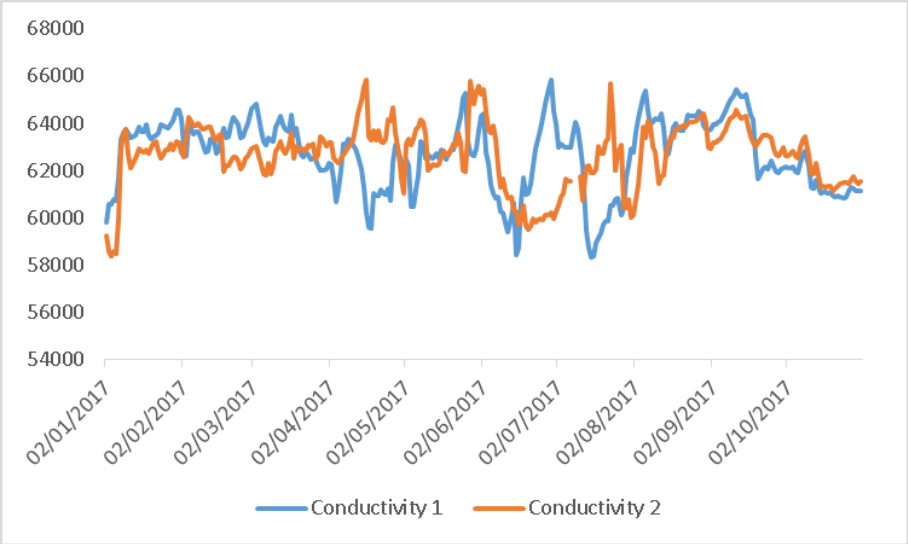
Time series of conductivity at Point 2 (minutes)



Typical values for the conductivity are 50000-60000.

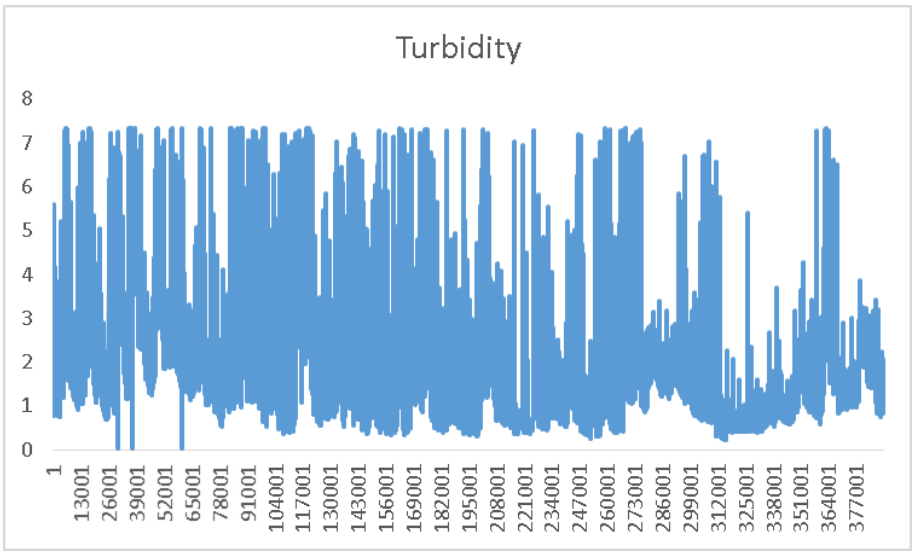
The conductivity at the entry point is high.

Time series of conductivity at Points 1 and 2 (days)

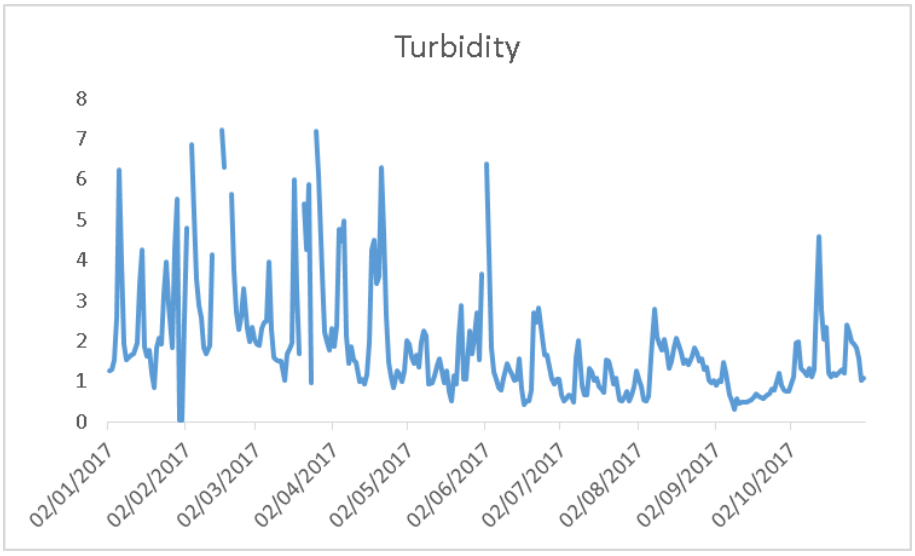


The conductivity at Points 1 and 2 match sometimes. Why???

Time series of turbidity at entry point (minutes)



Time series of turbidity at entry point (days)

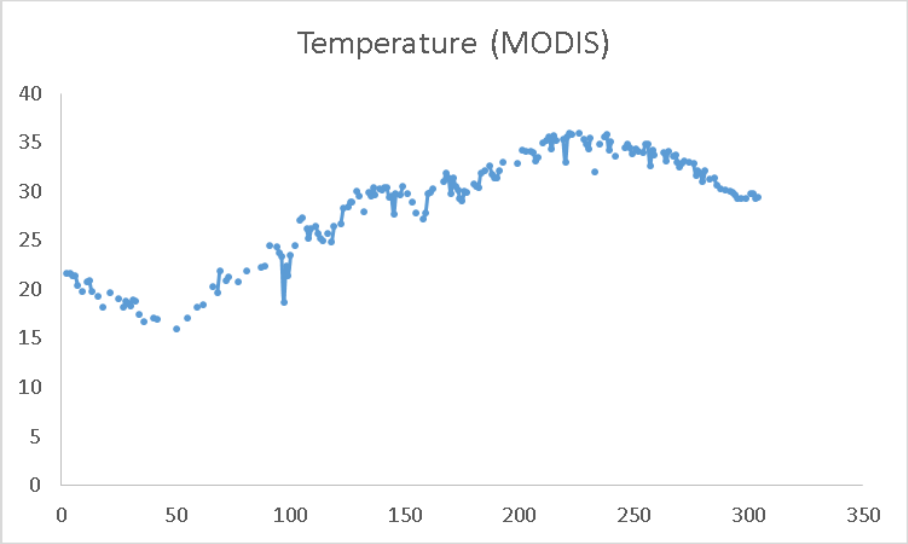


The turbidity has not a seasonal trend.

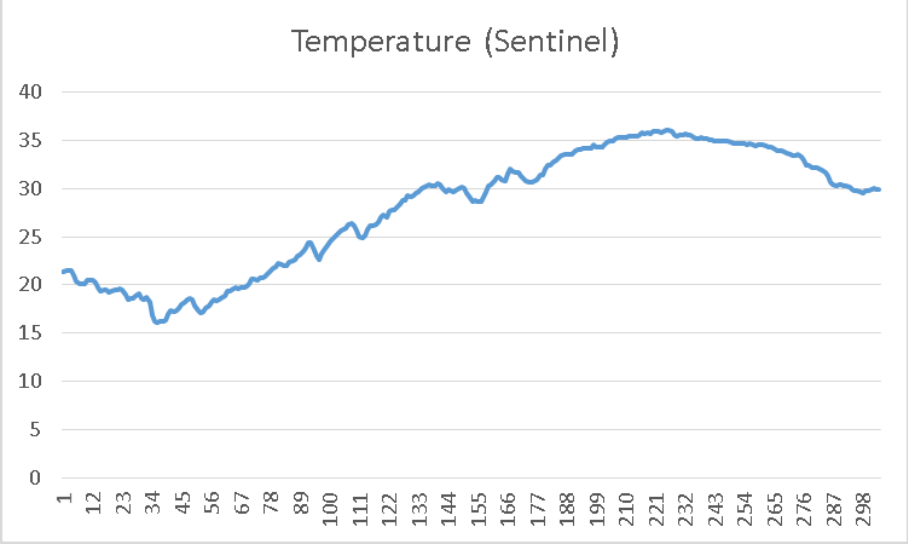
What variables does it depend on??? Meteorological variables?

3.2 Satellite data (entry point)

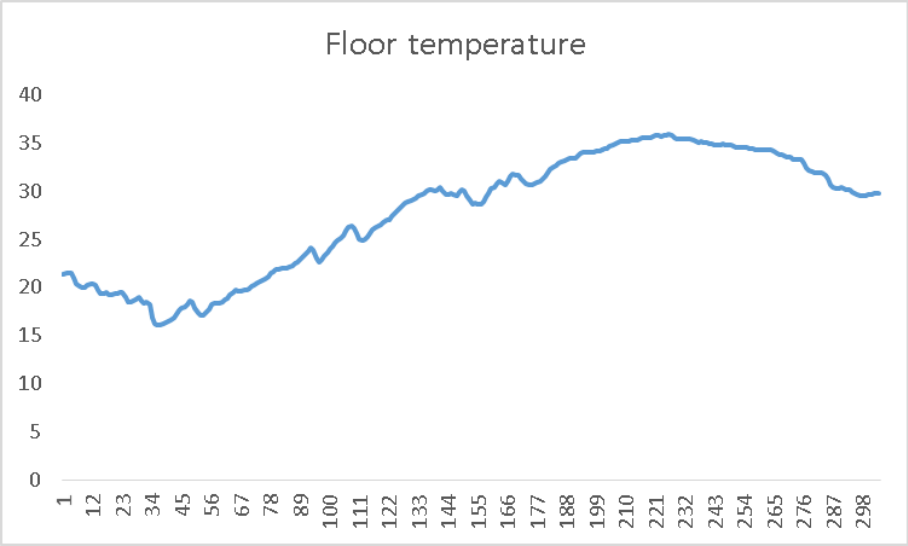
Time series of MODIS surface temperature (days) [Celsius]



Time series of Sentinel surface temperature (days) [Celsius]



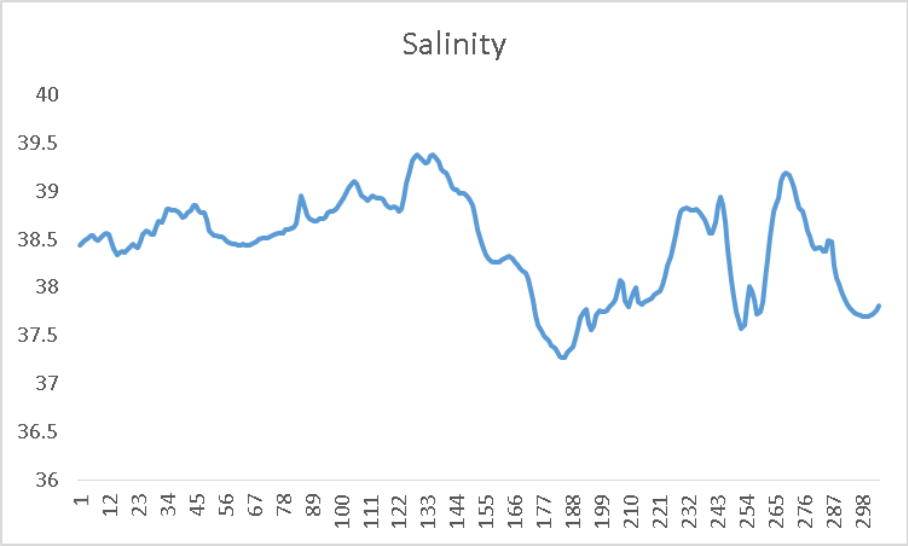
Time series of MODIS floor temperature (days) [Celsius]



The three temperatures have nearly the same graph. But, the Sentinel satellite have more data.

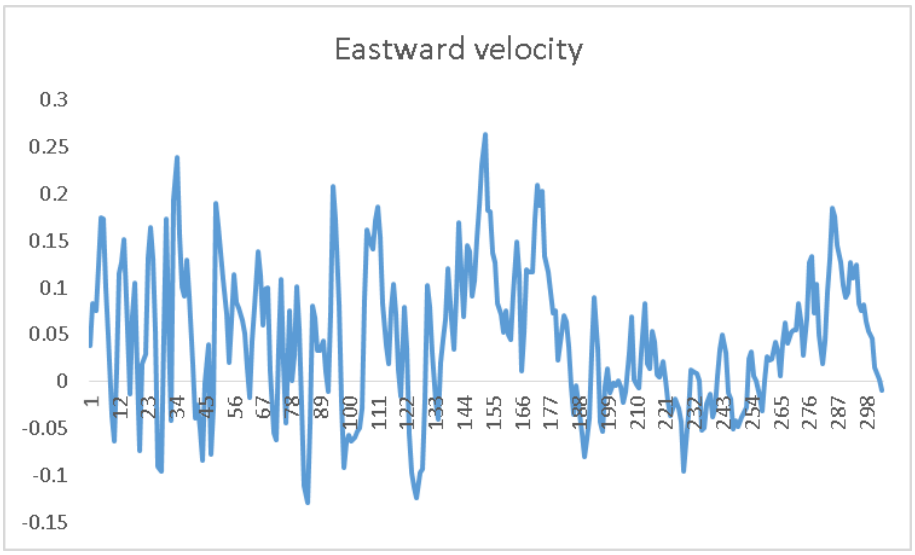
All of them are similar to the internal data.

Time series of Sentinel salinity (days) [practical salinity units]

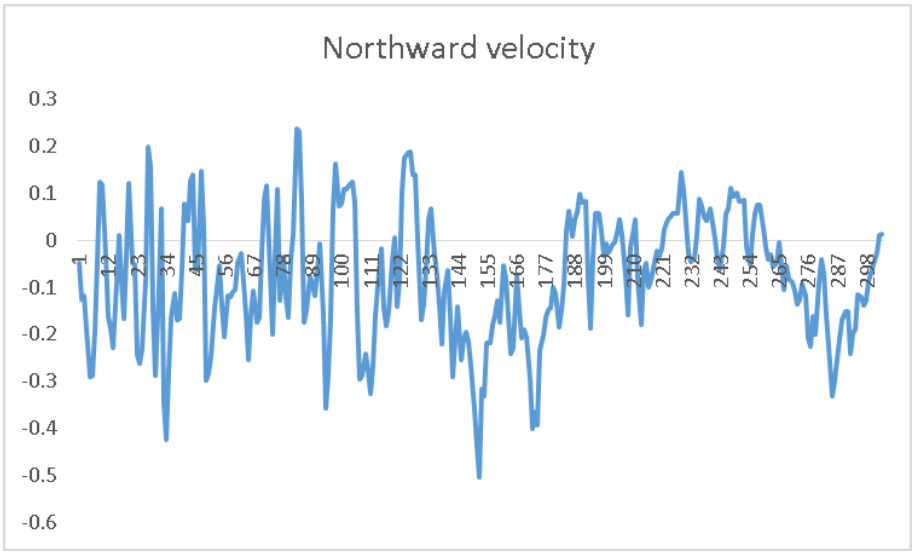


The salinity has no clear seasonality. A study of the influence of the rest of the variables in this one must be done.

Time series of Sentinel satellite water eastward velocity (days) [m/s]



Time series of Sentinel satellite water northward velocity (days) [m/s]



3.3 Satellite data (maximum depth)

@todo

4. Descriptive analytics

4.1 Plant data statistics

Variable	Units	Minimum	Maximum	Mean	Deviation
Temperature 1	Celsius	15.6	39.2	29.30	6.28
Temperature 2	Celsius	15.6	39.2	29.30	6.28
Conductivity 1	µS/cm	55493	67424	62669	1599
Conductivity 2	µS/cm	57599	68464	62660	1470
Turbidity	NTU	0.03	7.33	1.66	1.26

Both temperatures have the same statistics.

The conductivities have different statistics, but their mean is similar.

4.2 Satellite data statistics (entry point)

Variable	Minimum	Maximum	Mean	Deviation
chl_oc3	1.180949	27.33946	4.270516	7.319603
chlor_a	1.180949	27.33946	4.270516	7.319603
Ipar	0.001421	0.001742	0.001551	0.000136
Nflh	0.17889	0.436185	0.287197	0.089084
adg_443_gsm	0.1103	0.1349	0.1226	0.0123
bbp_443_gsm	0.02282	0.035	0.02891	0.00609
chl_gsm	0.127262	0.329561	0.228411	0.10115
a_412_giop	0.1998	0.3148	0.247175	0.041901
a_443_giop	0.1285	0.191	0.151825	0.023537
a_469_giop	0.0908	0.1287	0.10395	0.014703
a_488_giop	0.0731	0.0994	0.0817	0.010474
a_531_giop	0.0699	0.0829	0.07425	0.005123
a_547_giop	0.0709	0.0807	0.07415	0.003879
a_555_giop	0.0752	0.0838	0.078075	0.003389
a_645_giop	0.3229	0.3266	0.325025	0.001353
a_667_giop	0.4291	0.4358	0.43345	0.002641
a_678_giop	0.4561	0.4635	0.4608	0.002857
adg_443_giop	0.1022	0.1704	0.1325	0.024469
adg_s_giop	0.018	0.018	0.018	0
adg_unc_443_giop	0.0147	0.0468	0.025025	0.012744
aph_443_giop	0.0041	0.02	0.012975	0.005701
aph_unc_443_giop	0.0133	0.0437	0.022375	0.012373
bb_412_giop	0.01823	0.04099	0.026846	0.008891
bb_443_giop	0.01673	0.038355	0.024855	0.008463
bb_469_giop	0.015735	0.03651	0.023494	0.008142
bb_488_giop	0.01512	0.03532	0.02263	0.007926
bb_531_giop	0.013965	0.033005	0.020983	0.007487
bb_547_giop	0.013605	0.03225	0.020454	0.007337
bb_555_giop	0.013435	0.03189	0.020204	0.007265
bb_645_giop	0.0119	0.028505	0.017888	0.006566
bb_667_giop	0.011605	0.02782	0.017429	0.006418
bb_678_giop	0.01146	0.027495	0.017211	0.006349
bbp_443_giop	0.01449	0.03611	0.022618	0.008461
bbp_s_giop	0.628379	0.744637	0.692488	0.045365
bbp_unc_443_giop	0.001205	0.00253	0.001899	0.000543
Kd_490	0.1156	2.1832	0.344018	0.582582
Sst	15.965	35.98	28.61242	5.483987
Par	11.882	61.718	46.67001	11.2163
Pic	0.004598	0.010902	0.006343	0.001983
Poc	283.1999	914.5999	406.8726	180.091

a_443_qaa	0.0991	0.8229	0.237536	0.194806
adg_443_qaa	0.0778	0.6931	0.195518	0.165594
aph_443_qaa	0.0149	0.1234	0.035636	0.029206
bbp_443_qaa	0.015385	0.03348	0.022097	0.005056
Angstrom	0.8441	1.661	1.384764	0.279127
aot_869	0.0597	0.2247	0.115882	0.054343
Rrs_412	-0.00046	0.008022	0.003058	0.002726
Rrs_443	0.001662	0.013004	0.006805	0.003273
Rrs_469	0.002674	0.017436	0.009359	0.004167
Rrs_488	0.00311	0.019984	0.010892	0.00475
Rrs_531	0.006814	0.021128	0.011849	0.004263
Rrs_547	0.007054	0.020206	0.011345	0.00396
Rrs_555	0.006496	0.018296	0.010302	0.003543
Rrs_645	0.000892	0.002764	0.001441	0.000515
Rrs_667	0.000538	0.002018	0.001045	0.000379
Rrs_678	0.00063	0.00197	0.001139	0.000333
sst4	15.575	36.46	29.03907	5.659345
Zeu_lee	9.125004	24.185	19.09319	3.997856
So	37.26615	39.37803	38.4373	0.497999
Thetao	16.0604	36.04587	27.94634	6.260816
Uo	-0.1294	0.26307	0.045346	0.075309
Vo	-0.50356	0.236213	-0.08203	0.133365
Mlotst	8.087405	8.087405	8.087405	2.84E-14
Zos	-0.08911	0.347301	0.157866	0.072971
bottom	16.04062	35.89718	27.79602	6.251634
CHL	0.044394	7.156596	1.472607	2.076661

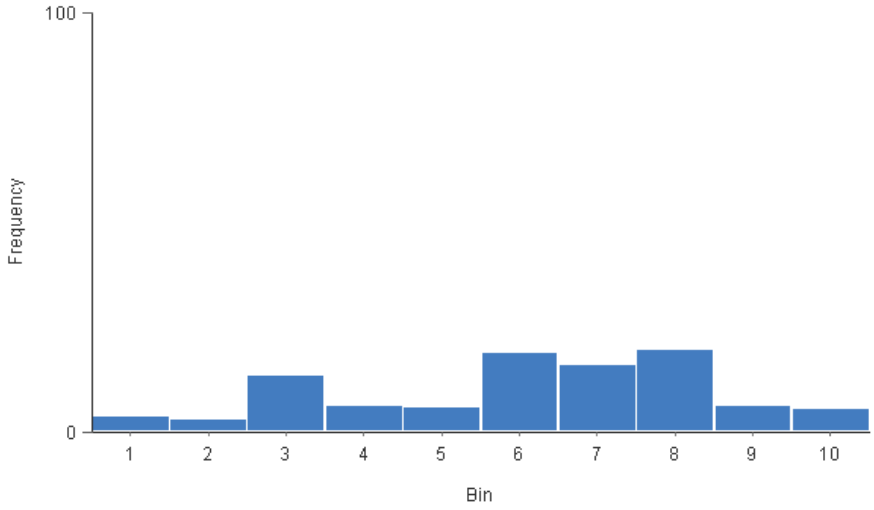
The standard deviation of all variable is very low. This is because we have a lot of missing values in this point.

4.3 Satellite data statistics (maximum depth)

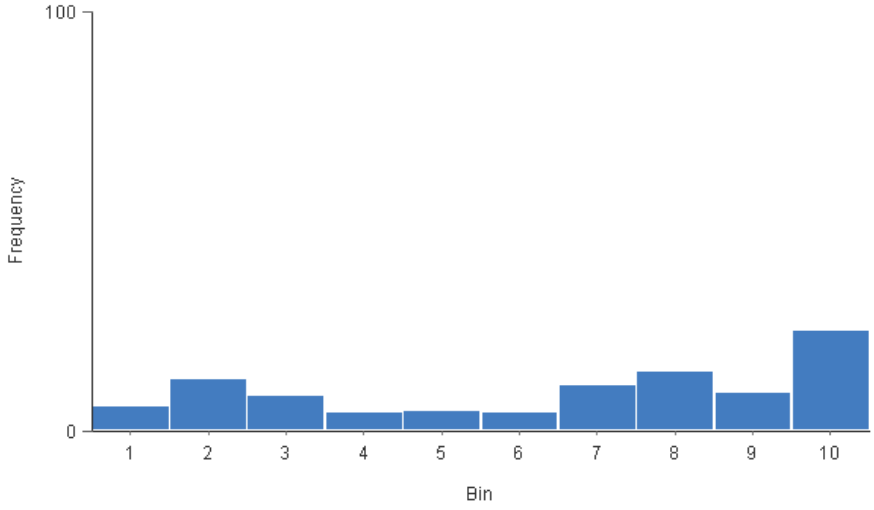
@todo

4.4 Satellite data distribution (entry point)

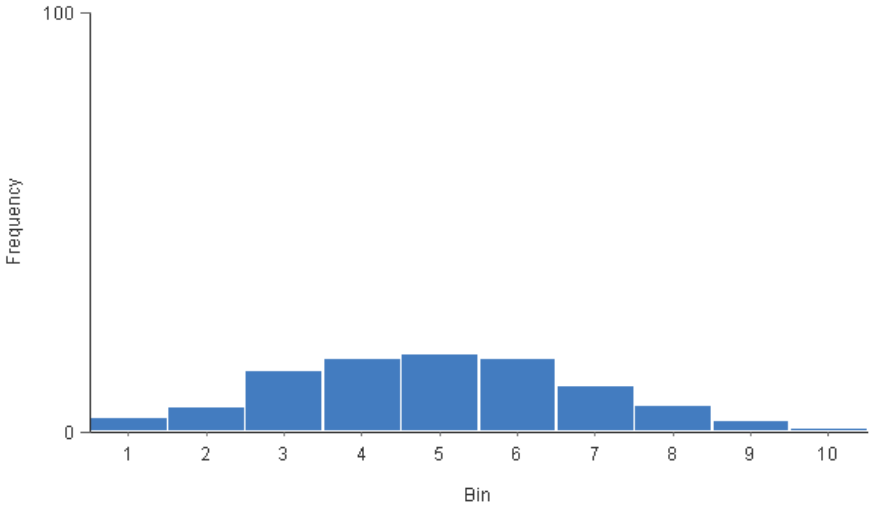
Salinity histogram



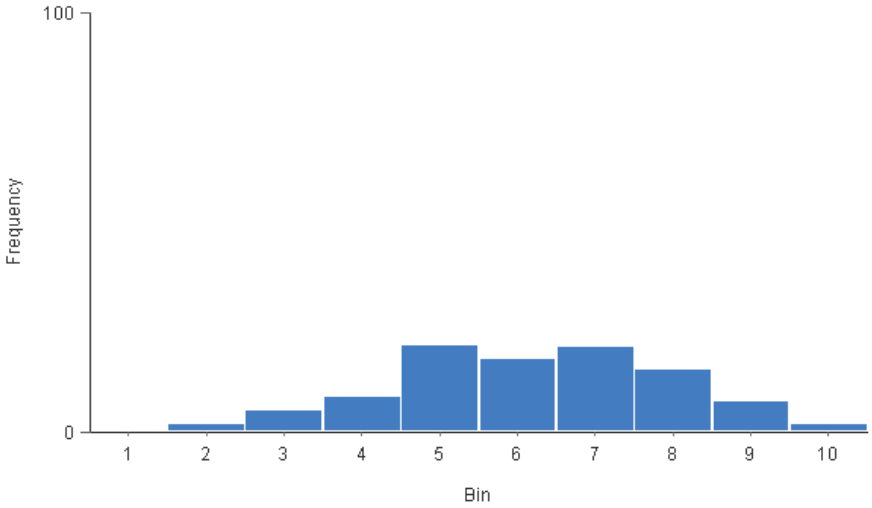
Sentinel surface temperature histogram



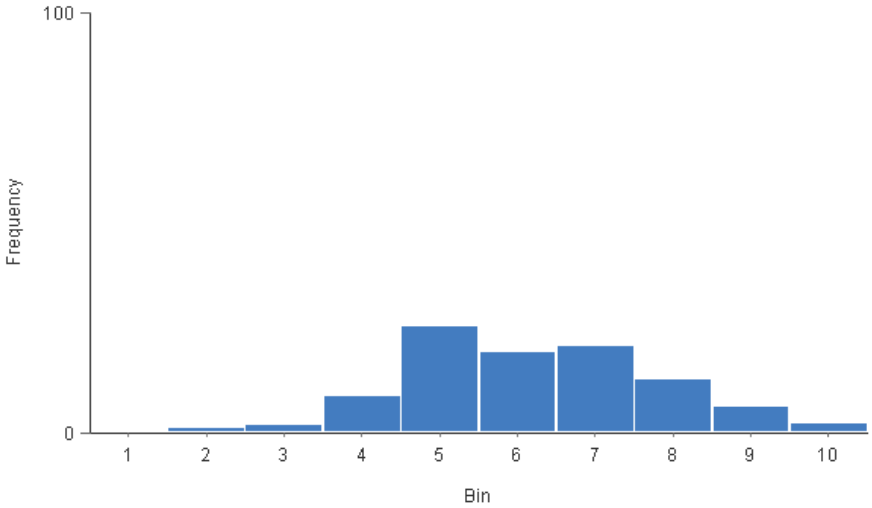
Eastward velocity histogram



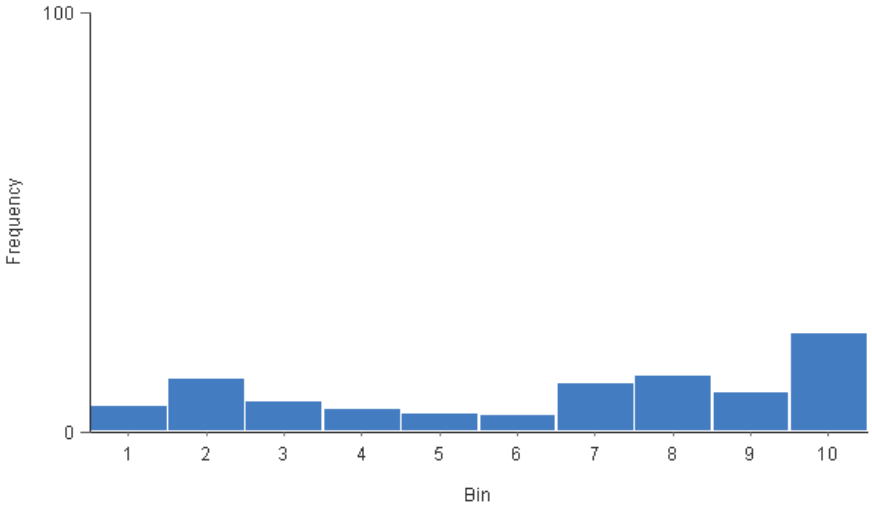
Northward velocity histogram



Sea surface height histogram



Floor temperature histogram



The Sentinel variables has no outliers because their distribution are similar to a normal or uniform distribution.

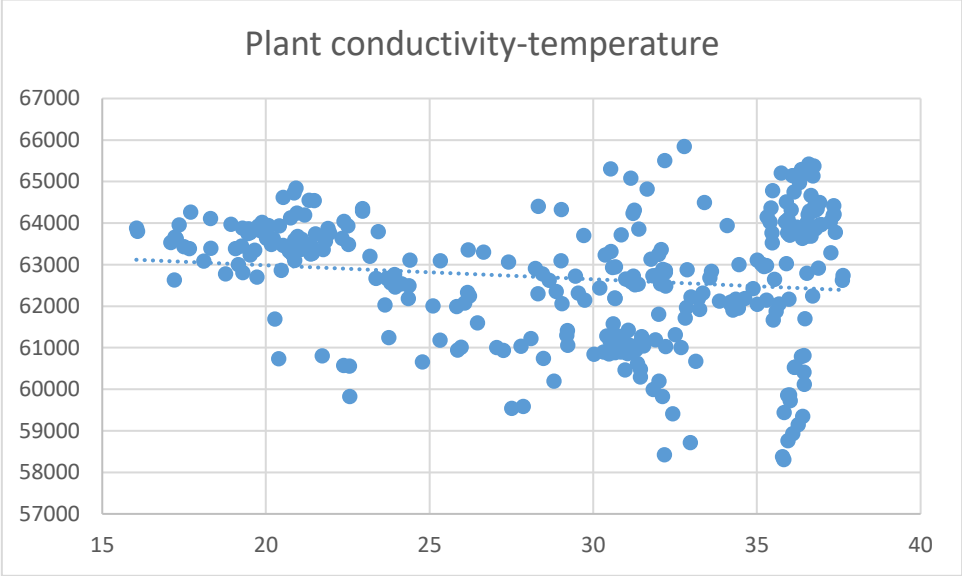
4.5 Satellite data distribution (maximum depth)

@todo

5. Diagnostic analytics

5.1 Plant scatter charts

5.1.1 Conductivity – Temperature

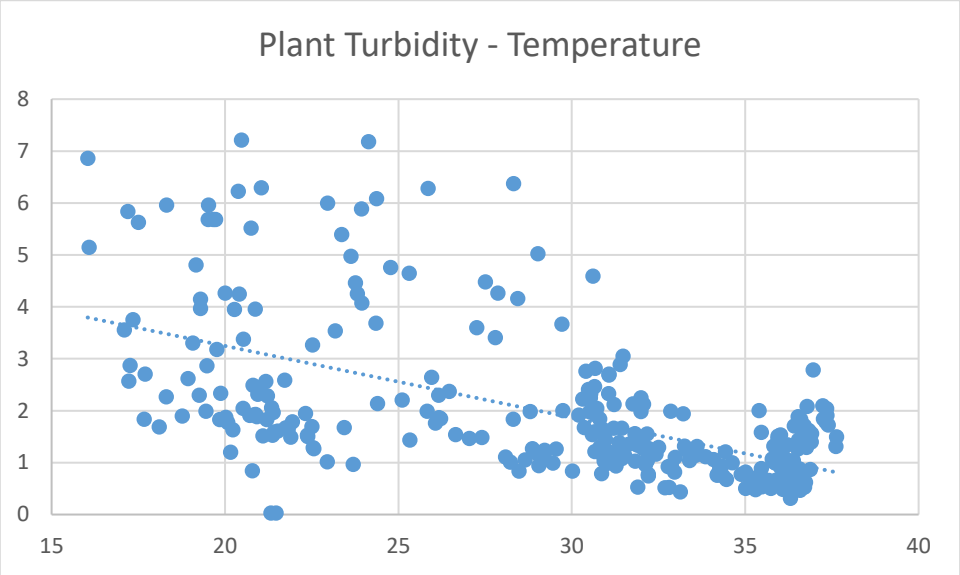


The conductivity is less scattered for low values of the temperature.

There is a small descending trend of the conductivity with the temperature.

What is the effect of algae bloom in conductivity?

5.1.2 Turbidity – Temperature

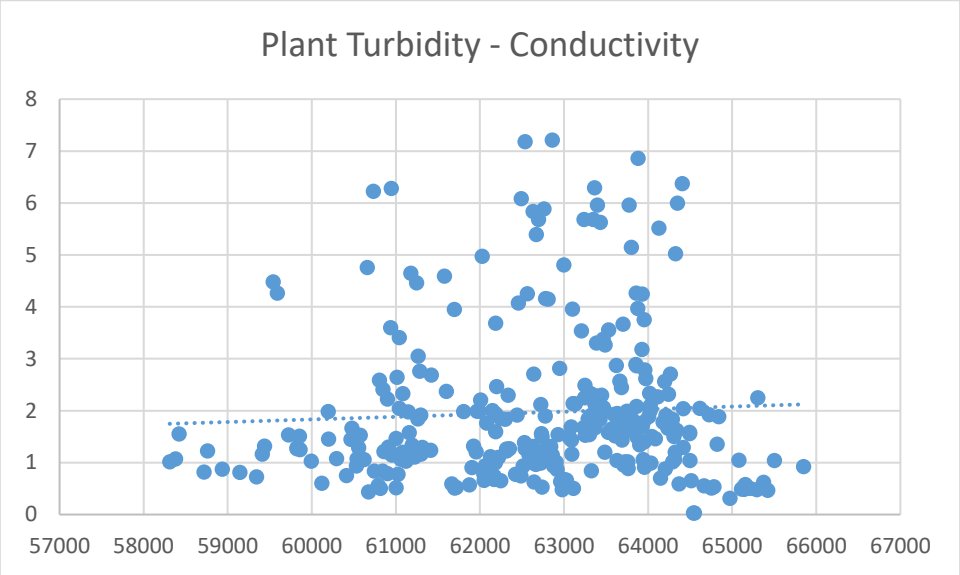


The turbidity is more scattered when the temperature is low.

There is a decreasing trend of the turbidity with the temperature.

Algae bloom appear with high temperatures.

5.1.3 Turbidity- Conductivity



The turbidity is more scattered for high conductivities.

There is a small increasing trend of the turbidity with the conductivity.

5.2 Plant correlations

The internal correlations are taken from the two capture points of the plant. We have two temperatures, two conductivities and a turbidity.

	Temperature	Conductivity	Turbidity
Temperature	1	-0.139	-0.589
Conductivity		1	0.051
Turbidity			1

More temperature – less conductivity (small).

More temperature – less turbidity (medium).

More conductivity – more turbidity (small).

Creemos que lo que puede estar dando la relación entre la temperatura y la conductividad y la turbidez pueda ser una termoclina.

La termoclina separaría la zona superficial con agua más cálida y más densa. Con el agua del fondo más fría y menos densa.

Al realizarse la inversión de la termoclina puede provocar que el agua más fría del fondo arrastre la turbidez hacia la toma. Por otro lado, con la conductividad (aunque a priori está asociada a la temperatura: más temperatura más conductividad) puede ser que en la inversión de la termoclina se arrastre también conductividad.

5.2.1 Conductivity

1. conductivity_ahead_1 - conductivity_lag_1: 0.935
2. conductivity_ahead_1 - conductivity_lag_2: 0.831
3. conductivity_ahead_1 - conductivity_lag_3: 0.736
4. conductivity_ahead_1 - temperature_lag_2: -0.162
5. conductivity_ahead_1 - temperature_lag_1: -0.161
6. conductivity_ahead_1 - temperature_lag_3: -0.161
7. conductivity_ahead_1 - turbidity_lag_3: 0.109
8. conductivity_ahead_1 - turbidity_lag_2: 0.102
9. conductivity_ahead_1 - turbidity_lag_1: 0.071

The correlations between the future conductivity with the past conductivities is very high and positive (bigger conductivity past, bigger conductivity future).

The correlations between the future conductivity with the past temperatures is small and negative (bigger temperature, smaller conductivity).

The correlations between the future conductivity with the past turbidities is very small and positive (bigger turbidity, bigger conductivity).

5.1.2 Turbidity

1. turbidity_ahead_1 - turbidity_lag_1: 0.761
2. turbidity_ahead_1 - temperature_lag_1: -0.572
3. turbidity_ahead_1 - temperature_lag_2: -0.545
4. turbidity_ahead_1 - temperature_lag_3: -0.525
5. turbidity_ahead_1 - turbidity_lag_2: 0.520
6. turbidity_ahead_1 - turbidity_lag_3: 0.422
7. turbidity_ahead_1 - conductivity_lag_3: 0.058
8. turbidity_ahead_1 - conductivity_lag_2: 0.035
9. turbidity_ahead_1 - conductivity_lag_1: 0.028

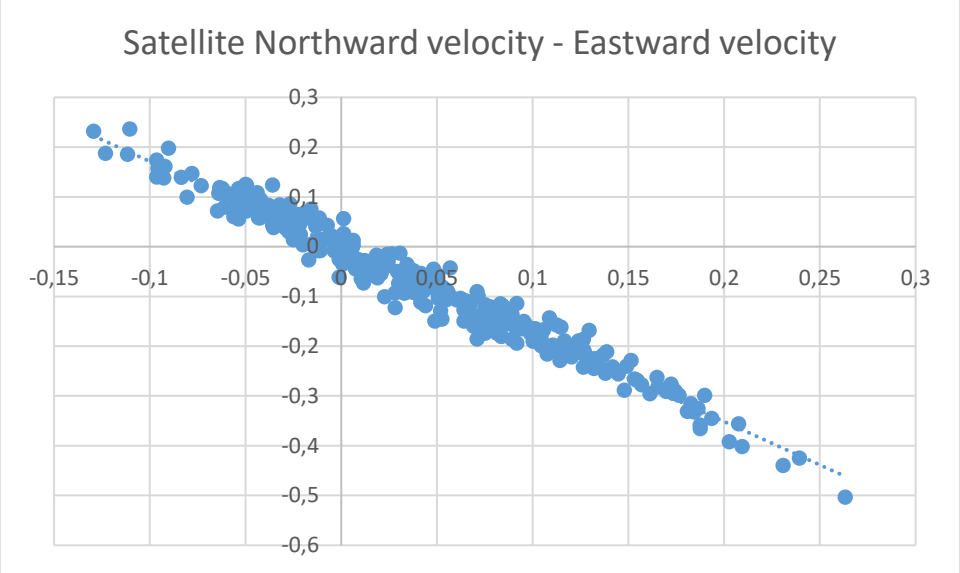
The correlations between the future conductivity with the past conductivities is very high and positive (bigger conductivity past, bigger conductivity future).

The correlations between the future conductivity with the past temperatures is small and negative (bigger temperature, smaller conductivity).

The correlations between the future conductivity with the past turbidities is very small and positive (bigger turbidity, bigger conductivity).

5.2 Satellite scatter charts (entry point)

5.2.1 Northward velocity – Eastward velocity

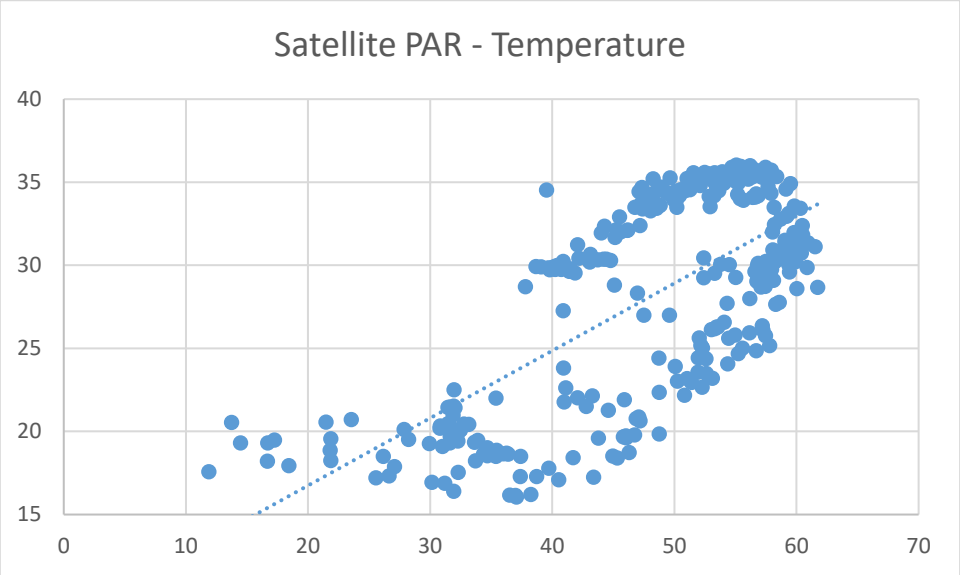


The latitude and longitude velocities are highly correlated:

- West – North
- East – South

The plant location is at the coast and in a very particular location, so the current might be forced by that?

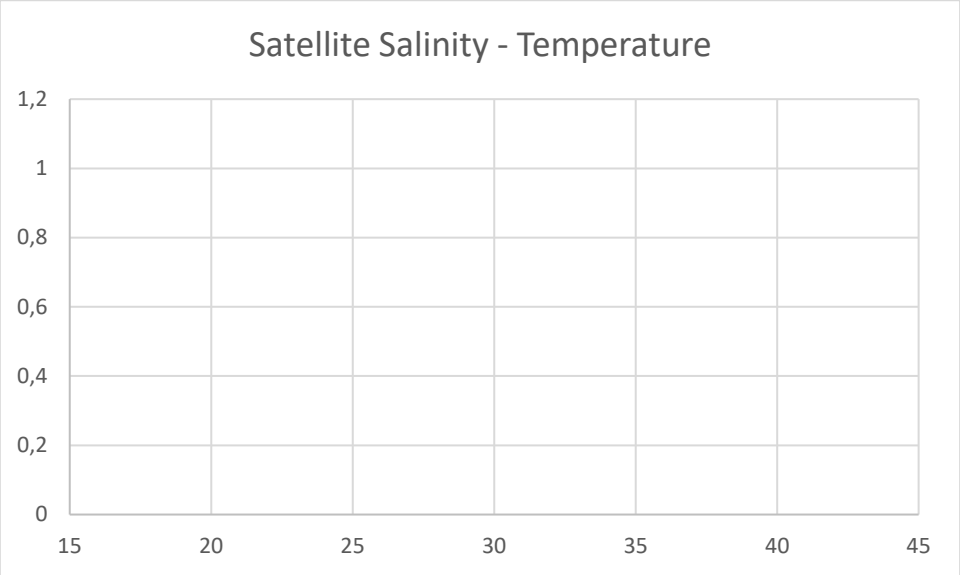
5.2.2 Photosynthetically Available Radiation- Temperature



The PAR increases with the temperature.

There are two branches for medium temperatures. Do they correspond to spring and autumn?

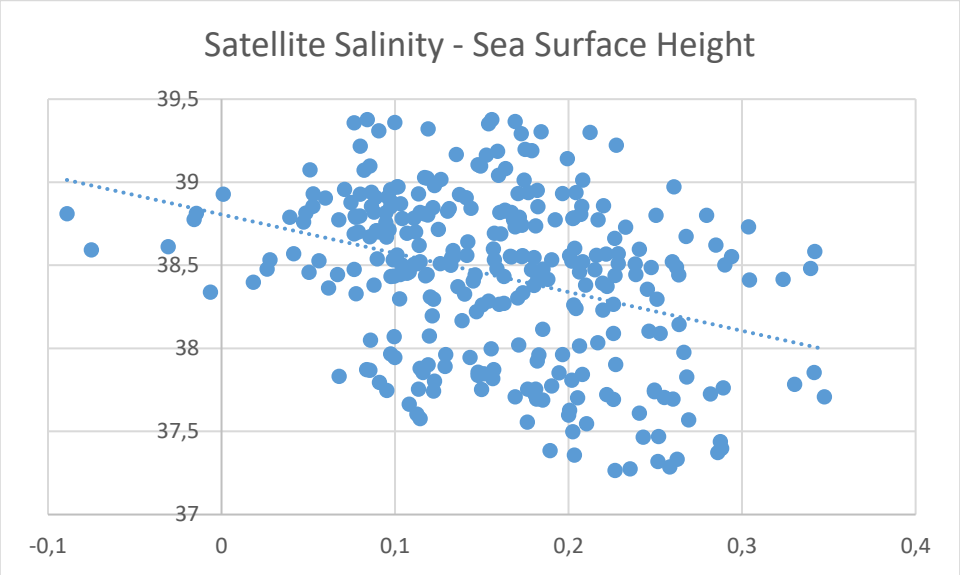
5.2.3 Salinity- Temperature



The salinity decreases with the temperature.

The data follows some trends. What is the explanation for that?

5.2.4 Salinity – Sea Surface Height



The salinity decreases with the sea surface height.

5.3 Satellite correlations (entry point)

1. temperature - sea_floor_potential_temperature: 0.999
2. 4um_sea_surface_temperature - sea_floor_potential_temperature: 0.995
3. 4um_sea_surface_temperature - temperature: 0.995
4. sea_surface_temperature - 4um_sea_surface_temperature: 0.995
5. sea_surface_temperature - sea_floor_potential_temperature: 0.992
6. sea_surface_temperature - temperature: 0.991
7. **eastward_velocity - northward_velocity: -0.983**
8. **photosynthetically_available_radiation - temperature: 0.703**
9. photosynthetically_available_radiation - 4um_sea_surface_temperature: 0.699
10. photosynthetically_available_radiation - sea_floor_potential_temperature: 0.699
11. sea_surface_temperature - photosynthetically_available_radiation: 0.694
12. **salinity - temperature: -0.354**
13. salinity - sea_floor_potential_temperature: -0.354
14. **salinity - sea_surface_height: -0.341**
15. 4um_sea_surface_temperature - salinity: -0.339
16. sea_surface_temperature - salinity: -0.326
17. temperature - eastward_velocity: -0.176
18. eastward_velocity - sea_floor_potential_temperature: -0.166
19. 4um_sea_surface_temperature - eastward_velocity: -0.159
20. sea_surface_temperature - eastward_velocity: -0.159
21. temperature - northward_velocity: 0.158
22. northward_velocity - sea_floor_potential_temperature: 0.149
23. sea_surface_temperature - northward_velocity: 0.145
24. 4um_sea_surface_temperature - northward_velocity: 0.144
25. photosynthetically_available_radiation - salinity: -0.143
26. sea_surface_height - sea_floor_potential_temperature: 0.115
27. temperature - sea_surface_height: 0.113
28. 4um_sea_surface_temperature - sea_surface_height: 0.113
29. northward_velocity - sea_surface_height: -0.112
30. sea_surface_temperature - sea_surface_height: 0.098
31. eastward_velocity - sea_surface_height: 0.079
32. salinity - eastward_velocity: 0.031
33. photosynthetically_available_radiation - sea_surface_height: -0.030
34. photosynthetically_available_radiation - northward_velocity: -0.027
35. salinity - northward_velocity: -0.020
36. date - northward_velocity: 0.014
37. date - eastward_velocity: -0.006
38. photosynthetically_available_radiation - eastward_velocity: 0.005

The highest correlations are among temperature variables.

There are many variables with very little correlation among them.

5.4 Satellite scatter charts (maximum depth)

@todo

5.5 Satellite correlations (maximum depth)

@todo

5.6 Plant-satellite correlations (entry point)

5.6.1 Temperature

@todo

5.6.2 Conductivity

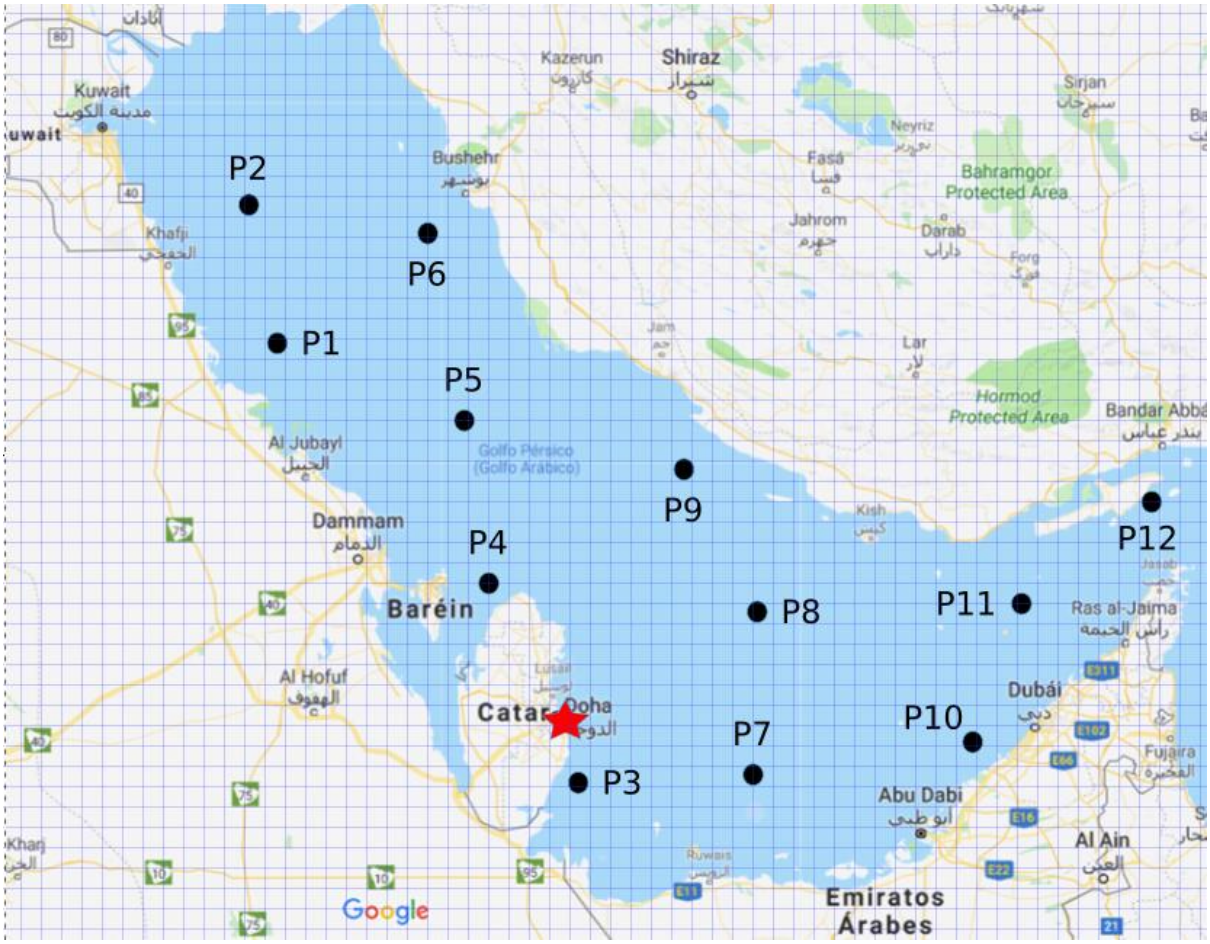
@todo

5.6.3 Turbidity

@todo

5.7 Plant-satellite correlations (defined regions)

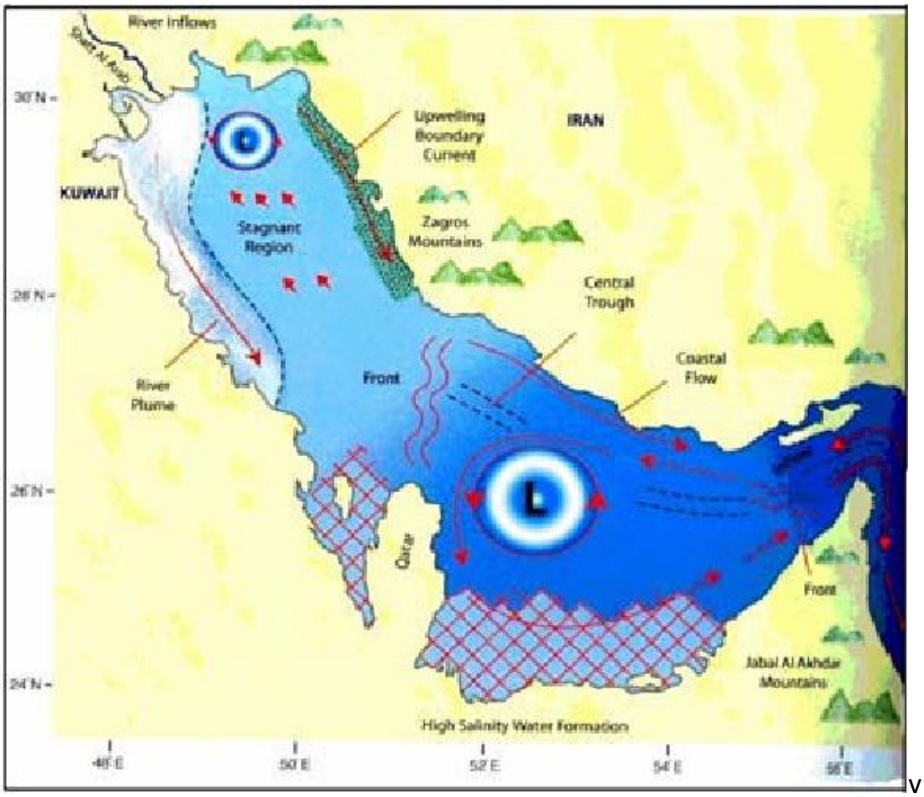
We have defined 12 regions in the Persian Gulf to see the geographical effect



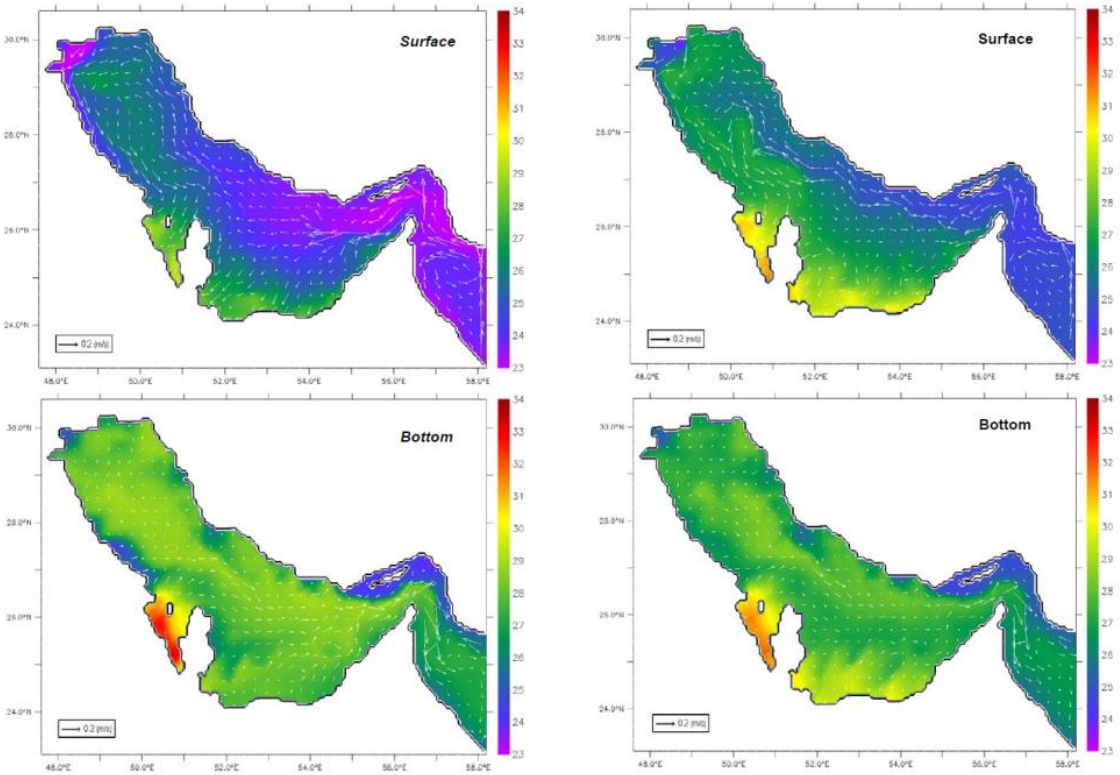
1.7.1. Next steps

We need to define a better representation of the zones of influence.

For that, we can use the general circulation of currents in the Persian Gulf.



Adjunto también esta imagen que nos puede dar información de la circulación de las corrientes vs la densidad en superficie y en profundidad. (Superficie es lo que captaría el satélite, y en profundidad seria lo analizado por las variables internas)

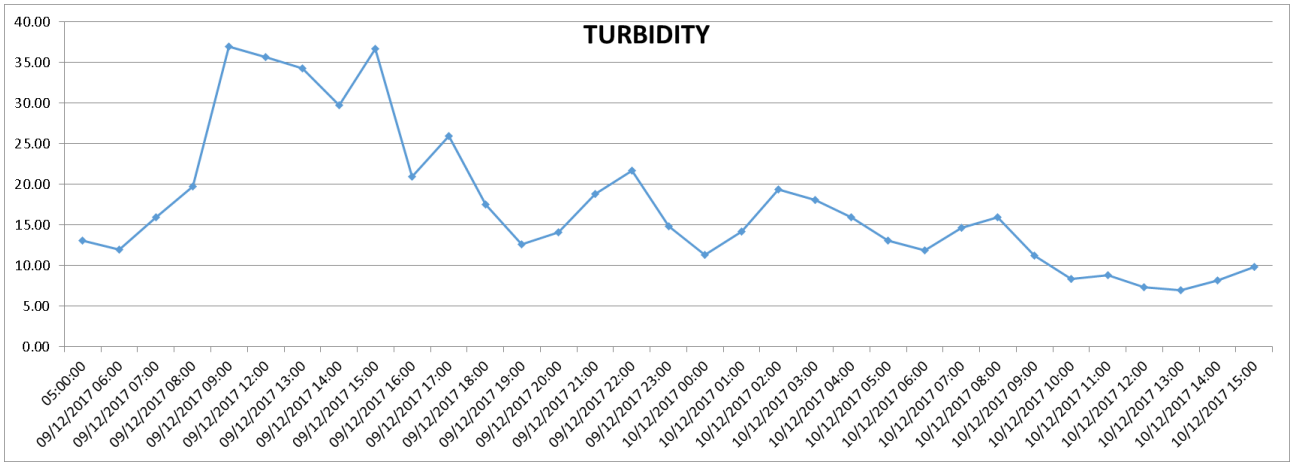


We can identify 5 different regions:

1. Strait of Ormuz, with high current.
2. Iran coast, with high depth.
3. UAE coast, with high salinity.
4. Kuwait coast, with a river plume.
5. North-west area, with little influence.

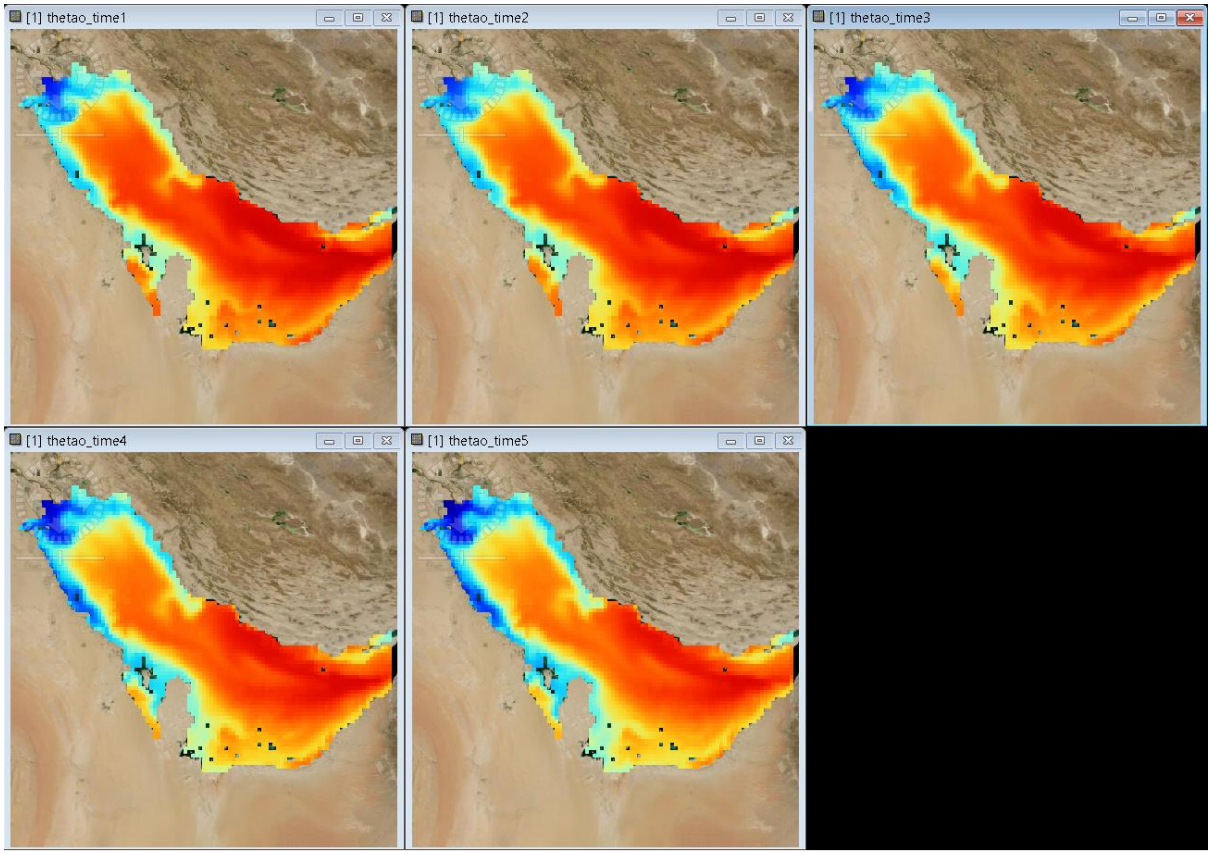
1.7.2. Turbidity problem

The day 09/12/2017 the plant registered a high level of turbidity with its measures.



We have analyzed the data from the satellites for that week.

Surface temperature in the Sentinel

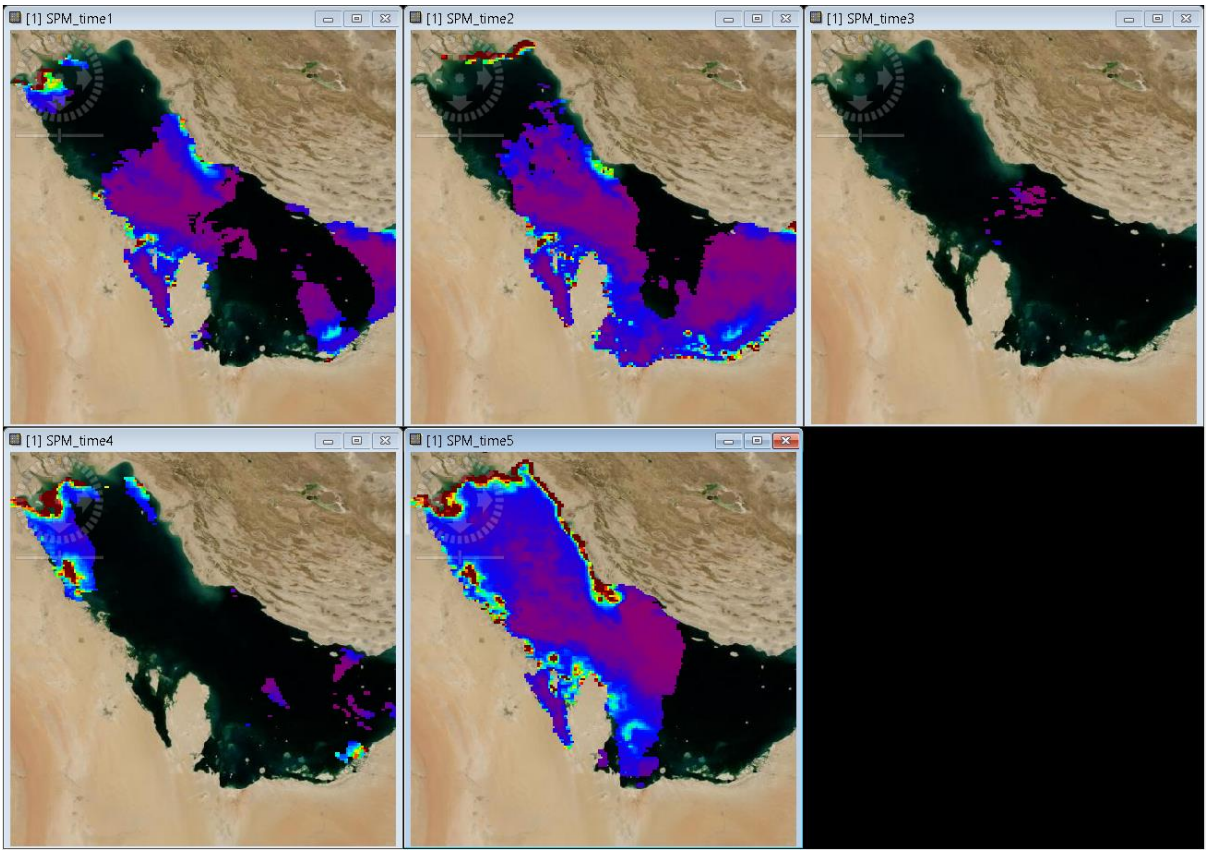


The image represents the temperature between the days 6 and 10 of December.

The last days we can see a lower temperature near the entry point.

The cold water comes from the west with the currents.

Suspended particle matter in the Sentinel



This image represents the suspended particle matter between the same days.

We can see that the day 8 and 9 of December there are nearly no data.
This could be because of the cloudy weather in the area.

Near the entry point the values of the SPM raise up to 2 g/m3.

6. Data Base

These are the collections for the current database.

Plant
- Plant name - Endogenous variables - Exogenous variables - Model paths - Model errors

Internal data
- Plant name - Date - Endogenous variables

External data
- Plant name - Date - Latitude - Longitude - Exogenous variables

Users
- Role - Accessible plants

We are using a Mongo data base for the storage of the data.

This database is able to retrieve data and calculate statistics and meshes.

The Mongo database will be migrated to Cosmos.

7. Predictive analytics

We have built models to predict the internal variables in the first entry point.

This model only have as input the variables taken by the satellite in the same point. The variables with more than 50 missing values has been removed too.

2.7 Temperature model

8. Front-end prototype

7.1 Design concepts

The system will consist of a graphic interface in the form of a web page. At the top of the page will be shown Acciona Agua logo. On the left side, there will be a button to choose the plant.

Below these, it will be divided into three buttons, one for obtaining variables, and another for predictions and the last one to create a new prediction system.

The first button will show in the body of the website the different satellites (in horizontal) that can be accessed and a description of these. Below, in drop-down menus, it will be the variables, the date of obtaining and the button to access them, the results will be displayed in a new window for each selected variable. It will show a map with the selected zone in the left part and the information referring to the variables in the right.

The second button, the predictive model, will show the prediction variables: temperature, conductivity and turbidity, the range of days for prediction and a button. When they are accessed, a window will open for each variable and a graph of the prediction will be shown on the left and information about it will be displayed on the right.

The third button will be for the start of the plant. If we access it will be shown in the body of the page the coordinates we want to predict, the satellites with a brief description of them and in a drop-down menu its variables, the time we want to go back to get the variables, a box to upload the file with the variables of the plant and the button to start the plant startup. If everything goes successfully, it will show on the page that everything went well and we will return to the initial state of entry in the page. If there were any error, the user would be informed.

Objeto

Documento borrador de inicio que nos puede ayudar a generar cada una de las actividades de desarrollo y necesidades de la plataforma de gestión y operación.

Podemos evolucionarlo un poco más como borrador donde están naciendo y dando forma a las ideas y posteriormente reconvertirlo o asociarlo al documento de diseño de la plataforma.

Este documento se irá manteniendo vivo hasta que el correspondiente entregable del diseño de la plataforma y sus funcionalidades esté terminado y cumpla con los requerimientos.

General issues

1. Autentificación:

a. ¿Cómo resolvemos este acceso?

i. Si la nube es corporativa TIC nos dará el acceso registrado y seguro

ii. Mientras no hay nube corporativa tenemos que ver como se realizará el acceso

2. Salida de datos:

a. La salida de datos se deberá poder mostrar de varias maneras:

i. Muestra en mapa de la predicción de alguna o varias de las variables a predecir

ii. Muestra en una gráfica (eje x tiempo) los valores de la predicción de una o unas determinadas variables

iii. Valor de acierto, coeficiente de error medio o KPI utilizado para la gestión del acierto y error en el modelo.

iv. La salida puede encadenarse como una gráfica con x valores reales más los n predichos.

3. Mapa:

a. Ver que App o API seleccionaremos para mostrar el mapa o generar coordenadas sobre un mapa para la localización de las instalaciones o coordenadas para una predicción

b. Propongo que uno de los campos visuales de la página de configuración sea mostrar en coordenadas y un espacio (cuadrado o rectángulo) la ubicación física de la planta o punto de trabajo. (Por ejemplo, a la derecha de la página de configuración podríamos ver el mapa que representa un cuadrado o rectángulo de trabajo o predicción donde dentro estarán las coordenadas de la planta.

4. Predicción:

a. Una opción básica será predecir únicamente las variables internas. Un ejemplo, si tenemos valores de Turbidez, que se pueda predecir solo la turbidez en un horizonte temporal.

b. La opción más compleja será la de predicción de variables utilizando otras variables (endógenas o exógenas).

c. Para poder decidir que variables son las más adecuadas para predecir se debería mostrar una pequeña matriz de correlación entre variables. De esta manera podremos saber que variables afectan de mayor manera en la progresión de las otras.

d. Tengo alguna duda sobre la predicción con modelos estacionales. Tendríamos que ver como se contempla esto en el modelo o modelos resultantes. La estacionalidad puede condicionar una predicción.

5. Colores de la plataforma:

a. Propongo utilizar los colores de ambas compañías y una base de colores adecuados para las combinaciones.

b. Azul/blanco: Artelnics

c. Rojo/Negro: Acciona

El sistema constará de una interfaz gráfica en forma de página web. En la cabecera de la página se mostrará el logo de Acciona Agua al lado en la parte izquierda estará un botón para elegir la planta, debajo de estos se dividirá en tres botones, uno para la obtención de las variables, otro para las predicciones y por último para crear un sistema nuevo de predicción.

Notas:

- Si va a ser una herramienta de uso repetitivo y de despliegue sugiero que repensemos este punto de selección de la planta.

- Propongo que sea un desplegable con las instalaciones donde se haya realizado la primera ingesta. Sería como un “dar de alta la instalación”. Paso previo a poder realizar ninguna actividad sobre esa actividad.

- Si consideramos que la planta “existe” entonces podrá ser seleccionada.

- Aquí hay un punto a revisar. Si se puede hacer predicción por coordenadas, se podría dar el caso en que “no estemos prediciendo” ningún valor en una instalación sino en una coordenada concreta.

- En la idea general es que este proceso de predicción e ingesta sea automatizado a lo largo del tiempo por lo que la carga “inicial” puede ser un protocolo de gestión y validación, antes de que pueda usarse como herramienta.

El primer botón mostrará en el cuerpo de la página web los distintos satélites (en horizontal) a los que se puede acceder y una descripción de estos, debajo en menús desplegables estarán, las variables, la fecha de obtención y el botón para acceder a ellas, los resultados se desplegarán en una nueva ventana para cada variable seleccionada, en ella se mostrará un mapa con la zona seleccionada en la parte izquierda y la información referente a las variables en la derecha.

Notas:

- Tenemos que diseñar la gestión de las peticiones de información e ingesta de los satélites.

- Puede ser una petición de usuario, una demanda o una petición automatizada que va alimentando el repositorio de información de la base de datos.

- Deben de quedar claros los datos accesibles de cada satélite, así como la capacidad de seleccionar uno o varios en función de cuántos queremos meter al visualizador y modelo predictivo.

- No tenemos que olvidarnos que también tendremos fuentes internas o de planta. Deberemos de gestionar de igual manera su selección e ingesta.

El segundo botón, el del modelo predictivo mostrará las variables de predicción temperatura, conductividad y turbidez, el rango de días para la predicción y un botón. Cuando se acceda a ellas se abrirá una ventana por cada variable y se mostrará en la parte izquierda una gráfica de la predicción y en la parte derecha la información referente a ella.

Notas:

- La variable o variables a “predecir” ha de poderse seleccionar en función del usuario (o fijo, pero definido por un superusuario).

- En función de cada planta habrá unas variables a predecir.

- Si el modelo y librerías son estándar, las variables a definir deberán ser seleccionables por el usuario de la herramienta de análisis.

- Tenemos por tanto una zona de variables seleccionables donde decimos que son las que queremos predecir

- Tenemos otra zona de variables donde decimos que son las que utilizaremos para hacer predicciones sobre las variables a predecir.

- Tendremos un elegible temporal para la predicción. En principio propongo que se establezcan 3 horizontes para la predicción, 1, 3 y 7 días.

El tercer botón será para el arranque de planta, si accedemos a él se mostrará en el cuerpo de la página las coordenadas que queremos predecir, los satélites con una breve descripción de ellos y en un menú desplegable sus variables, el tiempo que queremos retrotraernos para obtener las variables, un cajetín para subir el archivo con las variables de la planta y el botón para empezar el arranque de la planta, si todo sale con éxito se mostrará en la página que todo ha salido bien y volveremos al estado inicial de entrada en la página, si hubiera algún error se le informaría al usuario.

Notas:

- De cada fuente de datos (o variable) deberemos saber cuándo fue la última ingesta o el último valor obtenido. Esto puede ser bueno en las fuentes de datos en que se puede seleccionar el umbral temporal de selección.

- Tenemos que hacer una página de información básica de la plataforma:

o Número de variables almacenadas

o Espacio almacenado (GB)

o Espacio temporal almacenado

o Última predicción realizada

End user wireframe

Los requisitos funcionales del sistema propuesto deberían de cumplir:

1. Visualizar información de forma clara, concisa y concreta, abogar por una buena organización de la información.

2. Insertar información: debe de ser flexible a nuevas adiciones de datos, en la web de usuarios que puedan añadir nuevos ficheros de datos en formato Excel, en caso de existir ya la información en la base de datos el sistema deberá generar un mensaje preguntando si desea actualizar la información existente, control duplicados en base de datos.

3. Necesidad de poder modificar información, en caso de necesitar cambiar variables internas y externas en la página de usuario, con los checkbox existentes, así como en la web administración.

4. Eliminar información previa, en caso de equívoco, en la web administración.

5. Descargar información, se precisa disponer de un botón exportar, en ambas páginas web.

6. Posibilidad de buscar nuevas informaciones, en la página administrador.

7. Actualizar estadísticas automáticamente.

8. Eliminar estadísticas si el usuario lo desea.

9. Permitir autenticación corporativa.

10. Presentación del Wireframing, prototipo inicial rápido de la página web, a revisar posteriormente.

11. Flexibilidad y eficiencia de uso: el diseño debe de ser “web resposive”.

12. Información y diseño minimalista: evitar información o gráficos irrelevantes. Cada elemento debe distinguirse claramente del resto e, idealmente, todo debe estar al alcance de un click. El texto debe ser fácil de leer y estar bien organizado, sabiendo en todo momento qué está seleccionado, en que instalación nos encontramos y sobre que variables se está haciendo la predicción.

13. Adaptar el diseño a los usuarios: Presentar contenidos de forma lógica, utilizando iconos y botones claros, un lenguaje simple. Aplicando filosofía de menos es más.

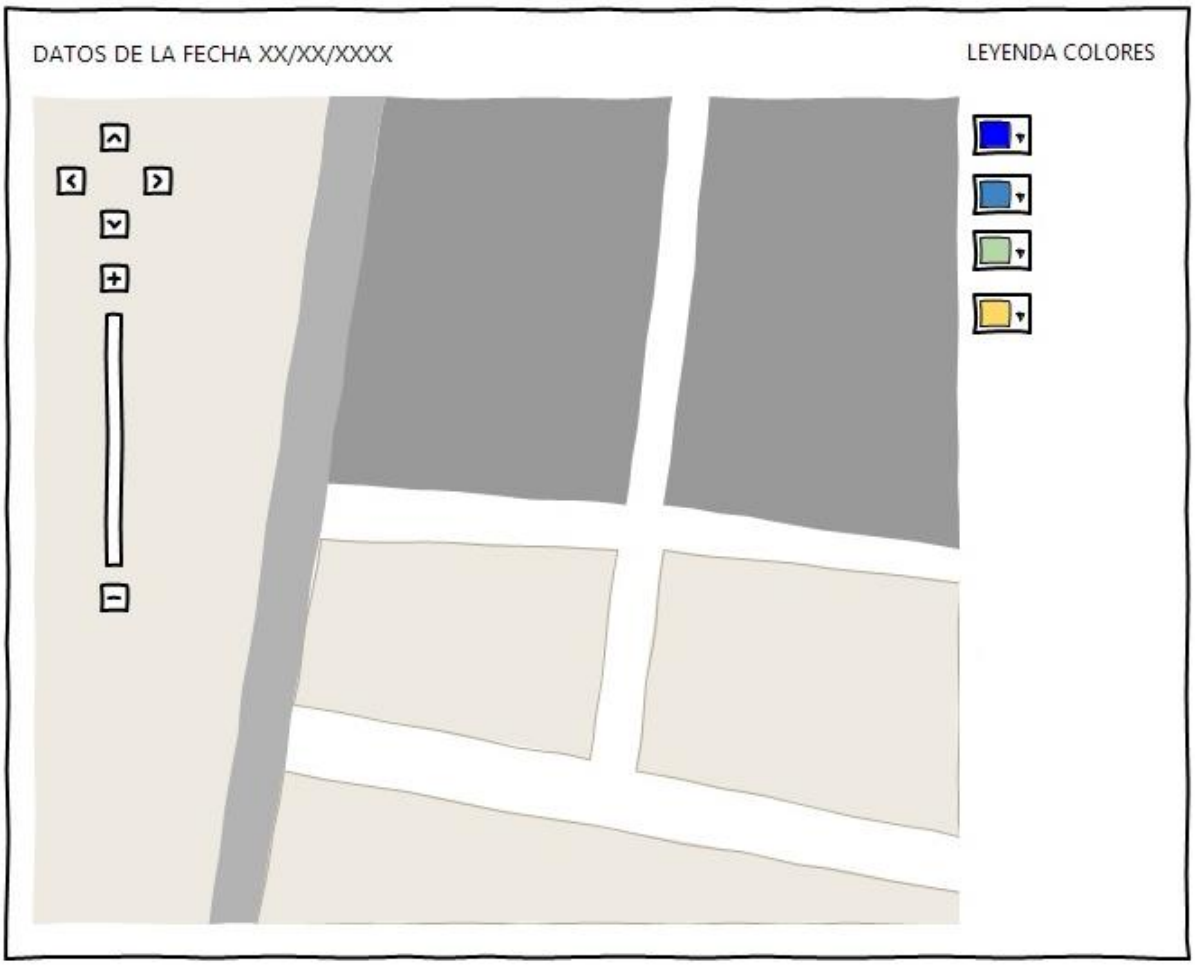
14. El sistema debe garantizar el control en el acceso, utilizando la autentificación de los usuarios para la administración del mismo, por tanto, debe de permitir asignar que usuarios pueden ver que instalaciones, que en función de los permisos verá en el árbol lateral aquellas a las que se les ha dado permisos.

Wireframing Usuario Final Jessica:

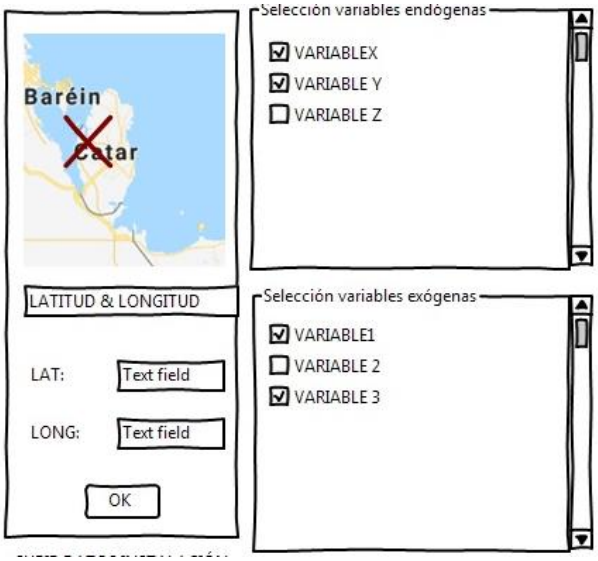
- El usuario tendrá acceso a las instalaciones que el administrador le haya concedido, por tanto, en la zona izquierda de navegación web solo aparecerán aquellas instalaciones a las que tenga acceso.

- El título principal y el contenido de la web (mapa, variables, etc.) llevan información de la instalación seleccionada, información que variará si el usuario clicka en otra.

TITULO DE LA APP-INSTALACIÓN SELECCIONADA: QATAR

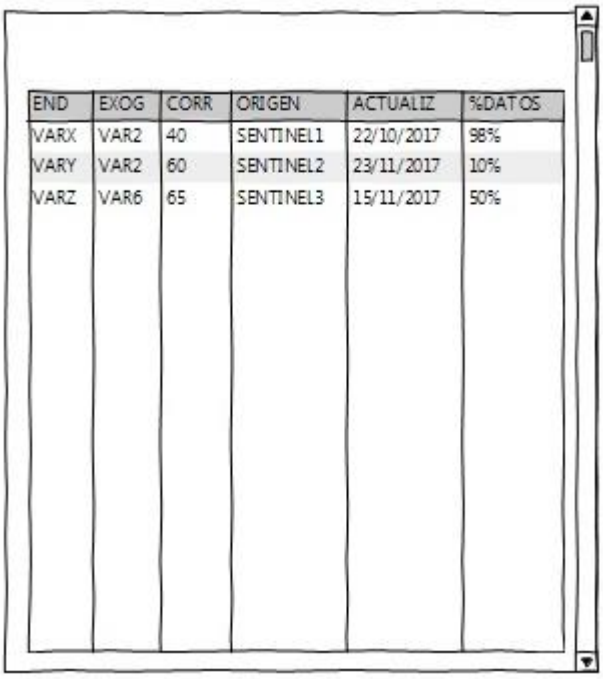


En la parte central se cargarán los datos de la instalación seleccionada en las coordenadas introducidas como geometría de la zona en la página de administración, junto con la leyenda de colores (que esto habrá que añadirlo en la página de administración).

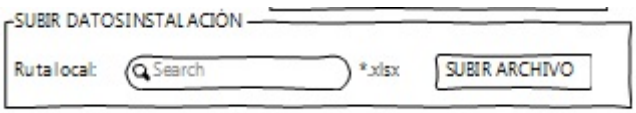


Además, se mostrará en detalle el punto donde se realiza el estudio, a modo de zoom, modificable por el usuario, por si hubiese cualquier error de localización

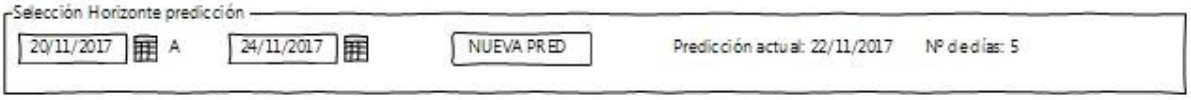
Y las variables correspondientes a la instalación seleccionada, donde el usuario podrá seleccionar para la realización de la predicción, esta decisión de añadir variables será gracias a la presencia de la siguiente matriz, ya que nos muestra información sobre la correlación de variables, orígenes de datos, fecha última actualización, % de información obtenida, etc.



Se permitirá subir ficheros localmente al usuario evitando duplicidad de información en la base de datos.



Así como reconfigurar un nuevo horizonte de predicción que se actualizará al clickar en el botón.

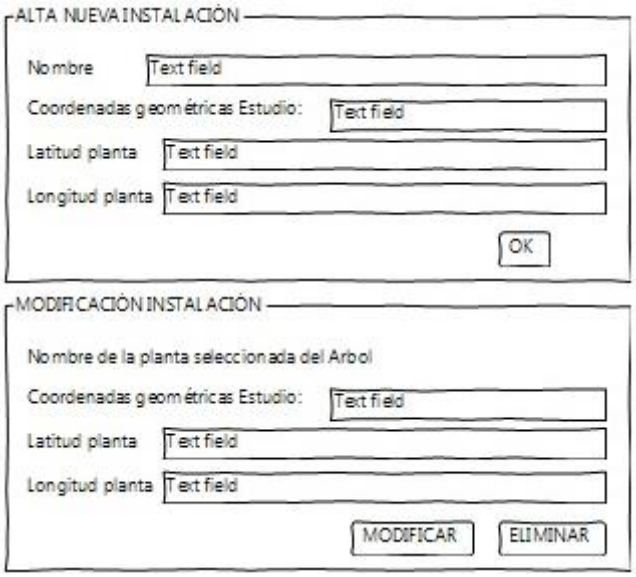


Y en la parte inferior se mostrarán los resultados, tanto las variables en bruto para poder exportarlo a csv, como la gráfica resultante y unas estadísticas del mismo.

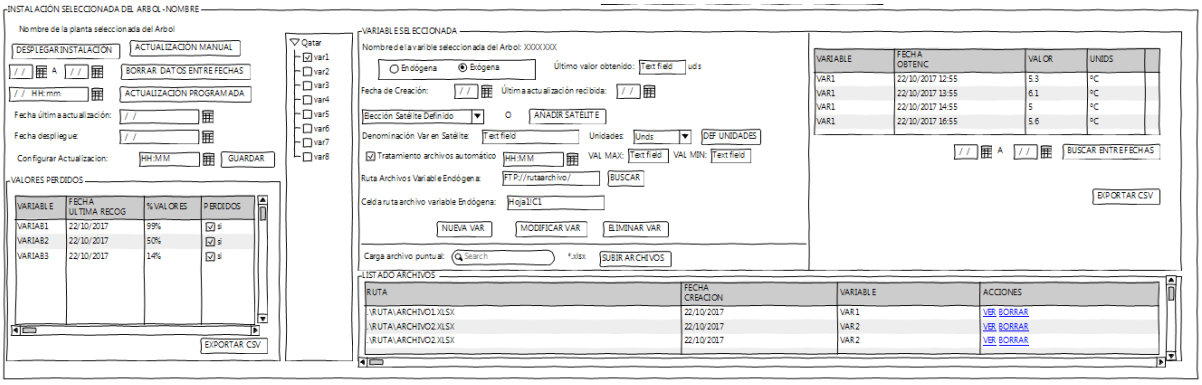
Administrator wireframe

- En la parte izquierda de la página se deberán mostrar el listado global de instalaciones definidas (estén o no estén desplegadas, una forma de diferenciarlas sería modificando el color del nombre en el árbol), deberá estar validado que no exista una instalación ya definida igual a una existente.

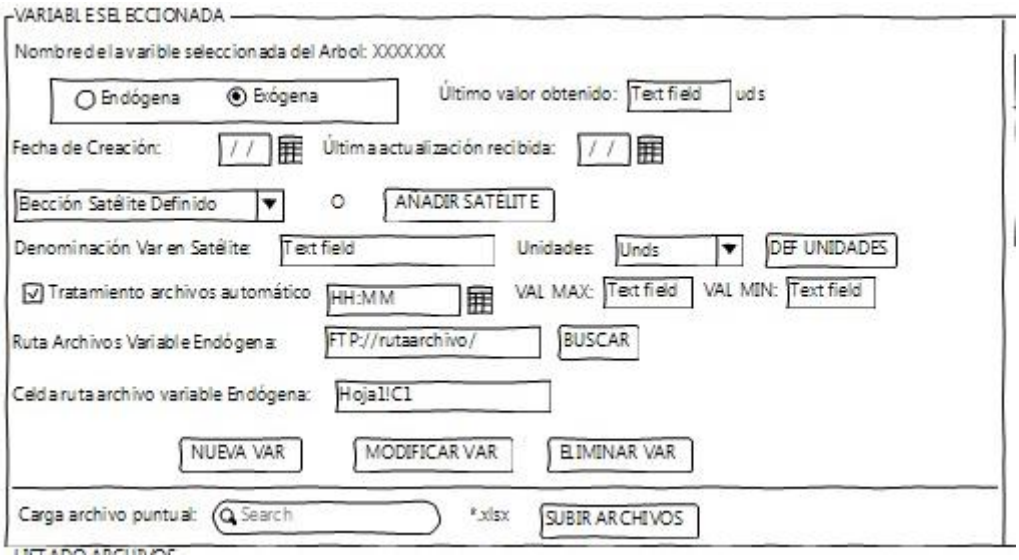
- Para realizar cualquier modificación o eliminación de una instalación se deberá clickar en el árbol izquierdo de la pantalla y se mostrará siempre el nombre seleccionado para conocer sobre qué instalación se realizarán las acciones, además como seguridad extra se deberá volver a preguntar al usuario si está seguro de las acciones a realizar.



- En la parte inferior de la pantalla, aparecen los detalles de la instalación seleccionada, donde el administrador podrá configurar las variables en mayor detalle, por tanto el árbol central donde aparecen las variables cambiará en función de la selección realizada en el árbol de instalaciones, en esta zona, se podrá configurar el borrado de datos al igual que el anterior punto y se deberá preguntar nuevamente al administración si está seguro de la acción a realizar, se podrá desplegar la aplicación, actualizar manualmente la recogida o programarla para que a partir de la fecha introducida cada 24 horas realice la recogida.

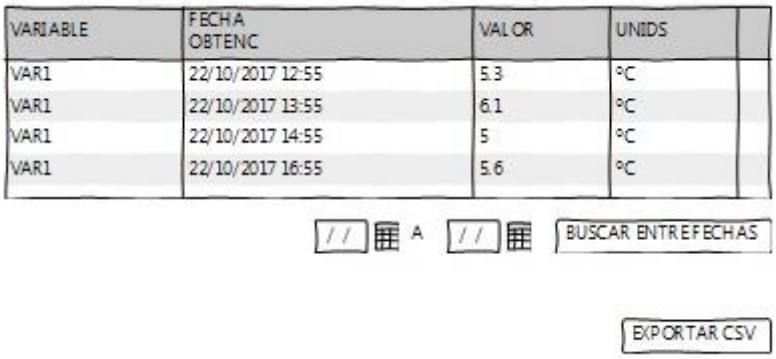


Se permite añadir satélites nuevos, definir tipología de la variable, último valor obtenido, fecha de creación, creación de unidades de variable, etc.



En esta zona además permitiría subir ficheros puntuales o automatizar esta tarea, es decir, definir una ruta y que el sistema a la hora configurada se encargue de escanear la ruta y tratar los ficheros, controlando la duplicidad de la información, los archivos aparecerán en un listado donde el administrador tendrá acceso a la ruta donde se han almacenado después de su tratamiento.

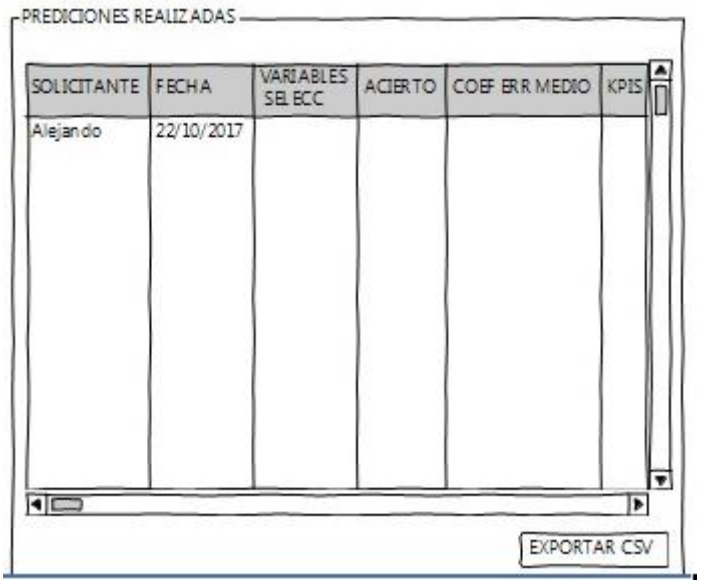
Se podrán realizar búsquedas de las variables entre fechas y exportar el resultado a csv , por si hay que realizar cualquier tratamiento de los datos para su estudio.



En la parte superior central de la pantalla será donde asigna el administrador los permisos a los usuarios ya definidos para las instalaciones de la aplicación.

A la derecha de los permisos, creo que es interesante mostrar información referente a la base de datos, variables, espacios, alarmas y rutas de acceso.

Seguidamente un histórico de las predicciones realizadas, creo que es útil para ver un resumen de las situaciones acontecidas por los usuarios, con exportación a formato CSV.



ANOTACION FINAL Jessica: Creo conveniente una gestión de errores de comunicaciones con los satélites, quizá en otra pestaña, además un log para poder ver donde puede estar fallando la recogida de datos y poder actuar en consecuencia, este punto habrá que dar una vuelta más.

Comentarios de Jessica:

Añadir nuevas variables satélite

Archivo de planta

Mensajes de advertencia en la interfaz de administrador.

9. Tools and libraries

Standard C++ 11

The API will be programmed in this language.

OpenNN

Neural network library used for the pre-processing of the data and the training of the model.

Python

A Python script will be used to download the data from the Sentinel satellite.

Graph Processing Tool

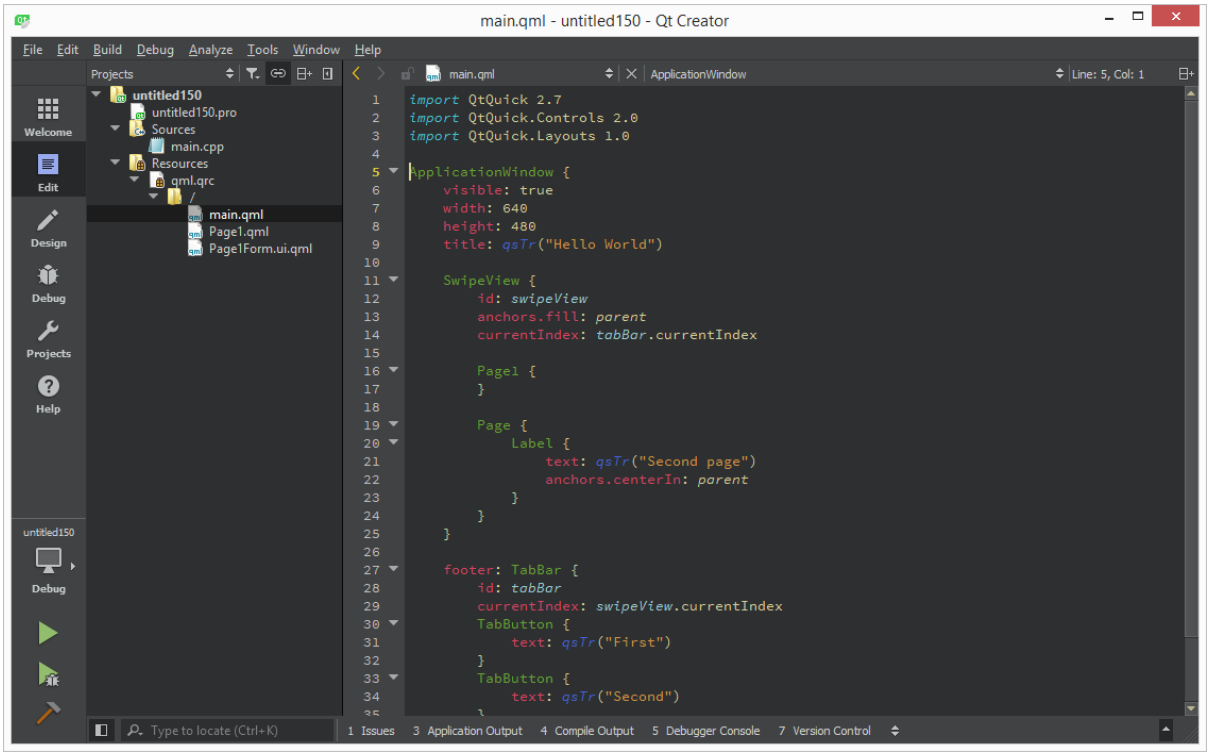
It is used to join the NETCDF data files and convert them in a csv file.

Qt Creator

Multiplatform IDE used to compile and create the API.

This environment can use the Qt libraries to ease some tasks of the API.

ireire



1.8. Microsoft Azure

Grupo de Recursos RG_AGUA_PoC_CECOI_Artelnics

Azure Active directory para la gestión de usuarios.

Control de versiones?

Pablo Peris pablop@microsoft.com

Poner a José y Marian en copia en los emails.

Daniel Tabuas databuas@acciona.com Manager Data Science

Mariano Muñoz mamunozm@acciona.com

Recursos:

- Blob Storage
- CosmosDB
- Maquina Virtual
- Website

Definir que es