Preparation of training data sets from ground truth data

Genki Tetrauchi, Akira Kozuka

November 30, 2021

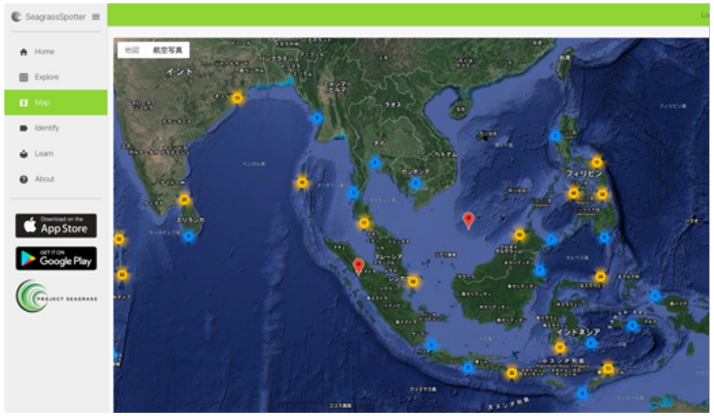
CEARAC Webinar

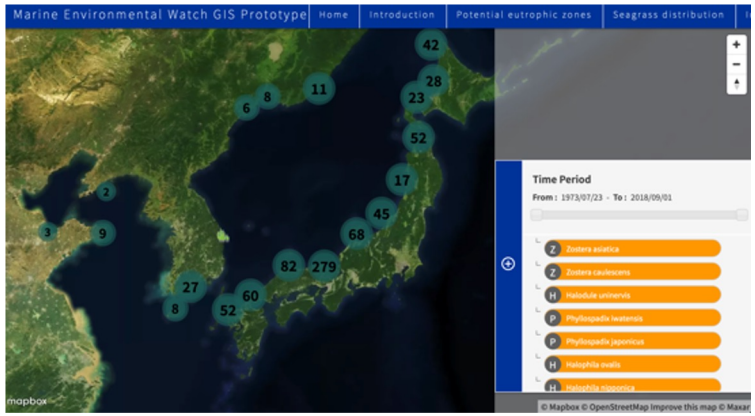
3 basic steps to prepare training data

- Collections of field information
  - Field sampling
  - Existing database or publication
- Defining substrate class
  - Seagrass, Seaweed, Sand...
- Creating features in defined substrate class
  - Google Earth Pro
  - QGIS

Collection of field information

Observing seafloor by under water video camera and recording location by GPS





Collecting information from existing database

Defining substrates classes from collected information

sparse seagrass

Sand

Dense seagrass

Seagrass with sargassum

Sargassum

Creating features

Features can be in form of points, lines and polygons in shp, kml, kmz, csv and geojoson

Points

Lines

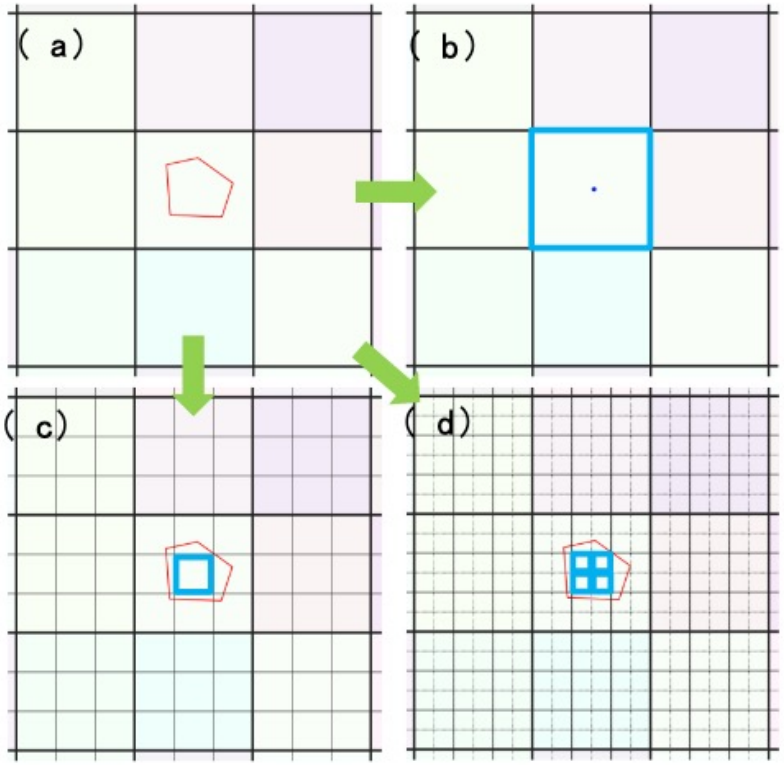
Polygons

Creating point, line and polygon features on Google Earth Pro

Something to remember in preparing training data

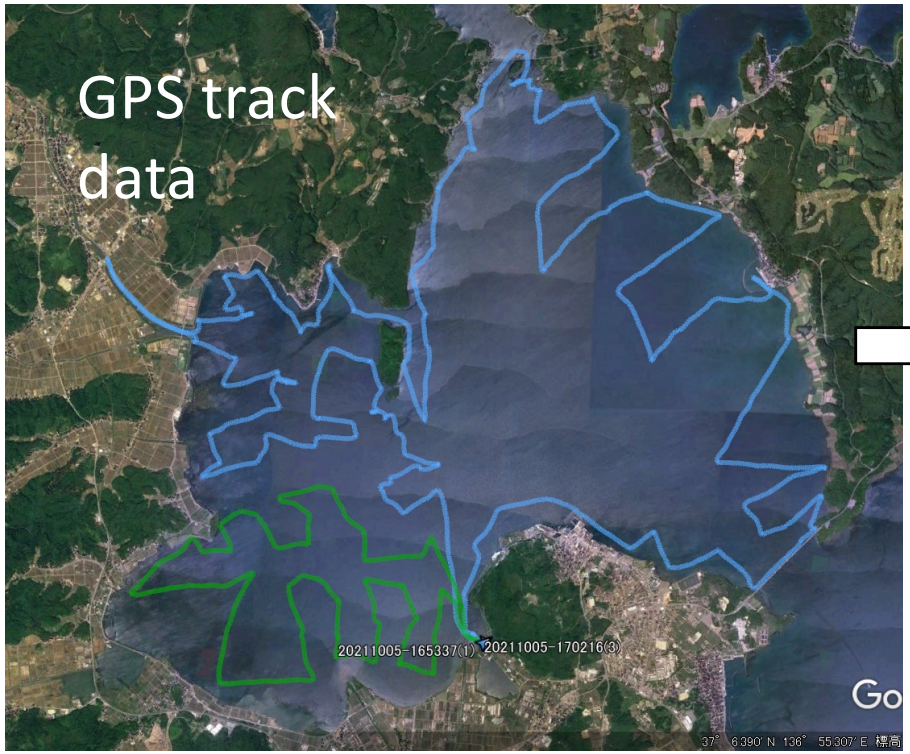
There are 4 types of sampling scale to be used

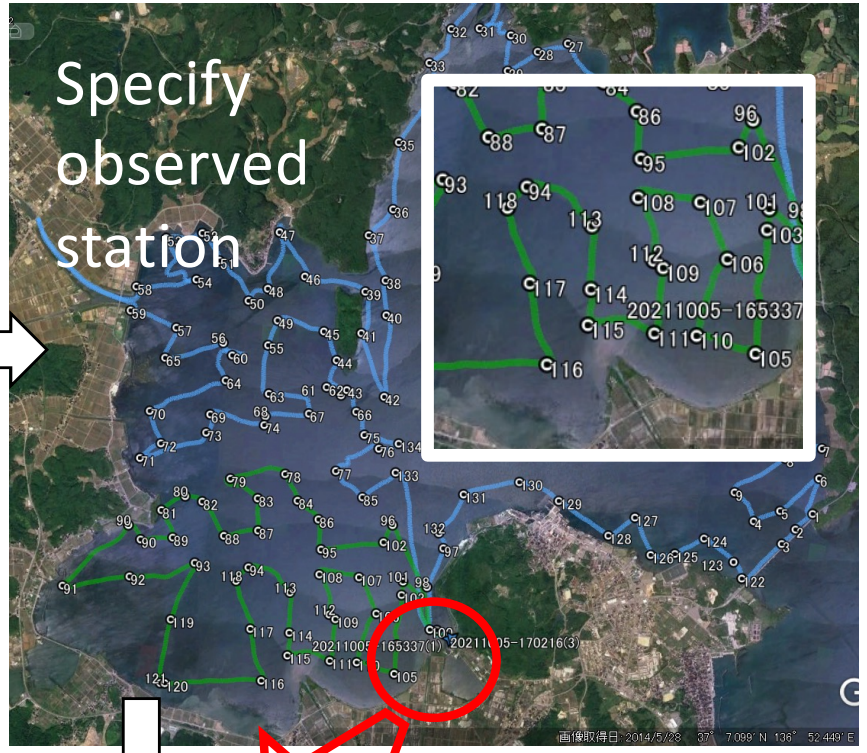
1. Original: same resolution of the satellite (e.g. Landsat 8 OLI = 30m)
2. 10m
3. 5m
4. 1m

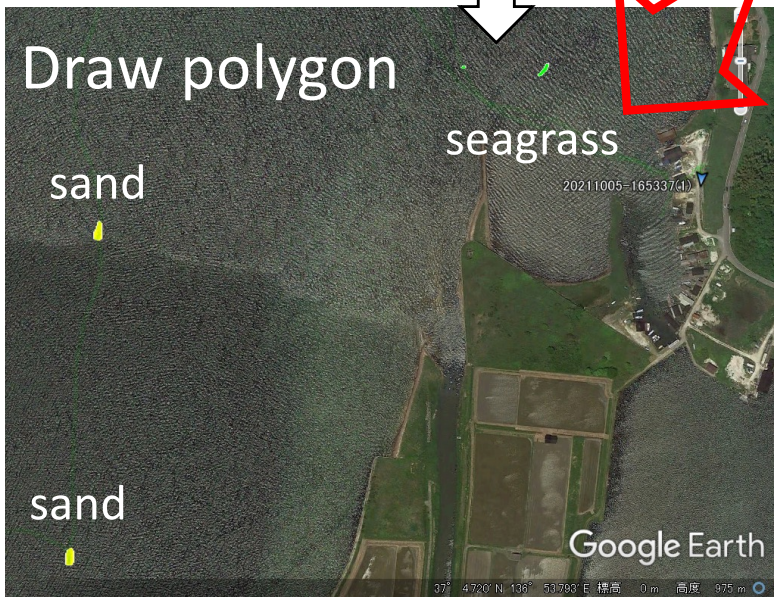


The above image shows how to work around when training data in a polygon format for classification is smaller than satellite resolution (here we use Landsat 8 OLI as an example); (a) training data will be ignored when the original sampling scale (30 m) is set, (b) polygon data are converted into point data so as to be used as training data, (c) training data will be used when sampling scale is changed to 10m, and (d) even 4 matching training data are found within the same satellite pixel, and they will be counted as one data

Example: Nanao bay







Make kmz files for each polygons

- 1, Equipment for field survey
- 2, Research plan
- 3, Field data processing
- 4, Drawing polygon data

1, Equipment for field survey

Necessary tools

- Digital watch, field note
- Underwater observation device: video camera, underwater drone, water glass
- GPS logger device: handy GPS, fishfinder, RTK-PS
- Depth measurement device: rope, fishfinder
- Small boat: 1t< for shallow area
- Battery: for video, GPS, tablet PC

Convenient tools

- GPS display device: handy GPS, tablet PC
- Operation record device: Video on boat, depth logger for underwater camera,
- Shade for watching monitor

Neccesary tool

- Digital watch
- Video camera
- GPS logger device

  At least 2 devices for safety

  Note the recording capacity

  For 10 hours, 36000 points(by 1 second)
- Fishfinder
- Small boat

GPS logger i-gotU GT-600

Fishfinder “Deeper”

Convenient tools

- depth logger for underwater camera

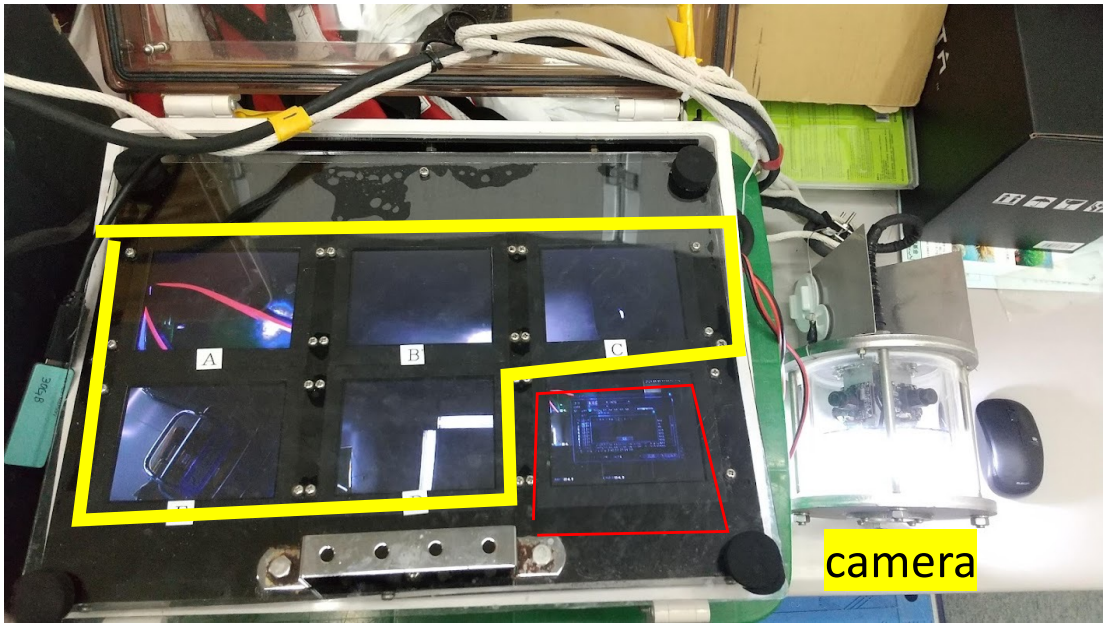
Depth logger”DEFI2-D”
JFE Advantech Co.,Ltd

About underwater video camera

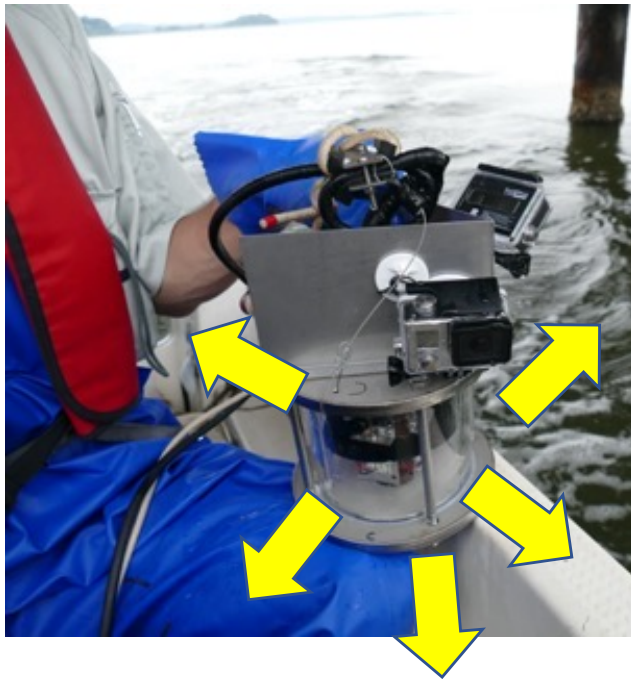
Custom-made underwater cameras which cost about 6000 USD

It shows five directions at the same time, allowing to see below and on all sides. It also records the video.

Camera of one direction is also good, recording function is desirable.



5 display screens and 1 operation screen



5 direction at the same time

Before field survey

Check the watch.

Adjust the watches of the survey equipment in seconds accurately .

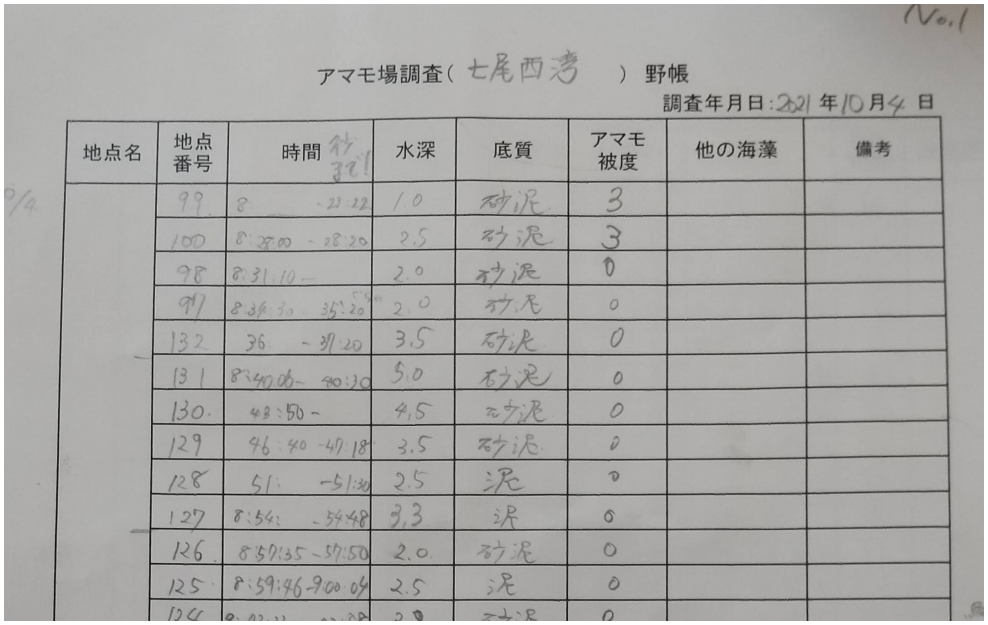
Digital watch, Video camera, GPS logger device, Depth logger

Register survey points in GPS display device.

To revisit past research points.

GPS logger is suggested to be set to record by 1 second interval if possible.

Field note



アマモ場調査（七尾西湾）野帳

調査年月日：2021年10月4日

地点名	地点番号	時間	水深	底質	アマモ被度	他の海藻	備考
	99	8 -21:22	1.0	砂泥	3		
	100	8:28:00 -28:20	2.5	砂泥	3		
	98	8:31:10 -	2.0	砂泥	0		
	97	8:34:30 - 35:20	2.0	砂泥	0		
	132	36 - 39:20	3.5	砂泥	0		
	131	8:40 - 40:30	5.0	砂泥	0		
	130	43:50 -	4.5	砂泥	0		
	129	46:40 - 47:18	3.5	砂泥	0		
	128	51: - 51:30	2.5	泥	0		
	127	8:54: - 54:48	3.3	泥	0		
	126	8:59:35 - 57:50	2.0	砂泥	0		
	125	8:59:46 - 9:00:04	2.5	泥	0		

- Water-resistant paper,
- Write with pencil

Record items

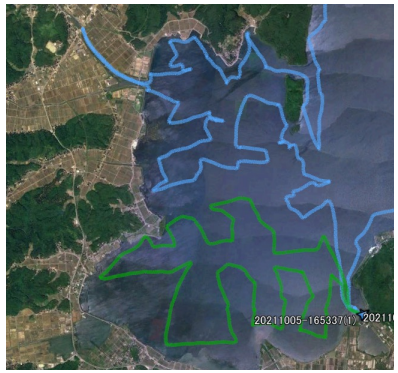
- Station number
- Time: start time of observation and end time of observations in seconds
- Depth
- Substrates: sand, muddy, rock and mix
- Density: 0~5
- Remarks: sargassum etc.

Density	Percentage
0	0
1	1-10
2	11-25
3	26-50
4	51-75
5	76over

Dense

Don't let the wind blow the paper lol

2, Research plan



- Number of observation stations

60 points are recommended as training data for one satellite image (Japan Fisheries Agency,2021).

In Nanao bay, we surveyed 134 stations for 15km^2: 9 stations per km^2.

- Time

  1 point takes a few minutes

In Nanao bay, it takes approximately 10 hours by 2 days for 93km cruise of 134 stations including resting time.

- Staff

  At least 3 person: Boat operator, camera operator, observer of video and recording note.

Field survey

Boat operator with handy GPS

camera operator and watching video

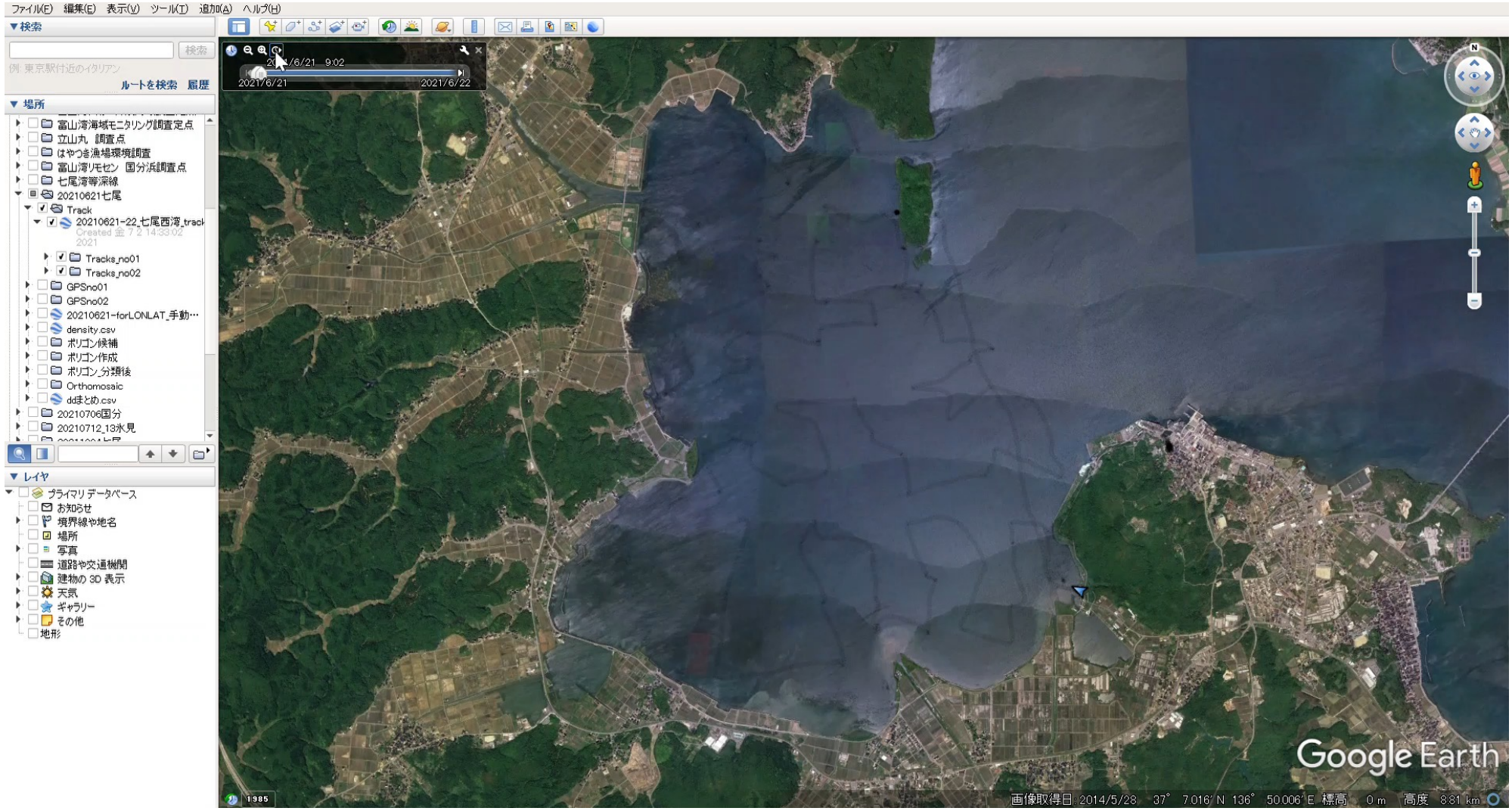
Strange hat for sun protection

Recorder with field note

Video image　Dense seagrass

Video image Muddy water

Cruise animation



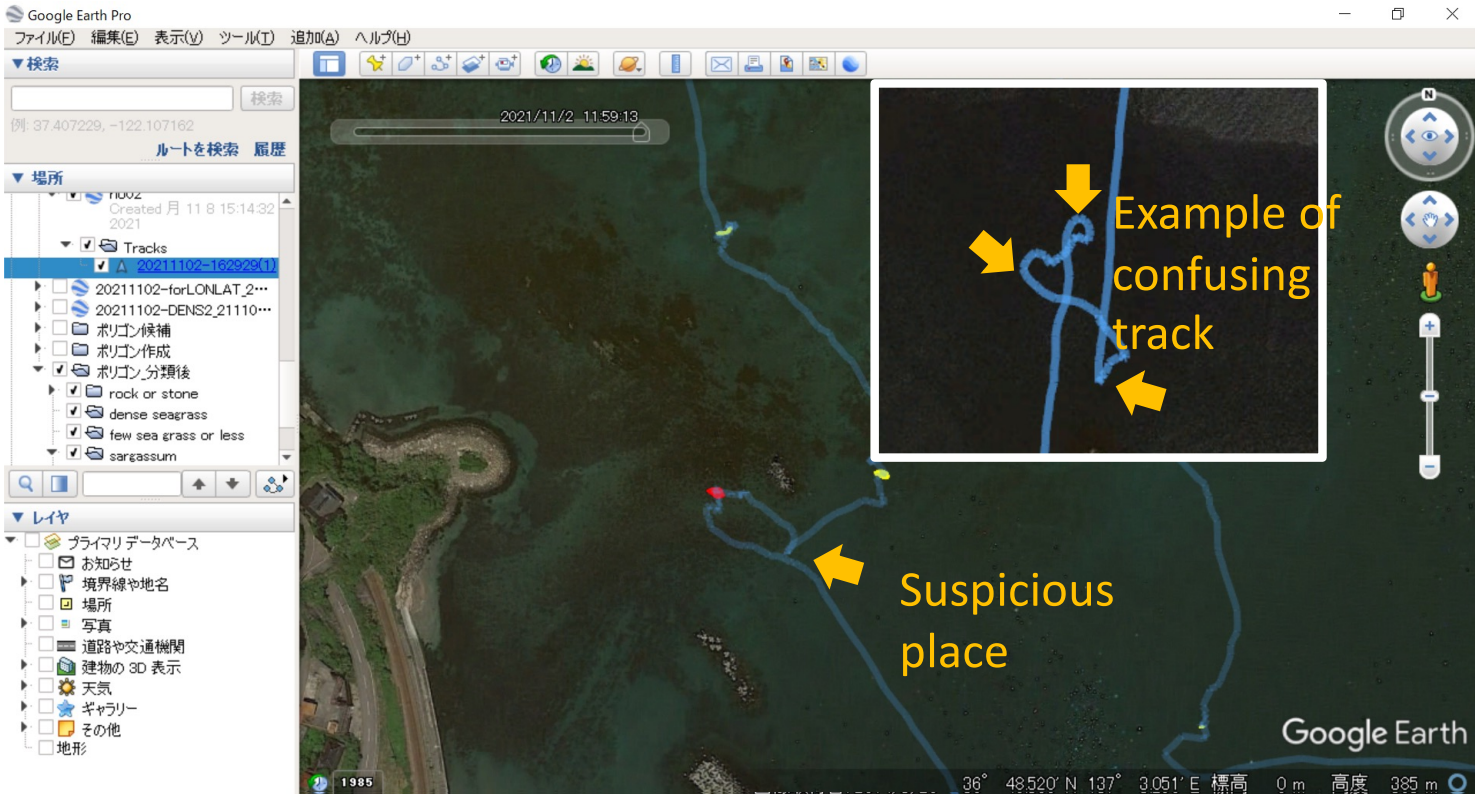
3, Field data processing

- Link the field note(csv file) and GPS information (csv file) based on the time in seconds for preparing training data.

Today, we will explain an efficient way for field data processing.

Exception Case : number of stations is a few

Draw GPS tracks and specify survey points by track appearance



Advantages : Operation is easy

Disadvantages : Not objective

Take many time to check suspicious place

and requires a lot of concentration and time

Case : stations over 50

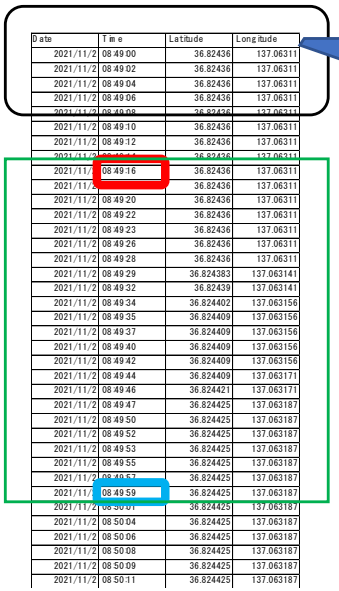
- Extract location data of observed points from all GPS data based on time from fieldnote.

Fieldnote data

station	Date	time_start	time_end	depth	substrates	density
1	2021/11/2	8:49:15	8:49:44	27	s	0
2	2021/11/2	8:56:56	8:57:10	19	s	0
3	2021/11/2	9:02:50	9:03:12	15	s	0
4	2021/11/2	9:09:40	9:09:55	12.5	s	0

Observation result

GPS data



Date	Time	Latitude	Longitude
2021/11/2	08:49:00	36.82436	137.06311
2021/11/2	08:49:02	36.82436	137.06311
2021/11/2	08:49:04	36.82436	137.06311
2021/11/2	08:49:06	36.82436	137.06311
2021/11/2	08:49:10	36.82436	137.06311
2021/11/2	08:49:12	36.82436	137.06311
2021/11/2	08:49:16	36.82436	137.06311
2021/11/2	08:49:20	36.82436	137.06311
2021/11/2	08:49:22	36.82436	137.06311
2021/11/2	08:49:23	36.82436	137.06311
2021/11/2	08:49:26	36.82436	137.06311
2021/11/2	08:49:28	36.82436	137.06311
2021/11/2	08:49:29	36.824383	137.063141
2021/11/2	08:49:32	36.82439	137.063141
2021/11/2	08:49:34	36.824402	137.063156
2021/11/2	08:49:35	36.824409	137.063156
2021/11/2	08:49:37	36.824409	137.063156
2021/11/2	08:49:40	36.824409	137.063156
2021/11/2	08:49:42	36.824409	137.063156
2021/11/2	08:49:44	36.824409	137.063171
2021/11/2	08:49:46	36.824421	137.063171
2021/11/2	08:49:47	36.824425	137.063187
2021/11/2	08:49:50	36.824425	137.063187
2021/11/2	08:49:52	36.824425	137.063187
2021/11/2	08:49:53	36.824425	137.063187
2021/11/2	08:49:55	36.824425	137.063187
2021/11/2	08:49:57	36.824425	137.063187
2021/11/2	08:49:59	36.824425	137.063187
2021/11/2	08:50:01	36.824425	137.063187
2021/11/2	08:50:04	36.824425	137.063187
2021/11/2	08:50:06	36.824425	137.063187
2021/11/2	08:50:08	36.824425	137.063187
2021/11/2	08:50:09	36.824425	137.063187
2021/11/2	08:50:11	36.824425	137.063187

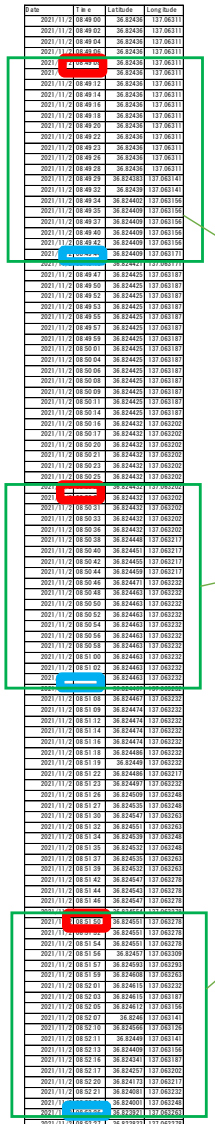
Extract GPS data based on time from fieldnote

Date	Time	Latitude	Longitude
2021/11/2	08:49:00	36.82436	137.06311
2021/11/2	08:49:02	36.82436	137.06311

Fieldnote data

station	Date	time_start	time_end	depth	substrates	density
1	2021/11/2	8:49:15	8:49:44	27	s	0
2	2021/11/2	8:56:56	8:57:10	19	s	0
3	2021/11/2	9:02:50	9:03:12	15	s	0

Observation result

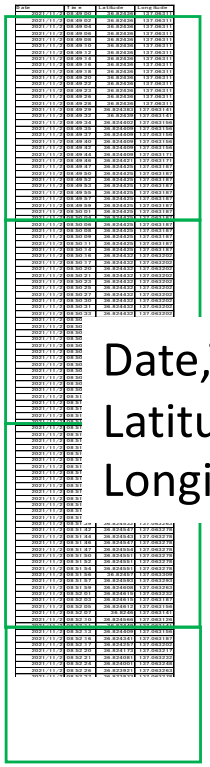


Extract data between start-time and end-time for each station

GPS track St.01

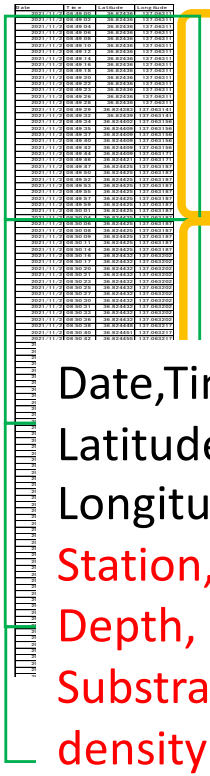
St.02

St.03



Date,Time, Latitude, Longitude

Match timestumps and merge observation result for stations



GPS track St.01

St.02

St.03

Observation result St.01

result St.02

Date,Time, Latitude, Longitude, Station, Depth, Substrates, density

t St.03

Save file as "**polygonbase_all**.csv"

※Extracting and merging data is processed automatically by R in NPEC.

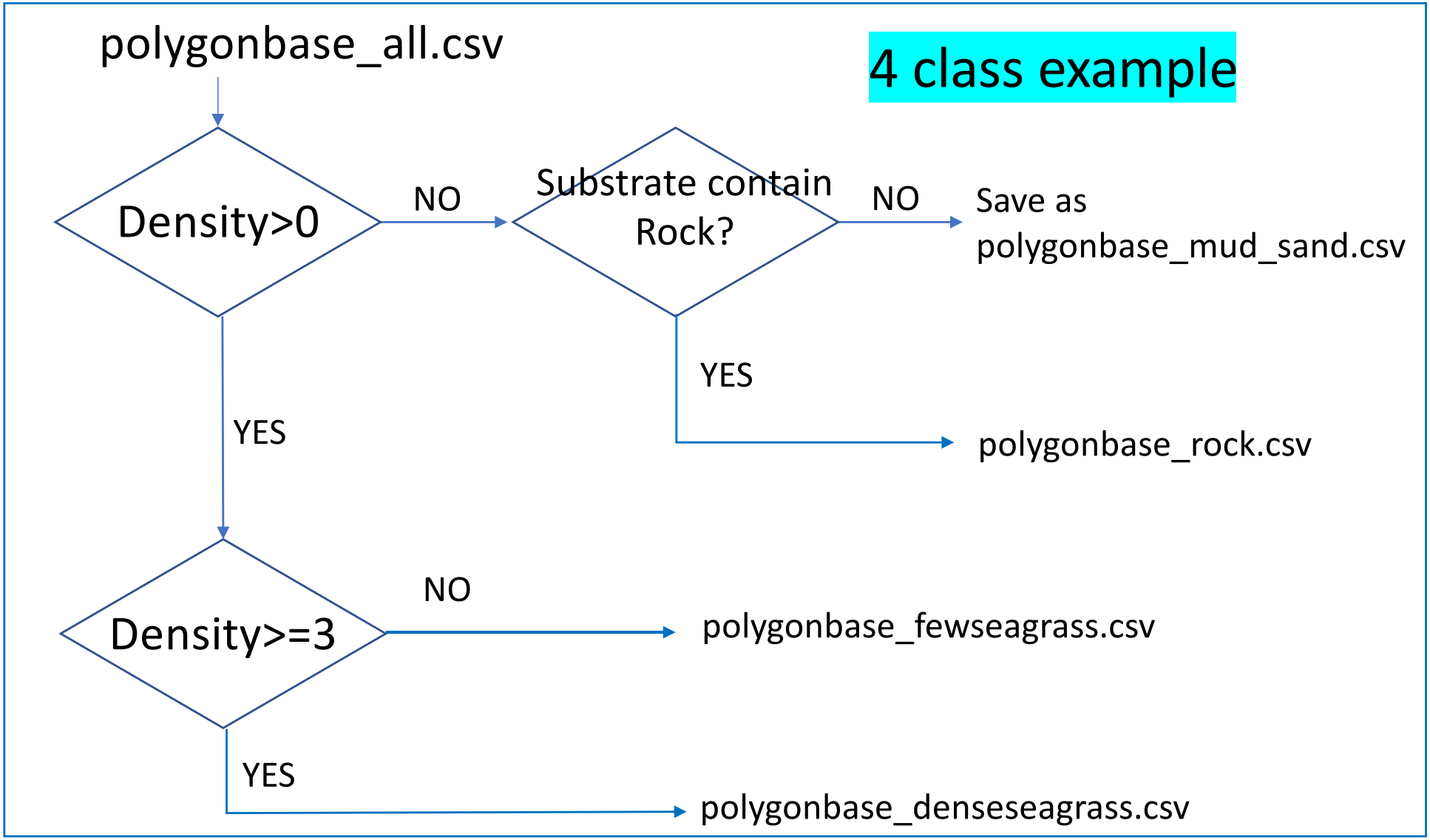
How training dataset would look like “polygonbase_all” data

station	Latitude	Longitude	Date	time	depth	substrates	density
99	37° 4.758'	136° 54.265'	2021/6/21	09:19:12	1	s	5
99	37° 4.76'	136° 54.267'	2021/6/21	09:19:45	1	s	5
99	37° 4.76'	136° 54.267'	2021/6/21	09:19:47	1	s	5
99	37° 4.76'	136° 54.267'	2021/6/21	09:19:52	1	s	5
99	37° 4.76'	136° 54.267'	2021/6/21	09:19:53	1	s	5
100	37° 4.762'	136° 54.207'	2021/6/21	09:21:49	1.5	s	5
100	37° 4.762'	136° 54.206'	2021/6/21	09:21:51	1.5	s	5
100	37° 4.761'	136° 54.205'	2021/6/21	09:21:53	1.5	s	5
100	37° 4.761'	136° 54.205'	2021/6/21	09:21:56	1.5	s	5
100	37° 4.761'	136° 54.205'	2021/6/21	09:21:58	1.5	s	5
100	37° 4.761'	136° 54.205'	2021/6/21	09:21:59	1.5	s	5
100	37° 4.761'	136° 54.205'	2021/6/21	09:22:01	1.5	s	5
116	37° 4.392'	136° 53.092'	2021/6/21	10:22:18	2	s/m	5
116	37° 4.393'	136° 53.091'	2021/6/21	10:22:20	2	s/m	5
116	37° 4.392'	136° 53.091'	2021/6/21	10:22:22	2	s/m	5
116	37° 4.391'	136° 53.091'	2021/6/21	10:22:25	2	s/m	5

Multiple location information for the same point
Latitude and longitude associated with features such as depth, substrates and density

Devide “polygonbase_all.csv” data before drawing on Mapping tool(eg. Google Earth Pro)

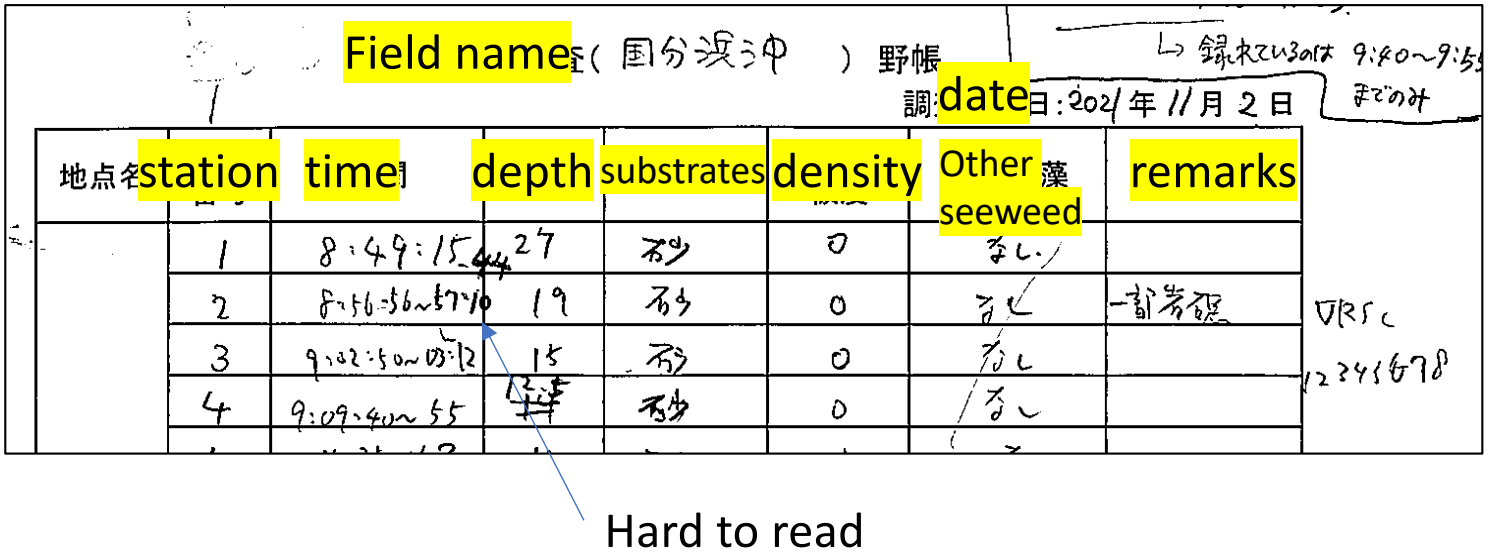
Split the file to the number of classes you want to create



Be aware of data transcription mistakes

- The more survey points you have, the more likely you are to make a mistake. Check at least two times.

Fieldnote original



Fieldnote converted to digital data

station	Date	time_start	time_end	depth	substrates	density
1	2021/11/2	8:49:15	8:49:44	27	s	0
2	2021/11/2	8:56:56	8:57:10	19	s	0
3	2021/11/2	9:02:50	9:03:12	15	s	0
4	2021/11/2	9:09:40	9:09:55	12.5	s	0

Advanced technique

- NPEC also use the information from the depth logger, since the field note sometimes includes mistakes and writing them and copy to digital data is hard work .

Depth logger”DEFI2-D”
JFE Advantech Co.,Ltd

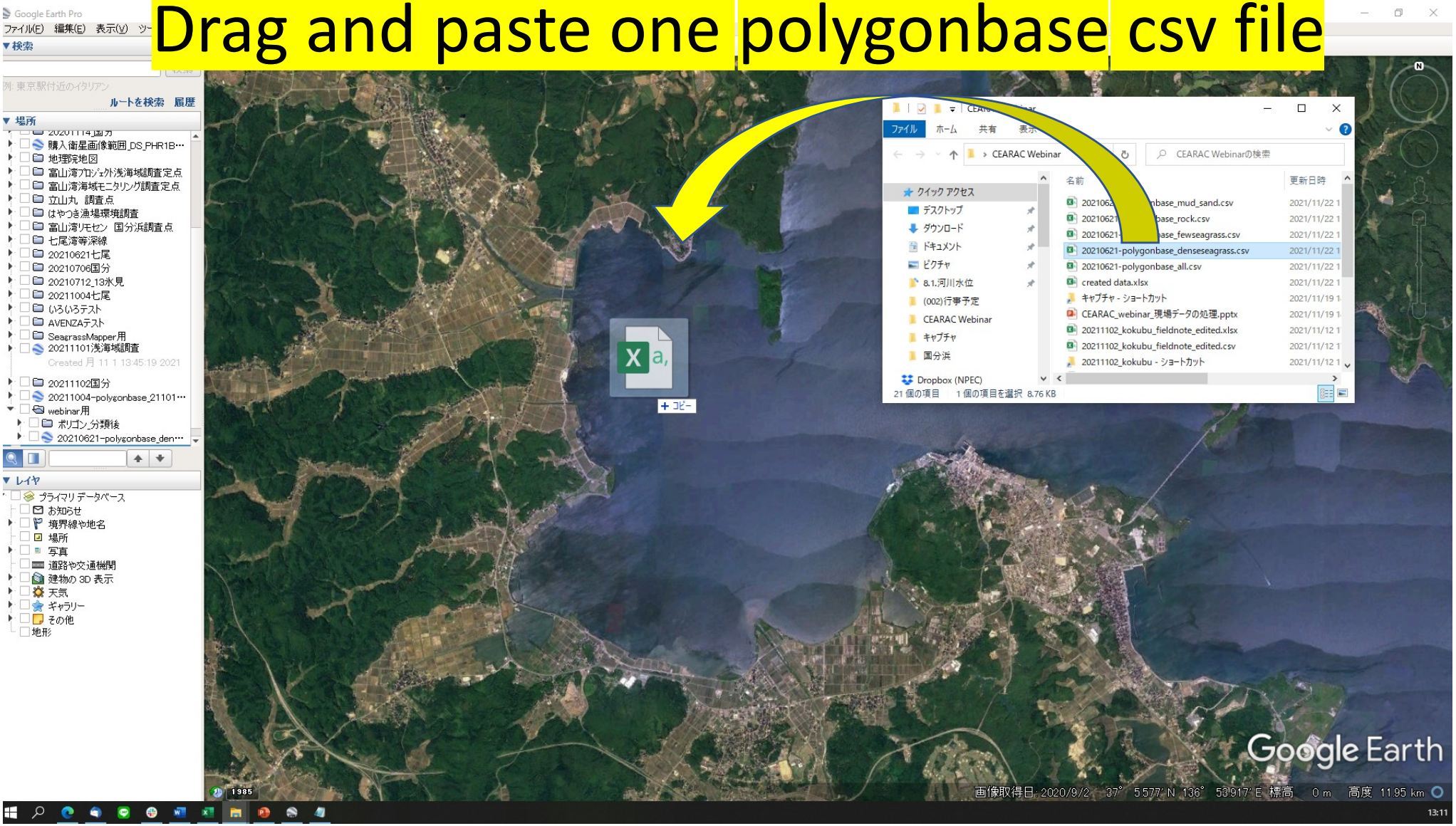
4, Drawing polygon data

outline

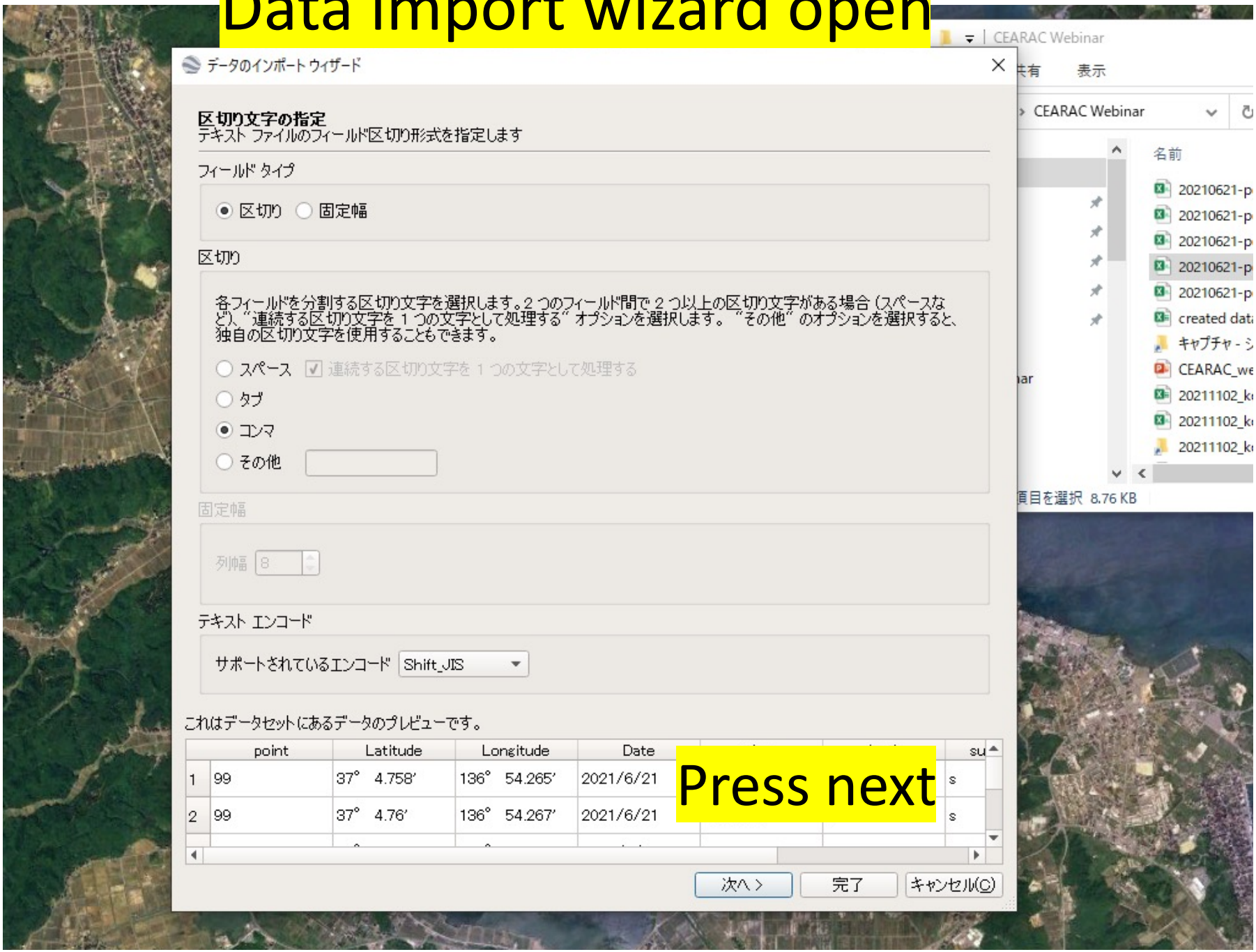
- Display points from csv file of each class and draw polygons.
- Polygons should be separated with different colours by each class.
- Output kmz files for each class.
- Finally kmz files of classes are ready to register by Seagrass trainer.

Mapping csv files
(eg. Google Earth Pro)

Drag and paste one polygonbase csv file



Data import wizard open

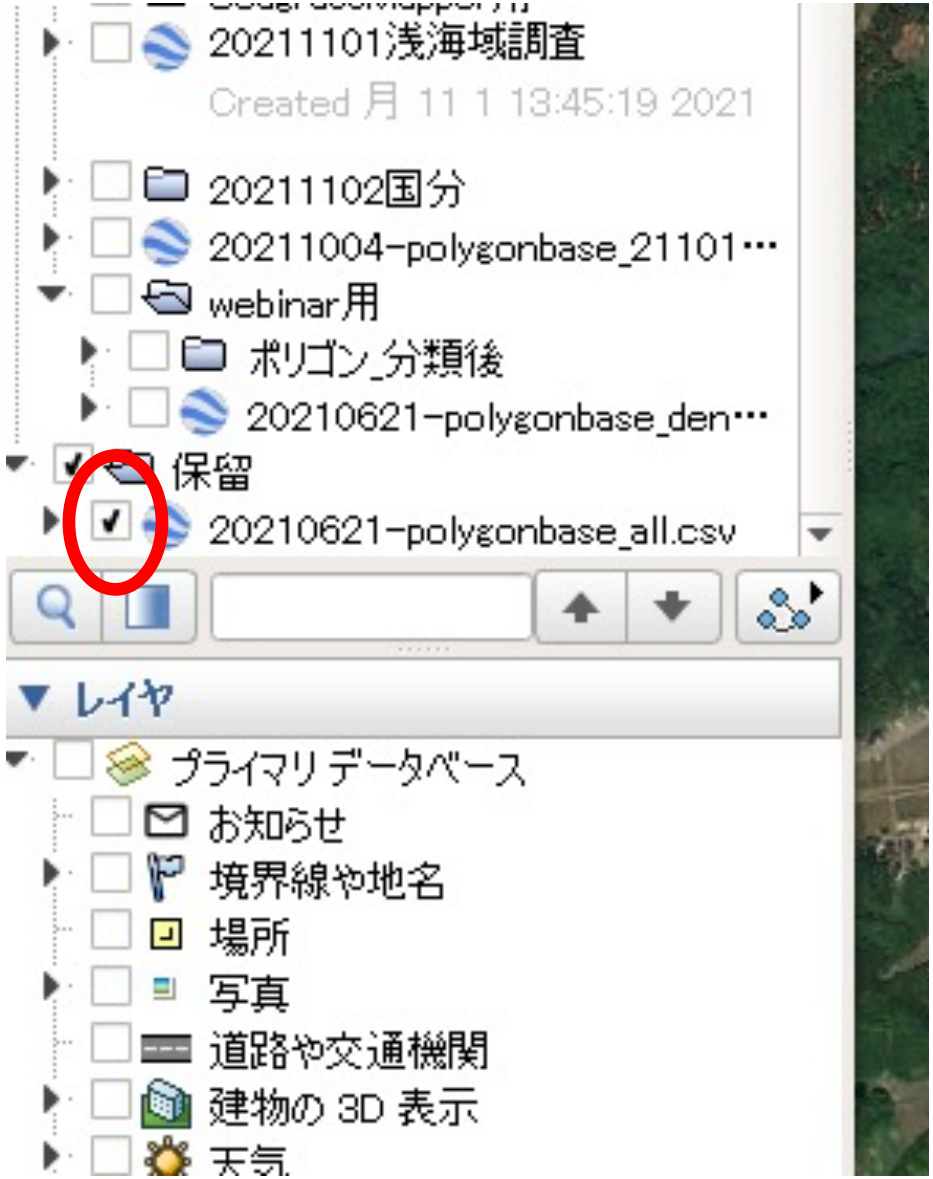


Press next

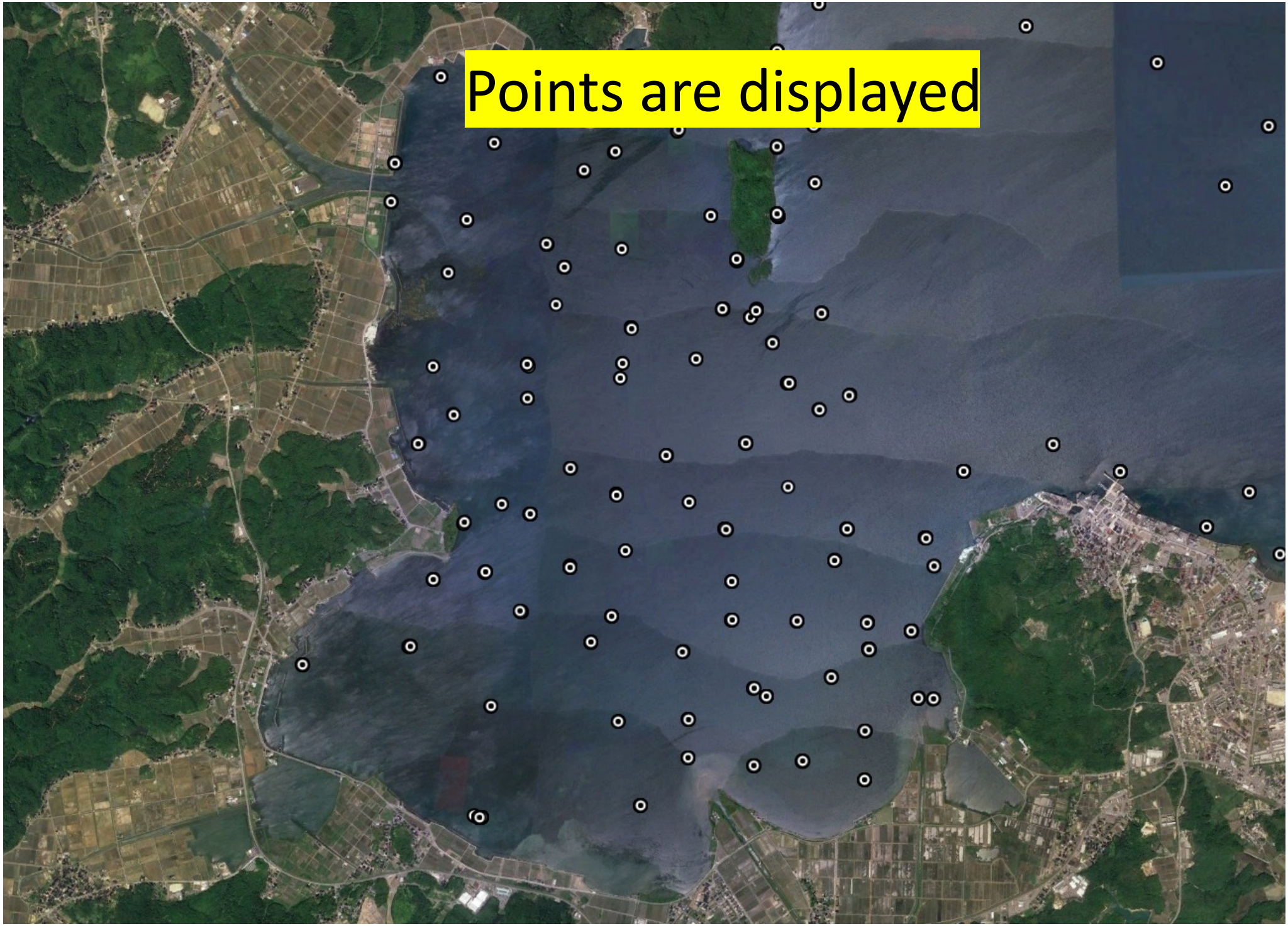

データのインポート ウィザード
緯度/経度フィールドを選択
このデータセットには緯度/経度情報は含まれていませんが、番地は含まれています
緯度フィールド Latitude
経度フィールド Longitude
“Longitude” and “Latitude” recognized automatically
これはデータセットにあるデータのプレビューです。
point Latitude Longitude Date time depth
1 99 37° 4.758′ 136° 54.265′ 2021/6/21 09:19:12 1
2 99 37° 4.76′ 136° 54.267′ 2021/6/21 09:19:45 1
3 99 37° 4.76′ 136° 54.267′
4 99 37° 4.76′ 136° 54.267′
Press next and finish
< 戻る
次へ >
完了
キャンセル(C)
CEARAC Webinar
共有
表示
CEARAC Webinar
項目を選択 8.76 KB

NO for now,

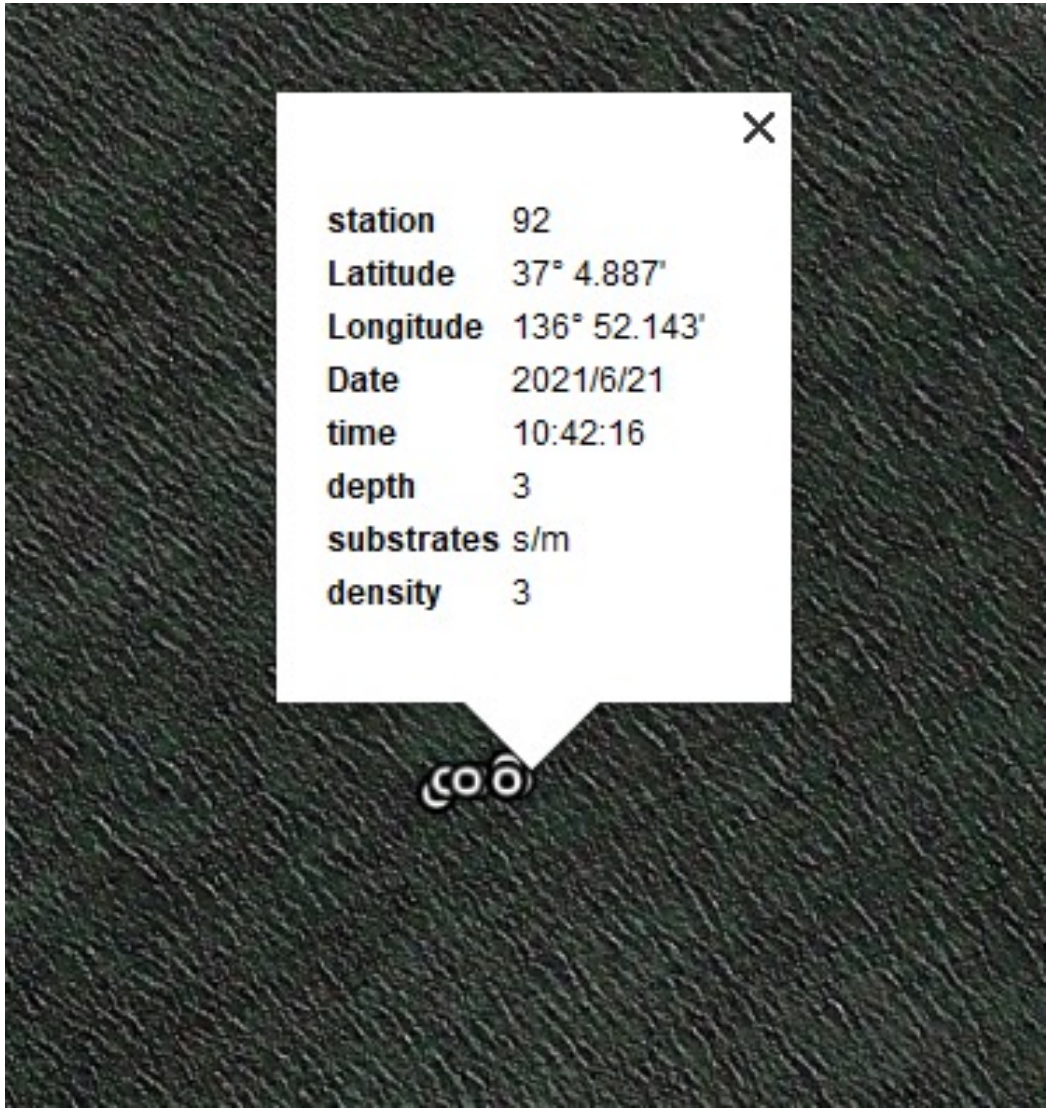
※ Style templates makes advanced display



Tick the “polygonbase_all.csv" box that is in Pending folder

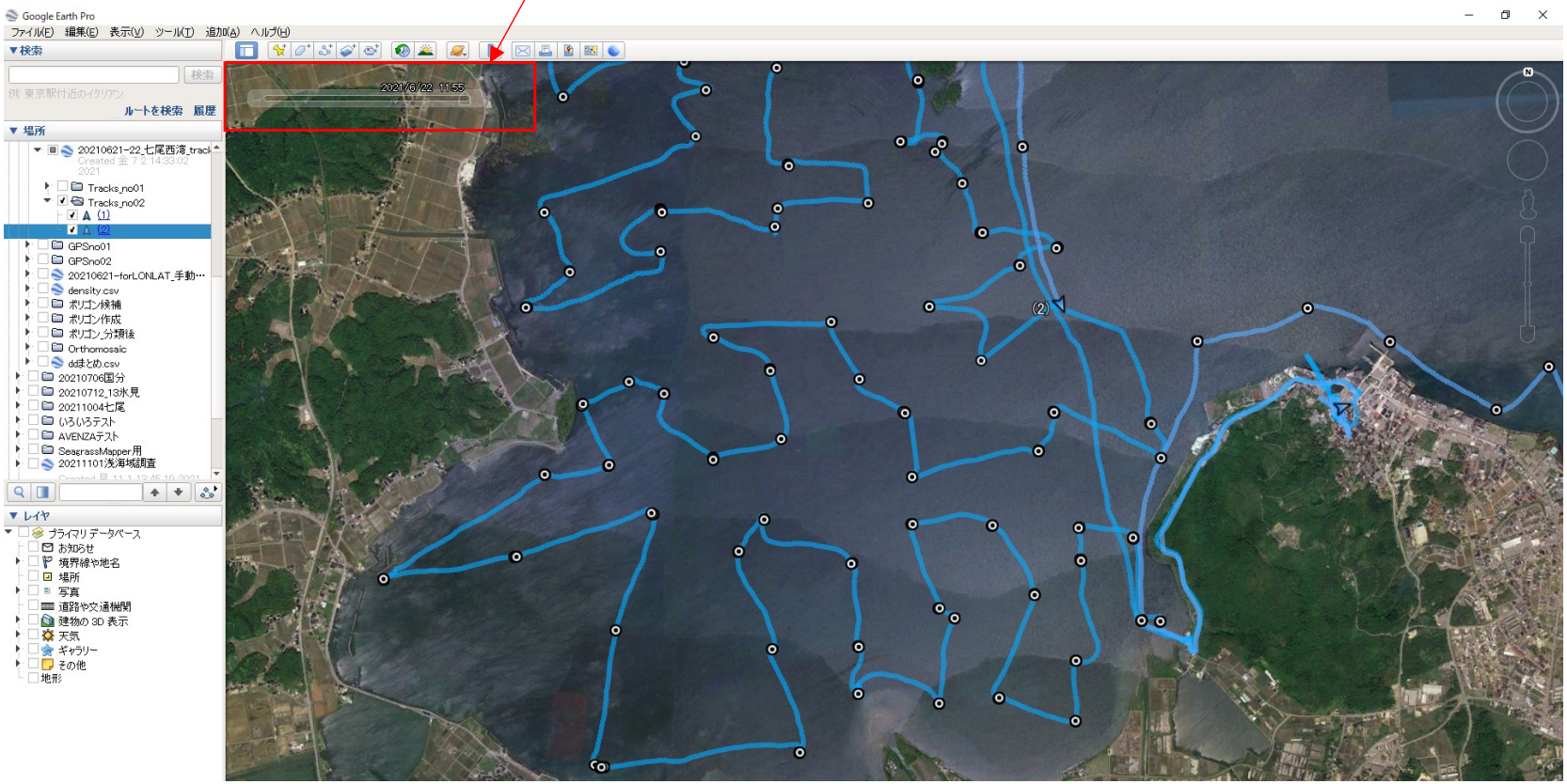

Points are displayed

Look at points in detail
Click a point and check the information

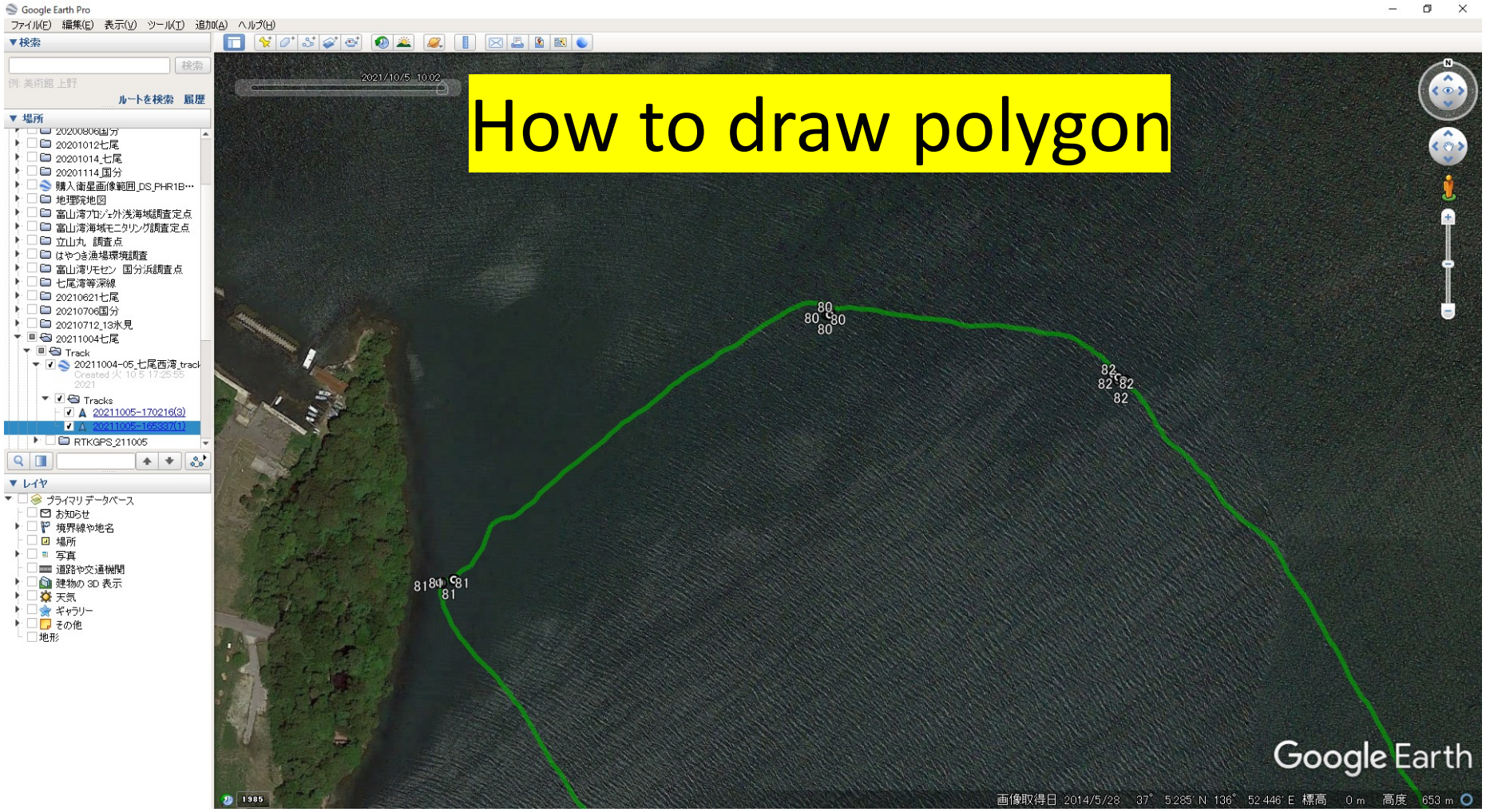


Check the overview of field data

- Display tracks and points, and check the location of the points of “polygonbase_all.csv”.
- To display track, drag gpx file in the same way.
- Note the time slider in the top left corner

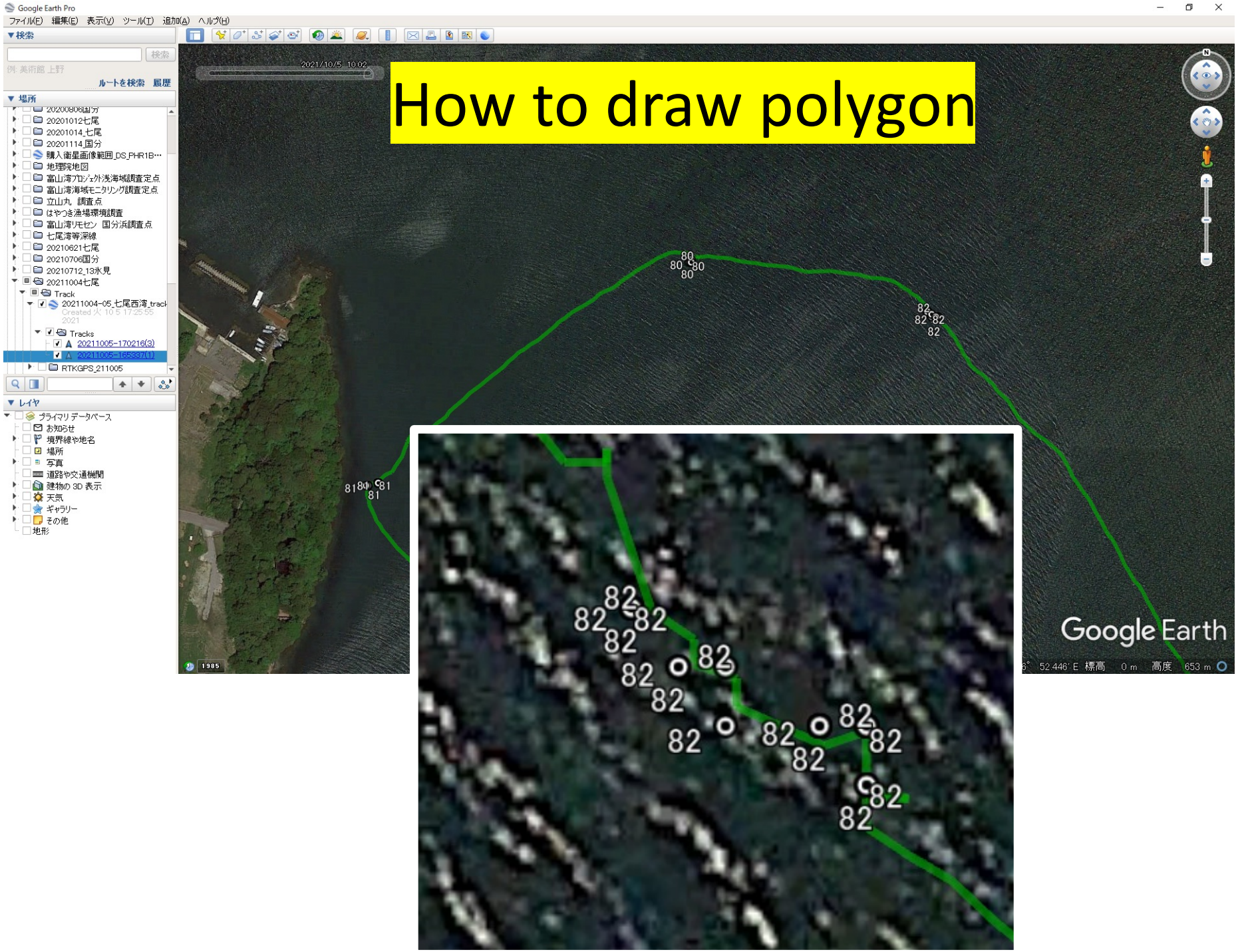


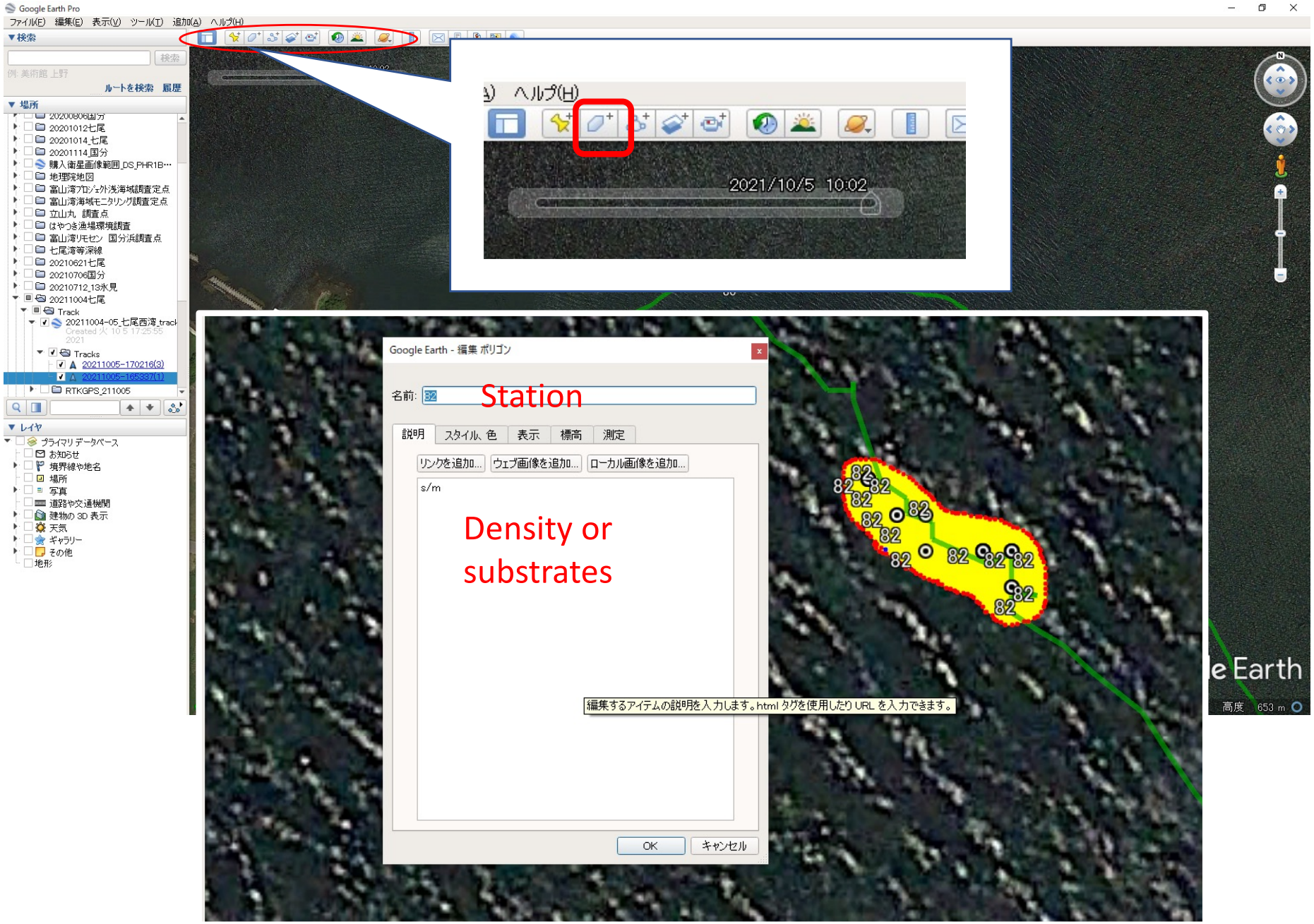

Wrong track and wrong estimated points
79
79
79
79
80
80
80
82
82
82
82
83
83
83
81
81
81
87
88
88
88
89
89
89



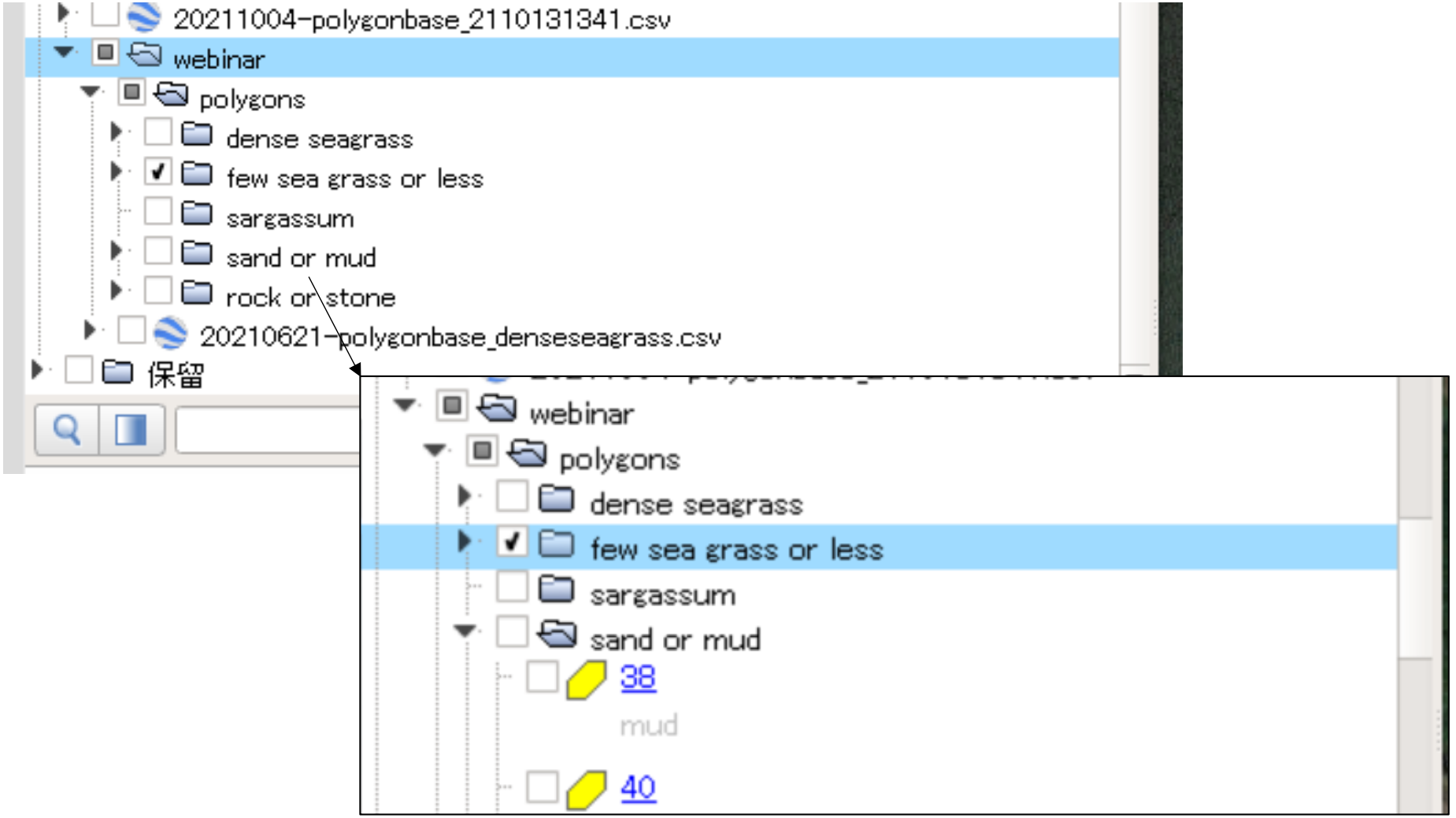
The dataset for polygonbase should be the classed one.

For example, polygonbase_mud_sand.csv.

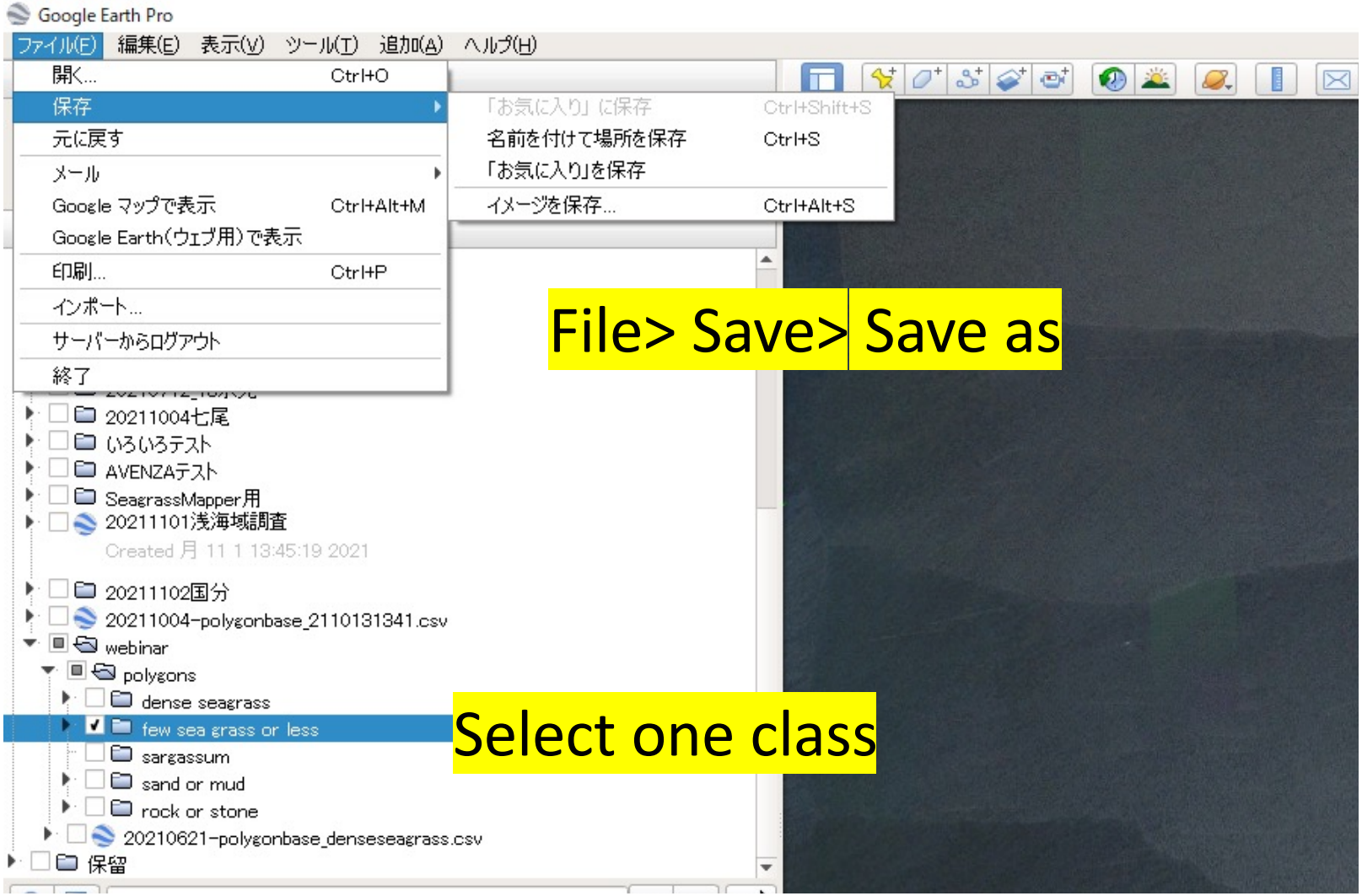

How to draw polygon
Google Earth Pro
ファイル(F) 編集(E) 表示(V) ツール(T) 追加(A) ヘルプ(H)
検索
場所
レイヤ
Google Earth

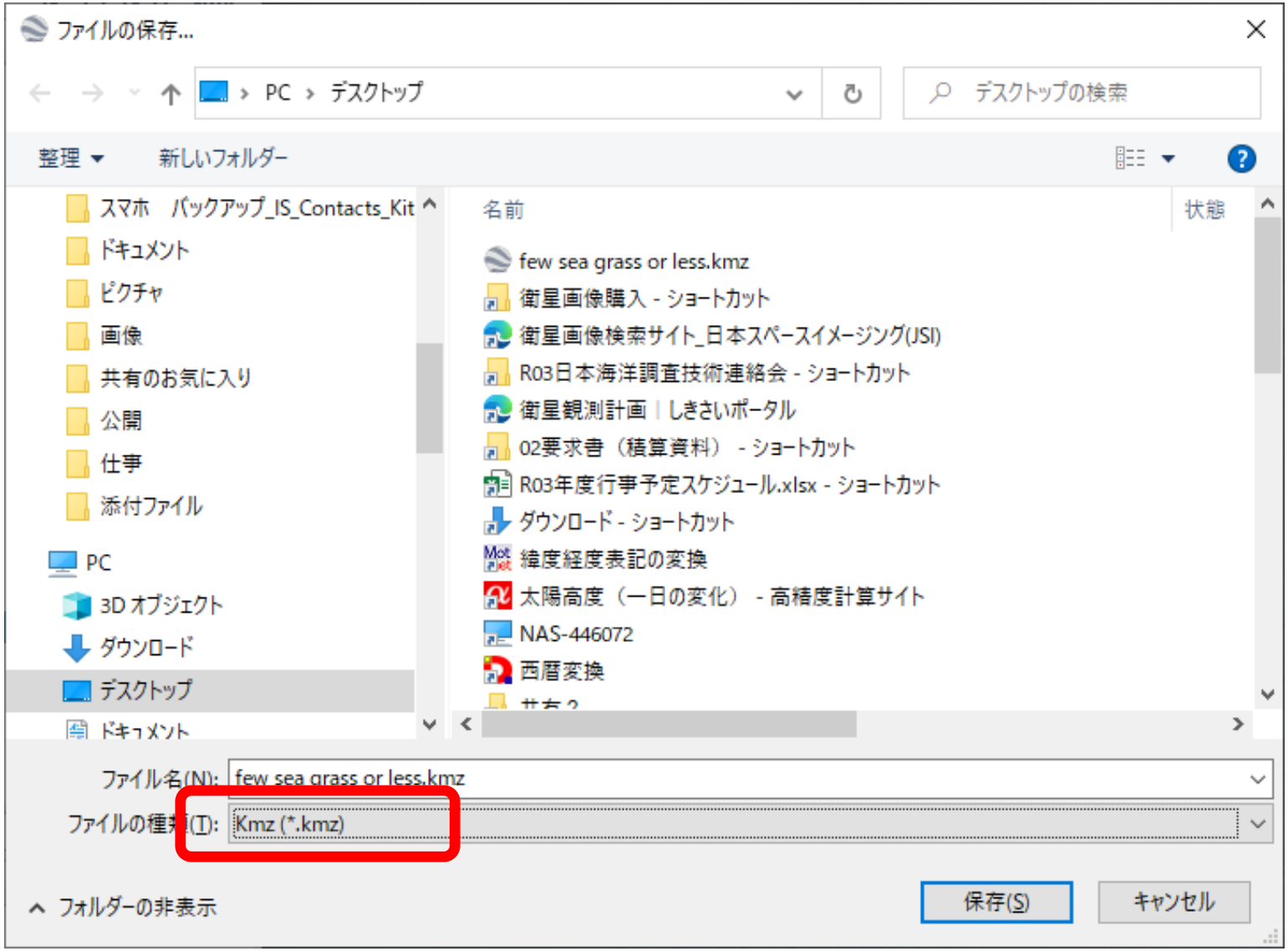

Google Earth Pro
ファイル(F) 編集(E) 表示(V) ツール(T) 追加(A) ヘルプ(H)
ヘルプ(H)
2021/10/5 10:02
Google Earth - 編集 ポリゴン
名前:
Station
説明
スタイル、色
表示
標高
測定
リンクを追加...
ウェブ画像を追加...
ローカル画像を追加...
s/m
Density or substrates
編集するアイテムの説明を入力します。html タグを使用したり URL を入力できます。
OK
キャンセル
82
高度 653 m

Create a folder with the name of each class and store the polygons in it



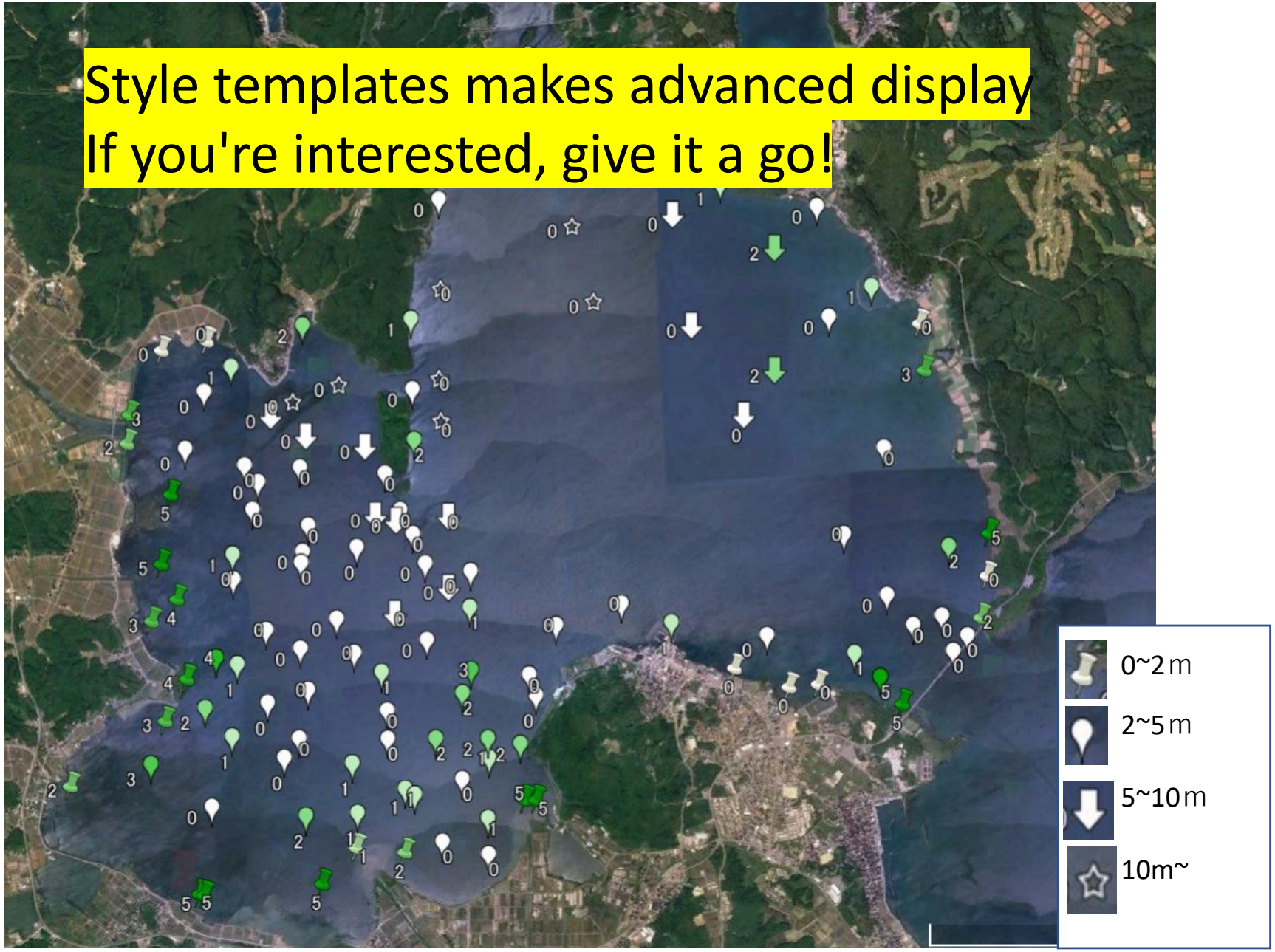
Save as kmz file for each class





Save same class polygons in each kmz file

Finaly, training dataset for Seagrass trainer is prepared


Style templates makes advanced display
If you're interested, give it a go!
0~2 m
2~5 m
5~10 m
10m~

Hands on

- Download data from https://u.pcloud.link/publink/show?code=kZawuAXZpcKfOOVxEs7RIkJxMtJtJHCGRNdk#returl=https%3A//u.pcloud.link/publink/show%3Fcode%3DkZawuAXZpcKfOOVxEs7RIkJxMtJtJHCGRNdk&page=login

- Drag and paste “20210621-polygonbase_denseseagrass.csv”
- Drag and paste “20210621-polygonbase_mud_sand.csv”

- Drag and paste “20210621-22_GPStrack.gpx”
- Display track and points to check the position
- Display only denseseagrass points
- Draw polygons (18 stations)
- Save as kmz file

- IF you have time, try to data of “mud_sand” (67 stations).