

Automation of Data Spike Flagging on TAO Refresh

Sathishkumar Samiappan
PhD Candidate in ECE
Mississippi State University



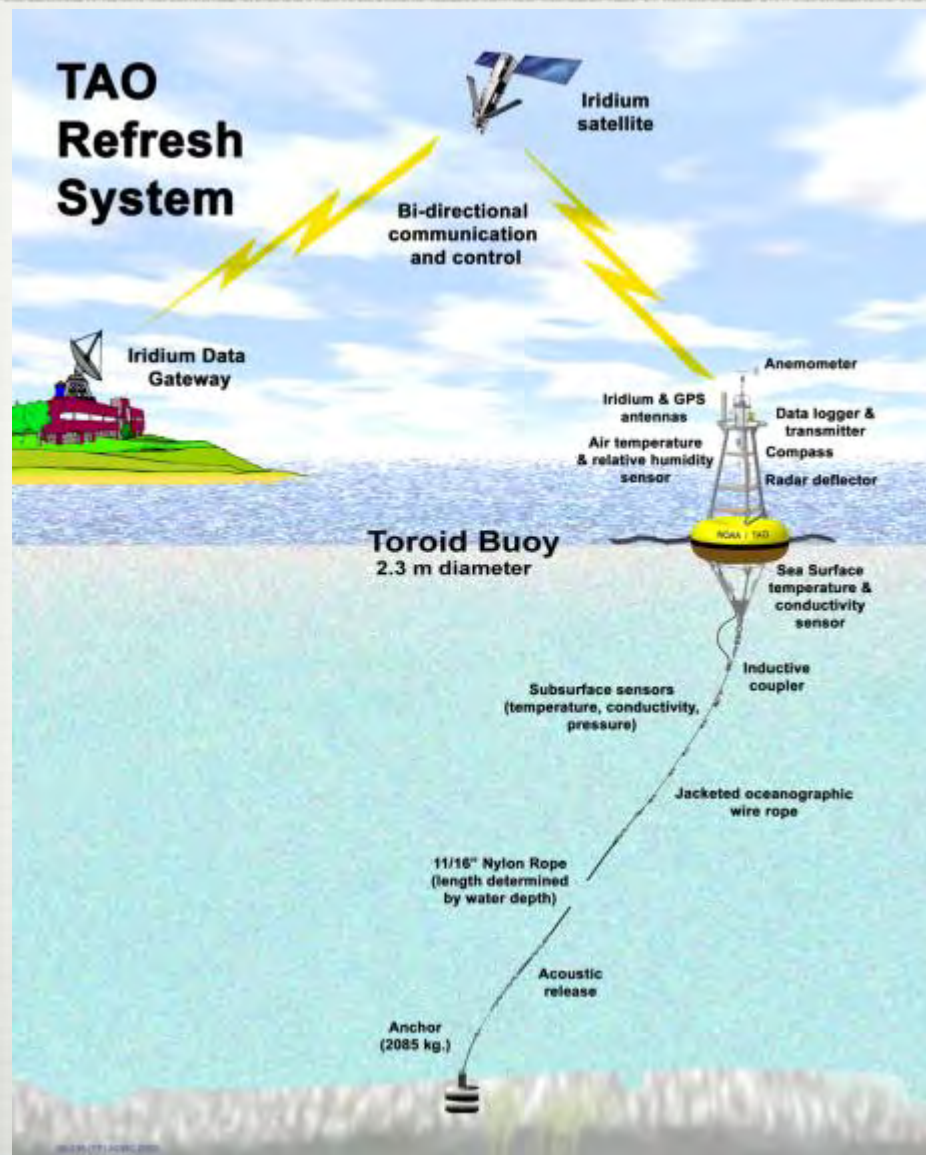
Outline

- Introduction - TAO
- TAO Refresh
- TAO Refresh Data Flow
- Quality Control
- Data Spikes
- Types of Data Spikes
- Proposed Auto QC & Auto QC Algorithms

Introduction

- TAO – Tropical Atmosphere Ocean project
- ≈ 55 moorings in the tropical pacific ocean
- To study and predict El Niño Southern Oscillation (ENSO)
- Sensors – sea surface & subsurface temperatures, salinity, water pressure, ocean currents, radiation, air temperature, wind direction, wind speed, rainfall, relative humidity and down welling radiation

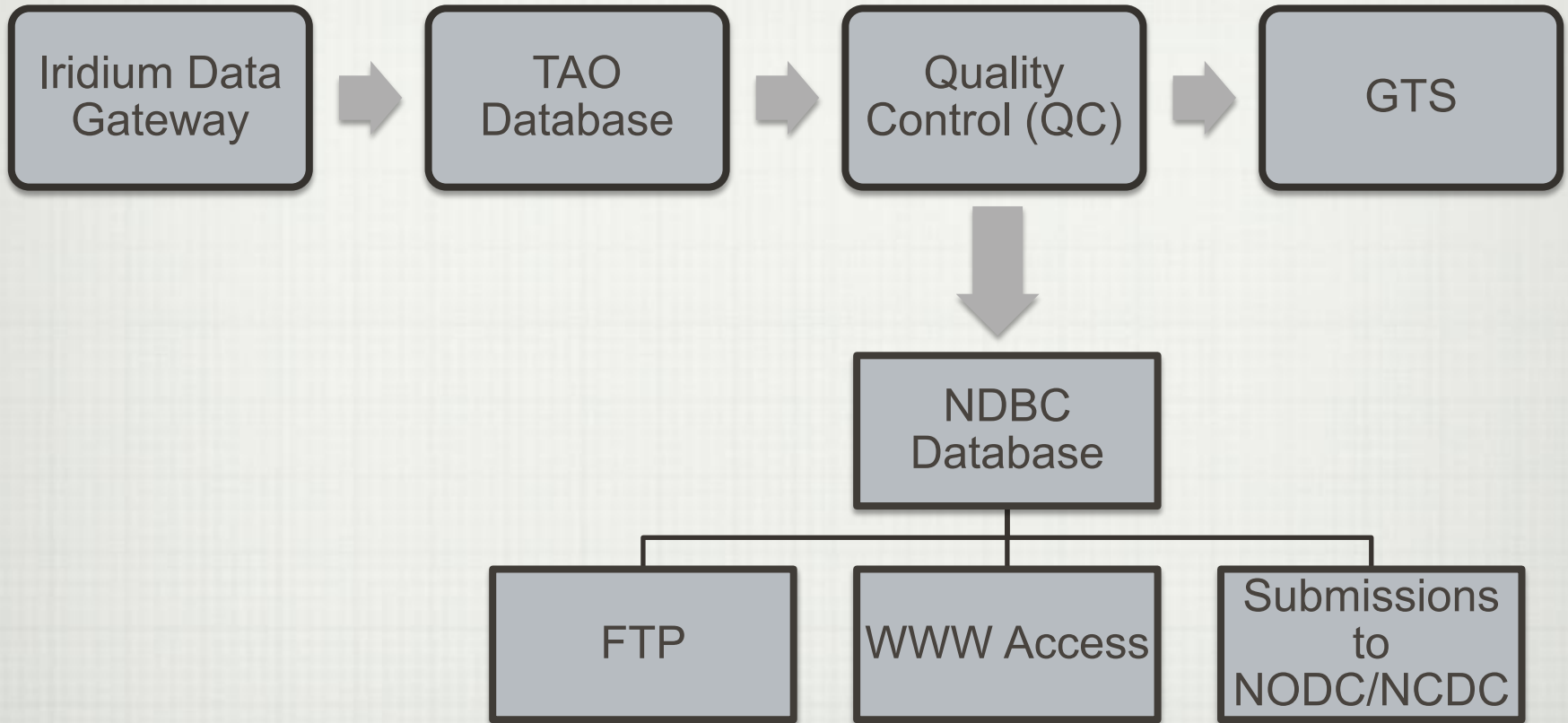
TAO Refresh



TAO Refresh Latitudes and Longitudes

	137E	147E	156E	165E	180W	170W	155W	140W	125W	110W	95W	
9N								DM037A				9N
8N						DM038A	DM047A					8N
5N				DM027A			DM046A	DM036A		DM048A		5N
2N					DM041A	DM026A			DM033A	DM049A	DM050A	2N
0					DM031B	DM039A	DM045A				DM051A	0
2S				DM028B	DM030A	DM040A						2S
5S							DM044A	DM035A	DM023B			5S
8S				DM029A	DM042A	DM043A	DM025A		DM034A		DM032A	8S
	137E	147E	156E	165E	180W	170W	155W	140W	125W	110W	95W	

TAO Refresh Data Flow



Quality Control

- Currently, an automatic gross range check is preformed on all incoming data before sending it to GTS.
- After few days, a manually QC data is updated on NDBC database
- Data sent to GTS is not modified
- Data on NDBC database is open and can be accessed through web, FTP and submitted to NODC/NCDC

Data Spikes

- A spike is an abnormality in the data values either due to sensor failure /error or transmission error.
- The data stored in the buoy on-board sensor memory is available when the buoy is recovered and this data can be compared with the data transmitted to understand the different transmission errors or spikes occurred and sensor failures.
- The data stored on-board sensor memory is commonly referred as sensor data or tube data and the data transmitted from buoy is referred as real-time data.

Data Spike Examples

5	Date/Time	25m	50m	75m	100m	125m	150m	200m	250m	300m	500m	Date/Time	25m
39	6/10/2013 5:30:00	28.7795	28.604	28.0846	26	23.7976	17.1615	12.1805	11.3612	10.2543	8.3622	6/10/2013 5:30:00	
40	6/10/2013 5:20:00	28.7596	28.6052	28.0654	26.7591	24.1029	17.1908	12.4652	11.3692	10.2543	8.2558	6/10/2013 5:20:00	
41	6/10/2013 5:10:00	28.7792	28.6197	28.1002	26.8304	24.2973	17.1908	12.1199	11.3578	10.0708	8.2111	6/10/2013 5:10:00	
42	6/10/2013 5:00:00	28.767	28.6218	28.0938	26.8752	24.3163	17.2254	12.1296	11.3691	10.2534	8.2752	6/10/2013 5:00:00	
43	6/10/2013 4:50:00	28.7538	28.6188	28.0773	26.7935	24.2491	17.1676	12.123	11.3861	10.6776	8.5307	6/10/2013 4:50:00	
44	6/10/2013 4:40:00	28.7444	28.5973	28.0378	26.6504	24.1755	17.1588	12.1483	11.5098	10.826	8.5467	6/10/2013 4:40:00	
45	6/10/2013 4:30:00	28.7488	28.5924	28.0364	26.7083	24.267	18.8227	12.4879	11.6412	10.8321	8.277	6/10/2013 4:30:00	
46	6/10/2013 4:20:00	28.7853	28.6236	28.0695	26.7561	23.6888	17.2203	12.2533	11.5737	10.6532	8.252	6/10/2013 4:20:00	
47	6/10/2013 4:10:00	28.7381	28.6068	28.053	26.6367	23.7307	18.4674	12.3202	11.6451	10.7757	8.3189	6/10/2013 4:10:00	
48	6/10/2013 4:00:00	--	--	--	--	23.5828	17.8444	12.3236	12.3236	--	--	6/10/2013 4:00:00	
49	6/10/2013 3:50:00	--	--	--	--	23.7158	18.8508	12.6384	12.6384	--	--	6/10/2013 3:50:00	
50	6/10/2013 3:40:00	--	--	--	--	23.8772	18.8262	12.6268	12.6268	--	--	6/10/2013 3:40:00	
51	6/10/2013 3:30:00	--	--	--	--	23.9323	18.8901	12.625	12.625	--	--	6/10/2013 3:30:00	
52	6/10/2013 3:20:00	--	--	--	--	24.0499	18.8994	12.6383	12.6383	--	--	6/10/2013 3:20:00	
53	6/10/2013 3:10:00	--	--	--	--	24.3451	18.7982	1	12.508	--	--	6/10/2013 3:10:00	
54	6/10/2013 3:00:00	28.7497	28.6078	28.033	26.4141	24.5635	18.8999	12.5095	11.6912	10.9156	8.5289	6/10/2013 3:00:00	
55	6/10/2013 2:50:00	28.7521	28.618	28.0347	26.3794	24.4352	18.8801	12.6385	11.7202	10.9468	8.5189	6/10/2013 2:50:00	
56	6/10/2013 2:40:00	28.7486	28.6072	28.0367	26.3671	24.4163	18.9077	12.4909	11.7178	10.9335	8.5382	6/10/2013 2:40:00	
57	6/10/2013 2:30:00	28.7582	28.6225	28.0399	26.7018	24.5566	18.8979	12.6235	11.6681	10.8673	8.5309	6/10/2013 2:30:00	
58	6/10/2013 2:20:00	28.7624	28.6276	28.0365	26.8277	24.7333	18.8939	12.6414	11.6729	10.8616	8.3757	6/10/2013 2:20:00	
59	6/10/2013 2:10:00	28.7407	28.6121	28.0356	26.8659	24.8689	18.9005	12.6176	11.6936	10.8578	8.2792	6/10/2013 2:10:00	
60	6/10/2013 2:00:00	28.7514	--	--	--	--	--	12.6222	--	--	--	6/10/2013 2:00:00	
61	6/10/2013 1:50:00	28.7581	--	--	--	--	--	12.609	--	--	--	6/10/2013 1:50:00	
62	6/10/2013 1:40:00	28.7437	--	--	--	--	--	12.5236	--	--	--	6/10/2013 1:40:00	
63	6/10/2013 1:30:00	28.7479	--	--	--	--	--	12.5536	--	--	--	6/10/2013 1:30:00	
64	6/10/2013 1:20:00	28.7453	--	--	--	--	--	12.6052	--	--	--	6/10/2013 1:20:00	
65	6/10/2013 1:10:00	28.7238	--	--	--	--	--	12.5446	--	--	--	6/10/2013 1:10:00	
66	6/10/2013 1:00:00	28.7439	28.6201	28.0558	27.5129	25.6841	--	12.5558	11.6249	10.9109	8.3494	6/10/2013 1:00:00	
67	6/10/2013 0:50:00	28.7356	28.6112	28.0516	26.9229	25.3908	--	12.8104	11.7239	10.9668	8.2961	6/10/2013 0:50:00	

Data Spike Examples

Real-time

Recent examples

5n140w – Most have already been hard flagged, so the missing messages are represented here as well (some of the 0s)

Spike+ immediately adjacent data value

26.09 °C
5.31 °C

26.16 °C
5.32 °C

26.25 °C
5.32 °C

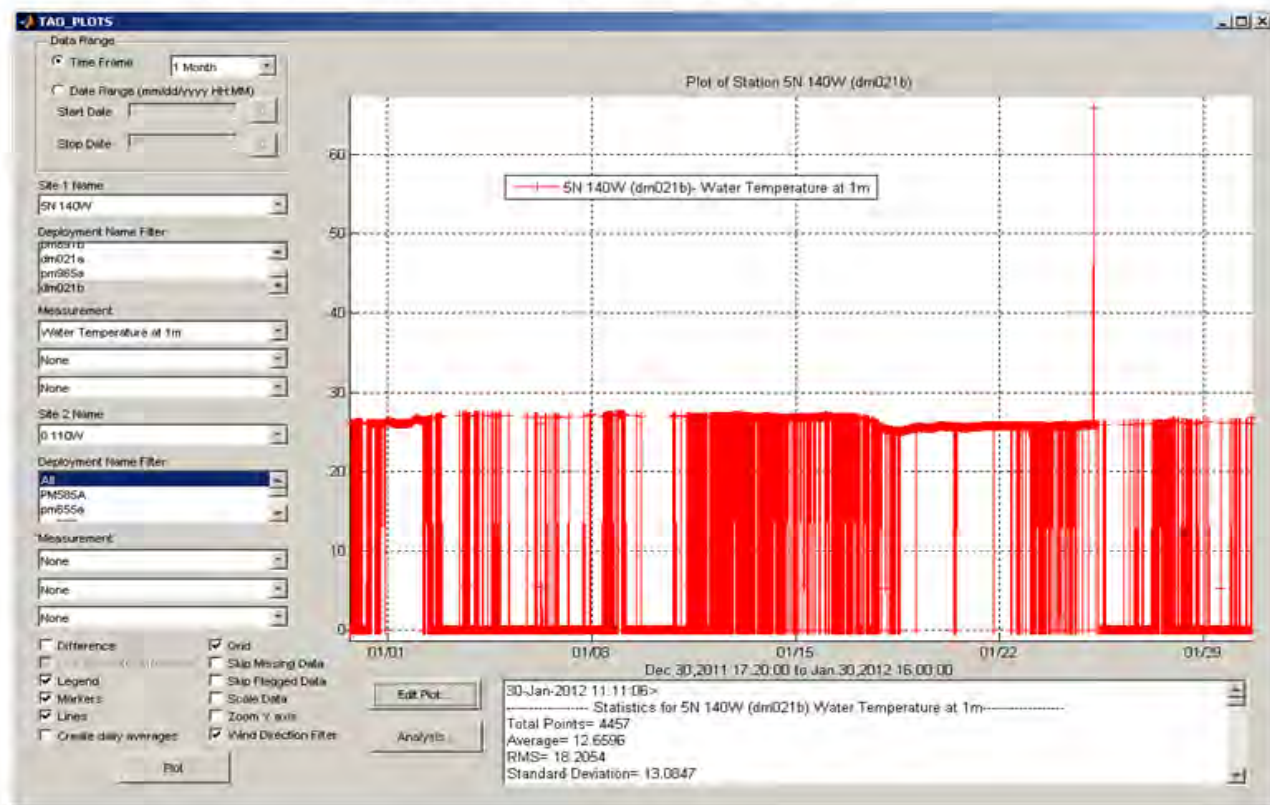
25.87 °C
65.87 °C
25.87 °C

25.82 °C
0.0 °C
25.83 °C

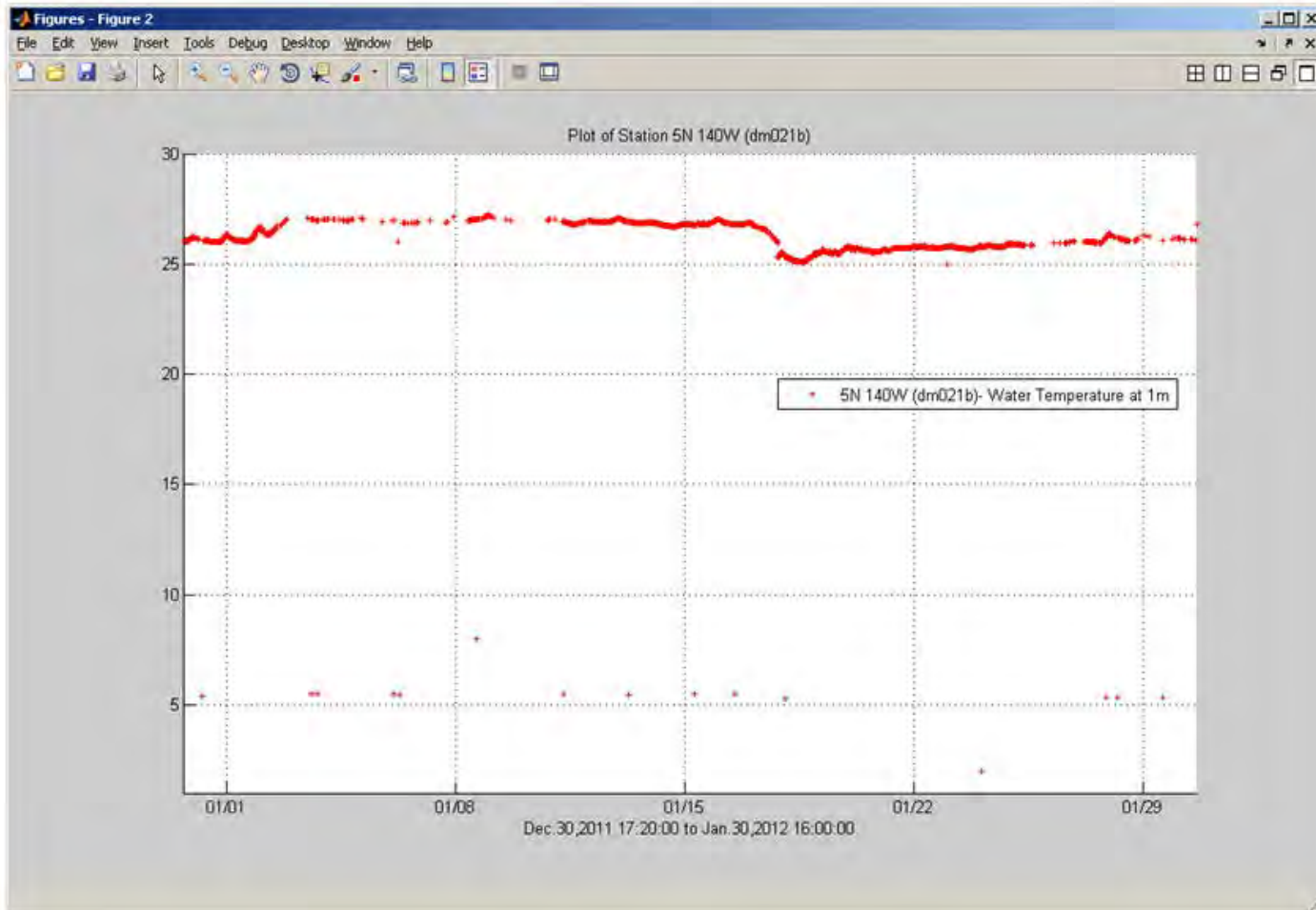
25.81 °C
2.0 °C
25.81 °C
25.8 °C
0.0 °C

25.37 °C
5.3 °C
25.34 °C

26.8 °C
5.48 °C
26.8 °C



Data Spike Examples



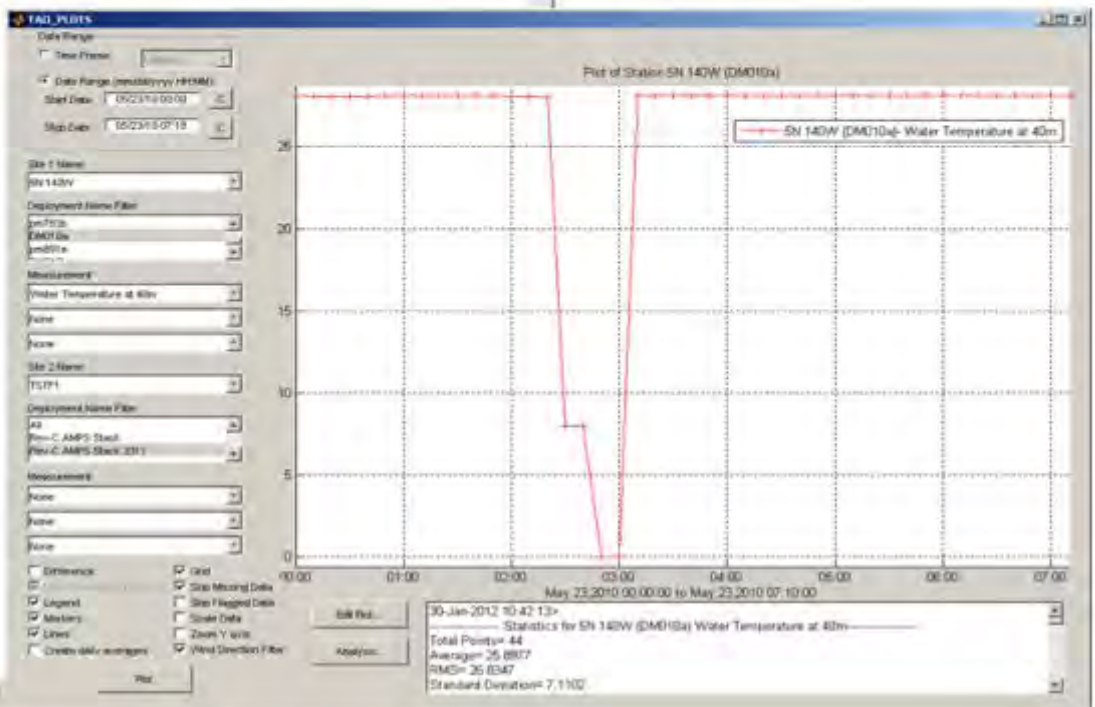
Data Spike Examples

https://comms02.ndbc.noaa.gov/refreshed/cache/DM010a_WATER_TEMPERATURE_1013920120130162137866716217.txt

Real-time

Date	Time	Temp
2010-05-23	03:20:00	28.0733
2010-05-23	03:10:00	28.0750
2010-05-23	02:40:00	8
2010-05-23	02:30:00	8
2010-05-23	02:20:00	28
2010-05-23	02:10:00	28.0644
2010-05-23	02:00:00	28.0697
2010-05-23	01:50:00	28.0748
2010-05-23	01:40:00	28.0781
2010-05-23	01:30:00	28.1021
2010-05-23	01:20:00	28.0808
2010-05-23	01:10:00	28.1046
2010-05-23	01:00:00	28.0756
2010-05-23	00:50:00	28.0735
2010-05-23	00:40:00	28.0662
2010-05-23	00:30:00	28.0645
2010-05-23	00:20:00	28.0651
2010-05-23	00:10:00	28.0688
2010-05-23	00:00:00	28.0738
2010-05-22	23:50:00	28.0677
2010-05-22	23:40:00	28.0672
2010-05-22	23:30:00	28.0793
2010-05-22	23:20:00	28.0785
2010-05-22	23:10:00	28.0912
2010-05-22	23:00:00	28.084
2010-05-22	22:50:00	28.0825
2010-05-22	22:40:00	28.0745
2010-05-22	22:30:00	28.0717
2010-05-22	22:20:00	28.041
2010-05-22	22:10:00	28.0651
2010-05-22	22:00:00	28.0621
2010-05-22	21:50:00	28.0717
2010-05-22	21:40:00	28.0679
2010-05-22	21:30:00	28.0701
2010-05-22	21:20:00	28.077
2010-05-22	21:10:00	28.0627
2010-05-22	21:00:00	28.073
2010-05-22	20:50:00	28.0721
2010-05-22	20:40:00	28.0445
2010-05-22	20:30:00	28.0708
2010-05-22	20:20:00	28.0685
2010-05-22	20:10:00	28.0674
2010-05-22	20:00:00	28.0455
2010-05-22	19:50:00	28.045
2010-05-22	19:40:00	28.0607
2010-05-22	19:30:00	28.0802
2010-05-22	19:20:00	27.7944
2010-05-22	19:10:00	27.7872
2010-05-22	19:00:00	27.8781

T40 sensor, taken from TAO Console
Realtime data telecommunicated from Tube



Data Spike Examples

TUBE

T40 sensor, taken from Tube download
Julian date 143 = 5/23

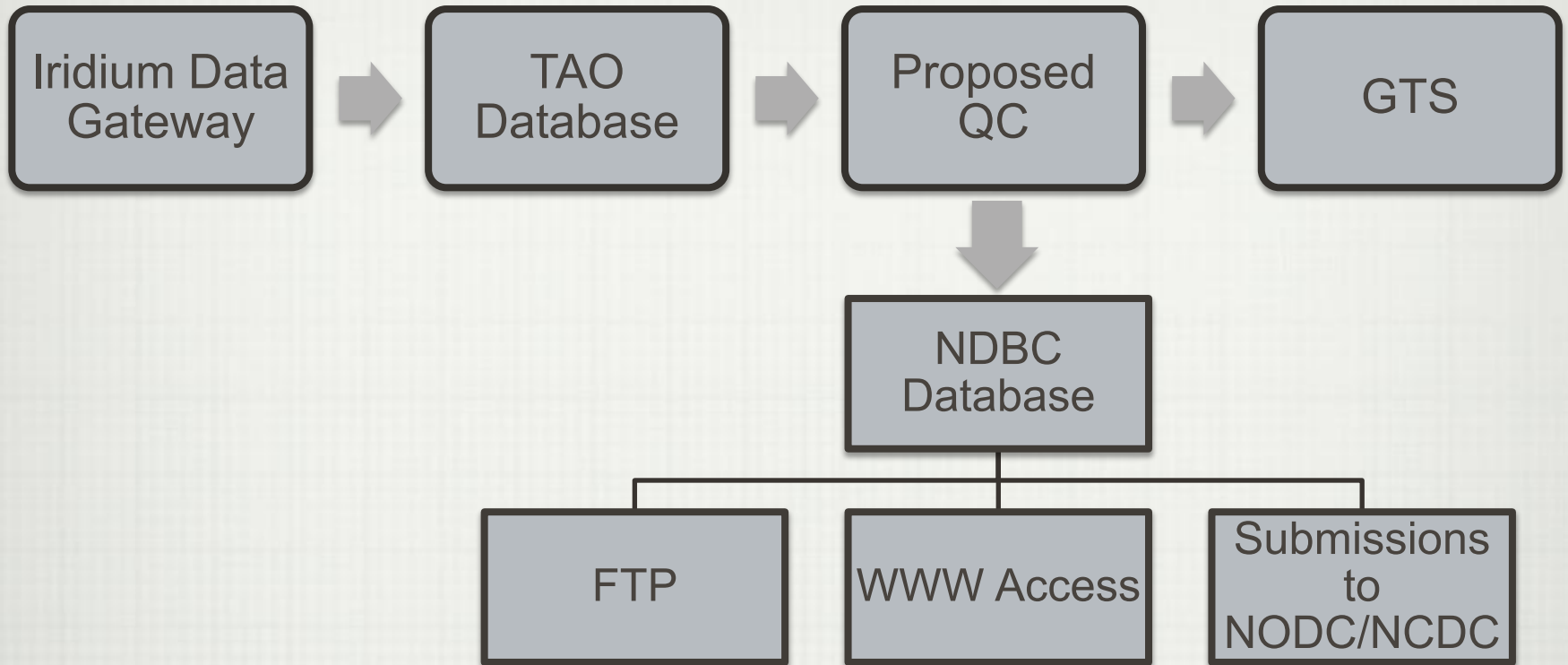
Julian Date	Temp 1	Temp 2	Temp 3	Temp 4	Temp 5	Temp 6	Temp 7	Temp 8	Temp 9	Temp 10	Temp 11	Temp 12
2010142222000	28.205	28.071	28.061	26.536	21.025	16.199	13.579	12.206	10.608	9.559	8.123	
2010142223000	28.183	28.078	28.072	27.196	21.648	16.191	13.524	12.170	10.609	9.575	8.129	
2010142224000	28.236	28.091	28.075	27.155	21.791	17.181	13.875	12.479	10.732	9.611	8.128	
2010142225000	28.249	28.089	28.083	27.200	22.115	17.120	13.763	12.302	10.626	9.593	8.118	
2010142230000	28.259	28.097	28.084	27.361	22.572	17.162	13.771	12.221	10.605	9.595	8.117	
2010142231000	28.255	28.111	28.091	27.513	22.518	17.026	13.720	12.223	10.607	9.589	8.118	
2010142232000	28.293	28.099	28.078	27.644	22.912	17.082	13.806	12.264	10.603	9.596	8.118	
2010142233000	28.262	28.107	28.079	27.668	23.544	17.265	13.932	12.470	10.644	9.624	8.122	
2010142234000	28.253	28.093	28.067	27.643	23.641	17.210	13.873	12.451	10.696	9.635	8.169	
2010142235000	28.235	28.082	28.068	27.620	23.922	17.291	14.092	12.538	10.831	9.643	8.222	
2010143000000	28.235	28.078	28.074	27.655	24.407	17.707	14.224	12.612	10.843	9.638	8.217	
2010143001000	28.238	28.095	28.066	27.834	25.123	18.443	14.286	12.610	10.813	9.627	8.192	
2010143002000	28.248	28.079	28.065	27.814	25.276	18.061	14.150	12.549	10.771	9.616	8.191	
2010143003000	28.270	28.082	28.064	27.841	25.313	18.284	14.248	12.594	10.799	9.622	8.226	
2010143004000	28.286	28.082	28.066	27.857	25.301	17.946	14.044	12.505	10.670	9.611	8.220	
2010143005000	28.258	28.078	28.073	27.806	24.856	17.454	13.949	12.464	10.660	9.591	8.201	
2010143010000	28.256	28.094	28.076	27.807	25.113	17.478	13.989	12.418	10.636	9.581	8.177	
2010143011000	28.259	1e35	28.085	27.838	25.369	18.068	14.136	12.474	10.698	9.593	8.174	
2010143012000	28.286	1e35	28.081	27.788	25.309	18.594	14.682	12.669	10.852	9.619	8.183	
2010143013000	28.295	1e35	28.102	27.966	25.525	18.417	14.303	12.597	10.844	9.627	8.184	
2010143014000	28.289	1e35	28.078	27.831	25.374	18.068	14.215	12.538	10.794	9.631	8.189	
2010143015000	28.222	1e35	28.077	27.902	25.590	18.521	14.348	12.594	10.842	9.634	8.194	
2010143020000	28.197	1e35	28.070	27.846	25.656	18.860	14.387	12.649	10.852	9.630	8.174	
2010143021000	28.199	28.074	28.064	27.821	25.687	19.182	14.521	12.623	10.843	9.602	8.119	
2010143022000	28.212	28.085	28.000	27.861	25.842	19.516	14.632	12.597	10.822	9.569	8.104	
2010143023000	28.228	28.076	8.000	27.873	25.803	19.341	14.628	12.636	10.826	9.568	8.103	
2010143024000	28.242	28.068	8.000	27.885	25.771	19.506	14.702	12.667	10.913	9.591	8.114	
2010143025000	28.224	28.125	0.000	27.859	25.711	19.713	14.713	12.675	10.959	9.603	8.112	
2010143030000	28.261	28.116	0.000	27.814	25.758	20.529	15.499	12.972	11.190	9.631	8.119	
2010143031000	28.247	28.088	28.075	27.743	25.796	20.637	15.569	13.070	11.325	9.645	8.154	
2010143032000	28.242	28.101	28.071	27.781	25.676	20.602	15.492	12.959	11.285	9.650	8.173	
2010143033000	28.200	28.107	28.076	27.845	25.891	20.815	15.595	13.078	11.350	9.663	8.182	
2010143034000	28.214	28.131	28.083	27.861	26.114	21.857	15.623	13.325	11.484	9.677	8.201	
2010143035000	28.193	28.097	28.091	27.825	25.942	21.893	15.619	13.428	11.563	9.692	8.239	
2010143040000	28.176	28.160	28.099	28.020	26.234	22.244	15.665	13.525	11.617	9.723	8.292	

Types of Data Spikes and Flags

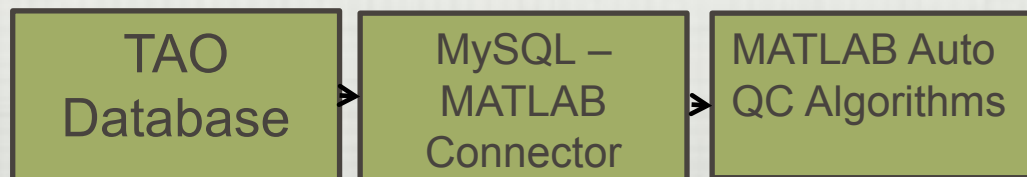
Type	Name
1	Missing Data
2	Deviation from actual value or outliers
3	Repeated data, truncated data and values shifted in time

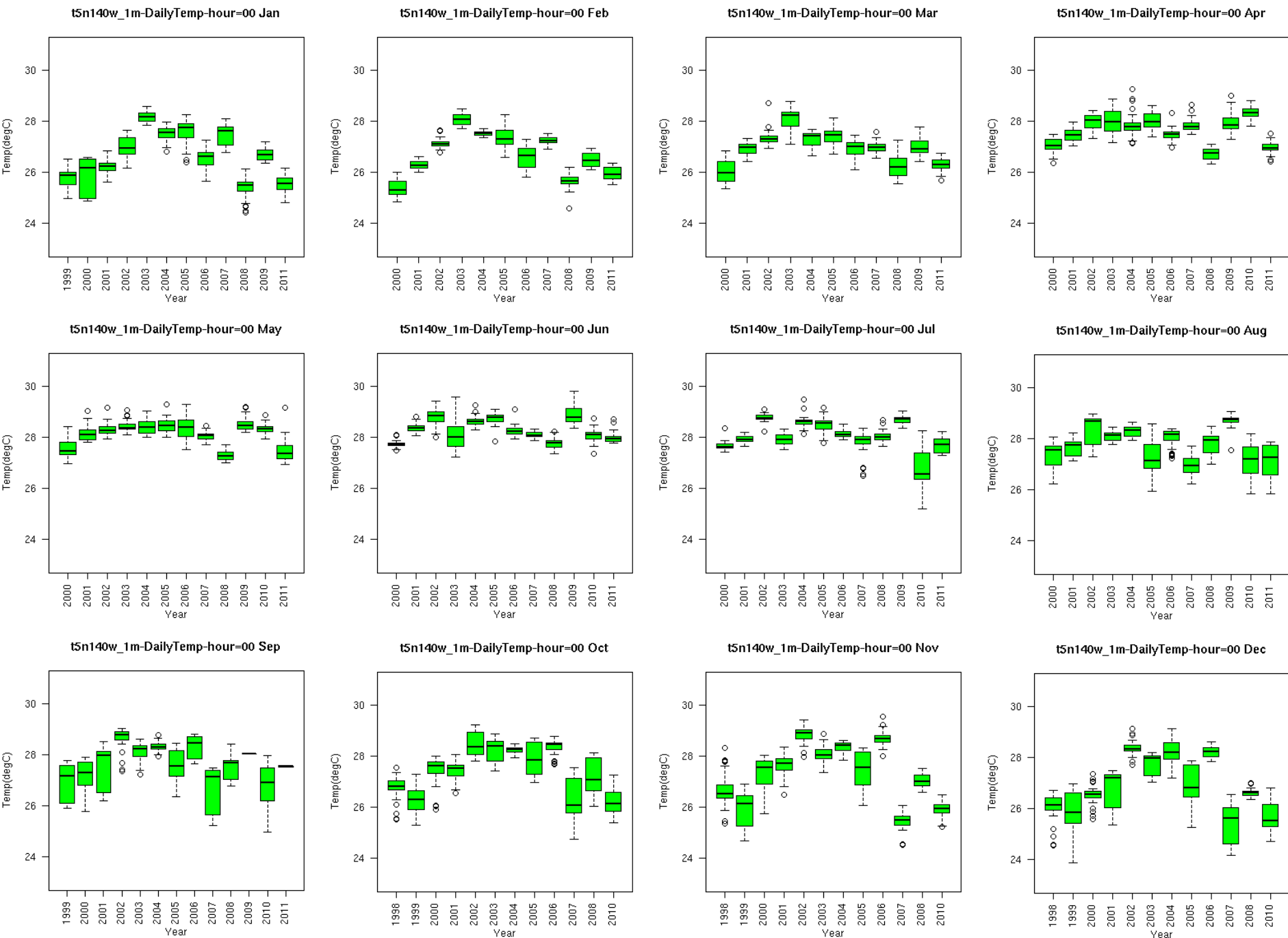
Flag Name	Condition
Good Data	When Confidence = 100%
Probably Good Data	When Confidence <100%
Questionable Data	Type -2 near thresholds
Bad Data	Type -2 or Type 3 or Way off threshold
Missing Data	Type - 1

Proposed Automatic QC

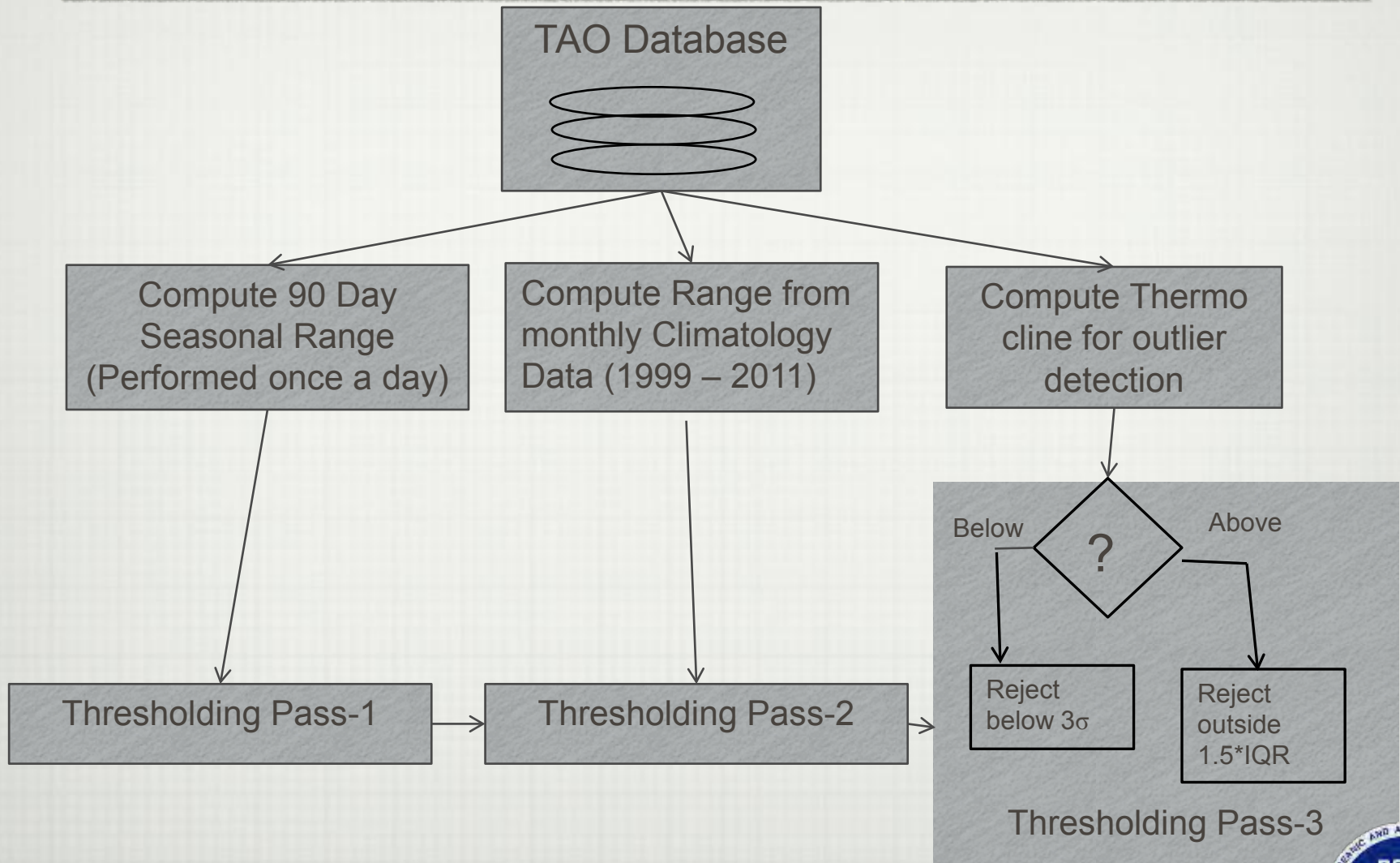


Proposed QC System





Automatic QC for Extreme Values



Automatic QC for Repeating values

125m	150m	200m	250m	300m	500m	Date/Time	25n
23.7976	17.1615	12.1805	11.3612	10.2543	8.3622	6/10/2013	5:30:00
24.1029	17.1908	12.4652	11.3692	10.2543	8.2558	6/10/2013	5:20:00
24.2973	17.1908	12.1199	11.3578	10.0708	8.2111	6/10/2013	5:10:00
24.3163	17.2254	12.1296	11.3691	10.2534	8.2752	6/10/2013	5:00:00
24.2491	17.1676	12.123	11.3861	10.6776	8.5307	6/10/2013	4:50:00
24.1755	17.1588	12.1483	11.5098	10.826	8.5467	6/10/2013	4:40:00
24.267	18.8227	12.4879	11.6412	10.8321	8.277	6/10/2013	4:30:00
23.6888	17.2203	12.2533	11.5737	10.6532	8.252	6/10/2013	4:20:00
23.7307	18.4674	12.3202	11.6451	10.7757	8.3189	6/10/2013	4:10:00
23.5828	17.8444	12.3236	12.3236	--	--	6/10/2013	4:00:00
23.7158	18.8508	12.6384	12.6384	--	--	6/10/2013	3:50:00
23.8772	18.8262	12.6268	12.6268	--	--	6/10/2013	3:40:00
23.9323	18.8901	12.625	12.625	--	--	6/10/2013	3:30:00
24.0499	18.8994	12.6383	12.6383	--	--	6/10/2013	3:20:00
24.3451	18.7982	1	12.508	--	--	6/10/2013	3:10:00
24.5635	18.8999	12.5095	11.6912	10.9156	8.5289	6/10/2013	3:00:00
24.4352	18.8801	12.6385	11.7202	10.9468	8.5189	6/10/2013	2:50:00
24.4163	18.9077	12.4909	11.7178	10.9335	8.5382	6/10/2013	2:40:00
24.5566	18.8979	12.6235	11.6681	10.8673	8.5309	6/10/2013	2:30:00
24.7333	18.8939	12.6414	11.6729	10.8616	8.3757	6/10/2013	2:20:00
24.8689	18.9005	12.6176	11.6936	10.8578	8.2792	6/10/2013	2:10:00
--	--	12.6222	--	--	--	6/10/2013	2:00:00
--	--	12.609	--	--	--	6/10/2013	1:50:00
--	--	12.5236	--	--	--	6/10/2013	1:40:00
--	--	12.5536	--	--	--	6/10/2013	1:30:00
--	--	12.6052	--	--	--	6/10/2013	1:20:00
--	--	12.5446	--	--	--	6/10/2013	1:10:00
25.6841	--	12.5558	11.6249	10.9109	8.3494	6/10/2013	1:00:00
25.3908	--	12.8104	11.7239	10.9668	8.2961	6/10/2013	0:50:00

Automatic QC for Repeating values

A
R
R
B

$$\frac{A + B}{2} - 0.01 < R < \frac{A + B}{2} + 0.01$$

- Repeating values are unmodified if it falls within the above range
- A and B are two different numbers & R is the repeating value
- R is considered okay if it satisfies the equation

Work Progress

Work Completed

- Implemented all the algorithms on matlab and tested individually on 8 datasets that spans 2 years.

Yet to be completed

- Combine all the algorithms along with database connection
- To develop a GUI console that enables modification of parameters


```
Editor - /home/sathish.samiappan/Documents/MATLAB/qc_main.m
File Edit View Go Cell Tools Debug Desktop Window Help
Stack
- 1.0 + - 1.1 x
1 - clear all;
2 - close all;
3 - cto;
4
5 - warning off;
6
7 & Path
8 - addpath('..:/src;/code');
9
10 % Get active deployments
11
12 database = 'tao02';
13 dbuser = 'samiapp';
14 pw = 'checker';
15
16 connectdb(database, dbuser, pw);
17 [DEPLOYMENT_NAME, DEPLOYMENT_SEQ] = get_active_deployments();
18 [SENSOR_SEQ, MOORING_DEPTH] = get_sensor_ids(DEPLOYMENT_SEQ);
19 [DEPLOYMENT_NAME, DEPLOYMENT_SEQ, MOORING_DEPTH, SENSOR_SEQ] = remove_empty(DEPLOYMENT_NAME, DEPLOYMENT_SEQ, MOORING_DEPTH, SENSOR_SEQ);
20
21
22
23
24 %sql('use DATA');
25 % query = sprintf('select MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG, MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG');
26 % [MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG] = query;
27
28 for each_deploy_seq = 1:length(DEPLOYMENT_NAME)
29     for each_sensor_seq = 1:length(MOORING_DEPTH)
30         query = sprintf('select MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG, MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG');
31         [MEASUREMENT_TIME, MEASUREMENT_VALUE, QUALITY_FLAG, RELEASE_FLAG] = query;
32     end
33 end
34
35 %sql('status');
36 %sql('close');
```

qc_main.m | get_active_deployments.m | connectdb.m | get_sensor_ids.m | remove_empty.m

[sathish.samiappan@... MATLAB 7.6.0 (R200... [shell_scripts] Editor - /hc

```
Editor - C:\Users\sathish.samiappan\Documents\MATLAB\TAO Spike\TAO Spike Analysis Files\main_code_v04.m
File Edit View Go Cell Tools Debug Desktop Window Help
Stack
- 1.0 + - 1.1 x
```

```
25
26
27 % Prompt for Input Options
28 % Get the information about the variable that need to be QCed and option
29 % for generating an excel file with outputs
30
31 prompt('Enter the Name of Variable to generate QC: ');
32 % Generate Excel file (see/90) ;
33 name = 'Input Options';
34 %sql('status');
35 %sql('close');
36
37 % Read the NetCDF File
38 % Read the netCDF selected netCDF variable, TIME and DEPIN variables and
39 % if the file is corrupted or wrong file format is selected show an error
40 % dialog and exit the program gracefully
41
42 %sql('status');
43 %sql('close');
```

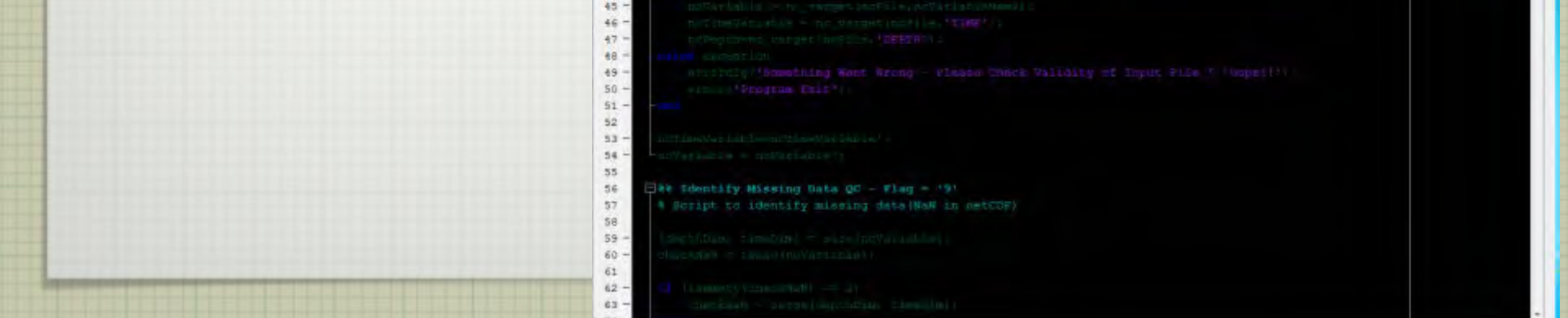
```
44 %sql('status');
45 %sql('close');
46 % Read the NetCDF File
47 % Read the netCDF selected netCDF variable, TIME and DEPIN variables and
48 % if the file is corrupted or wrong file format is selected show an error
49 % dialog and exit the program gracefully
50
51 %sql('status');
52 %sql('close');
53
54 % Read the NetCDF File
55 % Read the netCDF selected netCDF variable, TIME and DEPIN variables and
56 % if the file is corrupted or wrong file format is selected show an error
57 % dialog and exit the program gracefully
58
59 %sql('status');
60 %sql('close');
```

MATLAB 7.6.0 (R2008a)

Current Directory: /home/sathish.samiappan/Documents

Stack

Name	Value	Min	Max
DEPLOYMENT_NAME	<28x1 cell>		
DEPLOYMENT_SEQ	<28x1 double>	2086	2179
MEASUREMENT_TIME	[7.3545e+05; 7.3... 7.35... 7.35...		
MEASUREMENT_VALUE	[7.8841; 7.8897; 7... 7.78... 7.88...		
MOORING_DEPTH	<1x28 cell>		
QUALITY_FLAG	[3; 3; 3; 3; 3]	3	3
RELEASE_FLAG	[1; 1; 1; 1; 1]	1	1
SENSOR_SEQ	<1x28 cell>		
database	'tao02'		
dbuser	'tempqc'		
each_deploy_seq	1	1	1
each_sensor_seq	12	12	12
pw	'checker'		
query	'select MEASURE...		



References

- TAO project overview web page at
[“tao.ndbc.gov/proj_overview/proj_overview_ndbc.shtml”](http://tao.ndbc.gov/proj_overview/proj_overview_ndbc.shtml)
 - Unidata – NetCDF information at
[“http://www.unidata.ucar.edu/software/netcdf/”](http://www.unidata.ucar.edu/software/netcdf/)
 - OceanSITES ftp server “<ftp://data.ndbc.noaa.gov/data/oceansites/DATA/>”
 - NDBC threads server
[“http://dods.ndbc.noaa.gov/thredds/catalog/data/oceansites/DATA/catalog.html”](http://dods.ndbc.noaa.gov/thredds/catalog/data/oceansites/DATA/catalog.html)
 - NDBC threads server(TAO Legacy)
[“http://dods.ndbc.noaa.gov/thredds/catalog/data/tao/catalog.html”](http://dods.ndbc.noaa.gov/thredds/catalog/data/tao/catalog.html)
-

Acknowledgements

- Karen Grissom*, Dawn Petriatis* and Raymond Beets*
– For their overall ideas and support
- Jing Zhou* – For his support with MySQL database connectivity
- Yee H. Lau† and Patrick J. Fitzpatrick† – For their support in computing thresholds using climatology data

* - National Data Buoy Center

† - Geosystems Research Institute – Mississippi State University

Thank You



TAO Buoys on Map

