

Enhancing BGC data usability with Sensor profile audits

Josh Plant Ken Johnson & Tanya Maurer, MBARI

ADMT20, October 13-18, 2019

Laboratoire d'Océanographie de Villefranche-sur-Mer



How can we enhance BGC data usability?

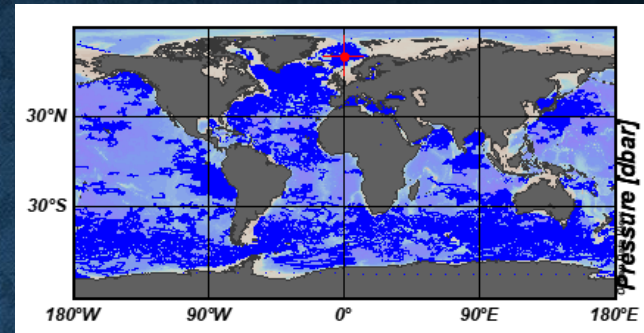
The easier the data is to access & use the more successful BGC Argo will be

Synthetic files have transformed global data access!!

Two ways to enhance usability:

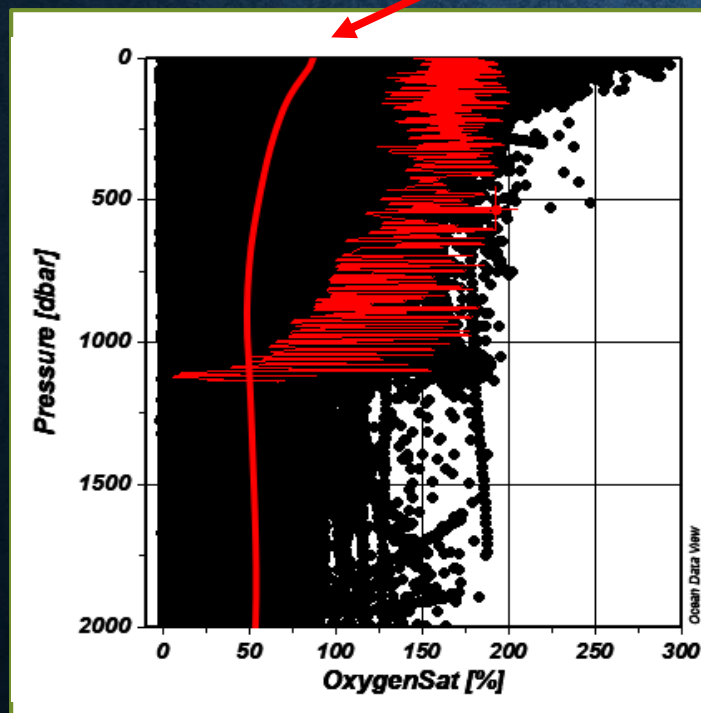
- Propagate more adjusted data to the GDAC's (more on this tomorrow)
- Flag obviously bad data as bad

All BGC Argo data (excluding QC = 4)

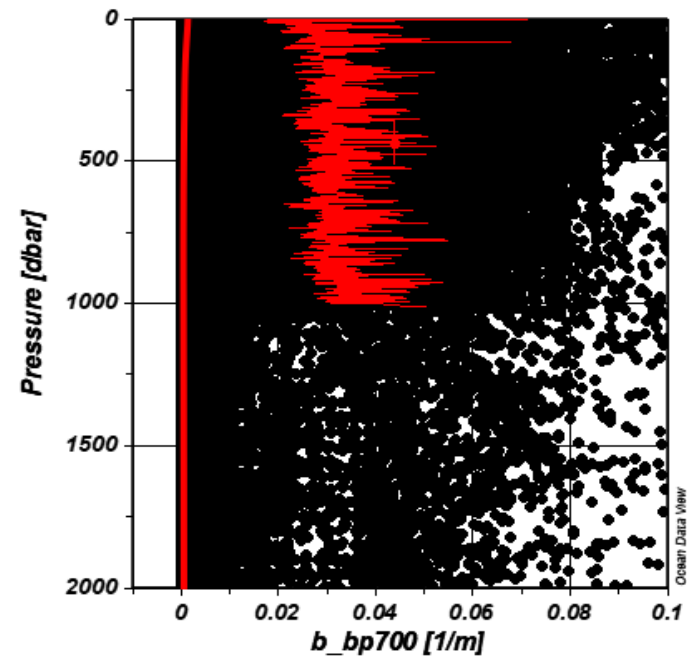


Fleet average profiles

DOXY



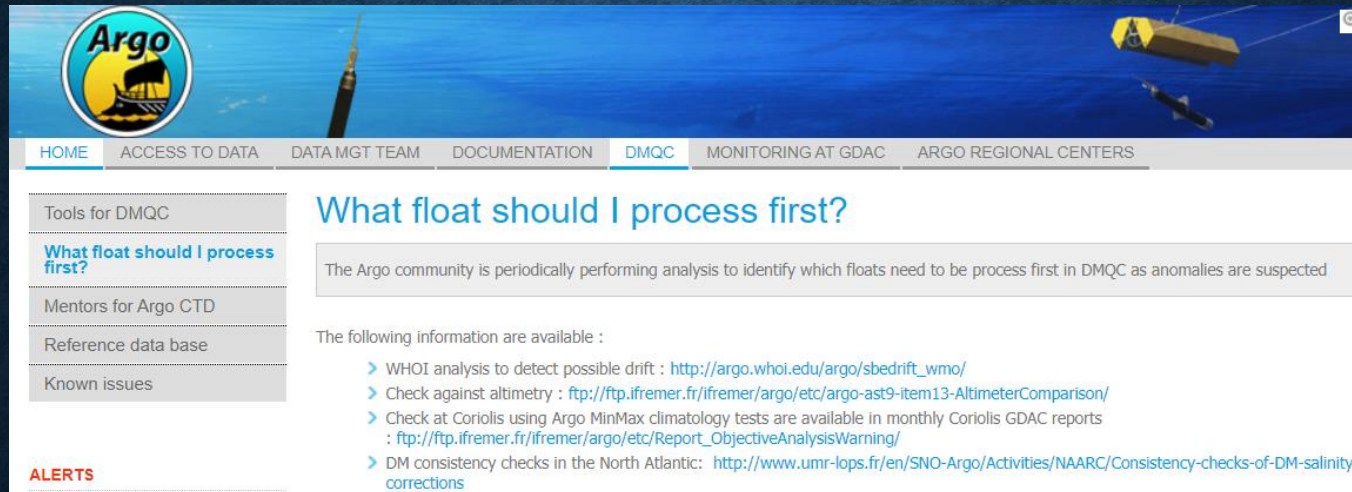
BBP700



Global Argo reports of anomalies

Core Argo does this to help DM operators QC T & S:

<http://www.argodatamgt.org/DMQC/What-float-should-I-process-first>



Christine Coatanoan-Girou

- T & S anomalies detected using objective analysis

Stephanie Guinehut

- T & S anomalies detected by comparing satellite sea level anomalies to float dynamic height anomalies
- We can do something similar for BGC Argo parameters

Global BGC Argo reports of anomalies

Compare sensor profile data to global statistics

Profile data vs reference data anomalies

- Oxygen – WOA2018
- NO₃, pH – CANYON-B or LIR's

Profile data vs global sensor statistics

- BBP700, CHL

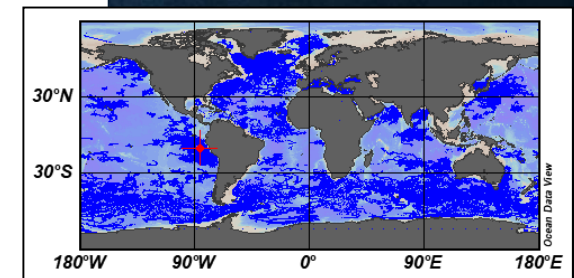
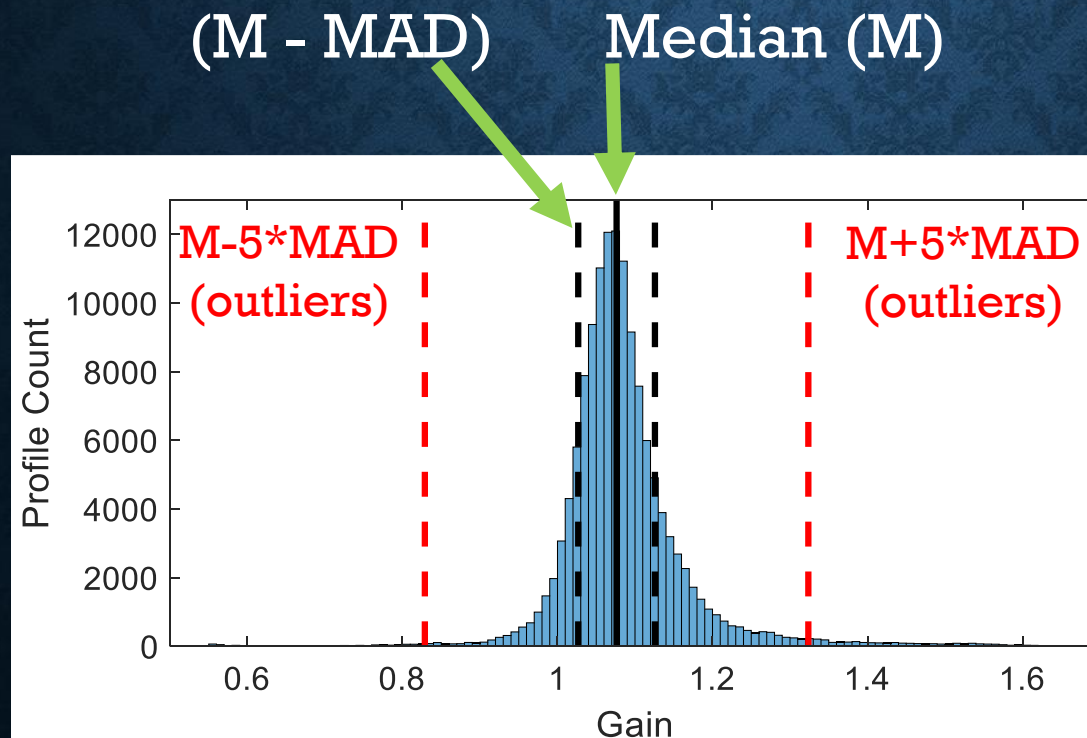
Detect outliers - Median Absolute Deviation (MAD)

- M = median of fleet test values (TV)
- $M_{anom} = \text{abs}(TV - M)$
- $MAD = \text{median}(M_{anom}) * B$ $B = 1.4826$
- $Z = M_{anom} / MAD$ (Outlier score)
- Choose a threshold for Z (using 5 for now)

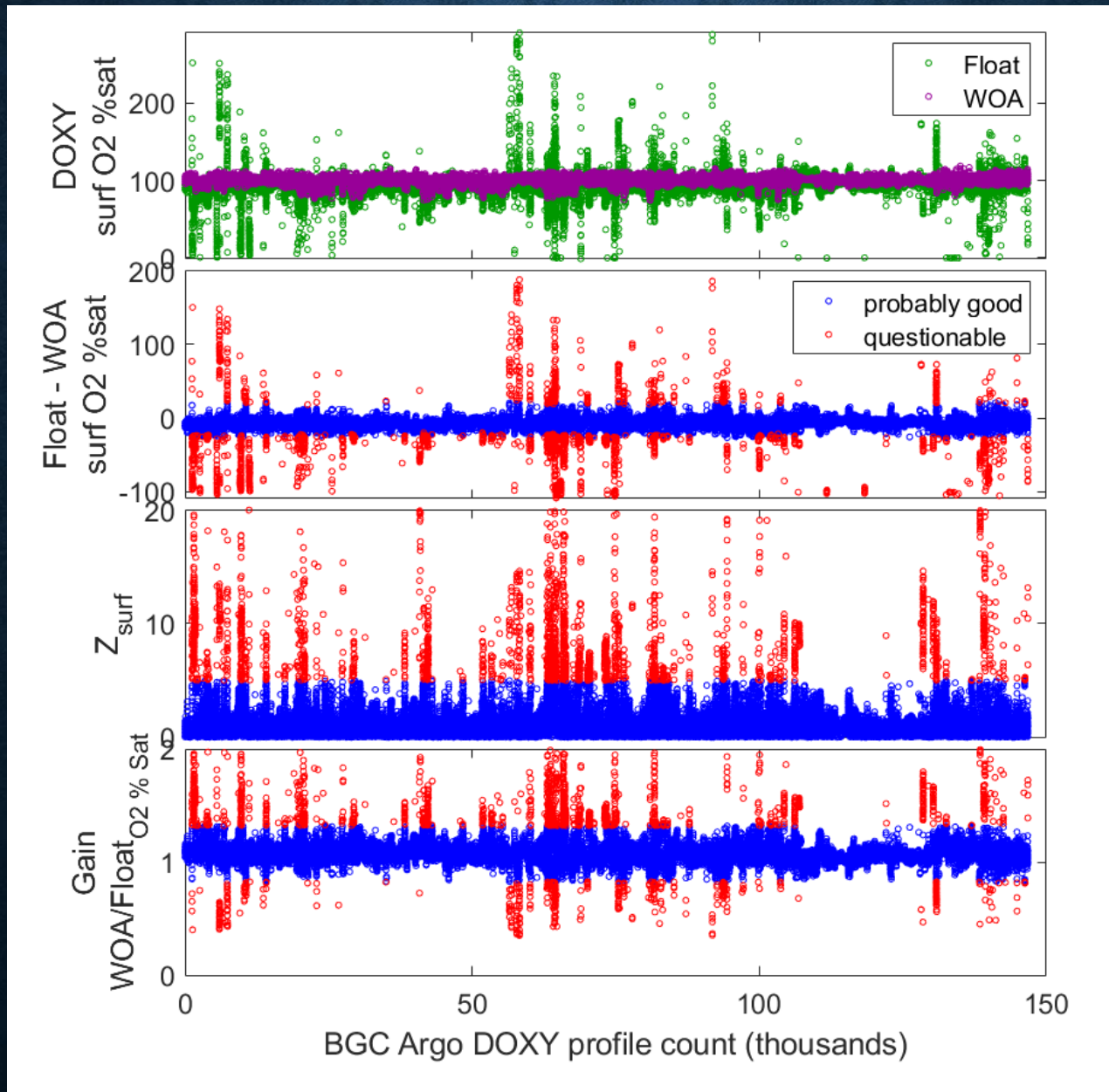
Global BGC Argo reports of anomalies

DOXY

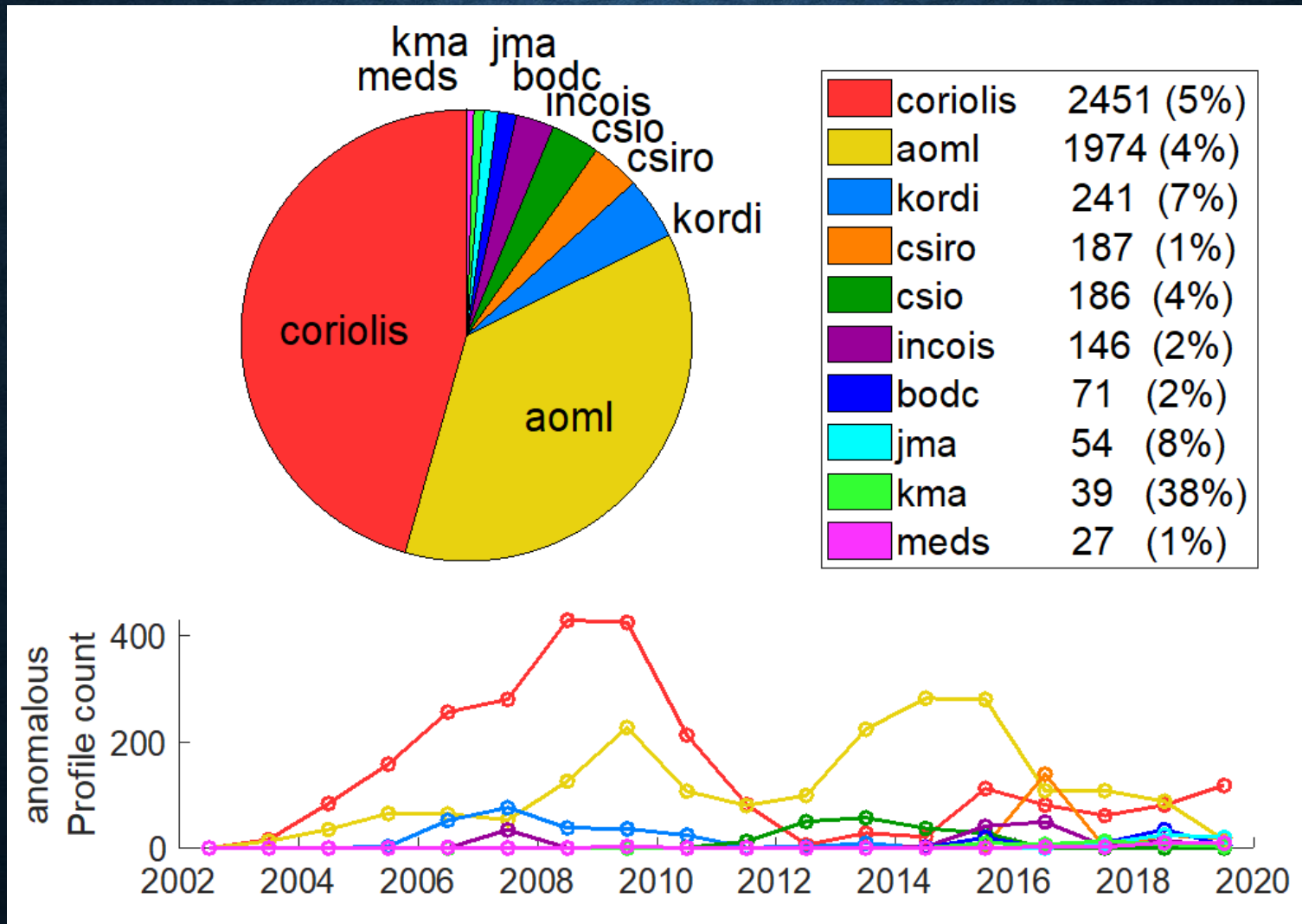
- 1086 floats & 146,000 profiles & ~18 years of data!
- Using DOXY not DOXY_ADJUSTED (only 40% corrected)
- Get WOA O₂ %sat profile for every float profile
- Test value = gain = [WOA O₂ %sat / DOXY % sat]_{surf}



Global BGC Argo reports of anomalies



Global BGC Argo reports of anomalies

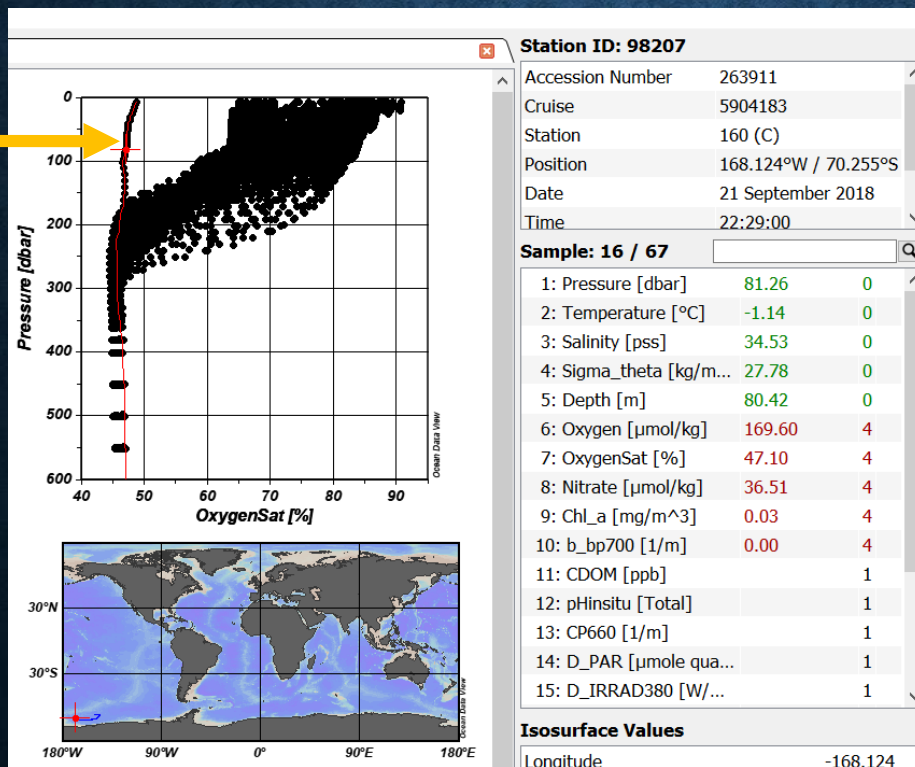


5904183 Cycle 160

Median deviation / MAD

13	//Reference:											
14	// Leys et al. (2013). Detecting outliers: Do not use standard deviation around the mean,											
15	// use absolute deviation around the median. http://dx.doi.org/10.1016/j.jesp.2013.03.013											
16	DAC	PI	WMO	cycle	sdn	lon	lat	float	WOA	Gsurf	Tsurf	Z_surf
1942	aoml	STEPHEN RISER	5904168	277	3/3/2019 22:20	179.8	-74.6	66.1	94.75	1.43	-0.85	7.22
1943	aoml	STEPHEN RISER	5904168	278	3/10/2019 18:52	179.8	-74.6	62.5	94.07	1.51	-1.09	8.67
1944	aoml	STEPHEN RISER	5904168	279	3/17/2019 15:47	179.8	-74.6	66.87	93.12	1.39	-1.28	6.39
1945	aoml	STEPHEN RISER, KENNETH JOHNSON	5904183	122	8/27/2017 10:13	-159.9	-71.9	68.3	90.79	1.33	-1.75	5.11
1946	aoml	STEPHEN RISER, KENNETH JOHNSON	5904183	160	9/21/2018 22:29	-168.1	-70.3	48.6	83.54	1.72	-1.59	12.99
1947	aoml	STEPHEN RISER, KENNETH JOHNSON	5904660	4	2/7/2016 17:57	-50.72	-59.9	72.4	98.8	1.36	0.69	5.83

Cycle
160



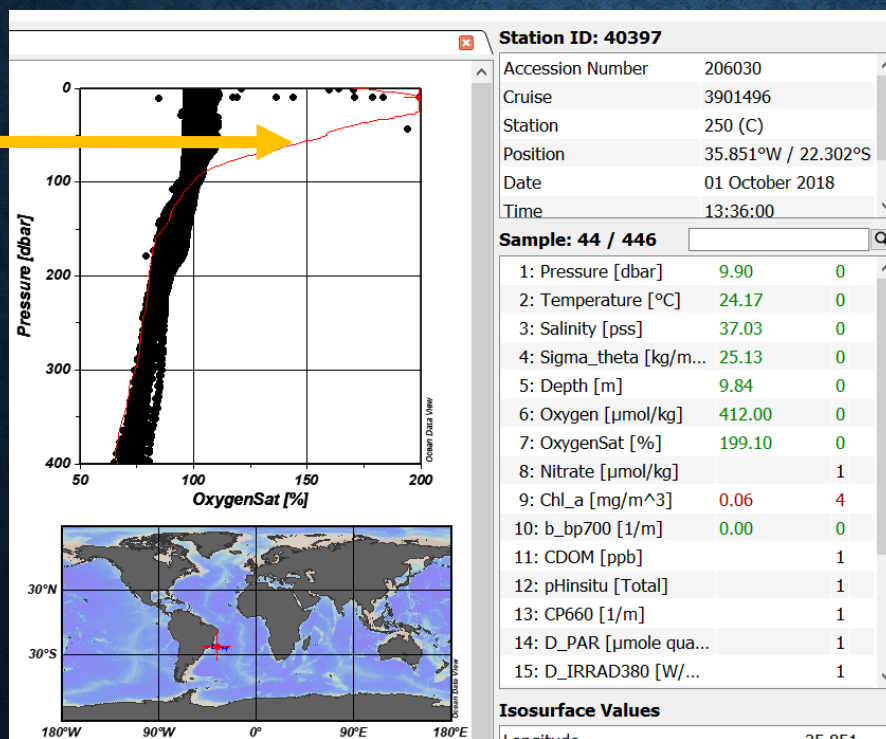
- Very low % O2 sat
- Profile should be marked bad
- APEX float – sensing foil blocked by gunk?

3901496 Cycle 250

Median deviation / MAD

13	//Reference:											
14	// Leys et al. (2013). Detecting outliers: Do not use standard deviation around the mean,											
15	// use absolute deviation around the median. http://dx.doi.org/10.1016/j.jesp.2013.03.013											
16	DAC	PI	WMO	cycle	sdn	lon	lat	float	WOA	Gsurf	Tsurf	Z_surf
1996	bodc	Giorgio Dall'Olmo	3901496	238	6/3/2018 13:35	-34.98	-21.5	170.5	100.42	0.59	25.53	9.85
1997	bodc	Giorgio Dall'Olmo	3901496	244	8/2/2018 13:35	-34.91	-21.6	159.5	100.98	0.63	24.41	8.96
1998	bodc	Giorgio Dall'Olmo	3901496	250	10/1/2018 13:36	-35.85	-22.3	199.1	102.63	0.52	24.17	11.33
1999	bodc	Giorgio Dall'Olmo	3901496	253	#####	-37.05	-22.3	136.5	103.43	0.76	24.13	6.44
2000	bodc	Giorgio Dall'Olmo	3901497	307	7/18/2018 12:39	-20.23	-31.9	129.7	97.58	0.75	17.66	6.55
2001	bodc	Giorgio Dall'Olmo	3901497	308	7/23/2018 12:34	-20.36	-32	128	97.85	0.76	16.92	6.3

Cycle
250



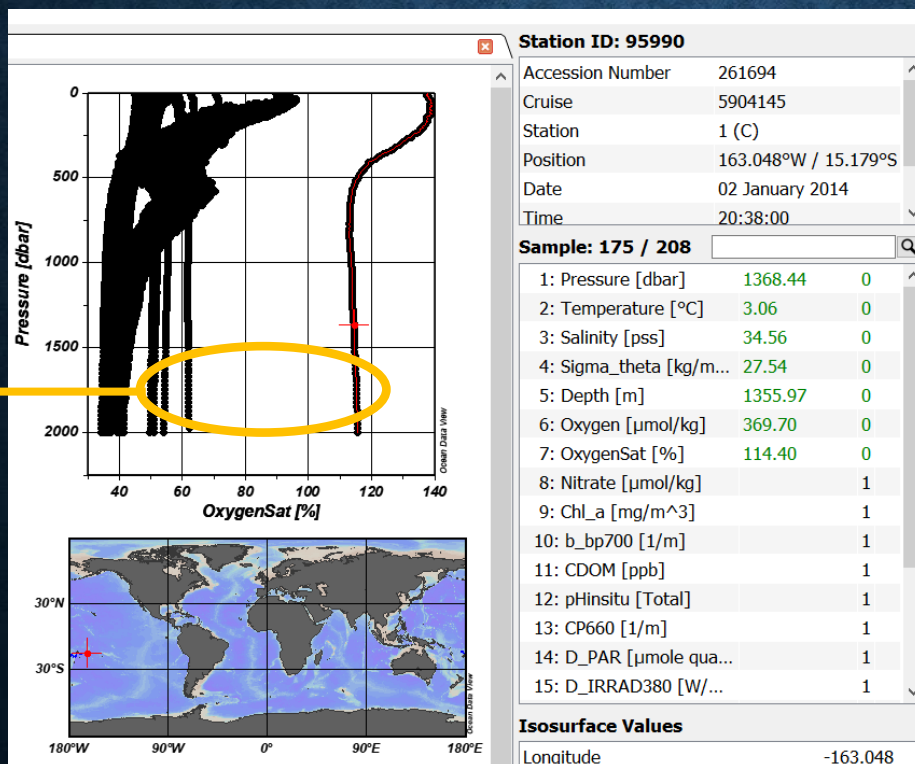
- Very high % O2 sat
- Profile is already marked bad EXCEPT for PRES = 9.9

5904145 Cycles 1-5

Median deviation / MAD

13	//Reference:											
14	// Leys et al. (2013). Detecting outliers: Do not use standard deviation around the mean,											
15	// use absolute deviation around the median. http://dx.doi.org/10.1016/j.jesp.2013.03.013											
16	DAC	PI	WMO	cycle	sdn	lon	lat	float O2surf %sat	WOA O2surf %sat	Gsurf	Tsurf	Z_surf
1686	aoml	STEPHEN RISER	5904126	11	9/23/2013 13:12	-138	50.7	66.73	102.68	1.54	13.38	9.35
1687	aoml	STEPHEN RISER	5904145	1	1/2/2014 20:38	-163.1	-15.2	137.7	100.1	0.73	28.43	7.06
1688	aoml	STEPHEN RISER	5904145	2	1/4/2014 8:03	-163.1	-15.2	70.66	100.32	1.42	28.56	6.94
1689	aoml	STEPHEN RISER	5904145	3	1/5/2014 7:08	-163.2	-15.2	62.73	100.48	1.6	28.49	10.62
1690	aoml	STEPHEN RISER	5904145	4	1/6/2014 7:19	-163.2	-15.2	58.93	100.72	1.71	28.58	12.79
1691	aoml	STEPHEN RISER	5904145	5	1/7/2014 6:54	-163.3	-15.2	57.35	100.86	1.76	28.45	13.79

Cycles 1-5



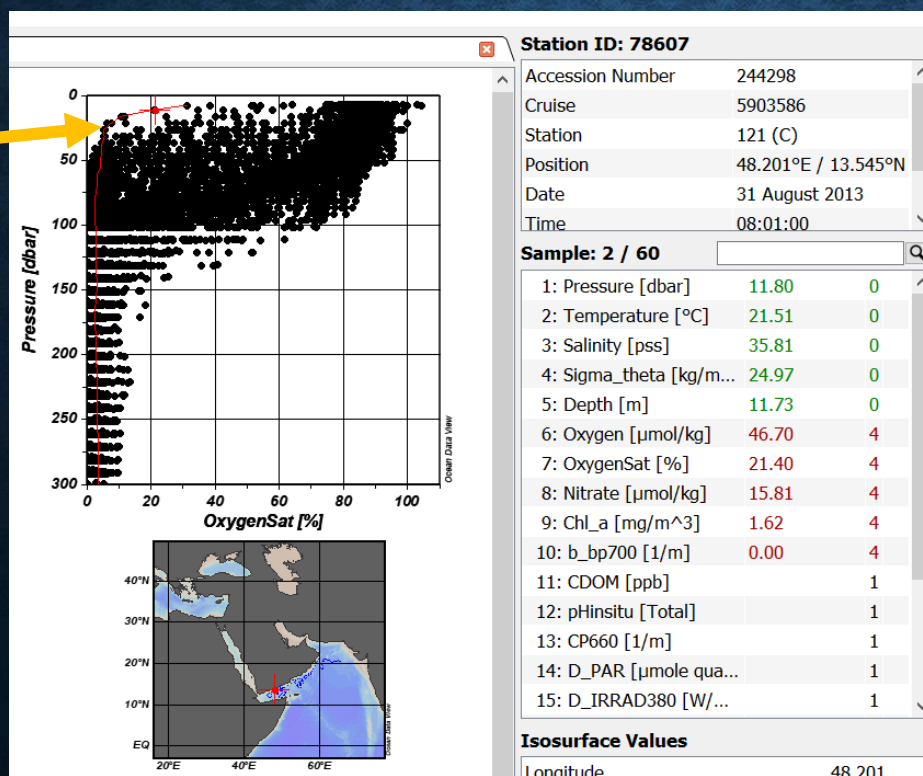
- Profiles should be marked bad or corrected for strong sensor drift

5903586 Cycle 121

Median deviation / MAD

13	//Reference:											
14	// Leys et al. (2013). Detecting outliers: Do not use standard deviation around the mean,											
15	// use absolute deviation around the median. http://dx.doi.org/10.1016/j.jesp.2013.03.013											
16	DAC	PI	WMO	cycle	sdn	lon	lat	float O2surf %sat	WOA O2surf %sat	Gsurf	Tsurf	Z_surf
1622	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	13	2/17/2012 5:26	64.33	20.6	79.43	106.88	1.35	24.06	5.44
1623	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	121	8/31/2013 8:01	48.2	13.5	21.07	86.91	4.13	21.3	61.62
1624	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	176	6/13/2014 0:43	49.54	14.4	76.23	104.76	1.37	25.7	6.02
1625	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	177	6/18/2014 6:07	49.45	14.4	71.2	102.96	1.45	24.48	7.47
1626	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	180	7/3/2014 21:20	49.41	14.4	56.27	97.14	1.73	24.23	13.14
1627	aoml	STEPHEN RISER , KENNETH JOHNSON	5903586	183	7/19/2014 12:53	49.4	14.4	42.87	91.9	2.14	22.43	21.57

Cycle
121



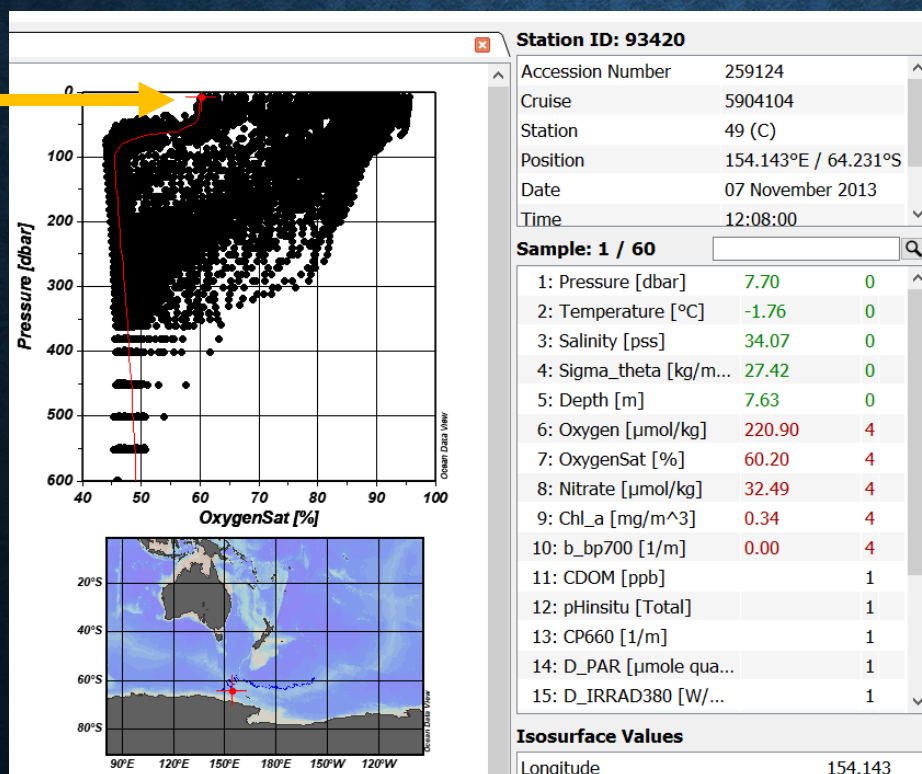
- Gulf of Aden, Arabian Sea
- Very High “Z score”
- Anomaly may be real
- WOA may not capture low O₂ upwelled waters

5904104 Cycle 49

Median deviation / MAD

13	//Reference:											
14	// Leys et al. (2013). Detecting outliers: Do not use standard deviation around the mean,											
15	// use absolute deviation around the median. http://dx.doi.org/10.1016/j.jesp.2013.03.013											
16	DAC	PI	WMO	cycle	sdn	lon	lat	float	WOA	Gsurf	Tsurf	Z_surf
1676	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	41	9/27/2013 8:03	152.9	-64.2	60.7	90.46	1.49	-1.77	8.37
1677	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	49	11/7/2013 12:08	154.1	-64.2	60.1	91.99	1.53	-1.77	9.18
1678	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	50	11/12/2013 16:04	154.3	-64.2	60.5	91.98	1.52	-1.75	8.98
1679	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	51	11/17/2013 19:34	154.5	-64.2	61.5	92.04	1.5	-1.74	8.49
1680	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	52	11/23/2013 2:30	154.6	-64.3	62.67	92.19	1.47	-1.69	7.98
1681	aoml	STEPHEN RISER , KENNETH JOHNSON	5904104	53	11/28/2013 8:50	154.8	-64.3	63.13	92.29	1.46	-1.74	7.79

Cycle 49



- 64S & $T_{surf} = -1.7$
- Under ice?
- WOA based on very little data
- Anomaly probably real

Summary

- Many ways to test for outliers
- MAD is simple & can be automated
- Same approach can be used for all BGC parameters
- Most of the DOXY data is good < 5% gets flagged!
- When adjusted parameters become well populated test them as well

Work flow:

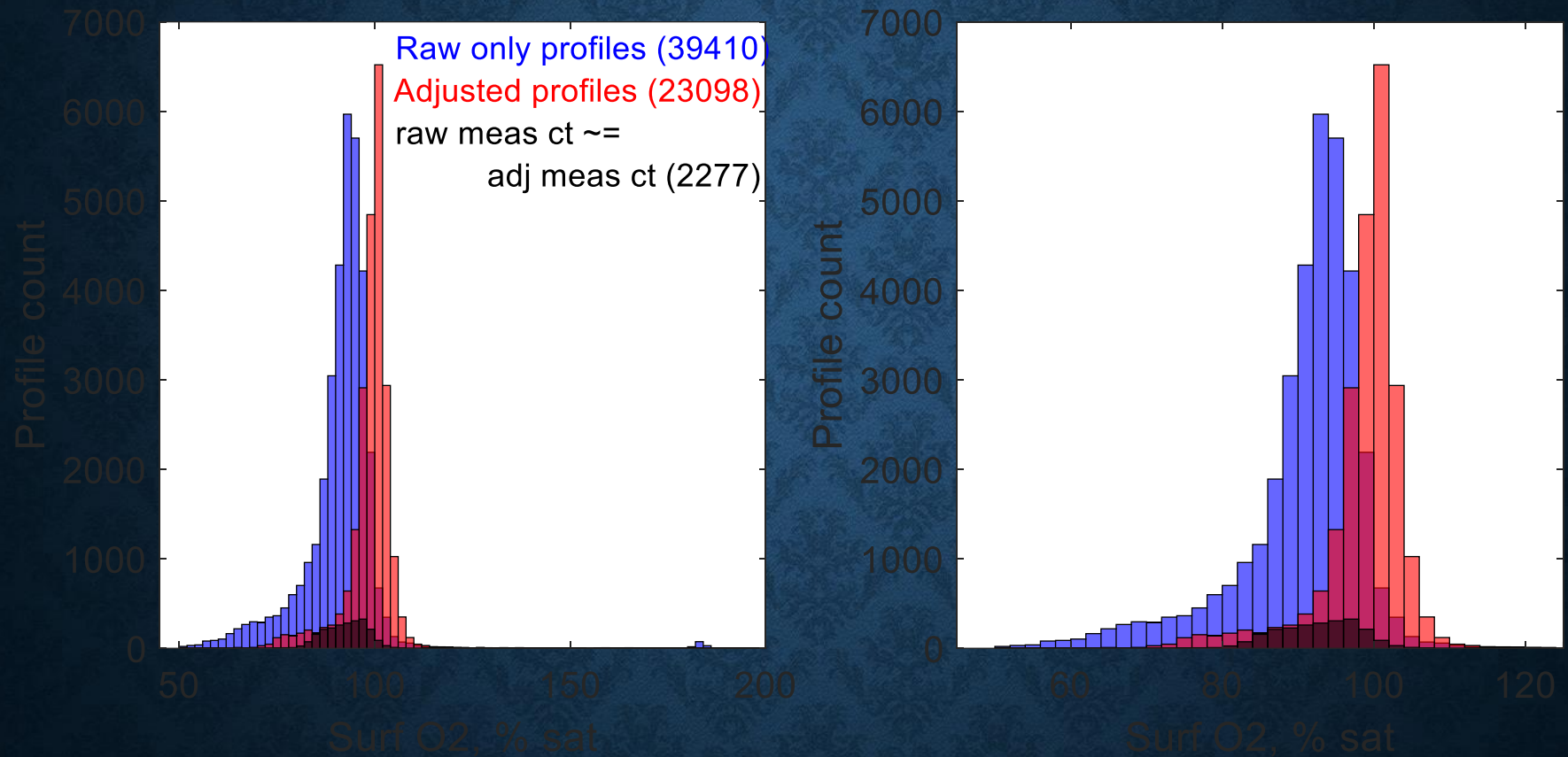
- Send out anomaly list
- DAC's /PI's adjust flag & re-submit data or confirm anomaly is real
- Real anomalies excluded from future reports

Side benefit:

- Average gain factor (WOA/Float %O₂ sat) generated for DOXY => $\text{DOXY} \times \text{Gain} = \text{DOXY_ADJUSTED}$

Questions?

BGC Argo GDAC Oxygen profile stats (from Sprof files!)



3 Problems:

- 1) Adjusted data not propagated to the GDAC – 60 % of profile
- 2) Bad data are not flagged as bad
- 3) Raw measurement count ~= adjusted measurement count