

## **PROFILE\_<PARAM>\_QC**

The PROFILE\_<PARAM>\_QC variables are defined to be “the worst QC flag of any level in the profile.” A concern was expressed that this definition will lead to a misinterpretation of the quality of the Argo data.

There was a long discussion on the DM mailing list regarding this issue. The one thing that is clear is that everyone that contributed thinks this variable is improperly defined. The following is a consolidation of these discussions and a recommendation for consideration.

Section A contains a partial compilation of the e-mail discussion. This can be referenced to see how these proposals were formulated and, especially, to see the dissenting opinions.

### ***Is this or a similar variable needed?***

There was no clear consensus of the group whether this variable is needed. Opinions were expressed that, if properly defined, this parameter may help users scan through the Argo data set more efficiently.

### **Proposal**

Retain an “overall quality” variable flag but redefine it to make it more useful.

### ***What should the variable be and how should it be defined?***

The general consensus was to use the percentage of good levels to indicate the overall profile quality. There was much discussion of how to represent the value.

First option: Encode the value into the existing PROFILE\_<PARAM>\_QC variable, which is limited to one character. (Changing its length is not straightforward.)

Second option: Introduce a new variable and eliminate the existing variable.

Both options are presented as proposals. We will need to choose one during the meeting.

### **Proposal #1: Redefine PROFILE\_<PARAM>\_QC**

```
char PROFILE_<PARAM>_QC(N_PROF);  
PROFILE_<PARAM>_QC:long_name = "Percentage of good data in the  
profile";
```

```
PROFILE_<PARAM>_QC:conventions = "A: 76-100% good; B: 51-75% good; C:
    26-50% good; D: 1-25% good; E: all bad";
PROFILE_<PARAM>_QC:_FillValue = "";
```

#### Advantages:

- o Simple to implement.
- o Smallest change to the data system

#### Disadvantages / Problems: (Similar to Proposal #2)

- o Might cause confusion to current users. However, changing from numbers to letters will make it clear to users that something has changed.
- o Existing files. We don't want to have to rewrite the entire history of data files. (Solution?)
- o Multi-profile files created by GDACs. How do we combine the new and old definitions in the multi-profile files. (Are GDACs willing to recompute values? Probably not.)

## Proposal #2: Introduce a new variable.

Delete PROFILE\_<PARAM>\_QC

Add:

```
int PROFILE_<PARAM>_QUALITY_PERCENTAGE(N_PROF);
PROFILE_<PARAM>_QUALITY_PERCENTAGE:long_name = "Percentage of good data
    in the profile";
PROFILE_<PARAM>_QUALITY_PERCENTAGE:conventions = "Argo user's manual";
PROFILE_<PARAM>_QUALITY_PERCENTAGE:_FillValue = 99999;
```

The value is computed as:

```
PROFILE_<PARAM>_QUALITY_PERCENTAGE = (# good levels / # QC'd levels) * 100
- Rounded to the nearest integer
- If the percentage is less than 1 it will be rounded to 1.
- If the percentage is larger than 99 but less than 100 it will be rounded to 99.
```

#### Advantages:

- o Variable name is explicit.
- o No confusion over redefinition of an existing variable.
- o Value is an explicit numerical value. No coding required.

#### Disadvantages / Problems: (Same as above)

- o Might cause confusion to current users. However, removing the existing variable will be a clear indication that something has changed.
- o Existing files. We don't want to have to rewrite the entire history of data files. (Solution?)
- o Multi-profile files created by GDACs. How do we combine the new and old definitions in the multi-profile files. (Are GDACs willing to recompute values? Probably not.)

## Rules for interpreting per-level QC codes:

If all QC flags are 0 (not QC'd) set the profile flag to 0 (zero).

If all QC flags are 9 (missing) set the profile flag to “\_FillValue”

Good QC flags: 1, 2, 5, 8

Bad QC flags: 3, 4

QC flag 0 or 9: Do not count level in total or as bad.

Since QC flags 6 and 7 are not used, they are not discussed.

## A. Comments of the Data Management Members

I could not realistically include all of the e-mails received regarding this issue. What I have tried to do is organize some of the relevant e-mail traffic to stimulate further discussion. I have specifically included all dissenting opinions.

<<<< *The message that raised the issue.* >>>>

**Dean Roemmich**            **2004-05-07**

Colleagues,

There has, in the past, been some discussion at ADMT meetings of the meaning of the "profile\_qc" flags. The practice at present is to set these flags to the maximum value of the qc flags of the individual data points. The justification for this has been, "that is the way the operational centers want it".

I'm somewhat concerned that this practice will label some perfectly good profiles as bad, and many partially good profiles as bad. Anyone looking at the statistics of Argo will reach the wrong conclusion about the percentage of bad profiles, and Argo users will reject good data.

An example is attached, consisting of the ncdump version of one profile. This is an instrument with one of the older SBE pressure sensors that drifted enough (9 dbar) so that the CTD came out of the water during its final depth bin. In the delayed-mode file, the adjusted pressures reflect the 9 dbar correction for drift, except for the shallowest value which is unchanged from the raw data since it carries a qc flag of 4.

Since the shallowest depth bin has qc flags of 4, the profile\_qc flags are also 4, in spite of the fact that there is a complete profile here with perfectly good data.

If the purpose of the profile\_qc flags is to allow users to quickly sort files into ones they want to use and ones they don't want to use, should we consider a future change to encourage more efficient use of the data? One way to accomplish this would be to have 2 sets of profile\_qc flags rather than one. One set would contain the maximum values for the profile and one set would contain the minimum values. In the present example, instead of PROFILE\_Psal\_QC = "4", we would have PROFILE\_Psal\_MAX\_QC = "4" and PROFILE\_Psal\_MIN\_QC = "1" . This would allow the user to decide whether or not he wants to sort through "partial" profiles.

**Ann Thresher**            **2004-05-09**

Dear Dean,

I've always liked this idea and been uncomfortable with the way the flag was originally set up. Most users want to pull in all profiles with ANY good data in them, and are not really interested in whether there is any bad data present (except as a warning that they can't use the profile as-is without paying attention to the individual QC flags). As it is,

the user is forced to read all profiles (and their QC) in order to find the ones that have useable data so the profile QC flag is pretty useless. Hence turning the definition around would make sense. If the data centres wanted/needed the opposite (the way it is now), erecting another variable as you have suggested solves the problem.

**Claudia Schmid**                      **2004-06-16**

All,

maybe this works:

- 1 - all good
- 2 - <25% bad data
- 3 - 25-75% bad data
- 4 - >75% bad data

Claudia

**Bob Keeley**                      **2004-05-17**

Dear Dean,

The profile QC flag represents the "worst" flag of the profile not the overall quality of the profile. With this understanding, users should not reject whole profiles because this flag is set to be something other than 1. Neither should they use this flag as a way to quickly count good or bad profiles. The reason for having the flag set the way it is, is to alert users of profiles that need special attention, those where problems have been found.

The vast majority of the data profiles in Argo or any other program always have one or more values that have a flag of 1 assigned at one or more levels. This means that virtually every profile will get a flag of 1 as the "best" flag found in the profile. This seems to me to be even less of a help to users wanting to know what to do with a profile.

I think Claudia's idea adds value but I would worry that using values of 1-4 would be interpreted as quality flags. Why not just report the percentage rather than a code? We could always specify that the reported percentage be rounded to the nearest 25% as in Claudia's example or nearest 10% or whatever we liked. If we use this idea, is there any reason to keep the presently assigned profile qc flag?

Regards, Bob

**Sylvie Pouliquen**                      **2004-05-27**

Dear Bob

If users were reading the user manual there will be no problem with this flag but as they ,most of the time, don't I think that it's misleading either with the present definition or the reserve one.

So I'm in favor of changing it to what Claudia proposed. But I agree with you that using 1-4 could be interpreted according to the present QC flag scale so why not just put the percentage of good data ( ie 87 means 87% of good data) in this field. It's easy to use by user, easy to generate and easy to patch at GDAC for the profiles that are already there.

**Takashi Yoshida**                      **2004-05-27**

> So I'm in favor of changing it to what Claudia proposed. But I agree  
> with you that using 1-4 could be interpreted according to the present QC  
> flag scale so why not just put the percentage of good data ( ie 87  
> means 87% of good data) in this field. It's easy to use by user, easy to  
> generate and easy to patch at GDAC for the profiles that are already  
there.

I don't think it is very easy to DACs and GDACs to put the percentages  
because length of PROFILE\_<PARAM>\_QC is short (one character). Do you want  
to expand the length? No, I don't. Putting a character, which corresponds  
the percentages of good data, seems good. If use of 1-4 is confusing, how  
about using A-D.

- A - all good
- B - <25% bad data
- C - 25-75% bad data
- D - >75% bad data

<<<<<<The following two e-mails starting the idea for  
the current proposal. >>>>>>

**Takashi Yoshida**                      **2004-06-17**

Dear all,

PROFILE\_<PARAM>\_QC issue:

I am still reluctant to change the length or type of PROFILE\_<PARAM>\_QC.

However, if many people think one character code is insufficient to  
describe the quality of a profile and consequently we have to make some  
modification on the profile format, I suggest to introduce a new  
variable which has more self-explaining name and to keep the existing  
global quality flag unchanged.

Changing the length of existing variable seems to be confusing because  
one of the characters can be misread by a user who don't notice the  
change.

The new variable could be  
integer or character\*3 (or 4) PERCENTAGE\_GOOD\_<PARAM>.

On a different topic:

We (Bob Keeley, myself and others) had a discussion about the profile  
flags (some of this discussion went to the mailing list). Basically we  
came to the conclusion that the flags 0-9 don't work well for overall  
profile flags.

**Claudia Schmid**                      **2004-06-16**

Here is what we propose:

The profile flag will be given as a percentage based on the length of the  
profile.

The rules for estimating the percentage are:

If the QC flag is 1, 5, 8 count the value as good  
If the QC flag is 2, 3, 4 count the value as bad  
If the QC flag is 0 or 9 do not count level as total or as bad

If all QC flags are 0 set the profile flag to 999  
If all QC flags are 9 set the profile flag to -99

Since QC flags 6 and 7 are not used, we excluded them.

If the percentage is less than 1 it will be rounded to 1.  
If the percentage is larger than 99 but less than 100 it will be rounded to 99.  
(That can be called fuzzy mathematics, but it does make some sense.)

Examples - a profile has:

50 times flag=1, 3 times flag=9 and 7 times flag=4. The overall QC flag will be 12 ( $7/57*100=12.2807$ ).

199 times flag=1, 0 times flag=9 and 1 time flag=4. The overall QC flag will be 1 ( $1/200*100=0.5$ ).

1 time flag=1, 0 times flag=9 and 199 times flag=4. The overall QC flag will be 99 ( $199/200*100=99.5$ ).

<<<<<<I left out some e-mails about certain details. They are reflected in the above proposal.>>>>>>

**Bob Keeley**            **2004-06-18**

Dear All,

I am trying to combine the different messages on profile qc flags. I think that this information is not a QC flag (the Argo flag scale is not appropriate for this).  
So I agree with Takashi to create a new integer variable that will contain the percentages described by Claudia.  
The existing PROFILE\_<PARAM>\_QC should be removed.  
Here is a proposal on the overall quality of argo profiles :

```
int PROFILE_<PARAM>_QUALITY_PERCENTAGE(N_PROF);
  PROFILE_<PARAM>_QUALITY_PERCENTAGE:long_name = "Percentage of good data
in the profile";
  PROFILE_<PARAM>_QUALITY_PERCENTAGE:conventions = "Argo user's manual";
  PROFILE_<PARAM>_QUALITY_PERCENTAGE:_FillValue = 99999;
```

<<<<<Rest of e-mail mirrored Claudia's suggestion from above.>>>>>

**Annie**                **2004-06-18**

PROFILE\_<PARAM>\_QC: If I may recount the origin of this debate ...

currently the value of this variable is the worst flag in the profile. Someone complained that this is misleading, and so other ideas were forwarded to make it less misleading ... 45%, 75%, A, B, C, D, etc. Now there are suggestions to add new variables to accommodate these ideas. Why don't we decide whether we actually need this in the first place?

Why do we need to give some overall judgement to a profile?  
I don't think we need such a thing. Different users with different needs will judge a profile differently. I suggest we simply remove PROFILE\_<PARAM>\_QC and not pursue anymore overall profile flags/good/bad/percentages, etc.