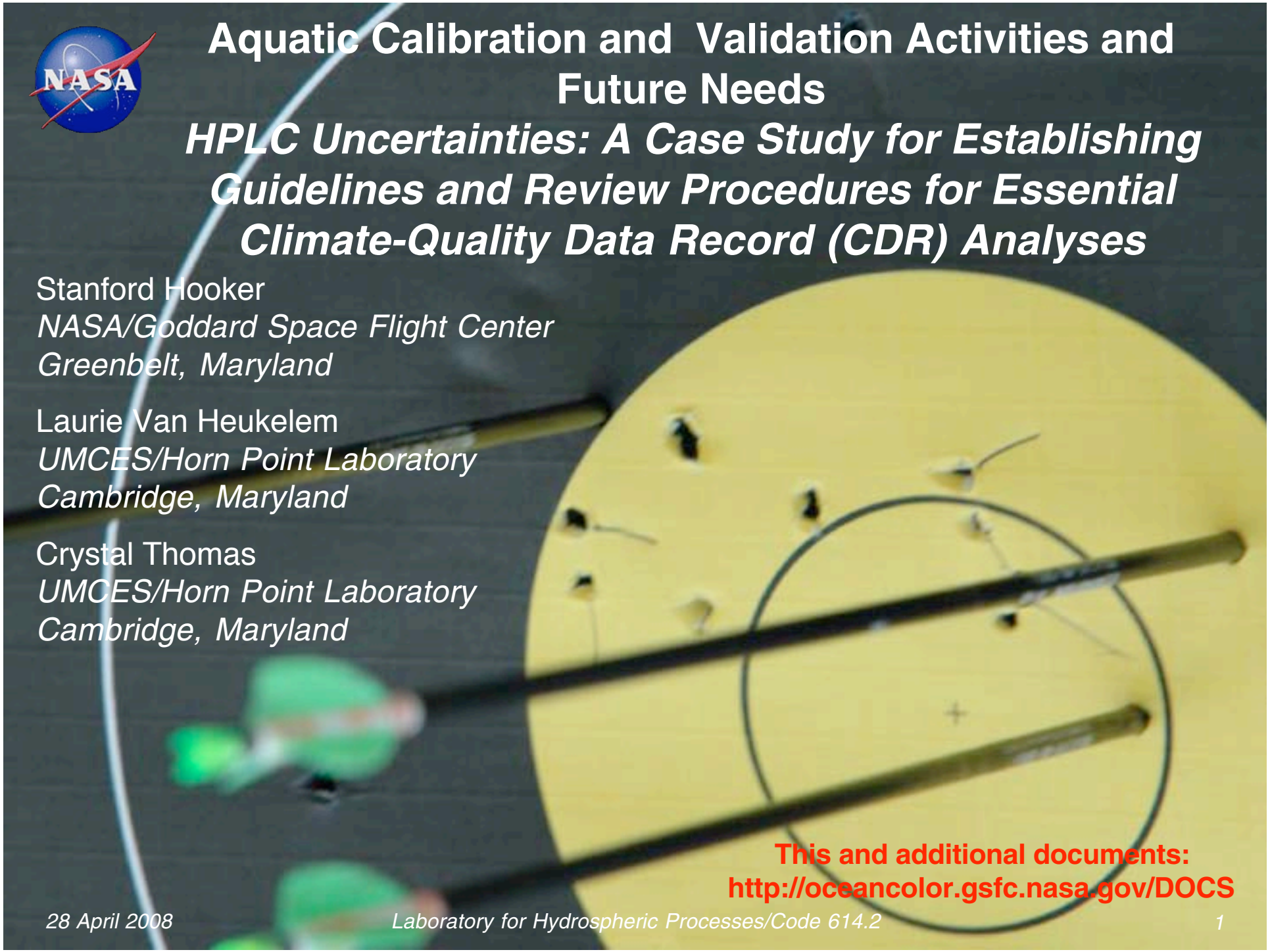
Aquatic Calibration and Validation Activities and Future Needs

HPLC Uncertainties: A Case Study for Establishing Guidelines and Review Procedures for Essential Climate-Quality Data Record (CDR) Analyses

Stanford Hooker
NASA/Goddard Space Flight Center
Greenbelt, Maryland

Laurie Van Heukelem
UMCES/Horn Point Laboratory
Cambridge, Maryland

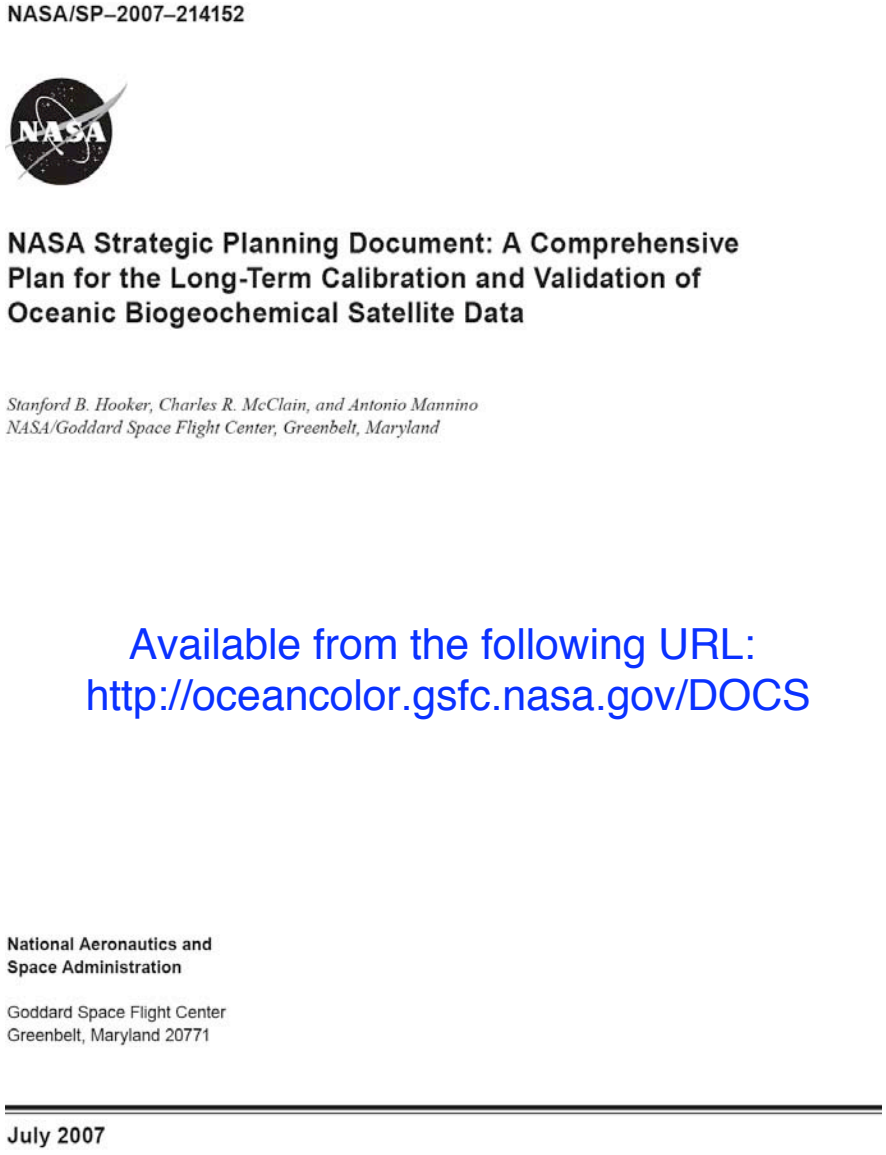
Crystal Thomas
UMCES/Horn Point Laboratory
Cambridge, Maryland



This and additional documents:
http://oceancolor.gsfc.nasa.gov/DOCS

A Strategic Plan for the Calibration and Validation of Ocean Color Satellite Data

The NASA Headquarters Ocean Biology and Biogeochemistry Program Manager has established a Calibration and Validation Office at the Goddard Space Flight Center, which is responsible for establishing a long-term capability for calibrating and validating oceanic biogeochemical satellite data. The activity is split into two components of equal stature: *calibration and validation* plus *satellite data processing*. The detailed elements of the activity are based on the tasks of the two main components and the current objectives of the Carbon Cycle and Ecosystems Roadmap. The former is distinguished by an internal core set of responsibilities and the latter is facilitated through an external connecting-core ring of competed or contracted activities.



NASA/SP–2007–214152

NASA Strategic Planning Document: A Comprehensive Plan for the Long-Term Calibration and Validation of Oceanic Biogeochemical Satellite Data

Stanford B. Hooker, Charles R. McClain, and Antonio Mannino
NASA/Goddard Space Flight Center, Greenbelt, Maryland

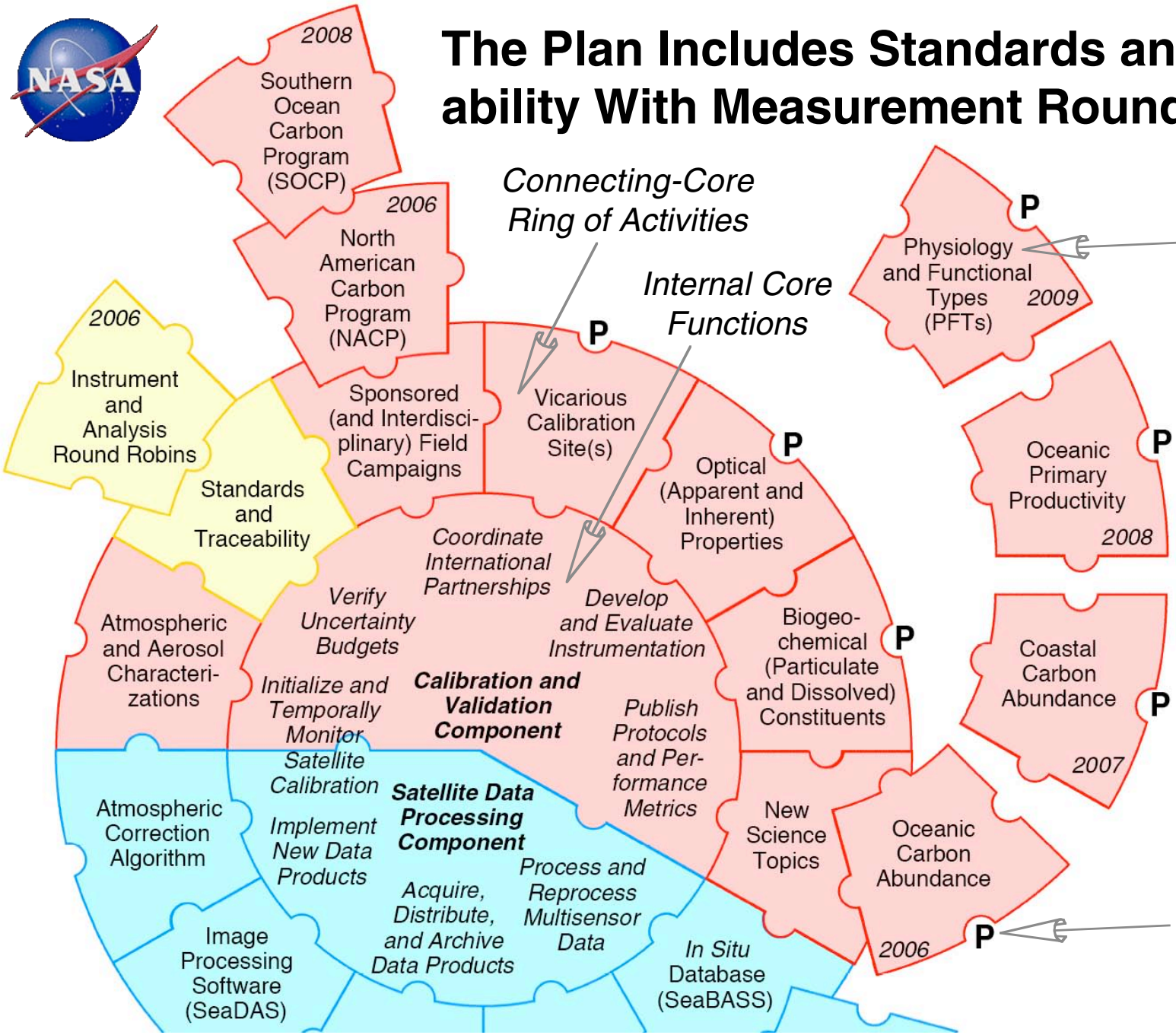
Available from the following URL:
http://oceancolor.gsfc.nasa.gov/DOCS

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

July 2007

The Plan Includes Standards and Traceability With Measurement Round Robins



New science topics are initiated and scheduled by *The Carbon Cycle and Ecosystems Roadmap* plus *The Advanced Science Plan*

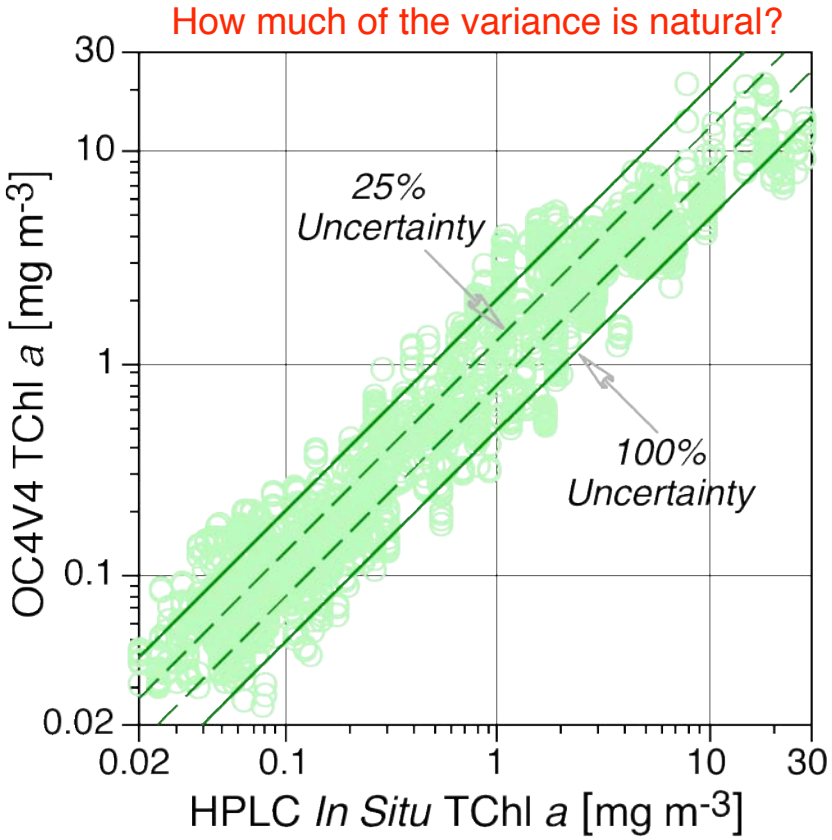
Protocols and performance metrics will be published for all measurements

The Motivation for the SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE) Activity

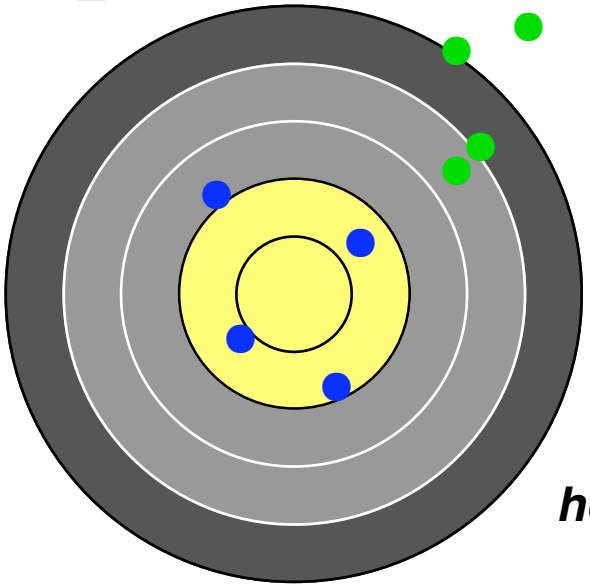
The primary motivation for the SeaHARRE activity was to determine whether or not the sea-truth requirement for ocean color remote sensing was being satisfied.

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Project requires agreement between the in situ and remotely-sensed observations of chlorophyll a concentration to within 35% over the range of 0.05–50.0 mg m^{-3}.

Starting with the SeaWiFS requirement, the SeaHARRE community has adopted (respectively) *validation and refinement accuracies of 15% and 10% for TChl a, and 25% and 15% for the remaining primary pigments* (defined later). A principal difficulty with validation work is estimating the uncertainty in the analysis of the natural samples. Round robins are a useful approach as long as *all of the participants use a validated method, which are equally capable of estimating a true result for each sample, and each sample is analyzed no differently than any other analyzed by the method.*



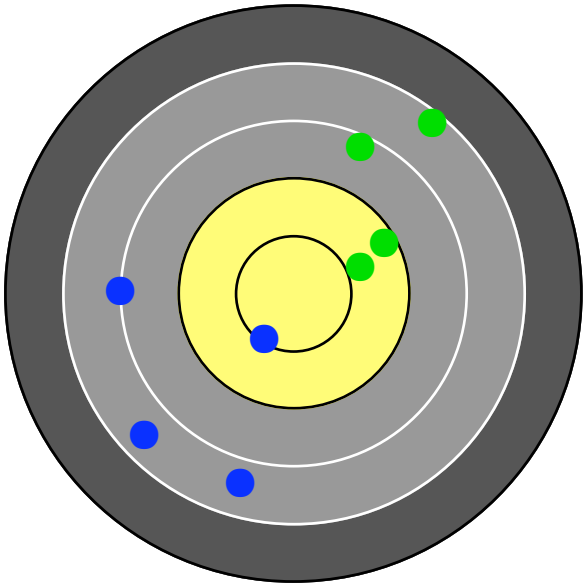
Computing Uncertainties in a Round-Robin Intercomparison Involving Knowns and Unknowns



Consider two groups competing for placement on an archery team. With respect to the *known* bull's-eye (left), the green archers are the least accurate. If a shot in a yellow or light gray circle is needed to make the team, all blue archers, but only one green archer, will qualify. *For the HPLC problem set, this situation is most similar to the analysis of laboratory standards.*

Accuracy is telling a story truthfully, and precision is how similarly the story is repeated over and over again.

Imagine the archers are not evaluated with respect to a known bull's-eye, because one does not exist (e.g., field samples), but are instead evaluated with respect to the average of all the shots (right). In this case, *all the archers qualify for the team*, and the green archers are slightly more accurate as a group than the blue archers. The latter is a recurring consequence of spreading the variance of bad results across all of the outcomes. To minimize this problem, a quality-assured subset is established (the blue archers) as the proxy for truth (the bull's eye).



The same shot pattern as before.

Pigment Products

The intercomparisons are performed on four different types of products: individual pigments, pigment sums, pigment ratios, and pigment indices. The individual pigments are further divided into primary, secondary, and tertiary pigments. *The primary pigments—which all participants must quantitate—are as follows*:

TChl *a*	Total chlorophyll *a*	Chlide *a* + DVChl *a* + Chl *a* (including allomers and epimers)
TChl *b*	Total chlorophyll *b*	DVChl *b* + Chl *b*
TChl *c*	Total chlorophyll *c*	Chl c_1 + Chl c_2 + Chl c_3
Caro	Carotenes	bb-Car + be-Car
Allo	Alloxanthin	
But	19 '-Butanoyloxyfucoxanthin	
Diad	Diadinoxanthin	
Diato	Diatoxanthin	
Fuco	Fucoxanthin	
Hex	19' -Hexanoyloxyfucoxanthin	
Peri	Peridinin	
Zea	Zeaxanthin	

Secondary pigments are used to produce primary pigments, and tertiary pigments are any other pigments for which at least three laboratories quantitated.

Higher-Order Pigment Products

The pigment sums, ratios, and indices are computed (or built up) from the primary pigments and are produced by all the laboratories:

Pigment Sums

PPC	Photoprotective carotenoids	Allo + Diad + Diato + Zea + Caro
PSC	Photosynthetic carotenoids	But + Fuco + Hex + Peri
PSP	Photosynthetic pigments	PSC + TChl *a* + TChl *b* + TChl *c*
TAcc	Total accessory pigments	PPC + PSC + TChl *b* + TChl *c*
TPig	Total pigments	TAcc + TChl *a*
DP	Total diagnostic pigments	PSC + Allo + Zea + TChl *b*

Pigment Ratios

TAcc / TChl *a*	TAcc to TChl *a* ratio	TAcc / TChl *a*
TChl *a* / TPig	TChl *a* to TPig ratio	TChl *a* / TPig
PPC / TPig	PPC to TPig ratio	PPC / TPig
PSC / TPig	PSC to TPig ratio	PSC / TPig
PSP / TPig	PSP to TPig ratio	PSP / TPig

Pigment Indices

mPF	Microplankton proportion factor	(Fuco + Peri) / DP
nPF	Nanoplankton proportion factor	(Hex + But + Allo) / DP
pPF	Picoplankton proportion factor	(Zea + TChl *b*) / DP

A Summary of SeaHARRE Participants

The SeaHARRE activity has emphasized public reporting of all results, international participation (purple), veteran analysts, plus the recruitment of new participants (green) and novice practitioners (yellow). There has also been specialized investigations of damaged samples (DS), reanalyses to better understand uncertainties (RA), the use of two methods for improved evaluation of differences between two methods (TM) using two common columns (C_8 and C_{18}), and comparison of *in situ* filter storage (FS) techniques (histopreps versus foils).

Code	Organization (and Country)	Principal Scientist	SH-1	SH-2	SH-3	SH-4
B	Bedford Institue of Oceanography (Canada)	Venetia Stuart				
C	Common. Scientific and Indust. Res. Org. (Australia)	Lesley Clementson				FS
D	DHI Water and Environment (Denmark)	Louise Schlüter				
F	USF/Florida Institute of Oceanography (USA)	Dave Millie				
G	NASA Goddard Space Flight Center (USA)	Mary Russ				
H	University of Maryland Horn Point Laboratory (USA)	Laurie Van Heukelem		DS		FS
J	Joint Research Centre (Italy)	Jean-François Berthon			RA	
L	Laboratoire d'Océanographie de Villefranche (France)	Hervé Claustre				FS
M	Marine and Coastal Management (South Africa)	Ray Barlow			RA	
N	Dalhousie University (Canada)	Claire Normandeau				
P	Plymouth Marine Laboratory (United Kingdom)	Jim Aiken				
S	San Diego State University/CHORS (USA)	Charles Trees			TM	FS
U	University of South Carolina (USA)	Jay Pinckney				

International participants shown in purple typeface.

SeaHARRE Method Diversity as a Function of Time (and for a TChl *a* range of 0.02 – 42.7 mg m^{-3})

The SeaHARRE field sampling has emphasized a large variety of ecosystems (oligotrophic gyres, wind driven upwelling, etc.), but most of them have been in open ocean (Case-1) waters. In addition, a diversity of methods have been used by SeaHARRE participants, but the majority of them have been based on C_8 columns. Consequently, there has been a recurring emphasis to add new practitioners who are using C_{18} methods, so method diversity can be maintained over time.

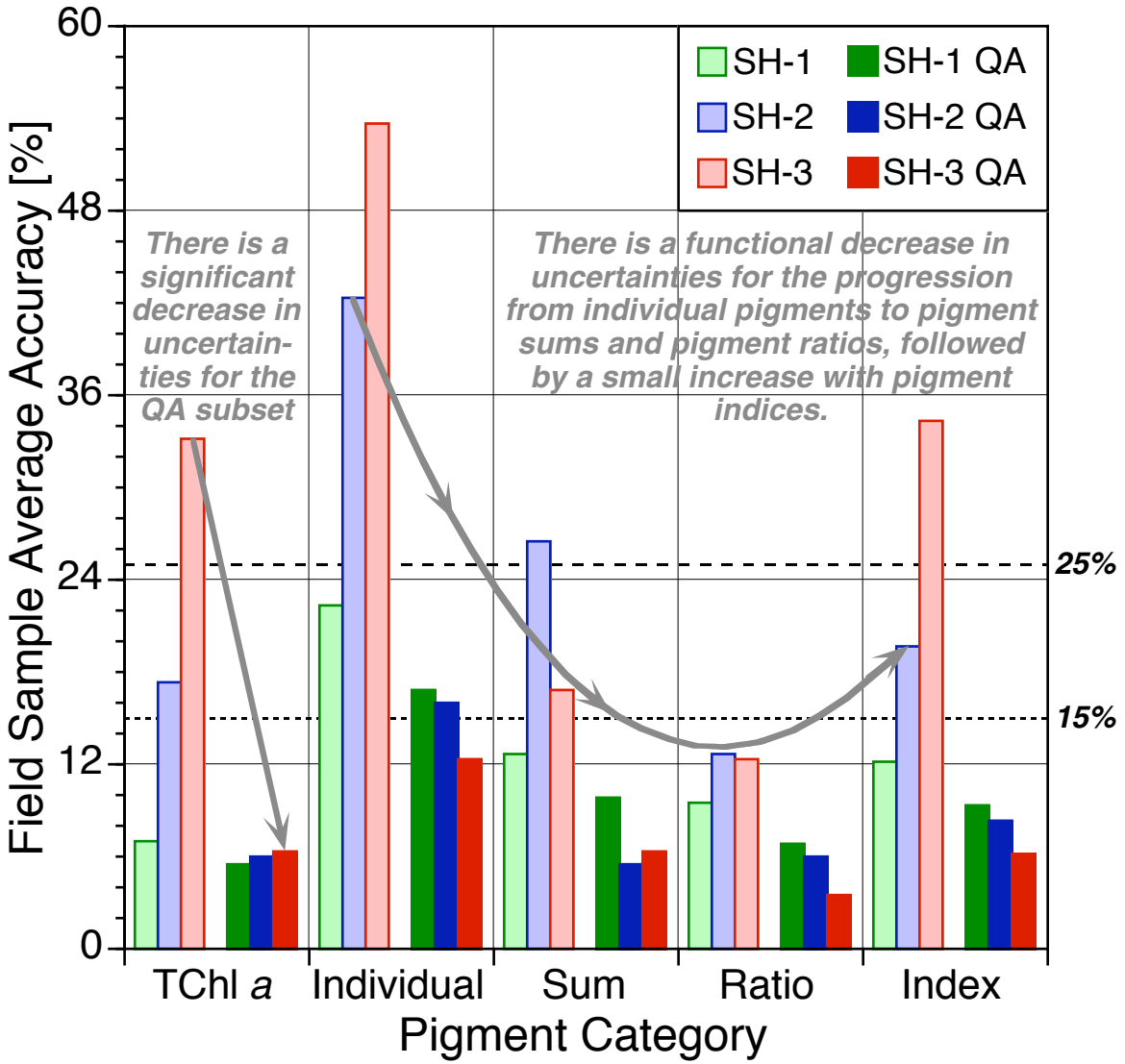
Column Type	Method Citation	Open Ocean (Case-1) SH-1	SH-2	SH-3	Coastal SH-4	Total Labs
C_{18}	Gieskes and Kraay (1989)		B			1
C_{18}	Wright et al. (1991)	J	C,D,S	J,S18	N,S	8
C_{18}	Pinckney et al. (1996)				F,U	2
C_8	Vidussi et al. (1996)	L	L			2
C_8	Barlow et al. (1997)	M	M,P	M		4
C_8	Van Heukelem and Thomas (2001)	H	H	C,D,H,L,S8	C,D,G,H,J,L	13

The laboratory codes for new participants are shown in red.

Despite an explicit effort to maximize method diversity and a strong initial desire to not have the SeaHARRE community produce a unified method, there has been a significant movement by the analysts to adopt one method — *Van Heukelem and Thomas (2001)* — more than any other. Consequently, a majority of the SeaHARRE data comes from the use of a C_8 method. This evolution was the reason for soliciting as many new practitioners using a C_{18} method as possible.

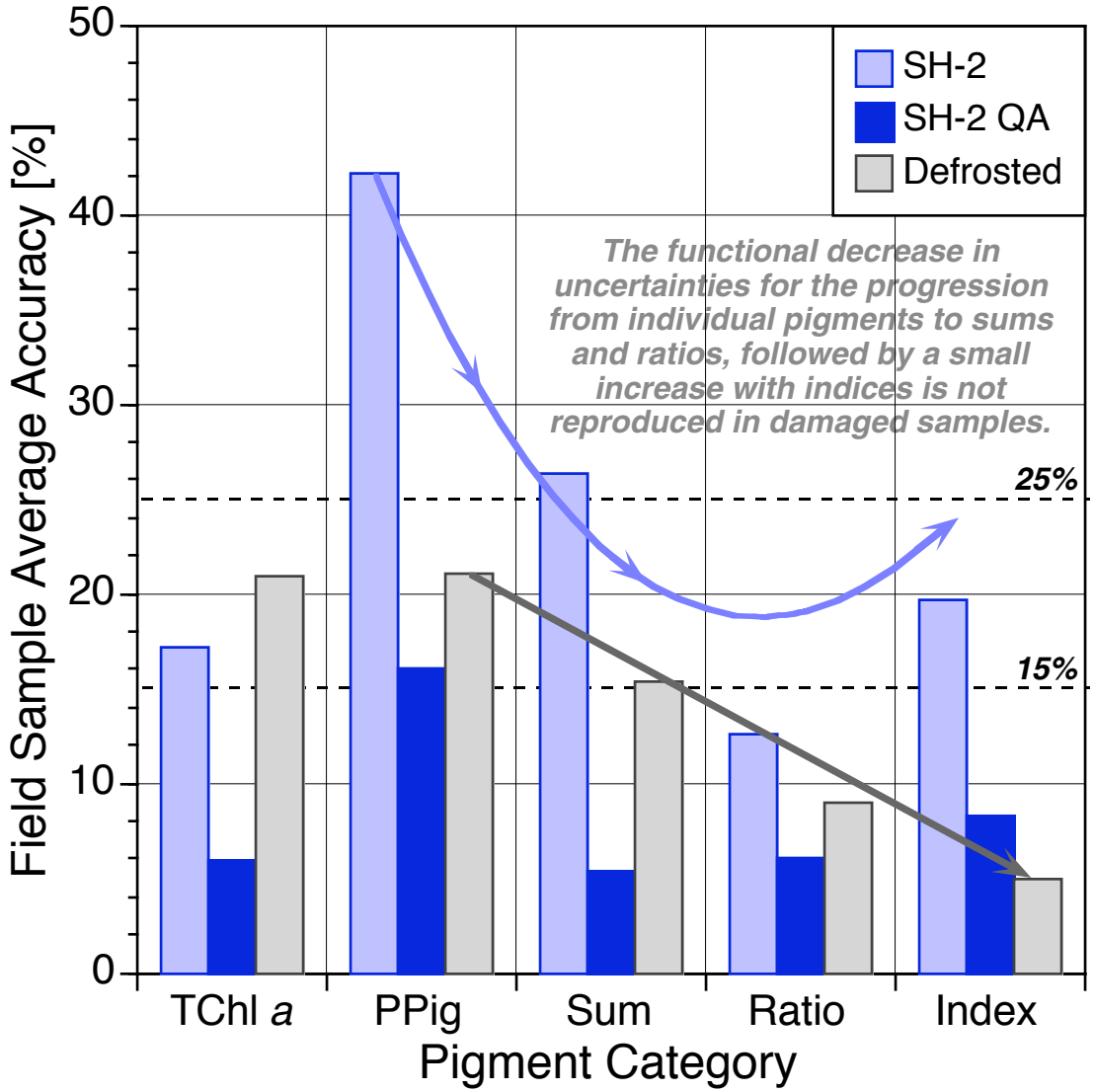
SeaHARRE Summary of Field Sample Accuracy (for HPLC TChl *a* Spanning 0.020 – 26.185 mg m^{-3})

The accuracy of the methods are primarily distinguished by the pigment categories and whether or not the methods were properly quality assured (dark bars) or not (light bars). The QA methods have the lowest uncertainties for TChl a and the individual primary pigments; they always meet the 15% and 25% validation requirement for the former and the latter (and they almost always satisfy the 15% refinement threshold). In addition, there is a functional decrease in the uncertainties for the progression from the primary pigments to the sums and ratios, followed by a small increase with the indices.



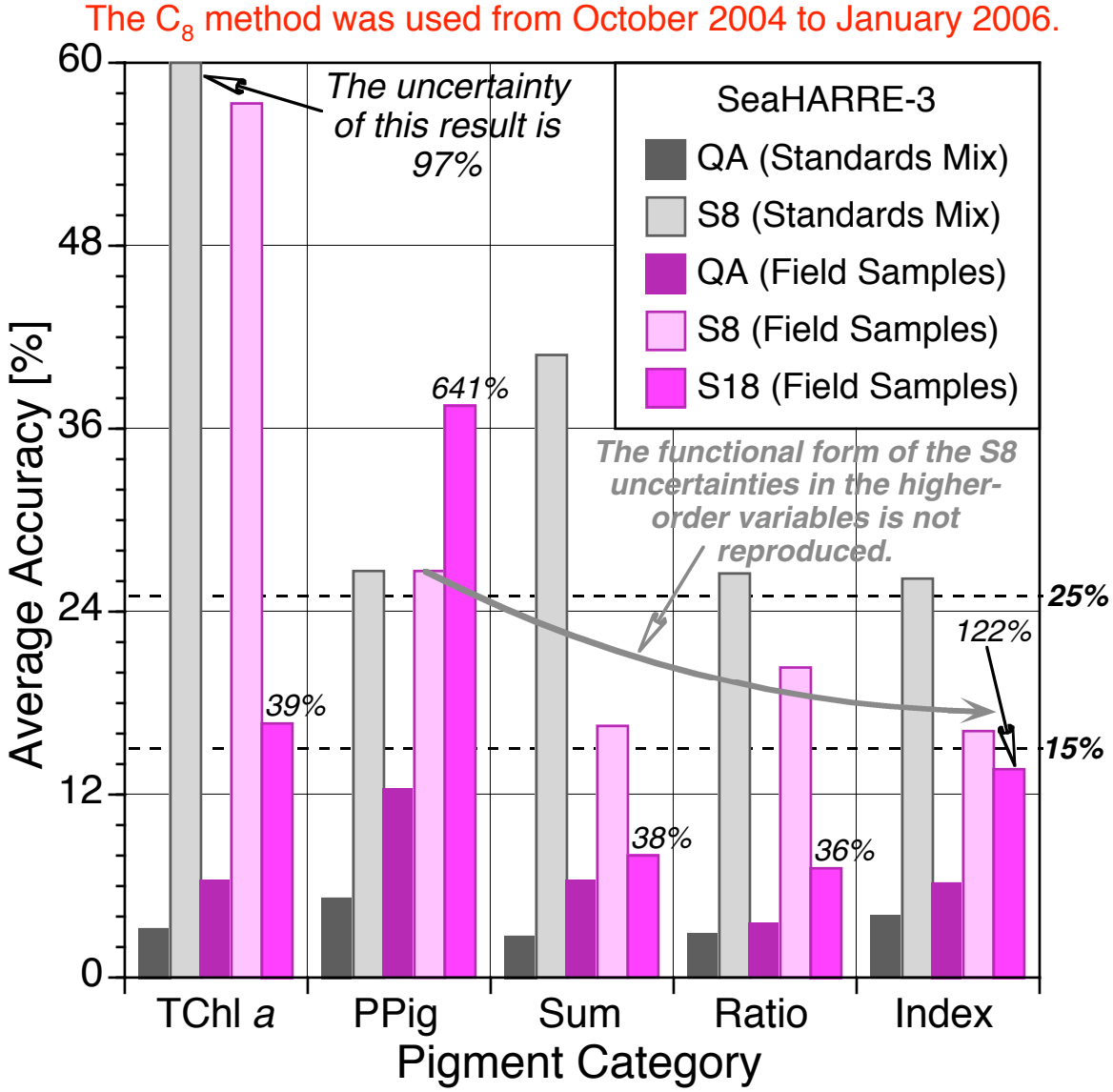
An Important Clue That Something was Wrong from an Unequivocally Damaged Set of Samples

During SeaHARRE-2, one of the QA laboratories analyzed a set of unequivocally damaged samples (they were defrosted during shipping). *The results showed a QA laboratory analyzing bad samples was superior to a method lacking a proper QA scheme and analyzing good samples*. This was confirmed by the precision data as well. An important aspect of the results, however, was that the functional form of the higher-order variables was not properly reproduced in the damaged samples. Consequently, this is now a test to look for aberrant data or analyses.



A Summary of the CHORS C_8 and C_{18} Results in SeaHARRE-3

The SeaHARRE-3 results are divided into methods that were properly validated or *quality assured* (dark bars) and those that were not (light bars). For the latter, the worst-case average result is shown above the bar. CHORS executed two methods based on a C_8 and a C_{18} column, denoted S8 and S18, respectively, and both had significant problems: *the new S8 method had poor TChl a and nearly adequate PPig results, while the old S18 method had the opposite*. Higher-order data products were not as degraded, but the functional form of the S8 uncertainties was aberrant.



The Major Events in Identifying and Understanding the Problems with the CHORS C_8 Data Analyses

A time line of the major events associated with CHORS C_8 problems are as follows:

- The *identification of aberrant data (uncertainties exceeding calibration and validation requirements) occurred in January 2006* as part of SeaHARRE-3.
- CHORS submitted a statistical analysis of the aberrant data, which was peer reviewed, but *the proposed correction scheme was rejected by the OBB Program Manager in February 2007*.
- *An HPLC team of experts (including an external chromatographer from a commercial EPA and FDA laboratory) was established in March 2007* and *an investigative plan was presented to the community in April 2007*.
- *The first team inquiries began in June 2007* at CHORS and determined that *the CHORS calibrations did not satisfy the requisite performance metrics*.
- *A so-called parametric correction scheme was established in July 2007 and approved by the external HPLC expert in August 2007*, but the investigations for *this work exposed an abnormality in the spectral properties of the red and blue wavelengths used to quantitate marine pigments*.
- Because the CHORS C_{18} and C_8 methods relied on basically the same, and in some cases less than optimal, calibration practices, *the investigative team was concerned that the older C_{18} results might be compromised*.

CHORS Pigment Calibration Problems Explained Using A C_8 Example

200507aDHI_V2.1.xls (SeaHARRE-3)

	MVChla 664nm		MVChl a	
	Concen.	Area	ng/inj	% Range
1	12.6	7287	0.92	0.3
2	31.1	16418	2.26	0.8
3	60.8	43318	4.41	1.6
4	89.1	53042	6.46	2.4
5	116.1	74190	8.42	3.1
6	212.8	119653	15.43	5.7
	MVChla 1/RF	631.7		

54% Difference

200412b HPLC Calib.xls

	MVChla 664nm		MVChl a	
	Concen.	Area	ng/inj	% Range
3	3.2	1252	0.23	0.1
4	8.0	5892	0.58	0.2
5	16.0	13177	1.16	0.4
6	23.9	22964	1.74	0.6
7	31.9	25242	2.31	0.9
8	79.8	77089	5.79	2.1
9	159.6	153442	11.57	4.3
10	239.4	249383	17.36	6.4
11	319.3	300164	23.15	8.6
	MVChla 1/RF	973.0		

Calibrations are usually done only after the linear dynamic range of the system has been determined, which in fact never occurred. Typically, a calibration begins close to the limit of detection to the top of the working range of the anticipated analyses (which must be within the linear dynamic range). For CHORS analyses, wherein worldwide samples were anticipated, a MVChl *a* calibration should span 2–270 ng (i.e., about 1–100% of the working range). Frequently, the CHORS calibrations spanned a very small concentration range, which were also too low in magnitude (yellow highlight). In some cases, this problem was exasperated by the arbitrary removal of one or more data points (orange highlight)—ostensibly in an effort to produce more consistent results—which still yielded grossly inadequate inverse response factors (737.7 average value).

The Parameters Describing the CHORS C_8 Quantitation Equation

The CHORS quantitation of pigments involves the use of volumes, peak areas, and inverse response factors (the calibration terms). The experiments conducted at from 22–25 June showed the following for the primary pigments: a) the volumetric terms *not involved with calibration* are probably known to within 1%; b) the peak areas associated with the natural samples and the internal standard were thought to be known to within 4–5% (but this is now known to be an underestimate that is further complicated by the inadequate working range in the calibrations), and c) the calibration process is inadequate for calibration and validation activities and has calibration-to-calibration variability exceeding 25% (presented in more detail later).

Volumes known to within 1%

Areas known to within 4%

Calibration variability can exceed 25%

$$C_{P_i} = \frac{V_m}{V_f}\,\frac{V_c}{V_s}\,\frac{\hat{A}_{c_1}}{\hat{A}_{s_1}}\,\hat{A}_{P_i}\,/\,F_{P_i}$$

P_i is a primary pigment.

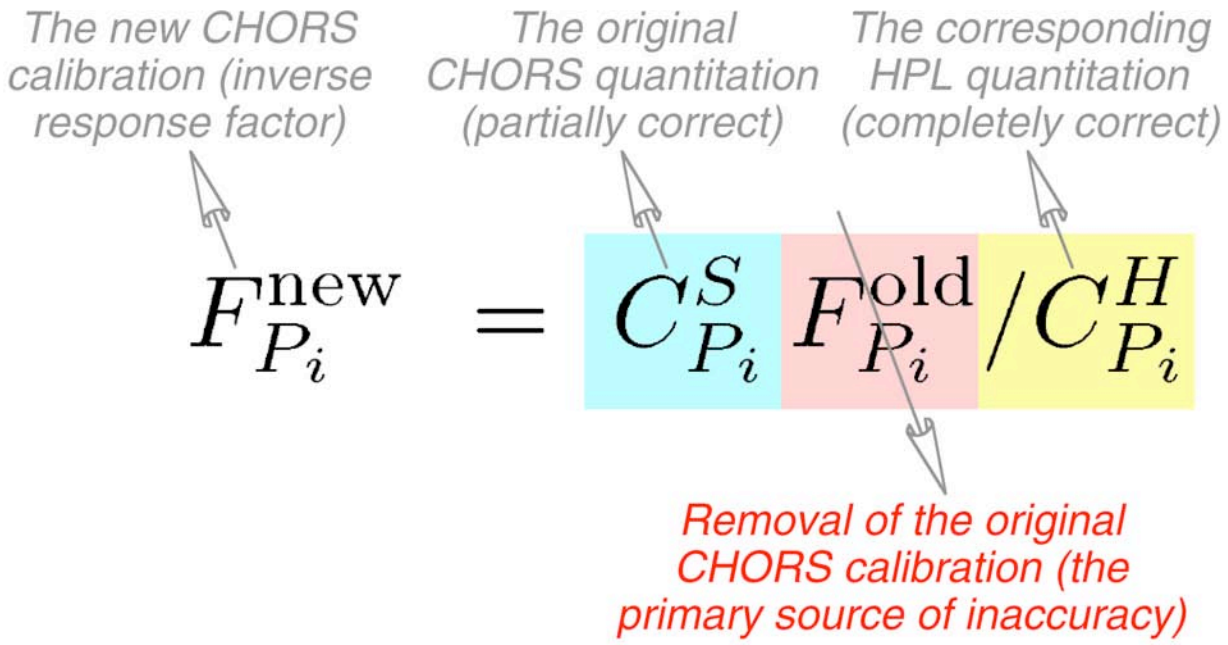
Equals 1 by definition

Determined to within 5%

The C_8 Parametric Correction Scheme

Because of the collection of duplicate seawater samples by A. Mannino and M. Russ (MR) as part of coastal investigations with S. Hooker (MODIS investigator), as well as the participation of CHORS in the SeaHARRE activity (SH-3), three data sets could be assembled wherein CHORS results could be compared to the corresponding analyses by HPL. Using the HPL results as truth (whose accuracies are within the current state of the art), the original CHORS calibration can be replaced with a *parametric* calibration derived from the HPL results. The HPL calibration uncertainty is to within 3% (95% confidence) over an 18-month time period.

The correction process involves matching the quantitated values from CHORS and HPL on a pigment-by-pigment and sample-by-sample basis. In each instance, the original CHORS calibration is removed and a new calibration (inverse response) factor is computed by assuming the HPL data are correct.



The Final C_8 Inverse Response Factors Derived Using the Parametric Correction Scheme

The final inverse RFs were derived using the following steps:

- Individual inverse RFs were computed from the three duplicate data sets for all pigments and all samples for which the concentration of the pigment was at least 0.05 mg m^{-3}.
- An average inverse RF was computed for each pigment from those duplicate data sets having an average concentration of at least 0.05 mg m^{-3} for the pigment.

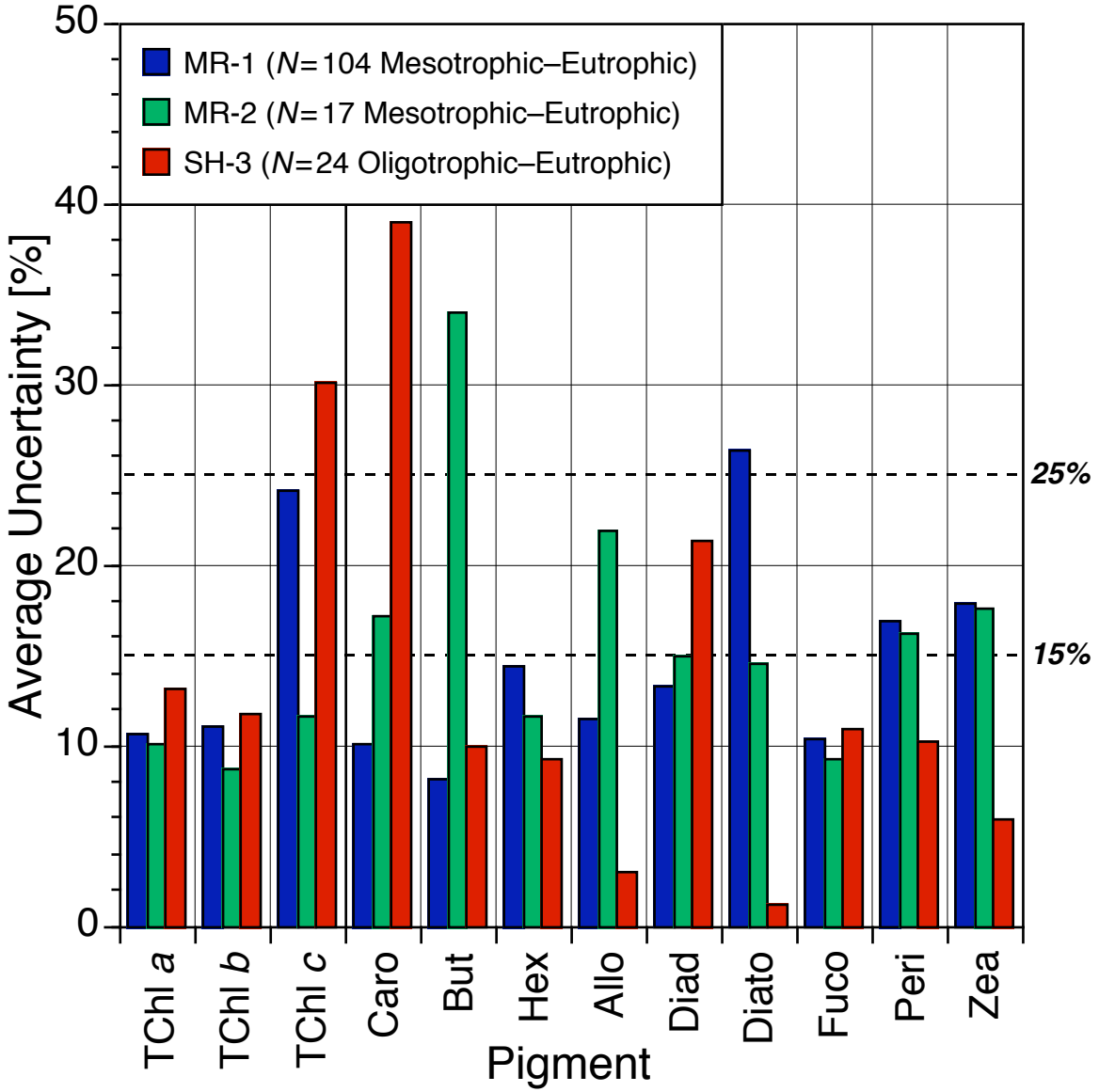
If an inverse RF could not be computed from two or three of the duplicate data sets, because of an excessive amount of data at low concentrations, the inverse RF was also computed by applying the parametric correction scheme to the data from the DHI Mix distributed during SeaHARRE-3. The pigments in the mix were all at concentrations ensuring large, well-formed peaks, but the mix did not contain all the pigments found in a natural sample, because it was formulated using phytoplankton cultures. *The inverse RF was computed by analyzing the potential relative RFs of the pigment from the various available data sets, and selecting the result with the best overall RRF value.* For some pigments (Caro, But, Diato, DVChl *a*, Neo, Phide *a*, Pras, and Viola), the final inverse RF was computed using only the parametric correction of the DHI Mix results .

The Estimated Accuracy of the C_8 Individual Primary Pigments (PPig) After Final Parametric Correction

The final inverse RF values are evaluated by using them to compute new CHORS data products and comparing them to the corresponding HPL data (i.e., the HPL data are the reference data in the uncertainty computations):

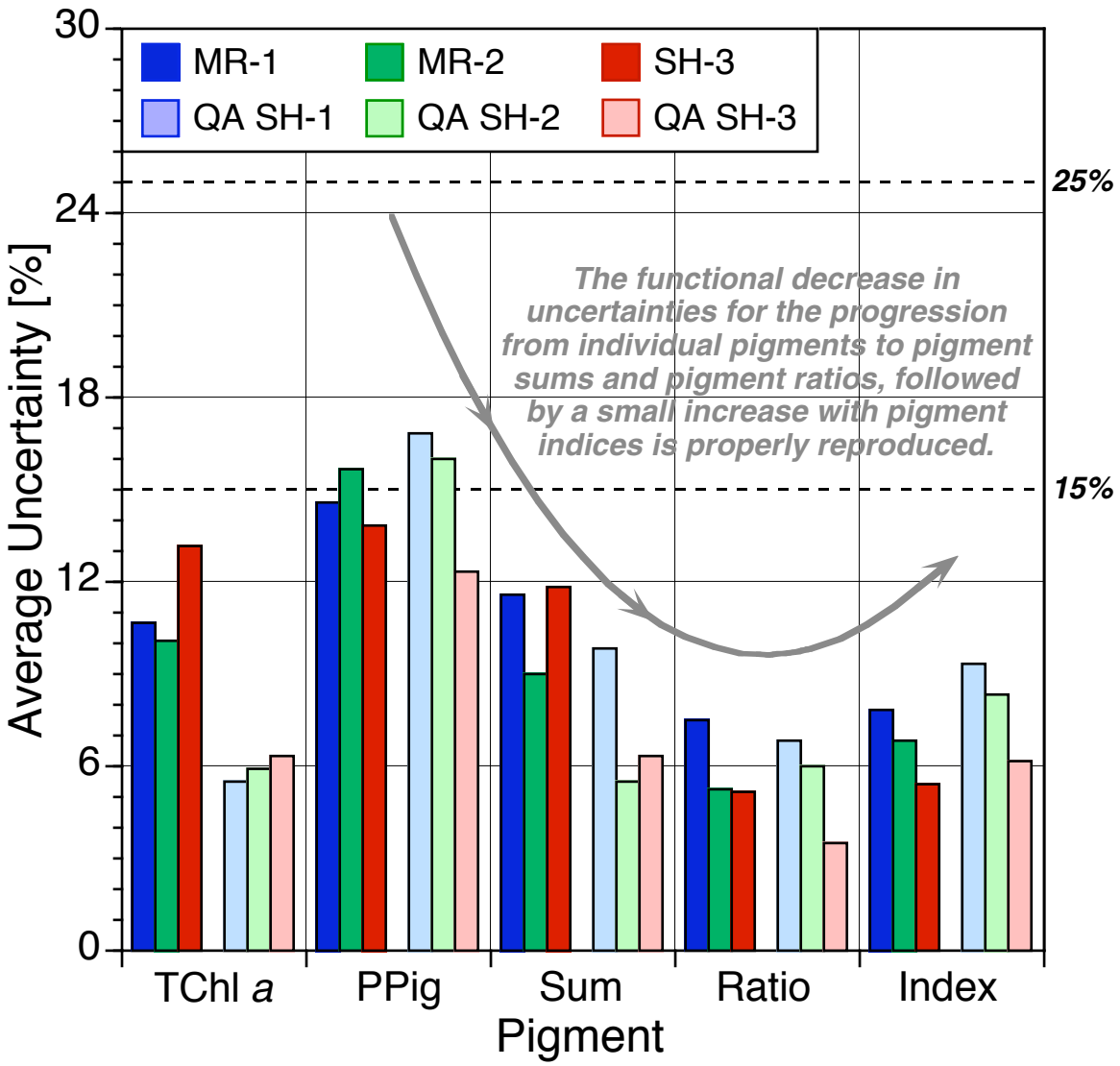
$$\psi_{P_i} = 100 \frac{|C^S_{P_i} - C^H_{P_i}|}{C^H_{P_i}}$$

Absolute values are used to prevent variance cancellation when computing averages (shown in the figure). The calibration and validation objective is an average uncertainty below 25% (less than 15% is desired for algorithm refinement activities).



The Estimated Accuracy of the C_8 Corrected Data Products (For TChl *a* Spanning 0.02–21.06 mg m^{-3})

One of the important tests of the efficacy of the parametric correction is whether or not the functional form of the uncertainties in the higher-order variables is reproduced (recalling that the CHORS SeaHARRE-3 results had aberrant relationships). The plot to the right shows the correct functional form for the corrected data and it also shows the average uncertainties are very nearly the same as the uncertainties obtained for the quality-assured laboratories for all three SeaHARRE activities. Some uncertainties are actually lower, which is an artifact of forcing agreement with HPL (which usually had the best results).



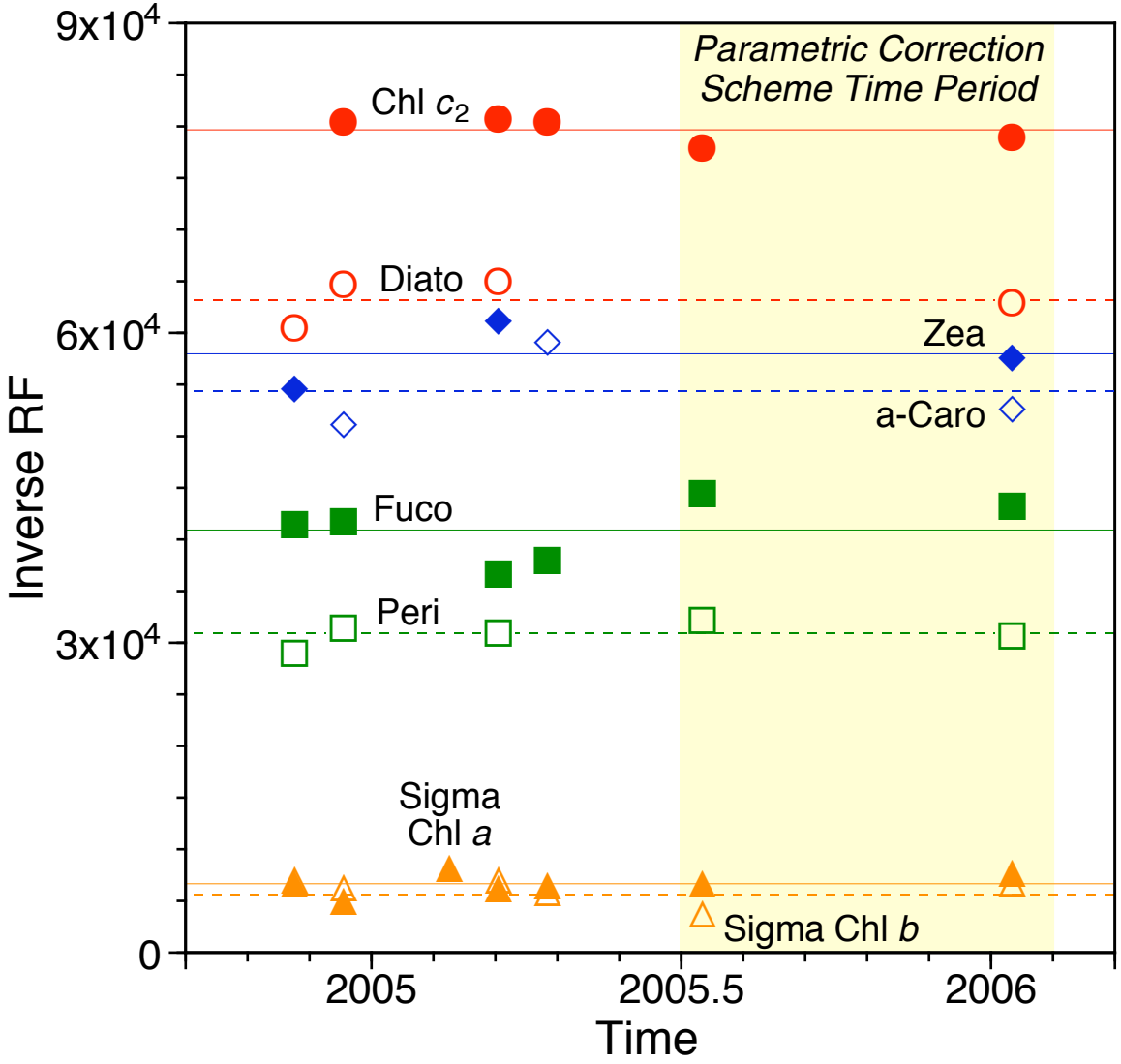
A Summary of Issues Identified During the First Investigation of CHORS C_8 Quantitation Problems

The first investigation of CHORS C_8 quantitation problems established a) the parametric correction scheme and showed it could correct the data to within calibration and validation requirements (always to within an average uncertainty of 25% and with few exceptions to within 15%); b) the CHORS calibration working range was inadequate and contributed significantly to calibration variability (at as much as the 50% level); c) some of the other CHORS calibration procedures were less than optimal (e.g., substandard pipette choices and the use of unscored vial caps) and degraded precision; and d) the inverse response factors for new calibration curves (executed using recommended improvements) agreed to within 1% of the response factors obtained from the parametric correction scheme for Chl *a*.

- *The original and the new calibration curves were not linear and there was a difference between the chlorophylls and the carotenoids. The nonlinearity was also seen in the C_{18} calibrations*.
- The data used in the parametric correction scheme were from July 2005 to January 2006, but the entire CHORS HPLC analysis time period spanned September 2004 to January 2006, so there is a need to demonstrate the system and data quality were unchanging over the whole time period.
- *The red-to-blue detector ratio during calibrations was not constant and was evident for both the C_8 and C_{18} methods*.

The Temporal Stability of the CHORS C_8 HPLC System: Inverse RFs at 10–15 ng inj^{-1}

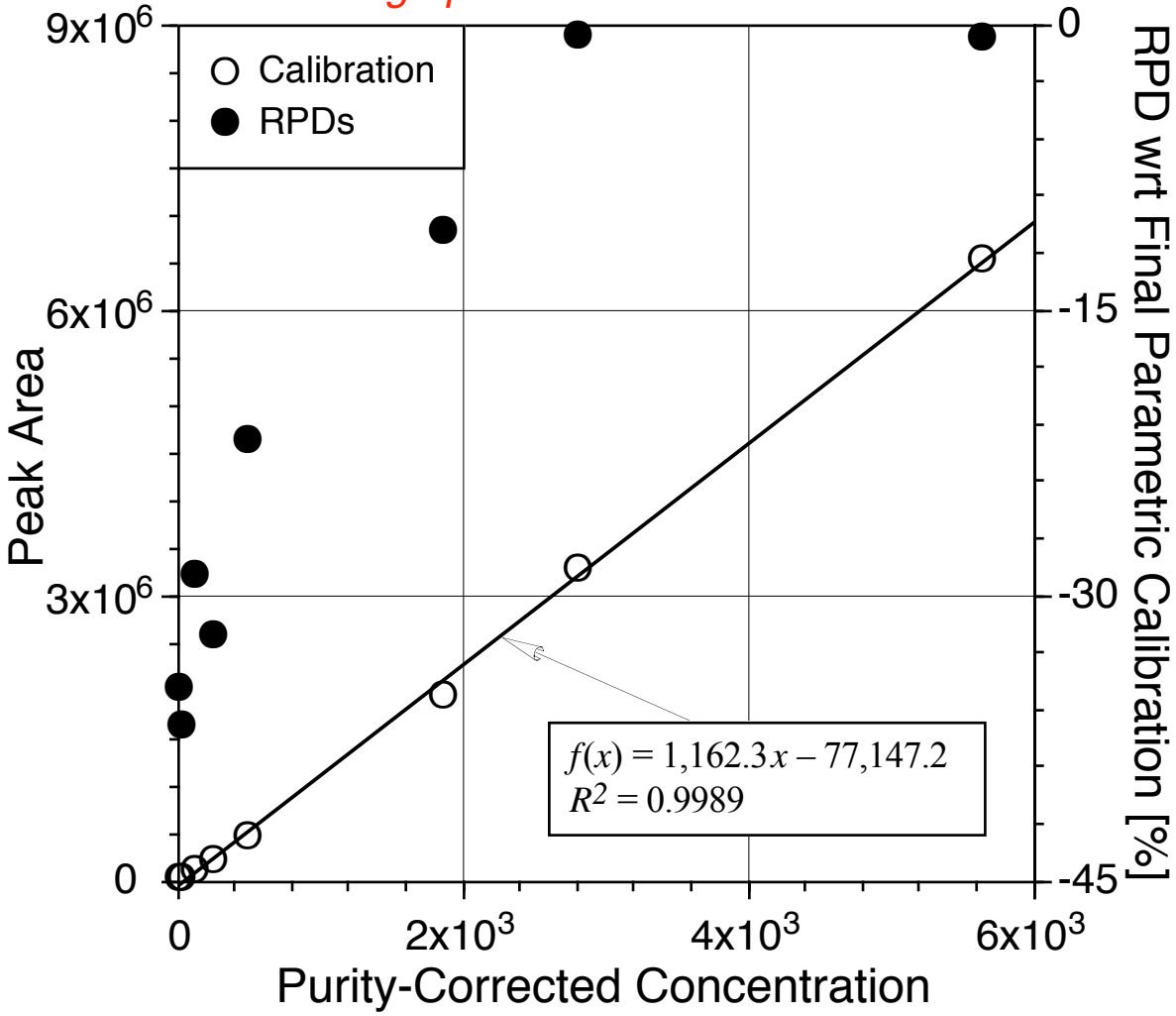
The DHI calibrations provide an extensive set of actual calibrations across the entire C_8 analysis time period. By selecting those data with a 10–15 ng inj^{-1} concentration, the high variability seen in CHORS calibrations can be avoided. These data show an average stability (CV) of approximately 1.5–8.0% across the analysis time period, except for the Chl *a* and Chl *b* calibrations, which have a stability of 16.6–28.5%. This higher variability was also seen in the more frequent Sigma calibrations, which had a stability of 16.3–20.6%. The reasons for this instability are explored later on.



The Effects of Using HPL Calibration Practices on Accuracy

HPL dilution procedures were used at CHORS to prepare eight Chl *a* solutions which were injected on the CHORS HPLC using cap septa that facilitate injection volume accuracy. The correlation coefficient (0.9989) appears good and the slope agrees with the parametric value (1,180.4). A better indicator of accuracy are the percent residuals of the data with respect to the final parametric calibration, which show a nonlinear response of the CHORS HPLC system (also seen in the large nonzero *y*-intercept). *If the CHORS working range is used, the slope is 811.8*.

Correlation coefficients exceeding 0.9999 are achievable for a quality-assured laboratory, as are average percent residuals to within 2%.



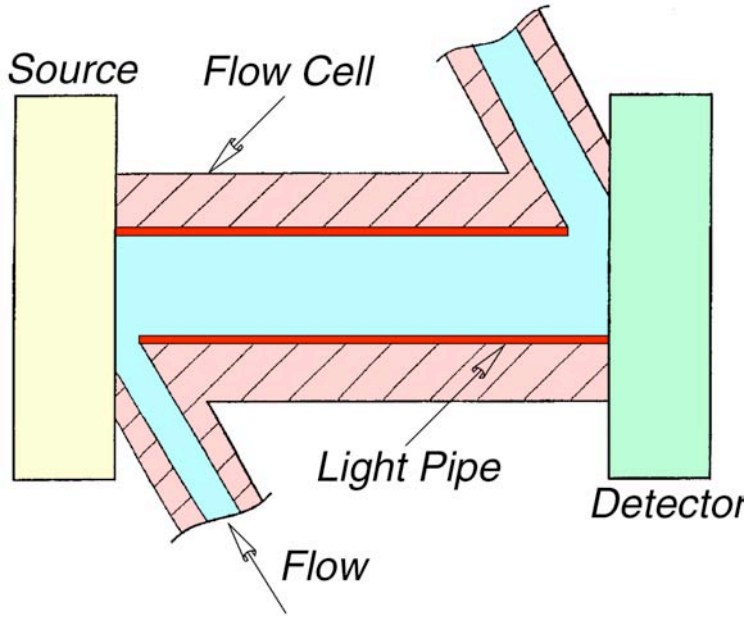
The Major Events in Identifying and Understanding Problems with the CHORS C_8 and C_{18} Data Analyses

The problem with the CHORS pigment analyses represents a case study in all the different ways a problem can go undetected and ultimately have a very significant impact on a program. The panel of HPLC experts reached the following conclusions:

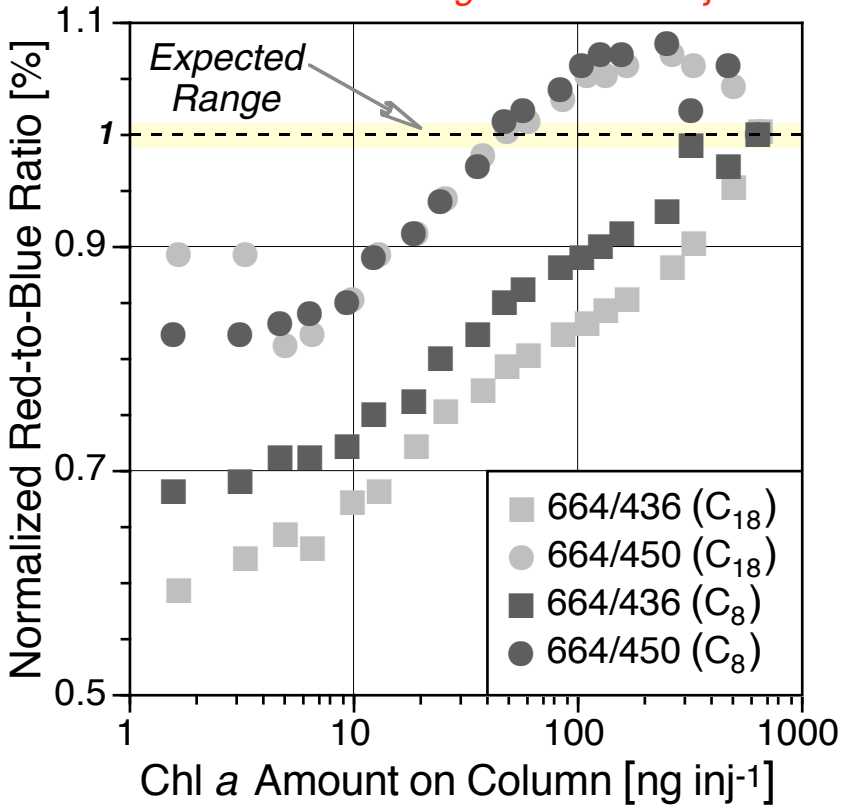
- *There is insufficient duplicate data (actually, almost none) to use the parametric correction scheme for the entire CHORS C_8 and C_{18} data set.*
- *An HPLC detector expert was hired in September 2007* to aid in the investigation of the abnormal spectral properties.
- The detector inquiries revealed a refractive index problem and ultimately a primary source for data degradation: *CHORS was using a flow cell design with a significant stray-light problem at the detector wavelengths being used for marine pigment quantitation, which was creating a nonlinear response when the extraneous light reached the detector*.
- *A series of experiments were conducted in October 2007 to verify the nonlinear response* was in keeping with the description of the expected performance of the flow cell as described in the U.S. and European patents.
- *A final workshop with all the principals, including CHORS, was held in November 2007* to agree on how to characterize the nonlinearity and then how best to implement a correction. *All the participants concurred on how to establish a correction scheme and the method for implementation*.

The Nonlinearity of the CHORS UV6000 Detector Invalidates All C_8 and C_{18} Results

The UV6000 flow cell (US patent 5,608,517) uses a thin polymer (dark red) to pipe light down the flow cell with an optimal response in the ultraviolet domain (190–300 nm):

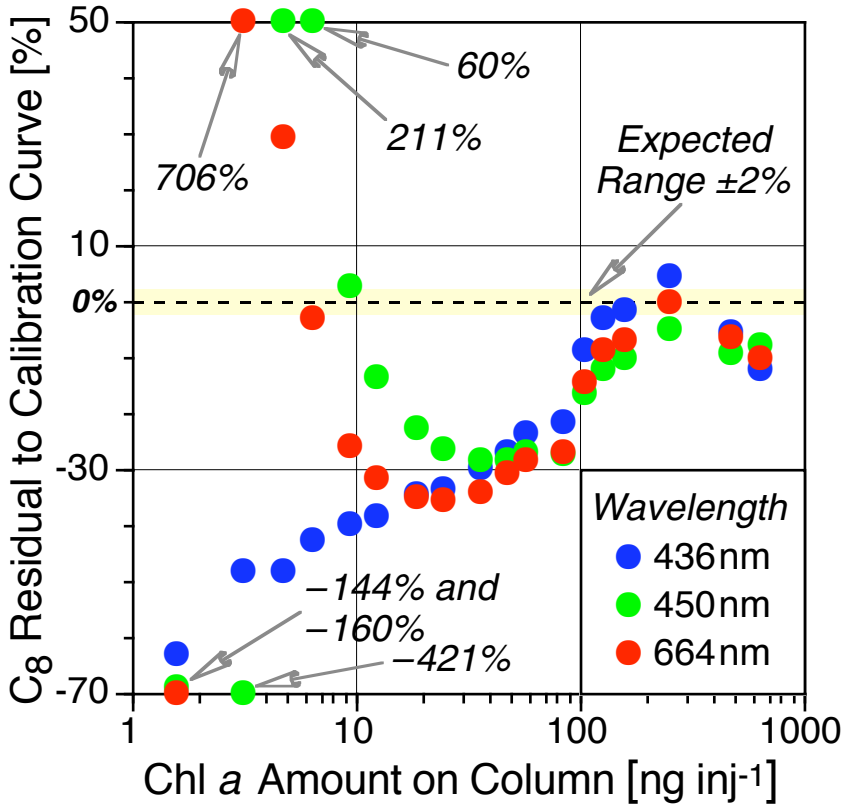


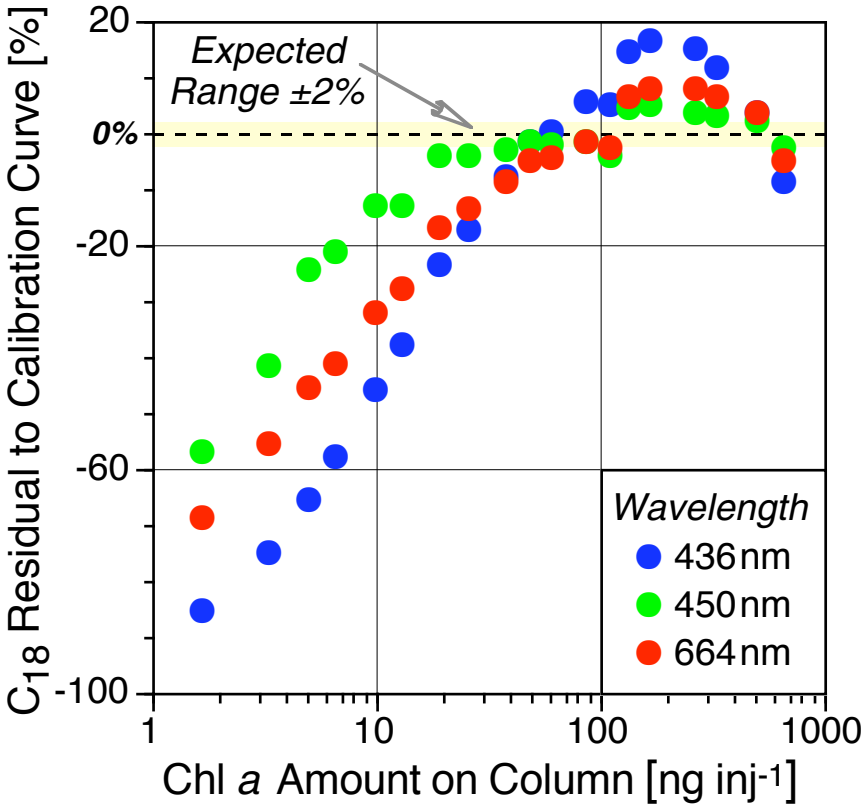
Normalization based on largest amount injected.



Nonlinearity is caused by two problems (US patent 6,281,975B): *a) light can be piped inside the cell wall so it never sees the sample, but is seen at the detector, and b) light is reflected back into the flow path, but still spends some time in the cell wall not interacting with the sample*. European patent 1,478,913C describes stray light issues from reflectance in the cell wall: *the characteristics of the polymer makes the material more opaque at 200 nm than at 600 nm*.

The Nonlinearity of the CHORS Detector Reinvestigated Using New C_8 and C_{18} Calibrations

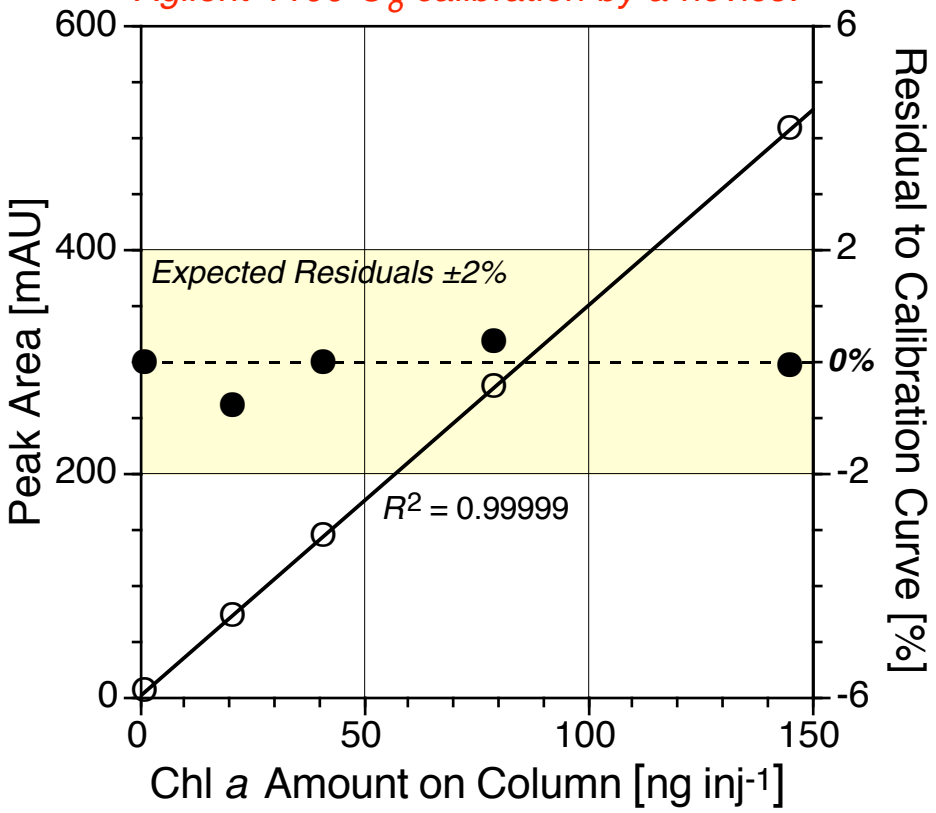


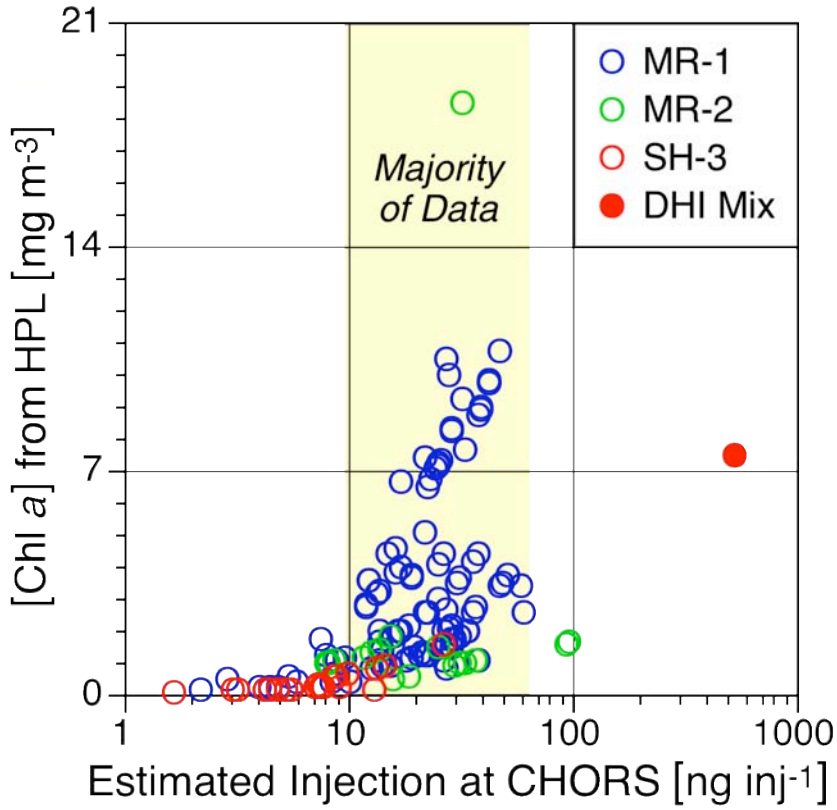


A CHORS calibration is based on immediately forcing through zero, rather than first confirming a negligible *y*-intercept (and average fit residuals to within 2%) before forcing through zero. If new CHORS calibrations are not forced through zero, the residuals show a much stronger nonlinearity. The two calibrations are similar in shape and amplitude until the effects of the large (and negative) *y*-intercept for the C_8 calibration are encountered (where the fit crosses the *x*-axis). .

Reality Checks: Novice Practitioner Results and the Importance of Pigment Loading

Agilent 1100 C_8 calibration by a novice.





C_8 calibration by a novice during SeaHARRE-4 on an Agilent 1100 system shows linear response with low percent residuals. The reason the parametric correction is not so negatively influenced by the nonlinearity is the majority of the duplicate data—*the average response of which establish the correction factors*—correspond to the area of largest change in the expression of the nonlinearity (yellow region).

Limitations of the Correction of the CHORS C_8 and C_{18} HPLC Data

The parametric correction scheme cannot correct all of the CHORS pigments with equal efficacy, because the HPL pigment set is different, and HPL pigment integration procedures were not the same as those used by CHORS—*particularly for pigments at low concentrations or with recurring coelution problems*—which influences both the duplicate data sets used to derive the corrections and the *in situ* samples to be corrected. When considering all the data from a nonlinear characterization and correction perspective, *some pigments will be very hard to correct, because of suboptimal laboratory and quantitation practices* (the marginal pigments are shown in green and the unacceptable pigments shown in red):

- *The uncorrectable pigments* (no overlap between CHORS and HPL): Gyro-diester, α-Caro, β-Caro, and DVChl *b*.
- *The tertiary pigments*: Lut, Neo, Phytin *a*, Phide *a*, Pras, and Viola.
- *The secondary pigments*: MVChl *a*, DVChl *a* (C_8 only), Chlide *a*, MVChl *b*, as well as, Chl c_1 (C_8 only), Chl c_2, and Chl c_3.
- *The primary pigments*: TChl *a*, TChl *b*, TChl *c*, Caro, But, Hex, Allo, Diadino, Diato, Fuco, Peri, and Zea (some of which are derived from the secondary pigments).
- *The higher order products*: pigment sums, ratios, and indices (which are all derived from the primary pigments).

PPig Uncertainties for SeaHARRE-4 (Coastal) Samples: the Quality-Assured (QA) Subset (A′)

Lab.	TChl *a*	TChl *b*	TChl *c*	Caro	But	Hex	Allo	Diad	Diato	Fuco	Peri	Zea	PPig
C	4	16	31	7	79	78	27	16	56	12	53	21	33
D	5	14	30	18	65	46	9	8	31	6	104	17	30
G	8	7	9	4	76	62	14	8	66	3	52	21	28
H	7	7	17	12	145	57	9	7	64	8	68	15	35
L	7	18	15	6	71	45	6	7	32	3	57	20	24
L'	8	23	14	4	93	46	12	12	32	5	57	21	27
A'	7	14	19	9	88	56	13	10	47	6	65	19	29
SH1	7	14	26	18	24	25	39	16	56	9	13	11	21
SH2	6	16	22	17	31	10	20	9	21	5	15	21	16
SH3	6	14	15	13	15	6	4	5	18	11	30	10	12

The SeaHARRE-4 results from QA laboratories show the PPig uncertainties are frequently within the requirements for calibration and validation activities, *15% for TChl a and 25% for all the other pigments* (black), and often meet the refinement objectives (blue). Many results exceed the threshold (red), however, and the PPig average is above 25% for the first time in any SeaHARRE activity (although all laboratories are within ±5.5%). The latter is due to the difficulties in quantitating But, Hex, Diato, and Peri. The problems with these pigments are caused by analyst-to-analyst differences in quantitating peaks in the presence of elevated contributions from degradation products and extraneous compounds not found in Case-1 waters.

PPig Uncertainties for SeaHARRE-4 (Coastal) Samples: Laboratories not in the QA Subset (A^+)

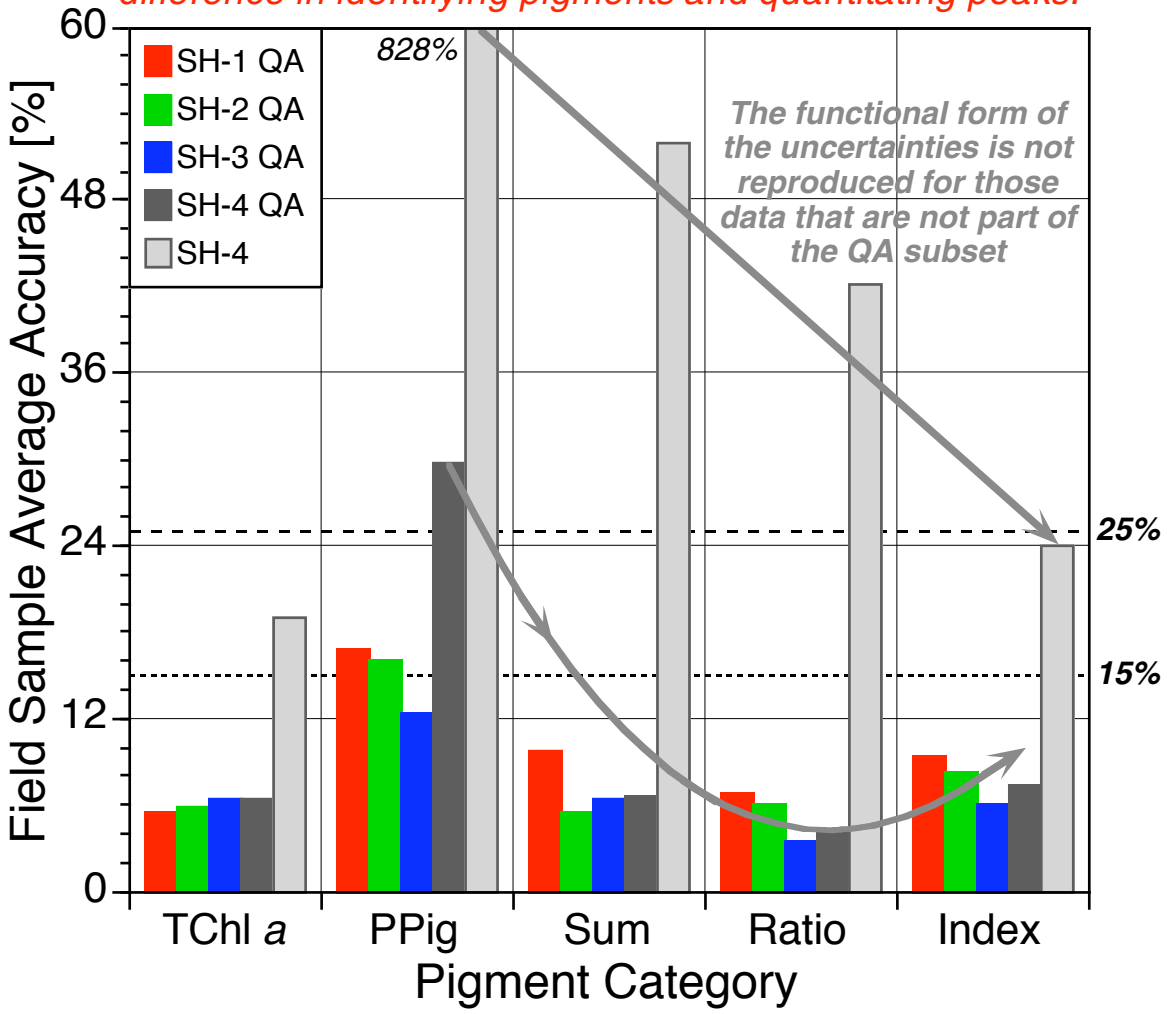
Lab.	TChl *a*	TChl *b*	TChl *c*	Caro	But	Hex	Allo	Diad	Diato	Fuco	Peri	Zea	PPig
F	25	20	55	1318	36471	123	30	18	87	33	93	57	3194
J	23	21	13	10	442	66	32	18	33	17	58	30	64
N	7	26	12	9	653	897	20	21	94	16	95	271	177
S	34	24	32	67	2750	3068	5	14	31	8	284	55	531
U	11	4	1328	325	76	64	37	32	66	27	69	32	173
A^+	20	19	288	346	8078	844	25	21	62	20	120	89	828
SH1	8	17	27	20	28	24	39	24	59	12	33	20	26
SH2	17	21	26	23	64	40	95	31	49	39	57	44	42
SH3	33	37	22	24	113	31	111	22	64	59	112	14	54

The details of the PPig uncertainties for the laboratories not in the QA subset are only occasionally within the requirements for calibration and validation activities (black) and rarely satisfy the refinement objective (blue). *Most of the results exceed the thresholds for calibration and validation (red), and the PPig average is significantly above 25%*. The primary reason for the latter is a large number of false positives and false negatives in pigment identification, with the latter producing the very large uncertainties for TChl *c*, But, and Hex. The CHORS results (S) confirm the extent of the problems identified in SeaHARRE-3, which are further degraded by the aforementioned peak identification problems—*primarily because there are no widely-accepted quantitation rules for pigments in complex chromatograms*.

SeaHARRE-4 Average Accuracy for the Primary Pigments and Higher-Order Variables

The accuracy of the higher-order variables (sums, ratios, and indices) for the QA subset are very similar to the other three round robins, even though accuracy for the PPig pigments does not meet the calibration and validation requirement. The reason for this result is the pigments responsible for degrading accuracy are at rather low concentrations, so they do not influence the higher-order products very much. For the results not in the QA subset, the notable results are the very high uncertainties and the absence of the functional form in the uncertainties.

The chromatographic complexity of the SH-4 coastal samples exposed the importance of analyst-to-analyst difference in identifying pigments and quantitating peaks.



Another CHORS Quantitation Problem Found in Both the C_{18} and C_8 Calibration and Field Data

2007 Calibration		*SIMBIOS-1*		*SIMBIOS-2*		*SeaHARRE-3*		*SeaHARRE-4*	
Sample	MVChl *a*	Sample	But	Sample	Fuco	Sample	Peri	Sample	Zea
Test-025	1232	SB1-T05	0.048	SB2-003	0.006	SH3-9a	0.024	SH4-D08	0.024
Test-026	1229	SB1-T09	0.024	SB2-008	0.006	SH3-9b	0.002	SH4-D23	0.024
Test-027	1785	SB1-T12	0.048	SB2-012	0.012	SH3-9c	0.002	SH4-D35	0.017
Average	1415	Average	0.040	Average	0.008	Average	0.009	Average	0.021
StanDev	320	StanDev	0.014	StanDev	0.003	StanDev	0.012	StanDev	0.004
CoeffVar	22.6	CoeffVar	34.0	CoeffVar	42.9	CoeffVar	133.2	CoeffVar	18.4

These data are all from the C_{18} method, but the same type of problem is found in the C_8 data.

An unexpected aspect of the most recent inquiries into the response of the CHORS system is the occurrence of large outliers during triplicate injections of a calibration standard. These anomalies have been seen in both the most recent C_8 and C_{18} calibrations. The question arises whether or not this was a feature of past inter-comparisons, because if it was, this is an aspect of the HPLC system that cannot be explained by the nonlinear response of the detector, and it would probably be impossible to detect in the analysis of natural samples (which are rarely done in triplicate). For those intercomparisons involving triplicates (SeaHARRE-2 was based on duplicates), the appearance of outliers is always found and it varies in magnitude. Furthermore, it is found in pigments whose detection and quantitation has been found to be similar to Chl *a* (e.g., Fuco).

The Current Status of the Effort to Correct the CHORS C_{18} and C_8 HPLC Analyses

CHORS pigment analyses represent many years of data starting in 1998 when the UV6000 was first purchased and ending with samples run at the end of 2007. The total number of samples involved is approximately 24,000, but not all are for NASA PIs, although some are still of interest to NASA because of the work involved (e.g., MOBY samples). The current correction status is as follows:

- NASA PIs were asked to accept a reduced set of pigments (the primary pigments) to speed up the process and minimize uncertainties.
- Requests for more pigments must be approved by the OBB Program manager.
- The CHORS technician (J. Perl) is contracted to a) deliver all relevant 2001–2007 data and records to NASA, b) collate all pigment calibrations, and c) help characterize the nonlinearity and explore other data problems (integration).
- Four international labs use the Thermo UV6000 and were notified of the nonlinear response.
- All Thermo UV6000 data have been removed from SeaBASS (CHORS, PML, and MCM).

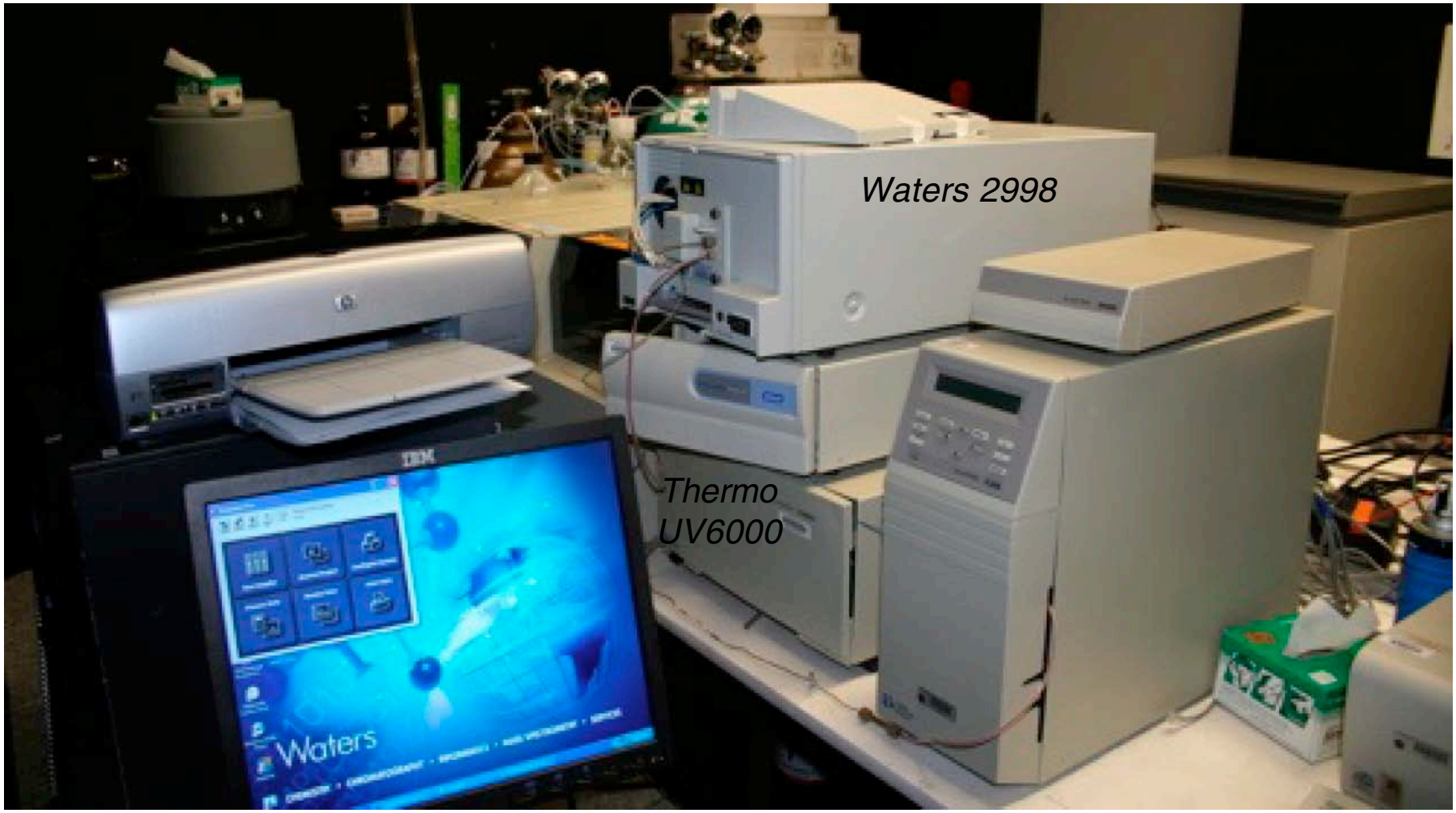
Year(s)	Method	NASA	Others
98-00	C_{18}		2,642
2001	SIMBIOS C_{18}	1,819	
2002	SIMBIOS C_{18}	3,986	
2003	SIMBIOS C_{18}	3,421	
2004	MODIS C_8	2,151	168
2005	MODIS C_8	4,965	792
2006	MODIS C_8	512	
2006	C_{18}		2,347
2007	C_8		667
2007	C_{18}		318
	Total	16,854	6,934

The Current Status of the Effort to Correct the CHORS C_{18} and C_8 HPLC Analyses (*cont*.)

An important aspect of the nonlinearity characterization problem is to *develop a methodology that other groups can implement without the need for hazardous or expensive compounds*. The reason this is important is *three other international laboratories that perform HPLC analyses for marine pigments with a UV6000 detector system have already been identified*. The specific tasks that have been undertaken to characterize the nonlinearity are as follows:

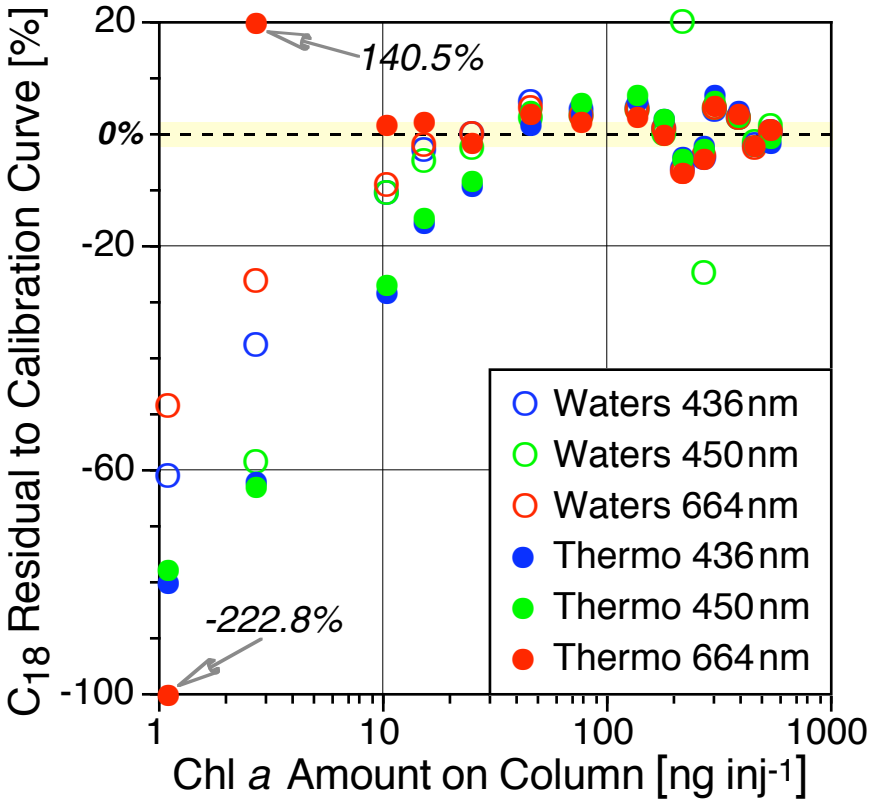
- Evaluate whether or not easily obtained commercial food dyes are useful compounds in establishing the nonlinear response function.
- Determine if the use of a second detector with an *unequivocally linear response* can be used in line with a UV6000 to characterize the nonlinearity.
- Compare the CHORS existing calibrations to the functional form of the nonlinearity to see if they have enough degrees of freedom to be individually fitted to the nonlinear function. The worry here is many of the CHORS calibrations have a very limited dynamic range.
- Establish surrogate QA parameters to determine the stability of the CHORS HPLC system across time periods when the CHORS calibrations cannot be used for nonlinear fitting (CHORS collected no QA or QC data).
- Work with the international labs to have an independent evaluation capability and to correct those data if necessary (only MCM has cooperated fully).

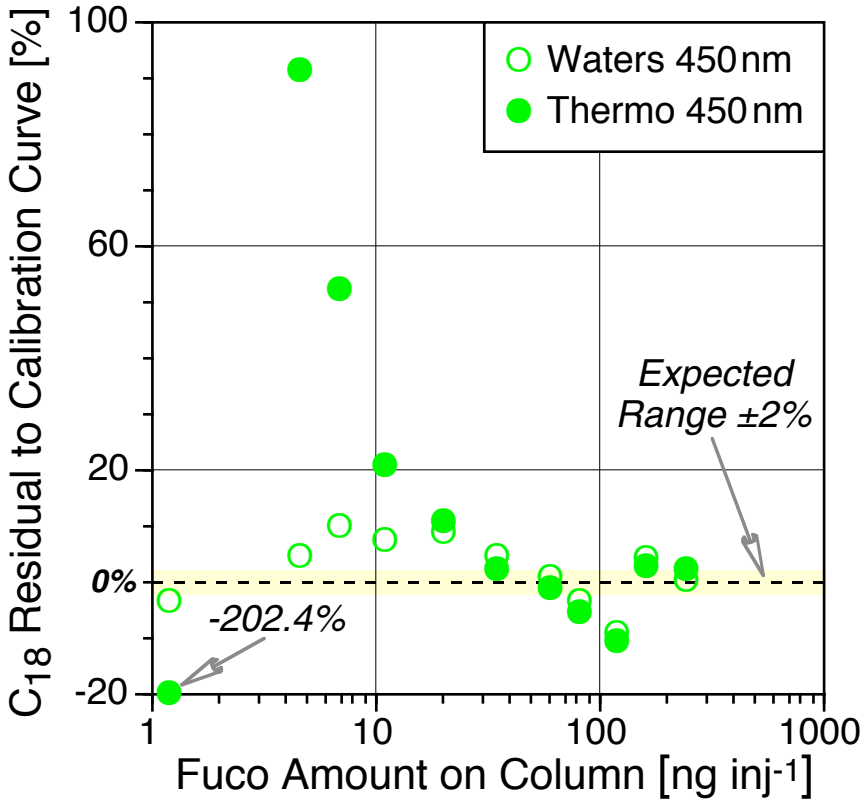
In-Line Use of a Waters 2998 Detector to Characterize the Nonlinearity of the Thermo UV6000



A Waters 2998 detector, which was just characterized by the manufacturer right before it was used, had a stated linearity of 2–5% residuals (the same specification is used in SeaHARRE, although 2% is typical) and was placed in-line in front of the Thermo UV6000 detector.

Characterization of the Nonlinear Response of the CHORS Detector Using a C_{18} Calibration



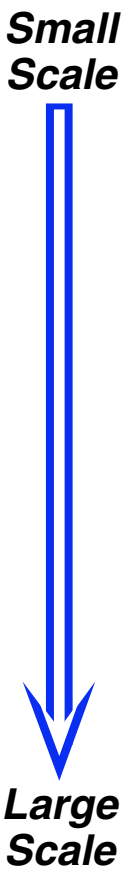


The left and right panels show the percent residuals for the calibration of Chl *a* and Fuco, respectively, using two in-line detectors: a new Waters 2998 and the CHORS Thermo UV6000 detector. Although both calibrations are similar, the Thermo detector is distinguished by larger residuals and extreme excursions from the basic pattern. *The fact that the Waters calibrations are not linear indicates the Thermo autosampler is also introducing a nonlinearity to the calibration process*.

The CHORS Problem Chronicles a Failure from the Smallest (Laboratory) to the Largest (Agency) Scale

The CHORS HPLC problem represents a case study in how undetected low- and high-level mistakes—*some made with the best of intentions*—can have a significant and negative impact on the quality of an entire program. The HPLC experts tasked with correcting the CHORS HPLC problem stressed the following conclusions:

- *Every protocol step must be strictly followed* to minimize uncertainties.
- If failures in laboratory procedures are to be detected, *the personnel need to be trained in good analytical chemistry practices*.
- A quality assurance plan—with well thought out QA and QC data—must be implemented by *personnel who are trained chromatographers*.
- Problems are inevitable, and early detection requires *an emphasis on the importance of accuracy by the cognizant project personnel*.
- Problems are more readily exposed if *the personnel involved are active participants in round robins and workshops*.
- *The advice of HPLC professionals must not be discounted or ignored,* ;particularly when dealing with the early detection of a problem.
- Proposed *solutions to a problem must be evaluated by scientists with a good understanding of method validation*.
- NASA *proposals must be reviewed by properly qualified panelists*.
- Oversight by *NASA should not rely too heavily on peer reviews*.



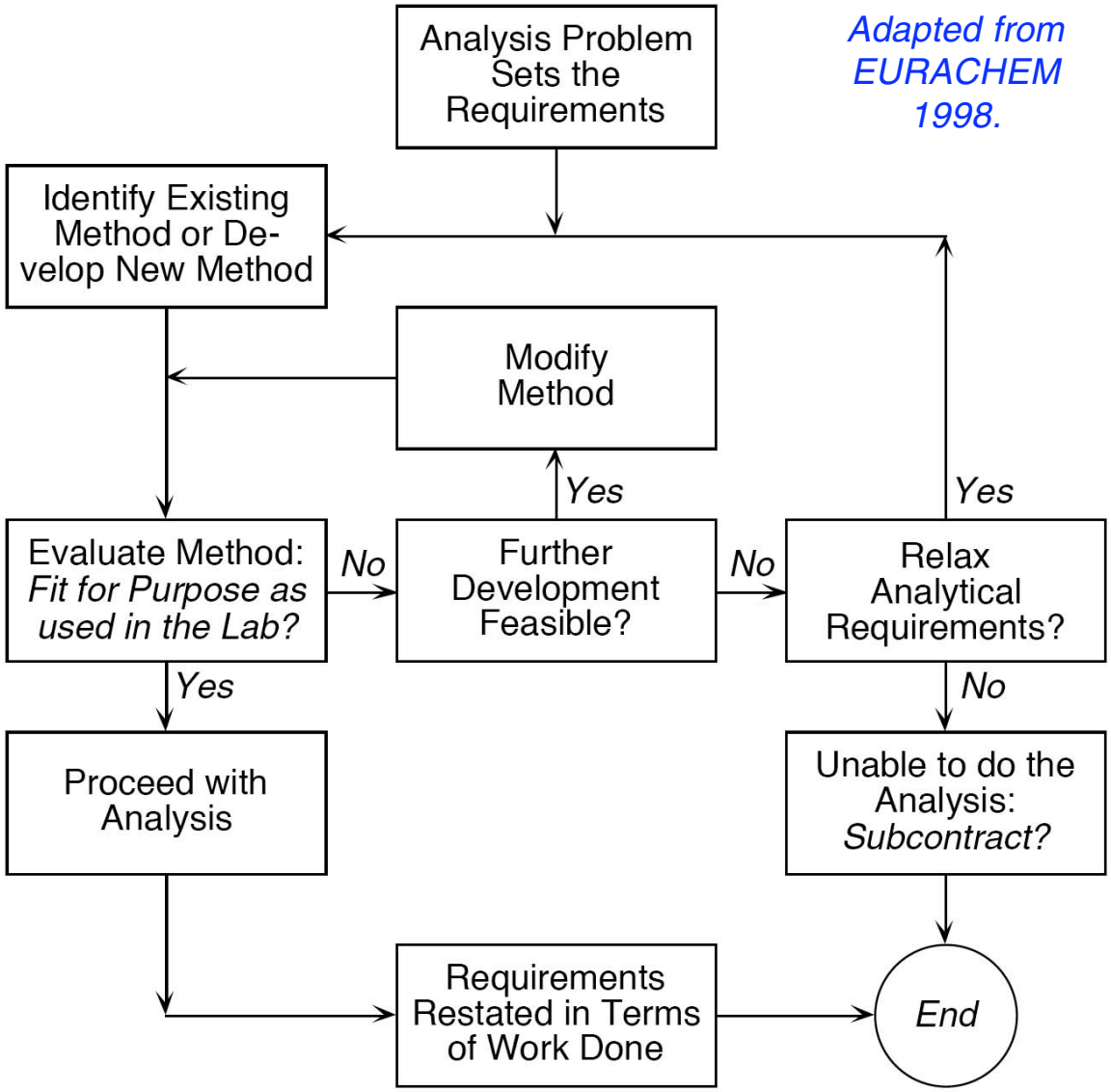
Establish and Improve Method Performance with a Quality Assurance Plan (QAP)

A QAP ensures an analytical method fulfills the specified performance metrics at all points during the analysis of samples. The components of a QAP are as follows:

- Method Validation
  - Defines the analytical requirement.
  - Determines if the method is suitable for the intended application.
  - Performed before method is used for samples.
- Standardized Procedures
  - Reduce uncertainties.
  - Minimize blunders.
- Quality Control (QC) Measurements
  - Implemented temporally.
  - Designed to describe uncertainty in calculation variables.
- Quality Assessment
  - Uses QC results to quantify expected limits of performance.
  - Round robins needed to evaluate accuracy with field samples.

An Evaluation of the Suitability of the Methods Being Used for CDR Analyses Needs to be Performed

An important lesson learned from the CHORS problem is the need to evaluate method suitability for CDR analyses. *It is common for methods to be used without proper evaluation, so there is a strong likelihood another parameter will suffer the same fate*. Method validation includes the evaluation of performance metrics, and fitness of purpose is determined by how the method performs when used by the analyst with the available equipment and facilities, and *not just by performance data collected by other practitioners*.



Adapted from EURACHEM 1998.

Method Validation Processes

The analytical requirement describes the compounds to be analyzed and identifies the accuracy objectives. The latter requires a set of performance parameters, which typically includes the following:

- Specificity;
- Limits of detection and quantitation;
- Working and linear ranges;
- Calibration;
- Accuracy and precision; and
- Ruggedness.

Quantitative assessment of performance parameters is performed during method validation. Later, during routine analysis of samples, many performance parameters can continue to be quantitatively evaluated and, after a sufficient number of observations, 95 and 99% confidence limits can be assigned for that method. For example, a laboratory may, during method validation, observe residuals for a calibration curve that average 3% and upon further re-calibrations, it may be observed that the residuals routinely average 2%, with an expectation that 95% of all residuals within the working range are within 5%. These assessments help individual laboratories determine if their methods are performing within expectations, but they also can be developed into a *performance metric* for assessing the potential for a method producing accuracy within specified limits.

HPLC Performance Metrics: An Example QA Capability for a CDR Analysis

The culmination of the SeaHARRE inquiries into using QA procedures to minimize uncertainties is a proposed set of performance metrics applicable to any HPLC method. The four different categories are arbitrary, and are used simply to provide a range of capabilities. Each category is assigned a weight and score, so the ultimate performance is based on summing the weights for each parameter, dividing by the number of parameters, and comparing the result to the category scores.

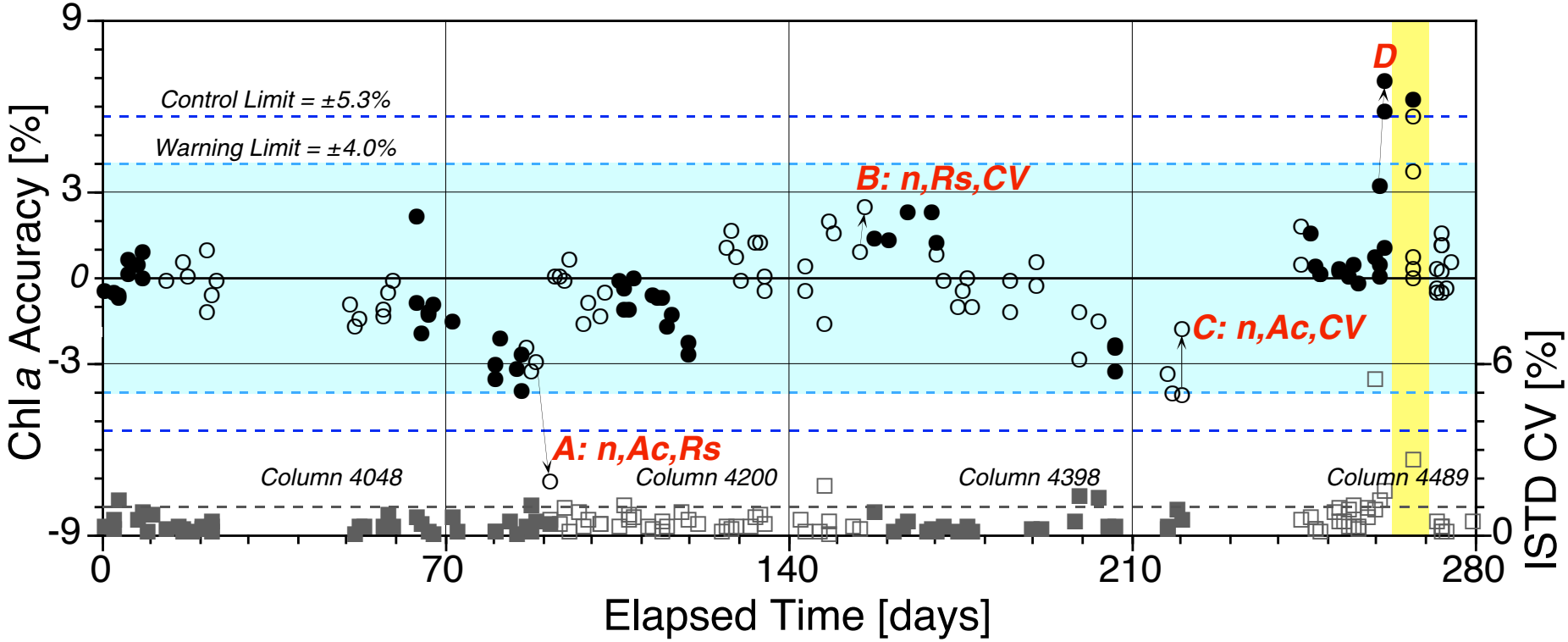
Performance Weight, Category, and Score		*TChl a*		*PPig*		*Separation*†		*Injection*‡ ($\bar{\xi}_{\text{inj}}$)		*Calibration*§	
		$\bar{\xi}$	$\lvert\bar{\psi}\rvert$	$\bar{\xi}$	$\lvert\bar{\psi}\rvert$	$\check{R}_s$	$\bar{\xi}_{t_R}$	*Perid*	*Chl a*	$\lvert\bar{\psi}\rvert_{\text{res}}$	$\bar{\xi}_{\text{cal}}$
1. Routine	0.5	8%	25%	13%	40%	0.8	0.18%	10%	6%	5%	2.5%
2. Semiquantitative	1.5	5	15	8	25	1.0	0.11	6	4	3	1.5
3. Quantitative	2.5	3	10	5	15	1.2	0.07	4	2	2	0.9
4. State-of-the-Art	3.5	$\leq$2	$\leq$5	$\leq$3	$\leq$10	$\geq$1.5	$\leq$0.04	$\leq$2	$\leq$1	$\leq$1	$\leq$0.5
Method H		*1*	*5*	*2*	*12*	*1.2*	*0.02*	*<1*	*<1*	*1.1*	*0.4*

† The $\check{R}_s$ parameter is the minimum resolution determined from a critical pair for which one of the pigments is a primary pigment. The retention time CV values presented here are based on sequential replicate injections of pigments identified in Mix C. In the absence of a diverse set of early- through late-eluting pigments, like Mix C, a practical alternative is to compute $\bar{\xi}_{t_R}$ based on Perid, Fuco, Diadino, Chl *a*, and $\beta\beta$-Car based on three sequential injections.

‡ The $\bar{\xi}_{\text{inj}}$ terms are calculated from the average of replicate injections of an early- and late-eluting pigment in the same run (Perid is chosen here to incorporate the possible effects of peak assymetry which is not presented as a separate parameter).

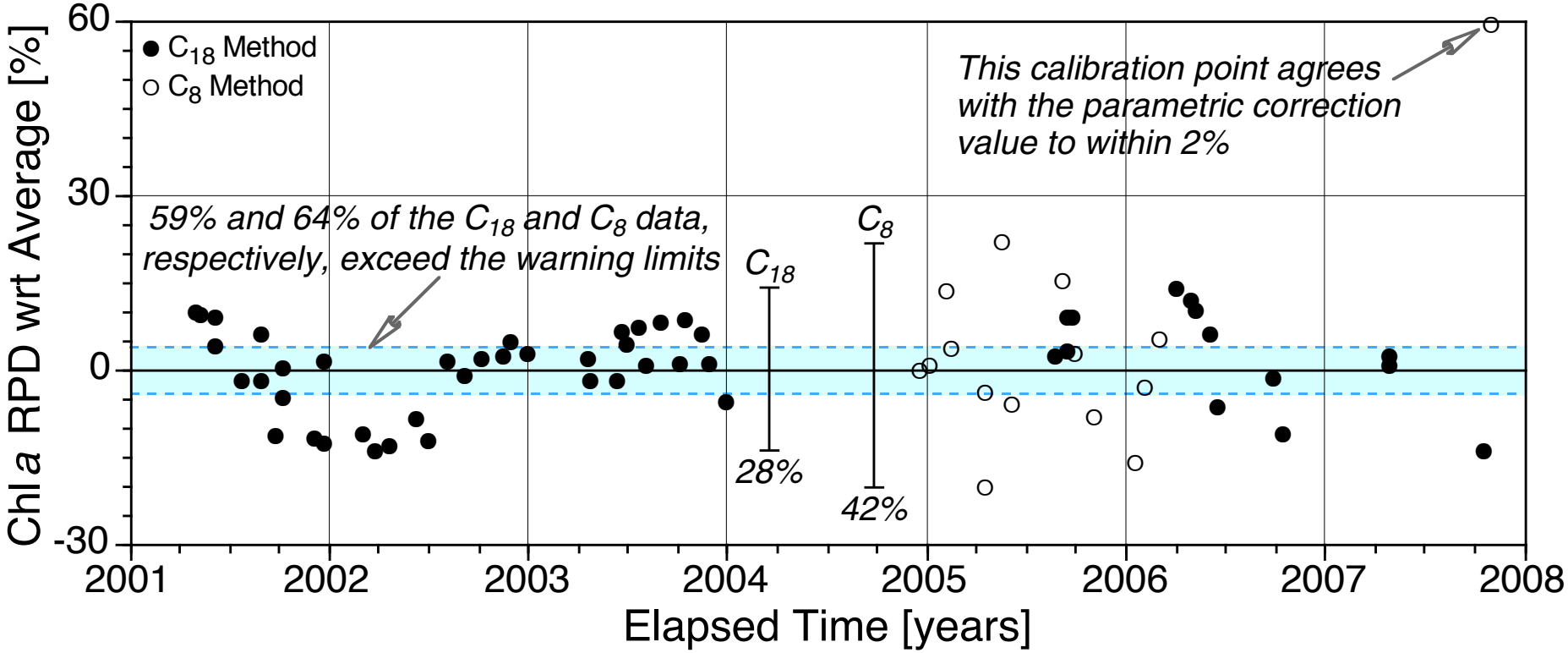
§ The $\lvert\bar{\psi}\rvert_{\text{res}}$ values presented here are based on calibration points within the range of concentrations typical of the SeaHARRE-2 field samples. To determine this metric for an arbitrary sample set, $\lvert\bar{\psi}\rvert_{\text{res}}$ is computed using those calibration points within the range of concentrations expected in the field samples to be analyzed (Sect. 1.5.5.5).

HPL Quality Assessment: QC Data Temporally Monitored on a Daily Basis



A, *B*, and *C* denote when columns were removed from service based on multiple criteria: *n* = aged column, *Ac* = poor Chl *a* QC accuracy, *Rs* = poor resolution between critical pairs (less than 1.0), and *CV* = poor ISTD precision (control limit exceeded). *D* indicates a hardware failure occurred, with a very high internal standard (ISTD) CV value, whereupon sample processing ceased during the days highlighted in yellow while repairs and testing were conducted. After the latter, sample processing continued with no change in the column.

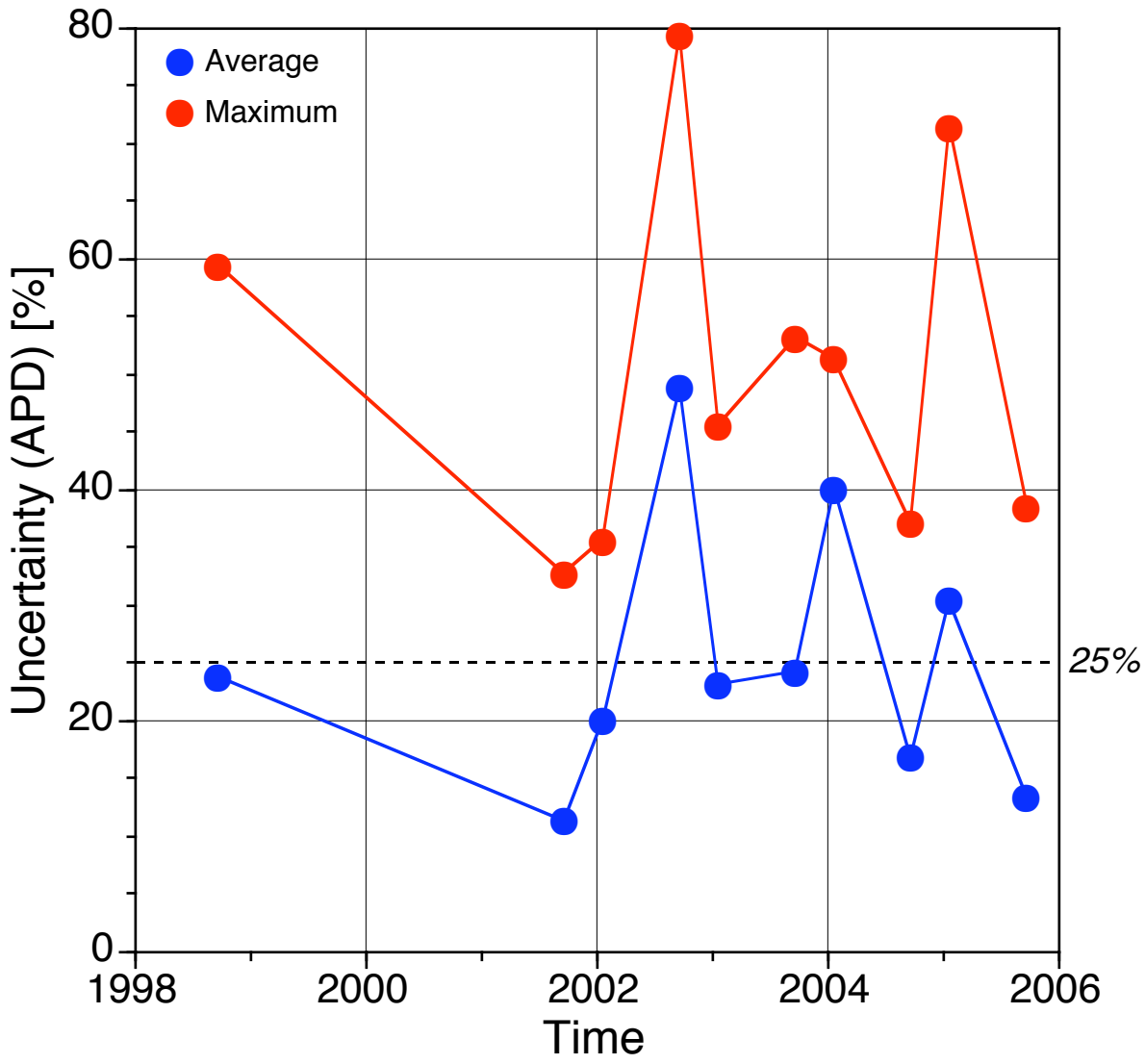
CHORS Quality Assessment: Temporally Monitored QC Proxy Data Recreated from Monthly Calibrations



The absence of QC data in the CHORS HPLC analyses means it is not possible to reconstruct a proper control chart for the analysis time period. The most temporally extensive available data are the Chl *a* calibrations, but these data are, of course, suspect, so some limitations are unavoidable. Nonetheless, if the average of the calibrations are used as a reference, the relative percent difference (RPD) of the actual values with respect to the average provides a proxy QC data set.

Accuracy Assessment of Field Samples Requires Carefully Executed Round Robins

Chesapeake Biological Laboratory conducts anonymous fluorometer round robins with water quality laboratories of the Chesapeake Bay. The mean and median chlorophyll *a* concentrations are determined and the distribution about the mean of individual results are computed. No effort is made to estimate the data quality associated with each participant's results and no inquiries into the sources of uncertainty are conducted. Despite many years of effort, no recurring improvement in overall quality has been achieved as has been seen in the SeaHARRE activity.



There is a Significant Source of Uncertainty for all HPLC Data—*Including Quality-Assured Data*

As pointed out in Jeffrey et al. (1997) and confirmed by all SeaHARRE activities, the variability in absorption coefficients for HPLC calibrations is an unwanted — *and unnecessary* — source of variance in pigment quantitations:

- In terms of the difference between the maximum and minimum values being used during the SeaHARRE-2 activity, *the average percent difference in absorption coefficients was about 10.8%* (the range was 0.0–58.8%).
- *Only 23% of the absorption coefficients matched the solvent used by the method, both in terms of type and concentration*. For the mismatched subset, 61% involved differing solvents (e.g., methanol versus acetone) and 16% involved differing concentrations (e.g., 90% versus 100% acetone).
- None of the chlorophyll pigments—including Chl *a*—are at the highest level of confidence, and some have no experimentally determined values.
- *Only two primary pigments*—Fuco and Diato—*have high-quality absorption coefficients* (purity checked with NMR) and both are for 100% acetone.
- Most of the absorption coefficients were determined in the 1960s and 1970s and predate the availability of NMR for purity checking.
- *The oldest absorption coefficient routinely in use is the 1938 Lutein value*.

Many absorption coefficients being used by HPLC analysts today are not the most recent, or the most reliable, or even supported by the peer-reviewed literature. Although citations exist, the endpoint is not always a laboratory experiment.

Establishing a Small Team of Scientists To Propose a Plan for Solving the Problem

Although getting the entire marine community to use the same set of absorption coefficients would remove the unwanted variance, it would not necessarily result in more accurate quantitations, because many pigments do not have experimentally determined absorption coefficients, or if they do, much of the work is dated and not of the quality required for calibration and validation activities.

To investigate how best to proceed in solving this problem, a small team of scientists agreed to meet and begin planning a solution process:

- Stanford Hooker (NASA/Goddard Space Flight Center, USA),
- Laurie Van Heukelem (Horn Point Laboratory, USA),
- Crystal Thomas (Horn Point Laboratory, USA),
- Louise Schlüter (DHI Water and Environment, Denmark),
- Lane Sander (National Institute of Standards and Technology, USA),
- Mary Russ (UMBC/Goddard Space Flight Center, USA), and
- Einar Egeland (Bodø University College, Norway).

Absorption Coefficients for HPLC Analyses: Chl *a* Pigments

A literature search of the experimental determinations of absorption coefficients applicable to HPLC analyses shows important omissions and data quality problems:

Rank	Pigment	100% Acetone	Date	Confi-dence	100% Ethanol	Date	Confi-dence	100% Methanol	Date	Confi-dence	90% Acetone	Date	Confi-dence
1	Chl a	84	1940	4	**83.4a**	**1965**	**3**	74.5	1941	3	90.8	1960	2
1	Chl a	92.6	1960	2				74.5	1965	2	89	1963	2
1	Chl a	91.5	1965	2				75	1968	2	89.7	1965	2
1	Chl a	91.8	1965	2				76.1	1972	2	87.67	1975	2
1	Chl a	88.15	1975	3				77.9	1978	3			
1	Chl a							75.95	1989	2			
2	DVChl a												
3	Chlide a												

A primary marine pigment or part of one

* = Important to human nutrition

54.36 in Rowan's book, 51.36 in original article

Purity of pigment probably checked by NMR

Purity of pigment not proven by NMR

Confidence Categories:

1 = Very confident (after 1985)
2 = Confident
3 = Somewhat confident
4 = Not confident
5 = Unknown confidence

a = 96% ethanol
b = 650 nm
c = 642 nm
d = 665 nm
e = 652 nm

Bold Black Calibration and validation quality

Red Typeface At least 25 years old.

Bold Red No absorption coefficient or at least 40 years old.

Bold Red More than 50 years old

Absorption Coefficients for HPLC Analyses: The Chlorophylls and Most Important Carotenoids

The two pigments with the most complete investigations are Chl *a* and Chl *b*, although the divinyl forms and degradation products are incomplete. Several carotenoids have no citable determinations of absorption coefficients:

Rank	Pigment	100% Acetone	Date	Confi-dence	100% Ethanol	Date	Confi-dence	100% Methanol	Date	Confi-dence	90% Acetone	Date	Confi-dence
4	Chl b	51.8	1940	3	45.9	1963	2	36.4b	1941	3	52.5	1960	2
4	Chl b	53.5	1960	2	**44.2b**	**1965**	**3**	44.5c	1978	4	54	1963	2
4	Chl b	**54.3**	**1965**	**2**				22.26d	1989	2	53.1	1965	2
4	Chl b							42.48e	1989	2	51.36	1975	3
5	DVChl b												
6	Fuco	166	1992	1	**114**	**1965**	**3**						
7	Allo												
8	Peri	134	1968	2	132.5	1968	2						
9	Zea *	**234**	**1966**	**2**	254	1938	3						
9	Zea *				248	1978	2						
10	Hex												
11	But												
12	Chl c2	37.2	1972	2							40.4	1972	2
13	Chl c3												
14	Chl c1	39.2	1972	2							44.8	1972	2
15	Diadino	223	1974	2	250	1977	2	225	1974	2			
16	Diato	272	1994	1									
17	b,b-Car *	250	1969	2	**262**	**1956**	**3**						
18	b,e-Car *	270	1969	2									

Absorption Coefficients for HPLC Analyses: Less Important (Marine) Pigments

Rank	Pigment	100% Acetone	Date	Confi-dence	100% Ethanol	Date	Confi-dence	100% Methanol	Date	Confi-dence	90% Acetone	Date	Confi-dence
19	Prasino												
20	Lutein *				255	1938	5						
21	trans-Neo				247	1965	3						
21	trans-Neo				224	1966	3						
21	trans-Neo				227	1972	3						
21	trans-Neo				247	1978	3						
21	**trans-Neo**				238	1992	1						
21	**cis-Neo**				233	1994	1						
22	Viola	244	1966	2	255	1957	3						
23	Phytin a												
24	Phide a												
25	b-Crypto *												
26	Lycopene *	345	1966	2									

A primary marine pigment or part of one

* = Important to human nutrition

54.36 in Rowan's book, 51.36 in original article

Purity of pigment probably checked by NMR

Purity of pigment not proven by NMR

Confidence Categories:

1 = Very confident (after 1985)

2 = Confident

3 = Somewhat confident

4 = Not confident

5 = Unknown confidence

a = 96% ethanol

b = 650 nm

c = 642 nm

d = 665 nm

e = 652 nm

Bold Black Calibration and validation quality

Red Typeface At least 25 years old.

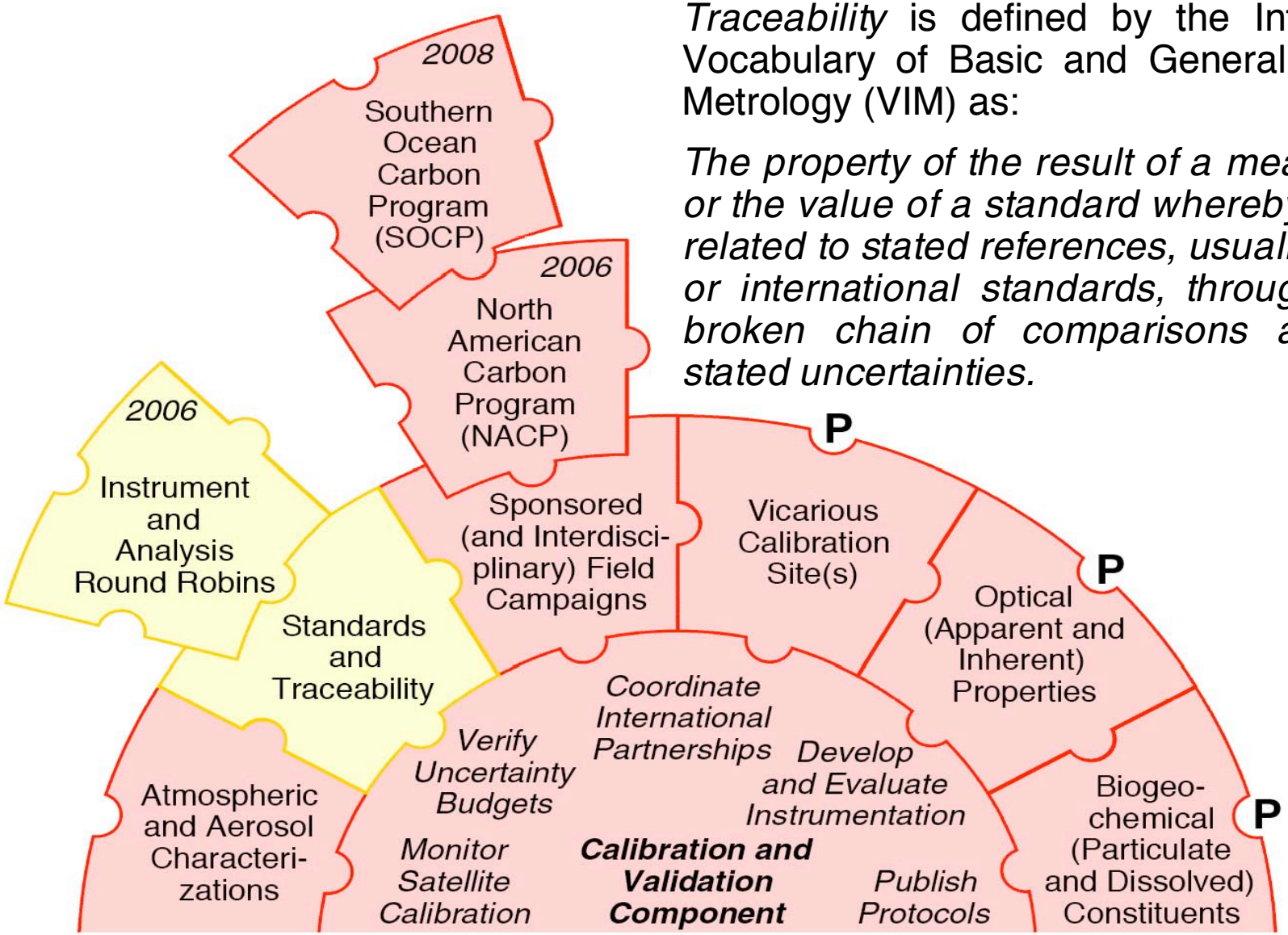
Bold Red No absorption coefficient or at least 40 years old.

Bold Red More than 50 years old

Traceability to a National Metrology Institute (NMI)—NIST for the U.S.—For a Parameter is Often Difficult

Traceability is defined by the International Vocabulary of Basic and General Terms in Metrology (VIM) as:

The property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.



Lessons Learned Establish the Need for an *Emphasis on Uncertainty Awareness*

Using the HPLC community as an example of biological and biogeochemical data, because it is the most studied, lessons have been learned that need to be part of the framework for establishing how a QA capability is implemented for CDR analyses:

- HPLC pigment *methods have been implemented with an incomplete understanding of the quantitative requirements for accuracy and precision*.
- Awareness of *uncertainty components and their magnitudes have not been adequately understood before methods were put into routine use*.
- For the analysis of natural samples, *a single laboratory that does not participate in round robins has an unknown accuracy and performance capability*.
- *Method performance can be properly evaluated and improved with round robins emphasizing an understanding of the sources of uncertainties*.
- The understanding of HPLC uncertainties began with the SeaHARRE activity, so *method capabilities occurred in reverse order: methods were first implemented, and then they were evaluated*.
- The result of this backward philosophy is *large pigment databases may or may not meet the accuracy requirements for their intended applications*.
- Most of the uncertainty and method performance analysis used open-ocean samples, *coastal waters are proving more challenging and rather different*.

The CHORS Problem is a Wake-Up Call—*NASA Must Implement Processes to Improve Data Quality*

The application of NASA data to global warming problems will result in conclusions and policy changes that may be every bit as important as the health-related decisions made by the FDA and EPA based on data under their oversight.

—Dr. John Dolan (8 October 2007)

- About 25–30 years ago, the FDA and EPA implicitly trusted laboratory data, then came the realization that errors occur and they can be significant.
- *What was assumed to be good science wasn't always good science when critically examined*, so guidelines were put into place regarding method validation requirements, system suitability, quality control, reporting limits, operator training, recordkeeping, as well as proof of proper maintenance, calibration, and change control of instrumentation.
- Those who lived through this process usually grumbled at the requirements, but in retrospect, *almost everyone—from the analysts to the laboratory directors—view the changes as being both good and necessary*.
- *Data quality and laboratory efficiency improved significantly*, which reduced costs, and the public image of the industry was substantially restored.
- The initial investment in these processes was big, but *the payoff was worth it*.
- There is a large body of procedural information available from other agencies, so *the development of a quality system does not have to start from scratch*.

Future Directions

Drawing on the accomplishments of the SIRREX, DARR, and SeaHARRE activities, and adding in the specific problems of the CHORS analyses, several important recommendations can be made for implementing a QA capability for CDR analyses:

- *Performance metrics and round robins need to be established for all analyses important to CDRs* (right now only AOPs and HPLC pigments have done this, with the latter being the most comprehensive and evaluated).
- The *performance metrics should include a sufficient diversity in a) the number of variables describing performance, to ensure methods can be adequately evaluated; and b) the different levels of accomplishment to improve the quality of all research endeavors*—and not just the most important, like calibration and validation for CDR analyses.
- All *analyses for CDRs must have a quality-assurance plan (QAP) that is approved by the program manager or cognizant project office*. The QAP must include a) method validation, b) standard operating procedures and protocols, c) appropriate training, d) QA of all data, and e) standardized record keeping (recording, rejection, change control, review, and archiving).
- Programmatic or project *oversight is needed to ensure inspections and compliance with the QAP*.

Future Directions (*cont.*)

Additional recommendations, more from an Agency perspective, are as follows:

- For those activities funded by the NASA Ocean Biology and Biogeochemistry (OBB) Program, *mandatory workshops with laboratory certifications will be conducted annually or every two years for any laboratory and technician that is conducting CDR analyses* (HPLC analyses are already compliant).
- To ensure proper control and review procedures are in place for all analyses essential in the production of CDRs, *a panel will be convened to make recommendations to NASA about a) implementing an oversight process with specific guidelines, and b) strengthening the peer-review process*.
- Similar sampling, laboratory, and analysis problems, both from a protocol and data quality perspective, might be discovered with data from other important measurements (e.g., IOPs and AOPs), if the laboratories involved were examined as closely as was done for HPLC. This means *the review panel should consider the widest possible context in their recommendations*.
- *The FDA and EPA recognized these problems 25 years ago and have designed and debugged many control procedures that can be transferred into the NASA program*, which would also allow the procedures to be thoroughly discussed before they were implemented.

Options to Consider

The main difficulty with the CHORS data is each attempt to peel back a layer of the problem has exposed a new problem. Options for any future effort are as follows:

1. Classify the data as being unsuitable for calibration and validation activities, remove them from SeaBASS (already done), and do no additional work. In a few years many sampling holes will be filled by ongoing research and the HPL contract. Individual PIs would have to determine the applicability of existing CHORS data to their research objectives (past and present).
2. Attempt to characterize the nonlinearities for Chl *a* (the most extensively calibrated CHORS pigment and probably the most important), establish a correction scheme, and correct the data. This will require new resources and personnel (CHORS stops working on this problem 31 May 2008).
3. Attempt to characterize the nonlinearities for the primary pigments, establish a correction scheme, and correct the data. This will require substantial new resources (reintegration of some pigments is likely).
4. Attempt to characterize the nonlinearities for all the pigments CHORS reported, establish a correction scheme, and correct the data. This will require very significant new resources (reintegration of many pigments).
5. A final option—*suggested by more than one PI*— is to ignore the problem and leave the data as is.

Questions to Address

The questions for the Breakout II sessions encompass more than the quantitation of HPLC pigments and are supposed to include all the aspects associated with aquatic calibration and validation activities. Under this broader mandate, the questions to be addressed are as follows:

- *What does the carbon cycle and ecosystems community expect of this effort* (in particular, what does the science demand)?
- *What are our biggest challenges in this area, and how do we address them* (e.g., vicarious calibration or the transition from an open-ocean perspective to one most likely dominated by coastal and near-shore processes)?
- *Is our list of identified data records complete, or is something missing* (if data records need to be added, should they be archived in SeaBASS, even if the only way to do that is to expand SeaBASS)?
- *Does the carbon cycle and ecosystems community need to establish priorities for these and other activities, and, if so, how should they be established*?