NASA Ocean Biology and Biogeochemistry
GSFC Calibration and Validation Office
1450 S. Rolling Road, Halethorpe, Maryland 21227
+1·410·294·7451

27 April 2009

Dear Colleagues:

This report was requested by the Ocean Biology and Biogeochemistry (OBB) program manager, Dr. Paula Bontempi. The primary purpose of the report is to summarize the investigations undertaken by the team established by NASA HQ to investigate data quality problems with the high performance liquid chromatography (HPLC) analyses of pigment concentrations in seawater samples produced by the San Diego State University (SDSU) Center for Hydro-Optics and Remote Sensing (CHORS). The investigative members of the team (hereafter referred to as The Team) and their affiliations are as follows:

Dr. Stanford Hooker, NASA Calibration and Validation Office (CVO);
Ms. Laurie Van Heukelem, Horn Point Laboratory (HPL);
Mr. Jason Perl, Center for Hydro-Optics and Remote Sensing (CHORS);
Dr. John Dolan, LC Resources; and
Mr. Ron Farnbach, Ron Farnbach Consulting.

The other persons who contributed to different parts of the overall process are Dr. Charles Trees (CHORS), Ms. Giulietta Fargion (CHORS), Ms. Crystal Thomas (HPL), Ms. Aimee Neeley (CVO), and Dr. Mary Russ (CVO).

For those unfamiliar with this problem or the complexities of HPLC analyses, review material is presented in the first four sections: Sect. 1 provides a background on the CHORS analyses; Sect. 2 presents the many aspects of method validation and explains why the CHORS results cannot be considered validated, Sect. 3 describes the components of a quality assurance plan and documents the CHORS results were almost always out of control, and Sect. 4 deals with the specific problem of the nonlinear response of the detector CHORS used.

There are new analyses and plots of previously discussed aspects of the overall problem within the first four sections, but they mostly reinforce in greater detail the prior analyses and discussions already made available to the community in other forums (see http://oceancolor.gsfc.nasa.gov/DOCS/ for relevant material). A summary of what has transpired since the combined Carbon Cycle and Ecosystems (CC&E) and Ocean Color Research Team (OCRT) meeting in April 2008 is presented in Sect. 5. Conclusions and recommendations are presented in Sect. 6, a summary, but detailed, chronology of the full problem set is provided in Sect. 7, and the cited references are provided in Sect. 8.

This report shows *CHORS did not validate the C_{18} or the C_8 method before either was placed into service to analyze field samples for NASA principal investigators (PIs)*, even

though the HPLC literature contained easily accessible method validation procedures and importance of implementing such more than a decade ago. The report also establishes there were so many sources of significant variance in the CHORS methodologies, the HPLC system was rarely operating within performance criteria capable of producing data of the requisite quality. CHORS appeared not to be cognizant of many uncertainty sources and repeatedly made decisions regarding hardware use and standard laboratory procedures that did not constrain uncertainties, but, in fact, exacerbated them. For example, a retrospective analysis of CHORS data reveals more than one uncertainty source was capable of contributing ten-fold greater variance than a validated method would expect. The amplified variance in the CHORS results not only degraded accuracy and precision, but, because a limit of quantitation (i.e., the point at which results can be unequivocally discriminated from noise) is based on the magnitude of variance in results, a large portion of the CHORS results may be below such a limit.

Particularly damaging to the objective of understanding the sources of uncertainty and correcting them was the fact that CHORS did not have a quality assurance plan or implement quality assurance (QA) and quality control (QC) capabilities—the absence of which during the analysis of samples means a retrospective effort to improve results is significantly thwarted, because method performance is unknown as a function of time. It is possible to conduct experiments to better characterize some of the variability with the CHORS protocols, but *more than 30 sources of uncertainty are identified in the following report*. Even if a significant effort is made, there is no guarantee that knowledge gained could be used to improve results, because the needed QA data are unavailable and the needed metadata were not always recorded. Finally, evaluating the efficacy of any proposed correction scheme is hindered by the lack of a large and diverse archive of sample filters and no routine analysis of duplicate filters by CHORS.

The comforting aspect of much of the variance in the CHORS methodology is it does not appear to have a trend—so large-scale averages of the results might very well be suitable for a variety of inquiries. For example, the OC4 algorithm does not exhibit large-scale changes if the CHORS data are included in the derivation of the fitting terms or not. Taking all of these elements into consideration, and remembering that it is the nature of science to build and improve upon the previous generation of results, it is the recommendation of The Team to a) not correct the data, b) put all the data that was removed from SeaBASS back into the database, and c) label the affected data with an appropriate warning, e.g., "*These data are not validated and should not be used as the sole basis for a scientific result, conclusion, or hypothesis—independent corroborating evidence is required*".

One of the difficulties in writing this report was trying to speak to several audiences. A very important reader is a current or future analyst, whether for HPLC, biogeochemistry, or any other analytical variable. The authors very much want to impress on analysts that regardless of how much time they spend executing a protocol, they have to maintain a healthy curiosity about what they are producing. CHORS had virtually all the same data the authors used to produce this report, they simply never looked at it from the point of view of asking probative questions. Conclusions for improving HPLC analyses are given

in Sect. 6, but they are rather easily adapted to a wide diversity of measurements. The lessons learned and the recommended future directions must extend beyond the need to populate NASA databases with good data and include the requirements for next-generation missions and the maintenance of climate-quality data records (CDRs). The latter will require *in situ* data with unprecedented quality, so several important recommendations are made about implementing QA and QC capabilities for CDR analyses.

Although the CHORS HPLC problem represents a case study in how undetected low- and high-level mistakes can have a significant and negative impact on the quality of an entire program, the full responsibility for the problem extends beyond a single laboratory. Regardless of what CHORS did incorrectly, their proposals were peer reviewed, as was their attempt to understand their problems once they were notified about them, and they were subjected to NASA oversight as part of the contract reporting process. None of those procedures, which are all associated with quality assessment, correctly identified CHORS analyses were significantly degraded or correctly identified the source of the problems.

This work has been a much more significant undertaking for the individuals who were recruited or volunteered their time than was first imagined when the CHORS HPLC quantitation problems were discussed at the beginning of 2006. In fact, as it turns out, evidence of the problem surfaced back in the middle of 2002, as documented in the chronology section. Unfortunately, this was not the only missed opportunity to minimize the impact associated with this problem. Consequently, the authors hope this document will be a clarion call to anyone responsible for analytical procedures and—most importantly—the entire ocean color community. What is needed is some introspection followed by frank discussions about programmatic changes to improve QA and QC procedures. Some initial recommendations as to how NASA can improve data quality while taking advantage of the lessons learned from this problem, as well as from other agencies that have had to do the same thing (e.g., EPA and FDA), are presented in Sect. 6.

Thank you for your patience as The Team worked through this very difficult problem. If you have any questions or comments, please address them to either Stan Hooker (stanford.b.hooker@nasa.gov) or Paula Bontempi (paula.bontempi@nasa.gov); the other members of The Team are no longer recurring participants in this activity (also documented in the chronology).

Stanford B. Hooker
Calibration and Validation Office (CVO)

Laurie Van Heukelem
Horn Point Laboratory (HPL)

1. **Analysis Background**

Table 1 shows the number of samples analyzed with the two HPLC methods (C_8 and C_{18}) used by CHORS. The time periods span the addition of a Thermo UV6000LP detector in March 1998 to an existing Thermo Separations Products Spectra System HPLC with a UV2000 detector, up until the last field samples were analyzed in 2007. The entries highlighted in blue are the samples quantitated as part of NASA contracts, first for the Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) project, which provided funding for a duplicate Thermo Separations Products HPLC system in July 2000, and then for the Moderate Resolution Imaging Spectroradiometer (MODIS) project. The annual total of samples are indicated in the two columns for NASA and other investigators (highlighted in yellow). The total number of samples involved is estimated to be a little less than 24,000, of which about 17,000 are samples from NASA PIs. Although not categorized as NASA samples, some of the other samples are of interest to NASA, because they include activities potentially important to the OBB program, for example, the analysis of samples from the Marine Optical Buoy (MOBY) site.

Table 1. A summary inventory of the CHORS HPLC analyses as a function of time and the method being used.

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

For the discussion presented here, however, the only explicit samples of interest are from 2001–2007, inclusive, because CHORS was not able to recover the calibration and chromatography files from 1998–2000 and submit them to The Team tasked with investigating the HPLC problems. Based on what The Team was able to determine, however, there is every reason to believe the majority of what was found deficient in the 2001–2007 time period is completely applicable to the 1998–2000 analyses.

The most recurring pigments involved in the discussions of the two CHORS methods are as follows:

Chlorophyll *a* (Chl *a*), which includes all allomers and epimers;
Divinyl chlorophyll *a* (DVChl *a*);
Chlorophyllide *a* (Chlide *a*);
Total chlorophyll *a* (TChl *a*), which is the sum of Chl *a*, DVChl *a*, and Chlide *a*;
Chlorophyll *b* (Chl *b*), which is the total chlorophyll *b* (TChl *b*) for these methods;
Chlorophyll c_1 (Chl c_1);
Chlorophyll c_2 (Chl c_2);
Chlorophyll c_3 (Chl c_3);
Total chlorophyll *c* (TChl *c*), which is the sum of Chl c_1, Chl c_2, and Chl c_3;

α-Carotene (α-Car);
β-Carotene (β-Car);
Carotenes (Caro), which is the sum of α-Car and β-Car;
Alloxanthin (Allo),
19′-Butanoyloxyfucoxanthin (But);
Diadinoxanthin (Diad);
Diatoxanthin (Diato);
Fucoxanthin (Fuco);
19′-Hexanoyloxyfucoxanthin (Hex);
Peridinin (Peri);
Zeaxanthin (Zea);
Lutein (Lut);
Prasinoxanthin (Pras);
Violaxanthin (Viola);
Phaeophytin *a* (Phytin *a*); and
Phaeophorbide *a* (Phide *a*).

Although there are some other (usually minor) pigments involved with the full analytical results for the two methods, the pigments listed here are routinely considered as part of the SeaHARRE activities, so there is a larger body of information to access for these pigments. There are other so-called higher-order data products that SeaHARRE analyses make use of (for example, the total accessory pigments), but these are not considered in this report.

In most of the ensuing investigations concerning the deficiencies in the CHORS protocols, expectations of quality are explicitly provided in terms of community-wide performance metrics or norms, which are frequently parameterized in terms of numerical thresholds or limits (e.g., the residuals to a calibration curve should be to within ±2% on average, for quality-assured results). It is important to remember departures from numerical thresholds or limits are a routine part of maintaining complicated analytical systems and are anticipated. Their impact on data quality is linked to the magnitude of the departure, how frequently it occurs, what investigative steps are taken to understand the cause(s), and then what corrections to the protocol are made to minimize any reoccurrence.

In the following presentations of the CHORS HPLC problems, the reader will be faced with two strongly contrasting information sets. One set is based on vigilantly monitoring control variables to detect the onset of inevitable degradations in method performance (e.g., caused by column aging), and then to quickly provide corrective measures. The other set contains a long time series of systematic and significant problems that originally were not properly investigated—*indeed, perhaps not even detected*—and, therefore, never adequately resolved. The contrast is inexorably tied to the QA and QC data that are supposed to be collected contemporaneously during sample analysis, because these data provide the metrics to detect problems and, thus, are some of the most important parameters to be used in determining the scope of the problem and how to pursue a correction. *The absence of QA and QC data provides the most extreme contrast, because it hinders the original analyst who is tasked with keeping a method in*

control and producing quality data, and the forensics investigator who is trying to reconstruct why a particular method was out of control: with no QA and QC data, both are denied the most powerful tools for successfully doing their jobs.

2. Method Validation

Method validation is a routine process that all analysts should follow to determine whether a method is suitable for its intended application (and requisite accuracy requirements) and is conducted before a method is put into service for field sample analysis. Representative topics evaluated during method validation are as follows (EURACHEM 1998): a) specificity[1], b) limit of detection and quantitation (LOD and LOQ, respectively), c) ruggedness[2], d) working and linear ranges, e) calibration, and f) accuracy and precision. The importance of the method validation topics and how they relate to producing quality-assured results is a well established part of the HPLC literature, which was brought together in a single volume by Jeffrey et al. (1997)—in fact, the book is sufficiently comprehensive that analysts refer to it as *The HPLC Bible*.

Although there has been a steady maturation and refinement of HPLC method validation since the influential work of Jeffrey et al. (1997), the basic principles are still as applicable today as they were more than a decade ago. Additional evidence of how HPLC analysts embraced—and continue to apply—the fundamental aspects of method validation is seen in the recurring use of the EURACHEM (1998) procedures, which were also established more than 10 years ago. Between the Jeffrey et al. (1997) book and the EURACHEM (1998) procedures, the most significant aspects of method validation have been, and continue to be, accessible to junior and senior HPLC practitioners alike (the EURACHEM procedures are available on the Web).

If method validation is done thoroughly and carefully, a validated method is capable of producing quality-assured results, that is, data satisfying the expected accuracy and precision capabilities. Conversely, if the work is done without the proper attention to detail, the interlinked nature of the many components involved will result in degraded performance in a large diversity of parameters. To some extent, the CHORS archives can be searched to reconstruct how insufficient method validation adversely affected their ability to provide results of consistent quality. To put such a retrospective analysis in perspective, however, it is helpful to first describe a typical, systematic sequence of events pertaining to method validation and then look at what can be discerned from the CHORS results. In the material presented here, accuracy and precision are considered first and the other topics are presented in subsequent sections.

---

[1] Specificity refers to the ability to accurately determine a pigment concentration in the presence of other components (e.g., other pigments, impurities, or degradation products), which are expected to interfere with the identification and quantitation of the pigment. Coelution is the most common form of a specificity problem, and it occurs over a wide range of severity if all pigments and methods are considered.

[2] Ruggedness, as adapted from EURACHEM (1998), refers to the degree of reproducibility in the results obtained from a method under a variety of representative test conditions (e.g., different days, different—but properly trained—analysts, different laboratories, different instruments, etc.).

2.1 Introduction

Typically, method validation would first include investigations into the simplest form of precision: replicate injections of the same standard all performed on the same day to confirm the requisite repeatability of the system can be obtained (for quantitative analysis, as defined by the SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE) community (Hooker et al. 2005), the injection precision should be less than 2% for Chl *a*). Second, the range in amounts injected over which acceptable results can be obtained (the working range) and whether the volumetric information (volumes filtered, extracted and injected) yield concentrations in sample extracts that fall within that working range are determined. Evaluations of working ranges also reveal whether the HPLC system yields a linear response or whether nonlinear calibration is required. Once these initial performance criteria are addressed, estimates of calibration precision over the short-term (repeatability precision) and long-term (reproducibility precision) are needed to determine frequency of recalibration and to establish a knowledge base for determining whether newly observed calibration factors are within the expected calibration variance.

With the above information established, informed decisions regarding appropriate quality control checks (injected on a daily basis, for example and which validate the accuracy of the calibration factors in use) can be made. For example, for linear systems, frequent injections (e.g., daily) of a Chl *a* standard solution representing a concentration that is midway in the working range is appropriate. For nonlinear systems, the need for full recalibration is more frequent (Snyder and Kirkland 1979). Accuracy with field samples and standards can then be evaluated by intercomparisons with other laboratories. Accuracy with field samples can be evaluated in house with spiked-recoveries, where known amounts of standard(s) are added to a sample extract and the amount “recovered” during analysis is compared to the amount added to the sample extract (Clesceri 1998 and Bidigare et al. 2005).

Precision with field samples is easily assessed by analyzing replicate filters over the short- and long-term, thus establishing the variance in results when filters are analyzed on the same day versus when they are analyzed on less frequent time intervals (e.g., months). Analysis precision is assessed by replicate injections of a sample extract at intervals that describe the minimum and maximum time a sample extract resides in the autosampler compartment prior to injection.

2.2 Method Validation

With regard to the validation introduction above, the most significant failure of the CHORS analyses of HPLC samples is

> *CHORS did not validate the C_{18} or the C_8 method before either was placed into service to analyze field samples for NASA PIs.*

The following discussions are organized according to accepted validation practices for chromatographic methods, as discussed in Sects. 2.0 and 2.1. The level of detail

documented by The Team and the reconstruction of what was (and wasn't done) by CHORS regarding method validation is crucial for determining a) if data correction is possible, b) how to implement a viable data correction scheme, and c) whether or not a useful estimate of the uncertainties associated with corrected data can be produced.

2.2.1 Accuracy

Because CHORS did not validate their methods, The Team reconstructed validation elements by reviewing the pertinent raw data, procedural descriptions, and the available metadata, in the hopes that uncertainties with variables in the CHORS quantitation equation could be assessed. Uncertainty assessments are necessary to determine the potential accuracy of a method, but for the investigations presented here, they were also part of trying to determine whether or not the CHORS results could be corrected.

CHORS used an internal standard for both the C_{18} and C_8 methods. Under normal circumstances, an internal standard is a powerful tool to improve and maintain accuracy. Like many laboratories, CHORS used the internal standard as part of the pigment extraction process. The CHORS quantitation equation for calculating the concentration of a particular pigment, C_P, is shown in the equation to the right (1), and is composed of the following terms: V_m is the mixed volume of extraction solvent and internal standard delivered to the filter for extracting pigments; V_f is the volume of seawater originally filtered; A_c is the peak area of the internal standard in the solvent solution before it is added to sample filters; A_s is the peak area of the internal standard in the sample extract, which is measured at the same time other pigments in the extract are quantified; A_P is the peak area of the pigment to be quantitated; and R_P is the response factor (RF) of the pigment (determined during the calibration of the pigment).

$$C_P = \frac{V_m}{V_f}\,\frac{A_c}{A_s}\,A_P\,R_P \qquad (1)$$

The only term in the quantitation equation (1) that is not in direct control of the HPLC analyst is V_f, because it is provided by the PI supplying the samples (who follows a sampling protocol to determine the appropriate volume of water to be filtered)—all the other terms represent controllable or accessible sources of uncertainty for the HPLC analyst. Laboratories producing quality-assured analyses make continuing efforts to understand the uncertainties in the individual terms for their particular quantitation equations. The latter involves the collection of specific QA measurements to ensure the uncertainties are to within method requirements. The follow-on discussions are pertinent to assessing uncertainties with the accessible terms for the CHORS methods: V_m, A_s, A_c, and A_P.

Uncertainties with the V_m term in (1) are assessed by calibrating the solvent-delivery device used to add the solvent/internal standard mixture to filters. The Team has no evidence showing CHORS calibrated for V_m other than during SeaHARRE-2, when all participants were asked to provide documentation of pipette calibration. The value for A_c is determined from replicate HPLC injections of the solvent–internal standard solution before it is added to samples, and in CHORS procedures, A_c was determined at the

beginning of a set of extractions. From metadata provided to The Team, it appears CHORS extracted large batches of filters at a time (more than could be analyzed in one day) and used the same A_c term with all samples in a batch (approximately 450 filters were in batches described here, which would require approximately 14 days of continuous HPLC operation). The stability of the internal standard in the solvent solution, represented by A_c, was checked by CHORS over the course of a single day of injections, but there is no evidence A_c stability over multiple days was checked or was re-verified on each day the solution was used. Metadata for each sample is incomplete or lacking pertinent temporal details, for example, date of extraction, date of analysis, storage conditions (and length of storage) of sample extracts prior to analysis and length of storage, length of time a sample resided in the autosampler compartment before injection, and procedures pertaining to a need for sample re-injection as a result of hardware failure.

In a properly controlled method, A_s/A_c is expected to remain fairly constant, with small deviations (e.g., ±2% fluctuations usually describe 95% of the variations). In general, A_s is slightly less than A_c, because the solvent added to the filter is diluted by water retained on the filter as part of the filtration process, or A_s may be slightly higher than A_c if evaporation occurs in the sample extract. At HPL, for example, the average A_s/A_c ratio between samples from three NASA investigators varied from 0.90–0.93 (0.93 is expected given the volume of liquid added to filters at HPL for extraction) and the average CV within each data set varied from 1.3–2.6% (for 704 samples). There is a practical (and theoretical) limit to the A_s/A_c ratio for properly validated methods, which is dictated by the usual amount of water on the filter and the volume of liquid added to the filter for extraction. Assuming the 0.2 mL value (for a 25 mm GF/F filter) suggested by Bidigare et al. (2003) is appropriate, and knowing the volume CHORS added to filters (4 mL), A_s/A_c should frequently approximate a value of 0.95.

For the purposes of investigating the CHORS A_s/A_c ratios, the data are presented in Fig. 1 as the relative percent difference (RPD[3]) between the observed value in the results and the 0.95 reference value. In both cases, A_c was held constant by CHORS for all analyses (about 450 samples). The average ratio is 0.75 and the CV is 12.1% for the C_8 method, which used Carotenal as the internal standard, and the average ratio is 0.93 and the CV is 14.0% for the C_{18} method, for which Cantha was the internal standard. The variance in CHORS As/Ac data are significantly greater than the expected range of approximately ±2% (shown by the yellow band in each panel). Also of note are the large outliers in the C_{18} data (the values on the *y*-axis limits are for data exceeding the ±30% plot boundaries) plus the sudden change in the basic A_s/A_c functional form in the C_8 method associated with the second group of data (shown in green) from about injection numbers 90–210 (in fact, these data correspond to about one full auto sampler compartment). The latter is particularly troubling, because it suggests a significant anomaly occurred in the routine analyses of these particular samples. Based on the

[3] The computation is RPD = 100(O-R)/R, where O is the observed value and R is the reference value. A positive RPD means the observed value was greater than the reference value, and a negative value means the opposite. Biases are indicated by a persistent expression of one sign or the other.

numerous problems CHORS had with hardware (Sect. 7), it is likely there was an interruption in the HPLC analyses and the sample extracts degraded.

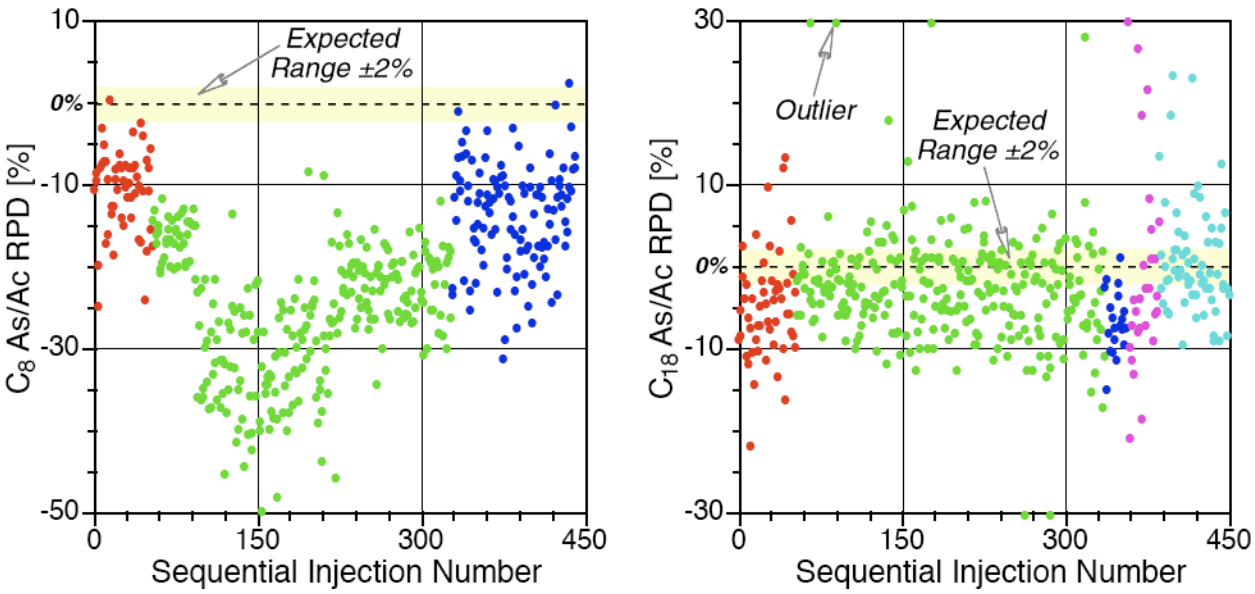

Fig. 1. The RPD in the As/Ac ratio with respect to a 0.95 reference value for the CHORS C_8 and C_{18} methods (left and right panels, respectively). The different colors correspond to different PIs.

The recurring low A_s/A_c ratio in C_8 results suggests the internal standard degraded in all samples. One way of establishing how severe such degradation might have been is to invert the logic associated with computing the A_s/A_c ratio and use the parameter to solve for the amount of water that would have to have been retained within a sample filter to produce the observed A_s/A_c ratio. If this computation is made, the amount of water that is needed averages 1.3 mL (with a 4.3 mL maximum) even though only 0.2 mL is expected for 25 mm GF/F filters. Alternate hypotheses, that the internal standard was unstable under the conditions used by CHORS or there was a specificity problem affecting the ability to accurately determine the A_s peak area, are more plausible.

An altered A_s/A_c ratio might be expected in the following abnormal and undesirable circumstances:

1. The injector inaccurately draws up enough volume during sample analysis, so A_s is reduced.
2. The internal standard is not stable during a) extraction, b) residence time in the autosampler compartment, or c) excessive storage while the sample is waiting for analysis (as might occur when a hardware failures occur), so A_s is reduced.
3. A naturally occurring compound co-elutes with the internal standard, so A_s is increased.
4. An adjacent pigment is not baseline resolved from the internal standard, so integration of the entire peak is not possible and the peak area of A_s is reduced.

5. The internal standard exhibits a peak area response in 100% acetone (the composition of CHORS solvent solution before it is added to filters) that is different from when it is in an aqueous–acetonic environment (as with sample extracts).
6. The A_s term is imprecise (or inaccurate) if the peak area is sufficiently small (or the peak is sufficiently broad) that discriminating the beginning and ending of the peak is difficult, especially if greater baseline disturbance is seen in sample extract injections than with standard injections.
7. The areas used for A_s and A_c are not within a linear range, and A_s is either reduced or increased relative to A_c.

It is plausible the suppressed A_s/A_c ratios observed with the Carotenal internal standard data for the C_8 method (Fig. 1) were the result of degradation. It is not known if the environmental influences that might have caused such a response would also have affected other pigments in the sample extracts, or if all pigments would have been affected equally. While it is possible to conduct experiments to evaluate degradation rates under varying circumstances, it would not be possible to reproduce the effects of hardware failures, for example. Also, metadata from CHORS is insufficient to know how to apply such knowledge about degradation rates to results of individual samples (e.g., how long an extract was stored before analysis, because of an equipment failure).

With the data set of 450 samples analyzed with the C_8 method at CHORS (Fig. 1), the frequency with which Caro was reported absent was 50%, 7%, and 0% for oligotrophic, mesotrophic, and eutrophic samples, respectively (as categorized by TChl *a* regimes of less than 0.1, 0.1–1.0, and greater than 1.0 $\mu g\ L^{-1}$, respectively). In fact, Caro was always reported as absent when TChl *a* was less than 0.05 $\mu g\ L^{-1}$. Caro is present in all algal divisions and because there is evidence of the internal standard peak area being suppressed with the C_8 method, it is likely Caro was found absent because of a generalized effect on peak area (which means other pigments in the chromatogram were also suppressed). It is plausible, therefore, that the oligotrophic samples and some of the mesotrophic samples are at risk of being below a quantitation limit. In comparison, at HPL it is atypical for Caro to be absent until TChl *a* values are less than 0.005 $\mu g\ L^{-1}$.

2.2.2 Specificity

Specificity involves the determination of the elution position of pigments to be quantified and interferences that have a potential for co-eluting with them. Peaks in the chromatogram of the various pigments must be sufficiently resolved—that is, separated in time from one another—for correct identification and quantitation. The adequacy of separation is quantified by the terms, resolution[4] and separation selectivity[5]. When developing a chromatographic method, the analyst needs to simultaneously control variables that primarily affect separation selectivity (e.g., mobile phase and stationary

[4] Resolution is equal to the distance between the two peak centers, divided by the average peak width (Snyder and Kirkland 1979).

[5] In a generalized definition, separation selectivity (or separation factor) refers to the position of the peak apex of one peak relative to another—it does not take into account peak widths.

phase composition) and peak width (e.g., column length, stationary phase particle diameter, and solvent velocity).

Peak position and peak width must be controlled to achieve adequate resolution for quantitative analysis by peak area, which is an important requirement for the CHORS methods, because all of the quantitations were based on peak area. Resolution and separation selectivity (documented as attainable during method validation) is expected to remain constant within specified tolerances under anticipated operating conditions. Frequent injections (e.g., daily) of a pigment mixture are used to temporally validate that resolution and selectivity (plus peak shape) remain sufficiently adequate for acceptable quantitation. Such monitoring is crucial because many factors can affect resolution and separation selectivity (Sect. 2.2.6).

CHORS provided no documentation that specificity parameters were defined and rigorously applied over time. The most damaging effects of this lack of validation are expressed in false-positive results—pigments are reported as present, when in fact they are not—keeping in mind that falsehoods in reporting are only identifiable in a round-robin environment. Reporting of false positives by CHORS with the C_{18} method has been observed in NASA round robins for many pigments, most notably (but not limited to) DVChl *a*, Pras, Diato, Viola, Lut, But, and Hex. Some types of false positives in CHORS results have also occurred with results of quality-assured laboratories, but for the latter, such events have been limited to difficulties with the identification of small peaks in otherwise complex chromatograms produced by concentrated sample extracts from eutrophic systems.

The persistent misidentification of large peaks in a chromatogram is indicative of method failure caused by undocumented co-elution with at least one, possibly more, other major component—such problems have been documented in CHORS C_{18} data for But and Hex in eutrophic, coastal water samples. In some cases, the peak that was misidentified was very large, approximating 40% of the peak area of the most dominant carotenoid in the chromatogram. It may be possible, with individual inspection of all peaks in all CHORS chromatograms, to determine the frequency of such types of false-positive reporting, but if the cause is determined to be from co-elution, improvements to quantitation may not be possible.

Another specificity problem is a false-negative result—pigments are reported as not found when in fact they are present. False-negative reporting for Caro by CHORS with the C_{18} method has been documented in more than one NASA round robin with a frequency as high as 80%. Causes for false-negative reporting are many, including an inability to adequately discriminate the beginning and ending of peaks that are broad and irregularly shaped, especially when the baseline is unstable. It is known that CHORS chromatograms exhibited these characteristics for the Caro pigments and that CHORS individually quantified α-Car and β-Car even though resolution between these two pigments was well below 1.0—the threshold below which quantitation by peak area is not recommended (Snyder and Kirland 1979).

The frequency of false-negative reporting in field samples can only be determined by individually inspecting all CHORS chromatograms and then carefully comparing the quantitation results to the limit of detection (LOD) and quantitation (LOQ). The latter are common thresholds analysts rely on to establish the smallest concentrations that can be reliably reported. The process for establishing a threshold is usually quantified by determining the detectability limits of the entire method, i.e., the skill of the analyst, the capabilities of the hardware, the calibration of any laboratory glassware, etc. The SeaHARRE community established a process based on determining the amount (in nanograms) of an injected pigment that fulfills a specified signal-to-noise ratio (SNR). The amount of pigment that results in an SNR of 3 is defined as the LOD and the amount of pigment that results in an SNR of 10 is defined as the LOQ.

It is important to remember the largest uncertainties occur with the reporting of falsehoods, so any procedures that are adopted during method validation to minimize the likelihood of a false positive or false negative is extremely important. In an era where phytoplankton community composition is hoped to be reconstructed from the pigment data, false reporting is particularly egregious, because it can remove or include the presence of entire species rather than simply modulate the abundance of a species.

2.2.3 LOD and LOQ

Prior to SeaHARRE activities, pigment analysts were typically rather naive about the consequences of quantifying pigments near a detection or quantitation threshold, even though recommendations for determining LOD and LOQ have been established for some time (Mantoura and Repeta 1997 and Bidigare and Trees 2000). Pigment analysts attending SeaHARRE working groups, and members of The Team in independent activities, have subsequently made improvements to reporting practices of pigments in low concentrations. Recurring efforts to inform data users of when pigments are quantified at detection-limited concentrations have been published (Hooker et al. 2005) and presented at scientific meetings (e.g., OCRT meetings).

CHORS did not determine LOD and LOQ values for the methods they used, so it is not known how much of their results are affected by the large uncertainties caused by quantitation of pigments near an LOD or LOQ. It may be possible to estimate, retrospectively, an LOD and LOQ based on a non-instrumental approach—a process which does not use signal to noise ratios, but rather uses data from replicate analyses to identify concentrations at which a value simply reflects imprecision in results. This approach requires that pigments are analyzed in replicate field sample filters (and replicate injections of a sample extract), but because CHORS did not routinely analyze either (except replicate filters during NASA round robins), a non-instrumental LOD and LOQ can only be established for data acquired during a round robin.

2.2.4 Precision

HPLC pigment analytical precision is described by the coefficient of variation (CV) of replicate injections of the same solution (usually expressed in percent), for which values less than 1% and up to 2% have been cited as routinely achievable (Mantoura and

Repeta 1997, Van Heukelem et al. 2002, and Hooker et al. 2005). Analytical precision for analyses of standards at CHORS was documented rather routinely, but it was uncharacteristically and persistently poor—even with the first calibration CHORS executed for analyzing NASA samples in April 2001. These seemingly random injection results were never flagged or investigated even though their adverse effect on precision was significant—in many cases, the aberrant injection was over 40% different from the other two injections in a set of triplicates, and in some instances, the difference exceeded 100%. A retrospective study conducted by The Team, following 121 observations of replicate injections of standard Chl *a* solutions conducted between 2001 and 2007 at CHORS, revealed an average injection precision of 4.3% and almost one-third of these observations exceeded 4.3% (the range is 0.1–20.9% among sets of replicate injections). For comparison, such averages at HPL are 0.4% with a 99% confidence limit of 2.2%. CHORS was able to meet this criterion only about 15% of the time for the data cited here. The aberrant data are numerous and can account for as much as one-third of the injections during a calibration procedure.

Overall method precision is described by the CV of pigments analyzed in replicate field samples. Standards of achievement by quality-assured laboratories in SeaHARRE round robins for replicate filters are 5% or better for the primary pigments and 3% or better for TChl *a*. The average CV across duplicate filters provided by multiple NASA PIs have been observed as 7% and 4%, respectively, for the primary pigments and TChl *a* in filters analyzed at HPL, remembering that added imprecision is expected when multiple PIs are preparing duplicate filters. CHORS did not routinely analyze replicate field sample filters, except during all but one NASA round robin. The CHORS average CV for the primary pigments among these round robins varied from 6–23%, with an overall average of 13%, which is substandard to the 5% routinely achievable by quality-assured laboratories in round-robin exercises. While there were times when CHORS exhibited method precision equivalent to quality-assured laboratories, the frequency with which this occurred is not known, because they did not routinely analyze duplicate filters for NASA PIs (although this is part of the sampling protocol).

A difficult aspect of CHORS imprecision is documented in the first SIMBIOS round robin CHORS participated in (SB-1). In SB-1, CHORS reported many pigments as present in at least one, but not all replicate filters, with the damaging effect that the average CV for all pigments was 46%. Another contributor to imprecision is the practice sometimes used by CHORS in which peak area integration was left to the discrimination of the auto-integration software—in other words, not all peaks were inspected for correct peak area integrations. With automated integrations, inaccurate peak areas are exacerbated when it is difficult to discriminate the beginning and ending of a peak, as occurs when peaks are small, asymmetrical or broad, when baseline wander and drift are present, and when “ghost” peaks and humps appear in the chromatograms. Problems with baseline instability were frequently documented by CHORS.

2.2.5 Working Ranges, Linear Ranges, and Calibration

Based on the volumetric information used most frequently at CHORS, a working range encompassing approximately 2–200 ng of Chl *a* is estimated to be realistic for the

analysis of samples from a world ocean sampling perspective (it is important to note CHORS never defined a working range for either the C_{18} or C_8 methods). The range in amounts injected varied considerably with a particular calibration curve, as seen in Fig. 2, wherein the response factor (pigment amount divided by peak area), at 436 nm is plotted as a function of the amount of pigment injected onto the column. Chl *a* standard data is shown (from the C_{18} method) and most standard dilutions were injected in duplicate or triplicate, except in 2006, where data from the calibration events presented appear more similar with regard to concentration, but each dilution was injected only once. The data set includes at least three calibration events per year, except for 2005 and 2007, which had only one and two, respectively, and no calibration data for 2004. The paucity of calibration data from 2004–2006 is associated with the time period when CHORS was using a C_8 method. The expected range in average variability for Chl a calibrations for a quality-assured laboratory is shown by the yellow band (±2%). The CHORS results significantly exceed this range and, in fact, do not satisfy a reasonably attainable range of ±5% (IUPAC 1997).

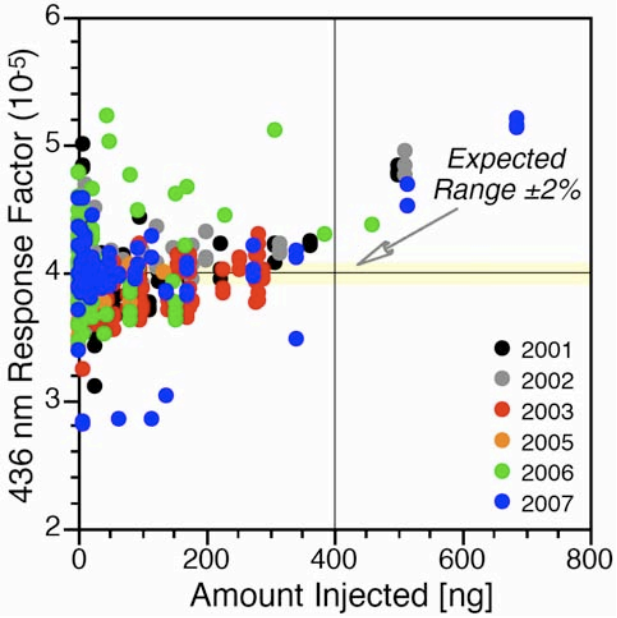

Fig. 2. The response factor (RF) for Chl *a* (C_{18} method) at 436 nm as a function of time (different colors). The expected range in RFs is shown by the yellow band.

The data presented in Fig. 2 were also used in computations of analytical precision statistics (Sect. 2.2.4), which show the full range in the magnitude of the RFs is almost a factor of two. With such a large amount of variance, it is very difficult to identify trends, e.g., whether response factors are stable across the range of concentrations needed for sample analysis and whether response factors are stable over time. Although there is an indication of a nonlinear response extending above 200–300 ng, the data below this threshold have sufficient variance that it is not possible to unequivocally determine whether or not the response is linear or if there is more than one nonlinear regime in the data—the data are simply too noisy. It is important to note that CHORS used these data mostly as is; there was an occasional light editing of the data, but it usually involved removing data from the higher concentration ranges, which frequently had some of the better statistical properties.

In a retrospective analysis to discriminate variance in calibration from variance among injections, The Team computed the average RF from the replicate Chl *a* injections as a function of the amount injected. To reduce the noise in the data, individual values in a replicate were compared to the average and clearly aberrant outliers were discarded. An additional refinement was to compare the resulting averages to their nearest

neighboring replicates and discard any outliers. Following the advice of King (1999) in evaluating the linear through zero range, a normalized RF was computed as a function of the amount injected. For this analysis, an average RF (spanning approximately 2–200 ng) was determined for each year, and because the annual average RFs were very similar and spanned a narrow range, 3.82–4.04 (10^{-5}), with a CV of 2.3%, an overall average RF across all years was used as the normalization reference. This 2.3% CV is an approximation of calibration variance across many years and would normally be considered a very good result, but it is important to remember this value was only produced after reducing the noise in the data caused by the poor analysis precision.

A plot of the individual normalized RFs is shown in Fig. 3 (which are based on the same data used in Fig. 2). The averaged (and edited) data are considerably less noisy and more linear, but the aforementioned nonlinear shift after 200 ng per injection is still seen. With regard to the annual RFs, more than three-fourths of the results within the working range are within ±5% of the average and they appear to have a linear relationship, but they are still noisier than expected for a properly validated method.

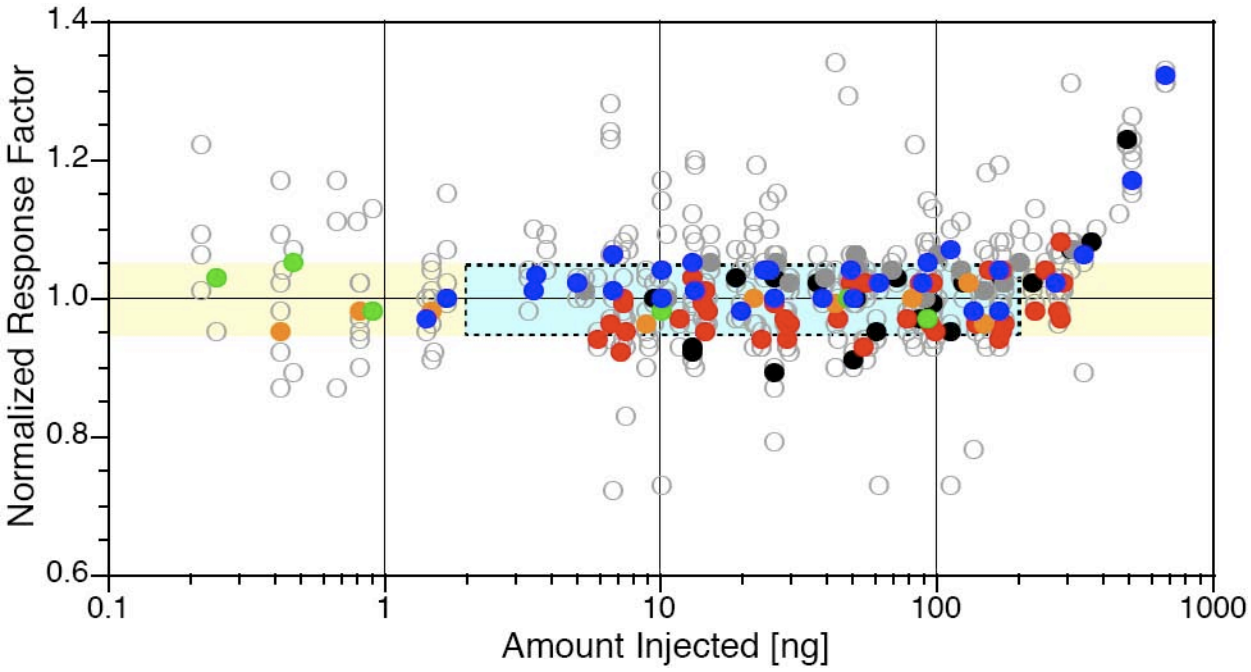

Fig. 3. The individual normalized RFs from Fig. 2. The averaged and edited normalized RFs are shown as solid circles and the remaining data are shown as open circles, and the averaged and edited normalized RFs are shown as solid circles (annual distinctions follow from the figure above). The 2–200 ng working range is indicated by the blue box with ±5% variability bounded by the dashed lines.

The yellow extension of the ±5% threshold outside the working range shows how mostly noise is added below the working range, and how the response becomes nonlinear above the working range. The ±5% threshold is suggested by IUPAC (1997) as reasonably attainable with most chromatographic systems (individual data points should

be within 5% with an expectation that they will be within 2% on average. This definition is in keeping with the SeaHARRE performance metric for quantitative analysis, which requires the average absolute residuals in a calibration should be to within 2% and have a negligible *y*-intercept (if a point with a higher residual is found, it can be discounted if its removal brings the calibration curve into compliance). Consequently, the expected range for the majority of response factors for a calibration curve should be within ±2% and all the data should be within ±5%.

If CHORS had recognized how damaging the imprecision in their calibration data was, and had taken steps to discover the source of the injection imprecision and corrected it, there is evidence they may have, otherwise, had stable calibration factors over the described working range for Chl *a*. It is also likely that they could have discriminated the nonlinear aspects of their calibration. It is important to remember these revelations were only possible, because of all the data that was rejected in the editing process. Unfortunately, the rejection criteria were only possible, because most of the Chl *a* calibrations were done in triplicate. The field samples were not analyzed in triplicate—in fact, CHORS rarely analyzed duplicates—so it is not possible to edit the field data to get a similar result.

2.2.6 Ruggedness

Ruggedness describes the ability of a method to yield consistent accuracy and precision as affected by changes in the operating environment. Most pertinent in the CHORS analysis is a failure to evaluate or provide sufficient documentation for subsequent, retrospective evaluation of the effects of:

1. The holding time between extraction date and analysis date;
2. Typical and maximum times a sample resided in the autosampler compartment;
3. Delays in the analysis of sample extracts (and procedures associated therewith) caused by hardware failures;
4. Exchanging hardware components; and
5. Imprecise column temperature control.

CHORS experienced frequent hardware failures of multiple components and power outages which would interrupt the typical analysis sequence, e.g., the normal time intervals between extraction of filters and HPLC analysis of extracts. Such time intervals are known to cause reduction in quantified amounts of some pigments—most notably the polar chlorophylls and Chl *b* (Hooker et al. 2005). Unfortunately, the CHORS record keeping does not allow accurate discrimination of such time intervals and whether testing was done to evaluate the effects of time delays in extract analysis. During testing of the C_{18} method in April 2008, for example, results of Chl *a* triplicate injections conducted after what could be considered a typical maximum residence time in the autosampler compartment, were 16% lower than when the same solution was injected (in triplicate) at the beginning of the sequence of analyses on that day—yet the same solution analyzed (in triplicate) halfway between these two endpoints were within 0.5% of the initial results. These types of inconsistencies have been very perplexing and

difficult for The Team to reconcile and are disheartening when considering the possibility of data correction.

Controlling the column temperature is important for maintaining consistent separation selectivity, peak shape, retention time reproducibility, and peak area reproducibility, which are all important to ensure correct peak identification and quantitation. CHORS implemented column temperature control for the C_{18} method by setting the air conditioner to maintain the room in which the HPLC was contained at 18°C. CHORS had frequent problems with air handling and power outages, so variations in this temperature were inevitable. Experiments at HPL during method development on three different C_{18} stationary phases (Van Heukelem and Thomas 2001) showed peak areas are frequently reduced at higher temperatures relative to the same analyses conducted at temperatures either 15 or 20°C degrees lower. Peak areas were always reduced for later eluting pigments (with a trend linking increased peak area reduction with later elution positions), but for Peri (an early eluting pigment), the peak area increased in one instance. The average peak area reduction per column ranged from 10—54% (for pigments tested on all columns, Peri, Cantha and Chl *a*).

The temperatures and the temperature differentials evaluated at HPL are greater than what CHORS would probably have experienced with their C_{18} method, but this cannot be ascertained because of incomplete record keeping at CHORS. In addition, it is not known if air handler failures caused CHORS to halt analyses. For comparison, at HPL a column temperature excursion of 0.8°C is considered a hardware failure and for each analysis conducted, column temperature is recorded.

3. Quality Assurance

A quality assurance plan (QAP) is a critical part of maintaining quality-assured results. A QAP describes QC measurements that are implemented on a frequent time scale and document, during the analysis of samples, the method is operating within expectations. After a sufficient number of QC analyses are conducted, the values are assembled into averages with 95% and 99% confidence limits, which are referred to as warning and control limits, respectively (WL and CL, respectively). The WL and CL values allow an analyst to be alerted when the chromatographic system is trending towards an out-of-control condition (e.g., when a QC parameter approaches a WL value) and to quickly determine when the system goes out of control (i.e., when a QC measurement exceeds a control limit). Such QC measurements are usually plotted as a function of time in a so-called *control chart* and are necessary to prove that expectations of performance identified during method validation hold true at all points in time during the analysis of samples.

Development of QC measurements is based on understanding uncertainties associated with variables in the calculation equation, such as injection volume, peak area (of standards, samples and internal standard), calibration factors, and extraction volume. The uncertainty associated with each of these variables can be assessed with appropriate QC measurements and, if the uncertainty in any one variable exceeds

performance expectations, then processing is halted and corrective action taken to assure the chromatographic system is, once again, in control. Sample processing is resumed once the system is operating within expectations.

A few examples of QC analyses within a QAP are as follows:

1. Daily analyses of Chl *a* calibration standards to validate the correctness of the Chl *a* calibration factor in use.
2. Daily injection of a mixture of pigments to document retention times (for purposes of pigment identification) and adequate resolution between pigments quantified.
3. Replicate injections of a sample extract at a time interval that describes the maximum amount of time a sample resides in the autosampler compartment prior to injection. This uncertainty is important to understanding the contribution of additional uncertainty associated with analysis of replicate filters and is also important to knowing daily analysis precision.
4. Duplicate filter analysis to describe overall method precision
5. Calibration of pipettes used for quantitative delivery of solvents (e.g. the internal standard solution, and for calibration dilutions).

CHORS did not present any data or evidence that any of these QC analyses were used during their analysis of HPLC samples with either the C_{18} or C_8 methods. Part of the reason for this omission was an inability of CHORS to provide a properly documented formulistic representation of their quantitation equation prior to the SeaHARRE-2 activity. Indeed, the exercise of producing the SeaHARRE-2 documentation resulted in the discovery that one of the CHORS quantitation terms was being applied twice. The bias this caused affected all pigment concentrations CHORS had quantitated up until that time and required on the order of a 5% correction.

CHORS participated in a variety of round robins from 2001–2007, inclusive. The intercomparisons covered a wide dynamic range in water types (coastal to open ocean) and TChl *a* concentrations. HPL was a participant in all of the intercomparisons, which were sponsored by the SIMBIOS (SB) and SeaHARRE (SH) activities: SB-1 involved eutrophic samples; SB-2 involved eutrophic and mesotrophic samples; SB-3 involved oligotrophic and mesotrophic samples; SH-2 involved eutrophic and mesotrophic samples; SH-3 involved oligotrophic, mesotrophic, and eutrophic samples; and SH-4 involved eutrophic and mesotrophic samples. The first three SeaHARRE samples were all collected in the open ocean, whereas the fourth SeaHARRE samples and the first two SIMBIOS samples were collected in coastal waters.

CHORS did not have a QAP and did not collect QC data, so a traditional control chart for the CHORS methods is not available. A proxy variable for a standard QC parameter can be computed, however, based on the RPD between the CHORS response factors with respect to the average value from all the calibrations for a particular pigment and method (the average is the reference value in the RPD calculations). Because CHORS executed two methods, there are necessarily two control charts, but their temporal distinction allows them to be presented together in Fig. 4. Although not presented in the same level of detail as the C_{18} method (Sect. 2.2.5), the C_8 method calibrations suffer

from many of the same problems, and an overall description of the calibrations is as follows: 49.6% of the original calibration data are for a pigment amount less than 15 ng, and 70.1% of the original calibration data less than 15 ng are for a pigment amount less than 2 ng. Consequently, the most significant amount of calibration data are for low concentrations.

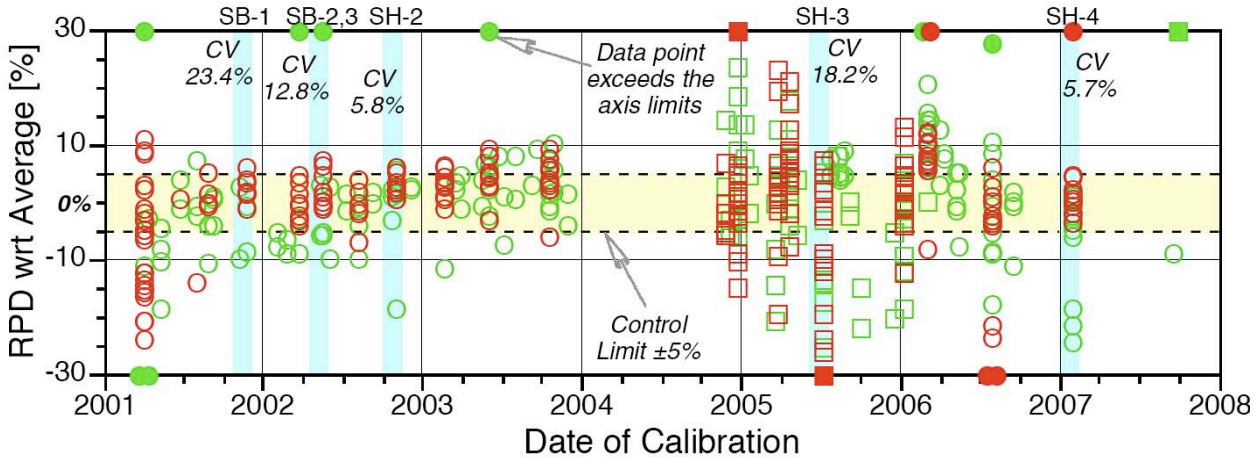

Fig. 4. A combined control chart for the CHORS C_{18} and C_8 methods (circles and squares, respectively). Points with deviations exceeding the *y*-axis limits are shown in solid. The time periods of the SIMBIOS (SB) and SeaHARRE (SH) round robins CHORS participated in are shown along the top and by the blue bands. Method precision during each round robin is shown by the CV entries closest to the blue band corresponding to the activity in question and range from 5.7–23.4%. The chlorophyll pigments are shown in green and the carotenoids in red. A reasonably attainable ±5% control limit is given by the yellow band and dashed lines.

To set a realistic boundary on what can be expected with cleaned up CHORS data, the CHORS calibrations for the control chart presented above are based on recomputing the calibrations to remove the most egregious problems caused by low concentration noise, the original manual editing to the data, and fitting errors. The other important improvement to the calibrations was removing the use of Chl *b* as an unproven relative RF to calibrate the carotenoids (discussed in more detail below)—*all of the calibrations are physical calibrations*. The data, therefore, are as good as can be expected without engaging in a calibration-by-calibration correction process and are significantly cleaner than the original CHORS RFs.

Like the CHORS calibration plots shown earlier (Sect. 2.2.5), the CHORS control chart exhibits an excessive amount of variance—almost half the data (46%) exceeds a ±5% control limit (the yellow band). Note the significant improvement in CHORS precision from their participation in round robins in 2001–2002, the degradation in precision associated with the incomplete attempts to bring up a C_8 method in 2004–2006, and the return to a more acceptable precision when the C_{18} method was readopted. Precision is

only part of the criteria used to evaluate a method, however, and accuracy, which is discussed in Sect. 5, is arguably more important.

The most dramatic aspect of the CHORS control chart is both the C_{18} and C_8 methods were never in control—all pigment values within ±5%—although, the C_{18} method was almost in control on a few occasions. For the latter, a period of superior performance is seen during the time period associated with the SeaHARRE-2 analyses, wherein almost all of the pigments are to within ±5% (the lone chlorophyll pigment at approximately -20% is Chl c_2). This small amount of encouraging performance is moderated by the incomplete calibration history of the pigments: *on average and ignoring 2004, because the C_{18} method was not used, 25% of the full C_{18} pigment set was calibrated less than once per year*.

4. Detector Nonlinearity

The CHORS UV6000LP detector contains a flow cell (US patent 5,608,517) with a thin polymer to pipe light down the flow cell (Fig. 5), which provides an optimal response in the ultraviolet domain (190–300 nm). 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. Anything influencing the absorption of the fluid in the flow cell (e.g., the solvent system being used or the pigment load) will necessarily influence the nonlinear response. 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. For the analysis of marine pigments, this means the nonlinearity effects are greatest for the pigments quantitated with red wavelengths and are exhibited to a lesser—but still significant—degree for pigments quantitated with blue wavelengths.

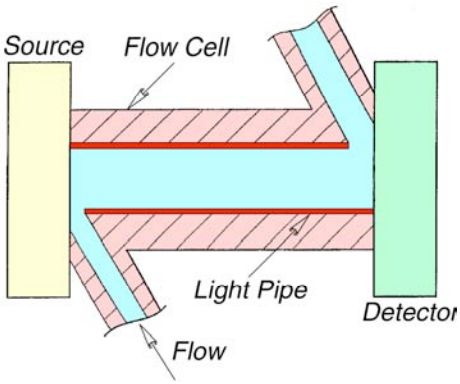

Fig. 5. A schematic of the UV6000LP flow cell, which is lined with a thin polymer (dark red) to pipe light down the flow cell (the LP designation indicates the light pipe).

The CHORS C_8 and C_{18} methods used very similar calibration procedures and a common set of detector wavelengths for quantitating carotenoids (450 nm) and the Phide *a*, Chlide *a*, and Phytin *a* degradation pigments (664 nm). There are differences for detecting the chlorophyll pigments, however. For the most important pigments, which directly or indirectly comprise the primary pigments, the differences mostly affect TChl *a*, because the final concentration is the sum of MVChl *a*, DVChl *a*, and Chlide *a*, and one or more of these constituents are detected at a red wavelength (664 nm). For the C_8 method, all of the TChl *a* constituents are detected with a red wavelength, so the

importance of the nonlinearity is expected to have a maximum effect. In comparison, the quantitation of TChl *a* with the C_{18} method involves mostly blue wavelengths; the exception is Chlide *a*, but it is usually the smallest contributor to TChl *a*, so the detector nonlinearity is expected to have a lesser effect.

CHORS calibrations were 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 CHORS calibrations are not forced through zero, the residuals exhibit a much stronger nonlinearity and rarely fall within the expected range of ±2%, as shown in Fig. 6. The calibrations for the two methods are somewhat 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). The spectral characteristics of the nonlinearity are not immediately as expected: the bluest and reddest wavelengths do not always have the smallest and greatest nonlinear response, respectively. As noted earlier, the nonlinearity depends on the pigment load. A subsequent analysis presented below will show how restricting the calibration range can change the nonlinearity and produce a response more in keeping with the generalized description first provided.

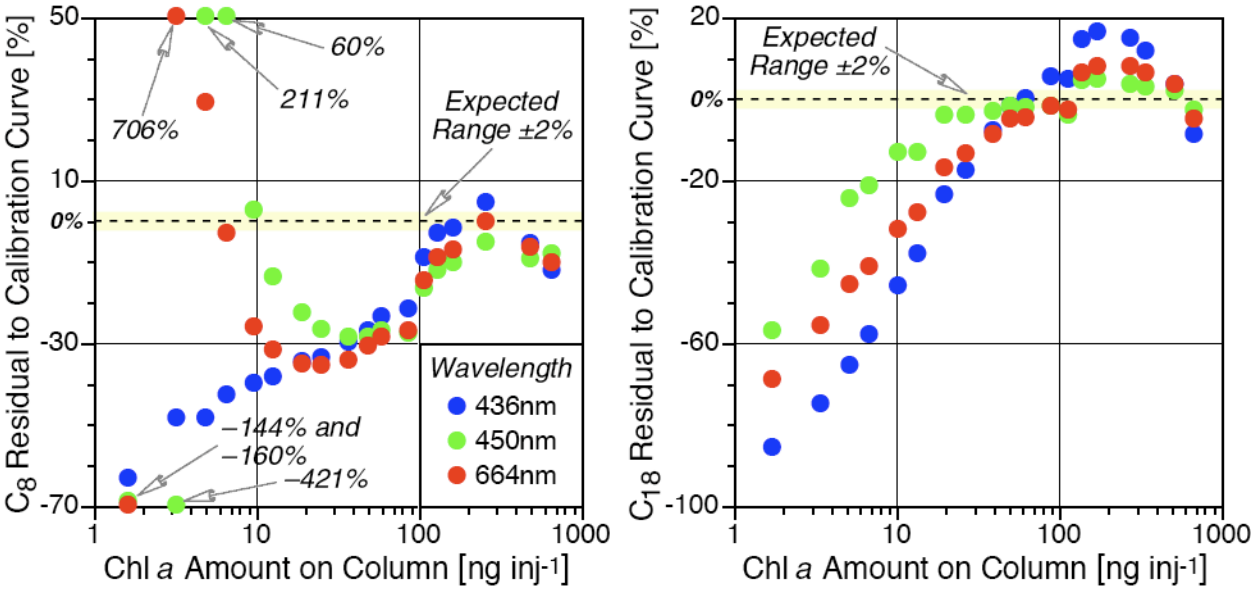

Fig. 6. The calibration residuals for the CHORS C_8 and C_{18} methods (left and right panels, respectively) based on a 2–200 ng working range and the wavelengths used for quantitation. The expected ±2% range for the residuals is shown by the yellow band.

The aforementioned "aberrant" CHORS calibration data points (Sect. 2.2) were first revealed during the high-resolution C_{18} calibration performed at CHORS after The Team visit in August 2007 (Sect. 7). The plots in Fig. 7 show this "outlier" behavior is seen in both the C_8 and C_{18} calibration data, which are rather similar in most respects: a) there are very few triplicate injections for which all three values (black circles) agree to within

2% (the yellow band in the plots); b) for those triplicates with an outlier, excluding the outlier from the average of the remaining duplicate usually results in a precision close to or within 2%; and c) the deviations of the outliers with respect to the average value of the duplicates are also plotted and can be very large with the positive excursions being the most notable in both cases (the outliers have similar magnitudes at all three quantitation wavelengths). It is important to note the outliers are primarily detectable, because these data are for calibration standards done as replicate injections—they would not be distinguishable in in field data.

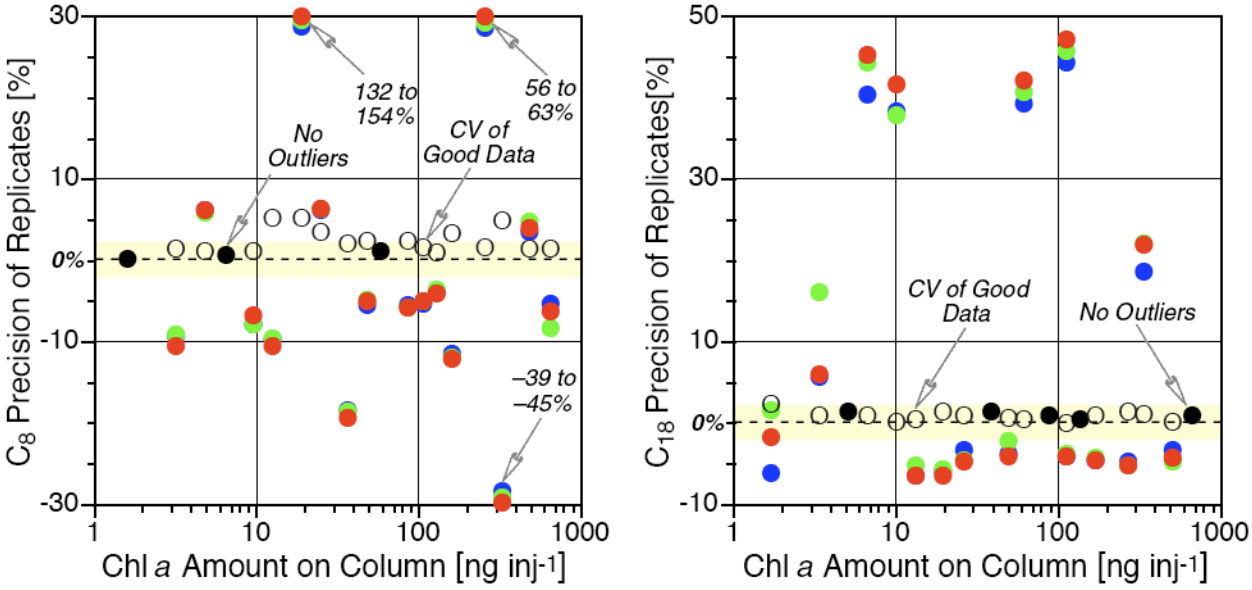

Fig. 7. The calibration precision (expressed as the CV) and "outliers" in the replicate injections for the CHORS C_8 and C_{18} methods (left and right panels, respectively). The precision of the replicates which had no outliers and with the outliers removed for the injections that had outliers are shown by the solid and open black circles, respectively. The outliers are shown as the colored circles and are seen as outliers at all three quantitation wavelengths when they occur (blue, green, and red correspond to 436, 450, and 664 nm, respectively). The yellow band shows the expected ±2% range in the precision for a quality-assured laboratory.

The sign of the outliers (positive or negative) is assumed to correspond to carry over or carry under, respectively, with respect to pigment concentration. That is, in the case of carry over, extra pigment material is being released, which increases the peak area and the concentration of the pigment; whereas for carry under, pigment material is being retained, which decreases the peak area and concentration of the pigment. In both cases, the mechanism(s) for retaining or releasing pigment material is not identified, because it is not known at this time. In addition, the two phenomena might have different explanations, e.g., the autosampler in one case and the type of column being used in the other.

5. Update

The CHORS technician resigned from the HPLC Team and terminated employment with San Diego State University (SDSU) in May 2008—6 months after being contracted to work on the correction process for at least one year, with renewable six-month options—so The Team had to spend the time available before the technician left getting all the CHORS data and laboratory notebooks into a single archive for eventual shipment to HPL or GSFC. Much of this work was anticipated in the original correction plan, but it had to be accelerated and executed over a shorter time period to meet the technician departure schedule, which means it significantly interfered with other aspects of the planned correction activity.

HPL had not wanted to be involved with the CHORS correction problem but did so beginning in May 2007, because an expert pigment analyst was needed. The enormity of the problem, however, was overwhelming HPL by April 2008, and because HPL personnel were concerned that sample processing in fulfillment of the HPL contractual obligation with NASA for annual pigment analyses would be inappropriately delayed, HPL stopped participating in the CHORS correction in May 2008 (with the exception of assisting in the preparation of this document).

The other activity that was completed right before the CHORS technician left was a parallel analysis of calibration samples with an unequivocally linear detector (a Waters model 2998), and this revealed a new source of nonlinearity in the CHORS HPLC equipment: most likely the autosampler. The latter can only be confirmed with much more thorough testing, so given the HPL and CHORS departures from The Team, GSFC researched and ordered an HPLC system in July 2008.

The review presented at the combined CC&E and OCRT meeting established a community consensus to correct at least the TChl *a* data. A firm schedule for this work could not be established, because of the loss of so many members of The Team. In addition, Greg Mitchell revealed an extensive inventory of duplicate filter samples stored in liquid nitrogen. A subset of these samples was established for future analyses to test the efficacy of the correction scheme, if one could be formulated.

With CHORS no longer participating in the correction process, all of the CHORS HPLC equipment was sent back to the government agencies that originally funded the procurement of the equipment (NOAA and NASA). The NASA components were received at GSFC in June 2008 and constituted the majority of what was being used to maintain a working system at that time. Basically, one system was kept up and running, and components from the other system were used to keep the selected system functioning. The only operational controller belonged to NOAA, and CHORS originally planned on sending it to NASA, but CHORS decided to send it back to NOAA after the NASA equipment was packed by the CHORS technician before the last day of employment at CHORS. NOAA agreed to send the controller to NASA, but it did not arrive until August 2008. Unfortunately, the CHORS HPLC equipment could not be set up, because the CVO Laboratory for Analyzing Bio-Optical Samples (C-LABS) was still

not approved by GSFC for the use of HPLC equipment even though the work order to make the laboratory compliant was submitted 31 October 2007.

The GSFC HPLC system to be used to characterize the CHORS system was delivered in September 2008, but the GSFC laboratory modifications were still not complete, because the GSFC facilities department had not been able to properly define and install the required venting modifications. NASA HQ and GSFC management decided the CVO had to relocate to an off-site facility to get a functional laboratory in September 2008. Although the CVO located a suitable site with compliant safety requirements in October 2008, delays from GSFC procurement and contracting officials delayed the CVO move to the new facility until 6 January 2009.

When the CHORS HPLC system was set up for the first time in early 2009, it was discovered that the autosampler lifting arm had been broken during shipment from CHORS to GSFC (most likely the repacking that was done to remove the NOAA controller compromised the original careful packing that was done by the CHORS technician). Replacement parts were obtained by the CVO in February 2009 by purchasing a used autosampler in an e-Bay on-line auction for $81 (plus shipping and handling). Basic functionality of the NASA part of the CHORS HPLC system was established in March 2009.

During the time period after the loss of both CHORS and HPL participation in the correction work being done by The Team, the CVO continued to investigate the nonlinearity problem while dealing with moving the CHORS equipment and trying to set up a functional laboratory. The synthesis of these more recent nonlinear investigations involves combining the plot of detector nonlinearity in terms of calibration residuals for the two methods (Sect. 4) with the concept of a sensible working range (2–200 ng) for the worldwide analyses CHORS was performing (Sect. 2.2.5).

Figure 8 shows the calibration residuals to all the high-resolution data for a linear calibration (not forced through zero) applied to the data corresponding to the working range. A relaxed ±5% range is shown in the plot to evaluate the maximum amount of data that might be found sufficiently suitable to proceed with this reduced—but appropriate—set of calibration data (remembering that a small number of data points with residuals in excess of ±2% is permissible, but a large number is not, because an overall average to within 2% must still be achieved for quantitative analysis). The C_8 data residuals are almost all outside the ±5% range for the residuals and exhibit a strong nonlinear structure. There is evidence of a possibly linear range from about 20–80 ng, but the data above and below this interval are clearly following a different functional form. Although many of the C_{18} data residuals are within the bounding box, they are not randomly distributed and exhibit a steady increase in residuals as the concentration decreases; at higher concentrations, the residuals are initially within the relaxed range, but then decrease to increasingly negative values as the concentration increases. In terms of the sources of uncertainty already discussed (internal standard, control chart, outlier injections, etc.), the uncertainty from the nonlinear response of the flow cell is of a similar order of magnitude.

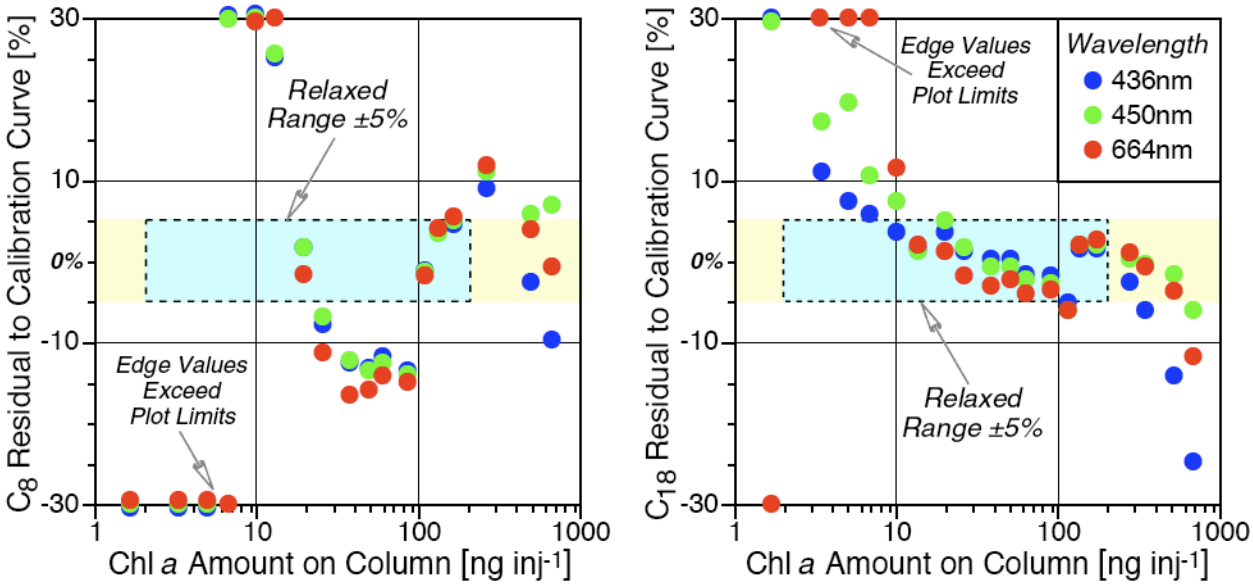

Fig. 8. The calibration residuals to all the high-resolution data for a linear calibration (not forced through zero) applied to the data corresponding to the 2-200 ng working range (delimited by the blue box bounded by the dashed outline). The full range of the data is shown by the yellow band extensions.

It is possible to establish piece-wise contiguous functions that better approximate the nonlinearity in the 2–200 ng range at a particular point in time. For the C_8 data above, for example, three linear segments with intersections at 19.5 and 74.9 ng, significantly improve the calibration. Unfortunately, the other sources of uncertainty render this effort moot, because they make the problem of determining a calibration and applying it to the original analyses multivariate. In this multivariate space, an effort to modify one variable without understanding its relationship to the others frequently yields contradictory lessons. For example, in one case, a seemingly sensible approach reduces the variance in a result (e.g., the residuals in a calibration), but when it is applied to another apparently appropriate case, the variance increases, because the causal linkages between the other variables are not understood and one of them is now predominant. Many of these variables are being approximated by creating proxies for what needed to be measured, but was not measured. This means a significant amount of time is spent trying to establish proxy variables and most eventually prove inadequate.

6. Conclusions and Recommendations

A summary of the CHORS uncertainties associated with the round robins they participated in is presented in Table 2. Uncertainties exceeding the allowed maximum for SeaHARRE quality-assured results are shown in red typeface. The SeaHARRE-2 results have an average uncertainty of 21.1%, which are the only results within the requirements for quantitative analysis as defined by the SeaHARRE community (and are highlighted in blue); all the other round robins show the CHORS methods did not

produce quality-assured analyses and sometimes by such a large margin, they cannot be considered of routine research value as defined by the SeaHARRE community.

Table 2. A summary of the CHORS uncertainties associated with the round robins they participated in. The SeaHARRE-3 results highlighted in orange are for the C_8 method; all other results are for the C_{18} method. Values shown in red exceed the requirements for quality-assured results (15% for TChl *a* and 25% for all other pigments).

Data	TChl *a*	TChl *b*	TChl *c*	Caro	But	Hex	Allo	Diad	Diato	Fuco	Peri	Zea
SB-1	19.8	53.6	29.4	91.5	22487.7	4618.1	22.3	26.0	30.4	11.9	28.3	22.3
SB-2	9.4	29.6	42.2	20.7	32175.8	2568.6	32.4	16.5	26.2	16.4	37.4	24.0
SB-3	20.6	31.7	34.7	46.5	6.3	13.2	33.7	20.6	53.8	13.6	19.0	21.5
SH-2	4.5	13.4	35.2	18.8	57.3	7.2	28.3	11.0	24.6	4.3	17.4	30.8
SH-3	16.5	17.2	11.1	57.4	23.3	19.7	20.3	7.6	182.4	32.1	40.6	18.4
SH-3	57.4	70.8	29.0	15.2	14.9	9.1	18.9	27.4	38.1	15.7	15.3	7.5
SH-4	34.1	23.9	32.3	67.3	2750.0	3067.5	5.1	13.6	30.6	7.5	284.1	55.1

Even if the most problematic pigments are ignored (But and Hex), the only time period when the CHORS results are acceptable for NASA calibration and validation activities is SeaHARRE-2 (remembering that omitting even one primary pigment when assessing the capabilities of a particular method is not permitted, but the huge uncertainties with But and Hex makes this a recurring question). In addition, the compliance rate for the number of pigments CHORS quantitated at the quality-assured level is no better than 67% (SeaHARRE-2) and is as low as 25% (SIMBIOS-1). Perhaps most troubling are the SeaHARRE-4 results, because for that activity, CHORS was completely aware of all the problems they were having with the C_{18} method and the results are rather similar to SIMBIOS-1.

Although quality-assured laboratories do have individual pigments with higher uncertainties than the others, they are restricted to a small number of pigments, and the overall average uncertainties are almost always within performance metric requirements. For example, the average uncertainty for the last three SeaHARRE round robins is 19.2% for the quality-assured laboratories. The corresponding CHORS value for the C_{18} method is 196.4%, and if But and Hex are ignored (to get a sense of how the rest of the pigments are performing), the value is 38.2%.

One of the most challenging aspects of the CHORS uncertainties is the changing patterns of compliance and noncompliance, with no one pigment always within quantitative analysis requirements, although Diad, Fuco, and Zea come close. This behavior is reminiscent of the variability already seen in the CHORS RFs (Sect. 2.2.5) and control chart (Sect. 3), which is to be expected, because one is inexorably linked to the other. Another problem with the CHORS results is the variability in the magnitude of the uncertainties from one round robin to the next. Again, only Diad, Fuco, and Zea are close to exhibiting the kind of stability typical of a quality-assured laboratory. A method that is in control will not have the kind of variability and swings in magnitude seen in the table above.

At this point, it is instructive to review a listing of the problems associated with the CHORS HPLC analyses, before making recommendations about how to proceed. This listing is taken from the discussions above, earlier documents already made available to the community, and details presented in the chronology (Sect. 7), with the most significant sources of uncertainty shown in bold typeface:

1. **CHORS did not validate the C_{18} or the C_8 method before either was placed into service to analyze field samples for NASA PIs**;
2. **There was no QAP and no daily, weekly, or monthly QA measurements** (except for Chl *a* and Chl *b* calibrations, which occurred about once every 1–2 months);
3. **There were no quantitative QA variables, so important parameters were characterized qualitatively** (e.g., mixed standards were used to determine retention time repeatability and chromatographic variables or other performance criteria were deemed satisfactory if they "looked good" or were "OK");
4. **There is a significant and unknown source of outliers in the quantitation process with approximately one-third of all triplicate injections exhibiting anomalous responses, many of which have significant deviations** (the outlier can more than 100% different than the remaining duplicate) and this problem was also seen in older data collected with a UV2000 detector, so the problem is not exclusive to the UV6000LP system;
5. **The UV6000LP has an optimized response in the ultraviolet (190–300 nm) domain for greater sensitivity, and a nonlinear response outside this range** (the nonlinearity increases from the blue to the red part of the spectrum), **but CHORS assumed the detector had a linear response for all wavelengths used to quantitate marine pigments with both the C_8 and C_{18} methods** (436, 450, and 664 nm);
6. Two international scientists who use a Thermo system with the UV6000LP detector have confirmed the nonlinearity problem, and one worked with a local Thermo technician to verify narrow linear response regimes can be found and properly calibrated, but it requires procedures CHORS rarely used, and when they were used, they were not fully implemented;
7. **A nonlinear response is also seen by an unequivocally linear detector (e.g., a Waters 2998) when it is placed in-line with the UV6000LP, so there is at least one other significant source of nonlinearity**;
8. **The detector has a substantial refractive index (RI) problem with the flow cell**, and the Thermo software does not provide an automated procedure to apply an RI correction, so the correction must be done manually (the RI correction improves the chromatograms and reduces the nonlinearity, but it is too labor intensive to implement retroactively for the thousands of chromatograms involved with the CHORS data set);
9. **Adequate baseline stability was not achieved with either the C_8 or the C_{18} methods** and was a continuing source of performance difficulties;
10. Pre-scored septa, which can improve draw-volume accuracy on HPLC vials, were not used when the viscosity of the solutions required it (i.e., during the use of the C_8 method), and experiments at CHORS by The Team revealed the

chlorophylls had poorer precision than the carotenoids in general, but changing septa improved precision by more than a factor of two on average;

11. Although detailed in some respects, the laboratory notebooks give very little information about routine changes in methodologies (e.g., extractions, changes in solvents, methods, standards, etc.);
12. The primary emphasis at CHORS was to run samples and not to maintain an on-going synthesis of method performance variables as a function of time (the oldest part of the archive was stored on a tape format that is no longer supported, so at some point it became physically impossible to maintain a firm connection with past performance);
13. The CHORS laboratory notebooks document how the system was plagued by an excessive number of errors, shutdowns, freezes, jams, restarts, and mechanical problems over the 1998–2007 time period (Sect. 7)—**the number of problems plus the amount and type of needed maintenance greatly exceeds what is normally seen in a quality-assured laboratory**;
14. Critical equipment was not on a maintenance agreement, so the CHORS technician became the primary service provider (Thermo service in San Diego was considered inadequate, but no professional alternative was investigated);
15. According to CHORS, the manufacturer has considered the HPLC system rather old for some time and is not able to fully maintain it, although Thermo claims otherwise (the HPL HPLC is older than either one used by CHORS and with proper maintenance and quality assurance oversight, this demonstrates that quality-assured results can be produced on older hardware);
16. Numerous power problems were identified during the course of the HPLC analyses from 2001–2007, but **there is no evidence any substantial power corrections were requested from or implemented by the San Diego State University (SDSU) facilities department** (the hosting organization for CHORS);
17. **The HPLC equipment was not connected to an uninterruptible power supply (UPS)** until The Team had one purchased in August 2007, so sensitive electronics were subjected to repeated and potentially damaging power surges, voltage reference fluctuations, and brownouts;
18. For the majority of the time, CHORS kept one HPLC system operational by swapping parts and components from a duplicate system (plus some occasional repairs from qualified professionals);
19. **When important parts or components were altered, repaired, or swapped out, the consequence on method performance was not quantified**—the recurring assessment was primarily qualitative (e.g., “looks good” or is “OK”);
20. **The range of concentrations expected in samples was not fully described by the calibration standards** (e.g., the linear range was not identified and a working range was not established);
21. CHORS did not monitor whether or not a sample injection was too concentrated, so if a sample contained a particularly large amount of pigment, it was not diluted and reinjected to ensure quantitation within a sensible working range;
22. **Routine linearity checks were not performed**;

23. **Duplicates from the PIs submitting samples were not required or enforced**, even though the field sampling protocols require the submission of duplicates (as part of the QA and QC requirements for good laboratory practices);
24. **From the very first calibration in April 2001, a) the calibration curves were not linear, but this went unnoticed, because the percent residuals to the calibration curve were not computed, and b) the *y*-intercepts were very large, but this went undetected, because all calibrations were forced through zero without inspecting them**;
25. **A time history of calibration response factors was not maintained and continuously evaluated once new calibrations were done, so the variability in calibrations was unrecognized**;
26. **Calibration curves were based on dilution intervals and qualitative response rather than a quantitative knowledge of the amount of pigment injected as a function of the corresponding peak area in the response of the system**;
27. The system was not calibrated for the measurement of many of the chlorophylls on a regular basis (e.g., for the C_{18} method from 2001—2007, Chlide *a*, Phytin *a*, and Phide *a* were calibrated four or fewer times each);
28. **The system was not calibrated for the measurement of the carotenoids on a regular basis, instead a theoretical calibration was used** (a physical calibration with Chl *b* was performed and then all other carotenoid calibration factors were computed based on previous observations of the relationship between Chl *b* calibration factors and those of the other carotenoids);
29. **It was never demonstrated that the Chl *b* calibration factors and its presumed relationship to all the other pigments were stable, accurate, or reproducible**;
30. A column heater was not used with any of the C_{18} analyses, so column temperature was basically the same as room temperature, which was not stabilized at the level needed for sensitive laboratory equipment, because of inadequacies with the air conditioning and building power systems;
31. **Calibration points were deleted based on visual inspection without supporting evidence based on statistical analyses** (e.g., percent residuals with respect to the fit);
32. **At different points in time, pigment quantitations were executed using automated integration features of the ChromQuest software**—the chromatograms were not individually inspected and quantitated in all cases, because "there was not enough time" to do so—the emphasis was getting samples processed (it is not known how many samples were subjected to automated integration);
33. **For both the C_8 and C_{18} methods, the internal standard had poor peak area precision**;
34. For the majority of the samples analyzed with the C_{18} method, the internal standard was a naturally-occurring pigment (Cantha), although it is an infrequently-occurring pigment;
35. In round robins, there were numerous occasions where CHORS reported pigments present in filters (and sometimes in great abundance) when in fact the

pigments were not present, and some pigments were reported absent when they were indeed present; and

36. **The frequent instrument problems suggest the analysis of the pigment extracts were frequently interrupted, which means they had to be stored—and, thus, were degrading in an undocumented fashion—while the problem at hand was addressed** (the uncertainties this causes will have a strong dependency on the types of pigments being quantitated and their original concentrations).

This is an extensive list of sources of uncertainty, and although all complicated analyses like HPLC will have issues that need to be addressed, a quality-assured laboratory is not expected to have a list this long containing so many practices that degrade method performance and that are not being properly addressed over time. Many of these entries are interconnected, so it will not be easy to isolate a problem and resolve it, because the entire approach quickly becomes multivariate. Adding to the difficulty are the connections to problems requiring extensive archival investigations, for example, estimating the frequency of the false-positive and false-negative reporting would require inspection of all peaks in all chromatograms with no guarantee that quantitation of pigments so affected could be improved (e.g., when a false-positive result is caused by co-elution problems).

It is important to remember the extent of the problem set has primarily been identified using the Chl *a* results, because there is the most extensive amount of data for this pigment. The other reason for looking at Chl *a* is it is arguably the easiest pigment to quantitate, because a) it is in every sample, b) it is usually the most abundant pigment, and c) it is a late-eluting pigment in all the methods discussed in this document (so the baseline is usually the simplest to interpret). With the exception of Chl *b*, all of the other pigments have much less data to work with and are more difficult to quantitate, so the problems and difficulties with interpreting the results will be at a minimum for Chl *a*—all the other pigments will be worse, and in some cases, significantly worse (for example, degradation pigments like Phytin *a* and Phide *a* are difficult to calibrate under the best of circumstances, because high-purity standards are not always available for purchase).

The list of problems given above documents why the CHORS system was almost always out of control: *there were so many sources of variance, the response of the system was going up and down on multiple time scales and amplitudes*. In such an environment, variances can sum or cancel depending on the direction they drive the response of the system. This means it is possible for the data to comply with the performance metrics of a quality-assured laboratory during one time period and then be noncompliant during another point in time—which is what has been observed. Compliance requires an alignment of many pluses and minuses, so it is rarely going to happen, but it can occur and was observed during the SeaHARRE-2 time period.

It is possible, therefore, that some PIs will have data that is probably suitable for calibration and validation activities, but most will not. In addition, this will have to be

applied and evaluated separately for each pigment, so the dimensionality of this statement is very large. Unfortunately, there is no single all-encompassing parameter that can be computed to figure out which of these two states applies to each PI and each pigment of interest. In the absence of such a parameter—and remembering that a separate determination is needed for each pigment—the results simply bounce back and forth between two end-member states of variance which have significant amplitude, as shown in the variability already documented with the internal standard (Sect. 2.2.1), the normalized RFs (Sect. 2.2.5), the control chart (Sect. 3), injection outliers (Sect. 4), and flow cell nonlinearity (Sect. 5). In most cases, the individual sources of variance span the ±10% to ±40% range, but, unfortunately, they also appear to go beyond those limits in many recurring situations, for example, instrument performance anomalies and pigment decay from hardware failures, false-positive and false-negative reporting, spurious peak area integration from automated integration tools, etc.

The comforting aspect of this variance is it does not appear to have a trend—in fact, many of the variance plots do not exhibit trends—so large-scale averages of the results might very well be suitable for a variety of inquiries. For example, the OC4 algorithm does not exhibit large-scale changes if the CHORS data are included in the derivation of the fitting terms or not. There are differences at the lower and upper ends of the fit, but this is to be expected because there is always a sensitivity at the end points from the effective boundary conditions of the nonlinear fitting which are dependent on changes in data distribution.

In terms of working towards a resolution of the CHORS HPLC data, it is important to remember *the ocean color community relies on many data sets for which the uncertainties are unknown*—and those data are still used without very much visible expression of concern or discomfort. What distinguishes the CHORS problem is the uncertainties are rather well known and they exceed what is expected. This new horizon of understanding triggers two actions: a) dissemination of that knowledge base (this report), and b) an agreed upon policy as to how the data can be used in the most appropriate manner (given below). It is expected that analytical procedures will evolve and improve over time, and that the need for uncertainty estimates will be honored. It is fully expected, for example, that retrospective analyses of HPL data will expose vulnerabilities and weaknesses. The important point is the appropriate method validation and QAP procedures are in place, so the quality of the data generated at all points in time can be evaluated.

The lack of method validation and a QAP, coupled with an absence of commonly achieved performance criteria, is not a unique failing—the SeaHARRE activities have documented such failings on more than one occasion. What is unique, however, is CHORS experienced that failing while being responsible for the pigment analyses of a very large research community. Compounding the difficulties was a lack of emphasis on proper training of personnel. It has been common practice for many SeaHARRE participants to attend chromatography courses offered by academia and the private sector, and to attend SeaHARRE workshops. CHORS

rarely did this. The primary emphasis at CHORS was on production without understanding the quality of the product being produced (numerous examples are presented in Sect. 7).

The CHORS technician did an extraordinary amount of work, but much of the effort did not focus on critical issues, such as finding the source of the poor precision exhibited during calibration or with the internal standard, conducting needed statistical analysis of the results, and identifying how laboratory practices influenced the statistical parameters. Investment in a UPS and a service contract with a quality company could have significantly improved the outcome. An effective and efficient plan for understanding sources of uncertainty is critical to achieving high-quality quantitative data. There is always a steady growth in uncertainty from the onset of establishing a pigment analysis (i.e., calibrating the first pigment) and submitting the final spreadsheet to the PI. For a quality-assured laboratory, uncertainties are basically held to 5% or less in order to have a fully intercompared average uncertainty of the primary pigments to within 25%. If the starting point for a laboratory involves uncertainties of 10—40%, the inevitable additive effects will quickly render the data rather useless for climate research or calibration and validation activities.

Taking all of these elements into consideration and remembering that it is the nature of science that the next generation of work builds upon the previous and inevitably improves upon it, it is the recommendation of The Team to do the following:

1. *Do not attempt to correct the data*. Correction is going to require significant inquiries into numerous (more than 30) sources of uncertainty with no reasonable prospect of successfully parameterizing and understanding most of them, because the needed QA and QC data were not collected. Furthermore, from a global database perspective, the data in question are being replaced by a steady stream of quality-assured analyses from other sources (e.g., HPL), so the uniqueness and importance of the CHORS analyses are diminishing over time.
2. *Put all the data that was removed from SeaBASS back into the database* (data from CHORS, PML, and MCM were taken out of SeaBASS, because they all used the UV6000LP detector).
3. Label the data associated with CHORS and PML analyses with an appropriate warning, e.g., *"These data are not validated and should not be used as the sole basis for a scientific result, conclusion, or hypothesis—independent corroborating evidence is required"*.

The last point requires some additional explanation. The only assessments of PML data quality determined PML was not producing quality-assured results. In addition, PML is not interested in participating in an investigation as to the extent of the problems associated with their use of a UV6000LP detector. MCM has been a reliable producer of high-quality data (as measured by SeaHARRE round robins), with the exception of a time period when extensive equipment failures degraded the results. Furthermore, MCM worked with The Team to confirm and clarify many

aspects of the nonlinearity problem and documented how the nonlinearity was being properly minimized during MCM calibrations.

The lessons learned and the recommended future directions must extend beyond NASA databases and include the requirements for next-generation missions and the maintenance of climate-quality data records (CDRs), which will require *in situ* data with unprecedented quality. Within that context, the CHORS problem is a wake-up call and NASA must implement processes to improve data quality. In a summary statement about the extent and importance of the CHORS problem, Dr. John Dolan noted the following (8 October 2007):

> *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*.

He went on to elaborate that 25–30 years ago, the FDA and EPA implicitly trusted laboratory data, but then they realized 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, record keeping, as well as proof of proper maintenance, calibration, and change control of instrumentation.

Those who lived through this process of improving the quality of FDA and EPA laboratory analyses usually grumbled at the requirements, but in retrospect, almost everyone—from the analysts to the laboratory directors—viewed 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 significant, 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 for NASA does not have to start from scratch.

The CHORS HPLC problem represents a case study in how undetected low- and high-level mistakes can have a significant and negative impact on the quality of an entire program. The HPLC experts tasked with correcting the CHORS HPLC problem stress the following conclusions for improving HPLC analyses (but they are rather easily adapted to a wide diversity of measurements):

- 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 practices;
- A QAP, with well thought out QA and QC data, must be implemented by personnel who are properly trained;
- 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 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; and
- Oversight by NASA should not rely too heavily on peer reviews—outside experts from related fields need to be represented.

The last two points require some clarification. The point is, regardless of what CHORS did incorrectly, their proposals were peer reviewed, as was their attempt to understand their C_8 problems once they were notified about them, and they were subjected to NASA oversight as part of the contract reporting process. None of those procedures, which are all associated with quality assessment, correctly identified CHORS analyses were significantly degraded or correctly identified the source of the problems.

Drawing now on the accomplishments of all the round-robin activities NASA has sponsored, and adding in the specific problems of the CHORS analyses, several important recommendations can be made for implementing a QA capability for CDR analyses:

1. 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 persistently evaluated).
2. 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).
3. All analyses for CDRs must have a 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).
4. Programmatic or project oversight is needed to ensure inspections and compliance with the QAP.
5. For those activities funded by the OBB Program, mandatory workshops with laboratory certifications should be conducted annually or every two years for any laboratory and technician that is conducting CDR analyses (HPLC analyses are already compliant).
6. To ensure proper control and review procedures are in place for all analyses essential in the production of CDRs, a panel should be convened to make recommendations to NASA about a) implementing an oversight process with specific guidelines, and b) strengthening the peer-review process.

7. 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.
8. 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 are implemented.

Although implementing these recommendations constitutes a very large undertaking, the present is an ideal time, because the community is between two important horizons of sensors and science: the end of the SeaWiFS and MODIS era, and the start of the ACE, GEO-CAPE, and HyspIRI era.

7. Chronology

A summary—but still detailed—chronological description of what transpired with the CHORS HPLC problem is constructed from the CHORS laboratory notebooks, presentations given to the community, debriefing documents and reports submitted to NASA HQ, personal notes of The Team members, and testimony of the scientists involved in the investigations. The material was condensed and edited by the authors (although some of the full-length documents are available on line at the following Web site: http://oceancolor.gsfc.nasa.gov/DOCS/). Color codes are assigned to the information sources as follows (gray denotes a general point or accomplishment important to the generalized problem):

CHORS PI (Chuck Trees)
CHORS Technician (Jason Perl)
HPL Chromatographer (Laurie Van Heukelem)
HPL Analyst (Crystal Thomas)
Expert on HPLC Chromatography (John Dolan)
Expert on HPLC Detectors (Ron Farnbach)
CVO Director and The Team leader (Stanford Hooker)

The chronological format allows readers to skip ahead to a particular point in time when they last understood the problem, and then reading beyond that will bring them up to date. If a reader is primarily interested in what was occurring for a particular identified group, reading only the corresponding colored text will bring the reader up to date with respect to what that group was doing.

1998

March CHORS adds a Thermo Separations Products UV6000LP photo-diode array (PDA) detector purchased using funding from NOAA in support of the MOBY activity to their existing Thermo HPLC system (hereafter referred to as the NOAA system). The rest of the system is composed of the following: a SN4000 controller, an SCM1000 degasser, a P4000 pump, an AS3000 autosampler, an FL3000 fluorometer, and a UV2000 detector. A primary reason for acquiring the new detector is the promise of

"increased sensitivity (factor of five over other detectors) from the use of a 50 mm flow cell and reduced flow cell volume (2 mL per 10 mm of path length)." It is hoped that "the large HPLC volumes filtered previously in oligotrophic waters (4–6 L) might be reduced to improve filtration time and enhance detection of some of the minor pigment compounds."

April HPL completes the HPLC sample analysis for the fourth Atlantic Meridional Transect (AMT) cruise. The first two AMT cruises were analyzed by CHORS with help from the Plymouth Marine Laboratory (PML) in Plymouth, United Kingdom (PML initiated and leads the AMT activity); and the third cruise was analyzed by the Marine and Coastal Management (MCM) group in Cape Town, South Africa. There are scheduling and logistical problems with continuing the analyses with MCM, and in the case of AMT-1 and AMT-2, there were problems and omissions in the data set, so an alternative HPLC laboratory was sought, and HPL agreed to analyze the AMT-4 samples.

December CHORS begins using the new UV6000LP detector with the NOAA system. The UV2000 detector is kept in-line as a second detector.

1999

April The concept of an HPLC round robin is proposed to *Laboratoire d'Océanographie de Villefranche-sur-mer* (LOV), HPL, and the Joint Research Centre (JRC); all three agree to participate. The MCM group is contacted as a fourth member, because they are sometimes doing the HPLC analyses for the AMT activity.

May The sampling protocols and basic procedures for an HPLC round robin are drafted at the *Productivité des Systèmes Océaniques Pélagiques* (PROSOPE) cruise planning meeting held at LOV.

July CHORS orders a second Thermo Separations Products HPLC Spectra System using funding from the SIMBIOS project (hereafter referred to as the NASA system). The system is composed of the following: P4000 pump ($8.4K), UV6000LP PDA detector ($9.1K), SN4000 controller and AS3000 autosampler ($10.9K), SCM1000 degasser ($1.8K), FL3000 fluorometer ($6.2K), and a DELL workstation ($3.3K). The total cost of the system is approximately $42K.

October The collection of the field samples for SeaHARRE-1 is completed, which are collected by the LOV during the PROSOPE cruise from the northwest Morocco upwelling to the Mediterranean Sea.

2000

February The data analyses for SeaHARRE-1 is completed, which are based on open-ocean (Case-1) samples spanning a wide dynamic range in TChl *a* concentration of approximately 0.04–2.09 mg m^{-3}. The SeaHARRE-1 results establish that TChl *a* and the so-called primary pigments can be quantitated with an uncertainty less than 10%

and 25%, respectively. The primary pigments are as follows: TChl *a*, TChl *b*, TChl *c*, Caro (the sum of α-Car plus β-Car), Allo, But, Diadino, Diato, Fuco, Hex, Peri, and Zea.

December CHORS receives a second Thermo Separations Products Spectra System HPLC with a UV6000LP detector.

2001

March The CHORS technician begins keeping a laboratory notebook. CHORS performs internal standard (Cantha) analyses on the NOAA system—there is no validation of the method to be used.

April CHORS calibrates for Chl *a*, Chl *b*, Chl c_2, Chl c_3, Allo, Anth, But, Croc, Diadino, Diato, Echin, Fuco, Hex, Lut, Neo, Peri, Pras, Viola, and Zea. Pump errors and an autosampler arm jam are noted.

May CHORS begins HPLC analysis of seawater samples for NASA PIs under a three-year contract with the SIMBIOS project. The Wright et al. (1991) method is used, which is based on a C_{18} column. The analyses are done on the NOAA system, and not on the new NASA system. CHORS calibrates for Chl *a* and Chl *b*.

June CHORS calibrates for Chl *a*, Chl *b*, α-Car, Allo, But, Chl c_2, Chl c_3, Diadino, Diato, Fuco, Hex, Lut, Viola, and Zea. Maintenance is performed on the pump and the piston seals are changed.

July CHORS swaps out the old NOAA pump, because of leaks from the pump head, and replaces it with the new NASA pump.

August CHORS calibrates for Chl *a* and Chl *b*.

October CHORS replaces the autosampler in the NOAA system with the autosampler from the NASA system.

September CHORS swaps the column from the NASA system into the NOAA system and performs maintenance on the autosampler arm. The extra in-line UV2000 detector is removed from the NOAA system. The HPLC system is calibrated for But, Diato, Fuco, Hex, and Zea.

October CHORS calibrates for Chl *a*, Chl *b*. The fluorescence detector from the NASA system is added to the NOAA system. The system is purged and the column equilibrated.

November The autosampler arm is lubricated and the needles are purged. CHORS calibrates the system for Chl *a* and Chl *b*.

December CHORS calibrates the system for Chl *a*, Chl *b*, α-Car, But, Chl c_2, Diadino, Diato, Fuco, Hex, Lut, Peri, and Zea.

2002

January CHORS replaces the lamp in the fluorescence detector. The first HPLC Workshop is hosted by GSFC. All the SeaHARRE-1 laboratories participate except MCM. The workshop establishes a) the need for performance metrics and a process to verify quality-assured results have been produced, b) the concept of reporting zeroes cannot be logically defended from the point of view of detectability and noise, and c) standardized data products are necessary.

February CHORS calibrates for Chl *a* and Chl *b*.

March CHORS replaces the PDA lamp in the NOAA system after more than 5,000 hours of use (the manufacturer recommends lamp replacement after 2,500 hours). CHORS calibrates the system for Chl *a*, Chl *b*, Allo, But, Chl c_3, Diadino, Fuco, Hex, Peri, Viola, and Zea.

April HPL helps organize a mini-round robin to investigate the uncertainties in the determinations of pigment concentrations by Plymouth Marine Laboratory (PML). PML is analyzing most of the pigments for the AMT activity. The results of the activity strongly suggest PML should participate in SeaHARRE-2, which they agree to do.

May CHORS replaces the PDA and flow cell in the NOAA system with the PDA and flow cell from the NASA system. The only checks to system performance are calibrations with Chl *a* and Cantha (the internal standard). CHORS calibrates the system for Chl *a*, Chl *b*, DVChl *a*, Allo, But, Chl c_2, Chl c_3, Diadino, Diato, Fuco, Hex, Lut, Neo, Peri, Phytin *a*, Pras, Viola, and Zea.

June CHORS experiences a large number of system restarts with the NOAA system.

July CHORS continues to experience a large number of system restarts with the NOAA system. HPL submits a document to the SIMBIOS project office discussing results of the first intercomparison with CHORS. Three mini-round robins between HPL and CHORS had been conducted, partly because a SIMBIOS investigator wished to continue having his samples analyzed at HPL rather than having them sent to CHORS and the SIMBIOS Project office was concerned the HPL analyses would not be of the requisite quality for calibration and validation work. The large differences in results between CHORS and HPL motivated HPL analysts to conduct experiments at HPL in hopes of explaining some of the differences. In addition to the field sample intercomparisons, replicate filters were subjected to CHORS and HPL extraction procedures at HPL. Vitamin E and Cantha internal standards were used simultaneously with filters subjected to the CHORS extraction procedures, and vitamin E alone was used with the HPL extraction procedures. These investigations were considered necessary because the average precision across replicate filters for all pigments reported by CHORS in the first

intercalibration exercise was 46% whereas it was 7% for HPL (CHORS uses Cantha as an internal standard, whereas HPL uses vitamin E). The efficacy of the CHORS and HPL internal standards in determination of extraction volume were investigated at HPL by extracting replicate field sample filters from each of four different sites. With the CHORS extraction procedure, the overall average extraction volume and precision determined from the Cantha internal standard was 4.22 mL and 4.7%, respectively; with vitamin E, the average extraction volume was 4.04 mL and the precision was 0.4%. Cantha, therefore, produced results 4.4% higher than the vitamin E results and with ten-fold poorer precision.

August CHORS installs a new column in the NOAA system, which is tested with pigment mixes. A spectrophotometer linearity test is also performed using Chl *a*. CHORS calibrates the system for Chl *a*, Chl *b*, But, Diadino, DVChl *a*, Fuco, Hex, Peri, and Zea. HPL submits a second document to the SIMBIOS Project Office describing the results of the pigment intercomparisons between CHORS and HPL. The average of the intercomparisons shows the CHORS determinations of TChl *a* are about 20.7% different from HPL. (In comparison, the average of the intercomparisons of HPL with respect to the other quality-assured laboratories in SeaHARRE-1 is 3.4% for TChl *a*.) Also, during the times of the CHORS HPL intercalibrations, HPL provided mixed calibration standards and a spiking solution of standards, so CHORS could conduct spiked recovery experiments. The CHORS technician was given instructions on how to do this work, but used up the spiking solution by analyzing it separately instead of using it to spike sample filters. The CHORS technician was told the data could still be important to analyze, but this work was not completed, because the technician "did not have time to pursue doing spiked recoveries" (and disclosed he didn't understand how to do them). During this time period, the CHOR technician also indicates automated integrations were used with some of the CHORS chromatograms, because "there was not enough time to inspect all peak area integrations." (The use of automated integration techniques has not been shown to produce quality-assured results—in fact, analysts from quality-assured laboratories have repeatedly indicated they should not be used if high-quality results are to be achieved.) HPL participates in a meeting convened by the SIMBIOS Project regarding the HPLC and CHORS intercomparisons. SIMBIOS project management believes a) it is irrelevant to place the SIMBIOS mini-round robin within the context of SeaHARRE results, so no representation of the SeaHARRE community (besides HPL) is allowed (in fact, the principal SeaHARRE data analyst is specifically not permitted to participate), and b) although there is a desire to understand the variability in the results, the biggest concern is to ensure the HPL data does not induce a bias with regard to data provided by CHORS. CHORS is particularly concerned about differences in results for DVChl *a*. CHORS uses a simultaneous equation and HPL chromatographically separates MVChl *a* and DVChl *a*. CHORS experiences a large number of system restarts with the NOAA system.

September CHORS continues to experience a large number of system restarts with the NOAA system.

October CHORS equilibrates and tests the column with Cantha and a pigment mix. The NASA autosampler is swapped in to replace the NOAA autosampler. The collection of the field samples for SeaHARRE-2 is completed, which are collected during the BENCAL cruise to the Benguela Current off the western coast of South Africa by LOV. CHORS is invited to participate in SeaHARRE-2, initially decides not to participate, but then ultimately does.

November CHORS continues to experience a large number of system restarts with the NOAA system, as well as an autosampler arm jam. CHORS calibrates the system for Chl *a*, Chl *b*, α-Car, But, Chl c_2, Chlide *a*, Diadino, Diato, Fuco, Hex, Peri, and Zea.

December CHORS purges and re-equilibrates the NOAA system.

2003

January CHORS replaces the NASA autosampler that was being used in the NOAA system with the original NOAA autosampler. CHORS replaces the NASA pump that was being used in the NOAA system with the original NOAA pump. CHORS replaces the NOAA PDA that was being used with the NOAA system with the PDA from the NASA system. CHORS replaces the NOAA pump that was recently put back into the NOAA system with the pump from the new NASA system, but then the parts are swapped out again. Despite all the parts swapping, no performance checks are made. At the end of the month, the operational system is composed of the following: NOAA degasser, NOAA pump, NASA autosampler, NASA PDA, and NOAA controller. CHORS calibrates for Neo, Viola, Pras, Phytin *a*, and Allo.

February CHORS sends the autosampler to Thermo for repairs, and the injector pod and check valves are replaced. CHORS has the pump serviced and the detector attenuators are adjusted. CHORS calibrates the system for Chl *a*, Chl *b*, Allo, But, Chl c_2, Diadino, Diato, DVChl *a*, Fuco, Hex, Lut, Peri, Pras, Viola, and Zea.

March CHORS performs lamp diagnostics and maintenance, and replaces the degasser on the NOAA system. The samples from February are run again. The UV2000 is put back in-line.

April The data analysis for SeaHARRE-2 is completed, which are based on open-ocean (Case-1) samples spanning a wide dynamic range in TChl *a* concentration of approximately 0.35–25.40 mg m^{-3}.

May An HPLC mini-workshop hosted by GSFC is held in Bethesda, Maryland. The workshop is organized to take advantage of international travelers attending a JGOFS meeting in Washington, DC. MCM and PML do not participate; CHORS participates part of the time. The workshop establishes the need for a symbology and lexicon to adequately document HPLC methods. The ensuing discussion reveals CHORS has been making a mistake in the quantitation of all their pigments: a water correction factor

has been added in twice. CHORS makes a new batch of internal standard (Cantha). A system equilibration is performed and test injections of Chl *a* and Chl *b* are performed.

June CHORS cleans the flow cells, runs detector diagnostics with methanol (MeOH), and tests the system with Cantha and a pigment mix. Calibrations are done for Chl *a*, Chl *b*, Phytin *a*, DVChl *a*, Chl c_3, Chl c_2, Peri, But, Fuco, Hex, Pras, Diadino, Diato, Allo, Zea, and Chlide *a*. After the calibrations are completed, the deuterium lamp in the NASA detector (which is being used in the NOAA system) fails and is replaced, but no calibrations are done.

July CHORS performs diagnostics and replaces the column in the NOAA system.

August The preparation syringe jams in the NOAA system.

September CHORS calibrates for Chl *a* and Chl *b*. Detector diagnostics and autosampler maintenance is performed. The system goes down, because solvent A runs dry. When the system is brought back up, there is a baseline hump that appears at a retention time of 16 min. Much of September is spent trying to understand the reason for the baseline hump, which Thermo believes is due to a contamination problem. Both UV lamps are replaced and two more Chl *a* and Chl *b* calibrations are performed.

October CHORS replaces the flow cell in the NASA detector with the flow cell from the NOAA detector (the NASA flow cell fell on the floor). The NOAA detector is subsequently swapped in, but the NOAA flow cell is used from the NASA detector. CHORS calibrates for Peri, But, Fuco, Neo, Hex, Viola, Diadino, Diato, Lut, Zea, Chl *b*, Chl *a*, DVChl *a*, and Cantha. After the calibrations, the pigment mix is run, pump diagnostics are made, and the autosampler and detector undergo maintenance. Later, the NASA detector is swapped back in with the NOAA flow cell, but the baseline is poor (the UV2000 is put in-line and "looks OK"). The NOAA detector is swapped back in to replace the NASA detector, but detector problems persist. Both lamps are replaced at the end of the month.

November CHORS notes the pressure of the pump is low due to a room temperature stability problem. The system is brought down, restarted, and looks "good." The second HPLC Workshop is hosted by LOV (Villefranche-sur-mer, France). BIO, MCM, and CHORS do not attend. The objectives for this workshop are a direct consequence of what has been learned from the first two round robins, what is currently imagined for the third, and the future direction of marine pigment research by the funding agencies. The objectives include the following: a) decide what to do about null detection; b) agree on what a blunder is, and how it should be handled; c) estimate the detection limits of all methods; d) recommend a standard set of pigments for all analysts to report (e.g., the primary pigments); e) establish a reference mixture; f) suggest changes to The Protocols (field and laboratory aspects); g) choose a course of action for resolving the problems with absorption coefficients; and h) consider whether or not to host an HPLC round robin emphasizing coastal samples. A presentation is given by HPL on determination of calibration linearity.

December The CHORS HPLC quality "looks good." The C_{18} method is brought down.

2004

January CHORS begins testing C_8 methods, so DVChl *a* and MVChl *a* can be chromatographically separated. Barlow et al. (1997) is tested first, and then the Goericke et al. (2000) method is tested. HPL suggests using vitamin E as the internal standard for the latter method.

February CHORS continues testing C_8 methods and concludes with the Van Heukelem and Thomas (2001) method.

March The CHORS technician makes note of baseline noise in the C_8 method.

April–May CHORS continues testing the Van Heukelem and Thomas (2001) C_8 method.

June HPL provides CHORS instructions as to how HPL makes the tetrabutyl ammonium acetate (TBAA) buffer, so CHORS can do the same. The NASA flow cell is replaced. Most of June is spent working on trying to bring up the HPL C_8 method, which requires a column heater, so an Alltech column oven model 631 is purchased.

July CHORS spends much of July testing injection methods (e.g., push versus pull) and working on the NASA autosampler, because of poor precision (which is greater than 6% and it should be less than 2%). Ultimately, the NASA autosampler is replaced with the NOAA autosampler, and then the rotor is replaced. The NOAA autosampler is sent back to Thermo for repair, and the NASA autosampler is swapped back into use. HPL suggests the persistent injection problems may be caused by the higher viscosity for the TBAA buffer used with the HPL C_8 method. The month ends with the syringes and six-port syringe valve being replaced in the NASA autosampler.

August CHORS swaps in the newly repaired NOAA autosampler to replace the NASA unit. A test of the repaired NOAA autosampler reveals the injector precision is 18.6% (it should be less than 2%). Different autosampler command options are tried to see if the precision can be improved through reprogramming. The Thermo dilution reproducibility test is received from Thermo and executed—the precision is less than 1%, which makes the CHORS technician believe the poor precision is a result of the interaction of the sample with the buffer used in the HPL method.

September CHORS swaps in the NASA UV6000LP to replace the NOAA UV6000LP, and then swaps the NOAA flow cell for the NASA flow cell, but the injector precision does not improve. The NOAA UV6000LP is swapped in to replace the NASA UV6000LP —the baseline is still noisy. The UV2000 is placed in line with the UV6000LP; the UV2000 shows a stable baseline and the UV6000LP does not. The NASA flow cell is swapped in to replace the NOAA flow cell. The NASA UV6000LP is swapped in to replace the NOAA UV6000LP, but the baseline is still noisy. The flow cell is purged with

water, then 20% nitric acid for 1 hour, and then rinsed with water, which is allowed to sit over the weekend. The flow cell is then flushed with MeOH, the system is flushed with MeOH, detector diagnostics are run with the oven on, and the baseline "looks good." The UV2000 won't trigger, but direct wiring to the autosampler solves the problem; then the pump won't trigger. The NOAA flow cell is swapped in for the NASA flow cell, but the baseline is still poor. The system goes down, because of a power failure—CHORS does not use an uninterruptible power supply (UPS) with the HPLC system—and then the autosampler pod jams.

October CHORS shuts down the system to replace the flow cell in the NOAA UV6000LP, but when the system is restarted, the UV2000 will not power up, so it is taken off line. The system is brought back up, but the power goes off to the entire lab (CHORS does not use a UPS with the HPLC system). When the system is restarted, the column pressure is high, so a new column is installed. The UV2000 is still down, the PDA baseline "looks good," but the baseline on the test mix "looks bad." The diaphragm is replaced on the membrane degasser. The detector intensities are checked, and the system "looks good," but the baseline is a little noisy. The system is restarted, and the baseline is "poor." The system is shut down and oil droplets are observed on the inlet peak line to the column. The pump diagnostics are "OK." The system is restarted and calibrated for Chl c_3, Chl c_2, Neo, Diato, Chl *a*, Phide *a*, But, Diadino, Chl *b*, Fuco, Allo, DVChl *a*, Pras, Zea, Phytin *a*, Chlide *a*, Viola, Lut, and α-Car. The system is stopped, because of a poor baseline. Consultations with Thermo determine the problem is not the pump. The flow cell is rinsed with water and then 20% nitric acid, followed by water and then MeOH. The baseline drifts with flow cell or no flow cell; low noise is seen with the former and some noise with the latter. The flow cell is replaced, but the detector diagnostics indicate the UV6000LP needs to be sent to Thermo for repair. The NASA UV6000LP is removed and replaced with the NOAA UV6000LP. A new column and pre-filter is installed. The system is checked with pigment mixes.

November Detector diagnostics check out "OK" with Thermo diagnostics. The system is restarted, but the internal standard baseline is bad (the precision is 20%); the Chl *b* baseline is also bad. The system is restarted after purging the entire system. The six-port solvent selection valve is observed to be leaking on the autosampler, so the system is shut down. The Goericke et al. (2000) method is revisited. The system is shut down for pump and lamp diagnostics, and both "look good." The CHORS technician notes poor baseline and peak shape, but the PI thinks both are "OK," so a calibration is performed. The baseline is still bad, so the NOAA pump is swapped in to replace the NASA pump. The autosampler stops at the vial 10 position, so the system is shut down, but after restarting the system, the baseline is still bad. The degasser is checked, but the noise is still bad with and without the flow cell. A new flow cell is installed, but there is still baseline drift and noise. The wiring for the tungsten lamp is adjusted to make sure there is a good connection. The seals on the outlet liquid end are replaced, a new column is installed, and a new tungsten lamp is installed. The baseline is still bad, and the pressure is too high. After reading Goericke et al. (2000) again, the CHORS technician modifies solvent A to 75% MeOH and 25% ammonium acetate (NH_4Ac). The

pressure reaches a maximum. The Van Heukelem and Thomas (2001) method is reinstalled, pigments are calibrated (Chl *a*, Chl *b*, Diadino, Fuco, Peri, Zea, But, Diato, DVChl *a*, and Hex), and samples are analyzed.

December The collection of the field samples for SeaHARRE-3 is completed, which are collected during the Biology and *In Situ* Optics of the South Pacific Experiment (BIOSOPE) cruise to the South Pacific gyre and the Chilean upwelling (Tahiti to Easter Island to Chile). The CHORS technician performs a method test and a calibration. The pump stops due to a power failure (CHORS does not use a UPS with the HPLC system). The system is calibrated for Chl c_3, Diato, Chl *a*, Neo, But, Diadino, Chl *b*, Phide *a*, Allo, Chl *a*, Fuco, β-Car, Gyro, Hex, Peri, Chl c_2, Phytin *a*, Pras, Zea, α-Car, Chlide *a*, Lut, and Viola.

2005

January The CHORS technician restarts the HPLC system after a power failure (CHORS does not use a UPS with the HPLC system). The system stalls and the detector is noted to be "flashing red." All components except the heater are turned off and restarted. A calibration of Chl *a* and Chl *b* is performed, but peak shape is poor. The system is shut down and a new column is installed. The system is calibrated for Chl c_3, Diato, Chl a, Neo, But, Diadino, Chl *b*, Phide *a*, Allo, DVChl *a*, Fuco, β-Car, Gyro, Hex, Peri, Chl c_2, Phytin *a*, Pras, Zea, α-Car, Chlide *a*, Lut, and Viola.

February The CHORS system goes down and is restarted, but the computer system freezes.

March CHORS installs a new column and pre-filter, performs a detector diagnostic check, and cleans, lubricates, and checks the alignment of the autosampler. The system is calibrated for But, Chl c_3, Hex, Lut, Chl c_2, Diadino, DVChl *a*, Fuco, Allo, Chlide *a*, Neo, Phytin *a*, α-Car, Diato, Phide *a*, Pras, β-Car, Peri, Viola, Zea, and Gyro.

April The CHORS technician notes the detector is not triggering, so the system is restarted. Later, the system is shut down, because the detector red light is on. An investigation reveals solvent A ran out, but the oven was on, so the column was probably "fried." A new column is installed and the autosampler is lubricated and purged. A calibration is performed for But, Chl c_3, Hex, Lut, Chl c_2, Diadino, DVChl *a*, Fuco, Allo, Chlide *a*, Neo, Phytin *a*, α-Car, Diato, Phide *a*, Pras, Peri, and Zea.

May CHORS restarts the system two times, because of problems with the autosampler.

June CHORS runs pump diagnostics and replaces the seals on the pump outlet; the tungsten detector lamp is replaced and detector diagnostics are checked. The second HPLC system is set up: the UV6000LP is serviced and new lamps installed, a new column is installed and equilibrated, a test run is made, the fittings are checked for leaks, and the system is purged overnight. The column is reconnected, the tubing is trimmed and reconnected to the flow cell, and the flow restrictor is reconnected. The

baseline still “looks bad,” so the system is purged with 100% MeOH. The column is re-equilibrated, but the MeOH baseline still “looks noisy.” The tungsten lamp and the flow cell are changed out, and the previous tungsten lamp is swapped back in, but the baseline still “looks bad.” The old column is swapped back in and pumped with 100% MeOH over the weekend. The tungsten lamp is replaced with a new one (the third one), but the baseline check is still “bad.” The system is restarted, the attenuators are adjusted, and the intensity is checked. The NOAA UV6000LP is put in-line with the NASA detector, and a baseline check is made—the baseline still “looks noisy.” The NOAA detector is put in-line with the NASA detector, and a baseline check is done again—the MeOH blank “looks OK.” The NASA system is shut down, the NOAA UV6000LP is returned to the NOAA system, and the power strip for the NASA HPLC system is moved to a different wall outlet (CHORS does not use UPS units for laboratory instruments). The NASA system is restarted, and the baseline check is “not bad.” Everything is shut down except the degasser and pump, and 100% MeOH is pumped through the system overnight. The degasser is suspected of being inoperable, because the gauge is not moving, so the degasser diaphragm is replaced and the column is replaced with a new C_{18} column. The system is purged and equilibrated. The NOAA UV6000LP is used for running diagnostics on the NASA system. A test calibration shows continuing baseline noise and inverter spikes. Thermo believes either a detector lamp is failing or the flow cell is contaminated. The NOAA UV6000LP is swapped in to replace the NASA UV6000LP. A baseline check with the same flow cell “looks good.”

July CHORS calibrates the NOAA (C_8) system for α-Car, Allo, β-Car, But, Chl *c*$_2$, Chl *c*$_3$, Chlide *a*, Diadino, Diato, DVChl *a*, Fuco, Gyro, Hex, Lut, Chl *a*, Chl *b*, Neo, Peri, Phide *a*, Phytin *a*, Pras, Viola, and Zea. CHORS calibrates the NASA (C_{18}) system for α-Car, Allo, β-Car, But, Chl *c*$_2$, Chl *c*$_3$, Chlide *a*, Diadino, Diato, DVChl *a*, Fuco, Gyro, Hex, Lut, Chl *a*, Chl *b*, Neo, Peri, Phide *a*, Phytin *a*, Pras, Viola, and Zea. A new C_{18} column is added to the NASA system with new connections to the autosampler and UV6000LP. Detector diagnostics are done on the NASA system, and the baseline check with MeOH “looks OK.” The NOAA system freezes and the UV6000LP did not start when the system was restarted. The NASA system stops overnight, and a leak is found in the autosampler injector valve.

August CHORS puts a new C_{18} column in the NASA system. The analysis of the SeaHARRE-3 samples begins on the NASA (C_{18}) system. A new C_8 column is put into the NOAA system. Pump and detector diagnostics are run on the NASA system. Both the NOAA and NASA systems are run at the same time with injections occurring within a few hours of each other. The NOAA (C_8) system is restarted due to high pressure and a the column is replaced with a new one. The NASA system freezes and is restarted, and then the NOAA system freezes and is restarted. Sigma calibrations are performed with Chl *a* and Chl *b*. The NASA (C_{18}) system shuts down and is restarted. The column in the NOAA (C_8) system is replaced with a new one. Agilent is contacted to see if the C8 columns are being made differently; Agilent says they have not changed the packing material in the column.

September CHORS puts a new C_8 column in the NOAA system. Calibrations are performed for Chl *a* and Chl *b*. Internal standard tests are done on the NOAA system, the system is restarted, and pump diagnostics are run. The NOAA (C_8) system is restarted, and the column is replaced.

October CHORS notes the UV6000LP does not start on injection. The NOAA system cannot be touched without the system crashing. Calibrations are done for Chl *a* and Chl *b*. The power to the laboratory fails, the pump is off with no flow to the system, so the column is "probably fried" (the CHORS laboratory instruments are not on a UPS). The column is flushed and tested; it "looks good," so the run is continued.

2006

January CHORS runs diagnostics on the NOAA (C_8) system, the autosampler is purged and prepped, and detector diagnostics are done with MeOH. New internal standard is made and checked, and the system is checked with a pigment mix. A Sigma calibration is done for Chl *a* and Chl *b*, and a DHI calibration is done for Chl c_3, Diato, Chl *a*, Neo, But, Diadino, Chl *b*, Phide *a*, Allo, DVChl *a*, Fuco, β-Car, Hex, Peri, Chl c_2, Phytin *a*, Pras, Zea, α-Car, Chlide *a*, Lut, and Viola. Samples are analyzed.

February CHORS runs more samples on the NOAA (C_8) system and does a calibration of Chl *a* and Chl *b*.

March CHORS decides to stop using the C_8 method until the problems with the method are resolved. The data analysis for SeaHARRE-3 is completed, which are based on open-ocean samples spanning a wide dynamic range in TChl *a* concentration of approximately 0.02–1.37 mg m^{-3}. The results establish the new CHORS C_8 method has uncertainties exceeding calibration and validation requirements. NASA HQ requests the data not be distributed publicly and to defer any representations to the wider SeaHARRE community until CHORS has an opportunity to respond. CHORS goes back to using the C_8 method. The system that is configured for this method is composed of the following components: NASA degasser, NASA pump, NOAA autosampler, NASA UV6000LP, and NASA controller. Pump and detector diagnostics are run, the column is equilibrated, the autosampler is tested, and the system is started up. A calibration of DHI pigments is started, but the system aborts the run during calibration. The system is shut down and restarted, and the problem appears to be with the autosampler, so the NASA autosampler is swapped in to replace the NOAA autosampler. The sample loop is changed to 100 µL, the syringes are prepped, and the system is restarted with blanks. The calibration is continued for α-Car, Allo, β-Car, But, Chl c_2, Chl c_3, Chlide *a*, Diadino, Diato, DVChl *a*, Fuco, Hex, Lut, Chl *a*, Chl *b*, Neo, Peri, Phide *a*, Phytin *a*, Pras, Viola, and Zea. Samples are analyzed for about a two-week contiguous period, but then the autosampler stops working. The system is restarted, but it stops again, because of an autosampler error. Thermo is consulted and the injector needle is replaced. The system is restarted, the repeatability of the autosampler is checked, but there is a system glitch. The system is restarted, but the baseline is drifting, the PDA spectrum is poor (probably

a flow cell problem), and the intensity is low. The system is flushed with MeOH, and the column, pre-filter, and flow cell are changed. Detector diagnostics are run, the system is checked, and it is restarted. The pump diagnostics are "OK," detector diagnostics are run again, and the system is restarted. There is a prominent solvent front, so the run is stopped.

April The flow cell from the NOAA UV6000LP is swapped in for the one in the NASA UV6000LP. The intensities "look good," and the baseline is "OK." A new lamp is installed in the UV6000LP, and a calibration is performed for Chl *a* and Chl *b*. The system stops with the same autosampler problem as before, so the system is shut down and restarted. The prep and sample syringes are flushed. The blanks "look good," so samples are analyzed. The system stops again, because of the same autosampler error. The system is restarted, but it stops for the same reason. The system is shut down and restarted, but the sequence has to be stopped due to a lack of sample vials. The system is shut down and restarted, which allows the remaining samples to be analyzed.

May CHORS starts the system, runs pump diagnostics, purges the autosampler lines, and runs detector diagnostics. Calibrations are done for Chl *a* and Chl *b*, which are repeated. HPL completes a proposal requested by NASA HQ entitled "Analysis of Phytoplankton Pigment Samples by High Performance Liquid Chromatography for NASA Investigators," which is reviewed and approved for funding. Funding for the HPL proposal is executed as an emergency procurement, so HPLC pigment analyses can resume as soon as possible.

June CHORS does a C_8 system test. A 200 µL loop is put in the autosampler, and pump and detector diagnostics are run. The month ends with another C_8 system test.

July CHORS runs a Chl *a* and Chl *b* mix, a no prep time test, and then continues with a C_8 buffer and sample interaction tests. The conclusion of the tests is the autosampler configuration makes no difference.

August CHORS replaces the C_{18} column with a new one. Pump and detector diagnostics are run (the pump is "OK"), and the system is tested. The system is calibrated for α-Car, Allo, β-Car, But, Chl c_2, Chl c_3, Chlide *a*, Diadino, Diato, DVChl *a*, Fuco, Hex, Lut, Chl *a*, Chl *b*, Neo, Peri, Phide *a*, Phytin *a*, Pras, Viola, and Zea.

September HPL begins analyzing HPLC samples for NASA investigators based on funding from the emergency contract. The CHORS system experiences a failure: the UV6000LP loses transmission due to a power outage (the CHORS laboratory equipment is not on a UPS). The system is restarted, but the pressure is too high. The system is restarted, Chl *a* and Chl *b* are calibrated, but the system is stopped, because the UV6000LP never triggered. Duplicate CHORS and HPL C_8 samples are identified, and all of them are from GSFC investigators or SeaHARRE activities. Although the purpose here is to find duplicates between CHORS and HPL, the search for duplicates

reveals there are probably no duplicates from the wider scientific community in their submission of samples to CHORS. The latter is a worrisome discovery, because the field sampling protocols require duplicates, but CHORS was not enforcing this requirement on the PIs submitting samples. The first part of the duplicate data set is sent to CHORS for match-up analysis.

October CHORS analyzes samples for the entire month. The collection of the field samples for SeaHARRE-4 is completed, which are collected during day cruises to the fjords and estuaries of Danish coastal waters. NASA HQ asks CHORS to produce a report summarizing the analysis of the aberrant SeaHARRE-3 C_8 quantitation problem and begins the process of selecting independent HPLC scientists to review the document.

November The remaining part of the duplicate data set is sent to CHORS for match-up analysis. This is not following the original concept of collecting the duplicate data which were supposed to be used to evaluate whatever correction scheme CHORS produced, but in the absence of CHORS actually producing a correction scheme, this becomes the de facto scenario.

December The first draft of the CHORS C_8 report is reviewed by NASA HQ.

2007

January The revised CHORS C_8 report, which is entitled the "HPLC Pigment Bias Report," is received by NASA HQ and sent to two reviewers.

February The comments from the reviewers (plus e-mail comments from one investigator) are provided to CHORS. The comments are incorporated into a final report, which is submitted by CHORS to NASA HQ. A summary of the conclusions of the report is as follows: 1. The CHORS C_8 method significantly overestimated MVChl *a*, DVChl *a*, and Chl *b*, but the other pigment compounds are asserted to "agree well with those measured by the 'A Group' during the SH3 exercise," and the overestimation is asserted to be constant throughout the year in which the C_8 method was used. 2. The uncertainty in the results is asserted to not be random, but the reason for the overestimation could not been determined. Because of the consistent nature of the bias, the report asserts an error in calibration or calculating concentrations from the peak areas would be the suspect, however "no obvious errors have been found in these calculations." An investigation into a methodological problem associated with mixing the buffer and sample together, prior to injection on the CHORS system, is considered, but the conclusion is this could only cause an increase in the calibration uncertainty by approximately 3–6%. 3. The fluorometric data collected for all MODIS samples are asserted to not have the biases found with the C_8 method, and a long-term analysis of fluorometric to HPLC ratios for a variety of field samples shows the data corresponding to the C_8 method have notably different ratios. 4. A log-linear regression approach is recommended to correct for the three identified biases—the report asserts that all the

other pigment data do not have a bias and agree well with the 'A Group' averages. The report concludes by noting, “there are some compounds that have differences, but these are difficult compounds to get agreements with other methods and laboratories.” The proposed CHORS correction scheme is rejected by the OBB Program Manager, who then asks the CVO to review the document and provide recommendations about how to proceed. The first response from the CVO is to clarify the most obvious mistakes in the CHORS report: a) the reference to the “A Group” is actually the “A’ Group,” that is, the quality-assured subset; b) the assertions that the other pigment data (i.e., the pigments other than MVChl *a*, DVChl *a*, and Chl *b*) do not have problems and “some compounds” that have differences are “difficult to get agreements with other methods and laboratories” is not completely supported by the SeaHARRE-3 results, which show the CHORS C_8 carotenoid results are 50% higher than the A’ (quality-assured) subset and the agreement of the laboratories involved is very good and rather uniform except for Peri. Most importantly, however, is the fact that the report does not address the poor results obtained with the CHORS C_{18} method in SeaHARRE-3. A comparison of the uncertainties for the A’ (quality-assured) subset versus the CHORS C_{18} results is as follows: the C_{18} chlorophylls average 15.0% percent versus 11.7% for the A’ group, and the C_{18} carotenoids average 44.6% versus 12.4% for the A’ group. The average uncertainty in the primary pigments (3 chlorophylls and 9 carotenoids) is 37.2% for the CHORS C_{18} results, but only 12.2% for the quality-assured subset. The two CHORS methods, therefore, are equally challenged, but for different reasons: the C_8 results have bad chlorophyll uncertainties, but good carotenoid uncertainties, whereas, the C_{18} results have the reverse. In both cases, the so-called “good” results have uncertainties that are larger than the quality-assured results and sufficiently so to be worrisome. For example, the uncertainty for the C_{18} TChl *a* analysis is 16.5%, which exceeds the 15% threshold established for quantitative analysis, and the uncertainty for the C_{18} DVChl *a* analysis is 106.9%.

March CHORS starts up the HPLC system with the C_{18} method after more than four months of being idled. The system is composed mainly of NASA components: the NASA degasser, the NASA pump, the NASA autosampler, the NASA UV6000LP, and the NOAA controller. System diagnostics are run and a calibration is initiated. The calibration is stopped, because of a bad tungsten lamp. The calibration is reattempted, but again the tungsten lamp is bad. A third calibration attempt is made and it is not successful. The NOAA UV6000LP is swapped in for the NASA UV6000LP. An old column is used to test Chl c_2 and Peri. A system check reveals the column is bad. A new column is ordered, and a new diaphragm is installed on the degasser. The new column is equilibrated and the system is checked. The UV6000LP has a fault, so the system is shut down and restarted. The system will not inject or trigger, so it is shut down again. The system is restarted, but the UV6000LP freezes. The NASA UV6000LP is swapped in to replace the NOAA UV6000LP, but the lamp with lesser hours on it is used. HPL addresses specific problems in the CHORS bias report directly to NASA HQ. The most salient points are as follows: a) CHORS did not properly validate either the C_8 or C_{18} method in preparation for SeaHARRE-3 or for the analysis of field samples during the time period that they conducted the side-by-side analyses (on both methods); b) the two

methods implemented by CHORS during SeaHARRE-3 performed well below standards they are capable of achieving (based on the SeaHARRE-2 results), which makes the issue of identifying a correction process a more complicated issue than simply determining what went wrong when they implemented the C_8 method; c) the report suggests the CHORS problems with the C_8 method of Van Heukelem and Thomas (2001) were associated with the TBAA buffer, because this buffer was observed to cause pigment losses. The latter deserves extra comment, because while it is true that pigments precipitate when organic solutions are too aqueous, it is not possible to address whether TBAA specifically contributed to CHORS poor results without knowing details that are not provided in the report (e.g., the purity of the buffer, whether the pH was properly adjusted, whether it was used while cold, etc.). Furthermore, precipitation of pigments would cause a reduction in concentration, not an overestimation (as was evident with the TChl *a* results for the C_8 method in SeaHARRE-3). A reviewer supports the suggestion that TBAA may be problematic and suggests further investigations to this effect, and also suggests the use of ammonium acetate buffer. Ammonium acetate buffer when combined with elevated column temperatures (as used by the C_8 method referred to here), has been observed at HPL to yield a very nonlinear Chl *a* response. The reviewer did not acknowledge a crucial point—CHORS knew they were having problems with the C_8 method and knowingly put it into service for the analysis of field samples for NASA PIs. The reviews conclude with a summary statement that the reviewer is "very glad that I never switched from the C_{18} method to the HPL C_8 method that only seems to produce reliable results when implemented on a HP Agilent HPLC system," which completely ignores the SeaHARRE laboratories who have implemented this method on different hardware (Shimadzu, Waters, and Agilent) and achieved quality-assured results (DHI, CSIRO, and LOV). NASA HQ arranges a meeting at GSFC where CHORS presents results. HQ decides to postpone the release of the document and to form an investigative team (already referred to as The Team), composed of scientists from the CVO, HPL, and CHORS. The Team is tasked with a) investigating the cause of the aberrant C_8 results by reviewing the implementation of the C_8 method on the CHORS system, including system performance, reproducibility, and uncertainty; and b) determining the best approach to correct the CHORS data, as well as, evaluating whether or not uncertainty estimates can be assigned to the corrected values.

April CHORS restarts the HPLC system and calibrates for Neo, Allo, But, Chlide *a*, Diadino, Diato, DVChl *a*, Fuco, Hex, Lut, Chl *a*, Chl *b*, Peri, Phide *a*, Phytin *a*, Pras, Viola, and Zea. An investigative plan is presented to the community at the OCRT meeting and includes the execution of the following tasks: a) a forensics activity to provide a clear description of what was done at CHORS to implement the C_8 method; b) an analysis of the QA and QC data collected during the execution of the CHORS C_8 and C_{18} methods as a function of time; c) the construction of a detailed time line of what errors occurred and at what points in time, so the QA data can be used diagnostically; d) an analysis of whether or not the underlying problem(s) can be corrected using the principles or parameters of the problem and not just the statistics of the data; e) an uncertainty analysis of the agreed upon correction scheme; and f) a review of the results obtained by The Team by an expert in HPLC chromatography.

May HPL and the CVO discuss how best to add an external chromatography expert to The Team. The agreement is to find someone outside the marine community, to maintain objectivity, but who has an international reputation and working presence in the larger HPLC community (e.g., someone associated with an appropriate journal).

June The first inquiries by The Team at CHORS determine the CHORS C_8 calibrations are substantially inadequate. The variation in the CHORS individual pigment calibrations (using DHI standards) averaged 11.6% and spanned 3.3–38.4%; for all the calibrations, the average is 16.6% spanning 9.7–48.7%. This variance was discernible early in the CHORS analyses of field samples, but was not investigated by CHORS. A method validated to produce quantitative or state-of-the-art results will have calibration variations less than 5%, and most practitioners producing data of this quality would stop using a method when variations exceeded 5%; analyses would not resume until the source of the variations was understood and corrected. An important contributor to the poor calibration results was an inadequate working range in the calibration points. Although there was not enough time to review all of the CHORS calibrations, no Chl *a* calibrations extended to 100% of the working range and the most trusted calibrations (using DHI standards) spanned 3.1–66.8% (the average was 24.6%). In addition, too many points below a reasonable lower limit (usually defined as 1% of the working range) had very large uncertainties and yet were unnecessarily included in the calibration dilution set. The HPL investigations of the C_8 calibrations expose an abnormality in the spectral properties of the red and blue wavelengths used to quantitate marine pigments. For validated methods with a well-established linear response, the ratio of the red-to-blue quantitation wavelengths is constant as a function of the amount of pigment in the calibrant, but for the CHORS C_8 data, the red-to-blue ratio is not constant. HPL solicits and receives data from two other SeaHARRE groups using C_8 methods on different hardware, DHI (Shimadzu) and CSIRO (Waters), and both of these groups have constant red-to-blue ratios in their calibration curves. This type of problem is usually associated strictly with the detector being used and not with the laboratory procedures. A so-called parametric correction scheme is established and evaluated by the CVO. The correction assumes the volumetric terms and peak area integrations quantitated by CHORS are largely correct, and then maps the HPL calibration from a database of GSFC and SeaHARRE duplicates analyzed by CHORS and HPL onto the corresponding CHORS peak areas, i.e., the HPL calibrations and peak areas are used to calibrate the CHORS peak areas. The CVO submits a procurement request to add the hourly services of John Dolan from LC Resources to The Team as the external chromatography expert. John has been in charge of an analytical laboratory for 20 years and has significant experience in method transfer and method validation, primarily from the perspective of FDA compliance.

July The OBB Program Manager is briefed about the principal findings for the June 2007 investigations of the CHORS C_8 analyses: a) the calibrations were substantially inadequate (the average variation in the CHORS calibrations is 16.6%); b) a linear dynamic range and a working range were not established (on average the calibrations only span about 25% of a sensible working range); c) a proposed parametric correction scheme based on duplicate CHORS and HPL analyses appears tenable (the

uncertainties for most of the corrected primary pigments are to within 15%; d) CHORS did not make any daily, weekly, or monthly quality assurance measurements; e) CHORS used a wide variety of less-than-optimal laboratory procedures; f) although the parametric correction scheme produced an inverse response factor that was very similar to the calibration performed by The Team at CHORS, there was a clear indication of a residual nonlinearity in the distribution of the data; g) there was a nonlinear relationship in the red-to-blue (664 nm/450 nm) ratio for the calibration data (and The Team suggested an outside expert was needed to investigate this further); and h) all of the findings of the inquiry to date—in particular, the parametric correction approach—should be reviewed by a third-party expert in chromatography. HPL submits a proposal entitled "HPLC Pigment Analyses to Support Ocean Biology and Biogeochemistry Research" to the EOS recompetition in the ROSES 2007 call, which is reviewed and selected by the peer-review panel for funding. The Team begins looking at all the CHORS calibration data, and it is clear the CHORS C_{18} and C_8 methods relied on basically the same less-than-optimal calibration practices. The Team is increasingly concerned that the older C_{18} results might be compromised by common poor laboratory practices and a significant detector problem. The CVO begins the search of adding an HPLC detector expert to The Team.

August The parametric correction scheme is approved by the external HPLC expert. The second visit by The Team to CHORS begins with detailed comparisons between the C_8 and C_{18} calibration procedures, and quickly establishes the CHORS C_{18} calibrations are also substantially inadequate. Subsequent findings include the following: a) the C_8 and the C_{18} calibrations are nonlinear, although the former exhibit the most nonlinearity, and the chlorophylls are worse than the carotenoids, because quantitation is based on detection in the red domain (the C_{18} Chl *a* calibration is based on blue wavelengths); b) the C_8 results appear to be further degraded by the requirement to mix in the vial with the Thermo injector (a hardware limitation that cannot be overcome), which might be producing chemical reactions that are reducing the amount of pigment injected onto the column or increasing the amount of pigment retained by the column—both effects result in lower pigment quantitations; c) as was seen with the C_8 method, the variance in the C_{18} calibrations renders a very large amount of data unsuitable for calibration and validation activities; d) the total number of samples requiring correction is estimated to be about 18,500; e) the possibility of using a parametric correction scheme is considered, but a database of sufficient duplicates cannot be identified, and there is the persistent problem that an unequivocal source of the nonlinearity seen in all CHORS calibrations has not been identified, so the applicability of the parametric correction scheme is unknown; f) the recommendation is made that all data reported with whatever correction scheme is adopted should be labeled as "Corrected Data," and should only be used if supporting data for an overall study is available—corrected data should not be used as primary data upon which to base policy decisions; g) possible future directions are considered, but no one path can be selected, because of the unknown source of the nonlinearity; and h) the future directions identify the importance of having the CHORS technician available for future work (funding for the CHORS technician ends at the end of November 2007), and the need for an HPLC detector expert. An HPLC detector expert, Ron Farnbach, is added to The Team. Ron is fortuitously based

in Temecula (California), which is very close to San Diego and has extensive experience with HPLC detector systems (he worked for Waters for many years). The OBB Program Manager is briefed about the principal findings for the August 2007 investigations of the CHORS C_{18} analyses.

September Power problems are diagnosed by The Team (the voltage between neutral and ground was floating) and improved in the laboratory CHORS uses for HPLC analyses. CHORS purchases and installs a 1.5KVA UPS for the HPLC system for the first time. In anticipation that the CVO might need to make supporting HPLC analyses, the CVO explores with GSFC Facilities what is required to allow HPLC analyses in C-LABS. GSFC Facilities and Safety determine that extra ventilation needs to be installed to remove organic solvent vapors over the HPLC equipment. A pure-water system is also needed and purchased, but GSFC Facilities will not let the CVO use an off-site commercial company to install the system, because of contractual requirements with the company already selected to perform on-site facilities support. The HPLC detector expert suggests using erbium to investigate the nonlinearity of the UV6000LP. The tests are mostly inconclusive, but they do reveal a substantial refractive index (RI) problem with the flow cell. The HPLC detector expert recommends a high-resolution (20 dilutions) Chl *a* calibration using the C_{18} method with all dilutions injected in triplicate. The CVO prepares an emergency procurement to hire the CHORS technician as a subcontractor, because there will be no viable NASA contractual vehicle at CHORS after 31 November 2007. There is no automated way to apply an RI correction with the Thermo software, but the CHORS technician is able to do it manually, and it improves the chromatograms and reduces the nonlinearity. The manual correction is so labor intensive, however, it will not be practical to implement if for the thousands of chromatograms involved with the CHORS data set. A high-resolution (20 dilutions in triplicate) C_{18} calibration of Chl *a* is performed to compare with the C_8 calibration performed in June 2007. The results from the new calibration are all "double checked" to confirm the areas are correct, because the detector expert notes precision problems in the quantitations. The CHORS technician reviews the data and concludes, "they looked weird." The CVO investigates further and finds precision is frequently very poor (over 20% for some triplicates), with individual triplicates having a precision as high as 23.6%. The average precision exceeds acceptable performance metrics, and is caused by one of the injections within a triplicate being anomalously high or low. As noted by the CHORS technician, the "bad points are random throughout the run" and they "seem like a lot of bad points." If the apparent "outlier" injections are removed, overall precision improves from 7.8 to 1.0%. The HPLC detector expert notes the ratio of the 436 to 664 nm absorbance is not constant for the C_{18} calibration curve, which means the C_8 and C_{18} response are rather similar in this aspect of the nonlinearity.

October The CVO submits a work order on 31 October 2007 to add the recommended ventilation to C-LABS for HPLC and a combustion oven, install a pure-water system, mount a gas canister rack, and rearrange some cabinetry. CHORS brings the HPLC system up running the C_8 method. Pump and detector diagnostics are done. A new high-resolution (20 dilutions in triplicate) C_8 calibration of Chl *a* is performed to compare with the C_{18} calibration performed in September 2007. The imprecision once again

exceeds acceptable performance, with individual triplicates having a precision as high as 53.2%. If the apparent "outlier" results are removed, overall precision improves from 8.7 to 2.0%. There is a clear indication of a nonlinear response at all wavelengths, as determined by the band-ratio analyses. The 664 nm response has the most structure (on the order of three separate regimes) and the largest deviations from linearity (almost 22%), and the 436 nm response has the largest extent of linearity. The C_{18} calibration from September 2007 exhibits similar properties: the 664 nm response has the most structure and largest deviations (almost 25%), and although the 436 nm response also has the largest extent of linearity, it has more variance than the corresponding data for the C_8 calibration. The data analysis for SeaHARRE-4 is completed, which are based on coastal (Case-2) samples spanning a wide dynamic range in TChl *a* concentration of approximately 1.90–42.70 mg m^{-3}. The SeaHARRE-4 results show the CHORS C_{18} method is significantly out of compliance in terms of being able to produce quantitative results: the average primary pigment uncertainties are 530.9%. If the two worst pigments (But and Hex) are totally removed, the average uncertainty drops to 55.4%. In addition, the average uncertainty for TChl *a* is 34.1%. All of these values exceed the performance metrics for quantitative analysis. The CVO, which had never executed an HPLC method, but had a properly trained chemist using good hardware and a willingness to adhere to the performance metrics and the protocols, had an average primary pigment uncertainty of 27.5% and an average TChl *a* uncertainty of 7.8%. In fact, the CVO results were ranked second for the laboratories satisfying quality-assured performance. The highest quality laboratories LOV (France), CVO (USA), DHI (Denmark), CSIRO (Australia), and HPL (USA) have an overall agreement to within ±5.4% of one another. The overall uncertainty in TChl *a* for these labs all agree to within ±1.7%. Despite strong recommendations from HPL and the CVO to not analyze any samples with the UV6000LP until the nonlinearity is understood, CHORS proceeds with analyzing approximately 600 samples for a non-NASA PI. CHORS analyzes samples for about two weeks, and then the computer freezes. The system is restarted, but it stalls, because it would not trigger. The system is shut down and pumped with MeOH. The system is restarted and brought back up. The third HPLC Workshop is hosted by DHI. All of the SeaHARRE-4 laboratories participate except USF and CHORS. The objectives of the workshop are as follows: a) establish an objective set of criteria for quantitating peaks with coelution problems or signal-to-noise problems—both of which are frequent features of small peaks, but not exclusive to small peaks—so the uncertainty budget is not dominated by false positives and false negatives; b) agree on reporting practices and the numerical resolution of the results; c) determine which pigments and higher-order products should be reported and whether or not the pigments should be classified (e.g., primary, secondary, and tertiary) with differing reporting or performance requirements for each classification; d) finalize the specification of performance metrics and what criteria should be applied to the agreed upon classification scheme; e) determine whether or not algal cultures should be part of SeaHARRE samples; and f) formulate the sampling plan for SeaHARRE-5 (the SeaHARRE-4 results make it clear that the HPLC community faces challenges that need to be addressed in the coastal analysis problem, so it seems sensible to emphasize coastal sampling again).

November All approvals for the modifications to C-LABS are obtained and GSFC Facilities confirms receipt of the work order. The HPLC detector expert suggests common food color dye might provide suitable absorption characteristics for the nonlinear characterization experiments. All the participants in the UV6000LP nonlinearity problem meet in Baltimore to review what has been learned in the last five months, since The Team was formed, and agree on a plan to correct the CHORS data set. The meeting ends with unanimous agreement concerning the following points: a) the UV6000LP has a nonlinear response, but its functional form is unknown, so that has to be determined from the existing calibration data and new higher-resolution calibrations; b) given that the UV6000LP was used with both the C_8 and C_{18} methods, the CHORS results should be considered invalidated, because the methodology used requires a linear response and the system was not able to provide such a response; c) nonlinear calibrations are not an unknown aspect of HPLC methods, so it should be possible to correct the calibrations and data, as long as the calibrations are extensive enough to describe the nonlinearity; d) the occurrence of large outliers during triplicate injections of a calibration standard requires investigation; e) NASA PIs should be queried to find out what pigments are the most important to them; and f) the inventory of the total number of samples involved must be completed, including data from the 1998–2000 time period.

December The CHORS technician begins working for the CVO as a subcontractor. The agreed upon tasks are as follows: a) duplicate all laboratory notebooks and send copies to the CVO; b) duplicate all electronic data sets associated with the production of HPLC results (chromatograms spreadsheets, final quantitations, etc.) on CD-ROM and send the copies to the CVO; c) submit a document describing how to access the information in the laboratory notebooks and the CD-ROM archive; d) make laboratory trials at the direction of the CVO to characterize the nonlinearity in the CHORS detector; e) participate in a training exercise to improve the quantitation of pigments; f) apply the correction methodology that the CVO develops, which will very likely require the re-quantitation of all CHORS chromatograms (but not necessarily all the pigments in each chromatogram); g) deliver the corrected chromatograms and quantitated results (plus any ancillary files) to the CVO on CD-ROM with a duplicate set retained for CHORS; and h) participate in documenting the correction methodology. The first tasks are to organize the laboratory notebooks into a single archive and photocopy them, determine whether or not food color dyes can be used for nonlinearity testing, and prepare a preliminary summary of the C_8 calibration data in order to determine the best way to organize the CHORS calibration data sets. CVO personnel meet with GSFC Facilities to go over the scope of the work for modifying C-LABS to accommodate the needed upgrades for HPLC analysis.

2008

January The food dye tests require fine tuning, because they elute right after the solvent front, and the chromatography software has trouble picking the correct spot to start the baseline. This makes it hard to repeat the same baseline from sample to sample. The C_8 calibration summary is revised and the new format will also be used for the C_{18}

calibration summary. The food dye tests appear sufficiently useful that a solid form of the dye are purchased, so quantitative relationships can be produced (right now relative relationships based on peak areas have been used). GSFC Facilities begins working on the design aspects of the C-LABS modifications. An international HPLC analyst who uses a UV6000LP is contacted to see if another practitioner might provide more insight into the nonlinearity problem.

February HPL begins analyzing HPLC samples for NASA investigators based on the new EOS Recompetition contract. While working on the oldest parts of the calibration archive (which are stored on a tape technology that is no longer supported), the CHORS technician discovers the first calibration in April 2001 was done with two detectors: the new UV6000LP and the old UV2000 detector. This latter is not supposed to have a light pipe flow cell and should be capable of producing linear calibrations. An analysis of the UV2000 data reveals the same problem with aberrant results seen with the September 2007 calibration: there is usually one anomalous injection within each triplicate, and the anomalous result can be as much as 12% different from the other two injections. The international HPLC analyst contacted in January provides an example calibration on a different UV6000LP. The calibration exhibits much of the same nonlinearity seen with the CHORS unit, and the percent residuals to the calibration curve are as high as 80%. The HPLC detector expert suggests replacing the food dye tests with the use of an unequivocally linear detector placed in-line with the UV6000LP. The food dye tests are proving difficult, because they elute too close to the injection front. It is believed that the second in-line detector will provide a) an independent assessment of the UV6000LP nonlinearity and b) a convincing test of whether or not there is more to investigate about the autosampler. The CHORS technician starts to look for a second detector with an established linear response. A second international HPLC analyst who uses a UV6000LP is contacted to see if another practitioner might provide more insight into the nonlinearity problem. Unfortunately, the group involved has replaced their Thermo instrumentation with another manufacturer and express no interest in investigating the problem.

March A third international HPLC analyst who uses a UV6000LP is contacted to see if another practitioner might provide more insight into the nonlinearity problem. GSFC Facilities confirms they are still working on the design aspects of the C-LABS modifications (more than four months after work order submission). The CHORS technician reaches an agreement with Waters to participate in the second in-line detector test.

April The CHORS technician completes the summary of all the C_8 and C_{18} calibrations. An unequivocally linear detector, a Waters 2998 that had just been calibrated at the factory, is put in-line with the UV6000LP detector, and the results show the linear detector does not have the expected linearity. There is a strong suspicion that the Thermo autosampler might also be a source of nonlinearity. The HPLC detector expert and the CHORS technician run tests on the autosampler which indicate the autosampler is operating within specifications. It should be noted that specifications cited by HPLC manufacturers pertain to single draw-and-inject type injections and not the complex

injector programming required for pigment analyses, so these tests that were run are not conclusive. The CHORS technician delivers all the CHORS laboratory notebooks and electronic data sets (approximately 75 CD-ROM disks) to HPL, along with a document describing how to access the information in the laboratory notebooks and electronic archives. The plan is to have the CHORS technician transfer as much knowledge as possible about using the Thermo software and accessing the data archive to the HPL analysts after the combined CC&E and OCRT meeting later in the month. The scientific community is briefed about the current status and future plans regarding the CHORS UV6000LP problem in a working group and in a plenary session of the combined CC&E and OCRT meeting (the presentations are available on-line at http://oceancolor.gsfc.nasa.gov/DOCS/). Greg Mitchell announces he has a very large number of duplicate filters—covering most of the analyses CHORS did of his HPLC samples—that were originally collected for mycosporine-like amino acid and phycoerythrin analyses he has not undertaken. The filters have been stored in liquid nitrogen and a subset could be made available for evaluating the nonlinear correction of CHORS results. HPL determines the work load for participating in the transfer of capabilities planned between HPL and CHORS is too time consuming. All of the CHORS laboratory notebooks and media are transferred to the CVO. The CHORS technician resigns from CHORS (and The Team) effective 31 May 2008—the stress of being the sole person responsible at CHORS is too much. Some of the portions of the C-LABS work order not associated with the ventilation and pure-water system are completed (more than 5 months after submission). In a meeting with the CVO, the international HPLC analyst contacted in March reports the local Thermo representative confirms the nonlinearity of the UV6000LP, but indicates the nonlinearity can be reduced (perhaps to acceptable levels) if a) the working range of the calibration (and, thus, sample analysis) is substantially restricted, and b) if the UV6000LP is optically tuned during each calibration. Neither of these are applicable to the CHORS procedures. HPL determines the enormity of the problem is overwhelming the personnel involved and the contractual obligation with NASA for annual pigment analyses.

May CHORS agrees the NOAA controller (S/N 034/01114-5) will be shipped with the NASA system, so the CVO will have a fully functioning system—the NASA controller (S/N 090/06227 is nonfunctional). The CVO sends a scientist to CHORS to complete the information transfer begun in April and to train with the CHORS technician to learn how to operate the UV6000LP system. All the functioning components are labeled, pictures of everything hooked up are stored on removable media, and the NASA components plus the NOAA controller are packed into boxes for shipment back to the CVO. HPL stops participating in the work The Team is doing. The CVO has the only recurring representation on The Team. The HPLC detector expert agrees to continue working problems on a case-by-case basis.

June CHORS decides their two Thermo HPLC systems should be sent back to their separate agencies of origin, so the NOAA controller (S/N 034/01114-5) goes to NOAA, which means the CVO will receive a UV6000LP system that will not work. The box containing the NASA autosampler is unpacked to remove the NOAA controller, because both controllers were packed with the NASA autosampler (so NASA would receive a

functional system). The CHORS UV6000LP system that SIMBIOS purchased arrives at GSFC. NOAA agrees to lend their working controller to the CVO. The CVO meets with GSFC Facilities to finalize the plan for the ventilation work (seven months after Work Order submission), and an additional meeting sets 25 July 2008 as a drop dead date to finish the work (in anticipation of the new HPLC equipment arriving in September).

July The CVO switches to Science Systems and Applications Incorporated (SSAI) for contracting support, because efforts to hire scientists with M.S. degrees proves impossible after five months of effort with the existing contractor. A position for an HPLC technician is posted.

August The CVO meets with GSFC Facilities to to review why the pure-water system has still not been installed. A candidate for the HPLC technician position is interviewed. Greg Mitchell sends an inventory of the duplicate samples he has and the ones he is willing to donate to the CHORS correction process.

September A commercial company is brought in to complete the installation of the pure-water system and GSFC Facilities completes the electrical work for the combustion oven more than 10 months after the Work Order was submitted (but the oven still cannot be used, because the ventilation work has not begun). GSFC management and NASA HQ agree the CVO needs to move off campus to a facility with a laboratory satisfying the requirements for the equipment the CVO needs to have operational (there are no spare laboratories at GSFC with the requisite ventilation capability). The new CVO HPLC system arrives at GSFC. The NOAA controller for the NOAA UV6000LP system arrives at GSFC. The HPLC technician interviewed in August is offered a position, but the start is deferred until 1 December, because of existing commitments with the candidate's present employer. Additional discussions take place with Haili Wang, who works with Greg Mitchell, about the inventory of the duplicate samples Greg is willing to donate to the CHORS correction process. HPL also participates in the duplicate analysis discussions, because if the duplicates are analyzed, the analysis will probably take place at HPL.

October The collection of the first phase of field samples for SeaHARRE-5 is completed, which are collected during day cruises to the rivers, estuaries, and coastal bays of New Hampshire. The CVO researches commercial properties in near vicinity to GSFC, but no suitable space is found (all require significant investments to bring the laboratories into safety compliance).

November A research facility and small business incubator called BWtech, which is part of the University of Maryland Baltimore County (UMBC), is found to have appropriate laboratory and office space, and is available immediately. The CVO is approved as a possible tenant, because GSFC is a research institute with ties to UMBC and SSAI is a Maryland small business with an emphasis on high-technology and research. Lease negotiations between BWtech and SSAI are started, and a lease starting date for early December is agreed to by both parties. The collection of the second phase of field

samples for SeaHARRE-5 is completed, which are collected during day cruises to the rivers, estuaries, and coastal bays of Tasmanian coastal waters.

December The new HPLC technician begins working at the CVO. Although the Contracting Officer (CO) for the SSAI contract approves the signing of the BWtech lease, this approval is rescinded before the lease is signed. GSFC Facilities puts forth a plan to make C-LABS compliant using a ventilation scheme they have already rejected, in a process that does not include Safety approval, and under a time schedule in contradiction with the facilities support contract. The CVO rejects the plan and again requests approval to move off site to BWtech. GSFC Procurement decides the move can go forward, but SSAI cannot sign the lease or arrange the move—these two functions must be carried out by a large logistics company headquartered in Las Vegas (Nevada).

2009

January SSAI signs a short-term lease with BWtech. The CVO moves to BWtech, which is located in Halethorpe (Maryland) close to Baltimore-Washington International (BWI) airport. A local moving company is used under contract to SSAI. The new laboratory is mostly operational by 21 January. An assessment of the NASA components of the CHORS HPLC capability reveals the autosampler was internally damaged when it was shipped from CHORS to the CVO: the lifting arm "hook" that picks up the sample vials is broken (this is part of the injector pod). This problem was not detected when the equipment was unpacked upon receipt of delivery at GSFC.

February The CVO participates in an e-Bay auction to obtain a used AS3000 from a laboratory in Massachusetts that was using the autosampler before it was removed from laboratory analyses. The photograph shows the injector lifting arm hook is not broken. The CVO wins the auction with a bid of $81 (plus shipping and handling). HPL agrees to help the CVO prepare a report to the community explaining a) what has transpired since the last update, and b) the final recommendations of The Team in terms of what to do about the CHORS HPLC quantitation problems.

March The NASA part of the CHORS Thermo Separations Products HPLC system is brought back on line by scavenging pod parts from the autosampler purchased in the e-Bay auction. This report is submitted to NASA HQ, it is reviewed by five members of the ocean color community, and revised by the authors.

8. References

Bidigare, R. R. and Trees, C.C., 2000: HPLC phytoplankton pigments: Sampling, laboratory methods, and quality assurance procedures. In: Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 2, J. Mueller and G. Fargion, Eds., *NASA Technical Memorandum 2000-209966*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 154–161.

Bidigare, R.R., Van Heukelem, L., Trees, C.C., 2003: HPLC phytoplankton pigments: sampling, laboratory methods and quality assurance procedures. In: Mueller, J.L., Fargion, G.S., McClain, C.L., Eds., Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volume V: Biogeochemical and Bio-optical Measurements and Data Analysis Protocols. *NASA Tech. Memo. 2003-211621/Rev4-Vol.V*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 5–14.

Bidigare, R.R., Van Heukelem, L., Trees, C.C., 2005: Analysis of algal pigments by high-performance liquid chromatography. In: Algal Culturing Techniques. Anderson, R., Ed., Elsevier Academic Press, Burlington, Massachusetts, 327–345.

Claustre, H., Hooker, S.B., Van Heukelem, L., Berthon, J.F., Barlow, R., Ras, J., Sessions, H., Targa, C., Thomas, C.S., van der Linde, D., Marty, J-C., 2004: An intercomparison of HPLC phytoplankton pigment methods using in situ samples: application to remote sensing and database activities. *Mar. Chem.,* **85,** 41–61.

Clesceri, L.S., Greenberg, A.E., Eaton, A.D., Eds., 1998: Part 1020, Quality Assurance, Section 1020C, Quality Assessment in Standard Methods for the Examination of Water and Wastewater. 20th ed. Baltimore (MD): American Public Health Association, American Water Works Association, Water Environment Federation, Baltimore, Maryland, 1–12.

EURACHEM Guide, 1998: The Fitness for Purpose of Analytical Methods, A Laboratory Guide to Method Validation and Related Topics, Mr. David Holcombe, Drafting Secretary for EURACHEM working Group, LGC, Queens Rd. Teddington, Middlesex, TW11 0lY, United Kingdom, http://www.eurachem.org/.

Hooker, S.B., H. Claustre, J. Ras, L. Van Heukelem, J-F. Berthon, C. Targa, D. van der Linde, R. Barlow, and H. Sessions, 2000: The First SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-1). *NASA Tech. Memo. 2000–206892, Vol. 14*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 42 pp.

Hooker, S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Barlow, R., Sessions, H., Schlüter, L., Perl, J., Trees, C., Stuart, V., Head, E., Clementson, L., Fishwick, J., Llewellyn, C., Aiken, J. (2005). The second SeaWiFS HPLC analysis round-robin experiment (SeaHARRE-2). *NASA Tech. Memo. 2005-212785*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 112 pp.

International Union of Pure and Applied Chemists (IUPAC), 1997: Compendium of Analytical Nomenclature: Definitive Rules, Chapter 18, quality assurance of analytical processes, Section 18.4.3.7. Blackwell Science, 3rd publication 1998 and Web edition: http://old.iupac.org/publications/analytical_compendium/

Jeffrey, S.W., Mantoura, R.F.C, Wright, S.W., 1997: *Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods*, UNESCO Publishing, Paris, 661 pp.

King, P.G., 1999: HPLC method development and validation: A direct procedure for determining the linear-through-zero range, *LC/GC,* **6,** 46.

Mantoura, R.F.C, Repeta, D.J., 1997: Calibration methods for HPLC. In: Jeffrey, S.W., Mantoura, R.F.C, Wright, S.W., Eds., Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods, UNESCO Publishing, Paris, 343–360.

Snyder, L.R., Kirkland, J.J., 1979: *Introduction to Modern Liquid Chromatography*, 2nd ed., Wiley, New York, 863 pp.

Van Heukelem, L., Thomas, C.S., 2001: Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J. Chrom. A.,* **910,** 31–49.

Van Heukelem, L., Thomas, C.S., Glibert, P., 2002: Sources of variability in chlorophyll analysis by fluorometry and high-performance liquid chromatography in a SIMBIOS inter-calibration exercise. *NASA Tech. Memo. 2002-211606*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 50 pp.