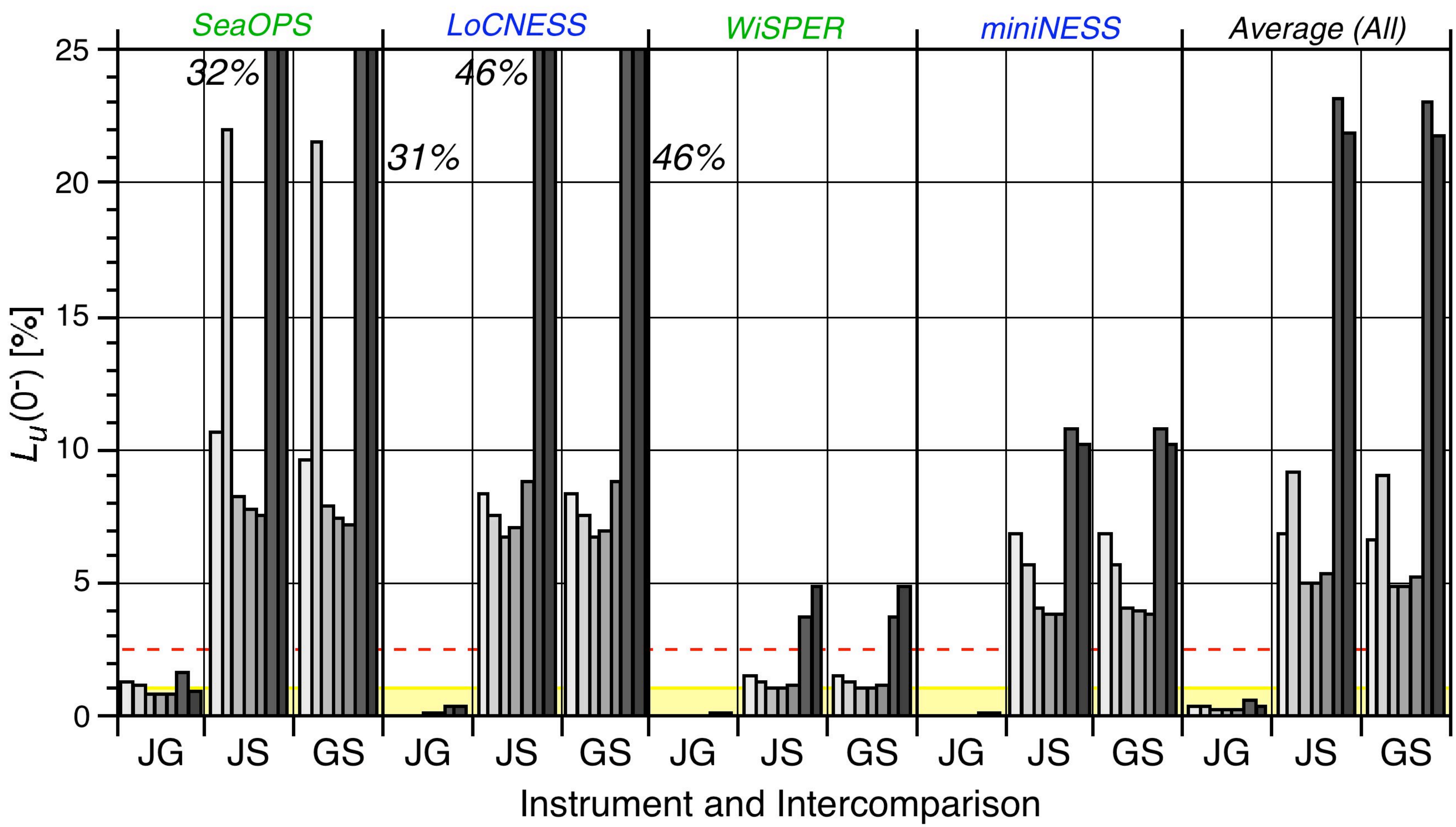
Introduction to the AOP Data Processor Workshop

Stanford Hooker
NASA/CVO
Baltimore, Maryland

Rationale for Specifying a Community-Maintained, Web-Based, Open-Source AOP Data Processor

One of the largest—and fortunately, one of the most accessible—sources of uncertainty is the data processing scheme used to convert raw field observations into usable products with geophysical units.



Benefits of a Community-Maintained, Open-Source, Web-Based AOP Data Processor

The artificial variance imparted by processor-to-processor differences frequently equals or exceeds the total uncertainty permitted in calibration and validation field activities, which is no more than a few percent (currently about 3.5%). The principal objective of this workshop is to specify the requirements of a community-maintained, open-source, Web-based interface for the Processing of Radiometric Observations of Seawater using Information Technologies, (PROSIT†). Envisioned outcomes beneficial to the NASA Ocean Biology and Biogeochemistry Program include the following:

- The definition and promulgation of standards for ocean color data formats, which facilitates the exchange of information between investigators.
- The exploitation of an open-source interface to provide a community-wide quality assurance capability and, thus, a self-consistent archive of AOP data from the contributions of different investigators from all around the world.
- The capability for automatically capturing raw data, plus all the needed ancillary data, as part of the data ingest module, as well as, the automated enforcement of the applicable data policy and the automated delivery of processed data to the appropriate database (SeaBASS).

† *From Latin, literally meaning "May it benefit".*

Benefits of a Community-Maintained, Open-Source, Web-Based AOP Data Processor (*cont.*)

- The promulgation among the ocean color community of data analysis techniques based on state-of-the-art information engineering solutions that allow a researcher to: a) interactively query the AOP database, b) validate existing ideas and formulate new hypotheses, c) identify abnormal conditions, and d) validate the algorithms used for generating ocean color data products and understand their range of applicability.
- The creation of a unique environment for the processing (or the reprocessing, when new techniques or refinements become available) and the analysis of in-water optical data, which allows: a) the evaluation of the instruments and protocols used for in-water radiometric measurements, b) a reduction in overall uncertainties by removing interlaboratory differences in data products, and c) resource allocation gained from not paying for the development and maintenance of multiple applications—likely, one for each laboratory—which make similar computations.
- The entire concept should be transportable to other types of processing, for example, IOPs.

AOP Specific Workshop Objectives

Assuming the basic architecture of the processor is open-source code, Web-based access, and community-maintained functionality, the needed specificity for what the workshop will accomplish comes mostly from establishing the required and desired features of what the processor will do. Additional objectives are derived from the inevitable link between the Ocean Optics Protocols, which provide a baseline (but dated) set of community-approved procedures for many of the features, plus the need to establish performance metrics for quality assurance:

- Specify the required features of the processor (basic functionality, data types and formats to be ingested, parameterizations and formulations for calculations, primary and ancillary data products, corrections to the various data types, data quality assessment and flagging, automated enforcement of data policy and submission to SeaBASS, etc.).
- Specify the desired features of the processor (training module, field transportable version, data product visualization, etc.).
- Specify any needed changes to the Ocean Optics Protocols (Case-1, Case-2, hyperspectral, fixed wavelength, winch and crane, free-fall, etc.).
- Specify performance metrics and quality levels for AOP processing, e.g., routine research, semi-quantitative analysis, quantitative assessment (calibration and validation), state of the art.

Performance Metrics

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

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

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

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

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

Workshop Agenda

There are six working groups, with two further divided into subgroups as follows:

- A Buoys
- B1 Winch and Frame Profilers
- B2 Free-Fall Profilers
- C1 Hyperspectral Sensors
- C2 Fixed-wavelength Sensors
- D Case-2 (Shallow) Waters
- E Case-1 (Deep) Waters
- F Performance Metrics

<table>
<tr><th>Time</th><th>13 January (Tue)</th><th>14 January (Wed)</th><th>15 January (Thu)</th></tr>
<tr><td>0830</td><td>Welcome</td><td>Welcome</td><td>Welcome</td></tr>
<tr><td>0845</td><td rowspan="2">Workshop Introduction (Hooker)</td><td>Legacy Processors (Siegel)</td><td rowspan="2">Working Group Report A and Discussion</td></tr>
<tr><td>0900</td><td rowspan="2">SeaBASS Lessons (Werdell)</td></tr>
<tr><td>0915</td><td>Survey Summary (Hooker)</td><td rowspan="2">Working Group Report B and Discussion</td></tr>
<tr><td>0930</td><td rowspan="2">Web-Based Processor (Hooker)</td><td rowspan="2">Practical Aspects of Calibration and Validation Quality (Bailey)</td></tr>
<tr><td>0945</td><td rowspan="3">Working Group Report C and Discussion</td></tr>
<tr><td>1000</td><td>CVO Processor (Hooker)</td><td rowspan="2">Hyperspectral Processing and Case-2 Considerations (Dana)</td></tr>
<tr><td>1015</td><td>ODU Processor (Zimmerman)</td></tr>
<tr><td>1030</td><td>Break</td><td>Break</td><td>Break</td></tr>
<tr><td>1100</td><td>UCSB Processor (Siegel)</td><td rowspan="6">Working Groups (A and B)</td><td rowspan="2">Working Group Report D and Discussion</td></tr>
<tr><td>1115</td><td>USF Processor (English)</td></tr>
<tr><td>1130</td><td>LOV Processor (Antoine)</td><td rowspan="2">Working Group Report E and Discussion</td></tr>
<tr><td>1145</td><td rowspan="2">SeaBASS Processor (Werdell and Bailey)</td></tr>
<tr><td>1200</td><td rowspan="2">Working Group Report F and Discussion</td></tr>
<tr><td>1215</td><td>Lamont Processor (Subramanium)</td></tr>
<tr><td>1230</td><td rowspan="3">Lunch</td><td rowspan="3">Lunch</td><td rowspan="3">Lunch</td></tr>
<tr><td>1300</td></tr>
<tr><td>1330</td></tr>
<tr><td>1400</td><td>Scripps Processor (Kahru)</td><td rowspan="6">Working Groups (C and D)</td><td rowspan="6">Plenary Discussions: Required and Desired Features of a Web-based Community Processor, Changes to the Ocean Optics Protocols, and Performance Metrics</td></tr>
<tr><td>1415</td><td rowspan="2">Biospherical Processor (Morrow)</td></tr>
<tr><td>1430</td></tr>
<tr><td>1445</td><td>Scripps Processor (Stramski)</td></tr>
<tr><td>1500</td><td rowspan="2">HOBI Labs Processor (Dana)</td></tr>
<tr><td>1515</td></tr>
<tr><td>1530</td><td>Break</td><td>Break</td><td>Break</td></tr>
<tr><td>1600</td><td>WHOI Processor (Sosik)</td><td rowspan="6">Working Groups (E and F)</td><td rowspan="6">Plenary Discussions: Workshop Report and Writing Assignments</td></tr>
<tr><td>1615</td><td>FURG Processor (Garcia)</td></tr>
<tr><td>1630</td><td>NOAA Processor (Wilson)</td></tr>
<tr><td>1645</td><td>Scripps Processor (Kozlowski)</td></tr>
<tr><td>1700</td><td rowspan="2">MLML Processor (Flora and Johnson)</td></tr>
<tr><td>1715</td></tr>
<tr><td>1730</td><td>Adjourn</td><td>Adjourn</td><td>Adjourn</td></tr>
</table>