Expanded Signal to Noise Ratio Estimates for Validating Next-Generation Satellite Sensors in Oceanic, Coastal, and Inland Waters

Raphael M. Kudela $^{1, *}$, Stanford B. Hooker $^{2}$, Liane S. Guild $^{3}$, Henry F. Houskeeper $^{4}$ and Niky Taylor $^{5}$

$^{1}$ Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA 95064, USA
$^{2}$ NASA Goddard Space Flight Center Code 616.2 Bldg. 28 Rm. W120D, Greenbelt, MD 20771, USA
$^{3}$ Biospheric Science Branch, Earth Science Division, NASA Ames Research Center, Moffett Field, CA 94035, USA
$^{4}$ Applied Ocean Physics & Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
$^{5}$ U.S. Geological Survey Western Geographic Science Center, 350 N. Akron Rd., Moffett Field, CA 94035, USA

* Correspondence: kudela@ucsc.edu; Tel.: +1-831-459-3290

Abstract: The launch of the NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) and the Surface Biology and Geology (SBG) satellite sensors will provide increased spectral resolution compared to existing platforms. These new sensors will require robust calibration and validation datasets, but existing field-based instrumentation is limited in its availability and potential for geographic coverage, particularly for coastal and inland waters, where optical complexity is substantially greater than in the open ocean. The minimum signal-to-noise ratio (SNR) is an important metric for assessing the reliability of derived biogeochemical products and their subsequent use as proxies, such as for biomass, in aquatic systems. The SNR can provide insight into whether legacy sensors can be used for algorithm development as well as calibration and validation activities for next-generation platforms. We extend our previous evaluation of SNR and associated uncertainties for representative coastal and inland targets to include the imaging sensors PRISM and AVIRIS-NG, the airborne-deployed C-AIR radiometers, and the shipboard HydroRad and HyperSAS radiometers, which were not included in the original analysis. Nearly all the assessed hyperspectral sensors fail to meet proposed criteria for SNR or uncertainty in remote sensing reflectance ($R_{rs}$) for some part of the spectrum, with the most common failures (>20% uncertainty) below 400 nm, but all the sensors were below the proposed 17.5% uncertainty for derived chlorophyll-a. Instrument suites for both in-water and airborne platforms that are capable of exceeding all the proposed thresholds for SNR and $R_{rs}$ uncertainty are commercially available. Thus, there is a straightforward path to obtaining calibration and validation data for current and next-generation sensors, but the availability of suitable high spectral resolution sensors is limited.

Keywords: signal-to-noise ratio; ocean color; coastal and inland waters; NDVI; kelp; chlorophyll

1. Introduction

Coastal, estuarine, and inland waters are optically complex and exhibit significant spatial and temporal variability. Ocean color remote sensing imagery is often used to derive relevant biogeochemical constituents within coastal and inland waters at appropriate space and time scales [1]. Beyond basic research, such data products are used to provide information about species composition and biodiversity, water quality, carbon cycling, and biogeochemical fluxes relevant to human and ecosystem health [2]. Historically, ocean color sensors have focused on the open ocean, which is typically optically simple (Case-1 waters; [3]) and exhibits greater spatial and temporal decorrelation scales compared to coastal and inland waters. The coastal zone serves as a dynamic transition between the oceanic and terrestrial biomes. It represents approximately 20% of Earth’s surface [4] but
provides the richest biodiversity and the highest fishery production, both of which are created from the highest primary production [4]. As such, it is an important target for remote sensing. Advances in satellite radiometry have greatly increased the availability of publicly available high spatial resolution data with the introduction of (e.g.) the Operational Land Imager (OLI; 30 m spatial resolution) aboard Landsat 8/9 and the MultiSpectral Instrument (MSI; 10, 30, 60 m spatial resolution) aboard Sentinel-2, while the launch of the Ocean Land Colour Instrument (OLCI) aboard Sentinel-3 provides a compromise between high spatial resolution (300 m) and more frequent return rate (~daily; [5]). Increasingly, the community of practice is now turning toward increased spectral resolution as well [6].

Water-leaving radiance \( L_W(\lambda), \mu W \text{ cm}^{-2} \text{ nm}^{-1} \text{ sr}^{-1}, \) or \( L_W \) is highly variable, ranging from extremely low values for clear, deep water to very bright values in optically shallow environments near the shoreline, in turbidity plumes, or in high productivity waters. Legacy satellite missions such as SeaWiFS, MODIS, and VIIRS provide ~1 km and ~daily resolution, which is generally not adequate for the coastal ocean [7]. Next-generation missions such as PACE, deploying the Ocean Color Instrument (OCI; [8]), and the NASA Surface Biology and Geology (SBG) study [9] provide more spectral resolution as well as measurements that extend into shorter wavelengths. These shorter wavelengths, extending into the ultraviolet (UV), are useful for the following: (a) discriminating red tides [10]; (b) identifying point sources for pollution [11]; (c) improving atmospheric correction [12], particularly in turbid coastal waters [13,14]; and (d) reducing the uncertainties in derived biogeochemical parameters at discrete wavelengths [15] or as part of end-member analyses [16]. Most commercial-off-the-shelf (COTS) spectrometers and calibration/validation (cal/val) sensors, however, lack UV bands or exhibit poor UV performance [17]. All these issues make aquatic remote sensing, as well as cal/val activities, challenging.

The “classic” airborne visible infrared imaging spectrometer (AVIRIS-C) provides both high spatial (~5–60 m) and spectral (~400–2500 nm) resolution, and has until recently been the platform of choice for simulating ocean color products [17]. It has been complemented by the next-generation airborne sensor AVIRIS-Next Generation (AVIRIS-NG). Unfortunately, in-water and above-water instrumentation suitable for ground truth of high spectral resolution imagery has decreased in availability [18]. Handheld sensors such as the Malvern Panalytical ASD instruments are considered the gold standard but are not automated and are therefore incapable of providing continuous spectral data in complex environments. The Aerosol Robotic Network-Ocean Color (AERONET-OC) sensors are automated, but are typically deployed in a fixed location [19,20].

As previously described [21,22], the remote sensing of coastal and inland waters often involves compromises between optimizing the SNR and maintaining a broad dynamic range for ocean sensors, and saturation over bright targets caused by limited dynamic range is a common issue. In contrast, sensors designed primarily for land (e.g., MSI, SBG) are optimized for observing bright targets and do not saturate but suffer from poor performance (low SNR) for dark targets, i.e., clear, deep water. As a compromise, SNR recommendations are variously set at an SNR of > 100–200 for shortwave infrared (SWIR), >600 for near infrared (NIR), and > 1000 for UV to visible (VIS), i.e., UV-VIS, bands. Muller-Karger et al. [2] recommended an SNR > 800 for the UV-VIS range. Wang and Gordon [23] specifically evaluated NIR and SWIR bands for atmospheric correction, and identified minimum requirements of ~200–300 for NIR and ~100 for SWIR, concluding that an SNR of ~600 and ~200 for NIR and SWIR are acceptable minimum thresholds. SNR values listed for at-sensor radiances can be difficult to interpret because of the wide dynamic range of typical at-sensor radiance \( L_{\text{typ}} \) values, particularly for coastal and inland waters, and [21] recommended presenting values in terms of \( L_W \), remote sensing reflectance \( R_{rs}, \text{sr}^{-1} \), or water-leaving reflectance \( \rho_W \), consistent with the PACE Science Definition Team report [24].

The goals of this manuscript are, first, to update [21] with sensors that were not included in the original analysis, in particular AVIRIS-NG. Second, this paper assesses whether existing commercially available airborne and in-water instrumentation provides
sufficiently high-quality estimates of $L_W$ (and therefore of $R_s$ and $\rho_W$). We provide guidelines for future cal/val instrument development to achieve an adequate SNR (and uncertainty) and, therefore, cal/val capability for next-generation sensors.

2. Materials and Methods

2.1. Field Sites and Targets

The sampling locations chosen for this analysis (Figure 1) were based on similar criteria as [21]. First, the sites provide $L_{typ}$ values that are representative of common inland and coastal water targets; with the exception of kelp, all the targets were spatially uniform water targets referred to as “bright” or “dark”, i.e., waters with enhanced scattering (bright) or absorption (dark). Second, quality-controlled data are available (see references in Section 2), and the sites are characterized well enough to assess their use as case studies for coastal and inland ocean color remote sensing. Third, the sites are both spatially extensive and optically homogenous for applying geostatistical methods to calculate the SNR [21,25,26]. Finally, datasets were excluded that were redundant with the previous analysis [21]. Based on those criteria, we identified five relevant sites as follows: San Francisco Bay, the Sub-Mesoscale Ocean Dynamics Experiment (S-MODE) site offshore of San Francisco, coastal Monterey Bay and Elkhorn Slough, the Santa Barbara Channel, the Gulf of Mexico, and coastal Hawaii (Table 1).

Figure 1. Data collection locations used in this analysis, plotted as solid black circles. (A) Coastal California and the Gulf of Mexico; (B) expanded view for California; (C) expanded view for the Gulf of Mexico; and (D) a dark target from Hawaii, with Marine Optical Buoy (MOBY) for reference.
Table 1. Summary of data collection sites and processing parameters.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Date</th>
<th>Resolution (m)</th>
<th>Spectral Range (Resolution) nm</th>
<th>Atmospheric Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-MODE, Monterey Bay, Elkhorn Slough</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-AIR</td>
<td>28-Oct-21</td>
<td>3.5</td>
<td>320–1640 (10) *</td>
<td>Not Necessary</td>
</tr>
<tr>
<td>PRISM</td>
<td>24-Jul-12</td>
<td>1.0</td>
<td>350–1054 (3.5)</td>
<td>ATREM</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HydroRad-3</td>
<td>23-Oct-19</td>
<td>95.8</td>
<td>350–850 (1.1)</td>
<td>Not Necessary</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HyperSAS</td>
<td>22-Apr-09</td>
<td>16.2</td>
<td>350–800 (3.5)</td>
<td>Not Necessary</td>
</tr>
<tr>
<td>Hawaii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRISM</td>
<td>17-Feb-17</td>
<td>16.2</td>
<td>350–1054 (3.5)</td>
<td>ATREM</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVIRIS-NG</td>
<td>17-Apr-21</td>
<td>7.9</td>
<td>380–2510 (5)</td>
<td>ATREM</td>
</tr>
</tbody>
</table>

* fixed wavelengths at 320, 340, 380, 412, 443, 490, 510, 532, 555, 589, 625, 670, 683, 710, 780, 875, 1020, 1245, 1640 nm.

2.2. In-Water, Airborne, and Satellite Sensors

Three airborne and two shipboard instrument suites were evaluated, respectively, as follows: AVIRIS-NG, PRISM, and Coastal Airborne In Situ Radiometers (C-AIR); plus the HOBI Labs HydroRad-3 and SeaBird Electronics (formerly Atlantic) HyperSAS (Table 1). AVIRIS-NG and PRISM are imaging spectrometers while the remaining instruments are spectrometers or multiple-waveband radiometers. AVIRIS-NG data were collected over the Santa Barbara Channel in support of the NASA Student Airborne Research Project (SARP) activity and were acquired as Level 2 (atmospherically corrected) data from JPL (https://avirisng.jpl.nasa.gov, accessed 28 August 2023). PRISM data were obtained during the PRISM validation campaign (Moss Landing Harbor, adjacent to Elkhorn Slough; prism.jpl.nasa.gov, accessed 12 September 2023) and the NASA COral Reef Airborne Laboratory (CORAL; [27]). Data from C-AIR were obtained as part of the NASA Coastal High Acquisition Rate Radiometers for Innovative Environmental Research (C-HARRIER) campaign in collaboration with the NASA Sub-Mesoscale Ocean Dynamics Experiment (S-MODE) experiment for offshore coastal waters, Monterey Bay, and Elkhorn Slough, California. Data from San Francisco Bay collected with the HydroRad-3 were obtained from a previous analysis [25] focusing on the derivation of Total Suspended Solids (TSS). HyperSAS data were collected in the Gulf of Mexico [28] and were obtained from the NASA SeaBASS repository [29]. Three of the four instrument suites collect data along-track, while AVIRIS-NG and PRISM collect a swath. For geostatistical analyses, we treated all the sensors as along-track. For each dataset, a homogenous along-track section used for the estimation of the SNR was identified by visual inspection of radiance or reflectance data at ~555 nm to avoid obvious features such as fronts, kelp canopy edges, and shadows.

2.2.1. AVIRIS-NG Imagery

Images for the Santa Barbara Channel were obtained from the JPL AVIRIS-NG site. The data were downloaded as Level 2 (L2) reflectance with a standard atmospheric correction applied [30]. Imagery was kept at native pixel and wavelength resolution (Table 2). Two regions were identified for analysis corresponding to a bright target (kelp bed) and dark target (clear Case-1 waters to the south of Santa Cruz Island). We compared the along-track SNR to a 2-dimensional data field [21] as well as multiple along-track datasets from the same swath and the results were comparable, so only the single along-track results are presented. $R_{ys}$ was calculated using the Thuillier [31] solar irradiance spectra interpolated to AVIRIS-NG wavelengths.
2.2.2. PRISM Imagery

Images for Moss Landing Harbor, adjacent to Elkhorn Slough, and coastal waters off Maui, Hawaii, were obtained from the JPL PRISM site and the CORAL site, respectively. The data were downloaded as L2 reflectance with a standard atmospheric correction applied [30]. The imagery was kept at native pixel and wavelength resolution (Table 2). Moss Landing Harbor was chosen as a bright target (elevated TSSs) and Hawaii was chosen as a dark target (clear Case-1 waters). The PRISM data exhibit “noise” due to the high spectral resolution, and the spectra and uncertainties can be improved with non-standard processing [30], but for this analysis the standard L2 data were used.

2.2.3. C-AIR Data

The C-AIR instrument suite is manufactured by Biospherical Instruments Inc. [32] and operated by NASA Ames Research Center. As part of the C-HARRIER campaign [32], the data were collected aboard a Twin Otter aircraft. C-AIR is an airborne radiometer instrument suite based on microradiometers with a flight campaign heritage since 2011. C-AIR consists of three 19-channel microradiometers [33], one fitted with a cosine collector to measure the global solar irradiance ($E_s; \mu W cm^{-2} nm^{-1}$), and two radiance instruments oriented to measure the indirect sky radiance ($L_i; \mu W cm^{-2} nm^{-1} sr^{-1}$) plus the total radiance from the surface ($L_T; \mu W cm^{-2} nm^{-1} sr^{-1}$). C-AIR is similar to the Compact-Airborne Environmental Radiometers for Oceanography (C-AERO) instrument suite [21] except for less advanced hardware (and firmware) components, the lack of radiance shrouds, and a 15 Hz sampling rate rather than as fast as 30 Hz for C-AERO [34]. In addition, the C-AIR data software is less sophisticated than C-AERO, with the former data acquisition being more properly representative of non-commanded data recording. For the latter, temporally updated laboratory metadata are combined with unique ancillary sensor records. Briefly, the combination permits preprocessing corrections to reduce uncertainties in data products [35] as follows: (a) time-dependent characterizations; (b) gain-stage transitions; (c) nonreal-time serial timing; (d) illumination geometry and normalizations; (e) planar aperture tilting; and (f) environmental or field characterizations (e.g., dark currents), as appropriate. Intercalibration corrections are not applied, because the absolute calibration of each radiometer is determined independently. Postprocessing corrections are likewise most advanced for the C-AERO data acquisition software, e.g., permitting glint discretization shown to remove positive bias (brightening) in above-water radiometric observations [36]. Lowest safe altitude (LSA) flights, ~30 m above the water surface and flown in the principal plane of the Sun to mitigate glint, were used for the data collection, negating the need for a full atmospheric correction [32]. The SNR was estimated using the $L_T$ sensor. It was not possible to calculate $R_{rs}$ for C-AIR because there were insufficient resources to fully process the data in compliance with the NASA Ocean Optics Protocols [37–41].

2.2.4. HydroRad-3 Data

The HydroRad-3 was deployed on the R/V Peterson in San Francisco Bay [25]. The HydroRad-3 recorded continuous radiometric measurements of $E_s$, $L_i$, and $L_T$. As described previously [25], the optical sensors were fixed 2 m above the main deck on the bow of the ship. The sensors were oriented such that a 100–130° average solar zenith angle was maintained during the data collection. $L_T$ was fixed 40° down from horizontal. While $R_{rs}$ was calculated [25], the $L_T$ data were used for the calculation of the SNR for consistency. The data were quality-controlled for suspect spectra [25] prior to analysis.

2.2.5. HyperSAS Data

The HyperSAS was operated aboard the R/V Cape Hatteras in the northern Gulf of Mexico as part of the GulfCarbon program [28,29]. The HyperSAS system included two HyperOCR-R radiance sensors for $L_T$ and $L_i$, plus two HyperOCR irradiance sensors for measuring $E_s$, covering both UV (350–400 nm) and visible (400–800 nm) wavelengths. The radiometers were mounted at 8 m height (above-water) with a 90° angle relative to the
ship’s heading. The $L_T$ and $L_i$ sensor viewing angles were $35^\circ$ from the nadir and $125^\circ$ from the nadir, respectively. The data were quality-controlled as described in [28]. As with the HydroRad-3, $L_T$ was used in this analysis for consistency.

2.3. Signal-to-Noise Ratio and Uncertainty Calculations

The SNR was estimated as the ratio of the mean signal for an invariant target to the standard deviation of the signal. The methodology has previously been described in detail [21]; we reproduce that description here as a background to the analysis. For orbital or airborne sensors designed to image water, the SNR is most often measured in the laboratory, using a spectrally uniform albedo of 5%, a reasonable reflectance for most aquatic targets [42]. Estimates of the SNR convolve multiple sources of uncertainty (noise). This can include instrument artifacts and uncertainties identified as part of the lab-based characterization of an instrument (discussed in greater detail by [43]). An alternative method, referred to as “geostatistical SNR”, calculates a semivariogram from field data that are spatially uniform. This provides an SNR that is relevant to the investigator [26] and can incorporate any pre- and post-processing, such as atmospheric correction schemes and any data reduction or quality control, if applicable. Semivariogram analysis requires that the data exhibit both isotropy and stationarity, and it is assumed that the data do not change spatial resolution [26]. The airborne data collection was at fixed altitude while the shipboard instruments were in fixed geometry relative to the target, but the assumption of isotropy may not have been rigorously met for all data. For the imaging sensors (AVIRIS-NG, PRISM), this was mitigated by testing for isotropy using the two-dimensional data, while explicitly using along-track data for analysis to be consistent across sensors. For the point-based sensors, it is not possible to test for isotropy without collecting data in multiple orientations. Since those data are not available, we analyzed multiple line segments and chose data that were qualitatively and quantitatively most uniform (see Section 2.2).

We applied a geostatistical SNR approach, generally following the methodology outlined in [25]. The SNR was determined for each wavelength by calculating the semivariance, $\gamma_h$, over the distance between pixel pairs ($h$) using the MATLAB (Mathworks Inc.) packages variogram and variogramfit [44]. A theoretical semivariogram was calculated with the data to estimate the nugget, sill, and range (c.f. [25, 45]). The SNR was then calculated as the mean signal ($z$, $L_T$ or reflectance) divided by the square root of the nugget variance ($C_0$):

$$\text{SNR} = \frac{z}{\sqrt{C_0}}.$$  

(1)

The nugget ($n$, or $C_0$) is the non-zero intercept, and it determines the degree of unresolved variability, or noise. Mathematically, this is the non-zero limit of $\gamma_h$ when $h$ approaches zero, where $h$ is the lag and $\gamma_h$ is the semivariance as a function of the lag. The range ($a$) determines how quickly in space the variability reaches a global maximum, while the sill ($C_1$) indicates the point beyond which pixel proximity does not correlate with the spatial structure of the data. This provides the total resolved variance. Multiple theoretical variogram models are available and the choice of model depends on the structure of the data. For this analysis, three theoretical models were used and are formulated as:

$$\gamma_h = C_0 + C_1 \left[ 1 - \exp \left( -\frac{h^2}{a^2} \right) \right]$$  

(2)

$$\gamma_h = C_0 + C_1 \left[ 1 - \exp \left( -\frac{h}{a} \right) \right]$$  

(3)
where (2) is a Gaussian, (3) is an exponential, and (4) is a bounded linear model. Following [25], the best R² value of the fit was used to select the model. Data were discarded for R² < 0.7 for all but the HyperSAS data, where the criteria had to be relaxed to < 0.5 to cover the full spectral range. For sensors other than HyperSAS the R² values exceeded 0.8–0.9 on average.

2.4. Uncertainties in $R_{ns}$ and Derived Geophysical Products

For a comparison to [21], we multiplied the percent noise of the mean signal by the $R_{ns}$ uncertainty for the HydroRad-3, HyperSAS, PRISM, and AVIRIS-NG (C-AIR was excluded because $R_{ns}$ was not calculated). Following [21,46], the same weightings in a 3-band configuration were used for calculating chlorophyll-a ($chla$) from the HydroRad-3, HyperSAS, and AVIRIS-NG (dark target). For kelp, the normalized difference vegetation index (NDVI) was calculated using AVIRIS-NG (bright target). The relevant algorithms are:

$$\log_{10}(chla) = a_0 + \sum_4 a_i \log_{10}(R_{bg})$$

(5)

where $R_{bg}$ is the ratio of maximum blue to green reflectance, and $a_0$–$a_4$ are weighting functions for instrument-specific bands (c.f. [47–50]) and for NDVI,

$$NDVI = \left( \frac{R_{NIR}^{ns} - R_{red}^{ns}}{R_{NIR}^{ns} + R_{red}^{ns}} \right)$$

(6)

where $R_{red}^{ns}$, $R_{NIR}^{ns}$ are the remote sensing reflectances from AVIRIS-NG at 652.55 nm (red) and 852.89 nm (NIR). Uncertainty was propagated using the standard OC3M band-ratio algorithm for chlorophyll retrievals following [43], and for NDVI for the kelp target following [21], where $\delta chla$ and $\delta NDVI$ are the calculated uncertainties with units of mg m$^{-3}$ for $chla$ and dimensionless units (−1 to 1) for NDVI.

3. Results

3.1. Geostatistical SNR

Previous analysis [21] provided SNR estimates for multiple satellite and airborne sensors for comparison to AVIRIS-NG, PRISM, C-AIR, HydroRad-3, and HyperSAS. Figure 2 shows a subset of those results for AVIRIS-C and C-AERO. C-AERO exhibited the highest SNR and lowest uncertainty across all wavelengths. AVIRIS-C exhibited rapid declines below ~450 nm and above ~700 nm with a peak SNR between ~490–700 nm depending on the target. As discussed in [21], an SNR of two provides the absolute minimum threshold for scientifically relevant data. The at-sensor signal for aquatic targets is assumed to be ~5% of $L_{typ}$. Therefore, for an at-sensor SNR of 400–1000, the atmospherically corrected SNR should be 20–50. Doubling those values to account for other sources of variance results in a minimum SNR of 40–100. Figure 2 thus provides examples of an airborne sensor that meets proposed standards (C-AERO) and an airborne sensor (AVIRIS-C) that does not consistently meet proposed standards.

Figure 3 provides the corresponding SNR for the airborne and shipboard datasets included in this analysis. AVIRIS-NG exhibits higher SNR compared to AVIRIS-C, but still falls below a threshold of 20 for blue-UV (<~450 nm) and for wavelengths beyond 750 nm, although those bands are of less interest for ocean color (but see [51]). C-AIR is consistently above both SNR thresholds in the UV and visible bands, but also drops off for wavelengths beyond 750 nm, particularly for dark targets (e.g., S-MODE). C-AIR exhibits about a decade of loss in the SNR compared to C-AERO. For the two shipboard sensors, the SNR peaks at ~555 nm, just at or below the proposed SNR threshold of 20. The SNR continues to drop into the UV and the NIR ranges.
Figure 2. SNR calculated from field measurements for C-AERO (circles) and AVIRIS-C (diamonds). Filled symbols are dark target results; open symbols are bright target results. The solid horizontal lines denote recommended uncertainty levels (values should be below the lines). Reproduced from [21].

Figure 3. SNR for dark and bright targets. (A) SNR for Santa Barbara Channel from AVIRIS-NG; (B) SNR for bright targets in Monterey Bay from C-AIR; (C) SNR from the S-MODE site (dark targets) from C-AIR. S-MODE-1 and S-MODE-2 are data flying into and out of the principal plane of the Sun, respectively; (D) SNR for Moss Landing Harbor (bright target) and Hawaii (dark target) from PRISM; (E) SNR for San Francisco Bay and Gulf of Mexico (both bright targets) from HydroRad-3 and HyperSAS. For each panel the proposed SNR thresholds (2, 20–50, 40–100; [21]) are indicated for reference.
3.2. Airborne and Satellite-Derived Uncertainty

Figure 4 provides the $R_{\text{rs}}$ spectra and corresponding data for AVIRIS-NG, PRISM, HyperSAS, and HydroRad-3 as $R_{\text{rs}}$ and uncertainty (sr$^{-1}$), and as percent environmental uncertainty. C-AIR was excluded because $R_{\text{rs}}$ was not available, but the results from [21] provide a reasonable estimate of C-AIR instrument uncertainty. Hooker et al. [52] provide guidelines for the generalized spectral properties of dark and bright water bodies in terms of peaks and spectral end members as follows: (1) a radiometrically dark water body has a central peak spanning the blue–green domains with the lesser amplitudes at shorter wavelengths containing secondary features and the substantially decreasing amplitudes at longer wavelengths containing an identifiable fluorescence peak in the red domain; and (2) a radiometrically bright water body has a dominant peak shifted toward the red–NIR domains with a usually identifiable fluorescence peak, plus secondary features (including lesser peaks) at shorter wavelengths, wherein with respect to a dark water body, the shortest wavelengths have significantly reduced amplitudes, but the longest wavelengths have elevated amplitudes. The generalized properties of dark and bright water bodies are not applicable to all the sampling circumstances and are provided to assist the reader in generally assessing the material provided in the figures without the need to state all the discernible features within or between figures. Consequently, only the most important differences relevant to the study are described below.

![Figure 4](image-url)

Figure 4. Data from (A) AVIRIS-NG (Santa Barbara Channel; black = dark target, gray = bright (kelp) target) for $R_{\text{rs}}$ (solid and dashed lines) and $R_{\text{rs}}$ uncertainty; (B) HyperSAS (black; Gulf of Mexico) and HydroRad-3 (gray; San Francisco Bay) for $R_{\text{rs}}$ (black and gray lines) and $R_{\text{rs}}$ uncertainty; (C) PRISM (Hawaii = black, dark target; Moss Landing Harbor = gray, bright target) and $R_{\text{rs}}$ uncertainty. Panels (D–F) show corresponding percent uncertainty with proposed thresholds as solid horizontal black lines (data should be below the line). Missing data at some wavelengths represent failure to converge on an acceptable fit for the variogram analysis. Panel (B) includes a dashed horizontal line for 0 $R_{\text{rs}}$ and $R_{\text{rs}}$ uncertainty; values below the dashed line represent negative calculated reflectances. Data from AVIRIS-NG (panels (A,D)) were truncated at 900 nm.
The PACE Science Definition team specified percent uncertainty ranges of 20% for 350–400 nm, 5% for 400–600 nm, and 10% for 660–710 nm [24]. AVIRIS-NG does not provide data from 350 to 400 nm and was generally below the thresholds for dark targets, but not for bright targets. PRISM does extend below 400 nm for the dark (but not bright) target and is below the 20% proposed uncertainty threshold, but exhibits variability in the UV range. In the visible spectrum, PRISM is not compliant for the dark target and fails the proposed threshold from ~400–450 nm for the bright target. HydroRad-3 was generally in compliance, except for the ~400–500 nm wavelengths. HyperSAS was well below the proposed thresholds for all the wavelengths, but exhibited negative reflectances below ~400 nm.

The Rs spectra were consistent with the expected targets; the PRISM and AVIRIS-NG dark (blue water) spectra peaked in the blue and steadily declined into the red and NIR ranges, while the bright (kelp) target exhibited a strong red edge feature from the surface canopy of the kelp bed. The Moss Landing Harbor, San Francisco Bay, and Gulf of Mexico targets (bright) exhibited very high Rs peaking between 550 and 600 nm, consistent with sediment-laden waters [25,28].

There was an unusual pattern in the AVIRIS-NG uncertainty, with both the absolute uncertainty and percent uncertainty rising in the red edge feature over the kelp, tracking the Rs signal. While still below the PACE-specified threshold, it suggests that despite choosing an apparently homogenous section of the kelp bed, there was pixel or subpixel variability that influenced the variogram. This is presumably due to the high spatial heterogeneity (e.g., [45]) associated with kelp beds at the native spatial resolution of AVIRIS-NG from this data collection. PRISM also showed spectral variability at some wavelengths consistent with narrow spectral bands being influenced by atmospheric absorption features [30] that tend to be smoothed out in lower spectral resolution data.

### 3.3. Derived Chlorophyll and NDVI

AVIRIS-NG and PRISM data were converted to Rs and HydroRad-3 and HyperSAS Rs were used to calculate chla and NDVI (for kelp). Although fully quality-controlled Rs were not available for C-AIR, we calculated δchla for the C-AIR dark targets (S-MODE) using a simplified above-water correction scheme following [47] with $\rho = 0.028$. Corresponding Rs uncertainties are not provided (as noted above) given the simplified correction, but the analysis provides a first-order estimate of instrument uncertainty for δchla using C-AIR over a dark target.

In-water validation data were not available for all the sites, so δchla and δNDVI are presented as percent uncertainty due to the estimated error introduced to the algorithm (OC3M or NDVI) associated with environmental variability and instrument performance, referred to hereafter as environmental uncertainty ($U_E$, %). Reported $U_E$ is, therefore, only an indication of the sensor performance as deployed and processed and does not account for potential uncertainties or biases in the underlying algorithm, nor does it capture total uncertainty. Statistics that compare satellite vs. in situ matchups result in a considerably higher mean error. For example, [49] reported an ~11% error for chla in the California Current (where in situ is considered correct), while [50] reported a Mean Absolute Error for OC3 applied to global MODIS matchups of 1.64, or about a 64% error.

For C-AERO, direct matchups between airborne and in situ data were available, while for San Francisco Bay and the Gulf of Mexico, in-water data that were spatially and temporally close but not true matchups were available. Those values provide $U_E$ between 8.1 and 55.2% (relative percent difference), respectively, comparable to [50]. With those caveats, Table 2 provides results from the relevant sensors. Kudela et al. [21] proposed a first-order threshold of between 2.5 and 17.5% uncertainty of the underlying algorithm for optically complex coastal and inland waters. With the exception of AVIRIS-C for δchla, all the sensors were under the relevant thresholds for $U_E$. The data demonstrate that $U_E$ generally tracks kelp health (more positive NDVI), with increasing uncertainty as NDVI values decrease. In contrast, for chla, there is no clear relationship between biomass and
with performance more closely associated with the sensor rather than environmental conditions. Of the sensors evaluated, C-AERO exhibited the highest SNR and lowest absolute $R_{\text{ns}}$ uncertainty metrics, with C-AIR providing quantitatively similar performance for $\delta \text{chl}_a$ over a dark target. AVIRIS-NG exhibited lower uncertainty for $\delta \text{chl}_a$ but not $\delta \text{NDVI}$ compared to AVIRIS-C. PRISM showed increasing $U_E$ for the dark target. All sensors except AVIRIS-C were below the proposed threshold of 17.5% $U_E$.

Table 2. Summary statistics for instrument uncertainty in retrieved biomass. OC3M estimates for $\text{chl}_a$ were classified for eutrophic status based on [50]. Measured $\text{chl}_a$ (mg m$^{-3}$), when available, are included in parentheses.

<table>
<thead>
<tr>
<th>Location</th>
<th>Sensor</th>
<th>Algorithm</th>
<th>$U_E$ (%)</th>
<th>Biomass (NDVI, $\text{chl}_a$)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santa Barbara Channel</td>
<td>AVIRIS-NG</td>
<td>NDVI</td>
<td>5.4</td>
<td>0.57</td>
<td>This Study</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td>AVIRIS-C</td>
<td>NDVI</td>
<td>0.7</td>
<td>0.97</td>
<td>[21]</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td>OLI</td>
<td>NDVI</td>
<td>0.9</td>
<td>0.80</td>
<td>[21]</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td>MSI</td>
<td>NDVI</td>
<td>8.2</td>
<td>0.34</td>
<td>[21]</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td>RapidEye-2</td>
<td>NDVI</td>
<td>4.1</td>
<td>0.82</td>
<td>[21]</td>
</tr>
<tr>
<td>Santa Barbara Channel</td>
<td>AVIRIS-NG</td>
<td>OC3M</td>
<td>2.4</td>
<td>1.17</td>
<td>This Study</td>
</tr>
<tr>
<td>Moss Landing Harbor</td>
<td>PRISM</td>
<td>OC3M</td>
<td>0.7</td>
<td>3.66</td>
<td>This Study</td>
</tr>
<tr>
<td>Maui, Hawaii</td>
<td>PRISM</td>
<td>OC3M</td>
<td>1.3</td>
<td>0.31</td>
<td>This Study</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>C-AERO</td>
<td>OC3M</td>
<td>0.1</td>
<td>8.08 (9.19)</td>
<td>[21]</td>
</tr>
<tr>
<td>Lake Tahoe</td>
<td>C-AERO</td>
<td>OC3M</td>
<td>0.2</td>
<td>0.49 (0.44)</td>
<td>[21]</td>
</tr>
<tr>
<td>Central California (S-MODE)</td>
<td>C-AIR</td>
<td>OC3M</td>
<td>0.2</td>
<td>0.21</td>
<td>This Study</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>OLCI</td>
<td>OC3M</td>
<td>1.1</td>
<td>17.71</td>
<td>[21]</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>AVIRIS-C</td>
<td>OC3M</td>
<td>40</td>
<td>3.58</td>
<td>[21]</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>OLI</td>
<td>OC3M</td>
<td>1.1</td>
<td>2.75</td>
<td>[21]</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>MSI</td>
<td>OC3M</td>
<td>4.1</td>
<td>3.52</td>
<td>[21]</td>
</tr>
<tr>
<td>Gulf of Mexico</td>
<td>HyperSAS</td>
<td>OC3M</td>
<td>2.3</td>
<td>1.90 (3.35)</td>
<td>This Study</td>
</tr>
<tr>
<td>San Francisco Bay</td>
<td>HydroRad-3</td>
<td>OC3M</td>
<td>10.5</td>
<td>2.84 (2.62)</td>
<td>This Study</td>
</tr>
</tbody>
</table>

4. Discussion

We updated the results from [21] to include current state-of-the-art imaging spectrometers (AVIRIS-NG and PRISM) as well as three COTS sensors capable of providing near-continuous measurements spatially (C-AIR, HyperSAS, HydroRad-3). In evaluating these and other sensors, one criterion for choosing to use a sensor is whether its performance is fit for purpose [15]. Ocean color remote sensing (broadly defined to include all aquatic remote sensing but referred to as “ocean color” for convenience) can be divided into three categories: calibration, validation, and research (CVR; [32,35]). Calibration has usually relied on fixed-location, custom-built radiometers at specific locations, e.g., MOBY [48], but Bouée pour l’acquisition de Séries Optiques à Long Terme (BOUSSOLE) used COTS radiometers [53]. Mobile autonomous platforms that meet calibration quality standards while extending observations into the UV and SWIR have also been demonstrated [35]. Requirements for validation are generally less rigorous and extend measurements to include both custom-built and COTS instrumentation including fixed platform [19], airborne [32], shipboard [11,54], and autonomous vehicles [35,55]. Research quality observations are the least restrictive and represent the vast majority of ocean color measurements. For calibration and validation, there are considerably fewer hyperspectral data, and even fewer data that extend into the UV [56,57] and SWIR [43] ranges.
Next-generation sensors extend observations into the UV range, primarily to improve atmospheric correction [12], although, as noted in the introduction, there are multiple biogeochemical parameters of interest that take advantage of UV bands [10,15,16,58–60]. While fixed-location UV sensors suitable for ocean color validation exist [11,48], the use of airborne, autonomous vehicle, or automated shipboard platforms for CVR exercises enables increased spatial extent at the temporal scale needed for match-ups to satellite observations as well as for characterizing dynamic, optically complex environments [32,35,52,55,61]. AVIRIS-C, AVIRIS-NG, and HyperSAS exhibit poor performance or lack data (either because of atmospheric correction failure or lack of bands) in the UV range. C-AIR, C-AERO, PRISM (for dark targets), and HydroRad-3 provide reasonable data (based on SNR, \( R_s \), uncertainty) but are of limited availability. The community of practice would benefit from access to, and support for, sensors with high UV fidelity that also provide adequate spatial and temporal coverage. At the other end of the spectrum, short-wave infrared bands are also underutilized. A recent analysis demonstrates that the technology used in C-AIR and C-AERO is fully capable of delivering absolute radiometric measurements of targets darker than sunlit waters (Figures 2 and 3A,D), and negating the common assumption of a “black pixel” in the NIR or SWIR region of the spectrum [51]. If the community of practice were to take advantage of those wavelengths, similar issues as for the UV bands become evident.

An instrument that meets calibration-quality performance is automatically capable of validation and research-level performance. Microradiometer technology has met this threshold [35] for in-water instruments and C-AERO is arguably also of calibration quality (Figure 2). C-AIR is based on the same technology but exhibits a lower SNR for the NIR and SWIR wavelengths (Figure 2), effectively placing it into the validation-quality category. AVIRIS-NG exhibit validation-quality performance, while HyperSAS and HydroRad-3 also meet the requirements for most wavelengths. Kudela et al. [21] noted that operational validation is possible for several sensors over turbid waters or for a reduced spectral range over blue or turbid waters (PRISM, MSI, OLCI, OLI, AVIRIS-C), to which we can add C-AIR, HyperSAS, and HydroRad-3. For dark targets, the available sensors for validation are limited if the full hyperspectral (UV-SWIR) range is required. Only C-AERO meets or exceeds all the proposed criteria, but it is multispectral rather than hyperspectral. Many of the sensors fail in the UV range. HydroRad-3 (at least in this analysis) shows a degraded performance from ~400–450 nm for a bright target, which would potentially translate to poor performance for dark targets as well. For all sensors, performance can be improved with spatial or spectral binning [62–65], but that negates the value of sensors with high spectral and spatial resolution [2].

For research, a reasonable criterion is that derived biogeochemical variables such as \( \text{chl}_a \) and NDVI exhibit better than 17.5% \( U_E \) for typical environments. While this study is limited in the number of targets, all the sensors meet this threshold, presumably in part because, as noted in [21], objectively poor (radiometrically) sensors often provide robust biogeochemical estimates when using band-ratio algorithms. Numerous publications have successfully applied all of these sensors to both inland and coastal waters, and the relative insensitivity of band-ratio algorithms to spatial and spectral resolution as well as radiometric performance is encouraging, providing a basis for the use of less expensive and lower performance sensors on novel platforms such as unoccupied airborne vehicles [6].

A question that is increasingly being asked is whether hyperspectral measurements are required or desirable for operational aquatic remote sensing [6]. While valuable, the large data volume [66] and corresponding field validation data [67] present challenges. There are anticipated benefits such as the ability to identify specific spectral signatures for (e.g.) phytoplankton groups [68,69] or trace gases such as methane [70,71], but a more common use is to locally tune band-ratio algorithms to achieve better performance [72,73] or to track spectrally dynamic features such as fluorescence line height [74] using a subset of fixed wavelengths. Recent analysis shows that global ocean hyperspectral data have comparable degrees of freedom compared to legacy sensors such as MODIS because of strong covariance across wavelengths [75]. While this may set an upper limit on the number
of parameters that can be independently extracted from hyperspectral imagery, it does not negate the use-case for the identification of spectrally narrow features or for band-switching algorithms. Nonetheless, the launch of PACE and SBG and the lack of readily available validation-quality hyperspectral instrumentation that can collect data at relevant spatial and temporal scales for inland and coastal waters is problematic. One potential solution is to combine high-fidelity multiwavelength point measurements with spectrometers [35] or to rely on numerical models to recreate full spectral data [76,77]. A third approach would be to rely on a variety of sensors with varying performance characteristics deployed as a sensor web [32,78,79].

This analysis focused exclusively on sensors amenable to the geostatistical determination of the SNR, which precludes the inclusion of fixed-location sensors [19] and in-water sensors on a variety of platforms [11,35,55]. While not discussed, it is implicit in proposed PACE validation activities that airborne sensors would be complemented with fixed-position, shipboard, and autonomous platforms [80]; the same validation methodology would also serve for other imminent hyperspectral missions [81]. This multi-modality approach has many advantages, including capturing a wider range of remote sensing targets than are available from calibration sites, the potential for improved $R_{rs}$ uncertainty [30] and cross-validation of airborne and in-water sensors [32], particularly for sensor web configurations. This analysis provides SNR and $R_{rs}$ uncertainty metrics for the sensors as used, which are often greatly degraded compared to ideal (engineering) estimates under controlled conditions. For example, PRISM is reported to have an SNR of 500 at 450 nm [80] while AVIRIS-NG was reported to have an SNR of 54–1114 (average of 345) for a scene-derived SNR from terrestrial targets [82]. In contrast, the estimates of HyperSAS $R_{rs}$ uncertainty of 10–15% under typical field conditions [82] were comparable to the HyperSAS and HydroRad-3 results from this analysis (Table 2).

The apparent discrepancy between the realized geostatistical SNR and ideal SNR can be caused by numerous factors since there are many sources of error such as instrument (hardware) noise, environmental variability, and incomplete or insufficient post-processing, e.g., incomplete removal of skylight and glint contamination. Data collection that follows existing protocols will reduce errors even for high-fidelity instruments. As an example, Figure 2 provides the SNR for dark and bright targets using C-AERO with data collection that followed strict metrology. Figure 3B,C provide the SNR for C-AIR over comparable dark and bright targets. The underlying hardware is essentially identical as used (the same sampling rate was used for both data collections), other than the inclusion of shrouds on C-AERO, which should reduce noise in the SWIR. The realized SNR is about an order of magnitude lower for C-AIR compared to C-AERO for the same geostatistical analysis, suggesting that the discrepancy is largely associated with the data collection (i.e., metrology), acquisition software, and post-processing; in the case of C-AIR for this study, support was available for data collection but not for full post-processing, highlighting the need for better end-to-end support for airborne missions to improve collection of high quality data.

Looking into the near future, this analysis demonstrates that next-generation remote sensing satellite platforms would benefit from next-generation calibration and validation technology given that existing instrumentation is of limited availability or no longer supported by the manufacturers. This has minor impacts on band-ratio based algorithms but suggests that the ocean color community is lacking in commercially available sensors (other than fixed-position deployments) that can provide adequate calibration and validation data across the full spectral range. While this issue is minimized for global ocean color, science assessing inland and coastal waters will suffer unnecessarily due to the lack of high-fidelity cal/val data.

5. Conclusions

We update a previous analysis of multiple sensors to demonstrate that presently operational shipboard, airborne, and satellite sensors are effective in the SNR and $R_{rs}$ uncertainty for typical coastal and inland targets. C-AERO continues to provide the highest
SNR and lower \( R_m \) uncertainties. AVIRIS-NG performs very well for aquatic targets but does not reach validation-quality data for all wavelengths at native resolution. Neither AVIRIS-NG nor PRISM can reliably provide data at the specified accuracy in the UV range. The derived biogeochemical products such as \( chla \) and NDVI are remarkably tolerant to poor radiometric performance and are retrieved with compliant accuracy for environmental uncertainty. Existing COTS technology (C-AERO) exceeds all radiometric requirements for the calibration and validation of multi-wavelength sensors extending into the UV, NIR, and SWIR ranges, while also exceeding all criteria for application to the clearest (darkest) inland (e.g., Lake Tahoe) and open-ocean systems.


**Funding:** This research was funded by the National Aeronautics and Space Administration (NASA) through the HyspIRI Airborne Preparatory Campaign, grant number NNX12AQ23G, the C-HARRIER Campaign, grant numbers NNX17AK99G, and NASA award 22-PACE22_2-0051. Partial funding for the analysis of inland water data was provided by California State Water Resource Control Board, project 16-044-270.

**Data Availability Statement:** Most data are publicly available as cited in the relevant sections or provided in the tables. C-AIR and C-AERO data have limited availability since funding was not available for full processing, and are available upon request.

**Acknowledgments:** We thank our collaborators on the HyspIRI Airborne Preparatory Campaign and the C-HARRIER campaign, and the NASA Student Airborne Research Program; we especially thank Jim Eilers and Jeremy Kravitz (NASA Ames Research Center), Kendra Hayashi and Charles Martin (University of California Santa Cruz), and Christopher Nykamp (United States Air Force) for the C-HARRIER field operations. Atmospherically corrected AVIRIS-NG data were kindly provided by the NASA Jet Propulsion Laboratory, and the NASA Student Airborne Research Program data collections were coordinated by Melissa Yang. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


11. Hooker, S.B.; Morrow, J.H.; Matsuoka, A. Apparent Optical Properties of the Canadian Beaufort Sea—Part 2: The 1% and 1 Cm Perspective in Deriving and Validating AOP Data Products. Biogeosciences 2013, 10, 4511–4527. [CrossRef]


27. Carpenter, R.; Dierssen, H.; Hochberg, E.; Lee, Z. CORAL. SeaWiFS Bio-optical Archive and Storage System (SeaBASS); NASA: Greenbelt, MD, USA, 2019. [CrossRef]


29. Lohrenz, S. GulfCarbon. SeaWiFS Bio-optical Archive and Storage System (SeaBASS); NASA: Greenbelt, MD, USA, 2013. [CrossRef]


41. Mueller, J.L.; Austin, R.W. Ocean Optics Protocols for SeaWiFS Validation, Revision 1; NASA: Greenbelt, MD, USA, 1995; p. 66.
51. Houskeeper, H.F.; Hooker, S.B. Extending Aquatic Spectral Information with the First Radiometric IR-B Field Observations. PNAS Nexus 2023, 2, pgad340. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.