## Uncertainty Estimation, Validation and Algorithm Maintenance Issues for the Spectral Derivative Pigments (SDP) Algorithm

Dave Siegel (ERI/UCSB), [Sasha Kramer](mailto:skramer@mbari.org)(MBARI) & Stéphane Maritorena (ERI/UCSB)

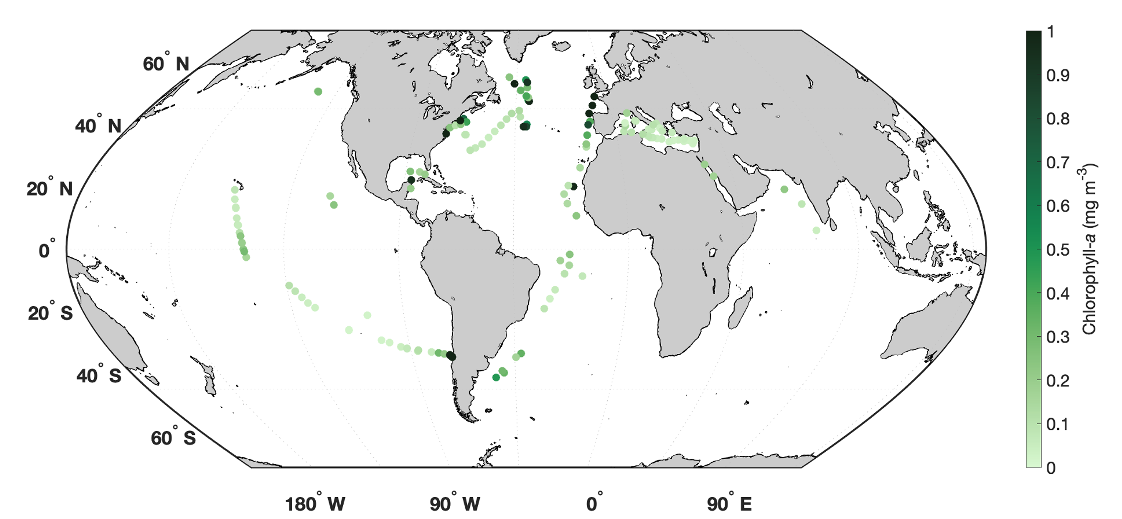
Draft: Jul 17, 2023

## The Spectral Derivative Pigments (SDP) algorithm

The Spectral Derivative Pigments (SDP) algorithm returns the concentration of thirteen phytoplankton pigment concentrations (in mg m-3), calculated using an empirical relationship derived from field data between analytical measurements of phytoplankton pigment concentrations and the second derivative of hyperspectral remote sensing reflectance residuals (Catlett & Siegel, 2018; Kramer et al., 2022). Implementation of the algorithm is contingent on the availability of hyperspectral remote sensing reflectance as an input. Hyperspectral remote sensing reflectance spectra are first modeled using an extension of the GSM bio-optical model (Maritorena et al., 2002). Reconstructed remote sensing reflectance then allows for the calculation of residuals between measured and modeled remote sensing reflectance spectra. The residuals should highlight spectral features associated with different phytoplankton group optical properties and can be diagnosed using spectral derivative analysis (Kramer et al. 2022). Field derived, spectrally-dependent coefficients are then used with the second derivative of the reflectance residuals to calculate pigment concentrations. Principal component regression analysis is applied to reduce the effects of collinearity among the spectral information and the individual pigments (Catlett & Kramer, 2018).

Uncertainty determinations are calculated as part of the principal components regression modeling, which involves 100 cross-validations to model pigments from the reflectance residual. During the cross-validation exercise, 75% of the original dataset was used for model development and 25% of the dataset was used for model evaluation. The model performance was evaluated using the mean absolute difference (MAD) following Seegers et al. (2018) and McKinna et al. (2021). From the 100 cross-validations, a mean pigment value was computed. The mean MAD, mean standard deviation, and mean R2 were also computed. The model is then validated with measured phytoplankton pigments from high performance liquid chromatography (HPLC), which also has reported measurement uncertainty. The relationship between measured and modeled accessory pigments ranged from R2 = 0.37 (zeaxanthin) to R2 = 0.70 (chlorophyll c1+c2). The SDP algorithm is presently being implemented in PACE’s pre-launch algorithm. All code for reconstructing hyperspectral remote sensing reflectance and modeling pigments is freely available at <https://github.com/sashajane19/Rrs_pigments>.

## Concerns with SDP

The performance of the SDP using field data is excellent for most of the biomarker pigments; however, it was constructed using relatively few (N=145), high quality, coincident, open ocean, field observations of hyperspectral Rrs and HPLC phytoplankton pigments (Figure 1; Kramer et al. 2022). The relatively sparse data set used to construct the SDP algorithm places severe limits on its applicability. A more expansive data set is critically needed to test the validity of the SDP algorithm applied to PACE Rrs spectra. 

**Figure 1**: Locations of pigment and Rrs stations used to develop the Spectral Derivative Pigment (SDP) algorithm. Colors show the total chlorophyll-*a* concentration measured by HPLC of each sample for context. Figure from Kramer et al. (2022).

The quality of the hyperspectral remote sensing reflectance (Rrs) in the 400-700 nm range will obviously be critical to the performance of the algorithm. That said, the SDP algorithm is a spectral derivative algorithm and uses the 2nd derivative spectrum to determine pigment concentrations (Kramer et al. 2022). As such, the effects of large spectral scale (≥ 50 nm) errors in the measured Rrs PACE-OCI spectrum due to imperfect corrections (atmospheric or other) will be minimized as is the nature of spectral derivative retrieval algorithms (Tsai and Philpot, 1998). However, noise or perturbations in the Rrs spectrum at smaller spectral scales (≤ 50 nm) can affect the shape of the residuals and its derivatives and thus the suite of pigments predicted by the algorithm. If these small spectral scale Rrs perturbations are quasi-random, spatial averaging can be performed to reduce their influence on the SDP data products. That said, the quality of the SDP products will be tied to the quality of PACE hyperspectral Rrs observations.

Other sources of uncertainty may arise with the use of HPLC data for validation. The dataset used for model development was based on a subset of the data compiled in Kramer and Siegel (2019), following criteria defined to maximize data quality. All samples used for model development were measured in the surface ocean (depths of 7 m or less) and analyzed at a small number of labs to reduce lab-dependent variability in the dataset. The labs measuring HPLC for this study participated in round-robin exercises to evaluate data quality (Hooker et al., 2012) and report precision and accuracy data for pigment measurements. We recommend that future calibration or validation of the SDP algorithm only use HPLC data from labs that report precision and accuracy as part of their data quality control and quality assurance. In this way, the uncertainty around the measured HPLC data can be quantified and reported.

A less-quantifiable source of uncertainty is the interpretation of phytoplankton pigments as community composition or taxonomy. There are many, well-documented sources of uncertainty and inaccuracy in converting phytoplankton pigments to taxonomy, which we will briefly summarize here (more details in Kramer and Siegel, 2019 and Catlett et al., 2022). Many phytoplankton taxa express the same accessory pigments (e.g., fucoxanthin is often used as a “biomarker” pigment for diatoms, but is also found in dinoflagellates, prymnesiophytes, pelagophytes, chrysophytes, dictyochophytes, and bolidophytes). Phytoplankton pigments may vary in composition or concentration due to environmental factors, such as light level or nutrient availability. Mixotrophic phytoplankton may express unexpected pigment composition due to acquisition of accessory pigments from other taxa during feeding. And finally, all phytoplankton pigments are correlated with each other and with total chlorophyll-*a* (Kramer and Siegel, 2019) which makes any pigment-based methods that assign taxonomy without assuming a lack of collinearity invalid.

## Validation and Maintenance Plans for the SDP Algorithm

A curated, global, high quality data set of surface ocean HPLC phytoplankton pigments, and when available in situ hyperspectral remote sensing reflectance (Rrs), is required to validate and iteratively improve the Spectral Derivative Pigment data product. The in situ observations of phytoplankton pigment concentrations need to be matched in time and space to PACE hyperspectral Rrs retrievals using established procedures. These data will be used to validate the performance of the SDP algorithm and can be used to iteratively improve the performance and wider applicability of the SDP algorithm. If coincident high quality, in situ hyperspectral remote sensing reflectance (Rrs) spectra are available, these data will be used to assess the performance of the reflectance residuals calculation.

The HPLC phytoplankton pigment concentrations need to be properly quality controlled and provided from a facility that is cross-compared with the NASA GSFC laboratory as mentioned above. Similarly, in situ hyperspectral Rrs spectra used in this validation effort need to be processed using PACE Science Team established procedures and all data, metadata and processing scripts (including original raw scans, calibration data, etc.).

As more paired HPLC and hyperspectral Rrs data become available, the SDP model will have to be updated to reflect the most accurate information for the global ocean. Our team plans to iterate with PACE staff at GSFC to evaluate model performance with new, incoming in situ data collected as part of the PACE calibration and validation efforts. Regional trends will also be evaluated and, if there are enough observations for a region, regional algorithms may be constructed. As coincident HPLC data are collected concurrently with PACE Rrs data, we will also evaluate the performance of the model with PACE data rather than in situ Rrs.

## References

Catlett, D. S., & Siegel, D. A. (2018). Phytoplankton Pigment Communities Can be Modeled Using Unique Relationships With Spectral Absorption Signatures in a Dynamic Coastal Environment. Journal of Geophysical Research: Oceans, 123(1), 246-264. <https://doi.org/10.1002/2017JC013195>.

Catlett, D., Siegel, D. A., Matson, P. G., Wear, E. K., Carlson, C. A., Lankiewicz, T. S., & Iglesias-Rodriguez, M. D. (2022). Integrating phytoplankton pigment and DNA meta-barcoding observations to determine phytoplankton composition in the coastal ocean. *Limnology and Oceanography*, 1-16. <https://doi.org/10.1002/lno.12274>

Hooker, S. B., Clementson, L., Thomas, C.S., Schlüter, L., Allerup, M., Ras, J., et al. (2012). The Fifth SeaWiFS HPLC Analysis Round‐Robin Experiment (SeaHARRE‐5), Rep., 1‐108 pp, NASA Goddard Space Flight Center, Greenbelt, Maryland.

Kramer, S. J., Siegel, D. A., Maritorena, S., & Catlett, D. (2022). Modeling surface ocean phytoplankton pigments from hyperspectral remote sensing reflectance on global scales. Remote Sensing of Environment, 270, 112879. <https://doi.org/10.1016/j.rse.2021.112879>.

Maritorena, S., Siegel, D. A., & Peterson, A. R. (2002). Optimization of a semianalytical ocean color model for global-scale applications. Applied Optics, 41(15), 2705-2714.

McKinna, L.I., Cetini ́c, I., Werdell, P.J., (2021). Development and validation of an empirical ocean color algorithm with uncertainties: a case study with the particulate backscattering coefficient. J. Geophys. Res. Oceans 126 (5), 1–21. https://doi.org/ 10.1029/2021JC017231.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26, 7404–7422. <https://doi.org/10.1364/OE.26.007404>.

Tsai, F. and Philpot, W., (1998). Derivative analysis of hyperspectral data. Remote sensing of environment, 66, 41-51.