Derivation of Marine Inherent Optical Properties

A Bayesian Approach

Erdem M. Karaköylü & Susanne E. Craig

NASA Ocean Biology Processing Group

erdem.m.karakoylu@nasa.gov — (301) 286-0501

Introduction

The advent of satellite oceanography has generated tremendous incite in our understanding of global biogeophysical processes in the ocean by providing a synoptic view of phytoplankton distribution dynamics. One of the principle obstacles in this endeavor is the signal contributed to the sensed light field by the atmosphere. Some of this contribution, such as Rayleigh scattering, is straightforward to correct for. However, light contribution by spatio-temporally variable aerosol distribution has so far been addressed by removing a modelled approximation of its contribution to estimate the water signal. This works well in the open ocean, but runs into trouble in coastal regions, where both marine and atmospheric layers are often optically complex.

Here, we propose circumventing the atmospheric complexity of coastal areas by using **top-of-the-atmosphere radiance (TOA)** as indirect model input. We develop, and compare alternative models to estimate **phytoplankton absorption (aph)** - a proxy for phytoplankton distribution - using a bayesian modeling framework.

Objectives

1. Estimate **aph** at 6 spectral bands from **TOA**.
2. Develop feature-selecting *Bayesian linear and nonlinear models*.
3. Use performance assessment and information theory for evaluation and model selection.

Materials and Methods

Three models were developed:

1. Linear regression with regularized horseshoe (**RHS**) prior[3]; hereafter, **Model 1**.
2. Linear regression with **rHSP** and 1^{st} order feature interactions; hereafter, **Model 2**.
3. Bayesian neural network [2]; hereafter, **Model 3**.

Model Development

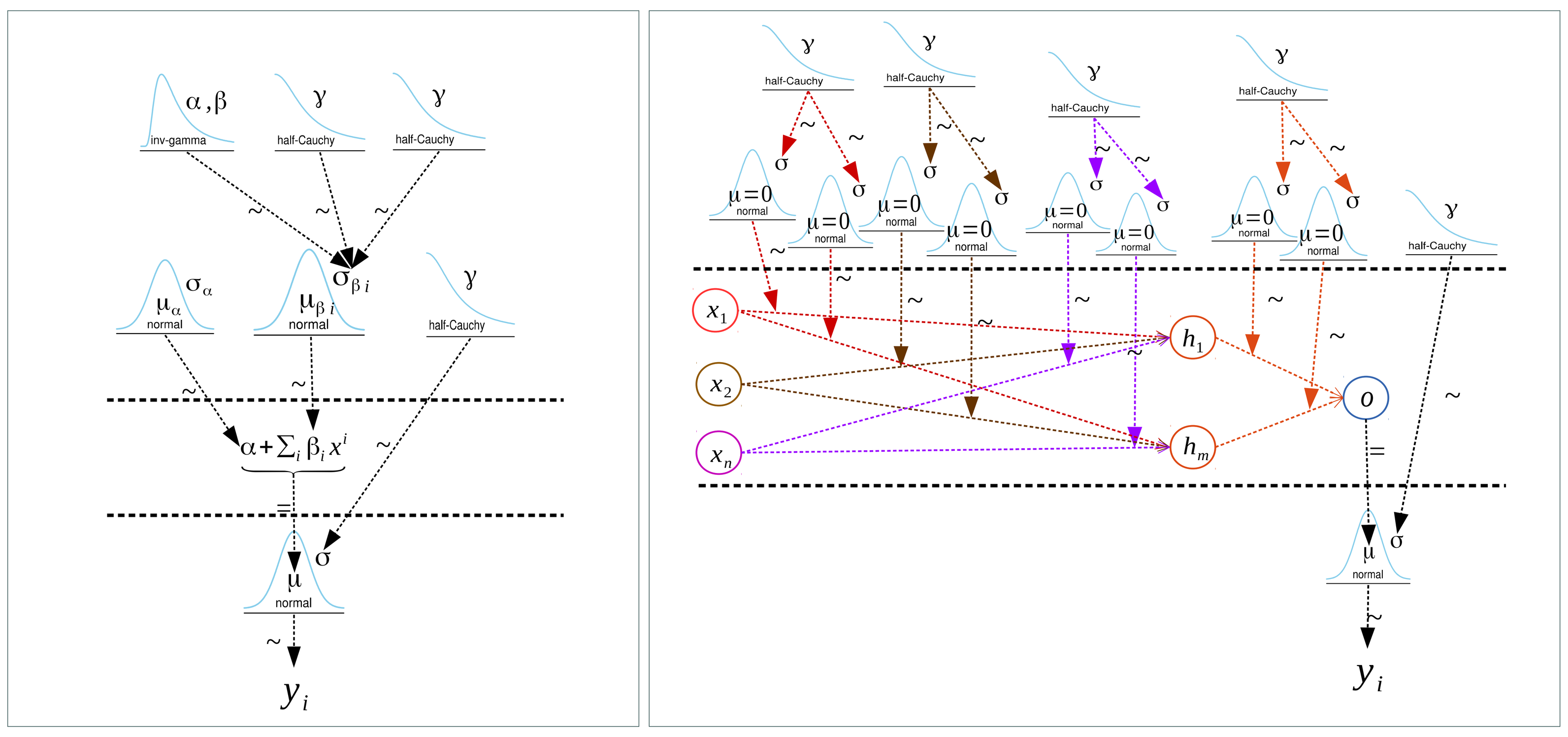

Figure 1: Inference diagram of Bayesian models used. Horizontal lines separate three conceptual groups; top → priors, middle → likelihood, bottom → outcome distribution. **Left:** Regression with horseshoe priors (Models 1 & 2). **Right:** Bayesian neural network (Model 3). Models shown here are hierarchical, built for automatic feature relevance determination.

Data Pre-Processing

Data consisted in satellite/in-situ data matchups as in [1], where TOA radiance was matched to in-situ measured phytoplankton absorption at 6 wavelengths; 411, 443, 489, 510, 555, and 670 nm. Pre-processing highlights: **(1)** principle components (**PC**) computed from correlated TOA radiance; **(2)** features include PCs, water temperature (**SST**), solar zenith angle (**SolZ**), depth (**Bathy**); **(3)** though not strictly required in a Bayesian setting, data was split into training/testing (out-of-sample) sets.

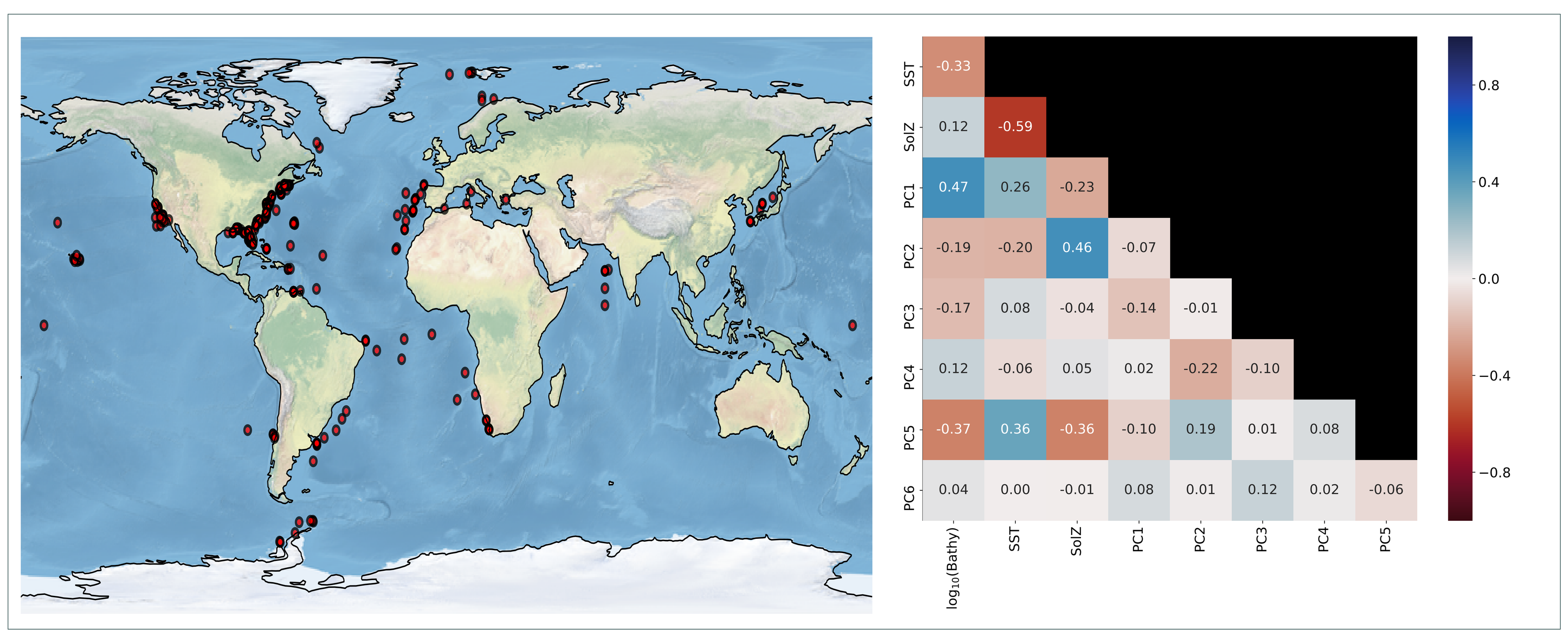

Figure 2: Left: in-situ sampling locations are mostly coastal; continental or insular. **Right**: Features used in all models.

Results

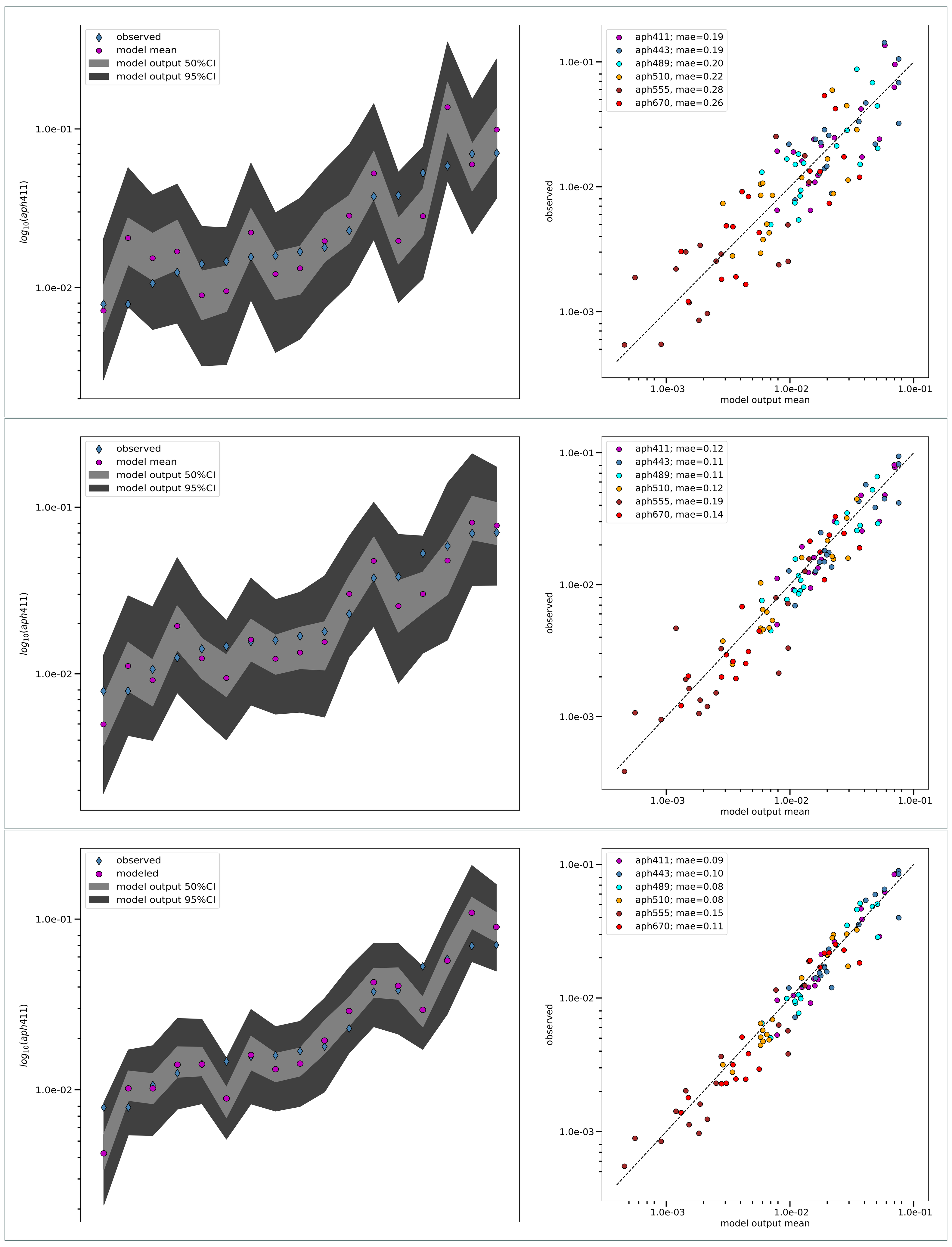

Figure 3: Top, middle and bottom panels correspond to Models 1, 2, and 3, respectively. **Left** plot shows out-of-sample observations of **aph** at 411nm, in relations to model posterior predictive mean, 50%, and 95% credibility interval, for aph at 411nm. **Right** plot shows out-of-sample against model predictions for all bands of **aph**.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
Model 3	**-183.21**	**29**	**0**	**0.98**	**21.51**	**0**
Model 2	-70.48	32.7	112.72	0	18.49	16.09
Model 1	-25.91	9.81	157.3	0.02	15.01	19.93

Table 1: **W**idely **A**vailable **I**nformation **C**riterion for models predicting **aph** at 411 nm. WAIC takes into account model complexity and the posterior distribution of a model to predict its performance on future (out-of-sample) data. **WAIC**: lower score predicts better performance; **pWAIC**: effective number of parameters - a measure of model flexibility ; **dWAIC**: difference with lowest WAIC ;**weight**: can be used when ensemble averaging similarly scored models when no clear winner is available; **SE**: standard error of WAIC estimate ; **dSE**: standard error of dWAIC. Here, **Model 3**, the Bayesian Neural Network model is predicted to be a more robust model, and proposed as sole model to be selected.

Conclusions

- Bayesian inference provides a principled modeling framework.
- Assumptions are explicit resulting in criticizable models that can be built to be comparable.
- By all measures, **Model 3** is predicted to be the better performing alternative.

Problems and Opportunities

The following items require attention:

- Tackling symmetry issues that complicates convergence in more complicated models.
- Intensified in-situ data collection for better model construction.
- Integration of this and other approaches into existing production systems.

References

[1] S. W. Bailey and P. J. Werdell. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102(1-2):12–23, 2006.

[2] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, 1996.

[3] J. Piironen and A. Vehtari. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11:5018–5051, 2017.