A Bayesian Approach to OC4

Erdem M. Karaköylü

October 13, 2018

Contents

1 Introduction

1.1 Background

- Necessity for estimating chlorophyll
- State of current chlorophyll algorithms
- Basic empirical form

$$log_{10}\left(chlor_a\right) = a_0 + \sum_{i=1}^{j} a_i log_{10}\left(\frac{max\left(Rrs\left(\lambda_{blue}\right)\right)}{Rrs\left(\lambda_{green}\right)}\right) \tag{1}$$

- Problems with current algorithms:
  - collinearity of inputs
  - poor performance in coastal
  - maximum likelihood estimation approach → increased odds of overfitting (lack of in-situ data vailability compared to satellite data makes it worse)

1.2 Proposed framework

1.2.1 Basis reduction via PCA

- PCA of Rrs to reduce overlap of information between predictor variables

1.2.2 Bayesian framework for chlorophyll estimation from remote sensing data

- transparent construction of models with explicit formulation of assumptions,
- assumptions/background information codified as priors,
- feasibility of priors verifiable before data collection via prior predictive checks
- built-in structure for selecting relevant features,
- posterior distribution as rich information structure from which to estimate parameter uncertainty as well as output prediction uncertainty,
- predictive ability of model assessed via posterior predictive checks
- multiple models encouraged by bayesian workflow,
- evaluation/comparison between models using both information about model complexity and posterior distribution (WAIC),

1.2.3 Reproducibility

- iterative process of bayesian framework relies on reproducibility for progress
- code available via github
- data available via osf

2 Methods

2.1 Model Development

2.1.1 Bayesian Linear Regression

- Order 1 regression for interpretable coefficients
- no interaction terms
- regularized horseshoe prior for feature selection

2.1.2 Bayesian Linear Regression with Interaction Terms

- generation of 1st order interaction terms
- allowing for both strong and weak heredity

2.1.3 Bayesian Neural Network

- Specific hierarchical structure for ARD
- HL1 4 NN with elu activation

2.1.4 Bayesian OC4 version as Baseline

2.2 Prior Predictive Checks

2.3 Data Acquisition/Exploration/Transformation

2.4 Model Fitting

2.5 Marginal Posterior of Coefficients → Feature Relevance Determination

2.6 Posterior Predictive Checks

2.7 Model Comparison Through Posterior Predictive Checks

3 Results