Comparing Biogeochemical Model Outputs using Neural Network Ensembles

Christopher Holder and Anand Gnanadesikan
Department of Earth and Planetary Sciences, Johns Hopkins University

Take-Home Messages

1. Neural network ensembles (NNEs) were able to predict outputs across three simulations of a biogeochemical model with different circulations.
2. NNEs were able to extract the same apparent relationships with light and nutrients across these three simulations.

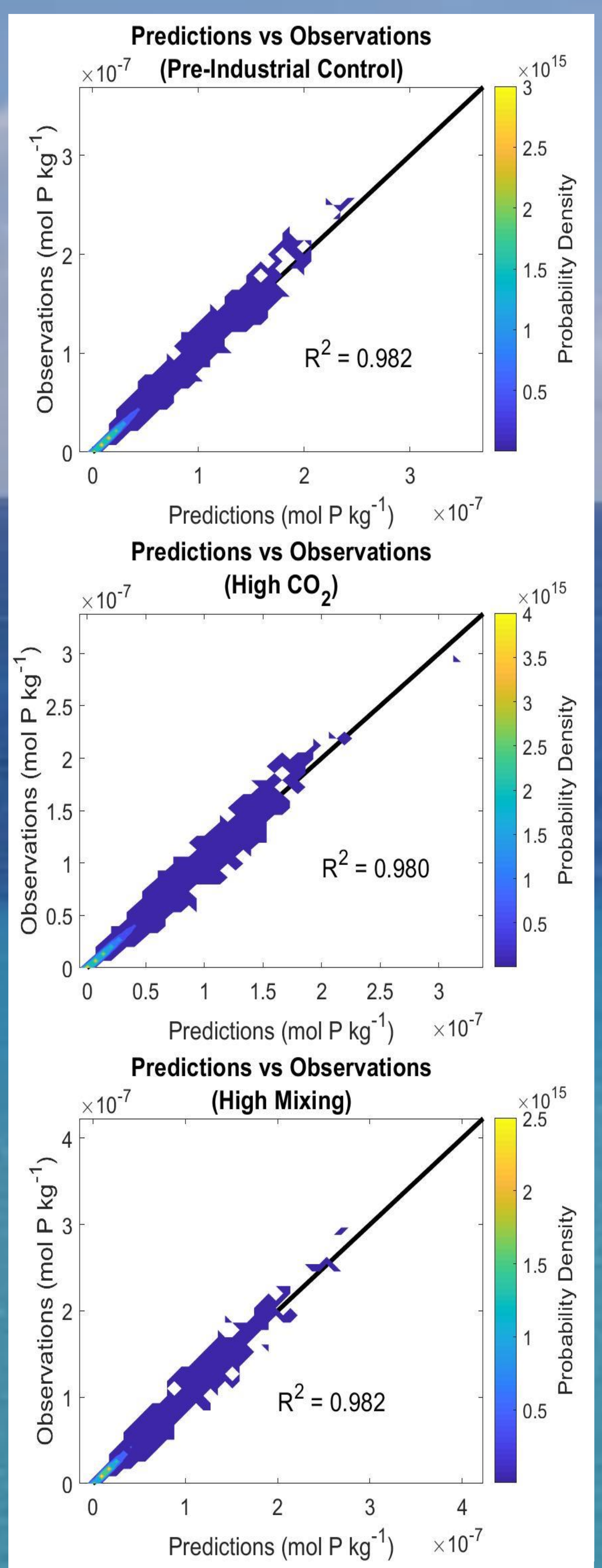

Fig. 1: An NNE was trained for each BLING simulation. The predictions on the x-axis are the predictions of each NNE. The observations on the y-axis are the actual BLING output values from that simulation.

Introduction

Earth system models (ESMs) show differences in their output because they encode different intrinsic relationships between biomass, light and nutrients as well as variations in the way those relationships interact with the physical portion of the respective model. Understanding and visualizing the reasons for the differences in output can be difficult given the time needed to run even a single simulation. To address the long computational times, neural networks have been used as emulators for ESMs which greatly reduce the computational time (Krasnopolsky et al. 2008, 2010, 2012). Recently, an effort has been made to investigate the use of neural networks for examining the relationships between light, nutrients, and biomass in a simple model (Holder & Gnanadesikan 2019). However, there have been few attempts to use neural networks for the purposes of visualizing and understanding differences between more complex ESM outputs.

We applied neural network ensembles (NNEs) to three versions of a biogeochemical model to address two questions: Can NNEs accurately reproduce the results seen in a realistic biogeochemical model? Can NNEs be used to evaluate whether models show the same apparent relationships when the differences between simulations are solely due to physical circulation?

Methods

Biogeochemical Model

We used the output from the biogeochemical model, BLING (Biogeochemistry with Light, Iron, Nutrients, and Gases; Galbraith et al. 2010). We used three versions of the BLING model in which each version was governed by the same biological components, but had differences in their inputs and physical parameters. The three versions consisted of a pre-industrial (PI) control, one with 4 times the amount of CO_2 as the PI (High CO_2), and one with 3 times higher horizontal mixing as the PI (High Mixing).

The single response variable was phytoplankton biomass and four variables were used as predictors, including concentrations for phosphate, iron, light, and temperature-dependence on growth rates.

Neural Network Ensembles

An NNE was trained for each version of the BLING model. Each NNE was the combination of the averaged results of 100 neural networks. Each individual neural network had an internal framework consisting of 4 nodes for the input layer, 25 nodes for the hidden layer, and 1 node for the output layer.

To assess NNE performance, the data was randomly split into 60% training data with the remaining 40% being used to assess NNE performance. Additionally, the NNE trained on the PI Control was used to predict the outcomes of the other two versions to assess whether the trained NNE was a robust emulator.

Visualizing Variable Relationships

Each NNE was given the same set of artificial observations in which one predictor was allowed to vary across its min-max range, while the other variables were held at a constant value.

Results

Can NNEs accurately reproduce the results seen in a biogeochemical model?

Each NNE was able to reproduce the outcome for the scenario on which it was trained with good agreement between the predictions and observations (Fig. 1).

Can NNEs be used to evaluate whether models show the same apparent relationships when the differences between simulations are solely due to physical circulation?

When the NNE that was trained on the PI Control version was asked to make predictions given the inputs of the other two versions (High CO_2 and High Mixing), the agreement between the predictions and observations remained (Fig. 2).

The sensitivity analysis also showed nearly identical apparent relationships for each of the three versions of BLING (Fig. 3).

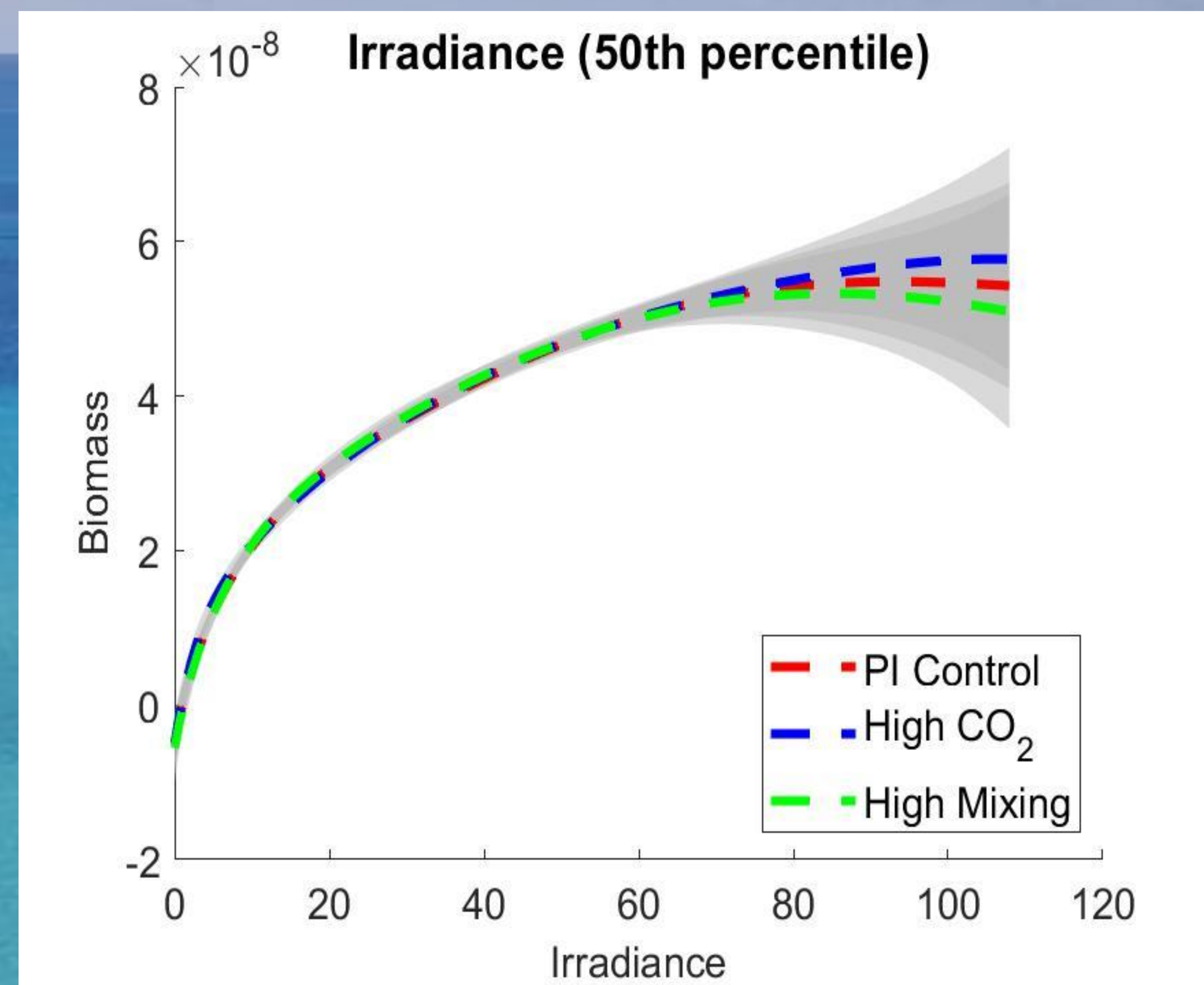

Fig. 3: Sensitivity curve for irradiance (light) vs biomass as determined by the NNE of each simulation. The other predictor variables were set at their 50th percentile values. The gray regions represent the standard deviation in the NNE predictions.

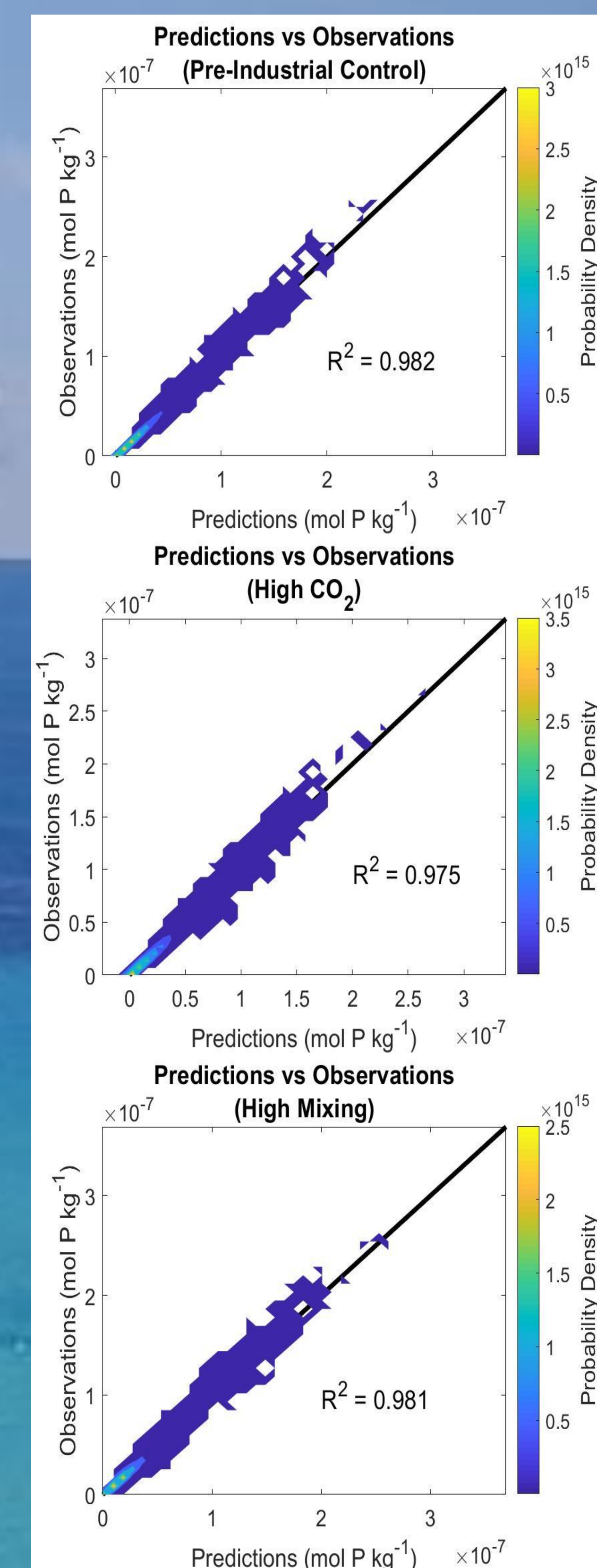

Fig. 2: The NNE that was trained on the PI Control was used to predict the outcomes of the other two simulations. The predictions on the x-axis are the predictions of the NNE (PI Control). The observations on the y-axis are the actual BLING output values from that simulation.

References

Galbraith ED, Gnanadesikan A, Dunne JP, Hiscock MR (2010) Regional impacts of iron-light colimitation in a global biogeochemical model. Biogeosciences 7(3): 1043-1064.

Holder C, Gnanadesikan A (Under Review) Understanding phytoplankton physiology using machine learning: Lessons learned from simple models. Submitted to Mar.Ecol.Prog.Ser. in April 2019.

Krasnopolsky VM, Belochitski AA, Hou YT, Lord SJ, Yang F (2012) Accurate and fast neural network emulations of long and short wave radiation for the NCEP global forecast system model.

Krasnopolsky VM, Fox-Rabinovitz MS, Hou YT, Lord SJ, Belochitski AA (2010) Accurate and fast neural network emulations of model radiation for the NCEP coupled climate forecast system: Climate simulations and seasonal predictions. Mon.Weather Rev. 138(5): 1822-1842.

Krasnopolsky VM, Fox-Rabinovitz MS, Belochitski AA (2008) Decadal climate simulations using accurate and fast neural network emulation of full, longwave and shortwave, radiation. Mon.Weather Rev. 136(10): 3683-3695.

Conclusions and Implications

The NNEs were able to reproduce the results across three versions of BLING (Fig. 1). Using an NNE trained on one version of BLING and used to predict the outcomes of the other versions showed excellent predictive capability (Fig. 2). As expected, this suggests that each version is governed by the same underlying relationships. This conclusion was further strengthened by the sensitivity analysis (Fig. 3) which showed nearly identical apparent relationships between each version.

These preliminary results suggest that NNEs could be used as a diagnostic tool to compare the outputs across different ESMs.

Acknowledgements

The authors would like to thank the Johns Hopkins University, Department of Earth and Planetary Sciences and the National Science Foundation Integrative Graduate Education and Research Traineeship.

Scan here for more information