Machine learning estimates of nitrogen fixation in the global oceans and comparison to other models

2018 OCB summer workshop
25-28, 2018 in Woods Hole

Weiyi Tang[1], Zuchuan Li[1,*] and Nicolas Cassar[1,2]

1. Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC
2. Laboratoire des Sciences de l'Environnement Marin (LEMAR), Institut Universitaire Européen de la Mer (IUEM), Brest, France

* Now at Applied Ocean Physics & Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA

Contact: **weiyi.tang@duke.edu**

Introduction

- Marine nitrogen (N_2) fixation is an important biogeochemical process, which supplies "new" nitrogen to the global oceans, supporting oceanic uptake and sequestration of carbon (1-2). Despite the central role of N_2 fixation, its controlling factors remain elusive and estimates of its magnitude vary substantially (3).
- Luo et al (2014) applied multiple linear regression to derive global N_2 fixation flux based on the global N_2 fixation database compiled by Luo et al (2012) (4-5). We revisit the estimates of Luo et al (2014) in light of the recent increase in N_2 fixation in broad regions of the world's oceans.
- Machine learning techniques have increasingly been applied to marine sciences, e.g. simulating global net community production (6).
- **Aim: To identify strong predictors of N_2 fixation, to predict N_2 fixation distribution using machine learning methods and finally to compare our estimates to the ones derived by other models**

Methods and data

1. **Global N_2 fixation dataset (Figure 1) is updated**, representing ~80% increase in the size of observations compared to Luo et al (2012) dataset.

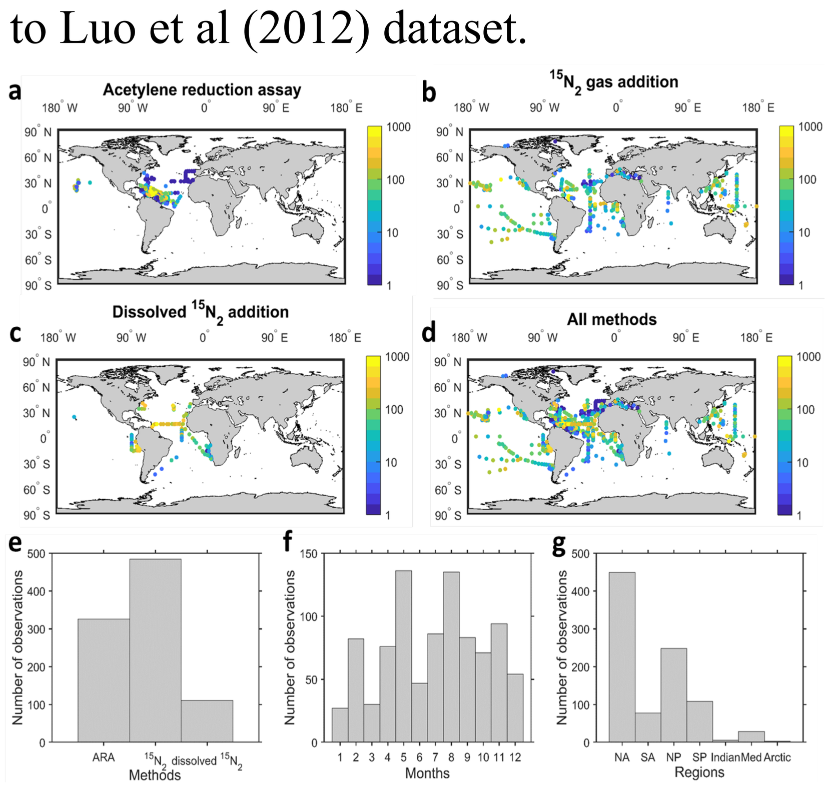

Figure 1. Maps (a-d) and frequency distributions (e-g) of field observations of N_2 fixation rates (unit: μmol N m^{-2} d^{-1}) as a function of measurement methods, months and regions. AR: acetylene reduction; NA: North Atlantic; SA: South Atlantic; NP: North Pacific; SP: South Pacific; Indian: Indian Ocean; Med: Mediterranean Sea; Arctic: Arctic Ocean.

2. **Depth-integrated N_2 fixation rates are matched with various environmental factors spatiotemporally (Figure 2).**
Daily: solar radiation; wind speed (NCEP/NCAR)
8-day: sea surface temperature, photosynthetically available radiation; chlorophyll_a concentration (NASA Ocean Color)
Monthly climatology: sea surface salinity; nutrients; oxygen concentration (WOA); mixed layer depth (7)
Data are binned into 2°×2° resolution after matching.

3. **Random forest (RF) and support vector regression (SVR) are applied to simulate N_2 fixation using compiled environmental factors.** Models were trained with randomly selected training dataset (70% of total) and evaluated using the test dataset (30% of total), shown in Figure 3.

4. **Other model outputs (Figure 4)**
CMIP5: CanESM, CNRM, GFDL, IPSL, MPI, CESM-BGC
Literature: Riche and Christian (2018); Jickells et al (2017); Paulsen et al (2017); Landolfi et al (2015); Luo et al (2014) (4, 8-11). All the model outputs are re-gridded into 2°×2° resolution.

Results

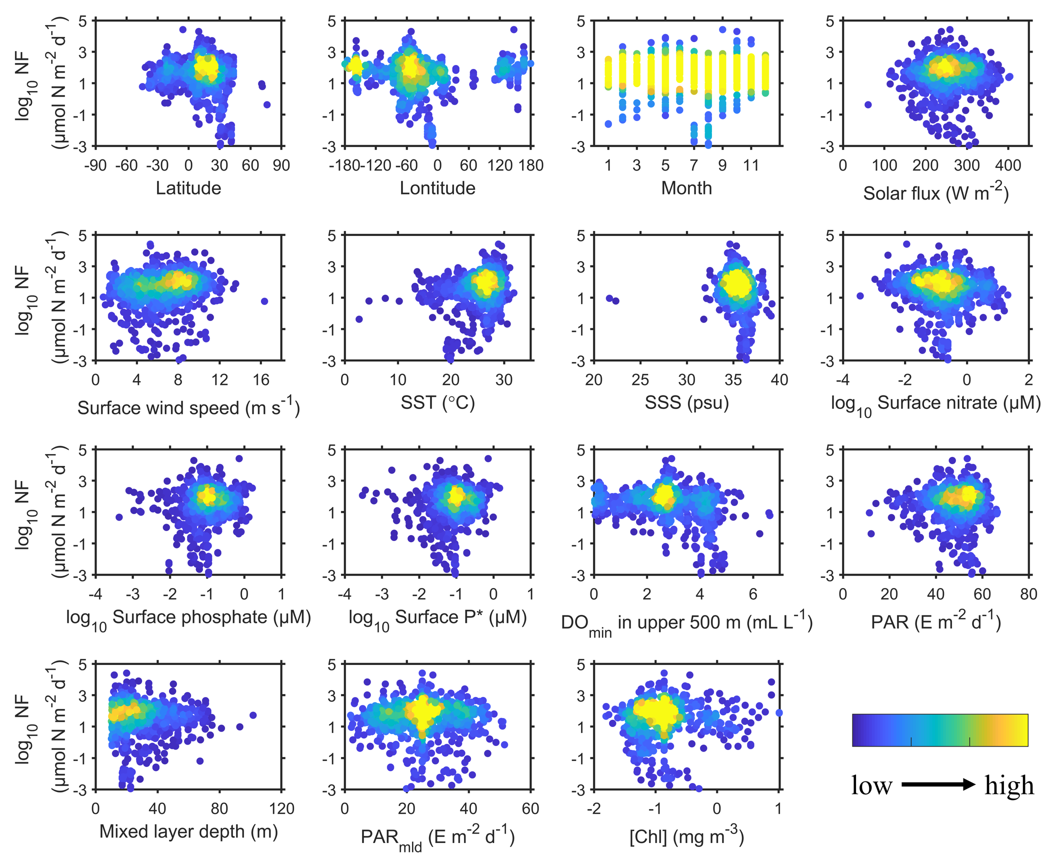

Figure 2. N_2 fixation rates vs environmental predictors. Points are color coded for density of observations (12).

No single predictor is strongly correlated with N_2 fixation rates at global scale.

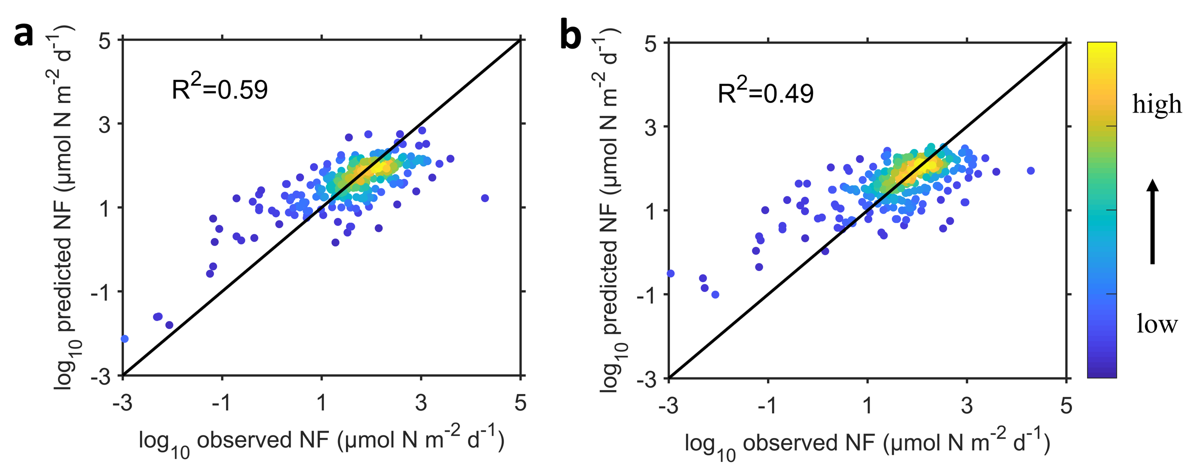

Figure 3. Comparison of observed and predicted N_2 fixation rates for test dataset using (a) random forest and (b) support vector regression. Points are color coded for density of observations (12).

Machine learning methods can predict observed N_2 fixation fairly well.

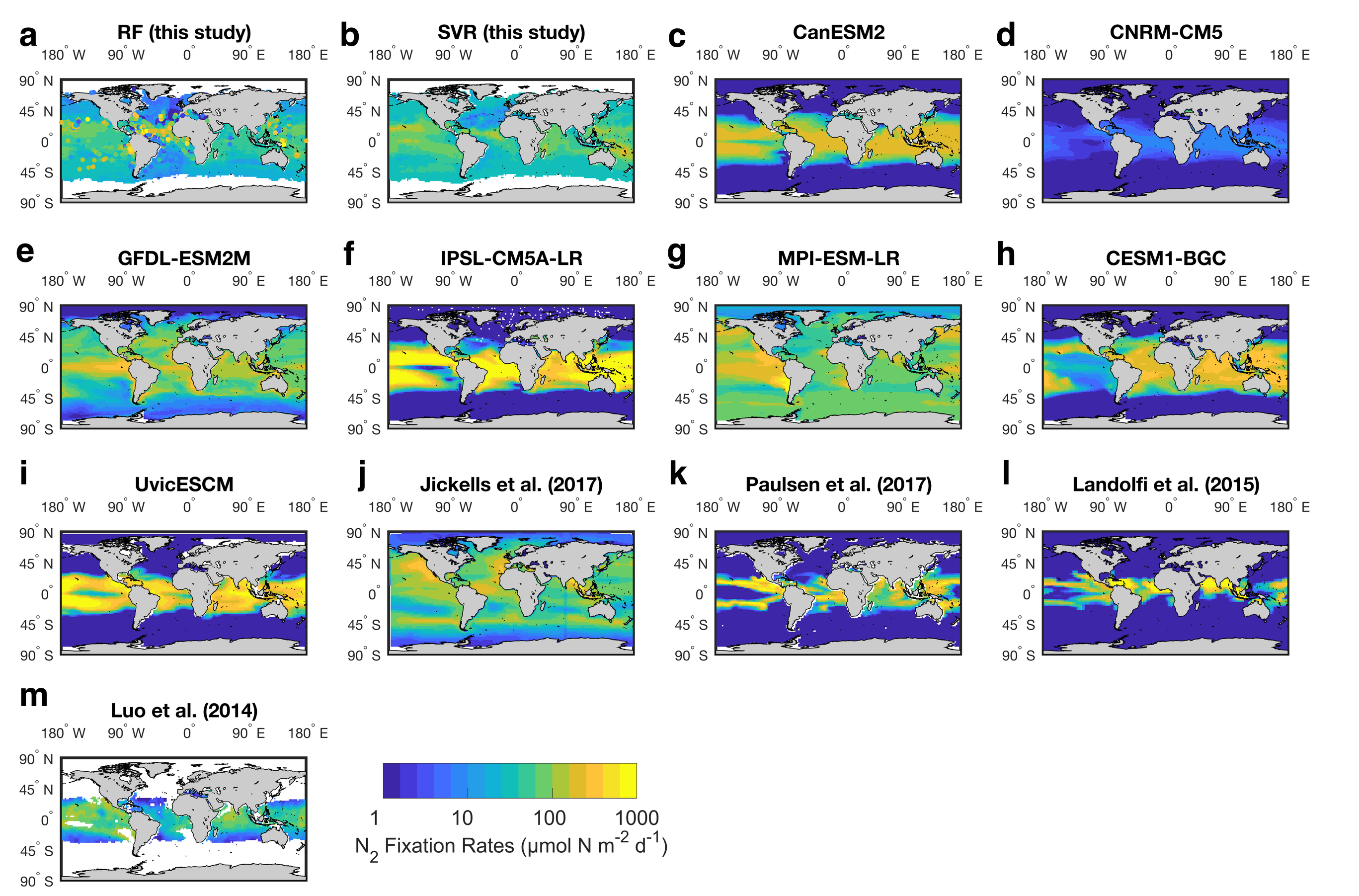

Figure 4. Global distribution of N_2 fixation rates estimated by different models. Daily N_2 fixation rates are calculated by summing monthly N_2 fixation rates and dividing by the number of days in a year. Observed N_2 fixation is overlaid on the prediction by random forest (a).

Large discrepancies exist among various models in terms of the predicted distribution and magnitude of marine N_2 fixation.

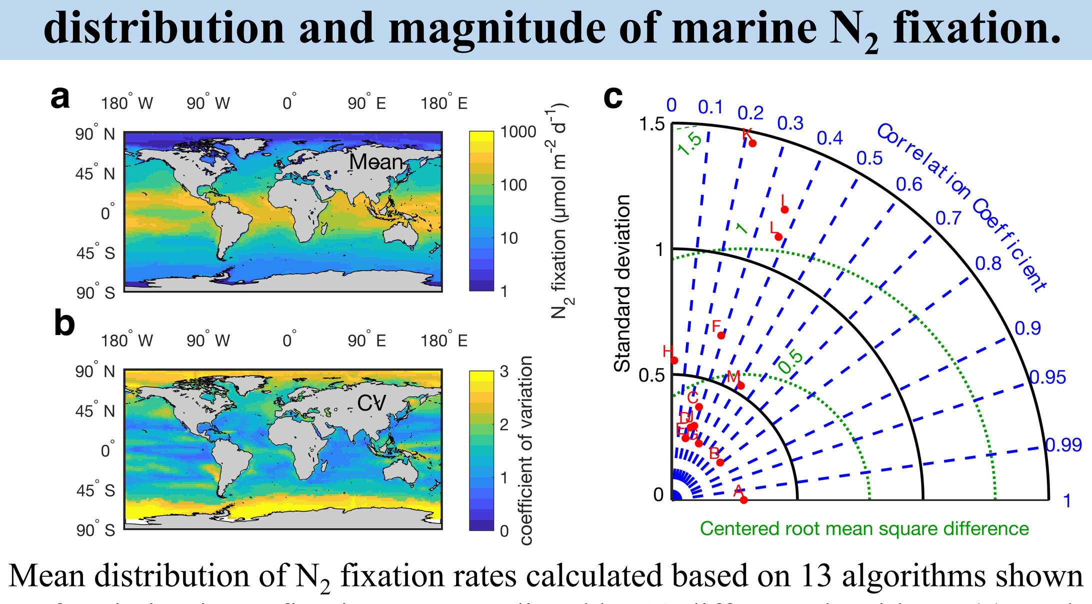

Figure 5. (a) Mean distribution of N_2 fixation rates calculated based on 13 algorithms shown in Figure 4. (b) Coefficient of variation in N_2 fixation rates predicted by 13 different algorithms. (c) Taylor diagram of N_2 fixation simulated by different models with the alphabetical order shown in Figure 4, with RF (a) as the reference model.

Model ensemble mean shows high N_2 fixation rates in the tropical oceans and largest uncertainty in the high latitudes.

Results

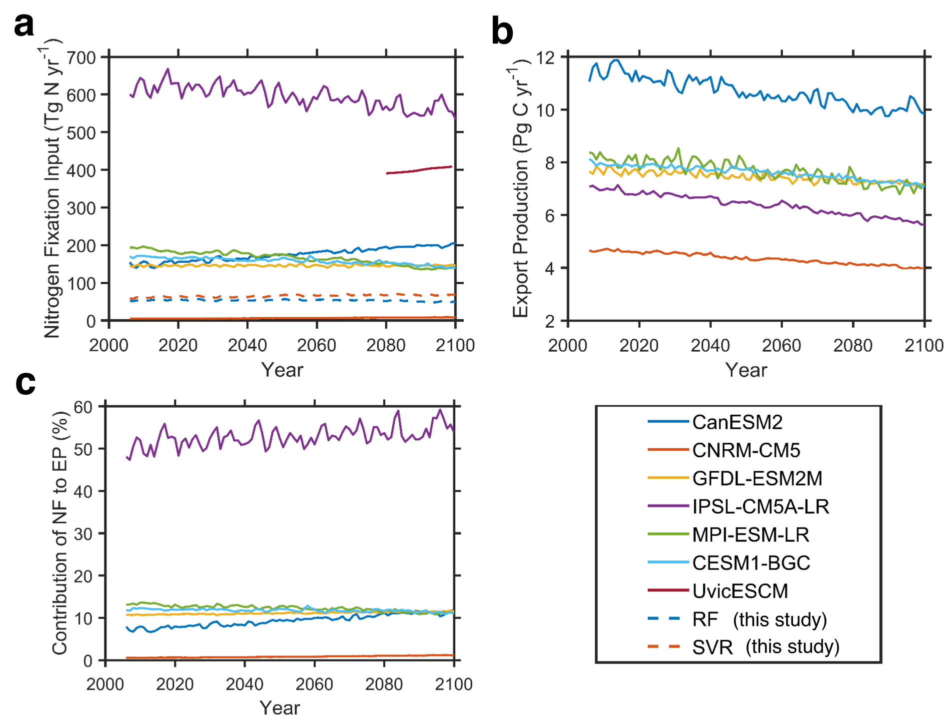

Figure 6. Projections of (a) N_2 fixation rates, (b) export production and (c) contribution of N_2 fixation to export production under RCP8.5.

Projections of future changes in N_2 fixation rates vary in the direction, let alone magnitude. This is in contrast to export production, projected to decrease by all models.

Caveats

- We combined data collected by various methods (AR, bubble and dissolved $^{15}N_2$ addition). Some of these methods may be biased.
- Mismatch of N_2 fixation rates with predictors in space and time: e.g. climatologies were used if contemporaneous predictors were not available.

Conclusions

- Weak correlation between N_2 fixation rates and single environmental factor suggests N_2 fixation may be controlled by a complex interplay of multiple factors.
- Modeled N_2 fixation fluxes by RF and SVR at 59 and 82 Tg N yr^{-1} respectively from 50°S to 50°N are in line with previous estimates but in the lower end of other models.
- Large uncertainties in model predictions argue for increased and more coordinated efforts to explore oceanic N_2 fixation using geochemical tracers, modeling, and observations over broad oceanic regions.

References

1. Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature, 387(6630), 272.
2. Karl, D. M., & Letelier, R. M. (2008). Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes. Marine Ecology Progress Series, 364, 257-268.
3. Gruber, N. (2016). Elusive marine nitrogen fixation. Proceedings of the National Academy of Sciences, 113(16), 4246-4248.
4. Luo, Y. W., Lima, I. D., Karl, D. M., Deutsch, C. A., & Doney, S. C. (2014). Data-based assessment of environmental controls on global marine nitrogen fixation. Biogeosciences, 11(3), 691-708.
5. Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., ... & Mulholland, M. R. (2012). Database of Diazotrophs in Global Ocean: Abundance, Biomass, and Nitrogen Fixation Rates. Earth System Science Data, 4(1).
6. Li, Z., & Cassar, N. (2016). Satellite estimates of net community production based on O_2/Ar observations and comparison to other estimates. Global Biogeochemical Cycles, 30(5), 735-752.
7. de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of Geophysical Research: Oceans, 109(C12).
8. Riche, O. G. J., & Christian, J. R. (2018). Ocean dinitrogen fixation and its potential effects on ocean primary production in Earth system model simulations of anthropogenic warming. Elem Sci Anth, 6(1).
9. Jickells, T. D., Buitenhuis, E., Altieri, K., Baker, A. R., Capone, D., Duce, R. A., ... & Lee, K. (2017). A reevaluation of the magnitude and impacts of anthropogenic atmospheric nitrogen inputs on the ocean. Global Biogeochemical Cycles, 31(2), 289-305.
10. Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. Journal of Advances in Modeling Earth Systems, 9(1), 438-464.
11. Landolfi, A., Koeve, W., Dietze, H., Kähler, P., & Oschlies, A. (2015). A new perspective on environmental controls of marine nitrogen fixation. Geophysical Research Letters, 42(11), 4482-4489.
12. Eilers, P. H., & Goeman, J. J. (2004). Enhancing scatterplots with smoothed densities. Bioinformatics, 20(5), 623-628.

Acknowledgement

We thank CMIP5, Olivier GJ Riche, James R Christian, Angela Landolfi, Hanna Paulsen for providing their model outputs. N. C. was supported by an NSF-CAREER grant (3331939) and supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19).