*Global Biogeochemical Cycles*

Supporting Information for

**Impact of predictor variables on estimates of global sea-air CO2 fluxes using an Extra Trees machine learning approach**

Rik Wanninkhof1

Joaquin Triñanes1,2,3

Denis Pierrot1

David R. Munro4,5

Colm Sweeney4

Amanda R. Fay6

1. Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, FL, USA.
2. Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine, Atmospheric and Earth Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL, USA.
3. Dept. of Electronics and Computer Science, Universidade de Santiago de Compostela, Santiago, Spain
4. Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
5. Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA
6. Columbia University and Lamont Doherty Earth Observatory, Palisades NY, USA

**Contents of this file**

Text Introduction

Figures S1 to S2

**Introduction**

The supplemental figures show the comparison of training the AOML\_ET with dataset deemed of good quality with an uncertainty in individual observation of less than 2 µatm (SOCAT dataset QC of A or B) with the default SOCAT dataset used to create the monthly 1˚ by 1˚ fields with uncertainties of < 5 µatm (SOCAT dataset QC A,B, C or D). For details on the SOCATQC criteria see: https://socat.info/wp-content/uploads/2020/12/2018\_SOCAT\_QC\_Cookbook\_final\_with-v2021-contact-details.pdf

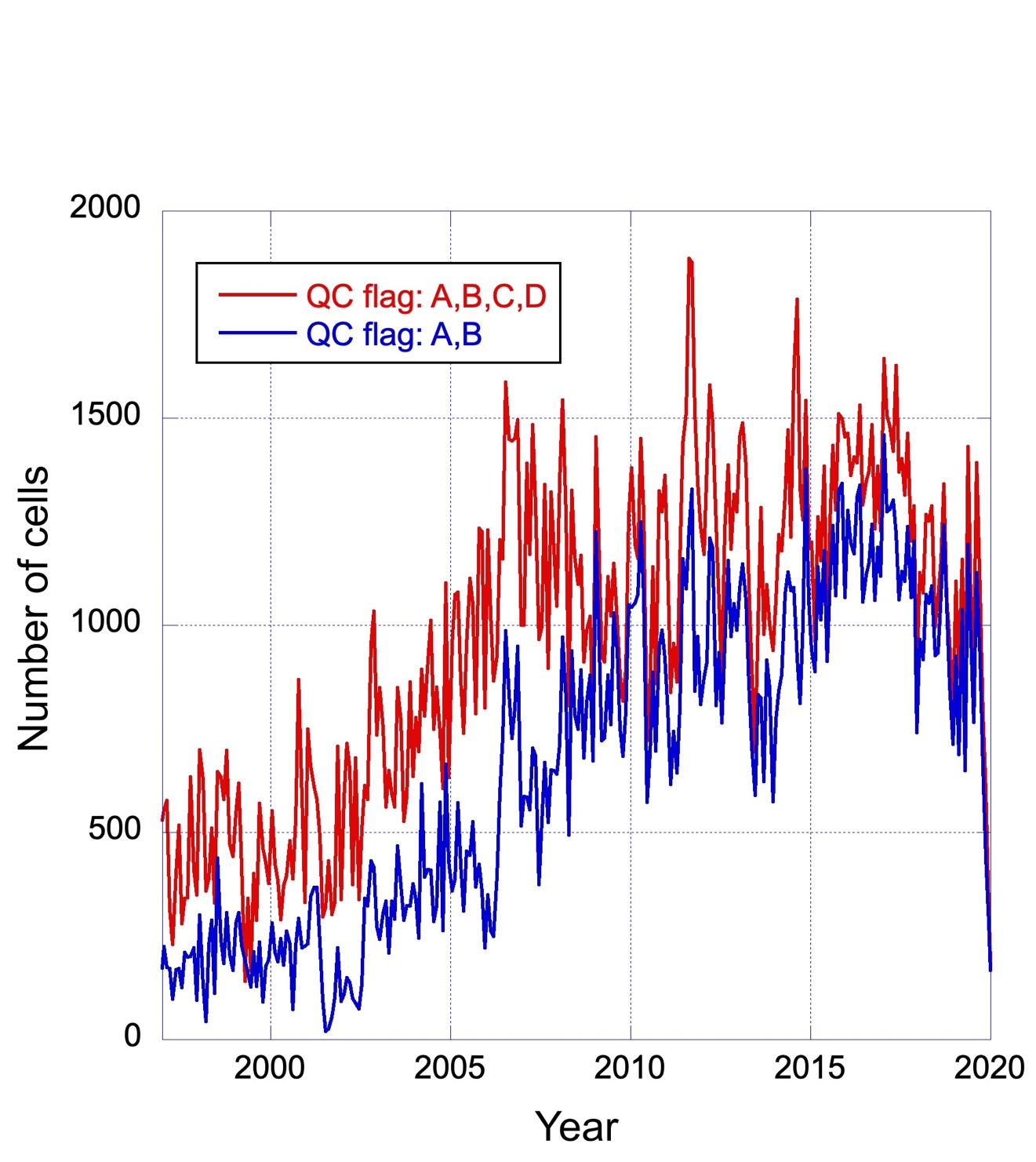


Figure S1. Number of monthly 1˚ by 1˚ grid cells with SOCAT QC flag A,B,C,D (uncertainty <5 µatm) (red line), and QC flag A, B (uncertainty <2 µatm) (blue line). From October 1997 through 2020 there were 264,496 cells with data (269 months) of which 180,160 (268 months) had A,B data.

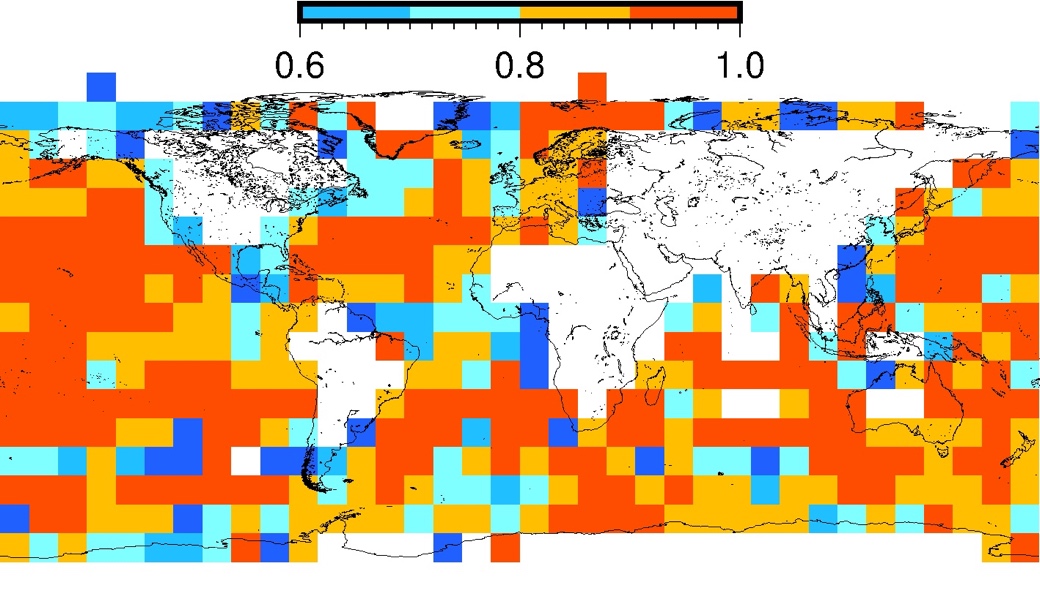


Figure S2. Pierson correlation of fCO2w in the monthly 1˚ by 1˚ using A-D and A, B data shown on 10˚ by 10˚ grid. Color scale with 0.1 gradations is provided at top of the map. Good correlations are seen over much of the open ocean, in part because much of the input data is the same. In coastal regions and areas with low data count correlations are less.