A Comparative Assessment of Gap-filling Techniques for Ocean Carbon Time Series

Jesse M. Vance[1], Kim Currie[2], John Zeldis[3], Peter Dillingham[4], Cliff S. Law[1,5]

5 [1]Department of Marine Science, University of Otago, Dunedin, 9016, New Zealand
[2]National Institute of Water and Atmospheric Research – University of Otago Centre for Oceanography, Dunedin, 9016, New Zealand
[3]National Institute of Water and Atmospheric Research, Christchurch, 8011, New Zealand
[4]Department of Statistics, University of Otago, Dunedin, 9016, New Zealand
10 [5]National Institute of Water and Atmospheric Research, Wellington, 6021, New Zealand

Correspondence to: Jesse M. Vance (jesse.vance@icloud.com)

Abstract. Regularized time series of ocean carbon data are necessary for assessing seasonal dynamics, annual budgets,
15 interannual variability and long-term trends. There are, however, no standardized methods for imputing gaps in ocean carbon
time series, and only limited evaluation of the numerous methods available for constructing uninterrupted time series. A
comparative assessment of eight imputation models was performed using data from seven long-term monitoring sites.
Multivariate linear regression (MLR), mean imputation, linear interpolation, spline interpolation, Stineman interpolation,
Kalman filtering, weighted moving average and multiple imputation by chained equation (MICE) models were compared using
20 cross-validation to determine error and bias. A bootstrapping approach was employed to determine model sensitivity to varied
degrees of data gaps and secondary time series with artificial gaps were used to evaluate impacts on seasonality and annual
summations and to estimate uncertainty. All models were fit to DIC time series, with MLR and MICE models also applied to
field measurements of temperature, salinity and remotely sensed chlorophyll, with model coefficients fit for monthly mean
conditions. MLR estimated DIC with a mean error of 8.8 µmol kg^{-1} among 5 oceanic sites and 20.0 µmol kg^{-1} among 2 coastal
25 sites. The empirical methods of MLR, MICE and mean imputation retained observed seasonal cycles over greater amounts
and durations of gaps resulting in lower error in annual budgets, outperforming the other statistical methods. MLR had lower
bias and sampling sensitivity than MICE and mean imputation and provided the most robust option for imputing time series
with gaps of various duration.

1 Introduction

30 Despite continued policy development aimed at combating climate change and declines in carbon dioxide (CO_2) emissions by
many countries over the last 10-15 years, global fossil fuel consumption continues to rise (Friedlingstein et al., 2019). We are
now in unchartered territory, with anthropogenic carbon emissions over the last two and half centuries eclipsing that of the
geological record of the past 66 million years, leaving the future of our marine and terrestrial ecosystems uncertain (Zeebe et
al., 2016). Our ability to predict future conditions, affect policy and effectively manage climate change relies on understanding

35   the feedbacks between climate, ecosystems, and biogeochemical cycles. To that end, the value of sustained time series observations have been well recognized for decades, as they are essential to characterizing processes, quantifying natural variability, identifying regime shifts and detecting long-term changes in our environment (Ducklow et al., 2009). Monitoring ocean carbon over the last three decades has revealed the decline in ocean pH concurrent with the uptake of 40% of anthropogenic CO_2 by the global ocean (Bates et al., 2014;DeVries et al., 2017). Quantification of the ocean carbon sink and

40   the impacts of ocean acidification remain actively researched given the significance of the ocean's role in controlling climate feedbacks as well as the ecological and economical importance of our marine systems (Kroeker et al., 2013;DeVries et al., 2019;Krissansen-Totton et al., 2018;Bernardello et al., 2014). Ocean carbon programs have led to a growth in surface pCO_2 data from 250,000 global measurements in 1997 to 13.5 million in 2019; however, continuity and coverage of this inorganic carbon data in space and time remains a challenge for understanding seasonal and interannual variability (Takahashi and

45   Sutherland, 2019;Takahashi et al., 1997).

1.1 Filling the gaps

Consistent sampling intervals for physical and biogeochemical parameters over several decades are critical for understanding ocean processes, establishing variability and detecting long-term changes (Henson et al., 2016). In addition to constraints arising from limitations in technology, logistics and funding, ocean science takes place in a particularly harsh environment

50   where data loss is a common occurrence. Whether from equipment failure, cancelled field campaigns, budget cuts, or a global pandemic, gaps in time series are ubiquitous and must be appropriately filled in order to carry out various statistical analyses and modelling applications which require serially complete data sets.

Machine learning techniques such as neural network methods, regression trees, and random forests have been widely used to fill gaps in meteorological and some oceanographic data, including surface ocean pCO_2 (Laruelle et al., 2017;Sasse et al.,

55   2013;Coutinho et al., 2018). While these methods are successful in the context of geospatial data, there remains little standardization in methods for imputing data gaps in oceanographic time series, particularly carbonate chemistry, at monitoring sites where there are not sufficiently close neighboring values (in time or space) that can be leveraged. Linear interpolation and mean imputation are among the most common methods for handling missing data over short to moderate time scales (Reimer et al., 2017;Kapsenberg and Hofmann, 2016;Currie et al., 2011), but comparative assessment and validation of

60   approaches overall is lacking. Gap-filling studies and standardization have been pursued in other terrestrial and atmospheric disciplines, such as eddy covariance carbon flux, solar radiation, air temperature, surface hydrology, and soil respiration (Moffat et al., 2007;Demirhan and Renwick, 2018;Zhao et al., 2020;Henn et al., 2013;Pappas et al., 2014), many of which focused on high temporal resolution data and imputing missing values over time scales from seconds to days. However it is

important that the imputation method not only focuses on minimizing error but also minimizing bias, as the preservation of

65   variance and trends is imperative for accurate analyses and understanding of climate (Serrano-Notivoli et al., 2019).

This study aims to identify the optimal gap-filling methods for carbonate time series by establishing which techniques perform with sufficiently low error and bias to assess seasonal and interannual variability of carbonate biogeochemistry and the biological and physical processes that determine it. Here we present an empirical multiple linear regression (MLR) model for estimating site-specific DIC concentration in the surface ocean using remotely sensed data products to fill gaps in field

70   measurement records. We compare this MLR approach to two other empirical methods and five statistical methods for time series imputation with the goal of informing best practice for gap-filling temporal ocean carbonate data. Although the focus here is on DIC time series, the principals of this study should extend to other carbonate parameters.

2 Materials and Methods

2.1 Field data

75   We used data from the Bermuda Atlantic Time-series (BATS, http://www.bios.edu/research/projects/bats/), Carbon Retention In A Colored Ocean (CARIACO, http://www.imars.usf.edu/cariaco), Firth of Thames (FOT, https://marinedata.niwa.co.nz/nzoa-on/), Hawaiian Ocean Time-series (HOT, https://hahana.soest.hawaii.edu/hot/), Kuroshio Extension Observatory (KEO, https://www.pmel.noaa.gov/ocs/KEO), Munida Time-series (MUNIDA, https://marinedata.niwa.co.nz/nzoa-on/), and Ocean Site Papa (PAPA, https://www.pmel.noaa.gov/ocs/Papa). These time

80   series present data describing significant ecological and environmental variability from different ocean basins and coastal regions (Fig. 1), which have been characterized in other studies (Bates et al., 2014;Fassbender et al., 2016;Fassbender et al., 2017;Zeldis and Swaney, 2018). Additionally, these time series have sufficient sampling frequencies and length of record to assess the monthly mean climatological conditions and seasonal cycle, so to allow inclusion of empirical imputation methods in this comparative assessment. Table 1 lists the site details including the carbonate parameters measured, the duration of the

85   time series, and the gap rate based on the expected sampling frequency for each of the seven sites.

All mixed layer temperature, salinity and dissolved inorganic carbon (DIC) data were averaged to monthly means for each time series site. For non-moored sampling sites with bottle sampling (BATS, CARIACO, HOT, Munida), monthly values were treated as the monthly mean condition. For each site the mixed layer depth was determined according to the temperature profile and a threshold of $DT > 0.2$ °C relative to 10 m depth (de Boyer Montégut, 2004). For sites that did not measure DIC directly,

90   the measured carbonate parameters were used with *in situ* temperature and salinity to calculate the DIC concentration using the function *carb* within the R package *seacarb* (Jean-Pierre Gattuso et al., 2012;Orr et al., 2018) with K_1, K_2 from (Lueker,

2000); K_f from (Dickson, 1979); and K_s from (Dickson et al., 1990); on the appropriate pH scale, where used, in R version 3.5.2 (R Core Team, 2020).

2.2 Remotely sensed data products

95 Monthly composites of satellite-derived surface ocean chlorophyll (O'Reilly et al., 1998) from MODIS (4 km resolution) data were paired with field data from each site except FOT. The mean surface chlorophyll was taken from a ~20 km^2 cell surrounding each of these sampling locations. For FOT, surface chlorophyll was estimated from monthly composite of VIIRS data (750 m resolution), with the mean from a ~ 4 km^2 cell surrounding the mooring used in this case given the greater spatial heterogeneity in this semi-closed coastal system.

2.3 Estimation of DIC with MLR

100

DIC, pCO$_2$ and other carbonate parameters have been successfully estimated in a variety of marine systems using multiple linear regression (MLR) approaches (Bostock et al., 2013;Velo et al., 2013;Hales et al., 2012;Lohrenz et al., 2018). In addition, empirical estimates of pCO$_2$ using remotely sensed chlorophyll and sea surface temperature (SST) have proven useful for investigating seasonal and interannual dynamics across spatial gradients, particularly in coastal systems where sustained

105 observations may be limited (Hales et al., 2012;Lohrenz et al., 2018). We investigated using an MLR model to estimate DIC from remotely sensed chlorophyll, SST and salinity in order to fill gaps in the seven monthly time series data. Parametric correlation matrices of DIC with remote chlorophyll, in situ SST and salinity showed significant linear correlation (Table 2), across most sites, with temperature having the strongest and most consistent correlation with DIC.

DIC at time t can be estimated using MLR relationships described in the form of Equation 1.

110 $$E(DIC_t) = \alpha + \beta_1 Chl_t + \beta_2 T_t + \beta_3 S_t, \tag{1}$$

where DIC has units of µmol kg^{-1}, Chl has units of mg m^{-3}, T has units of ^{o}C, and S has units of psu and the coefficients α and β_1 through β_3 are the regression coefficients fit using a generalized linear model with a Gaussian error distribution and link function. The sensitivity to each predictor variable was assessed by selectively omitting chlorophyll, temperature, and salinity from the model fit.

2.4 Imputation of DIC time series

115

Six general methods were compared for imputing DIC time series: classical, interpolation, Kalman filtering, weighted moving average (WMA) and regression and multiple imputation by chained equations (MICE). To apply the six methods, it was assumed that the gaps in the time series were data 'Missing at Random' (Little, 2002) that is, not missing systematically (i.e., as caused by seasonal sea ice cover or season-specific sampling regimes). Given this assumption, these methods can be used

120 to handle data gaps with limited biasing.

The primary goal was imputing timeseries at monthly resolution to investigate variability and trends over seasonal, interannual and decadal timescales. Therefore, random sampling and persistence methods were not considered as these methods can lead to distortion of seasonal structure in the time series. Within the 6 methods chosen, 8 models were evaluated. These imputation
125   models vary in complexity and flexibility and represent a range that have been widely applied to time series data, with 6 of the 8 models utilizing formalized packages (Demirhan and Renwick, 2018;Moritz, 2017). These methods limit overfitting and can be implemented with relative ease and low computational cost. Artificial data gaps were created as described below (Section 2.5) for the time series from each site in order to assess the performance of each method. In addition to the MLR model described by Equation 1, alternate models are described next.

130

The classical (and simplest) method applied was mean imputation, where missing values were replaced by the monthly climatological average. The climatological mean was taken as the monthly averaged means across the duration of the time series, which was over 1-2 decades in most cases. Linear interpolation was used to estimate missing values by drawing a straight line between existing values in the time series and using the slope of each of these segments to determine the value of
135   DIC at a time point(s) between known values. Spline interpolation utilized piecewise cubic polynomials to fit a curve with knots at ξ_K, $K = 1,2\ldots k$, to the data, providing more flexibility with the ability to interpolate between each point of the training data. Stineman interpolation was developed to provide the flexibility of polynomials while reducing unrealistic estimations during abrupt changes in slope within the time series (Stineman, 1980) (see Demirhan and Renwick (2018) for algorithm details). Kalman filtering was implemented using a structural model. In this case a linear Gaussian state-space model was fit
140   to the univariate time series by maximum likelihood based on decomposition (Demirhan and Renwick, 2018). A single weighted moving average model was evaluated. Missing values were replaced by weighted average of observations in the averaging window with size $k = \pm 2$ and weighting was exponential such that the exponent increases linearly to the ends of the window, here ¼, ½… ½, ¼.

145   Multiple Imputation by Chained Equations (MICE), also known as fully conditional specification (FCS) and sequential regression multivariate imputation, was applied to time series data with artificial gaps and fit using the mice library (Van Buuren, 2011) (cite1) in R version 3.5.2 (Team, 2020) (cite2), with function call *mice(data = TimeSeries$data, m = 5, method = "pmm", maxit = 20)*, where m is the number of multiple imputations, method is predictive mean matching and maxit is the maximum number of iterations. This method progresses through the following steps: 1) missing values are filled by random
150   sampling from the observations for a given variable; 2) the first variable with missing values is regressed against all other variables, while using only those with observed values; 3) moving iteratively, the remaining variables are regressed against the others but now including imputed values fitted by the regression models (White et al., 2011). This process is repeated according to the set iterations, in this case 20, to allow stabilization and convergence of the results. Regression models used in MICE allow for both linear and nonlinear relationships across variables, making this method very flexible.

2.5 Model performance and comparison

155

Each imputation model was evaluated using two schemes that assessed model performance and sampling sensitivity.

2.5.1 Cross validation

Leave one out cross validation (LOOCV) was chosen to assess the predictive error of the MLR model as well as the standard error for each imputation method. In this approach a single observation ($DIC_{t=1}$) is held out for validation while the remaining

160 observations ($DIC_{t=2} \dots DIC_{t=n}$) are used for training the model. This process is repeated *n*-1 times, allowing each data point in the time series to be treated as both training data and testing data, thus maximizing the efficiency when the data sets are of modest sampling size. Predicted DIC values and model parameters determined in each iteration were collated for the time series and performance statistics were evaluated on the total output.

2.5.2 Bootstrap sampling sensitivity

165 A bootstrapping approach was used to evaluate the sensitivity of the imputation models to the amount of data gaps in each time series. For each year of input data in the time series, artificial gaps were created by random removal of 1:8 monthly samples resulting in data gaps of 8.33%, 16.67%, 25.00%, 33.33%, 41.67%, 50.00%, and 66.67%. Random sampling was replicated 1000 times for each gap amount to ensure that an even distribution of sampling combinations was evaluated to assess the impacts of degree of data gaps on imputation error. Only years with 12 monthly samples were used to evaluate the

170 sampling sensitivity in order to ensure consistency. It should be noted that most data sets used in this study do not have monthly mean data available for all years. Table 3 shows which years of data were used from each site and the distribution of years across sites.

2.5.3 Statistical performance metrics

The performance of each model was evaluated by comparing the predicted DIC values to the observed DIC measurements.

175 The performance metrics included the coefficient of (multiple) determination (R^2) for indicating correlation; the root mean square error (RMSE), the relative root mean square error (RRMSE), and the mean absolute error (MAE) for establishing the distribution of individual errors; and the bias error (BIAS) for indicating bias induced on annual sums. Performance metrics were calculated according to Equations 2-6, where o_i and p_i denote the individual observed and predicted values respectively.

$$R^2 = \frac{\{\sum(p_i - \bar{p})(o_i - \bar{o})\}^2}{\sum(p_i - \bar{p})^2 \sum(o_i - \bar{o})^2} \tag{2}$$

180

$$RMSE = \sqrt{\frac{1}{N}\sum(p_i - o_i)^2} \tag{3}$$

$$RRMSE = \sqrt{\frac{\sum(p_i - o_i)^2}{\sum(o_i)^2}} \tag{4}$$

$$MAE = \frac{1}{N}\sum|p_i - o_i| \tag{5}$$

$$BIAS = \frac{1}{N}\sum(p_i - o_i) \tag{6}$$

2.6 Error propagation, seasonality and annual summation

185   To evaluate the propagation of error associated with each imputation model we calculated the net annual air-sea CO_2 flux at BATS and compared the observed and imputed time series using 3 synthetic gap schemes: bimonthly, as well as 3-month and 6-month sequential gaps. We used a Monte Carlo method to estimate combined uncertainty of the net annual CO_2 flux (Fassbender et al., 2016).

2.6.1 Calculation of air-sea CO_2 flux

190   The flux of CO_2 ($\mu mol\ m^{-2}\ d^{-1}$) across the air-sea interface where a positive value denotes flux from the sea to the air was calculated according to Equation 7:

$$F_{CO_2} = k \times K_0 \times [pCO_{2,ocean} - pCO_{2,atm}] \tag{7}$$

Where k is the gas transfer velocity coefficient (Wanninkhof, 1992), K_0 is the temperature- and salinity-dependent solubility of CO_2 in seawater (Weiss, 1974), $pCO_{2,ocean}$ and $pCO_{2,atm}$ are the partial pressure of CO_2 in the surface ocean and the

195   atmosphere respectively. The ocean pCO_2 was calculated from DIC and TA collected at BATS using the R package *seacarb* (see Section 2.1). The $pCO_{2,atm}$ was calculated according to Equation 8:

$$pCO_{2,atm} = xCO_{2,atm}(p_{baro} - p_{H_2O}) \tag{8}$$

Where $xCO_{2,atm}$ is the atmospheric concentration of CO_2, p_{baro} is the barometric pressure at sea level and p_{H_2O} is the vapor

200   pressure of water at the sea surface temperature and salinity (Zeebe and Wolf-Gladrow, 2001).

The monthly mean molar fraction of atmospheric CO_2 was measured at Mauna Loa (https://www.esrl.noaa.gov/gmd/ccgg/trends/data.html). Wind speed was calculated from the mean zonal and meridional components of the NCEP reanalysis of daily surface winds across a 20 km^2 cell

205   (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.pressure.html). The vapor pressure of water was calculated using barometric pressure by taking the daily mean of the 6-hourly FNMOC Sea Level Pressure data product across a 100 km^2 cell (https://data.noaa.gov/dataset/dataset/fnmoc-sea-level-pressure-360x180-6-hourly1).

2.6.2 Monte Carlo simulations

Given that the air-sea CO_2 flux depends non-linearly on the sea surface temperature and salinity, the wind speed, barometric

210   pressure, and the pCO_2 in the atmosphere and the surface ocean, a Monte Carlo approach provides a straightforward means

of determining uncertainty. We followed the approach of (Fassbender et al., 2016) in which the independent input values were varied randomly in a Gaussian distribution about the given time series values $\pm 3u_c$, where u_c is the standard uncertainty of the input variable. This standard uncertainty was taken as either the measurement uncertainty or the combined standard uncertainty that accounted for measurement uncertainty and uncertainty associated with derived values, such as pCO_2 from

215   DIC and TA, and the imputed DIC values. The calculations of the air-sea flux were performed n = 1000 iterations using these input values and the overall combined uncertainty was taken as $U_{c,t} = 3\sigma_t$, where σ_t is the standard deviation of the outputs for each point in the time series.

3 Results

3.1 Seasonal and interannual variability across sites

220   Box and whisker plots (Fig. 2) show the seasonal climatology and interannual variability for DIC across the sites tested. The bar plots in Fig. 2 show the seasonal amplitude, which was taken as the difference between maxima and minima of the climatological monthly means, and the interannual variability, which was taken as the standard deviation of the monthly means. The amplitude of the seasonal cycle of DIC spanned 11 – 90 µmol across sites, while interannual variability ranged from 8 – 22 µmol. The seasonal cycles, including amplitude, timing and interannual variability illustrate diversity among the test sites

225   so enabling robust assessment of the empirical MLR model for surface layer DIC and other imputation methods.

3.2 DIC estimation by MLR

Fig. 3 shows the performance of the MLR model to estimate DIC using the available time series data from each site ($N = 897$). The cross validated MLR exhibited an R^2 of 0.9352 with an RMSE of 12.04 µmol kg^{-1}, RRMSE of 0.59%, MAE of 8.76 µmol kg^{-1} and bias of 0.030 µmol kg^{-1}. The high R^2 and low error and bias indicate that the MLR model worked well for prediction

230   of DIC from remotely sensed chlorophyll, temperature, and salinity across different ecosystems. The predictions and errors for the data from each site are provided in Table 4, which includes the means of the model coefficients and their standard deviations for the N iterations of LOOCV per site.

The MLR performed best at PAPA with a RMSE of 4.85 µmol kg^{-1}. The greatest error was associated with the CARIACO and

235   FOT coastal sites, however, most of the predicted values still fell within 1% of observed DIC. When the sites were separated into oceanic (BATS, HOT, KEO, PAPA and MUNIDA) and coastal (CARIACO, FOT) categories, the RMSE was 8.75 µmol kg^{-1} and 19.97 µmol kg^{-1} respectively. When comparing the predictive accuracy of the MLR to the DIC variability at each site (Fig. 4), the interannual variability is strongly correlated ((R) = 0.8532, p < .02) to the RMSE while the seasonal amplitude has no apparent impact ((R)= 0.0771, p > .8), meaning the error in the predictions is most strongly related to interannual

240   variability at each site.

To assess the sensitivity of the MLR to the predictor variables, the model was adjusted by selectively removing predictor variables and refitting the model. The changes in RMSE per site due to the omission of a given variable are shown as an anomaly in the tile plot of Fig. 5. BATS exhibited the greatest sensitivity to chlorophyll relative to other sites; FOT, HOT and

245 KEO were relatively more sensitive to the effect of salinity; and temperature omission had the greatest impact for CARIACO, KEO, MUNIDA, and PAPA. The mean effects of variable omissions are given in Table 5, which indicates that collectively temperature had the greatest impact among the predictor variables on the predictive error. This was consistent with the expectations resulting from the correlation matrix provided in Table 2. The selective omission of predictor variables indicates that salinity contributes the most to the bias error although the bias error was low (<0.1) across all sites.

250 **3.3 Performance of imputation methods**

The results of the LOOCV (Fig. 6 and Table 6) indicated that each of the imputation models performed reasonably well with only 12% of all residuals exceeding 1% error and zero estimated DIC values exceeding 5% error. Differences in the performance of the models to fill singular gaps were generally minor. Table 6 shows the performance metrics for the cross validated models across all sites. Overall, the MICE and MLR models exhibited the highest R^2 and lowest error (MAE, RMSE

255 and RRMSE), followed by Linear Interpolation, Mean Imputation, Stineman Interpolation, Kalman Filtering, Weighted Moving Average and Spline Interpolation in order of increasing RMSE. MLR exhibited the least amount of bias, while Mean Imputation exhibited the greatest amount of bias.

There was considerable variability among the performance of each method across sites. The key in Fig. 7 indicates the RMSE

260 normalized to the mean value for comparing the relative error across sites and methods. The individual cross-validated errors for each imputation method per site are given as the numerical value in each tile of the figure. Fig. 7 provides further evidence that CARICO and FOT exhibit the greatest error overall, while BATS and HOT exhibit the lowest error.

3.4 Sampling sensitivity

Sampling sensitivity was assessed by the RMSE for randomized artificial gaps totaling 8.33%, 16.67%, 25.00%, 33.33%,

265 41.67%, 50.00%, and 66.67%. The randomized approach resulted in a mixture of sequential and non-sequential gaps, while bootstrapping achieved equivalent representation of all months for each assessment. Fig. 8a shows boxplots of the RMSE for each imputation method as a function of percent of data missing at each site. Spline interpolation resulted in much greater error and necessitated separate scaling. There was significant variability in both the performance of different imputation methods within sites and within imputation methods across different sites. In general, mean imputation and MLR converge on a

270 maximum error once data gaps reached 20-40%, whereas the error for other imputation models is positively correlated with the percent of data missing. While the performance of the cross validated Kalman filtering model did not differ greatly from the other interpolation methods, Fig. 8A indicates it leads to a greater number of outliers overall, in particular at BATS, KEO

and PAPA. Spline interpolation also resulted in a high number of outliers, with the most extreme error over other methods. Fig. 8B shows the median error as a function of the percent of data missing with a loess fit. The general lack of a strong
275 correlation shown by Mean imputation and MLR exhibit the least amount of sensitivity to the number of data gaps in the time series. The MICE model shows the highest level of sensitivity to the percent of data missing despite performing very well under the LOOCV and low numbers of data gaps.

3.5 Time series gaps

The imputed secondary time series synthesized with artificial gaps at bimonthly, and sequential 3-month and 6-month durations
280 that were used for assessing imputation error on seasonality and annual sums are shown in Fig. 9a. The time series indicated that most methods performed well to impute bimonthly data gaps. The positive bias error exhibited by mean imputation in the LOOCV analysis was evident in the bimonthly timeseries as well, and it had the highest RMSE at 7.90 μmol kg^{-1}. The MICE model, which outperformed other methods in the LOOCV analysis, including at BATS, had the second highest error at 7.10 μmol kg^{-1} for bimonthly imputation. However, mean imputation and MICE performed equally well, with the second lowest
285 RMSE for the 3-month gap time series. One-way ANOVA indicated that the differences in imputation error between methods for both the bimonthly and 3-month gap time series are not significant (Table 7). However, it is notable that the MLR model exhibited an RMSE that was less than half of the other methods at 3.10 μmol kg^{-1}. The 3-month gap length does not result in significant divergence from seasonal variability when using linear, spline and Stineman interpolation and also Kalman filtering and weighted moving average.

290

There was however greater variability among imputation errors across methods for the 6-month gap time series (Fig. 9b). This was further evidenced by the 1-way ANOVA, with a p-value < 0.0001. The MLR model had the lowest imputation error at 5.02 μmol kg^{-1} followed by mean imputation and MICE with errors of 7.95 and 8.19 μmol kg^{-1} respectively. Spline interpolation exhibited the greatest error at 29.18 μmol kg^{-1}, while the remaining models had errors of approximately 20 μmol
295 kg^{-1} (Table 7). The 6-month gap length appears long enough to force linear, spline and Stineman interpolation as well as Kalman filtering and weighted moving average models to diverge from the seasonal variability observed in the time series. Spline interpolation dramatically overestimated the maximum DIC and seasonal signal in 1998 in the 6-month gap time series.

3.6 Annual summation of air-sea CO$_2$ fluxes at BATS

Fig. 10 shows the net annual air-sea CO$_2$ fluxes calculated form BATS observations collected 1998-2001 compared to the CO$_2$
300 flux calculated from time series imputed using each of the models for the bimonthly, 3-month gap and 6-month gap secondary time series. In general, the models tended to overestimate the flux when the observed sums were lower (1998-1999) and underestimate the flux when the observed sums were higher (2000-2001). The anomaly observed from the poor fit using the

spline interpolation model for the 6-month gap time series resulted in a large positive flux which was opposite in sign but nearly equal in magnitude to the observations during 1998.

305

The absolute percent difference in the annual summation of air-sea CO_2 flux using each imputation model for the bimonthly, 3-month gap and 6-month gap time series are shown in Fig. 11. The relative error of the MLR was similar to MICE and mean models for 6-month gaps but was consistently lowest across all other time series. However, consistent with the time series imputation, 1-way ANOVA indicates that only the mean errors for the 6-month gap time series are significantly different. At

310   this gap duration, the MICE method, mean imputation and MLR model perform significantly better than other approaches.

3.7 Error propagation and uncertainty

Fig. 12 shows the uncertainty of the air-sea CO_2 flux as a function of the estimated flux value. The distribution of the values for each imputation method indicated little difference for bimonthly gaps. However, the dependence of uncertainty on imputation method increased for 3-month and 6-month gaps. Here, the MLR had the least sensitivity of uncertainty to the

315   magnitude of the flux and performed twice as well as all other methods over the 3-month gaps. Mean imputation and MICE performed nearly as well as MLR over 6-month gaps.

4 Discussion

4.1 MLR estimation of DIC

The development of remote sensing and MLR-based approaches for carbonate chemistry have been used extensively for

320   extrapolating over broad spatial and temporal scales to investigate regional to basin scale phenomena (Bostock et al., 2013;Hales et al., 2012;Evans et al., 2013;Lohrenz et al., 2018;Juranek et al., 2011;Alin et al., 2012). Remote sensing applications have focused primarily on predicting pCO_2 and estimating air-sea flux in coastal waters to better understand the seasonal and spatial heterogeneity of carbon sources and sinks and their implications for regional and global carbon budgets (Hales et al., 2012;Lohrenz et al., 2018). Many MLR models that predict carbonate parameters have been developed using

325   large observational data sets that include either dissolved oxygen (O_2) (Juranek et al., 2009;Kim et al., 2010;Alin et al., 2012;Bostock et al., 2013) or nitrate (NO_3) (Evans et al., 2013) as a predictor variable along with temperature and salinity. MLR models that incorporate O_2 and NO_3 can perform particularly well in coastal environments where ecosystem metabolism has a dominant effect the carbonate chemistry (Alin et al., 2012, Juranek et al., 2009). However, there are currently no remotely sensed O_2 and NO_3 data products and the chances of glider or float data being available at a given time series site to coincide

330   with a gap in carbonate measurements are limited. The MLR model presented herein serves as a method for imputing missing DIC values in time series using remotely sensed chlorophyll and temperature and in-situ salinity. Model-based estimates of salinity could be used when in-situ data are not available; however, the error associated with reanalysis salinity data products, such as the Mercator Ocean Global Reanalysis (GLORYS), the Hybrid Coordinate Ocean Model (HYCOM), the Climate

Forecast System Reanalysis (CFSR), and the Bluelink Reanalysis (BRAN), would need to be assessed for a given location and

335   included in the uncertainty budget (de Souza et al., 2020)

The variability in the MLR model coefficients indicated that the relationships between DIC, chlorophyll, temperature and salinity were location-specific and cannot be spatially extrapolated to different water masses and ecosystems. This was indicated by the variability seen among the correlations of predictor variables to DIC across sites and clearly evidenced by the

340   differences in model performance between the coastal sites (FOT and CARIACO) and the oceanic sites. However, when the MLR was trained with sufficient observations to capture the seasonal cycle, it could predict DIC with error that was far less than the natural variability over seasonal and interannual time scales and was typically on the order of, or better than the variability on monthly time scales. The RMSE of $4.85 - 10.67$ µmol kg^{-1} at the oceanic sites is consistent with other MLR studies which have ranged from ~4-11 µmol kg^{-1} (Evans et al., 2013;Juranek et al., 2011;Bostock et al., 2013), while the RMSE

345   at coastal sites (FOT and CARIACO) of approximately 20 µmol kg^{-1} is larger than exhibited in a California Current study (Alin et al., 2012). The Alin study, like others (Juranek et al., 2009;Juranek et al., 2011), estimated DIC based on O_2 and density, incorporating a multiplicative relationship. While O_2 may improve the performance of MLR approaches, particularly in biologically active coastal environments, the MLR model here only utilized remotely sensed chlorophyll and temperature and therefore only applied to the surface layer. O_2 and CO_2 may become decoupled in the surface layer due to varying time

350   scales for air sea gas exchange, making O_2 a less reliable predictor variable for surface concentrations of DIC (Juranek et al., 2011). Despite somewhat higher RMSE in coastal environments relative to the results of Alin et al. (2012), the MLR model here exhibited predictive error that is still less than 1% at such sites. We contend that this is an acceptable error in such environments for the purposes of gap filling and temporal extrapolation.

4.2 DIC time series imputation

355   Cross validation of the imputation models evaluated in this study indicated that each of these models have reasonably low (typically <1%) error when imputing a single value at monthly timescales. This was similar to other comparative gap-filling studies that focused on higher temporal resolution data and imputing missing values over time scales from seconds to days (Moffat et al., 2007;Zhao et al., 2020;Demirhan and Renwick, 2018). For the assessment of annual budgets in the studies of (Zhao et al., 2020;Moffat et al., 2007), the error associated with the imputation methods was similar to the uncertainty in the

360   fluxes across sites (Lavoie et al., 2015). As a result, the choice of imputation model yielded limited improvement on the accuracy of budget estimates. By contrast, we found that the relative uncertainty for annual air-sea CO_2 fluxes at BATS (3.5%) was much less than the percent error for the annual flux values associated with each imputation method (Fig. 11), indicating the importance of the choice of imputation model for a given data set. Evaluation of the BATS time series with synthetic gaps showed that selection of imputation method can have significant effects on the calculated timing, magnitude and structure of

365   seasonal variability.

While all methods had similar performance imputing a bimonthly time series, which was consistent with the cross-validation analysis, at a sequential gap duration of 3 months the linear, spline, and Stineman interpolation, Kalman filtering and weighted moving average methods all began to affect the shape and timing of calculated seasonal DIC dynamics. Despite the mean

370 errors and 1-way ANOVA indicating that there was no significant difference in the performance of these methods with gaps of 3 months or less, it is clear visually that mean imputation, MICE and MLR did better in reproducing the timing of seasonal changes. This became even more apparent for the time series with 6-month gaps, where important seasonal dynamics were lost or misrepresented using the linear, spline, and Stineman interpolation, Kalman filtering and weighted moving average methods.

375

In addition to the importance of retaining seasonal dynamics, the timing of the gap within a given year had significant impact on the error in the annual sum of the CO_2 flux. For example, in 1998, the 3-month gap was associated with the seasonal maximum and resulted in an underestimate of the DIC and an overestimate of CO_2 flux. However, the 3-month gaps in 1999-2001 were associated with the shoulder seasons and did not have as large an impact on the flux error. As noted above, at the

380 6-month gap duration, the timing within the year had a greater impact on the over- and underestimation of DIC for most methods, resulting in much greater error in the CO_2 flux. The annual fluxes in Fig. 10 show the stability of mean imputation, MICE and MLR compared to other methods evaluated in this study. This was further supported by the fact that the mean absolute percent error for these three techniques increased less than with other methods, when the sequential gap length was increased (Fig. 11). This was also evidenced by the uncertainty of the CO_2 fluxes that were determined from time series imputed

385 using mean imputation, MICE and MLR remained much lower than other methods when gap durations increased (Fig. 12).

The stability of mean imputation, MICE and MLR was expected, because they are based on climatological and empirical relationships rather than the other statistical approaches evaluated here. The bootstrapping assessment of sampling sensitivity for each method provided additional insight into how the imputation methods performed at randomized data missingness rates.

390 Linear and Stineman interpolation, and weighted moving average had responses similar to each other in terms of the median error and range of outliers in response to varied rates of missingness in the data, while spline interpolation produced a far greater range of outliers for all sites (over 5 times greater at FOT and CARIACO). This was also exhibited in the BATS time series assessment where the flexibility of the spline interpolation led to a tendency to overestimate seasonal maxima and minima, as observed in other comparative studies (North and Livingstone, 2013). Stineman interpolation performed better than

395 basic spline interpolation by providing greater constraint, but no better than linear interpolation, despite the increased flexibility. Interestingly, MICE performed very well at lower percentages of data missing and led to relatively low error in estimating the annual budget, yet it is highly sensitive to the percent of data missingness. However, outliers produced by MICE were constrained by the observational range because it is an empirical model. Outliers were most tightly constrained when using mean imputation and MLR given these empirical approaches are based on the climatology. This was shown in Fig. 8A

400 which illustrated how error variability decreased with increasing percentages of data missing for mean imputation and MLR.

Though the sampling sensitivity for each imputation model varied across sites (Fig. 8B), MLR exhibited the lowest sensitivity and overall error and bias for imputing missing DIC data.

The results presented here indicate that care should be taken when considering what method to use to fill data gaps in ocean
405   carbon time series, with criteria for selection including the percent of missing data, gap lengths and site characteristics. In general, the empirical models performed better than statistical models evaluated in this study. Mean imputation provides a stable and straightforward approach to filling longer gaps but may lead to higher bias in annual budgets, possibly impacting the interpretation of interannual variability and long-term trends. MICE appeared to be well suited to environmental time series data that have covariate parameters such as the correlation between DIC, chlorophyll, temperature and salinity. This could be
410   extended to other nutrients such as phosphate and nitrate as well as dissolved oxygen in order to train the models used in MICE. MICE also offers the opportunity to impute data gaps over multiple variables in larger time series data sets. Our MLR model provides a stable option that performs well over all rates of data missingness once it is sufficiently trained with field data. This MLR approach also provides the benefit of utilizing remotely sensed and modelled data products in the absence of covariate field data. This MLR model consistently performed with lower error and uncertainty than other models in this study
415   and should be considered when assessing best practices for imputing ocean carbon time series.

5 Conclusions

This study provides the first comparative assessment of several gap-filling methods that may be applied to ocean carbon time series. Regularized carbonate time series data are necessary for understanding seasonal dynamics, annual budgets, interannual variability and long-term trends in the ocean carbon cycle and changes to the ocean carbon sink, which are of particular
420   importance in the face of global climate change. Our assessment indicates that the amount and distribution of gaps in the data should be a determining factor in choosing an imputation method that optimizes uncertainty. Imputed values, however, cannot be treated as measurements and the uncertainty of imputation methods must be included in the overall uncertainty budget of broader ocean carbon analyses. The results presented above indicate the performance and behavior of select empirical and statistical approaches and the methods used provide a simple approach for estimating uncertainty using the RMSE of DIC
425   predicted by a given imputation method.

This study provides evidence that DIC can be estimated with an empirical MLR approach that uses remotely sensed chlorophyll and temperature, and in-situ salinity. This method performs consistently well across 7 disparate ecosystems in oceanic and coastal environments, but the model coefficients are unique to the water mass and ecosystem and further study is needed to
430   assess the spatial extent over which regional extrapolation is still valid. However, when using this method to impute data gaps in carbonate time series, it performs better than several options, particularly for larger gaps. We conclude that when trained with sufficient field data (e.g., captures the seasonal cycle and some interannual variability), this empirical MLR method

accurately predicts DIC from remotely sensed data and provides the most robust option for imputing gaps over a variety of data gap scenarios.

435 **Acknowledgments, Samples, and Data**

The authors thank the following for long-term contributions to ocean carbon time series and access to high quality data: the Institute for Marine Remote Sensing team for making data from the Cariaco Basin publicly available; the members of the Bermuda Institute of Ocean Sciences for making data from the Bermuda Atlantic Time Series publicly available; the School of Ocean and Earth Science and Technology at the University of Hawaii Manoa for making data from the Hawaiian Ocean

440 Time-series publicly available; NOAA's Pacific Marine Environmental Laboratory for making data from Ocean Station Papa and the Kuroshio Extension Observatory publicly available; and to New Zealand's National Institute for Water and Atmospheric Research for providing data from the Munida Time Series and the Firth of Thames. The data sets and processing code used for the analyses in this study can be found under the figshare project: A Comparative Assessment of Gap-filling Techniques for Ocean Carbon Time Series, (https://figshare.com/account/home#/projects/100349). The research presented

445 herein was supported financially by NIWA Research Scholarship Grant C 17959. The authors declare no conflicts of interests associated with this research.

References

Alin, S. R., Feely, R. A., Dickson, A. G., Hernández-Ayón, J. M., Juranek, L. W., Ohman, M. D., and Goericke, R.: Robust empirical relationships for estimating the carbonate system in the southern California Current System and application to CalCOFI hydrographic cruise
450 data (2005-2011), Journal of Geophysical Research: Oceans, 117, n/a-n/a, 10.1029/2011jc007511, 2012.
Bates, N., Astor, Y., Church, M., Currie, K., Dore, J., Gonaález-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santa-Casiano, M.: A Time-Series View of Changing Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO2 and Ocean Acidification, Oceanography, 27, 126-141, 10.5670/oceanog.2014.16, 2014.
Bernardello, R., Marinov, I., Palter, J. B., Sarmiento, J. L., Galbraith, E. D., and Slater, R. D.: Response of the Ocean Natural Carbon Storage
455 to Projected Twenty-First-Century Climate Change, Journal of Climate, 27, 2033-2053, 10.1175/jcli-d-13-00343.1, 2014.
Bostock, H. C., Mikaloff Fletcher, S. E., and Williams, M. J. M.: Estimating carbonate parameters from hydrographic data for the intermediate and deep waters of the Southern Hemisphere oceans, Biogeosciences, 10, 6199-6213, 10.5194/bg-10-6199-2013, 2013.
Coutinho, E. R., Silva, R. M. d., Madeira, J. G. F., Coutinho, P. R. d. O. d. S., Boloy, R. A. M., and Delgado, A. R. S.: Application of Artificial Neural Networks (ANNs) in the Gap Filling of Meteorological Time Series, Revista Brasileira de Meteorologia, 33, 317-328,
460 10.1590/0102-7786332013, 2018.
Currie, K. I., Reid, M. R., and Hunter, K. A.: Interannual variability of carbon dioxide drawdown by subantarctic surface water near New Zealand, Biogeochemistry, 104, 23-34, 10.1007/s10533-009-9355-3, 2011.
de Boyer Montégut, C.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, Journal of Geophysical Research, 109, 10.1029/2004jc002378, 2004.
465 de Souza, J. M. A. C., Couto, P., Soutelino, R., and Roughan, M.: Evaluation of four global ocean reanalysis products for New Zealand waters–A guide for regional ocean modelling, New Zealand Journal of Marine and Freshwater Research, 55, 132-155, 10.1080/00288330.2020.1713179, 2020.
Demirhan, H., and Renwick, Z.: Missing value imputation for short to mid-term horizontal solar irradiance data, Applied Energy, 225, 998-1012, 10.1016/j.apenergy.2018.05.054, 2018.
470 DeVries, T., Holzer, M., and Primeau, F.: Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning, Nature, 542, 215-218, 10.1038/nature21068, 2017.

DeVries, T., Le Quere, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., Landschutzer, P., Lenton, A., Lima, I. D., Nowicki, M., Schwinger, J., and Seferian, R.: Decadal trends in the ocean carbon sink, Proc Natl Acad Sci U S A, 116, 11646-11651, 10.1073/pnas.1900371116, 2019.

475 Dickson, A. G., Wesolowski, D. J., Palmer, D. A., and Mesmer, R. E.: Dissociation Constant of Bisulfate Ion in Aqueous Sodium Chloride Solutions to 250 oC, The Journal of Physical Chemistry, 94, 7978-7985, 1990.

Dickson, A. G.: The estimation of acid dissociation constants in seawater media from potentiometric titrations with strong base., Marine Chemistry, 7, 101-109, 1979.

Ducklow, H. W., Doney, S. C., and Steinberg, D. K.: Contributions of long-term research and time-series observations to marine ecology
480 and biogeochemistry, Ann Rev Mar Sci, 1, 279-302, 10.1146/annurev.marine.010908.163801, 2009.

Evans, W., Mathis, J. T., Winsor, P., Statscewich, H., and Whitledge, T. E.: A regression modeling approach for studying carbonate system variability in the northern Gulf of Alaska, Journal of Geophysical Research: Oceans, 118, 476-489, 10.1029/2012jc008246, 2013.

Fassbender, A. J., Sabine, C. L., and Cronin, M. F.: Net community production and calcification from 7 years of NOAA Station Papa Mooring measurements, Global Biogeochemical Cycles, 30, 250-267, 10.1002/2015gb005205, 2016.

485 Fassbender, A. J., Sabine, C. L., Cronin, M. F., and Sutton, A. J.: Mixed-layer carbon cycling at the Kuroshio Extension Observatory, Global Biogeochemical Cycles, 10.1002/2016gb005547, 2017.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S.,
490 Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth System Science
495 Data, 11, 1783-1838, 10.5194/essd-11-1783-2019, 2019.

Hales, B., Strutton, P. G., Saraceno, M., Letelier, R., Takahashi, T., Feely, R., Sabine, C., and Chavez, F.: Satellite-based prediction of pCO2 in coastal waters of the eastern North Pacific, Progress in Oceanography, 103, 1-15, 10.1016/j.pocean.2012.03.001, 2012.

Henn, B., Raleigh, M. S., Fisher, A., and Lundquist, J. D.: A Comparison of Methods for Filling Gaps in Hourly Near-Surface Air Temperature Data, Journal of Hydrometeorology, 14, 929-945, 10.1175/jhm-d-12-027.1, 2013.

500 Henson, S. A., Beaulieu, C., and Lampitt, R.: Observing climate change trends in ocean biogeochemistry: when and where, Glob Chang Biol, 22, 1561-1571, 10.1111/gcb.13152, 2016.

Jean-Pierre Gattuso, Jean-Marie Epitalon, Heloise Lavigne, James Orr, Bernard Gentili, Mathilde Hagens, Andreas Hofmann, Jens-Daniel Mueller, Aurélien Proye, James Rae, and Soetaert, K.: seacarb. In: Seawater Carbonate Chemistry, 2012.

Juranek, L. W., Feely, R. A., Peterson, W. T., Alin, S. R., Hales, B., Lee, K., Sabine, C. L., and Peterson, J.: A novel method for determination
505 of aragonite saturation state on the continental shelf of central Oregon using multi-parameter relationships with hydrographic data, Geophysical Research Letters, 36, 10.1029/2009gl040778, 2009.

Juranek, L. W., Feely, R. A., Gilbert, D., Freeland, H., and Miller, L. A.: Real-time estimation of pH and aragonite saturation state from Argo profiling floats: Prospects for an autonomous carbon observing strategy, Geophysical Research Letters, 38, n/a-n/a, 10.1029/2011gl048580, 2011.

510 Kapsenberg, L., and Hofmann, G. E.: Ocean pH time-series and drivers of variability along the northern Channel Islands, California, USA, Limnology and Oceanography, 61, 953-968, 10.1002/lno.10264, 2016.

Kim, T.-W., Lee, K., Feely, R. A., Sabine, C. L., Chen, C.-T. A., Jeong, H. J., and Kim, K. Y.: Prediction of Sea of Japan (East Sea) acidification over the past 40 years using a multiparameter regression model, Global Biogeochemical Cycles, 24, n/a-n/a, 10.1029/2009gb003637, 2010.

515 Krissansen-Totton, J., Arney, G. N., and Catling, D. C.: Constraining the climate and ocean pH of the early Earth with a geological carbon cycle model, Proc Natl Acad Sci U S A, 115, 4105-4110, 10.1073/pnas.1721296115, 2018.

Kroeker, K. J., Kordas, R. L., Crim, R., Hendriks, I. E., Ramajo, L., Singh, G. S., Duarte, C. M., and Gattuso, J. P.: Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming, Glob Chang Biol, 19, 1884-1896, 10.1111/gcb.12179, 2013.

520 Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., and Regnier, P.: Global high-resolution monthly pCO$_2$ climatology for the coastal ocean derived from neural network interpolation, Biogeosciences, 14, 4545-4561, 10.5194/bg-14-4545-2017, 2017.

Lavoie, M., Phillips, C. L., and Risk, D.: A practical approach for uncertainty quantification of high-frequency soil respiration using Forced Diffusion chambers, Journal of Geophysical Research: Biogeosciences, 120, 128-146, 10.1002/2014jg002773, 2015.

525 Little, R. J. A., Rubin, D. B.: Statistical Analysis with Missing Data, 2nd ed., John Wiley & Sons, Inc., Hoboken, New Jersey, 381 pp., 2002.

Lohrenz, S. E., Cai, W. J., Chakraborty, S., Huang, W. J., Guo, X., He, R., Xue, Z., Fennel, K., Howden, S., and Tian, H.: Satellite estimation of coastal p CO 2 and air-sea flux of carbon dioxide in the northern Gulf of Mexico, Remote Sensing of Environment, 207, 71-83, 10.1016/j.rse.2017.12.039, 2018.

Lueker, T. J., Dickson, A. G., Keeling, C. D., : Ocean pCO2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1
530   and K2 : validation based on laboratory measurements of CO2 in gas and seawater at equilibrium, Marine Chemistry, 70, 105-110, 2000.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agricultural and Forest Meteorology, 147, 209-232, 10.1016/j.agrformet.2007.08.011, 2007.

535   Moritz, S., Beielstein-Bartz, T.: imputeTS: Time Series Missing Value Imputation in R, 2017.

North, R. P., and Livingstone, D. M.: Comparison of linear and cubic spline methods of interpolating lake water column profiles, Limnology and Oceanography: Methods, 11, 213-224, 10.4319/lom.2013.11.213, 2013.

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C.: Ocean color chlorophyll algorithms for SeaWiFS, Journal of Geophysical Research, 103, 24,937-924,953, 1998.

540   Orr, J. C., Epitalon, J.-M., Dickson, A. G., and Gattuso, J.-P.: Routine uncertainty propagation for the marine carbon dioxide system, Marine Chemistry, 207, 84-107, 10.1016/j.marchem.2018.10.006, 2018.

Pappas, C., Papalexiou, S. M., and Koutsoyiannis, D.: A quick gap filling of missing hydrometeorological data, Journal of Geophysical Research: Atmospheres, 119, 9290-9300, 10.1002/2014jd021633, 2014.

Reimer, J. J., Cai, W.-J., Xue, L., Vargas, R., Noakes, S., Hu, X., Signorini, S. R., Mathis, J. T., Feely, R. A., Sutton, A. J., Sabine, C.,
545   Musielewicz, S., Chen, B., and Wanninkhof, R.: Time series pCO2 at a coastal mooring: Internal consistency, seasonal cycles, and interannual variability, Continental Shelf Research, 145, 95-108, 10.1016/j.csr.2017.06.022, 2017.

Sasse, T. P., McNeil, B. I., and Abramowitz, G.: A new constraint on global air-sea CO2
 fluxes using bottle carbon data, Geophysical Research Letters, 40, 1594-1599, 10.1002/grl.50342, 2013.

Serrano-Notivoli, R., Tomas-Burguera, M., Beguería, S., Peña-Angulo, D., Vicente-Serrano, S. M., and González-Hidalgo, J.-C.: Gap Filling
550   of Monthly Temperature Data and Its Effect on Climatic Variability and Trends, Journal of Climate, 32, 7797-7821, 10.1175/jcli-d-19-0244.1, 2019.

Stineman, R. W.: A consistently well-behaved method of interpolation, Creative Computing, 6, 54-57, 1980.

Takahashi, T., Feely, R. A., Weiss, R. F., Wanninkhof, R. H., Chipman, D. W., Sutherland, S. C., and Takahashi, T. T.: Global air-sea flux of CO2: an estimate based on measurements of sea-air pCO2 difference, Proc Natl Acad Sci U S A, 94, 8292-8299, 1997.

555   Takahashi, T., and Sutherland, S. C.: Global ocean surface water partial pressure of CO2 Database: Measurements performed during 1957-2018 (LDEO Database Version 2018), NOAA National Centers for Environmental Information, Silver Springs, MD, 2019.

R: A Language and environment for statistical computing. : https://www.R-project.org/, 2020.

Van Buuren, S., Groothuis-Oudshoorn, K.: MICE: Multivariate Imputation by Chained Equations in R, Journal of Statistical Software, 45, 10.18637/jss.v045.i03, 2011.

560   Velo, A., Pérez, F. F., Tanhua, T., Gilcoto, M., Ríos, A. F., and Key, R. M.: Total alkalinity estimation using MLR and neural network techniques, Journal of Marine Systems, 111-112, 11-18, 10.1016/j.jmarsys.2012.09.002, 2013.

Wanninkhof, R.: Relationship between wind speed and gas exchange over the ocean, Journal of Geophysical Research, 97, 10.1029/92jc00188, 1992.

Weiss, R. F.: Carbon dioxide in water and seawater: the solubility of a non-ideal gas, Marine Chemistry, 2, 203-215, 1974.

565   White, I. R., Royston, P., and Wood, A. M.: Multiple imputation using chained equations: Issues and guidance for practice, Stat Med, 30, 377-399, 10.1002/sim.4067, 2011.

Zeebe, R. E., Ridgwell, A., and Zachos, J. C.: Anthropogenic carbon release rate unprecedented during the past 66 million years, Nature Geoscience, 9, 325-329, 10.1038/ngeo2681, 2016.

Zeebe, R., and Wolf-Gladrow, D.: CO2 in Seawater: Equilibrium, Kinetics, Isotopes, 1st ed., Elsevier Science, 360 pp., 2001.

570   Zeldis, J. R., and Swaney, D. P.: Balance of Catchment and Offshore Nutrient Loading and Biogeochemical Response in Four New Zealand Coastal Systems: Implications for Resource Management, Estuaries and Coasts, 41, 2240-2259, 10.1007/s12237-018-0432-5, 2018.

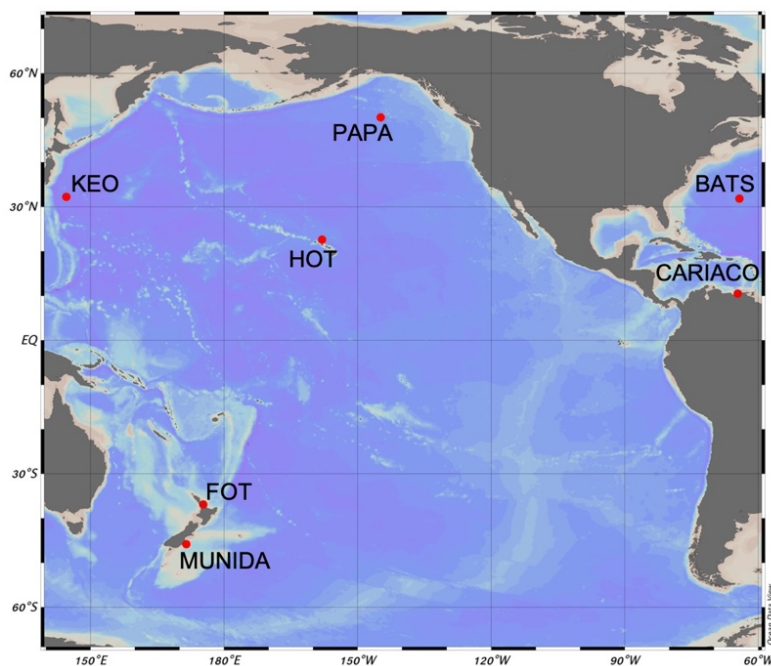
Zhao, J., Lange, H., and Meissner, H.: Gap-filling continuously-measured soil respiration data: A highlight of time-series-based methods, Agricultural and Forest Meteorology, 285-286, 10.1016/j.agrformet.2020.107912, 2020.

575

Figure 1. Location map of seven ocean carbon time series sites utilized for estimating DIC using an empirical multiple linear regression model and other empirical and statistical approaches for imputing carbonate time series, including Bermuda Atlantic Time-series (BATS), Carbon Retention In A Colored Ocean (CARIACO), Firth of Thames (FOT), Hawaiian Ocean Time-series (HOT), Kuroshio Extension Observatory (KEO), Munida Time-series (MUNIDA), and Ocean Site Papa (PAPA). See Table 1 for
580   **additional information about each sampling site.**

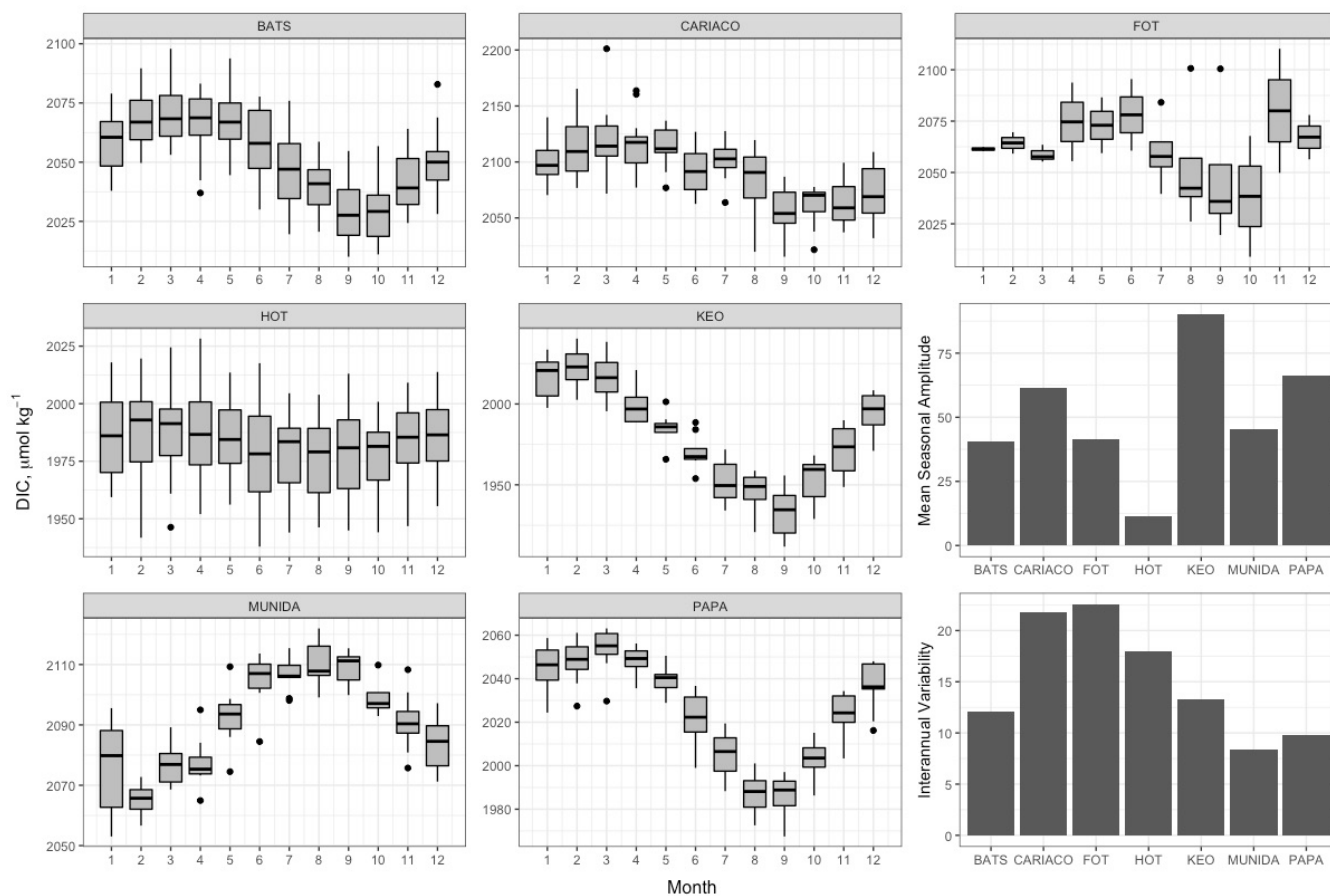



Figure 2. Box and whisker plots of monthly mean DIC concentration in the mixed layer at each site, and bar plots showing the seasonal amplitude and interannual variability. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

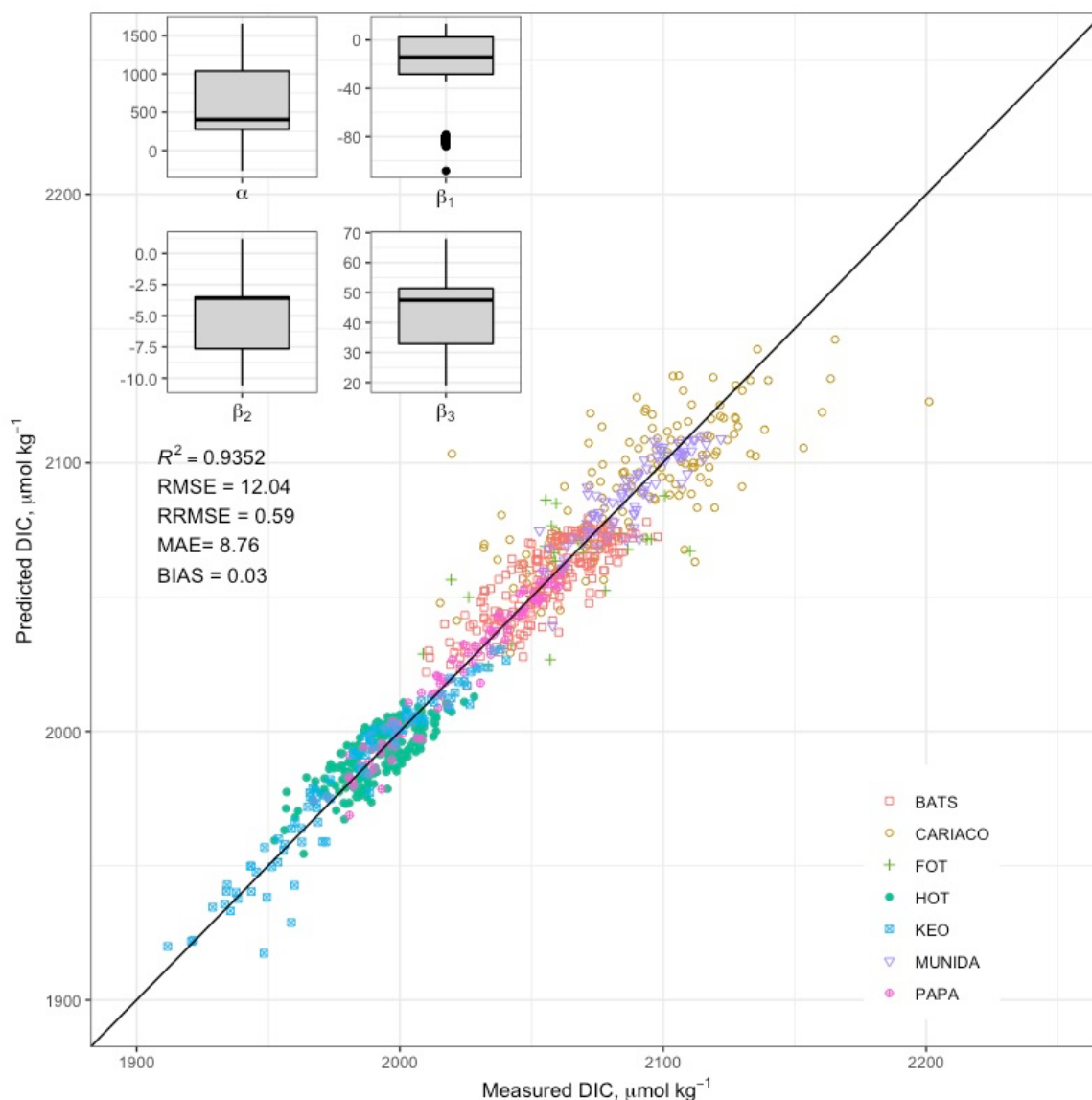



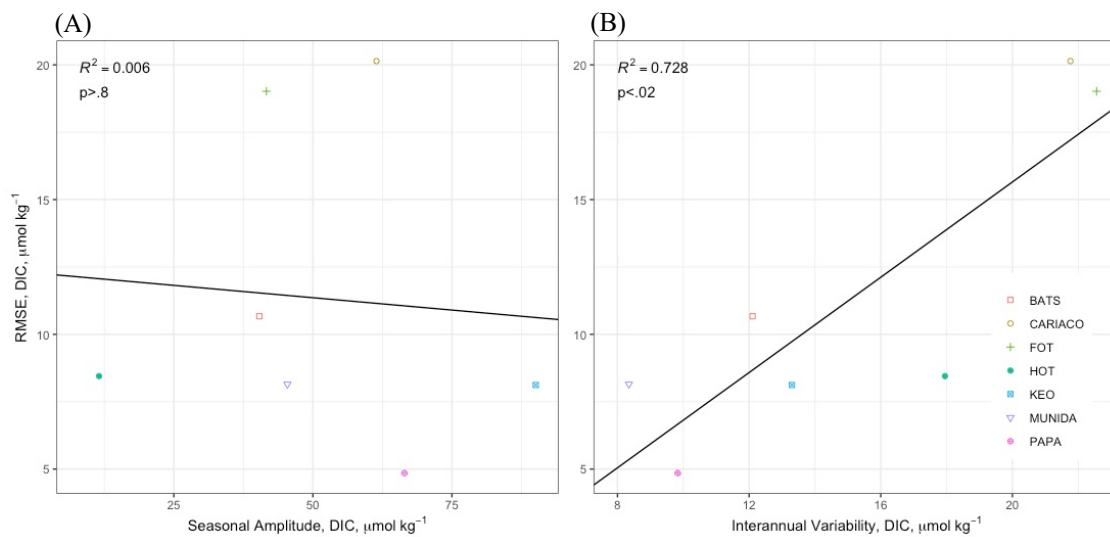
Figure 3. Comparison of predicted and measured DIC using a multiple linear regression model based on measured temperature, salinity and remotely sensed chlorophyll from sites: Bermuda Atlantic Time-series Study (BATS); Carbon Retention In A Colored Ocean (CARIACO); Firth of Thames (FOT); Hawaiian Ocean Time-series (HOT); Kuroshio Extension Observatory (KEO); Munida Time-series (MUNIDA); Ocean Site Papa (PAPA). Box and whisker plots for predictor variable coefficients α, β_1 β_2 and β_3 are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

595

Figure 4. Correlations between RMSE and (A) seasonal amplitude and (B) interannual variability across sites.

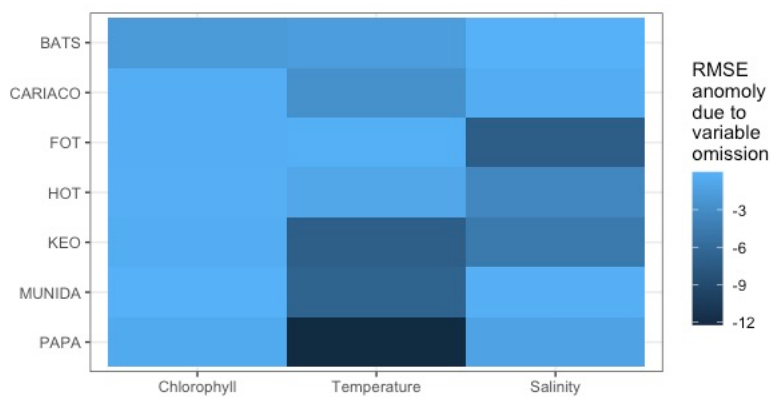



Figure 5. Tile plot showing the change in RMSE per site due to the selective omission of input variables and refitting of the MLR.

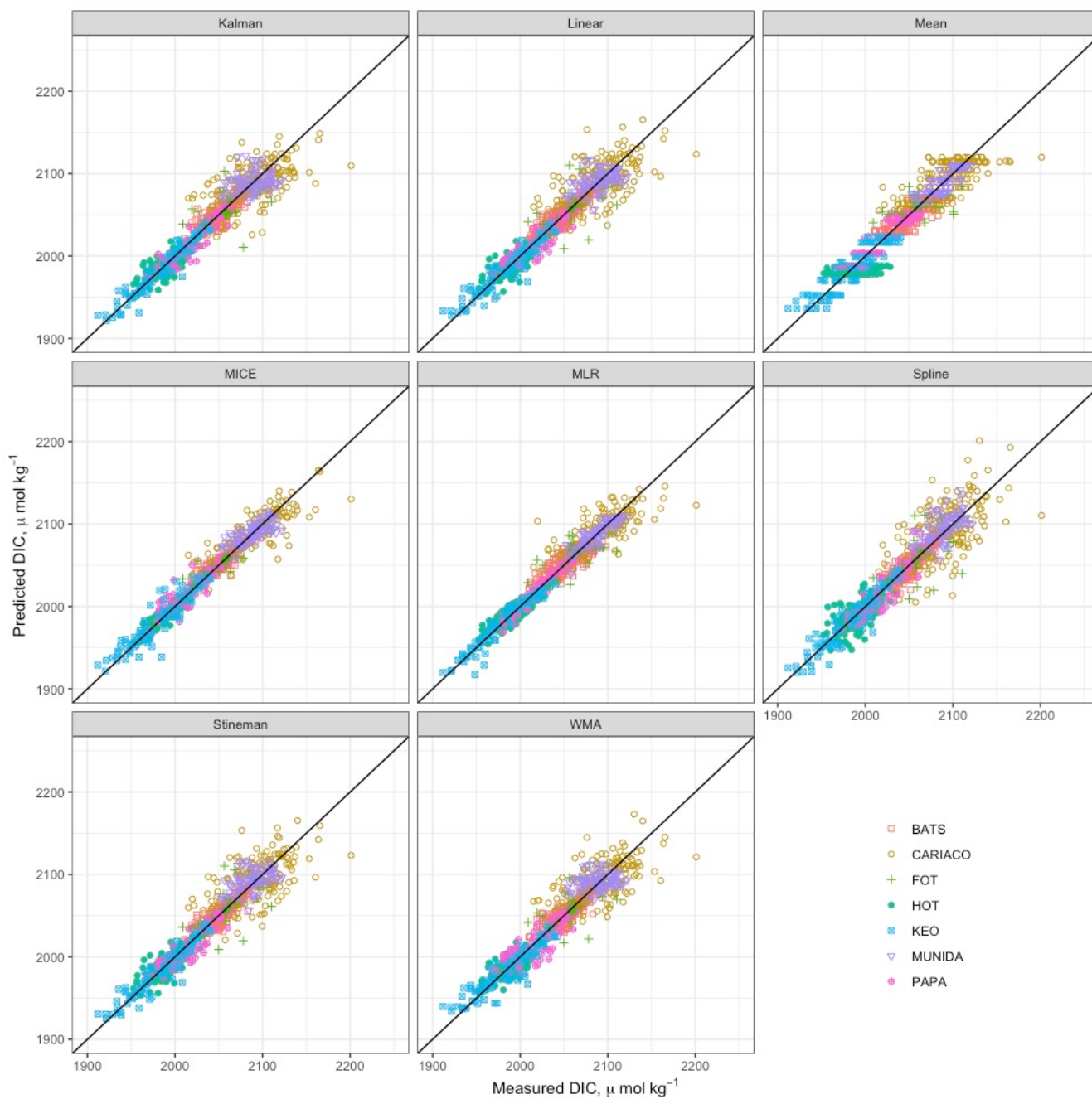



Figure 6. Estimates of DIC based on each imputation model using Leave One Out Cross Validation, for all years >1997 containing >4 monthly samples.

600

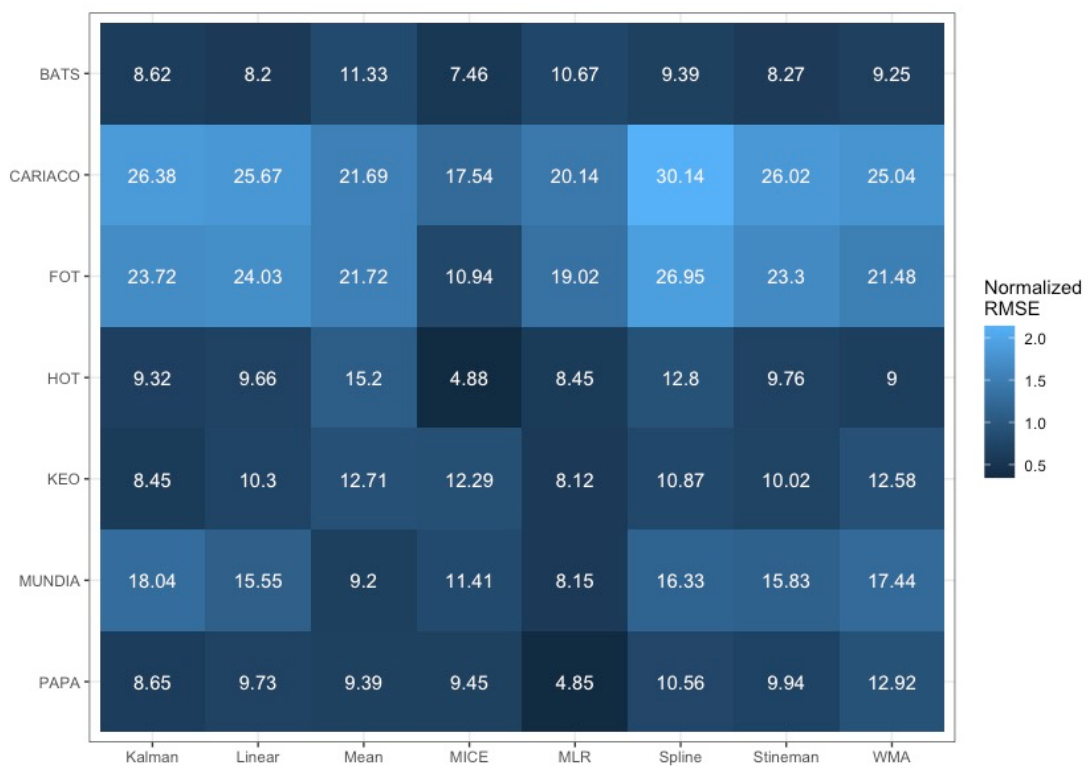



Figure 7. Tile plot showing the RMSE (white text in tiles) for each cross validated imputation methods at each site. Tiles are colored according to RMSE normalized to the mean value across all methods and sites.

(A)

(B)

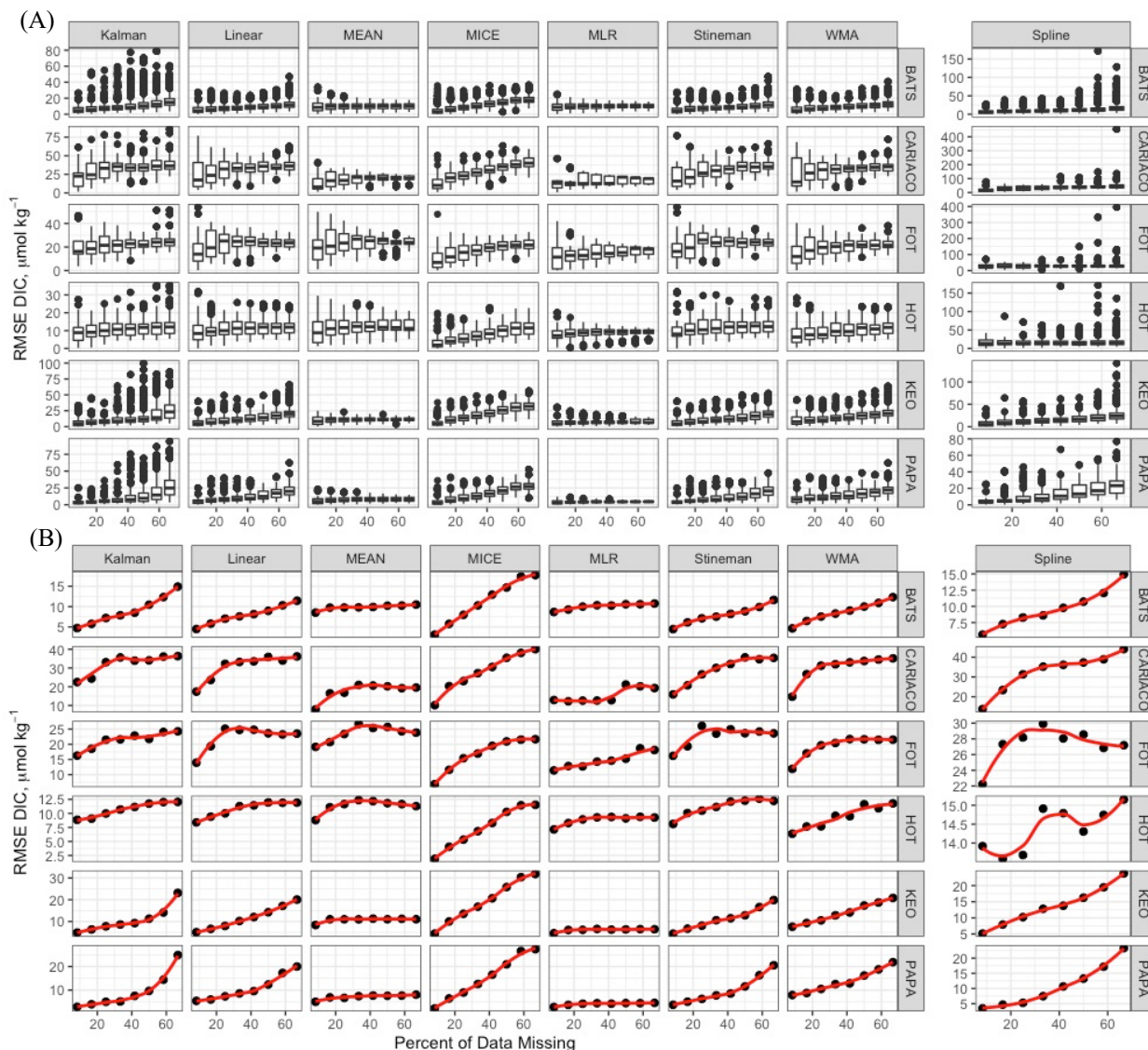



Figure 8. (A) Boxplots of RMSE for each gap assessment corresponding to 8.33%, 16.67%, 25%, 33.33%, 41.67%, 50%, 58.33% and 66.67% data missing rates. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots). (B) Loess fit (red line) of the median error for each gap assessment.

610

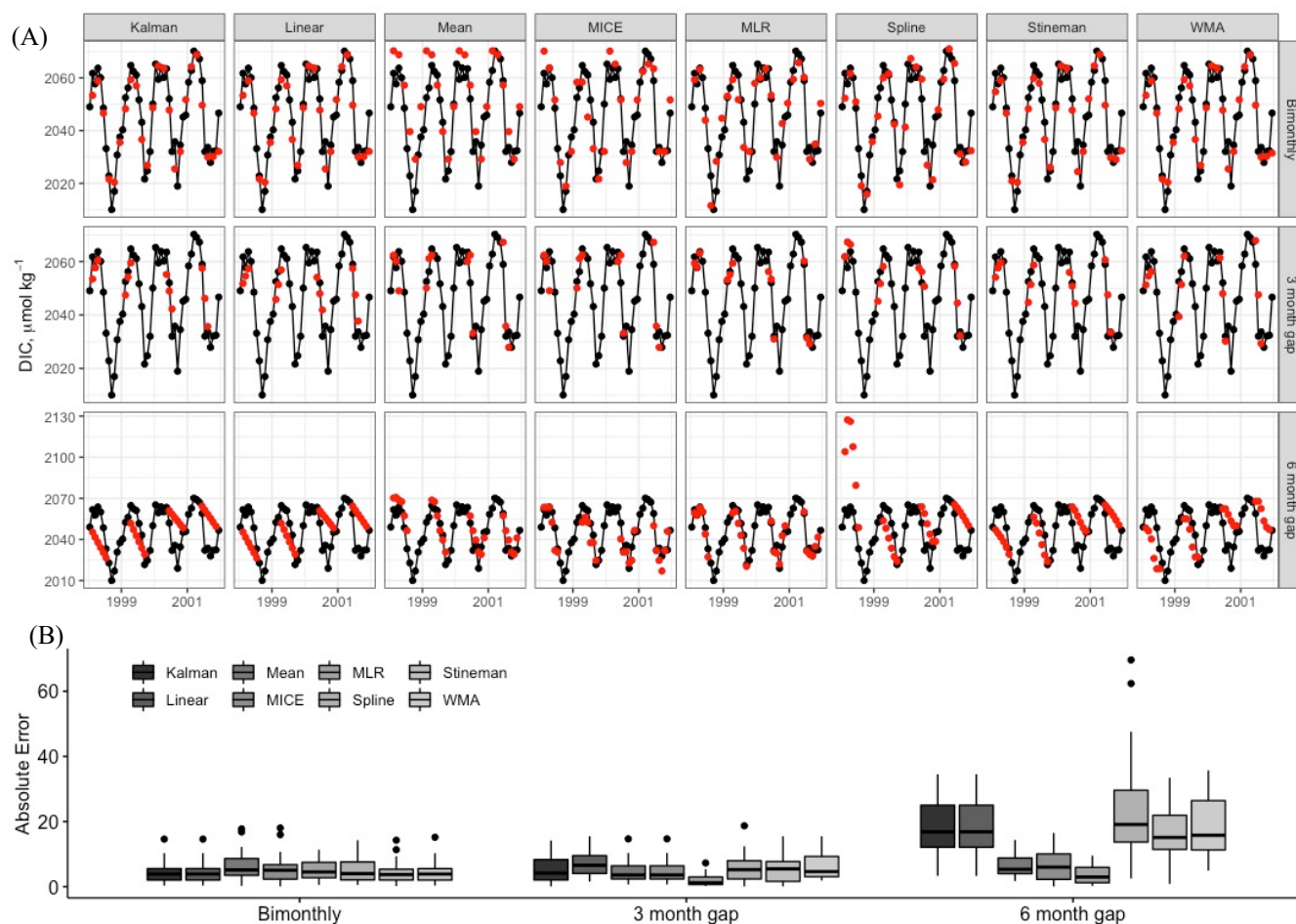



Figure 9. (A) Time series of DIC measurements at BATS (black) with artificial gaps representing bimonthly sampling, 3-month and 6-month sequential gaps. Predicted DIC values for each artificial gap are shown (red) for each imputation model. **(B)** Box and whisker plots of the absolute imputation error for each method are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

615

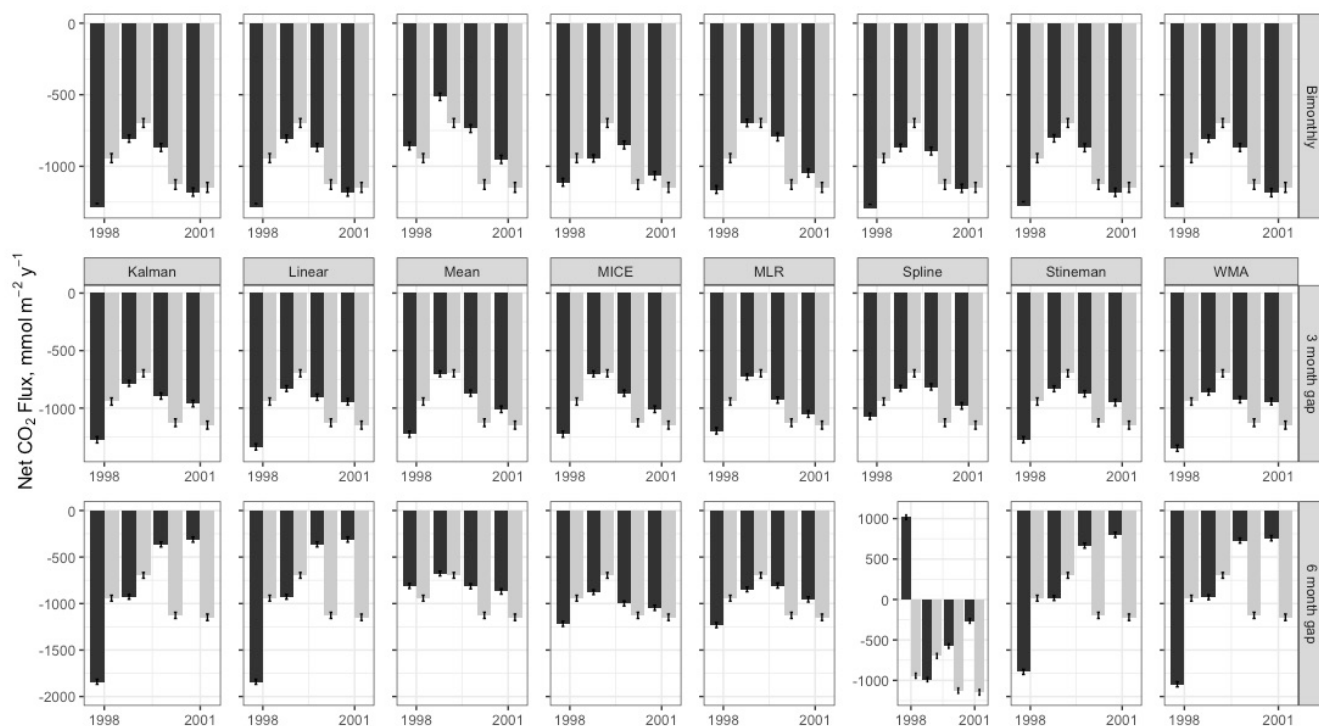



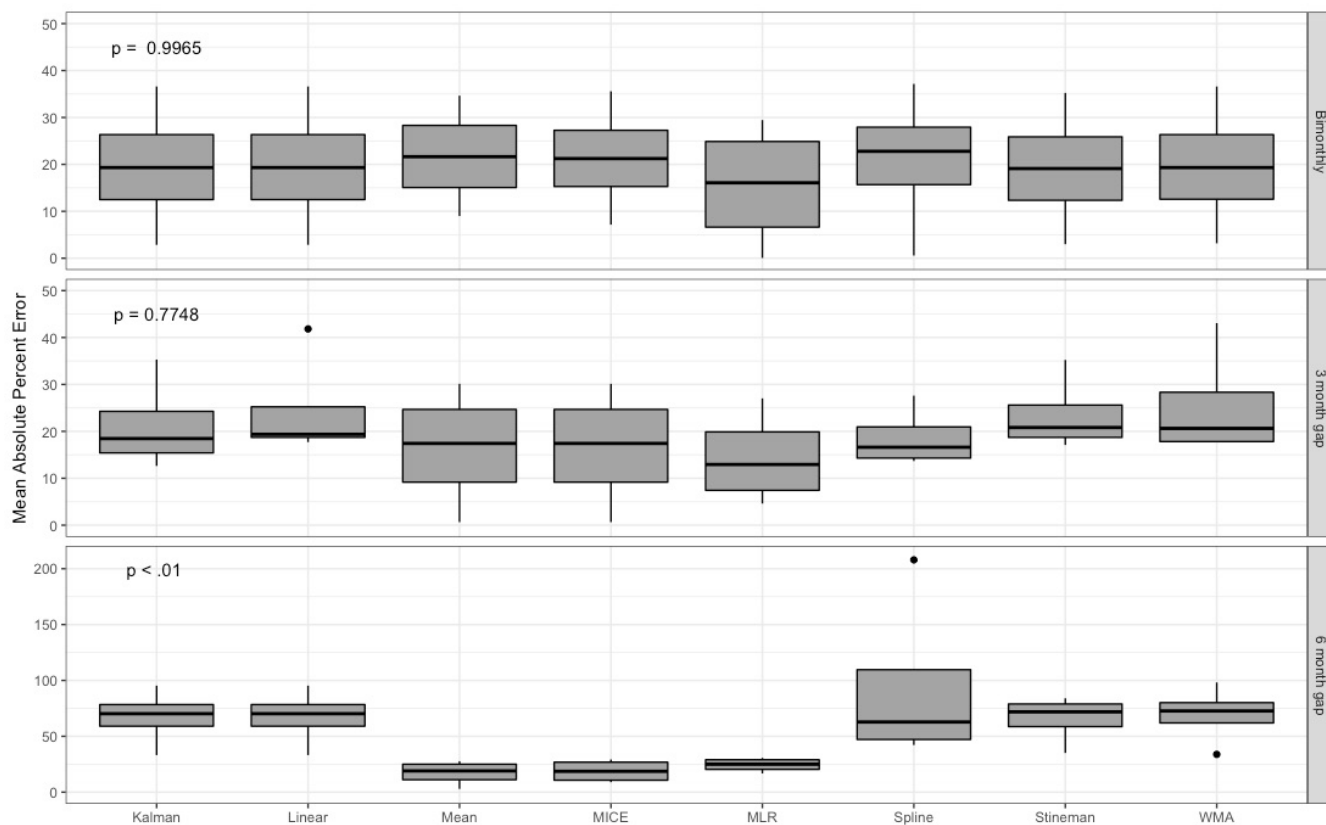
Figure 10. Annual summation of air-sea CO₂ flux using observations (grey) and imputed bimonthly, 3-month gap and 6-month gap time series using each model (black).

620

Figure 11. Absolute percent difference between annual CO$_2$ flux observed at BATS 1998-2001 and the flux estimated from imputed DIC values over bimonthly, 3-month and 6-month sequential gaps per year.

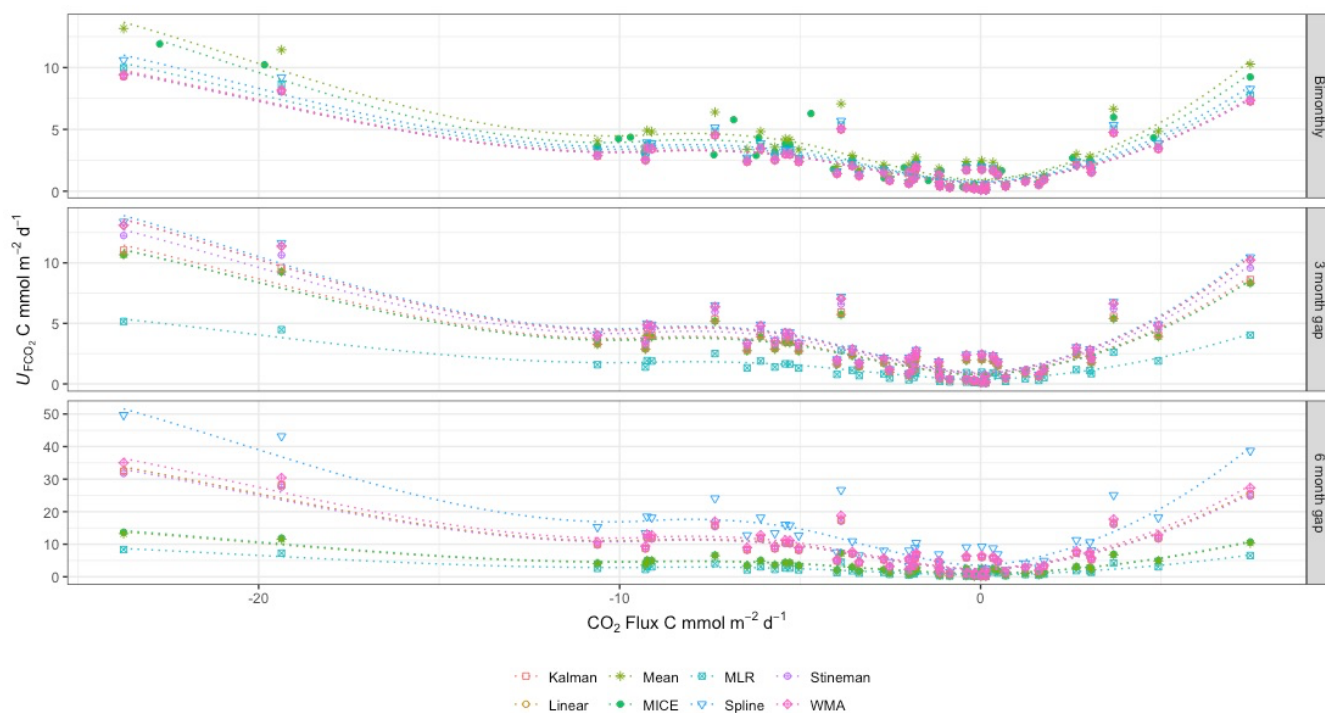



Figure 12. Flux uncertainty for mean monthly air-sea CO2 flux estimates for imputed bimonthly, 3-month gap and 6-month gap
time series at BATS over the period 1998-2001.

Table 1. Information about each sampling site with ocean carbonate time series used in our analyses, including Bermuda Atlantic Time-series (BATS), Carbon Retention In A Coloded Ocean (CARIACO), Firth of Thames (FOT), Hawaiian Ocean Time-series (HOT), Kuroshio Extension Observatory (KEO), Munida Time-series (MUNIDA), and Ocean Site Papa (PAPA). DIC = dissolved inorganic carbon. TA = total alkalinity. pCO_2 = partial pressure of carbon dioxide. pH = -log[H^+]. Gap rate based on expected sampling frequency.

Site Type	Time series Site	Sampling Region	Location	Time series Duration	Sampling Frequency	Gap Rate	Carbonate Measurements
Sampling Site	BATS	Sargasso Sea	31.88°N, 64.26°W	1983 - present	[1]monthly	4%	DIC/TA
	HOT	North Pacific	22.67°N, 158°W	1988 - present	[2]monthly	15%	TA/pH
	CARIACO	Cariaco Basin	10.5°N, 64.67°W	1995 - present	monthly	16%	TA/pH
	MUNIDA	South Pacific	45.8°S, 171.5°E	1998 - present	[3]bimonthly	5%	pCO_2/TA
Mooring	PAPA	North Pacific	50.13°N, 144.83°W	2007 - present	3 hours	26%	pH/pCO_2
	KEO	North Pacific	32.25°N, 144.56°E	2004 - present	3 hours	18%	pH/pCO_2
	FOT	New Zealand Coast	36.88°S, 175.32°E	2015 - present	15 minutes	59%	pH

Web addresses for site information and data access:
BATS: http://www.bios.edu/research/projects/bats/
HOT: https://hahana.soest.hawaii.edu/hot/
CARIACO: http://www.imars.usf.edu/cariaco
Munida: https://marinedata.niwa.co.nz/nzoa-on/
Papa: https://www.pmel.noaa.gov/ocs/Papa
KEO: https://www.pmel.noaa.gov/ocs/KEO

FOT: https://marinedata.niwa.co.nz/nzoa-on/
*Sampling began in 1998, mooring installed in 2015
[1]BATS sampling target is at least monthly
[2]HOT sampling target is approximately monthly
[3]Munida sampling is typically bimonthly, varying with conditions and additional coordinated voyages

Table 2. Pearson correlation coefficients between DIC and chlorophyll, temperature and salinity in the surface layer across test sites. Asterisks indicate weak correlations (threshold = 0.3).

Site	Pearson Correlation Coefficient		
	Chlorophyll	**Temperature**	**Salinity**
KEO	0.49	-0.91	0.87
BATS	0.48	-0.73	0.65
PAPA	-0.34	-0.97	0.73
FOT	-0.22*	0.24*	0.74
HOT	0.1*	-0.51	0.74
CARIACO	0.53	-0.77	0.58
MUNIDA	-0.37	-0.87	0.32

670

Table 3. Years with 12 monthly samples per site. *Actual sampling interval greater than monthly

Time-Series Site	Years With 12 Monthly Samples	N Years
BATS	1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2004, 2005, 2007, 2008, 2012, 2013	17
HOT	1998, 2004, 2006	3
CARIACO	2008	1
MUNIDA	NA*	0
PAPA	2015, 2016, 2017	3
KEO	2009, 2010, 2014, 2015, 2016	5
FOT	2016	1

675

680

Table 4. Results of cross validated MLR model for estimating DIC at each individual site, and at grouped oceanic (BATS, HOT, KEO, MUNIDA, PAPA) and coastal (FOT, CARIACO) sites, including the mean and standard deviation of each coefficient for *N* 685 **LOOCV iterations.**

Site	RMSE	RRMSE	R^2	MAE	BIAS	N	α	β1	β2	β3
BATS	10.67	0.52	0.6611	8.93	0.017	208	401.65±13.75	-13.48±1.56	-3.53±0.03	47.53±0.36
CARIACO	20.14	0.96	0.5861	14.94	0.015	153	1446.46±40.07	2.50±0.10	-10.16±0.12	24.37±1.02
FOT	19.02	0.92	0.3958	15.13	0.099	28	718.32±47.59	8.30±2.53	0.47±0.35	37.93±1.26
HOT	8.45	0.42	0.6178	7.40	0.029	204	276.44±9.51	-82.88±2.25	-3.47±0.04	51.44±0.26
KEO	8.12	0.41	0.9330	6.12	0.061	90	-208.45±16.79	-27.85±1.01	-4.61±0.03	66.36±0.48
MUNIDA	8.15	0.39	0.7564	6.48	0.029	109	1069.11±65.27	4.77±1.05	-7.69±0.08	32.00±1.89
PAPA	4.85	0.24	0.9631	3.74	0.035	94	799.13±17.96	-16.47±0.52	-6.55±0.02	39.82±0.55
Oceanic	8.75	0.43	0.9567	7.09	0.030	671	412.04±356.85	-34.86±32.81	-4.54±1.53	48.35±9.35
Coastal	19.97	0.95	0.6078	14.97	0.028	181	1333.82±267.23	3.40±2.32	-8.52±3.86	26.47±5.03

Table 5. Model results for selective omission of input variables.

Variable Omitted	RMSE	RRMSE	R^2	MAE	BIAS
none	12.044	0.591	0.9352	8.764	0.030
chlorophyll	12.106	0.594	0.9345	8.849	0.005
temperature	15.526	0.762	0.8923	11.871	0.013
salinity	13.998	0.687	0.9124	10.285	0.022

Table 6. Performance metrics for cross validated imputation models across all sites.

Model	RMSE	RRMSE	R^2	MAE	BIAS
Mean	15.07	0.74	0.905	11.65	2.586
Linear	14.79	0.72	0.895	10.09	-0.141
Spline	17.64	0.86	0.874	12.21	-0.252
Stineman	15.28	0.75	0.900	10.51	-0.036
Kalman	15.40	0.76	0.896	10.49	0.570
WMA	15.50	0.76	0.896	11.01	0.069
MLR	12.04	0.59	0.935	8.76	0.030
MICE	10.92	0.54	0.949	7.08	-0.297

690

Table 7. RMSE for each imputed time series and p-value from 1-way ANOVA for each time series.

Missingness	Kalman	Linear	Mean	MICE	MLR	Spline	Stineman	WMA	*p*
Bimonthly	5.59	5.59	7.90	7.10	5.99	6.36	5.55	5.65	0.591
3-month gap	6.63	7.86	6.41	6.41	3.10	8.04	7.36	7.86	0.207
6-month gap	19.31	19.31	7.95	8.19	5.02	29.18	18.90	20.80	<0.0001

695