



Supplement of

An empirical MLR for estimating surface layer DIC and a comparative assessment to other gap-filling techniques for ocean carbon time series

Jesse M. Vance et al.

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

The copyright of individual parts of the supplement might differ from the article licence.

Below are additional figures not provided in the manuscript which were part of analyses in this study.

1 Firth of Thames (FOT) Alkalinity – Salinity Relationship and Model Fit

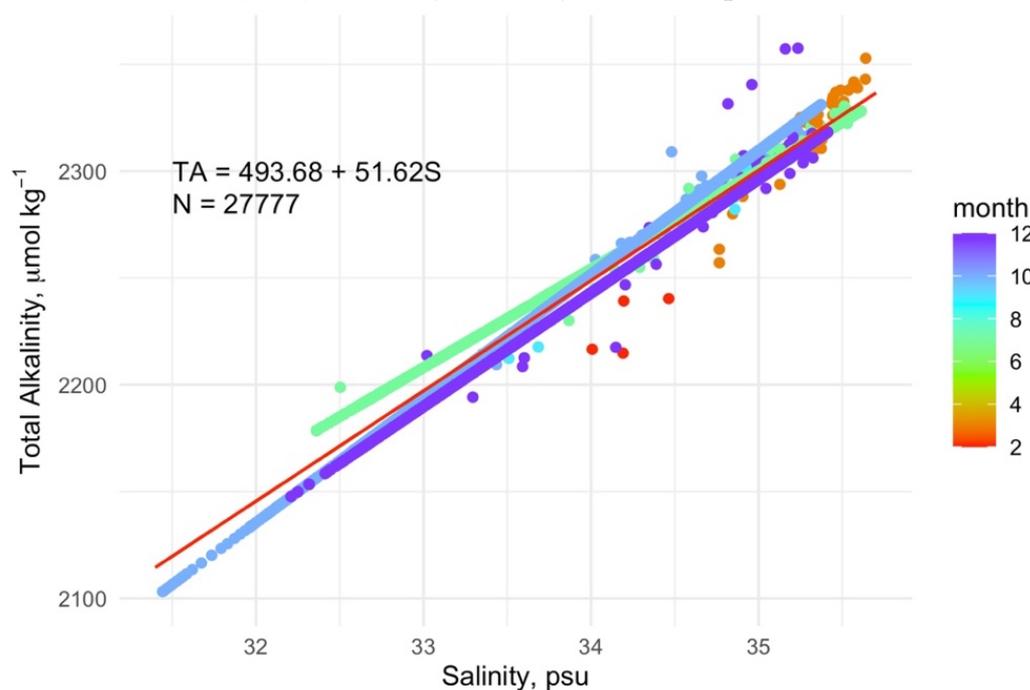


Figure 1. Linear regression between salinity and alkalinity measured in the Firth of Thames. Includes unpublished data from bottle sampling and shipboard measurements taken from all sampling campaigns carried out by the National Institute for Water and Atmospheric Research (NIWA). Points are colored according to month, and indicate seasonal sampling campaigns, indicating some seasonal variability in the salinity-alkalinity relationship between winter (green), summer (blue), autumn (purple), and spring (orange).

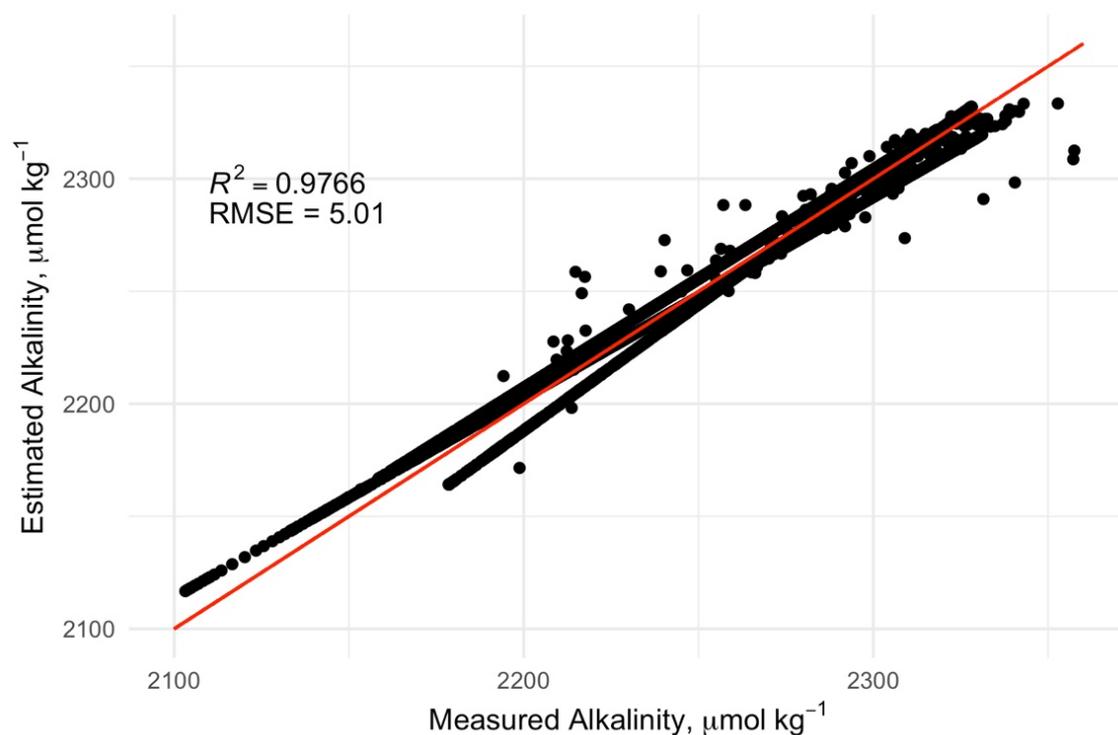


Figure 2. Linear regression of measured alkalinity and alkalinity estimated from salinity, with correlation and root-mean-square error provided.

2 FOT Remotely Sensed Chlorophyll Data Comparison Between MODIS and VIIRS.

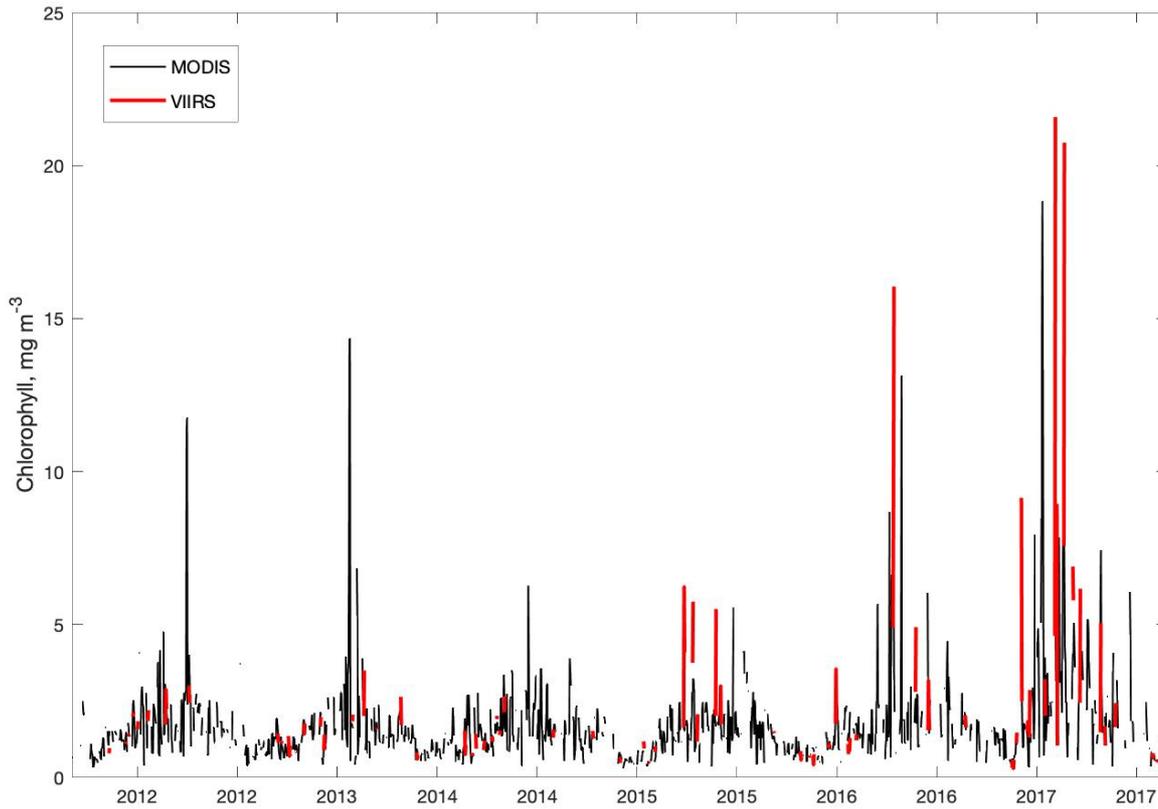


Figure 3. Time series of daily values for remotely sensed chlorophyll in a 4km² cell at the location of the outer FOT mooring, illustrating the difference in observations between platforms for the study site: 21.9% for MODIS data vs. 72.7% for VIIRS data.

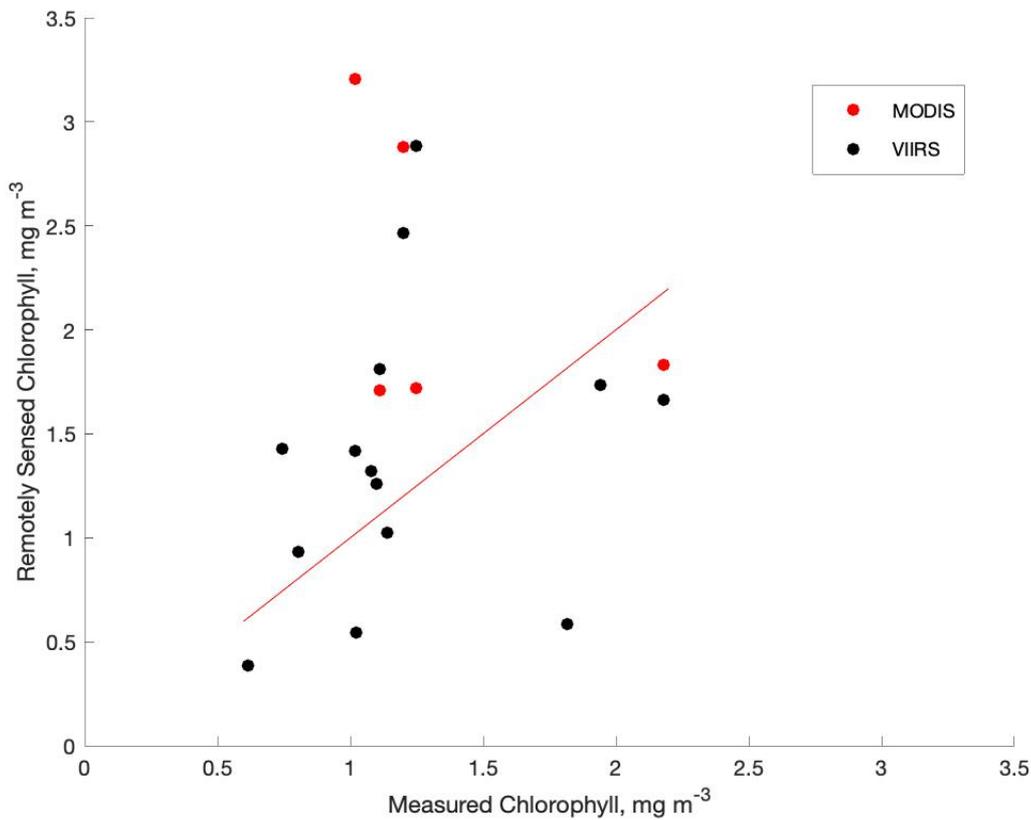


Figure 4. Regression between chlorophyll measured at the outer FOT mooring and remotely sensed chlorophyll estimated by MODIS (red) and VIIRS (black).

3 GLORYS Temperature and Salinity Model Performance per Test Site

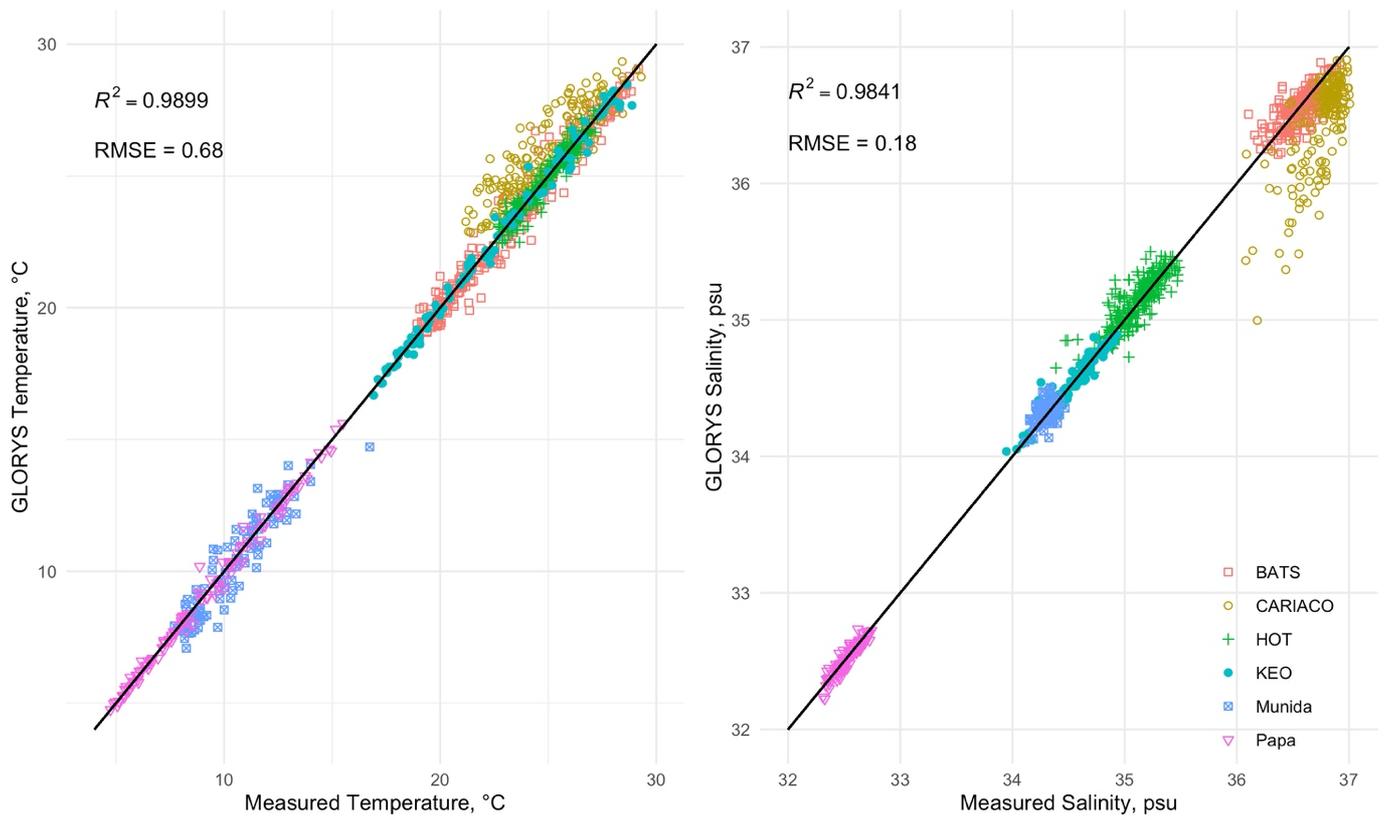


Figure 5. Linear regressions between measured and modelled temperature (left) and salinity (right) pooled across test sites (BATS, CARIACO, HOT, KEO, Munida, and Papa), with correlation and root-mean-square errors indicated for each.

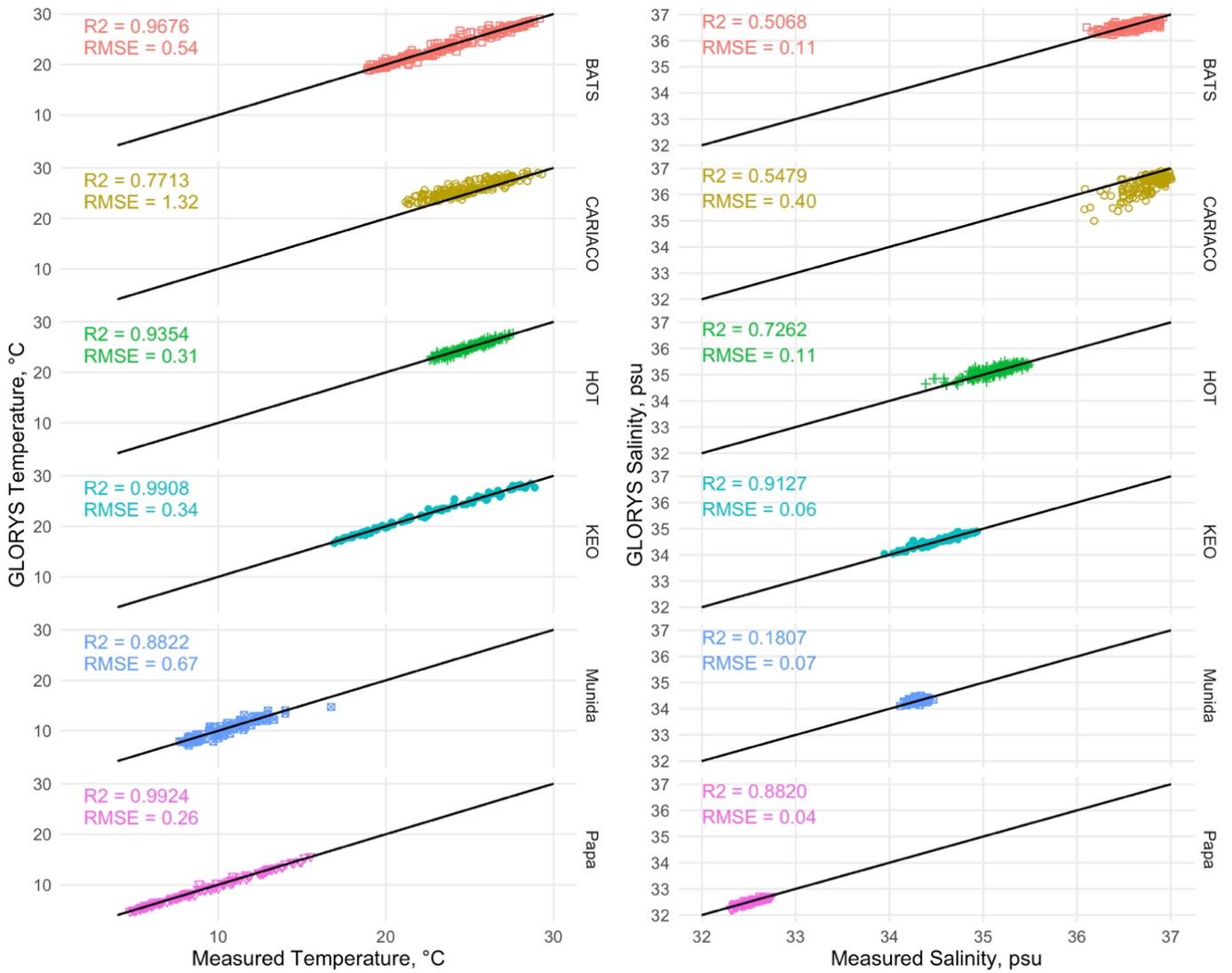


Figure 6. Linear regressions between measured and modelled temperature (left) and salinity (right) for individual test sites (BATS, CARIACO, HOT, KEO, Munida, and Papa), with correlation and root-mean-square errors indicated for each.

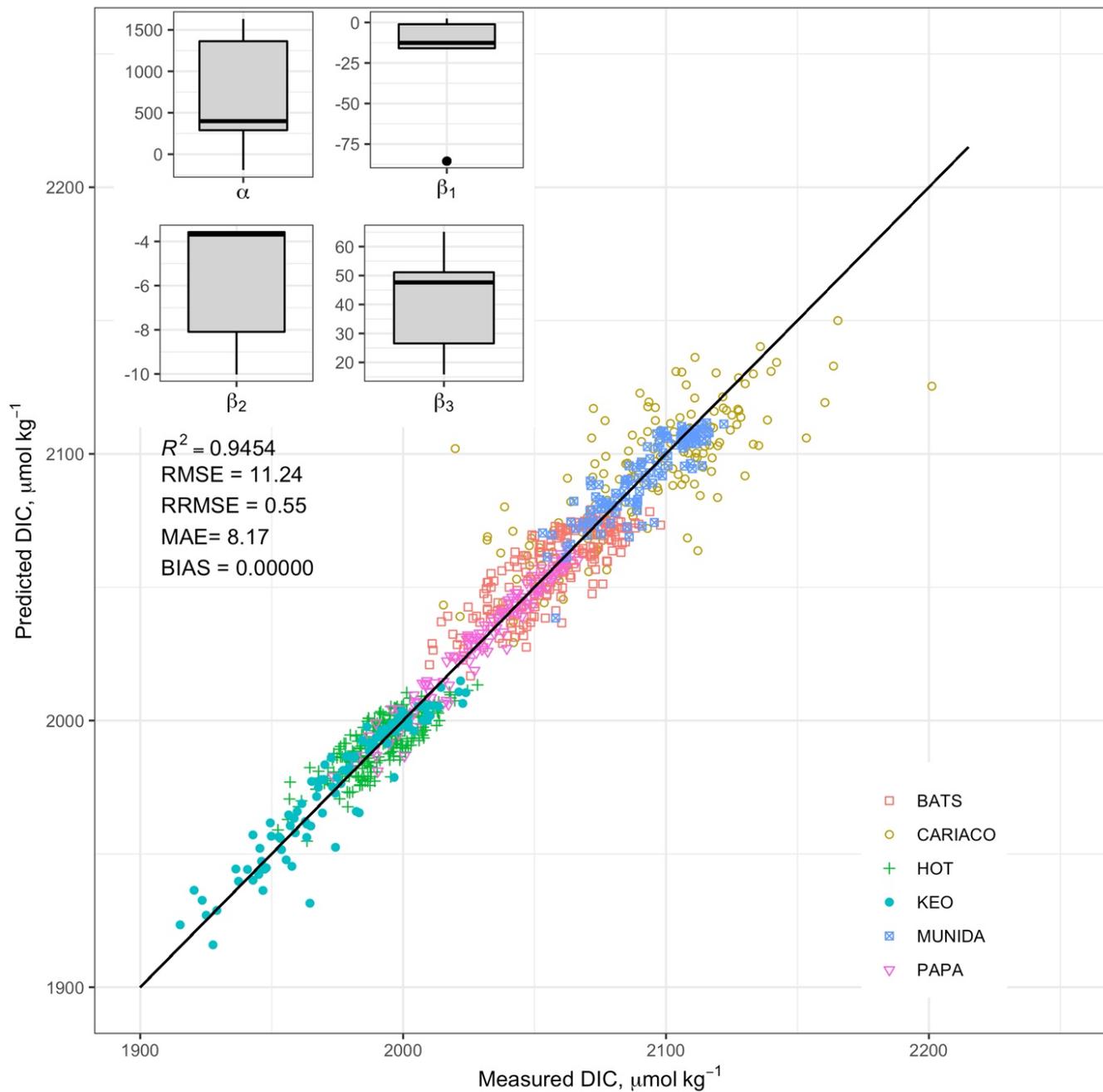


Figure 7. Composite of predicted and measured DIC values using a multiple linear regression model based on temperature, and salinity from GLORYS reanalysis data products with remotely sensed chlorophyll from sites: Bermuda Atlantic Time-series Study (BATS); Carbon Retention In A Colored Ocean (CARIACO); 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).

4 Imputed Time Series

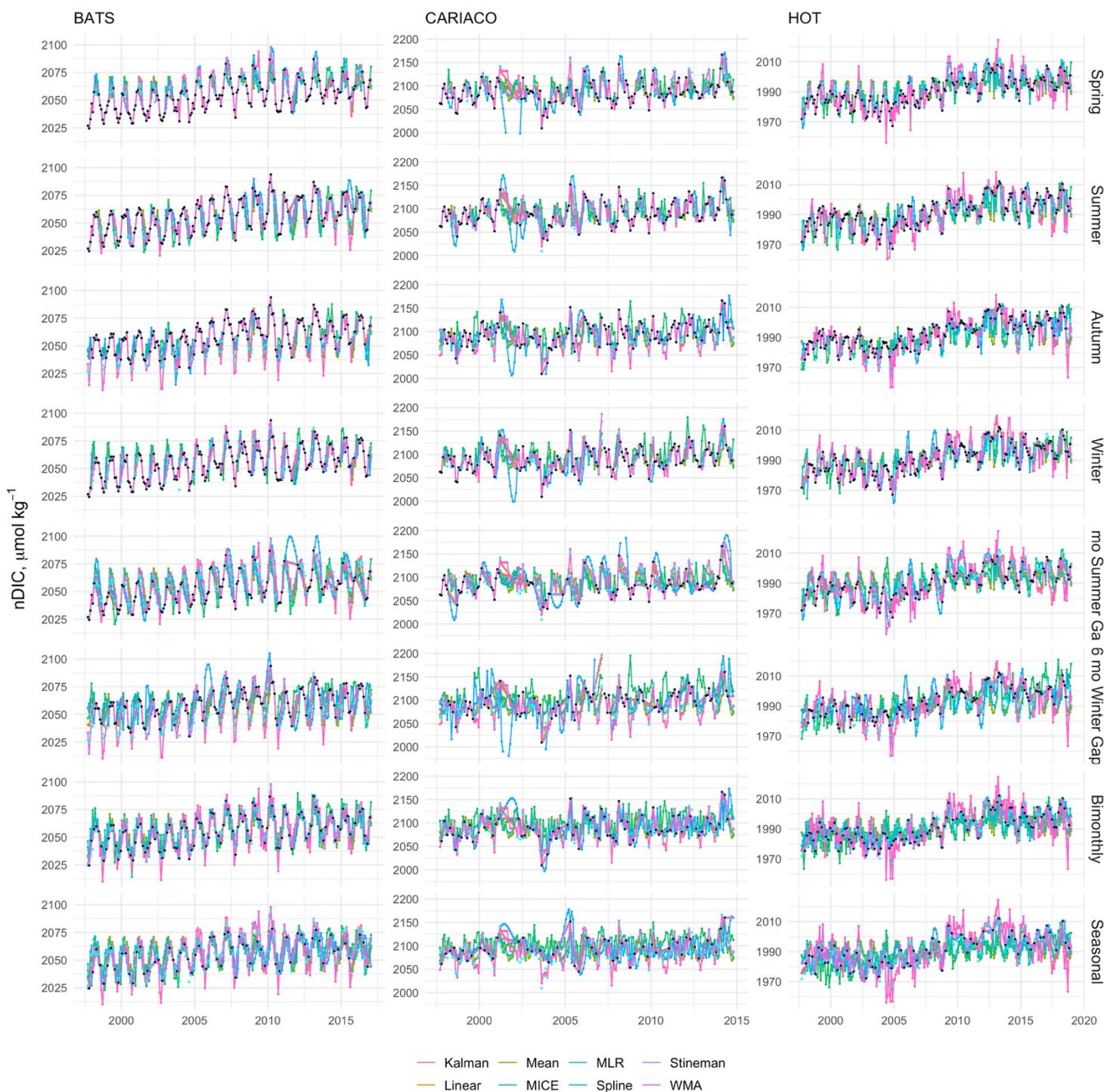


Figure 8. Composites of reconstructed time series of nDIC measurements from BATS, CARIACO, and HOT. Observations were selectively removed using eight gap filters: 3-month sequential seasonal filters for Spring, Summer, Autumn, and Winter; 6-month sequential gaps centered on summer and winter; and bimonthly (odd months) and seasonal (1 max, 1 min. and 2 transition samples) sampling regimes and gaps were filled using Kalman filter with a state space model, linear interpolation, mean imputation, empirical multiple linear regression (MLR), multiple imputation by chained equations (MICE), spline interpolation, Stineman interpolation and exponential weighted moving average (WMA). Training observations are shown as black points, while testing data (removed observations) are shown as cyan points.

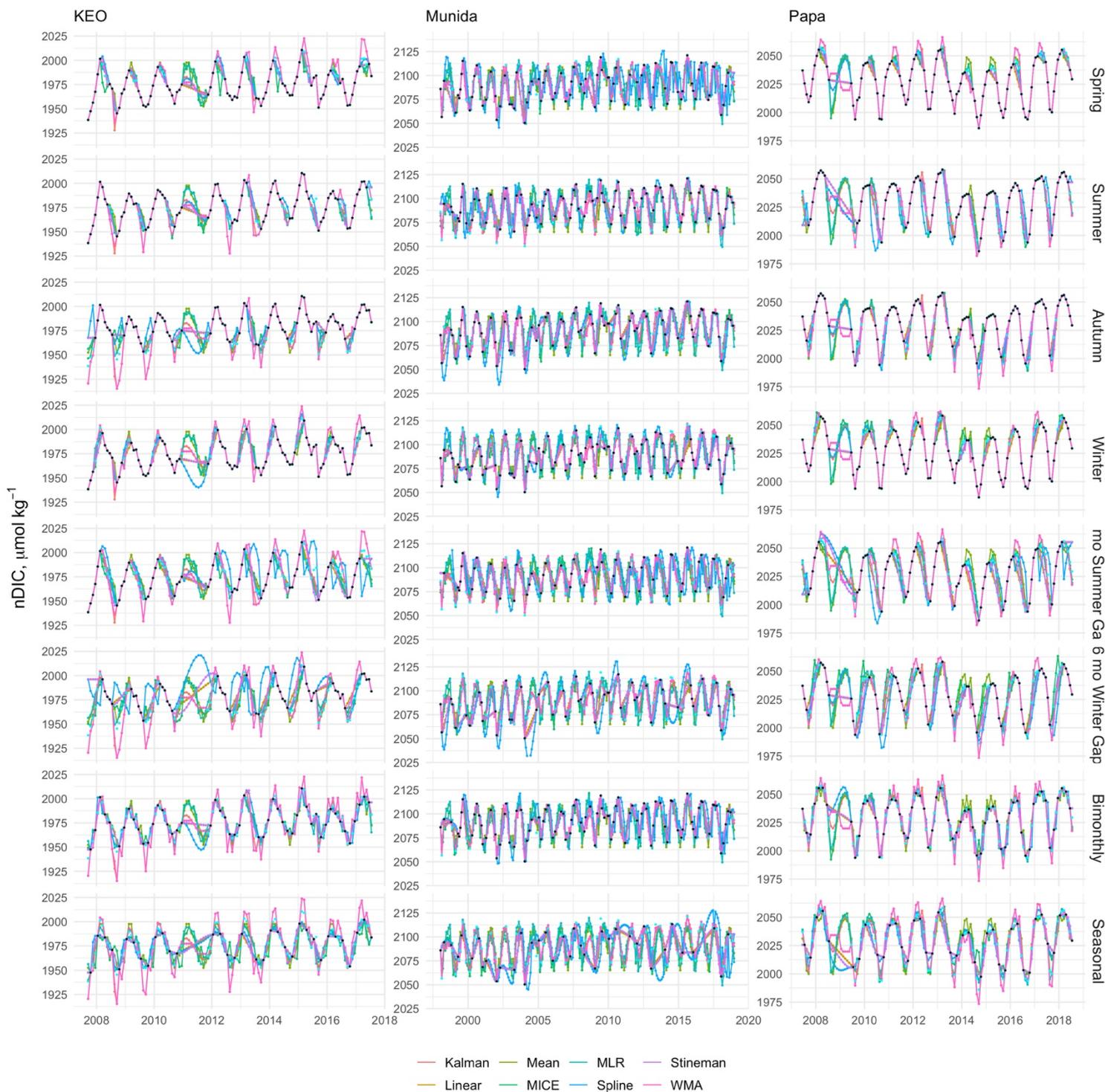


Figure 9. Composites of reconstructed time series of nDIC measurements from KEO, Munida, and Papa. Observations were selectively removed using eight gap filters: 3-month sequential seasonal filters for Spring, Summer, Autumn, and Winter; 6-month sequential gaps centered on summer and winter; and bimonthly (odd months) and seasonal (1 max, 1 min. and 2 transition samples) sampling regimes and gaps were filled using Kalman filter with a state space model, linear interpolation, mean imputation, empirical multiple linear regression (MLR), multiple imputation by chained equations (MICE), spline interpolation, Stineman interpolation and exponential weighted moving average (WMA). Training observations are shown as black points, while testing data (removed observations) are shown as cyan points.

5 Pooled Trend Errors

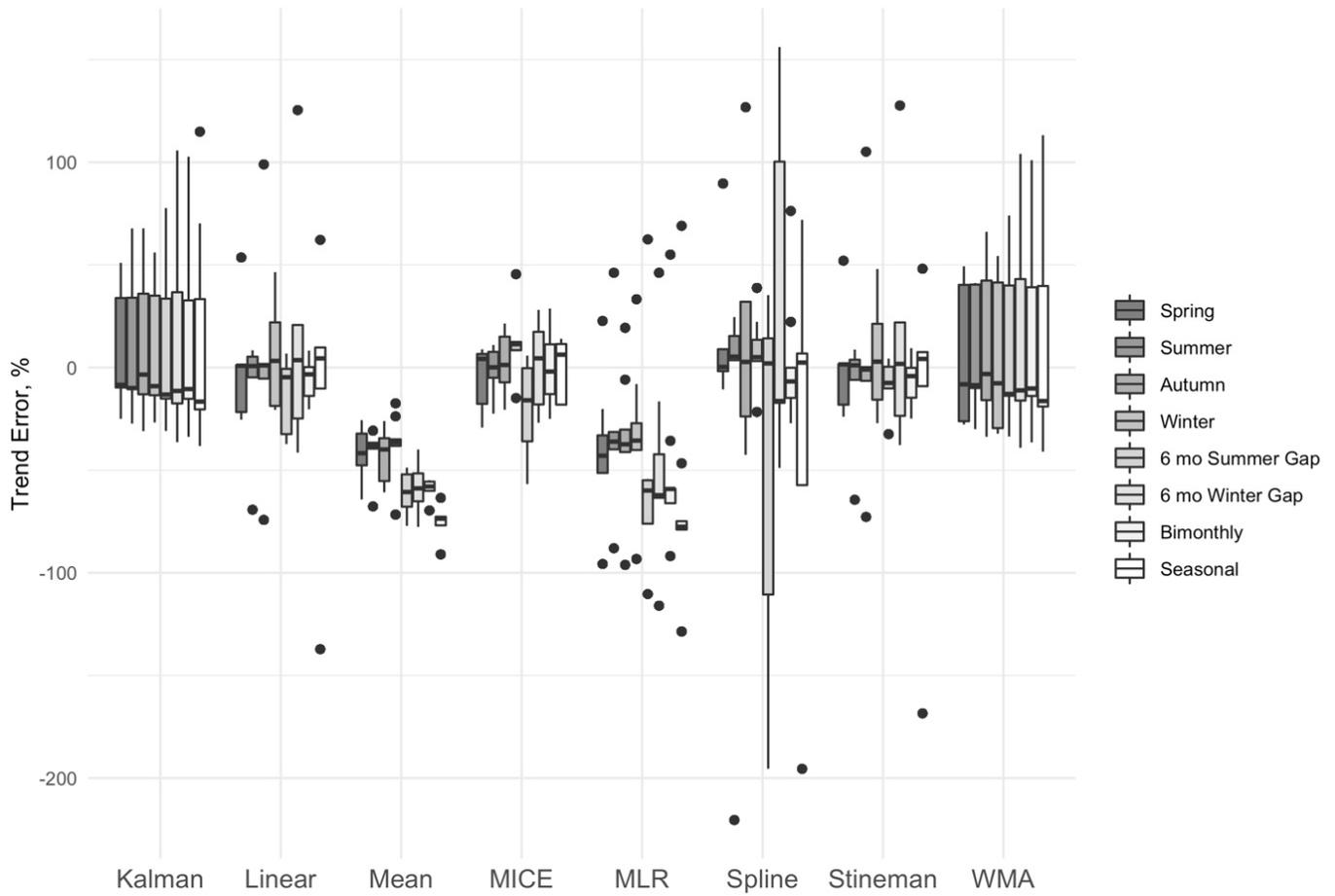


Figure 10. Box and whisker plots of percent error between trends in observed and reconstructed time series pooled across sites using each imputation model for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

6 Pooled Interannual Variability Errors

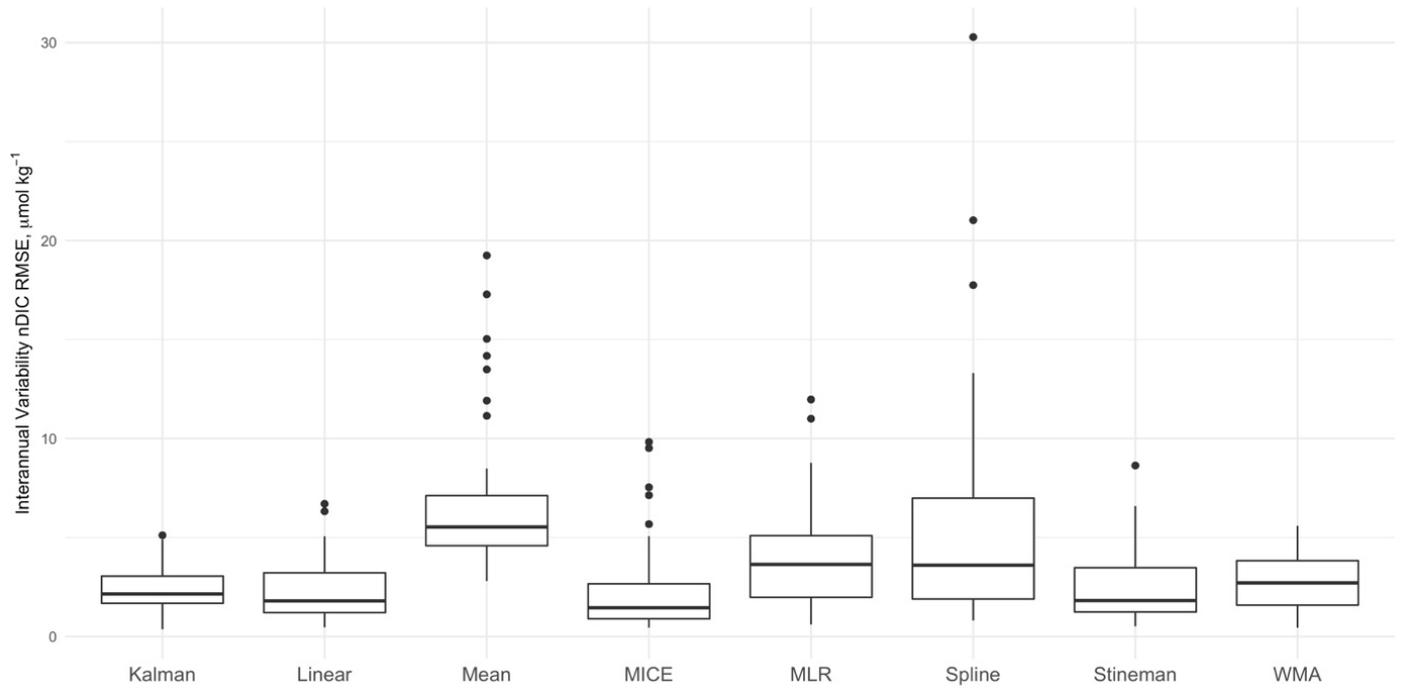


Figure 11. Box and whisker plots of root-mean-square error between observed and reconstructed interannual variability pooled across sites and gap filters using each imputation model. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

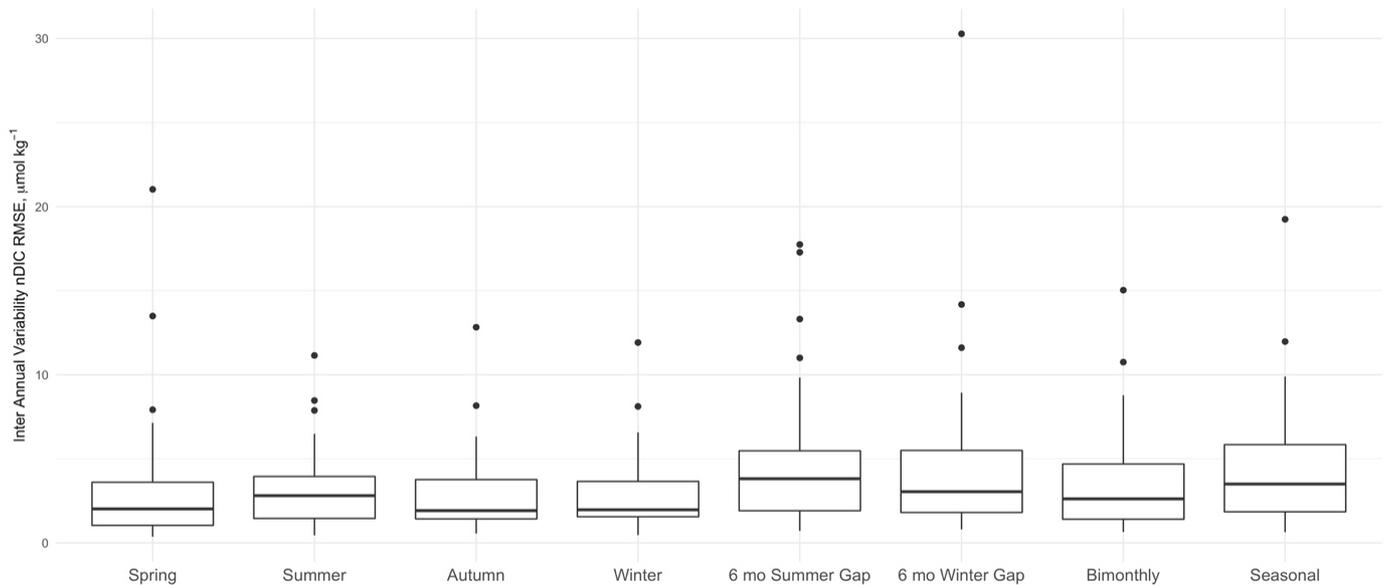


Figure 12. Box and whisker plots of root-mean-square error between observed and reconstructed interannual variability pooled across sites and method for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

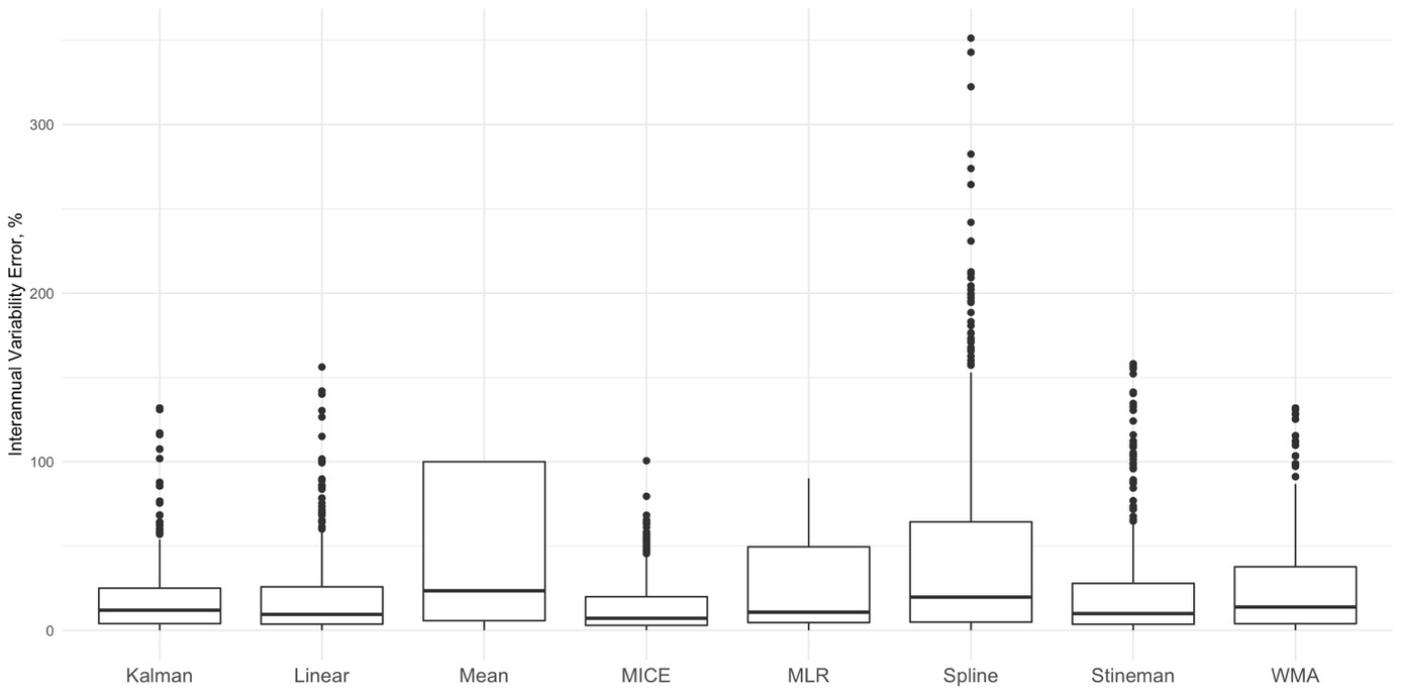


Figure 13. Box and whisker plots of percent error between observed and reconstructed interannual variability pooled across sites and gap filters using each imputation model. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

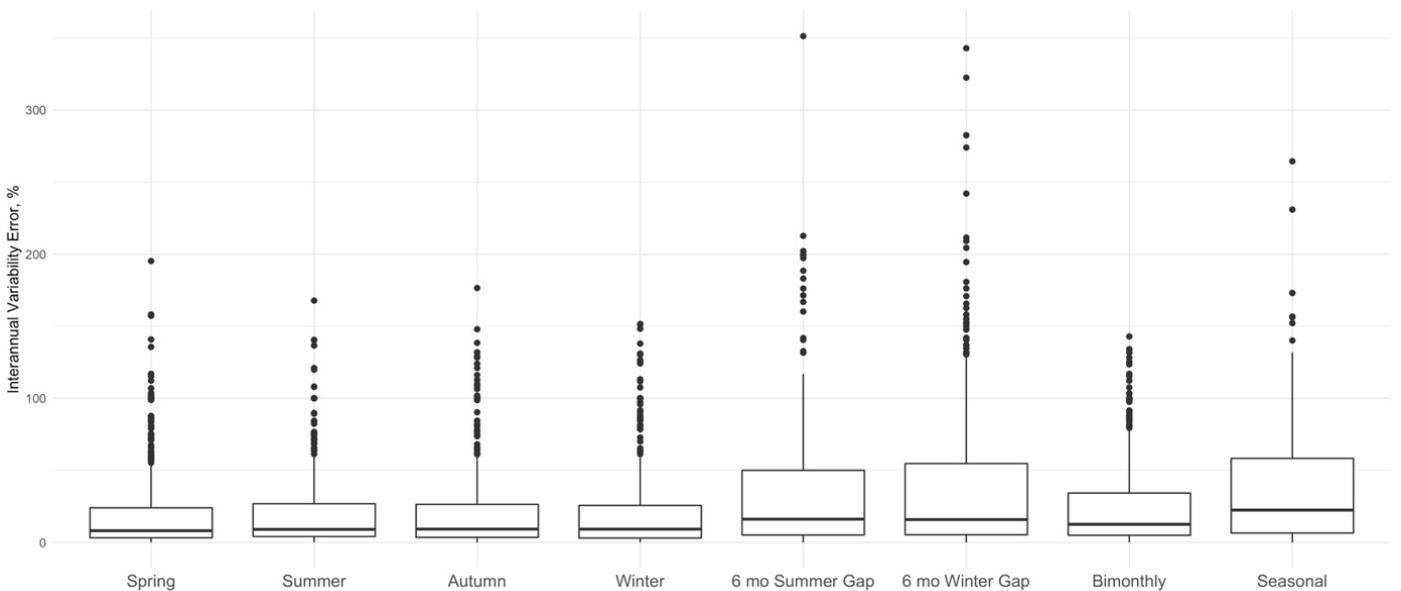


Figure 14. Box and whisker plots of percent error between observed and reconstructed interannual variability pooled across sites and method for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

6 Pooled Annual Mean Errors

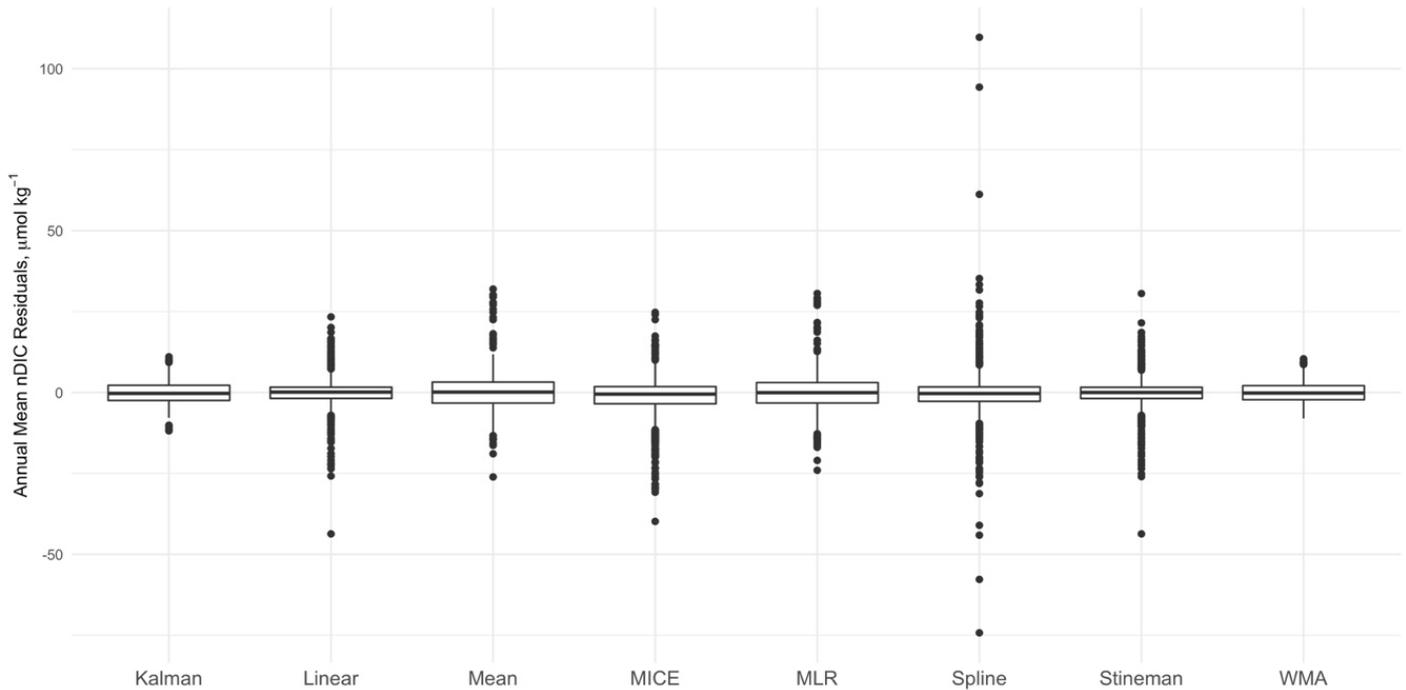


Figure 15. Box and whisker plots of residuals between annual means calculated from observed and reconstructed time series pooled across sites and gap filters using each imputation model. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

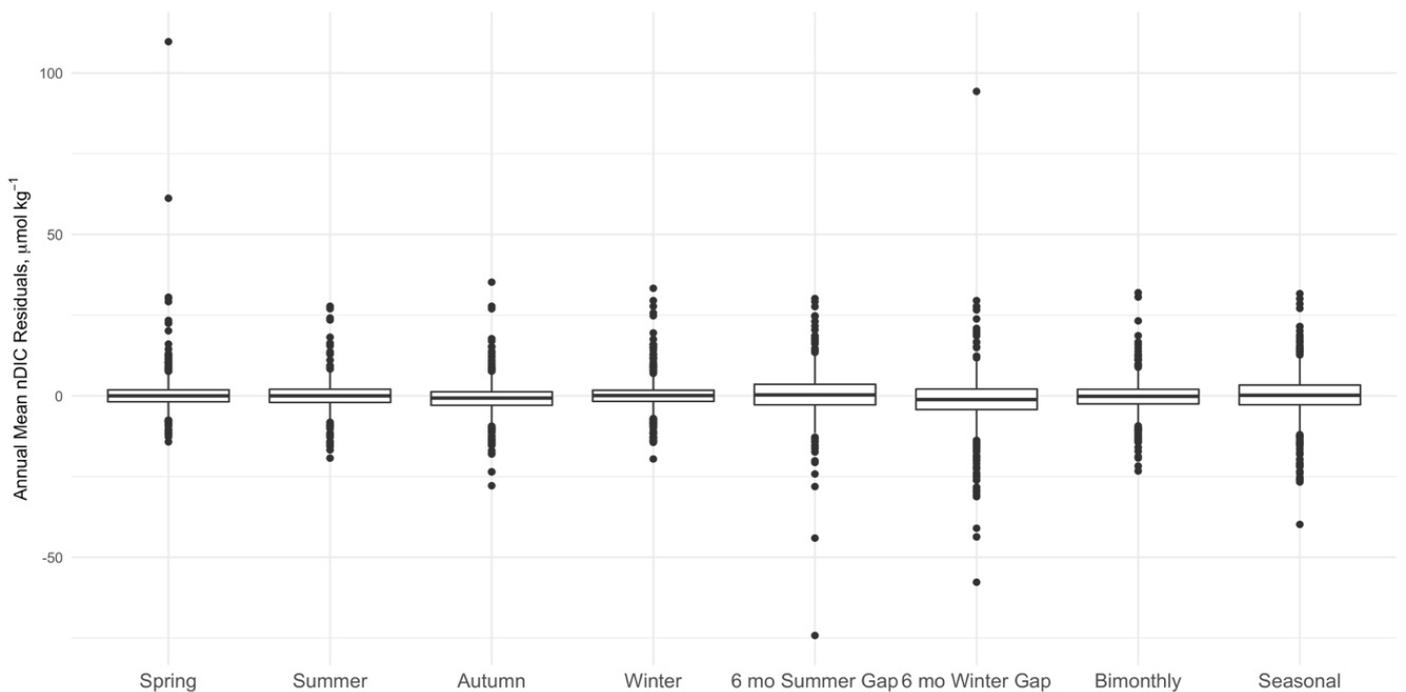


Figure 16. Box and whisker plots of residuals between annual means calculated from observed and reconstructed time series pooled across sites and method for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

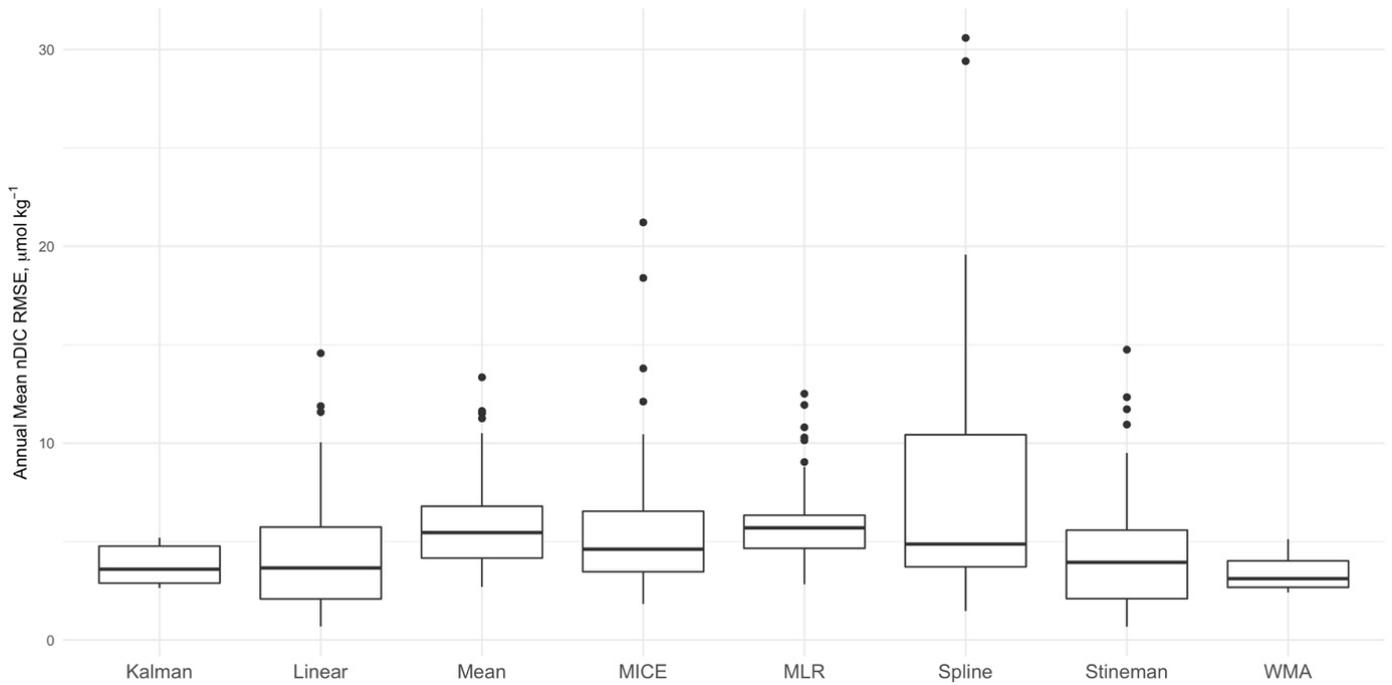


Figure 17. Box and whisker plots of root-mean-square errors of annual means calculated from observed and reconstructed time series pooled across sites and gap filters using each imputation model. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

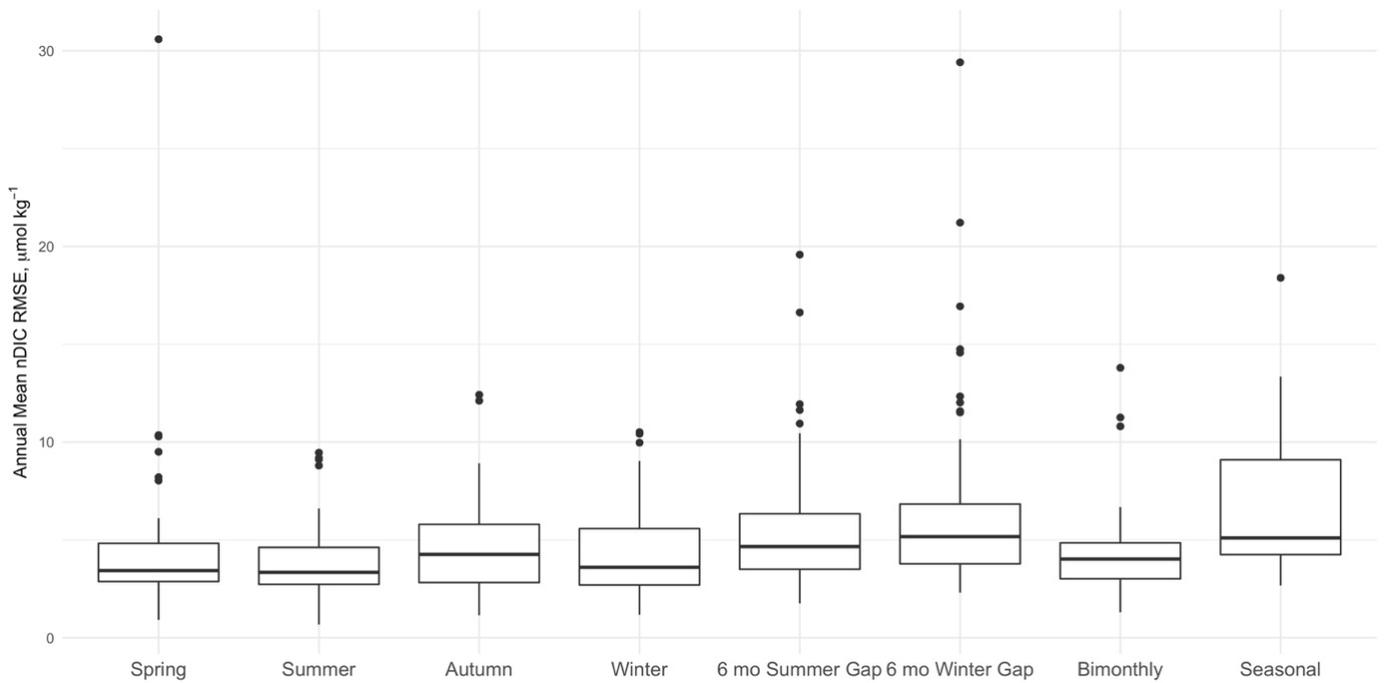


Figure 18. Box and whisker plots of root-mean-square errors of annual means calculated from observed and reconstructed time series pooled across sites and method for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

7 Pooled Seasonal Errors

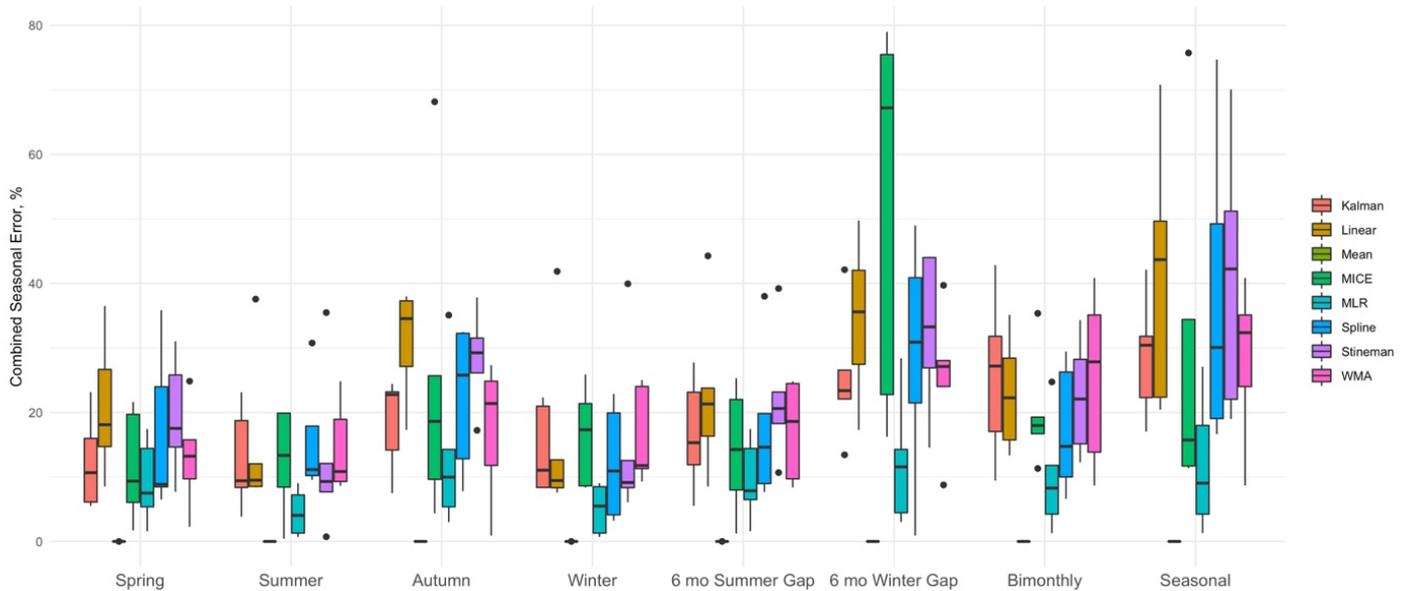


Figure 19. Box and whisker plots of combined seasonal errors calculated from reconstructed time series pooled across sites for each imputation model and gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

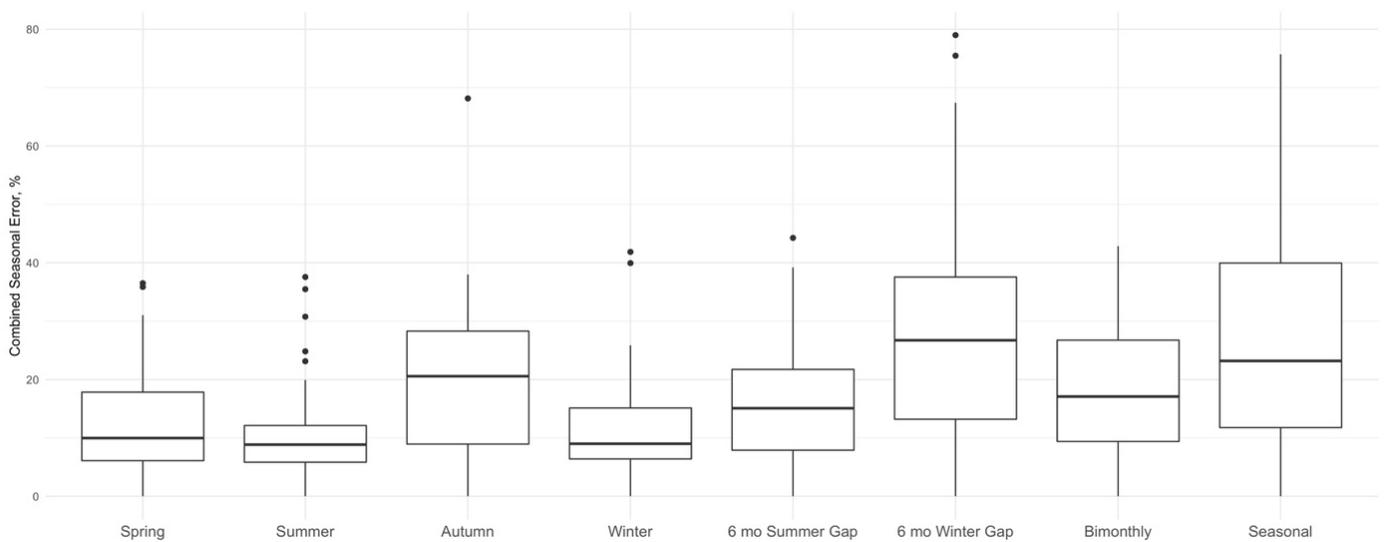


Figure 20. Box and whisker plots of combined seasonal errors calculated from reconstructed time series pooled across sites and method for each data gap filter. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).

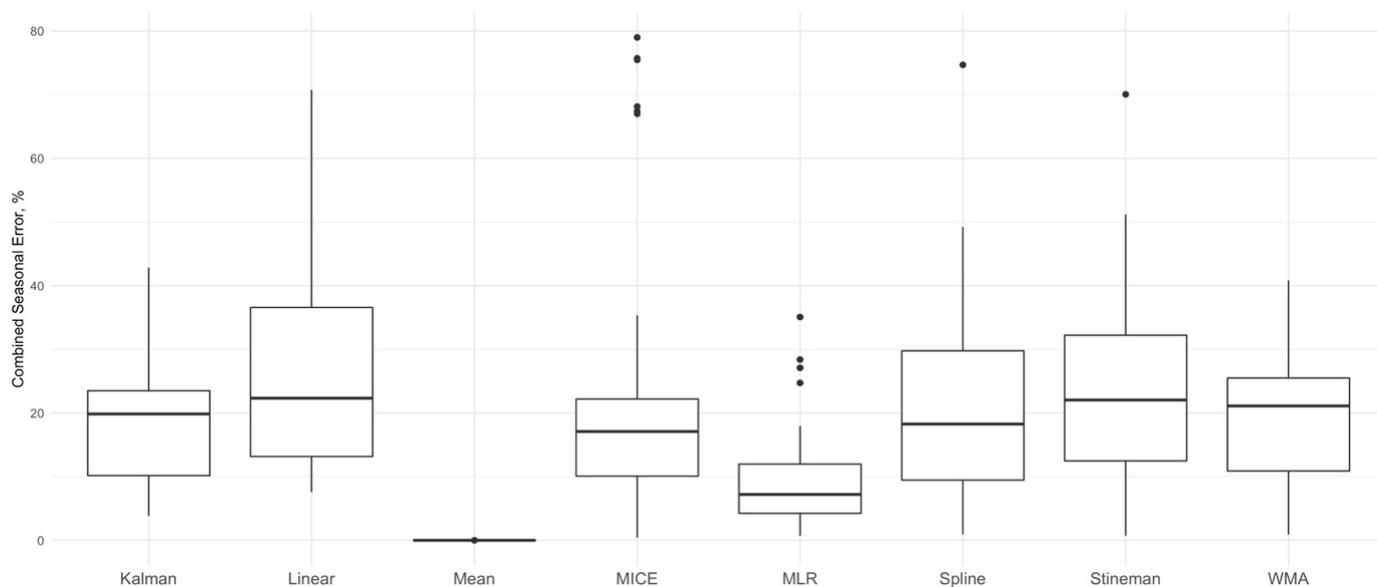


Figure 21. Box and whisker plots of combined seasonal errors calculated from reconstructed time series pooled across sites and gap filters using each imputation model. Box and whisker plots are composed of the median (solid line), lower and upper quartiles (box), the minimum and maximum values beyond the 25th and 75th quantile but < 1.5 interquartile range (whiskers) and values > 1.5 interquartile range (dots).