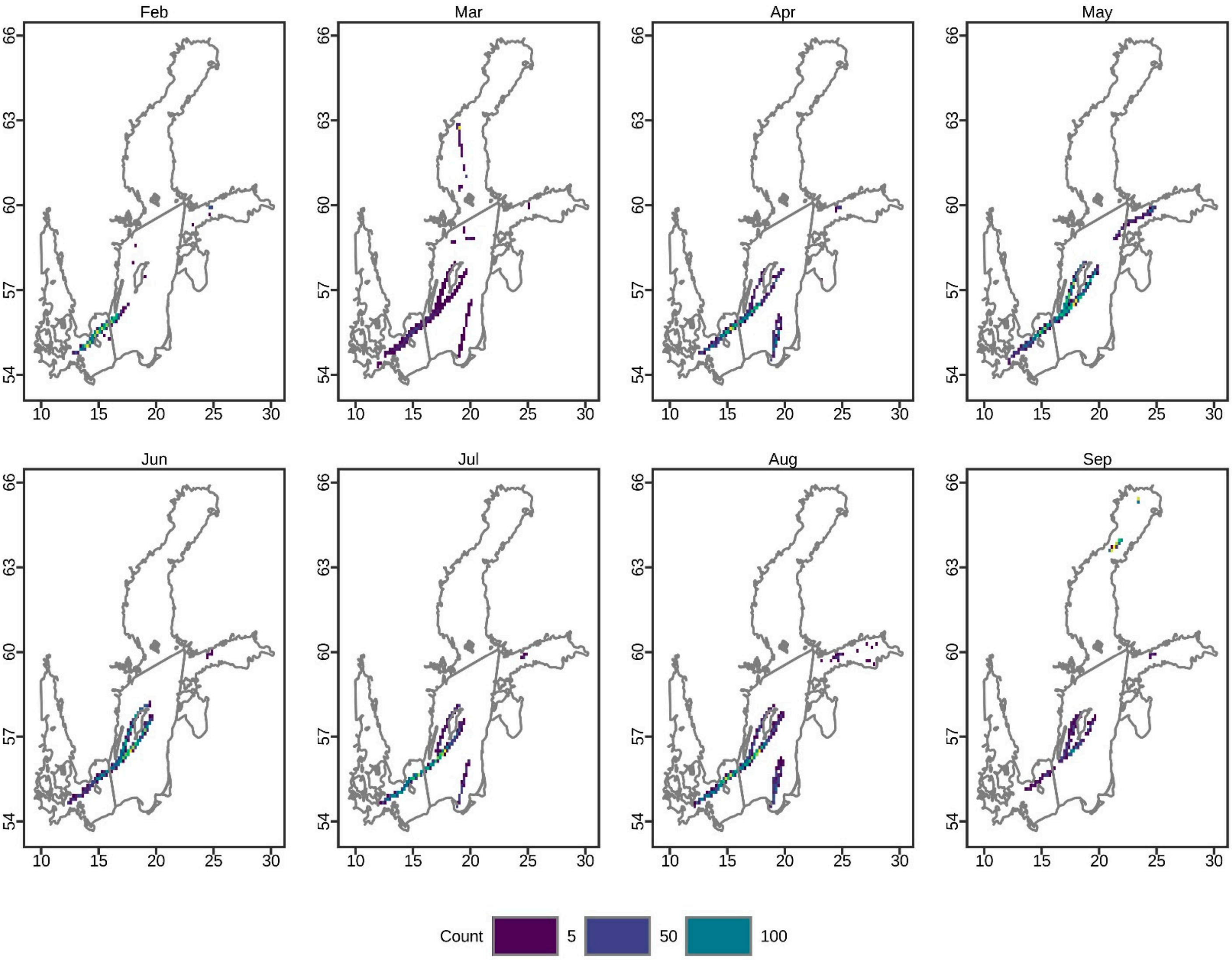
Supplementary to "Remote sensing supported sea surface pCO_2 estimation and variable analysis in the Baltic Sea"

1. Spatiotemporal distribution of the in-situ data used in this study

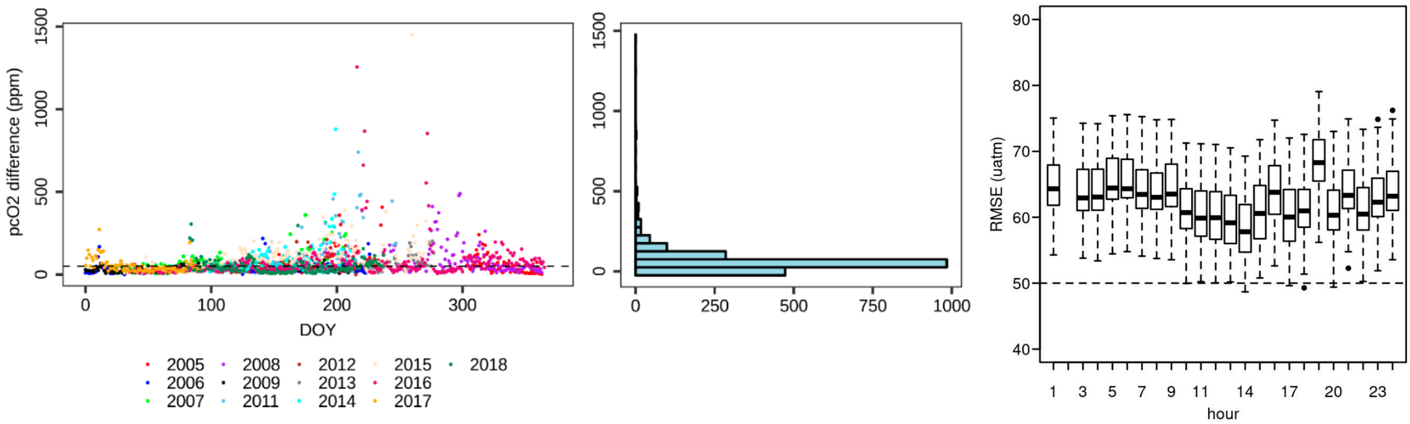


S1. Spatial and temporal distributions of the in-situ data used for training and validating the pCO_2 estimate.

2. Diurnal effect

We analyzed the pCO_2 measurements from the static Östergarnsholm site to see the diurnal difference of sea surface pCO_2. The results showed that the diurnal difference of pCO_2 at the Östergarnsholm were up to 1200 µatm (Scheme.1 left) and mainly between 30 and 60 µatm (Scheme.1 middle).

The pCO_2 estimated with models trained only with pCO_2 measurements from vessel was validated with the pCO_2 measurements at Östergarnsholm site. 50 random forest models were trained with in-situ data from the months randomly selected. The RMSEs of the models showed remarkable diurnal cycle, and the minima were between 9:00 and 14:00 o'clock. It proved the necessity to narrow down the time period of in-situ data to match the remote sensing images.

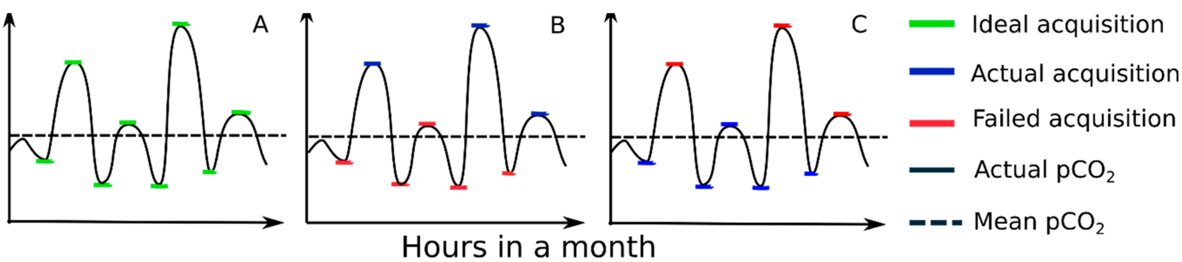


Scheme.1 Diurnal effect on the pCO_2 estimate. Left: Diurnal differences of the pCO_2 at the Östergarnsholm site; middle: the histogram of pCO_2 diurnal difference; Right: RMSEs of random forest models when valdiated with in situ pCO2 measurents at the Östergarnsholm site. Dashed lines in the subfigures are at 50 μatm.

3. Upwelling

3.1. Conceptual mechanism of upwelling effect on monthly pCO_2 estimate

Ideally, the monthly mean pCO_2 derived from monthly variables reveals actual situation if the remote sensing images in a month captured both during upwelling and beyond upwelling (**Error! Reference source not found.** A). However, if the remote sensing images in a month are dominated by these capturing upwelling, then using the oceanic variables derived from those remote sensing images leads to overestimated pCO_2 (**Error! Reference source not found.** B). If the remote sensing images in a month are dominated by these mapping the sea beyond the time of upwelling, then using the oceanic variables derived from those remote sensing images would produce underestimated pCO_2 (**Error! Reference source not found.**C).



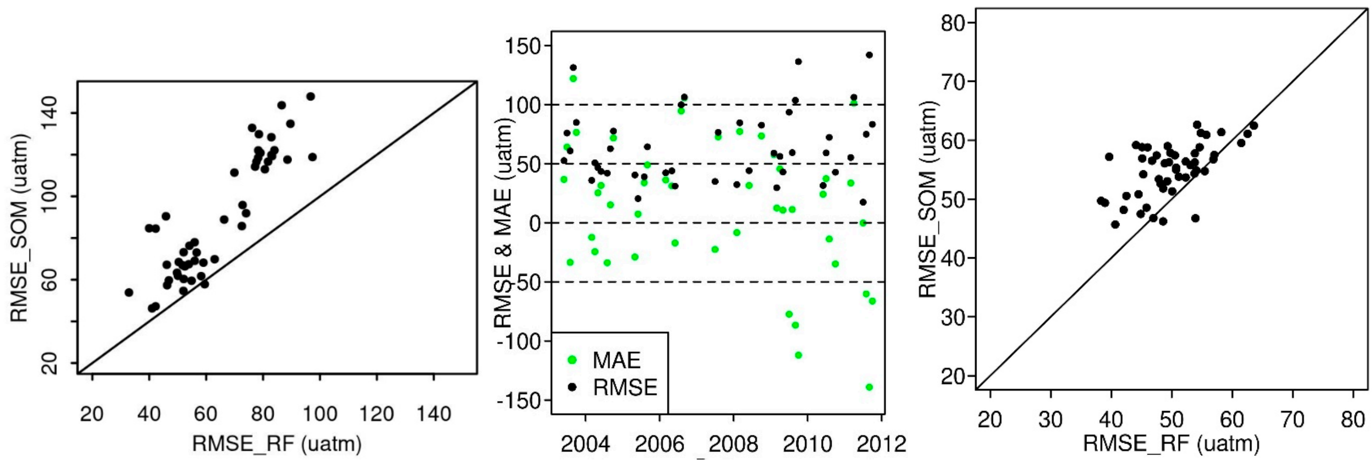
Scheme 3. Scenarios where the upwelling affects the pCO_2 estimate from remote sensing images. A: ideal situation without misestimation; B: upwelling causes overestimated pCO_2; C: upwelling causes underestimated pCO_2.

3.2. Eliminating the upwelling effect in the pCO$_2$ estimate

The training data were the in-situ pCO$_2$ from 2/3 months of random selection, the complementary of training data served as validation data. Random forest and SOM models were constructed 50 times with identical training data. The RMSE of the models were distributed in a large range, namely, random forest models were in 20-100 µatm and SOM models were in 40-140 µatm (**Error! Reference source not found.** left). Upwelling affected the pCO$_2$ estimate in the Baltic Sea regardless of the methods.

Subsequently, random forest model was separately trained with in-situ data from each of the months and validated with the rest. The RMSE and mean absolute error (MAE) of the models showed that upwelling can cause both overestimate and underestimate in pCO$_2$ (**Error! Reference source not found.** middle). The months with significant upwelling effect were 2003-09, 2006-09, 2006-08, 2009-07, 2009-09, 2009-10, 2011-04, 2011-08, 2011-09, 2011-10. Nearly all in autumn time when the upwelling prevailing in the Baltic Sea.

After the upwelling affect image were removed, 50 random forest and SOM models were again constructed, with the same training data selection as in S.3 left. The result showed that the ranges of the RMSEs of the both random forest and SOM models were narrowed down to 40-70 µatm (**Error! Reference source not found.** right).



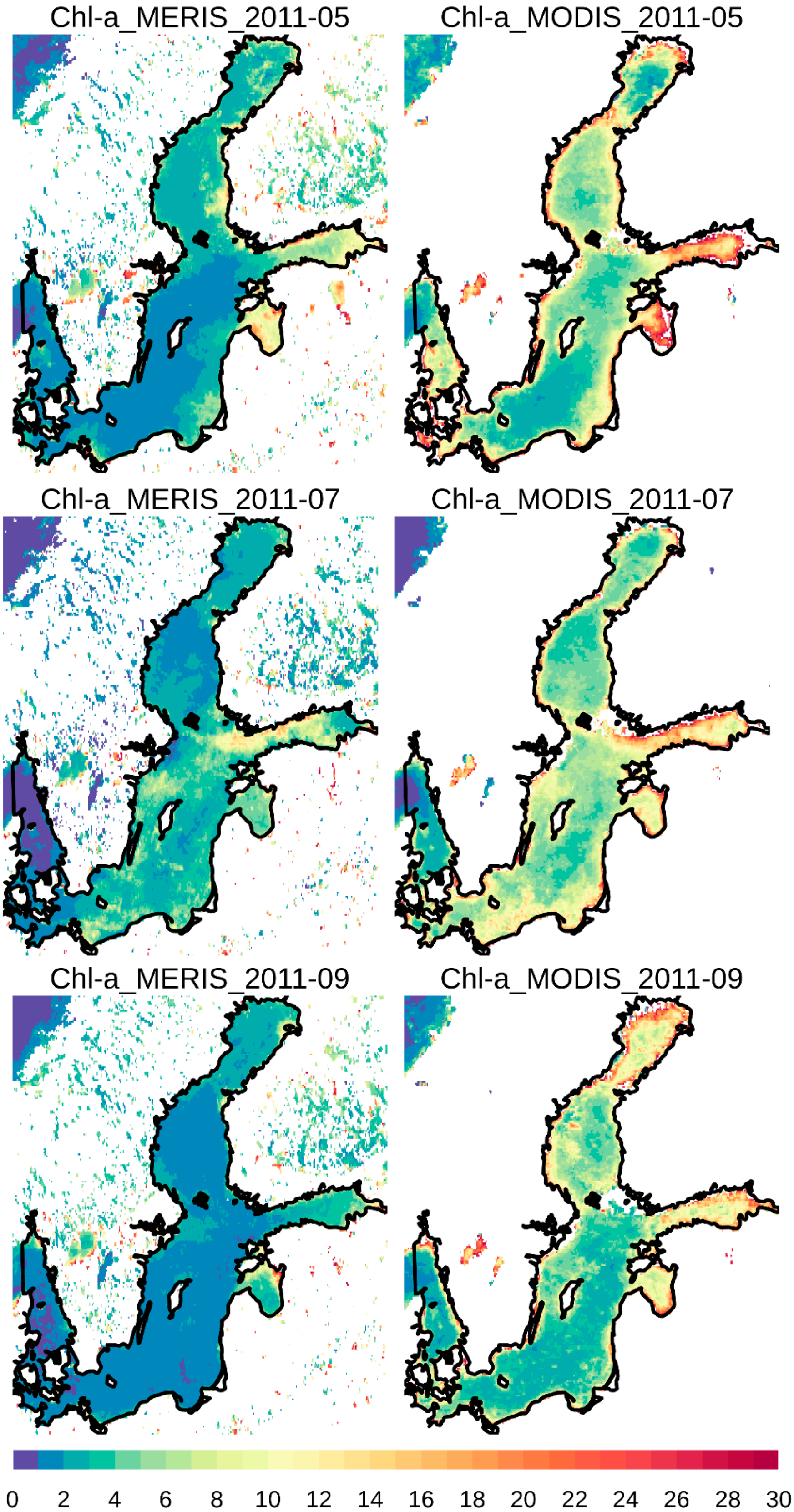
Scheme 4. The effect of upwelling in the pCO$_2$ estimate with remote sensing image. Left: RMSE random forest and SOM from the same training and validation data, with upwelling affected months included; Middle: the RMSE and MAE of all the images as served as validation data; Right: RMSE and MAE from random forest and SOM from the same training and validation data, restively, with upwelling affected months excluded.

4. Chl-a and a$_{CDOM}$ from MODIS and MERIS

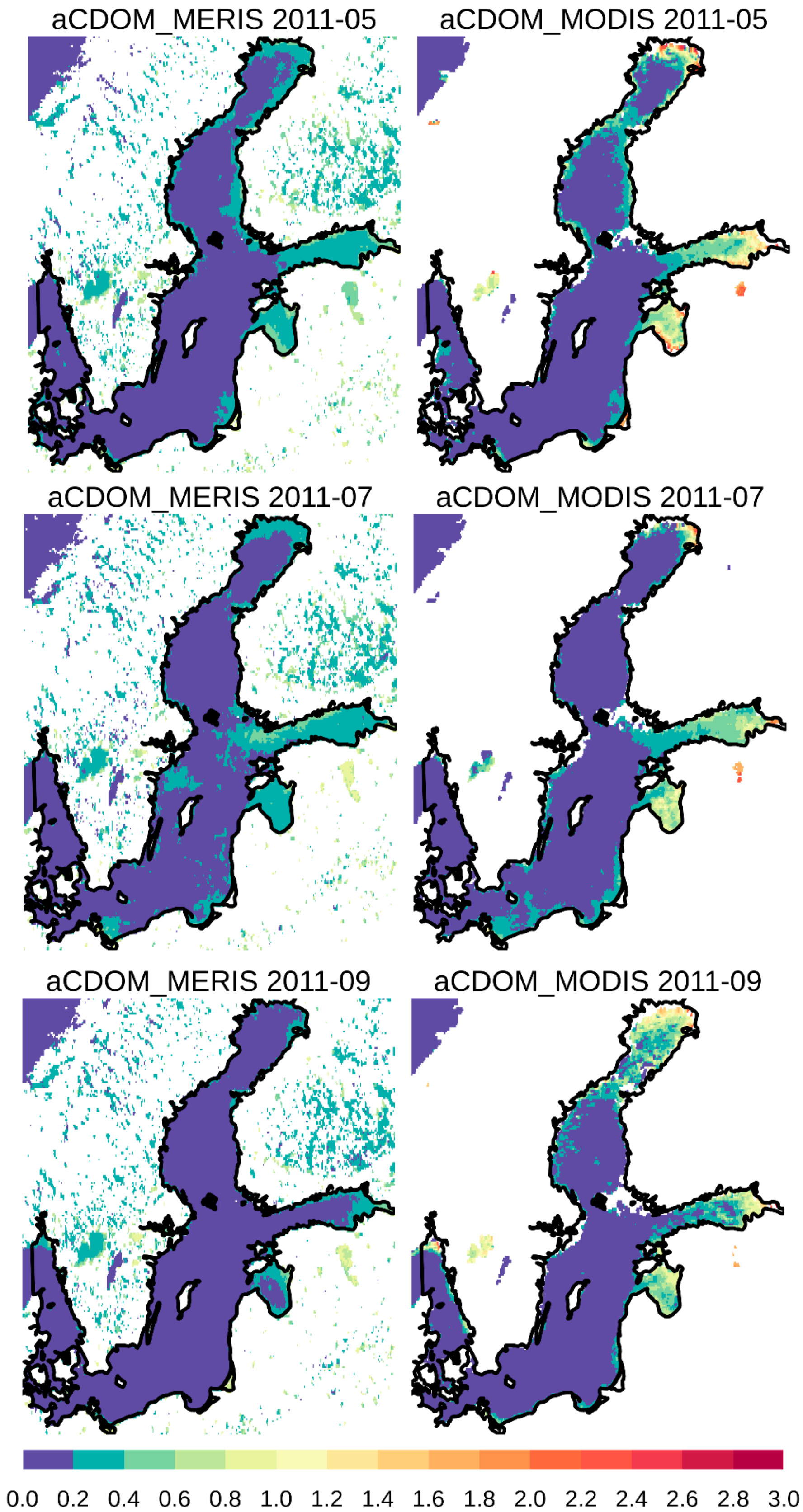
4.1. Visual comparison

The two Chl-a maps showed similar Chl-a concentration pattern in the Baltic sea in May, July and September 2011, the typical time before, during and after intensive biological activities in this sea. The Chl-a concentrations derived from MODIS images were in a larger range (0-30 ug/L) than Chl-a concentrations derived from MERIS (0-10 ug/L), particularly in the gulf and coastal areas (Scheme 2).

a$_{CDOM}$ from MERIS and MODIS showed the similar spatial patterns to that of the Chl-a concentration (**Error! Reference source not found.**), but the latter showed higher CDOM in the deep gulfs.



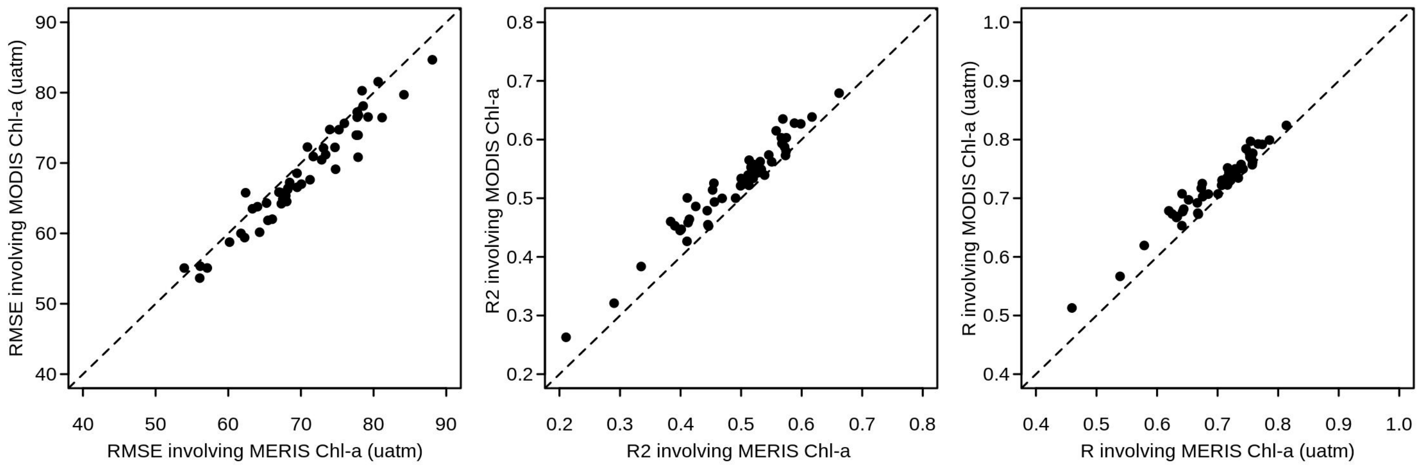
Scheme 2. The monthly mean product of Chl-a derived from MODIS and MERIS images in May, July and September 2011 mapping the Baltic Sea.



Scheme 6. aCDOM from MODIS and MERIS in the Baltic Sea.

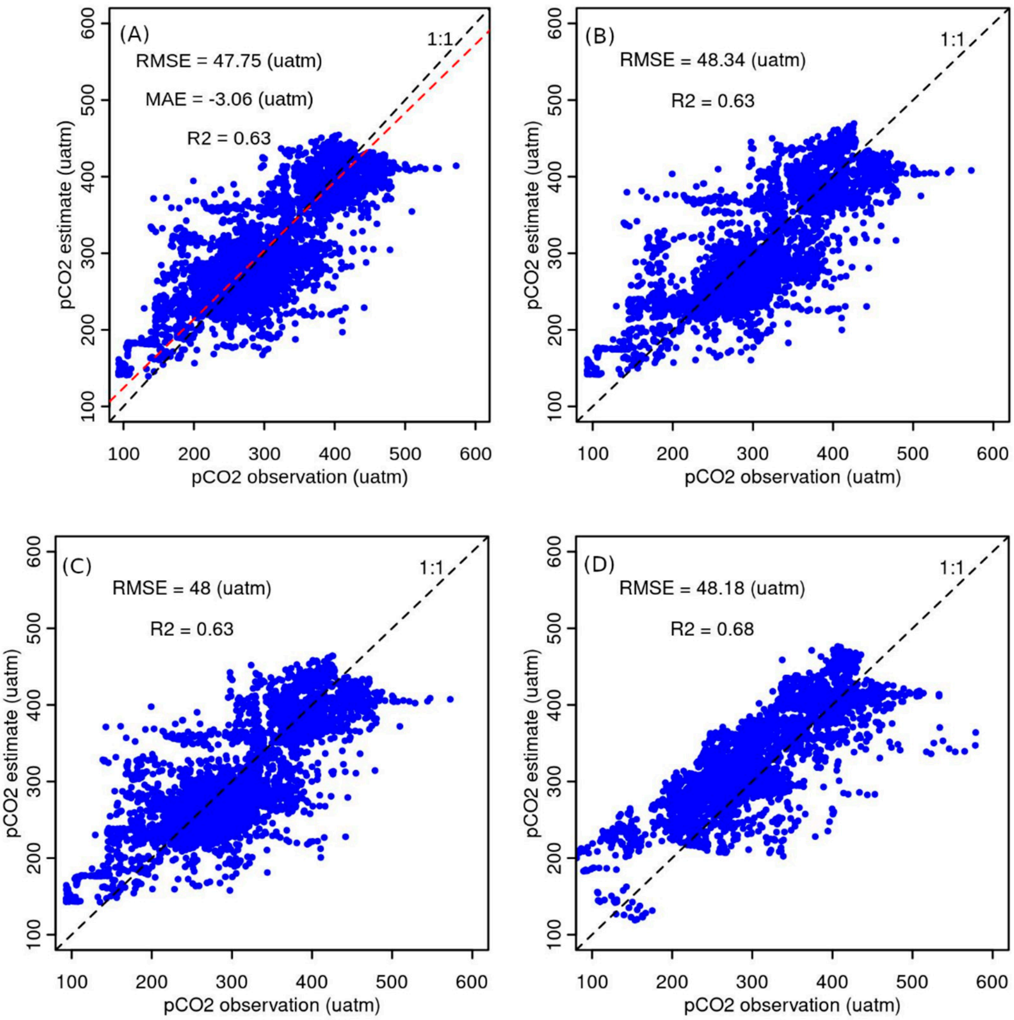
4.2. Comparison of MERIS Chl-a and MODIS Chl-a in sea surface pCO_2 estimation

Random forest models trained with the respective participation of MERIS Chl-a and MODIS Chl-a in the identical combination of other input variables. The result showed that using MODIS Chl-a or MERIS Chl-a would produce similar sea surface pCO_2 estimation in the Baltic Sea.



Scheme 7. The performance differences between of Chl-a from MODIS and Chl-a from MERIS in the pCO_2 estimate; The dashed lines are the 1:1 line. Left): RMSE, Middle) the coefficient of determination (R^2); right) Correlation between the estimated and observed pCO_2.

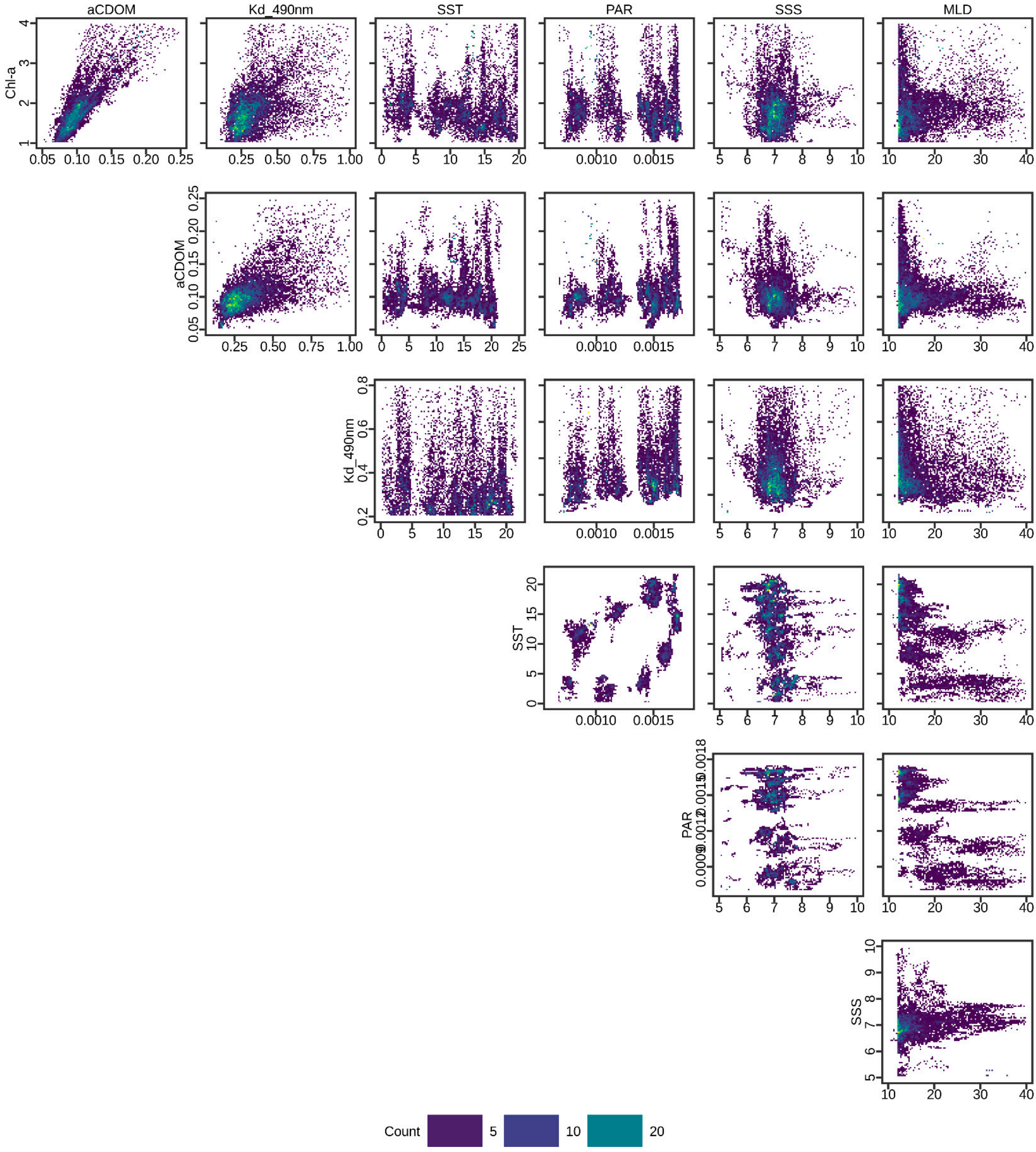
5. Alternative of the final model for pCO2 estimate in the entire Baltic Sea



Scheme 8. Alternative of the final model for pCO2 estimate in the entire Baltic Sea. A): Performance of the model Table 2. measurements in odd months of even years (e.g. March 2002) and even months of odd years (e.g. April 2003) and validated with the remaining in-situ data; B): Performance of the model using the exchanged training and validation data in the mode in subfigure A; C): Performance of the model trained with the in-situ pCO_2 measurements in odd months of odd years(e.g. May 2003) and even months of even years (e.g. April 2002) and validated with the remaining in-situ data; D): Performance of the model exchanged training and validation data in the mode in subfigure C.

6. Relationships between the co-located variables

6.1. Relation between variables co-located to the in-situ pCO2 measurements in the Baltic Sea



Scheme 9. Relationship between variables in the Baltic Sea.