A Geographic Weighted Regression Approach for Improved Total Alkalinity Estimates in the Northern Gulf of Mexico

Padmanava Dash*, Madhur Devkota, Andrew E. Mercer, and Shrinidhi Ambinakudige

Department of Geosciences, Mississippi State University, Mississippi State, MS 39762, USA

*Corresponding author: Padmanava Dash (pd175@msstate.edu)

Highlights:

- Algorithms were developed to estimate total alkalinity in northern Gulf of Mexico
- Use of chlorophyll *a* addressed the biological and chemical complexities
- Geographically weighted regression produced the best estimates

Abstract

Total alkalinity (TA) is one of the important parameters to show the intensity of seawater buffer against ocean acidification. TA dynamics in the northern Gulf of Mexico (N-GoM) is significantly affected by the Mississippi River. An empirical TA algorithm is offered here which accounts for the local effects of coastal processes. *In situ* data collected during numerous research cruises in the N-GoM were compiled and used to develop TA algorithms using sea surface temperature (SST) and sea surface salinity (SSS) as explanatory variables. After improving the coefficients and functional form of this algorithm, chlorophyll *a* (Chl-*a*) was included as an additional explanatory variable, which worked as a proxy for addressing the pronounced effects of biological forcing on coastal waters. Finally, a geographically weighted regression algorithm was developed in the form $TA=\exp[Xo+X_1(SSS-35)^2+X_2(SSSxSST)^{1/2}+X_3 chl\text{-}a]$ to address spatial non-stationarity, which produced improved estimates of TA in the N-GoM.

Keywords: Total Alkalinity; Northern Gulf of Mexico; Ocean Acidification; Geographic Weighted Regression

Introduction

About 25% of the anthropologically generated atmospheric CO_2 is absorbed by the world oceans [*Quere et al,* 2016]. This has resulted in an increase in the concentration of dissolved inorganic carbon and decrease in average ocean pH, also known as ocean acidification (OA) [*Sun et al.,* 2012]. A gradual decrease in carbonate ion concentrations, aragonite saturation states (Ω_{ar}), and calcite saturation states (Ω_{ca}) leads to reduced calcification rates and ultimately calcium carbonate dissolution [*Cooley,* 2009; *Gledhill et al.,* 2009], a key concerning aspect of ocean acidification. Reduced calcification rates negatively impact the calcifying organisms in the ocean and also have important implications for global carbon cycle [*Raven et al.,* 2005].

Thirty-five years of observation concluded that the rate of alteration in seawater carbonate system parameters varies geographically with the greatest variability observed in the North Pacific Ocean region [*Takahashi et al.,* 2006]. This variability likely results from local oceanographic processes which contribute additional carbon into the system by lateral mixing of waters from marginal seas, input from coastal areas and rivers, upwelling of subsurface waters, and biological activities [*Takahashi et al.,* 2006]. Similarly, there are a myriad of factors that affect coastal acidification in the northern Gulf of Mexico (N-GoM), primarily owing to heterotrophic respiration of terrestrial and autotrophically produced organic carbon [*Cai,* 2003; *Hu et al.,* 2017] mainly due to nutrient loading by the Mississippi River as well as outflow from other rivers [*Cai et al.,* 2011; *Lohrenz & Cai,* 2006]. Hence, the variability in the carbonate system parameters in the N-GoM coastal waters is notably higher than the pelagic central Gulf of Mexico due to pronounced coastal biogeochemical cycling.

The partial pressure of CO_2 (pCO_2), total alkalinity (TA), pH, and dissolved inorganic carbon (DIC) are the principal carbonate chemistry parameters used to document the status

and trends of ocean acidification. The carbonate system is bivariate and can be fully constrained if any two of the four measurable carbonate system parameters are known. TA coupled with pCO_2 are frequently used (*e.g.*, NOAA's Ocean Acidification Product Suite [*Gledhill et al.,* 2008]) to estimate pH, Ω_{ar}, and Ω_{ca} using the CO2SYS Program [*Lewis & Wallace,* 1998] employing the carbonate equilibrium equations [*e.g.*, *Mehrbach et al.,* 1973]. Thus, TA is one of the important parameters that aids in deriving other carbonate system parameters [Bever et al, 2021]. TA is one of the important parameters to show the intensity of seawater buffer against ocean acidification and is defined (Dickson, 1981) as the number of moles of hydrogen ion equivalent to the excess of proton acceptors in 1 kilogram of sample ($\mu mol\ kg^{-1}$). Due to the prevalence of carbonate rocks in the drainage basin of the Mississippi River, the TA values of the Mississippi River water are usually high (Raymond and Cole, 2003), which has increased significantly in the past half century as a result of increased flow from higher rainfall and land-use changes that have altered vegetation cover (Keul at al., 2010). Keul et al. (2010) observed a non-conservative behavior of TA in the Mississippi River plume region, which they attributed to changes in microbial recycling across salinity gradients. Additionally, non-conservative mixing with waters from local estuaries, such as Terrebonne Bay and smaller inputs from other estuaries decrease the TA values in the coastal waters. This decrease is considerably less due to the high alkalinity content of the Mississippi River and other rivers draining to the Gulf of Mexico (Keul et al., 2010).

Millero et al. [1998] developed a set of algorithms for estimating salinity-normalized TA in global oceans using historical TA measurements and TA measured during global carbon surveys during 1990s. *Lee et al.* [2006] updated those algorithms with five polynomial algorithms for estimating TA in five major oceanographic regimes using Sea Surface Salinity (SSS) and Sea Surface Temperature (SST) as explanatory variables. These oceanographic regimes include the Subtropics, the Equatorial upwelling Pacific, the North Atlantic, the

79 North Pacific, and the Southern Ocean. The N-GoM region is close to zone 1 (i.e. the
80 Subtropics) as described by *Lee et al.* [2006]. However, the calibration dataset of the TA
81 algorithm for zone 1 lacked *in situ* data from the N-GoM region, which likely explains the
82 bias of the algorithm for the N-GoM region.

83 The Mississippi River is the largest river influencing N-GoM, with an average
84 discharge of about $1.35 \pm 0.2 \times 10^4$ m^3/s based on 64 years of United States Geological
85 Survey discharge data [*Hu et al.,* 2005]. At the Mississippi River outfall and shelf area, high
86 net pCO_2 is evidenced due to heterotrophic production of CO_2 from land-derived organic
87 carbon [*Cai,* 2003]. It is further influenced by autotrophically produced organic carbon
88 caused by high nutrient loading leading to CO_2 drawdown in shelf surface water [*Ternon et*
89 *al.,* 2000; Huang et al., 2015] and heterotrophy of autotrophically produced carbon [*Cai &*
90 *Lohrenz,* 2006], which acts as an additional pCO_2 source [*Lohrenz & Cai,* 2006]. *Guo et al.*
91 [2012] reported that the net autotrophic production of carbon at the Mississippi River plume
92 is even higher than the Amazon system. The nutrient-rich Mississippi River water coupled
93 with outflow from other rivers influences carbon dynamics of the N-GoM coastal ecosystem
94 [*Lohrenz & Cai,* 2006], causes bottom water hypoxia in the region [*Rabalais et al.,* 2002],
95 and impacts the marine ecosystem, affecting fisheries and coral reefs in the N-GoM [*Hu et*
96 *al.,* 2003; *Salisbury et al.,* 2001; *Walker,* 1996; *Inoue et al.*, 2008]. Thus, the N-GoM coastal
97 region is biologically and chemically complex and is significantly different from the rest of
98 the subtropical ocean. In past research, chlorophyll *a* (chl-*a*) has been used as a proxy for
99 primary production in estimating pCO_2 [*Lohrenz & Cai,* 2006]. Thus, the incorporation of
100 chl-*a* as an explanatory variable has great potential to increase TA algorithm performance.

101 It is difficult to account for the high degree of spatial and temporal variability in
102 carbonate system parameters in the coastal waters using an algorithm that was developed for
103 the open ocean due to the biological and chemical complexities of the coastal waters, as well

as the variable freshwater influence. Spatial variability in the relationships between parameters is referred to as spatial nonstationarity [*Brunsdon et al.,* 1996]. Geographically weighted regression (GWR) is a statistical modeling technique that attempts to address the spatial nonstationarity [*Brunsdon et al.,* 2010], the spatial variation of complexities in ocean chemistry and acidification processes in coastal waters. The GWR approach generates a set of coefficients for each point in space empirically using the provided functional form for the regression equations. In the case of estimating carbonate system parameters, the GWR approach has great potential in accounting for the spatial variability of the influence of biogeochemical processes on the explanatory variables.

The objective of this effort was to develop a revised empirical algorithm for estimating TA in the coastal waters of the N-GoM by proposing an improved functional form incorporating chl-*a* as an additional explanatory variable, and utilizing GWR to address spatial nonstationarity in the N-GoM domain.

Materials and Methods

In situ TA, SST, and SSS data, and satellite-derived chl-*a* were used for the development and testing of the algorithms in this study. The algorithms were later applied to satellite derived SST, SSS, and Chl-*a* imagery to illustrate the efficacy of the algorithms in showing the spatial distribution of TA using the derived algorithm. A total of 247 observations of TA, SST, and SSS each were assimilated from the *in situ* data collected during numerous research cruises from July 2007 to September 2013 in the N-GoM (Table 1; Fig. 1). The areal extent of these observations is about 474,219 sq km rendering an average of one observation per 41 km x 41 km area. *In situ* SSS and SST values were used to estimate TA using each of the five algorithms given by *Lee et al.* [2006] and estimated TA values were compared with *in situ* measured TA values. As the N-GoM is closer to the zone 1, it was expected that the TA estimation error would be lowest with this algorithm. However, the

lowest error was found when the algorithm that was developed for zone 4 was used, with zone 1's algorithm performing second best (Table 2). While error was smallest with the zone 4 algorithm, the zone 1 algorithm was selected due to its geographic proximity to the study region and its simpler functional form than the zone 4 algorithm. The zone 4 algorithm incorporates longitude as an exploratory variable, which would be redundant in the presented GWR algorithm. The zone 1 algorithm coefficients were modified to investigate potential improvement in TA estimations in the N-GoM. This was done by maintaining the functional form of the algorithm but updating the algorithm coefficients by cross-validation.

Next, the functional form of the algorithm was modified to lower the standard errors of the coefficients in the algorithm by introducing Chl-*a* as an additional explanatory variable. Chl-*a* data were derived from a time-series of MODIS satellite imagery, processed using the SeaDAS software package v7.4 with the default atmospheric correction. Out of the 247 locations with TA observations, MODIS derived chl-*a* were available for only 78 locations (Fig. 1) due to cloud cover and atmospheric correction failure at certain coastal locations. This reduction in observations led to a study area reduction to 359,397 sq km with an average of one observation per 67 km x 67 km area.

Using the data from these 78 locations, the TA algorithm was modified by adding a linear term for chl-*a*, along with the modified functional forms for SSS and SST. Finally, geographically weighted regression (GWR) was carried out with SSS, SST, and chl-*a* as explanatory variables and TA as the dependent variable. While the 78 data points constituted a limited dataset, the distribution of observations was still relatively uniform. Due to the paucity of data, GWR was chosen to improve the algorithm since GWR allows coefficients to vary continuously over the study area, and a set of coefficients are estimated for each location – typically on a grid using a weighted approach so that the coefficient rasters account for the spatial heterogeneity. The open source GWR4 Semiparametric GWR/GWGL Modeling Tool

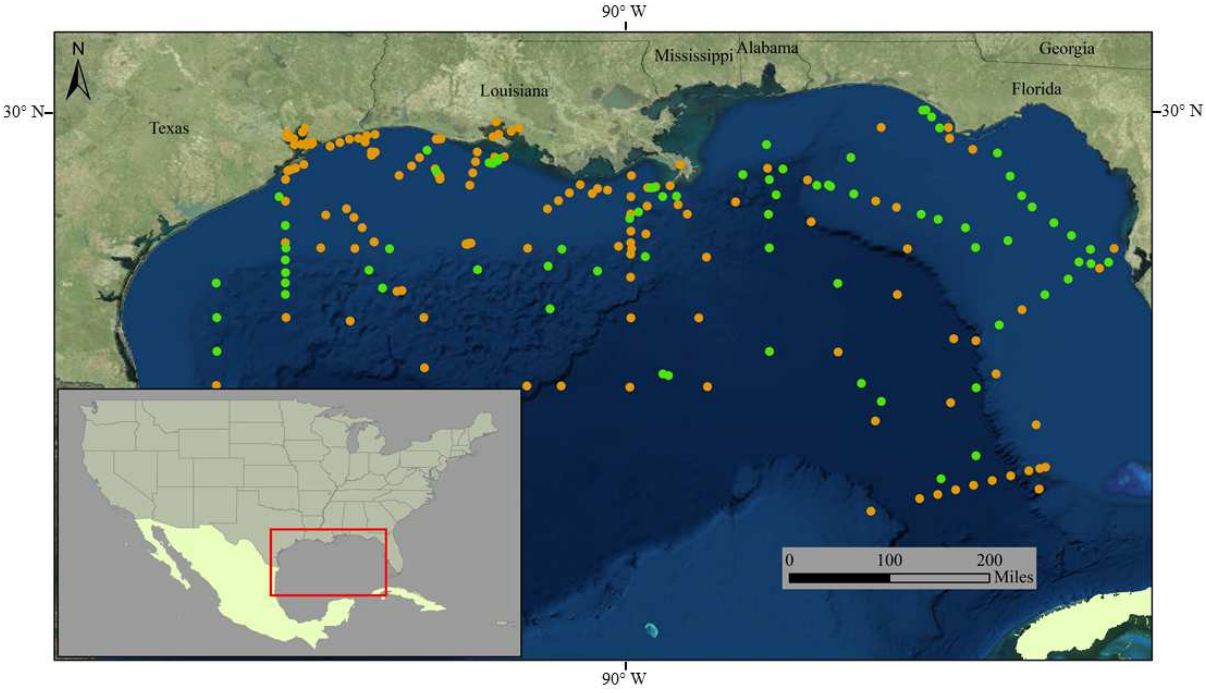


Fig. 1. *In situ* observations of Total Alkalinity (TA), Sea Surface Salinity (SSS), Sea Surface Temperature (SST) and chlorophyll *a* (chl-*a*) used in this study. The green points represent the availability of chl-*a* in addition to TA, SSS, and SST, while orange points represent the availability of TA, SSS, and SST.

(version 4.09; 2016) developed by the National Centre for Geo-computation, Ireland and Ritsumeikan University, Japan, was used to carry out GWR with an Adaptive Gaussian Kernel (AGK) with cross-validation as the selection criteria [*Brunsdon et al.,* 1996]. During the GWR algorithm development, weights for each data point were computed based on the relative distance between the point under consideration and the rest of the observations using AGK [*Charlton and Fotheringham,* 2009]. The rate of decrease in the weights was calculated using the AGK. This kernel function, unlike Fixed Kernel, uses dynamic bandwidth distance where the bandwidth-distance is automatically chosen with a smaller dynamic bandwidth distance for the regions with a higher number of observations and higher dynamic bandwidth distance for the regions with a lower number of observations. The AGK was chosen because of non-uniform spatial distribution density of observations in the dataset.

All these algorithms were developed using all available *in situ* data and validated using a 10-fold cross-validation approach. For validation, the dataset was randomly divided

171 **Table 1.** Datasets used in this study

Cruise Name	Vessel Name	Data Collection Date	Organization
RB-07-05	GOMECC 1 - R/V Ronald H. Brown	July 2007	NOAA-AOML
Gordon Gunter - GU0802	R/V Gordon Gunter	May 2008	NOAA-AOML
NODC Accession 0081040*	Pacific Celebes	December 2007; April-May 2008; February, March, July 2009	NOAA
Tara Oceans Expedition	Tara Oceans Consortium**	August 2009	PANGAEA
Lophelia II – Cruise 3	R/V Ronald Brown	August-September 2009	NOAA
NF10-13	NOAA ship Nancy Foster	July 2010	NOAA
NCEI Accession 0144622[#]	Weatherbird II	May 2011	NOAA-NCEI
NCEI Accession 01574614[##]	PELICAN	September 2013	NOAA

172 * NODC Accession 0081040, Hydes et. al. (2011);

173 ** Tara Oceans Consortium; PANGAEA, https://doi.org/10.1594/PANGAEA.839230, Picheral et al. (2014);

174 [#] National Centers for Environmental information (NCEI) Accession 0144622, Barrera et al. (2016);

175 [##] NCEI Accession 0157461, Salisbury et al. (2016)

176 into ten subgroups. An algorithm was developed using the data from the nine subgroups and

177 validated using the 10th subgroup. This was repeated nine times so that each subgroup could

178 be used for validation and in the calculation of the residuals. Subsequently, mean absolute

179 error (MAE) and root mean square error (RMSE) were calculated for the estimated TA and

180 **Table 2.** Deviations in total alkalinity (μmol kg^{-1}) observed when Lee et al. (2006)

181 algorithms were applied to the Gulf of Mexico dataset

Parameter	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
RMSD*	272.6	481.7	308.1	243.7	320.6
%RMSD	11.4	20.1	12.8	10.2	13.4
MAD[#]	183.9	339.4	176.5	166.9	164.4
%MAD	7.7	14.2	7.4	7.0	6.9

182 * RMSD = Root Mean Square Deviation
183 [#] MAD = Mean Absolute Deviation

184 the median RMSEs and MAEs were compared between the algorithms. This approach
185 allowed the use of all the data for algorithm development and validation. Finally, TA images
186 were generated by using the developed algorithms with SSS from the Soil Moisture Active
187 Passive (SMAP) sensor, and SST and chl-*a* from the Aqua MODIS sensor.

188 **Results**

189 A set of algorithms are given by *Lee et al.* [2006] for estimating global distribution of
190 surface alkalinity using SSS and SST as explanatory variables. For the purpose of evaluating
191 the existing algorithm performance in N-GoM, the algorithm developed for zone 1 was tested
192 to estimate TA using *in situ* SSS and SST data. The *Lee et al.* [2006] algorithm for zone 1 is
193 given by:

194 $$TA = 2305 + 58.66\,(SSS-35) + 2.32\,(SSS-35)^2 - 1.41\,(SST-20) + 0.04\,(SST-20)^2 \quad (1)$$

195 Estimated TA values were compared with *in situ* measured TA for quantifying the
196 algorithm accuracy within the N-GoM domain. This algorithm produced a median RMSE of
197 272.6 μmol kg^{-1} with a %RMSE of 11.4% and a MAE of 183.9 μmol kg^{-1} with a %MAE of
198 7.7% (Tables 1 & 3). The distribution of the residuals as percent error (calculated as [*in situ*
199 TA–estimated TA)/*in situ* TA] x 100) illustrates that the TA values were mostly
200 underestimated by the *Lee et al.* [2006] zone 1 algorithm (Fig. 2a & 2b). Fig. 3a shows the
201 residual percentages on a map, revealing the highest TA estimate deviations (range: 10-80%)
202 were primarily located in coastal waters. In the central GoM, the deviation of estimated
203 values was less (< 2%) from the *in situ* measured values. Hence, the *Lee et al.* [2006]
204 algorithm seems well suited for the central GoM but performs poorly near the coast. To
205 address this bias, the *Lee et al.* [2006] zone 1 algorithm was updated using the coastal *in situ*
206 data to modify the algorithm coefficients. The resulting modified algorithm was:

$$TA = 3299.394 + 7.449\,(SSS-35) - 0.083\,(SSS-35)^2 - 250.062\,(SST-20) + 15.784\,(SST-20)^2 \quad (2)$$

A comparison of this algorithm with the zone 1 algorithm of *Lee et al.* [2006] revealed that the effect of SSS on TA is different in the N-GoM region relative to the subtropical region as the magnitude of SSS coefficients were smaller in Eq. 2 compared to Eq. 1. Additionally, the effect of SST on TA was more pronounced in the N-GoM region than the subtropics, as evidenced by the larger SST coefficients in the updated algorithm. The modification produced notable improvements, as the median RMSE was 94.5 µmol kg^{-1} with a range from 76.5 to 112.4 µmol kg^{-1} for this algorithm (Table 3), which is a 65% improvement over the zone 1 algorithm provided by *Lee et al.* [2006]. Moreover, the residuals and comparison of *in situ* and estimated TA showed significant improvements (Fig. 2c & 2d). However, the coefficient for the third term $(SSS-35)^2$ had a very high standard error (error for X_2 = 91%; Table 3) suggesting high algorithm instability with this coefficient.

Initially, a regression without including the salinity parameter (with only two terms of SST) was carried out. This resulted in an algorithm with high coefficient standard errors; 30% for the first SST term and 33% for the second SST term and a lower adjusted R^2 (0.13). Similarly, a regression with SSS alone also did not yield a good algorithm. To address the issue of higher coefficient standard errors and lower adjusted R^2, several functional forms were tested after including both SSS and SST. Ultimately, an exponential equation with modification in the parameters (comprising one SSS and two SST terms) produced the best algorithm with lower RMSE, lower standard error, and higher adjusted R^2, which is given by:

$$TA = \exp\left[8.053 - 0.000156\,(SSS-35)^2 - 0.0562\,(SST-20)^2 + 0.000272\,(SST-20)^3\right] \quad (3)$$

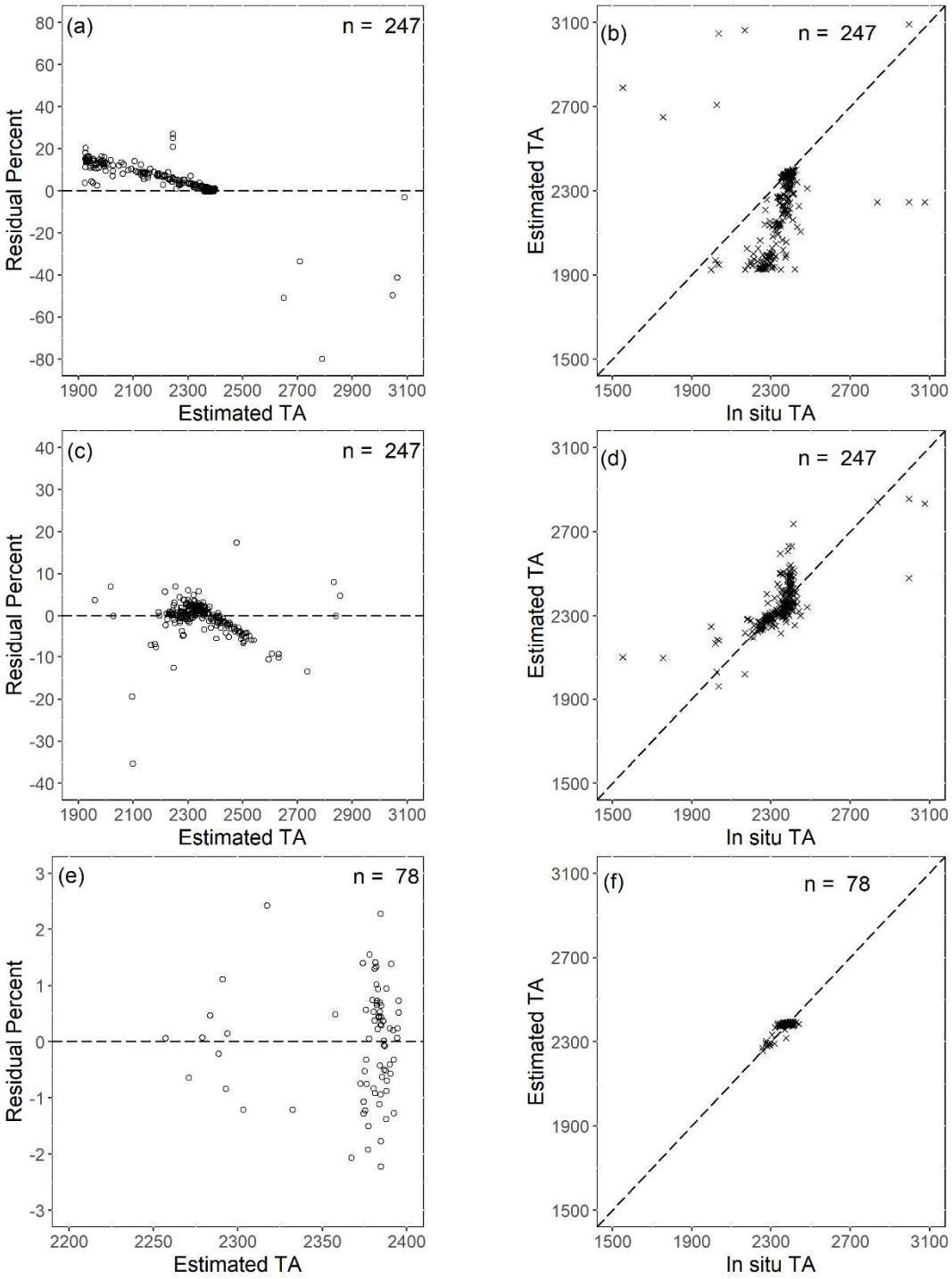
Modification in the functional form decreased the standard error for the coefficient of the second and third terms from 27% to 10% and 91% to 8%, respectively (Table 3). However, there was a decrease in R^2 and an increase in the RMSE (Table 3). As such,

chlorophyll *a* (chl-*a*) was added as an additional explanatory variable to the algorithm. The correlation between chl-*a* and TA was found to be -0.74 which indicates it may be a confounding variable that requires addressing. After adding chl-*a*, various modifications to the functional form were tested and it was found that an algorithm with SSS and SST interaction term produced the most efficient algorithm when chl-*a* was included in the algorithm. Among all the algorithms, the one with the highest R^2 (0.61), lowest RMSE (22.6 μmol kg^{-1}) and lowest standard error for the coefficients (Table 3) was chosen, which is given by:

$$TA = \exp [7.727 - 0.000385 (SSS-35)^2 + 0.00162 (SSS \times SST)^{1/2} - 0.00212 \text{ chl-}a] \quad (4)$$

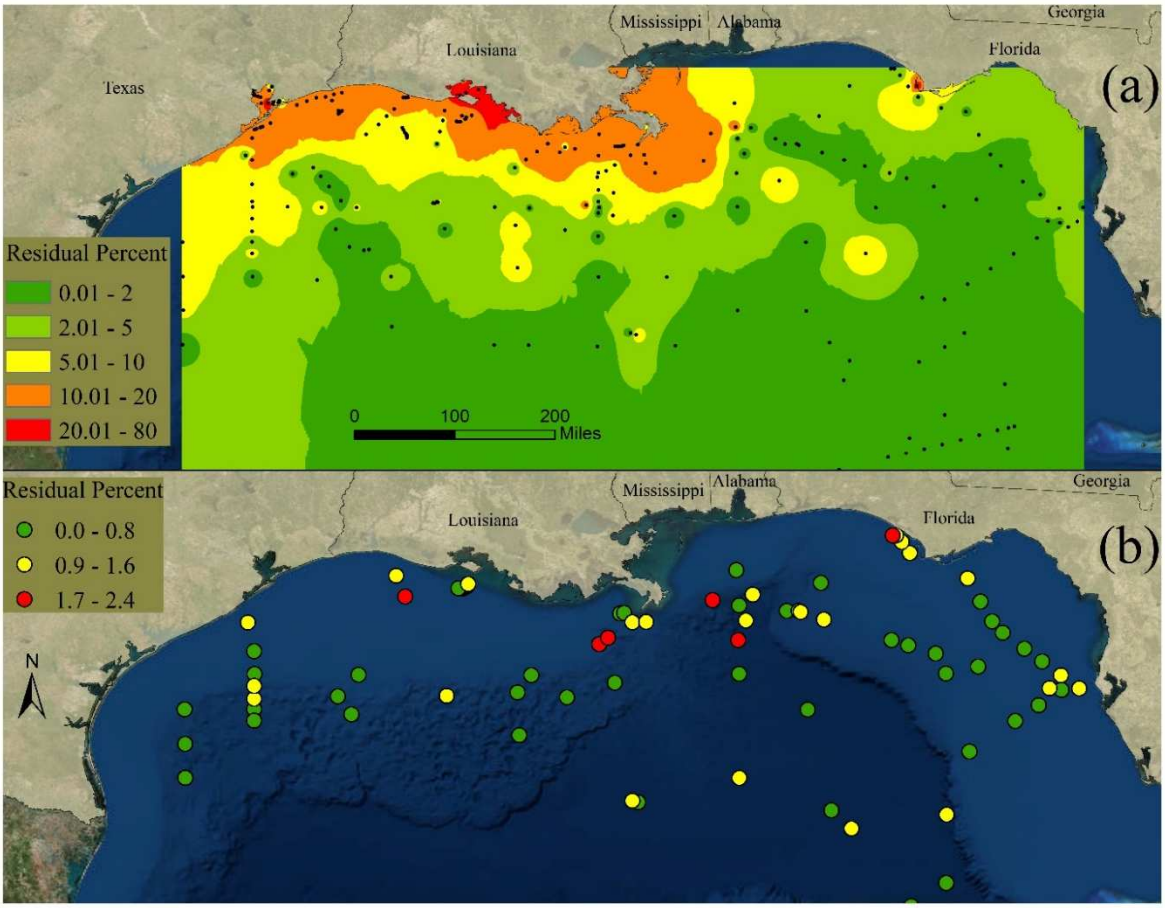
The inclusion of chl-*a* as an explanatory variable improved the algorithm considerably, however, the adjusted R^2 was still less than that of Eq. 2 (Table 3). Next, a GWR algorithm was developed using the above functional form that includes chl-*a* as the additional explanatory variable. The output from the GWR approach was the value of the coefficients at each point. However, the coefficients were noisy, so the cubic spline surface interpolation was performed to develop smooth raster surfaces of the coefficients over the spatial domain.

The raster images for the four coefficients showing the range of values in the N-GoM are shown in Fig. 4 and given as an Arc GIS map package as supplementary material. For example, a resulting raster surface for the second coefficient 'X_1' [the $(SSS-35)^2$ term) ranged between 0.000350 to 0.000410 (Fig. 4). The coefficients in coastal locations differed significantly from those in the central GoM. These differences suggest the spatial variation in the importance of the explanatory variables in TA estimations, which likely arise from local effects of biogeochemical processes that affect carbonate chemistry. An adjusted R^2 of 0.69 was obtained for the GWR algorithm (Table 3), which is the highest adjusted R^2 obtained in this study, indicating that the algorithm best fits the data. The residuals were also randomly distributed for this algorithm (Fig. 2e). Notably, most points were closer to the 1:1 line (Fig.



257

258 **Figure 2**. (a) Residuals of *Lee et al.* [2006] zone 1 algorithm (Eq. 1) as percent error, (b) comparison of *in situ*
259 and estimated total alkalinity (TA) using *Lee et al.* [2006] zone 1 algorithm (Eq. 1), (c) residuals of the TA
260 algorithm with modified coefficients (Eq. 2) as percent error, (d) comparison of *in situ* and estimated TA using
261 Eq. 2, (e) residuals of TA derived using geographically weighted regression (GWR) algorithm as percent error,
262 (f) comparison of *in situ* and estimated TA obtained using the GWR algorithm developed in this study.



263

264 **Figure 3**. Residual percentages for **(a)** *Lee et al.* [2006] zone 1 algorithm, where the residuals were classified
265 into five classes and the classes are shown using spline surface interpolation, **(b)** the geographically weighted
266 regression (GWR) algorithm developed in this study.

267 2f) for the GWR algorithm relative to that of other algorithms (Fig. 2b and 2d), suggesting
268 the GWR algorithm is in good agreement with the *in situ* TA values. Additionally, there is an
269 absence of a spatial trend in the GWR residuals (Fig. 3b) when compared to the residuals of
270 *Lee et al.* [2006] zone 1 algorithm (Fig. 3a). Clearly, the GWR approach significantly
271 reduced spatial nonstationarity in the TA estimates, deeming this approach more valid.

272 All the algorithms were applied to MODIS-derived SST and chl-*a*, and SMAP-
273 derived SSS images of February 10, 2016 to obtain TA images by multiplying the coefficient
274 rasters with the SST, SSS, and chl-a images as per the functional form of the algorithm
275 provided in Eq. 4. The higher SSS coefficient magnitudes in the *Lee et al.* [2006] zone 1

276 **Table 3**. Summary statistics of the existing and developed algorithms.

	***Lee et al.* [2006] algorithm (Eq. 1)**	**Algorithm with modified coefficients (Eq. 2)**	**Algorithm with modified functional form (Eq. 3)**	**Algorithm with chl-*a* as an additional explanatory variable in addition to SSS & SST (Eq. 4)**	**Geographic Weighted Regression algorithm**
MAE (μmol kg^{-1})	183.9	Median = 57.1 (Range = 47.4, 66.8)	Median = 65.9 (Range = 55.3, 76.4)	Median = 19.2 (Range = 17.1, 21.3)	Median = 19.3 (Range = 17, 22.1)
%MAE (%)	7.7	Median = 2.1 (Range = 1.8, 2.5)	Median = 2.7 (Range = 2.4, 2.9)	Median = 0.7 (Range = 0.6, 0.9)	Median = 0.8 (Range = 0.6, 0.8)
RMSE (μmol kg^{-1})	272.6	Median = 94.5 (Range = 76.5, 112.4)	Median = 102.1 (Range = 90.9, 113.3)	Median = 22.6 (Range = 20.1, 25)	Median = 22.7 (Range = 20.1, 24.8)
%RMSE (%)	11.4	Median = 3.5 (Range = 3.1, 3.9)	Median = 5.1 (Range = 4, 6.2)	Median = 0.89 (Range = 0.76, 1.02)	Median = 0.89 (Range = 0.76, 1.0)
Adjusted R^2	NA	0.66	0.59	0.61	0.69
Standard Error (%)	NA	X_o:1.6 X_1:27 X_2:91 X_3:7 X_4:7	X_o:0.2 X_1:10 X_2:8 X_3:10	X_o:0.4 X_1:23 X_2:59 X_3:48	X_o:0.4 X_1:22 X_2:55 X_3:42

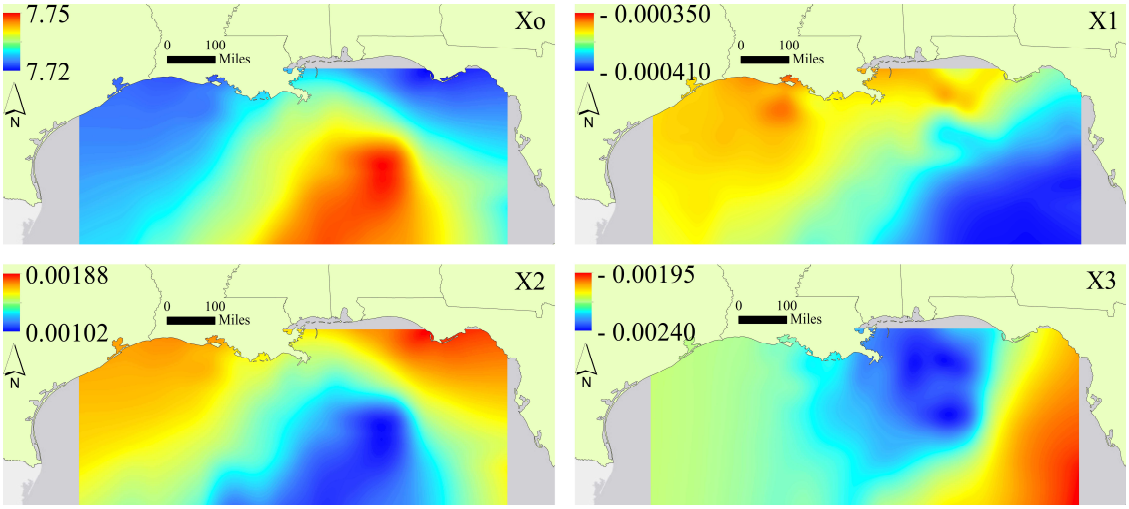


Figure 4. Range of values of the four coefficients X_0, X_1, X_2, and X_3 in the geographic weighted regression (GWR) algorithm for the northern Gulf of Mexico for the equation $TA = \exp [X_0 + X_1 (SSS-35)^2 + X_2 (SSS \times SST)^{1/2} + X_3$ chl-*a*]. These rasters are provided in the Arc GIS map package with this submission. Multiplying those rasters to a set of sea surface temperature, salinity, and chlorophyll a rasters as per the provided functional form of the algorithm will produce the corresponding total alkalinity image.

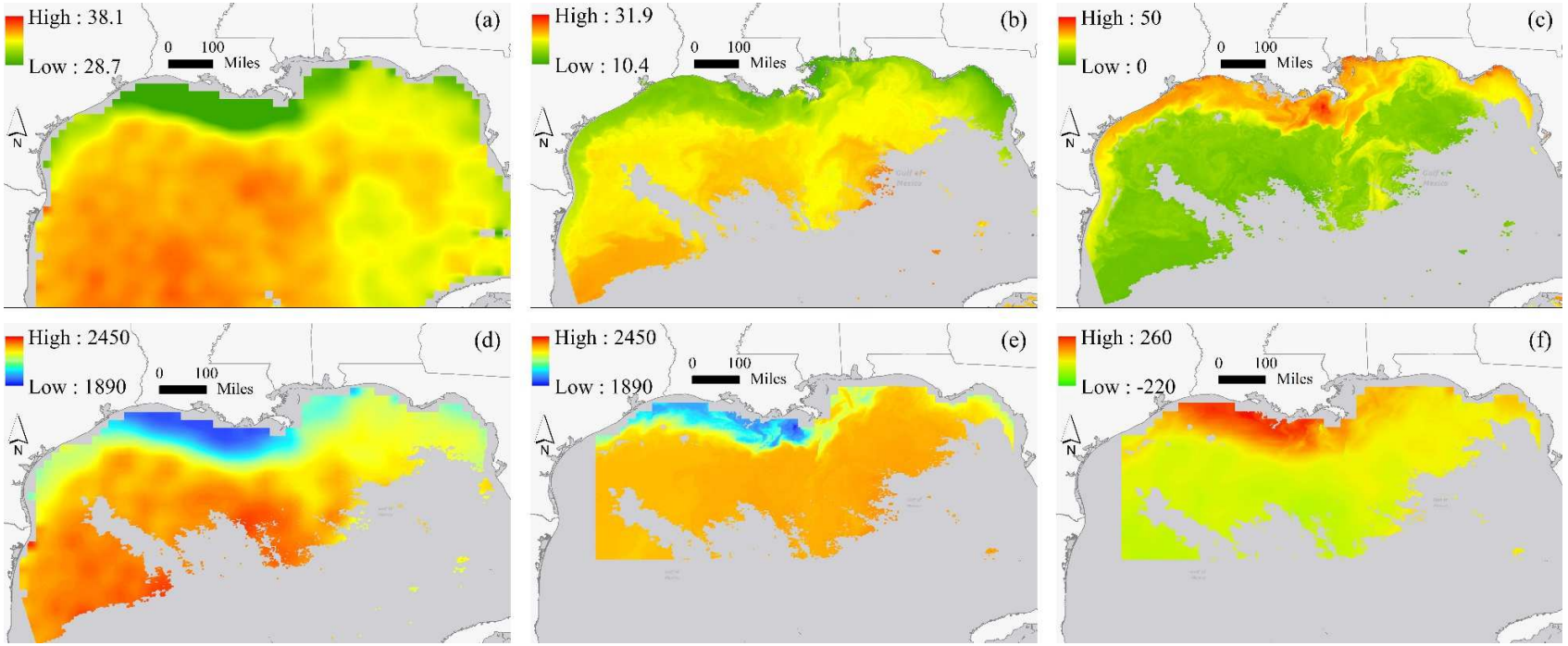
algorithm suggest TA estimated by that algorithm was highly influenced by SSS. The resulting TA image (Fig. 5d) appears very similar to the SSS image (Fig. 5a) and is strongly correlated at 0.97. With the improved GWR algorithm, the resulting TA image displays the influence of SST, SSS, and chl-*a* on TA (Fig, 4e). Specifically, due to high chl-*a* concentrations near the Mississippi River outfall (red color in Fig. 5c), the TA values are estimated to be low (blue in Fig. 5e). Thus, the coastal region showed improved estimates. Differences between the GWR algorithm and the *Lee et al.* [2006] zone 1 algorithm (Eq. 1) with chl-*a* is displayed in Fig. 5f, and the improvement offered by GWR is clearly demonstrated. The *Lee et al.* [2006] algorithm underestimated TA in the coastal waters (shown in deep blue in Fig. 5d and red in Fig. 5f), and over-estimated

293 in the areas away from the coastal region (green color in Fig. 5f) compared to the image
294 generated using the GWR algorithm.

295 **Discussion**

296 When the *Lee et al.* [2006] algorithm, which was developed for subtropical open
297 ocean waters, was applied to the N-GoM dataset, the residuals showed a clear trend and
298 the TA was mostly underestimated (Fig. 2a & 2b). The highest residuals were located in
299 coastal waters (Fig. 3a), likely a result of the lack of *in situ* data from the N-GoM in the
300 *Lee et al.* [2006] algorithm calibration. Errors were reduced by 65% when the algorithm
301 coefficients were updated using *in situ* data from N-GoM, but the algorithm coefficients
302 had high standard errors after the update. To alleviate this issue, an exponential
303 functional form was used, which decreased the standard errors in the coefficients
304 significantly (Table 3).

305 Ocean chemistry is significantly different in the N-GoM, relative to the
306 subtropical Atlantic, owing to the pronounced effects of coastal biogeochemistry heavily
307 influenced by the Mississippi River and outflow from other rivers which supply nutrients
308 and alkalinity [*Cai et al.,* 2011; *Lohrenz & Cai,* 2006]. *Schneider et al.* [2007] was able
309 to obtain a working algorithm for TA estimation in the Mediterranean Sea using SSS
310 alone. However, for the N-GoM region, use of only SSS or only SST or even SSS and
311 SST together resulted in high RMSEs possibly because of the influence of the Mississippi
312 River and outflow from other rivers. Thus, chl-*a* was introduced as an additional
313 explanatory variable in the TA algorithm to serve as a proxy for primary production
314 along with an interaction term between SST and SSS. More recently, *Fine et al.* [2017]
315 investigated global TA derived using *Lee et al.* [2006] algorithms and reported that the



316

317 **Figure 5**. Satellites derived products and estimated sea surface total alkalinity (TA) in µmol kg^{-1} for February 10, 2016, **(a)** sea surface salinity (SSS) in

318 parts per thousand, **(b)** sea surface temperature (SST) in °C, **(c)** Chlorophyll a (chl-*a*) in µg/L, **(d)** TA estimated using *Lee et al.* [2006] zone 1 algorithm

319 (Eq. 1), **(e)** TA estimated using the geographically weighted regression (GWR) algorithm, **(f)** Difference between TA estimated using GWR and *Lee et al.*

320 [2006] zone 1 algorithm.

errors in estimated TA in North Atlantic and Pacific were due to higher observed SST and SSS in large percentages than the ranges given for *Lee et al.* [2006] algorithms. An underestimation of TA was observed in this study for the same reason in the N-GoM. Hence, a regional parameterization of the algorithm using observed values in the N-GoM improved the algorithm considerably by reducing the median RMSE from 272.6 to 22.6 μmol kg^{-1}.

However, the standard error in the coefficients increased slightly as the sample sizes were significantly reduced (247 to 78 observations) when chl-*a* data were included. While the introduction of chl-*a* offered improvements in the TA estimates in the coastal region, GWR was tested to minimize spatial nonstationarity. The resulting GWR algorithm provided several interesting observations, one clear observation was the difference in coefficient magnitudes between coastal areas and the central GoM. These differences resulted from the biological and chemical complexities of the coastal waters as compared to the central GoM and was not revealed without using the GWR approach (Fig. 4 & 5).

In the development of the algorithms, it was clear that the *Lee et al.* [2006] zone 1 algorithm relied heavily upon SSS. Contrarily, the improvements based on the *Lee et al.* [2006] algorithm offered here incorporated the effects of SSS, SST, and chl-*a* in its calculations of TA. These results support the work by *Keul et al.* [2010] which showed that the TA-salinity relationships from *Lee et al.* [2006] do not apply to shelf waters effectively. Adding the chl-*a* variable produced more accurate TA estimates relative to *Lee et al.* [2006] in the N-GoM. The adjustments offered by the GWR algorithm in the coastal regions of the study domain were also very apparent when comparing it side-by-side with the image generated by applying the *Lee et al.* [2006] zone 1 algorithm (Figs. 4d and 4e). Overall, the estimated TA obtained by the GWR algorithm corresponded well to the findings of *Keul et al.* [2010] and the expected TA in the N-GoM. Nevertheless, a denser observation distribution would address local spatial effects more accurately. While observation density

346 was low, the distribution of observations was relatively uniform, suggesting the algorithm
347 coefficients derived from the GWR approach are able to produce reasonable estimates of TA
348 in the N-GoM. Mississippi River and other small rivers supply TA into the N-GoM, and there
349 is significant seasonal variability in TA and nutrient levels in the N-GoM. The relationship of
350 TA with SSS, SST, and Chl-*a* is dependent on mixing. Thus, the GWR approach and the use
351 of three variables, SSS, SST, and Chl-*a* in the algorithm account for this variability to a great
352 extent, making this approach potentially applicable to other coastal systems around the world.

353 **Conclusions**

354 The *Lee et al.* [2006] TA algorithm, which was developed for the subtropical oceans
355 did not perform efficiently in estimating TA in the N-GoM coastal regions likely owing to the
356 biological and chemical complexities in the coastal waters caused by the outflow of the
357 Mississippi River and other smaller rivers. An improved algorithm was formulated by using
358 *in situ* data from the N-GoM region and by introducing satellite-derived chl-*a* in the
359 algorithm. Prior to including chl-*a*, the coefficients and the functional form of the *Lee et al.*
360 [2006] TA algorithm were modified with decreased standard errors. The addition of an
361 interaction term between SSS and SST, and the addition of chl-*a* as an additional explanatory
362 variable resulted in an increase in adjusted R^2 with least RMSE and MAE. Finally, the GWR
363 approach offered the most reliable and valid algorithm with highest adjusted R^2 and least
364 error. *Lee et al.* [2006] global TA algorithm was improved by introducing satellite-derived
365 chl-*a* as an additional explanatory and by addressing spatial nonstationarity using GWR.
366 Notable improvement in estimated TA was observed in the coastal N-GoM region with this
367 approach as it considered both biological complexities and spatial heterogeneity. An Arc GIS
368 map package has been provided with four rasters with coefficients for the GWR algorithm as
369 supplementary material. To create TA images, one would simply multiply these raster images
370 to satellite derived SST and Chl-*a* images and either satellite or model (e.g., HYCOM)

derived SSS as per the provided functional form of the algorithm, which can easily be performed using Arc GIS or any other image processing software or program. Although the algorithm was successfully improved, future research utilizing a larger dataset from a longer period covering more locations is still essential to continue to improve TA estimates in N-GoM or other complex coastal ecosystems especially over the river-influenced coastal margins.

Acknowledgements

This research was partially supported by a NOAA grant NA11OAR4320199 to Dr. Padmanava Dash. We sincerely thank Dr. Dwight Gledhill, Deputy Director, NOAA ocean acidification program and Dr. Xinping Hu, Associate Professor, Department of Physical and Environmental Sciences, Texas A&M University, Corpus Christi for reviewing the manuscript and suggesting substantial improvements. *In situ* data used in this study was provided to the community by NOAA-AOML and NOAA-NCEI through PANGAEA. The satellite data products used in this study were provided to the community by NASA through the Ocean Biology Processing Group, GSFC, NASA and were processed using NASA's SeaWiFS Data Analysis System (SeaDAS).

References

Barrera, K. E., Knorr, P.O., Daly, K. L., Robbins, L. L. (2016). Partial pressure (or fugacity) of carbon dioxide, dissolved inorganic carbon, pH, alkalinity, temperature, salinity and other variables collected from Surface underway observations using Carbon dioxide (CO2) gas analyzer, Shower head chamber equilibrator for autonomous carbon dioxide (CO2) measurement and other instruments from WEATHERBIRD II in the Coastal Waters of Florida and Gulf of Mexico (NCEI Accession 0144622). Version 1.1. NOAA National Centers for Environmental Information. Dataset. [Accessed: June 2017]

395 Bever, A. J., Friedrichs, M. A. M., St-Laurent, P. 2021. Real-time environmental forecasts of
396 the Chesapeake Bay: Model setup, improvements, and online visualization,
397 *Environmental Modelling & Software*, 140, 105036

398 Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted
399 Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*,
400 28(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

401 Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (2010). Geographically Weighted
402 Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*,
403 28(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

404 Cai, W.-J. (2003). Riverine inorganic carbon flux and rate of biological uptake in the
405 Mississippi River plume. *Geophysical Research Letters*, *30*(2).
406 https://doi.org/10.1029/2002GL016312

407 Cai, W.-J., Hu, X., Huang, W.-J., Murrell, M. C., Lehrter, J. C., Lohrenz, S. E., … Gong, G.-
408 C. (2011). Acidification of subsurface coastal waters enhanced by eutrophication.
409 *Nature Geoscience*, *4*(11), 766–770. https://doi.org/10.1038/ngeo1297

410 Cai, W. J., and S. E. Lohrenz (2006), Carbon, nitrogen, and phosphorous fluxes from the
411 Mississippi River and the transformation and fate of biological elements in the river
412 plume and the adjacent margin, in Carbon and Nutrient Fluxes in Continental Margins:
413 A Global Synthesis, edited by K. K. Liu et al., Springer, New York.

414 Cooley, S. (2009). Ocean acidification. *Oceanography*, *22*(4), 172–181.
415 https://doi.org/10.5670/oceanog.2009.106

416 Dickson, A. G. (1981) An exact definition of total alkalinity and a procedure for the
417 estimation of alkalinity and total inorganic carbon from titration data, *Deep Sea*

418 *Research Part A. Oceanographic Research Papers*, *28*(6), 609-623.

419 Fine, R. A., D. A. Willey, and F. J. Millero (2017), Global variability and changes in

420 ocean total alkalinity from Aquarius satellite data, Geophys. Res. Lett., 44, 261–267,

421 doi:10.1002/2016GL071712.

422 Gledhill, D. K., Wanninkhof, R., Millero, F. K., & Eakin, M. (2008). Ocean acidification of

423 the Greater Caribbean Region 1996-2006. *Journal of Geophysical Research: Oceans*,

424 113(10), 1–11. https://doi.org/10.1029/2007JC004629

425 Gledhill, D., Wanninkhof, R., & Eakin, C. M. (2009). Observing Ocean Acidification from

426 Space. *Oceanography*, *22*(4), 48–59. https://doi.org/10.5670/oceanog.2009.96

427 Guo, X., W.-J. Cai, W.-J. Huang, Y. Wang, F. Chen, M.C. Murrell, S.E. Lohrenz, L.-Q.

428 Jiang, M. Dai, J. Hartmann, Q. Lin, and R. Culp. 2012. Carbon dynamics and

429 community production in the Mississippi River plume. Limnology and Oceanography

430 57(1): 1–17.

431 Hu, C., Nelson, J. R., Johns, E., Chen, Z., Weisberg, R. H., & Mu, F. E. (2005). Mississippi

432 River water in the Florida Straits and in the Gulf Stream off Georgia in summer 2004,

433 *32*(September 2004), 1–5. https://doi.org/10.1029/2005GL022942

434 Hu, C., Muller-Karger, F. E., Biggs, D. C., Carder, K. L., Nababan, B., Nadeau, D., &

435 Vanderbloemen, J. (2003). Comparison of ship and satellite bio-optical measurements

436 on the continental margin of the NE Gulf of Mexico. *International Journal of Remote*

437 *Sensing*, *24*(13), 2597–2612. https://doi.org/10.1080/0143116031000067007

438 Hu, X., Li, Q., Huang, W. J., Chen, B., Cai, W. J., Rabalais, N. N., and Turner, R.E. 2017.

439 Effects of eutrophication and benthic respiration on water column carbonate chemistry

440 in a traditional hypoxic zone in the Northern Gulf of Mexico. *Marine Chemsitry*, 190,

441 33-42.

442 Huang, W.-J., Cai, W.-J., Wang, Y., Lohrenz, S. E. & Murrell, M. C. 2015. The carbon
443 dioxide system on the Mississippi River-dominated continental shelf in the northern
444 Gulf of Mexico: 1. Distribution and air-sea CO2 flux. J. Geophys. Res.: Oceans 120,
445 1429–1445.

446 Hydes, D., Z. Jiang, M.C. Hartman, J.M. Campbell, S.E. Hartman, M.R. Pagnani, B.A. Kelly-
447 Gerreyn (2011). Dissolved inorganic carbon, alkalinity, temperature, salinity and
448 DISSOLVED OXYGEN collected from the profile and discrete sample observations
449 using Alkalinity titrator, CTD and other instruments from the Pacific Celebes in the
450 Alboran Sea, the Arabian Sea and others from 2007-06-11 to 2012-03-18 (NODC
451 Accession 0081040). Version 3.3. National Oceanographic Data Center, NOAA.
452 Dataset.

453 Inoue, M., Park, D., Justic, D., Wiseman, W. J. 2008.A high-resolution integrated hydrology–
454 hydrodynamic model of the Barataria Basin system, *Environmental Modelling &*
455 *Software*, 23, 9, 1122-1132

456 Keul, N., Morse, J. W., Wanninkhof, R., Gledhill, D. K., & Bianchi, T. S. (2010). Carbonate
457 chemistry dynamics of surface waters in the Northern Gulf of Mexico. *Aquatic*
458 *Geochemistry*, *16*(3), 337–351. https://doi.org/10.1007/s10498-010-9091-2

459 Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., … Key, R. M.
460 (2006). Global relationships of total alkalinity with salinity and temperature in surface
461 waters of the world ' s oceans, *33*, 1–5. https://doi.org/10.1029/2006GL027207

462 Lohrenz, S. E., & Cai, W. (2006). Satellite ocean color assessment of air-sea fluxes of CO2 in
463 a river-dominated coastal margin, *33*, 2–5. https://doi.org/10.1029/2005GL023942

Mehrbach, C., Culberson, C. H., Hawley, J. E., & Pytkowicx, R. M. (1973). Measurement of the Apparent Dissociation Constants of Carbonic Acid in Seawater At Atmospheric Pressure. *Limnology and Oceanography*, *18*(6), 897–907. https://doi.org/10.4319/lo.1973.18.6.0897

Millero, F. J., L. Lee, and M. P. Roche (1998), Distribution of alkalinity in the surface waters of the major oceans, *Mar. Chem.*, 60, 111–130.

Picheral, Marc; Searson, Sarah; Taillandier, Vincent; Bricaud, Annick; Boss, Emmanuel; Ras, Josephine; Claustre, Hervé; Ouhssain, Mustapha; Morin, Pascal; Coppola, Laurent; Gattuso, Jean-Pierre; Metzl, Nicolas; Thuillier, Doris; Gorsky, Gabriel; Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2014): Vertical profiles of environmental parameters measured on discrete water samples collected with Niskin bottles at station TARA_143 during the Tara Oceans expedition 2009-2013. PANGAEA, https://doi.org/10.1594/PANGAEA.839230. [Accessed: June 2017]

Le Quéré C, Andrew, R.M, Canadell, J.G, et al. Global Carbon Budget 2016. Earth System Science Data 2016; 8: 605-49.

Rabalais, N. N., Turner, R. E., & Wiseman, W. J. (2002). Gulf of Mexico Hypoxia, A.K.A. "The Dead Zone." *Annual Review of Ecology and Systematics*, *33*(1), 235–263. https://doi.org/10.1146/annurev.ecolsys.33.010802.150513

Raven, J., Caldera, K., Elderfield, H., Hoegh-Guldberg, O., Liss, P., Riebesell, U., … Quinn, R. (2005). Ocean acidification due to increasing atmospheric carbon dioxide. *The Royal Society*, (June), 60. Retrieved from www.royalsoc.ac.uk

Raymond, P. A., Cole, J. J. (2003) Increase in the export of alkalinity from North America's largest river, *Science*, 301:88–91

487 Schneider, A., D. W. R. Wallace, and A. Kortzinger (2007), Alkalinity of the Mediterranean
488 Sea, Geophys. Res. Lett., 34, L15608, doi:10.1029/2006GL028842.

489 Salisbury, J. E., Campbell, J. W., Meeker, L. D., & Vörösmarty, C. (2001). Ocean color and
490 river data reveal fluvial influence in coastal waters. *Eos*, *82*(20).
491 https://doi.org/10.1029/01EO00117

492 Salisbury, J., C. Hunt and A. Mannino (2016). Partial pressure (or fugacity) of carbon
493 dioxide, dissolved inorganic carbon, pH, alkalinity, temperature, salinity and other
494 variables collected from Surface underway, discrete sample and profile observations
495 using Alkalinity titrator, CTD and other instruments from PELICAN in the Coastal
496 Waters of Louisiana, Coastal Waters of Texas and Gulf of Mexico from 2013-09-09 to
497 2013-09-22 (NCEI Accession 0157461). Version 1.1. NOAA National Centers for
498 Environmental Information. Dataset. [Accessed: June 2017]

499 Sun, Q., Tang, D., & Wang, S. (2012). Remote-sensing observations relevant to ocean
500 acidification. *International Journal of Remote Sensing*, *33*(23), 7542–7558.
501 https://doi.org/10.1080/01431161.2012.685978

502 Takahashi, T., Sutherland, S. C., Feely, R. A., & Wanninkhof, R. (2006). Decadal change of
503 the surface water pCO2 in the North Pacific: A synthesis of 35 years of observations.
504 *Journal of Geophysical Research: Oceans*, *111*(7), 1–20.
505 https://doi.org/10.1029/2005JC003074

506 Ternon, J. F., C. Oudot, A. Dessier, and D. Diverres (2000), A seasonal tropical sink for
507 atmospheric CO2 in the Atlantic ocean: The role of the Amazon River discharge, *Mar.*
508 *Chem.*, 68, 183–201.

509 Walker, N. D. (1996). Satellite assessment of Mississippi River plume variability: Causes
510 andpredictability. *Remote Sensing of Environment*, *58*(1), 21–35.