

We highly appreciate the reviewers for all their valuable comments and suggestions. We have included our responses in blue. References and Supplementary Figures (Fig. S) are after the response sections.

1. Response to Dr. Fennel

1.1 Validation

1.1.1. Physical validation:

The authors provide no validation of the physical model. If there are previous publications in which this is reported, it would be fine to refer to those. Otherwise some physical model validation should be provided.

We agree that the manuscript needs some physical-component validation. We have done comparisons that show a good agreement between model outputs and observation of SST, coastal sea level anomalies, eddy kinetic energy, and surface salinity (See Figs. S1 to S5). In addition to those variables, in the revised manuscript version we will provide validation for time series of salinity and vertical profiles of temperature-salinity, in both coastal and oceanic domains.

1.1.2. Chlorophyll patterns:

On page 6 (first paragraph) the authors state that the model underestimates mean satellite chlorophyll by factors between 2.5 to 3. These are rather large deviations in the mean. They then give reasons for why the satellite can't be trusted. There are two problems with this: first, the simpler models of Fennel et al. (2011) and Laurent et al. (2012) reproduced satellite chlorophyll without such large biases; and, second, if chlorophyll cannot be trusted it shouldn't be used in validation. However, satellite derived chlorophyll is essentially the only data set that is used in this manuscript to validate the model.

We understand the concern pointed out by the reviewer regarding to the large deviation in the mean of model and satellite chlorophyll in the coastal region. To support better our results, we will extend the validation analysis for chlorophyll, including a comparison between model outputs and in situ observations. We are still working in this validation analysis, but preliminary comparisons indicate that model chlorophyll tends to underestimate in situ chlorophyll in the Louisiana-Texas inner shelf (see Figures S6-S7). An estimation of the differences between model and in-situ chlorophyll will be reported in the revised manuscript version.

Regarding to the first related problem (a simpler model reproduces better the chlorophyll variability), we agree that the Fennel model matches better the chlorophyll pattern in the Louisiana-Texas shelf, as shown by Fennel et al. (2011) and Laurent et al. (2012). However, the Fennel model tends to overestimates satellite chlorophyll in the open ocean region, which can be noted in the chlorophyll patterns reported by Xue et al. (2013) [see their Figure 8]. To evaluate better to what degree the chlorophyll patterns from Fennel and our 18-component model (referred to as GOM) differ, we run the Fennel model for the period 1999-2004, using the same parameters values used by Fennel et al. (2011).

The comparison between model-derived and satellite chlorophyll patterns (Fig. S8 and S9) indicates that Fennel model overestimates satellite chlorophyll in most of the open ocean region (bottom depth >200 m). This model bias displays a seasonal pattern, with the strongest deviations occurring in winter and early spring (Dec-Mar) (the overestimation factor can be >3; Fig. S8c). On the other hand, the chlorophyll from the GOM model reproduces better the mean satellite condition and temporal variability in the open ocean region, showing no large seasonal deviation. In the Mississippi Delta and Texas shelf, there is a clear difference between the long-term mean of Fennel and GOM (Fennel matching well the SeaWiFS mean, Fig. S8a-b). However, the Fennel-SeaWiFS and GOM-SeaWiFS correlation coefficients are pretty similar (in terms of correlation, Fennel does better than GOM in the Texas shelf, and GOM does better than Fennel in the Mississippi Delta), indicating a similar model ability to reproduce the temporal variability of chlorophyll. We will add this comparison within the result section.

Regarding to the second problem (if chlorophyll cannot be trusted it shouldn't be used in validation), previous studies have shown that although satellite chlorophyll can largely overestimate in situ chlorophyll in regions influenced by river runoff (until by 300%), there is a significant correlation between satellite and in situ chlorophyll (as example, see Figure 1.10 in Nababan, 2005). Besides, the bias associated to optically complex waters occurs mainly in the coastal region but not in Open Ocean. These two aspects, added to the fact that satellite sensors provide large spatial and temporal coverage to properly describe seasonal and interannual variability, makes the model-satellite comparison of chlorophyll valid.

1.1.3. Biomass and nutrients:

There should be some validation of the biogeochemical model with in situ observations of phytoplankton biomass and nutrients. Such observations are available in the NODC and GOMRI databases. Profile comparisons for the open Gulf should be included.

We agree that the manuscript requires validation for the biological and chemical components. Therefore, a validation of nutrient and plankton biomass, using the NODC and GOMRI databases, will be provided.

1.1.4. Diatom and nanophytoplankton:

Perhaps the largest omission, given the objective of the study, is that there is no validation of the different phytoplankton groups. I recognize that it is hard to get good data sets for this purpose, but there are some algorithms that can be used to separate satellite-derived chlorophyll into different size groups (see Hirata et al. 2011, Mouw et al. 2017 and references therein).

As Dr. Fennel indicates, it is difficult to obtain data set to validate the diatom and nanophytoplankton components. We will not include a satellite-based analysis on phytoplankton group biomass pattern, as time limitations preclude performing this type of analysis (which we think can be the topic for an independent paper). Although biomass patterns of phytoplankton functional groups are not well documented in the region, few studies have reported data on size-fractionated chlorophyll, and phytoplankton group

abundance/biomass (Bode and Dortch, 1996; Nelson and Dortch, 1996; Dagg and Breed, 2003; Zhao and Quigg, 2014) that can be compared with our model. Figure S10 shows model and observed diatom to total chlorophyll ratio for two coastal stations (A and B) over the Louisiana shelf, taken from data reported by Zhao and Quigg (2014). In station B, both model and observed chlorophyll ratio are similar. In station A, the model ratio overestimates the observed ratio, especially in August. Both simulated and observed ratios show greater values during April than during August. Though differences between model and observations are evident, the standard deviation bars indicate large variability, which can be linked to strong mesoscale variability in the Mississippi delta region (e.g. Marta-Almeida et al., 2013). We considered that the model does a reasonable work simulating those reported chlorophyll ratios, and is qualitatively consistent with the seasonal pattern. We will include Figure S10 in the validation section of the revised paper version.

1.2. Terminology

With regard to terminology: In section 3.4 the authors define the “biological term” as the balance between phytoplankton production and biological losses. This is the same as Net Phytoplankton Growth, a widely used term in biological oceanography. It is not only unnecessary to redefine this as a new term, but also potentially confusing. Then the authors state that the balance of the biological and physical terms determines the change in net phytoplankton growth. This is wrong. Net phytoplankton growth is equal to what they defined as the biological term. The balance between the biological and physical terms is the local rate of change of phytoplankton.

We highly appreciate the term names clarification, and understand that is preferable avoid any confusing terminology in the budget analysis. We will modify the term names following Dr. Fennel indications.

1.3. Improving description of the results and the context in the existing literature:

1.3.1. In the last sentence of the abstract, they claim that their study shows the importance of representing large and small plankton in order to describe PP patterns. This is not supported by the results presented. On the one hand, there is no validation of the contributions of large and small phytoplankton to biomass and PP (see 1 above). On the other hand, there is no comparison to simulated phytoplankton abundance and PP from a model with only one phytoplankton group. Simpler models exist that, in fact, reproduce chlorophyll from satellite more accurately than this model in the Mississippi plume region (see comment 1.2).

We agree that that sentence is not supported by the result presented and will be removed from the abstract.

1.3.2. In the second to fourth sentences of the discussion the authors make statements about their results that are not supported. “Inclusion of two phytoplankton components allowed for realistic representation...” is not accurate as simple models arguably

reproduced this better (see comment 1.2). “The good agreement between model outputs and observations of chlorophyll...” is a questionable statement (see again comment 1.2).

We recognize that Fennel model match better the satellite chlorophyll patterns than our model in the coastal regions, and that we need to tone down those statements. However, it is worth to note that our model reproduces better than Fennel model the satellite chlorophyll pattern in the deep ocean region (see answer to comment 1.1.2), so we will comment about this in the discussion.

1.3.3. With respect to phosphorus (P) the authors seem to be diminishing previous findings in an effort to justify why their model does not include P. On page 3 (line 8, sentence starting with “Although: : :”) they seem to suggest that previous studies (specifically Laurent et al. 2012) suggest P limitation to be unimportant. This is not the conclusion of Laurent et al. (2012) nor of the follow-up studies by Laurent and Fennel (2014) and Fennel and Laurent (2017), which are consistent with the observational studies by Sylvan et al. (2006, 2007). All these studies do suggest the P limitation is critically important in the region influenced by the Mississippi River plume. Saying that P limitation is “moderate” while N and Si limitation are “critical” seems disingenuous. To be clear, I do not object to the fact that P is neglected in this model. All models are simplifications. It would be fine to state that their model neglects P, although it has been shown to be important in a portion of the model region. In the Discussion (end of first paragraph) it would be appropriate to be more forthcoming about previous studies on P limitation.

We understand the point indicated by the reviewer and recognize that phosphorous limitation can be critical in the region influenced by the Mississippi River. We will modify the statement in the ‘Model description’ section to:

“Although previous modeling studies have indicated the existence of phosphate limitation near the MS-A deltas during May-July (Sylvan et al. 2006, 2007; Laurent et al., 2012; Laurent and Fennel, 2014), we focus here on the role of N and Si, as observational studies suggest that N and Si can modulate phytoplankton composition (Dortch and Whitley, 1992; Nelson and Dortch, 1996; Lohrenz et al., 1997; 2008; Rabalais et al., 2002; Zhao and Quigg, 2014).”

In addition, we will be more forthcoming about previous studies on P limitation, incorporating within the discussion the studies by Laurent et al. (2012), Laurent and Fennel (2014), Fennel and Laurent (2017), and Sylvan et al. (2006, 2007).

1.3.4. The statement in the Discussion (last sentence starting on page 11) about consistency with the dilution-recoupling hypothesis of Behrenfeld seems a bit cavalier. No detailed analysis in support of this statement was presented in this manuscript. The authors may want to consider the study by Kuhn et al. (2015), which used the same data set as Behrenfeld, and later papers by Behrenfeld where he backtracked himself somewhat from his early paper (Behrenfeld et al. 2013).

We agree with the reviewer that more analyses need to be done to support the consistency between the dilution-recoupling hypothesis and our model results. Since this aspect is beyond the paper goals, we decided to remove the dilution-recoupling hypothesis part from the discussion.

Other comments (not in order of importance):

1.4. P1, Line 13: Suggest inserting “improving” after “tools for”

The change will be done accordingly.

1.5. P1, Line 14: Suggest removing “However”

The change will be done accordingly.

1.6. P1, Line 19, sentence starting with “The model results show...” and following sentences in the abstract. Because diatoms in the model are strongly silica-limited doesn’t necessarily mean they are in reality. Making inferences about reality from the model requires that the model accurately reproduces reality, which in this case is hard to prove. The authors certainly haven’t (see my comments about validation). I would suggest that here and throughout the remainder of the abstract and manuscript the authors are more precise in their language. It is fine to say “diatoms in the model are silica limited” or some variation thereof. And “Simulated nanophytoplankton are...” rather than “Nanophytoplankton are...”

Agree with the suggestion. We will precise better our result’s statements.

1.7. P1, Line 27: Suggest replacing “vertical diffusion” with “turbulent vertical diffusion” or “vertical mixing.” Diffusion typically refers to molecular diffusion which acts on too small scales to make any difference to the processes considered here.

Agree with the suggestion, change will be done accordingly.

1.8. P1, Line 27, sentence starting with “This study highlights the...” This is an overstatement not supported by the results actually presented in this manuscript. See major comment 3.

This statement will be removed.

1.9. P2, Line 9: ...because of deleterious impact on coastal ecosystems.” The authors should provide one or more references in support of this statement, or modify it. I would like to challenge them to find a study that shows deleterious impacts on the ecosystem in the northern Gulf of Mexico (I am not aware of one). There are studies about specific aspects of the ecosystem, which would be fine to cite if sentence is slightly modified.

Following the reviewer suggestion, we will modify this statement, mentioning the specific

ecosystem aspects that are negatively impacted by bottom hypoxia, such as individual growth and metabolism (Rosas C, et al., 1998; Craig and Crowder, 2005) and stock catchability (Craig, 2012).

1.10. P3, Line 5, sentence beginning with “New modelling efforts...” I object to the logic of this statement. Adding complexity to biogeochemical models is not in itself a worthwhile undertaking. It has to be motivated by the scientific questions (e.g. one might be interested in species succession). Sentence should be reformulated accordingly.

The sentence will be reformulated to:

New modeling efforts are required to examine spatiotemporal biomass patterns of phytoplankton functional groups across the northern and deep GoM regions.

1.11. P3, Line 8: “...diatoms require...” Citing a modeling study (Kishi et al.) in support of a general statement about diatom traits seems inappropriate. There are more appropriate references. I suggest the authors look up publications by Elena Litchman and collaborators. She has worked extensively on documenting phytoplankton functional traits.

We agree with the suggestion. We will include Litchman and Klausmeier (2008) as reference for the diatom traits [Litchman, Elena, and Christopher A. Klausmeier. "Trait-based community ecology of phytoplankton." Annual review of ecology, evolution, and systematics 39 (2008): 615-639.]

1.12. P4, Line 30: Which basin does “basin-scale” refer to here?

It refers to the Atlantic Ocean. We will precise the statement in the revised manuscript version.

1.13. P5, Line 18: “...randomly selected year” Which year?

We will modify the sentence to make the spin-up procedure clearer:

A 40-year model spin-up was completed before starting the historical simulation. To run the model spin-up, we used the basin-model boundary conditions and ERA surface fluxes of randomly selected years from the 1979-2014 period, following Lee et al. (2011). After the spin-up, the model was run continuously from January 1979 until December of 2014, with monthly averaged fields saved.

1.14. P5, Line 23: Stating the model “reproduces” the observations is an overstatement. It would be more appropriate to say they agree qualitatively.

The sentence will be modified to:

“The spatiotemporal patterns of model and satellite chlorophyll agree qualitatively (Fig. 2).”

1.15. P5, Line 31: The authors should make it much more clear upfront that these are anomalies (i.e. that the bias was removed).

We will indicate that the series in Figure 2 correspond to anomalies with the long-term mean removed.

2. Response to Referee #2

2.1. My major concern is associated with the validation of the coupled physical biogeochemical model:

First, there is no physical validation presented in the paper, despite that the authors have emphasized the importance of physical processes on the net phytoplankton growth. Has the physical validation work been done and/or published elsewhere? If yes, it is important to summarize that here in some way. If not, I think it's worthwhile to do some extra work on physical validation to make the presented results here more convincing considering how important the physics is controlling the biogeochemical cycling in this region (e.g., the mixing and transport by riverine waters to northern GoM, Loop Current and eddy interactions to deep GoM, etc.). For example, the simulated spatial extent of the high chlorophyll river plume in northern GoM is narrower than that observed in satellite (visually viewed from Fig. 2), could it be associated with the distant transport of riverine nutrients?

We agree that a validation of the physical model component is required. In the revised manuscript version, we will provide model-observations comparisons for SST, coastal sea level anomalies, eddy kinetic energy, salinity, salinity, and T-S vertical profiles in coastal and oceanic domains (see answer to Dr. Fennel, section 1.1.1 and Figures S1-S5).

Second, the validation of biogeochemical (BGC) model doesn't seem sufficient to me. The BGC validation in the paper primarily relies on comparing model simulated and satellite observed surface chlorophyll. While the model overall reproduces the dominant seasonal and spatial patterns in satellite chlorophyll, it significantly underestimates the coastal chlorophyll both in magnitude (2.5-3 times lower in the model) and spatial extent. The authors attribute the mismatch to satellite overestimating in situ observations of chlorophyll in northern GoM. If true, it would be useful to also include comparisons between simulated and in situ observations of chlorophyll in the paper for justification. In addition, while satellite chlorophyll observations have the advantage for model validation due to its spatial and temporal coverage, they are limited to the first optical depth that could hardly represent the plankton dynamics in subsurface water (e.g., the deep chlorophyll maxima). Hence a good complement to the validation might be including comparison to chlorophyll profiles, which to my knowledge is available in GoM during the model simulation period (e.g., the bio-optical profiling float results presented in Green et al., 2014). Also, there are relatively 'abundant' observations, apart from chlorophyll, in the northern GoM, such as those provided by Mechanisms Controlling Hypoxia (MCH) program (<http://hypoxia.tamu.edu/field-program>), in situ observations of

primary production (Lehrter et al. 2009), and water column community respiration rates (Murrell et al. 2013). These datasets might improve the BGC validation in coastal region where satellite chlorophyll is considered to have higher uncertainty.

We agree that the biogeochemical model requires extra validation and we are working now in the comparison between model and observations for nutrients, plankton biomass, and primary production, using the NODC, GOMRI, and other database available, for both coastal and oceanic domains, which will be included in the revised paper version.

2.2. One novelty of this work is that the model includes two phytoplankton types and two zooplankton types that complement the previous modeling work in GoM that mostly only includes one phytoplankton and one zooplankton type. While the additional complexity added to the BGC model is more faithful in representing the lower-trophic level dynamics in real system, it also adds more complexities and challenges in calibrating and validating the model. With respect to calibration, have the parameter values shown in Table 1 (especially those with *) been informally or formally tuned or optimized? Are the conclusions presented here sensitive to the selected parameter values? I think providing more information/comments on these would be helpful to others. The additional complexity of the BGC model also adds difficulties in model validation, e.g., the model-data chlorophyll comparison alone cannot tell how reasonable the model simulates each type of phytoplankton group as it could not distinguish the contribution from small- and large-size phytoplankton groups. How has the added complexity benefit us to understand the plankton dynamics in this region? Does the presented model do a better job than the previous modeling work that only include one phytoplankton type (e.g., compared with Xue et al. 2013)? I think readers would appreciate with a bit more discussions/comments on these.

*The selected parameters are within ranges reported in previous studies, with * indicating minor variations from reported values. We agree with the reviewer that more information and comments on the model parameters are helpful to other biogeochemical modelers, and we will include them in the discussion of the new revised version. In addition, we will include a comparison between our biogeochemical model (refer to as GOM model) and Fennel model. This comparison reveals that although Fennel's chlorophyll match better the long-term mean of satellite chlorophyll in the coastal regions, GOM does better reproducing the seasonal chlorophyll patterns in the open ocean, with no significant seasonal bias (see answer to Dr. Fennel, section 1.1.2 and Figures S8-S9). We will run additional experiments to evaluate the sensitivity of the simulated chlorophyll/phytoplankton patterns to changes in model parameters. Specially, we will examine to what degree the improved representation of seasonal chlorophyll patterns in the open ocean region depends on the parameterization of phytoplankton growth (nutrient-limitation and maximum growth rate), zooplankton grazing (micro- and mesozooplankton parameters) and chlorophyll to carbon ratios.*

2.3. Specific comments:

Page 4, Line 6: Would it be more appropriate to list an observational rather than a modeling work (Xue et al., 2013) as a reference?

We will modify the citation to Green and Gould (2008).

Green, R. E., and R. W. Gould (2008), A predictive model for satellite-derived phytoplankton absorption over the Louisiana shelf hypoxic zone: Effects of nutrients and physical forcing. Journal of Geophysical Research: Oceans, 113(6): 1–17, <https://doi.org/10.1029/2007JC004594>

Page 4, Line 14: delete one ‘to’ either in front of the ‘.’ or after the number.

The change will be done accordingly.

Page 4, Line 22: Why listed MODIS SST here? Has it been used anywhere in the paper?

We apologize for this mistake. However, the revised manuscript version will include a comparison between model SST and MODIS, so a description of the MODIS data will be included in the Data section.

Page 4, Line 28: Horizontal diffusivity is non-zero here, but it seemed to be neglected when analyzing the role of advection and diffusion in section 3.4.

We did not include horizontal diffusivity term in the budget analysis because is about 2 orders of magnitude smaller than the advection and vertical diffusion terms, so its contribution can be neglected. We will mention this aspect in the revised paper version.

Page 4, Line 30: Does the basin-scale model also include biogeochemistry and provide BGC initial conditions? If not, how do you specify them? Could you also provide more information on how you specify open boundary conditions? Has tide been included?

The basin model specifies the boundary and initial condition for both the physical and biogeochemical model. Tides were not included in the model. The revised manuscript version will have an extended description of all the boundary condition aspects.

Page 5, Line 18: Where were the boundary conditions and surface fluxes extracted from? the basin-scale model?

The basin model provides the boundary conditions, and the surface fluxes are from ERA-interim (same surface forcing as in the basin model). We will clarify those aspects in the revised version.

Page 6, Line 23: ‘mean production values’, is it spatial or/and temporal mean? Maybe also provide the standard deviation if available, since the primary production is highly variable?

Standard deviation values of primary production will be provided in the revised manuscript version

Page 7, Line 29: change ‘ranges’ to ‘range’?

The change will be done accordingly.

Page 8, Line 26: In the text, it's switching between 'summer' (or winter) and 'months' back and forth. Could you specify the summer and winter months at the first time they appear?

We will make the modification accordingly.

Page 10, Line 22-24: This statement is a bit exaggerated to me since the validation is on chlorophyll, a combination of two phytoplankton groups, that how well each type of phytoplankton is simulated by the model is not directly validated.

We agree with the reviewer. We will rewrite this paragraph, indicating the agreement and differences between model and observed chlorophyll patterns. Besides, we will discuss about the comparison between Fennel and our 18-component model.

Fig.2: the lower limit of the color bar is missing? From 0? What does the gray contour line represent? 200m isobath?

The colorbar ranges from 0.05 to 5 mg m⁻³. We will include the lower limit in the revised Figure 2 version. The contour gray line represents the 200 m isobath. We will mention this in the Figure 2 legend, and also will include '200 m' as contour labels.

Fig.8: should be '...in panels a-b depict...'

Change will be done accordingly.

3. Response to Referee #3

3.1. Validation of the physical model

The paper stated that the boundary conditions were from a HYCOM model, yet the model (ROMS)'s own performance regarding circulation and T/S fields was not evaluated, without which, I would have a big question mark about the results presented in the manuscript;

We agree that the physical model component needs to be validated. We have completed part of this validation, which shows a good agreement between observed and simulated patterns of SST, sea level anomalies, eddy kinetic energy, and surface salinity (see section 1.1.1 in answer to Dr. Fennel and Figures S1-S5). In the revised manuscript version, we will also include comparisons of surface salinity time series, as well as vertical profiles of temperature and salinity, in coastal and oceanic domains.

3.2. Validation of the biogeochemical model

The author evaluated their model's performance via a comparison against satellite data and admitted that their model underestimated the Chl-a. And unfortunately, these satellite data were the only source used for model evaluation. How about the model's performance on nutrient and plankton groups? Without such information, it is hard to conclude that the model could at least represent the nutrient and biological cycle in the Gulf;

We agree that the biogeochemical model requires extra validation. To evaluate the model performance, we are working now in comparing pattern in nutrients, plankton biomass,

primary production, and chlorophyll with observations from NODC, GOMRI, and other database available. This validation will be included in the revised paper version.

3.3. Given that point 1) and point 2) were addressed, I could not find the benefit of introducing the new plankton group (2 phytoplankton and 3 zooplankton vs. 1 phytoplankton and 1 zooplankton by Fennel et al. 2011), which, indeed, could be the most important contribution of this study

We understand the reviewer concern regarding to the benefit of introducing a more complex representation of lower trophic levels, as the first manuscript version did not include any contrast with outputs from simpler biogeochemical models. In the revised manuscript version, we will include a comparison between chlorophyll patterns derived from our 18-component model (refers to as GOM) and Fennel's model, which reveals that GOM does better reproducing the seasonal chlorophyll patterns in the open ocean region, with no significant seasonal bias (see section 1.1.2 in answer to Dr. Fennel and Figures S8-S9). This result strongly suggests that our model allows a better representation of phytoplankton dynamics compared to previous model results in the open ocean region. To complement this comparison, we will perform new sensitivity experiments, examining to what degree the improved representation of seasonal chlorophyll patterns in the open ocean region depends on the parameterization of phytoplankton growth (nutrient-limitation and maximum growth rate), zooplankton grazing (micro- and mesozooplankton parameters) and chlorophyll to carbon ratios.

References

Behrenfeld, M. (2010) Abandoning Sverdrup's critical depth hypothesis on phytoplankton blooms, *Ecology* 91(4), 977-989.

Bode, A. and Dortch, Q. (1996) Uptake and regeneration of inorganic nitrogen in coastal waters influenced by the Mississippi River: spatial and seasonal variations. *J. Plankton Res.* 18, 2251– 2268.

Craig, J.K. (2012) Aggregation on the edge: Effects of hypoxia avoidance on the spatial distribution of brown shrimp and demersal fishes in the northern Gulf of Mexico. *Mar Ecol Prog Ser* 445:75–95

Craig J.K, Crowder L.B (2005) Hypoxia-induced habitat shifts and energetic consequences in Atlantic croaker and brown shrimp on the Gulf of Mexico shelf. *Mar Ecol Prog Ser* 294:79–94.

Dagg, M.J., and Breed, G.A. (2003) Biological effects of Mississippi River nitrogen on the northern Gulf of Mexico—a review and synthesis, *Journal of Marine Systems*, 43(3), 133-152.

Fennel, K., Hetland, R., Feng, Y., and DiMarco, S. (2011) A coupled physical-biological model of the Northern Gulf of Mexico shelf: model description, validation and analysis of phytoplankton variability, *Biogeosciences*, 8(7), 1881.

Fennel, K. and Laurent, A. (2017) N and P as ultimate and proximate limiting nutrients in the northern Gulf of Mexico: Implications for hypoxia reduction strategies, *Biogeosciences Discuss.*, <https://doi.org/10.5194/bg-2017-470>, in review.

Laurent, A., Fennel, K., Hu, J., and Hetland, R. (2012) Simulating the effects of phosphorus limitation in the Mississippi and Atchafalaya River plumes, *Biogeosciences*, 9(11), 4707-4723.

Laurent, A., and Fennel, K. (2014). Simulated reduction of hypoxia in the northern Gulf of Mexico due to phosphorus limitation. *Elem Sci Anth*, 2.

Litchman, E., and Klausmeier, C.A. (2008) Trait-based community ecology of phytoplankton, *Annual review of ecology, evolution, and systematics* 39: 615-639.

Marta-Almeida, M., Hetland R.D., and Zhang, X. (2013) Evaluation of model nesting performance on the Texas-Louisiana continental shelf. *J.Geophysical Res.-Oceans*, 118: 1-16. doi: 10.1002/jgrc.20163.

Nababan, B. (2005) Bio-optical variability of surface waters in the Northeastern Gulf of Mexico. PhD thesis, College of Marine Science, University of South Florida, Tampa, FL, 145 pp.

Nelson D.M., and Dortch, Q. (1996) *Silic acid and silicon limitation in the plume of the Mississippi River: evidence from kinetic studies in spring and summer*, *Marine Ecology Progress Series*, 136, 163-178.

Rosas C, et al. (1998) *Effect of dissolved oxygen on the energy balance and survival of Penaeus setiferus juveniles*. *Mar Ecol Prog Ser* 174:67–75.

Sylvan, J.B., Dortch, Q., Nelson, D.M., Brown, A.F.M., Morrison, W., and Ammerman, J.W. (2006) *Phosphorus limits phytoplankton growth on the Louisiana shelf during the period of hypoxia formation*, *Environ. Sci. Technol.*, 40, 7548–7553, doi:10.1021/es061417t.

Sylvan, J. B., Quigg, A., Tozzi, S., and Ammerman, J. W. (2007) *Eutrophication-induced phosphorus limitation in the Mississippi River Plume: evidence from fast repetition rate fluorometry*, *Limnol. Oceanogr.*, 52, 2679–2685, doi:10.4319/lo.2007.52.6.2679.

Xue, Z., He, R., Fennel, K., Cai, W. J., Lohrenz, S., and Hopkinson, C. (2013) *Modeling ocean circulation and biogeochemical variability in the Gulf of Mexico*, *Biogeosciences*, 10(11), 7219.

Zhao, Y., and Guigg A. (2014) *Nutrient limitation in Northern Gulf of Mexico (NGOM): phytoplankton communities and photosynthesis respond to nutrient pulse*, *PloS one* 9.2, e88732.

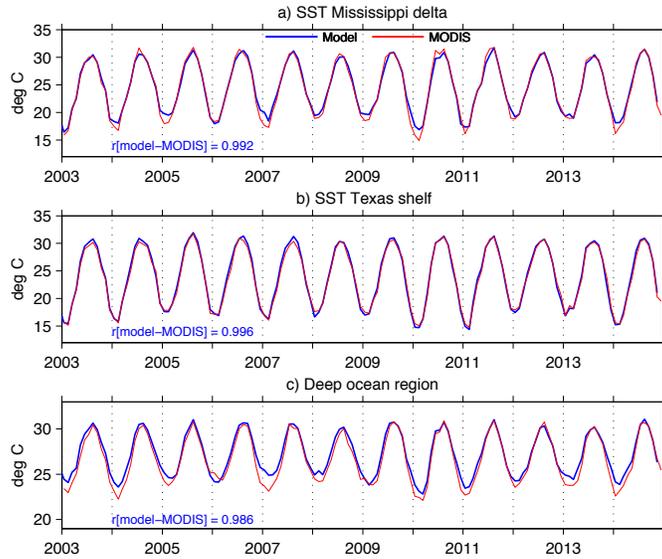


Figure. S1. Monthly time series of SST derived from model outputs and MODIS for the Mississippi delta, Texas shelf, and Deep Gulf regions. Correlation coefficient between model and MODIS series are indicated at each panel.

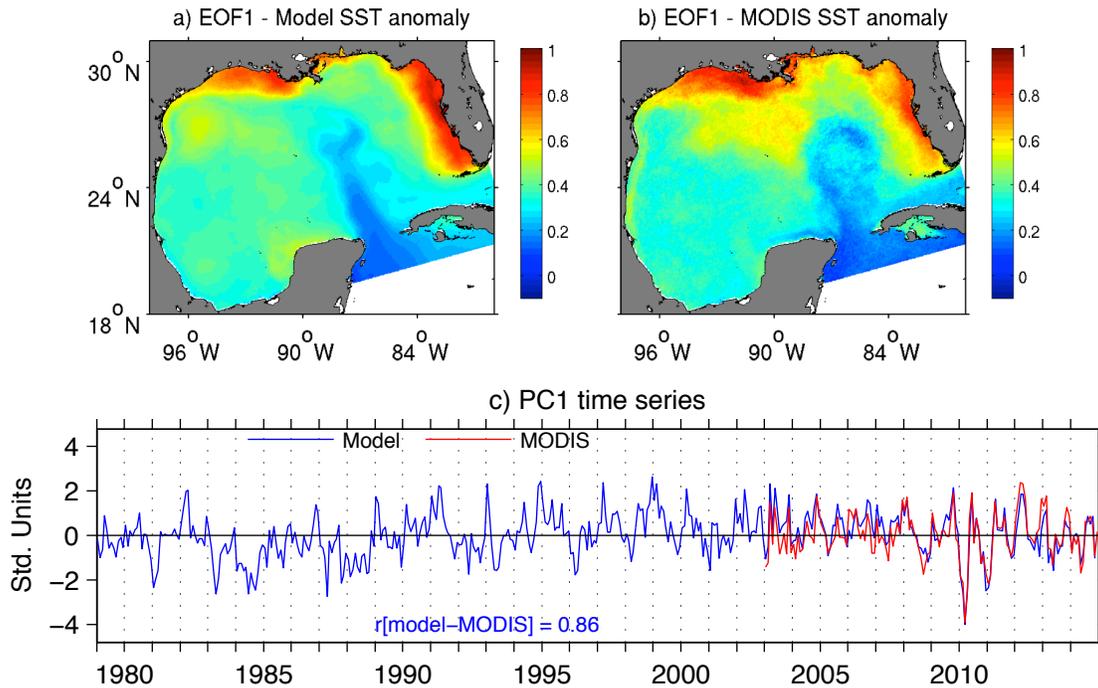


Figure. S2. a-b) First Empirical Orthogonal Function (EOF1) of model and MODIS SST anomalies (seasonal cycle removed). c) First Principal Component time series (PC1) of model and MODIS SST anomalies. Correlation coefficient between model and MODIS PC1 is indicated in panel c.

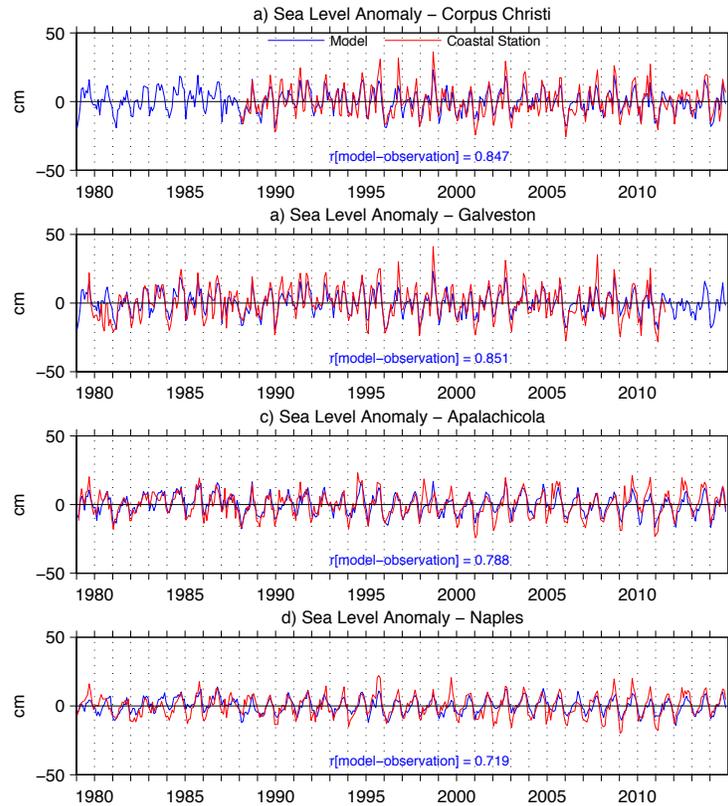


Figure S3. Monthly sea level anomaly derived from model (blue) and coastal observations (red) at a) Corpus Christi ($27^{\circ} 35' N$, $97^{\circ} 13' W$), b) Galveston ($29^{\circ} 17' N$, $94^{\circ} 47' W$), c) Apalachicola ($29^{\circ} 43' N$, $84^{\circ} 58' W$), and d) Naples ($26^{\circ} 7' N$, $81^{\circ} 48' W$). Coastal sea level observations were retrieved from the Sea Level Center, University of Hawaii (<https://uhslc.soest.hawaii.edu>). Correlation (r) between modeled and observed time series is indicated at each panel.

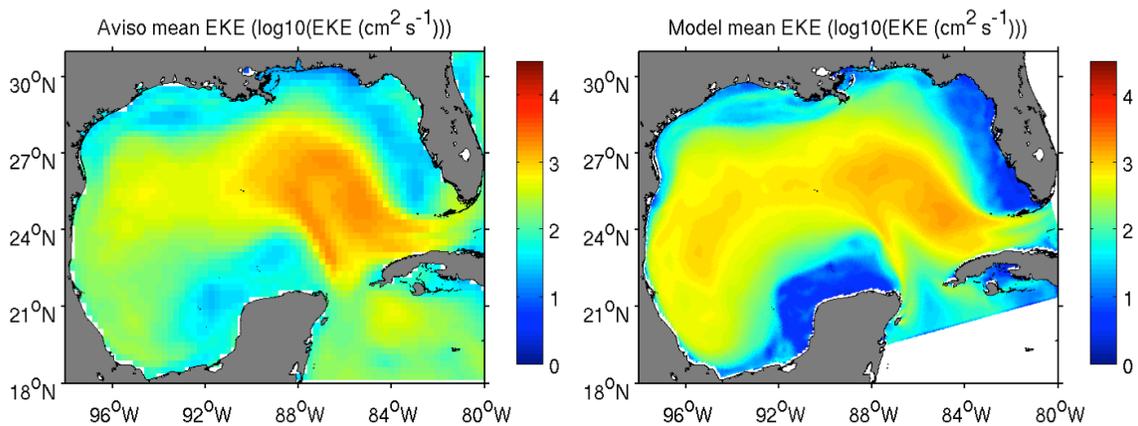


Figure S4. Mean Eddy Kinetic Energy (EKE) derived from AVISO sea surface height (left) and model (right) during period 1993-2010.

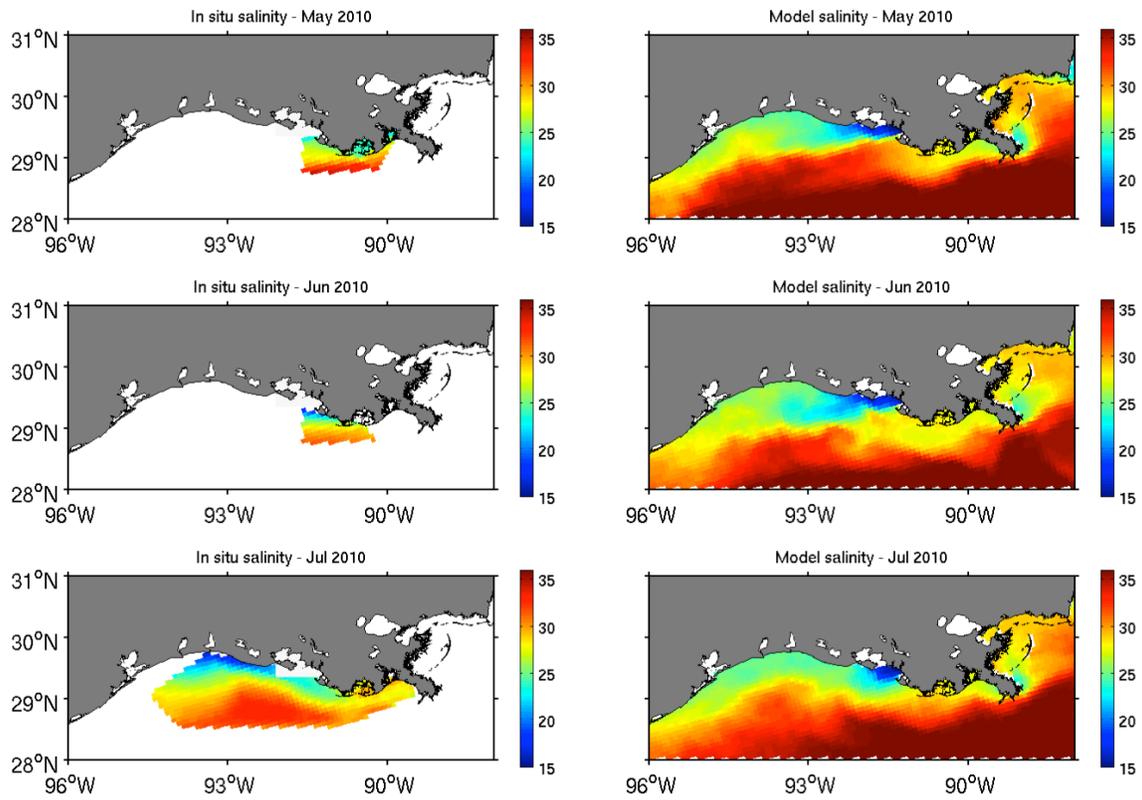


Figure S5: Comparison between in situ (left) and modeled (right) surface salinity during May-July of 2010. In situ observations are from the Louisiana Universities Marine Consortium, available in the Gulf of Mexico Research Initiative (<http://gulfresearchinitiative.org>).

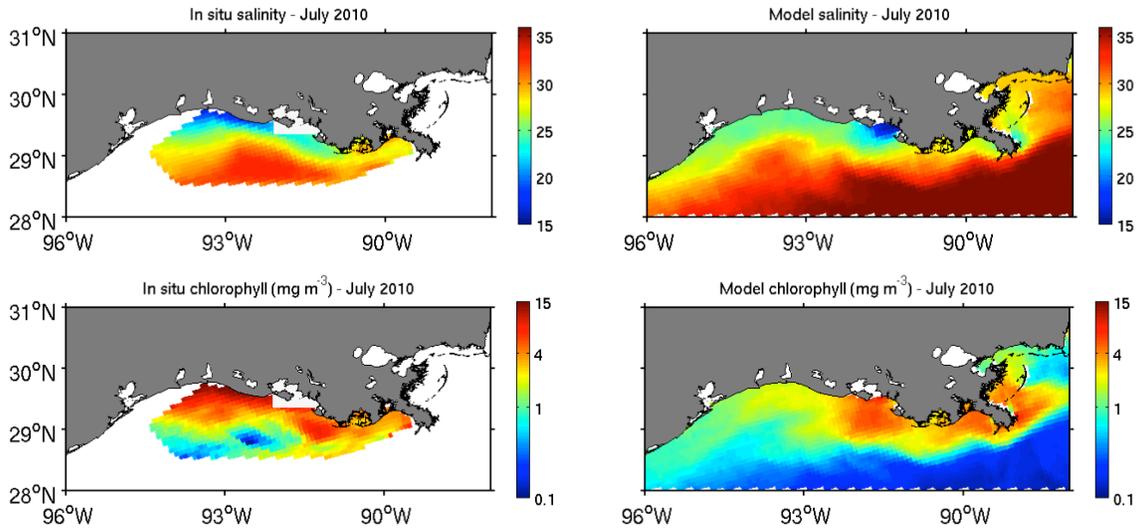


Figure S6: Comparison between in situ and model surface field of salinity (top panels) and chlorophyll (bottom panels) during July of 2010. In situ observations are from the Louisiana Universities Marine Consortium, available in the Gulf of Mexico Research Initiative (<http://gulfresearchinitiative.org>).

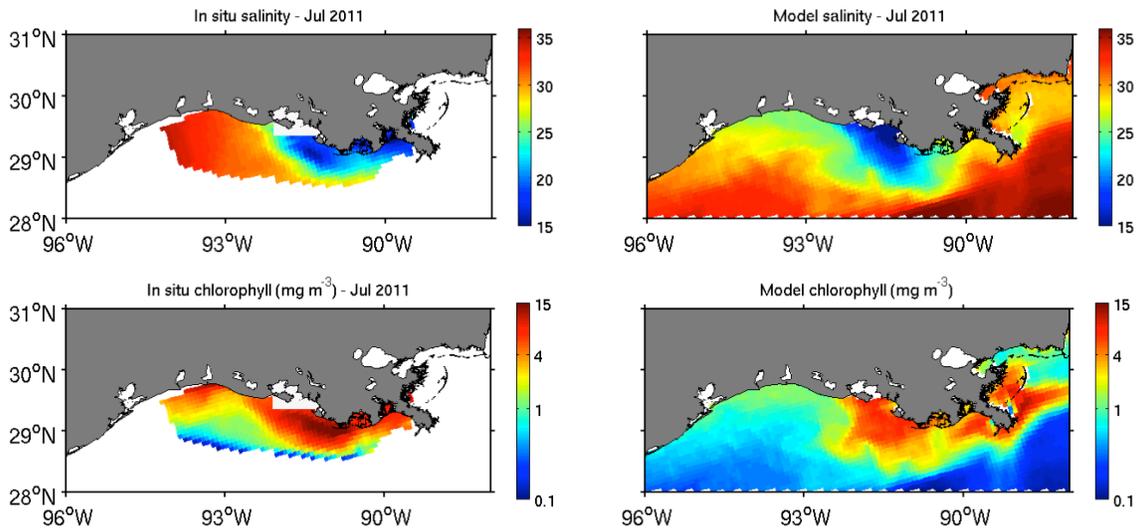


Figure S7: As Fig. 1 but for July of 2011.

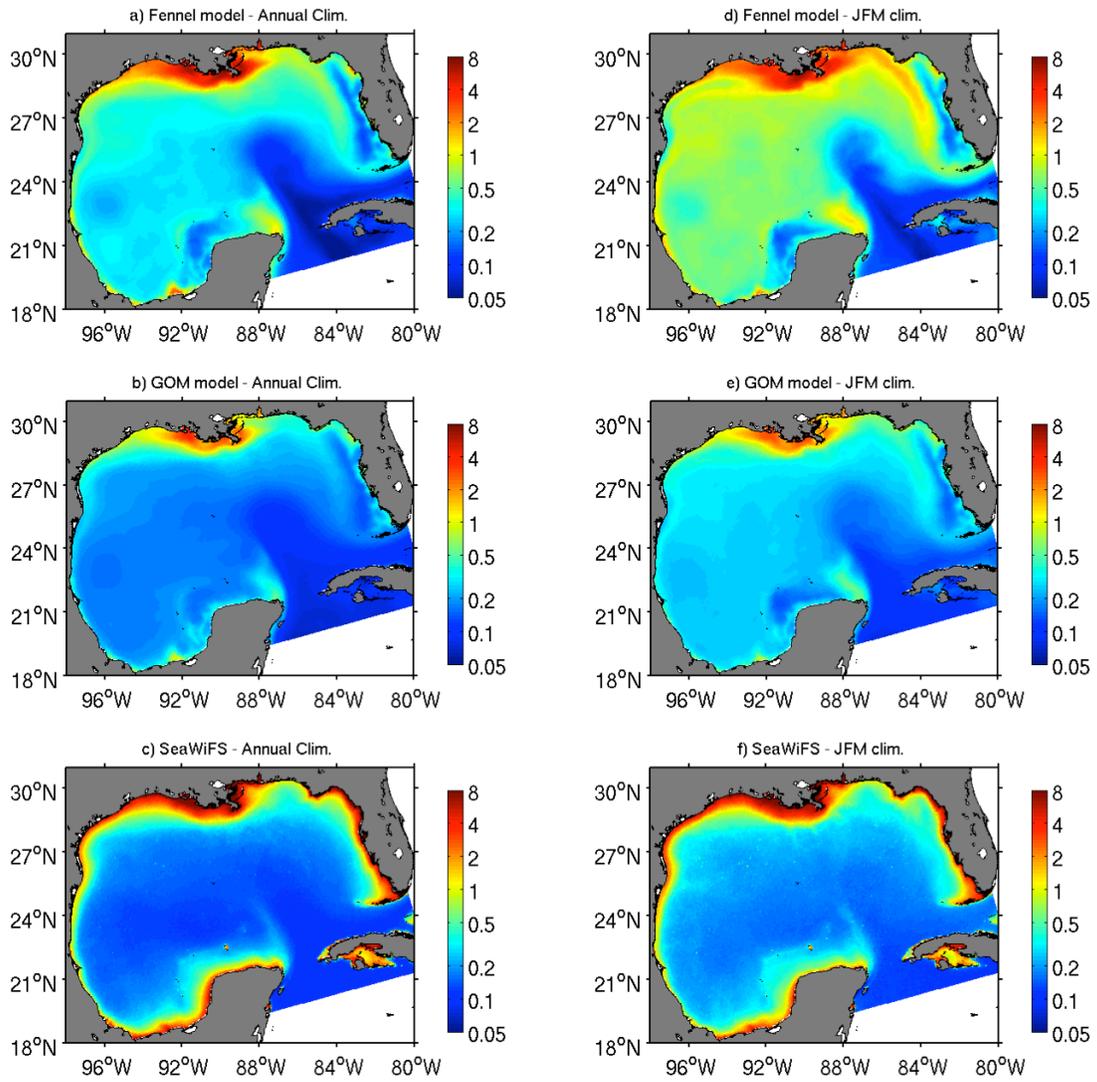


Figure. S8. a-c) Surface chlorophyll climatology derived from Fennel model (a), GOM model (b), and SeaWiFS (c) during 1999-2005. d-f) as in a-c but for Jan-March only.

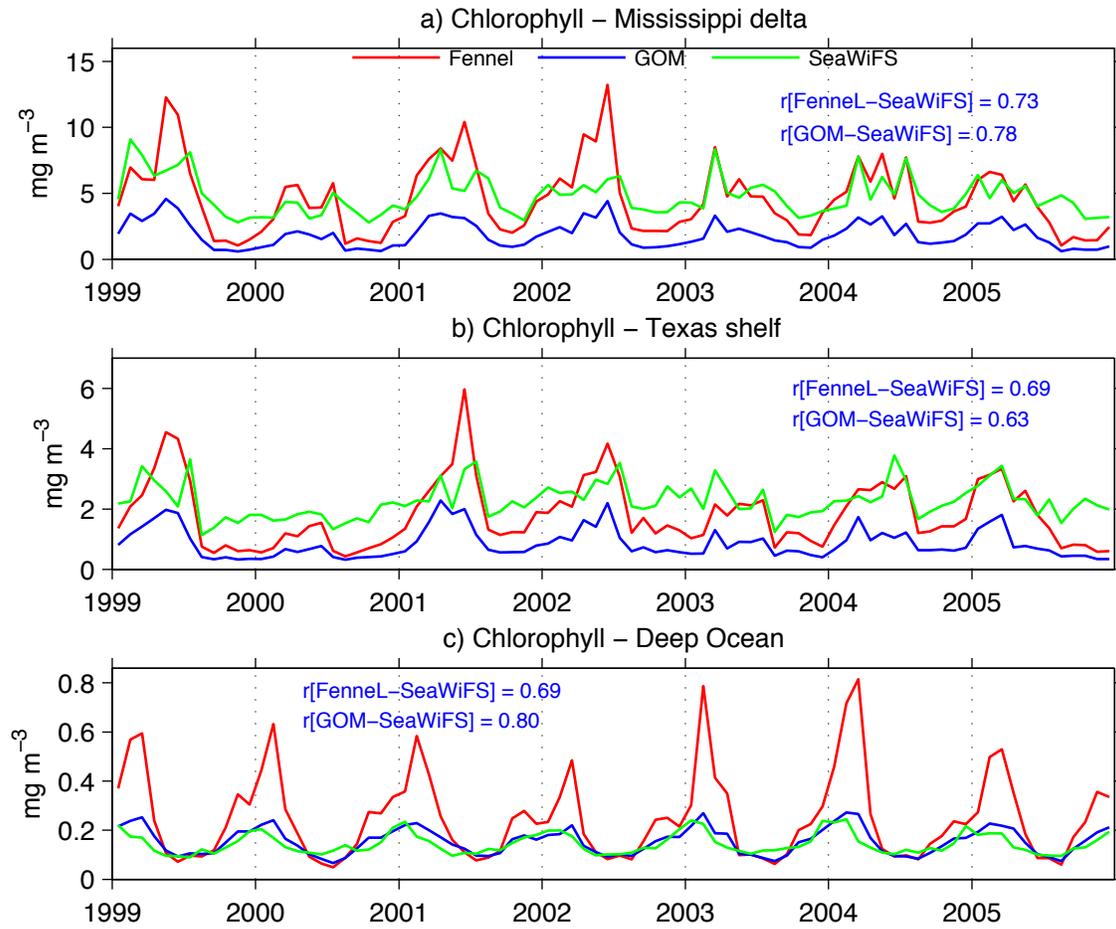


Figure. S9. Monthly surface chlorophyll time series derived from Fennel model, GOM model, and SeaWiFS satellite observations in the northern shelf. Correlation coefficients between model and SeaWiFS time series are indicated at each panel.

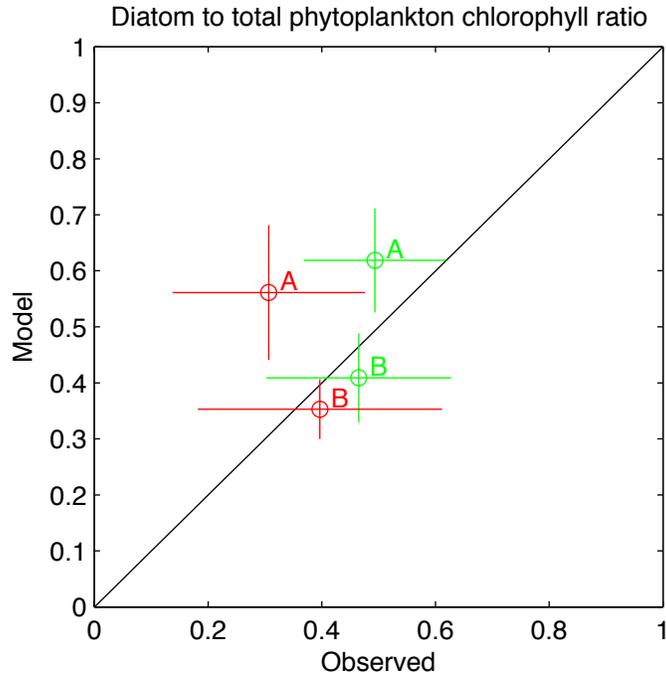


Figure S10. Diatom's chlorophyll to total chlorophyll ratio derived from model and observations. The observed ratios were reported by Zhao and Quigg (2014) for two coastal stations on the Louisiana shelf (stations A and B, located at 29.04°N-89.56°W and 28.59°N-92.00°, respectively) during April (green) and August (red) during 2010-2012. Vertical and horizontal bars depict ± 1 standard deviation.