Author response to C. Hunt review

" This manuscript uses in-situ total alkalinity and physical/biogeochemical measurements (temperature, salinity, chlorophyll fluorescence, and nitrogen) to both test a global equation for alkalinity retrievals (Lee et al. 2006) and to develop localized equations for nine coastal sites. I think this topic is potentially of significant interest to others, and I think the authors have a very nice data set to exploit. However, I think the analysis is somewhat superficial, and I finished the paper wanted a lot more analysis and discussion that what is provided. I will address these points below, but I encourage the authors to further expand their work. The paper is generally well-written, but I list some specific comments which may improve readability."

Response: The authors thank the reviewer for expressing their concerns. This has led to major revisions, which can be viewed in the attached draft manuscript. This is the best way in which we can express such major changes.

Major Comments

"My major reservation about this work is the depth of analysis. This mostly relates to the results from the nine study sites. As noted in the manuscript, these sites are scattered along a very long coastline, distributed across a wide range of latitudes, and presumably represent contrasting conditions from the interactions of offshore ocean water and unique terrestrial and estuarine inputs and transformations. However, except for the discussion of the Yongala site (where the regression results were weakest), little to no information is presented to describe how these sites differ. The one citation referencing the sites (Page 5 Line 5, "Lynch, Morello et al. 2014) appears to be missing from the References. The reason I am left wanting more information about the physical settings is that a number of the regression coefficients are quite similar. By eye, it seems that under Base Model 2 the sites Kangaroo Island, Maria Island, and North Stradbroke Island have nearly the same regression line: is this true? If so, is this coincidence, or are there commonalities between these sites that might explain their similar results? Interestingly, the Ningaloo and Port Hacking Bay Base Model 2 results seem similar to each other, and they are on opposite sides of the continent! Considering Base Model 4, one might group Kangaroo Island, Ningaloo, and Rottnest Island, which is a much different cohort. Is there a statistical way to cluster the sites together to look for spatial trends?"

Response: Thank you for this suggestion. We have undertaken a cluster analysis of the sites to show how they relate and provide a detailed discussion in the revised manuscript. As a result the below figure and recommendations have been incorporated. Please note that a new statistic, mean absolute error (MAE) has been worked into the text. Also the missing reference has been added.

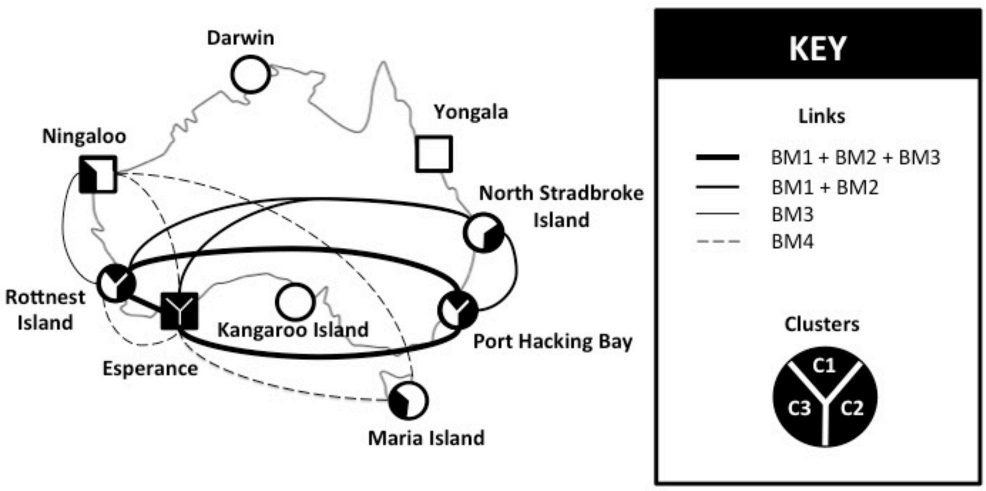

Figure 1: Results from K-S tests, as described in Section 2.5. Links symbolise that the TA distributions at a particular NRS can be modelled by regressions trained from connected NRS, to significantly similar distributions. The shape of the node represents the minimum model of each NRS; circles indicate BM3 is the minimum model, while squared indicate BM4 is the minimum model.

Table 3: Author recommendations for the modelling of TA in the locality of the nine NRS. Recommendations are based on a critical analysis of models using a number of statistical results, as reasoned in the text. Results are presented alongside minimum model predictions (from AIC) and the maximum effect on MAE, when BM2 is instead employed.

NRS	Scale	Recommended BM	Max effect on MAE using BM2	Reasoning
Darwin	Regional	BM3	4.08	Minimum model with agreeing RSE and MAE
Esperance	Regional	BM4	-0.23	Minimum model with agreeing RSE and MAE
Kangaroo Island	Regional	BM4	0.80	Minimum model with agreeing RSE and MAE
Maria Island	Regional	BM4	0.32	RSE and MAE contradict minimum model and indicate that BM4 is the best model
Ningaloo	Regional	BM4	0.06	Minimum model with agreeing RSE and MAE
North Stradbroke Island	Regional	BM3	0.27	Minimum model with agreeing RSE and MAE
Port Hacking Bay	Regional	BM3	0.04	Minimum model with agreeing RSE and MAE
Rottnest Island	Regional	BM3	-0.02	RSE and MAE contradict minimum model and indicate that BM4 is the best model, although the use of BM2-4 is comparable.
Yongala	Regional	BM2	NA	RSE and MAE contradict minimum model and indicate that BM2 is the best model.
C1	Synoptic	BM2	NA	RSE and MAE contradict minimum model and indicate that BM2 is the best model. This makes sense as the three members have different minimum models but are all successfully modeled by BM2 at a regional scale.
C2	Synoptic	BM2	NA	The Esperance NRS displays a different minimum model, so it is not advised to use this cluster at with BM3.

“The authors describe the site-by-site results as ‘regional’, but how far might that region extend around each site? How can these results be applied to locations between the study sites?”

Response: The sampling effort is not designed to answer these questions, so interpolation cannot be performed with confidence. This point is discussed in the discussion. Two interesting studies of particular interest to this question characterise a footprint of the NRS monitoring system on different time scales for different physical variables (Oke and Sakov 2012, Jones et al. 2015). However, as discussed, the role of non-conservative processes which is largely influenced by the benthos and cannot be traced along the distribution of physical parameters. Thus, extrapolation between NRS sites cannot be done with confidence, but these studies will be useful if Australian coastal waters can be divided into zones based on the non-conservative influences of each region.

“One recent paper by Carter et al. discusses a method for interpolating alkalinity data between station which may be helpful. Again, some understanding of what makes the study sites alike or different would help me understand how applicable these equations may be in other places. I also think the authors should read the paper by Alin et al. (2012), which may provide more insight for this work. Those authors also used multiple linear regression techniques to model alkalinity (and other carbonate system parameters) from physical/biogeochemical data at coastal sites.”

Response: The Alin et al. 2012 paper methods cannot be applied to this work, as the spatial resolution of their data is much higher in their region of study. The study area of the Alin et al. (2012) paper corresponds to the footprint scale of one of the NRS stations. The distance in coastline by which they interpolate is not comparable to the region which we would be interpolating in our study. The Alin et al paper has enough spatial resolution to discuss spatial heterogeneity fully and has concluded that it is larger than the sampling resolution so interpolation is valid. Here, we have not been able to reach this conclusion, as the spatial heterogeneity is smaller than the sampling resolution.

Specific Comments

“P2L9-10 the word “threatens/threatened” is repeated in one sentence.”

Response: Changed “Ocean acidification threatens calcifying marine organisms by hindering calcification rates, weakening the structural integrity of coral reefs and other ecosystems”

“P2L12 what are the synoptic scales of interest?”

Response: We have edited the text to include a definition: “The synoptic scales of interest is any scale that includes more than the locality of one NRS”

“P3L2- Define CO2 (and format the subscript)”

Response: Changed

“P3L21-22- the phrase “contribute [to] *a significant [presence of] calcifying fauna” seems pretty awkward”*

Response: Changed

“The World Heritage-listed Ningaloo Reef system and remote reef systems of the Kimberley and Pilbara coasts in Western Australia are other examples of Australia’s vulnerable coral habitats. Elsewhere, sponges, bryozoans, molluscs and crustaceans contribute to a significant presence of vulnerable calcifying fauna, including some commercially significant species of abalone and scallop.”

“P3L26- This line is also pretty awkward”

Response: Changed

“Understanding and quantifying distributions of total alkalinity (TA), the proton deficit of seawater relative to neutrality, is an indication of how much carbon dioxide seawater can hold. Waters with higher TA are less prone to rapid change in ocean pH, as they have a higher proton deficit to “consume” the protons generated from CO_2 uptake, potentially offering refuge for marine biodiversity in the face of OA.”

“P3L29-30- “TA is conservatively related to salinity”- this is an overstatement”

Response: Changed to "Salinity is a conservative tracer within a water mass, meaning that it only experiences changes due to mixing of different water masses or through the addition or removal of freshwater. This property is often exploited through the construction of linear relationships between salinity and TA in a region for the prediction of TA."

"P3L36- don't forget organic matter respiration too"
Response: Changed

"P4L24- again, what are the synoptic scales of interest?"
Response: Addressed

"to predict TA in Australian coastal waters at regional (within the locality of the NRS) and synoptic (algorithms that combine at least 2 NRS) scales."

"P5L23-24- how useful is an integrated phytoplankton biomass over the entire water column, if discrete alkalinity/salinity/temperature pairs are used? Will this affect the statistics, if the same CHL value is used for multiple alkalinity samples? This seems risky."

Response: The authors acknowledge that this is a methodological error. We do have initial analysis performed using measured CHL values rather than integrated CHL values. Please see an updated draft attached in supplementary material. This has significantly changed the results, with respect to the determination of minimum models, but not the over-all message of the paper.

"P5 and P6: the Sections "Linear Regression (LR) Analysis","Open ocean model", and "Statistical analysis" are all numbered 2.2- shouldn't they be 2.2, 2.3, and 2.4?"
Response: Changed

"P6L3- these equations are listed with little in the way of introduction. Can the authors set them up more in the text before listing them? "
Response: Changed, please see below"

LR is well recognised as a useful predictive tool for spatial extrapolation, particularly in comparison to neural networks which are proven to have less predictive power in extrapolation (Lefèvre et al., 2008). Given the goal of enabling predictions of TA in areas of sparse *in situ* measurements, we restricted the range of input variables to those available with broad coverage from satellite Earth observation, namely T, S, and CHL. Additionally, a fourth BM that included nitrate (N) rather than CHL was included for comparison, which can be measured *in situ* using autonomous sensors. This variable choice accommodates the conservative three end-member mixing model presented in Fig.1, in addition to testing for variability due to primary production and other non-conservative coastal processes.

General and regional models for the prediction of TA were constructed from LR analysis using the four base models (BM) shown below and the lm() function in R. General models refer to those derived from a combined dataset collected from all nine NRS. Regional models refer to those derived from data collected from singular NRS. In total, 40 models were derived (the 4 base models applied to 1 general coastal model and 9 regional models).

………(equations)

"P7L10- the term "minimum model" is a little confusing to me. It makes me think this is the minimum set of input parameters needed to accurately estimate alkalinity. Perhaps this is a statistical term I am not familiar enough with, but it seems the minimum model is just the one with the lowest AIC numbers- correct? But it may be perfectly reasonable to still use the other regression models, depending on the input data available and the user's goals."

Response: Correct, the minimum model is the model with the lowest AIC value. The manuscript has been revised to highlight this term more clearly. Again, you are correct in saying that it is perfectly reasonable to still use the other regression models and other parameters which assess robustness should be considered. See adaption below.

«3. The **Akaike Information Criterion (AIC)** measures the relative quality of statistical models and is particularly useful when models with different numbers of variables are being compared. In calculating AIC there is a trade-off between the goodness-of-fit and the complexity of the model, adding an extra level of analysis compared to RSE. The minimum model, the model that minimises information loss, can then be determined as the model with the lowest AIC value. Using AIC values, the **relative probability of minimising information loss (RPMIL)** for each model can also be determined which normalises differences in AIC according to the number of observations collected. This allows a more intuitive and robust method for comparing models, by determining probabilities that another model is actually the minimum model given infinite data points were collected.»

"P8- This section is so brief, it feels somewhat like an afterthought. Have the authors considered combining this with the next section into a dual Results and Discussion section? This might result in better flow from topic to topic."
Response: The authors have edited the manuscript for flow. There is more analysis in the results section, and we hope that it has addressed this point.

"P9L10- again, the paper by Alin et al. (2012) and other related works undermine this argument that little coastal regression work has been done."
Response: Perhaps this is an overstatement. The text has been edited to address this.

"P9L14-29- much of this material seems like it should be in the Introduction or perhaps Data and Methods sections- it seems out of place here."
Response: Changed. We have moved discussion on variable choice and transformation to the methods section.

"P9L39- are the "decreases" described here decreasing AIC values? Unclear from the text."
Response: Changed for clarity

"P10L37-40. This argument seems a little shaky. Seasonality in river discharge would also result in a salinity seasonality as well, not just in alkalinity. Are the authors implying that the alkalinity concentrations in river discharge vary seasonally? If so, what is the mechanism for this? Also, the authors state that this seasonality cannot be measured by remote sensing, but isn't one of the prospects held out by this paper the potential of new remotely-sensed salinity products? Why would remotely sensed salinity not show this seasonality?"

Response: As mentioned in the text, the residuals at this station do coincide with a freshening in salinity, which is what lead us to the conclusion that the river input was the cause of error. At the Yongala NRS, there are mixing processes changing TA that are being explained by salinity, this produces the TA-S linear mixing line that has a freshwater endpoint and an oceanic endpoint. Perturbation of the system by some riverine derived variability in TA adds another direction of change to the TA, moving the intercept of the mixing line up and down but keeping the oceanic end-member constant. Thus you need an extra variable that captures this change because it is deviant from the TA-S mixing line (see below figure which has now been worked into the text). Temperature cannot be used to account for all the variability in riverine input, in this case, as the changes seen are not occurring on an annual seasonal cycle, but rather appear to be changing inter-annually. The analysis we ran shows that neither CHL nor NO3 can capture this seasonality, and it is still evident in the residuals. Upon further investigation the trends seen in the Yongala NRS residuals coincide with major flooding events, and hare highest in a period of high river flow (Logh et al. 2015) . The massive perturbations experienced by these events is a second, larger mode of variability compared to seasonal summer/winter changes in riverine input. Additionally, to answer your question, larger changes in TA would be seen in a flooding event, compared to salinity, so salinity cannot fully capture such a change. This consideration had been re-worked into the discussion.

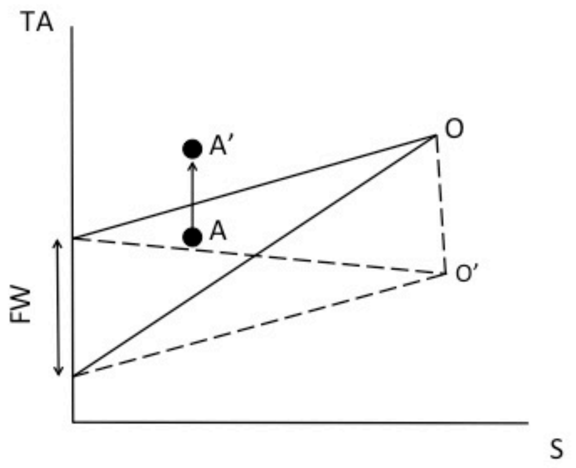

Figure 2: A depiction of a two end-member mixing model that contains an open ocean end-member (O) and a variable fresh-water end member (FW). Solid lines indicate different conservative mixing lines. Point A lies in the conservative mixing region, however an arrow indicates how it can be perturbed away from conservative mixing predictions by a non-conservative change in TA to point A'. Thus, there are three distinct modes of variability; FW variability, conservative mixing, and non-conservative changes in TA. The mixing of two oceanic end members in coastal regions it also very likely, further complicating the problem, and extending the region of variability, indicated by the addition of O' in the model, and associated mixing lines (dashed lines).

"P11L18- change to "in the future""
Response: Changed

"P11L23-38- Figure 6 might undermine several of this paper's points. For one, it only uses the general coastal model, since the authors did not attempt to interpolate their regional results over the entire domain. Also, Figure 6 shows that the actual locations of the stations where data were collected are masked out by a land mask. How sure are the authors that applying this coastal model is appropriate, given that the model data were collected in an area where remotely sensed measurements are not even available. I acknowledge that it is good to show the potential application of these relationships, but this raises the question of site selection in this study."
Response: This figure has been removed from the manuscript

"P13L2- Again, I am confused about the use of the term "minimum" Previously the authors state that at least temperature should be included with salinity in these equations. However, this seems to contradict that recommendation by saying that the inclusion of salinity, temperature, and CHL or N are the minimum. Also, can the authors quantify, perhaps in terms of umol alkalinity error, how much better it is to include CHL or N?"
Response: Changed "minimum sets of variables" to "minimum models". This information can be found by comparing RSE and MAE across the two models. The results show that the robustness of each model, comparatively, it is regionally dependent. We have clarified our recommendations, as presented above, with a summarised reasoning for each conclusion, as presented in the text.

“Figure 1- what is the inclusion of the 1000m isobath intended to show? Why 1000m?”
Response: The isobath was included to indicate the shape of the coastal zone and continental shelf.

“Figure 2- the text in this figure is very small, and might not show up well in the final version”
Response: Changed

“Figure 3 and Figure 5- please insert a space into the µmol kg^{-1} labels”
Response: Changed

“Figure 5- the caption says results from four models are shown, when only three are shown”
Response: Changed

“Table 1- I’d appreciate another table, perhaps describing the salinity, temperature, alkalinity, CHL and N data for each site. Perhaps just some basic statistics such as range, mean, standard deviation etc.”
Response: This has been added to the manuscript. See below:

Table 1: Latitude and mean distributions of parameters at each NRS. Means are presented for each variable with associated standard deviations

NRS	Lattitude	TA	S	T	CHL	N
Darwin	-12.4	2265 (40)	34.07(0.68)	28.78(2.53)	0.688(0.687)	0.590(0.348)
Esperance	-33.9333	2337 (11)	35.62(0.14)	18.26(1.69)	0.267(0.058)	0.378(0.155)
Kangaroo Island	-35.8322	2355(11)	35.84(0.19)	16.78(1.57)	1.083(1.561)	0.324(0.197)
Maria Island	-42.5967	2324 (7)	35.31(0.15)	14.49(2.02)	2.369(1.316)	0.666(0.492)
Ningaloo	-21.99	2281(8)	34.80(0.13)	25.77(2.41)	0.514(0.449)	0.380(0.141)
North Stradbroke Island	-27.345	2324(12)	35.48(0.21)	22.56(2.45)	2.917(2.432)	0.263(0.313)
Port Hacking Bay	-34.1192	2326(9)	35.47(0.13)	18.92(2.02)	2.439(2.271)	0.688(0.408)
Rottnest Island	-32	2327(14)	35.49(0.22)	20.73(1.50)	0.514(0.328)	0.306(0.125)
Yongala	-19.3085	2296(32)	35.19(0.60)	25.86(2.37)	0.308(0.206)	0.248(0.145)

“Figure S1- Might this be better shown in a table? Or at least could this information be briefly described in the text? This figure seems a bit superfluous to include.”

Response: Changed. We have presented the study period as a range in table form as suggested.
“Table S1-S4- while these tables contain a lot of information, they are also really the heart of the paper’s analysis. Seems a little strange to bury them in the Supplementary Information.”

Response: The key message of the paper is not the regression models themselves, but rather their ability to predict TA at regional and continental scales. It was considered to include tables in the text, however it was decided that the parameters would only be considered and used by a minority of readers. Thus, AIC values and

RSE values were chosen to be better displayed in the text as these are the parameters required to formulate the results of the paper.

"Also, do the terms Intercept, S, T etc. in these tables refer to the terms d, a, b respectively in the model equations listed on Page 6 of the main text? If so, please use consistent terms between the two."

Response: Changed.

References

Jones, E.M., Doblin, M.A., Matear, R. and King, E.: Assessing and evaluating the ocean-colour footprint of a regional observing system, *J. Marine Syst.*, *143*, pp.49-61, 2015.

Lough, J.M., Lewis, S.E. and Cantin, N.E.: Freshwater impacts in the central Great Barrier Reef: 1648–2011. Coral Reefs, 34(3), 739-751, 2015.

Oke, P.R. and Sakov, P.: Assessing the footprint of a regional ocean observing system. *J. Marine Syst.*, *105*, 30-51, 2012.