Referee # 1

Major Comments

Part of the paper is based on the analysis of long-term data-sets from the USGS, going back to the 1940's. The authors compare their own data with the recent USGS (Figure 7), which validates the quality of the recent USGS data. But this does not necessarily mean that the old data are of the same quality, meaning the derived trends over the decades could be methodological. Please add in the discussion, some elements on the methods of TA analysis, data quality check, and any other element that might be useful to show that over the last 70 years the USGS data-set is of uniform quality and that the observed changes are real rather than methodological.

Response:

1) Good point. Frequently, the methodology of scientific methods change over time. To examine this issue, we compared current USGS protocol for alkalinity measurements to previous USGS guidelines and practices to evaluate the impact that methodological changes might have on observed alkalinity trends. Thus, we have added a second paragraph to section '4.3 Historical trends in estuarine alkalinity' as shown below:

"While numerous studies across the world indicate a shift towards increasing alkalinity in estuarine waters, the impact of methodological changes on measured values over time cannot be neglected. Conveniently, USGS has published a series of manuals, both past and present, discussing the analytical procedures and methods followed during specialized work in water resources investigations (Woods, 1976; Fishman et al., 1989; Radke et al., 1998). Historically, the USGS measured alkalinity as fixed endpoint titrations on unfiltered samples, and commonly reported values as concentrations of bicarbonate (Clarke, 1924). By 1984, the USGS also began conducting fixed endpoint and incremental titrations on filtered samples (Raymond et al., 2009; Kaushal et al., 2013). Presently, USGS performs several variations of tests that describe the alkalinity of water samples including standard alkalinity, acid neutralizing capacity, and carbonate alkalinity. Samples are measured using either a standard buret, micrometer buret, or by an automated digital titrator (Fishman et al., 1989; Radke et al., 1998). Micrometer burets offer higher accuracy and precision than standard burets while automated titrators are more preferred due to convenience and durability (Radke et al., 1998). Fixed endpoint titrations are generally less accurate than inflection point titrations, especially in low carbonate waters or areas with high organic and noncarbonated contributions to alkalinity (Radtke et al., 2008). Such methodological changes, however, would result in an underestimate of alkalinity if there is any (Kaushal et al., 2013). Thus, our conclusion of an increasing alkalinity trend in the Delaware River water will still hold and can be a conservative estimate. Such alkalinity increase has been observed throughout many river and estuarine systems (Raymond et al., 2003; Raymond et al., 2009; Duarte et al., 2013; Kaushal et al., 2013; Stets et al., 2014)."

2) In addition, we described and expanded on the specific USGS alkalinity parameter codes used to clarify water quality data and analytical procedures. We added Table 4 to show the exact *parameter codes used during this analysis. The following section was added to the first paragraph in section 4.3.*

"The extensive and routine collection of water samples conducted by USGS allows us to explore long term trends in alkalinity (from the mid-20th to early 21st century) in the Delaware and Schuylkill rivers (USGS stations 01463500 and 01474500, respectively). For USGS alkalinity values, we use similar approaches as conducted in Stets et al., 2014. We combine 8 various parameter codes that include alkalinity, acid neutralizing capacity (ANC), or HCO_3^- (Table 4). Alkalinity and ANC follow identical electrometric procedures except that alkalinity samples are filtered while ANC samples are not."

Table 4. USGS parameter codes used during analysis

Parameter Code	Parameter Description	Total Count	Percentage of Total Count
00410	Acid neutralizing capacity, water, unfiltered, fixed endpoint titration, field	920	28.5
00419	Acid neutralizing capacity, water, unfiltered, inflection-point titration, field	25	0.8
00440	Bicarbonate, water, unfiltered, fixed endpoint titration, field	1529	47.4
00450	Bicarbonate, water, unfiltered, inflection-point titration, field	25	0.8
00453	Bicarbonate, water, filtered, inflection-point titration, field	86	2.7
29801	Alkalinity, water, filtered, fixed endpoint titration, laboratory	133	4.1
39086	Alkalinity, water, filtered, inflection-point titration, field	283	8.8
90410	Acid neutralizing capacity, water, unfiltered, fixed endpoint titration, laboratory	224	6.9

Minor Comments

P2 L 16: to the list of processes that control CO_2 in rivers, you could mention inputs from wetlands (Abril et al. 2014).

Response: Agreed. We have added wetlands to the list of controlling processes as shown below.

"The majority of carbon fluxes in inland waters involve inputs from soil-derived carbon, chemical weathering of carbonate and silicate minerals, wetlands, dissolved carbon in sewage waste, and organic carbon produced by phytoplankton in surface waters (Battin et al., 2009; Tranvik et al., 2009; Regnier et al., 2013; Abril et al., 2014)."

P3 L 10: I assume that this statement is based on some sort of analysis of numbers, so could you please state the range and central value of the area and the residence time of the estuaries from the cited studies.

Response: Yes, we used the estuarine classification groups as described in Dürr et al., 2011, where surface areas were estimated based on geographical information system analysis. Further, Borges and Abril et al., 2011 listed the criteria for each of the estuarine classification types. We have added the ranges to clarify the differentiation between typical "large" and "small" estuarine systems as shown below.

"Further, the majority of past estuarine CO_2 studies have focused primarily on small estuarine systems (typically within 1 – 100 km in length and less than 10 m in depth) with rapid freshwater residence times (10^{-3} – 10^{-1} yr) (Chen and Borges, 2009; Cai, 2011; Borges and Abril, 2011, Dürr et al., 2011)."

P3 L 11: Please define the criteria (threshold value?) and quantity (surface area? discharge? drainage areas? Length?) to distinguish "large" and "small" estuaries.

Response: Please see above.

P7 L 12-13: The correlation between TA fluxes and discharge is due to auto-correlation. If you plot AxB versus B, you'll always generate a good correlation (Berges 1997), especially if B changes over several orders of magnitude (unlike A).

Response: We are aware of this issue and its potential concern. As TA flux is defined as concentration multiplied by discharge, one would expect a solid correlation between the two variables. We offer the following explanation to clarify our approach. Here, in Fig. 4 we plot TA against river discharge to determine the correlation between the two variables (strong negative correlation). We used this relationship paired with high frequency USGS discharge records to estimate high resolution TA at the river end-members. Then, we calculated the annual TA flux. We did not directly use $F = TA*Q$ to plot against Q to determine the annual TA flux. We believe the difference between our approach and what the reviewer has mentioned as auto-correlation is that our statistically significant correlations in Fig. 4 is between TA and discharge and not between flux and discharge (subsequently the regression is not directly driven by discharge, though geochemically they are related).

P 10 L 25: The finding that intertidal marshes have little influence on the CO_2 dynamics of the Delaware is quite interesting and would contradict the main conclusion and among the opening statements of the Cai (2011) paper: "It is demonstrated here that CO_2 release in estuaries is largely supported by microbial decomposition of highly productive intertidal marsh biomass".

Response: Agreed. In our revision, we further expand on the interesting contrasts with the small southern estuaries emphasized in Cai (2011). The impact of intertidal marshes on estuarine CO_2 dynamics can be significant, particularly in small estuarine systems. In this study, we did not sample the sub-estuaries within nor areas near the perimeters of the bay, but instead were limited to sampling within the main channel of the estuary. We note while the Delaware River is *only a medium size river, the Delaware Bay is one of the largest bays in the U.S. eastern coast and its hydrodynamics is largely controlled by the exchange with the ocean (residence time of 1-3 months). Previous studies that conducted cross bay transects, sampling at various depths, over diel cycles, and along tributaries, found that except near the shoreline where suspended sediment and chlorophyll concentrations were high, general cross-bay gradients were erratic and comparatively small (Culberson et al., 1987; Lebo et al., 1990; Sharp et al., 2009). While significantly more research and data are needed, we suggest that due to the much broader geographical size of the Delaware Bay, that except near shallow waters the flushing of intertidal marshes has a minor impact on overall surface water* pCO_2 *and* CO_2 *flux dynamics in the system (as opposed to in small estuaries where marshes may have significant impacts on estuarine* CO_2 *degassing fluxes).*

P 13 L 6: This discussion seems to contradict the Introduction (P3 L10) that previously studied estuaries have a "short residence time"

Response: We see why this might illustrate contrasting ideas and perhaps we need to restructure/reemphasize key points to clarify the main objective of this paragraph. In the Introduction, we stated that most of the previously studied estuaries have a "short residence time" but the Scheldt is an exception. Here, we want to contrast the differences in ecosystem metabolism between estuaries with short versus long residence times. The Scheldt Estuary has a long, freshwater residence time similar to that of the Delaware Bay. We would like to use prior metabolic findings in the Scheldt Estuary to serve as a model for net ecosystem metabolism in the Delaware estuary and potentially other large estuarine systems with long residence times. We revised our text to make this point clear.

"Unlike in most previously studied estuaries, freshwater residence times in the Scheldt Estuary and Delaware Bay are generally long ranging from about one to a few months (Gay and O'Donnell, 2009; Borges and Abril, 2011) … Thus, we suspect that in other estuarine systems with long freshwater residence times (i.e. the Delaware Estuary), much of the DIC produced by NEP is most likely removed to the atmosphere rather than exported to the sea."

Figure 1: axis legends have a different font from all of the other figures, it is advisable to have a uniform font in all figures.

Response: Good point. We have revised it so that all fonts are uniform.

Figure 4: Two decimals for R^2 are sufficient. In some figures, numbers in axis legend have thousands separated by comma, but not in others. It is advisable to make this uniform. In some figures, the axis name is "alkalinity", in others it is "TA". It is advisable to make this uniform.

Response: We have reduced to two decimal points for R^2 *values. We also fixed the number issue in the axis legends and used 'alkalinity' as the uniform axis name in all figures.*

Figures 6 and 7: It is odd that one of the data-sets is named after one of the authors ("Cai"), I suggest that the data set is named "this study", something neutral and a bit more modest.

Response: Agreed. We have changed it to "This Study".