

Interactive comment on “Climatological variations of total alkalinity and total inorganic carbon in the Mediterranean Sea surface waters” by E. Gemayel et al.

Anonymous Referee #1

Received and published: 20 September 2015

Gemayel et al. present an interesting study regarding the sea surface total alkalinity and total inorganic carbon in the mediterranean sea. To-date our knowledge regarding the carbonate system is limited due to the sparsity of available observations, hence I very much appreciate the effort of the authors to gather available observations and perform this basin scale study. The authors investigate the spatial distribution as well as seasonal variability and nicely explain their findings.

The manuscript is well structured, well written and nicely relates the findings of this manuscript to previous studies. I do however believe that the authors need to substantially improve their currently too short methods section. Please find specific points below in the major and minor comments sections.

C564

Major comments:

The authors need to add more detail regarding the 10-fold cross validation technique. E.g: how are the subsets (training, testing) chosen? randomly?

How is the data distribution between training and testing established? On page 1504 last line the authors report 375 truing data and 115 testing data for the total alkalinity and on page 1505 lines 1-2 they report 381 training data and 45 testing data. I struggle to understand how these numbers add up? Are the distributions between training and testing data different for alkalinity and total inorganic carbon?

The authors report that the algorithm was applied for polynomials of 1-3 (page 1504 line 19), however the authors do not explain why? Would it not be possible that a 4th order polynomial could further improve the total inorganic carbon fit?

The authors use the established relationships to estimate alkalinity and DIC where there are no data, hence it is important to show that the algorithm does not overfit the data but is capable of extrapolating data, which is currently only partly done. E.g. one sign of overfitting would be if there is a substantial difference between the RMSE and mean difference between the residuals of the training set compared to the testing set. A table would help to illustrate this. Furthermore, it is somehow worrisome that the different algorithms from table 3 lead to such different results, as they are all developed for different regions, but do not seem to have a good predictable power in the Mediterranean.

Minor comments:

I was very confused to see a reference to equation 1 in the text but I could not find the equation in the text, but rather had to look for it in table 1. It would help the reader if you could put equations in the text

Please clarify what you mean by summer and winter? E.g: is summer the average of the months of June, July and August?

C565

On page 1507 line 13 the authors mention the effect of biology, however, biology is not included in the polynomial fit. Why? You could e.g. use satellite derived biological proxies.

Page 1507 line 6: "... presents a significant improvement ... " please provide some information on how the significance has been tested.

Interactive comment on Earth Syst. Dynam. Discuss., 6, 1499, 2015.