RC C8075 by H. Pearl (Referee) hans_paerl@unc.edu Received and published: 28 November 2015

In this study, the authors clearly and convincingly demonstrate, using remote seining data, that climatic changes with resultant impacts on water column dynamics (temperature, wind speed, transparency) that have occurred since the late 1970's-early 1980's in the Baltic Sea region have led to significant changes in ecology, physiology and overall trophic conditions of the Baltic Sea proper. In particular, increases in temperature, combined with longer warm periods, coupled to decreases in wind speed shave significantly widened the "window" for surface-dwelling cyanobacterial blooms in the Baltic Sea. Interestingly, turbidity has increased as well during this period, most likely as a result of higher cyanobacterial production rates. In some respects this represents a positive feedback on the cyanobacterial populations as well, because as surface dwelling populations, they can circumvent higher turbidity in the water column and as a results this favors their competitive capabilities relative to eukaryotic phytoplankton communities. These are all highly significant and important findings that are worth publishing. The manuscript is well-written, data were thoroughly analyzed and (where possible) statistically tested. The results are clearly and convincingly discussed in the context of impacts of climate change on cyanobacterial bloom physiology and ecology. the results also have considerable management implications. In the art line of the abstract the authors state: "The seasonality of a number of abiotic and biotic variables for which data are available has changed drastically during the last few decades. These changes are likely to have major effects on many aspects of Baltic ecosystems." What "ecosystems" are the authors referring to, or are that eferring to the Baltic Sea "ecosystem" as a whole? P. 12, lines 10-20, the authors state: "The changes in the timing of Ked490 and CHL are related and reflect the increased turbidity and decreased water transparency in the Baltic Sea. An obvious consequence of the increased Ked490 is that less light reaches depths below the surface. While this analysis was done for the central Baltic with small areas of benthic photosynthesis, we can assume that benthic communities in the coastal areas must also be experiencing significantly reduction in light due to the decreased water transparency. We can hypothesize that the increased turbidity and decreased water transparency are related to increased phytoplankton concentrations and increased bacterial production. Likely effects on the rest of the ecosystem including commercially important fisheries should be further evaluated.' The fact that dominant cyanobacterial bloom genera in the Baltic Sea (i.e. Anabeana, Aphanizomenon, Microcystis are highly buoyant may play a relevant role in their ability to overcome recent higher levels of water column turbidity (much of it exerted by the blooms). This should probably be emphasized more directly here. Otherwise, I have no major issues with or corrections to add to the paper and recommend publication with minor revisions. This manuscript will be broadly appreciated!

Thanks for these encouraging comments. We agree on the "ecosystems vs. ecosystem" and have changed several cases of "ecosystems" to "the Baltic Sea ecosystem". Thanks for this interesting idea on the possible relationship between turbidity and the buoyancy regulation in cyanobacteria. We added a sentence "That dominant cyanobacterial genera in the Baltic Sea, especially Aphanizomenon and Nodularia, are buoyant may both contribute to the higher turbidity and favor them over non-buoyant phytoplankton."

RC C8292 by H. Pearl (Referee) hans_paerl@unc.edu Received and published: 6 December 2015

Once the authors have had a chase to revise the manuscript according to my and perhaps other comments, it should be published.

Agreed.

RC C8624 by R. Stumpf ; Received and published: 21 December 2015

This paper, "Changing seasonality of the Baltic Sea", is an excellent complement to the authors' previous work on cyanobacteria (Biogeosciences, 2014). The evaluation of the various data sets shows changes in the system, some indicative of trends, some may indicate "state" changes, and some may suggest short term aberrations. I trust that the data presented in the figures will be available as supplements to this paper, as much analysis could be done with them. The paper is an important contribution to the science and an excellent demonstration of how to use remotely sensed environmental data to address an ecological problem. One interesting example is that the irradiance (figure 3) shows the longest delays in the 1980s (Figure 3), when warming (SST) appeared most delayed (Figure 4). This is expected when we see it, but it indicates some interesting questions for climate change.
We agree.

If climate change alters cloud cover, will the SST change more or vary more? While satellite chlorophyll is problematic, which the authors clearly and carefully discuss, the data in Figure 9 shows a striking pattern of timing, particularly relative to the cyano blooms (Figure 10). Figure 9 uses field data and satellite data to show timing changes that are quite large and unlikely to be explained by simple uncertainty in the data. The authors present a combined data set to argue convincingly that the change in timing of the cyanos is likely due to changes in SST.
Thank you for this assessment.

On the question of trend vs "state change". The SST phenology could argue either way. After 1996, the Baltic warmed to 12C always before day 170, where until that year 1/3 of the years did not see 12C until after day 170. Similarly 17C was reached before day 195 all but 2 years after 1996, but the warming was frequently later than that before 1996. This could be a trend, and I would not delete the trend line.
The same thing occurs at the end of season. In all but 2 years after 1996, the Baltic did not cool below 17 until after day 230, and more dramatically, warm water persisted until after day 245, which never occurred before 1996. Same thing with 12C, day 275 was not reached until 2000 (and not any of the previous years). Is the Baltic simply warming, or did the the system shift in 1996? The result is the same, but a provocative question.
We thank the reviewer for encouraging comments and for the intriguing idea of separating between a trend and a state change. We are not aware of any statistical tools how to distinguish these two but it is certainly an interesting problem to look into in the future. For the time being we are looking for trends in the data.

Some Technical comments. Switching between satellites should be noted, while they don't appear to have an effect, except for the change from AVHRR to SeaWiFS for bloom area. (I looked first for

discontinuities at years when satellites would change, and didn't see anything obvious.). The SST change could possibly change, but the Pathfinder set does cross-match AVHRR and MODIS rather meticulously. Still there is the possibility that satellite algorithms may cause some changes.
As the reviewer mentions, switching between different satellite sensors is a major technical problem in interpreting time series including multiple sensors. This was the central topic in our previous paper (Kahru and Elmgren, 2014) that consolidated results from a series of different satellite sensors. As we showed in that paper, switching from the low-sensitivity AVHRR sensors to the high-sensitivity ocean color sensors involves some unavoidable differences between the detected cyanobacteria accumulations, particularly in detecting weak accumulations. However, the main interannual differences are caused by major accumulations that are dense and well defined. We were able to show that the processed time series of FCA is consistent between different sensors. With SST the impact of the sensor to sensor differences is expected to be much smaller as the sensors are much more similar. Considering that the changes in SST cover the range of many degrees C, the observed changes are therefore much bigger that the probable difference between different sensors.

The one outstanding point is that uncertainties in timing are not discussed or estimated. This is important for the statistical confidence, particularly as this paper will become a reference on some of these methods. Clouds are the obvious factor in uncertainty. Mostly overcast conditions can lead to delays in the day of interest (e.g., the date that the threshold was passed may have been overcast). Partly cloudy can also cause an error. For example, if clouds cover the southern part of the Baltic for a day or more, the averaged temperature will be biased to the northern Baltic, and may be colder, leading to a delay in the warming temperature threshold. The error may be slightly larger for AVHRR, but the authors should be able to provide some discussion on this. One key reason to capture the uncertainty is to determine trend slopes that are real. The slope should be greater than that caused by the uncertainty. For example, on Figure 5, if the error is +/- 3 days, than the slope needs to be at least twice that to be real (so more than +/- 0.2 day/year over the 30 years). That is a perfectly reasonable way to present whether the trend is real. A rigorous assessment of uncertainty in date is not needed, just a brief description of what that might be, as many trends are much larger than those uncertainties. Slope confidence, by the way, is captured by Figure 3b and Figure 5. These figures provide good metrics of confidence in the slope estimates. The consistency across the
plots indicate that the slope errors are probably small. There are numerous ways to estimate slope confidence (jackknife, Bayseian, etc.), but that would be an exhausting point, and use of these plots is reasonable.
The uncertainty in the timing estimates is an important and complex question. We have thought a lot about that topic and even designed some Monte Carlo experiments to create statistical estimates. The uncertainty in timing is a function of many variables and parameters like the uncertainty of a single measurement, the uncertainty of the spatially (partially) averaged measurement, the uncertainty in timing dependent on the speed of change in the variable, etc. At this moment this work is a still in early planning. Considering that most changes in timing were quite drastic and follow a clear pattern in time and as a function of the threshold value, we are confident that the uncertainties are much smaller than the detected change and that the detected trends are real. We have discussed the uncertainty problem in the revised manuscript.

P-values, by the way, are not valid for these data sets and should be replaced with estimates of significant slope based on uncertainties or scientific significance (like the previous paragraphs). Null hypothesis P is the probability that a sample taken from a population that has no trend will return a trend at least as large as what was observed. This paper is working with populations, so P has no value. While P looks like it is telling us what we want to know, it is not. Nicholls 2001 (Bulletin Amer. Meteor. Society) describes the problem nicely. (In spite of what we've been led to believe, P-values are rarely appropriate for any data set, and a few journals have even decided to refuse to publish P-values, preferring other metrics).

The trend on Figure 8 (chlorophyll) is a bit dodgy, and probably better to leave off. There is a trend in the lowest observed winter "chlorophyll", but an overall trend will require a different statistic, like a seasonal Theil-Sen slope (to address the seasonality). That's not necessary.

We appreciate the reviewer's informative criticism on our use of P-values in our statistical analysis. We have left them out in the revised manuscript. Instead, we use the nonparametric Sen slope and significance estimates of the slope of the least squares linear regression. We added the following text to the Methods: "The existence of trends and their significance was evaluated with the nonparametric Sen slope (Sen, 1968) and the Mann-Kendall test using 95% significance level (Salmi et al., 2002). In parallel, 95% confidence limits of the least squares linear regression slope (as implemented in NMath numerical libraries, http://www.centerspace.net) were also used."

The paper sets the stage for environmental modeling of the cyano blooms in the Baltic. I hope the authors tackle that.

We have added a discussion of the use of coupled biogeochemical-physical models to evaluate some of the observed features. The models discussed support our observations and show similar patterns of changing phenology. However, the changes in the models are driven primarily by temperature. Currently the models discussed have relatively primitive structure compared to the complexities in intra-species life cycles and inter-species interactions. However, some newer models are beginning to include aspects of those. We are planning such model experiments for the future. We have added the following paragraph: "Even though the Baltic Sea is one of the marine areas in the world best covered by observations, the frequency of long-term sampling is insufficient for a confident detection of similar phenological indicators using measurements at sea. Instead, comparisons can be made with mathematical models simulating ecosystem dynamics over decades with high temporal resolution (*e.g.* Eilola *et al*., 2011; Hense *et al*., 2013; Meier *et al*., 2012). For instance, BALTSEM (BAltic Long-Term large Scale Eutrophication Model, Savchuk et al., 2012) produces quite similar tendencies in the Central Baltic Sea (Fig. 1b). The prolongation of the vegetative season from about 190 days in the beginning of 1970s to about 230 days in the middle of 2010s has been accompanied by a tripling of the net primary production and shift of the annual biomass maximum from spring to summer. Due to earlier warming and delayed cooling, the duration of the period with simulated surface water temperature exceeding 14 ˚C, the assumed threshold for initiation of nitrogen fixation by cyanobacteria in the model, has increased from about 75 days in 1982 to about 110 days in 2014, i.e.

by 35 days (cf. Fig. 5b). The center of timing of simulated cyanobacteria development has become 17 days earlier (cf. Fig. 10)".