Summary: The manuscript “Climate-scale changes of the semidiurnal tide over the North Atlantic coasts from 1846 to 2018”, by Pineau-Guillou et al., evaluates 9 long tide data sets on both sides of the Atlantic to investigate whether there is evidence for basin-scale perturbations to tides caused by climate variations or climate change. It is found that M2 for 3 gauges in the northeastern Atlantic follow a similar pattern as the decadally filtered NAO index. No clear correlation is found in the western Atlantic. It is noted, however, that rates of M2 change go negative at many of the 9 stations after around 1990.

Evaluation: It is an interesting idea to try to discern whether there are coherent, basin-scale variations in M2 in the northern Atlantic, and one that has proved challenging to find in the past (to my knowledge). It would be quite an interesting result if M2 patterns caused by climate variability only show up in long data sets. Therefore, if properly done, it would be interesting to convincingly prove (or disprove) the hypothesis that (for example) long-term NAO patterns affect tides coherently. However, as presently conceived and presented, am not convinced that the authors have really shown that this is occurring (or disproven it). Issues include:

1. There is no attempt at a statistical correlation between the tide records and climate records such as the NAO index. While the tidal records in Europe superficially follow a similar trend as the filtered NAO, it is quite possible for data sets with few degrees of freedom (here, a decadal median) to resemble each other by random chance. For this reason, it is important to do some sort of significance testing and report statistics such as R2 and the p-value. Similarly, would be good to verify that the 6 US East coast records show a statistically insignificant correlation.

2. Overall, it would seem to me that the data selection is incomplete. On the European side, there are some Dutch coastal records that predate 1920, such as Hoek van Holland and Delfzijl (maybe also others; perhaps check if the GESLA data set has them). Similarly, there are additional records on the US East Coast that could be used. These include Fernandina (1897-present; available from NOAA) and Sandy Hook (Available since ~1910 from NOAA) (Note that Baltimore (1902-present) and Philadelphia (1901-present) also exist, but are not coastal stations). Moreover, there are a number of M2 estimates extending into the 1800s in New York Harbor (Talke et al. 2014 supplement and Chant et al., 2018), Boston (Talke et al., 2018), Eastport (Ray and Talke, 2019), and Long Island Sound (Kemp et al., 2017). Some additional estimates of tidal range at coastal stations deep into the 1800s are found in Talke & Jay (2020). These can be used as a proxy for M2 if divided by ~2.

3. Further, it’s not clear how significant the data cut-off of 1920 is. If 1930 or 1940 were used, for example, how might results or conclusions change? A later cutoff would enable inclusion of a number of additional US East Coast stations (e.g., Mayport, Fort Pulaski, Sewells Point, Willets Point, Providence, Eastport, etc). Perhaps the same is true for Europe. How much would conclusions or patterns change if a slightly less restrictive date-cutoff were used? It would be good to use a statistics-driven reason for

the date cutoff (e.g., degrees of freedom in a correlation analysis, or something like that), and to check the effect of relaxing the cutoff. Regardless, my guess, based off of Figure 2 of Talke & Jay 2020, is that patterns of M2 change might be even less coherent if more data are used. This is in part because there are so many local processes that can affect tide gauges, even those at or near the outlet of an estuary (for example, Mayport, Lewes, Charleston, Fernandina, Sewells Point, Boston, and Sandy Hook in the US; Cuxhaven in Germany; and the Saint John gauge in Canada are all "coastal" gauges that are actually at the mouth of or within an estuary). On the other hand, there is an interesting and not completely explained decrease in tidal range and M2 at Sandy Hook, The Battery, and Boston from the mid and early 1800s until the 1920s, and then an increase. In other words, it's possible there is a larger Northwest Atlantic Signal, in addition to local processes. However, while models such as Schindelegger et al. 2018 are able to see coherent trends, it has been challenging to see it in the data (as mentioned in this manuscript as well, also in Ray 2006).

4. A similar data comment holds for the trend-switch which is observed around 1990. It would be quite interesting if this is a coherent signal throughout the Atlantic. To prove, would suggest that there are many more gauges that are available that could be used to test this hypothesis. Again, a trend switch in 6 out of 9 gauges need not be statistically significant or could be argued to be local in nature. However, would be more significant if you had 30 or 40 gauges and found a similar percentage shift between the 1950 to 1990 trend, vs. 1990 to the present, That would be quite interesting. Also, the effect of moving the date (1985 or 1995 instead of 1990) might be worth investigating. If results depend on start date, the interpretation becomes less clear (and vice versa).

5. The paper would also be improved by digging more deeply into mechanisms. Some discussion of how the NAO affects sea-level (and therefore, perhaps M2) is made, but it is quite qualitative. There are, however, a number of process-based studies that look into tidal changes at local, regional, and oceanic scale (see the Haigh et al. 2020 and Talke &Jay, 2020 reviews for references). The Atlantic is known to be near resonance, and there could be coupling between the shelf and the deep ocean (Arbic and Garret papers). Tide changes in the Gulf of Maine (and for that matter, Long Island Sound) can in theory radiate out to the larger ocean (e.g., Godin 1993). Also, it is known that the M2 amphidrome in confined, shallow seas moves with sea-level changes (see the references in the Haigh et al. review, or the Lee et al. 2017 and Ross et al. 2017 papers on the Chesapeake). Based on this, also perhaps on basin scale modeling (e.g., Schindelegger et al. 2018), what sorts of coherent patterns might be expected in the Atlantic based on historical sea-level rise? How much might this be spatially variable based on sea-level change caused by long-term NAO patterns? What might be the magnitude of the signal? The Schindelegger et al. 2018 paper shows that coherent changes across the basin are possible with sea-level rise, but are relatively small, for a given increment of sea-level change. Also, they show that some locations in the Atlantic are anti-correlated. How much of a sea-level change is needed, roughly, before a coherent longterm signal is findable in tide gauge data (given noise in data, etc)? One needs to know whether it's even possible (by the mechanisms listed) to obtain a secular coherence in basin-scale variability.

In other words, a more clear hypothesis of what a basin scale shift in M2 tides might look like and whether it is detectable (given current understanding of processes) might help with the interpretation. What sort of excursion in M2 would you expect the NAO to cause, based on how it affects sea-level? Setting up a hypothesis with specific criteria that can be proven or disproven might help. It's ok, in my opinion, to have a paper with a non-detect result (a possible outcome here). However, since the set of non-detect papers is infinitely large, the added value could come from adding scientific or statistical insight into the problem. A great example of this is the Haigh et al. 2014 paper which showed that one would need to wait a couple decades before being able to analyze recent sea-level acceleration at an acceptable level of confidence. Is something similar true here? My qualitative guess is that it might be hard to see a basin scale coherence in data, but that regional scale effects that are driven by a similar process like sea-level fluctuations can perhaps be detected (see for example the Devlin et al. papers).

Basically, the paper would be improved by more specifically investigating what sort of change is needed (and what sort of data quality/signal to noise is needed) before it might become possible to discern coherent climate effects on tides across the entire basin.

Detailed comments:

Line 14—Would also cite the review of Talke & Jay, 2020, since the historical changes in tidal range shown therein are relevant to this paper.

Line 19—"Long-term changes in tidal constituents are rather small" Would modify this to specify "at coastal stations". As shown in Talke & Jay 2020 (and refs therein), the secular change at many estuary and tidal river gauge stations is huge.

Line 20 "still poorly understood"-- Not sure I would say this. Some of the mysteries are being solved (see the review papers), while some issues remain. Maybe rephrase?

Line 46-48—Check grammar—grammar of list Is not quite right.

Line 60-63—There are some M2 results for 19th century US stations that you could/should use. See for example the supplement of Talke et al., 2014 or Chant et al., 2018 for New York and Sandy Hook. See Ray & Talke (2019) for Eastport and Portland. See also Talke et al., 2018 for Boston. Finally, there are multiple tidal ranges shown in the Talke & Jay 2020 review paper. These can be divided by two to get an estimate of M2 over time.

Also, Sandy Hook data from around 1910 is available at the NOAA site. The datum is wrong, but that shouldn't matter for tidal analysis.

Table 1—Since you are using Cuxhaven, why not also use some of the Dutch stations? In the records I have, Hoek van Holland starts in 1900, and Delfzijl starts in 1876. Maybe there are earlier ones as well—see for example the Hollebrandse 2005 thesis. You could check if they are in the Gesla dataset and/or contact the Dutch. Data used to be available at waterbase.nl, but not sure that works anymore.

Table 1—Why is Fernandina (1897-present) not used?

Table 1—Unclear what the meaning of mean sea-level is. What is the datum? Why not include the trend, rather than an absolute measurement (which is not necessarily meaningful).

Line 69: "This constraint resulted in excluding between 1 and 9 years". Unclear what you mean. You mean for each station?

Line 70 seasonal variation—where? Again, non-coastal stations will see more variability. Also, in the North Sea the change is higher (e.g., Graewe et al. 2014). So, maybe be specific and mention the Atlantic.

Line 74-91—This would seem to be pretty standard nodal correction theory. Unless you can explain what is unique about your approach, would suggest greatly condensing this and simply citing an older study that discusses this in more detail

Line 98-110—What is the rational for using the more complex method here? As you later state, it doesn't lead to significantly different results. I guess it's interesting that it can work for small time series. Do you see any evidence of changes to nodal cycle? There are a few papers on this recently. However, am not convinced that there is a physical reason for these observations, vs. just statistical noise. Could be something to look into, though—if there is a coherence between nodal cycle variations in the western and eastern Atlantic, would be worth commenting on. Otherwise, not sure that you need the complex approach to nodal cycle characterization.

Line 117, Equation 5—Can you explore/motivate the use of the standard deviation a bit more? A potential issue is that sigma may also reflect errors in the gauge data (e.g., timing errors, etc). Some exploration would be good as to whether this is a factor. For example, does sigma change as a function of time? If there is a decrease in sigma around 1990 or 1995—when new digital gauges started being used, at least in the US—then it might indicate that instrumental issues are potentially affecting your results. See for example Zaron & Jay, 2014, who concluded that some constituent trends in the Pacific are spurious. Using some sort of method to validate the causes of sigma would therefore be good. How can we be sure that a few years of non-optimal data are not biasing sigma? The method of Zaron & Jay, 2014 could be used, or the method used by Talke et al. 2018 to assess timing errors could be used (see their supplement).

Figure 2—This figure may have some educational/explanatory value, but it's not really a new result. One could consider removing.

Line 128 maybe remove "are essential, as they"? Doesn't' really add much to sentence

Line 156—"is no linear trends" should be "is no linear trend" or "are no linear trends"

Line 158 "curve is flattening" should be "curve flattens"

Line 159 Remove "yet" in "yet noticed"

Line 163-164—Not sure that the lack of an astronomical explanation automatically implies a solid earth-ocean-atmosphere coupling system cause. Am not even sure what is meant by that. A few more logical steps are needed before a reader can believe that

Figure 3: Charleston is a harbor city with a channel that has probably been subjected to dredging, though I haven't looked into it extensively. Can you discuss how/whether this impacts results?

Line 165—The Delfzijl station starts in 1876, so would be worth comparing to Cuxhaven. It's probably somewhat impacted by long term changes to the Ems estuary tides. Then again, Cuxhaven is probably influenced by the large change to Elbe tides. See for example Winterwerp et al. 2013. In general, there are quite a few papers out of Germany (e.g., Jensen et al. 2003, 2005 conference papers, and maybe Mudersbach et al. 2013 (?) that discuss a big increase in tidal range from about 1960 to the 1990s on the German coast. More recently, I've been told this has slowed or reversed (though I'm not sure there is a paper on that yet). Another good reference is the Hollebrandse 2005 Master's thesis on Dutch gauges.

Line 166—The Talke & Jay (2013, 2017) paper and report are good references for sea-level/tide data archaeology, as are Peauvreau 2008 and some of the papers by Marta Marcos.

Line 172—Again, note the Winterwerp et al. 2013 paper that includes the Elbe. There are probably some German references too. The Talke & Jay 2020 and Haigh et al. 2020 reviews discuss tidal resonance (see also references therein).

Line 182—The Ray (2006) and Ray & Talke (2019) papers discuss change in M2 trend in the 1980s in the Gulf of Maine—maybe reference.

Line 186—There are many other papers that have explored Gulf of Maine resonance besides Ray & Talke—that is not perhaps the best example. See e.g. the discussion and references in the Talke & Jay review, in addition to the Garret and Godin reference.

General comment: The Godin 1993 reference, and also for that matter the Arbic and Garret and Arbic et al.papers, are interesting because they discuss how resonance on a small scale (Gulf of Maine, Continental Shelf) can affect the larger Atlantic. See also the Platzman papers on resonance from the 1970s. All this could/should be discussed and investigated, since it gets at the idea that there might be a mechanism through which western and eastern Atlantic tides could be coupled. Is there reason to believe there might be? In a sense, this is an implicit hypothesis that is being investigated here, through correlation with climate indices. However, it would be helpful to motivate and explore physical mechanisms as well. Further, it might be helpful to explicitly pose a hypothesis in the introduction, such as "is there any evidence for correlated/coupled changes in tides that might provide evidence for cross-Atlantic connectivity"?

Line 190-194—Why not use the Eastport data points from Ray & Talke, or at least discuss? The composite Pulpit Harbor/Bar Harbor data set might also be worth discussion. Boston is a possibility, too, though it is influenced by local processes as well (see the Talke et al. 2018 paper…).

Line 198 –It would be good to compare Atlantic City to Sandy Hook and The Battery (see Talke et al, 2014 and Chant et al., 2018). In fact, the case of Sandy Hook and The Battery/Governors Island are interesting, since there is a marked decrease from the 1860s until the 1920s or 1930s, and then an increase. Chant et al. (2018) show an even bigger change in nearby Newark Bay, though the 19th century data there are based on very short time series. In any case, the results are sort of consistent with the results at Brest. Dredging may have at least somewhat caused the 20th century amplification (see Ralston et al., 2019), and work at the channel mouth may have cause the early 20th century changes (Marmer, 1935).

Also, Boston showed a similar, large decrease in tidal range through the 1920s, then an increase. While this is likely in large part local, Talke et al. 2018 did note that it's similar to the pattern observed at Sandy Hook.

Line 218-219—Would also look into/discuss amphidrome changes. See the Haigh et al. review and references therein.

Line 218-223—These are very short paragraphs and not that well developed. Some more thought would be good. For example, "The trends have to be interpreted carefully" is perhaps an obvious statement (hopefully there is not a case when it is ok to interpret trends haphazardly…).

Line 226—Misleading statement. Ray & Talke (2019) are referencing other results when they state that MSL rise only partly explains trends. Furthermore, they only focused on Gulf of Maine. Would instead look into some of the studies that have more carefully looked at SLR effects, such as Schindelegger et al. 2018 or Greenberg et al. 2012.

Line 228 "than mean sea" should be "as mean sea"

Line 230-235—The Pickering papers are for large sea-level rise scenarios, but don't retrospectively look at 20th century rise (if memory serves). Hence, is it a fair comparison? There are probably some papers or reports that discuss reasons for North Sea changes more thoroughly—please look into and review.

Figure 6—One could include the Portland sea-level data point from Talke & Ray 2019. If you include the Battery, then a longer data set is possible. Not sure however if this graph is needed or is critical for the story. It is not really a result of this study, just a replotting of other results. There is no clear analysis of how tides might be influenced by SLR—it's basically just a literature review.

Line 247-250. Wouldn't storminess also impact tidal constituents, at least on the shelf or in a harbor? I think there are some references on that. I came across a Pugh reference at some point for the Irish Sea, if memory serves. The Graewe et al. 2014 reference also discuss this for the North Sea, I think. In any case, wind stress and wave breaking and these sort of things represent an input of turbulent kinetic energy and could in theory affect tides at some stations, if there are climate-based shifts in storminess. In the context of this paper, Talke et al. 2014 showed that the probability of large storm tides in New York goes up when the NAO is negative. There are also known NAO effects in Europe (see the Woodworth et al. (2007?) paper). Does this matter for tides? Might be something to at least investigate.

By the way, it's not clear to me that a measured decrease in M2 during periods of stormy weather is a real change in M2. Another (perhaps not mutually exclusive) explanation is that depth changes during storms alter the phase speed of the tides, such that they arrive a bit earlier than usual (See for example Horburgh and Wilson 2007). A period with a lot of ups and downs in mean sea-level is going to cause lots of phase speed variations, more spectral spread (cusping), but decreased amplitude. Just as timing errors can cause a decrease in measured M2 (see Zaron & Jay 2014), so would changes in phase speed.

Line 254 "possible role of stratification" for what? Would clarify, e.g., something like "possible role of stratification on secular tidal trends"

Line 255 "between these processes"—what processes? Maybe be specific.

Line 266—How can the NAO decrease globally? It is specific to the North Atlantic.

Line 278—"Pushed southern" should be "pushed southerly".

Line 289—Might be good to discuss the role of wind earlier. See notes above.

Line 291-292—Maybe, but there is quite a bit of variance in all the plots and it seems like a couple curves looking similar could easily happen from random chance. Unless you can figure out the statistical robustness of these results, would perhaps avoid ascribing M2 behavior at a few locations to NAO.

Line 332—The Devlin et al. papers discuss correlations between sea-level anomalies and tidal anomalies, and possible reasons for them. In a way, you are trying to do something similar, but over a larger time scale. However, there is little statistical correlation or significance testing done here. Would suggest this be done.