Review of Garel and Ferreira, “Multiyear high-frequency physical and environmental observations at the Guadiana Estuary”, ESSD

This data set, or rather these 28 separate data files, might prove useful, as the authors contend, to research, international comparative science on estuaries and hydrologic and ecologic modeling. But not as presented in this paper. The authors give us the metadata and geographic descriptions but not the necessary value added in terms of compilation, temporal sequencing, or informed quality control to make our use of the data easier and more successful. They describe in their abstract “substantial data gaps” - expected no doubt in environmental time series and quite evident in Figure 2 - but leave it to the user to find the useful bits. My exploration of one of the ‘best’ data periods suggests that, when combined, gaps or failures in water quality, water velocity and river discharge data dominate the time series. I can not say with confidence that the 2008 to 2014 time period advertised in the abstract, roughly 50 months including described ‘out of water’ intervals, contain more than a few months of useful data.

I believe the authors need to confront and disprove, to their and to users satisfaction, this null hypothesis: the combined data sets contain no more than 6 useful months of data. I pose the question in this manner precisely so that they can quantitatively disprove it. Otherwise we have no basis on which to accept their statement “prolonged (months-long) periods of observations during contrasted external forcing conditions are available”.

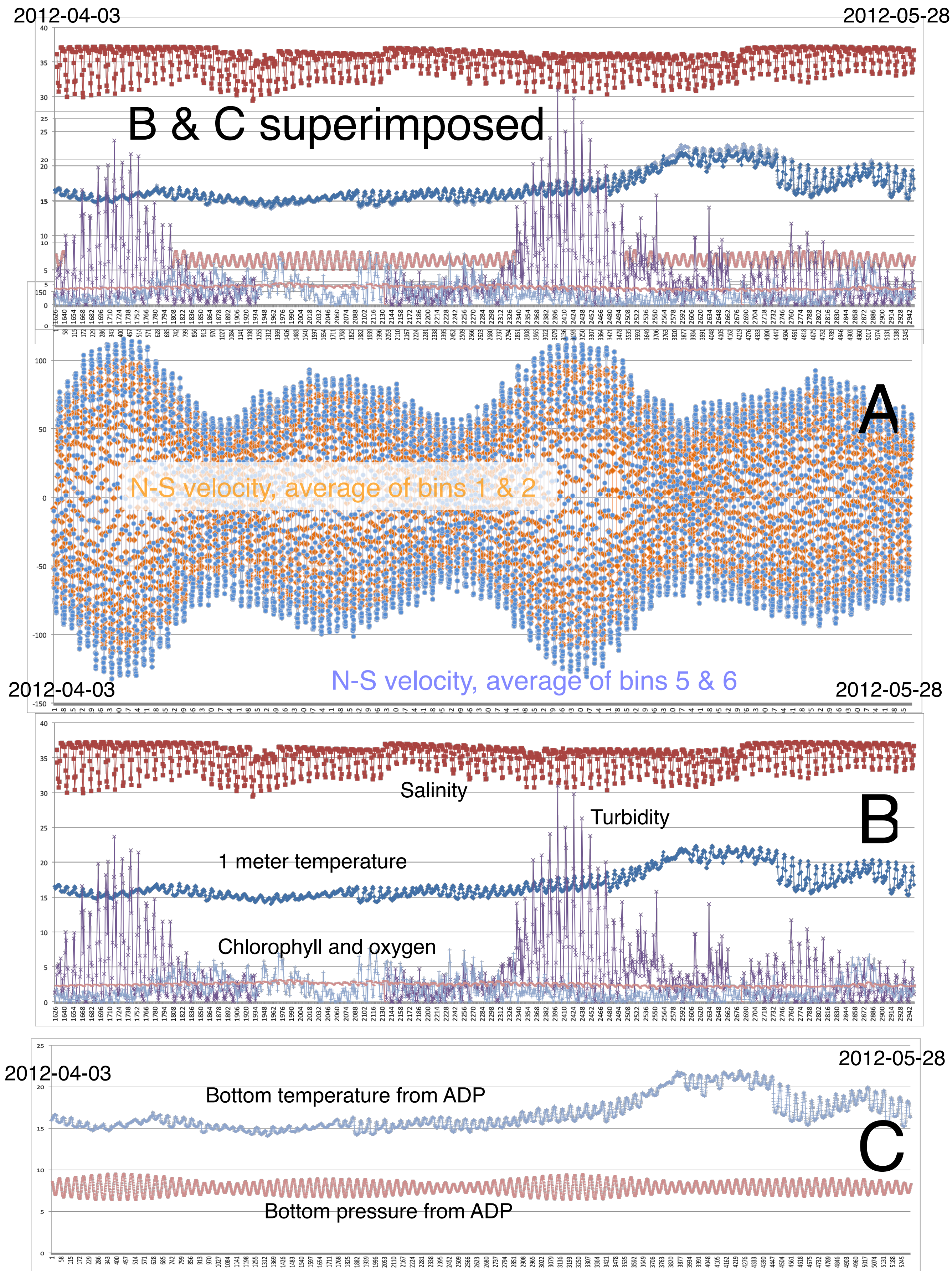
The authors have provided good access to and very good metadata in each file. As demonstrated below, and despite inconveniences, I could access and open every file. The water quality and discharge data exist as .tab files on the Pangaea landing page. Perhaps because of their much larger size the ADP files require an additional downlink step, via a doi copied from a spreadsheet. The ADP files download individually from tape.

The ADP files exist as .txt. On a Mac, the .tab files do not open without renaming to .txt. I see no reason why the authors should not change all data types to .txt.

I wanted to evaluate an extract of the estuarine data but I found Figure 2 not at all useful (specific recommendations below). Reading the water quality and ADP time periods from Table 1, I decided to try for the time period 2012-01-26 (based on water quality) to 2012-12-12 (based on ADP), nominally 11 months.

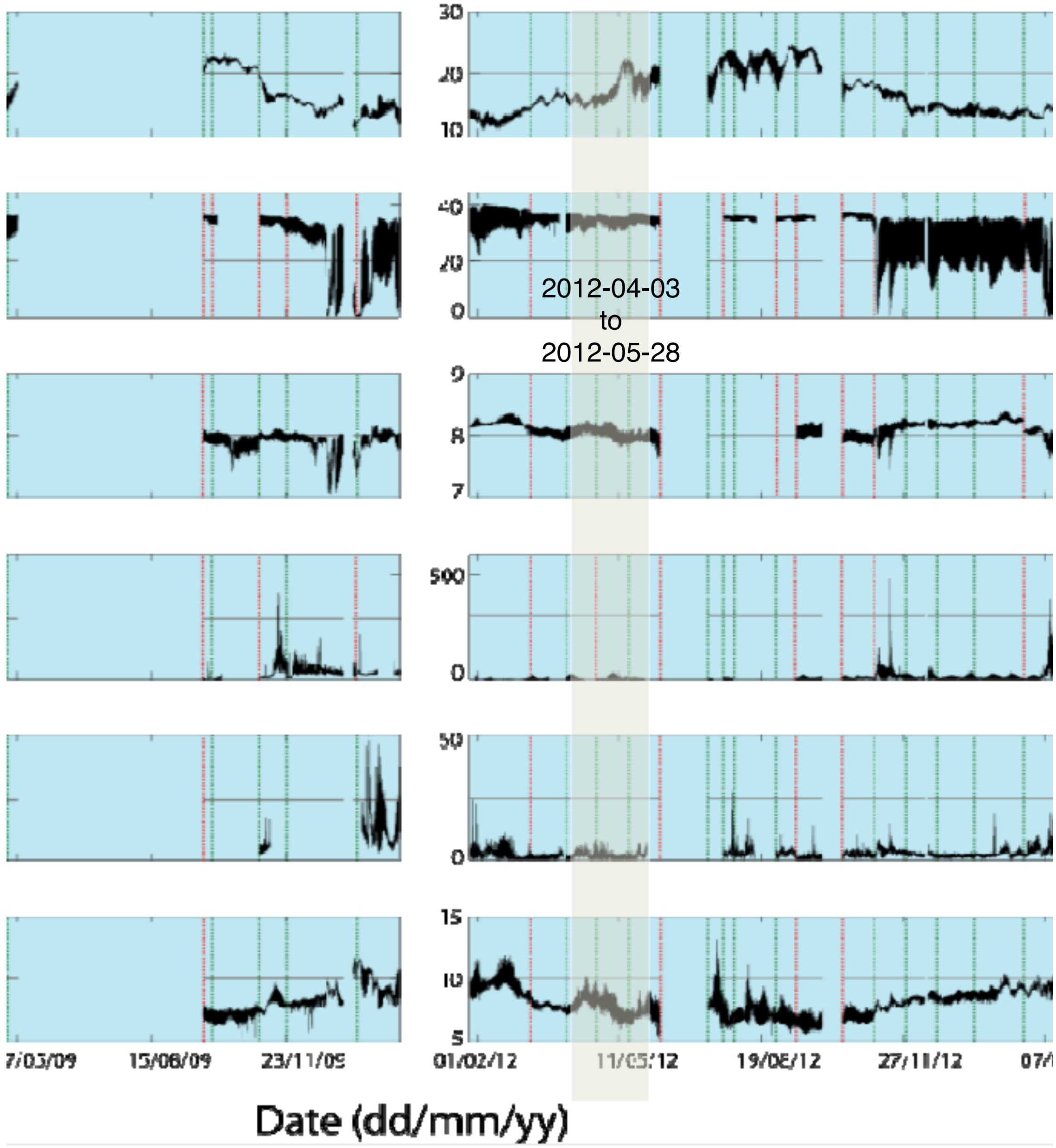
The authors tell us as users that the water quality occur hourly while the ADP data record 4 per hour - a second inconvenience (specific recommendations below). Unfortunately, the two types of data files use different mutually incompatible date-time conventions - a third inconvenience.

Attempting to use as much data as possible but while also adhering to the data quality flags (ADP) or delimiters (water quality) - again different, a further inconvenience - I ended up with roughly 6 weeks of useful continuous data, 2012-04-03-1300 to 2012-05-28-1700. To gain a quick look at this data I ignored vertical (upward) and eastward velocities, ignored pH, and scaled chlorophyll and oxygen back to mg (the authors present the data in micrograms). To look separately for an estuarine circulation of bottom inflow with surface outflow I averaged the two lowest ADP bins (vn_1 and vn_2) and the two highest bins with reasonably few bad data flags (vn_5 and vn_6). Only 12 records out of 5300 had bad (0) data flags in both bins 5 and 6. I cleared all bad data (/) from the water quality files. From Excel I exported these files as pictures, made the backgrounds transparent, and plotted them below.



Data extracted for 2012-04-03-1300 to 2012-05-28-1700. A - average of vn_1 + vn_2 (orange) and vn_5 + vn_6 (blue). B - water quality. C - bottom temperature and pressure from ADP.

Once I produced a useful extract, I plotted that time period on Figure 2 (below).



I found no useful river discharge data from either station for this time period.

Data assessment

Because of the dismal quality of Figure 2 I needed to make this extract myself. I believe the authors should have done this, not me. Here, unlike Figure 2, we can actually view the "high-frequency physical and environmental observations" as advertised.

The ADP pressure and temperature data look good: few or no gaps and very reasonable values. The surface temperature data, from the water quality file, also look very good. In the superimposed images one can see this close coherence of water temperature data from separate sensors.

The velocity data (only N-S) look very clean, two very clear tidal cycles, about as one might expect (I plotted these in cm sec^{-1}). The authors - based apparently on arbitrary criteria - describe this time period as low flow. Nevertheless, one can find hints of the expected estuarine circulation. The velocities for bins 5 and 6 show stronger southward signal while the bottom two bins show closer to net zero flow.

For water quality data I consider turbidity noisy as expected but acceptable (and showing tidally-driven resuspension?), likewise for chlorophyll with oxygen at reasonable values throughout. I consider the salinity data as highly unreliable probably for exactly the reasons the authors describe: biofouling and other factors leading to calibration drift. Personally I would not accept this salinity data for any useful analysis. Without valid salinity data, or at least to find some correlation with freshwater discharge, one would wish to have river discharge data from at least one station but we unfortunately had none for this time period.

How should we evaluate this data? What constitutes a minimum useful data subset? The lack of reliable salinity data seems quite troubling - but not surprising to many of us who have tried in situ unattended conductivity measurements - and salinity at a single depth and point does not provide sufficient data to constrain circulation or mixing in any case. For these data then this user would suggest a minimum acceptable data set to include temperature from bottom and surface, pressure, E and N velocities, turbidity, oxygen and perhaps pH or chlorophyll (although the latter becomes less reliable in periods of high turbidity and also suffers some of the same biofouling and other contamination as the conductivity cell). Absent salinity, one needs river discharge.

Accepting my definition of useful data, how much useful useful data exists across these data sets? Unfortunately, the paper as presented gives us no mechanism to answer this question. Panels (a) and (b) of Figure 2 have offset, inconsistent, asynchronous and unreadable time scales. Figure 3 has yet another time scale. Note that if we used salinity as a usefulness criteria, and based on what we see in my high resolution data extract, we might conclude from the blurred but discouraging salinity signals widely spread across Figure 2 that the data time series has very little useful data.

Issues and recommendations

(I base all of these issues and recommendations on the premise that the central data providers, not the separate data users, need to make these improvements!)

1) <u>Time stamps and recording intervals</u> - we need a uniform ISO time format across all files. We need all data reduced or reprocessed to a standard 1 hour data frequency.

2) <u>Data errors and data flags</u> - having worked with this subset of data I have zero confidence in the data quality procedures applied to this data. First, we encounter two separate validity indicators: 1 and 0 flags in the ADP files but /designators in the water quality data. Second, the authors seem to have accepted, in the case of ADP data, some software-generated or generic data quality assessments (beam amplitude, beam noise, beam strength) without any actual informed data analysis. They have reliable bottom pressure from which to determine water

depth so they could apply differential flags or corrections based on water depth for the upper bins? Across multiple bins they could do nearest in time or nearest in space checks? Most of the ADP data that I eliminated due to 0 flags looked in fact quite reasonable and consistent (but of course I avoided bins with high error rates). Still it seems that the authors could do a smarter situational error estimate rather than simply using machine parameters. For the water quality data I removed everything flagged with '/' and many of those data looked quite bad, especially unrealistically and erroneously high. But meanwhile, as my analysis shows, the salinity data look highly questionable throughout but carry no error or uncertainty designators. Overall I think the authors have accepted default sensor-based and software-based error detection algorithms without any informed analysis based on this deployment in this situation.

3) <u>Overall data return</u> - Accepting biofouling, sediment contamination, data logger failures, broken electrical connections, adverse weather, engineering changes, malicious disruptions, power failures, bad calibrations - all the real-world factors of environmental monitoring in a shallow and exposed estuary - one could assume that over 6 years of calendar time and perhaps 4 or 4.5 years of deployment time that one achieved a data return of 50%. If we accept the abnormal salinity patterns identified in my analyses as indicators of bad data, then a quick perusal of Figure 2 suggests perhaps very few periods of successful data returns. Even ignoring salinity we seem to have very long periods of far less than 50% return. Or perhaps better returns, but we really can't tell! We need a full data availability data quality time series indicator for all data sources (water quality, ADP and discharge) on one uniform and readable time sequence. Figure 2 provides no assistance whatsoever in this regard. We need this for three reasons! First, to help the authors show overall success rather than predominant failure. Second, to give users a view of the overall data stream so that they can easily find the useful sections. And third, so that the authors can extract a subset of data from one of the 'good' periods and prepare a figure or figures, much better than mine, to actually show the quality of the data.

In addition we need some summary statistics: overall measurement time periods, overall possible combined (ADP, water quality and discharge) output, actual output. For that task I issue again my null hypothesis challenge: disprove the null hypothesis that the combined data sets contain no more than 6 useful (the authors can develop their own definition or definitions of useful) months of data. Disprove that and improve the quality assessments and then we could have useful -very useful - data.

The time axes for the water quality (a) and velocity (b) panels in Figure 2 do not coincide.

Figure 2 labels far too small to read.

Figure 2 ADP not useful, no net flow?

ACP does not resolve estuarine circulation, in at depth and out at surface?
For the extracted segment below, column integrated Vnorth shows -0.03 m/s with high SD and range of +1 to -1. Bottom two bins show much smaller net southward with about the same min and max. Two upper bins with valid data show stronger southward flow with min (south) distinctly large than max (north). In plot, near bottom bins show closer to net zero flow while surface shows southward displacement.

Text does not use ISO dates while files do.