We thank the referee for taking the time to read the manuscript and for making comments. We believe the provided suggestions/comments will help to enhance the manuscript.

NOTE: The original comments by the referee have been numbered 1-44, and red text has been used for the response by the authors.

Comments:

1. *The presentation of the data manipulations and suggested corrections appear very diagnostic and data-driven. At the same time, there are plausible physical causes to most of them, which the authors present, too (e.g., ambient light effect on FETs; temperature effect on E*). It's likely mainly a question of style, but my preference would be to always start from the sensor knowledge to explain an observed mismatch and then suggest corrections, rather than an "our data didn't fit, so we made it fit" approach and then defending these corrections with theory after the fact.*

As the ISFET sensor was supposed to take into account temperature and pressure changes in the environment (Shitashima *et al*., 2002; Shitashima, 2010), with a good level of accuracy (Shitashima *et al*., 2013), ISFET pH measurements were not expected to differ significantly from the ship based pH measurements, particularly when considering past observations of pH in this region over a similar timescale (see comment 2. Below). We believe that the observed disagreement between ISFET sensor and ship-based pH data should be considered as a result, which should be described first, followed by a discussion of the sensor, and lastly a proposed correction/course of action. We will make this logical progression clearer in the revised manuscript, by explaining the generally accepted views at the beginning, and pointing out where our results differ from this.

2. *More importantly, the goal of most (all?) the corrections seems to be to reduce the glider pH variability to the level seen in shipboard pH samples? This misses the point. Continuous, autonomous observations can very well be more variable than discrete measurements, in particular if the continuous measurement series captures time/ spatial scales of variability or events that simply go undetected with coarser discrete sampling. In that regard, the drift correction to pH(@14 °C) is a critical point that needs more detail and potentially a second look (details further below).*

We agree with the referee that variability seen by the glider should be different to ship-based measurements; this after all is the advantage of using gliders. We did see a larger variability in the higher resolution profiles from the glider, compared with ship measurements. However the magnitude of the variability seen in the glider-borne sensor was well outside the likely maximum range of pH variability seen in the literature (as well as the ship measurements in this study). This is particularly true when looking at measurements obtained at greater depths where we would not expect large changes in pH. For example, pH measurements (T = 25˚C) presented by Alvarez *et al*., (2014) taken from a hydrographic transect in the western Mediterranean Sea over a similar timescale to our deployment, varied between 7.87 and 7.93 at depths greater than 100 m, whereas pH measured by the glider's ISFET sensor (Fig. 4d) ranged between 7.97 and 8.18, which is roughly three times larger. This suggested that the measured range of the ISFET pH was incorrect (i.e. drifted) and required correcting.

A paragraph will be added to the introduction describing past observations of *in situ* high resolution pH in general (e.g. Hofmann *et al.*, 2011), and hydrographic measurements specifically from the northwest Mediterranean Sea (e.g. Alvarez *et al.*, 2014) in order to provide readers a background of typical pH variability.

The drift correction is discussed further in the response to comments 20 & 21.

3. *There is a lack of detail on the sensor used and its handling. This limits the utility and impact of the present study. Relevant information need to be added.*

More information will be added to the manuscript. This is detailed in the response to comments 9.1-9.4.

4. *P1L4: "Northwestern Mediterranean Sea" suggests a basin scale study and is maybe a bit a too generic description of the deployment location, i.e., a transect of just 100 km off the Sardinian coast? Similarly, P1L14 and P1L16: "this region" is not well defined (I didn't know what it actually refers to), so I would suggest to closer specify the study region.*

We agree with the referee that 'Northwestern Mediterranean Sea' represents a larger area than that observed by the glider. Instead we now refer to the region as the 'Sardinian Sea'.

5. *P2L9: Why not use ppm for the mole fraction?*

The term 'ppm' is an ambiguous unit (Schwartz and Warneck, 1995) and we prefer to use 'µmol mol^{-1}' as this is more specific.

6. *P2L16: Potentially add the relevance of pH changes/anthropogenic CO_2 invasion to the study region/Northwestern Mediterranean?*

Sentences discussing this will be added to the introduction.

7. *P2L17: What is stochastic variability?*

We will simplify this sentence for increased clarity.

8. *P3L11: Last sentence is irrelevant to the presented study.*

This sentence will be removed from the manuscript.

9. *P3 ISFET and glider sensors:*
   9.1 What was the source of the sensor? Is it commercial/ semi-commercial/custom-built?

The sensor is currently under trial and not commercially available. The sensor was custom-built by Kiminori Shitashima (Tokyo University of Marine Science and Technology, Japan) based on an ISFET sensor from Hitachi ULSI Systems Co., Ltd. ten years ago. It has previously been used in studies by Shitashima *et al.*, (2008) and Shitashima *et al.*, (2013).

9.2 Was the ISFET unit a commercial product (Honey- well?)?

The ISFET unit was made by Kiminori Shitashima via special order.

9.3 On what material support is it mounted (important to assess the pressure tolerance)?

The housing of the unit was made from acrylic resin, and the ISFET and CL-ISE were moulded with epoxy resin in the housing.

9.4 Is the packaging of the ISFET into a sensor a commercial/semi-commercial/ custom-built one?

The housing of the unit was custom-built and is not commercially available.

9.5 This is essential information to put it into context of other studies with (other) ISFET pH sensors and directly affects the impact of this.

Information provided in the responses to comments 9.1-9.4 will be added to the manuscript.

10. *P3 ISFET and glider sensors: The handling of the pH sensor needs to be described in detail.*

10.1 Was there any temperature or pressure compensation/calibration (in particular on E) other than described later in the manuscript?*

No temperature and pressure calibrations/compensations other than the corrections specified in the manuscript were performed on the ISFET data. A sentence stating this will be added to the manuscript.

10.2 Were salinity/Cl^- changes taken into account (as suggested by equation 1) for the calculations?

Yes, salinity changes were taken into account when undertaking the calculations. This will now be described in the manuscript.

10.3 How was the ISFET and the reference electrode stored before deployment: in NaCl solution, artificial seawater, Mediterranean seawater, at what salinity, how long before deployment? ...

The ISFET and reference electrode were stored in a bucket of seawater for an hour before the deployment of the glider. The salinity of this water was about 38.05. This information will be added to the manuscript.

11. *P3L16: "the [other] retrieved data were of very poor quality". Any ideas why?*

The data retrieved from these sensors could not be used due to quality issues. It is unclear why there was a problem with measurements obtained by the stand-alone p(CO_2) sensor. However, we think the regular on/off cycling of electricity to the integrated dual sensor in between sampling did not allow it to function properly. This information will be added to the manuscript. A sentence outlining the authors' recommendation to potentially improve the integrated dual sensor will also be added to the conclusion section.

12. *P4L18: I don't understand the figures. "DIC and A_T differed by 3.1 and 2.5 µmolkg^{-1}, respectively" means that the second sample was always higher than the first one? I would hope that the difference between replicates would average around zero, otherwise this sounds like a serious methodological issue? I assume the authors refer either to the average absolute difference between replicates or the standard deviation between replicates?*

We understand that this could be confusing, and the referee is correct in assuming that we used the average absolute difference between replicates, with the value to the right of the '±' symbol representing

the standard deviation of these absolute differences. However, we now think it will be better to list the mean standard deviation of the replicate samples in the manuscript.

13. *P4 last sentence and first sentence on P5: This is unclear:*
    13.1 *How many casts were performed? (Should probably be mentioned in section 2.3 and/or P5L18)*

We agree that this should be further clarified as 'casts 24-51' is not specific. We will mention in the text that eight casts were performed.

    13.2 *Why are there several standard deviations for a "standard deviation of the mean DIC/A_T (averages over all casts)"? I kind of get the idea to split it into surface (top 150 m) and deep values, but that only gives me two values. Instead, I see two ranges of standard deviations? Looking at the figures (4d), it seems like the data were aggregated into depth bins and – likely – the ranges are the numbers for the respective depth bins shallower than 150 m and deeper? This has to be explained in the text.*

This has in part been addressed in the response to Review 1 – comment 13.

To calculate these ranges of standard deviations, the values from all profiles of a given variable (e.g. DIC,A_T, pH_s…) were sorted into 10 m depth bins down to a maximum depth of 1000 m. The mean and standard deviation was calculated for each one of these 10 m bins using the assorted data within. This produced two arrays; 100 x mean values and 100 x corresponding standard deviation values between the surface and 1000 m depth. Thus, the quoted standard deviation ranges (e.g. for the top 150 m) were defined using the minimum and maximum standard deviation calculated from these bins within the depth range (e.g. 15 out of 100 binned standard deviations for 150 m). We will make sure to explain clearly what these standard deviation ranges represent, and how they were calculated, in the updated manuscript.

    13.3 *Depending on the size of the depth bin, the depth gradient can become an important contributor to the standard deviation. Say all glider dives are identical, the standard deviation of the top 150 m would still be much higher than the bottom 150 m because of the higher depth variability near the surface compared to depth. Same for P5L9 and other statements like this (e.g., P7)*

The size of the depth bins was chosen to take into account the vertical pH gradient. Hence, for example, for the top 150 m of the water column, there are 15 standard deviation values. It is for the reason highlighted by the referee that a range of standard deviation values was given, rather than one value for the entire selected depth range (e.g. top 150 m).

More information (including the size of the depth bins) will be described in the manuscript, as mentioned in the response to Review 1 – comment 12.

14. *P5L14: What about the magnitude of diel variations? Because that's essentially what is looked at here.*

There is not much in the literature specifically considering the diel variability of pH within the top 1000 m

of the water column close to this part of the Mediterranean Sea. However, we will add references here that describe diel variations in pH at other location (e.g. Hofmann *et al*., 2011).

15. *P6L6-9: Is there experience from other autonomous deployments (floats?) in the literature that could be used?*

The authors have not been able to find in the literature experience of correcting ISFET pH measurements for ambient light on a glider/float.

16. *Section 3.3 Correcting pH for drift, temperature, and pressure: Can you give more details about your corrections (equations, magnitude/values of m and c) to make it reproducible for others? Please also comment whether the temperature and pressure slopes are comparable to other findings (in particular P6L24 and Johnson et al. 2016)?*

More information will be added to this section. This will include the offset equation, delta pH equation, and temperature and pressure correction equations, incorporating the calculated slope and intercept coefficient values. These will be compared with other findings, such as Johnson *et al*., 2016.

17. *P6L27: "unrealistic scale" is unclear. Please specify or rephrase (large range?). In addition, scale is ambiguous here since it could refer to the different pH scales (total, seawater, ...).*

This sentence will be re-written in a clearer way, and 'scale' will not be used.

18. *P6L28/P6L30/P7L4/P1L8: What is it, a time-varying or a constant offset? Please be consistent to avoid confusion (or simply remove the constant in P6L30/P7L4?).*

It is a depth-constant time-varying offset, as highlighted by comment 19 below. This offset will now be consistently referred to as a 'depth-constant time-varying' offset.

19. *P6L28: "depth-constant" (uniform with depth) instead of "constant-depth" (applied to the same depth level)?*

We now refer to a 'depth-constant time-varying' offset. See comment 18 above.

20. *P6L30: The density gradient was weak, the pH gradient, too? You don't want to have a gradient in your variable.. Can the depth of ϑ = 14 °C be made visible in one of the plots to get an idea of the depth range?*

The mean depth where offset values were calculated was 188 (±105) m generally below the thermocline. The majority of these offset values were obtained using pH values below 100 m depth where pH gradients were weaker, but some offset values were calculated at depths between 75 and 100 m where pH gradients were greater (see Fig. 4d/Fig. 6). However, the relationship between the calculated offset values and variability in other parameters (e.g. salinity and dissolved oxygen) was insignificant (see comment 21 below), suggesting the calculated offsets were mostly representative of instrumental drift rather than physical and biogeochemical variability. We will add a few sentences explaining these points in the manuscript.

The depths where the offset values were calculated (i.e. where T = 14˚C) will be displayed in the updated version of the manuscript.

21. *P6/P7 offset drift correction: How does the time evolution of the offset look like? Is it linear, exponential, or at least smooth (could be added to Figure 7)? If not, then what the authors measure is in fact not the pH but a pH anomaly relative to pH_S(14 ˚C), i.e., they remove the environmental variability of pH_S(14 ˚C) from their pH_g data.*

The time evolution of the offset is varying with time and is essentially a pH anomaly relative to pH_s(14 ˚C). As highlighted by the referee, this could be interpreted as environmental variability. We agree that it would be good to show the time evolution of the offsets, and a figure displaying offset pH values, salinity, and dissolved oxygen concentration (at 14 ˚C) with time will be added to the manuscript.

Weak relationships (r^2 = 0.2) were found between variability in the pH offset values and variability in salinity and $c(O_2)$. This analysis suggests a small proportion of the variability in pH offset values can be attributed to changing environmental conditions, and that the calculated pH offset values are mostly representative of instrumental drift. Sentences discussing this new figure and the relationships between pH offset values and salinity, and $c(O_2)$, will be added to the manuscript.

22. *P7L5: If you derive the temperature correction from a subset with similar temperature gradients in the surface, is it applicable for the entire deployment/dives with different temperature gradients? Temperature certainly plays a role for these dives, too, but does it follow along the same relation? A look at figure 6 suggests that the selected stations cluster on one side of the corrected profiles, i.e., there is a bias? (Which might also cause some portion of the high surface variability in pH_{gTPc}?)*

The referee is right that the '$pH_{g\ TPc}$ Sel.' data points were generally situated to the right of '$\mu pH_{g\ TPc}$' within the top 100 m of the water column. We have decided to now use measurements from all dives (i.e. no subset) for the temperature and pressure corrections, with light affected measurements removed from the top 50 m of the water column during the day. We think this will be a more robust approach. Figures 6 & 7 will be updated to display all dives with slopes and coefficients, and the corresponding text will be modified.

23. *P7L13: And excluding daytime dives?*

As mentioned in comment 22, pH data affected by light within the top 50 m will now be excluded from the correctional procedures. These excluded measurements will be scattered on Fig. 6 for reference, and a sentence explaining this exclusion will be added to the updated manuscript in the relevant section.

24. *P7L14 vs. P7L21: in situ or potential temperature??*

'Potential temperature' on P7L21 was a mistake. It will be changed to '*in situ* temperature'.

25. *P7L28: "to achieve a match within the pH repeatability of the discrete samples" That's not the point of continuous vs. discrete measurements. A higher variability in continuous data can easily be real.*

We agree that a large range in pH variability measured continuously could be real in some cases, as was presented at some locations by Hofmann *et al*., 2011. However, as discussed in comment 2, the magnitude

of the variability observed by the glider-borne sensor was well outside the likely maximum range of pH variability seen in the literature (as well as the ship measurements in this study). This was particularly clear when looking at deep measurements, as glider-borne measurements varied within a range three times larger than the range measured by ship during a past expedition in the Mediterranean Sea when looking at measurements collected on a similar timescale (Alvarez *et al.*, 2014).

26. *P7L26: Indeed. Did you try any laboratory experiments with your pH sensor to confirm a temperature dependence (and salinity- and pressure dependence, if possible)? At least the temperature aspect should be easily feasible and would add significantly to solidify the correction approach.*

We agree that this would improve our understanding of the ISFET sensor. However, it was not possible to test the ISFET sensors under laboratory conditions.

27. *P7L29: Can you comment on the uncertainty of your corrections and how that might affect your data? A linear temperature correction for ISFETs seems to be well- established, pressure corrections seem to be handled differently (e.g., this work, Johnson et al. 2016)?*

A few sentences will be added to the manuscript commenting on the differences between our corrections and those used in other papers, such as Johnson *et al*., 2016.

28. *P8L3: "at some locations": Imprecise, please specify (East/West/coastal/...?)*

The longitudinal ranges of these locations will be added to this sentence.

29. *P8L5: Don't you have any data to support the DCM depth for your study? It seems like there were (at least) 12 gliders and two research vessels deployed..it should be possible to find (even an uncalibrated) Chlorophyll a fluorometer on a CTD among them..?*

Fluorescence was measured by the ship's CTD instrument, and an increase can be seen at the depths where oxygen and pH increases, supporting the notion that this is the DCM depth. Fluorescence measured by the ship will be described in the manuscript.

30. *P8L17: "The spatial variability of these two regions differed for each time period" is unclear. Can you extend on this (what time periods; any relation of changing extend with displacement of isopycnals/water masses/SSHA)?*

The reader will be referred to Fig. 9 for the specific time periods, and comments will be made on the relation between pH spatial variability and changes in temperature and salinity (i.e. water mass properties), and isopycnals.

31. *P8L18: "at a range of depths": Please specify. Were values similar along isopycnals E/W and the depth differences are just inclined density surfaces?*

The depth will be specified as 'deeper than 100 m'. It seemed these parameters followed isopycnal surfaces at a range of points in time and space, which is particularly clear in the top 200 m. This will be discussed in the manuscript.

32. *P8L23: Which time periods? (Maybe specify in section 2.1?)*

Again, the reader will be referred to the time periods labelled in Fig. 9.

33. *P8L28: Sentence unclear to me. (Intrusion instead of encroachment?)*

We have replaced 'encroachment' with 'intrusion' as suggested.

34. *P8, section 3.4: This section describes the data and depth structure (first paragraph), it describes the East-West differences in the transect (second paragraph), and it discusses circulation aspects to explain mainly the physical oceanography data (third paragraph). What I think is missing in a section entitled "pH variability" is a biogeochemical discussion how to interpret the East West differences in pH. Is it related to a coastal/offshore gradient, to different preformed pH/DIC/A_T/O_2 concentrations in the respective water masses, to a gradient in nutrient supply and/or respiration (again: coastal/offshore gradient or likely water mass effect), ...? All these questions remain unanswered. (Potentially, part of the depth structure discussion of the first paragraph could be merged with this "fourth" paragraph.)*

This section will be reorganised as suggested by the referee, and a biogeochemical discussion will be added to the manuscript.

35. *P9L4: Do you have any ideas/reason/speculation what caused the drift? The ISFET unit? E*? How could you reduce the drift in the first place or is it impossible to avoid?*

As the referee suggested, we can speculate that the drift was likely related to the interface potential between the two n type silicon parts (source and drain) being affected. However, as it was not possible to test the ISFET sensor for drift, and that many variables were involved, it is difficult to identify the true cause of the drift.

It is possible the drift may have been caused by the lack of proper conditioning before the deployment. The ISFET was switched on and left in a bucket of seawater for just one hour, contrary to some weeks as suggested by others (e.g. Bresnahan et al., 2014). Putting aside that our sensor differed from the Honeywell Durafet sensor described by Bresnahan et al., 2014, and that the salinity sensitivity of our ISFET sensor was small, perhaps more time was needed for the ISFET to stabilise in seawater prior to deployment.

To determine the true cause of the drift, in future two ISFET sensors should be tested in laboratory conditions within a bridge arrangement circuit to attempt to isolate possible factors contributing to drift.

A discussion concerning the possible cause of the drift will be added to the manuscript.

36. *P9L7-9: Again, a lab temperature study would solidify this result.*

We agree with the referee that this would be useful, but a lab temperature test was unfortunately not possible.

37. *Fig 1: A distance scale in the left figure, too, would be nice.*

A distance scale will be added to the left panel in Fig. 1.

38. *1 Fig 1: What about the ca. 15 km North/South displacement between water samples and glider*

path for the match of water samples to glider dives? I might have missed it, but did you describe in your methodology how you matched glider dives to ship hydrocasts?

We calculated the offset, ΔpH, and ΔpH_{Tc}, using the mean profile of the ship pH measurements ('μ pH_s' - blue profile in Fig. 6). This was decided as the pH_s standard deviations were relatively small when compared with standard deviation values of the ISFET glider data. Mean pH_s is indicated in Sect. 3.3 on P6L29, P7L11, and in Fig. 3. However, this was not specified within the caption for Fig. 7 which may have caused confusion. Equations for ΔpH and ΔpH_{Tc} will be added to the manuscript, and the caption for Fig. 7 will be modified.

38.2 Shortest distance? Along equal longitude? The bathymetry suggests quite some difference at the same longitude close to the coast, so that a "distance from the coast" or "equal bottom depth" might be more adequate/give a better match?

We thank the referee for suggestions on how to match the ship hydrocasts with the glider's measurements, however we did not match the individual bottle casts with glider measurements as described in comment 38.1 above.

39. *Fig 4: What about a left/right grouping of water samples (DIC, A_T, pH_S; left top to bot- tom) and CTD/glider data (ϑ, S, pH_g; right top to bottom)? This would avoid confusion about the legend next to 4c. Also, the legends could be placed inside the subpanels to gain some space (in particular to better see the subsurface maximum in pH_S)?*

We plan to split Fig. 4 into four separate figures:

1. c(DIC), A_T, and derived pH
2. temperature, salinity, fluorescence, and $c(O_2)$ from the ship CTD measurements
3. pH_g vs. pH_s measurements
4. glider vs. ship measurements of temperature, salinity, and $c(O_2)$.

We think splitting Fig. 4 into separate figures, and visualising the spatial variability better as transects will improve the manuscript. We will take the advice (i.e. saving space) of the referee on board when creating the figures.

40. *Fig 5: Maybe rename the y axis labels of panels b-e and the variables in the figure caption by ΔX instead of X to emphasize the anomaly?*

We agree this would be better. This figure will be updated.

41. *Fig 6: "offset drift correction" and 40 m? (Fig 7: Make consistent with in situ / potential temperature of the correction description.)*

The caption for Fig. 6 will be modified, and the description of the temperature correction will be consistent with Fig. 7.

42. *Fig 8 and 9: Why did you split the plots into two figures? In my view, they would be more sensible as one (pH data together with its context). If space is a concern, you could think about removing the x axis labels and ticklabels for the upper panels since they are identical (as you did for the y axis labels and ticklabels for the center and right panels).*

We will combine Fig. 8 and 9 as suggested.

43. *Minor: I would also appreciate a distinction between "the sensor"/"the ISFET sensor"/ "the ISFET pH sensor" and "the ISFET". The first refers to the ISFET including the packaging (housing, electronics, ...) the authors used (i.e., their experimental sensor) while the second refers to the type of sensing probe (a transistor)/its working principle that can be shared by many different pH sensors but the one discussed here. It seems that in quite a few instances where "The ISFET ..." is used, it merely refers to "Our ISFET pH sensor ..." rather than to all ISFETs.*

A distinction will be made between "the sensor"/"the ISFET sensor"/ "the ISFET pH sensor" and "the ISFET".

44. *Typos: P4L15: ...Scripps Insititution of Oceanography, USA, ... P5L33: FET-based sensors P7: "Tc" is sometimes italic and sometimes not*

These typos will be fixed in the updated manuscript.