Response to referee review of bg-2016-54

Reviewer #2

General Comments

The paper under review for biogeosciences presents a literature review combined with laboratory and field tests to evaluate the application potential of three widely used automated equilibration systems to continuous long-term or underway $p\mathrm{CO}_2$ measurements. The paper is generally well-written and easy to follow, but I found some grammar and sentence structure issues and I am not a native speaker. I also found some inconsistences between the figure and the results descriptions (see specific comments). In addition, some objectives of the study were not achieved.

As you wrote “This study aims to review advantages and disadvantages of widely used $p\mathrm{CO}_2$ equilibration methods and automated equilibration systems that can be used for continuous monitoring of highly variable $p\mathrm{CO}_2$ across”. The ‘’short review” of this paper with the advantages and disadvantages of widely used $p\mathrm{CO}_2$ equilibrators is not a novelty for continuous aquatic $p\mathrm{CO}_2$ measurements. The studies of Santos et al. 2012 and Webb et al. 2016 (including others) presents laboratory step experiments on six different equilibrators to constrain CO_2 equilibration time constants and short reviews of the equilibration technique, including shower-head, marble and membrane type equilibrators.

I think you must focus on the new information that this paper provide about improvements in the aquatic $p\mathrm{CO}_2$ measurements, which are the long-term deployment of the equilibrators under various field conditions and biofouling with the membrane systems. You must to describe the equilibrator systems with details (the systems were poorly described).The figure 1 and the text did not present details of the measurement-systems, and I think this is very important. In addition, some tests were not performed for the marble equilibrator. I think that is important provide one or two tables with the field and laboratory test results.

You compare the drifts of the $p\mathrm{CO}_2$ results for the membrane equilibrators comparing the relationship between pH and $p\mathrm{CO}_2$ during successive 4-day monitoring periods following maintenance. However, pH measurement method is missing in the Method section. Since it is used to evaluate the reliability of measured $p\mathrm{CO}_2$, it must be evaluated more rigorous. Apparently, the problematic of long-term monitoring of $p\mathrm{CO}_2$ is still unsolved (the drifts of the results are very high if is not applied continuous maintenance of the measuring system).

- **Response overview:** We appreciate your constructive comments and suggestions. We will revise our manuscript incorporating all your comments and suggestions. The revised manuscript will be checked again by a native English editor to improve accuracy and readability. Some of your and another reviewer’s major comments are overlapped, so the same overview of our common responses is provided as below.

  (1) Review: There was a common critique on the novelty of our literature review; the review was evaluated as “somewhat limited” or “not a novelty”. We agree to your comment that equilibrator systems have been reviewed and assessed in other studies, but we would like to ask your attention to the fact that our review is the first effort that compares application potentials of the three gas equilibration systems for both underway and temporally continuous $p\mathrm{CO}_2$ measurements. To our knowledge, there have been rare efforts to review the three systems from theory to applications focused on freshwater systems. For examples, excellent assessments by Santos et al., (2012) and Webb et al., (2016) focused on the response time of various equilibration systems using laboratory experiments but lacked details on theoretical/technical backgrounds, power requirements, maintenance, and so on. We expect that this introductive review would help researchers initiating $p\mathrm{CO}_2$ monitoring study obtain both theoretical and practical information. However, if the editor and reviewers want us to remove or reduce the review section, we will follow the suggestion; we could incorporate the essential contents into the introduction section or keep only focal review components (e.g., applications of gas equilibration systems to continuous measurements) in a separate, but reduced review section.

(2) Additional monitoring data: In response to the comments on the lack of measurements by the marble-type equilibrator in comparing the performance of the three equilibration systems, we will include additional field measurements that would be useful when comparing the performance (e.g., response time) of the three systems.

(3) Methodological details: More detailed descriptions on our gas equilibration systems, together with other in-situ measurements such as pH, and analytical procedures and QC procedures, will be added in the Methods section, Table 1, and Figure 1.

(4) Target water systems: We used inland waters in the title because we also considered estuarine waters in literature review and our field study. For example, our study site includes a tidal reach of the Han River estuary (e.g., sites 10–12 in Figure 3 where underway investigation and long-term monitoring were conducted). We would like to keep this term, but will switch to "freshwater" if the editor and reviewers want us to focus on freshwater systems.

Specific Comments

Line 11: Replace for emissions.

- **Response:** The "evasion" will be replaced by "emission".

Line 18: ''. . .upper detection limit of the sensor". What is this limit?

- **Response:** We also think that this sentence would be confusing without more detailed descriptions, so it will be rewritten like "…for the river section where $p\text{CO}_2$ varies within the sensor detection range"

Line 17: The overall results suggest that the equilibrators are better suited for relatively short underway measurements than long-term deployment. Why? Do you have suggestions to improve the equilibration systems in order to long-term $p\text{CO}_2$ monitoring? I think you must discuss better this point.

- **Response:** Our suggestions based on additional discussion will be added in the revision. The sentence in the abstract will rewritten as follows:

  The overall results suggest that the fast response allows the equilibrator systems to capture large spatial variations in $p\text{CO}_2$ *during relatively short underway measurements, while technical challenges such as clogging and desiccant maintenance should be carefully addressed for their long-term deployment over several days to weeks. The membrane-enclosed sensor can be an alternative tool for long-term continuous measurements, if membrane biofouling can be overcome using anti-fouling measures such as copper-mesh covering.*

Line 26: First sentence confuse.

- **Response:** The sentence will be rewritten as follows:

The importance of carbon dioxide (CO_2) *emission from inland waters has increasingly been recognized in the global carbon cycle research.*

Line 27: I think the “respiration” is more adequate.

- **Response:** The change will be made in the revision (The emission of carbon (C) resulting from the respiration in inland waters…).

Line 26: I think “emission” or “degassing” is better than evasion.

- **Response:** The change will be made in the revision (The emission of carbon (C) resulting from the respiration in inland waters…).

Please review all the references. I found some mistakes.

- **Response:** We will thoroughly check any mistakes in the cited references.

Line 30: You wrote CO_2 and CH_4. I think is better write dioxide carbon (CO_2) and methane (CH_4).

- **Response:** The change will be made in the revision.

Line 35 and 36: Confused. You must explain better the principles of direct and indirect measurements. This sentence is not clear and not sufficient.
Line 36: “. . . between water and air and gas transfer. . .”? This is not clear, please rewrite.

- **Response:** The principles of direct and indirect measurements are presented in separate paragraphs (see Lines 35–49 and Lines 50–61, respectively). To clarify, the sentence will be changed as follows:

  The emission of CO_2 *can be determined either by directly measuring the transfer of* CO_2 *across the water-air interface or by estimating the flux based on (1) differences in the partial pressure of* CO_2 *(*pCO_2*) between water and air and (2) gas transfer velocity.*

Line 39-44: Please read and include information of Lorke et al. (2015) paper. There are important considerations about the floating chamber measurements and improvements on this technique to application for running waters.

- **Response:** Thank you for the useful reference. The paper will be cited as follows:

  Disrupted natural turbulence inside the floating chamber could cause the risk of overestimating CO_2 *flux (Vachon et al., 2010), especially when the chamber is anchored at a fixed spot (Lorke et al., 2015). Lorke et al. (2015) recently proposed improved floating chamber designs that minimize the bias of gas transfer velocity, including a freely drifting chamber on running water and an anchored chamber with a close contact over water surface. However, validation and further technical improvements are still required for practical applications of these proposed chamber systems.*

Line 50: from pCO_2 measurements.

- **Response:** The change will be made in the revision.

Line 51 and 52: You can also calculated pCO_2 from dissolved inorganic carbon (DIC) and total alkalinity (TA) and ancillary parameters. Please include this information. You can read Dickson (2010) to include more accurate statements about the indirect calculations of pCO_2.

- **Response:** The change will be made in the revision.

> *The pCO_2 can also be estimated using two of the three variables in the carbonate equilibrium model - pH, alkalinity, and dissolved inorganic C (DIC) (Lewis et al., 1998; Dickson et al., 2007, 2010).*

Lines 61 and 62: SOCAT?

- **Response:** The "Surface Ocean CO_2" will be added in the revision.

Line 70: delete "from polluted waterways" ?

- **Response:** The change will be made in the revision.

> *the evasion of CO_2 from "polluted waterways" in urbanized watersheds → the evasion of CO_2 from urbanized watersheds*

You did not present the results of the tests (you must insert one or two tables with the results of the field and laboratory tests).

- **Response:** The revised manuscript will contain all necessary data incorporating reviewer comments and suggestions.

Figure 1: This information is not sufficient. Please, provide more details about the measurement systems of pCO_2. For example, see the Figure 1 in Frankignoulle et al. (2001). As you stated that "Our review and cross-validation tests focus on three automated equilibration systems: spray- and marble-type equilibrators and a membrane-enclosed sensor (Table 1)" you must provide more details about the functioning and details of these three systems.

- **Response:** More detailed descriptions will be added in the revised Figure 1 and Table 1, including the photos of three systems that will be included in the figure or as the supplementary information.

Line 121 – 124: Bakker et al. (1996) measuring pCO_2 in estuarine waters, found "Frequent blockage of the showerhead of the equilibrator with algal material", adding some problems to the measurements. I would like to see some discussion about this problem with the equilibrators.

- **Response:** The paper will be cited in discussing potential sources of system failure due to blockage and contamination.

Lines 129 – 138: A figure with more details of the systems can better elucidate this section of the paper.

- **Response:** Specific features of equilibration systems that we used will be presented in Figure 1 and Table 1 with further details provided in the supplementary information.

Lines 150-152: Do you have some suggestions to turn the equilibrations systems (marble type and showerhead) more automated for long-term monitoring? Please discuss possible improvements that are necessary for long-term monitoring.

- **Response:** Some maintenance and technical recommendations will be provided in the discussion section.

Line 154-156: Again, here add one figure can better illustrate how is the passive membrane CO_2 equilibration systems, providing details for easy reproducibility.

- **Response:** The change will be made in the revision.

Line 159 – 161: "There are a small number of commercially available membraneeenclosed sensor systems (e.g., eosGP, Eosense Inc., Canada; Mini-Pro CO2, Pro-Oceanus Systems Inc., Canada)". What are the lower and upper detection limits of these sensors? They can be applied in aquatic systems where the $p\mathrm{CO}_2$ values can easily be higher than 10,000 ppmv?

- **Response:** The detection ranges can be adjusted by manufacturer, from low range (e.g., 0–1000 ppm) to high range (e.g., 0–20%). The sensors are able to cover the range of $p\mathrm{CO}_2$ in inland waters when the detection ranges are appropriately set and calibrated. This wide detection range will be mentioned in the same sentence.

Lines 220-222: A range of 2000 ppmv is "high" in inland waters. Then, this type of system (membrane-enclosed senor) could not be used in some environmental conditions. In addition, I would like to see some discussions about these overestimations.

- **Response:** The cited paper did not elaborate on potential sources of overestimation in the high $p\mathrm{CO}_2$ range, so we cannot provide any explanation here, but we will discuss in more detail when we evaluate our own measurement results from the membrane-enclosed sensors.

Lines 233-235: This information is not sufficient. You must provide details of the instruments.

- **Response:** Aforementioned more detailed information about the systems will be added in the revision.

Line 242 – 245. Are you sure that the unit is "mm"? One acrylic tube with this measure is very small, and I think cannot be filled with glass marbles. For example, in Frankignoulle et al. (2001) the vertical Plexiglas measures were: height 80 cm; diameter 10 cm.

- **Response:** Yes. We devised a smaller system modified from the original marble-type equilibrator of Frankignoulle et al. (2001). The smaller chamber was filled with smaller marbles (diameter: 10 mm) than the larger 20–30 mm marbles used for the bigger chamber. Other researchers have also used small systems for portability; for instance, Abril et al., (2006) used a chamber of 8 cm in diameter and 60 cm in height. We compared small-sized and original equilibrators in laboratory and we found no significant differences in performance between them. We will note this information in the revision.

Line 250-255: Provide a detailed picture of the complete system.

- **Response:** As mentioned before, the change will be made in the revision.

Line 252: Despite the fact that Johson et al. (2010) provided details of the membrane enclosed sensor, this is not sufficient for publish in biogeosciences. Your work must yield descriptions of the equilibration systems, both in text and in figures. Your third objective was "to compare the accuracy and maintenance requirements of three selected equilibration systems (a spray- and a marble-type equilibrators and a membraneenclosed CO_2 sensor) for field applications in a series of laboratory and field crossvalidation tests". I think that your objective is not just this, rather, I think that is also describe with details these three selected equilibration systems.

- **Response:** Again, we will add more information about the system in text and figure in the revision.

Line 256-257: "The CO_2 analyzers and sensors were calibrated in the laboratory using CO_2 gases of known concentrations (0, 500, 500, and 10,000 ppm) immediately before each laboratory or field test." Why two concentrations of 500? Did you make the calibration after the field test to see the drift of the sensors?

- **Response:** The mistake of the wrong CO_2 gas concentration will be corrected (500 → 5,000 ppm). Yes, the sensors were also checked following deployment and we found little drift for the deployment periods from which the presented results were obtained.

Line 262: About the laboratory test, why you did make the first test just to the membrane-enclosed sensor? This test is not well explained, please rewrite.

- **Response:** The sentence will be rewritten to describe the objectives and procedures of the tests in more detail.

Line 275: you did not perform the field test for the marble-type equilibrator. One of your objectives was not realized, since you compared the field tests just for 2 equilibrators. Despite the fact that you assumed that both types would exhibit similar results based on the laboratory results, this cannot be true in field conditions.

- **Response:** We will add more field data obtained using the marble-type equilibrator in the revision. We hope it would satisfy your expectation.

Line 295: "The response time was determined as the full time (*t*100) or 95% of the full time (*t*95) it took to a final pCO_2 level that represents pCO_2 values exhibiting less than 1 % of coefficient of variation (CV) for 2 min." The full time (*t*100) is unusual for calculations of equilibration time.

- **Response:** As described in our response to another reviewer's comment, the response time is usually calculated based on the exponential decay or e-folding curve fitting. However, the ideal exponential decay curve did not represent the various response patterns observed in our field tests conducted in various water types and under various initial conditions. Therefore, we used our own criteria to determine the response time as the time required for pCO_2 to reach a stabilization point, as presented in the manuscript. We will provide more detailed descriptions and discussion of response time determination. This approach allowed us to obtain both t_{100} and t_{95}, as we presented both t_{95} and t_{100} results in the text. For Figure 4, t_{100} values were selected because they provided slightly higher R^2 and lower P-value for regression fittings than t_{95}. If you want us to replace t_{100} with t_{95} data, we will revise the manuscript accordingly.

Line 303: How were the prior tests of boat speed effect?

- **Response**: We determined the boat speed based on our own field tests and Crawford et al. (2010). Three systems exhibited comparable results until boat speed approaches the selected speed. This will be described in more detail in the revision.

Line 314: I think the upper detection limit of the membrane sensor are low and cannot be applied in several inland waters where the natural variations of pCO_2 are very higher than these limits.

- **Response**: The detection ranges can be adjusted by manufacturer, various detection ranges, from low ranges (e.g., 0–1000 ppm) to high ranges (e.g., 0–20%). The sensor we used and other commercially available sensors can cover the wide range of pCO_2 in inland waters when the sensors with proper detection ranges are well calibrated. This detail will be specified.

Line 320-322: This section is confuse. Please rewrite. What preliminary tests did you perform? Why the power supply and air flow dehydration were easiest for the spraytype equilibrator?

- **Response**: The sentences will be rewritten as follows:

  In preliminary tests the spray-type equilibrator was easier to maintain the desiccant that was used to dehydrate air flow compared to the marble-type equilibrator. Logistic considerations also let us opt for the spray-type equilibrator in spot measurements at 12 sites.

Line 324: Why you did not test the marble-type equilibrator for the long-term measurements?

- **Response**: Our tests with both equilibrators were not successful for the same reason as described for the spray-type equilibrator. In our study sites, electric power supply was not available to run equilibrator systems for long-term periods. This will be mentioned more clearly in the revision.

Line 338-342: Move to results section.

- **Response**: The change will be added in the revision.

Line 345: Did you test the normality of data set? If not follow a normal distribution, you cannot apply the t-test. You must apply the non-parametric tests as Wilcoxon, for example.

- **Response**: Yes, we checked the normal distribution using Shapiro-Wilk test.

Line 349: What are the pH-pCO_2 relationships? Not clear in the text. Since it is used to evaluate the reliability of measured pCO_2, it must be evaluated more rigorous.

- **Response:** Following sentences will be added.

  The analysis was based on the assumption that robust pH-pCO_2 relationships would be expected from the carbonate equilibrium model if there were no artifact effects such as sensor biofouling. Temporal changes in pH-pCO_2 relationships were examined to assess biofouling-induced deviations from the robust pH-pCO_2 relationship.

Lines 363 – 365: The results in figure 2 shows that for low pCO_2 values the coefficient of variation calculated from the compared measurements were higher than 10%. However, in the text you not explain why this occurs for low pCO_2 values. Line 364: “The CV values were smaller than 5% at all sites except site 3 and 8. . .”. The graph did not show this. Sites 1, 2, 4, 7, 10, 11, 12 CV < 10%. Sites 5, 6, 9 CV < 5%. Sites 3 and 8 CV > 10%. For only 3 sites the CV values were smaller than 5 %, please correct.

- **Response**: We double checked and the modified sentence and an additional sentence explaining the large CV observed at two sites will be added as follows:

  The CV values were smaller than 5% at seven sites and < 10 % at three sites. At two sites where CV values were > 10, even slight differences in measured values between the three methods compared to the results observed at the other sites resulted in relatively large CVs due to the lower pCO_2 values than the atmospheric pCO_2.

Lines 373-374: “The response time of the spray-type equilibrator falls within the usual range of response times reported for the spray- (8 min; Santos et al., 2012). . .” Not really. You tests were approximately 4 time more rapid than that reported by Santos et al., 2012 for the spray-type equilibration. You should point some suggestions to explain this difference. Why you did not perform the equilibration time test for the marble-type equilibrator? Where are the results?

- Additional field data using the marble-type equilibrator will be included in the revision. Following change will be made in the revision.

  Mean $t_{95\%}$ and $t_{100\%}$ for the spray-type equilibrator was 1 min 31 s and 2 min 36 s, respectively, with no difference between standing and flowing waters. The response time of the spray-type equilibrator was shorter than the response time reported for a spray-type equilibrator (8 min; Santos et al., 2012), but similar to the response times of marble-type equilibrators (2 – 3 min; Frankignoulle et al., 2001; Abril et al., 2014). The difference in the response time might result from various factors including the different levels of pCO_2 (~100–5000 µatm in this study vs. > 10,000 µatm in the Santos et al.), equilibrator size (251 vs. 1963 cm^{-3}), length of air circuit, and spray nozzle performance.

Line 384: The figure 4 not showed a logarithm curve, rather, showed a linear tendency.

- **Response**: The curve appear linear because of the log-scaled X axis. To clarify, “note a log scale for the x-axis” will be added in the figure caption.

Line 385: “. . .with steeper increases observed for the membrane-enclosed sensor, particularly in flowing waters.” Is not the contrary? The steeper increase seems to be to the standing waters (Figure 4; red circles)."

- **Response**: Yes, the increase was steeper in the standing waters. The sentence will be corrected in the revision.

Line 394: and expected range of $p\mathrm{CO}_2$ levels.

- **Response**: The change will be added in the revision.

Line 405–407: “although it failed to respond to rapid $p\mathrm{CO}_2$ increases from the relatively low value at the confluence (11:57) to the concentration peak (12:25) due to the limited detection range.” Not just this. If you look at 12:15 and at 12:30, the deviation seems substantial also when the $p\mathrm{CO}_2$ values decreased abruptly. You just discussed the deviation when the $p\mathrm{CO}_2$ rise, and not when the $p\mathrm{CO}_2$ decrease.

- **Response**: The sentence will be revised according to your comment.

  In contrast to the long response times observed for discrete measurements at the 12 sites (Figure 4), the membrane-enclosed sensors reasonably performed well across most of river sections where pCO_2 changed gradually (Figure 5). However, the measurements by the membrane-enclosed sensor exhibited substantial deviations from the measurements by the spray- and marble-type equilibrators where pCO_2 changed abruptly during the period from12:20 to 12:35.

Line 430 - 432: What is the explanation to this drift? Not explained in text.

- **Response**: We will provide more explanations in the revised discussion.

  During the field test, extraordinary algal blooms occurred as a combined result of a severe drought, warm temperatures, and high loads of nutrients discharged from water treatment facilities and polluted tributaries draining the Seoul metropolitan area. Chlorophyll-a concentrations increased from 21.1 mg m^{-3} on June 2nd to 46.7 mg m^{-3} on July 2nd (Water Information System of Korea; http://water.nier.go.kr). The “membrane sensor” might have been *more prone to biofouling of the sensor membrane by planktonic and associated bacterial communities than the other sensor protected by the Cu-mesh screen. Enhanced production or consumption of CO_2 around the sensor membrane might have amplified diurnal fluctuations of pCO_2, leading to an increasing divergence between two sensor measurements as the time from maintenance increased.*

Lines 432-434: "The duration during which relative differences of day-averaged pCO$_2$ between the two sensors. . ."? I did not understand this section.

- **Response**: The sentence will be slightly revised in the revision.

  Relative differences in daily means of pCO_2 between the two sensors remained within 10 % during 5, 2, and 7 d since the routine maintenance on the 153th, 169th, and 182th day of the year, respectively.

Line 436: How you measure pH? What is the accuracy of the method? As you used the relationship pH-pCO$_2$ to examine the increasing biofouling effects with progressing time following the maintenance day, you must provide this information. Also, you need to show that the pH sensor not drift with time.

- **Response**: We used YSI 6820, as written in Line338. The accuracy of the pH probe was double checked on site with pH buffers and concurrent pH measurements using a portable pH meter (Orion 5-Star, Thermo Scientific, USA). This detail will be provided in the revised method section.

Line 438: How the biofouling can produced additional CO_2 molecules? Explain in the text the process.

- **Response:** Explanation will be added in the revised text.

  If additional CO_2 molecules were produced by enhanced microbial activities in the biofilm formed over the surface of the sensor membrane, this could disturb the usual pH-CO_2 relationship that can be explained by the carbonate equilibrium (Nimick et al., 2011).

Line 442-443: You pointed that "the method validation would require concomitant pCO$_2$ measurements using other equilibration methods". You had all the possibilities to validate this method, but not did, i.e., you had large pCO$_2$ variations and you had three equilibrator types.

- **Response:** We could not validate our sensor measurements with concurrent measurements using other methods. This limitation, together with further discussion of method validation, will be added in the revision.

Line 446: "Repeated maintenance visits at short intervals of 3 – 5 d may be required for a long term deployment of the sensor without antifouling measures in an inland water site with high levels and large diurnal fluctuations of pCO$_2$." This is difficult depending of the study are. Do you have other suggestions?

- **Response:** That's the reason we recommended the antifouling practice. Short-interval maintenance would be very difficult to implement in many cases; therefore, antifouling measures should be prepared for long-term observation at eutrophic waters. Our suggestion was presented in Lines 450–460. The paragraph will be rewritten for clarity in the revision.

Line 460: Interesting result. Can you plot the graph showing these results for oligotrophic waters?

- **Response:** The results will be added as an additional figure or supplementary material in the revised manuscript.

Table 1. Insert one column with the equilibration time for each method.

- **Response:** The change will be made in the revision.

For figures 2, 3, 4, 5, 6, 7, and 8 I think is better a white fill, without grades and with black contours.

- **Response**: The figures were drawn using default setting provided in ggplot2 of R. The default gray background theme was based on studies of visual impacts as stated by Wickham (2009; p. 141; original text provided below). Unless the gray background disturbs readability of the figures, we would like to respect the developer's intent. Nevertheless, we are open to revise the theme of the figures.

  This (very light grey background with white gridlines) follows from the advice of Tufte (1990, 1997, 2001, 2006) and Brewer (1994a); Carr (1994, 2002); Carr and Sun (1999). We can still see the gridlines to aid in the judgement of position (Cleveland, 1993b), but they have little visual impact and we can easily "tune" them out. The grey background gives the plot a similar colour (in a typographical sense) to the remainder of the text, ensuring that the graphics fit in with the flow of a text without jumping out with a bright white background. Finally, the grey background creates a continuous field of colour which ensures that the plot is perceived as a single visual entity. (Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York, p. 141)