Review of the "Evaluating the Arabian Sea as a regional source of atmospheric CO2: seasonal variability and drivers" by Alain de Verneil1, Zouhair Lachkar1, Shafer Smith2, and Marina Lévy3

This study analyzes the spatial and seasonal distribution of the air-sea CO2 exchange in the Arabian sea. This study is relevant and original both because, on the one hand, it is carried out in the Arabian Sea; region poor in data and poorly known. On the other hand, this study covers a large panel of aspects both in space and time. They first validated their model to observational pCO2 data and climatologies in addition to other variables such as SST, SSS DIC. The spatio-temporal variability of the CO2 flux is then analyzed for the entire domain of the AS as well as for different subregions with a quantification of the wind speed and pCO2 contribution. A further details analysis of the pCO2 spatio-temporal variability is also performed with a finer decomposition into processes such as temperature, circulation (upwelling effect), biology, …

However, this study presents a lot of information that goes a bit in all directions with a lot of sentences without justifications behind their strategies (e.g., line 106 value of 2 µatm) and a lack of quantitative values to support the text in the results and discussions sections.

I recommend this publication in "Biogeosciences" after major revision.

1

Lines 25-31: clearly need quantitative values to support the text.

Lines 25-26: "*and up to the present has on average acted to uptake excess anthropogenic CO2 (Ciais et al., 2013; Khatiwala et al., 2009)*" Should add a value + references need to be updated (e.g., GCP)

Line 27: "*The Arabian Sea (AS) is a region of the ocean that has been found to naturally release CO2 to the atmosphere (Sarma et al., 1998), mitigating the ocean's role in moderating atmospheric CO2 accumulation*." Reference 1998 is very old. What about e.g., Cao et al 2020? What is the value of the AS source of Sarma? Without a value, it is difficult to evaluate the contribution of the AS.

Line 65: "*The need is further emphasized when quantities such as pCO2 can be utilized as a proxy for other things, such as community compensation depth (Sreeush et al., 2019b).*" Not clear, need to be developed if you mention that in the text.

Line 66: "*However, most recent studies…*" Should add references. And most recent "modelled" studies or "observational" studies? Not clear

Line 71-72: "*Sarma et al. (2013) found that coupled ocean biogeochemical models underestimated the air-sea CO2 flux.*" should add "in the AS."

line 75: "*peak of flux*". Which flux? Air-sea co2 exchange? Advecting carbon flux? Not clear

Line 83: should say "air-sea CO2 flux."

Line 84: "*the quantification of the air-sea flux will focus on the contributing factors of pco2 …*". Spatial contribution or seasonal contribution? Not clear

Line 84: should say "deltapco2 and wind" instead of pco2 and wind. What about the contribution of atmospheric pco2?

Line 84: should say "sea surface temperature". Same for salinity

Line 85: "*which of course also varies from region to region within the AS*". Should say "for the entire domain of the AS as well as its spatial heterogeneity within the AS".

Line 87: "*biological production/respiration*". What about calcification remineralization or Nitrification denitrification?

Line 91: "*resolved by the data*" Not clear which data? from which study?
line 91 In the beginning of the introduction, you refer that data in the AS are very old (20 years ago) and here you refer to study from 2015. Not clear

Line 92: « The rest of the paper » should use another sentence/to be reformulate.

2

2.1

Question: How many data in SOCAT and LDEO database? Monthly? Daily? annual values? what period of study? What is the spatial resolution? Do you include the coastal datasets?

Question: paragraph from line 110: Why in addition to the Socat and LDEO observation data, do you also add the climatology of tk09 and L15? what more do they bring? And why you used the climatology of TK09 since others and more recent are available (see e.g., Rodenbeck et al., 2015 which compare different pco2 climatologies). I do not understand why you choose TK09?

Question: Is the coastal domain included in your definition of the AS domain? if yes (e.g., line 157 you use the word coast), the pco2 climatology of Laruelle et al 2017 and Roobaert et al (2019) for the coastal flux should be considered here? You refer to TK09 and L15 but what about the coastal regions?

Comment: In the text, a clarification is need on the period in which you are going to work + spatial resolution + temporal and how you have treated the different datasets to switch, for example, to seasonal. Why are some databases used only for model validation while others for the flux estimation (table 1)?

Comment: This manuscript needs a paragraph about the spatial and temporal pco2 and FCO2 uncertainties.

Comment: I suggest moving section 2.4.3 before 2.4.1

Line 98: "*pco2*" should specify that it is sea surface pco2

Line 100: "*model pCO2 is calculated from DIC and TA*" how?

Line 101: "*in situ un-gridded data merged from SOCAT v. 2019*" why you do not discuss about this dataset in the Introduction. In the introduction, you said that there is in the AS only observational data from 20 years ago (JGOFS) and here you discuss about LDEO and SOCAT. Not clear

Line 103: "*SOCAT data was preferred to LDEO observations*" why?

Line 103: "*large overlap*" how many data? you should be more quantitative and more precise in the words used.

Line 105: "*… using reported sea surface temperature (SST) and S data included in the products*" S= sea surface salinity? which equation do you use to convert fco2 to pCO2? Not clear how fugacity and partial pressure are related to SST and S

Line 106: why a value of 2 µatm yr-1. Should be justified.

Line 109: "*The gridded products*" monthly?daily?annual grid? For which period?

Line 115: Which products/datasets do you used to calculate the air-sea flux?

2.2

Line 121: "*System-AGRIF (ROMS-AGRIF) 3.1.1.*" version 3.1.1? what means AGRIF?

Line122: "*Previously used in the region (Lachkar et al., 2016), the…*" should be **by** Lachkar et al.,

Line 125: "*advection of tracers"* in repeat twice in the same sentence

Line 125: what is the K-profile?

Line 126: what is KPP?

Line 127: 33° W or E?

Line 127: which previous studies? Need references.

Line 127: sometimes you speak at present sometime in future → same comment for all text.

Line 130: what is the name of the OBGCM model?

Line 131: "*Biological parameters for the model are the same as those used in Gruber et al. (2011)*." That is to say? Which parameters? Should be enumerated.

Line 136: what about nitrification and denitrification?

Line 144: "*climatological mode*". Not clear, what is the period?

Line 145: "*surface salinity*". In the text before it was S. Not clear

Line 146: SST is from 1985-1997 but then you make a FCO2 comparison for another period. Still confused about the period of this study. What is the spatial resolution of SST?

Line 147: Why do you need wind stress? The period of the wind stress is 1999-2009, different than SST. Not clear again.

Line 148: "*T, S*" same comment than before. SST or T .. not clear

2.3

Line 157-158: need reference from the literature to support summer monsoon and enhanced biology.

Line 162: same comment need reference for winter monsoon NPP

Figure 1: what is the period of the study?

2.4

2.4.1

Line 168: "*The proximate variables impacting pCO2 in the*" should be "*..that affect pco2 change in the"*

Line 170: What is the depth of the water column where the pco2 decomposition into processes is performed in the model?

Line 171: In equation 1, terms on the RHS are not described in the text and not clear when you said "*deviation from an average.*"

Line 181: how partial derivative are calculated? Which coefficient are used?

2.4.2

Line 197-198: what about the river flux? Did you include the contributions from rivers?

Line 202: Eq 6 what are "New" and "Reg" and why CaCO3precip.remin has the same sign than PPnew+reg (both with a -)?

2.4.3

Line 226: "*five-year average*" why five? Not clear

3

3.1

Question: Why did you analyze the variables in depth in Fig. S1 and Fig. S2 and not just the spatial distribution at the sea surface as it is performed for Fig. S3? What is the reason for the choice to do in depth?

Question: I do not understand between figure 2 and figure 3 for the data spatial distribution. Are they the same data between both figures? It seems there is more surface covered by the data on figure 2 than when they are divided into seasons (figure 3)?

Comment: Lots of imprecision on the form and problems with the numbering of the panels.
e.g., Figure 3. « *Seasonal surface pCO2 (µatm) from data (a-d)* » panels are not ok between the figures and the caption.

Comment: the pco2 Seasonality has already been analyzed by the previous figures and compared to the SOCAT database spatially and temporally with a Taylor diagram. Why do a comparison to LA15 again? What more does this bring? not clear in the text

Comment: I do not understand the added value of the Taylor diagram. What more does this figure add? moreover, you do not explain the various terms of the taylor diagram such as the Pearson coefficient. Why is it not done on the other variables, especially when you go directly to line 255: Since the model successfully replicates other tracers, physics, and biological processes". A little fast, no? In addition, the seasonality of these processes is not analyzed.

Comment: You start with fig S1 to S3 discussing variables then you discuss about pco2 and then you discuss again with variables (line 255). Not clear. Should be first pco2 then variable or vice versa

Line 230 "*The implementation of ROMS-AGRIF presented here has been used in previous studies (Lachkar et al., 2016).*" And? What is the conclusion of this sentence? Is it in the AS?

Line 233: Add panels with the mismatch (model minus data) in FIG S1 to S3 would be useful to highlight difference between model-data. Same comment for figure 3.

Fig. S 2. "*Transects, similar to Fig. S1, but with salinity instead of temeprature*" should be temperature.

Lines 230-233: e.g. "*Reproduce well the …*" need to add quantitative values.

Line 235: "*Regarding pCO2, in situ data from the merged SOCAT/LDEO database shows that average binned pCO2 values in the region are positive for most of the AS (Fig. 2a). The ensemble of observations show that 90% of _pCO2 observations are positive, indicating positive flux to the atmosphere (Fig. 2a, inset).*" These two sentences should be merged in one.

Line 239: "pCO2 value…" we do not know if it is modelled pco2 value or observational pco2 value. Maybe you could use "data-based pco2" and "modeled-based pCO2" everywhere in the text.

Line 242: Again, this paragraph needs quantitative values. In the pco2 model description, there is no value in the text. Use values instead of "Is lowest".

Fig. 4 "*Taylor diagram of modeled vs. observed surface pCO2, both in total and seasonal sub-sampling. Data are from merged SOCAT and LDEO databases, corrected to year 2005. Distance **frmo** origin ..*" should be from

Line 260: "*close to reported measurement error*" give a value.
Line 255: "*tracers, physics, and biological processes*" which are?
Line 257: Why the SST from SOCAT is used here from the socat database while in Fig. S1 it derives from the WOA? Not clear

Line 258: "*sst*" sometime sst sometime T. Not clear. Same comment for every variable e.g. Fig S4, SAL instead of S or SSS. Be consistent all text and figures.

Fig. S4. same comment SST or temperature… not clear

Line 267: Figure 5, I suggest merging the three plots in one and use monthly boxplot for instance for median, 5th and 95th. Easier for the reader.

Line 277: "*competiting product*" Not clear. what that mean? Observational? Model? Not clear

3.2
3.2.1

Comment: need to be more quantitative.
Comment: You should be careful when you use the term "dpco2" for change in pco2 spatial or temporally. You also use dpco2 for the air-sea CO2 gradient. Not clear e.g., Fig. 6 and Fig.2.
Comment: maybe specify in the text what corresponds to a positive anomaly and a negative anomaly on pco2

Line 286: T, S alk are variable but FW is a process. Maybe salinity instead of FW?

Line 284: In fig 6, maybe add numbers to regions for each box, then in panel b also add the number instead of name. Easier for the reader to say in panel b where is the regions in panel a

Line 286: "*The cold SST structure largely overlaps the stronger summer monsoon winds*." Not clear

Line 301: "*The previous section outlines why certain geographic regions within the AS have overall high or low pCO2 values*" This does not explain why. Just how the variables change within different regions but not in terms of process the section does not explain why. It DESCRIBES the pCO2 changes associated with the variables but not processes; pay attention to the word used. Question: what are the processes that explain the spatial distribution of pco2? Not clear

3.2.2

Comment: Again, this section needs numbers to support the text. E.g., what is the typical amplitude of the respective seasonal signal?

Question: How do you explain that the important contribution of DIC in the upwelling regions is not seen on the seasonality of the entire domain (Fig. 7a)

Line 302: "*the decomposition of factors affecting monthly*" should be "variable" instead of "factors"?

Line 303: Figure 7. "*Timeseries of pCO2 anomalies (µatm)*" units should be µatm month-1

Line 304: "pCO2" should be called "pco2 anomaly"

Line 305: "*DIC acts*" DIC is not a process. Should be « change in pCO2 associated to DIC change»

Line 307: you first discuss about Fig.7a in the text for the entire domain. Then you discuss about the oligotrophic central region. Hence, Fig.7d should be Fig. 7b. The different panels in figure 7 should be in order of what they are called in the text.

Lines 320-326 Maybe this section should be in the discussion section instead of the result section.

Line 328: "*is controlled by the physical processes of surface forcing, mixing and advection*" not clear what means the term "surface forcing"?

Line 330: Figure 8. "*Timeseries of DIC fluxes (PgCyr−1)*" why the DIC flux is in Pg C yr-1? Should be inPg C month-1? Why do you call this variable "DIC flux" instead of "delta DIC" such as "deltapco2" before?

3.3

Comment: Again, this section needs numbers and more precise.

Line 330: "*Budgets of DIC fluxes in the upper 20 m (Fig. 8) show that two major processes dominate, vertical circulation (light blue lines) and net biological processes (green lines)." I'm not sure that the term "budget" is appropriate here. This figure shows the monthly decomposition of DIC change into processes which is not a budget. Also, it is not clear when you said, "the two major processes dominate". They do not dominate; they show the largest seasonal variations but they not dominate the final seasonal signal.*

Line 330: Fig. 8 complicated to compare one region over another. All panels should be the same yaxis scale.

Line 332: "*acts as a source of DIC*" when? All time? Which month? Should be more precise.

Line 341: "*escape is mostly smaller than biological flux*," not clear. Smaller of what? In term of seasonal amplitude? For a specific month? Again, need to be more precise in the text.

Line 351-357 Maybe this section should be in the discussion section instead of the result section.

3.4

Comment: Again, this section needs more value to support results than only a qualitative description.

Comment: I propose to change the order of the different sections. First, what governs the variability of FCO2, then pCO2. So, I suggest to first start with 3.1, then 3.4 then 3.2 and 3.3.

Line 370: what about the contribution of atmospheric pco2?

Line 372: Figure 10 "*(a) Anomaly of air-sea CO2 flux during summer monsoon JJAS (molCm−2yr−1. Summer flux anomaly contributions due to (b) wind, (c) pCO2, and (d) cross-terms in Eqn. (8)".* Could say what means a positive/negative anomaly. Also (molCm-2yr-1).

Line 374: "*The _pCO2 contribution to*" dpco2 or pco2? Not clear. What about the atmospheric CO2 contribution?

Line 378: Figure 11. "*Seasonal CO2 flux anomaly (purple) for winter DJFM monsoon (top-left), springAM(top-right), summer monsoon JJAS (bottomleft), and fall ON (bottom-right). Contributors to the flux are solubility/winds (k_,blue), pCO2 (red), and cross-terms (yellow)." Results are shown for entire domain and sub-regions.*" Problem with panel "top left" "top right" in the caption and the figure.

Line 378: Figure 11. It is complicated based on this figure to see which process controls the seasonal flux variability. This figure represents the budget for each season. It would be easier to represent the seasonal flux variability into processes using the same strategy than for the seasonal pCO2 anyalsis with seasonal profiles as realized for pco2 (Fig 7 and 8).

Line 378: "*The seasonal flux anomalies for all regions throughout the year are displayed in Fig. 11"* should be "aggregated by season" instead of "*throughout* the year".

Line 393-395: Maybe this section should be in the discussion section instead of the result section.

4

Comment: Lots of repetition of results in the discussion section. Should be reduced. I also propose to reorganize the discussion section in 1) model vs data, 2) spatial analysis (FCO2 and pCO2) and processes, 3) seasonal analysis (FCO2 and pCO2) and processes and include in 3) the point 4.3

Question: according to your results, does the seasonal FCO2 and pCO2 variabilities and processes diverge between the open ocean and the coastal domain in the AS?

Line 406: what is the order of magnitude of the bias of Valsala? same order as you? Need quantitative values.

Line 428: "*Since upwelling regions are limited in geographic extent near the coast, capturing their high pCO2 values can be difficult for other approaches, such as TK09 with its coarse grid*". Same comment than before. Why did you use the TK09 climatology instead of others finer spatial resolution climatologies. Also, since the upwelling is close to the coast, what about the coastal climatology of Laruelle et al. (2017)?

4.2

Line 434-445: should be in section 4.1

Line 434: "from previous studies" observational or modelled studies? Not clear

Line 446: what are the consequences on this seasonal mismatch on your decomposition into processes?

Line 462: "*influence of outflows from marginal seas*" how?

Line 470: "*the dominance of temperature is clear*" spatially or seasonally?

Line 471: Be careful when you compared processes on different time scale (seasonal vs interannual)

4.3

Line 497: "*and so no biologically-induced decrease of pCO2 occurs*". Maybe add "in the final pCO2 signal".

4.4

4.4.1

Line 512: "*where uptake of CO2 takes*" should be "atmospheric CO2 uptake takes"

Line 518: Fig. 9d instead of Fig. 10d

Line 521: "*The fact that model CO2 flux peaks in summer despite a wide-ranging spring peak in pCO2*" For the entire region? Which region? Not clear

Line 535: need more reference for the use of different k-parametrization and wind speed on the flux calculation (e.g., wanninkhof, Ho, Roobaert, …)

Line 521-535: Need more comparison to other studies. e.g, you could compare your results with the study of Roobaert et al. (2019) that decompose the CO2 flux seasonality into processes in the region.

4.4.2

Question: What is the real purpose of this section 4.4.2? is the objective to calculate an uncertainty on the flux? if so, why only focused on the choice of pco2? why you chose TK09 instead of that of Rodenbeck (2014) which is finer in terms of spatial resolution and more recent. What about the uncertainty associated with the other terms involved in the flux calculation? the wind? its spatial and temporal resolution? the k-parametreization? Etc…

Question: This study highlights that the effect of the wind governs the seasonal variability of the flux. Why is it important to do, as it is performed here, a finer pCO2 decomposition into processes?

Line 540: "*Considering the important role of winds*" on the seasonality? Spatiality? Not clear

Line 542: Fig 12: should be "Tg C month-1" instead of "TgC"

Line 546: "*Despite differing pCO2 seasonality*". In this figure only LA15 is referred. Not the other products. So you can not refer as something common between all pco2 products

Line 560: «*1) gridded data-based pCO2 products will under-estimate the upwelling zone maxima of pCO2 and CO2 flux during the summer*" how can you conclude that since you said before that your model produces higher pCO2 values compared to other studies. What is the value/study for this conclusion? Not clear

5

Line 573: 160 while in line 550 it is between 57-120. Not clear.

Line 574: need references.