

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2021-328-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Review of Sharp et al „A monthly surface pCO₂ product for the California Current Large Marine Ecosystem”

Anonymous Referee #1

Referee comment on "A monthly surface pCO₂ product for the California Current Large Marine Ecosystem" by Jonathan D. Sharp et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-328-RC1>, 2021

Sharp and co-authors use a random forest regression approach to map the sea surface pCO₂ for the California current system at a high 0.25x0.25 degree resolution. The authors extensively test their approach and compare it with global mapped products from the literature and conclude that their regional approach provides a substantially better representation of the sea surface pCO₂, particularly when compared to local measurements.

I believe this is a novel approach and an important analysis and should therefore be published. While it is not particularly surprising that this new local approach outperforms global approaches from L17 and L20, the message is important and clear: Don't rely on global approaches if you are interested in local features.

I am particularly impressed about the effort the authors put into evaluating their approach. Withholding independent observations or full years really guarantees that the observations seen by their machine learning method are truly independent. This should indeed be the standard – well done.

I only have a few minor comments listed below:

.) I am missing a rationale why random forest regression was chosen? It clearly was a reasonable choice, but I was wondering whether there was a particular motivation to pick this method out of the many available?

.) Uncertainty: Firstly, I believe the authors overestimate the uncertainty. As shown e.g. by Landschutzer et al 2014, the larger scale (or the full region) error is also dependent on autocorrelation features of the error. Say e.g. your gridding uncertainty for the open

ocean is $4.8\mu\text{atm}$ (based on the std of each grid cell). The standard error for the entire region investigated, however, would scale with the number of uncorrelated grid cells, such that the error $E = \text{std}/\sqrt{N}$ where N is the number of independent grid cells. For an entirely random (independent) sample of grid cells, N might be very large and the error very small (since grid cell errors cancel each other out). For a dependent (correlated) sample of grid cells, N might be small. So, in a nutshell, by not accounting for the N effect, the authors overestimate their uncertainty. This should be stated

.) Additionally, the authors do a lot of effort calculating the uncertainty, but do not always display it in their figures – Figure 2 and B3 should have an error-shading (corresponding to measurement errors – plus gridding errors for the mapped products) as well, so the reader can see whether differences are within the respective uncertainty.

.) Lines 11-12: Not necessarily – other methods (not relying on the pCO_2) exist as well

.) Lines 22-23: “alternative global products” – I agree with the sentence but would remove “alternative”. Global products are not an alternative for regional efforts, but target a different research question (e.g. the global ocean CO_2 uptake)

.) Lines 32-33 – It should be noted that the 25% refer to the annual uptake

.) Line 66: The Rodenbeck product actually covers the coastal ocean, but due to its coarse resolution it is not considered a coastal product

.) Line 110: It is more relevant to state how many SOCAT observations exist in the study region

.) Line 155 – Table 1: Probably not an issue, but are there any missing chlorophyll values in the study region?

.) Line 198: “most similar pCO_2 observations” – what does this mean and how is this determined?

.) Line 365 and following: fair point to use SOCATv4 and SOCATv5, but this only partly compensates for the fact, that the RFR-CCS method still was trained with a newer dataset (and the global estimates with observations from the globe)

.) Lines 389-394: I agree with point 1 but disagree with point 2 (for reasons outlined in Gregor et al 2019). Of course local phenomena will only be better represented if a local reconstruction approach is used, but I have my doubt that exploring new methods will overcome this issue – The authors can easily test this by applying their approach to the full coastal domain and then compare the error statistics in the study region.