

The Carbon Cycle in Mexico: past, present and future of C stocks and fluxes.

Murray-Tortarolo et al.

Reviewer 1

We are very grateful with the support the reviewer has expressed to our paper. It is encouraging to have an in-depth review of our manuscript We provide our responses in bold-type to the comments below.

- 1) One question is about which scale is adequate for a “comprehensive” understanding of the C balance of a country: is 1x1o grid enough or we should aim to higher spatial resolution?

Although we agree that a finer resolution analysis could improve our understanding of the C balance in Mexico, most current models and the driving data needed to force them are on either a 0.5x0.5 or 1x1 resolution or higher. In addition, increasing the spatial resolution may not necessarily yield better results, as finer-scale processes may or may not affect larger scale processes (e.g. landslides usually have local scale effects). However, DGVMs are moving in this direction, i.e. application at finer spatial scales and working towards accounting for disturbances at the landscape scale.

Nonetheless, we feel there is an error in the interpretation of our use of the word “comprehensive”. What we meant by “comprehensive” was the need for an analysis that includes multiple independent sources of evidence and takes into account multiple processes, e.g. the C stocks and C fluxes, plus interpretations in terms of the driving mechanisms behind them, which we feel we achieved. Very few studies in Mexico, and only very recently, have

started to measure larger scale C fluxes and certainly not at the country scale. We will clarify this on the text and remove the term “comprehensive”.

- 2) As I said, this is an interesting study, but I think the methodology has to clearly define the datasets used and account for the uncertainty in upscaling of information and the quality of the data (beyond standard deviations on the estimates from models).

We fully agree with the reviewer and are grateful for this observation. We will add a full description of the uncertainties and limitations of all the datasets we used, along with links to the original data. We will also include a table where we state the origins of each dataset, its advantages and main limitations for this particular study.

- 3) This country has produced and updated gridded data and newer datasets (produced during the last 6-10 years), but these have not been used in the current study (e.g., de Jong et al 2010, MexFlux, Mex-LTER, CONAFOR, INEGI, and CONABIO information). Newer datasets (and those revised for AQ/QC) could be more accurate than those used by the authors; mainly those that describe vegetation types, C flux data, LUC, and soil C. I understand that the authors used datasets that were freely available, but I want to bring to their attention that this country is doing a large effort on standardization and updating of datasets.

Although we agree with the reviewer that other products are available for the country, these are not publically available and have not been published for evaluation in international journals. Additionally, some of the datasets the reviewer mentions (e.g. MexFlux, Mex-LTER) are not national products, but site-level. Analyzing the patterns for individual sites goes beyond the scope of our work.

- 4) That said, I recommend being cautious with the interpretation of the results. I suggest rephrasing the title, results, and discussion to include the main goal of this study: This study is a “numerical experiment with currently freely

available data”. These results, may or not, change if new and revised country-specific data is used from vegetation types, to ecosystem fluxes, to soil carbon, but need to be tested with other datasets. The main limitation of the study is that the results are not systematically tested with available “land-truth” data, and in several cases cannot be corroborated (as stated in the manuscript).

We feel that the expression “numerical experiment with currently available data” does not represent our study correctly, but we take the point that we need to be clear on exactly the type of evidence presented. We show consistency across multiple datasets for C stocks and fluxes, we also proposed mechanisms of change over the past century (attribution of change) and show potential changes over the future for several RCPs. We computed NBP from different sources (literature, models and atmospheric inversions) and the discrepancy across them. We therefore state that we primarily use data from numerical models (DGVMs and atmospheric inversions), necessary for future projections, but complemented with ground-based information over the historical period.

Regarding the “land-truth” data, it is not possible to compare our results for the last century and the future, as it is not available. For the present, we included data from the national forest inventory for biomass and soil C, which we consider to be “land-truth”. Also, we compared our estimates with available site level information, which is not widespread either. This is especially true concerning C fluxes estimations, only recently published (eg. Reimer et al. 2015, PLOS One), and which we would include in a new version of our manuscript. We agree that our analysis would benefit from analyzing changes-over-time in those stocks, but such data is not yet available for the country. That said, we will add a section on limitations of our study, where we will state that our results need to be reassessed in the light of more available data.

5) I call for transparency, reporting of uncertainties and assumptions, and to be careful on the interpretation of the results. A simple way to address these

issues is to mention that is a numerical exercise that needs to be validated with future data and that there are X uncertainties associated with this work. This is of critical importance because there may be convergence on the results from the approaches because all of them give the answer because of the “wrong” reason. Therefore, error propagation and uncertainties must be included in this study.

We have already addressed the comments regarding our work as a numerical exercise and about the uncertainties in points 3 and 4, above

We strongly disagree with the argument that all approaches could provide the same answer because they are all “wrong”. For example, it is highly unlikely that datasets on biomass C from sources as different as models, satellite and field inventories, generated with completely different methodologies, can yield similar results because of similar biases.

I also encourage the authors to include hypotheses into their questions of the past, present and future. We cannot test the future but we can test the hypotheses for the past and the present. I am mainly interested in reading about why the effect of CO₂ fertilization is larger than LUC during the last 100 years.

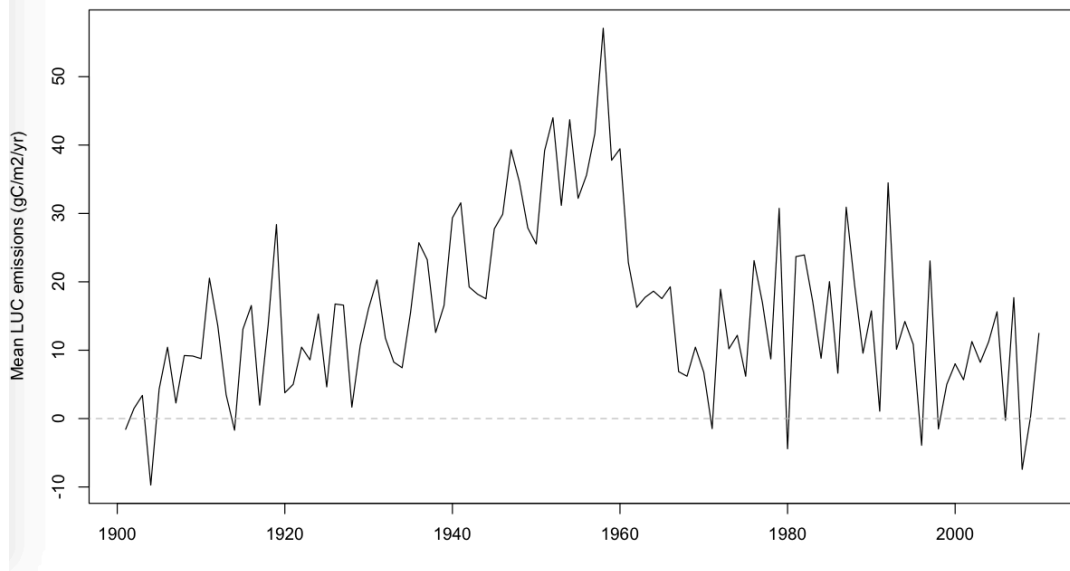
We appreciate the reviewer’s idea on proposing hypotheses for our work, however, we feel this goes beyond the scope of our study, which was not designed to test specific hypotheses. In addition, Reviewer 2 states that our questions are clear and our objectives are well defined.

The effect of CO₂ increasing vegetation NPP and biomass has been well characterized in laboratory and field experiments (e.g. Norby et al., 2005; Ciais et al., 2013) and was proposed as a major driver of change in land-C over 20 years ago (Friedlingstein et al., 1995). Over the last 100 years CO₂ increased by about 40%, which means that vegetation should have gained C as well. On the other hand, LUC in Mexico has shown a general decline over time in the last 50 years (with a maximum peak over 1940-1960), leading to

smaller LUC emissions and to areas with forest re-growth. As a result, it may not be surprising that the CO₂ fertilization effect is greater than the effect of LUC, which is also consistent with global patterns (CO₂ effect (~+3PgC) is almost 3 times larger than LUC (~-1PgC). We appreciate there are active discussions on the relative attribution of large-scale carbon sinks in the research community (see earlier comment #3 about getting the right answer for the wrong reason, which is applicable to small net fluxes, like NBP, derived from opposing large gross fluxes), therefore we would include this in the discussion on uncertainties. Nevertheless, these results represent a synthesis of our current understanding from state-of-the-art large-scale numerical models.

- 6) During this period there have been policies and practices that have mandated and changed landscapes for a large part of the last century for Mexico. How the authors include this socioecological influence in the LUC past and future estimates in this country with large spatial heterogeneities?

The DGVMs were forced with LUC from the HYDE (cropland area) dataset and some used in addition a consistent Hurtt et al. (2003) dataset for wood harvest and land-use transitions. These datasets use a mixture of process-based and FAO country statistics to calculate the transformation of forest, agricultural areas, pastures and natural grasslands to other categories. As our results show, the implementation of the HYDE dataset in the models, do display a similar spatial pattern to the agricultural areas of the country from other estimates (e.g. INEGI), and also show a temporal pattern expected for the country (i.e., high LUC over 1940-1960 that slowly decreases over time) (see figure below). We will add a sentence stating that we use one realization of historical land-use and land cover change based on HYDE, which is consistent with other estimates (e.g. INEGI for temporal changes in agricultural extent) and will add the figure below as supplementary information.



7) Finally, there is no attempt to include disturbances that are common in this country and substantially influence carbon stocks (e.g., fires, hurricanes, land slides) in the past, present (e.g., the 2005 hurricane season) and future. Here, I list several comments that I hope the authors find useful to improve the manuscript.

Some DGVMs do include a mechanistic representation of wildfire (we will include a figure as supplementary material to specify this). One of the main effects of hurricanes on the C-cycle is the addition of high volumes of water, which is included in the forcing data we use. The impacts of disturbance (wind throw, land-slides) are not currently represented in DGVMs; however, as these are important processes particularly at the landscape scale, which are not widely represented in large-scale DGVMs, these will need to be included in future assessments. Nevertheless, their potential effects on the carbon cycle are implicitly included in atmospheric inversions from CO₂ monitoring stations.

8) The authors discussed that previous studies that have placed Mexico as a source of C may be biased as they have derived this conclusion for estimating Carbon fluxes from biomass change only. The authors also state that these

previous approaches did not take into account soil C dynamics, the effect of CO₂ fertilization on GPP or the impacts of climate change. How certain are the calculations of the authors for LUC and soil carbon that make this study less biased than previous ones? As the current results stay there is a large mismatch with soil C estimates, so how this approach reduces bias?

We agree with the reviewer that the way we phrase previous results sound as if they were biased. What we actually meant is that their results are correct using their methodological approach, but they are incomplete, since they need to include additional processes not taken into account (i.e. to define whether a location is a net source or sink for carbon one needs to include the soils and the fluxes, not only the biomass stocks).

Our work is based on processed-based models that include several components of the land-C and the hydrological cycles. As a result, we can compile a value for NBP that takes into account the dynamics of the vegetation and the soil, plus their interaction with climate, CO₂ and LUC. Although the values for the C-estimates show some level of mismatch, these are C-stocks, while NBP comes from the C fluxes (i.e. measuring changes in biomass or soil C-stocks does not necessarily account for NBP). One key piece of evidence that shows our results to be less biased is the close value obtained for NBP from multiple sources, e.g. atmospheric CO₂ inversions and current literature estimates.

- 9) There is a large mismatch in scales representing the past, present, and future. The past are 99 years, but the present are just 5 years (2005-2009), and the future takes into account years 2010 to 2015 and extend into 2100. I think the authors need to reconsider the time scales presented here; mainly the present that I would argue at least should consider 20 years (1995-2015) which are the years that Mexico has had forest inventories (see de Jong et al 2010, CONAFOR 2015) and could be used to validate the study. Therefore, the future should be at least after 2016. It will be very interesting to see if the results for question 1 change if the analysis is done with the last 20 years rather than with 5 years. This is very important because during the last 15

years Mexico has implemented several policies and programs that have affected carbon dynamics and available measurements to corroborate modeling efforts (see de Jong et al 2010, Vargas et al 2012).

The time scale for the past includes the last century (100 years and not 99) and the future expands for the 21st century after our analysis on the present (90 years). It is unlikely that the patterns we find for the next century will change if we add 5-10 extra years, as they are strongly determined by what happens over the period 2050-2100. Hence, we can talk about “full centuries” about the past and the future in our study.

Concerning the time-scale for the present (2000-2005), we agree that we should use additional years (as we did for NBP), however, we are constrained by the extent of the land data. We will include additional information on the spatial patterns for C in the vegetation and soil for the period 1990-2010 for the available products and will compare with the results for the shorter time-period.

Comments

The questions are definitively interesting, but as results cannot be corroborated then several issues should be consider:

- 10) First, I think it will be crucial to disclaim the assumptions and the uncertainties of the study in the introduction and then in the results section. This is important because the current results contrast to several past publications and the authors state that previous results may be biased (page 1205).

We fully agree that a clearer definition of data uncertainty is needed in the text; we will also discuss briefly the methodology used by the different authors to get their estimations and the limitations of each.

- 11) Second, I think it is important to clearly state the hypothesis of this study. For example, why CO₂ fertilization on GPP could be more important for carbon dynamics than land use change? This issue is of critical importance in a country with large heterogeneity and with high land use change that was motivated by government policies; mainly during the first part of the past

century (i.e., the “past” period in this study). It is critical not to ignore the history of land use change in this country that precedes satellite information and “present” inventories.

Please see points 6 and 7, above.

12) The authors wrote in the introduction that doing this study for Mexico is important because is a country with large heterogeneity, but in the methods section (page 1206) the authors reduce heterogeneity from 10 categories (derived from Ramankutty and Foley 1999) to five. Furthermore, a 1x1o grid largely reduces and combines heterogeneity. I understand that this was done for simplicity, but it somehow contradicts the motivation of the study in this heterogeneous region. This will add uncertainty that it is not accounted for and will need explanation. Also why the authors did not use country-specific information of land cover (see CONABIO and INEGI information).

The Reviewer is correct, but we might add that the large heterogeneity not only arises from the vegetation land cover, but also from the climatic and orographic particularities of the country (all taken into account in our forcing). In addition, the usage of our vegetation dataset and the reduction to 10 land categories has the advantage that it can be applied to any other regions as it matches the general PFTs used by other models. Additionally it helps synthesizing the information into well-defined vegetation functional types that facilitate the nationwide analysis. Nonetheless, we agree that a further explanation on how the heterogeneity arises and a better justification of our methodology is needed, which we would add.

13) I encourage the authors to clearly state the assumptions and include calculations of uncertainty in their estimates. I totally agree that there is limited available information, but assumptions on the methodology must be transparent when doing a comprehensive country-specific estimate. There are several issues that need to be disclaimed and taken into account.

Please see point 2.

14) DGVMs, Earth system models, atmospheric inversions: I fully understand and agree on the application of these approaches, and the numerical experiment for DGVMs (Simulations 1-3), but the authors need to clearly explain that these are simulations that are hard to evaluate in this country. For example, there is the issue of past LUC that I discussed earlier and it will be very interesting to read about how this was addressed. The product MOD17 has not been tested in ecosystems of these country. It is unclear the performance of the VOD to calculate aboveground biomass in this country (as I understand that the authors used the same algorithm that Liu et al 2011, 2013). The MTE was parameterized with worldwide towers, but not country-specific information, this is important because the authors state that they use available flux tower data (see MexFlux for country-specific data).

We fully agree with the reviewer that more transparency is needed regarding the uncertainty of our data and we would add this to the text.

15) The source of the soil carbon data set is unclear. I was not able to find these data in the cited reference (SEMARNAT 2002). To the best of my knowledge it seems that the cited dataset is derived from road inventories (maybe provided by Carlos Ortiz) and the scale of 1:250,000 is not correct. To the best of my knowledge this data could have been taken from a 1:1,000,000 scale from Cruz-Cardenaz et al (2014) but transparency and clarity is needed. De Jong et al (2010) used a systematic nation-side grid (see de Jong et al 2010 for comparison using INEGI data from a systematic grid) and there is INEGI systematic data for nation-wide purposes. Second, there is no information about the quality assurance and quality control for that data as the cited reference is usually used for data on soil degradation (SEMARNAT 2002). Finally, the dataset is in %C but the authors used the FAO map to extract bulk density and transform the data set values to mass. This creates several issues: a) which soil depth was used?, b) which is the uncertainty added when using this generic information?, c) how the authors upscaled the information from 4000 sampling point to the national scale?, How does the authors combined the soil C field data with DGVMs and FAO dataset (page12508)? These

challenges influence the fact that there are large differences for estimations on soil C across the used approaches, but is difficult to evaluate as methods do not explain in detail this approach. d) Finally, how point-measurements and dataset were re-gridded to a 10x10 grid? Was this done by simple linear interpolation? How errors were propagated into this approach.

We will clarify the origins of this dataset (provided by Carlos Ortiz) and how it was post-processed. Regarding the specific questions: a) we used a 20 cm depth, as stated in the text, b) although we added a layer of uncertainty when adding the FAO dataset, the uncertainty added by using bulk density is relatively small and to our knowledge is the best way to proceed in order to get realistic values, c and d) we upscaled the values using a bilinear interpolation as stated in the text, this is a standard procedure to convert point to gridded data. Additionally, most of the country was well covered by the soil sampling (as shown in the supplementary material) with the exception of the desert areas in the North (however, here C values are small, hence less relevant for the country-level C balance). We will add additional information on the origins of the data and how it was processed (for all datasets). Certainly, it is difficult to have a “good” estimate for soil C stocks at the country scale due to soil heterogeneity, and the use of different data sets and methodologies aids in constraining the estimate.

16) The spatial correlations between the numerical products are high (between 0.91 and 0.97). Are these correlations high because of the “correct” reason, or they are high because all these products make the same “mistakes”. I think that the authors need to discuss this issue and also test for spatial correlations with the available measurements (e.g., inventory data). These analyzes applied for all the grid-point data and the total socks (eg., aboveground biomass, soil carbon, ecosystem fluxes (GPP)) but not for the site available information.

Please see point 5. In addition, we feel that doing site-level analysis goes beyond the scope of our paper.

17) Page 12512: There are estimates in situ measurements of GPP using flux towers across Mexico (MexFlux). Page 12513 on the fact of NPP

measurements the authors only cite two studies (one by Martinez-Yrizar 1996 and another by Garcia-Moya 1992). These are two relatively old studies and following them there have been multiple studies on NPP, NEE, biomass and soil carbon across the country. Just a few examples are works by Masera et al 2003, de Jong et al 1999, Masek et al 2011, Saynes et al 2005, Jaramillo et al 2003, Hughes et al 1999, Vargas et al 2008, Lawrence et al 2002, de Jong et al 2013, Delgado-Balbuena et al 2013, and many others). The authors partially review the information available across some ecosystems in Mexico, but I encourage them to review the wealth of information that has been contributed by many researchers.

We agree with the Reviewer about in situ measurements of GPP and there is now another study (cited above) at the scale of the Baja California peninsula that will be cited in our manuscript. We also thank the reviewer for pointing out relevant literature. However, we want to bring up ~~his attention~~ two to key issues: 1) most of the literature suggested was already included in our paper, and 2) other studies do not measure productivity directly, but proxies of NPP (e.g. litterfall), which is why we only included those two papers. Nonetheless, we will add an extra paragraph for the NPP section citing papers that are proxies of NPP.

18) I think that the manuscript needs to end with a section of limitations and considerations. For example, some of these points are emphasized on the challenges that models have to represent drought responses (page 1217) considering that water limited ecosystems cover 40% of Mexico. How this uncertainty is incorporated for past, present and future? Other issues that should be discussed are QA/QC of available data and error propagation to include uncertainties that goes beyond standard deviations of modeling approaches. This section will be of critical importance because the authors end the manuscript stating that “the methodology proposed here can be used to analyze the full-C cycle of regions elsewhere”. This is important because the authors are not systematically testing the results with “ground truth” information across Mexico.

We are grateful with the reviewer for this comment, which we feel, will benefit the paper a great deal. We fully agree and we'll add a section on limitations and considerations at the end of the paper.

19) Finally, to add transparency and to ignite research on the “comprehensive understanding of the C balance in Mexico” all datasets used (biomass, soil carbon, climate, etc) must be published along this manuscript. This will be the only way the scientific community can move towards testing approaches and comparisons to reduce uncertainties in the regional-to-global carbon balance. Sadly, we are not in the position of making all datasets available on the paper, as we did not generate the information and some require permissions to be used. We will add the reference webpage to all the datasets we used and we'll make the post-processed data available on request (we'll add a link to the paper for this) to guarantee we don't violate the terms and conditions of each dataset. However, previous versions of all the modeling dataset we used are freely available (which we will also state on the paper).

We thank the reviewer for all the key issues raised. We're sure we'll have a much better manuscript because of this.

Reviewer 2

The paper addresses a set of straightforward questions and goals within a modeling and data assimilation framework regarding the Carbon cycle in Mexico. It aims at providing nationwide estimates of carbon stocks (vegetation and soils) and investigates how gross primary production is affected by land and climate changes relying on the experience of a robust modeling and data assimilation community. It is very appealing to see an effort for estimating nationwide values of gross primary production as it invokes mechanistic comprehension between land processes and climate variability;

We thank the positive feedback from the reviewer and for taking the time to provide us with key points that need to be addressed.

However the paper fails on giving confidence on how land data information for parameterization was utilized. Despite that several data sources are acknowledged, I have a hard time understanding how such information was implemented in the data assimilation scheme and in particular how Mexico's unique features were considered. I believe the reader will be benefited if the authors give a brief but significant description of Mexico's singularities regarding carbon cycling (i.e. orographic features and strong seasonality to mention some). This would be of value to construct a stronger discussion that lays out paths to constrain this initial numbers in further efforts and might pose the study as an example useful to generalize on C Cycle processes in complex and dynamic terrains (i.e. expand arguments on Page 12505 L14-21)

We fully agree with the reviewer and we'll add additional information on the particular characteristics of the country that make it so heterogeneous. We'll then link this to possible ways to constrain the values of the C-cycle components over complex and dynamic terrains.

Two issues are of particular concern: 1) For the Model Tree Ensemble (MTE) authors need to specify what type of flux data was used since this is central to this product and gives mechanistic description, while Mexico's flux data is just starting to arise in the literature.

We'll add additional information on how the datasets were constructed and the uncertainty behind them. Please see our response to reviewer 1 in point 1, above.

2) Since an important argument in the manuscript is the effect of temperature on C

cycle via effects in heterotrophic respiration (Rh) the authors need to be more specific on how Rh is calculated and incorporated in the MRT estimates since this parameter is central to assess variation through time. Although the numbers presented in the discussion for particular land covers are somewhat consistent with the very little field evidence that they compare with, the uncertainty due to the coarse resolution of the modeling scheme remains significant and the paper itself provide little information on the particular strengths and weaknesses of their approach.

We fully agree with the reviewer. We'll include additional information on how Rh is calculated in the models, its effects on MRT and ways to constrain its values.

Insights to pay attention to this issues comes when we see the discrepancies in the estimates for the "drylands" (a term preferred over grassland for this cover type) which accounts for a large portion of the land cover in the country, for almost half of the GPP, that is very sensitive to drought as expressed in this study).

We'll strengthen the discussion regarding the role of drylands and the uncertainties around the C cycle over low-C areas.

As such the paper will be benefited by arguments on means for improvement, for example the use better data sources that are certainly available in Mexico (i.e. INEGI/CONABIO cartography to establish land covers among others).

We'll include a paragraph on limitations and considerations at the end of the paper to clarify to the reader what additional information is needed in the country.

I celebrate the effort to zoom in into Mexico's unique and relevant role in the C cycle with this modeling scheme and invite further efforts to constraint the dimension of stocks and fluxes with new and available country's since, no doubt, knowledge to generalize in carbon cycle processes for tropical, subtropical and drylands would benefit our understanding of global patterns as we face climate change.

We are grateful with the reviewer for the strong support shown to our paper, for the important issues and recommendations. We feel that our paper will improve greatly by addressing all comments.