Modeling Interannual Variability of the Ocean and Land Carbon Sinks

Sally Matson

March 2021

Word count: 2679

Over 50% of anthropogenic CO_2 emissions are sequestered each year by the land and ocean sinks, making them a critical force in mitigating climate change. For my dissertation I will use machine learning to develop a model for interannual variability (IAV) of the land and carbon sinks using climactic variables as input. I will use this model to analyze which regions, environmental variables, or combinations therein drive IAV. This work will help explore how future climate changes might impact the sinks, which has important implications in understanding and preparing for the future.

1 Background, Related Work, and Justification

1.1 Introduction

Each year over the past decade, the ocean and land sinks have removed an average of 23% and 31% respectively of annual anthropogenic CO_2 emissions from the atmosphere (Friedlingstein et al. 2020). Figure 1 shows that the land and ocean sink have both sequestered increasingly more carbon since 1959. This upward trend is largely attributable to rising anthropogenic CO_2 emissions from fossil fuels and land use change. As seen in Figure 1, both the land and ocean sink show significant interannual variability (IAV). IAV of the sinks, and especially the land sink, is the main driver of IAV of atmospheric CO_2 growth (Ciais et al. 2014). IAV for the land sink is on the order of 2 GtC/year and for the ocean sink is a few tenths of a GtC/year (Friedlingstein et al. 2020).

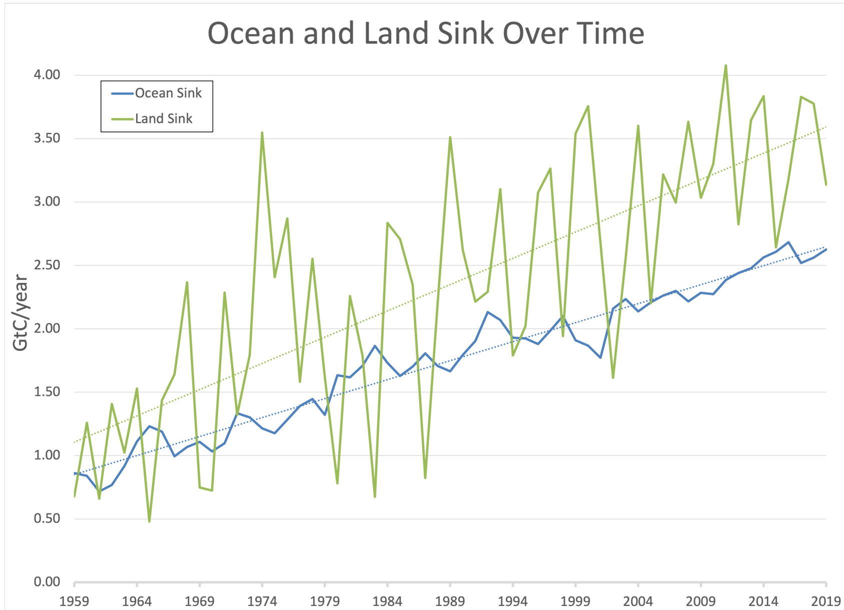

Figure 1: The amount of anthropogenic carbon removed from the atmosphere each year (GtC/year) by the land and ocean sinks according to the Global Carbon Budget (Friedlingstein et al. 2020).

1.2 Land Sink

Plants removes CO_2 from the atmosphere through photosynthesis; the net CO_2 removed is called the land sink. The Global Carbon Budget (Friedlingstein et al. 2020) tracks the major sources of anthropogenic CO_2 over time as well as their partition among the atmosphere, land sink, and ocean sink. The land sink is calculated using the multi-model mean of 17 dynamic global vegetation models (DGVMs). The sink increased from 1.3$\pm$0.4 GtC/year in the 1960s to 3.4$\pm$0.9 GtC/year in the decade from 2010-2019. The increased land sink over time is potentially due to increased CO_2 and nitrogen concentrations, as well as climate changes that increase growing season length in Northern areas (Friedlingstein et al. 2020).

1.2.1 Land Sink Interannual Variability

There has been significant research into the environmental drivers and specific regions which have the largest impact on the land sink IAV. Multiple works suggest that semiarid regions, those in which water is often the limiting factor to plant photosynthesis, contribute the most to land sink IAV. Temperature and soil moisture have emerged as critical environmental drivers but the relationship is less certain.

Ahlström et al. (2015) assessed the relative impacts of different ecosystems on the terrestrial carbon sink. They used a DGVM to model net biome production (NBP, net carbon sequestered by land) over space and time, and then analyzed the relative NBP of six land cover classes. They found that while tropical forests sequester the most carbon of any category (26%), the uptake of carbon in semi-arid ecosystems has the biggest impact on both the increasing trend of the sink over time and the IAV. They further examined the impact of temperature and precipitation on NBP IAV by connecting GPP anomalies to weather variability. They found that the carbon uptake of semi-arid ecosystems is higher under cool and wet conditions. Poulter et al. (2014) use three methods (DGVM, CO_2 inversion, and satellite observation analysis) to examine carbon uptake over space and time, as well as attempt to explain the trends with environmental drivers. Considering the especially large land sink in 2011, they found that 79% to 87% of anomaly in global net primary production (NPP) is attributable to semi-arid regions. They found that high precipitation, increased atmospheric CO_2, and a 'memory' effect from the previous year were important climactic drivers. Fan et al. (2019) studied above-ground carbon (AGC) fluxes in tropical areas using vegetation optical depth methods from satellite imagery. They found a strong correlation between tropical AGC fluxes and atmospheric CO_2 growth rates. Further, semiarid tropical biomes contributed 55.5% of IAV in the AGC fluxes.

Jung et al. (2017) used emerging machine learning methods to research land sink IAV. They examined spatio-temporal patterns in the IAV of GPP, terrestrial ecosystem respiration (TER, total respiration from land), and net ecosystem exchange (NEE = GPP - TER), and analyzed the environmental drivers. They used machine learning algorithms to construct 0.5° grids of GPP and TER from various gridded inputs and compared these to simulations using DGVMs. They detrended GPP and TER to examine IAV and found that temperature and moisture drove NEE IAV on different scales. Temperature drives it at the global scale, however on a local individual-grid cell level water availability was the most important factor.

A recent review on the topic from Piao et al. (2020) concluded that climactic variation drives the IAV of the land sink. The dominance of semiarid regions toward IAV is relatively solidified, however the environmental drivers are less certain. It is clear that both temperature and soil moisture (often using precipitation as a proxy) are critical climactic metrics, however Piao et al. suggest that developing a better understanding of the relationship between these variables is an opportunity for further research.

1.3 Ocean Sink

The oceans are considered a carbon sink because CO_2 can dissolve in ocean water, thus removing it from the atmosphere. This process is mediated by differing partial CO_2 concentrations in the ocean and atmosphere. The Global Carbon Budget estimates the ocean sink using an ensemble of global ocean biogeochemistry models (GOBMs) and validats against IPCC figures (Friedlingstein et al. 2020). The mean ocean sink was 1.0$\pm$0.4 GtC/year in the 1960s and 2.5$\pm$0.6 GtC/year from 2010-2019. The increasing size of the ocean sink is mainly due to the increased atmospheric concentration of CO_2 and is limited by the rate of surface water transport to the deeper ocean (Ciais et al. 2014). Seasonal trends are attributable to annual changes in sea surface temperature and biological activity, two processes which have opposing impacts and thus help cancel each other out (Heinze et al. 2015).

1.3.1 Ocean Sink Interannual Variability

The ocean sink IAV is much smaller than the land sink IAV, and there is less dedicated research on this topic. However, there is research on which climactic factors are drivers in carbon uptake variability. Le Quéré et al. (2010) studied the effect of anthropogenic climate change on the ocean CO_2 sink. While the ocean sink has increasingly sequestered more carbon per year due to the increasing atmospheric CO_2 concentration, they found that due to this effect alone the ocean sink should have grown 20% more than it did. They attributed a significant part of this gap to changes in climactic conditions. Their models estimate that 60% of the gap is specifically attributable to changes in wind which in turn change the ocean circulation, e.g. wind patterns in the equatorial Pacific led to more upwelling which caused more CO_2 to outgas. A further 20% of the gap is due to rising sea surface temperatures, as CO_2 is more soluable in cold water. These results imply that wind and sea surface temperature are the climactic drivers with the most important role in dictating changes in ocean CO_2 uptake thus should be studied when considering ocean sink IAV.

Wanninkhof et al. (2013) used multiple ocean models in order to examine variability and trends in ocean CO_2 sequestration. They found high IAV in regions that have been especially impacted by climate reorganizations, including the North Atlantic, equatorial Pacific and Indian Ocean, and Southern Ocean. This finding supports the idea that climactic variation contributes to ocean IAV, a critical conclusion for the feasibility of this project.

1.4 Role of El Niño

A significant amount of IAV in the land and ocean can be explained by the El Niño Southern Osscilation (ENSO) (Ciais et al. 2014). Jones et al. (2001) used a global climate model with both ocean and terrestrial components to examine the impacts of ENSO on carbon cycle variability. During an El Niño event the ocean becomes a net CO_2 sink because of reduced upwelling of cold, carbon-rich water in the Pacific region. The tropical land sink becomes a net source during El Niño years because GPP is reduced due to warmer temperatures and decreased precipitation. In total, the land process is more extreme than the ocean so overall there is less carbon sequestration in El Niño years, leading to a higher atmospheric CO_2 increase. The opposite happens during La Niña years. Given this known relationship, strong El Niño or La Niña years should be of specific focus during model validation to ensure that the model can fit the variation.

1.5 Justification

There are remaining knowledge gaps in understanding the climactic drivers for both the land and ocean sink IAV. They are often researched separately which limits the connections that can be drawn between the two sink processes. For the land sink, developing a stronger link between precipitation, temperature, and IAV remains important (Piao et al. 2020). A review on land sink trends and IAV by Niu et al. (2017) emphasizes the need for multiple independent modeling projects for climate IAV with different methods in order to build a more robust understanding. Most studies focus on DGVMs for considering land sink IAV; using machine learning models that have no fundamental assumptions about the carbon cycle process will also potentially allow for different trends to reveal themselves and thus identify new areas of impact or exploration. IAV is less prominent in the ocean sink thus it has been the target of less specific research. Examining climactic drivers of change in ocean dynamics and specifically how they relate to IAV has the potential to contribute to understandings of ocean processes. Specifically, investigating and validating findings that use

GOBMs will be interesting.

2 Aims

For my project, I will develop a machine learning algorithm that uses climactic variables (including temperature, precipitation, SST, wind speeds, and more) to predict the size of the land and ocean carbon sinks. I will draw heavily on research of the physical properties of the carbon cycles to direct the data inputs, evaluation, and final model experiments. I will leverage techniques in machine learning that can intuit temporal relationships and retain 'memory.' After the model has been developed and refined, I will conduct experiments using different subsets of data, both differing variables as well as manipulating variables by extracting regional information. Through this analysis I will not only develop an algorithm that reliably predicts the amount of carbon sequestered by the land and ocean, but I will also try to determine which specific environmental factors, combinations of factors, or regions contribute to IAV.

3 Data & Methods

3.1 Data

The initial data sources are:

- **Land and ocean sinks**: I have monthly land and ocean sink data from 13 DGVMs (1700-2019) and 9 GOBMs (1958-2019) that report total carbon sequestered (GtC/month) by either land or ocean in the North Extropics, South Exatropics, and Tropics.

- **Atmospheric CO_2 concentration**: NOAA ESRL provides monthly CO_2 concentration and growth rates from Mauna Loa since 1980.

- **Weather data**: The CRU TS4 dataset (Harris et al. 2020) includes variables for precipitation, cloud cover, wet day frequency, frost day frequency, etc. This dataset is a gridded 0.5° product with monthly temporal resolution from 01/1901-12/2019.

- **Temperature**: The HadCRUT5 dataset (Morice et al. 2020) is 5° resolution from 01/1850-12/2018.
- **Wind**: The NOAA Blended Sea Winds Dataset (NOAA n.d.) reports vector winds and wind stresses on a 0.25° grid with six-hour, daily, and monthly temporal resolution.
- **Sea surface temperature**: The NASA Earth Observation provides a 0.1° gridded product with day and night sea surface temperature on a monthly temporal resolution (NASA n.d.).

Further datasets will be considered to incorporate advanced nuance to the model potential:

- **Weather extremes**: I plan to incorporate the HadEX4 dataset from the UK Met Office which quantifies different weather extremes. The dataset is gridded 1.25° by 1.875° with either monthly or annual resolution dependent on the variable. Potential variables to include are cool/warm spell durations, consecutive wet/dry days, and growing season length.
- **Land type**: The MODIS MCD12C1 dataset (Friedl & Sulla-Menashe 2015) is a 0.5° gridded dataset in which each grid square denotes the land cover type. This is the class cover used by Ahlström et al. (2015).

Given that most of these datasets are gridded (i.e. retain spatial information), the variables can be aggregated on a global, hemispheric, or regional basis. This is critical as many studies emphasize the role of specific regions and systems to the carbon sink.

3.2 Methods

I plan to use machine learning (ML) to model a time-series of the annual size of the carbon sinks. Jung et al. (2017) used ML methods from Tramontana et al. (2016) to reconstruct a gridded (spatio-temporal) map of NEE using climactic inputs. Tramontana et al. use 11 different ML algorithms to reconstruct land-atmosphere fluxes over space and time. They emphasize that part of the value of ML is when used as a complement to process driven models (such as DGVMs or GOBMs) because the ML algorithms have no physical assumptions; instead, multi-variate patterns can emerge from the data alone. They found that the algorithms were relatively comparable in accuracy, indicating that the choice of specific algorithm might not be crucial to the efficacy.

Grange et al. used a random forest machine learning algorithm to use meteorology variables for predicting PM2.5 Grange et al. (2018). They extended these methods toward time-series prediction, which would be an interesting experiment in this application. New methods for time-series forecasting in other domains show that long short term memory (LSTM) neural networks are increasingly effective over traditional methods (Siami-Namini et al. 2018). Poulter et al. showed that 'memory' has an impact on the terrestrial sink (Poulter et al. 2014), yet other research does not focus on this concept. LSTMs are able to learn both short and long-term dependency which would implicitly amplify this potentially important yet under-researched factor.

4 Schedule

I have already gathered and processed a significant amount of the core data and developed initial model infrastructure. For the remaining time I will build robust data, model, and evaluation infrastructure (March 22 - May 23); work on model performance and accuracy in conjunction with the literature and carbon cycle dynamics (May 24 - June 27); and perform final experiments and finish the written report (June 28 - August 5). In more detail:

1. ~~**Train current model using variability**: Develop a finished prototype for the land-sink model. *(March 22 - March 28)*~~

2. **Finish initial data collection and processing, work on model infrastructure**: Collect all data necessary for the land and ocean core models. Develop necessary functions to use all of the data in a seamless and easy manner, create demos. *(3 weeks, March 29 - April 18)*

3. **Incorporate more data**: Access and process remaining data. Incorporate into the models and reevaluate. *(2 weeks, April 19 - May 2)*

4. **Initial model evaluation**: Develop standardized model assessment, evaluation, and reporting techniques. Create methods to compare model results to those in the literature. *(3 weeks, May 3 - May 23)*

5. **Core model development**: Iterate on the model methods, data input choices, and performance in conjunction with literature. *(5 weeks, May 24 - June 27)*

6. **Final model developments and writing**: Run final simulations for paper results, refine literature review and incorporate in any final details, write dissertation text. *(3 weeks, June 28 - July 18)*

7. **Final writing and submission preparation**. *(2.5 weeks, July 19 - August 5)*

8. **Submission**: August 5.

5 Potential Problems

Efforts to understand IAV in the sinks through modeling tend to focus either on the land sink or the ocean sink because the carbon cycle properties of both are so different. Developing parallel models offers novelty as well as an opportunity to find connections between the two systems. However, it could also diminish the time, specificity, and attention paid to both systems individually. Further, it is possible that ML models will simply not work well in this application or that global sink values are to generalized to predict from the data. If this happens I will use knowledge of the physical properties of earth systems to consider why the methods do not work, and through this analysis try to develop alternative hypotheses.

6 Conclusion & Impact

If this project is successful it will be a novel modeling method of land and carbon sink IAV and contribute to greater understanding of their environmental and regional drivers. Both Niu et al. (2017) and Piao et al. (2020) suggest that developing new methods and understanding in this area is an important research focus. This research has broader impacts as quantifying how natural carbon sinks respond to different climactic variations can help indicate how they might change under future climate change scenarios. These predictions also have useful and important applications in carbon budgets (and remaining carbon budgets), policy development, and more. Finally, developing a machine learning algorithm to predict the size of the sinks will help validate against the findings from large DGVMs and GOBMs and will provide a quick, easy alternative to these bigger and more complex models.

References

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K. et al. (2015), 'The dominant role of semi-arid ecosystems in the trend and variability of the land co2 sink', *Science* **348**(6237), 895–899.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M. et al. (2014), Carbon and other biogeochemical cycles, *in* 'Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change', Cambridge University Press, pp. 465–570.

Fan, L., Wigneron, J.-P., Ciais, P., Chave, J., Brandt, M., Fensholt, R., Saatchi, S. S., Bastos, A., Al-Yaari, A., Hufkens, K. et al. (2019), 'Satellite-observed pantropical carbon dynamics', *Nature plants* **5**(9), 944–951.

Friedl, M. & Sulla-Menashe, D. (2015), 'Mcd12c1 modis/terra+aqua land cover type yearly l3 global 0.05deg cmg v006', Accessed from `https://doi.org/10.5067/MODIS/MCD12C1.006`.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S. et al. (2020), 'Global carbon budget 2020', *Earth System Science Data* **12**(4), 3269–3340.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E. & Hueglin, C. (2018), 'Random forest meteorological normalisation models for swiss pm 10 trend analysis', *Atmospheric Chemistry and Physics* **18**(9), 6223–6239.

Harris, I., Jones, P. & Osborn, T. (2020), 'Cru ts4.04: Climatic research unit (cru) time-series (ts) version 4.04 of high-resolution gridded data of month-by-month variation in climate (jan. 1901-dec. 2019)', *Centre for Environmental Data Analysis* .

Heinze, C., Meyer, S., Goris, N., Anderson, L., Steinfeldt, R., Chang, N., Quéré, C. L. & Bakker, D. C. (2015), 'The ocean carbon sink–impacts, vulnerabilities and challenges', *Earth System Dynamics* **6**(1), 327–358.

Jones, C. D., Collins, M., Cox, P. M. & Spall, S. A. (2001), 'The carbon cycle response to enso: A coupled climate–carbon cycle model study', *Journal of Climate* **14**(21), 4113–4129.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P. et al. (2017), 'Compensatory water effects link yearly global land co 2 sink changes to temperature', *Nature* **541**(7638), 516–520.

Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rödenbeck, C. & Sutherland, S. C. (2010), 'Impact of climate change and variability on the global oceanic sink of co2', *Global Biogeochemical Cycles* **24**(4).

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P. & Simpson, I. (2020), 'Hadcrut.5.0.0.0: Ensemble near-surface temperature anomaly grids and time series', *Centre for Environmental Data Analysis* .

NASA (n.d.), 'Aqua/modis sea surface temperature', Accessed at: `https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MYD28M`.

Niu, S., Fu, Z., Luo, Y., Stoy, P. C., Keenan, T. F., Poulter, B., Zhang, L., Piao, S., Zhou, X., Zheng, H. et al. (2017), 'Interannual variability of ecosystem carbon exchange: From observation to prediction', *Global ecology and biogeography* **26**(11), 1225–1237.

NOAA (n.d.), 'Blended sea winds dataset', Accessed at: `ncdc.noaa.gov/data-access/marineocean-data/blended-global/blended-sea-winds`.

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., Friedlingstein, P. & Sitch, S. (2020), 'Interannual variation of terrestrial carbon cycle: Issues and perspectives', *Global change biology* **26**(1), 300–318.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y. et al. (2014), 'Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle', *Nature* **509**(7502), 600–603.

Siami-Namini, S., Tavakoli, N. & Namin, A. S. (2018), A comparison of arima and lstm in forecasting time series, *in* '2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)', IEEE, pp. 1394–1401.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G. et al. (2016), ‘Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms’, *Biogeosciences* **13**(14), 4291–4313.

Wanninkhof, R., Park, G.-H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., McKinley, G. A., Lenton, A. et al. (2013), ‘Global ocean carbon uptake: magnitude, variability and trends’, *Biogeosciences* **10**(3), 1983–2000.