**Initial Project Proposal**

Greenhouse gases have played a significant roll in climate change by trapping solar radiation from escaping from the earth. Thus, the globe warms up rapidly and that has caused a series of destructive impacts. A commonly known effect of climate change is sea level rise caused by the melting ice in the polar. The higher sea level has been an unneglectable threat to coastal cities, which are home to 50% global population (Hansen et al. 2016). Climate change also has made the hurricanes more devastating by providing more fuels for the hurricanes – more water vapor due to the higher ocean temperature (Anthes et al. 2006). More importantly, human health has been at risks because climate change increases the transmission of some insect-borne diseases like Malaria and Dengue Fever (Martens et al. 1997) and the food insecurity caused by more frequent drought and flood (Betts et al. 2018)

The rapid increase of greenhouse gases in the atmosphere is mainly due to anthropogenic carbon emissions. Fossil fuel combustion is the largest contributor, while deforestation and forest degradation are the second largest carbon source, accounting for 12 – 20% of global anthropogenic CO2 emissions (van der Werf et al. 2009). Forests reduce the carbon concentration in the atmosphere by absorbing carbon dioxide from the air. However, deforestation and forest degradation not only erase the reduction of carbon concentration, but even increase the concentration because the dead vegetation releases the carbon accumulated within their biomass via decomposition or fire (Hurtt et al. 2002).

To mitigate the negative impacts of climate change, 25 states in the U.S. together formed the United States Climate Alliance (USCA). The goal is to “reduce greenhouse gas emissions by at least 26-28 percent below 2005 levels by 2025” according to the USCA website. In order to reach the goal, USCA collaborates with my advisor, Dr. George Hurtt, to track the current carbon stock and calculate the carbon budget for the future years. The Ecosystem Demography (ED) model developed by Dr. Hurtt is capable of both estimating how much carbon is contained in vegetation currently, as well as projecting how much more carbon would be absorbed in the future years (Moorcroft et al. 2001; Hurtt et al. 2002). The ED model can simulate the life cycle of plants but cannot anticipate the land cover change. The model can calculate how much carbon is stored within an acre of forest, but what if that forest is removed? That may cause inaccurate results and thus lead to misleading carbon budgets for the states in the USCA. To avoid such a problem, I would like to develop a product which projects the permanent forest loss of each year. The product would be able to estimate where the loss happens and thus to increase the accuracy of the model.

Due to the time span for the project, I would like to choose the state of Maryland as my study area. If the product is tested to be reliable, I would like to expand my study area to the whole 25 states of the USCA. The main dataset I plan to use is the National Land Cover Dataset (NCD). NLCD publishes a national land cover map in every five years. Now, the available datasets are those published in 2001, 2006, 2011, and 2016. The datasets are raster data acquired by Landsat, and the pixel value is the encoded land cover type, such as value 11 is open water, 31 is barren land. For each dataset, the size is 16.8 GB. However, since my region of interest is Maryland, I will clip Maryland out using ArcMap and the resulting raster should be much smaller. The limitations of my data are the spatial resolution and the categorical data type. Since Landsat images have a 30m spatial resolution, it is difficult to identify trees in the backyard or along the road. In addition, since the data value is categorical, if a pixel has 20% trees and 80% water, the trees are ignored. Those would cause errors in calculating carbon for land cover change. For example, if a forest pixel with 100% of trees in 2001 only contains 70% of trees in 2006, but the pixel remains forest, we would underestimate the loss. If a forest pixel with 100% of trees in 2001 changes to an urban pixel in 2006 but actually 30% of the trees remains, we would overestimate the loss.

To start the project, at first, I will compare the four datasets to identify the non-permanent forest loss, which is to find the pixels of forest gain and the previous land cover type at the forest gain. For example, if there are many forest gains happen at pixels which were shrubland previously, then when a pixel of forest changes to shrubland, it is still possible that the pixel would change back to the forest. Therefore, we could say that forest loss to shrubland is not a permanent loss. Then, it would be easy to identify the permanent forest loss: I will find the pixels of forest loss, identify the land cover types that forest changes to, and exclude land cover types which indicate non-permanent loss of forest found in the last step. Therefore, there will be a list of land cover types that cause permanent forest loss. The urban land cover is expected to be on the list, because urban is characterized by its impervious surface, which is very hard for trees to grow on.

The next step is to use machine learning methods to analyze where the potential permanent forest loss is located. The possible spatial dependencies are the surrounding land cover types, the population density, the elevation and slope, and the distance to roads. I will primarily test the surrounding land cover types. If that does not have a statistically significant result on the permanent forest loss, I will add the other elements one by one until getting a satisfactory result. I plan to use the four datasets to “project” the forest loss in 2018 and 2019, so that I can use available datasets to test my projection. The NLCD land cover dataset which will be published in 2021 is the best option, but since it is not available now, I will use the forest loss dataset produced by Global Forest Watch (GFW). It is the only known dataset that contains comprehensive and up-to-date forest loss data. In addition, the GFW dataset is also acquired by Landsat, the same satellite used to create the NLCD datasets.

I will write my own code to compare the datasets but use a few Python packages to do the later steps. Since my data are raster data, I will use Rasterio and/or GDAL for raster processing. The Scikit-learn package will be used for the machine learning part and the pysal package for spatial analysis. I may use other datasets if needed.

A challenge I envision is the machine learning part. I never used Scikit-learn or did any machine learning before, so it could be difficult for me at the beginning. And I am still thinking whether to use the neural network method or the random forest method to do it. I definitely need to do more research on machine learning methods. Another challenge is the validation of my product. The GFW data may not be able to fully test my product because GFW only shows forest loss in forest areas, but my product may show forest loss outside the “forest areas” defined by GFW, like trees close to roads.

**Citation**

Anthes, Richard A, Robert W Corell, Greg Holland, James W Hurrell, Michael C MacCracken, and Kevin E Trenberth. 2006. “Hurricanes and Global Warming- Potential Linkages and Consequences.” *Bulletin of the American Meteorological Society* 87 (5): 623–28.

Betts RA, Alfieri L, Bradshaw C, Caesar J, Feyen L, Friedlingstein P, Gohar L, et al. 2018. “Changes in Climate Extremes, Fresh Water Availability and Vulnerability to Food Insecurity Projected at 1.5°c and 2°c Global Warming with a Higher- Resolution Global Climate Model.” *Philosophical Transactions. Series a, Mathematical, Physical, and Engineering Sciences* 376 (2119). https://doi.org/10.1098/rsta.2016.0452.

Hansen J, Sato M, Kharecha P, Hearty P, Ruedy R, Kelley M, Russell G, et al. 2016. “Ice Melt, Sea Level Rise and Superstorms: Evidence from Paleoclimate Data, Climate Modeling, and Modern Observations That 2 °c Global Warming Could Be Dangerous.” *Atmospheric Chemistry and Physics* 16 (6): 3761–3812. https://doi.org/10.5194/acp-16-3761-2016.

Hurtt, G. C, S. W Pacala, P. R Moorcroft, J Caspersen, E Shevliakova, R. A Houghton, and B Moore. 2002. “Projecting the Future of the U.s. Carbon Sink.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (3): 1389–94.

Martens, WILLEM J. M, THEO H Jetten, and DANA A Focks. 1997. “Sensitivity of Malaria, Schistosomiasis and Dengue to Global Warming.” *Climatic Change* 35 (2): 145–56. https://doi.org/10.1023/A:1005365413932.

Moorcroft, P. R, G. C Hurtt, and S. W Pacala. 2001. “A Method for Scaling Vegetation Dynamics: The Ecosystem Demography Model (Ed).” *Ecological Monographs* 71 (4): 557–85.

van der Werf, G. R, D. C Morton, R. S DeFries, J. G. J Olivier, P. S Kasibhatla, R. B Jackson, G. J Collatz, and J. T Randerson. 2009. “Co2 Emissions from Forest Loss.” *Nature Geoscience* 2 (11): 737–38. https://doi.org/10.1038/ngeo671.