Stat 522

Final Paper

Sampling for the soil carbon: Applying to the soil carbon density change in Pastaza-Marañon peatlands, Peru under certain climate scenario in the next 100 years

Sirui Wang

1. Background

Tropical peatlands or wetlands cover approximately 441,025 km^2 and store a large quantity (about 88.6 Pg C) of soil carbon (SC) [Page et al., 2004, 2011; Rieley et al., 2008]. These ecosystems occupy ~11% of the global peatland area and contribute ~15-19% to the total global peat SC pool [Page et al., 2011]. Tropical peatlands are mainly distributed in Southeast Asia (~56%, 247,778 km^2), and South and central America (~23%, 107,486 km^2) [Page et al. 2011]. The major difference between the peatland ecosystems and the non-peatland ecosystems (sometimes referred as uplands) is that peatlands or wetlands are waterlogged. The water may come from the adequate rainfall or the groundwater discharge. The main interest in studying soil is to estimate the soil carbon (SC) density or stock. In other words, to know the amount of the SC density of a certain pixel or region is of great significance in the fields of atmospheric, environmental and ecosystem research. The soil will accumulate SC if the C input is greater than the C decomposition per time step (month, year, etc). The C input is mainly from the litterfall (mainly dead plant tissues), and the C decomposition is mainly caused by the microbial activity, often referred as the heterotrophic respiration. This works for both the peatlands and the non-peatlands ecosystems. However, unlike the non-peatlands, the peatlands are often waterlogged. The sufficient water separates the peat soil into two zones: the unsaturated zone above the water table, and the saturate zone below the water table. In the unsaturated zone, due to the enough amount of oxygen, the heterotrophic respiration is largely dominated by the aerobic respiration. In contrast, below the water table, due to the lack of oxygen, the respiration is dominated by the anaerobic respiration. The efficiency of the aerobic respiration is significantly higher than the anaerobic respiration, indicating that under waterlogged condition, the peatlands will have less amount SC decomposed. Therefore, peatlands can accumulate much more SC than non-peatlands, although it may take several thousands of years.

The climate change in the future may have huge impacts on the SC change. Current studies are largely focusing on the climate effects on the SC accumulation in boreal and tropical regions. In my previous study, we quantified the C accumulation in peatland and non-peatland ecosystems in the Pastaza-Marañon foreland basin in Peruvian Amazonia from 12,000 years before present to the end of the 21st century using a process-based peatland biogeochemistry model (see Wang et al., 2016a; 2016b for model description and details).

Here, we will use the SC density of the study region from our model output as the population and do the sampling. The data are not from any website or publication, but from our research (Wang et al., 2017, in review in Geophysical Research Letter).

2. Target Population

The target population is defined as the SC accumulation amount of the Pastaza-Marañon foreland basin (PMFB) study area in the next 100 years under a certain future climate scenario [IPCC 2014]. Such scenario is called RCP 2.6 (see https://svs.gsfc.nasa.gov/4110 for details). For the study region, the future air temperature was predicted to increase 0.5 ℃ and mean annual precipitation increase ~260 mm. The population is the SC change / accumulation (in $\mathrm{kg\ C\ m^{-2}}$) simulated by our model. The population contains 37749 pixels covering the whole study region (PMFB) at a resolution of 1.6×1.6 km. The region dominated by four different vegetation types: palm swamp (PS), open peatland (OP), pole forest (PF), and flooded rainforest (FF). The former three types are peatland ecosystem and the last one FF is non-peatland ecosystem. The spatial distribution of the four vegetation types and their simulated SC accumulation amounts are presented below (Figure 1 and Table 1).

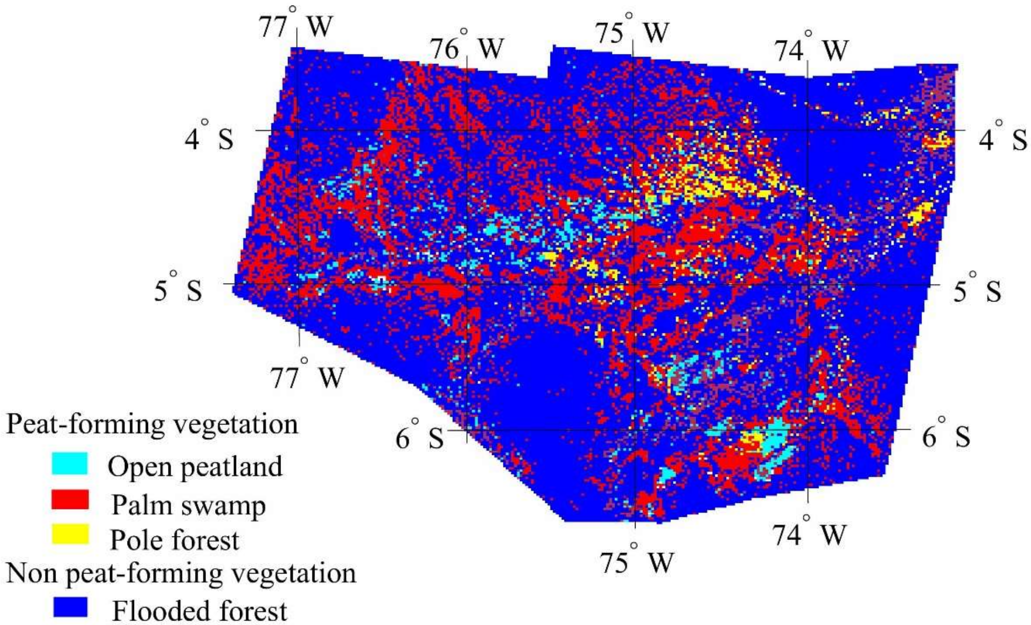

Figure 1. Distribution of peat- and non peat-forming vegetation in the PMFB. The map has been resized from 90 m×90 m to 1.69 km×1.69 km. Colors represent different vegetation categories: open peatland (light blue), palm swamp (red), pole forest (yellow), and flooded forest (blue). See the Figure 4 of Draper et al. [2014] for original map made using three support vector machine supervised classifications. The occasionally flooded forest and seasonally flooded forest were categorized into flooded forest in this study.

Table 1. The properties of the four ecosystem types (target population) of the study area.

Ecosystem types	Area (km^2)	Mean SC accumulation ($kg\ C\ m^{-2}$)	Variance ($kg^2\ C\ m^{-4}$)	Standard deviation ($kg\ C\ m^{-2}$)	Number of corresponding pixels
Palm swamp	27308	2.35	6.97	2.64	8511
Open peatland	3922	1.34	1.1	1.05	1303
Pole forest	2919	1.03	2.52	1.59	1022
Flooded forest	76935	0.14	4.8×10^{-4}	0.02	26913
Total	111084	0.7	2.53	1.59	37749

3. Sampling

Our goal is to estimate the overall mean of the simulated SC accumulation amount in the next 100 years.

3.1 SRS

We first use a SRS to draw a sample of size 1000 to estimate the population mean.

$$\hat{\bar{y}} = \frac{1}{n} \sum_{i=1}^{1000} y_i$$

The variance of the sample is

$$s^2 = \frac{1}{1000 - 1} \sum_{i=1}^{1000} (y_i - \bar{y})^2$$

The SE of the estimate is

$$SE(\hat{\bar{y}}) = \sqrt{\frac{\left(1 - \frac{n}{N}\right) s^2}{n}}$$

If we replace the sample variance with the population variance (2.53), we get

$$SE(\hat{\bar{y}}) = \sqrt{\frac{\left(1 - \frac{n}{N}\right) S^2}{n}} = \sqrt{\frac{\left(1 - \frac{1000}{37749}\right) 2.53}{1000}} = 0.049$$

The 95% CI of the population mean is

$$[\hat{\bar{y}} \pm 1.96 \times 0.049]$$

3.2 Stratification

If we look at the distribution of the population, we know that there are four vegetation types. That means we are able to consider the vegetation type as the stratum and use the stratification to draw a sample with the size of 1000. Assume that we know the population variance within each strata (Table 1) and assume the cost of each strata is the same. We will use Neyman allocation.

We use the equation

$$n_h = (\frac{N_h S_h}{\sum_{l=1}^{4} N_l S_l})n$$

to get the optimal sample size within each stratum.

The sampling frame is shown below (Table 2).

Table 2. Properties of stratification using Neyman allocation.

Stratum	PS	OP	PF	FF
S_h^2	6.97	1.1	2.52	4.8×10^{-4}
N_h	8511	1303	1022	26913
Prop	0.936	0.023	0.041	2.04×10^{-4}
n_h	936	23	40	1

Interestingly, we find that although most pixels are dominated by FF (non-peatlands), due to the very little variation within such ecosystem type, we only need to select one pixel to represent the whole type. This is because the non-peatlands have much lower SC density compared with the peatlands. Future climate will have the least effect on it as concluded in previous studies.

The population mean is

$$\overline{y_{str}} = \sum_{h=1}^{4} \frac{N_h}{N} \overline{y_h}$$

The SE is

$$SE(\overline{y_{str}}) = \sqrt{\sum_{h=1}^{4} \frac{\left(1-\frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2 S_h^2}{n_h}}$$

$$= \sqrt{\frac{\left(1-\frac{936}{8511}\right)\left(\frac{8511}{37749}\right)^2 6.97}{936} + \frac{\left(1-\frac{23}{1303}\right)\left(\frac{1303}{37749}\right)^2 1.1}{23} + \frac{\left(1-\frac{40}{1022}\right)\left(\frac{1022}{37749}\right)^2 2.52}{40} + \frac{\left(1-\frac{1}{26913}\right)\left(\frac{26913}{37749}\right)^2 4.8\times 10^{-4}}{1}}$$

$$= 0.026$$

The 95% CI of the population mean is

$$[\overline{y_{str}} \pm 1.96 \times 0.026]$$

3.3 Two-stage Cluster Sampling

If we look at Figure 1, we can see that the study area is separated by the latitude and longitude grid lines. There are totally 18 sub-areas. Sometimes, it will be time-consuming if we apply the model to the whole study region considering there are too many pixels. In this study, even if we used the Purdue RCAC Conte Cluster to run the model, and the model itself is MPI version (which means the region can be separated into parts and each part can be submitted as a job. We submitted multiple jobs at a time), it still took us 4 to 5 hours to finish the simulation (the time step is monthly from 4000 years before present day to 2100 AD). Thus, we may want to just apply the model to a relatively small region to conduct the simulation.

Here, we assign three PSUs which are all located < 6° S (Figure 1). The sizes of the PSUs are: 1580, 1869, and 695. Then we randomly select the SSUs using the proportional allocation. We get the sizes of SSUs: 381, 451, and 168.

To estimate the population mean, we use

$$\hat{\bar{y}}_r = \frac{\sum_{i=1}^{3} M_i \bar{y}_i}{\sum_{i=1}^{3} M_i}$$

where

$$\bar{y}_1 = \sum_{j=1}^{381} \frac{y_{1j}}{381}$$

$$\bar{y}_2 = \sum_{j=1}^{451} \frac{y_{2j}}{451}$$

$$\bar{y}_3 = \sum_{j=1}^{168} \frac{y_{3j}}{168}$$

The SE of the population mean is (by assuming N is large).

$$SE(\hat{\bar{y}}_r) = \sqrt{\frac{1}{n\bar{M}^2} \frac{\sum_{i=1}^{3} M_i^2 (\bar{y}_i - \hat{\bar{y}}_r)^2}{n-1}}$$

$$= \sqrt{\frac{1}{1381^2 \times 1000} [\frac{1580^2(\overline{y_1} - \hat{\bar{y}}_r)^2}{999} + \frac{1869^2(\overline{y_2} - \hat{\bar{y}}_r)^2}{999} + \frac{695^2(\overline{y_3} - \hat{\bar{y}}_r)^2}{999}]}$$

Since we do not know the exact value of the sample means within each SSU, only the formula is given.

4. Comparisons

We apply three methods: SRS, stratification using Neyman allocation, and two-stage cluster sampling. Which one is more efficient?

The design effect for the stratification is

$$\sqrt{DEFF} = \frac{0.026}{0.049} = 0.53$$

This means that the stratification almost doubles the precision compared to SRS if using a sample size of 1000. The reasons are:

1. Non-peatlands have much smaller variation. We only need to take 1 out of 26913 pixels.
2. The PS is widely spread on the map and it has relatively large variation. Therefore we adjust the weight of this stratum.

We may not be able to estimate the design effect of the two-stage cluster sampling. However, we have the confidence to say that the cluster sampling is less efficient than the SRS due to the following reasons.

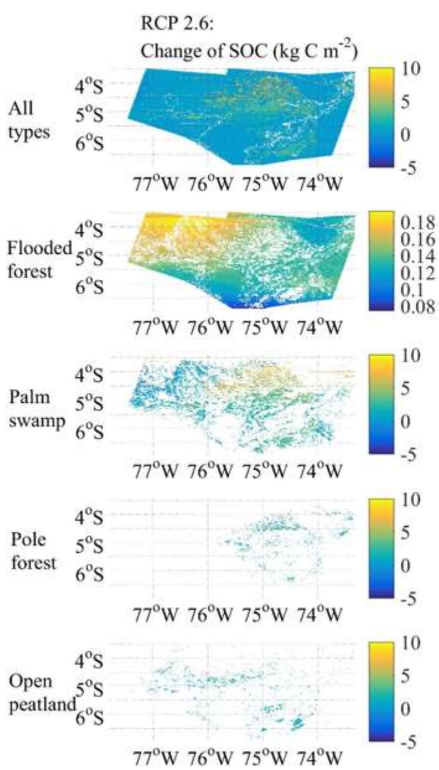

Figure 2. Changes of SOC density from 2014 to 2100 AD under RCP 2.6, future climate scenario of flooded forest, palm swamp, pole forest, open peatland, and their combination in the PMFB

If we look at Figure 2, we can see that what we have sampled using two-stage cluster sampling are the areas < 6° S. Those areas have lower SC density (deep blue color) compared to other areas, especially the northwestern regions, where higher SC density (light yellow color) is mainly located. The spatial distribution of the SC density has very large variations, which leads to a high ICC, if using formula

$$\widehat{ICC} = 1 - \frac{M}{M-1}\frac{\widehat{SSW}}{\widehat{SSB} + \widehat{SSW}}$$

The variation between PSUs (e.g., between different regions) is large from Figure 2.

In this case, we may confidently think that if using two-stage cluster sampling of 1000 pixels, we may get the lowest precision compared with SRS and stratification.

5. References

Change, I. C. (2014). Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK and New York, NY.

Draper, F. C., Roucoux, K. H., Lawson, I. T., Mitchard, E. T., Coronado, E. N. H., Lähteenoja, O., Montenegro, L. T., Sandoval, E. V., Zaráte, R., & Baker, T. R. (2014). The distribution and amount of carbon in the largest peatland complex in Amazonia. Environmental Research Letters, 9(12), 124017.

Page, S. E., Wűst, R. A. J., Weiss, D., Rieley, J. O., Shotyk, W., & Limin, S. H. (2004). A record of Late Pleistocene and Holocene carbon accumulation and climate change from an equatorial peat bog (Kalimantan, Indonesia): implications for past, present and future carbon dynamics. Journal of Quaternary Science, 19(7), 625-635.

Page, S. E., Rieley, J. O., & Banks, C. J. (2011). Global and regional importance of the tropical peatland carbon pool. Global Change Biology, 17(2), 798-818.

Rieley, J. O., Wüst, R. A. J., Jauhiainen, J., Page, S. E., Wösten, H., Hooijer, A., Siegert, F., Limin, S. H., Vasander, H., & Stahlhut, M. (2008), Tropical peatlands: carbon stores, carbon gas emissions and contribution to climate change processes. Peatlands and Climate Change. International Peat Society, Vapaudenkatu, 12(40100), 148-182.

Wang, S., Zhuang, Q., Yu, Z., Bridgham, S., & Keller, J. K. (2016a). Quantifying peat carbon accumulation in Alaska using a process-based biogeochemistry model. Journal of Geophysical Research: Biogeosciences, 121(8), 2172-2185.

Wang, S., Zhuang, Q., & Yu, Z. (2016b). Quantifying soil carbon accumulation in Alaskan terrestrial ecosystems during the last 15000 years. Biogeosciences, 13(22), 6305.

Wang, S., Q. Zhuang, O. Lähteenoja, F. C. Draper, and H. Cadillo-Quiroz (2017), A shift from a carbon sink to a source in Pastaza-Marañon peatlands under a changing climate, Geophysical Research Letters (in review).