A non-dogmatic critique of West et al. (2023), "Action needed to make carbon offsets from forest conservation work for climate change mitigation," *Science* 381: 873-877

August 25, 2023

Hereby we present the results from a technical review of the recently published study authored by West and collaborators. We found multiple serious methodological deficiencies, which make its conclusions patently unreliable and doubtful at best. The main failings identified include the following:

Sample size

There are currently 93 REDD projects registered in the Verra Registry, and 124 more under development or in the process of registration. Even though the authors claim to have analyzed 26 of those projects, in reality, they omitted two of them (see table S1) due to their inability to construct suitable synthetic controls. Thus, the authors only analyzed 24 different REDD projects registered in the VCS. Extrapolating the study's conclusions to all carbon offset projects and to the baseline methodologies they used is clearly unwarranted when only about a quarter of the projects have been examined.

Dogmatic reliance on a particular methodological approach

West et al. (2023) uncritically imply that the baselines they constructed using abstract synthetic controls are the correct ones, as opposed to those constructed by the project proponents following the VCS methodologies and based on on-the-ground information and experience. The opposite possibility—namely, that baselines based on synthetic controls are wrong—is uncritically disregarded.

Despite their growing use in economics, political science, social research, and, more recently, nature conservation, synthetic control methods pose potential pitfalls. As pointed out by several authors (Ferraro and Hanauer 2014; Baylis et al. 2015; Samii 2016), the use of synthetic controls, particularly in the nature conservation field, rests on untested and even untestable assumptions as the counterfactuals for REDD projects are, by definition, unobservable. As has also been shown (e.g., Schleicher et al. 2019), the credibility of this methodological approach depends heavily on how it is implemented. In particular, the results can be highly sensitive to the way control units are selected and weighted. Small changes in model specification can lead to substantially different synthetic controls (Schleicher et al. 2019; Desbureaux 2021). Moreover, the flexibility in choosing control units and model specification implies that the researcher has to make a number of discretionary decisions that would lead to different results, from which the most "convenient" ones can be cherry-picked (Samii 2016; Ferman et al. 2020). Similar or equivalent criticisms are made by West et al. (2023) of the methodologies used by the projects examined.

Thus, synthetic control methods are just alternative methodological approaches that bear no philosophical or practical superiority to other properly implemented approaches. Moreover, the way West et al. (2023) implemented this methodological approach shows serious deficiencies, which make their conclusions unreliable and doubtful.

Synthetic controls

West et al. (2023) constructed, for each project area (PA) examined, a set of synthetic controls (SC) that would serve as counterfactuals to compare the deforestation and carbon emissions in the PA (the "treatment") against those in its corresponding synthetic control areas ("without treatment"). Significant flaws are evident in the overly simplistic way that those synthetic controls were constructed. This makes the results from this study doubtful.

While the use of synthetic controls in conservation work is a novel approach that is becoming more common, several authors (e.g., Schleicher et al. 2019; Desbureaux 2021) have pointed out their difficulties with the approach, in addition to some of its shortcomings. A crucial issue is to ensure that synthetic controls are constructed in such a way that they are indeed similar to the case/treatment sites. In order to do this, the relevant covariates should be carefully identified, and the similarity thresholds carefully set. Differences in these steps can lead to substantial differences in the results, and it is easy for researchers to change or modify matching features or thresholds to yield (i.e., cherry-pick) desired results. Clearly, these recommendations were not taken into consideration in the study by Thales West et al.

West et al. (2023) constructed their synthetic controls by selecting, from among an initial set of 1,000 circular areas randomly placed across the entire country, those that exhibited similar characteristics to the corresponding project area, as well as an average annual deforestation rate (prior to project implementation) similar to that observed in a 10-kilometer-wide buffer around the project area. As shown by tables S2 and A1–A26, the characteristics (covariates) selected by the authors to claim that the synthetic controls were similar to their corresponding project areas only include physical features (e.g., slope, elevation, distance to urban centers) and generic vegetation features (e.g., tree coverage, coverage of timber concessions, deforestation extent) of the sites. Clearly, their synthetic controls excluded the key factors that drive deforestation and land-use change in a given location, such as forest type, agricultural practices, access to markets, access to technology, and geography.

By constructing synthetic controls in this way, and as figures S3 and S4 clearly show, each project area was compared to some "synthetic control" located hundreds of kilometers away, likely covered by a different forest type, and also likely exposed to different deforestation pressures. For instance, Schleicher et al. (2019) have shown that "in the Peruvian Amazon … there are few land areas with no formal or informal land-use restrictions and the land-use designations often overlap.... This hinders isolation of a particular treatment type (e.g., government-controlled PAs or conservation concessions) … and identification of appropriate control units (e.g., unprotected land without land-use restrictions." Just compare in figure S4 the eight projects in the Peruvian Amazon that were examined by West et al. (2023) with the various land-use designations and restrictions shown in figure 2 of Schleicher et al. (2019). Similarly, figure S4 shows that projects (e.g., projects with the following ID numbers in the Verra Registry: 1392, 1395, 1396, and 856) in the Colombian Pacific ecoregion were compared to "synthetic controls" located in the Amazon ecoregion, and even in the Andean ecoregion, under clearly very different conditions.

According to the authors and without any substantive justification, the requirement that synthetic controls should exhibit deforestation rates similar to those observed in the 10-kilometer buffer around the project area would ensure that the selected controls had been exposed to similar levels of deforestation pressure as the project sites and, thus, they could replicate the historical deforestation pattern in the project areas. This assumes then that the mere deforestation rate observed in a randomly selected site, likely very distant from but exhibiting similar physical characteristics (distance to roads, elevation, slope, etc.) to the project area, indicates that it has been exposed to the same deforestation drivers and pressures as the project area, thus ignoring the underlying, locale-specific socio-economic processes that actually determine deforestation and land-use change.

Despite the very loose requirements for selecting control areas, these methods proved to be hardly able to construct suitable synthetic controls for the projects examined. This is clearly seen in the very few areas that made up the synthetic control of each project. Tables A1–A26 and figure S5 show that each REDD+ project was compared to a few (1–9) circular areas located across the entire country. Two of the projects were even compared to just one single control area. Why should such a control site make for a better counterfactual than a reference region carefully selected by the project proponent according to the methodology requirements and based on careful analysis, local knowledge, and on-the-ground information?

The study claims to have validated the synthetic controls so constructed by showing (fig. S6) that the pre-project deforestation rates are similar in the synthetic controls and the corresponding project area. Figure S5 shows that this was not entirely true, as there were large disparities in the deforestation patterns of the synthetic controls of most projects. Moreover, under Verra's rules, a REDD project is required to be 100 percent forested on its start date, which means that the deforestation rate in the project area itself is zero or near zero and that in its surrounding buffer area is also likely to be low or, at most, modest. The consequence of the authors' approach is that they likely selected geographic areas that may have had zero to low threat of deforestation, as figure S6 suggests. In contrast, REDD project areas are selected because they are exposed to some significant threat of deforestation, as revealed and documented through a detailed analysis of deforestation drivers in the region surrounding the project area.

For these reasons, the control areas selected by West et al. (2023) are, clearly, not comparable to the project areas examined, and therefore are not aligned with a fundamental tenet of the synthetic control approach.

Despite these evident methodological shortcomings, West et al. (2023) plainly stated that their findings show that "Only a minority of projects achieved statistically significant reductions in comparison with ex post counterfactuals," that only "Eight of the … 26 project sites showed some evidence of additional reductions in deforestation compared with their individual SCs," and that "the weight of evidence suggests that voluntary REDD+ projects … achieved much less avoided deforestation than forecast by project developers" and "the unrealistic ex ante baselines adopted by many projects likely resulted from the use of methodologies that systematically fail to produce credible counterfactuals for the REDD+ interventions, compromising the evaluation of the projects' effect on mitigating deforestation and thus carbon emissions." Such statements assume that the synthetic controls are absolutely true or, at least, better than those constructed by the project developers based on historical trends.

West et al. (2023) recognized that the implementation of dynamic methods like the SC to construct ex post deforestation baselines for REDD+ interventions can create "implementation and monitoring challenges." For instance, they stated, "given the biophysical heterogeneity of tropical regions, and limited data, ideal control areas for the project sites may not always exist or be possible to identify" and "dynamic baselines may still fail to account for all relevant determinants of deforestation owing to data constraints." And yet, the authors dismissed these shortcomings and caveats, assuming that their synthetic controls provide a better counterfactual than those constructed based on historical trends by the project developers. In fact, they did not.

Data used

West et al. (2023) used only satellite imagery to estimate deforestation trends in their synthetic controls. In contrast, VCS methodologies require that satellite imagery analyses be complemented with extensive on-the-ground analysis including a detailed consideration of local circumstances and socio-economic processes. The limitations of using satellite imagery are exacerbated by the authors' choice of which satellite imagery to use.

The authors used Hansen et al. (2013)'s tree cover loss data taken from the Global Forest Watch platform, developed by the World Resources Institute, to estimate annual deforestation over the period of 2001 to 2020 for the REDD projects and their synthetic controls. The use of these data for this purpose is highly questionable. First, as the Global Forest Watch portal itself states, its dataset shows tree cover extent and its changes over time, with tree cover being defined as "all vegetation taller than 5 meters in height. 'Tree cover' is the biophysical presence of trees and may take the form of natural forests or plantations existing over a range of canopy densities." For this reason, this dataset bears key limitations that are described in detail on the Global Forest Review webpage.

Second, Hansen et al. (2013) improved their methodology by using finer resolution data and an improved analytical method such that "these changes lead to a different and improved detection of global forest loss. However, the years preceding 2011 have not yet been reprocessed in this manner, and users will notice inconsistencies as a result…. The integrated use of version 1.0 2000–2012 data and updated version 1.7 2011–2019 data should be performed with caution."

Notwithstanding this limitation, the authors used data from 2001 to 2020 with no adjustment to make allowance for such changes, which makes their analysis and results questionable.

Third, in many forest types, Hansen et al. (2013)'s data yield inconsistent results, either overestimating or underestimating forest cover change. This inability to detect real forest loss in a control area calls into question the validity of comparing deforestation rates estimated from this dataset with those estimated by the projects based on satellite imagery plus on-the-ground information and verification. Further, as the accuracy of the Global Forest Watch dataset has not been assessed at the local level, and it might vary significantly between localities, deforestation estimates using this data set are unreliable.

For these and other reasons, several studies, including some by the World Resources Institute itself (e.g., Harris et al. 2018; Bos et al. 2019; Chen et al. 2020), have explicitly stated that Hansen et al. (2013) data should not be used off-the-shelf to estimate deforestation or for REDD purposes, although they might be used for those purposes provided suitable adjustments are made first. Such elementary but crucial caution was neglected by West et al. (2023), rendering their results doubtful and their conclusions questionable.

The authors briefly acknowledged this shortcoming: "Many remote sensing studies highlight the differences in deforestation rates detected by [Global Forest Change] and officially recognized by governments.... Such differences emerge from different mapping methodologies and definitions of deforestation and forest degradation." However, the authors ignored the implications of this shortcoming by simply stating, "since the [Global Forest Change] data is used to map deforestation rates both inside the project sites and in the control areas, both before and after project implementation, it is reasonable to use this database to estimate REDD+ impacts in all study countries." The fact is that deforestation rates estimated using these data were also—and critically—compared to those obtained by the project proponents based on more suitable data plus field information and verification.

Additional analyses

In order to check the robustness of their results, the authors conducted project-specific and cross-project analyses for all the projects in a country and varied how they selected potential controls. Such analyses, however, merely implied merging and reshuffling the same data used for the project-level analyses with no additional or independent information. Thus, they carry with them the same failures of the initial data and, not surprisingly, led to similar results and confirmed the results from the project-level analyses.