Part 2 Question & Answer Session

Please type your questions in the Question Box. We will try our best to answer all your questions. If we don't, feel free to email Erika Podest (erika.podest@jpl.nasa.gov) or Sean Healey (sean.healey@usda.gov). This document will be shared to the training webpage within one week.

1. **Question: GEDI and OBIWAN seem very useful for estimating biomass change, but high-stakes carbon decisions require strong evidence of additionality, uncertainty control, and validation. In your view, what is the current practical boundary between using OBIWAN for scientific monitoring and using it for operational carbon-credit or policy verification?**

   Answer: Ultimately, decisions about the maturity/reliability of carbon verification tools depends upon either the party that uses them to report progress on committed forest-related mitigation or (in the case of offset markets) the credit-issuing body responsible for choosing among alternative standards. The Emission Factor x Activity Data paradigm is frequently accepted in this context; it has a clear statistical foundation and addresses relevant question of biomass change and additionality. However, that method also tends to treat carbon stocks in undisturbed forests as static, and it requires substantial human interpretation that can be difficult to standardize from year to year.

   OBIWAN addresses some of these problems while maintaining a straightforward statistical foundation. Especially in the US, where calibration with FIA's systematic long-term forest inventory eliminates many biases (according to the validation material presented here), it seems to me that OBIWAN should be considered a legitimate alternative to the Emission Factor x Activity Data approach. In other parts of the world, calibration using GEDI waveforms not used in model-building provides a notable benefit in terms of accuracy (according to our validation), but not at the level available in the US (and other countries with similar NFIs). Speaking for myself in answer to your question, the line between the best available science and a tool for operational carbon monitoring may depend upon the ability to calibrate OBIWAN change estimates with systematic and repeated inventory measurements.

2. **Question: Can the integration of GEDI-derived biomass structure, Landsat/Sentinel land-cover history, and SIF-based photosynthetic activity improve estimates of watershed-scale carbon-capture and biomass-change dynamics under climate and disturbance pressures?**
   Answer: Landsat is available over time and is moderately correlated with biomass, which is why it is currently used by OBIWAN. If better-correlated imagery were available across time, that would improve estimates of biomass change. Climate pressures are not monitored by OBIWAN. Disturbance pressures in a retrospective way are monitored.

3. **Question: Can the Obiwan API be used for Urban prediction?**
   Answer: We have not tested it in urban areas, but OBIWAN is not optimized to work in areas where building might "fool" GEDI's lidar observations. They can look like trees. We would not suggest this as an application. Even if you mask out building footprints, the GEDI location error would cause issues.

4. **Question: How do uncertainty approximation approaches differ between MDN-based probabilistic modeling and OBIWAN-based biomass-change estimation, and can these approaches be integrated to improve watershed-scale carbon and ecosystem decision support?**
   Answer: OBIWAN operates under a model-based inferential framework. I am not familiar with MDN-based probabilistic modeling, so I'm not sure how it might be integrated with OBIWAN. OBIWAN is appropriate for watershed carbon trends.

5. **Question: In slide 11, what does the black stripe line mean?**
   Answer: The black dashed line in these figures represents the 1:1 line. The closer to that line an estimate of change might be, the less biased that estimate is.

6. **Question: I know that over steep slopes, the LIDAR waveform tends to stretch, this can lead to overestimating canopy height. Do the current versions of Relative Height metrics (such as RH100 and RH98 in the L2A product) include any slope-correction algorithms, or have they successfully mitigated this slope-stretching effect?**

Answer: GEDI's footprint is much smaller than the earlier ICESAT footprint, an improvement intended to minimize effects of differential topography. GEDI provides RH values using 6 different ground-finding algorithms. More detail and documentation on algorithms: https://www.earthdata.nasa.gov/data/catalog/lpcloud-gedi02-a-001#documents-and-resources. We don't have a way to trace individual returns within the footprint.

7. **Question: Could you share more applied cases of satellite data fusion, especially examples where multiple datasets such as GEDI, Landsat, Sentinel, SIF, or SAR are integrated to improve biomass, carbon, or ecosystem monitoring?**

   Answer: This paper (Duncanson et al., freely available preprint) describes fusion of both ICESAT-2 lidar data and GEDI data with imagery from Landsat and Sentinel-2: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5784282. This application applies model-based inference methods similar to OBIWAN's.

8. **Question: How is calibration extrapolated to every Landsat pixel when using GEDI/FIA?**

   Answer: Remeasured FIA data gives us a simple linear correction at the landscape level to changes predicted from Landsat data alone. Application of that correction is intended to remove biases across large areas at the cost of increasing the RMSE of predictions at the pixel level. Remember, OBIWAN is a tool for estimating biomass and biomass change across large areas – the focus is not pixel-level error. Calibration is ____ across the population.

9. **Question: The underestimation/over estimation f high/low values sounds like a "classic" problem with "heteroscedastic" data [unequal variances]. There are statistical corrections for this, lie "variance-stabilizing transformations," e.g., ln(y). Have these approaches been applied to this problem?**

   Answer: The focus in model-based inference is removing heteroscadasticity condition on the predictors instead of with respect to true values. The paper below describes this apparently counter-intuitive focus. The "classical" linear calibration functions applied here are intended to remove bias within the context of this inference paradigm.

Ståhl, G., Gobakken, T., Saarela, S., Persson, H. J., Ekström, M., Healey, S. P., ... & McRoberts, R. E. (2024). Why ecosystem characteristics predicted from remotely sensed data are unbiased and biased at the same time–and how this affects applications. *Forest Ecosystems*, 11, 100164.

10. **Question: To see the temporal changes in the optical imagery, do we also use the hyperspectral remote sensing? Because the spectral information in the Amazon rainforest could be very similar for each pixel due to the dense canopy.**

    Answer: At the moment, we are only using (multispectral) Landsat, as it is available in a globally consistent way over the last several decades. As you suggest, I am not sure that hyperspectral imagery would offer much additional benefit by way of detecting forest structure attributes occurring beneath the canopy top.

11. **Question: I used the Giovanni Toolkit. I noticed most regions are jammed during extraction. I hope this is not the case here as well.**

    Answer: NASA Goddard's Giovanni Toolkit does not operate on the same platform as the OBIWAN API. Most of OBIWAN is hosted on Google assets.

12. **Question: When using Landsat information for the annual biomass estimation, what variables are used? If NDVI or any vegetation indices are used, have you considered the annual variation in vegetation indices which are not reflected in biomass (e.g., NDVI can increase largely immediately in one year but a large biomass increase for forest might not be possible).**

    Answer: The Landsat features used to model biomass are anniversary surface reflectance values derived from the CCDC time series model (which accesses all clear pixels throughout the year). This algorithm generally smoothes some year-to-year variation unless there is evidence for a surface state change; in such cases, a new time series function is fitted.

13. **Question: In the demo notebook, the Obiwan API is a mockup for testing over Alabama. Is there a generic API?**

Answer: The API used in the Alabama notebook is the same API that will eventually be used across the world. We will notify ARSET as more areas are added, and ARSET will distribute important updates to registrants of this training.

14. **Question: What are the API characteristics: are credentials needed (for GEE for example), is it free/free with credits/pay as you request, is there a maximum request size, ...? Can we access the API directly or only by using the notebook?**

    Answer: The notebook gives an example of the syntax used in calls to the API. Those calls may be made through any number of services. The idea is that a developer (for a company, country, or market) can build those API calls into a data system that meets local needs WRT visualization, confidentiality, spatial scale, baseline scenarios, etc. Costs for those calls will eventually be the responsibility of the local service, but the vast majority of computing expense has already been incurred by our project, with generous support from Google, in creating and storing the needed bootstraps. IO costs for summarizing across bootstraps should normally be quite small.

15. **Question: In the hierarchical bootstrapping framework, how does spatial autocorrelation in Landsat predictors affect the independence assumption between bootstrap samples, particularly when estimating biomass change over small areas of interest?**

    Answer: OBIWAN and GEDI apply model-based inference in a large area context. Spatial autocorrelation of errors is not considered in the variance estimators we use, and we therefore do not recommend application of OBIWAN over small areas, where effects of spatial autocorrelation might be significant.

16. **Question: For teams looking to build automated workflows or pipelines around this, do you provide a formalized developer documentation or an Open spec page for OBIWAN endpoint?**

    Answer: I should have given users more specifics in this direction. A current version of the API documentation is here: https://obiwan-api-5026633953.us-west1.run.app/docs.

17. **Question: Thanks for the presentation. The presenter mentioned that bootstraps are available for the entire USA. Does it mean that the model only run for the USA? What are the implications that the bootstraps are only for the USA?**

    Answer: Estimates may only be made where we have created bootstraps that represent uncertainty in the respective models. Currently, that is only in the US and in Nepal, We are deepening the stack of bootstraps in the US, and will soon be expanding to other countries. ARSET will distribute periodic updates related to this process.

18. **Question: Can this biomass estimation method be used as a base for yield estimation for specific crops?**

    Answer: If this question refers to field crops, I do not think this fusion of lidar data and Landsat time series is the best available option.

19. **Question: What is the recommended slope threshold for masking out GEDI shots in steep terrain when estimating AGB with L4A products (for example max slope 20degree), and is there a preferred DEM (such as Copernicus DEM or NASADEM for generating that slope mask? Does OBI-WAN or any tools handle this correction internally, or should it be applied before querying the API?**

    Answer: The OBIWAN app precomputes bootstraps needed to make biomass estimates (and calculate uncertainty related to those estimates). It is not possible to customize the shots used by OBIWAN for this purpose; we use all high-quality shots and do not apply slope-only filters. As a sampling mission, GEDI does not want to specifically remove hillside forests from the sample. GEDI's relatively small footprint (25m) hopefully minimizes errors related to variable within-footprint topography.

20. **Question: What is the timeline for releasing pre-computed bootstrap maps globally?**

    Answer: We do not yet have a definitive timeline for expansion of bootstrap assets. We will communicate through the ARSET team to keep you posted.

21. **Question: Are there plans to use Landsat from the 1980s-1990s to extend the time series predictions?**

Answer: Currently, no. It remains a possibility, though, in places where baselines from those decades are of particular interest.

22. **Question: Does calibration also consider the age effect of forest?**
    Answer: We have experimented with time since disturbance (using GFW) as a variable addressed by calibration functions. In some cases, it is promising, but we currently do not operationally use age.

23. **Question: Is there any harmonization between different Landsat sensors?**
    Answer: The Landsat CalVal team at the USGS puts considerable effort into maintaining cross-calibration across Landsat missions.

24. **Question: Can we use OBIWAN in Google Earth Engine?**
    Answer: Important components of OBIWAN already use GEE functions, as the bootstrap assets are hosted by Google. Several GEDI products are maintained as assets on Earth Engine (e.g., https://developers.google.com/earth-engine/datasets/catalog/LARSE_GEDI_GEDI04_A_002_MONTHLY).

25. **Question: Is the calibration option in the API notebook example based on the L4A or FIA based calibration graphed in the slides?**
    Answer: We should have specified. Because this is in the US, we were able to implement the FIA-based calibration.

26. **Question: How can we calibrate OBIWAN data with other available AGBD data (e.g., field data or national FIA) to develop locally adjusted estimates?**
    Answer: This open-access paper gives a nice example (from Paraguay) of creating a nationally specific GEDI biomass model: https://iopscience.iop.org/article/10.1088/1748-9326/acdf03/meta.

27. **Question: What information/records are needed if we want to customize the model for other regions/forest-ecosystems? Any recommendations?**
    Answer: The above paper (Bullock et al.) spatially matched GEDI shots to field-measured biomass. The default GEDI biomass models did not fit very well in Paraguay because S. American models had been fit mostly in the Brazilian Amazon. Local fitting led to significant improvement. Field measurements of biomass density can also be used to fit a calibration function similar to what we described for FIA.

28. **Question: Can you speak to any biases with slope or mountainous terrain in general? Is this considered in bootstrapping errors?**

    Answer: As mentioned earlier, we use all high-quality waveforms, which in steep areas likely introduces errors. We do this because we don't want to eliminate hillside forests from GEDI's sample. Bootstrapping theoretically could represent this uncertainty, but in the current implementation, we do not vary predictions as a function of slope.

29. **Question: Can this biomass estimation method be used for specific crop yield estimations? For example, wheat, rice, sugarcane, or tobacco.**

    Answer: As mentioned above, I believe there are remote sensing products that are better tuned to detect crop production.