
Aboveground carbon biomass estimate with Physics-informed deep network

Juan Nathaniel*
Columbia University
NY, USA

Levente J. Klein
IBM Research
Yorktown Heights, NY, USA

Campbell D. Watson
IBM Research
Yorktown Heights, NY, USA

Gabrielle Nyirjesy
Columbia University
NY, USA

Conrad M. Albrecht
German Aerospace Center
Weßling, Germany

Abstract

The global carbon cycle is a key process to understand how our climate is changing. However, monitoring the dynamics is difficult because a high-resolution robust measurement of key state parameters including the aboveground carbon biomass (AGB) is required. Here, we use deep neural network to generate a wall-to-wall map of AGB within the Continental USA (CONUS) with 30-meter spatial resolution for the year 2021. We combine radar and optical hyperspectral imagery, with a physical climate parameter of SIF-based GPP. Validation results show that a masked variation of UNet has the lowest validation RMSE of 37.93 ± 1.36 Mg C/ha, as compared to 52.30 ± 0.03 Mg C/ha for random forest algorithm. Furthermore, models that learn from SIF-based GPP in addition to radar and optical imagery reduce validation RMSE by almost 10% and the standard deviation by 40%. Finally, we apply our model to measure losses in AGB from the recent 2021 Caldor wildfire in California, and validate our analysis with Sentinel-based burn index.

1 Introduction

Aboveground carbon biomass (AGB) is an important component to monitor carbon cycle on the local [1, 2] and global scale [3, 4]. A recent state-of-the-art light detection and ranging (LiDAR) mission from NASA’s Global Ecosystem Dynamics Investigation (GEDI) generates a global yet non-continuous sparse measurements of vegetation parameters, including AGB estimates with a 60-meter along-track and 600-meter across-track gaps between footprints [5]. Here, we apply a Physics-informed deep network to generate a 30-meter resolution, wall-to-wall continuous AGB estimate in CONUS for the summertime (June-August) 2021 period. The model is trained on more than one million GEDI footprints using a combination of radar and optical imagery [6]. In addition, we incorporate a measure of photosynthetic intensity as captured by solar-induced fluorescence (SIF)-based gross primary production (GPP). SIF-based GPP is one of the key physical parameters regulating AGB [7, 8]. We validate our results using field-based AGB observations and evaluate their consistency across climate zones. Finally, we use our high resolution AGB map to assess the impact of wildfire, in terms of how much carbon biomass had been lost to the environment.

*Corresponding author (jn2808@columbia.edu)

2 Methodology

This section provides a detailed description of the data, models, and evaluation metrics used, and are summarized in Figure 1.

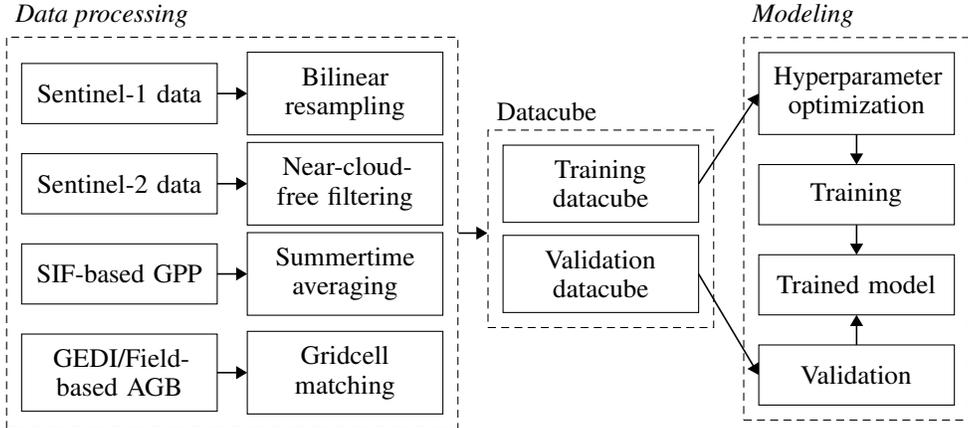


Figure 1: Schematic overview that summarizes our data-processing steps at the leftmost panel and datacube generation during training and validation steps. The modeling process is iterated across different data setups (SIF/S1/S2, S1/S2, and S2-only) in an ablation study.

2.1 Data Processing

GEDI is one of the current state-of-the-art spaceborne LiDAR missions to capture a detailed measure of ecosystem structure [9]. However, the data generated is spatially sparse. Therefore, we attempt to produce a dense 30-meter resolution AGB estimate using multiple input features including radar and optical hyperspectral imagery from Sentinel-1 and Sentinel-2, as well as SIF-based GPP from the OCO-2 mission [7]. We use the vertical co-polarization (VV) and vertical/horizontal cross-polarization (VH) band of Sentinel-1, and the entire 12 bands of Sentinel-2. Furthermore, we produce near-cloud-free optical imagery by taking the median value of non-cloudy pixel as defined by the scene classification layer (SCL) of Sentinel-2. Finally, we perform grid cell geospatial matching to combine the dataset together. The validation sites are situated in the Northwestern America [10] and New England [11].

2.2 Model and Experimental Setup

We benchmark our deep network model against linear regressor (LR), random forests (RF) [12], and extreme gradient boosting (XGBoost) [13] algorithms due to their reported robustness in climate-related tasks [14]. Specifically, we implement a masked variation of UNet [15] due to the sparsity of our target AGB variable [16]. The models are optimized using a randomized grid search and a k-fold cross validation (k=5) approaches. We split the data into a 90-10 training-testing set, where set here refers to pixels for ML-based models and tiles for deep networks. We use Adam optimizer [17] with a learning rate of 0.01 and root-mean-squared error (RMSE) as our loss function. For the deep network, we use a size 32 batching and implement a collection of image augmentations including random horizontal and vertical flipping, as well as cropping to a 512x512 image size. The full deep network modeling workflow is illustrated in Figure 2.

3 Results and Discussion

This section summarizes and validates our results, showcasing the application of our AGB 30-meter resolution (Figure 3) to assess how much carbon has been released to the environment from a major wildfire event.

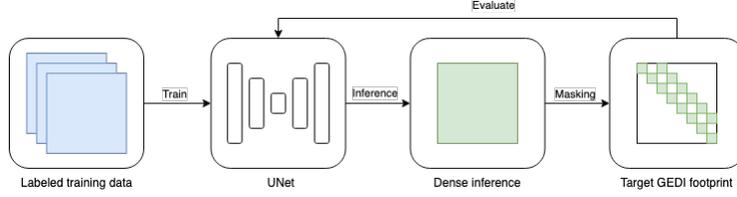


Figure 2: Our modeling workflow starting from training a multi-channel dataset using UNet and evaluating with sparse GEDI footprints

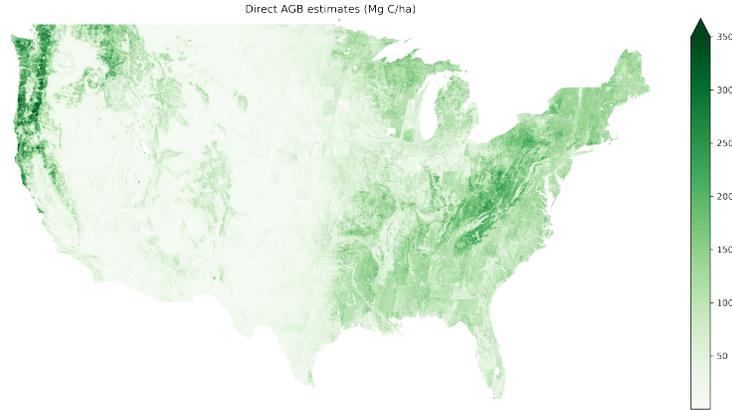


Figure 3: A 30-meter resolution AGB dense estimate for CONUS in the year 2021 summertime with Physics-informed masked variation of UNet

3.1 Model Performance and Ablation Study

As summarized in Table 1, UNet has the lowest validation RMSE of 37.93 ± 1.36 Mg C/ha, as compared to 81.95 ± 0.01 Mg C/ha, 53.37 ± 0.05 Mg C/ha, and 52.30 ± 0.03 Mg C/ha for linear regressor, gradient boosting, and random forest algorithms respectively. Furthermore, models that learn from SIF-based GPP, in addition to radar and optical hyperspectral imagery have lower validation RMSE of 37.93 ± 1.36 Mg C/ha than 41.99 ± 3.23 Mg C/ha in the latter case. This error accounts for 20-40% of the average AGB in CONUS, which is estimated by [18, 5, 19] to be around 100 and 200 MgC/ha. Nonetheless, UNet still exhibits larger uncertainty across model runs than the other ML-based models.

Table 1: Evaluation RMSE (Mg C/ha) for different combinations of inputs and models

Model	Inputs	Testing	Validation
Linear Regressor	SIF/S1/S2	66.07 ± 0.06	81.95 ± 0.01
	S1/S2	66.46 ± 0.10	84.33 ± 0.00
	S2-only	67.10 ± 0.11	90.99 ± 0.03
XGBoost	SIF/S1/S2	56.66 ± 0.06	53.37 ± 0.05
	S1/S2	57.35 ± 0.05	54.74 ± 0.03
	S2-only	57.82 ± 0.02	54.81 ± 0.26
RF	SIF/S1/S2	57.16 ± 0.05	52.30 ± 0.03
	S1/S2	58.05 ± 0.03	54.72 ± 0.06
	S2-only	58.12 ± 0.02	54.88 ± 0.18
UNet	SIF/S1/S2	48.83 ± 0.19	37.93 ± 1.36
	S1/S2	49.30 ± 0.18	41.99 ± 3.23
	S2-only	50.35 ± 0.43	45.93 ± 2.25

3.2 Consistency Check Across Climate Zones

Finally, we perform a consistency check to ensure that our AGB estimate agrees with expectations from literature. Figure 4 illustrates the distribution of AGB estimate at pixel-level across the top-6 (by area size) Köppen-based climate zones in CONUS [20]. As expected, the highest AGB estimates are observed at the dry summer temperate and humid subtropical regions (~ 70 - 300 MgC/ha), while the lowest ones are at the arid and semi-arid regions (~ 5 - 15 MgC/ha) [21, 22, 23, 19].

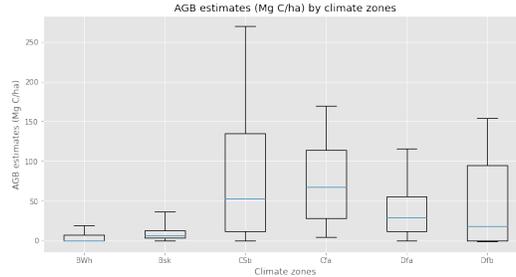


Figure 4: AGB estimates across different climate zones in CONUS - *BWh*: arid, *Bsk*: semi-arid, *Csb*: dry summer temperate, *Cfa*: humid subtropical, *Dfa*: humid continental, *Dfb*: humid continental

3.3 Application: Wildfire Impact Assessment

Lastly, we apply our AGB estimate to evaluate how much AGB has been lost after a major wildfire event [24]. We analyzed the Caldor wildfire event at California in the year 2021. The left panel in Figure 5 highlights the difference between AGB after (June 2022) and AGB before (June 2021) the fire, while the right panel indicates a normalized burn ratio (NBR) as derived from Sentinel-2 bands: $\left(\frac{B08-B12}{B08+B12}\right)$; lower NBR value suggests a burned area. We observe a close relationship between impact (AGB loss) and intensity (NBR). A robust AGB loss estimate, therefore, could help institutions assess fire risk mitigation strategies and forecast potential fire hazard [25], among many other.

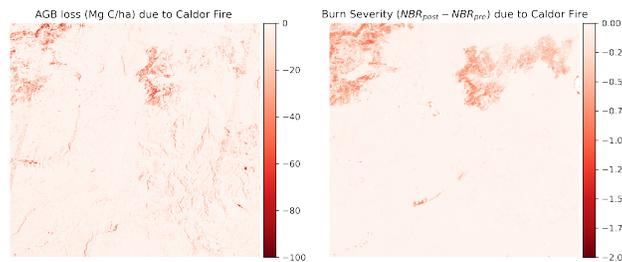


Figure 5: Left panel: AGB loss due to Caldor wildfire; Right panel: NBR index where lower value indicates more burning

4 Conclusion

In conclusion, we have demonstrated the transformation of sparse GEDI measurements into continuous map using multimodal sensing of optical and radar satellites and how the addition of physical parameter helps improve performance across models. And the use of a deep network, specifically the masked variation of UNet, further improves performance. We showed that our AGB estimate agrees with previous literature in terms of its consistency across climate zones. Lastly, we showcased how our AGB estimate can be used to assess wildfire impact, among many other interesting yet critical applications including for the evaluation of carbon credits and conservation projects for primary forests.

References

- [1] C. M. Albrecht, C. Liu, Y. Wang, L. Klein, and X. X. Zhu, “Monitoring urban forests from auto-generated segmentation maps,” *arXiv preprint arXiv:2206.06948*, 2022.
- [2] J. A. Anaya, E. Chuvieco, and A. Palacios-Orueta, “Aboveground biomass assessment in colombia: A remote sensing approach,” *Forest Ecology and Management*, vol. 257, no. 4, pp. 1237–1246, 2009.
- [3] Y. Zhang and S. Liang, “Fusion of multiple gridded biomass datasets for generating a global forest aboveground biomass map,” *Remote Sensing*, vol. 12, no. 16, p. 2559, 2020.
- [4] L. Duncanson, J. Armston, M. Disney, V. Avitabile, N. Barbier, K. Calders, S. Carter, J. Chave, M. Herold, T. W. Crowther, *et al.*, “The importance of consistent global forest aboveground biomass product validation,” *Surveys in geophysics*, vol. 40, no. 4, pp. 979–999, 2019.
- [5] R. Dubayah, J. Armston, J. Kellner, L. Duncanson, P. Healey, S.P. and Patterson, S. Hancock, H. Tang, M. Hofton, J. Blair, and S. Luthcke, “Gedi l4a footprint level aboveground biomass density, version 1.,” <https://doi.org/10.3334/ORNLDAAAC/1907>, 2021.
- [6] N. Lang, W. Jetz, K. Schindler, and J. Wegne, “A high-resolution canopy height model of the earth,” *arXiv preprint arXiv:2204.08322*, 2022.
- [7] X. Li and J. Xiao, “Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: A global, fine-resolution dataset of gross primary production derived from oco-2,” *Remote Sensing*, vol. 11, no. 21, p. 2563, 2019.
- [8] X. Wang, J. M. Chen, and W. Ju, “Photochemical reflectance index (pri) can be used to improve the relationship between gross primary productivity (gpp) and sun-induced chlorophyll fluorescence (sif),” *Remote Sensing of Environment*, vol. 246, p. 111888, 2020.
- [9] R. Dubayah, J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, *et al.*, “The global ecosystem dynamics investigation: High-resolution laser ranging of the earth’s forests and topography,” *Science of remote sensing*, vol. 1, p. 100002, 2020.
- [10] P. Fekety and A. Hudak, “Annual aboveground biomass maps for forests in the northwestern usa, 2000-2016,” *Oak Ridge, TN: National Laboratory Distributed Active Archive Center*, doi, vol. 10, 2019.
- [11] H. Tang, L. MA, A. Lister, J. O’Neil-Dunne, J. Lu, R. Lamb, R. Dubayah, and G. Hurtt, “Lidar derived biomass, canopy height, and cover for new england region, usa, 2015,” *ORNL DAAC*, 2021.
- [12] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, *et al.*, “Xgboost: extreme gradient boosting,” *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [14] J. Nathaniel, “Bias correction of global climate model using machine learning algorithms to determine meteorological variables in different tropical climates of indonesia,” 2020.
- [15] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [16] R. Zhang, C. Albrecht, W. Zhang, X. Cui, U. Finkler, D. Kung, and S. Lu, “Map generation from large scale incomplete and inaccurate data labels,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2514–2522, 2020.
- [17] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [18] T. Hu, Y. Su, B. Xue, J. Liu, X. Zhao, J. Fang, and Q. Guo, “Mapping global forest aboveground biomass with spaceborne lidar, optical imagery, and forest inventory data,” *Remote Sensing*, vol. 8, no. 7, p. 565, 2016.
- [19] M. Santoro, O. Cartus, N. Carvalhais, D. Rozendaal, V. Avitabile, A. Araza, S. De Bruin, M. Herold, S. Quegan, P. Rodríguez-Veiga, *et al.*, “The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations,” *Earth System Science Data*, vol. 13, no. 8, pp. 3927–3950, 2021.

- [20] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, “Present and future köppen-geiger climate classification maps at 1-km resolution,” *Scientific data*, vol. 5, no. 1, pp. 1–12, 2018.
- [21] N. L. Harris, D. A. Gibbs, A. Baccini, R. A. Birdsey, S. De Bruin, M. Farina, L. Fatoyinbo, M. C. Hansen, M. Herold, R. A. Houghton, *et al.*, “Global maps of twenty-first century forest carbon fluxes,” *Nature Climate Change*, vol. 11, no. 3, pp. 234–240, 2021.
- [22] Y. Yang, J. Fang, Y. Pan, and C. Ji, “Aboveground biomass in tibetan grasslands,” *Journal of Arid Environments*, vol. 73, no. 1, pp. 91–95, 2009.
- [23] J. Chave, M. Réjou-Méchain, A. Búrquez, E. Chidumayo, M. S. Colgan, W. B. Delitti, A. Duque, T. Eid, P. M. Fearnside, R. C. Goodman, *et al.*, “Improved allometric models to estimate the aboveground biomass of tropical trees,” *Global change biology*, vol. 20, no. 10, pp. 3177–3190, 2014.
- [24] W. Zhou and L. Klein, “Monitoring the impact of wildfires on tree species with deep learning,” *arXiv preprint arXiv:2011.02514*, 2020.
- [25] E. Rodrigues, B. Zadrozny, and C. Watson, “Wildfire risk forecast: An optimizable fire danger index,” *arXiv preprint arXiv:2203.15558*, 2022.