NetZeroCO_2, an AI framework for accelerated nature-based carbon sequestration

Ademir Ferreira da Silva
IBM Research - Brazil
Rio de Janeiro, Brazil
ademir.ferreira@br.ibm.com

Juan Nathaniel
Columbia University
New York City, USA

Ken C. L. Wong
IBM Research - Almaden
San Jose, USA

Campbell Watson
IBM TJ Watson Research Center
Yorktown Heights, USA

Hongzhi Wang
IBM Research - Almaden
San Jose, USA

Jitendra Singh
IBM Research - India
Gurgaon, India

Alexandre Alkmim Chamon
IBM Research - Brazil
Rio de Janeiro, Brazil

Levente Klein
IBM TJ Watson Research Center
Yorktown Heights, USA

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020359

***Abstract*—Nature-based carbon sequestration is currently the most viable solutions to extract CO_2 from the atmosphere and convert it into carbon. Oceans, soils and forests have the potential to capture and store large amount of carbon for decades. There is an ongoing debate about the permanence of the carbon sequestered by nature-based processes and the precise techniques required to monitor these carbon pools. Remote sensing plays a crucial role in the large scale observations of the Earth surface and provides a scalable method to monitor land use that can affect carbon sequestration. Optical spectral information and radar signals are the best candidates as proxy data to quantify and monitor the change in carbon sequestered. Here we outline the design of an AI enabled framework to monitor, verify, and quantify carbon sequestration in nature-based carbon sequestration processes.**

***Index Terms*—remote sensing, carbon sequestration, LiDAR, soil organic carbon, urban forests**

I. Introduction

Nature-based carbon sequestration (NBCS) is a set of natural processes that capture carbon from the atmosphere and trap it in carbon pools as biomass, soil, deep ocean, and underground reservoirs. NBCS can be easily adopted and deployed on a global scale if economic and social incentives exist. For example, with regenerative agriculture practices the soil carbon can be increased in a few years, compared with other NBCS solutions that takes decades, with the additional benefits of increased crop yield and soil health. Furthermore, carbon sequestration in soil represent 25% of the total carbon sequestration potential for NBCS [1].One concern of NBCS deployment for the carbon trade market is the reversibility of the carbon sequestration process which ideally should capture and hold carbon for decades. Reversal can happen through deforestation, wildfires, soil erosion and many other factors. In order to ensure that reversal is not taking place, there is a need for tools and platforms that can continuously monitor the land cover and generate early warning in case carbon sequestration is jeopardized, for instance, by identifying early the reversal processes and reporting the extents of inflicted areas. Global carbon maps can be assembled from continuous satellite coverage of the Earth to capture the spatial and temporal variations across eco-regions, land covers, topographies and climate zones. Satellites can acquire multi-spectral, hyper-spectral, radar or light detection and ranging (LiDAR) data but they do not measure directly the carbon sequestered in forest or soil. Satellite data need to be correlated with other data sets, like LiDAR, soil chemical sampling, weather, etc. Typically, LiDAR instruments acquire vegetation height information, and allometric models are used to convert height into biomass. For soil organic carbon (SOC) quantification, discrete soil samples need to be collected and their spectral data correlated with the SOC values. There are several well known research institutions that analyse and store soil samples globally, like the International Soil Reference and Information Centre (ISRIC) [2]. This methodology can provide a snapshot of the soil carbon storage, but can not be easily extended to monitor year to year changes in SOC.

In this work, we propose an AI driven framework, called NetZeroCO_2, for monitoring, quantification, and validation of NBCS processes. The goal is to enable an ecosystem of applications, such as, carbon credit quantification, net zero (carbon neutral) validation, regenerative agriculture monitoring and sustainable impact studies, by providing trustworthy and verifiable models. The framework aims to quantify, monitor and/or validate carbon sequestration through nature-based processes such as forests, coastlines, wetlands and croplands, either above ground or below ground on a global scale. The key differentiation of the NetZeroCO_2 framework is the ability to make seamless combinations of physical models, remote sensing and local field measurements with AI models that are reliable and interpretable across multiple spatial and temporal scales.

II. Related work

Biomass, SOC, rock carbonates, fossil fuels, together with the ocean and atmosphere are the most important natural carbon storage reservoirs. The carbon sequestration potential of these approaches is outlined in Table I.

CO_2 fluxes are continuously exchanged between the atmosphere, soil, vegetation, and oceans moving carbon be-

TABLE I
CARBON STORAGE POTENTIAL FOR NATURE BASED CARBON SEQUESTRATION APPROACHES

Carbon Pool	Methods	Carbon Storage PgC	Annual Increase PgC
Forests	Planting and restoring forest	381 [3]	4.6 [3]
Soils	Land management changes	2500 [4]	1.85 [4]
Oceans	Algae grows	38000 [5]	2.5 [6]

*PgC: Pentagrams of Carbon

tween these reservoirs through several natural mechanisms such as photosynthesis, respiration, CO_2 mineralization and decomposition of living organisms [5] [7]. Due to the subtle nature of CO_2 and massive ranges of carbon flux exchanges, point-wise measurements like local sensors or flux towers are difficult to scale and require combination with modeling to extrapolate measurements and predict time dependent carbon fluxes in locations where data may not exist. However, the number of satellites acquiring daily observations on the Earth surface has significantly extended in the last decade. Satellite data are used to classify land cover and understand human activities and continuously detect changes on the ground. Indeed, satellites are used to monitor and alert deforestation from Indonesia [8] to Brazil [9], Africa [10] and more. The land classification enables monitoring of both intact forests and individual trees, including tree species identification, tree health and tree geometrical factors [11] [12].

Satellite based observation detects light reflection from the ground, and it is sensitive of topsoil chemical and physical composition. Carefully chosen spectral reflectance bands can be used to quantify SOC storage when areas are not covered by vegetation. One of the emerging approaches for NBCS is quantification of SOC using multi-spectral or hyper-spectral imagery where reflectance from the soil is correlated with local SOC measurements [13] and data are interpolated using topography, land cover and land use practices. Some satellite signals can even penetrate the surface for a small distance, as in the case of Synthetic Aperture Radar, and they can identify morphological changes [14].

Hyper-spectral imaging collects spectral data between visible to near infrared (VIS-NIR) scale (400 nm to 2500 nm) in hundreds of spectral channels. For instance, Stenberg et al. [15] describe a visible–near infrared spectroscopy for soil application, with focus on SOC. In their review, they list several works which describe an 80-90% accuracy using VIS-NIR spectroscopy for SOC. Methodologies combining remote sensing and weather data with AI techniques can increase precision of SOC quantification. For example, Tiwari et al. [16] used a machine learning neural network algorithm to predict the amount of SOC through hyper-spectral satellite data. They used 65 reference points in an area of 189 km^2, and get 92% of accuracy. The work was validated only on a farm and scaling their techniques required a big data platform that can automate the task.

Live forests are responsible for a major portion of the carbon flux exchange between land and atmosphere with estimated carbon extraction of 4.6 PgC/year with significant regional variations [3]. Regenerative agriculture and afforestation can be explored to increase carbon sequestration and enable removal of 25% of the overall CO_2 emitted into the atmosphere [17] if more than one trillion trees are planted, although determining the right planting locations and forest type without damaging existing ecosystems and preserving the way of life for many communities are still unresolved issues [18]. Current approaches on quantifying the carbon storage in trees are based on allometric equations developed for large ecoregions and generic types of trees [19], with the main challenge in identifying individual trees and calculating the carbon stored above and below ground based on their age, growth rate and species characteristics.

National monitoring initiatives such as the Natural Resource Conservation Service (NRCS) – USA [20], Empresa Brasileira de Pesquisa Agropecuária (EMPRAPA) – Brazil [21], Commonwealth Scientific and Industrial Research Organisation (CSIRO) – Australia [22] are focused on understanding carbon capture on local scales with highly optimized and custom parameterized models. Unfortunately scaling this solution outside of the studied area is still challenging due to the absence of community agreed tools and methodologies to implement, monitor and validate the carbon capture. For example, considering the global nature of the problem, the optimization requires to be accurate from scales of centimeters (tree leaves) to continental scale (ecosystems) and be able to capture processes that happen from seconds to centuries. Such problems can be only solved only by integration of heterogeneous data sources and leveraging big data technologies to seamlessly process and model Petabytes of data.

III. NetZeroCO$_2$, an AI framework

One challenge of enabling a carbon trading market is the lack of tools and methods to reliably quantify the carbon sequestered by nature-based solutions. Direct measures of carbon sequestered in soil require manual soil sample collection and lab analysis, while for trees an uproot of trees and incineration are needed to calculate biomass and CO_2 generation. Both these methods are slow, expensive, and hard to address global scale and alternative methods like satellite data infused with AI methods are more suitable to tackle this challenge, being the choice of this work.

Global continuous satellites imagery like Sentinel and Landsat are generating yearly Petabytes of data that enables almost weekly monitoring of every location on Earth. Geospatial platforms like IBM PAIRS [23], [24], Google Earth Engine [25] and AWS Earth provision data access and computational resources to massive geospatial data. While most platforms store data in a file storage system, the PAIRS platform indexes every pixel and harmonizes data across multiple spectral bands. One advantage of pixel-based indexing of the data is the possibility to quickly filter out locations that share similarity or differences and carry out simple statistics on data.

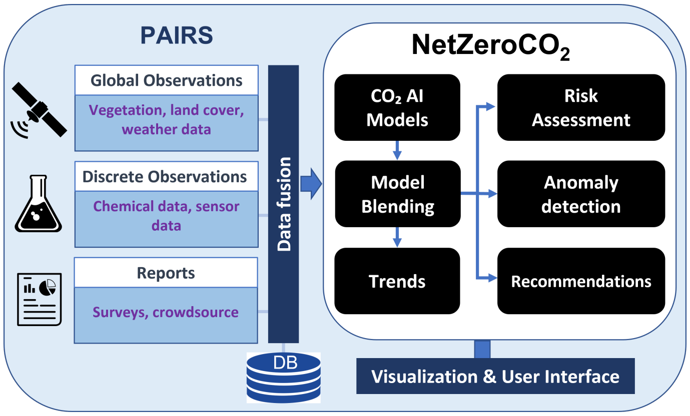

Fig. 1. NetZeroCO_2 framework architecture for quantifying nature-based carbon sequestration.

The NetZeroCO_2 framework is built on top of PAIRS and other enabling IBM technologies to allow rapid analysis of multi-modal data such as LiDAR, radar, multi-spectral and hyper-spectral satellite data, soil sample chemical results, weather data, survey-based results, etc. Due to the multi-dimensionality of the data, spanning from discrete measurements like soil sample chemical results, to global observations as three-dimensional LiDAR data, weather data, or even getting survey inputs, all datasets are georectified and aligned in space and time. The NetZeroCO_2 framework comprises of the following modules, as depicted in Figure 1:

• CO_2 AI models: The ability to automate the AI learning process can improve the speed and flexibility of learning. Traditional AI systems require data download for creating training data sets, this task can be achieved using API calls from PAIRS database which has more than 5 Petabytes of data currently available. Furthermore, the data is prepared in machine learning feature vectors, being readily available to use in AI learning models. The AI modules also contain both machine learning and deep learning models. Learning can be on pixel-based approaches like random forest, support vector machine or it can be object and image-based learning, like U-NET.

• Model blending: Once AI models are trained, they can be blended with existing models created by other parties that quantify carbon sequestration. Model blending may be required to better capture local variability or minimize the overall error and uncertainty of the final predictions.

• Trends: Currently, two models are integrated: Above Ground Biomass (AGB) estimate and SOC quantification, as described in the following section. The framework is designed to calculate and validate baseline carbon pools on a defined area and track changes of carbon sequestration.

• Recommendations: Based on land use cover for a given locations, the framework can process the best composition of forestry and crop cover to maximize carbon sequestration by recommending reforestation or carbon farming practices.

• Anomaly detection: The sequestered carbon can be quantified at different times and spatial resolutions, dependent on the

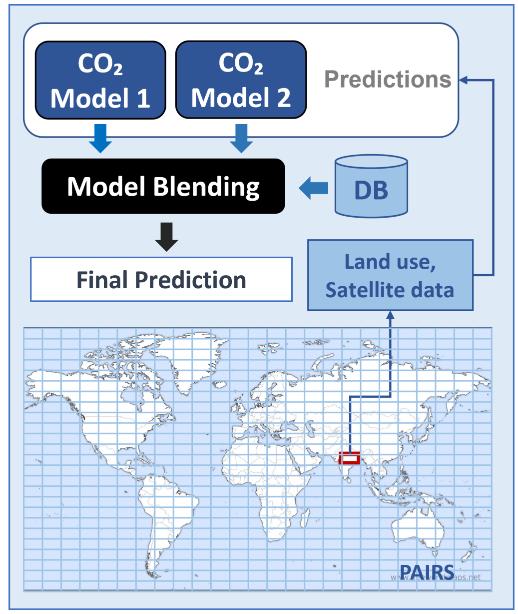

Fig. 2. AI based model selection for reliable carbon credit quantification and prediction

applications, for example, weekly basis carbon flux exchange or yearly Greenhouses Gas (GHG) emission at country level for regulatory purposes. These maps can be also used for change detection of biomass loss due to wildfires [26], [27].

• Risk assessment: With high social desire for a rise of green economy, a quantitative rating on NBCS projects, like reforestation projects, are very welcome. Risk assessment may be calculated based on regional surveys, local crowd-sensing data or even with historical Normalized Difference Vegetation Index (NDVI) trends for the area of interest.

The NetZeroCo_2 is ideally suited as an AI framework to enable carbon quantification, to monitor and to verify carbon accumulation or loss using machine learning models and multi-modal data. Some use cases are developed to tackle multiple challenges related to carbon quantification: i) Verify carbon credits and long-term climate risks; ii) Monitor carbon sequestration trends and predict future trajectories with quantified uncertainties; iii) Carbon credit trading on blockchain; iv) Carbon credit rating based on climate variables; v) Environment, Social & Governance reporting support using current satellite-based observations; and vi) Validate NetZero compliance. The quantifiable results provided by NetZeroCO_2 can be leveraged by industries that rely on land use like agriculture, transportation, mining, forestry, and energy.

One practical use case for the NetZeroCo_2 framework is to combine models in order to provide the best available carbon sequestration information for a given area. For instance, the framework is based on nested geospatial grids where the resolution of grid cells ranges from the sub-meter to hundreds of kilometers, dependent on the native data resolution. A grid resolution can be chosen based on the region of interest and the availability of geospatial data to quantify the sequestered car-

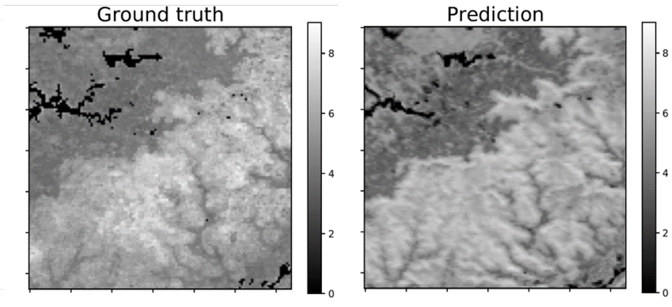

Fig. 3. Ground truth and prediction of soil organic carbon (g/kg).

bon. Once an area of interest is chosen, the data available for that region is extracted from PAIRS [23]. The multi-spectral data is used to identify the dominant land cover. Land cover classification can be carried out using pixel-based classification using Random Forest or Support Vector Machine methods. Once the land use is identified, the overall equivalent CO_2 can be calculated using standard methods, like Intergovernmental Panel on Climate Change (IPCC) emission factors (model M1 in Fig 2). This model can be a carbon sequestration baseline on a globally agreed quantification methodology. Other types of databases such as hyper-spectral, radar and LiDAR data can be also used to quantify carbon sequestration using a set of custom AI models. These internal models (depicted as model M2) can be run on PAIRS using the Climate Impact Modelling Framework (CIMF) method to leverage cloud computation [28]. Besides, some predefined rules can be applied to promote the best possible prediction between models. For example, in a scenario where local experts concluded that, in a specific cell of the grid, a SOC concentration cannot be achieved. Thus, considering a predefined threshold value decided by these local experts, the framework automatically selects from both models the most plausible SOC result for that cell.

IV. NetZeroCO$_2$ applications for carbon sequestration

The proposed framework is designed to develop multiple NBCS applications leveraging a common data and modelling stack. Currently, we are focusing on two key applications, described below in detail.

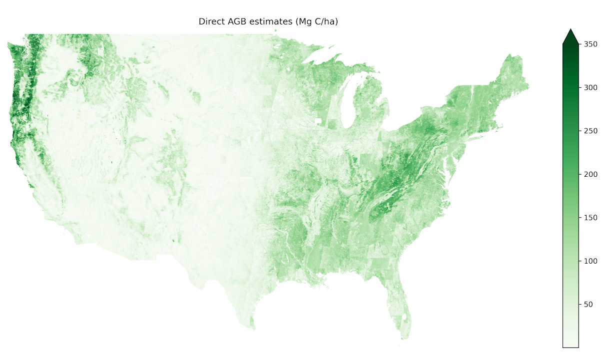

Fig. 4. Above ground biomass - 30 m resolution (MgC/ha)

A. Soil organic carbon estimation

As per estimates, soils can sequester 1.85 PgC/year from the atmosphere if best soil management practices can be implemented [4]. SOC is considered the main reserve of soil carbon, being also an important component in the earth carbon cycle. SOC can be enhanced with a variety of methods and practices, such as forest management, agroforestry, and conservation tillage on croplands [29]. Estimating SOC at a global scale is a challenging task due to variability in soils, complex biochemical processes, and different management practices [30]. Traditionally, soil is sampled at different depths and lab analyzed to measure SOC, but manual sampling is impractical to monitor SOC change at a global scale. In contrast, satellite data can provide a complementary, scalable, and cost-effective alternative. Remote sensing through satellites has the potential to enable measurement, reporting, and verification of SOC across the globe, with year-to-year tracking of carbon storage and carbon cycle disruptions when carbon sequestration practices are implemented [31]. Figure 3 shows the ground truth and SOC quantification using using NetZeroCO$_2$ framework.

B. Biomass estimation

Nature through photosynthesis converts CO_2 to biomass. Trees are a natural sink for atmospheric carbon dioxide when entrapped into the tree's biomass. Besides forests, trees in urban areas constitute a significant storage of carbon with direct impact on a city's carbon footprint assessments. One of the practical challenges in NBCS considering biomass is effective quantification [32]. There is ongoing interest to exploit satellite imagery to model carbon stored in trees. Modeling is commonly based on dimensional scaling models where tree height is related to canopy diameter, and tree biomass is correlated with carbon storage. These scaling models termed allometric equations are popular within the remote sensing community [33] [34]. Previous studies did demonstrate the utility of allometric equations [32] and the methodology got adopted to large geographies and ecoregions [34]. Uncertainty quantification in addition to estimates of carbon sequestered must be part of carbon sequestration modeling. To identify the height and distribution of vegetation, LiDAR is a commonly used remote sensing approach.

Access to high quality LiDAR data may not be always possible due to significant financial resources required to conduct surveys. Exploiting open source, non-classified point cloud data may be an alternative in combination with satellite data. High quality LiDAR may serve as a source to calibrate multi-spectral satellite images across small geospatial patches. After calibration, the use of these images in other similar areas serves to estimate the biomass and also calibrate tree growth models [35]. To scale such models globally, multi-modal data processing (LiDAR, satellite, ortho-imagery, radar, and crowd-sourcing data) is required to generate consistent maps across different geographies, and to identify the best available data for a given geolocation and analytic task [23]. Figure 4 shows

TABLE II

SOC ESTIMATION WITH SATELLITE IMAGES. RMSE, MAPE, AND SSIM REPRESENT ROOT MEAN SQUARED ERROR, MEAN ABSOLUTE PERCENTAGE ERROR, AND STRUCTURAL SIMILARITY, RESPECTIVELY. BEST RESULTS ARE HIGHLIGHTED.

Metric	RMSE (g/kg)	MAPE (%)	SSIM
Random forest	2.50±2.37	45.40±46.30	0.07±0.08
Modified V-Net	2.00±1.99	29.79±21.11	0.17±0.12
Fourier neural operator	**1.96±1.96**	**28.32±18.29**	**0.18±0.13**

an AGB estimate of the Contiguous United States (CONUS) area with 30 m resolution using the NetZeroCO$_2$ framework.

V. RESULTS AND DISCUSSION

A. SOC estimation

For model development to estimate SOC, six spectral bands from the Moderate Resolution Imaging Spectroradiometer (MODIS) [36], i.e., blue (459–479 nm), green (545–565 nm), red (620–670 nm), near infrared (841–876 nm), and shortwave infrared (SWIR1: 1230–1250 nm, SWIR3: 2105–2155 nm), were used as the predictors. The SOC data from SoilGrids [30] in the top 5 cm were the predictands. Sampling from the regions of USA, Mexico, and Canada, a total of 3059 samples were generated. The dataset was partitioned into 50% for training, 20% for validation, and 30% for testing, and identical partitions were used in all experiments. All images were adjusted to 128 x 128 pixels with a spatial resolution of 500 m.

Instead of using traditional machine learning methods such as random forests and support vector machines, we study the use of convolutional neural networks (CNN) on SOC estimation. Table II shows performance of machine learning models. The modified V-Net [37] is a CNN model originally proposed for image segmentation, and the Fourier neural operator [38] is a CNN model based on Fourier transform. The results show that the deep learning models outperformed the random forest. Fig. 3 shows that the SOC prediction using the Fourier neural operator has the best performance compared to the ground truth.

B. Biomass estimation

To compose a data model for biomass estimation, we collected data of CONUS area from four types of satellites: i) LiDAR satellite, GEDI [39]; ii) radar satellite, SENTINEL-1 [40]; iii) multi-spectral satellite, SENTINEL-2 [40], and Solar Induced Fluorescence (SIF) data adapted from the MODIS satellite [41].

• LiDAR: The Global Ecosystem Dynamics Investigation (GEDI) instrument uses high resolution laser ranging observations of the 3-dimensional structure of the Earth. In this work, we were interested in L2A and L4A pre-processed data level. L2A level provides ground elevation, canopy top height and relative height metrics. L4A level is the reference data to estimate biomass and corresponds to the footprint above ground biomass density. For this satellite we selected a period from 12/01/2020 to 08/31/2021. In total the data was collected around 15K GEDI measurement points, being 14354 points from L4A level and 620 points from L2A level.

• Radar: SENTINEL-1 is a satellite operating day and night performing C-band synthetic aperture radar imaging, enabled to acquire imagery regardless of the weather. We used L1-IW pre-processed data level in Interferometric Wide swath mode (IW), comprising complex imagery with amplitude and phase, for both VV and VH polarization, with spatial resolution of 5 x 20 m.

• Multi-spectral: SENTINEL-2 aims at monitoring variability in land surface conditions. We used L2A pre-processed data level for all bands (12 bands). L2A gave us a bottom-of-atmosphere reflectance. Both for SENTINEL-1 and SENTINEL-2 we selected a period from 06/01/2021 to 08/31/2021, being collected 23118 areas.

• SIF: Solar-Induced chlorophyll fluorescence satellite measurements are based on a phenomenon where, by the photosynthesis process, plant chlorophyll emits two-peak wavelengths in spectrum between 650–800 nm [42]. We chose SIF-GPP, a Gross Primary Productivity (GPP) index, for assessing plant productivity [43]. In this experiment we used aggregated monthly data in period from June/2021 to August/2022, with spatial resolution of 0.05 degrees.

TABLE III

EVALUATION RMSE (MG C/HA) FOR DIFFERENT COMBINATIONS OF INPUTS AND MODELS

Model	Inputs	Testing	Validation
Linear Regression	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
Random Forest	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
U-Net	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

All satellite data were combined for each latitude and longitude creating a ”data cube”. As mentioned, we considered for data training the CONUS area. We applied a filter discarding points in the data cube that do not contain data for all satellites, ending with 2402 areas, representing a volume of 358 G bytes of data training. It was used 20% of this dataset as testing data. We performed root mean square error (RMSE) analysis to evaluate the performance of machine learning models, table III. It was compared the performance of Linear Regression, Random Forest, XGBoost and U-NET models in 3 different scenarios: i) S2-only: multi-spectral data only (SENTINEL-2); ii) S1/S2: multi-spectral and radar data (SENTINEL-1 and SENTINEL-2) and iii) SIF/S1/S2: identical to the previous one with inclusion of SIF data. By comparison, the method with better performance, the lowest RMSE, combines all available datasets, SIF/S1/S2. It this case, U-NET was the best performing model, with 48.83 +- 0.19 Megagram of carbon per hectare (MgC/ha), in the test set. This is a CNN technique designed for finding patterns on images that also greatly reduce normally required annotated training samples [44]. U-Net was configured with 4 layers with padding of 1, kernel size 2 and double convolution for downscale and upscale hidden layers.

Calibration of carbon sequestration models in NBCS is an ongoing challenge as direct measurements are hard and expensive to carry out. AI driven models in combination with satellite imagery can capture spatial and temporal variations due to the dense coverage of Earth. NetZeroCO_2 serves both as a framework to quantify carbon sequestration in soil and forestry as well as a monitoring tool for carbon credit integrity. Using AI ensemble learning it can combine existing open source and custom models to determine the best prediction of carbon sequestration at multiple spatio-temporal scales. The framework will eventually allow both researchers, developers and practitioners to develop emerging applications in the area of NBCS.

VI. Acknowledgment

We would like to thanks Bruce Elmegreen, Mathias Steiner, Hendrik Hamann, Tanveer Syeda-Mahmood, and Conrad Albrecht for stimulating and inspiring discussions and guidance.

References

[1] D. A. Bossio, S. C. Cook-Patton, P. W. Ellis, J. Fargione, J. Sanderman, and P. e. a. Smith, "The role of soil carbon in natural climate solutions," *Nature Sustainability*, vol. 3, no. 5, pp. 391–398, 2020.

[2] T. Hengl, J. Mendes de Jesus, G. B. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger *et al.*, "Soilgrids250m: Global gridded soil information based on machine learning," *PLoS one*, vol. 12, no. 2, p. e0169748, 2017.

[3] L. Xu, S. S. Saatchi, Y. Yang, Y. Yu, J. Pongratz, A. A. Bloom, K. Bowman, J. Worden, J. Liu, Y. Yin *et al.*, "Changes in global terrestrial live biomass over the 21st century," *Science Advances*, vol. 7, no. 27, p. eabe9829, 2021.

[4] R. J. Zomer, D. A. Bossio, R. Sommer, and L. V. Verchot, "Global sequestration potential of increased organic carbon in cropland soils," *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

[5] R. Lal, "Sequestration of atmospheric co 2 in global carbon pools," *Energy & Environmental Science*, vol. 1, no. 1, pp. 86–100, 2008.

[6] J. Hauck, M. Zeising, C. Le Quéré, N. Gruber, D. C. Bakker, L. Bopp, T. T. T. Chau, Ö. Gürses, T. Ilyina, P. Landschützer *et al.*, "Consistency and challenges in the ocean carbon sink estimate for the global carbon budget," *Frontiers in Marine Science*, vol. 7, p. 571720, 2020.

[7] H. Riebeek *et al.*, "The carbon cycle," *NASA Earth Observatory*, vol. 16, 2011.

[8] B. A. Margono, P. V. Potapov, S. Turubanova, F. Stolle, and M. C. Hansen, "Primary forest cover loss in indonesia over 2000–2012," *Nature climate change*, vol. 4, no. 8, pp. 730–735, 2014.

[9] D. F. d. Silva, I. Muraoka, and E. C. Garcia, "Thermal control design conception of the amazonia-1 satellite," *Journal of Aerospace Technology and Management*, vol. 6, pp. 169–176, 2014.

[10] P. V. Potapov, S. A. Turubanova, M. C. Hansen, B. Adusei, M. Broich, A. Altstatt, L. Mane, and C. O. Justice, "Quantifying forest cover loss in democratic republic of the congo, 2000–2010, with landsat etm+ data," *Remote Sensing of Environment*, vol. 122, pp. 106–116, 2012.

[11] E. Ayrey and D. J. Hayes, "The use of three-dimensional convolutional neural networks to interpret lidar for forest inventory," *Remote Sensing*, vol. 10, no. 4, p. 649, 2018.

[12] M. Immitzer, C. Atzberger, and T. Koukal, "Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data," *Remote sensing*, vol. 4, no. 9, pp. 2661–2693, 2012.

[13] R. V. Rossel, T. Behrens, E. Ben-Dor, D. Brown, J. Demattê, K. D. Shepherd, Z. Shi, B. Stenberg, A. Stevens, V. Adamchuk *et al.*, "A global spectral library to characterize the world's soil," *Earth-Science Reviews*, vol. 155, pp. 198–230, 2016.

[14] A. Bartsch, B. Widhalm, P. Kuhry, G. Hugelius, J. Palmtag, and M. B. Siewert, "Can c-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra?" *Biogeosciences*, vol. 13, no. 19, pp. 5453–5470, 2016.

[15] B. Stenberg, R. A. V. Rossel, A. M. Mouazen, and J. Wetterlind, "Visible and near infrared spectroscopy in soil science," *Advances in agronomy*, vol. 107, pp. 163–215, 2010.

[16] S. K. Tiwari, S. K. Saha, S. Kumar *et al.*, "Prediction modeling and mapping of soil carbon content using artificial neural network, hyperspectral satellite data and field spectroscopy," *Advances in Remote Sensing*, vol. 4, no. 01, p. 63, 2015.

[17] J.-F. Bastin, Y. Finegold, C. Garcia, D. Mollicone, M. Rezende, D. Routh, C. M. Zohner, and T. W. Crowther, "The global tree restoration potential," *Science*, vol. 365, no. 6448, pp. 76–79, 2019.

[18] B. B. Strassburg, A. Iribarrem, H. L. Beyer, C. L. Cordeiro, R. Crouzeilles, C. C. Jakovac, A. Braga Junqueira, E. Lacerda, A. E. Latawiec, A. Balmford *et al.*, "Global priority areas for ecosystem restoration," *Nature*, vol. 586, no. 7831, pp. 724–729, 2020.

[19] N. Picard, L. Saint-André, and M. Henry, "Manual for building tree volume and biomass allometric equations: from field measurement to prediction," *Manual for building tree volume and biomass allometric equations: from field measurement to prediction, FAO; Food and Agricultural Organization of the United Nations (2012)*, 2012.

[20] S. Wills, T. Loecke, C. Sequeira, G. Teachman, S. Grunwald, and L. T. West, "Overview of the us rapid carbon assessment project: sampling design, initial summary and uncertainty estimates," in *Soil carbon*. Springer, 2014, pp. 95–104.

[21] J. Polidoro, M. d. L. Mendonça-Santos, J. Lumbreras, M. Coelho, A. d. Carvalho Filho, P. Da Motta, W. d. Carvalho Junior, J. de Araujo Filho, G. Curcio, J. Correia *et al.*, "Programa nacional de solos do brasil (pronasolos)." *Embrapa Solos-Documentos (INFOTECA-E)*, 2016.

[22] E. Bui, B. Henderson, and K. Viergever, "Using knowledge discovery with data mining from the australian soil resource information system database to inform soil carbon mapping in australia," *Global biogeochemical cycles*, vol. 23, no. 4, 2009.

[23] L. J. Klein, F. J. Marianno, C. M. Albrecht, M. Freitag, S. Lu, N. Hinds, X. Shao, S. B. Rodriguez, and H. F. Hamann, "Pairs: A scalable geo-spatial data analytics platform," in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015, pp. 1290–1298.

[24] S. Lu and H. F. Hamann, "Ibm pairs: Scalable big geospatial-temporal data and analytics as-a-service," *In Handbook of Big Geospatial Data*, pp. 3–34, 2021.

[25] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote sensing of Environment*, vol. 202, pp. 18–27, 2017.

[26] J. Nathaniel, L. J. Klein, C. D. Watson, G. Nyirjesy, and C. M. Albrecht, "Aboveground carbon biomass estimate with physics-informed deep network," *arXiv:2210.13752*, 2022.

[27] W. Zhou and L. J. Klein, "Monitoring the impact of wildfires on tree species with deep learning," *rXiv:2011.02514*, 2022.

[28] B. Edwards, P. Fraccaro, N. Stoyanov, N. Bore, J. Kuehnert, K. Weldemariam, and A. Jones, "Climate impact modelling framework," *arXiv:2209.12080*, 2022.

[29] K. Lorenz and R. Lal, "Soil organic carbon sequestration in agroforestry systems. a review," *Agronomy for Sustainable Development*, vol. 34, no. 2, pp. 443–454, 2014.

[30] L. Poggio, L. M. De Sousa, N. H. Batjes, G. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter, "Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, 2021.

[31] P. Smith, J.-F. Soussana, D. Angers, L. Schipper, C. Chenu, D. P. Rasse, N. H. Batjes, F. Van Egmond, S. McNeill, M. Kuhnert *et al.*, "How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal," *Global Change Biology*, vol. 26, no. 1, pp. 219–241, 2020.

[32] S. Brown, "Measuring carbon in forests: current status and future challenges," *Environmental pollution*, vol. 116, no. 3, pp. 363–372, 2002.

[33] T. Jucker, J. Caspersen, J. Chave, C. Antin, N. Barbier, F. Bongers, M. Dalponte, K. Y. van Ewijk, D. I. Forrester, M. Haeni *et al.*, "Allometric equations for integrating remote sensing imagery into forest

monitoring programmes," *Global change biology*, vol. 23, no. 1, pp. 177–190, 2017.

[34] J. M. Omernik and G. E. Griffith, "Ecoregions of the conterminous united states: evolution of a hierarchical spatial framework," *Environmental management*, vol. 54, no. 6, pp. 1249–1266, 2014.

[35] L. J. Klein, W. Zhou, and C. M. Albrecht, "Quantification of carbon sequestration in urban forests," *arXiv preprint arXiv:2106.00182*, 2021.

[36] C. Justice, J. Townshend, E. Vermote, E. Masuoka, R. Wolfe, N. Saleous, D. Roy, and J. Morisette, "An overview of modis land data processing and product status," *Remote sensing of Environment*, vol. 83, no. 1-2, pp. 3–15, 2002.

[37] K. C. L. Wong, M. Moradi, H. Tang, and T. Syeda-Mahmood, "3D segmentation with exponential logarithmic loss for highly unbalanced object sizes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, ser. LNCS, vol. 11072, 2018, pp. 612–619.

[38] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," in *International Conference on Learning Representations*, 2021.

[39] "Gedi ecosystem lidar," https://gedi.umd.edu/data/documents/, accessed: 2022-10-29.

[40] "sentinelhub," https://www.sentinel-hub.com/, accessed: 2022-10-29.

[41] "Global ecology," http://data.globalecology.unh.edu/data/Mean/, accessed: 2022-10-29.

[42] M. Meroni, M. Rossini, L. Guanter, L. Alonso, U. Rascher, R. Colombo, and J. Moreno, "Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications," *Remote sensing of environment*, vol. 113, no. 10, pp. 2037–2051, 2009.

[43] 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.

[44] 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*. Springer, 2015, pp. 234–241.