Data Descriptor

Terrestrial Carbon Storage Estimation in Guangdong Province (2000–2021)

Wei Wang , Yueming Hu , Xiaoyun Mao * , Ying Zhang , Liangbo Tang , Junxing Cai

Posted Date: 24 February 2025

doi: 10.20944/preprints202502.1820.v1

Keywords: Terrestrial ecosystems; Carbon storage; Spatiotemporal patterns; Convolutional neural network; Guangdong Province

 doi:10.20944/preprints202502.1820.v1

Data Descriptor

Terrestrial Carbon Storage Estimation in Guangdong Province (2000–2021)

Wei Wang [1,2,3], Yueming Hu [4], Xiaoyun Mao [1,*], Ying Zhang [5], Liangbo Tang [2,3] and Junxing Cai [2,3]

[1] College of Resources and Environment, South China Agricultural University, Guangzhou, Guangdong Province, 510,640; gdlucc@stu.scau.edu.cn

[2] College of Biology and Agriculture, Shaoguan University, Shaoguan, Guangdong Province, 512005;tangliangbo@sgu.edu.cn(L.T.),CJX@sgu.edu.cn(J.C.)

[3] Guangdong Provincial Engineering Research Center for Efficient Utilization of Water and Soil Resources in Northern Guangdong

[4] College of Tropical Crops, Hainan University, Haikou, Hainan Province, 995149@hainanu.edu.cn

[5] Tangshan Vocational and Technical College, Tangshan, Hebei Province, 063000;zy529@stu.scau.edu.cn

* Correspondence: xymao@scau.edu.cn

Abstract: (1) Terrestrial ecosystems are critical carbon sinks, and accurate assessment of their carbon storage is vital for understanding global carbon cycles and formulating climate change mitigation strategies.; (2)This study integrated vegetation indices, meteorological factors, land use data, soil/vegetation types, field sampling, and a convolutional neural network (CNN) model to estimate the carbon storage of terrestrial ecosystems in Guangdong Province; (3) Total carbon storage increased by 0.11 Pg from 2000 to 2021, with vegetation carbon gains (+0.19 Pg) offsetting soil carbon losses (-0.08 Pg), the latter primarily driven by reduced soil carbon in forest ecosystems; (4) Northern and eastern Guangdong exhibit high potential for enhancing carbon storage capacity, which is crucial for achieving regional carbon peaking and neutrality targets. **Dataset:** DOI number or link to the deposited dataset in cases where the dataset is published or set to be published separately. If the dataset is submitted and will be published as a supplement to this paper in the journal Data, this field will be filled by the editors of the journal. In this case, please make sure to submit the dataset as a supplement when entering your manuscript into our manuscript editorial system. **Dataset License:** license under which the dataset is made available (CC0, CC-BY, CC-BY-SA, CC-BY-NC, etc.)

Keywords: terrestrial ecosystems; carbon storage; spatiotemporal patterns; convolutional neural network; Guangdong province

1. Summary

This dataset comprises estimates of total terrestrial ecosystem carbon storage, vegetation carbon storage, soil carbon storage, total carbon density, vegetation carbon density, and soil carbon density in Guangdong Province, China. The data were estimated using a convolutional neural network (CNN) model integrating multisource data, including remote sensing data, meteorological data, land use/cover data, vegetation and soil types, and field sampling data. The sampling campaign was supported by the National Key R&D Program of China [Grant No. 2023YFD1900100]. Public access to this dataset will contribute to advancing regional carbon cycling research and further enhance the accuracy of terrestrial ecosystem carbon storage estimation.

- **Dataset DOI**: 10.5281/zenodo.14835471
- **Temporal Coverage**: 2000, 2005, 2010, 2015, 2018, 2021
- **Geographic Coverage**: Guangdong Province, China (20.13°–25.31°N, 109.68°–117.20°E)
- **Data Format**: GeoTIFF (raster), CSV (tabular), Shapefile (vector)

 doi:10.20944/preprints202502.1820.v1

2. Data Description

Table 1. Data Sources.

Data Type	Temporal Coverage	Spatial Resolution	Source
Field data	2018,2021	--	Field sampling and surveys
LUC	2000-2021	30 m	GLC_FS30, doi:10.12237/casearth.64d094d1819aec27a589a856
VEG	--	1 km	Resource and Environment Science Data Center (www.resdc.cn)
SOIL	--	1 km	HWSD2.0, doi:10.4060/cc3823en
TEMP/PRE	2000-2021	1 km	Resource and Environment Science Data Center (doi:10.12078/2022082501)
RESI, NPP, NDVI, EVI	2000-2021		MODIS data processed via Google Earth Engine (GEE)
DEM		30m	ASTER GDEM V3 (www.gscloud.cn)

2.1. Land Use/Cover Data

- Source: GLC_FS30 (Global Land Cover Fine Classification Product)
- Resolution: 30 m
- Temporal Span: 2000–2020 (extended to 2021 via temporal interpolation)
- Processing Steps:

  Reclassified into 10 forest, 9 shrub/grassland, and 4 cropland subtypes.
  Aligned to WGS 1984 UTM Zone 49N coordinate system using ArcGIS Pro 3.0.

  Access: DOI:10.12237/casearth.64d094d1819aec27a589a856

2.2. Remote Sensing Indices

- Variables: NDVI, EVI, RESI (Remote Sensing Ecological Index), NPP (Net Primary Productivity)
- Source: MODIS products (MOD13Q1, MOD17A3H) via Google Earth Engine (GEE)
- Resolution: 250 m (NDVI/EVI), 500 m (NPP)
- Processing:

  Vegetation growing season (April–October) maximum value compositing.
  Masked for cloud cover using QA bands.
  Access: NASA Earthdata (requires GEE API access)

2.3. Meteorological Data

- Variables: Mean annual temperature (TEMP), total annual precipitation (PRE)
- Source: Resource and Environment Science Data Center (RESDC)

  Resolution: 1 km (spatially interpolated from station data)

- Method: Thin-plate spline interpolation with elevation correction.
- Access: DOI:10.12078/2022082501

2.4. Soil and Vegetation Data

- Soil Type: HWSD2.0 (Harmonized World Soil Database v2.0)
- Resolution: 1 km
- Key Parameters: Organic carbon density (0–30 cm depth).
- Vegetation Type: RESDC Vegetation Atlas of China
- Classification: 12 vegetation subtypes (e.g., subtropical evergreen broadleaf forest).

 doi:10.20944/preprints202502.1820.v1

- Access: FAO HWSD | RESDC

2.5. Field Sampling Data

- Soil Samples: 2,316 sites (0–30 cm depth, organic carbon measured via dry combustion).
- Vegetation Samples: 1,264 sites (aboveground biomass measured by destructive sampling).
- Quality Control:
- Outliers removed using ±3σ threshold.
- Spatial representativeness validated via Thiessen polygon analysis.
- Access: Restricted (available upon request for academic use).
- Numbered lists can be added as follows:

2.6. Data Processing Workflow

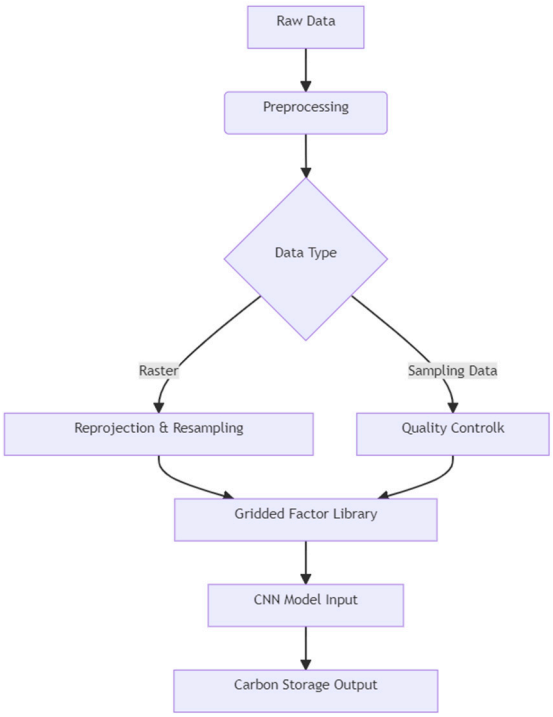

Figure 1. Data Processing Workflow.

2.7. File Structure:

Carbon storage and carbon density.rar
- scd_gd_00p.tif (Soil carbon density_GuangDong_2000)
- ...
- vcd_gd_00p.tif (Vegetation carbon density_GuangDong_2000)
- ...
- tcd_gd_00p.tif (Total carbon density_GuangDong_2000)
- ...
- scs_gd_00p.tif (Soil carbon storage_GuangDong_2000)
- ...
- vcs_gd_00p.tif (Vegetation carbon storage_GuangDong_2000)
- ...
- Tcs_gd_00p.tif (Total carbon storage_GuangDong_2000)
- ...

 doi:10.20944/preprints202502.1820.v1

3. Methods

All variables for 2000–2021 were preprocessed to unify coordinate systems, clip study area boundaries, and assemble a time-series carbon storage factor database.

Field sampling data were filtered using ArcGIS geostatistical tools to remove outliers (e.g., values beyond ±3σ). A 500m×500m grid was overlaid across the province, with factor values extracted at grid centroids to generate carbon storage base maps.

Convolutional Neural Networks (CNNs)—a deep learning architecture specialized in grid-structured data processing—automatically extract spatial features through hierarchical learning Our model comprised [1–3]:

- **Input layer**: 500m-resolution multisource data grids.
- **Convolutional layers**: 3 layers with 32–64 filters to capture spatial patterns.
- **Pooling layers**: Max-pooling for dimensionality reduction.
- **Fully connected layers**: 2 layers mapping features to carbon storage values.
- **Output layer**: Predicted vegetation/soil carbon densities.

Author Contributions: W.W.: writing—original draft, software, visualization, and methodology. Y.H.: supervision, resources, funding acquisition. X.M.: project administration, and writing—review and editing. Y.Z.: data curation. L.T. software, visualization. J.C.: validation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China [grant number 2023YFD1900100]; the Shaoguan Science and Technology Plan Project [grant number 220531134531827].

Data Availability Statement: Data are available and can be provided upon request.

Acknowledgments: We are thankful to the anonymous reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Lei L, Xu B Q, Gao Q J, et al. Extended-range forecasting method of summer daily maximum temperature in the Yangtze River Basin based on convolutional neural network. Transactions of Atmospheric Sciences, 2022,45(06):835-849. doi:10.13878/j.cnki.dqkxxb. 20211101001.
2. Zhou L T, Yan Z Y, Gu X F, et al. Global Sensitivity Analysis for CNN Based Deformation Prediction of A Cohesive Structure of A Sluice and Pumping Station. Water Resources and Power,2024,42(08):119-122. doi: 10.20040/j.cnki.1000-7709.2024.20231382.
3. Gong A, Zhang H. Reservoir Fluid Identification Model Based on Wavelet Transform and CNN-Transformer. Journal of Xi'an Shiyou University (Natural Science Edition) ,2024,39(04):108-116. doi:10. 3969/j. issn.1673-064X.2024. 04.015..