Institute of Urban Environment,
Chinese Academy of Sciences
1799 Jimei Road, Xiamen, 361021, Fujian, China

中國科學院
CHINESE ACADEMY OF SCIENCES

Mar 16th, 2026

Ref: Cover Letter for Manuscript Revision

Dear Editor:

We thank the reviewers for their insightful comments on our manuscript. All 14 comments from 2 Reviewers are valuable and very helpful to improve our paper. We have revised the manuscript accordingly.

In the revised manuscript, we placed particular focus on clarity and technical depth raised by Reviewer 1. Specifically, we have added figures and textual descriptions to better illustrate the identified gaps. In response to Reviewer 2, we conducted a comprehensive comparative analysis between our dataset and Global Renewables Watch. This includes a detailed evaluation of current global wind turbine mapping methodologies and data, supplemented by new figures and tables that analyze correlations and clarify discrepancies across different datasets.

Our responses to the comments and the corresponding changes to the manuscript are provided below in *blue*, in which the changes in the reply correspond to lines in the clean version of the manuscript. For detailed changes, please also check the revised manuscript with tracked changes.

Apart from those major improvements, we have also carefully proofread this manuscript for several runs. Please see the following detailed responses to your valuable comments, and we look forward to hearing your thoughts.

Yours sincerely,

Dr. Peng Wang

Email: pwang@iue.ac.cn

Point-by-point response to Reviewer 1

#C1.1

Spelling and Grammar Issues
ZZ 45-46
Original: "The codes and dataset … is available."
Correct: "The code and dataset … are available."
ZZ 83
Original: "there is a geospatial wind turbine dataset for 2020 is introduced."
Correct: "A geospatial wind turbine dataset for 2020 was introduced."

Response:

Thank you very much for your comments. We have corrected the spelling and grammar errors as you suggested (lines 51-52, line 79), and conducted a full-text review to address similar issues (line 118, line 438-439).

"The codes and dataset of the global onshore wind turbines are available at …" (lines 51-52)

"At the global scale, a geospatial wind turbine dataset for 2020 is introduced " (line 79)

"…, combining with validation through…" (line 118)

"Compared to the current datasets of available global onshore wind turbines, our dataset provides…" (line 438-439)

#C1.2

Unclear Description: OSM Query
The query string in line 129 (["generator: source"="wind"]) appears syntactically incorrect.
OpenStreetMap commonly uses:
generator: source=wind
or power=generator combined with generator: source=wind
The manuscript should clearly state the exact Overpass query used.

Response:

Thanks very much for pointing this out. We revised the description of the OSM query operation accordingly (lines 128-129).

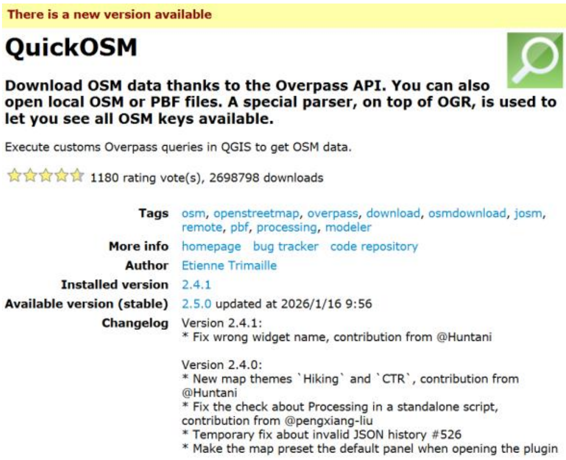
" …through the QuickOSM plugin (based on the Overpass API) in QGIS software with the query parameter: generator: source=wind." (lines 128-129)

#C1.3

Major Comment: Missing Methodological Detail on OSM Extraction
The manuscript does not specify whether nodes, ways, or relations were extracted.
It is unclear whether both power=generator and generator: source=wind were used.

Response:

Thank you very much for your valuable comment. As mentioned in #C1.2, we used generator: source=wind for OSM data extraction. The query is executed using the QuickOSM plugin in QGIS (as shown in the figure below), which leverages the Overpass API. Accordingly, the data were downloaded via the query interface within the QGIS software. We further clarify the data extraction process by adding a description of this plugin in the manuscript (line 129).



" …through the QuickOSM plugin (based on the Overpass API) in QGIS software with the query parameter: generator: source=wind." (line 129)

The handling of offshore turbines is not described.

Response:

Thanks very much for pointing this out. As this study focuses on onshore wind turbines, the initial data extraction was spatially constrained using OSM land polygons. Consequently, offshore wind turbines were excluded from the dataset (lines 132-134).

"Since we focus on onshore wind turbines, OSM land polygon derived from https://osmdata.openstreetmap.de/data/land-polygons.html is used to define the study extent and refine the dataset." (lines 132-134)

No link to the exact Overpass script is provided.

Response:

Thanks very much for your comment. As clarified in #C1.3, the data were queried and downloaded using the QuickOSM plugin within QGIS software, which provides a pre-packaged interface for querying.

There is no information on whether wind farms, meteorological masts, or power-line

infrastructure were filtered out.
These omissions significantly limit reproducibility and must be clarified.

Response:

Thanks very much for raising this issue. Given our focus on the individual wind turbine level, we utilized the Overpass query generator: source=wind filters for nodes representing wind turbines as point features. We further clarify it in the Methods section (lines 128-132).

"…through the QuickOSM plugin (based on the Overpass API) in QGIS software with the query parameter: generator: source=wind. Given our focus on the individual wind turbine level, we utilized this query filter for nodes representing wind turbines in the format of point features." (lines 128-132)

#C1.4

Major Comment: Unexplained OSM Error Rate
The manuscript states a "10% error rate in OSM's global wind turbine dataset" but provides no methodological explanation.
Missing information:
How was this error rate calculated? What was the validation procedure?

Response:

Thanks very much for pointing this out. We apologize for the omission of this explanation. We have added a detailed description of the error rate calculation for the OSM data in the Methods section (lines 257-266). This includes a thorough explanation of the computational methodology and the validation workflow. Besides, we further incorporated an analytical discussion of these error rates based on our updated data in the Results section (lines 287-291).

"Based on our updated wind turbine dataset, we evaluated the data accuracy and errors within the OSM wind turbine records. We calculated omission and commission errors using a spatial proximity analysis in ArcGIS Pro with a 30-meter tolerance buffer. We applied a 30-m buffer to our generated points and performed a spatial selection on the OSM reference points to calculate the omission error. And the OSM points not captured within these buffers were classified as omissions. Conversely, to calculate the commission error, we buffered the OSM points and identified our generated points that fell outside these zones. The respective error rates were derived by dividing the count of omitted or committed points by the total number of OSM turbines. Finally, these two rates were summed to provide a total error rate." (lines 257-266)

"The calculated discrepancy is yielded from the omission and commission error rates of 14.4% and 4.1%, respectively. It is worth noting that this error rate represents global averages, significant regional variations could exist as the OSM data fluctuate across different countries due to varying mapping efforts." (lines 287-291)

Was the error rate spatially or regionally variable?

Response:

Thank you very much for raising this issue. The error metrics reported were calculated based on the total number of wind turbine locations globally. We acknowledge that spatial heterogeneity exists, as the accuracy and completeness of OSM data inevitably vary across regions.

To clarify the specific boundaries and limitations of this error assessment, we have added a detailed explanation in the Results section (lines 289-291).

"It is worth noting that this error rate represents global averages, significant regional variations could exist as the OSM data fluctuate across different countries due to varying mapping efforts." (lines 289-291)

Were commission and omission errors distinguished?

Response:

Thank you very much for your comment. As stated above, we conducted a quantitative assessment of the OSM data by calculating both commission and omission errors to derive a composite error rate. Accordingly, we have expanded the Methods section to include a detailed description of the analytical procedures and the computational workflow used to determine these error metrics (lines 259-266).

"We applied a 30-m buffer to our generated points and performed a spatial selection on the OSM reference points to calculate the omission error. And the OSM points not captured within these buffers were classified as omissions. Conversely, to calculate the commission error, we buffered the OSM points and identified our generated points that fell outside these zones. The respective error rates were derived by dividing the count of omitted or committed points by the total number of OSM turbines." (lines 259-266)

Was the result compared to existing studies with likewise approaches, e.g. https://www.mdpi.com/2220-9964/14/6/232

Response:

Thanks very much for your valuable comment. While we initially provided comparisons with several existing studies and public datasets, we acknowledge they may not have been sufficiently comprehensive.

We have downloaded the data from the provided links and benchmarked our results against this dataset as you suggested. These updates, including expanded discussions and comparative analyses, have been integrated into the Introduction (lines 88-90), the Results (lines 302-304, lines 308-310), and Table 1.

Besides, we further conduct comparisons with current global-scale wind turbine datasets, including Dunnett et al. (2020) and Global Renewables Watch (lines 320-334).

"...there are geospatial datasets for the United States (Rand et al., 2020), Germany

(Manske et al., 2022), Italy (Smeraldo et al., 2020), and South Africa (Kleebauer et al., 2025)." (lines 88-90)

"...along with official and research-based turbine inventories from the United States (Rand et al., 2020), Italy (Smeraldo et al., 2020), Germany (Manske et al., 2022), and South Africa (Kleebauer et al., 2025)." (lines 302-304)

"The consistency between our estimates and official records for temporally comparable years is high, with discrepancies of less than 1.8% in the United States and less than 3.4% in South Africa." (lines 308-310)

Table 1. *Comparison of open-source datasets of onshore wind turbines with our results.*

Scope	***Time***	***Number***	***Ours (2024)***
Dunnett et al. (2020)	*2020*	*33,514*	*416,532*
Global Renewables Watch (Robinson et al., 2025)	*2024(Quarter 2)*	*375,197*	*416,532*
United States (Rand et al., 2020)	*2024*	*75,781*	*74,052*
Germany (Manske et al., 2022)	*2021*	*28,156*	*29,971*
Italy (Smeraldo et al., 2020)	*2020*	*8,729*	*10,591*
South Africa (Kleebauer et al., 2025)	*2025*	*1,487*	*1,483*

"We further benchmark our dataset against the current global-scale wind turbine datasets, including Dunnett et al. (2020) and Global Renewables Watch (Table 2).. Results show that our dataset contains the largest number of identified onshore turbines while maintaining nation-level coverage and land type classification compared to Dunnett et al. (2020). In terms of data records, the Global Renewables Watch is updated to the second quarter of 2024 with 375,197 wind turbines and includes a limited number of offshore wind turbines that are not comprehensively. Our dataset focuses on onshore turbines and incorporates additional updates by the end of 2024. Methodologically, the Global Renewables Watch requires massive training datasets and substantial computational resource budget exceeding 650 V100 GPU hours to process around 14 terapixels of satellite imagery (Robinson et al., 2025). In contrast, our hybrid framework utilizes medium-to-high resolution imagery to enable global-scale updates with significantly lower computational demands. By leveraging publicly available platforms, this framework lowers the barrier to entry through a cost-effective and resource-efficient alternative." (lines 320-334)

#C1.5

Major Comment: Insufficient Documentation of Random Forest Sampling
Missing details include:
From which land cover classes were negative samples drawn?

Response:

Thank you very much for your valuable comment. We have added a detailed description of the land-use categories for the negative samples in the Methods section (lines 185-188).

"The negative samples are obtained via globally uniform random sampling to ensure spatial objectivity. The resulting dataset encompasses diverse land-cover categories including grasslands, bare land, cropland, and forests." (lines 185-188)

Was spatial autocorrelation considered?

Response:

Thanks very much for raising this issue. We employed a global-scale uniform random sampling to mitigate spatial autocorrelation. This ensured sufficient spatial separation and geographic diversity among samples, minimizing spatial dependency and maintaining sample independence. We have added this explanation in the Methods section (lines 178-180).

We further provide a visualization of the positive and negative samples to demonstrate these spatial distribution characteristics (Figure 4).

"We employ a global-scale uniform random sampling strategy. This ensured sufficient spatial separation and geographic diversity among samples, minimizing spatial dependency and maintaining sample independence." (lines 178-180)

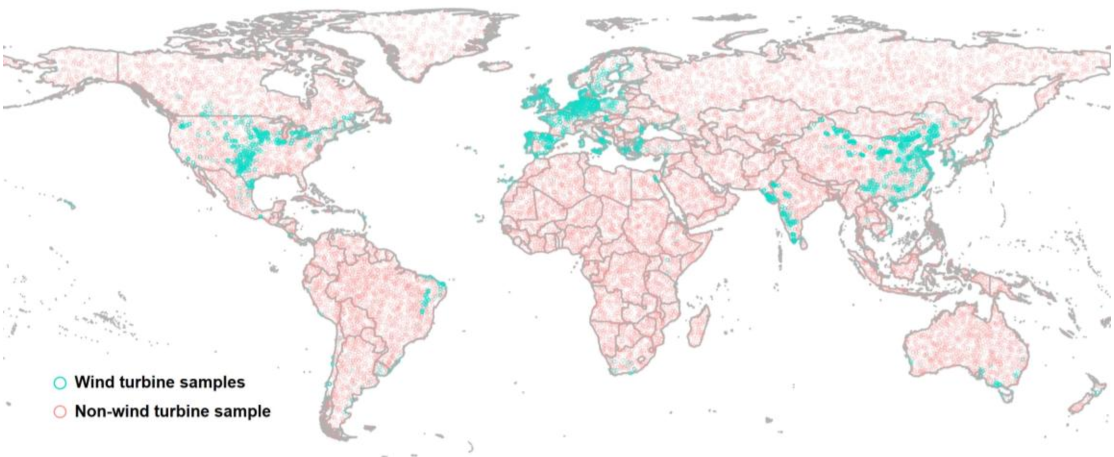


Figure 4. *Spatial distribution of wind turbine (positive) and non-turbine (negative) training samples for machine learning.*

How was it ensured that negative samples were not within 30 m of existing turbines?

Response:

Thanks very much for your comment. To implement spatial constraints, we established exclusion buffers around all existing turbines. During the sampling process, these buffered areas were masked out to ensure that no negative samples were generated within a 30-meter proximity of a positive sample. We have added this detailed description to the Methods section (lines 180-183).

"Besides, we apply a 30-meter buffer around all existing wind turbine locations (positive samples). These buffered areas are then masked out from the global sampling pool to ensure that no negative samples are drawn within this exclusion zone." (lines 180-183)

Was the global distribution of positive and negative samples balanced?
The Random Forest sampling workflow requires a clear methodological description.

Response:

Thank you very much for your comment. We utilized a 1:2 ratio of positive to negative samples for machine learning model training, which were generated through uniform random sampling. To further demonstrate their spatial balance on a global scale, we have visualized the geographic distribution of both sample sets in Figure 4 as shown above.

#C1.6

Minor Comment: Sentinel-1/2 Features and Missing GEE Scripts
The manuscript lists processing steps, but details are missing, included:
Exact satellite collections used.
Time span and temporal compositing strategy.
Cloud masking method (e.g., QA60 for Sentinel-2).
Preprocessing steps such as resampling, mosaicking, and normalisation.
Referencing Zenodo alone is insufficient; the core processing steps must appear in the manuscript.

Response:

Thanks very much for your valuable comment. We added a paragraph in the Methods section detailing the Sentinel-1/2 features and providing the GEE scripts. Specifically, we further added a detailed description of the satellite data collections used (lines 198-203), the cloud-masking method (lines 203-205), the temporal coverage and compositing strategy, and the detailed preprocessing steps (lines 205-209).

"We constructed a comprehensive feature set for machine learning based on Sentinel-1 and Sentinel-2 satellite imagery integrated via the GEE platform. We utilized the Sentinel-1 Ground Range Detected (GRD) dataset (COPERNICUS/S1_GRD), extracting the VV and VH polarization bands. We employed the Sentinel-2 Surface Reflectance collection (COPERNICUS/S2_SR_HARMONIZED), which includes the visible and near-infrared (NIR) bands. To ensure data quality, we applied the QA60 band for cloud masking in Sentinel-2 images. Both datasets were processed using a median reducer across the entire year of 2024 to generate cloud-free, representative composites. All spectral bands and backscatter coefficients were then normalized to a range of [0, 1] to mitigate effects from illumination conditions and sensor characteristics. Finally, these processed layers were stacked into a unified feature collection to serve as input for the machine learning models." (lines 198-209)

Point-by-point response to Reviewer 2

#C2.1

Overall Assessment

This manuscript presents a 2024 global inventory of 379,595 onshore wind turbines derived through a hybrid workflow integrating OpenStreetMap (OSM), high-resolution Google Earth imagery, Sentinel-1/2 data, deep learning (ResNet-18), and Random Forest classification. The topic is timely and relevant, and the resulting dataset could potentially support renewable energy planning, biodiversity assessments, and land-use analysis at multiple scales.

Despite its potential value, the manuscript raises significant concerns regarding benchmarking against existing global datasets, spatial completeness, false positives, and validation rigor. In its current form, the study does not convincingly demonstrate that the proposed dataset represents a substantial advancement over the most up-to-date global wind turbine inventories. These issues should be carefully addressed before the manuscript can be considered for publication.

Response:

Thank you very much for your constructive feedback. We appreciate your insightful comments and recognition of the potential value of our work. These suggestions are valuable in improving the quality of our manuscript. Detailed point-by-point responses to each comment are provided below.

#C2.2

Major Comments

The authors state that their dataset represents a tenfold expansion over the 2020 global wind turbine inventory containing 33,514 turbines. While this numerical comparison is technically correct, it does not constitute an appropriate benchmark. The manuscript does not adequately compare the proposed dataset with the Global Renewables Watch dataset (https://github.com/microsoft/global-renewables-watch?tab=readme-ov-file#dataset-download), which already contains 375,197 wind turbines and is publicly available. Given that this resource is both recent and widely recognized, it should serve as the primary point of reference rather than a 2020 dataset.

Response:

Thanks very much for your valuable comment. We acknowledge that using the 2020 inventory as the sole benchmark was insufficient given the rapid advancements in this field. Accordingly, we have incorporated a comparative analysis of the Global Renewables Watch dataset regarding its methodology and data structure across the Abstract, Introduction, and Results sections (lines 45-47, lines 85-88, lines 306-308, lines 320-334, Table 2, Figure 9).

"This dataset represents a tenfold expansion over global wind turbine inventories as of

2020, and updates 42,955 more onshore wind turbines compared to the Global Renewables Watch based on lower computational requirements." (lines 45-47)

"...Microsoft and Planet's Global Renewables Watch platform employs deep learning for global wind and solar monitoring (Robinson et al., 2025), but it demands massive computing resources for data updates." (lines 85-88)

"The wind turbine counts of GonshoreWT2024 closely align with Global Renewables Watch (375,197 wind turbines), with a 9.9% variance." (lines 306-308)

"We further benchmark our dataset against the current global-scale wind turbine datasets, including Dunnett et al. (2020) and Global Renewables Watch (Table 2).. Results show that our dataset contains the largest number of identified onshore turbines while maintaining nation-level coverage and land type classification compared to Dunnett et al. (2020). In terms of data records, the Global Renewables Watch is updated to the second quarter of 2024 with 375,197 wind turbines and includes a limited number of offshore wind turbines that are not comprehensively. Our dataset focuses on onshore turbines and incorporates additional updates by the end of 2024. Methodologically, the Global Renewables Watch requires massive training datasets and substantial computational resource budget exceeding 650 V100 GPU hours to process around 14 terapixels of satellite imagery (Robinson et al., 2025). In contrast, our hybrid framework utilizes medium-to-high resolution imagery to enable global-scale updates with significantly lower computational demands. By leveraging publicly available platforms, this framework lowers the barrier to entry through a cost-effective and resource-efficient alternative." (lines 320-334)

#C2.3

Furthermore, the Global Renewables Watch dataset includes additional attributes such as construction year, which are absent from the authors' dataset. This omission weakens the claim of producing a more comprehensive inventory. A direct and systematic comparison with Global Renewables Watch is therefore essential. The authors should quantify the spatial overlap between the two datasets, explain methodological differences, and clarify in what specific ways their dataset improves upon existing products in terms of coverage, accuracy, update frequency, or attribute richness. Without this comparison, the assertion of substantial advancement remains insufficiently supported.

Response:

Thanks very much for your insightful comment. We added a section and a table for a thorough comparison with global-scale wind turbine datasets, including Dunnett et al. (2020) and the Global Renewables Watch dataset in the Results section. Specifically, we contrast the methodological differences (lines 320-334) and perform a critical assessment of their comparative advantages, including data coverage and attributes (Table 2).

"We further benchmark our dataset against the current global-scale wind turbine

datasets, including Dunnett et al. (2020) and Global Renewables Watch (Table 2). Results show that our dataset contains the largest number of identified onshore turbines while maintaining nation-level coverage and land type classification compared to Dunnett et al. (2020). In terms of data records, the Global Renewables Watch is updated to the second quarter of 2024 with 375,197 wind turbines and includes a limited number of offshore wind turbines that are not comprehensively. Our dataset focuses on onshore turbines and incorporates additional updates by the end of 2024. Methodologically, the Global Renewables Watch requires massive training datasets and substantial computational resource budget exceeding 650 V100 GPU hours to process around 14 terapixels of satellite imagery (Robinson et al., 2025). In contrast, our hybrid framework utilizes medium-to-high resolution imagery to enable global-scale updates with significantly lower computational demands. By leveraging publicly available platforms, this framework lowers the barrier to entry through a cost-effective and resource-efficient alternative." (lines 320-334)

Table 2. *Comparisons with current global-scale wind turbine datasets.*

Scope	***Technology***	***Time***	***Number***	***Onshore number***	***Land type***	***Nation***	***Construction year***	***Updating algorithm***
Dunnett et al. (2020)	*Onshore and part Offshore*	*2020*	*33,514*	*33,240*	*No*	*Yes*	*No*	*No*
Global Renewables Watch (Robinson et al., 2025)	*Onshore and part Offshore*	*2024 (Quarter 2)*	*375,197*	*373,577*	*Yes*	*Yes*	*Yes*	*Yes*
Ours (GonshoreWT2024)	*Onshore*	*2024*	*416,532*	*416,532*	*Yes*	*Yes*	*No*	*Yes*

#C2.4

I downloaded both datasets. A visual comparison between the two datasets reveals substantial spatial inconsistencies. A considerable number of turbines present in Global Renewables Watch are missing from the authors' dataset. Conversely, the authors' dataset includes many turbines that do not appear in Global Renewables Watch. Among these newly detected turbines, a noticeable proportion appear to be false positives. These discrepancies raise concerns about both omission and commission errors.

The manuscript reports very high classification metrics, but these results appear to be derived primarily from internal validation procedures. Independent cross-dataset validation is necessary for a global infrastructure product of this scale. The authors should quantify omission and commission rates relative to Global Renewables Watch and perform targeted manual verification of turbines that are unique to their dataset. Regional accuracy assessments would also be more informative than reporting only global averages, particularly given the strong geographic heterogeneity in wind turbine siting environments.

Response:

Thank you very much for your valuable comments. We appreciate your comparing our dataset with the Global Renewables Watch dataset and for highlighting the potential discrepancies between the two products.

Differences between the two datasets are expected because they were developed using different data sources, detection strategies, and update timelines. As elaborated in #C2.3, the Global Renewables Watch dataset is currently released up to the second quarter of 2024 and also includes a limited number of offshore turbines. Our dataset focuses on onshore turbines and incorporates additional updates by the end of 2024.

First, we conducted a spatial distribution correlation analysis to compare our dataset with the Global Renewables Watch and Dunnett et al. (2020) (lines 343-360). Specifically, the global study area was partitioned into 2°× 2°grid cells (Fig.1 below). We then quantified the wind turbines within each cell and generated scatter plots to evaluate their consistency (Figure 9). Pearson's r^2 is calculated to quantify the correlation between the datasets.

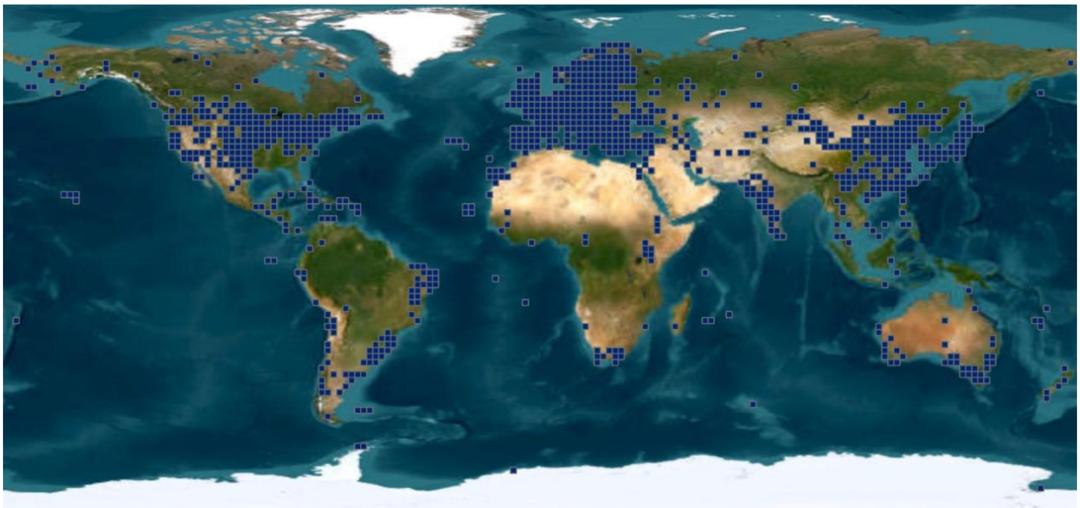


Fig.1 2° × 2° grid cells covering global regions with wind turbines

"We also conducted a spatial distribution analysis to assess the correlation between our dataset and existing benchmarks, including the Dunnett et al. (2020) and Global Renewables Watch. Specifically, the global study area was partitioned into 2° × 2° grid cells to calculate turbine counts. We then employed scatter plots to evaluate the spatial consistency of our dataset relative to Dunnett et al. (2020) and the Global Renewables Watch (Figure 9). Subsequently, Pearson's r^2 is calculated to quantify the correlation between the datasets. Our dataset shows a moderate correlation with Dunnett et al. (2020) (Figure 9a), with a Pearson's r^2 of 0.4, primarily due to the significantly expanded coverage of our dataset. In contrast, our dataset shows a high correlation with Global Renewables Watch with a Pearson r^2 of 0.93, indicating a high degree of geospatial consistency (Figure 9b)." (lines 334-347)

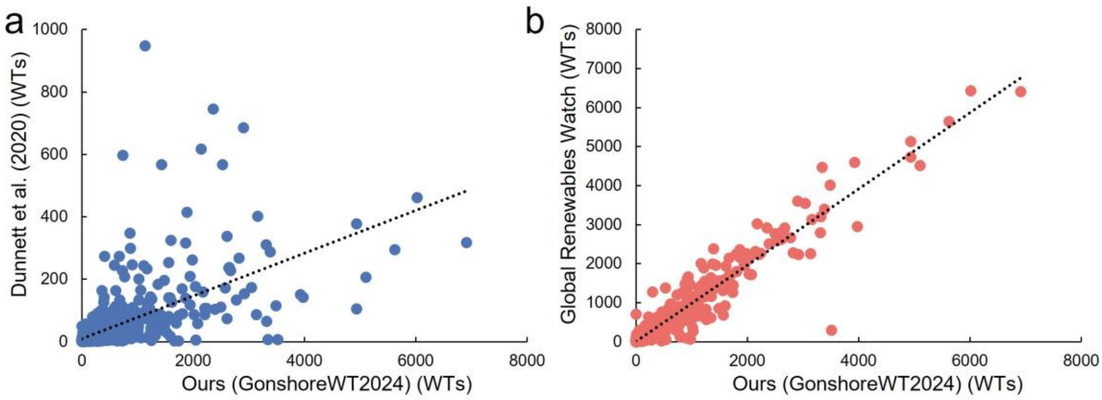


Figure 9. Correlation comparison of wind turbines on global grids (2°× 2°) between our dataset (GonshoreWT) and Dunnett et al. (2020), Global Renewables Watch. ***(a)*** *Distribution of grid-level wind turbine counts between our dataset (GonshoreWT) and Dunnett et al. (2020).* ***(b)*** *Distribution of grid-level wind turbines between our dataset (GonshoreWT) and Global Renewables Watch.*

Then, we further quantified the mutual global underreporting between Global Renewables Watch and ours, which is around 20%. Specifically, Global Renewables Watch has 72,304 more different wind turbines than ours, and we have 80,532 more different wind turbines than theirs. We spent a significant amount of time conducting manual verification to quantify omission and commission rates of our dataset in these wind turbines.

During the manual verification, we found that Global Renewables Watch exhibits omission and commission errors, as shown in Fig. 2 below. Final verification shows a 59% validity rate (43,011/72,304) for unique turbine entries of Global Renewables Watch, compared to a 92% validity rate (74,458/80,532) for ours. We further updated our dataset to a final global count of 416,532 wind turbines based on manual verification. Accordingly, we added the manual verification process and results in the manuscript (lines 348-356).

"Additionally, we quantified the mutual global underreporting between Global Renewables Watch and ours is around 20%. Global Renewables Watch has 72,304 more different wind turbines than ours, and we have 80,532 more different wind turbines than theirs. We conduct manual verification to quantify omission and commission rates of our dataset in these wind turbines. Final verification shows a 59% validity rate (43,011/72,304) for unique turbine entries of Global Renewables Watch, compared to a 92% validity rate (74,458/80,532) for ours. We further updated our dataset to a final global count of 416,532 wind turbines based on the manual verification." (lines 348-356)

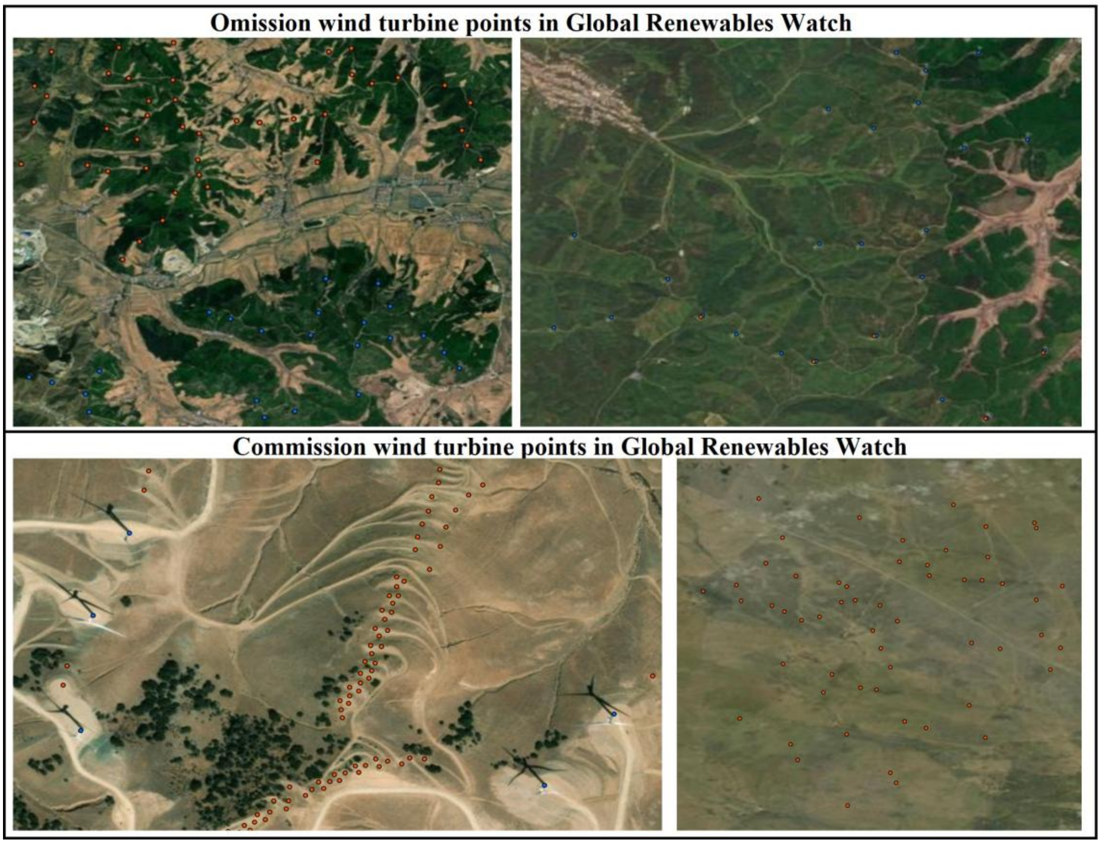


Fig. 2 Example of omission and commission errors in Global Renewables Watch relative to our dataset. Blue points represent our dataset, while orange points denote the Global Renewables Watch dataset.

#C2.5

An additional concern relates to the acquisition and use of Google Earth imagery for training deep learning models. Google's terms of service and licensing agreements generally restrict bulk downloading, scraping, and redistribution of imagery, particularly for automated analysis or machine learning training, unless explicit authorization has been obtained. Even if enforcement actions are unlikely, the legal and ethical responsibility remains with the data producers. The fact that similar practices may have appeared in other published studies does not necessarily imply compliance with platform policies.

Response:

We sincerely appreciate your reminder regarding the legal and ethical responsibilities associated with the use of Google Earth satellite imagery.
We take these concerns very seriously. To further ensure full compliance with platform policies, we have removed all redistributed data involving Google Earth imagery from our dataset link. The updated dataset is at https://doi.org/10.5281/zenodo.18984175. These changes and the corresponding updated data availability statement are detailed in the revised manuscript (lines 419-434).

The imagery was accessed and utilized for non-commercial, academic research purposes. We utilized these images solely to train and validate our deep learning model for wind turbine detection, which was accessed via platforms such as Google Maps. We understand the restrictions outlined in platform policies, and we have ensured that our study does not involve the redistribution or commercial resale of the original Google Earth imagery.

"These open-access data resources could help promote transparent and just sustainable wind energy development, and enable detailed feature extraction and spatial analysis for future wind energy research. The global onshore wind turbine dataset (GonshoreWT2024) is freely available from the Zenodo website at: https://doi.org/10.5281/zenodo.18984175 (Shujun et al., 2025).

The dataset includes:

- *A comprehensive global inventory of 416,532 onshore wind turbines in the format of a geospatial shapefile. The dataset includes geolocation coordinates for all wind turbines, along with corresponding nation (Field: 'Nation') and land use classification (Field: 'landtype') for each wind turbine.*

The code file includes:

- *A PyTorch-based ResNet-18 implementation for classifying onshore wind turbines in Google Earth images, including codes for model architecture and pre-trained weights.*
- *The GEE-based code for the Random Forest model, including sample point splitting (training/test sets) and model training."* (lines 419-434)

#C2.6

More broadly, a truly state-of-the-art global wind turbine dataset should ideally integrate existing authoritative sources rather than operate independently of them. A stronger contribution would involve building upon Global Renewables Watch as a baseline and incorporating newly detected turbines identified by the proposed workflow. Preserving useful attributes such as construction year and land cover would substantially increase scientific value. Including a data source field (e.g., OSM, Global Renewables Watch, or the authors) to indicate provenance and a confidence score or quality flag would further improve transparency and usability. Such integration would significantly enhance the comprehensiveness and reliability of the dataset.

Response:

Thanks very much for your constructive suggestions. As described in #C2.3 and #C2.4, we first compiled a comprehensive table (Table 2) that integrates multiple existing data sources, summarizing their respective methodologies and key attributes, including data origins and characteristics. In addition, we conducted detailed correlation comparisons among the datasets (Figure 9) and further examined and corrected the omission and commission errors in our dataset.

#C2.7

Finally, the manuscript should focus more explicitly on validating newly detected turbines that are absent from existing inventories. These turbines represent the main added value of the proposed method, yet they are also the most likely source of false positives. The authors should isolate these unique detections, conduct stratified manual validation across multiple regions, and report false positive rates specifically for this subset. Clarifying whether classification thresholds were optimized using independent data would also strengthen methodological credibility. Reducing false positives to a minimum is particularly important for infrastructure datasets that may inform ecological impact assessments and policy decisions.

Response:

Thank you very much for your valuable suggestion. As detailed in #C2.4, to further examine potential omission and commission errors, we overlaid the additional wind turbine points detected in our dataset relative to Global Renewables Watch on satellite imagery for manual verification in the Results section (lines 344-352).
The classification metrics in training the models are determined using an independent validation dataset to avoid potential bias from the training data. We have added this clarification in the Methods section (lines 243-245).

"Additionally, we quantified the mutual global underreporting between Global Renewables Watch and ours is around 20%. Global Renewables Watch has 72,304 more different wind turbines than ours, and we have 80,532 more different wind turbines than theirs. We conduct manual verification to quantify omission and commission rates of our dataset in these wind turbines. Final verification shows a 59% validity rate (43,011/72,304) for unique turbine entries of Global Renewables Watch, compared to a 92% validity rate (74,458/80,532) for ours. We further updated our dataset to a final global count of 416,532 wind turbines based on the manual verification." (lines 344-352)

"...as based on an independent test and validation set (Congalton, 1991; Goutte and Gaussier, 2005) to ensure the model's generalizability and avoid over-optimization on training data." (lines 243-245)

#C2.8

Minor Comments
The manuscript contains several grammatical and stylistic issues that require correction. For example, the phrase "current methods remain inadequate monitoring" is awkwardly constructed, "which balancing representativeness" contains a grammatical error, and "Compared current datasets" is an incomplete sentence.

Response:

Thanks very much for your valuable comments. We apologize for the grammatical and stylistic issues in the original version. We have corrected the identified errors and carefully reviewed the entire manuscript to ensure similar issues have been rectified (lines 37-38, lines 148-149, line 405).

"…current methods remain inadequate for monitoring the fast-growing wind turbine deployment." (lines 37-38)

"…which balances representativeness with computational constraints during training." (lines 148-149)

"Compared to the current datasets of …" (line 405)

"The codes and dataset of the global onshore wind turbines are available at …" (lines 45-46)

"At the global scale, a geospatial wind turbine dataset for 2020 is introduced " (line 83)

"…, combining with validation through…." (line 118)