



BREAK THE GRIDLOCK! WIRE FOR GROWTH!



21 - 23 OCTOBER 2025
8:00 am – 5:00 pm



CAPE TOWN, SA
CTICC 2

2025 PRESENTATION

A Wind Turbines Dataset for South Africa: OpenStreetMap Data, Deep Learning Based Geo-Coordinate Correction and Capacity Analysis

Stefan Karamanski

22 October 2025



SPEAKER **OVERVIEW****Stefan Karamanski**

- **B.Eng Electrical and Electronics (Robotics) cum laude**
- **M.Eng Mechanical (Renewable and Sustainable Energy) cum laude**
- **Researcher within the Energy Centre at the Council for Scientific and Industrial Research (CSIR)**
- **Areas of expertise:**
 - Wind energy (onshore and offshore)
 - Solar PV
 - Hydrogen energy
 - Smart grids
 - Battery energy storage systems
 - Spatial downscaling
 - Wind forecasting
 - Resource assessment



Article

A Wind Turbines Dataset for South Africa: OpenStreetMap Data, Deep Learning Based Geo-Coordinate Correction and Capacity Analysis

Maximilian Kleebauer, Stefan Karamanski, Doron Callies and Martin Braun

Special Issue

Advances in AI-Driven Geospatial Analysis and Data Generation (2nd Edition)

Edited by

Dr. Levente Juhász, Prof. Dr. Hartwig H. Hochmair and Dr. Hao Li



<https://doi.org/10.3390/ijgi14060232>

Article

A Wind Turbines Dataset for South Africa: OpenStreetMap Data, Deep Learning Based Geo-Coordinate Correction and Capacity Analysis

Maximilian Kleebauer ^{1,2,*}, Stefan Karamanski ³, Doron Callies ^{1,2} and Martin Braun ^{1,2}

¹ Department of Energy Management and Power System Operation, University of Kassel, 34121 Kassel, Germany; doron.callies@ee.fraunhofer.de (D.C.); martin.braun@ee.fraunhofer.de (M.B.)
² Energy Meteorology and Geo-Information System, Fraunhofer IEE, 34117 Kassel, Germany
³ Energy Supply and Demand Group, Council for Scientific and Industrial Research, Stellenbosch 7600, South Africa; skaramanski@csir.co.za
* Correspondence: maximilian.kleebauer@ee.fraunhofer.de

Abstract: Accurate and detailed spatial data on wind energy infrastructure is essential for renewable energy planning, grid integration, and system analysis. However, publicly available datasets often suffer from limited spatial accuracy, missing attributes, and inconsistent metadata. To address these challenges, this study presents a harmonized and spatially refined dataset of wind turbines in South Africa, combining OpenStreetMap (OSM) data with high-resolution satellite imagery, deep learning-based coordinate correction, and manual curation. The dataset includes 1487 turbines across 42 wind farms, representing over 3.9 GW of installed capacity as of 2025. Of this, more than 3.6 GW is currently operational. The Geo-Coordinates were validated and corrected using a RetinaNet-based object detection model applied to both Google and Bing satellite imagery. Instead of relying solely on spatial precision, the curation process emphasized attribute completeness and consistency. Through systematic verification and cross-referencing with multiple public sources, the final dataset achieves a high level of attribute completeness and internal consistency across all turbines, including turbine type, rated capacity, and commissioning year. The resulting dataset is the most accurate and comprehensive publicly available dataset on wind turbines in South Africa to date. It provides a robust foundation for spatial analysis, energy modeling, and policy assessment related to wind energy development. The dataset is publicly available.

Keywords: wind turbine location; renewable energy; deep learning; geo-coordinate correction; OpenStreetMap

1. Introduction

Wind energy is one of the fastest-growing renewable energy sources in the world. In 2023, wind energy recorded its highest ever growth: in a single year, of new onshore capacity and over 11 GW of offshore wind capacity were added. Total installed capacity worldwide exceeded the symbolic milestone of 100 GW and is expected to reach 2 TW before the end of this decade if current trends continue [1]. In addition, the International Energy Agency (IEA) forecasts that wind energy could meet more than 20% of global electricity demand by 2050, provided that ambitious climate protection measures are implemented to renewable energy sources presents major challenges. Accurate mapping of wind turbine locations and meta-information on the turbine character-

acteristics with class imbalance, such as the detection of sparsely distributed wind turbines in large aerial images. The Fraunhofer IEE has previously demonstrated its capability to significantly improve detection performance for rare objects compared to conventional one-class models in other applications [2]. In addition, RetinaNet achieves competitive results in common object detection benchmarks (e.g., COCO dataset), while offering simpler training requirements compared to two-stage detectors like Faster R-CNN. RetinaNet is an object detection model that combines classification and regression within a unified architecture. It integrates several well-established DL techniques to enable high-precision object localization and classification. A key component is the Residual Network (ResNet) backbone, a variant of Convolutional Neural Networks (CNNs), which utilizes skip connections between layers to facilitate residual learning and improve gradient flow in deep networks [25,26]. To handle multi-scale object detection, a Feature Pyramid Network (FPN) is employed on top of the backbone. The FPN uses a top-down architecture with lateral connections to generate semantically rich feature maps of multiple scales [11]. This allows the network to detect objects at varying sizes effectively. The classification subnetwork is trained using the Focal Loss, which was specifically developed to address the problem of class imbalance between background and foreground objects in dense detection tasks [27]. Unlike the standard cross-entropy loss, Focal Loss introduces a modulation factor to down-weight easy examples and focus training on hard negatives. The Equation (1) for Focal Loss is:

$$L_{FL}(p) = -\alpha(1-p)^{\gamma} \log(p) \quad (1)$$

where p is the model's estimated probability for the true class, α is the weighting factor for class imbalance ($\alpha = 0.25$), and γ is the focusing parameter ($\gamma = 2.0$). This formulation ensures that well-classified examples receive less weight, allowing the model to focus on misclassified or more difficult samples. The regression subnetwork is responsible for predicting bounding boxes around detected objects. From the ResNet-1.1-Loss function, which combines the benefits of L1 and L2 losses and is less sensitive to outliers. This loss was originally introduced in the Fast Region-based Convolutional Network (Fast R-CNN) architecture [12]. The regression loss ℓ_{reg} is introduced in Equation (2), is computed for the predicted bounding box triple $b = (x_1, y_1, x_2, y_2)$ and the ground truth box $b_{gt} = (x_{gt1}, y_{gt1}, x_{gt2}, y_{gt2})$:

$$\ell_{reg}(b, b_{gt}) = \sum_{i=1}^4 \max(0, 1 - |b_i - b_{gt,i}|) \quad (2)$$

The Smooth L1 function itself is defined by the following Equation (3):

$$\text{smooth}_{L1}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq \frac{1}{2} \\ |x| - \frac{1}{4} & \text{otherwise} \end{cases} \quad (3)$$

The resulting parameter γ was set to its commonly used default value of 2.0. This loss formulation enables stable training and effective bounding box regression even in the presence of noisy labels. RetinaNet outputs bounding boxes with predicted aspect ratios of 1:2, 1:3, and 2:1 [28].

Model performance is evaluated using the Average Precision (AP) metric, as defined in Equation (4). Following the Common Objects in Context (COCO) detection framework, a prediction is considered correct if its Intersection over Union (IoU) with the ground truth exceeds 50%. To calculate AP, two key metrics are first needed, Precision and Recall:

accurately represents the tower's exact ground location. As illustrated in Figure 3, several samples are depicted to exemplify their suitability. The training is divided into two parts. First, all 12,000 samples automatically derived from the data preprocessing are used, whereas in the second training, the number of samples is reduced to 7000 highly suitable samples by manual filtering. All other parameters remained the same for both the first and second training: 100 epochs, 100 steps, 80% training and 10% independent validation, and 10% test dataset.



Figure 2. The method of the static cutting of the training images is shown. The black lines represent the cutting edges, the red dots the coordinates of the wind turbines.

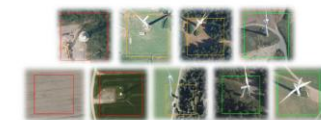


Figure 3. Samples based on their suitability for training. Due to incorrect position or poor image resolution, the turbines that are clearly visible but were rejected for fine-tuning are marked in green. The images marked in green contain turbines whose tower is not visible.

3.2. Deep Learning Approach

Several object detection frameworks were compared, including Faster R-CNN, YOLOv3, and RetinaNet, to find the best balance between high detection accuracy and computational efficiency.

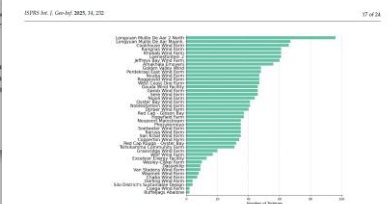
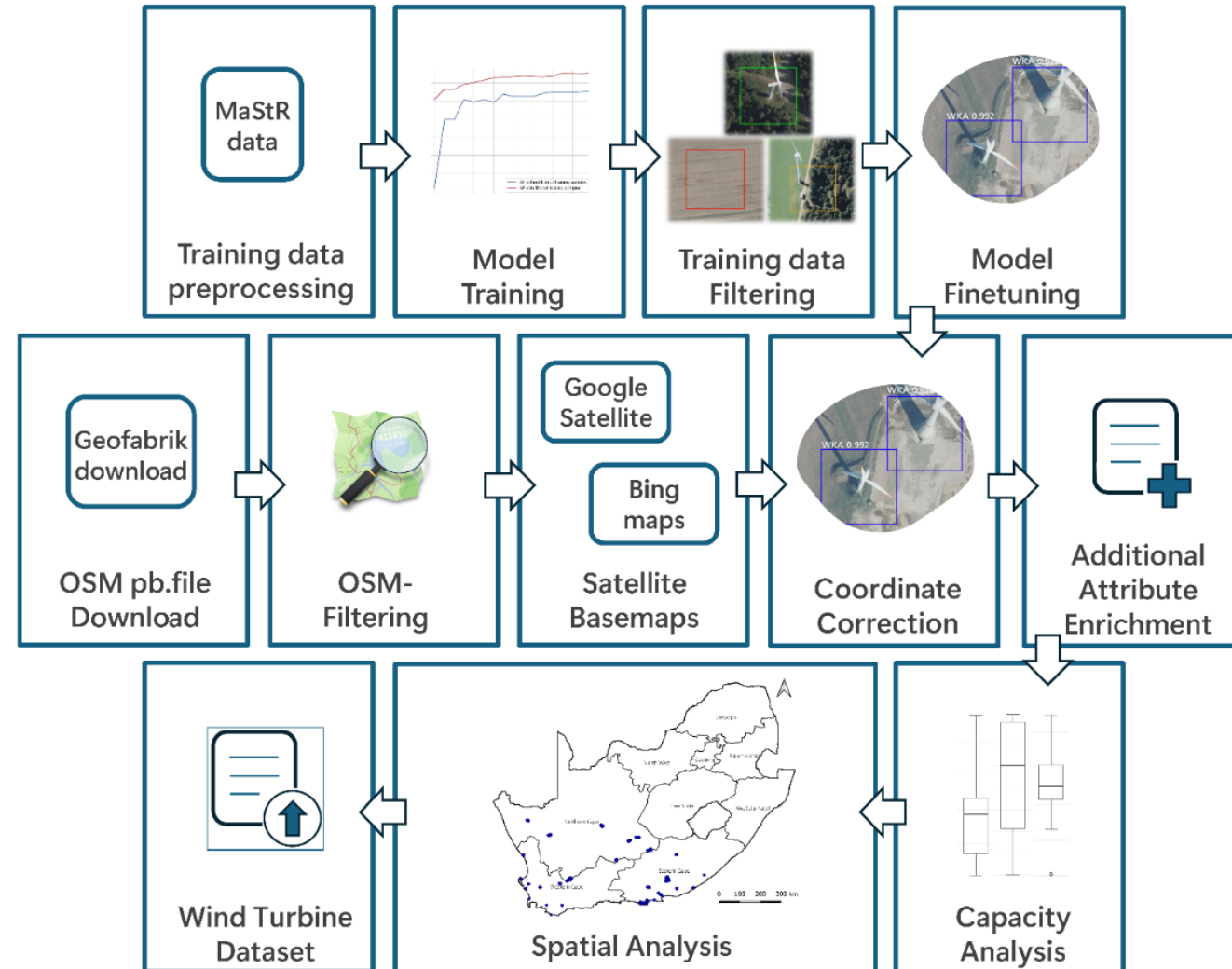
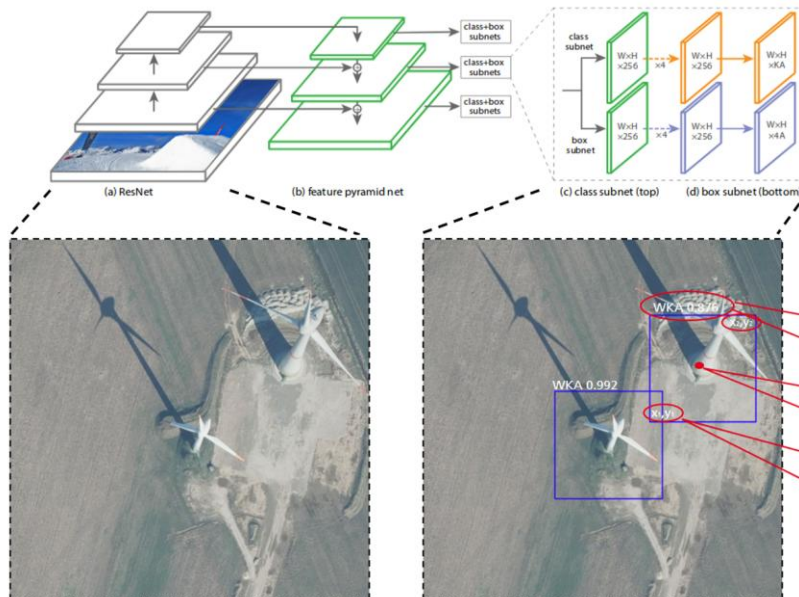


Figure 14. This figure shows the number of wind turbines per wind farm. Wind farms with more than 10 turbines are shown at the top, while smaller farms with fewer turbines are listed further down.

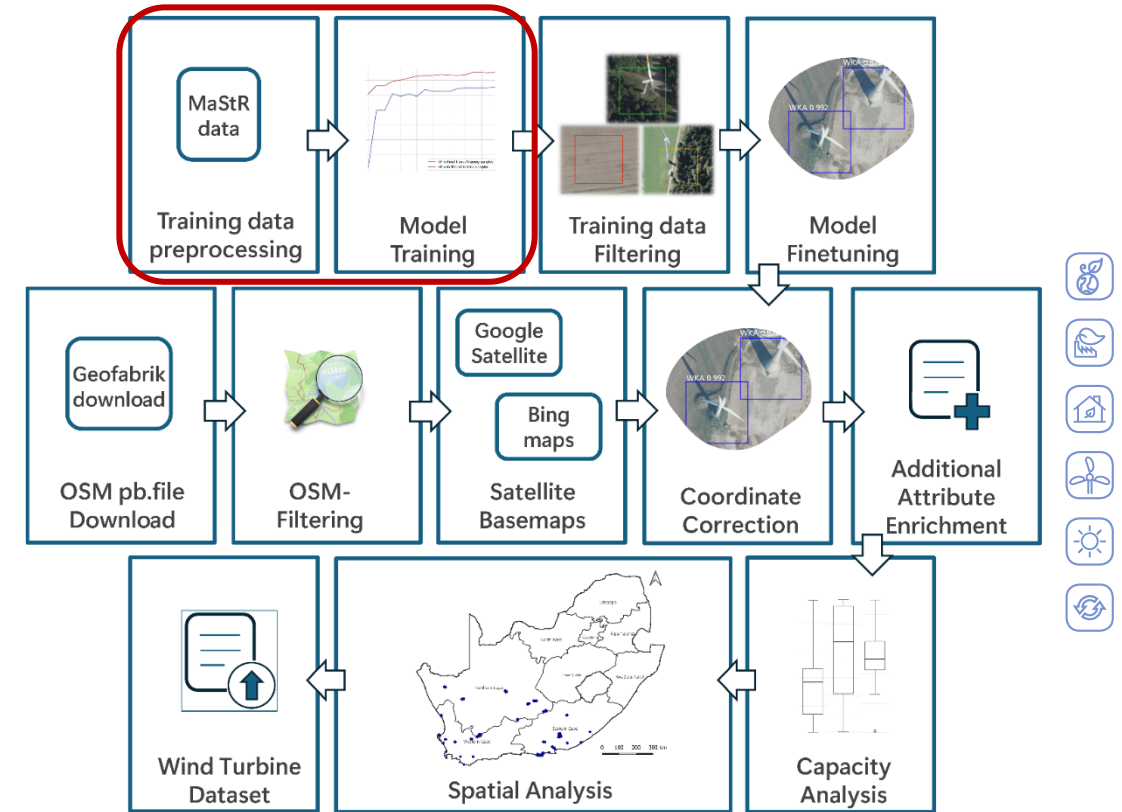
Figure 15. The capacity per wind turbine in megawatts (MW) for the wind farms in South Africa is shown. The values are sorted in ascending order so that wind farms with a lower capacity per turbine are shown at the bottom and wind farms with a higher capacity at the top.

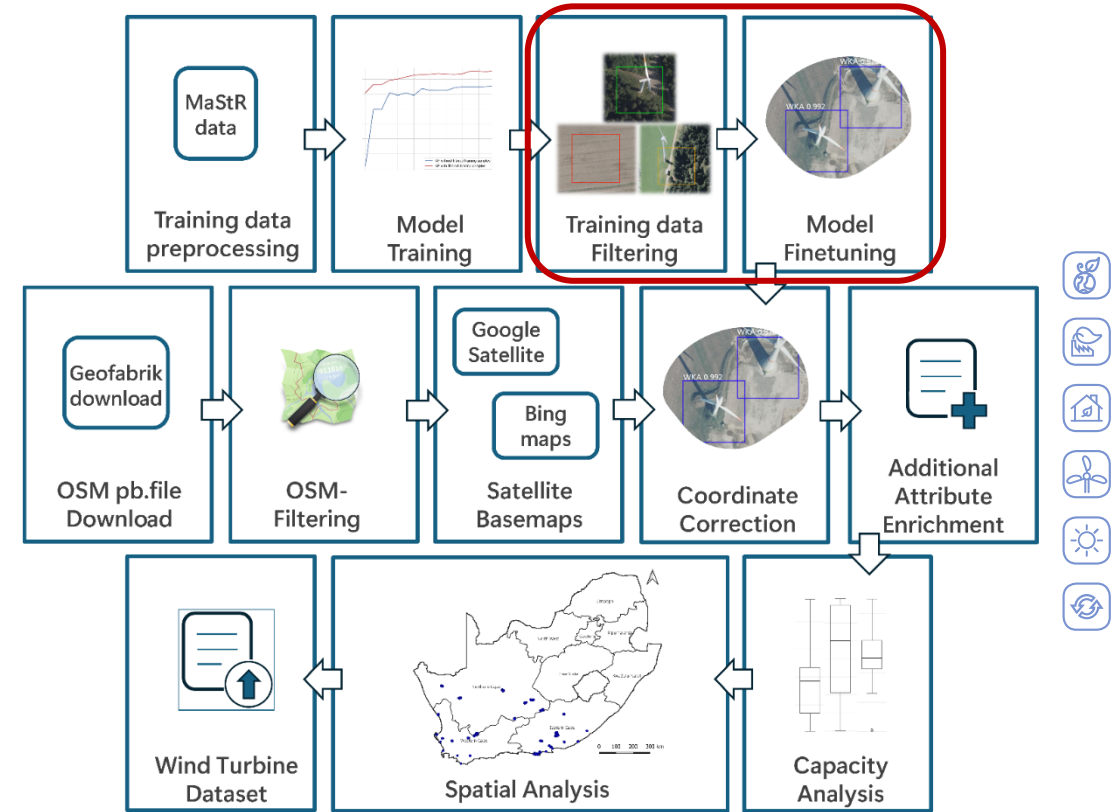
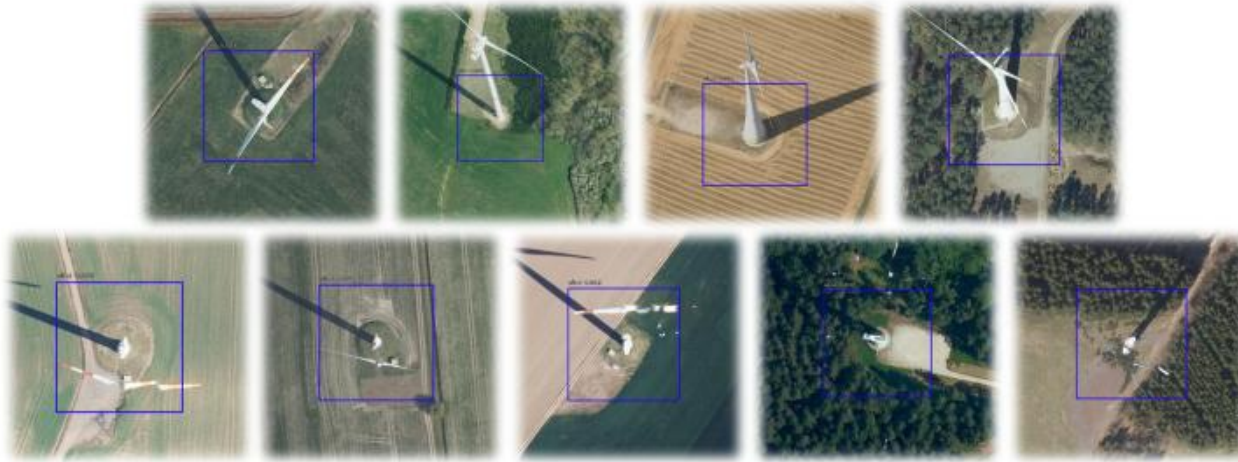
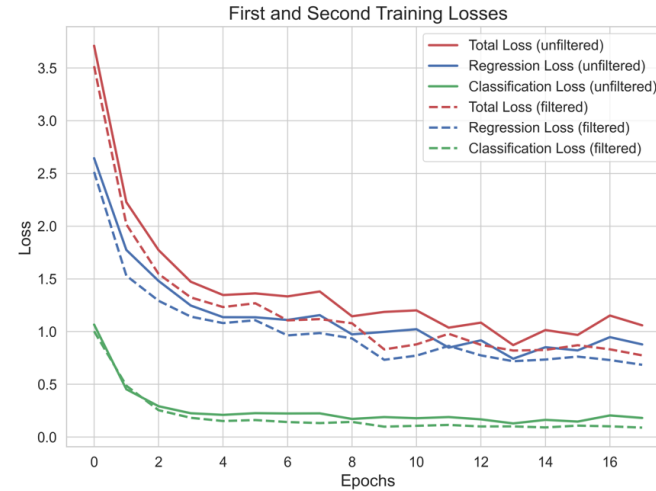
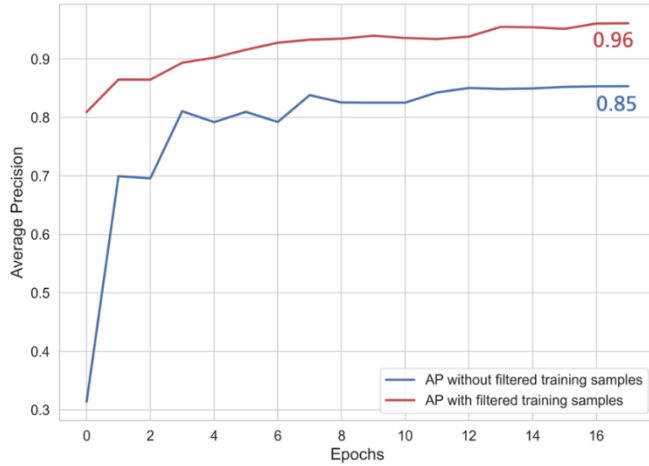
The following section of the results focuses on the spatial distribution of wind turbines in South Africa. The installed wind power capacity is concentrated in just three of the country's nine provinces: Northern Cape, Eastern Cape, and Western Cape. Table 1 provides a summary of wind energy infrastructure at the provincial level.





Objekttabelle
Bildname
Vorhersagegenauigkeit
Klassenname
Easting
Northing
X ₁
Y ₁
X ₂
Y ₂







Download OpenStreetMap data for this region:

South Africa

[one level up]

The OpenStreetMap data files provided on this server do **not** contain the user names, user IDs and changeset IDs of the OSM objects because these fields are assumed to contain personal information about the OpenStreetMap contributors and are therefore subject to data protection regulations in the European Union.

[Extracts with full metadata](#) are available to OpenStreetMap contributors only.

Commonly Used Formats

- [south-africa-latest.osm.pbf](#), suitable for Osmium, Osmosis, imposm, osm2pgsql, mkgmap, and others. This file was last modified 1 day ago and contains all OSM data up to 2025-05-13T20:20:35Z. File size: 347 MB; MD5 sum: [7e63e92a842c8bd4f8ee3c8f49bb7f34](#).
- [south-africa-latest-free.shp.zip](#), yields a number of ESRI compatible shape files when unzipped. ([Format description PDF](#)) This file was last modified 1 day ago. File size: 851 MB; MD5 sum: [f9aaa56785fd9cddde1dcb9b9bf86d90a](#).

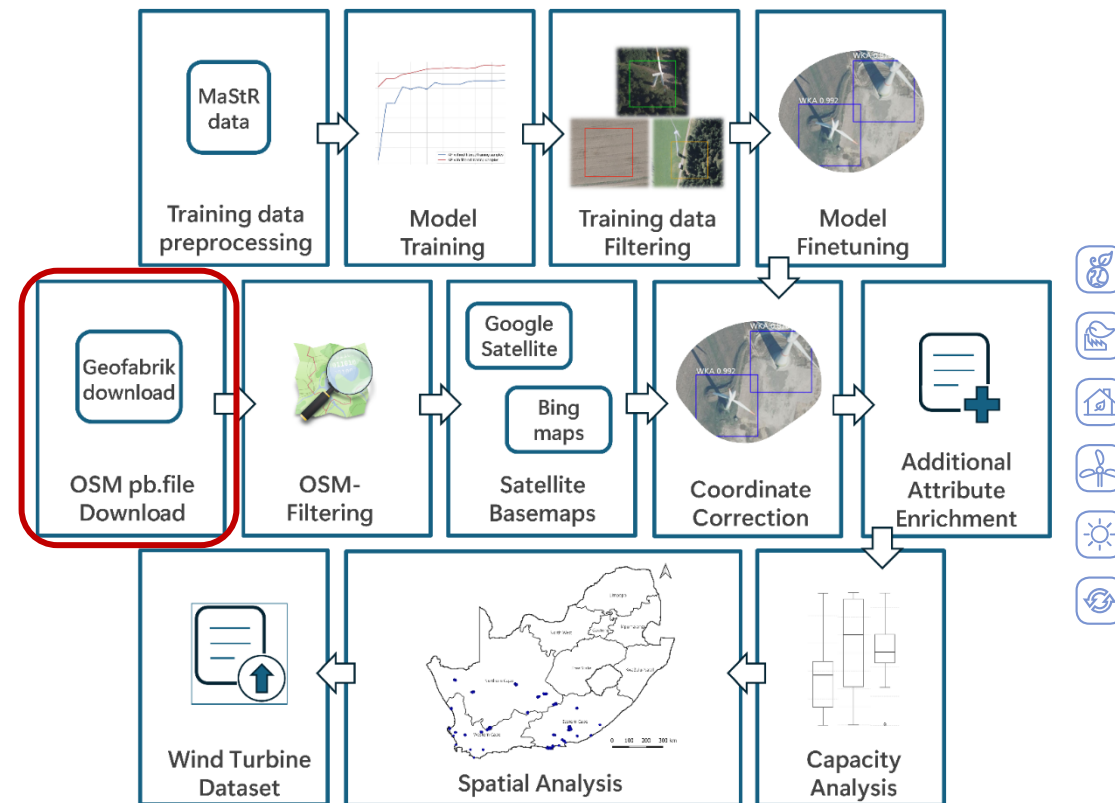
Other Formats and Auxiliary Files

- [south-africa-latest.osm.bz2](#), yields OSM XML when decompressed; use for programs that cannot process the .pbf format. [Deprecated - not updated any more](#). MD5 sum:

GEOFABRIK⁺downloads



🇬🇧 Not what you were looking for? Geofabrik is a consulting and software development firm based in Karlsruhe, Germany specializing in OpenStreetMap services. We're happy to help you with data preparation, processing, server setup and the like. [Check out our web site](#) and contact us if we can be of service.





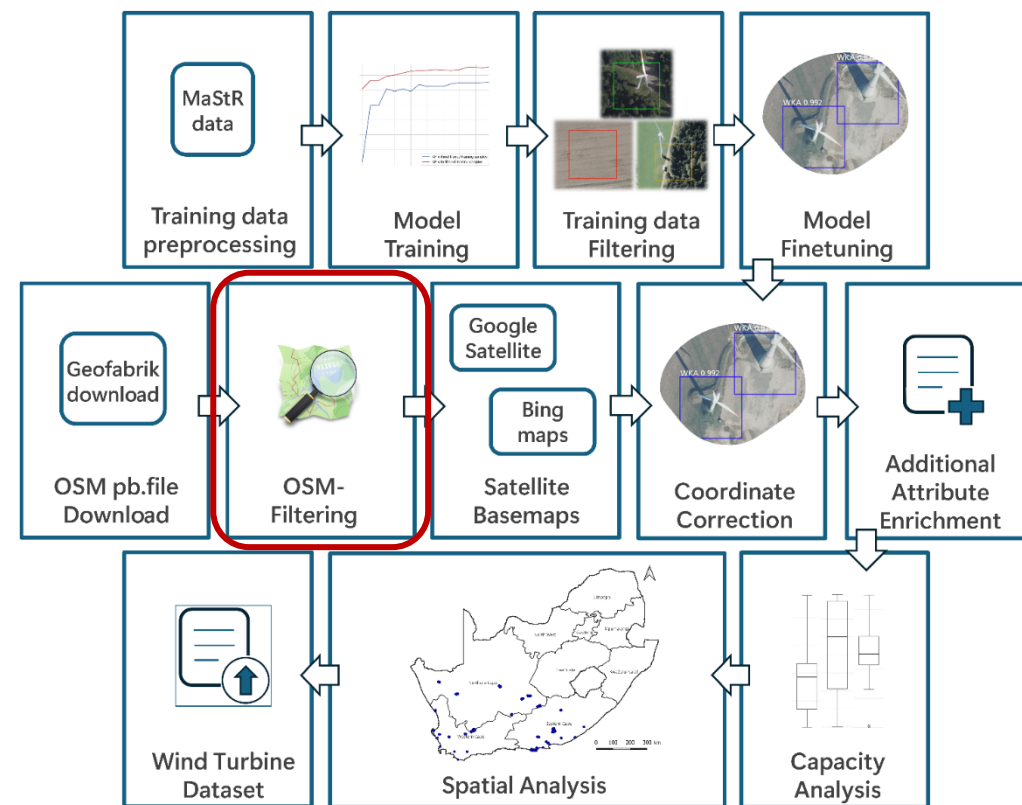
Software Metapapers

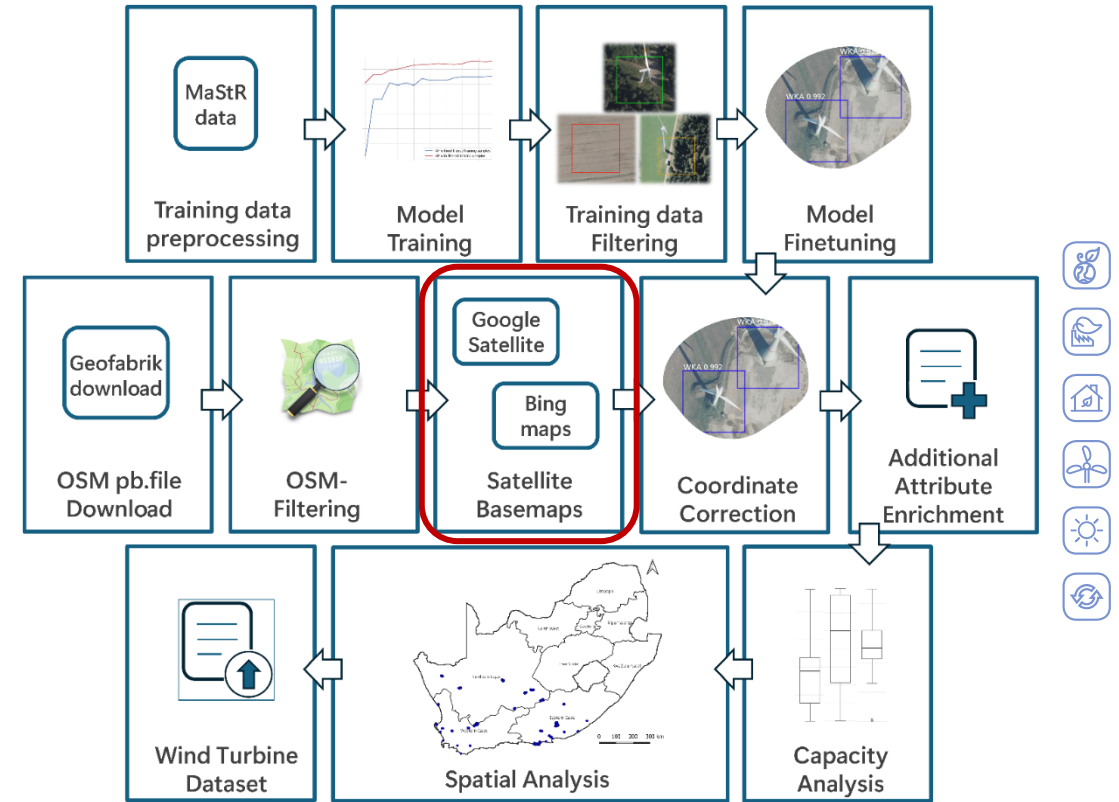
esy-osmfilter – A Python Library to Efficiently Extract OpenStreetMap Data

Adam Pluta , Ontje Lünsdorf

- **Prefilter:** The 'prefilter' is used to identify nodes, ways, and relations tagged with attributes like "power": ["generator", "plant", "solar", "photovoltaic"] to capture all relevant renewable energy installations.
- **Blackfilter:** A 'blackfilter' is applied to exclude certain types of infrastructure that are not of interest, such as those associated with fossil fuels or hydro-based generation. Examples include ("generator:source", "gas"), ("generator:method", "combustion"), and ("generator:source", "coal").
- **Whitefilter:** A 'whitefilter' is also used to ensure that elements explicitly tagged with ("power", "generator") are retained in the dataset.

This process provides a refined dataset that filters out non-relevant elements and focuses on renewable energy facilities, improving the quality and relevance of the geospatial analysis.





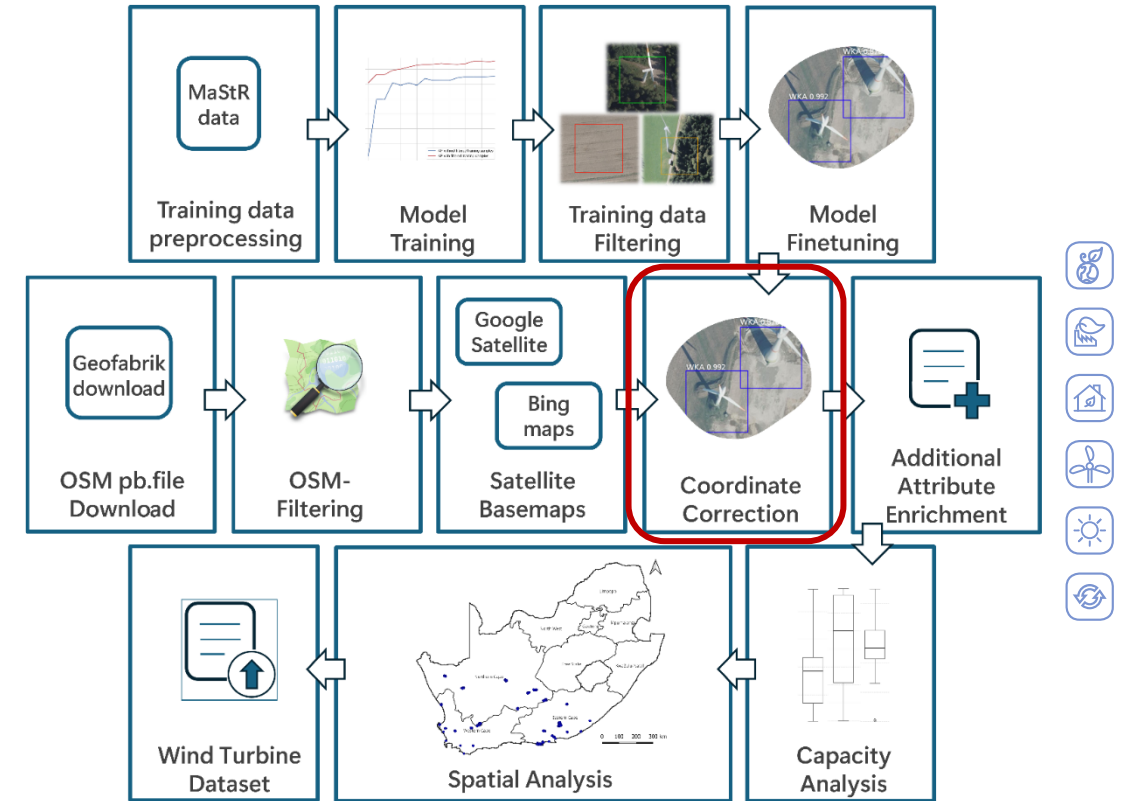
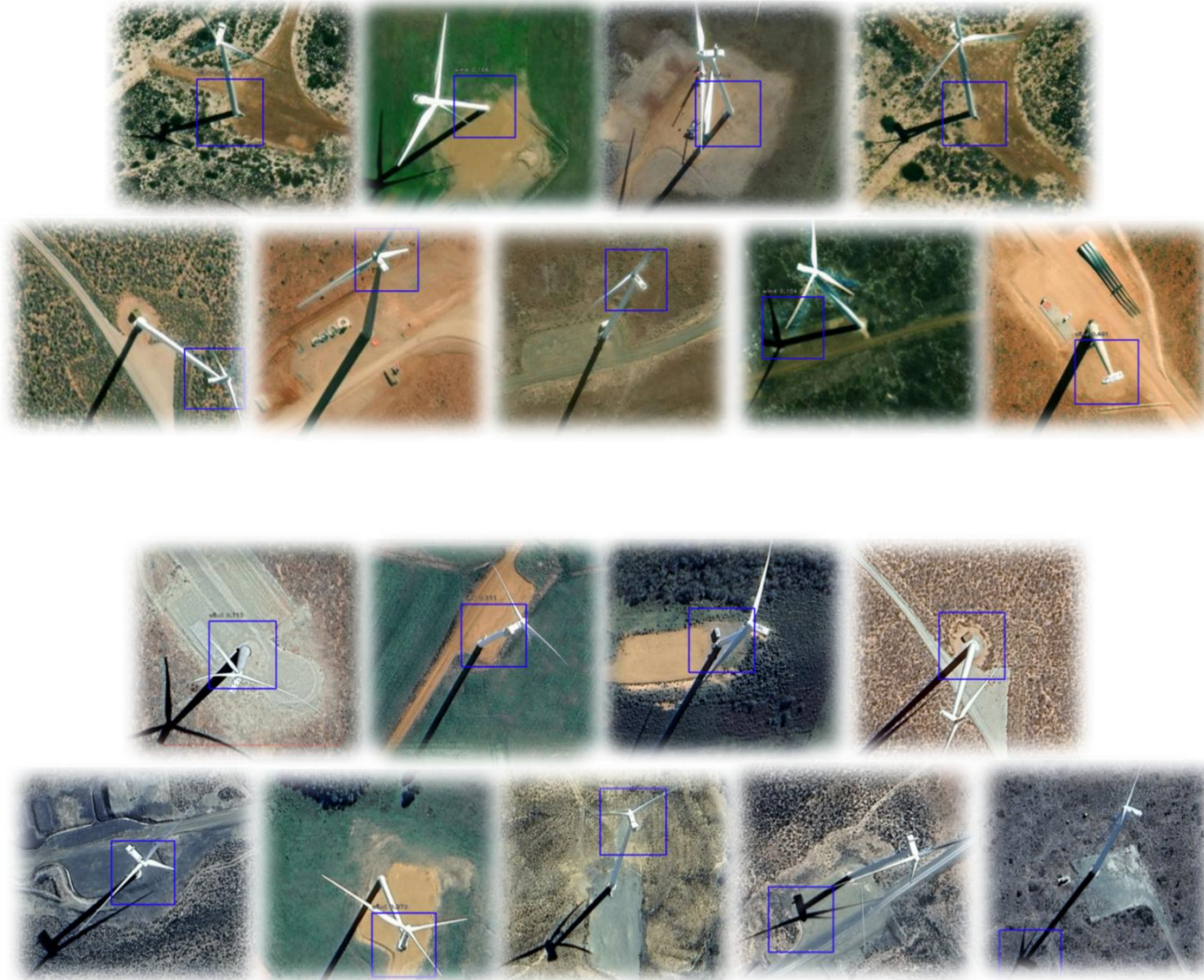
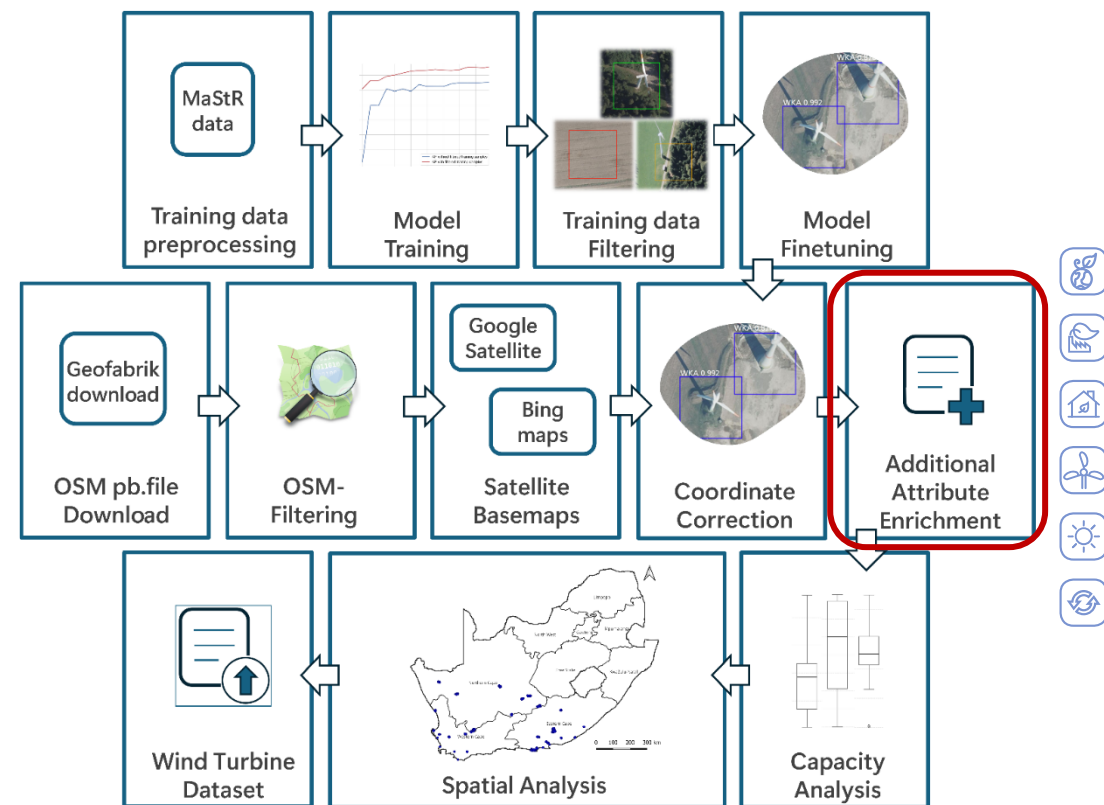




Table 3. Summary of Wind Turbines in South Africa, including the commissioning year, number of turbines, total capacity, capacity per turbine, and turbine type for each wind farm.

Name of Farm	Comm. Year	Turbines	Tot. Cap. (MW)	Cap./Turbine (MW)	Turbine Type
Amakhala Emoyeni	2016	56	134.4	2.4	Nordex N117/2400
Buffeljags Abalone	2012	2	0.13	0.065	Horizontal Axis Turbine
Chaba Wind Farm	2015	7	21.5	3.075	Vestas V112-3.075
Coega Wind Farm	2010	2	3.6	1.8	General Electric GE2.5XL
Cookhouse Wind Farm	2014	66	138.6	2.1	Suzlon S88/2100
Copperton Wind Farm	2021	34	102	3.15	Acciona AW-3150/125
Darling Wind Farm	2008	4	5.2	1.3	Fuhrländer FL 1250/62
Dassieklip	2015	9	27	3	Sinovel SL 3000/90
Dorper Wind Farm	2014	40	100	2.5	Nordex N100/2500
Excelsior Energy Facility	2020	13	32.5	2.5	Goldwind GW121/2500
Garob Wind Farm	2021	46	145	3.15	Nordex AW125/3150
Golden Valley Wind	2020	48	120	2.5	Goldwind GW121/2500
Gouda Wind Facility	2015	46	138	3	Acciona AW-3000/100
Grassridge Wind Farm	2016	20	60	3	Vestas V112/3000
Hopefield Farm	2014	37	66.6	1.8	Vestas V100-1.8
Jeffreys Bay Wind Farm	2014	60	138	2.3	Siemens SWT-2.3-101
Kangnas Wind Farm	2020	61	140	2.3	Siemens SWT-2.3-108
Karusa Wind Farm	2021	35	147	4.2	Vestas V136-4.2
Khobab Wind Farm	2017	61	140	2.3	Siemens SWT-2.3-108
Loeriesfontein 2	2017	61	140	2.3	Siemens SWT-2.3-108
Longyuan Mulilo De Aar 2 North	2017	96	144	1.5	Guodian UP86/1500
Longyuan Mulilo De Aar Maanh.	2016	67	100	1.5	Guodian UP86/1500
Noblesfontein Wind Farm	2014	41	73.8	1.8	Vestas V100-1.8
Nojoli Wind Farm	2016	44	88	2	Vestas V100-2.0
Noupoort Mainstream	2016	35	80.5	2.3	Siemens SWT-2.3-108
Nxuba Wind Farm	2020	47	140	3	Nordex AW 125/3150
Oyster Bay Wind Farm	2021	41	140	3.45	Vestas V117-3.45
Perdekraal East Wind Farm	2020	48	110	2.3	Siemens SWT-2.3-108
Phezkumoya	2025*	35**	140	4	Vestas V136-4.0



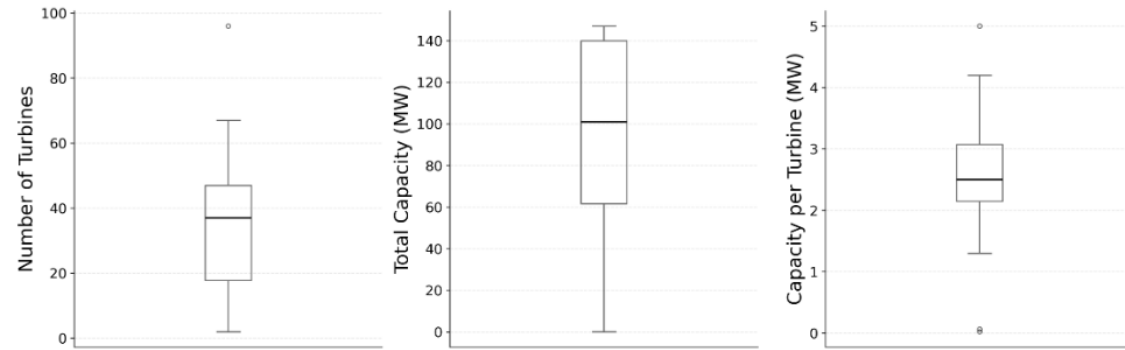


Figure 11. Summary statistics of key parameters of South African wind farms. Number of turbines per wind farm (left side), total installed capacity (MW) per wind farm (in the middle), and capacity per turbine (MW) (on the left). The boxplots contains the median, interquartile range, and outliers in the dataset

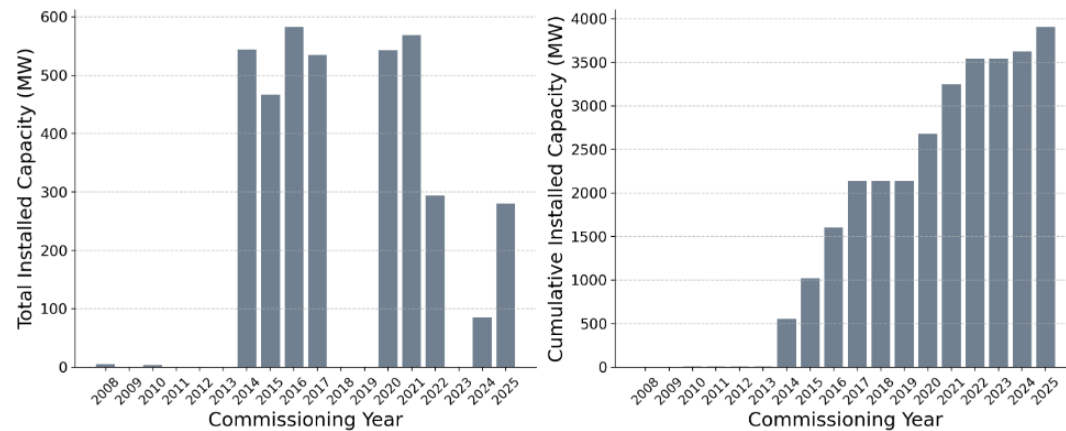
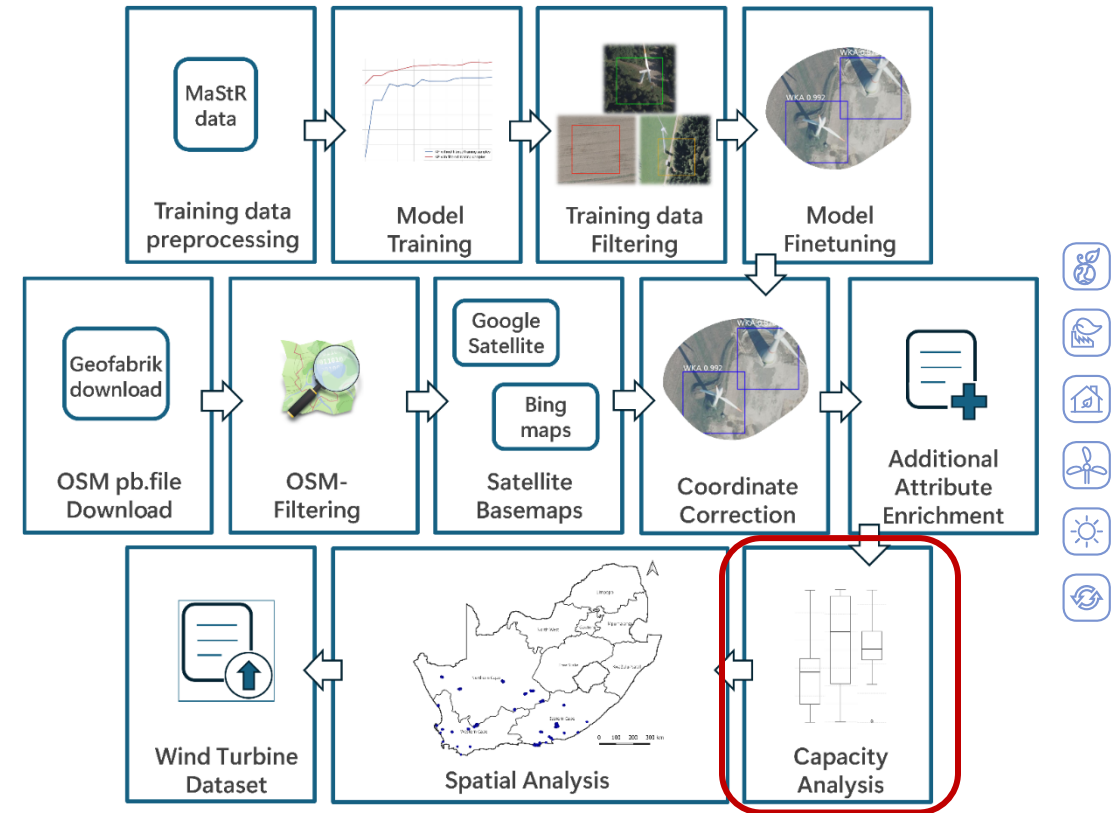


Figure 12. Development of wind power capacity in South Africa by year. The annual installed wind power capacity from 2008 to 2025 is shown on the left-hand side, and the cumulative installed capacity on the right-hand side.



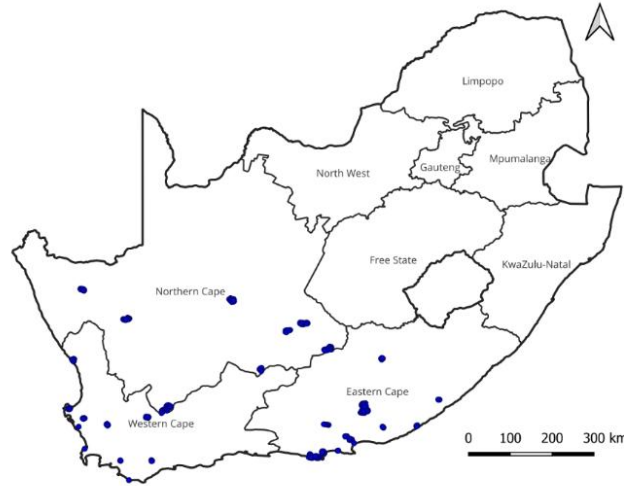


Figure 16. Spatial distribution of all existing wind turbines in South Africa, marked in blue, highlighting their locations across the country.

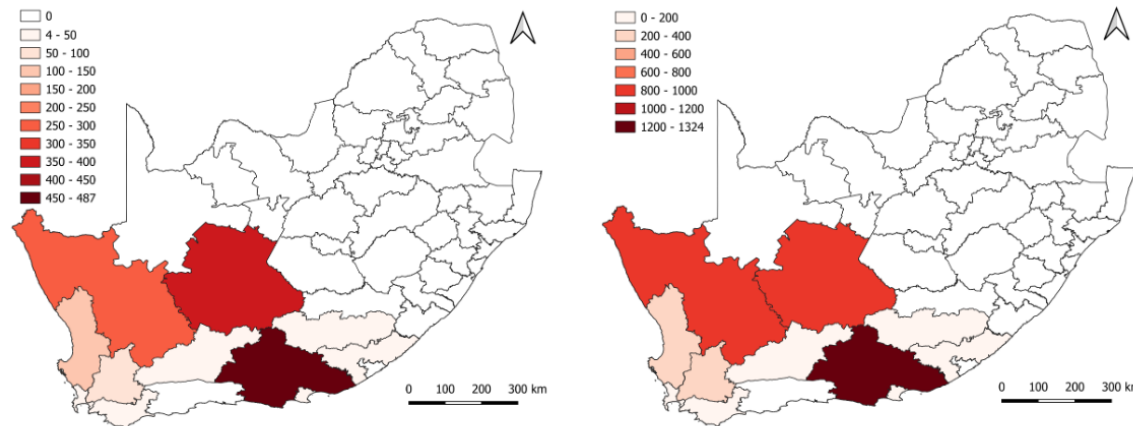


Figure 17. Spatial distribution of wind energy infrastructure by municipality. Map (on the left side) displays the number of wind turbines, while map (on the right side) shows the total installed wind capacity (MW). The patterns reveal significant regional clustering, with a small number of municipalities concentrating the majority of infrastructure.

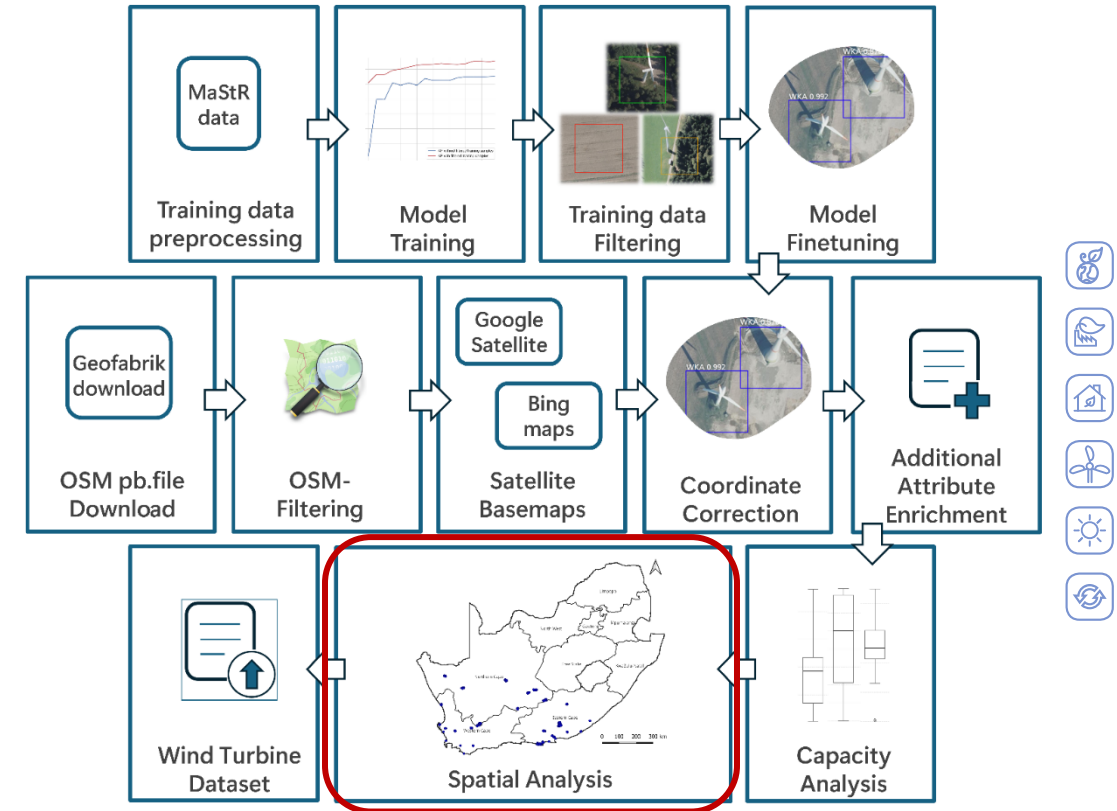
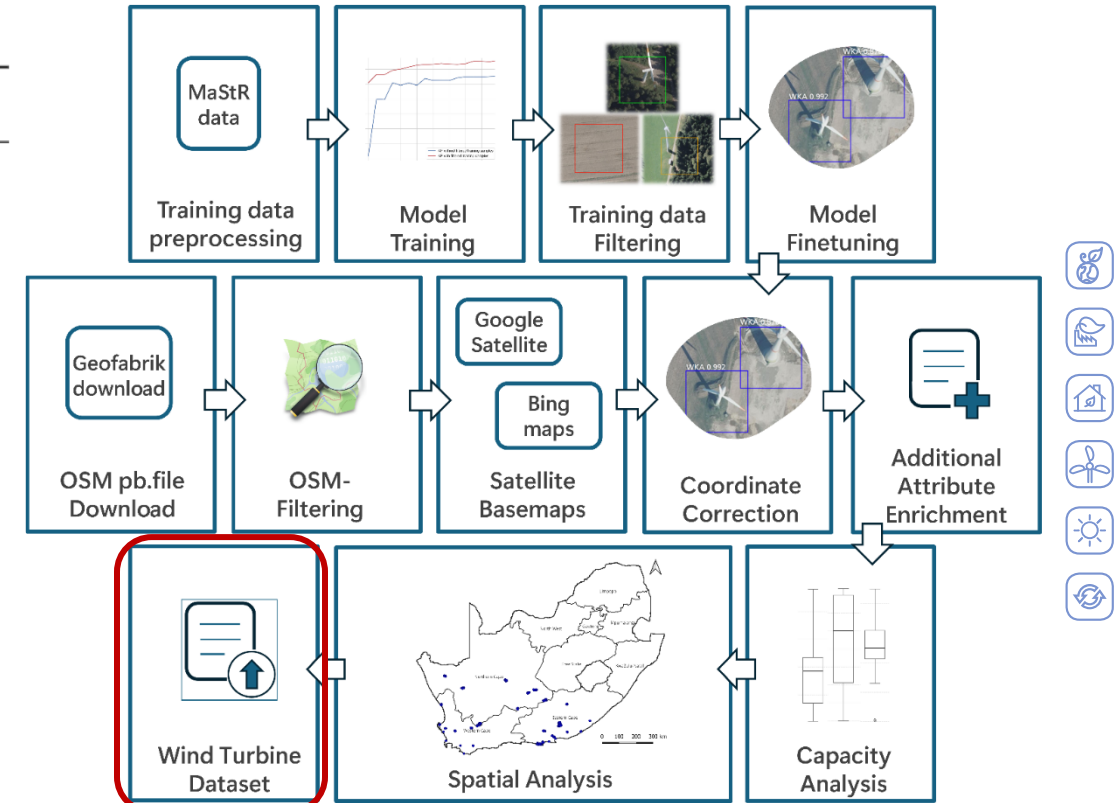
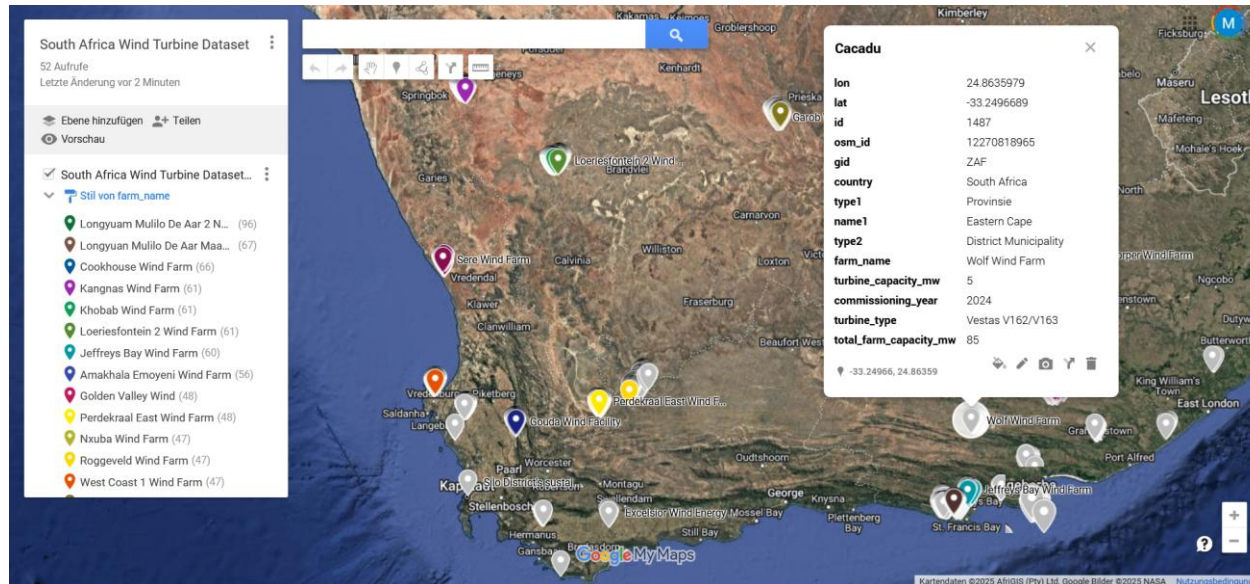




Table 3. Summary of Wind Turbines in South Africa, including the commissioning year, number of turbines, total capacity, capacity per turbine, and turbine type for each wind farm.

Name of Farm	Comm. Year	Turbines	Tot. Cap. (MW)	Cap./Turbine (MW)	Turbine Type
Amakhala Emoyeni	2016	56	134.4	2.4	Nordex N117/2400
Buffeljags Abalone	2012	2	0.13	0.065	Horizontal Axis Turbine
Chaba Wind Farm	2015	7	21.5	3.075	Vestas V112-3.075
Coega Wind Farm	2010	2	3.6	1.8	General Electric GE2.5XL
Cookhouse Wind Farm	2014	66	138.6	2.1	Suzlon S88/2100
Copperton Wind Farm	2021	34	102	3.15	Acciona AW-3150/125
Darling Wind Farm	2008	4	5.2	1.3	Fuhrländer FL 1250/62
Dassieklip	2015	9	27	3	Sinovel SL 3000/90



Online Dataset:



Follow **SAWEA** on social media



BREAK THE GRIDLOCK! WIRE FOR GROWTH!



21 - 23 OCTOBER 2025
8:00 am – 5:00 pm



CAPE TOWN, SA
CTICC 2

2025 PRESENTATION

THANK YOU FOR LISTENING!

CONTACT DETAILS

Stefan Karamanski

Council for Scientific and Industrial Research (CSIR)

Researcher

skaramanski@csir.co.za

Publication:

