Mining Satellite Imagery for Offshore Wind Energy

Cristian C. Noriega Monsalve, Aparna S. Varde
School of Computing; Clean Energy and Sustainability Analytics Center; Montclair State University, NJ, USA
(noriegamonsc1 | vardea)@montclair.edu, ORCID ID: 0000-0002-3170-2510 (Varde)

***Abstract*— This work mines big data in Sentinel-1 satellite images to unveil geographical patterns in offshore wind energy. We leverage unsupervised machine learning to extract insights from a 44GB open access dataset for decision support in wind farm orientations to guide stakeholders. It has broader impacts of overcoming climate change by enhancing renewable energy.**

Keywords*— *Climate Change, Clustering, Geospatial Big Data, Image Mining, Ocean Wind Field, Radar, Renewable Energy, Satellite Data, Unsupervised Learning, World Geodetic System

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825812

I. Introduction

The earth system operates daily, monthly, and annually, with the sun as its main energy source. Interactions between solar radiation and the atmosphere drive weather patterns; uneven heating of Earth's surface creates temperature and pressure variations among air masses, affecting coasts and motivating studies [1]. In offshore wind energy, identifying wind patterns along New Jersey to Massachusetts shorelines can help energy generation via optimal wind turbines localization and orientation. To the best of our knowledge, our study is unique in harvesting large datasets from satellite images and mining them to unveil wind patterns geographically for long-term decision support in offshore wind farm and energy estimation.

II. Methods and models

A. Remote Sensing

The Sentinel-1 mission comprises a constellation of two satellites, launched by ESA (European Space Agency). Sentinel-1A was launched on April 3rd, 2014, and Sentinel-1B was launched on April 25th, 2016; the latter ceased operations on August 23rd, 2022. Sentinel-1 is in a near-polar orbit with a 12 day repeat cycle. Both Sentinel-1A and Sentinel-1B share the same orbit plane with a 180° orbital phasing difference; with both satellites, the repeat cycle is 6 days. This mission use Synthetic Aperture Radar (SAR) active sensors, emitting C-band pulses toward earth's surface and capturing backscatters free from atmospheric interactions, despite time and weather conditions. With a high spatial resolution of 1km for the Sentinel-1 Ocean Wind Field (OWI) Level-2 product effectively monitors ocean surfaces and offshore wind assessment [2]. An example of a Sentinel-1 Level-2 dataset appears in Fig 1.

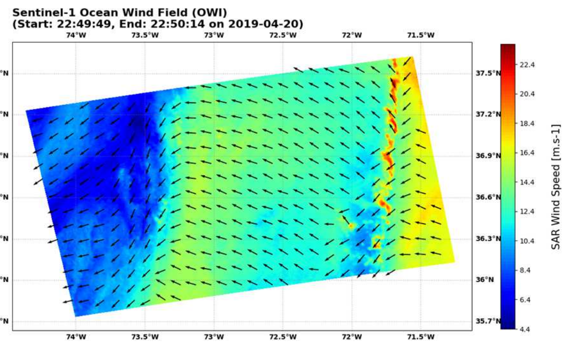

Fig. 1 Dataset example from Sentinel-1 Level-2

B. Data Harvesting

Sentinel-1 Level-2 products are open-access in ASF (Alaska Satellite Facility). The products include georeferenced data on wind speed and direction, based on the World Geodesic System 1984 (WGS84). We download 5,773 products from the start of the mission to September 2024, with each product containing over 20,000 data points. We harvest the data with PCA (Principal Components Analysis), synopsized in Fig. 2.

C. Geospatial Data Transformation for 4D Analysis

The spatial-temporal-direction is a four-dimensional (4D) construct. The geographic velocity vector is split into a horizontal movement along the east axis and a vertical movement along the north axis, converting the radial direction of the wind into a Cartesian representation. Meanwhile, the WGS84 is projected as a Coordinate Reference System (CRS) using the Lambert Conformal Conic (LCC) projection. The southernmost (34.4566 °N) and northernmost (43.2517 °N) points define the minimum and maximum latitudes, respectively. The LCC projection preserves the angles and shapes, making it highly suitable for representing spatial relationships accurately. Hence, this feature extraction ensures compatibility with Euclidean distance metrics.

D. Unsupervised Machine Learning

Unsupervised learning identifies similarities and differences by clustering unlabeled datasets [3]. Leveraging the k-means algorithm, which utilizes Euclidean distance to measure proximity between these 4D points, facilitates appropriate grouping. We employ the elbow method to determine the optimal number of clusters k. This technique evaluates the sum of squared distances (inertia) between points and their corresponding cluster centroids for varying values of k. As the value of k increases, the inertia decreases, indicating improved compactness within clusters. However, beyond a certain point, the rate of improvement diminishes, forming an "elbow" in the curve. For this dataset, the elbow occurs at $k=5$, suggesting that 5 clusters provide the best balance between minimizing inertia and avoiding overfitting. The clustering can be well-visualized in Fig.5 with 5 colors corresponding to 5 clusters. This maps to the value of $k=5$ learned by the elbow method.

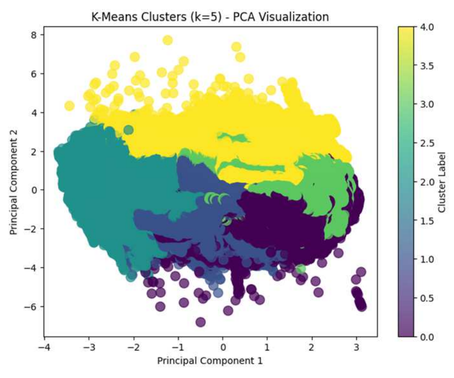

Fig. 2. Harvesting Sentinel-1 data with PCA

III. Related work

Remote sensing with reanalysis dataset from the European Centre for Medium-Range Weather Forecasts can enable wind speed and energy density estimations over the Mediterranean Sea [4]. The analysis presents Sentinel high-resolution spatial coverage and capability to monitor temporal wind variations crucial for energy forecasting. However, challenges for SAR data interpretation complexity, and integration of ground datasets underscore a need for interdisciplinary expertise. A study in Ireland [5] validates Sentinel-1 SAR Level-2 data to estimate surface wind speed and average wind power against in-situ data from weather buoys and coastal stations using 1,544 match-up points from May 2017-2019. Despite a consistent underestimation of wind speeds by 0.4 m/s, satellite data has strong correlations with the in-situ measurements *(R>0.92)* and reliable average wind power estimations, with errors of ~10% for buoys and ~5% for coastal stations. Seasonal analysis highlights stronger winds in winter and autumn, aligning with North Atlantic cyclonic patterns, while the northwest coast exhibits the highest wind speeds. Yet, spatial analysis reveals biases influenced by proximity to coastlines, emphasizing a need for more careful interpretation. Sentinel-1 SAR data with machine learning offers offshore wind resource assessments at turbine hub heights, as shown in a study in Dutch coast [6]. Areas of 70 km^2 use SAR data and Doppler wind LiDAR measurements for validation. Machine learning models correct SAR surface wind speeds, integrating geometrical parameters and metadata of the SAR sensor, and buoy data, achieving minor bias of 0.02 m/s. Corrected speeds are extrapolated to 200m hub heights using meteorological inputs from numerical weather models. High-resolution SAR wind power maps coastal wind gradients, surpassing a few models. It enhances site assessments and risk management by correcting SAR data and generating detailed wind fields. Some gaps in this study highlight the potential to address challenges in offshore wind resource mapping to support efficient farm development and sustainable energy planning. Research in the Norwegian Arctic [7] assesses Sentinel-1 Level-2 Ocean Wind Field (OWI) products to study offshore wind conditions comparing with in-situ observations and reanalysis datasets (ERA5, NORA3, CARRA). Sentinel-1 demonstrates strong correlations with these datasets, and the wind direction accuracy is consistent across datasets. The high-resolution dataset in Sentinel-1 of 1 km provides spatial advantage over reanalysis datasets (2.5-31 km) for localized assessments, though reanalysis offered broader temporal coverage. The findings support Sentinel-1's integration into offshore wind energy strategies, especially enhancing wind resource evaluation in Arctic regions, and its unique characteristics in wind distribution modeling. It motivates more studies in the overall field, such as our work here.

Our research in this paper is orthogonal to the literature. It supplements existing work [4-7] through its novelty and uniqueness in terms of mining complex satellite imagery with unsupervised learning, aiding long-term decision support in offshore wind farm and energy estimation. Moreover, since the learning is unsupervised in our work, it does not require pre-labeled training datasets along with predefined notions of correctness. In many domains, e.g. environmental computing, such huge training data with ground truth from experts can be hard to obtain. Hence, our study contributes much here. It thrives on work by our team [8-14] and others [15-16], mining complex big data with domain knowledge in environmental computing and other areas, to enhance decision support.

IV. Experiments and results

Over 10 years, wind speed data has right-skewed distributions with *mean=7.44m/s* and *median=6.86m/s*, indicating that speeds above 20m/s are rare, and that the most frequent wind direction is southwest. Fig. 3 illustrates a few seasonal trends.

Fi

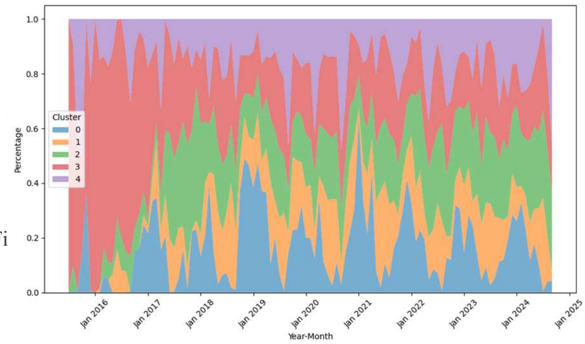

Fig. 3. Illustration of seasonal trends discovered by k-means

Our analysis with k-means clustering reveals that 18.7% of wind speeds, categorized in cluster 0, have a northwest frequency with a mean speed of 10.90 m/s and a median of 10.80 m/s along the Atlantic coasts from NJ to MA. Furthermore, cluster 0 peaks in winter and is at its lowest around the summer solstice (as seen in the blue area in Fig. 4). Furthermore, cluster 1 has the closest mean speed (7.78 m/s) to historical data and is the only cluster that does not contain data points near the shoreline. It is bounded by latitudes 40.5°N to 35°N and longitudes 73.5°W to 64.5°W. This is well-illustrated in Fig. 5 here.

Moreover, clusters 2 and 3 have lower mean wind speeds of 5.79 m/s and 5.33 m/s, respectively, and together represent 47% of the data. These clusters highlight a geographic division extending from (41°N, 73.5°W) to (39°N, 70.5°W) (see Fig. 6). Although clusters 1, 2, and 3 overlap between latitudes 39°N to 40°N and longitudes 72°W to 70.5°W, the area in cluster 1 shows higher frequencies during the summer season. All the clusters exhibit a frequent southwest wind direction, maintaining stable frequency throughout the years.

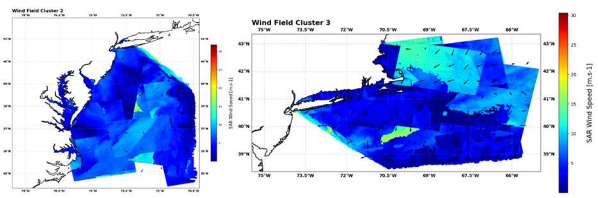

Fig. 4. Geographic distribuiton of cluster 2 and 3

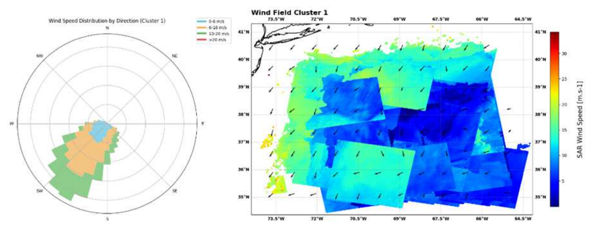

Fig. 5. Wind speed and direction in cluster 1. *Left:* Southwest direction seen as frequent. *Right:* Datapoints plotted geographically

When frequent wind directions occur in areas near the shoreline, the mean speed remains stable at 5.79 m/s, as shown in clusters 2 and 3. However, as the area extends further offshore, cluster 1 highlights regions with higher wind speeds, offering critical insights for offshore wind farm optimization and enhancing renewable energy projects. The main findings from this analysis reveal that wind speeds are stronger moving northwards and that such trends are more frequent during the winter months. Cluster 1, which represents 25.1% of the data, illustrates the potential for high ocean wind speeds towards the southwest and is the largest cluster identified.

Likewise, such analysis can yield valuable insights to support decisions about wind farm orientations and other factors, by many of the concerned stakeholders. Energy estimation from the renewable sources is also facilitated. Such work can thereby have a long-term impact of combating climate change.

V. CONCLUSIONS AND ROADMAP

This work mines complex satellite images to draw inferences helpful in decision support for offshore wind energy. Vital initial findings are that wind has higher speeds going north and can be more frequent in winter (for the areas analyzed). Such inferences can aid wind farm planning by stakeholders. Key findings from this analysis reveal that wind speeds are stronger moving north and are more frequent during winter. Cluster 1, which represents 25.1% of the data, illustrates the potential for high ocean wind speeds towards the southwest and is the largest cluster identified.

Enhancing the patterns unveiled by k-means to forecast and track wind speed and direction in real-time can potentially improve energy estimation. Accurate energy estimation is beneficial for offshore wind stakeholders, as it supports scheduling maintenance during low wind speeds for easier access, forecasting when populations reliant on this clean energy may need alternative sources, and identifying periods of surplus energy generation to store it for future use. Inferences such as these can guide decisions on wind farm orientations and distribution in offshore wind projects. Such decisions by stakeholders can have broader impacts in combating climate change through the development of more and better renewable energy projects. Future work will include reanalysis using in-situ data, leveraging machine learning to fill in a few gaps caused by the limitations of remote sensing technology. With these gaps addressed, forecasting machine learning models can further enhance energy estimation.

Our study uniquely highlights the vital role of clustering for big data mining on complex satellite imagery. As more training data is available along with the notions of ground truth, supervised learning with classifiers might be explored in the future to predict best / average case performance for optimizing energy output.

This work has broader impacts in the realm of climate change. It can guide the assessment and enhancement of solutions, thus helping much in the environmental computing area.

ACKNOWLEDGMENTS

The authors acknowledge the New Jersey Wind Institute Fellowship Program by NJEDA, a statewide initiative in NJ. We thank the Clean Energy and Sustainability Analytics Center (CESAC) at Montclair State University. In addition. Dr. Aparna Varde acknowledges NSF MRI grant 2018575 here. Another part of her research is supported by a grant from NOAA.

REFERENCES

[1] Gonzalez-Moodie, B., Daiek, S., Lorenzo-Trueba, J., & Varde, A. S. "Multispectral drone data analysis on coastal dunes". IEEE International Conference on Big Data (Big Data), 2021, pp. 5903-5905.

[2] SentiWiki. Sentinel-1.https://sentiwiki.copernicus.eu/web/sentinel-1

[3] Swarnkar, S. K., Patra, J. P., Kshatri, S. S., Rathore, Y. K., Tran, T. A. (Eds.). *Supervised and unsupervised data engineering for multimedia data*. John Wiley & Sons, 2024.

[4] Majidi Nezhad, M., Shaik, R. U., Heydari, A., Razmjoo, A., Arslan, N., & Astiaso Garcia, D. "SWOT Analysis for Offshore Wind Energy Assessment Using Remote-Sensing Potential," Applied Sciences, 2020, 10(18):6398.

[5] L. de Montera, T. Remmers, R. O'Connell, and C. Desmond, "Validation of Sentinel-1 offshore winds and average wind power estimation around Ireland," Wind Energ. Sci., 2020, 5(3): 1023-1036.

[6] L. de Montera, H. Berger, R. Husson, P. Appelghem, L. Guerlou, and M. Fragoso, "High-resolution offshore wind resource assessment at turbine hub height with Sentinel-1 synthetic aperture radar (SAR) data and machine learning," Wind Energ. Sci., 2022, 7(4):1441-1453.

[7] E. Khachatrian, P. Asemann, L. Zhou, Y. Birkelund, I. Esau, and B. Ricaud, "Exploring the Potential of Sentinel-1 Ocean Wind Field Product for Near-Surface Offshore Wind Assessment in the Norwegian Arctic," Atmosphere, 2024, 15(2):146.

[8] Pawlish, Michael J., and Aparna S. Varde. "A decision support system for green data centers." 3rd workshop on Ph. D. students in information and knowledge management. ACM CIKM, 2010, pp. 47-56.

[9] Karthikeyan, Divydharshini, Aparna S. Varde, and Weitian Wang. "Transfer learning for decision support in Covid-19 detection from a few images in big data." IEEE International Conference on Big Data (Big Data), 2020, pp. 4873-4881.

[10] Varde, A., Rundensteiner, E., Javidi, G., Sheybani, E., & Liang, J. "Learning the relative importance of features in image data". IEEE ICDE, 23rd international conference on data engineering (workshops), 2007, pp. 237-244.

[11] Varde, A. S., Takahashi, M., Rundensteiner, E. A., Ward, M. O., Maniruzzaman, M., & Sisson Jr, R. D. "Apriori algorithm and game-of-life for predictive analysis in materials science". International Journal of Knowledge-based and Intelligent Engineering Systems, 2004, 8(4):213-228.

[12] Shrestha, S., Varde, A. S. "Roles of the Web in Commercial Energy Efficiency: IoT, Cloud Computing, and Opinion Mining". ACM SIGWEB journal, 2023(Autumn), 1-16.

[13] Pawlish, M., Varde, A. S., Robila, S. A. "Cloud computing for environment-friendly data centers". IEEE ICDM, International workshop on Cloud data management, 2012 (pp. 43-48).

[14] Puri, M., Du, X., Varde, A. S., de Melo, G. "Mapping ordinances and tweets using smart city characteristics to aid opinion mining". WWW, The Web Conference (companion vol), 2018, pp. 1721-1728.

[15] Lausch, A., Schmidt, A., & Tischendorf, L. "Data mining and linked open data–New perspectives for data analysis in environmental research". Ecological Modelling, 295, 5-17.

[16] Zhang, Z., Li, J. *Big data mining for climate change*, Google Books (books.google.com), 2019.