

A multi-resolution weather dataset for the Southwestern South Atlantic (2017-2018)

Luan C. V. Silva¹, Livia Sancho¹, Mauricio S. Silva¹, Elisa Passos³, Larissa F. R. Jacinto¹, Rebeca S. Lyra^{1,2}, Nilton O. Moraes^{1,2}, Carina S. Bock¹, Douglas M. Nehme¹, Raquel Toste¹, Jacques Honigbaum¹, Rodrigo S. Luna¹, Carlos H. Beisl⁴, Patricia M. Silva¹, Adriano O. Vasconcelos¹, Rian C. Ferreira¹, Fernando A. Rochinha¹, Luiz P. F. Assad^{1,2}, Alvaro L. G. A. Coutinho¹, Laura S. B. S. Leite¹, and Alexandre G. Evsukoff¹

¹COPPE, Universidade Federal do Rio de Janeiro

²Instituto de Geociências - IGEO, Universidade Federal do Rio de Janeiro

³Faculdade de Oceanografia, Universidade do Estado do Rio de Janeiro

⁴Geospatial Petroleum

ABSTRACT

The Southwestern South Atlantic (SWSA) is a critical region for understanding climate dynamics and assessing renewable energy potential, yet high-resolution meteorological data remain limited. Here we present a multi-resolution weather dataset spanning February 2017 to November 2018, combining atmospheric simulations from the Weather Research and Forecasting model with wind field retrievals from Sentinel-1A/B Synthetic Aperture Radar (SAR) processed using the CMOD5 geophysical model. WRF outputs were generated at 30-minute intervals for three nested domains (9-km, 3-km, and 1-km resolution) through 975 short-term simulations. SAR/CMOD5 wind fields are provided at 500 m and 1 km resolution across 104 acquisition dates. Model validation results show strong agreement between WRF and SAR/CMOD5 10-m wind speed daily spatial averages, with more than 93% of acquisition days having both RMSE and MAE below 3 m/s. On a per-pixel basis, over 91.5% of the residuals lie within ± 3 m/s. In situ wind measurements from the Itajaí buoy were used to validate both data sources, and also yielded satisfactory results. The dataset supports various applications, including regional climate studies, wind energy resource assessment, and machine learning development for weather forecasting and downscaling. Usage examples to support practical use are provided.

1 Background & Summary

The South Atlantic plays a pivotal role in the global climate system. It is the only ocean basin with a net equatorward heat transport, directly influencing heat redistribution, carbon storage, and global climate variability (1). The limited understanding of local physical processes and climate variabilities in the Southwestern South Atlantic (SWSA) is largely attributed to sparse observational data and inaccuracies in long-term model integrations (2).

Brazil is already one of the world's largest producers of renewable electricity, with hydropower accounting for over half of its electricity generation, leveraging the extensive river systems (3). However, the region also has significant potential for other sources of large-scale renewable energy generation. Brazil's coast has potential for offshore wind energy estimated at ≈ 3 TW and more than 14,800 TWh of average annual electricity production (4; 5) and wave energy, with the potential to reduce CO₂ emissions by approximately 44.52 million tons per year if replacing gas-fired thermal energy (6). According to Vinhoza & Schaeffer (7), few studies have comprehensively assessed Brazil's offshore potential, and those that exist have primarily focused on gross potential, often neglecting technological, environmental, and social constraints. While global models do provide coverage of the region, increasing spatial resolution significantly enhances the representation of meteorological features such as wind variability and extreme events (8; 9).

Beyond improving our understanding of SWSA weather dynamics, a multi-resolution weather dataset also provides rich training material for artificial intelligence (AI) methods. Thanks to booming computing power, it is now possible to simulate climate systems at high resolution. This, combined with a surge in climate observations from weather satellites, creates a higher data volume that allows for the application of advanced machine learning methods to improve climate modeling and prediction (10). However, data limitations are still recognized as a fundamental challenge in weather forecasting (11; 12). AI can reduce both cost and time for delivering high-resolution weather data (13).

In this study, we present a dataset covering the southeastern coast of South America and the adjacent South Atlantic

Ocean was created by integrating numerical simulations and satellite observations at multiple spatial resolutions. High-temporal-resolution (30 min) output was generated by the Weather Research and Forecasting (WRF) model (14) through 975 independent runs-each lasting approximately 5.5 hours-for a combined compute time of about 5,362.5 hours. SAR images from Sentinel-1A/B C-band sensors (15; 16) were processed with the CMOD5 algorithm to estimate 10-m wind fields. Wind fields from the numerical and satellite models were validated against height-corrected in situ buoy measurements. This dataset spans from February 2017 to November 2018 and provides exciting possibilities for AI applications, including forecast emulation (17; 18) and cross-scale downscaling (19). The different data sources available allow observation-constrained bias correction (20; 21), and the wide range of weather variables allows multivariate analysis (22), leading to improved results and better exploitation of correlations among various atmospheric parameters.

2 Methods

2.1 Study area

The study area is situated in the southwestern Atlantic, with a focus on the southern coast of Brazil and its adjacent offshore regions. To ensure adequate spatial coverage and resolution over the study region, a system of nested computational domains was configured. The simulation used a one-way nesting strategy with three domains, as shown in Figure 1a. The outermost domain (D01), shown in blue, covers a broad area at a 9 km grid resolution; an intermediate domain (D02), in black, with a grid resolution of 3 km, is embedded within D01 and serves as a bridge by receiving input from D01 and passing boundary information to the innermost domain. The innermost domain (D03), shown in red, focuses on the core study area, which was configured with a grid resolution of 1 km (Figure 1a).

Figure 1b displays the area of the wind speed computed from the SAR acquisition frames overlapping the study area, serving as a reference for data validation along with the Itajaí meteo-oceanographic buoy location. This combination of nested atmospheric domains and satellite coverage enables a detailed examination of wind variability from regional to local scales, with particular focus on the high-resolution D03 area.

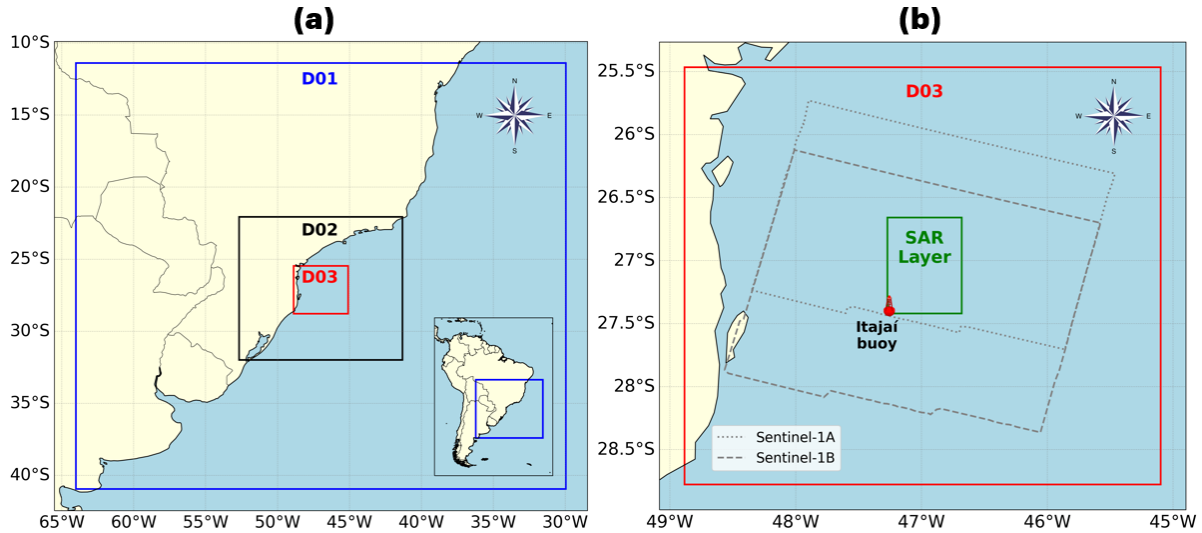


Figure 1. Study area. (a) Three-nested domains configured for the Weather Research and Forecasting model (WRF); D01 with 9 km resolution in blue, D02 with 3 km resolution in black, and D03 with 1 km resolution in red. (b) WRF model's most refined grid (D03), selected SAR region and the location of the Itajaí meteo-oceanographic buoy.

2.2 The Weather Research and Forecasting (WRF) model

The atmospheric simulations were conducted using version 4.3.3 of the WRF model, configured with the Advanced Research WRF (ARW) dynamic core. This core integrates the full set of Navier-Stokes equations, thus avoiding the hydrostatic assumption and allowing for a detailed representation of atmospheric flows. The vertical discretization employs a terrain-following hybrid sigma-pressure coordinate. The model framework adopts the Arakawa staggered C-grid (23) and utilizes finite differences to solve the governing equations (24). Staggered grids solve quantities in different regions of the grid, such as corners, centers, or faces, which means that dimensions can vary at one point depending on the variable. The mass, thermodynamic, scalar, and chemical variables are calculated in the center of the grid cell. At the same time, the U and V components of horizontal velocity

are normal to the corresponding faces of the cell (14). Thus, all dimensions with the suffix "stag" are located on the normal to the corresponding faces of the cell, while the others are located in the center of the cell.

Table 1. WRF model configuration.

Description	Configuration
Timing	
Simulation period	06:30h 06/02/2017 to 06h 18/11/2018
Time step	48s
Output Frequency	30 min
Maps and grids	
Map projection	Mercator
Horizontal grid spacing	9, 3 and 1 km (361 x 355)
Grid Points	361 x 355 (all grids)
Vertical levels	45
Nesting strategy	One-way
Topography data	GMTED2010 30 s (1 km) (25), GMTED2010 30 s (1 km) (25) and SRTM 9 s (90 m) (26)
Land Use and Land Cover data	MODIS 30 s (1 km) (27), MODIS 30 s (1 km) (27), and MAPBIOMAS 30-m (28)
Forcing strategy	
Initial and Boundary Conditions	FNL GFS (0.25 km) + SST from ROMS model on D03
Boundary Conditions Frequency	6 hours
Initialization	00 hour
Runs duration	2 days + 6 hours (54 hours, 109 times)
Spin-up time	6 hours
Physical parametrization schemes	
Shortwave radiation	RRTMG (29)
Longwave radiation	RRTMG (29)
Cumulus parametrization	Multi-Scale Kain-Fritsch (30)
Microphysics	Morrison Double-Moment (31)
Surface Layer	Revised MM5 scheme for the surface layer (32)
Land surface model	Noah-MP (33)
Planetary boundary layer	Yonsei University (YSU) (34)
Turbulence Diffusion Option	2

As a consequence, the number of grid points for the "stag" variables is one unit larger than the other variables located in the center of the cell grid. In this way, the variables XLAT and XLONG store the coordinate values located at the centers of the grid cells, while the variables XLAT_U, XLONG_U, XLAT_V, and XLONG_V store the coordinate values located at the respective faces of the grid cells. Detailed information can be found on the WRF website¹.

Physical processes, including radiation balance, microphysical processes, convection, turbulence, and surface interactions, are incorporated through parameterization schemes. In the vertical dimension, all domains shared the same discretization, composed of 45 vertical levels distributed throughout the atmospheric column. Close to the surface, enhanced resolution was applied with four vertical levels approximately 200 meters above mean sea level, located at 26, 60, 103, and 160 meters, respectively. The nested domains were centered at 27.12°S and 46.99°W, using the Mercator projection. This positioning ensures dynamic consistency of the physical processes within the model and avoids inconsistencies or artificial relaxation effects at the boundaries of D03 relative to the core area of interest.

The physical options of the model were selected based on previous evaluations and applications in similar contexts (35; 36; 37). The chosen schemes are the following: Morrison Double-Moment for cloud microphysics (31), Multi-Scale Kain-Fritsch for convective processes (30), Rapid Radiative Transfer Model for General Circulation Models (RRTMG) for both longwave and shortwave radiation (29), Noah-MP for land surface modeling (33), Revised Mesoscale Model (MM5) scheme for the surface layer (32), and the Yonsei University scheme for boundary layer dynamics (34). Horizontal diffusion was treated in physical space to ensure better accuracy over complex terrain. Overall model configuration is displayed in Table 1.

Another important change in the data concerns the precipitation variables (RAIN and RAINNC). Originally, the precipitation data produced by WRF are cumulative over the entire duration of the one run (every 54 hours in the present case); however, the precipitation values made available are cumulative only between time intervals (30 minutes). Some variables can be derived indirectly from others, as shown in Table 2, which presents a relevant subset of such relationships.

¹https://amps-backup.ucar.edu/information/configuration/wrf_grid_structure.html

Table 2. Variables derived from WRF outputs and their formulas.

Variable	Formula
Atmospheric pressure (P_{atm})	$P + PB$
Geopotential height	$PH + PHB$
Height above mean sea level (z)	$\frac{PH+PHB}{g}$, $g = 9,81 \text{ m/s}^2$
Potential temperature (θ)	$T + 300$
Air temperature (T_{air})	$\theta \left(\frac{P+PB}{100} \times \frac{1}{1000} \right)^{\frac{R}{c_p}}$, $\frac{R}{c_p} = 0,2854$

The simulations utilized initial (IC) and boundary conditions (BC) from the final analysis (FNL) provided by the National Centers for Environmental Prediction (NCEP), which were derived from the Global Forecast System (GFS2). This dataset offers global atmospheric fields with 0.25° horizontal resolution (approximately 24 km) and a temporal frequency of 6 hours. The GFS model has undergone continual development since the 1980s, with the most recent upgrade documented in 2021². To improve the terrain representation in the boundary data, topographic information from the Shuttle Radar Topography Mission (SRTM) (26) and land cover data from the MapBiomas project (28) were incorporated. This approach follows methodological refinements proposed in recent works (38; 39). Additionally, to improve the air-sea interactions on the innermost domain (D03), Sea Surface Temperature (SST) results from a hydrodynamic model were implemented. More details can be found in item 2.2. A summary of the main features of the simulations is presented in Table 1.

Sea Surface Temperature (SST) boundary condition. The hourly SST boundary conditions, with ≈ 1 km of horizontal resolution, were obtained from a Regional Ocean Modelling System (ROMS) simulation. ROMS is a widely used three-dimensional free-surface hydrostatic model that solves the primitive equations in a discretized space using an Arakawa C grid (40). The hydrodynamic simulation was performed on a regional grid with 40 vertical layers and encompasses the region between 25.3°S and 28.9°S in latitude and 45.0°W and 49.0°W in longitude, i.e., including the WRF's D03 domain. The bathymetry was extracted from the Earth TOPOgraphy (ETOPO) 2022 (41) global model. The simulation was performed using the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5) (42) as surface forcing and the daily results from the Copernicus Marine Environment Monitoring Service's GLORYS12V1 (Global Ocean Physics Reanalysis) product (43) as IC and lateral BC. Astronomical tides from the TOPEX/Poseidon TPXO7.2 global model were also used as forcing at open boundaries (44; 45). SST results were validated using the Operational Sea Surface Temperature and Ice Analysis dataset (46) and in situ data from the National Buoy Program (PNBOIA³) network.

2.3 Wind field estimation from Synthetic Aperture Radar (SAR) images

SAR imagery. The wind field was estimated using data from the Sentinel-1A and Sentinel-1B satellites, provided free of charge through the European Space Agency (ESA) Sentinel portal⁴. The data were acquired in Interferometric Wide Swath (IW) mode, featuring a swath width of approximately 250 km and a spatial resolution of $5\text{-m} \times 20\text{-m}$ (ESA, 2024). The original datasets are available at Level-1 processing in Ground Range Detected (GRD) format, which consists of focused SAR data that have been detected, geocoded, and projected onto the ground range using the WGS84 ellipsoid model. The pixel values represent radar backscatter intensity in this format, with phase information discarded.

For the present study, a total of 104 SAR images were acquired over the offshore region of interest. The temporal availability of validated reference data from the Itajaí metoceanographic buoy guided image selection. Although the nominal temporal resolution of Sentinel-1 is 12 days per satellite, the combined coverage of Sentinel-1A and Sentinel-1B over the region of interest results in an effective revisit time of approximately 6 days for descending orbits. However, the image frames of the two satellites are spatially offset. Figure 1b illustrates the spatial configuration of the Sentinel-1A and Sentinel-1B acquisition frames. The Sentinel-1A frame is deliberately offset relative to the Sentinel-1B frame, enabling both enhanced temporal coverage and reduced imaging gaps.

The coverage area of the Sentinel-1A frame is approximately $44,558 \text{ km}^2$, while that of Sentinel-1B is about $52,660 \text{ km}^2$. The overlapping region between the two frames spans $32,500 \text{ km}^2$, corresponding to 73.0% and 61.7% of the Sentinel-1A and Sentinel-1B frames, respectively. Notably, the Itajaí metoceanographic buoy is located near the southern edge of the Sentinel-1A frame, whereas it is positioned closer to the center of the Sentinel-1B frame. This spatial discrepancy may introduce edge effects in the wind field estimates derived from the Sentinel-1A data, potentially influencing comparisons with buoy measurements. To mitigate these effects and ensure consistency in the spatial analysis between the two acquisitions, a subset was defined based on

²https://emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php

³<https://www.marinha.mil.br/chm/dados-do-goos-brasil/pnboia>

⁴<https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data>

the intersection area between the Sentinel-1A and Sentinel-1B scenes (SAR Layer in Figure 1b). This subset focuses on the central portion of the overlapping region and includes the buoy location, thereby reducing spatial mismatches and allowing for a more robust comparison of wind estimates across sensors.

In SAR imagery, ships appear as strong backscatters and are represented as bright targets against the ocean surface (Figure 2). Masking these high-reflectivity objects prevents such targets from introducing artifacts into the wind speed retrieval process (47). Without this masking step, ship signatures may lead to erroneous wind speed estimates, manifesting as outliers in the derived wind field.

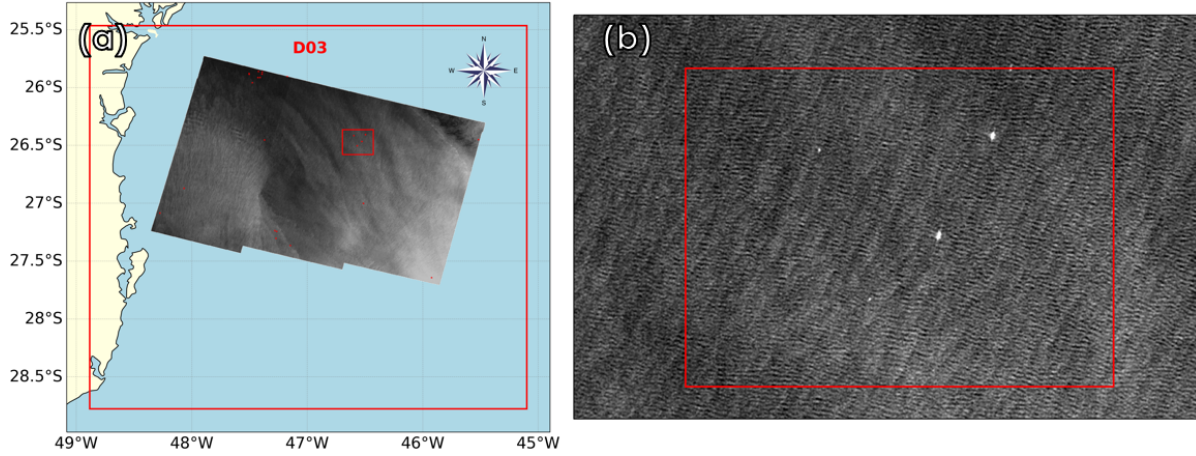


Figure 2. (a) Bright object detection mask produced by the SNAP platform, highlighting potential ship targets; (b) detailed view of the corresponding objects identified in the SAR image.

A series of specific functions was developed to detect and mask anomalous reflectors, thereby mitigating this issue. Ship detection is based on identifying point-like intensity peaks that significantly exceed the surrounding pixel values. In contrast, oil slick detection focuses on regions exhibiting anomalously low backscatter intensities.

The primary function implemented for this purpose, named *ship_detection*, is designed to identify both types of anomalies. By default, pixels with values above the 99.7th percentile of the image intensity distribution are classified as ships, whereas pixels below the 0.01st percentile are flagged as potential oil slicks. These threshold values are configurable and can be adapted to suit specific application requirements.

Following anomaly detection, the *dilation_mask* function is applied to expand the initially identified pixel regions. This dilation step is crucial for eliminating discontinuities and forming cohesive, contiguous masks. The operation uses a default square structuring element of 3×3 pixels, though this parameter can be adjusted depending on the desired level of mask expansion.

To reconstruct masked regions in the SAR images, we employed an image inpainting function, a technique commonly used for restoring degraded or missing areas in digital images, such as damaged photographs. This function provides two available methods: Terrain Elevation Dependent Attenuation (TELEA) and the Navier-Stokes method. The TELEA method utilizes the Fast Marching Method to propagate known pixel values into the masked regions iteratively (48). The Navier-Stokes method, on the other hand, applies partial differential equations grounded in fluid dynamics to guide the reconstruction process (49). A comparative analysis was conducted to evaluate the performance of both techniques within the context of wind field estimation from SAR images. The Navier-Stokes method was selected as the default approach due to its superior performance in maintaining image continuity and minimizing edge artifacts. Nevertheless, the inpainting method remains a configurable parameter within the implementation, allowing users to adapt the choice according to specific application requirements.

The CMOD5 geophysical model. The subsequent stage of the methodology involves applying the CMOD5 geophysical model function (GMF), developed by Hersbach in 2003 (50). This empirical model is designed to estimate near-surface wind speed and direction from the normalized radar cross-section (NRCS or σ^0) obtained from C-band SAR measurements.

As wind interacts with the ocean surface, it generates small-scale roughness patterns that tend to align with the wind direction. These surface features modulate the radar backscatter signal, a relationship well documented in the literature (e.g., Wackerman et al. (51)). The CMOD5 model captures this relationship by expressing σ^0 as a function of three key variables: wind speed at 10 meters above sea level (u_{10} , in m/s), the relative angle between the radar look direction and wind direction (Φ , in degrees), and the local incidence angle of the radar beam (θ , in degrees). Formally, the model is expressed as

$\sigma^0 = f(u_{10}, \Phi, \theta)$, where f is a non-linear empirical function calibrated using extensive datasets from scatterometers, SAR observations, buoys, and numerical weather prediction models.

Wind direction is first estimated using the u and v components of wind vectors provided by the WRF model. These components are interpolated to match the spatial resolution defined for the SAR-based analysis. The wind direction is computed for each grid cell and subsequently compared with the direction retrieved from SAR imagery using the CMOD5 model. The wind speed field is derived by applying the CMOD5 geophysical model to the Sentinel-1 SAR data. This process yields two wind speed products: one derived from the WRF model and the other obtained directly from the SAR imagery through inversion of the CMOD5 function.

3 Data Records

The dataset acquired in this data descriptor is available at NSF NCAR Research Data Archive⁵. The dataset is composed of: (1) WRF weather output with selected variables (Table 4), (2) SAR/CMOD5 wind spatial retrievals, (3) in situ buoy data (single location). Table 3 shows the grid sizes of in the dataset for WRF and SAR/CMOD5 layers for the available spatial resolutions. Note that the WRF grid dimensions listed in Table 3 refer to the number of mass-point (C-grid) locations in the horizontal (X-Y) plane-that is, the unstaggered model grid. These values represent the size of a single horizontal slice of mass centers in the WRF output.

Table 3. Grid dimensions for WRF and SAR/CMOD5 10-m wind speed data at different spatial resolutions.

Resolution	WRF	SAR/CMOD5
500 m	–	164×114
1 km	354×360	82×57
3 km	354×360	–
9 km	354×360	–

WRF data WRF output is provided for three nested domains (D01, D02, D03) spanning 00:30 UTC 06 February 2017 through 06 UTC 18 November 2018 at 30-minute temporal resolution. Table 4 lists the WRF variables in the dataset. The first seven are coordinates, while the rest are data variables. For more details, refer to the WRF Users' Guide⁶.

As shown in Table 5, the WRF files that compose this dataset have been reorganized into two groups: "4D" variables, which contain individual three-dimensional variables (e.g., wind components, temperature, pressure) that evolve in all three spatial dimensions plus time; and the remaining data variables, mostly two-dimensional surface fields (e.g., land-sea mask, 2-m temperature, surface pressure). All variables have the corresponding coordinates and the data is stored in the NetCDF (.nc) format.

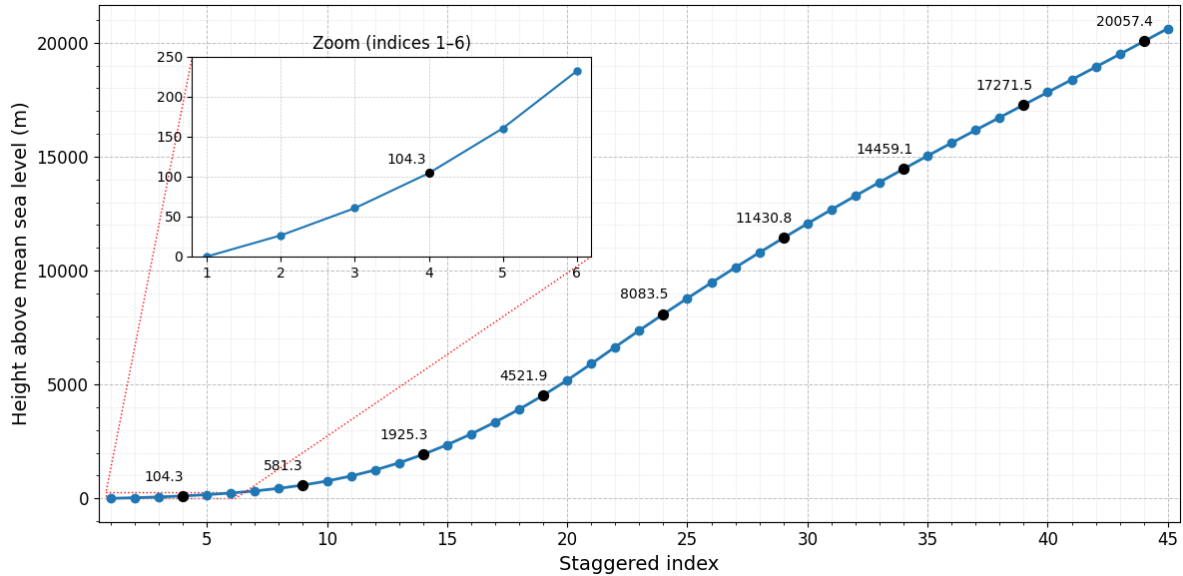
Bilgili, M., Alphan, H. & Ilhan, A. (2022) reported that the 2020 mean hub height of offshore wind turbines was 104.03 m (52). Figure 3 shows the corresponding heights at a range of vertical levels, calculated using the formula given in Table 2, with the highest levels exceeding 2,000 m. However, the full set of "4D" variables required to represent every model level produces very large files. Many climate-related applications-such as assessing offshore wind-energy potential-do not require altitudes this high. Accordingly, we limit this dataset to the first 6 staggered and 5 centered vertical layers, substantially reducing file size. The data with all layers is available upon request.

⁵<https://rda.ucar.edu/>

⁶https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/contents.html

Table 4. Description of WRF output variables available in the dataset.

Variable	Description and units	Dimensions
Times	Reference date (calendar date)	1D (Time)
XLAT	Latitude values of grid cell center points (degree_north)	2D (south_north, west_east)
XLONG	Longitude values of grid cell center points (degree_east)	2D (south_north, west_east)
XLAT_U	Latitude values of grid cell zonal faces (degree_north)	2D (south_north, west_east_stag)
XLONG_U	Longitude values of grid cell zonal faces (degree_east)	2D (south_north, west_east_stag)
XLAT_V	Latitude values of grid cell meridional faces (degree_north)	2D (south_north_stag, west_east)
XLONG_V	Longitude values of grid cell meridional faces (degree_east)	2D (south_north_stag, west_east)
LU_INDEX	Land use index (categorical)	2D (south_north, west_east)
HGT	Terrain height above sea level (m)	2D (south_north, west_east)
LANDMASK	Land-sea mask (1 = land, 0 = water)	2D (south_north, west_east)
ZNU	Vertical coordinate at full (mass) levels (m)	2D (Time, bottom_top)
ZNW	Vertical coordinate at half (staggered) levels (m)	2D (Time, bottom_top_stag)
T2	2-m air temperature (K)	3D (Time, south_north, west_east)
TH2	2-m potential temperature (K)	3D (Time, south_north, west_east)
SST	Sea surface temperature (K)	3D (Time, south_north, west_east)
PHB	Base-state geopotential ($\text{m}^2 \cdot \text{s}^{-2}$)	3D (bottom_top_stag, south_north, west_east)
PB	Base-state pressure (Pa)	3D (bottom_top, south_north, west_east)
PSFC	Surface pressure (Pa)	3D (Time, south_north, west_east)
RAINC	Cumulative convective precipitation ($\text{kg} \cdot \text{m}^{-2}$)	3D (Time, south_north, west_east)
RAINNC	Cumulative non-convective precipitation ($\text{kg} \cdot \text{m}^{-2}$)	3D (Time, south_north, west_east)
U10	10-m U-component of wind ($\text{m} \cdot \text{s}^{-1}$)	3D (Time, south_north, west_east)
V10	10-m V-component of wind ($\text{m} \cdot \text{s}^{-1}$)	3D (Time, south_north, west_east)
T	Perturbation potential temperature (K)	4D (Time, bottom_top, south_north, west_east)
PH	Perturbation geopotential ($\text{m}^2 \cdot \text{s}^{-2}$)	4D (Time, bottom_top_stag, south_north, west_east)
P	Perturbation pressure (Pa)	4D (Time, bottom_top, south_north, west_east)
U	U-component of wind ($\text{m} \cdot \text{s}^{-1}$)	4D (Time, bottom_top, south_north, west_east_stag)
V	V-component of wind ($\text{m} \cdot \text{s}^{-1}$)	4D (Time, bottom_top, south_north_stag, west_east)
W	Vertical velocity ($\text{m} \cdot \text{s}^{-1}$)	4D (Time, bottom_top_stag, south_north, west_east)

**Figure 3.** Approximate (pressure-dependent) height above mean sea level at each WRF staggered vertical index, computed as $(PH + PHB)/g$, where PH and PHB are the perturbation and base-state geopotential respectively, and g is the gravitational acceleration.

As an illustrative case, Figures 4 and 5 present the near-surface meteorological fields simulated across the three model

Table 5. Summary of WRF NetCDF files.

File Name	D01 (GB)	D02 (GB)	D03 (GB)	Description
data_{domain}_3D.nc	51.85	47.53	42.98	2D and 3D data variables
data_{domain}_P.nc	25.92	24.54	22.52	Perturbation pressure on 4D grid.
data_{domain}_PH.nc	39.88	37.74	35.05	Perturbation geopotential on staggered 4D bottom-top grid.
data_{domain}_T.nc	30.94	29.18	25.08	Perturbation potential temperature on 4D grid.
data_{domain}_U.nc	45.99	45.38	43.54	x-wind component on staggered 4D east-west grid.
data_{domain}_V.nc	45.72	44.93	42.33	y-wind component on staggered 4D north-south grid.
data_{domain}_W.nc	56.14	55.51	51.18	z-wind component on staggered 4D bottom-top grid.
Size (GB)	296.44	284.81	262.68	-
Total size (GB)		843.93		-

domains for the experiment initialized on August 1st, 2017. These figures specifically depict the spatial distribution of 2-meter air temperature and 10-meter wind.

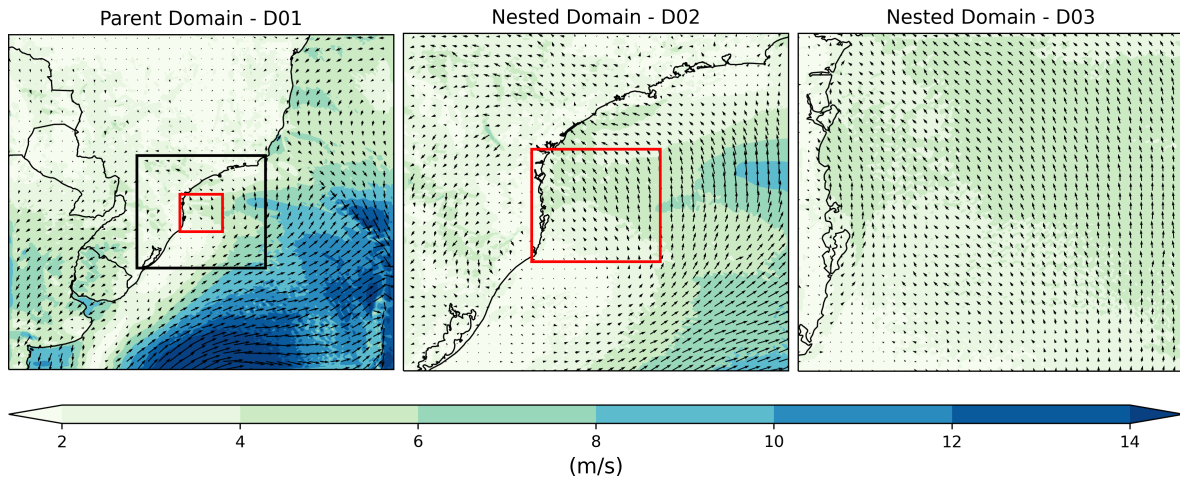


Figure 4. Maps of wind fields at 10 meters from the 9km (D01), 3km (D02) and 1 km-grid (D03) of the August 1st, 2017 run of the WRF model.

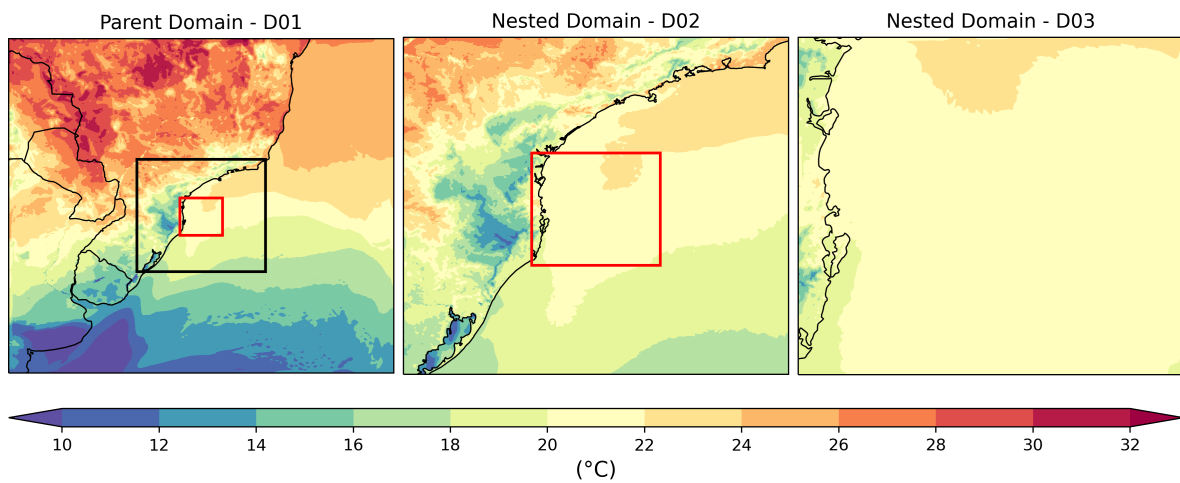


Figure 5. Maps of 2-meter temperature fields from the 9km (D01), 3km (D02) and 1 km-grid (D03) of the August 1st, 2017 run of the WRF model.

SAR/CMOD5 data All processed SAR/CMOD5 data are consolidated into a single NetCDF file containing 104 acquisition dates between 07 February 2017 and 11 November 2018. SAR-derived wind fields computed via CMOD5 inversion after masking and inpainting. The variables present in the SAR/CMOD5 data are shown in Table 6. Two data files were created based on the spatial resolution (Table 7). Figure 6 shows examples of wind speed fields derived from the model at both resolutions.

Table 6. Description of SAR/CMOD5 variables available in the dataset.

Variable	Description and units	Dimensions
Times	Date of SAR image acquisition (calendar date)	1D (Time)
XLAT	Latitude values (degree_north)	2D (south_north, west_east)
XLONG	Longitude values (degree_east)	2D (south_north, west_east)
XSAT	Satellite identifier	1D (satellite)
U10	10-m U-component of wind ($\text{m}\cdot\text{s}^{-1}$)	3D (Time, south_north, west_east)
V10	10-m V-component of wind ($\text{m}\cdot\text{s}^{-1}$)	3D (Time, south_north, west_east)
WIND_SPD	Wind speed magnitude ($\text{m}\cdot\text{s}^{-1}$)	3D (Time, south_north, west_east)
WIND_DIR	Wind direction (degrees from north)	3D (Time, south_north, west_east)

Table 7. Summary of SAR/CMOD5 NetCDF files.

File Name	Size	Description
data_SAR_500m.nc	7.47 MB	SAR/CMOD5 wind speed fields at 500-m spatial resolution.
data_SAR_1000m.nc	29.83 MB	SAR/CMOD5 wind speed fields at 1-km spatial resolution.

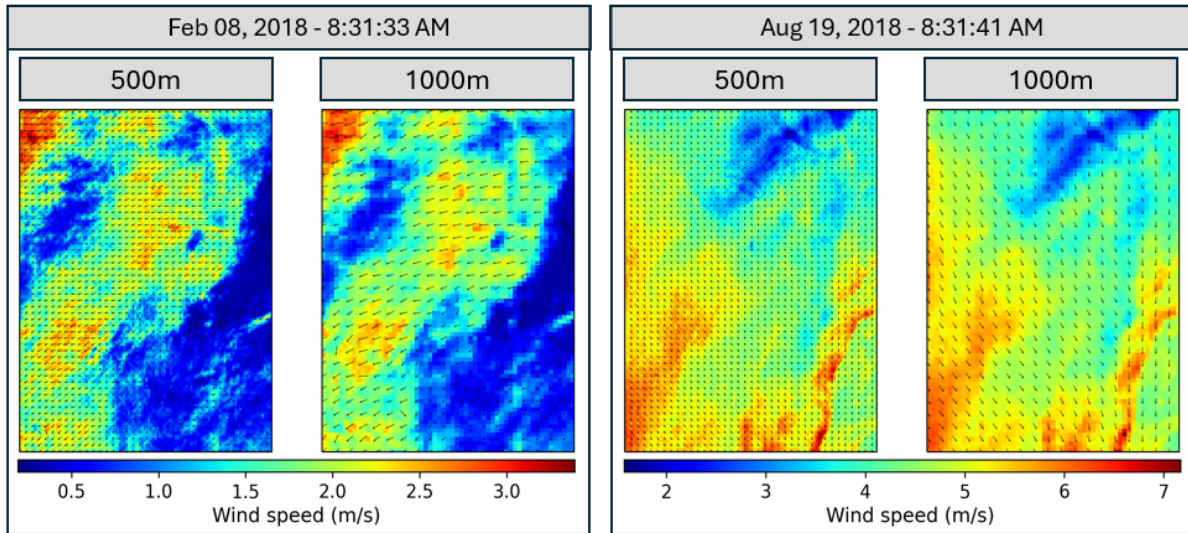


Figure 6. SAR/CMOD5 Wind speed field maps derived from the SAR images of February 8, 2018 and August 19, 2018 in 500m and 1km spatial resolution.

Itajaí oceanographic buoy In situ observations from the Itajaí oceanographic buoy (PNBOIA program) are also used to evaluate the consistency of these results. The observations include date, time, geographic coordinates, wind speed, and wind direction. However, the observations are at approximately 4-m height. To compare them to 10-m wind fields available in WRF and SAR/CMOD5 outputs, we used the logarithmic profile law (Equation 1) (53):

$$u(z) = u(z_{\text{ref}}) \frac{\ln\left(\frac{z}{z_0}\right)}{\ln\left(\frac{z_{\text{ref}}}{z_0}\right)}. \quad (1)$$

where z is the target height for wind speed estimation [m], z_{ref} is the reference height at which the speed $u(z_{\text{ref}})$ was measured [m], $u(z)$ is the wind speed at height z [m/s], and z_0 is the surface roughness length [m]. The surface roughness length

was estimated following the ERA5 methodology (42). Given the relevance of the buoy data for the region-located within the WRF and SAR/CMOD5 study area-these measurements are also provided in tabular format, with one row per observation time. The data is stored in the file `itajai_buoy.csv`, and the variables (columns) are listed in Table 8.

Table 8. Description of buoy variables available in the dataset.

Variable	Description and units
Time	Date and time of observation
Latitude	Latitude position of buoy (degree_north)
Longitude	Longitude position of buoy (degree_east)
Fsr	Surface roughness length [m]
Wdir	Wind direction (degrees from north)
Wspd	Wind speed ($\text{m}\cdot\text{s}^{-1}$)
is_SAR	SAR comparison flag (Subsection 4.2)

4 Technical Validation

The data products presented in this dataset were evaluated using two approaches over the full observation period. First, we compared the 10-meter wind fields from the WRF model and the SAR/CMOD5 retrievals. Then, we validated both datasets using in situ wind measurements from the Itajaí oceanographic buoy by comparing them with the nearest grid points in the WRF and SAR/CMOD5 data.

4.1 WRF and SAR/CMOD5 grid comparison

The wind speed data produced by the innermost WRF numerical model domain were compared with the results obtained by the SAR/CMOD5 model in the days/hours of the acquisition of the images. The validation was carried out in two different approaches: Global analysis of the grid points and analysis of the temporal mean. The 1-km SAR data and D03 WRF domain were used. Since both have the same spatial resolution and SAR/CMOD5 interpolated values to match WRF grid, we have 82×57 values at each time (Table 3) and 104 available SAR images. Therefore, there is a total of total of $N = 486,096$ records.

Global analysis For the comparison, WRF model outputs corresponding to the times closest to the SAR image acquisitions were used. The SAR images analyzed in this study were acquired between 08:31:26 and 08:32:14 UTC (HH:MM:SS). Since the WRF model provides outputs at 30-minute intervals (i.e., at 00:00 and 30:00 UTC each hour), the 08:30:00 UTC output on the respective acquisition days was selected for comparison. Table 9 shows that the mean residual between SAR/CMOD5 and WRF (SAR/CMOD5 - WRF) is low (0.145 m/s), and the 5th and 95th percentiles show that 90% of the residuals lie within the range $[-3.297, 2.366]$, reflecting overall agreement between the datasets. However, the standard deviation of the residuals (1.782 m/s) suggests notable local discrepancies.

Table 9. Statistics of the WRF, SAR/CMOD5 data and their difference

	WRF	SAR/CMOD5	Residuals
Minimum	0.080	0.200	-10.635
Maximum	16.806	19.775	11.541
Mean	6.719	6.506	-0.213
Standard deviation	2.813	2.687	1.782
1 st percentile	1.334	0.580	-5.022
5 th percentile	2.246	2.427	-3.297
95 th percentile	11.380	11.349	2.366
99 th percentile	14.324	14.104	3.647

The distributions are shown in the Figure 7. The Figure 7a shows the distributions of wind intensity values in each pixel for the WRF and SAR/CMOD5 models. The distributions are shown in Figure 7. Figure 7a displays the relative frequency distribution of wind speed values (per pixel) for both WRF and SAR/CMOD5 data. The distributions exhibit similar overall patterns, with the highest frequencies occurring in the [4-6) and [6-8) m/s intervals for both models. SAR/CMOD5 shows a heavier right tail, while WRF has more values in the lower range ([0-4) m/s), consistent with the results in Table 9.

Figure 7b shows the cumulative distribution of absolute residuals between SAR/CMOD5 and WRF. Over 90% of the absolute residuals are below 3m/s, indicating good agreement between the distributions. This is supported by the error metrics (54), with a RMSE (Eq.2) of 1.795 m/s and a MAE (Eq.3) of 1.388 m/s.

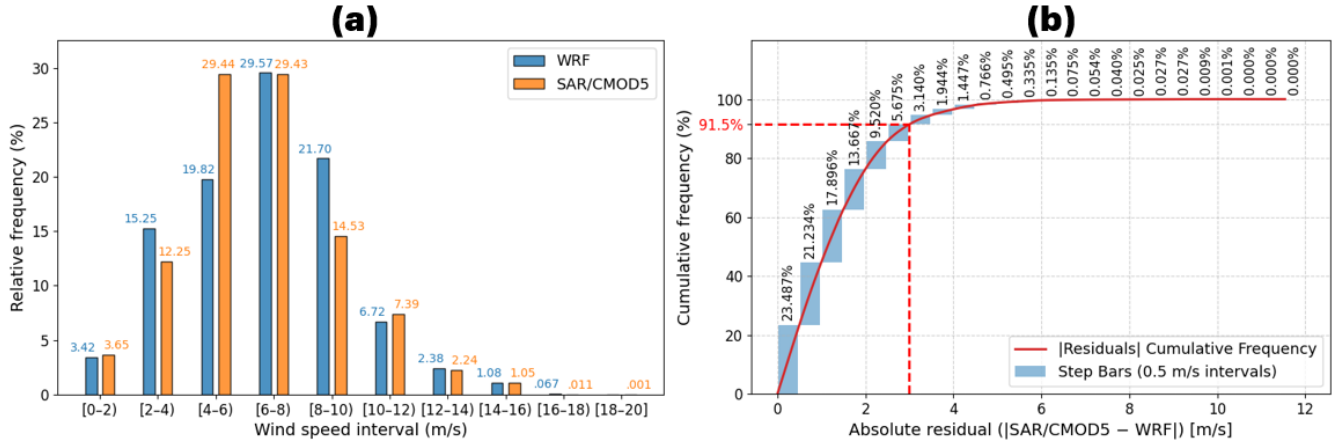


Figure 7. (a) Comparison of 10-m wind speed data from WRF and SAR/CMOD5 at the times of SAR image acquisitions. (a) Relative frequency distributions of average wind speeds for WRF (blue) and SAR/CMOD5 (orange) data across 2 m/s intervals. (b) Cumulative distribution of absolute residuals, including a step-bar representation of the incremental contributions within 0.5 m/s bins and the cumulative frequency curve.

$$\text{root mean squared error (RMSE)} = \sqrt{\frac{1}{N} \sum_{t=1}^N \text{residual}(t)^2} \quad (2)$$

$$\text{mean absolute error (MAE)} = \frac{1}{N} \sum_{t=1}^N |\text{residual}(t)| \quad (3)$$

Temporal mean In the second approach, WRF and SAR/CMOD5 wind speeds were compared based on the daily mean across all pixels ($82 \times 57 = 4674$) within the region. The upper panel of Figure 8 presents the daily mean wind speeds from both models over the 104 SAR acquisition days, while the lower panel displays the corresponding daily error metrics. As shown in the figure and summarized in Table 10, at the stricter 1 m/s threshold, over a third of the days (35.6%) have MAE values within the limit, whereas only 21.2% meet this criterion for RMSE. This indicates stronger local deviations that significantly affect the RMSE, which is more sensitive to outliers. Nonetheless, over 93% of the days exhibit RMSE and MAE below 3 m/s. Overall, the results indicate good agreement between the models in terms of daily spatial averages.

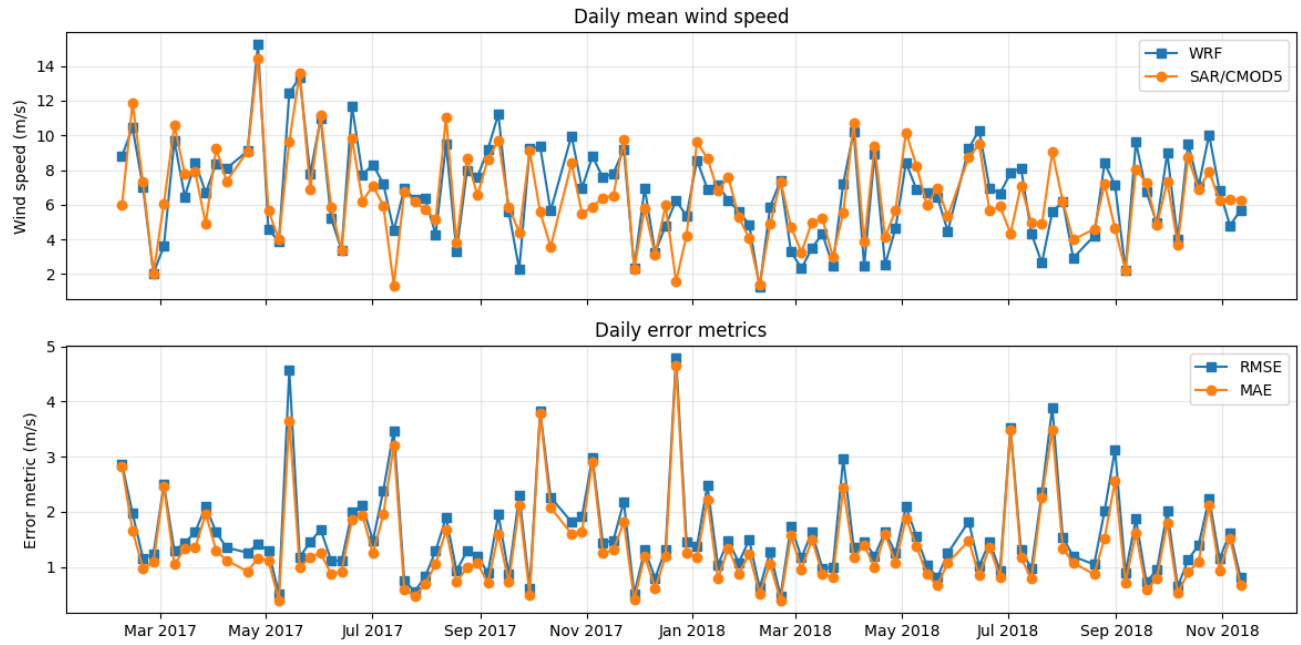


Figure 8. Daily mean wind speed at 10-m (m/s) from WRF and SAR/CMOD5 over 104 SAR acquisitions, spaced at 6-day intervals between February 2017 and November 2018. The upper panel shows the time series of daily spatial averages from both models. The lower panel presents the corresponding RMSE and MAE error metrics for each day.

Table 10. Proportion of days used in the comparison with RMSE and MAE below given thresholds.

Threshold (m/s)	RMSE \leq Threshold	MAE \leq Threshold
1	22/104 (21.2%)	37/104 (35.6%)
2	81/104 (77.9%)	88/104 (84.6%)
3	97/104 (93.3%)	98/104 (94.2%)
5	104/104 (100.0%)	104/104 (100.0%)

4.2 Buoy comparison

To evaluate the accuracy of the WRF model and SAR/CMOD5 wind speed estimates, we compare them with measurements from the Itajaí buoy. The buoy coordinates vary slightly over time, but for consistency, we used the fixed position of 27°24.35'S, 47°15.93'W. Wind speed values from the buoy were adjusted to represent the 10-meter height.

WRF The WRF simulation data spans from 2017-02-06 06:30 UTC to 2018-11-18 06:00 UTC. Buoy observations are recorded hourly, at hh:00:00 UTC. Within the entire WRF simulation period, there are 59 missing records in the buoy dataset: one isolated missing hour on 2018-04-15 14:00 UTC and one continuous gap of 58 hours from 2018-04-30 01:00 UTC to 2018-05-02 10:00 UTC (inclusive). After accounting for these gaps, there are 15,497 timestamps with overlapping WRF and buoy wind speed data. Since the closest grid point in the WRF data was used for comparison, 15,497 is also the number of observations used in the comparison.

Figure 9 compares the wind speed distributions from the WRF model and the Itajaí buoy. Panel (a) shows the relative frequency of wind speed values, binned at 2 m/s intervals, revealing distinct distributions—most notably, a heavier right tail in the buoy data. Despite these differences, the cumulative distribution of absolute residuals in panel (b) indicates that over 80% of the absolute residuals fall below 3 m/s. The calculated RMSE and MAE are 2.538 m/s and 1.956 m/s, respectively.

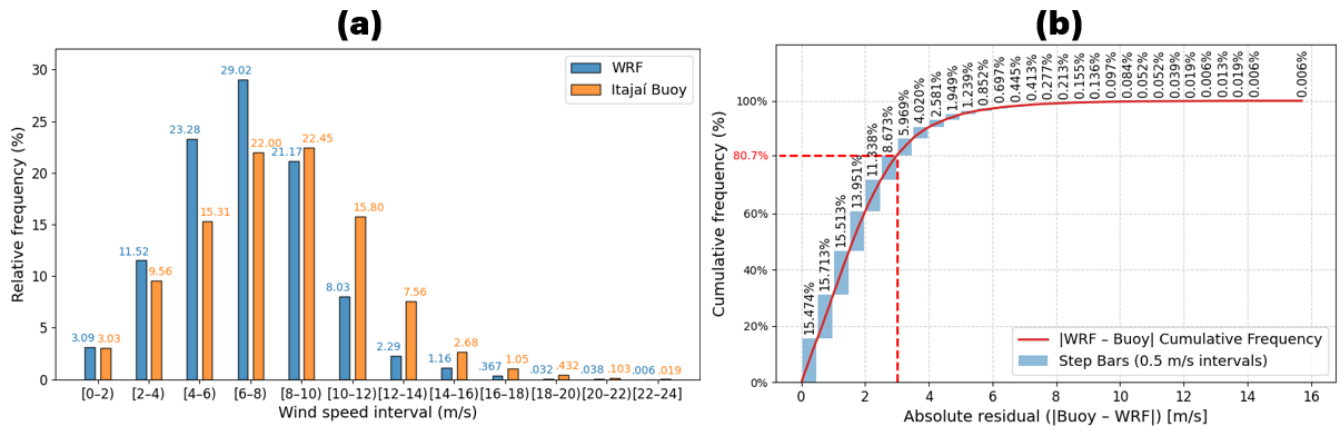


Figure 9. (a) Comparison of 10-m wind speed data from WRF and SAR/CMOD5 at the times of SAR image acquisitions. (a) Relative frequency distributions of average wind speeds for WRF (blue) and Itajaí buoy (orange) data across 2 m/s intervals. (b) Cumulative distribution of absolute residuals, including a step-bar representation of the incremental contributions within 0.5 m/s bins and the cumulative frequency curve.

SAR/CMOD5 The buoy provides hourly observations (hh:00:00), and since SAR acquisitions during the study period occurred between 08:31:26 and 08:32:14 UTC, the closest hourly record-at 09:00 UTC-was used for comparison. Similarly, the nearest grid point in the SAR data was selected. Figure 10 shows good agreement between the SAR-derived wind speeds and the buoy observations across the entire time series of 104 SAR acquisitions, spaced at 6-day intervals. MAE of 1.57 m/s and RMSE of 2.03 m/s support this result, with the higher RMSE value reflecting the influence of larger individual deviations between SAR and buoy measurements.

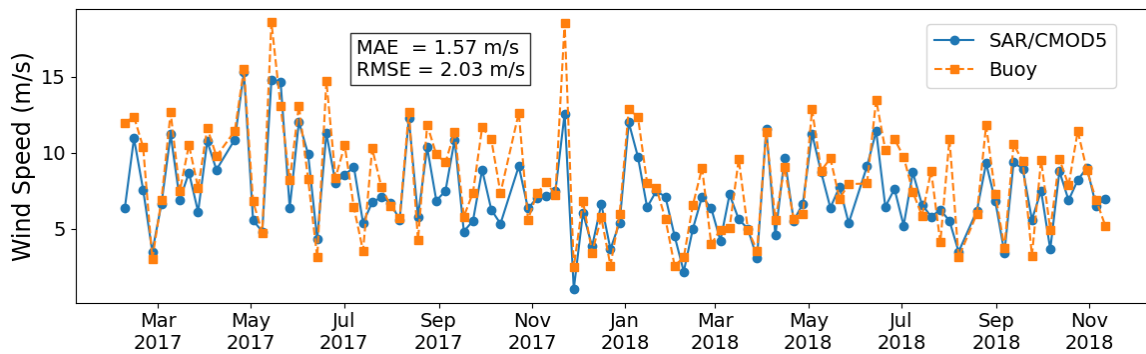


Figure 10. Time series of wind speed (m/s) from SAR/CMOD5 closest grid point to the Itajaí buoy and buoy measurements adjusted to 10-meter height. The plot includes 104 SAR acquisitions at 6-day intervals between February 2017 and November 2018.

5 Usage Notes

The dataset files are provided in NetCDF (.nc) format, a widely used standard for array-oriented scientific data. This format allows for easy manipulation and analysis using libraries such as `xarray` in Python. To work with the dataset, you can load a file named "filename.nc" using the following code snippet:

```
import xarray as xr
dataset = xr.open_dataset("data_d01_U.nc")
u_wind_component = dataset["U"]
```

The code behind Section 4, along with example scripts demonstrating key workflows (e.g., loading WRF output, computing heights, and generating figures), is available on GitHub (<https://github.com/luancvieira/Southwestern-South-Atlantic-weather-dataset>).

Code Availability

The dataset was produced using the Weather Research and Forecasting (WRF) model, which is accessible via its Git repository <https://github.com/wrf-model/WRF>. Further information can be obtained from the model's official website <https://www2.mmm.ucar.edu/wrf/users>. SAR data processing relied on the XSAR (<https://github.com/umr-lops/xsar/tree/develop>) and XSARSEA (<https://github.com/umr-lops/xsarsea>) Python libraries, which offer tools for reading, processing, and analyzing Sentinel-1 SAR measurements for ocean and coastal applications.

References

1. Chidichimo, M. P. *et al.* Energetic overturning flows, dynamic interocean exchanges, and ocean warming observed in the south atlantic. *Communications Earth & Environment* **4**, 10 (2023).
2. Gramscianinov, C. B. *et al.* Recent changes in extreme wave events in the south-western south atlantic. *State Planet* **1**, 1–12 (2023).
3. Catolico, A. C. C., Maestrini, M., Strauch, J. C. M., Giusti, F. & Hunt, J. Socioeconomic impacts of large hydroelectric power plants in brazil: A synthetic control assessment of estreito hydropower plant. *Renewable Sustainable Energy Reviews* **151**, 111508 (2021).
4. Azevedo, S. S. P. d., Junior, A. O. P., Silva, N. F. d., Araújo, R. S. B. d. & Júnior, A. A. C. Assessment of offshore wind power potential along the brazilian coast. *Energies* **13**, 2557 (2020).
5. González, M. O. A., Santiso, A. M., de Melo, D. C. & de Vasconcelos, R. M. Regulation for offshore wind power development in brazil. *Energy Policy* **145**, 111756 (2020).
6. Bastos, A. S., Souza, T. R. C. d., Ribeiro, D. S., Melo, M. d. L. N. M. & Martinez, C. B. Wave energy generation in brazil: A georeferenced oscillating water column inventory. *Energies* **16**, 3409 (2023).
7. Vinhoza, A. & Schaeffer, R. Brazil's offshore wind energy potential assessment based on a spatial multi-criteria decision analysis. *Renewable Sustainable Energy Reviews* **146**, 111185 (2021).
8. Hadjipetrou, S. & Kyriakidis, P. High-resolution wind speed estimates for the eastern mediterranean basin: A statistical comparison against coastal meteorological observations. *Wind* **4**, 311–341 (2024).
9. Ryu, J., Kim, H., Wang, S.-Y. S. & Yoon, J.-H. Increasing resolution and accuracy in sub-seasonal forecasting through 3d u-net: the western us. *EGU sphere* **2025**, 1–18 (2025).
10. Bracco, A. *et al.* Machine learning for the physics of climate. *Nature Reviews Physics* **7**, 6–20 (2025).
11. Pagano, T. C. *et al.* Challenges of operational weather forecast verification and evaluation. *Bulletin American Meteorological Society* **105**, E789–E802 (2024).
12. Chen, L. *et al.* Machine learning methods in weather and climate applications: A survey. *Applied Sciences* **13**, 12019 (2023).
13. Wu, Y. & Xue, W. Data-driven weather forecasting and climate modeling from the perspective of development. *Atmosphere* **15**, 689 (2024).

14. Skamarock, W. C. et al. A description of the advanced research wrf model version 4.3. Tech. Rep., NCAR (2021). [10.5065/1dfh-6p97](#).
15. Hadjipetrou, S. et al. Evaluating the suitability of sentinel-1 sar data for offshore wind resource assessment around cyprus. *Renewable Energy* **182**, 1228–1239 (2022).
16. Moreira, A. et al. A tutorial on synthetic aperture radar. *IEEE Geoscience Remote Sensing Magazine* **1**, 6–43, [10.1109/MGRS.2013.2248301](#) (2013).
17. Li, L., Carver, R., Lopez-Gomez, I., Sha, F. & Anderson, J. Generative emulation of weather forecast ensembles with diffusion models. *Science Advances* **10**, eadk4489 (2024).
18. Wang, J., Balaprakash, P. & Kotamarthi, R. Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model. *Geoscientific Model Development* **12**, 4261–4274 (2019).
19. Oh, M., Lee, J., Kim, J.-Y. & Kim, H.-G. Machine learning-based statistical downscaling of wind resource maps using multi-resolution topographical data. *Wind Energy* **25**, 1121–1141 (2022).
20. Zhang, J., Gao, Z. & Li, Y. Deep-learning correction methods for weather research and forecasting (wrf) model precipitation forecasting: a case study over zhengzhou, china. *Atmosphere* **15**, 631 (2024).
21. Wang, F. & Tian, D. On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics* **59**, 3451–3468 (2022).
22. Zhu, X. et al. Weather2k: A multivariate spatio-temporal benchmark dataset for meteorological forecasting based on real-time observation data from ground weather stations. *arXiv preprint arXiv:2302.10493* (2023).
23. Mesinger, F. & Arakawa, A. Numerical methods used in atmospheric model, volume 1. Tech. Rep., International Council of Scientific Unions-World Meteorological Organization (1976).
24. Lundquist, K. A., Chow, F. K. & Lundquist, J. K. An immersed boundary method for the weather research and forecasting model. *Monthly Weather Review* **138**, 796–817, <https://doi.org/10.1175/2009MWR2990.1> (2010).
25. Danielson, J. J. & Gesch, D. B. Global multi-resolution terrain elevation data 2010 (gmtd2010). Tech. Rep. (2011). [10.3133/ofr20111073](#). Report.
26. Survey, U. U. S. G. Usgs eros archive-digital elevation-shuttle radar topography mission (srtm) 1 arc-second global (2023). Accessed on October 2023.
27. Friedl, M. A., Gray, J. & Sulla-Menashe, D. Mcd12q2 modis/terra+aqua land cover dynamics yearly l3 global 500m sin grid version 6, [10.5067/MODIS/MCD12Q2.006](#) (2019). Accessed on 2025-06-04.
28. Souza, C. M. et al. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing* **12**, [10.3390/rs12172735](#) (2020).
29. Iacono, M. J. et al. Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal Geophysical Research: Atmospheres* **113**, <https://doi.org/10.1029/2008JD009944> (2008). <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD009944>.
30. Zheng, Y., Alapaty, K., Herwehe, J. A., Genio, A. D. D. & Niyogi, D. Improving high-resolution weather forecasts using the weather research and forecasting (wrf) model with an updated kain-fritsch scheme. *Monthly Weather Review* **144**, 833 – 860, <https://doi.org/10.1175/MWR-D-15-0005.1> (2016).
31. Morrison, H., Thompson, G. & Tatarskii, V. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review* **137**, 991 – 1007, <https://doi.org/10.1175/2008MWR2556.1> (2009).
32. JimÁñez, P. A. et al. A revised scheme for the wrf surface layer formulation. *Monthly Weather Review* **140**, 898 – 918, <https://doi.org/10.1175/MWR-D-11-00056.1> (2012).
33. Niu, G.-Y. et al. The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements. *Journal Geophysical Research: Atmospheres* **116**, <https://doi.org/10.1029/2010JD015139> (2011). <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD015139>.

34. Hong, S.-Y., Noh, Y. & Dudhia, J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review* **134**, 2318 – 2341, <https://doi.org/10.1175/MWR3199.1> (2006).
35. Dragaud, I. C. D. V. *Análise dos mecanismos físicos que governam o escoamento atmosférico na região costeira do estado do Rio de Janeiro*. Ph.D. thesis, Universidade Federal do Rio de Janeiro (2021).
36. Soares da Silva, M. *et al.* Assessment of meteorological settings on air quality modeling system-a proposal for un-sdg and regulatory studies in non-homogeneous regions in brazil. *Environmental Science Pollution Research* **30**, 1737–1760, [10.1007/s11356-022-22146-1](https://doi.org/10.1007/s11356-022-22146-1) (2023).
37. de Souza, L. S. *et al.* Evaluation of cumulus and microphysical parameterization schemes of the wrf model for precipitation prediction in the paraíba do sul river basin, southeastern brazil. *Pure Applied Geophysics* **181**, 679–700, [10.1007/s00024-023-03419-3](https://doi.org/10.1007/s00024-023-03419-3) (2024).
38. Pedruzzi, R. *et al.* Update of land use/land cover and soil texture for brazil: Impact on wrf modeling results over paraíba do sul river basin, sãõ paulo. *Atmospheric Environment* **268**, 118760, <https://doi.org/10.1016/j.atmosenv.2021.118760> (2022).
39. Jacinto, L. d. F. R. *Sistema Integrado de Modelagem Ambiental dos Fenômenos de Transporte na Área de Influência de Centrais Nucleares*. Master's thesis, Programa de Pós-Graduação em Meteorologia, PPGM, Universidade Federal do Rio de Janeiro (2020).
40. Shchepetkin, A. F. & McWilliams, J. C. The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling* **9**, 347–404 (2005).
41. MacFerrin, M., Amante, C., Carignan, K., Love, M. & Lim, E. The earth topography 2022 (etopo 2022) global dem dataset. *Earth System Science Data* **17**, 1835–1849, [10.5194/essd-17-1835-2025](https://doi.org/10.5194/essd-17-1835-2025) (2025).
42. Hersbach, H. *et al.* Era5 hourly data on single levels from 1940 to present, copernicus climate change service (c3s) climate data store (cds) [data set] (2023).
43. Jean-Michel, L. *et al.* The copernicus global 1/12° oceanic and sea ice glorys12 reanalysis. *Frontiers Earth Science* **9**, [10.3389/feart.2021.698876](https://doi.org/10.3389/feart.2021.698876) (2021).
44. Egbert, G. D., Bennett, A. F. & Foreman, M. G. G. Topex/poseidon tides estimated using a global inverse model. *Journal Geophysical Research: Oceans* **99**, 24821–24852, [10.1029/94JC01894](https://doi.org/10.1029/94JC01894) (1994).
45. Egbert, G. D. & Erofeeva, S. Y. Efficient inverse modeling of barotropic ocean tides. *Journal Atmospheric Oceanic Technology* **19**, 183–204 (2002).
46. Good, S. *et al.* The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing* **12**, [10.3390/rs12040720](https://doi.org/10.3390/rs12040720) (2020).
47. Dagestad, K.-F. *et al.* Wind retrieval from synthetic aperture radar-an overview. In *4th SAR Oceanography Workshop: Advances in SAR Oceanography* (European Space Agency, 2013).
48. Telea, A. An image inpainting technique based on the fast marching method. *Journal Graphics Tools* **9**, 23–34 (2004).
49. Bertalmio, M., Bertozzi, A. L. & Sapiro, G. Navier-stokes, fluid dynamics, and image and video inpainting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, I–I (IEEE, 2001).
50. Hersbach, H. Cmod5-an improved geophysical model function for ers c-band scatterometry. (2003).
51. Wackerman, C. C., Pichel, W. G. & Clemente-Colon, P. Automated estimation of wind vectors from sar. In *Proceedings of the 12th Conference on Interactions of the Sea and Atmosphere* (2003).
52. Bilgili, M., Alphan, H. & Ilhan, A. Potential visibility, growth, and technological innovation in offshore wind turbines installed in europe. *Environmental Science Pollution Research* **30**, 27208–27226 (2023).
53. Emeis, S. *Wind energy meteorology: atmospheric physics for wind power generation* (Springer, 2018).
54. Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT Press, 2012).

Acknowledgements

The authors would like to thank TotalEnergies for providing financial support for this project. We would also like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for providing scholarships to the authors.

Author contributions statement

Conceptualization: All authors. **Methodology:** All authors. **Data Curation:** Livia Sancho; Mauricio S. Silva; Elisa Passos; Larissa F. R. Jacinto; Rebeca S. Lyra; Nilton O. Moraes; Carina S. Bock; Douglas M. Nehme; Raquel Toste; Carlos H. Beisl; Patrícia M. Silva; Adriano O. Vasconcelos; Rian C. Ferreira. **Visualization:** Luan C. V. Silva; Patricia M. Silva; Larissa F. R. Jacinto; Livia Sancho; Mauricio S. Silva; Elisa Passos; Adriano O. Vasconcelos. **Writing – Original Draft Preparation:** Livia Sancho; Mauricio S. Silva; Elisa Passos; Luan C. V. Silva; Larissa F. R. Jacinto; Raquel Toste. **Validation:** Luan C. V. Silva; Rodrigo S. Luna; Mauricio S. Silva; Livia Sancho; Elisa Passos; Jacques Honigbaum; Fernando A. Rochinha; Alvaro L. G. A. Coutinho; Alexandre G. Evsukoff. **Writing – Review & Editing:** Luan C. V. Silva; Mauricio S. Silva; Livia Sancho; Elisa Passos; Alexandre G. Evsukoff; Luiz P. F. Assad. **Supervision:** Fernando A. Rochinha; Luiz P. F. Assad; Alvaro L. G. A. Coutinho; Laura S. B. S. Leite; Alexandre G. Evsukoff. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.