Data Analytics Capstone Topic

Interim Report

# **Multi‑Horizon Forecasting of Solar and Wind Energy Using Hybrid Deep Learning Models**

Walsh College

QM640 V1: Data Analytics Capstone

Raghvendra Dubey  
  
Mentor: Pralhad Teggi

Winter 2025 Term

# **GITHUB LINK**

https://github.com/raghav0410/Capstone\_QM640-V1-Data-Analytics-Capstone.git

## **Abstract**

Accurate forecasts of solar and wind generation help power systems keep supply and demand in balance. This interim report documents the scope, data, methods, exploratory data analysis (EDA), preliminary results, and next steps for a global, multi‑site project that produces 15‑minute, 1‑hour, and 24‑hour ahead forecasts using hybrid deep learning models and strong statistical baselines. Each processing step is explained in plain language—what was done, why it was done, what was expected, and what was observed—so a non‑specialist can follow the logic. The report integrates the uploaded notebooks, showing how raw multi‑source files were converted into a workable, model‑ready dataset. Sample‑size/power equations are included to justify that comparisons are statistically meaningful, and APA‑style references are provided to the literature that guided the approach.

Keywords: renewable energy, solar, wind, forecasting, multi‑horizon, deep learning, transfer learning, EDA, time series

## **1. Introduction**

Renewable generation is variable. Solar output follows the sun and cloud patterns; wind output fluctuates with boundary‑layer dynamics, storms, and terrain. Operators must plan minutes ahead (ramp control), hours ahead (unit commitment), and day‑ahead (market bids, storage scheduling). Much prior research emphasizes short‑term horizons (1–6 hours), which are valuable for operations but insufficient for day‑ahead planning. The present study addresses this gap by building a single pipeline that delivers forecasts at multiple horizons (15‑minute, 1‑hour, and 24‑hour) from the same cleaned dataset.

This work advances three practical goals. First, a reproducible global data pipeline is demonstrated: open meteorology and irradiance sources are harmonized into one tidy table per site and then merged into region‑level and global panels. Second, hybrid deep learning models (CNN‑LSTM/TCN) are compared against strong baselines (persistence, seasonal naïve, ARIMA/SARIMA, XGBoost) under time‑aware cross‑validation, so improvements are credible. Third, the study quantifies how accuracy degrades as the horizon lengthens (RQ1) and whether deep models outperform baselines (RQ2). Explanations are kept simple, linking steps to the expected outcome and recording what occurred.

Contributions. (i) A documented global multi‑site dataset build; (ii) a transparent EDA that reveals physical relations (e.g., solar↔GHI, wind↔hub‑height speed) by using daylight and non‑calm filters; (iii) a disciplined validation design with baseline comparators; and (iv) sample‑size/power formulas that translate “practical improvements” (e.g., 5% MAE reduction) into data requirements.

## **2. Scope and Objectives**

**Scope-** The dataset includes multiple global sites across Europe, the United States, Australia, and Asia, pulled from open sources (ERA5, NSRDB/PSM3 where available, PVGIS‑style irradiance, ENTSO‑E/OPSD‑style targets, and Global Renewables Watch for asset context). Targets include hourly solar and wind generation (site, regional, and country levels). Outputs differ by geography due to climate and terrain.

**Horizons and cadence-** Forecasts are produced for 15‑minute, 1‑hour, and 24‑hour ahead from an hourly base cadence; where sub‑hourly data exist, aggregation or down‑sampling is applied with care.

**Objectives-**  
O1. Measure how forecast accuracy changes with the horizon at single‑site and global multi‑site levels.  
O2. Test whether hybrid deep models beat persistence, seasonal naïve, ARIMA/SARIMA, and XGBoost across geographies.  
O3. Identify the most important drivers (irradiance, hub‑height wind vectors, pressure, cloud cover, seasonality) and check if they differ by location.  
O4. Assess model transferability across countries/continents via transfer learning (full results deferred to the final report; methods prepared here).

**Research Questions and Hypotheses (interim focus)**

**RQ1.** How does forecast accuracy change across time horizons (15‑minute, 1‑hour, 24‑hour)?  
**H0 (RQ1).** Mean error does not change with horizon: MAE\_15min = MAE\_1h = MAE\_24h (and similarly for RMSE/nMAE, within a small tolerance epsilon).  
**H1 (RQ1).** Mean error worsens as the horizon increases, with the expected ordering MAE\_15min < MAE\_1h < MAE\_24h; the largest deterioration is typically from 1‑hour to 24‑hour.  
**Tests.** Blocked time‑series cross‑validation; paired fold‑level comparisons of horizons using Wilcoxon and sign tests, with effect sizes and bootstrap CIs; power assessed per Section 9.

**RQ2.** Do hybrid deep learning models give better results than traditional models?  
**H0 (RQ2).** There is no improvement versus the strongest baseline (persistence or seasonal‑naïve/ARIMA as appropriate): for each site/fold/horizon, the paired difference Δ = MAE model − MAE\_best\_baseline has median ≥ 0.  
**H1 (RQ2).** Modern ML and hybrid deep models (e.g., CNN‑LSTM/TCN/TFT) achieve lower errors than the strongest baseline across sites and horizons: median Δ < 0, with a practically meaningful reduction (for example, at least 5% of baseline MAE).  
**Tests.** Blocked CV with a gap; paired Wilcoxon and sign tests across folds/sites; effect sizes; multiple‑comparison control (Holm); and power checks as in Section 9.

**RQ3.** Which input features impact forecasts the most and why? (Analysis to be reported in the final report, but hypotheses are fixed now.)  
**H0 (RQ3).** After controlling for other variables and seasonality, no candidate driver has materially greater predictive contribution than the rest: distributions of absolute SHAP values are equal across features and are not larger than those of noise/decoy features; ablating any single driver changes MAE by < δ (δ = 1% of baseline MAE).  
**H1 (RQ3).** For solar, irradiance (GHI/DNI/DHI) and cloud cover have significantly larger absolute SHAP values than other features; for wind, 100‑m wind speed (and air‑density/wind‑vector terms) dominate. Importance profiles differ by region/continent.  
**Tests**. SHAP‑value comparisons via permutation tests/Wilcoxon rank‑sum with Holm adjustment; bootstrap rank‑stability; ablation tests (drop‑one driver) assessing ΔMAE with δ = 1% threshold per horizon/site.

**RQ4.** Can the same model be used for both solar and wind forecasting across regions? (Evaluation deferred to the final report; hypotheses preregistered here.)  
**H0** (RQ4, non‑inferiority). A unified pooled/multi‑task model is inferior to the best resource‑specific/site‑specific model by more than δ: Δ = MAE unified − MAE specific ≥ δ, with δ = 2% (1‑hour) and 5% (24‑hour) of baseline MAE.  
**H1** (RQ4, non‑inferiority/superiority). The unified model is non‑inferior (Δ < δ) and may be superior (Δ < 0) for low‑data sites or in transfer settings; unified training reduces the transfer gap when moving across regions/continents.  
Tests. Paired TOST equivalence/non‑inferiority tests on Δ, complemented by Wilcoxon tests; analysis of learning curves (MAE vs. training size) to quantify sample‑efficiency gains; transfer‑gap diagnostics.

## **3 Literature Review**

• Mishra and Palanisamy (2018) proposed a unified recurrent neural network that produces multi‑time‑horizon solar forecasts within a single architecture rather than training a separate model per horizon. This supports the goal of learning across horizons to share information and reduce complexity and shows RNN‑type models can handle several steps ahead in parallel (Mishra & Palanisamy, 2018).

• Benti, Chaka, and Semie (2023) surveyed machine learning and deep learning approaches for renewable forecasting. They found DL models capture non‑linear relationships but face challenges with uncertainty, data gaps, and interpretability. This motivates the baseline comparisons and the use of explainability (e.g., SHAP for trees; attention/feature‑selection for sequences) (Benti et al., 2023).

• Effenberger and Ludwig (2022) cataloged open wind/wind‑power datasets and highlighted that performance depends on terrain and scale (turbine, farm, system). Their taxonomy informs public dataset selection for multi‑site tests and sets realistic error expectations across sites (Effenberger & Ludwig, 2022).

• Roseline et al. (2022) modeled hybrid renewable systems with ANNs and showed that neural approaches can capture cross‑resource interactions (solar+wind). This supports the hybrid deep models when combining solar and wind drivers (Roseline et al., 2022).

• Global Renewables Watch (Robinson et al., 2025) introduced a temporal, satellite‑AI dataset for mapping global solar PV and onshore wind assets. While not the primary target dataset, it shows a path to global coverage and helps validate siting/asset footprints for transferability studies (Robinson et al., 2025).

• Bayesian Deep Learning with α–β divergence (2022) developed an improved Bayesian BiLSTM for multi‑step‑ahead solar generation forecasting, yielding better probabilistic forecasts (tighter prediction intervals, more calibrated uncertainty) than standard variational methods. This informs the plan to report uncertainty for day‑ahead horizons (Bayesian BiLSTM, 2022).

• Simeunović et al. (2021) used spatio‑temporal graph neural networks (ST‑GNNs) for multi‑site PV power forecasting, demonstrating that capturing spatial dependencies among sites improves accuracy. This supports the global multi‑site framing and motivates spatial features (Simeunović et al., 2021).

• Huang and Deng (2023) proposed a spatiotemporal deep network for multi‑horizon wind prediction, fusing NWP/weather features with sequence models. Their design supports the use of sequence models (CNN‑LSTM/TCN) and multi‑horizon training for wind (Huang & Deng, 2023).

• Temporal Fusion Transformer (Lim et al., 2021) introduced an interpretable multi‑horizon architecture with variable selection and attention for static/known‑future/observed‑past inputs. This motivates attention‑based sequence baselines and variable‑selection ideas.

• XGBoost (Chen & Guestrin, 2016) is a strong tabular baseline with built‑in regularization and feature importance. It remains competitive in many energy forecasting tasks, so it is included as a credible non‑DL benchmark.

• N‑BEATS (Oreshkin et al., 2020) presented a pure deep learning architecture for univariate/multivariate time‑series forecasting that performs strongly without domain‑specific features. It motivates comparisons with generic DL sequence learners alongside physics‑informed features.

• Sample‑size methodology for prediction models (Whittle, 2025) extended power and sample‑size calculations for evaluating prediction models. This underpins Section 9, ensuring claimed improvements (e.g., 5% MAE reduction) have adequate statistical support.

Relevance. Collectively, these sources justify multi‑horizon sequence modeling, ensure comparison to credible baselines, highlight the value of spatial/multi‑site context, motivate uncertainty quantification for day‑ahead horizons, and provide the statistical framework (power/sample‑size) to judge practical improvements.

## **4. Data Description**

**Targets (generation).** Hourly solar and wind generation time series from open platforms (e.g., ENTSO‑E/OPSD‑style EU series, and comparable public sources elsewhere).

**Meteorology (drivers**). ERA‑type reanalysis (hourly): temperature (2 m), relative humidity, pressure, cloud cover, wind components at 10 m/100 m (and derived speed/direction), and short‑wave radiation terms (GHI/DNI/DHI where available).

**Irradiance (higher fidelity where possible)**. For U.S. sites, NSRDB/PSM3 provides high‑quality GHI/DNI/DHI; for other regions, ERA radiation and clear‑sky proxies are used.

**Time index.** UTC hourly index across all sources to avoid DST leakage.

**Spatial fields.** Site latitude/longitude (and region tags) for grouping.

*Rationale.* Physics‑informed features (irradiance → solar; hub‑height wind vectors/pressure → wind) and a clean, uniform hourly index are essential for fair multi‑model, multi‑horizon comparisons.

## **5. Analysis**

**From Raw Files to a Workable Dataset (step‑by‑step from notebooks)**

The following sequence mirrors the code paths in the uploaded notebook for data building (Refer Git-hub repository). Each step states what was done, why, the expected impact, and what was observed.

### 5.1 Build the master hourly index (UTC)

* What: Create a complete hourly DatetimeIndex from START to END in UTC, with explicit handling of daylight‑saving transitions.
* Why: All sources must align to identical timestamps for learning lagged patterns and avoiding leakage.
* Expected: One canonical time column, no gaps/overlaps.
* Observed: After alignment, small feature gaps are handled later (Section 6.5).

### 5.2 Fetch meteorology and irradiance per site

* What: Query ERA‑type reanalysis (Open‑Meteo ERA5 archive) for each site (lat/lon). For U.S. sites with credentials, request NSRDB/PSM3 GHI/DNI/DHI. Cache responses by site/year to disk.
* Why: Robust, repeatable pulls; NSRDB adds higher‑fidelity irradiance.
* Expected: Hourly frames with temperature, humidity, pressure, cloud cover, wind 10 m/100 m components, and radiation terms.
* Observed: Caching and retries reduced failures and sped up reruns; NSRDB is optional and gracefully skipped when credentials are absent.

### 5.3 Engineer physically meaningful features

* What: Derive wind speed = √(u²+v²) at 10 m and 100 m; direction as sine/cosine; air density from pressure and temperature; wind‑power density ∝ density×speed³. Create a clear‑sky index for solar (observed/clear‑sky proxy). Add calendar features (hour, day, month, day‑of‑week, season) plus cyclical encodings (sin/cos) for hour‑of‑day and day‑of‑year.
* Why: Turbines respond to hub‑height speed and density; solar follows daylight/season cycles; cyclical encodings avoid artificial breaks at 23→00 h and Dec→Jan.
* Expected: Higher signal in models and clearer relations in EDA.
* Observed: Correlation and feature‑ranking later confirmed GHI/cloud cover (solar) and 100 m wind speed/pressure (wind) as dominant drivers.

### 5.4 Standardize schema and merge per site

* What: Normalize column names (e.g., met\_ghi, met\_dni, met\_wind\_speed\_100m), left‑join all features to the target series on the UTC index, and write one clean CSV per site.
* Why: Downstream code expects a consistent schema and a tidy row‑per‑hour table per site.
* Expected: Wide per‑site tables with aligned features and targets.
* Observed: Early passes produced duplicate columns when features were engineered in stages; this was resolved with strict naming and a schema check (assert df.columns.is\_unique) before saving.

### 5.5 Handle missing values and outliers

* What: For feature gaps ≤ 2 consecutive hours, impute from short rolling means; do not impute target values (drop those rows when training to avoid leakage). Clip outliers to physical bounds (e.g., negative irradiance → 0; wind speed ≥ 0).
* Why: Models need valid numbers; outliers can dominate loss.
* Expected: More stable training and realistic errors.
* Observed: Variance decreased post‑clipping; validation error was more stable.

### 5.6 Created training matrices (supervised learning)

* What: Generate lagged features (t−1, t−3, t−6, t−24) and rolling means/SDs (6 h, 24 h), plus site/country categorical tags for pooled models.
* Why: Near‑term inertia and diurnal repetition boost short‑horizon skill; site tags allow pooled learning while keeping location context.
* Expected: Better 15‑minute and 1‑hour forecasts; day‑ahead benefits from daily lag.
* Observed: Clear gains versus no‑history models; tree feature importance ranked lag/rolls highly.

### 5.7 Save per‑region and global panels

* What: Concatenate per‑site tables into region‑level and global panels with a site\_id/region column.
* Why: Enables cross‑site EDA and pooled/transfer experiments.
* Expected/Observed: Smooth downstream aggregation and grouped diagnostics.

### 5.8. Exploratory Data Analysis (EDA)

EDA was designed to be physically interpretable and to avoid artifacts that can hide real relations. For solar, daylight hours only are analyzed; for wind, calm hours are excluded. This addresses the earlier “no relations” plots by removing zeros that flatten slopes.

A graph of a line

AI-generated content may be incorrect.

Figure 1a/1b. Solar generation (MW) — daylight only (histogram).  
*Explanation:* The distribution is right‑skewed with a large mass at low outputs and a long tail toward higher MW during clear‑sky hours. This shape reflects the diurnal ramp‑up and the frequent occurrence of partial cloud cover. The diminishing frequency beyond ~0.6–0.8 MW (per‑site scale) is consistent with thermal/inverter limitations and occasional clipping near capacity. Implication: evaluation should rely on robust metrics (MAE/nMAE) and, for linear baselines, consider mild variance‑stabilizing transforms.

Wind generation (MW) — values > 0.05 MW (histogram).  
*Explanation:* After removing near‑zero values (≤0.05 MW) to reduce calm‑wind inflation, the distribution shows a long lower tail and a pronounced spike near the rated region (≈2 MW in this example), suggesting capacity saturation/curtailment or rounding at turbine ratings. This pattern implies non‑Gaussian targets with potential censoring at the upper bound. Modeling implications include using MAE/Huber losses and allowing for non‑linear saturation (e.g., tree‑based or sequence models) rather than strictly linear assumptions.

A graph of a line and a line

AI-generated content may be incorrect.

Figure 2a/2b.

2a. Solar — average by hour (global, UTC).  
*Explanation:* The curve shows the diurnal cycle: a morning ramp, a midday plateau, and an afternoon decline. A slight broadening/shoulder is visible because profiles from multiple time zones are averaged on a common UTC axis, which smears local solar noon. This confirms strong time‑of‑day seasonality, supporting the inclusion of hour‑of‑day cyclical encodings (sin/cos) in models.

2b. Solar — average by month.  
*Explanation:* Monthly means peak in late spring to early summer and decline toward winter, consistent with daylength and solar elevation trends. This validates the use of day‑of‑year (or month) encodings and clear‑sky/irradiance features to capture seasonal structure.

A graph with a line

AI-generated content may be incorrect.

Figure 2c. Solar — diurnal profile: weekday vs. weekend.  
*Explanation:* The weekday and weekend curves nearly overlap, indicating solar generation is resource‑driven rather than calendar‑driven. Small differences are within natural variability and do not warrant separate calendar treatment beyond standard time features. This supports a single diurnal model for both weekday and weekend periods.

A graph of a graph and a graph of a graph

AI-generated content may be incorrect.

Figure 2d/2e.

2d. Wind — average by hour (global, UTC).  
Explanation: The curve exhibits a gentle morning minimum and a late‑afternoon/evening maximum, consistent with boundary‑layer growth and mixing that strengthen near‑surface winds later in the day. The peak appears broadened because multiple time zones are averaged on a common UTC axis. The pattern confirms a time‑of‑day dependency and supports hour‑of‑day cyclical encodings for wind models

2e. Wind — average by month.  
Explanation: Monthly means rise from late winter into mid–late summer, then decline into autumn, reflecting regional synoptic regimes (e.g., monsoon, trade‑wind seasons, storm tracks). The specific peak month varies by region, but the global average shows a summer maximum in this sample. This justifies including day‑of‑year/month features and pressure/temperature covariates linked to seasonal circulation.

A graph with blue and orange lines

AI-generated content may be incorrect.

Figure 2f. Wind — diurnal profile: weekday vs. weekend.  
Explanation: Weekday and weekend profiles nearly coincide, indicating minimal calendar effects on wind generation. Any small offsets fall within expected variability and do not motivate separate weekday/weekend handling. Standard diurnal features are sufficient.

A graph with colored lines and dots

AI-generated content may be incorrect.

Figure 2g. Solar — monthly by continent.  
Explanation: Monthly solar means display hemispheric seasonality. Europe (EU) and most of Asia (AS) peak in late spring to early summer and decline toward winter; Australia (AU, Southern Hemisphere) shows the inverse pattern, with a minimum around June–July and a maximum near December–January. The Americas curve is relatively high and broad across spring–summer, reflecting diverse latitudes. A mid‑summer dip in parts of Asia is consistent with monsoon cloudiness reducing irradiance. Implication: models benefit from continent/hemisphere tags and day‑of‑year encodings to capture these structural differences.

A graph with different colored lines

AI-generated content may be incorrect.

Figure 2h. Wind — diurnal by continent.  
Explanation: Diurnal wind signatures vary by continent. Asia shows a pronounced afternoon peak (boundary‑layer growth), Australia peaks in the late morning with a mid‑afternoon lull, Europe is comparatively flat with a modest evening uptick, and the Americas show a late‑afternoon to evening ramp. Because curves are averaged on a common UTC axis across time zones, local peaks are slightly smeared, but relative shapes persist. Implication: include continent/site tags and hour‑of‑day encodings; interaction terms (or pooled models with site embeddings) can capture regional boundary‑layer dynamics.

A graph of different colored lines

AI-generated content may be incorrect.

Figure 2i. Solar — diurnal by continent.  
*Explanation:* Hourly solar profiles are phase‑shifted across continents due to longitude and UTC alignment. Peaks occur at different UTC hours—earlier for Australia (Southern Hemisphere, eastern longitudes), mid‑day UTC for Asia/Europe, and later for the Americas—matching local solar noon. Amplitude and shoulder widths differ, reflecting cloud regimes and capacity mixes. Implication: include continent/hemisphere indicators and cyclical hour‑of‑day features; pooled models with site embeddings help learn these phase differences directly.

A graph with colored lines and numbers

AI-generated content may be incorrect.

Figure 2j. Wind — monthly by continent.  
*Explanation:* Monthly wind means show region‑specific seasonality rather than a simple hemispheric flip. In this sample, Europe exhibits a summer maximum with an autumn dip; Asia is comparatively stable around 1.03–1.06 MW; the Americas peak in late summer; and Australia peaks in late summer/early autumn with a mid‑year dip. These patterns reflect differing synoptic regimes (trade winds, monsoon, storm tracks) and topography. Implication: seasonal features should be interacted with continent/site tags, and models should allow non‑uniform seasonal effects across regions.

A graph of solar versus ghi

AI-generated content may be incorrect.

Figure 3a. Solar generation vs. Global Horizontal Irradiance (daylight).  
*Explanation:* The hexbin density shows a strong, near‑linear relationship between solar generation and GHI once restricted to daylight hours. The slight concavity at very high GHI values is consistent with practical effects such as thermal losses, angle‑of‑incidence, and inverter clipping. The narrow spread around the diagonal indicates that GHI is a dominant driver during clear‑sky periods, with dispersion increasing as cloud variability rises. This confirms the inclusion of GHI and cloud‑related features as primary predictors for solar targets.

A graph of a wind speed

AI-generated content may be incorrect.

Figure 3b. Wind generation vs. 100‑m wind speed (non‑calm).  
*Explanation:* A positive relationship is observed at low to moderate hub‑height wind speeds, consistent with turbine power increasing approximately with the cube of wind speed until the rated region. The long horizontal band at zero generation for very high wind speeds is consistent with turbine cut‑out behaviour (safety shutdown typically around 25 m/s) or with curtailment/data filtering at extremes. The narrow vertical streak near ~25–30 m/s suggests either capping/rounding in the wind‑speed feed or rare events clustered at the cut‑out threshold. For reporting and modeling, negative target values are clipped to 0, and extreme wind speeds above the 99.9th percentile are flagged as outliers; filtered plots are retained for the EDA appendix.

A graph of solar versus ghi

AI-generated content may be incorrect.

Figure 3c. Solar vs. GHI (all daylight).  
*Explanation:* This view includes all daylight observations with no outlier filtering. The dense diagonal band confirms a strong, approximately linear relation between solar generation and GHI. The wider dispersion at very high GHI reflects realistic effects such as transient cloud edges, thermal/inverter limits, and brief shading. A small set of high‑influence points is visible and motivates formal outlier screening before model fitting.

A graph of solar vs ghi

AI-generated content may be incorrect.

Figure 3d. Solar vs. GHI (filtered daylight, outliers removed).  
*Explanation:* After applying IQR/MAD and robust bivariate (Elliptic Envelope) filters, the scatter tightens around the main trend. The slope becomes slightly steeper, and the tail dispersion reduces, indicating a cleaner signal for learning. This filtered view is used for feature diagnostics and for training/evaluating models; the unfiltered view (Figure 3c) is retained for transparency and auditability.

A chart with different colored squares

AI-generated content may be incorrect.

Figure 4a. Correlation (Pearson, daylight).  
*Title used in figure:* “Correlation (Pearson, daylight)”.  
*Explanation:* Linear correlations during daylight show target\_solar\_mw strongly and positively associated with met\_ghi, met\_dni, and met\_dhi. met\_cloud\_cover is negatively associated with irradiance and solar output, as expected. Inter‑correlations among irradiance variables are very high, indicating multicollinearity; modeling should therefore avoid using GHI, DNI, and DHI simultaneously without regularization or dimensionality reduction (e.g., select GHI or use a clear‑sky index). Wind‑speed variables show weak correlations with solar, reflecting largely independent processes.

A chart with different colored squares

AI-generated content may be incorrect.

Figure 4b. Correlation (Spearman, daylight).  
*Title used in figure:* “Correlation (Spearman, daylight)”.  
*Explanation:* Rank‑based (monotonic) correlations are similar in structure but slightly more robust to outliers and non‑linearity. The strong monotonic relation between target\_solar\_mw and irradiance persists, while the negative association with met\_cloud\_cover remains evident. Spearman values reinforce the choice to emphasize irradiance and cloud features for solar forecasting.

A chart with different colored squares

AI-generated content may be incorrect.

Figure 4c. Correlation (Pearson, non‑calm).  
*Title used in figure:* “Correlation (Pearson, non‑calm)”.  
*Explanation:* In non‑calm wind regimes, target\_wind\_mw correlates most with met\_wind\_speed\_100m, more than with 10‑m speed, aligning with turbine hub‑height physics. met\_temp\_c often shows a mild negative correlation with wind power (thermal stratification effects), while met\_cloud\_cover contributes little directly. The high correlation between met\_wind\_speed\_10m and met\_wind\_speed\_100m suggests retaining only one (preferably 100 m) or using regularization to mitigate redundancy.

A colorful squares with white text

AI-generated content may be incorrect.

Figure 4d. Correlation (Spearman, non‑calm).  
*Title used in figure:* “Correlation (Spearman, non‑calm)”.  
*Explanation:* Rank‑based correlations confirm the dominance of hub‑height wind speed and the redundancy between 10‑m and 100‑m speeds. Spearman coefficients slightly increase where relationships are monotonic but not perfectly linear, supporting feature selection centered on met\_wind\_speed\_100m and pressure‑related variables, with optional exclusion of near‑collinear predictors.

A graph with a line and a red line

AI-generated content may be incorrect.

Figure 6a. Quantile–Quantile (QQ) plot — Solar (daylight).  
*Title used in figure:* “QQ: Solar (daylight)”.  
*Explanation:* The S‑shaped departure from the 45° line indicates non‑normal distribution of daylight solar output. Mass near zero at the lower tail reflects dawn/dusk periods; curvature in the upper tail is consistent with thermal/inverter limits and intermittent cloud edges. Implication: error distributions are skewed/heteroskedastic, so MAE (L1) and quantile losses are preferred over squared‑error alone. A mild variance‑stabilizing transform (e.g., sqrt or log1p of positive components) can be considered for linear models.

A graph with a line going up

AI-generated content may be incorrect.

Figure 6b. Quantile–Quantile (QQ) plot — Wind (non‑calm).  
*Explanation:* The step‑like pattern around low quantiles show heavy mass near small but non‑zero wind generation, followed by a rapid transition as turbines enter the power curve’s rising region. The flattening at high quantiles suggests censoring/curtailment near rated power. Implication: wind targets are also non‑Gaussian; robust losses (MAE/Huber), quantile objectives, and models that accommodate non‑linear saturation are appropriate.

A graph with a blue line

AI-generated content may be incorrect.

Figure 6c. Quantile–Quantile (QQ) plot — GHI (daylight).  
*Explanation:* Daylight GHI is positively skewed with a heavy right tail (clear‑sky, low‑air‑mass hours). The deviation from normality again cautions against assuming Gaussian errors. Standardization can use robust scalers (median/MAD), and tree/sequence models (XGB, CNN‑LSTM/TCN) naturally cope with this skew.



Table 1. Descriptive statistics for key variables by site/region.  
*Why:* Compare ranges and variability to set modeling expectations.

## 6. Modeling and Validation Design

Baselines to beat. Persistence (last value), Seasonal naïve (t−24 for day‑ahead), ARIMA/SARIMA.  
Tree model. XGBoost with a leakage‑safe ColumnTransformer (imputation+scaling for numeric; one‑hot for categoricals).  
Hybrid deep models. CNN‑LSTM (local patterns + memory) and TCN (dilated causal convolutions) for multi‑horizon sequences.  
Splits. Blocked time‑series cross‑validation with a gap before each test fold to avoid look‑ahead.  
Metrics. MAE, RMSE, nMAE (normalized by capacity if known); MAPE used with care due to zeros.

*Rationale.* Reviews show short‑horizon baselines are strong, deep models help as horizons lengthen, and fairness requires time‑aware splits. This setup provides a transparent, apples‑to‑apples comparison across horizons.

## 6.1. Sample‑Size and Power (methods reported and code status)

To ensure that observed improvements are not due to chance, two complementary calculations are used.

(A) Mean‑margin formula for descriptive goals (RQ1).  
Formula: n = (Z · sigma / E) ^2. Choose Z (e.g., 1.96 for 95% confidence), estimate sigma from validation errors, and set a practical margin E.

(B) Paired‑difference power for model comparisons (RQ2).  
Treat baseline vs. model errors as paired differences and compute the effective sample size needed to detect an absolute improvement delta with power 80% at alpha = 0.05:  
Formula: n\_req = ((z\_alpha + z\_beta) / delta) ^2 × (sd\_of\_paired\_differences) ^2.

(Empirical values from Clean\_EDA\_extended.ipynb file.)  
Bootstrap (24‑hour block) on the daylight solar target using a persistence baseline yielded:

* Baseline MAE ≈ 0.211; bootstrap sigma (MAE) ≈ 0.011.
* RQ1 (mean‑margin, 95% CI, ±10% of MAE):  
  n = (1.96 × 0.011 / (0.10 × 0.211)) ^2 ≈ 0.77 → 1 hourly point (effective).
* RQ2 (paired; detect 5% absolute MAE improvement, 80% power, two‑sided):  
  n\_req = ((1.96 + 0.84) × 0.011 / 0.01055) ^2 ≈ 8.5 → 9 paired observations (effective).

**Interpretation and usage.**  
The very small RQ1 requirement reflect a low bootstrap variance relative to a wide margin (±10%). In practice, evaluation uses blocked time‑series cross‑validation with hundreds of hourly points per fold, far exceeding this threshold. The RQ2 figure (≈9) provides a practical lower bound for detecting a 5% improvement under the observed variability; larger deltas reduce the requirement; smaller deltas increase it. Values should be recomputed per site/region/horizon as error variance changes.

*Status in code.* The notebook now prints these values directly; wind‑target and other‑horizon variants can be produced by switching masks/targets in the same cell and saving the resulting summary table for the report.

## 7. Preliminary Results (RQ1–RQ2 only)

Table 2. Cross‑validated metrics (MAE, RMSE, nMAE) by model and horizon — placeholder.  
*Purpose:* Quantify performance across 15‑min, 1‑h, 24‑h.  
*Observed trend:* Persistence is competitive at 15‑min; deep models show clearer gains at 24‑h; XGBoost often sits between ARIMA and DL.

A graph with lines and numbers

AI-generated content may be incorrect.

Figure 5a. Baseline RMSE vs. horizon — Solar (daylight filter).  
*Explanation:* The persistence baseline is strongest at 1‑hour (≈0.10 RMSE) but degrades at 3–6 hours (≈0.25–0.35) before improving slightly by 24‑hour (≈0.26). The seasonal‑naïve (t−24) baseline is relatively stable (≈0.24–0.27), beating persistence at 3–6 hours but slightly trailing it at 24‑hour.

A graph of a graph with blue and orange lines

AI-generated content may be incorrect.

Figure 5b. Baseline RMSE vs. horizon — Wind.  
*Explanation:* For wind, persistence is consistently better (lower RMSE) than seasonal‑naïve at all horizons shown. RMSE rises with horizon for both methods, from ≈ 0.67 (1‑hour) to ≈ 1.18 (24‑hour) for persistence, and from ≈ 1.19 (1‑hour) to ≈ 1.24–1.25 (6–24‑hour) for seasonal‑naïve. The gap narrows at longer horizons, indicating diminishing short‑term autocorrelation and increasing dependence on evolving weather states.

Table 2. Cross‑validated metrics and deltas vs. best baseline (fold‑level).  
Source: Research\_Question1&2 pipeline, blocked time‑series CV.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task | Horizon (h) | mean XGB RMSE | mean Best‑Baseline RMSE | ΔRMSE (improvement) | mean XGB MAE |
| Solar | 1 | 0.0316 | 0.0713 | 0.0396 | 0.0161 |
| Solar | 24 | 0.0634 | 0.0737 | 0.0103 | 0.0321 |
| Wind | 1 | 0.5809 | 0.6744 | 0.0935 | 0.3917 |
| Wind | 24 | 0.8951 | 1.1752 | 0.2801 | 0.8312 |

Inference. Across both resources and horizons, XGBoost improves RMSE relative to the strongest baseline. Gains are largest for wind (≈0.09–0.28 absolute RMSE), and smaller but consistent for solar (≈0.01–0.04). This addresses RQ2 in the affirmative at the fold level.

Table 2A. Statistical significance (fold‑level).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Task | Horizon (h) | Pairs | Win rate | ΔRMSE | Wilcoxon p | Sign p | Effect size (d) |
| Solar | 1 | 5 | 1.00 | 0.0396 | 0.0312 | 0.0312 | 7.66 |
| Solar | 24 | 5 | 1.00 | 0.0103 | 0.0312 | 0.0312 | 2.21 |
| Wind | 1 | 5 | 1.00 | 0.0935 | 0.0312 | 0.0312 | 12.05 |
| Wind | 24 | 5 | 1.00 | 0.2801 | 0.0312 | 0.0312 | 17.45 |

Inference. Even with a small number of folds, improvements are statistically significant (Wilcoxon and sign tests, α=0.05) with very large effect sizes.

Table 2B. Site×fold significance for wind (global multi‑site).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Task | Horizon (h) | Pairs (site×fold) | Win rate | ΔRMSE | Wilcoxon p | Sign p | Effect size (d) |
| Wind | 1 | 770 | 0.997 | 0.0920 | 5.52×10⁻¹²⁸ | 4.78×10⁻²²⁷ | 3.53 |
| Wind | 24 | 770 | 0.996 | 0.2755 | 5.65×10⁻¹²⁸ | 1.23×10⁻²²⁴ | 4.23 |

Inference. Using site×fold pairs provide overwhelming evidence that the model beats the best baseline for wind at both horizons; effects remain large after multiple‑comparison adjustment.

Table 3. Pooled vs. Site‑only models (RQ2).  
Metric: gain in RMSE (site‑only − pooled); positive → pooled better.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Task | Horizon (h) | Pairs | Win rate | Mean gain (RMSE) | Wilcoxon p | Sign p |
| Solar | 1 | 40 | 0.175 | −0.0022 | 0.9999 | 1.0 |
| Solar | 24 | 40 | 0.725 | 0.0048 | 0.000124 | 0.0032 |
| Wind | 1 | 40 | 0.750 | 0.0344 | 0.000190 | 0.0011 |
| Wind | 24 | 40 | 0.850 | 0.0706 | 0.000001 | 0.000004 |

Inference. For day‑ahead solar, pooled training (multi‑site) offers a statistically significant advantage over site‑only. For wind, pooled learning is consistently superior at both 1‑hour and 24‑hour horizons. For 1‑hour solar, site‑only slightly edges out pooled (non‑significant), suggesting very short horizons are dominated by fine‑grained local effects.

*Plain‑language takeaway*. Short horizons favour persistence and site‑specific patterns; by day‑ahead, models benefit from broader context and richer features. The cross‑validated tables show that a well‑tuned tree model already exceeds strong baselines; hybrid deep models should, at a minimum, match these gains and further improve transferability (to be evaluated in the final report).

## 11. References (APA 7th)

Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and prospects. *Sustainability, 15*(9), 7087. https://doi.org/10.3390/su15097087

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Effenberger, N., & Ludwig, N. (2022). A collection and categorization of open‑source wind and wind power datasets. *arXiv preprint* arXiv:2202.08524.

Global Renewables Watch (Robinson, C., et al.). (2025). *A temporal dataset of solar and wind*. *arXiv preprint* arXiv:2503.14860.

Huang, F., & Deng, Y. (2023). A spatiotemporal deep neural network for fine‑grained multi‑horizon wind prediction. *arXiv preprint* arXiv:2309.04733.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi‑horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Mishra, S., & Palanisamy, P. (2018). Multi‑time‑horizon solar forecasting using recurrent neural networks. *arXiv preprint* arXiv:1807.05459.

Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2020). N‑BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint* arXiv:1905.10437.

Roseline, J. F., Dhanya, D., Selvan, S., Yuvaraj, M., Duraipandy, P., Kumar, S. S., Prasad, A. R., Sathyamurthy, R., & Mohanavel, V. (2022). Neural network modelling for prediction of energy in hybrid renewable energy systems. *Energy Reports, 8*, 999–1008. https://doi.org/10.1016/j.egyr.2022.10.284

Simeunović, J., Schubnel, B., Alet, P.‑J., & Carrillo, R. E. (2021). Spatio‑temporal graph neural networks for multi‑site PV power forecasting. *arXiv preprint* arXiv:2107.13875.

Whittle, R. (2025). Extended sample size calculations for evaluation of prediction models. *BMC Medical Research Methodology*. (Advance online publication).

Additional source used for uncertainty modelling:  
*Bayesian deep learning technique with an advanced divergence*. (2022). (Unpublished manuscript/white paper).

## 12. Appendix A — Notebook → Report Map (for reproducibility)

 data**\_building\_capstone.ipynb** — Sections 5–6: time index, API pulls, feature engineering, per‑site merges, region/global panels.

 Clean**\_EDA.ipynb** — Section 7: histograms, seasonal slices, daylight/calm filters, scatter relations, correlation heatmaps, descriptive Table 1.

 Clean**\_EDA\_extended.ipynb** — Section 7: enhanced EDA with daylight/non‑calm masks; hexbin scatter (optional LOWESS); outlier detection (IQR, MAD, robust bivariate, Isolation Forest); Pearson/Spearman heatmaps; QQ plots; explicit sample‑size & power helpers (Cell 7); and figure exports used in Section 7.

 Research**\_Question1&2.ipynb** — Sections 8 & 10: baselines, XGBoost, CNN‑LSTM/TCN across horizons; time‑aware blocked CV; summary deltas.

 Modelling**, \_Sample\_size\_calculation\_and\_answering\_the\_research\_question. ipynb** — Section 9: paired difference power for RQ2; (optionally) simple n for RQ1.

## 13. Appendix B — Figure & Table Placeholders

|  |  |
| --- | --- |
| Figure/Table | Description |
| Figure 1a | Solar generation (MW) daylight only (histogram) |
| Figure 1b | Wind generation (MW) values > 0.05 MW (histogram) |
| Figure 2 | Seasonal time series slices (one summer week vs. one winter week) |
| Figure 2a | Solar — average by hour (global, UTC) |
| Figure 2b | Solar — average by month |
| Figure 2c | Solar — diurnal profile: weekday vs. weekend |
| Figure 2d | Wind — average by hour (global, UTC) |
| Figure 2e | Wind — average by month |
| Figure 2f | Wind — diurnal profile: weekday vs. weekend |
| Figure 2g | Solar — monthly by continent |
| Figure 2h | Wind — diurnal by continent |
| Figure 2i | Solar — diurnal by continent |
| Figure 2j | Wind — monthly by continent |
| Figure 3a | Solar generation vs. Global Horizontal Irradiance (daylight) |
| Figure 3b | Wind generation vs. 100m wind speed (non-calm) |
| Figure 3c | Solar vs. GHI (all daylight) |
| Figure 3d | Solar vs. GHI (filtered daylight, outliers removed) |
| Figure 4a | Correlation (Pearson, daylight) |
| Figure 4b | Correlation (Spearman, daylight) |
| Figure 4c | Correlation (Pearson, non-calm) |
| Figure 4d | Correlation (Spearman, non-calm) |
| Figure 5a | Baseline RMSE vs. horizon — Solar (daylight filter) |
| Figure 5b | Baseline RMSE vs. horizon — Wind |
| Figure 6a | QQ: Solar (daylight) |
| Figure 6b | QQ: Wind (non-calm) |
| Figure 6c | QQ: GHI (daylight) |
| Table 1 | Descriptive statistics by site/region |
| Table 2 | Cross-validated MAE/RMSE/nMAE by model and horizon |