Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Day-Ahead Forecasting of Sunshine and Wind for a Dust- and Heat- Prone Hyper-Arid Mini-Grid: Machine Learning against Persistence and SARIMA at Agadez, Bilma and N'Guigmi, Niger (2015-2024) A real-data benchmark with regime-aware features, cross-site transfer and conformal calibration

Mohamed Moumouni Guero

mohamed.moumouni@uddm.edu.ne

POLYTECH Maradi, University of Maradi

Research Article

Keywords: solar resource forecasting, wind forecasting, gradient boosting, SARIMA, quantile regression, conformal prediction, transfer learning, conditional value at risk, mini-grid operations, Sahel

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11245814/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Day-Ahead Forecasting of Sunshine and Wind for a Dust- and Heat-Prone Hyper-Arid Mini-Grid: Machine Learning against Persistence and SARIMA at Agadez, Bilma and N'Guigmi, Niger (2015-2024)

A real-data benchmark with regime-aware features, cross-site transfer and conformal calibration

Moumouni Guero Mohamed

[1]Department of Mechanical and Energy Engineering, POLYTECH Maradi, Dan Dicko Dankoulodo University of Maradi, Maradi, Niger

[2]Innovatech Smart Energies, Maradi, Niger

Grammarly and ChatGPT were used for language editing and reference formatting. During revision, Anthropic's Claude was used to restructure the manuscript to the journal's format and to partially check the references. The author conceived the study, analyzed and interpreted the data, verified all content and figures, and takes full responsibility for the manuscript.

*Corresponding author. E-mail : mohamed.moumouni@uddm.edu.ne. ORCID: 0000-0002-4183-8972

Abstract

Sizing a mini-grid well still leaves open how to run it the next morning. Using the quality-controlled 2015-2024 record of three Nigerien stations (Agadez, Bilma, N'Guigmi), we forecast daily sunshine one to seven days ahead and 3-hourly wind speed three to 72 hours ahead with gradient boosting (HGB) and a compact neural network against persistence and SARIMA, training on 2015-2022, testing on 2023-2024, and confirming the comparison over six rolling years (2019-2024). Both beat persistence by a wide margin (RMSE skill 18-49 percent for wind, 15-30 percent for sunshine), but the margin over SARIMA is thin and site-dependent: HGB beats SARIMA for wind at every lead at Agadez in 2023-2024, though it narrows over the rolling years, is most consistent at Bilma, and is not significant at N'Guigmi. Heat-stress and wind-regime

features change RMSE by about 1 percent or less for wind and up to 5.4 percent for sunshine, short of the hypothesised Harmattan gain; cross-site pooling lowers Agadez's 24-h wind RMSE by 3.1 percent when data are scarce. Raw quantile forecasts at Agadez are miscalibrated (67-72 percent coverage for nominal 80 percent intervals); conformal recalibration improves this to 76-82 percent but drifts across years. For an illustrative 100 kWp PV plus 30 kW wind plant, a better forecast lowers the CVaR95 battery buffer at Agadez from 342 to 235 kWh, narrowed further by propagating wind quantiles through the power curve. The gains are real but modest, site-dependent, and the honest measure of what an operator can count on tomorrow.

Keywords: solar resource forecasting; wind forecasting; gradient boosting; SARIMA; quantile regression; conformal prediction; transfer learning; conditional value at risk; mini-grid operations; Sahel

1. Introduction

Prior technical design and sizing studies for hybrid photovoltaic-wind mini-grids in Niger have established how to size the generation and storage capacity for a given site [1], but sizing a plant well still leaves open how to run it from one day to the next, the question this article addresses. Deep learning has become the default tool for short-term renewable power forecasting almost everywhere solar and wind capacity has grown fast enough to need one. Hybrid convolutional-recurrent architectures in particular have repeatedly outperformed single-family models, from direct normal irradiance forecasting in Ghardaia, on the Algerian side of the Sahara, to hourly photovoltaic forecasting for Johannesburg [2,3]. More recently, a cross-location comparison of temporal-fusion transformers and LSTM networks has benchmarked long-horizon solar forecasting across several arid sites [4]. Over the past two years the field has moved toward attention-based temporal modelling, and a 2026 review of transformer-based renewable energy forecasting credits attention with better handling of long-range dependencies than recurrent networks alone, once training data are sufficient [5,6]. A parallel strand of the literature has moved past point forecasts toward quantile and other probabilistic formulations, on the reasonable grounds that a grid operator needs to know how wrong a forecast is likely to be, not only what it expects to happen [7,8].

As far as we know, none of that progress has been demonstrated in a Sahelian hyper-arid setting dominated by dust and heat, the setting studied here. West African applications remain concentrated in Nigeria and use largely conventional models: support vector machines and standard neural networks across Nigerian climatic regions, and a geospatially enhanced neural network paired with gradient-boosted trees for solar radiation, neither with dust- or heat-specific inputs [9,10]. Reviews of machine-learning wind forecasting make a similar point: most published gains come from data-rich mid-latitude sites [11], a pattern also visible in decomposition-based multi-step wind-power forecasting elsewhere [12]. The gap matters because two well-documented mechanisms are likely to degrade a naive forecast here, yet they are rarely evaluated together in this setting. One is photovoltaic thermal derating, which rises with the same heat that drives cooling demand. The other is dust dimming and soiling, which follow a wind-regime structure and which long-term North African and Niamey studies show can remove a measurable share of photovoltaic output [13,14].

The second gap concerns data volume and, as it turned out, the observing schedule itself. A decade of daily or three-hourly station observations is a small dataset by deep-learning standards, and reviews of transfer learning for wind and photovoltaic forecasting identify data scarcity at under-instrumented sites as a central open problem [15,16]. Pretraining on a data-rich source and fine-tuning on a data-scarce target has produced measurable gains in offshore wind, in hybrid temporal-convolutional and attention models, in daily solar plant forecasting and in open frameworks such as SolNet [17,18,19,20,21,22]. The three stations studied here are a natural test case: N'Guigmi's wind record is nearly complete (97.6 % of 3-h slots), Bilma's is intermediate (89.4 %) and Agadez's is the most gap-affected (63.3 %). One of the first findings of this article is that Agadez's gaps are not random. The station reports almost no wind at 21, 00 and 03 h local time, so its wind record is a daytime record, not a sparse one.

This article makes five contributions, all computed on the measured record. First, it benchmarks two machine-learning models that were actually trained, histogram gradient boosting and a compact multilayer perceptron, against persistence and SARIMA for every station and horizon, and it reports the horizons and sites where the learned models lose. Second, it engineers heat and wind-regime features from established heat-stress and wind-regime physics and measures their contribution with an ablation stratified by regime, with bootstrap

confidence intervals. Third, it tests cross-site transfer to Agadez under three data-availability scenarios. Fourth, it treats uncertainty as an output: quantile forecasts are checked for calibration, recalibrated with split-conformal prediction where needed, and the day-ahead forecast miss is converted into a CVaR95-based battery buffer using the PV and wind energy models of Section 3.8. Fifth, it asks how much of this survives a stricter evaluation: the model comparison is repeated over six rolling test years with block-bootstrap confidence intervals and Diebold-Mariano tests, conformal coverage is checked by sub-period, and the wind quantiles, rather than the point forecast alone, are propagated through the power curve.

Three hypotheses were fixed before the analysis. H1: the learned models achieve positive skill against both persistence and SARIMA at every horizon from 3 to 72 h, with the margin narrowing as the horizon lengthens. H2: heat and wind-regime features reduce forecast error more during Harmattan dust-prone conditions than during background conditions. H3: cross-site pooling or fine-tuning reduces forecast error at Agadez relative to a model trained on Agadez alone, with the largest gain in the most data-scarce scenario. Section 4 answers each hypothesis with the measured data, and not all of the answers were anticipated.

2. Study area and data

The three stations are Agadez Aéro (16.97 °N, 7.99 °E, about 502 m), Bilma (18.69 °N, 12.92 °E, about 357 m) and N'Guigmi (14.25 °N, 13.11 °E, about 279 m), spanning the transition from the Aïr foothills to the Kawar oasis and the Lake Chad basin. The only data source is the station file used for this study, covering 1 January 2015 to 31 December 2024. It contains daily maximum and minimum temperature (TMAX, TMIN), daily maximum and minimum relative humidity (UMAX, UMIN), daily sunshine duration n (hours), station pressure and 3-hourly wind direction and speed at 00, 03, ..., 21 h local time (UTC+1). The same quality-control rules are applied at all three stations: physical bounds; TMIN > TMAX and UMIN > UMAX set to missing (three cases: Agadez on 9 June and 27 September 2024, Bilma on 27 May 2024); and sunshine longer than the astronomical day length N set to missing (Agadez on 5 March 2022, 40.3 h against N = 11.73 h; Bilma on 31 July 2015, 13.9 h against 12.85 h). Pettitt and SNHT tests on annual means detect no break

significant at 5 % for temperature, sunshine or wind. There is a borderline signal at Agadez around 2018-2019 for TMAX and sunshine ($p \approx 0.07$-0.09), which we note but do not treat as a confirmed inhomogeneity. One further feature of the sunshine record matters for forecasting. Agadez reports exactly 0.0 h of sunshine on 179 days (Bilma 31, N'Guigmi 30). At Agadez 131 of these days fall in May-July, and at all three stations they come with higher afternoon humidity and a smaller diurnal temperature range than average, as expected for days when dense dust haze or convective cloud keeps direct irradiance below the 120 W m^{-2} sunshine threshold. They are therefore kept. Their frequency at Agadez is uneven, however (46 days in 2015 and 40 in 2017 against 3 in 2019), and this coincides with the borderline 2018-2019 signal mentioned above, so a recording change cannot be excluded.

The file contains no pyranometer irradiance, no aerosol optical depth, no visibility and no measured power output. The two forecast targets are therefore the two resource variables measured directly: daily sunshine duration, as the solar proxy, and 3-hourly 10-m wind speed. Wind speed is reported in whole units and its unit is not documented in the file; following regional synoptic practice it is treated as m/s, an assumption revisited in Section 6. Wherever a result is expressed in kWh (Section 3.8, Figure 14), the energy is derived from these observations through the physical models described in Section 3.8. It is a modelled production figure, not a measured one.

Figure 1 shows the structure of the record that shapes the forecasting problem. Sunshine is 91.2 % complete at Agadez, 98.1 % at Bilma and 96.5 % at N'Guigmi. Wind is 63.3 %, 89.4 % and 97.6 % complete, but the missing slots are concentrated in specific hours. Agadez has 18 reports at 00 h, 108 at 03 h and 225 at 21 h in ten years, against more than 3,500 in each daytime slot. Bilma lacks most 00 h reports. The two targets also differ strongly between stations (Figure 2, Table 1). Mean sunshine is 8.0 h/day at Agadez, 9.1 h/day at Bilma and 8.9 h/day at N'Guigmi, with the minimum in the monsoon months at Agadez (June, 5.4 h/day) and N'Guigmi (August, 7.6 h/day). Bilma, which the monsoon barely reaches, has its minimum in December (8.3 h/day). All three stations peak in November (9.7-10.2 h/day). Mean reported wind is 4.84 at Agadez, 1.77 at Bilma and 1.08 at N'Guigmi. Calm reports (speed 0) make up 2.9 %, 40.8 % and 56.8 % of observations respectively. At every station wind peaks in the late morning (09-12 h), when the nocturnal inversion breaks

down and momentum is mixed down from the low-level jet, consistent with an earlier correlation analysis of the meteorological drivers of the N'Guigmi wind regime [23].

Models are trained on 2015-2022 (eight years) and tested on 2023-2024 (two years). The split is strictly chronological: a training sample must have its target time on or before 31 December 2022, and a test sample must have its forecast origin on or after 1 January 2023, so no future information reaches the training set. All tuning (boosting iterations, network epochs, regime thresholds, climatologies) uses an inner split, fitting on 2015-2020 and validating on 2021-2022. The test years were never used for any choice. Because two test years cannot separate a model difference from inter-annual variability, the comparison with the statistical references is repeated with a rolling origin over 2019-2024 (Section 3.9).

All input series of this study were built from the station file within one processing chain, and every step that depends on the data is fitted on training years only. Wind speed and direction are placed on a regular 3-hourly grid of 29,224 slots per station; missing slots are left missing, never interpolated, and the models handle them either natively (HGB, SARIMA) or by training-period median imputation with missingness indicators (MLP). Daily wind aggregates (mean, maximum, 06-18 h mean, north-east share and share of reports at or above 6 m/s) are computed from whichever slots of the day are present. Sunshine duration is used as reported after the quality control described above; no gap is filled and no irradiance series is substituted for it. Day length and extraterrestrial radiation follow FAO-56 [24]. The derived heat and regime features of Section 3.2, the sunshine and wind climatologies and the regime thresholds are estimated on the training years of each experiment, and in the rolling-origin evaluation they are re-estimated for every fold. Station pressure, which is 77.3 % complete at N'Guigmi, enters the models with its gaps and is filled with the training-period monthly mean only in the air-density term of the energy model (Section 3.8). What was removed, reconstructed or assumed for each series is summarised in Table A1 (Appendix A).

Table 1. Station record used in this article (2015-2024). Completeness is the share of expected values present after quality control. Sample counts are for the sunshine 1-day and wind 24-h models (training 2015-2022 / scored test forecasts 2023-2024).

Station	Location	Sunshine	Wind	Mean	Mean wind	Calm	Sunshine n	Wind n

		complete (%)	complete (%)	sunshine (h/day)	(m/s)	reports (%)	train/test	train/test
Agadez	16.97 N, 7.99 E, 502 m	91.2	63.3	8.00	4.84	2.9	2603 / 724	14837 / 3633
Bilma	18.69 N, 12.92 E, 357 m	98.1	89.4	9.15	1.77	40.8	2886 / 694	21125 / 4954
N'Guigmi	14.25 N, 13.11 E, 279 m	96.5	97.6	8.94	1.08	56.8	2826 / 694	22916 / 5587

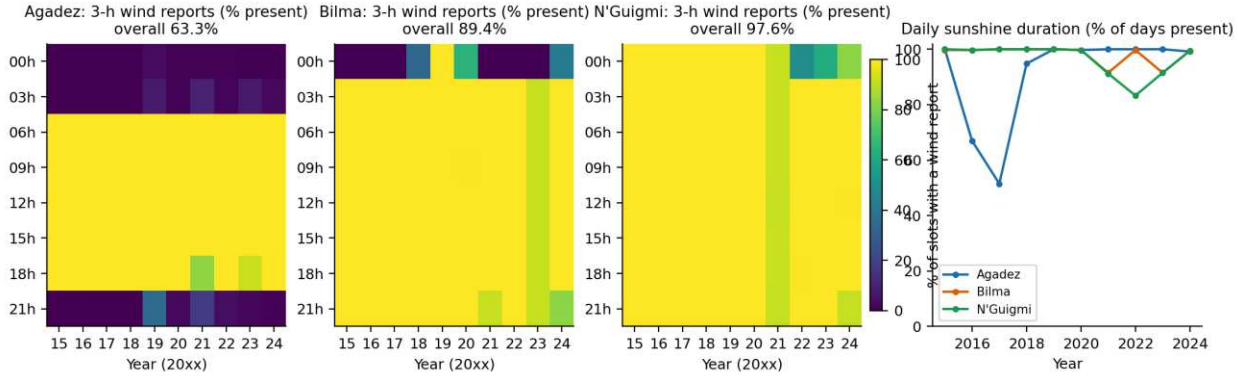


Figure 1. Data availability in the station file, 2015-2024. Left three panels: percentage of 3-hourly slots with a wind report, by local hour and year, for Agadez (overall 63.3 %), Bilma (89.4 %) and N'Guigmi (97.6 %). Right: percentage of days with a sunshine-duration value, by year. n = 29,224 expected 3-h slots and 3,653 expected days per station.

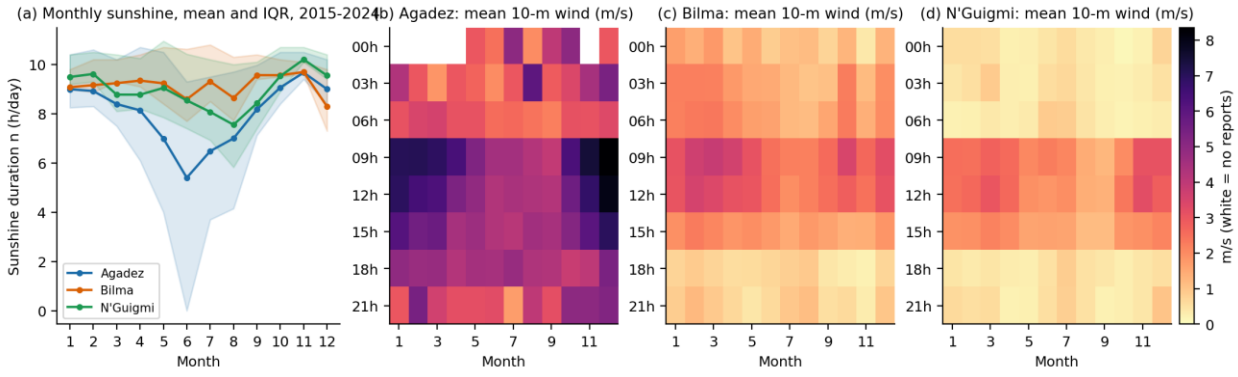


Figure 2. Climatology of the two forecast targets, 2015-2024. (a) Monthly mean daily sunshine duration with inter-quartile range (shaded), all three stations (n = 3,332 to 3,585 days per station). (b-d) Mean reported 10-m wind speed by month and 3-hourly local slot; white cells have no reports (Agadez night-time slots).

3. Methods

3.1 Forecast targets, horizons and problem formulation

For each station, target and lead time h, a separate model maps information available at the forecast origin t to the target value at t + h (the direct multi-horizon strategy [25]):

$$\hat{y}(t+h) = f_h(x(t);\ \theta_h), \quad h \in \{1, \dots, 7\} \text{ days (sunshine)}, \quad h \in \{1, 2, \dots, 24\} \text{ three-hour steps (wind)} \quad (1)$$

The wind leads computed are 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60 and 72 h. The direct strategy avoids feeding a model its own errors, which would be awkward at Agadez, where recursive forecasts would have to run through night slots that are almost never observed. Wind forecasts are issued from every 3-h origin in the test period and scored only where the target slot was observed, so Agadez wind skill is effectively daytime skill.

3.2 Base and derived meteorological features

The feature vector x(t) is built only from information observed at or before t. Daily covariates enter the 3-hourly wind model with a one-day lag (day d - 1 for any origin on day d), because TMAX, UMIN or the day's mean wind are not complete until the day has ended. Features are grouped as in Table 2. The base set contains lagged values of the target (the last seven days for sunshine; the last 24 h, plus 48 h, 72 h and the same slot on previous days, for wind), rolling means and spread, calendar and astronomical terms (day-of-year harmonics, the target slot's hour, day length N and extraterrestrial radiation Ra from FAO-56 [24], and a training-period climatology of the target), and conventional meteorology (TMAX, TMIN, UMAX, UMIN, diurnal temperature range, station pressure and its day-to-day change).

The derived set adds two physically motivated feature groups. The heat group follows standard heat-stress physics for this climate. Afternoon wet-bulb temperature is computed from TMAX and UMIN with the Stull approximation [26],

$$T_w = T \cdot \operatorname{atan}[0.151977(RH + 8.313659)^{0.5}] + \operatorname{atan}(T + RH) - \operatorname{atan}(RH - 1.676331) + 0.00391838\, RH^{1.5} \cdot \operatorname{atan}(0.023101\, RH) - 4.686035 \quad (2)$$

together with the simplified WBGT approximation used operationally by the Australian Bureau of Meteorology [27],

$$\mathrm{WBGT} \approx 0.567\,T + 0.393\,e + 3.94, \quad e = (RH/100) \cdot 6.105 \cdot \exp[17.27\,T/(237.7 + T)] \ \text{(hPa)} \quad (3)$$

which is known to drift from physically based WBGT in very hot conditions and is used here only as a predictor, not as an exposure estimate [28]. The group also contains a near-dawn vapour pressure (from TMIN and UMAX, a tracer of monsoon moisture), a hot-day flag (TMAX above the calendar-month 90th percentile of the training years) and a heat-wave flag (three or more consecutive hot days). The wind-regime group combines a circular encoding of wind direction with the threshold-wind reasoning used for dust mobilisation. Direction is encoded as unit-vector components rather than a scalar bearing,

$$u = \sin\theta, \quad v = \cos\theta \qquad (4)$$
(θ measured clockwise from north, direction the wind blows from; $u = v = 0$ for calm)

and the group adds the daily share of reports from the north-east quadrant (0-90°) and the daily share of reports at or above 6 m/s. The 6 m/s value is a nominal 10-m uplift threshold at the low end of the range derived for Saharan sources [29]. It is a proxy for dust mobilisation, not a measurement of it. Finally, each day is assigned a regime by k-means clustering (fitted on the training years) of the daily mean direction components and the near-dawn vapour pressure. Aerosol optical depth, which would refine this regime classification, is not available in this file. The silhouette score is highest for k = 2 at all three stations (0.52, 0.37 and 0.54 for Agadez, Bilma and N'Guigmi, against 0.46-0.49, 0.28-0.33 and 0.49-0.52 for k = 3-4). The drier, north-easterly cluster is labelled Harmattan-type, and its days are split into strong- and weak-wind days at the station-specific 75th percentile of training-period daily mean wind (6.8, 2.9 and 1.4 at Agadez, Bilma and N'Guigmi). The Harmattan-type share of days is above 90 % from November to February at all stations and falls to 0-12 % in August, which is the physically expected seasonal cycle. These regime labels enter the models only as origin-day features; target-day labels are used only to stratify errors in the ablation (Section 4.5). This weather-feature-engineering approach follows the broader practice of integrating station meteorology directly into learned energy-forecasting models [30].

Table 2. Feature groups. "Base" groups form the reference feature set of the ablation; "derived" groups are added in the full set. Daily covariates enter the wind model with a one-day lag.

Group	Content (sunshine model / wind model)	Set
Target lags	n at days t..t-6, n/N, anomaly / wind at 0-24 h, last report and its age	Base
Older lags, same-slot history	- / wind at 48 h and 72 h, same slot on the previous one and two days	Base
Rolling statistics	3-, 7-, 14-, 30-day means and 7-day SD / 24-h and 72-h means, 24-h maximum	Base
Calendar and astronomy	day-of-year harmonics, month, wet-season flag, N and Ra of target day, climatology / target and origin hour harmonics, day of year, wet-season flag, month-by-slot climatology	Base
Conventional meteorology	TMAX, TMIN, UMAX, UMIN, DTR (days t, t-1), pressure and its change	Base
Heat indices	Stull wet-bulb at TMAX/UMIN, simplified WBGT, near-dawn vapour pressure, hot-day and heat-wave flags	Derived
Wind-direction and regime proxies	daily mean, max and daytime wind, direction vector (u, v), NE-sector share, share of reports >= 6 m/s, k-means Harmattan flag, strong-Harmattan flag / plus u, v at lags 0-2 and NE flag of the latest report	Derived

3.3 Reference models

Two references are mandatory for every result. For wind, persistence is the last observation no older than 24 h. Because wind has a strong diurnal cycle, a diurnal persistence (the same slot on the most recent day available at the origin) is also reported. For sunshine, the analogue of clear-sky-index persistence is relative-sunshine persistence, which keeps the ratio n/N and rescales it to the target day's astronomical day length:

$$\hat{n}(t+h) = \left[\frac{n(t)}{N(t)}\right] \cdot N(t+h) \tag{5}$$

The statistical reference for wind is a seasonal ARIMA with a daily period of s = 8 three-hour steps,

$$\phi(B)\,\Phi(B^8)\,(1-B^8)\,y_t = \theta(B)\,\Theta(B^8)\,\varepsilon_t \tag{6}$$

fitted with the statsmodels state-space implementation [31]. Its Kalman filter handles missing observations exactly, so the Agadez night gaps need no imputation [32]. Orders were selected by AIC on the training period over $p \in \{1, 2\}$, $q \in \{0, 1\}$, $P \in \{0, 1\}$ and $Q = 1$, and the selected model is SARIMA(2,0,1)(1,1,1)_8

at all three stations. For daily sunshine a 365-day seasonal difference is impractical, so the annual cycle is carried by Fourier regressors and the residual by an ARMA process:

$$n_t = \beta_0 + \sum_{k=1}^{2}[\beta_k \sin(2\pi k \cdot \mathrm{doy}/365.25) + \gamma_k \cos(2\pi k \cdot \mathrm{doy}/365.25)] + \eta_t, \quad \phi(B)\,\eta_t = \theta(B)\,\varepsilon_t \quad (7)$$

with (p, q) chosen by AIC over {1, 2, 3} × {0, 1, 2}: ARIMA(2,0,2) at Agadez and Bilma, ARIMA(3,0,2) at N'Guigmi. Parameters are estimated once on 2015-2022 and then frozen. The filter is run through 2023-2024, and each h-step forecast is obtained by propagating the filtered state from its origin, ŷ(t+h|t) = x(t+h)'β + Z T^h a(t|t), which uses only information available at t. At the last training origin this reproduces statsmodels' own multi-step forecast to within 10^{-6}.

3.4 Machine-learning models actually implemented

A one-dimensional convolutional front end, an LSTM layer [33] and additive temporal attention, pretrained on decades of reanalysis, would be a natural architecture for this problem. That architecture was not trained in this study. The learning curves of Section 4.8 show that even a 64-32 network overfits the station record within a few epochs, and no reanalysis archive was used here that could have served for pretraining. We therefore report two model families that were trained end to end on the station data with scikit-learn 1.8 [34]. We make no claim about a convolutional-recurrent or attention network. Gradient-boosting and hybrid boosting-based architectures have nonetheless proved competitive with deep recurrent networks for operational energy forecasting elsewhere [35].

The first family is histogram gradient boosting (HistGradientBoostingRegressor), an additive ensemble of shallow regression trees fitted stage-wise to the negative gradient of the loss [36]:

$$F_M(x) = F_0 + \nu \sum_{m=1}^{M} f_m(x), \quad f_m = \mathrm{argmin}_f \sum_i \left[-\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F} - f(x_i)\right]^2 \quad (8)$$

with learning rate ν = 0.05, at most 15 leaves per tree, at least 80 samples per leaf and an L2 penalty of 1.0. The number of stages is M = 100 for wind and M = 60 for sunshine. These values were read off the inner validation curves, where the minimum lay between 49 and 138 iterations for wind and between 28 and 61 for

sunshine (Figure 12). Missing inputs are routed natively by the trees, which suits a record with structured gaps. Recent benchmarks also show tree ensembles outperforming deep networks on medium-sized tabular problems of exactly this kind [37].

The second family is a multilayer perceptron (MLPRegressor) with two hidden layers of 64 and 32 ReLU units, trained with Adam [38] (learning rate 10^{-3}, mini-batches of 256, L2 penalty 10^{-3}) on standardized inputs, with median imputation plus missingness indicators and a standardized target:

$$\hat{y} = w_3^{\mathrm{T}}\,\mathrm{ReLU}(W_2\,\mathrm{ReLU}(W_1 z + b_1) + b_2) + b_3, \quad z = \mathrm{clip}\left(\frac{x-\mu_{\mathrm{train}}}{\sigma_{\mathrm{train}}}, -5, 5\right) \tag{9}$$

Clipping the standardised inputs at ±5 SD was added after the first inner-validation run. Isolated extreme readings (wind reports up to 36 at N'Guigmi, where the 99th percentile is 7) had pushed the unclipped network to a validation RMSE worse than persistence for N'Guigmi sunshine (3.36 against 2.83 h/day). The number of epochs was set per station and target from the inner validation minimum (Figure 12): 5 to 10 epochs for wind and 7 to 21 for sunshine, after which validation error rises steadily. With about 1,900 to 2,200 daily training samples per station, the 64-32 network already has more than 6,000 weights, and the validation curves show it overfitting within a few passes. That behaviour argues against moving to a larger architecture on this dataset rather than for it. Table 3 lists every setting used.

Table 3. Models and settings actually used. Every value was chosen on the 2015-2020 / 2021-2022 inner split; the 2023-2024 test years were not used for any choice.

Component	Setting	How chosen
Persistence (wind)	last report no older than 24 h; diurnal persistence = same slot, previous available day	fixed
Persistence (sunshine)	relative sunshine n/N carried to target day (Eq. 5)	fixed
SARIMA (wind)	(2,0,1)(1,1,1) with s = 8 at all stations, Kalman filter with gaps	AIC over 8 candidate orders
ARIMA + Fourier (sunshine)	(2,0,2) Agadez, Bilma; (3,0,2) N'Guigmi; 2 annual harmonics + constant	AIC over 9 orders
Gradient boosting (HGB)	scikit-learn HistGradientBoostingRegressor; learning rate 0.05, 15 leaves, min. 80 samples/leaf, L2 = 1; 100 stages (wind), 60 (sunshine); one model per lead	inner validation curves (optimum 28-138 stages)

Neural network (MLP)	scikit-learn MLPRegressor, 2 hidden layers (64, 32), ReLU, Adam, lr 1e-3, batch 256, L2 1e-3; median imputation + missingness flags, inputs standardised and clipped at +/-5 SD; 5-10 epochs (wind), 7-21 (sunshine)	inner validation minimum per station and target
Quantile HGB	19 levels 0.05-0.95, pinball loss, 200 stages; wind on de-discretised scale	fixed
Conformal recalibration	split CQR: fit 2015-2021, calibrate 2022, test 2023-2024	fixed
Transfer	station-normalised wind; pooled HGB with site flag; residual HGB (50 stages); MLP fine-tuned 10 epochs	fixed (not tuned on test)
Not implemented	CNN-LSTM with temporal attention, reanalysis pretraining, L2-SP fine-tuning (not trained in this study; see Section 3.4)	-

3.5 Probabilistic forecasts and calibration

Quantile forecasts are obtained with the same gradient-boosting learner under the pinball loss [39], one model per level $\tau \in \{0.05, 0.10, ..., 0.95\}$ (19 levels, 200 stages), with crossing quantiles removed by sorting:

$$L_\tau(y, q) = \max[\tau(y - q),\ (\tau - 1)(y - q)] \tag{10}$$

Wind is reported in whole units with up to 57 % calms, and in a first run this tie structure made the N'Guigmi quantile models collapse to a constant median of zero. For the probabilistic wind forecasts we therefore work on a de-discretised scale: each report is replaced by the reported integer plus uniform noise on [-0.5, 0.5) ([0, 0.5) for calms), which treats the report as a rounded speed. Training, calibration, the climatological reference and evaluation all use this scale, following the randomisation used to assess forecasts of discrete data [40]. Coverage on the reporting scale (rounded quantiles against integer reports) is shown alongside for transparency. Calibration is assessed by the observed frequency $P(y \leq q_\tau)$ across the 19 levels (a reliability diagram) and by the prediction interval coverage probability,

$$\mathrm{PICP} = \frac{1}{n}\sum_t \mathbf{1}\left[q_{\mathrm{lo}}(t) \leq y(t) \leq q_{\mathrm{hi}}(t)\right] \tag{11}$$

for the nominal 80 % ($q_{0.10}$ to $q_{0.90}$) and 90 % ($q_{0.05}$ to $q_{0.95}$) intervals, together with their mean width. Overall probabilistic skill is summarised by the continuous ranked probability score, approximated through its quantile-score representation on the 19-level grid [41]:

$$\mathrm{CRPS}(F, y) = \int [F(x) - \mathbf{1}(x \geq y)]^2 dx = 2 \int_0^1 L_\tau\,(y, q_\tau)\, d\tau \approx \frac{2}{19} \sum_k L_{\tau_k} \left(y, q_{\tau_k}\right) \quad (12)$$

The discretisation omits the tails beyond 0.05 and 0.95, so the absolute CRPS values are slightly low. The same grid is applied to the reference, so the skill score CRPSS = 1 - CRPS/$\mathrm{CRPS}_{\mathrm{clim}}$ is consistent. The reference is the training-period climatological quantiles of the target for the same calendar month (and, for wind, the same 3-h slot). Where the raw intervals are miscalibrated, split-conformal quantile regression [42] is applied. Quantile models are refitted on 2015-2021, conformity scores are computed on 2022, and the test interval is shifted by their finite-sample quantile:

$$E_i = \max[q_{\mathrm{lo}}(x_i) - y_i,\ y_i - q_{\mathrm{hi}}(x_i)], \quad \hat{Q} = E_{(\lceil (n+1)(1-\alpha) \rceil)}, \quad C(x) = \left[q_{\mathrm{lo}}(x) - \hat{Q},\ q_{\mathrm{hi}}(x) + \hat{Q}\right] \quad (13)$$

3.6 Cross-site transfer to Agadez

Agadez is the target site; Bilma and N'Guigmi are the source, a pooling strategy in the spirit of studies that have examined how aggregating or transferring information across sites and levels changes forecast accuracy [43]. Because the wind climates differ by a factor of four in mean speed, wind targets and all wind-valued features are expressed in station-normalised form,

$$z_s = \frac{w - \mu_s}{\sigma_s} \quad (14)$$

with μ_s and σ_s the station's 2015-2022 mean and standard deviation, and forecasts are transformed back to physical units before scoring. Sunshine is used in physical units, since the station means differ by only about one hour. Three Agadez training scenarios are compared: one year of data (2022), two years (2021-2022) and the full training record (2015-2022). The methods are a local model trained on Agadez alone (HGB and MLP), a zero-shot source HGB trained on Bilma and N'Guigmi only, a pooled multi-site HGB trained on

source and Agadez rows with a target-site flag, and two fine-tuning variants. For the tree model, fine-tuning is residual boosting: a second, smaller ensemble (50 stages) is fitted to the Agadez residuals of the source model,

$$F_{\text{Agadez}}(x) = F_{\text{source}}(x) + g(x), \quad g = \text{HGB fitted to } \{x_i,\ y_i - F_{\text{source}}(x_i)\}_{i \in \text{Agadez}} \tag{15}$$

For the network, the source-trained MLP keeps its weights and input scaling and is trained for ten further epochs on the Agadez data. An L2-SP penalty toward the pretrained weights was not implemented. All scenarios are scored on the same Agadez 2023-2024 test samples.

3.7 Evaluation

Point accuracy is reported as RMSE and MAE, and skill against a reference as

$$\text{SS} = 1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{reference}}} \tag{16}$$

so that a negative value means the model is worse than that reference. The contribution of each feature group g (Table 2) is measured by grouped permutation importance on the test set [44]. All columns of the group are shuffled jointly across test samples, which keeps within-group correlation intact, and the increase in RMSE is averaged over five permutations:

$$I_g = \text{E}_\pi\left[\text{RMSE}\left(f\left(X_{\pi,g}\right),\ y\right)\right] - \text{RMSE}(f(X),\ y) \tag{17}$$

This permutation-importance analysis stands in for an attention-weight map, which the models used here do not produce; it is a property of the trained HGB model, not of an attention mechanism. Uncertainty on the ablation results is estimated with a paired day-block bootstrap (1,000 resamples of whole days, so that the eight 3-h errors of a day stay together) [45].

3.8 Operational risk translation

The day-ahead set-up assumes a forecast issued at 18:00 local time on day D - 1, the last slot observed at all three stations. The PV forecast is the 1-day sunshine forecast for day D. It is converted to global horizontal

irradiation with the Angström-Prescott relation, using the FAO-56 default coefficients because the file contains no independent irradiance reference against which to calibrate them locally [24]:

$$H = \left(a_s + b_s \cdot \frac{n}{N}\right) \cdot \frac{Ra}{3.6} \ \text{(kWh m}^{-2}\,\text{d}^{-1}\text{)}, \quad a_s = 0.25, \ b_s = 0.50 \tag{18}$$

DC energy per kWp follows a NOCT-based thermal model in daily form,

$$E_{\text{PV}} = C_{\text{PV}} \cdot H \cdot PR_0 \cdot [1 + \gamma(T_{\text{cell}} - 25)], \quad T_{\text{cell}} = T_{\text{air}} + \frac{\text{NOCT}-20}{800} \cdot G_w \tag{19}$$

with PR_0 = 0.85 (wiring, inverter, mismatch and availability; soiling excluded), γ = -0.004 K^{-1}, NOCT = 45 °C, G_w = 1.234 times the mean daytime irradiance (the irradiance-weighted mean of a half-sine day) and T_{air} = TMAX, following the Sandia/NOCT family of module-temperature models [46]. For the forecast, the TMAX of D - 1 is used, since that is what an operator would know. The wind forecast is the set of 3-h forecasts issued at 18:00 for the 06, 09, 12, 15 and 18 h slots of day D (leads of 4 to 8 steps). Each is extrapolated to a 30-m hub with the log law and assumed station roughness lengths (z_0 = 0.03, 0.01 and 0.05 m), then passed through a cubic power curve with a density correction from station pressure and daily mean temperature:

$$V_{\text{hub}} = V_{10} \cdot \frac{\ln(30/z_0)}{\ln(10/z_0)}, \quad P_{\text{WT}} = C_{\text{WT}} \cdot \frac{V^3 - V_{\text{ci}}^3}{V_r^3 - V_{\text{ci}}^3} \cdot \frac{\rho}{\rho_0} \quad \text{for } V_{\text{ci}} \le V < V_r \tag{20}$$

with generic small-turbine values V_{ci} = 3, V_r = 12 and V_{co} = 25 m/s. Each slot represents 3 h, so the wind window covers 06-21 h, the only window observed at all three stations. The day-ahead shortfall of an illustrative plant with C_{PV} = 100 kWp and C_{WT} = 30 kW is $S_D = \max(0, \hat{E}_D - E_D)$. Its tail is summarised with the conditional value at risk [47]:

$$\text{CVaR}_{95}(S) = \text{E}[S \mid S \ge \text{VaR}_{95}(S)] \tag{21}$$

and converted into a nominal battery buffer with an assumed battery round-trip efficiency of 0.92 and an assumed depth of discharge of 0.8:

$$B = \frac{\text{CVaR}_{95}(S)}{\text{DoD} \cdot \eta_{\text{dis}}}, \quad \eta_{\text{dis}} = \sqrt{0.92} \tag{22}$$

3.9 Robustness of the evaluation

A single two-year test period cannot separate a genuine difference between models from year-to-year variability. The comparison between HGB and the statistical references is therefore repeated with an expanding-window rolling origin [48]. For each test year Y from 2019 to 2024, every training-dependent quantity (the calendar-month 90th percentiles behind the hot-day flag, the sunshine and wind climatologies, the k-means regime centroids and the strong-wind threshold) is re-estimated on data up to 31 December of Y - 1. HGB is refitted with the settings of Table 3 on samples whose target time falls on or before that date, and forecasts are issued from every origin in year Y. The SARIMA and ARIMA-Fourier references keep the orders selected in Section 3.3, but their parameters are re-estimated on the same expanding window before the filter is run through year Y. The first fold trains on four years (2015-2018). The 2023 fold reproduces the fixed-origin forecasts of Section 4.2 exactly, which confirms that the two protocols are consistent, and scoring uses the same sample definition as Table 4.

Uncertainty on the skill of HGB against a reference is estimated with a moving-block bootstrap over days [45]. Blocks of 28 consecutive days are resampled 2,000 times, so that the eight 3-h errors of a day and most of the synoptic persistence stay together, and the skill score is recomputed from the resampled sums of squared errors. Whether the difference in squared error is significant is assessed with the Diebold-Mariano test [49], applied to the daily mean loss differential $d_t = e^2_{ref,t} - e^2_{HGB,t}$ with a Bartlett long-run variance and the small-sample correction of Harvey et al. [50]:

$$\mathrm{DM} = \sqrt{\frac{T+1-2k+k(k-1)/T}{T}}\ \frac{\overline{d}}{\sqrt{\hat{\sigma}^2_{\mathrm{LR}}/T}} \qquad \hat{\sigma}^2_{\mathrm{LR}} = \hat{\gamma}_0 + 2\sum_{j=1}^{L}\left(1-\frac{j}{L+1}\right)\hat{\gamma}_j \tag{23}$$

where T is the number of test days, $\hat{\gamma}_j$ the autocovariance of d_t at lag j, k the lead time in days (rounded up) and L = max(k - 1, ⌊4(T/100)^(2/9)⌋), which gives L = 6 for the two-year test and L = 7 for the six-year rolling test. The statistic is compared with a Student t distribution with T - 1 degrees of freedom. Thirteen wind leads and seven sunshine leads are tested per station, so the p-values are also adjusted with the Holm procedure [51]

within each station and target. For the rolling-origin results, the number of years in which HGB beats the reference is reported together with a Wilcoxon signed-rank test on the six annual RMSE differences. Three levels of evidence are then kept apart. A numerical improvement is a lower RMSE on the test sample. A statistically significant improvement requires a Holm-adjusted p-value below 0.05 and a bootstrap interval that excludes zero. An operationally relevant improvement is one that reduces the battery buffer of Eq. 22 by more than its own sampling uncertainty, estimated with the same block bootstrap on the daily shortfall series.

Split-conformal coverage rests on the exchangeability of calibration and test scores [42], which a record with inter-annual changes need not satisfy. Coverage of the recalibrated intervals of Section 3.5 is therefore reported separately for 2023 and 2024 and for the dry (October-May) and wet (June-September) seasons, and the conformity scores of the calibration year are compared with those of each test year with a two-sample Kolmogorov-Smirnov test. Two variants that adapt to change are also evaluated: a sequential update that recalibrates on 2023 before forecasting 2024, and an annual rolling scheme that, for Y = 2020-2024, fits the quantile models on data up to Y - 2, calibrates on Y - 1 and tests on Y, with features rebuilt at each cutoff. Online schemes designed explicitly for distribution shift exist [52] but were not implemented here.

Finally, the energy translation of Section 3.8 passes point forecasts through a power curve that is convex between cut-in and rated speed, and the output at the expected speed is not the expected output. For the day-ahead wind window, quantile HGB models with the settings of Section 3.5 (19 levels, 200 stages, de-discretised target, training 2015-2022) were fitted for the 18:00 origins at leads of 4 to 8 steps. Their quantiles are interpolated linearly on a grid of 1,000 probability levels τ_i, with linear extrapolation into the tails beyond 0.05 and 0.95 and speeds clipped at zero, which gives a predictive quantile function $\hat{Q}_{t(\tau)}$.The power curve of Eq. 20 is non-decreasing below cut-out, which none of the forecast speeds reaches (the highest extrapolated 10-m quantile is 19.5 m/s, against about 21 m/s for a 25 m/s hub-height cut-out), so the quantiles of turbine output are the outputs of the speed quantiles, and the expected output of slot t is

$$\mathbb{E}[P_t] = \int_0^1 P\left(\hat{Q}_t(\tau)\right) \mathrm{d}\tau \approx \frac{1}{1000} \sum_{i=1}^{1000} P\left(\hat{Q}_t(\tau_i)\right) \tag{24}$$

PV production is treated in the same way with the one-day sunshine quantiles, the PV model of Eq. 19 being increasing in n over the observed range. Adding the five wind slots and the PV day into a daily production distribution requires an assumption about their dependence, which marginal quantile forecasts do not provide. Two bounding cases are therefore used: comonotonic slots (quantiles added level by level, with no diversification) and independent slots (4,000 Monte Carlo draws per margin). A day-specific reserve is then defined from the 5 % quantile of daily production E_D,

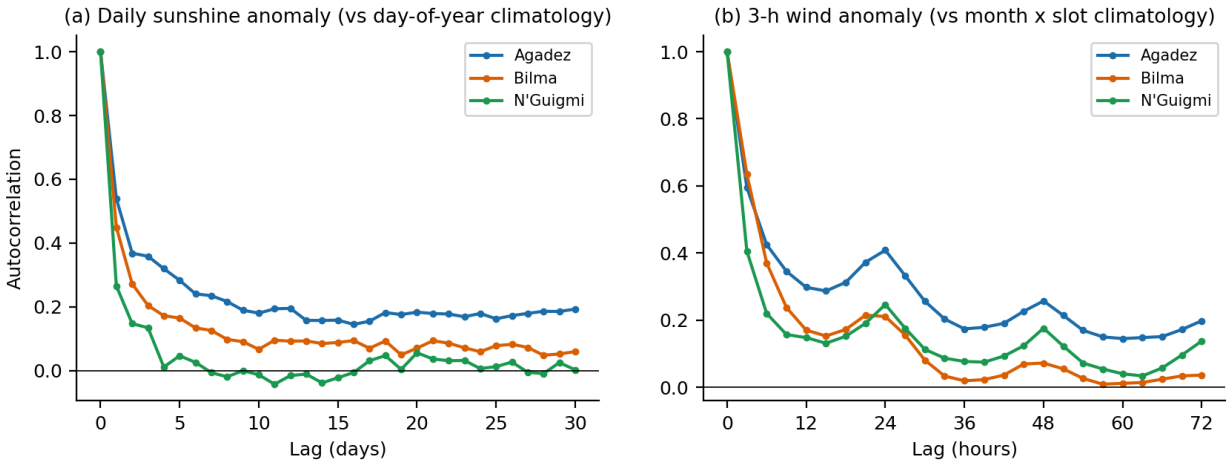
$$R_D = \mathbb{E}[E_D] - Q_{E_D}(0.05), \qquad S_D = \max(0,\ \mathbb{E}[E_D] - E_D) \tag{25}$$

and its reliability is judged by the share of days on which the realized shortfall S_D exceeds R_D, against a nominal 5 %.

4. Results

4.1 Structure of the record and predictability

Before any model is trained, the anomaly autocorrelations already set expectations (Figure 3). Day-to-day persistence of sunshine anomalies is moderate at Agadez (lag-1 autocorrelation 0.54) and Bilma (0.45) and weak at N'Guigmi (0.27), and it is close to zero within a week everywhere (0.24, 0.13 and -0.01 at lag 7). Once the month-by-slot climatology is removed, wind anomalies keep a lag-1 (3 h) autocorrelation of 0.41-0.64, a 24-h autocorrelation of 0.21-0.41 and little beyond 72 h. Two features of the forecasting problem follow from this. At all leads a large share of the predictable signal is the deterministic diurnal and seasonal cycle, which any competent model will capture. What remains after the first day is weakly autocorrelated noise, so large gains over a good statistical model should not be expected, whatever the architecture.



402

403 **Figure 3.** Autocorrelation of target anomalies, 2015-2024. (a) Daily sunshine duration minus its smoothed day-of-year
404 climatology, lags 0-30 days. (b) 3-hourly wind speed minus its month-by-slot climatology, lags 0-72 h (pairwise-complete;
405 Agadez values rest mainly on daytime pairs).

406 ## 4.2 Deterministic skill against persistence and SARIMA

407 Figure 4 shows wind skill against last-observation persistence. All three models beat persistence at every lead
408 and station, by 20-36 % at Agadez, 18-36 % at Bilma and 15-49 % at N'Guigmi. The saw-tooth pattern with
409 minima at 24, 48 and 72 h is not a model artefact. At those leads persistence coincides with diurnal persistence
410 (same slot, one day earlier) and becomes a much stronger reference. Diurnal persistence alone (dashed line) is
411 itself worse than the learned and statistical models at every lead. For sunshine (Figure 5) the models beat
412 relative-sunshine persistence by 15-30 %.

413 The comparison that decides H1 is the one with SARIMA (Figure 6, Table 4). For wind, HGB has a lower
414 RMSE than SARIMA at every lead at Agadez, by 4.3 % at 3 h, 3.5 % at 24 h and 0.6 % at 72 h. At Bilma its
415 RMSE is lower up to 18 h (by 5.2 % at 3 h) and slightly higher from 21 h onward (by 0.4 % at 24 h and up to
416 1.7 % at 48 h). At N'Guigmi it is higher at every lead, by 1.2 % at 3 h and 3.8-3.9 % at 48-72 h. These are
417 numerical differences on a single two-year sample; Section 4.11 tests which of them are statistically
418 significant and whether they hold in other years. The MLP is weaker than HGB almost everywhere: its skill
419 against SARIMA ranges from +3.2 % (Bilma and Agadez, 3 h) to -7.6 % (N'Guigmi, 48 h). For sunshine the
420 three approaches are practically indistinguishable. HGB's skill against the ARIMA-Fourier model lies

between -2.9 % and +2.5 %, and the largest gain is at Bilma one day ahead (1.911 against 1.961 h/day), which does not remain significant once multiple testing is accounted for (Section 4.11). The MLP is 1.4-8.4 % worse than the statistical model at every station and lead.

H1 is therefore rejected as stated. The learned models do beat persistence everywhere, but their advantage over SARIMA is confined to the windiest, most diurnally driven site (Agadez) and to short leads at Bilma, and only part of it is statistically significant (Section 4.11). At N'Guigmi, where more than half the 3-h reports are calms and the useful signal is little more than a daytime pulse, the seasonal ARIMA's explicit daily period appears to be the better description in 2023-2024, although the rolling-origin results show that this ranking is not stable from year to year. The narrowing of the margin with lead time, the second half of H1, does hold at Agadez and Bilma.

Table 4. Test-period RMSE by station, target and lead, and HGB/MLP skill scores.

Skill scores (%) of HGB against persistence and of HGB and MLP against SARIMA; negative skill = worse than the reference. All models scored on the same samples.

Station	Target	Lead	n	Persistence	SARIMA	HGB	MLP	SS HGB vs pers.	SS HGB vs SARIMA	SS MLP vs SARIMA
Agadez	Wind (m/s)	3 h	3638	3.410	2.404	2.300	2.329	+32.6	+4.3	+3.1
Agadez	Wind (m/s)	24 h	3633	3.338	2.685	2.591	2.664	+22.4	+3.5	+0.8
Agadez	Wind (m/s)	72 h	3602	3.635	2.808	2.791	2.827	+23.2	+0.6	-0.7
Agadez	Sunshine (h/day)	1 d	724	2.885	2.434	2.442	2.468	+15.4	-0.3	-1.4
Agadez	Sunshine (h/day)	3 d	722	3.340	2.598	2.672	2.719	+20.0	-2.9	-4.6
Agadez	Sunshine (h/day)	7 d	718	3.701	2.672	2.721	2.726	+26.5	-1.8	-2.0
Bilma	Wind	3 h	4961	2.078	1.711	1.621	1.656	+22.0	+5.2	+3.2

	(m/s)									
Bilma	Wind (m/s)	24 h	4954	2.420	1.917	1.925	1.954	+20.5	-0.4	-1.9
Bilma	Wind (m/s)	72 h	4912	2.652	1.970	1.993	2.038	+24.9	-1.2	-3.5
Bilma	Sunshine (h/day)	1 d	694	2.393	1.961	1.911	2.024	+20.1	+2.5	-3.2
Bilma	Sunshine (h/day)	3 d	690	2.702	2.039	2.042	2.152	+24.4	-0.2	-5.5
Bilma	Sunshine (h/day)	7 d	682	2.663	2.089	2.134	2.157	+19.9	-2.1	-3.3
N'Guigmi	Wind (m/s)	3 h	5594	1.996	1.315	1.331	1.373	+33.3	-1.2	-4.4
N'Guigmi	Wind (m/s)	24 h	5587	1.742	1.389	1.411	1.445	+19.0	-1.6	-4.1
N'Guigmi	Wind (m/s)	72 h	5565	1.835	1.444	1.499	1.529	+18.3	-3.8	-5.9
N'Guigmi	Sunshine (h/day)	1 d	694	3.029	2.290	2.291	2.381	+24.4	-0.0	-4.0
N'Guigmi	Sunshine (h/day)	3 d	690	3.043	2.311	2.319	2.505	+23.8	-0.3	-8.4
N'Guigmi	Sunshine (h/day)	7 d	682	3.365	2.360	2.360	2.523	+29.9	-0.0	-6.9

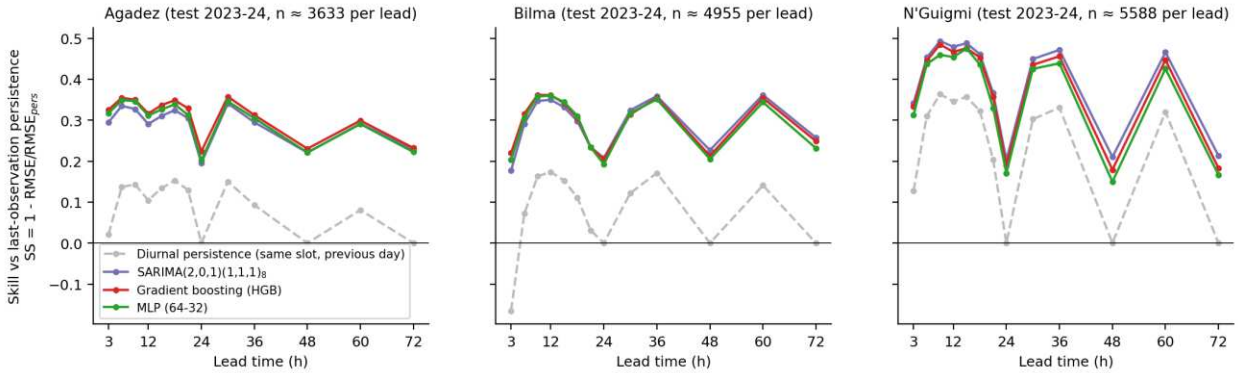


Figure 4. Skill of 3-hourly wind forecasts against last-observation persistence, by lead time and station.

SS = 1 - RMSE/RMSE_pers, by lead time (3-72 h) and station, test period 2023-2024 (n ≈ 3,600, 4,950 and 5,590 scored forecasts per lead at Agadez, Bilma and N'Guigmi). Dashed: diurnal persistence (same slot, most recent available day).

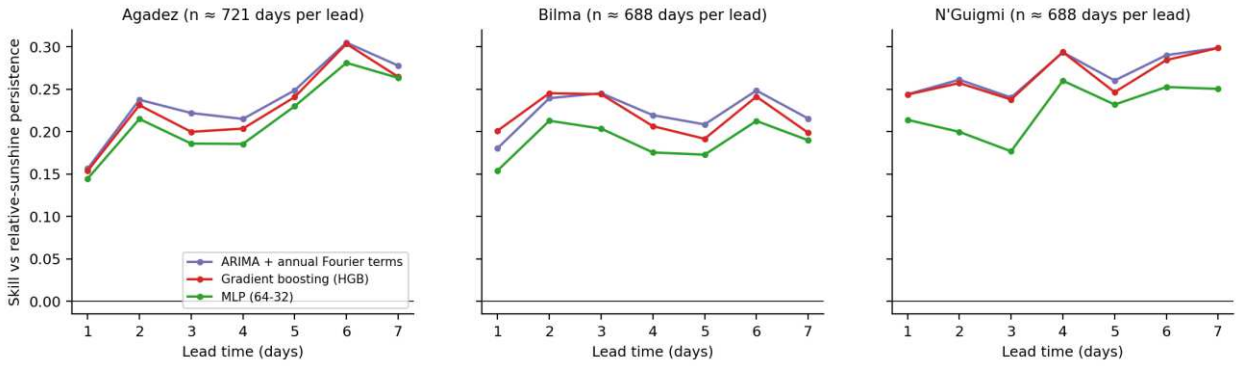


Figure 5. Skill of daily sunshine-duration forecasts against relative-sunshine persistence (Eq. 5), by lead time (1-7 days) and station, test period 2023-2024 (n ≈ 682-724 days per lead).

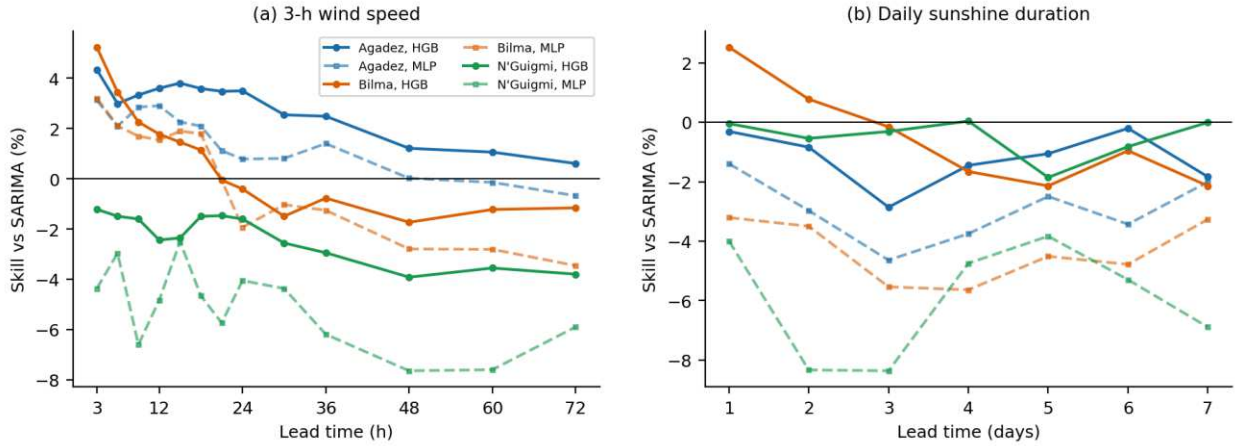


Figure 6. Skill of the machine-learning models against the statistical reference, all three stations.

SARIMA(2,0,1)(1,1,1)_8 for wind, ARIMA with annual Fourier terms for sunshine, all three stations on one graph; solid: HGB, dashed: MLP. Negative values mean the learned model is worse than SARIMA. Test period 2023-2024.

4.3 Error distributions by lead time

Figure 7 gives the full distribution of HGB errors (forecast minus observation). Mean biases are small, between -0.25 and +0.31 for wind and between -0.28 and +0.33 h/day for sunshine, but none of the distributions is symmetric. Wind errors have a long negative tail at every station (skewness between -1.4 and -

4.6): strong-wind events, the ones that matter most for dust uplift and turbine output, are systematically under-forecast. The positive side is short at Bilma and N'Guigmi because a forecast can overshoot a calm observation only by its own, usually small, value. The 5th percentile of the error moves from -2.6 to -3.7 m/s between 3 and 12 h at Bilma and changes little after that, whereas at Agadez it keeps widening slowly, from -3.7 at 3 h to -5.0 at 72 h. Sunshine errors are skewed the other way (skewness 1.0 to 1.8), with a 95th percentile of 4.2 to 5.9 h/day. Days that turn out much duller than expected, through dust or convective cloud, are not anticipated, while unexpectedly bright days are rarer and smaller. This asymmetry is why the shortfall tail in Section 4.10 is much larger than the mean error would suggest.

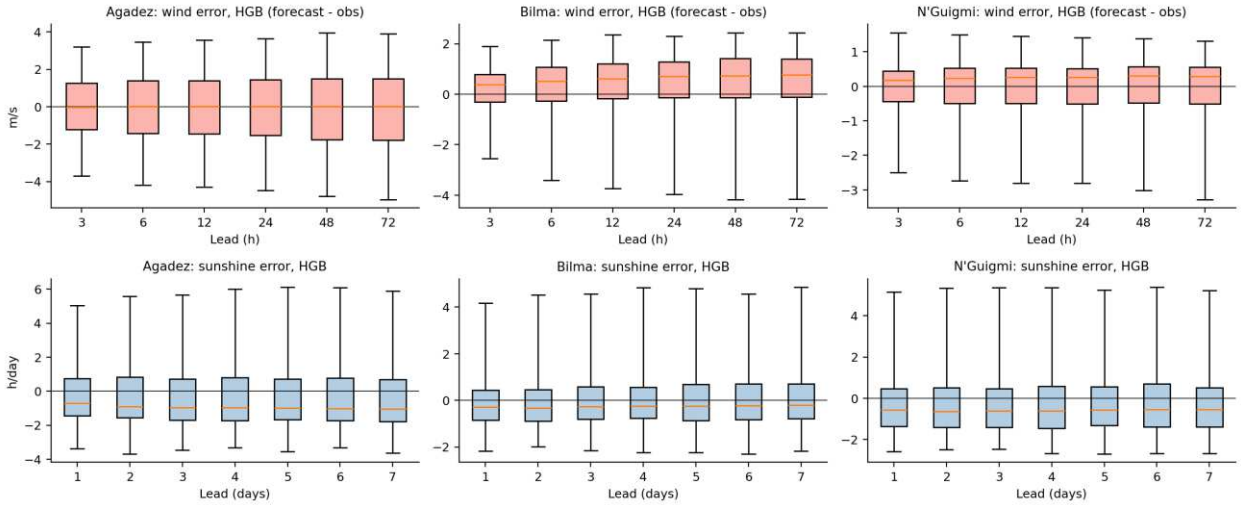


Figure 7. Distribution of HGB forecast errors (forecast minus observation) by lead time, test period 2023-2024. Top: 3-hourly wind speed at leads 3-72 h. Bottom: daily sunshine duration at leads 1-7 days. Boxes: inter-quartile range; whiskers: 5th-95th percentiles; outliers not shown.

4.4 Example forecast trajectories with prediction intervals

Figure 8 shows two test windows chosen objectively, as the windows whose median-forecast MAE is closest to the test-period mean, so that they are typical rather than flattering. The N'Guigmi wind window falls in the Harmattan season (late January to early February 2023). On the reporting scale, 90 % of the reports in this window fall inside the 10-90 % band. The quantile forecasts reproduce the late-morning pulse and the calm nights well. They also show the main limitation of a model trained on station data alone: the median follows

the typical diurnal amplitude, while the strongest reports (6-7 m/s) sit at or above the upper edge of the band, because nothing in the inputs announces a surge a day ahead. For Agadez sunshine one day ahead, the interval is narrow on clear days and widens sharply after an overcast or dusty day, which is the behaviour an operator would want. The dullest days still fall below the 10th percentile more often than they should, consistent with the calibration results of Section 4.9.

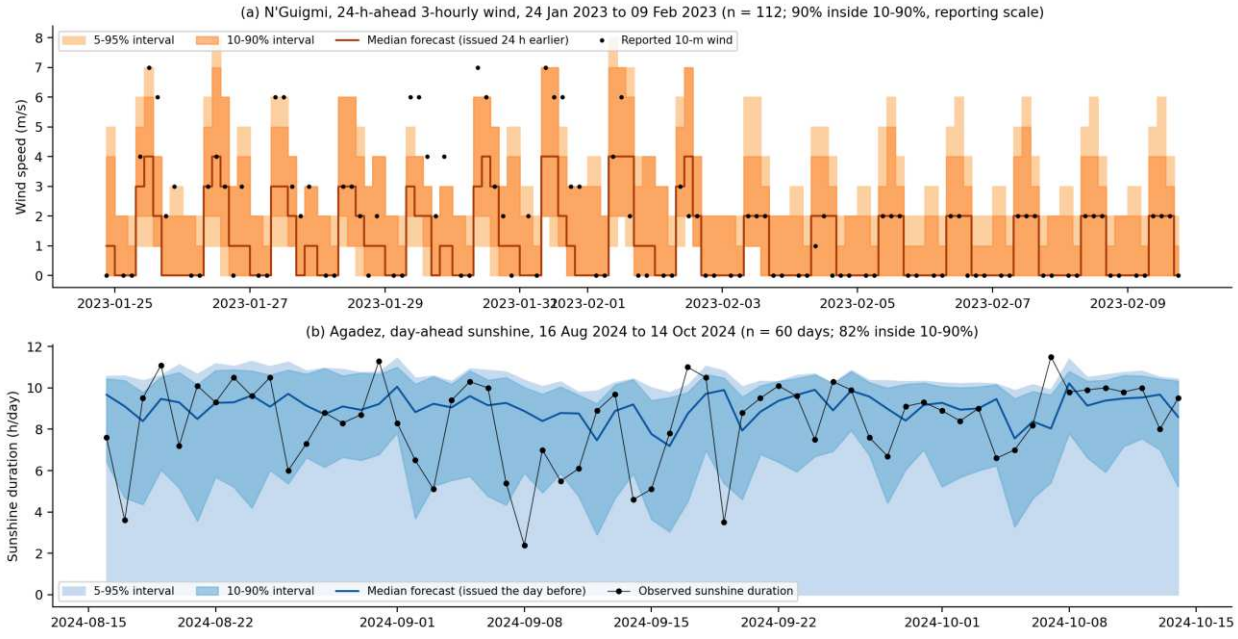


Figure 8. Example day-ahead forecasts with prediction intervals from quantile gradient boosting, test period. (a) N'Guigmi 3-hourly wind forecast 24 h ahead over a 14-day window (n = 112 slots), with quantile bounds rounded to the reporting resolution of whole units (calm = 0). (b) Agadez daily sunshine forecast 1 day ahead over a 60-day window. Windows selected as those whose median-forecast MAE is closest to the test-period mean MAE.

4.5 Ablation: value of heat and wind-regime features

Figure 9 compares HGB trained on the base features with HGB trained on base plus derived features, stratified by the regime of the target day, for wind 24 h ahead and sunshine one day ahead. The differences are small. For wind they lie between -1.0 % and +0.8 % in every station-regime cell. The largest gain is in the strong-Harmattan regime at Agadez (+0.75 %, 95 % bootstrap interval of the RMSE difference -0.001 to +0.053 m/s), and at N'Guigmi the derived features make strong-Harmattan forecasts slightly worse (-0.87 %,

interval -0.031 to -0.005 m/s). For sunshine the only differences whose bootstrap interval excludes zero point in opposite directions. At Agadez the derived features degrade strong-Harmattan sunshine forecasts by 5.4 % (n = 160 days, interval -0.124 to -0.046 h/day). At Bilma they improve monsoon-type days by 1.8 % (n = 203, interval +0.016 to +0.080 h/day), where the moisture tracer probably helps anticipate convective cloud.

H2 is not supported. With the variables this file provides, the heat and regime features do not concentrate their value in dust-prone Harmattan conditions, and in the one clearly significant case they hurt. The likely reason is that the file contains no aerosol, visibility or irradiance measurement. The regime label and the strong-wind share describe the wind that can raise dust, not the dust itself, and most of what they encode is already in the lagged target and the calendar terms. An aerosol optical depth series would be needed for a fair test of H2, and regimes defined from aerosol optical depth cannot be built from this file.

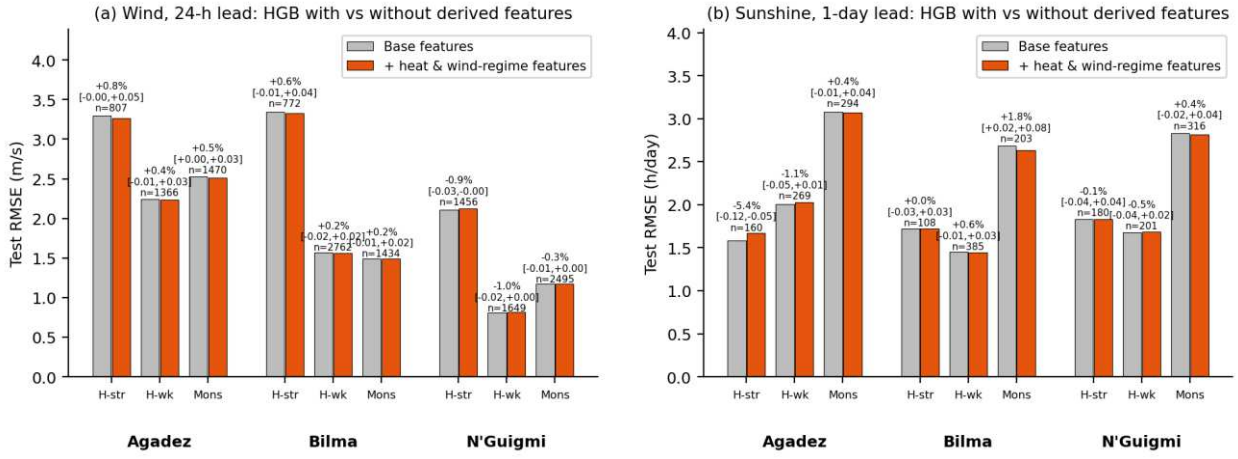


Figure 9. Ablation of the derived heat and wind-regime features. Test RMSE of HGB with base features only (grey) and with base plus derived features (orange), stratified by the regime of the target day (H-str: Harmattan-type, strong wind; H-wk: Harmattan-type, weak wind; Mons: monsoon-type). (a) Wind, 24-h lead. (b) Sunshine, 1-day lead. Labels give the relative gain, the 95 % day-block bootstrap interval of the RMSE difference (base minus full, 1,000 resamples) and the number of test samples.

4.6 Feature importance by forecast horizon

Figure 10, a permutation-importance analysis standing in for an attention map and in the spirit of explainable-AI approaches to energy forecasting [53], shows the RMSE increase when each feature group is shuffled on

the test set. For wind, calendar and diurnal terms dominate at every lead and station (0.23-0.55 m/s). Recent lags matter strongly at 3 h at Agadez (0.47 m/s) and Bilma (0.63 m/s) and fade by 12-24 h at Bilma and N'Guigmi. At Agadez they keep some weight up to 24 h (0.41 m/s), consistent with synoptic Harmattan surges lasting more than a day at that site. The direction and regime group reaches 0.09 m/s at Agadez at 3 h and stays at or below 0.02 m/s elsewhere. Heat indices never exceed 0.01 m/s. For sunshine, lagged sunshine dominates the 1-day forecast at Agadez (0.36 h/day) and Bilma (0.21 h/day). At N'Guigmi conventional meteorology carries more (0.16 h/day at one day), which fits a site where humidity and temperature signal the monsoon's arrival better than yesterday's sunshine does. Heat indices and wind proxies stay at or below 0.03 h/day at every lead. The feature importance thus tells the same story as the ablation.

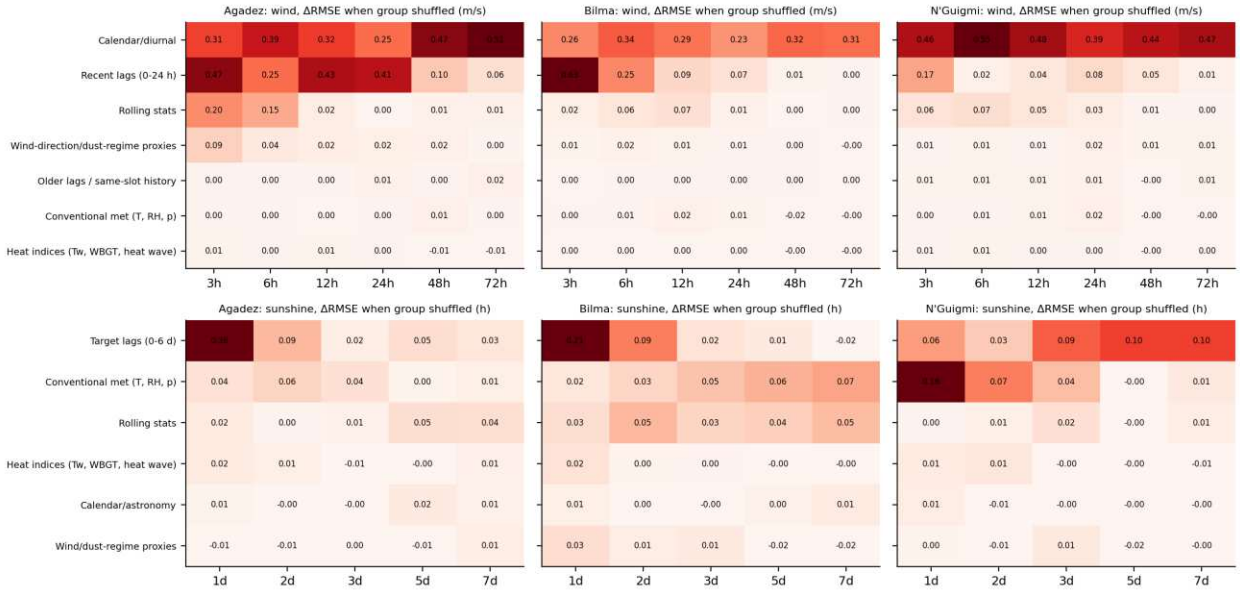


Figure 10. Grouped permutation importance of the HGB models by lead time.

Increase in test RMSE (2023-2024) when all columns of a feature group are shuffled jointly (mean of five permutations). Top: wind (m/s) at leads 3-72 h. Bottom: sunshine (h/day) at leads 1-7 days. Groups as in Table 2.

4.7 Cross-site transfer to Agadez

Figure 11 answers H3. The zero-shot source model, which has never seen Agadez data, is worse than even a one-year local model for 3-h wind (2.522 against 2.386 m/s) and for sunshine (2.520 against 2.504 h/day). The

Agadez wind climate is different enough that source data cannot simply replace local data. Combined with a little local data, however, the source helps. With one year of Agadez training data, pooled multi-site HGB lowers the 24-h wind RMSE from 2.714 to 2.631 m/s (-3.1 %) and the sunshine RMSE from 2.504 to 2.458 h/day (-1.8 %). Residual fine-tuning lowers the 3-h wind RMSE from 2.386 to 2.368 m/s (-0.8 %), and fine-tuning the source MLP improves on the local MLP by 2.9 % for 3-h wind and 1.4 % for 24-h wind. With two years of Agadez data the gains shrink (-1.4 % for 24-h wind, -1.1 % for sunshine), and with the full eight-year record they vanish for wind (±0.5 %). For sunshine, pooling keeps a 1.0 % advantage even with the full record.

H3 is supported, with an important qualification. The ordering predicted by H3 holds, with the largest gain in the scarcest scenario, but the gains are a few percent. The local one-year model is itself only 3.2 % worse (3-h wind) and 4.3 % worse (24-h wind) than the full-record local model. Agadez's wind record is incomplete by hour, not by year: its daytime slots are nearly complete over ten years, so the data-scarcity problem that transfer learning is meant to solve is milder than the headline completeness of 63 % suggests. Night-time Agadez wind, the part that really is missing, cannot be evaluated at all.

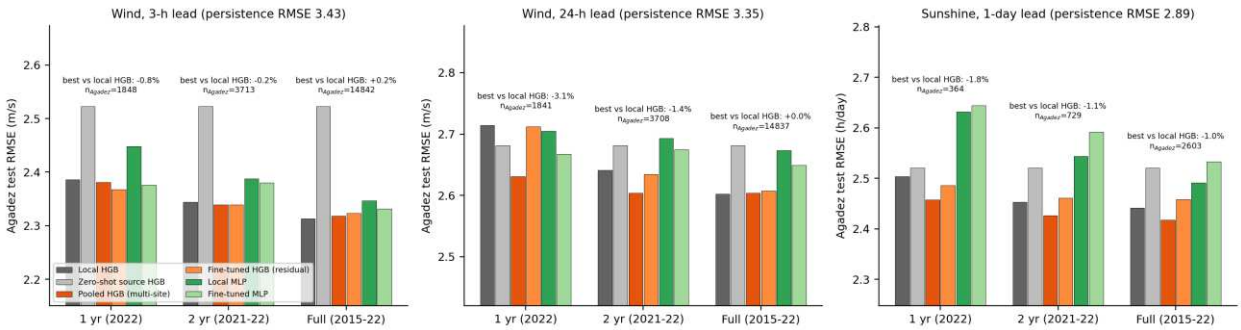


Figure 11. Cross-site transfer to Agadez (source: Bilma and N'Guigmi). Agadez test RMSE (2023-2024) for six methods under three Agadez training scenarios (1 year, 2 years, full 2015-2022 record). Panels: wind at 3-h and 24-h leads (station-normalised training, errors in physical units) and sunshine at 1-day lead. Labels give the best non-local method's change relative to the local HGB and the number of Agadez training samples.

4.8 Learning curves

Figure 12 reports the learning curves of the models actually trained. The MLP's validation error reaches its minimum after 4 to 21 epochs and then rises steadily while training error keeps falling, a clear sign of overfitting even for a 64-32 network. Gradient boosting behaves better. Its validation curve has a shallow minimum between 28 and 138 stages and then a long plateau, so the fixed choices of 100 (wind) and 60 (sunshine) stages sit on the plateau. Validation error falls only slowly with the length of the training record. For 24-h wind it drops from 2.61 to 2.42 m/s at Agadez, 2.14 to 2.03 m/s at Bilma and 1.35 to 1.29 m/s at N'Guigmi between one and six years. For sunshine it drops from 2.63 to 2.49, 1.90 to 1.81 and 2.17 to 2.11 h/day. The curves flatten after about three to four years. More years of the same station variables would bring modest improvements; new information, such as irradiance, aerosol or numerical weather prediction inputs, would be needed for a step change.

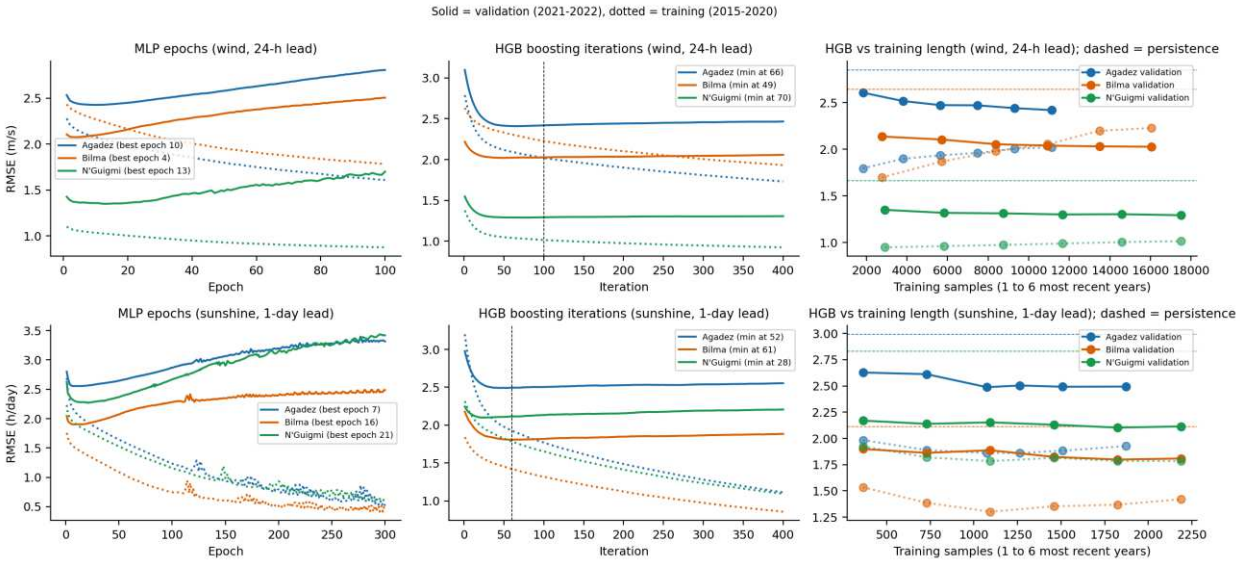


Figure 12. Learning curves on the inner split (fit 2015-2020, validation 2021-2022). Top row: wind, 24-h lead; bottom row: sunshine, 1-day lead. Left: MLP training (dotted) and validation (solid) RMSE by epoch. Centre: HGB RMSE by boosting stage (vertical line: stages used in the final models). Right: HGB validation RMSE against the length of the training record (1 to 6 most recent years; dotted: training RMSE; dashed horizontal: persistence).

4.9 Calibration of the probabilistic forecasts

Figure 13 and Table 5 evaluate the quantile forecasts. All nine station-case combinations have positive CRPS skill against the climatological quantiles. The skill is 0.13-0.24 for wind 3 h ahead, falls to 0.04-0.09 at 24 h, and ranges from 0.16 (Agadez) and 0.15 (Bilma) to only 0.03 (N'Guigmi) for sunshine one day ahead. Calibration differs by site. For wind on the continuous scale, the raw 80 % intervals are close to nominal at Bilma (coverage 0.81 at both leads) and slightly narrow at N'Guigmi (0.76-0.78). At Agadez they are clearly too narrow (0.72 at both leads), and the reliability curves there bend away from the diagonal in both tails. On the reporting scale, rounded interval bounds cover 0.81-0.93 of the integer reports, because a rounded interval absorbs whole speed classes. Sunshine intervals are too narrow at every station (raw 80 % coverage 0.67, 0.77 and 0.73). The Agadez sunshine 90 % interval is unusually wide (10.7 h on average) because its lower bound sits at zero: the zero-sunshine days described in Section 2 make up about 5 % of the Agadez record (149 of them in the training years), and the q_0.05 model learns to allow for them.

Split-conformal recalibration (Table 5, right panel of Figure 13) does what it is designed to do where the raw intervals are too narrow. For Agadez the 80 % coverage rises from 0.72 to 0.76 (wind 3 h), from 0.71 to 0.76 (wind 24 h) and from 0.66 to 0.82 (sunshine), at the cost of intervals 9-18 % wider. Sunshine coverage also rises to 0.80 at Bilma and 0.82 at N'Guigmi. For Bilma wind the conformal correction is essentially zero, since those intervals were already calibrated. At Agadez wind remains below nominal after recalibration (0.76), which points to a change in error behaviour between the 2022 calibration year and the test years; Section 4.12 locates it in 2024. The recalibrated intervals are the better basis for an operational reserve rule, provided their coverage is monitored. Sizing a reserve on the raw Agadez intervals would leave 28-34 % of outcomes outside the nominal 80 % band instead of 20 %.

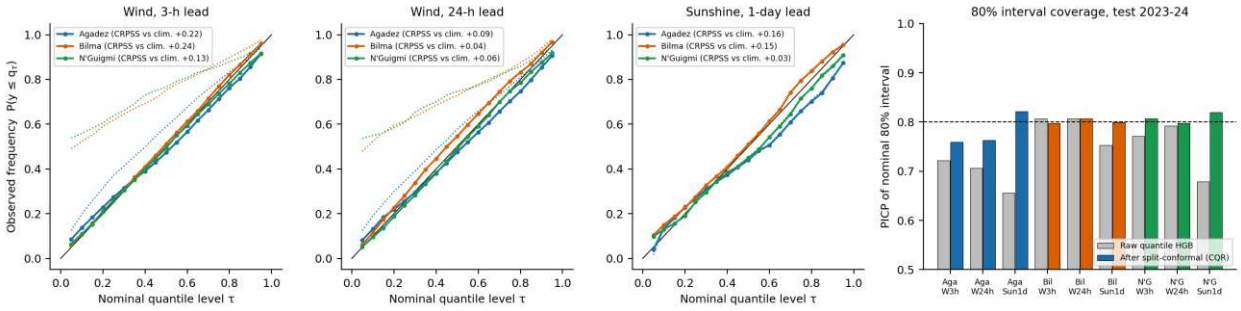


Figure 13. Calibration of quantile gradient-boosting forecasts, test period 2023-2024. Left three panels: reliability diagrams (observed frequency of $y \le q_\tau$ against nominal level τ; dotted: de-discretised observations) for wind at 3-h and 24-h leads and sunshine at 1-day lead; solid: continuous (de-discretised) scale, dotted: wind on the reporting scale (rounded quantiles against integer reports; rounded bounds of 0 absorb all calms, hence the offset at low τ) and sunshine with 0.1-h jitter; legend gives the CRPS skill against training-period climatological quantiles. Right: coverage of the nominal 80 % interval before (grey) and after (colour) split-conformal recalibration on 2022.

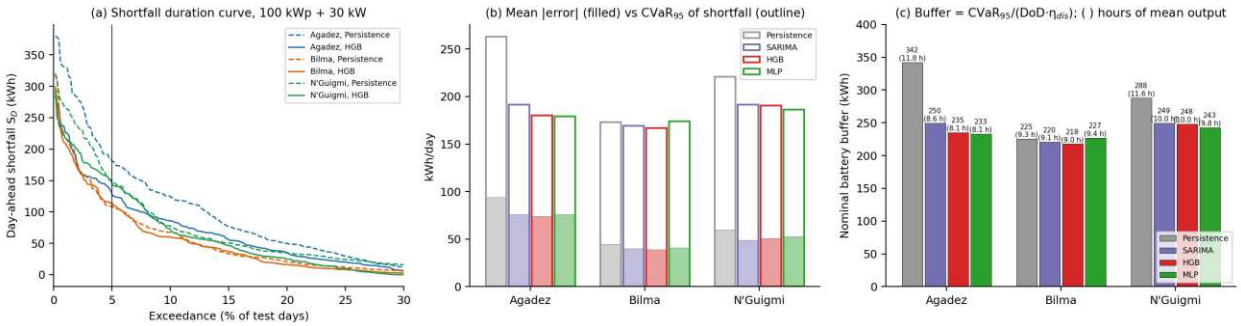
Table 5. Probabilistic forecasts, test period 2023-2024. PICP of the nominal 80 % and 90 % intervals from the raw quantile HGB (fitted 2015-2022) and after split-conformal recalibration (CQR, fitted 2015-2021, calibrated on 2022); mean 80 % width (raw / CQR, m/s or h/day); quantile-grid CRPS of the model and of training-period climatological quantiles, and CRPS skill. Wind evaluated on the de-discretised scale.

Station	Case	PICP80 raw	PICP80 CQR	Width80 raw / CQR	PICP90 raw	PICP90 CQR	CRPS	CRPS clim.	CRPSS
Agadez	Wind 3 h	0.719	0.759	4.38 / 4.73	0.830	0.875	1.238	1.579	+0.216
Agadez	Wind 24 h	0.722	0.763	5.09 / 5.57	0.824	0.893	1.432	1.575	+0.091
Agadez	Sunshine 1 d	0.673	0.821	4.30 / 4.95	0.873	0.917	1.233	1.466	+0.159
Bilma	Wind 3 h	0.807	0.797	2.93 / 2.94	0.900	0.897	0.671	0.879	+0.237
Bilma	Wind 24 h	0.805	0.807	4.13 / 4.13	0.906	0.916	0.843	0.881	+0.043
Bilma	Sunshine 1 d	0.772	0.799	3.06 / 3.25	0.852	0.901	0.906	1.067	+0.151
N'Guigmi	Wind 3 h	0.764	0.807	2.12 / 2.40	0.852	0.894	0.573	0.661	+0.132
N'Guigmi	Wind 24 h	0.780	0.797	2.37 / 2.63	0.865	0.883	0.617	0.658	+0.063
N'Guigmi	Sunshine 1 d	0.729	0.819	4.01 / 4.39	0.811	0.884	1.146	1.181	+0.029

4.10 Operational CVaR and battery buffer

Figure 14 and Table 6 translate the day-ahead forecast errors into energy for the illustrative 100 kWp + 30 kW plant. Modelled PV production averages 435, 434 and 447 kWh/day at Agadez, Bilma and N'Guigmi. The turbine delivers 96 kWh over the 06-21 h window at Agadez, a daytime capacity factor of 21 %, but only 11 kWh (2-3 %) at Bilma and N'Guigmi, where the reported winds rarely exceed the 3 m/s cut-in once extrapolated to hub height. Sunshine errors therefore dominate the shortfall at Bilma and N'Guigmi, and they make up a large share of it at Agadez too. The shortfall distribution is strongly skewed. At Agadez with HGB, the mean daily shortfall is 22 kWh (Table 6), while the CVaR95 is 180 kWh, about eight times larger. A buffer sized on average forecast performance would therefore miss the days that matter.

At Agadez, where wind forecasting has real skill, the nominal buffer (Eq. 22) falls from 342 kWh with persistence to 250 kWh with SARIMA and 235 kWh with HGB (233 kWh with the MLP), from 11.8 to 8.1 hours of mean plant output. HGB's 6 % gain over SARIMA comes almost entirely from the wind term: the CVaR95 of the wind shortfall alone is 120 kWh with SARIMA and 92 kWh with HGB. Whether this gain exceeds its sampling uncertainty is examined in Section 4.11. At Bilma the buffers cluster between 218 and 227 kWh for all non-naive forecasts. At N'Guigmi, SARIMA, HGB and the MLP give 250, 248 and 243 kWh against 288 kWh for persistence. The MLP's slightly lower buffer there comes from a lower PV shortfall tail, although its mean PV error is higher, a reminder that the tail and the mean do not have to agree. The practical reading for mini-grid sizing is that forecast quality justifies a smaller operational reserve mainly at the windy site, and that at sunshine-dominated sites the improvement from any non-naive forecast over persistence (about 14 % at N'Guigmi, 2-3 % at Bilma) matters more than the choice between SARIMA and machine learning.



608

609 **Figure 14.** From forecast error to battery buffer for an illustrative PV-wind plant.

610 Illustrative 100 kWp PV + 30 kW wind plant, day-ahead forecasts issued at 18:00 on D - 1, test period 2023-2024 (n = 691-694
611 days per station with both PV and complete 06-18 h wind forecasts). (a) Duration curve of the daily shortfall S_D (forecast
612 minus modelled production, positive part) for persistence (dashed) and HGB (solid). (b) Mean absolute daily energy error
613 (filled) and CVaR95 of the shortfall (outline) by forecast model. (c) Nominal battery buffer B = CVaR95/(DoD·η_dis), with the
614 equivalent hours of mean plant output in brackets.

615 **Table 6.** Day-ahead energy error and battery buffer for the illustrative PV-wind plant.

616 Illustrative 100 kWp PV + 30 kW wind plant (modelled production from observed sunshine and wind, Section 3.8), test period
617 2023-2024. MAE: mean absolute daily error of forecast PV and 06-21 h wind energy; S: daily shortfall; buffer B from Eq. 22;
618 hours = B-equivalent CVaR95 in hours of mean plant output.

Station	Forecast	PV MAE (kWh/d)	Wind MAE (kWh/d)	Mean S (kWh/d)	VaR95 S (kWh)	CVaR95 S (kWh)	Buffer B (kWh)	Hours of mean output
Agadez	Persistence	51.9	71.2	32.2	179.5	262.7	342	11.8
Agadez	SARIMA	48.9	53.1	26.5	143.9	191.6	250	8.6
Agadez	HGB	49.7	48.7	21.9	125.2	180.3	235	8.1
Agadez	MLP	51.0	50.0	23.0	127.2	179.0	233	8.1
Bilma	Persistence	39.1	12.8	18.0	107.3	172.8	225	9.3
Bilma	SARIMA	34.1	11.0	16.8	105.3	169.1	220	9.1
Bilma	HGB	33.6	10.8	16.8	111.0	166.9	218	9.0
Bilma	MLP	35.9	10.7	18.0	113.3	174.0	227	9.4
N'Guigmi	Persistence	54.5	12.5	25.1	146.9	221.0	288	11.6
N'Guigmi	SARIMA	44.4	9.8	20.6	141.2	191.4	249	10.0
N'Guigmi	HGB	44.8	10.2	20.5	141.8	190.6	248	10.0

N'Guigmi	MLP	47.8	9.9	22.0	144.1	186.3	243	9.8

4.11 Rolling-origin evaluation and statistical significance

Figure 15 and Table 7 show what two test years hide. For Agadez wind, HGB has a lower RMSE than SARIMA in five of the six rolling test years at every lead up to 24 h, with gains of 1.0-4.5 % in those years. The exception is 2021, when HGB is worse at every lead, by 2.0-6.9 %. The loss has an identifiable cause. On 19-21 November 2021 the Agadez 15-h reports reach 26-30 m/s on three consecutive days, far above anything in the 2015-2020 training years, which contain no report of 20 m/s or more. A tree ensemble cannot predict beyond the range of its training targets, so HGB stays near 9 m/s, whereas the linear SARIMA follows the persistent surge (17-22 m/s). At the 3-h lead, the twenty forecasts with the largest excess squared error account for more than the whole of HGB's 2021 deficit. Whether these reports are real surges or recording errors cannot be decided from the file; they pass the physical-bounds test and were kept. Pooled over 2019-2024, the Agadez gain is 1.4-2.2 % up to 24 h, about half of the 2023-2024 value, falls below 1 % at 30-36 h and turns slightly negative beyond. MAE tells the same story (Table 7). Bilma is the more stable case: HGB beats SARIMA at 3, 6 and 9 h in all six years (pooled 4.6, 3.8 and 2.8 %), while from 36 h onward it is slightly worse in most years (pooled -0.6 to -0.8 %). N'Guigmi changes sign between periods. HGB is marginally better than or equal to SARIMA in 2019-2022 at leads up to 21 h and worse in 2023-2024, so the losses reported in Section 4.2 are not a stable property of the site. For sunshine (Figure 15d), the one-day HGB forecast beats the ARIMA-Fourier model in all six years at Agadez and N'Guigmi and in five at Bilma, by 1.5-2.1 % pooled; at two days and beyond the sign changes from year to year.

Table 8 and Figure 16 separate the three levels of evidence defined in Section 3.9. In the fixed 2023-2024 test, the Agadez wind gain is statistically significant from 3 to 30 h (Holm-adjusted Diebold-Mariano $p \leq 0.033$, bootstrap intervals clear of zero). At 36 h the unadjusted p-value is 0.013 but the adjusted value is 0.053, and from 48 h on the gain of 0.6-1.2 % is numerical only, with intervals spanning zero. At Bilma the gain is significant at 6 h after adjustment ($p = 0.003$) and at 3 h only before it ($p = 0.007$, adjusted 0.082); the losses

from 21 h onward are not significant. None of the N'Guigmi losses is significant (unadjusted p = 0.08-0.34), and no sunshine difference survives the Holm adjustment, the closest being the Bilma one-day gain (p = 0.017, adjusted 0.12) and the Agadez three-day loss (p = 0.035, adjusted 0.25). The rolling-origin record changes the picture in both directions. With about 2,150-2,190 test days instead of 700-730, the Bilma wind gain becomes significant up to 15 h, and the one-day sunshine gains at Agadez (2.1 %, adjusted p = 0.008) and N'Guigmi (2.0 %, adjusted p = 0.034) become significant although they were invisible in 2023-2024. The Agadez wind gain, on the other hand, remains significant after adjustment only at 6 and 12 h, largely because the 2021 year lowers the pooled mean and widens the intervals.

Operational relevance is judged on the battery buffer of Section 4.10 (Figure 16d). Relative to persistence, HGB lowers the Agadez buffer by 107 kWh (95 % interval 64 to 149 kWh) and the N'Guigmi buffer by 40 kWh (12 to 66 kWh); SARIMA alone achieves 93 kWh (47 to 137 kWh) and 38 kWh (11 to 64 kWh). These reductions are both significant and large enough to matter for battery sizing. At Bilma even the gain over persistence (8 kWh, -7 to +24 kWh) is not significant, because sunshine errors dominate the shortfall there. Relative to SARIMA, the 15 kWh saving at Agadez (6 %) has an interval of -0.1 to +28 kWh and is positive in 97 % of the bootstrap resamples, so it is probable but not established at the 5 % level; the MLP's 16 kWh saving over SARIMA has an interval of 5 to 28 kWh. At Bilma and N'Guigmi the HGB-SARIMA differences (3 and 1 kWh) lie well inside their intervals (-5 to +10 and -7 to +7 kWh). The evidence therefore supports three statements of decreasing strength. Both machine learning and SARIMA beat persistence in a way that matters operationally at Agadez and N'Guigmi. HGB is significantly more accurate than SARIMA for short-range wind at Agadez and Bilma. Only at Agadez does this accuracy gain plausibly translate into a smaller reserve, by about 6 %, a figure that should be regarded as indicative until it is confirmed on a longer record.

Table 7. Rolling-origin evaluation of Agadez wind forecasts, expanding window, test years 2019-2024.

Model parameters and all training-dependent features re-estimated on data up to 31 December of the previous year. RMSE and MAE of HGB / SARIMA (m/s) and RMSE skill SS of HGB against SARIMA (%) at 3-, 24- and 72-h leads (1,797-1,972 scored forecasts per year and lead). Pooled: all six test years, with the 95 % moving-block bootstrap interval of SS (28-day blocks, 2,000 resamples). The 2023 forecasts are identical to the fixed-origin forecasts of Section 4.2.

Test year	3 h RMSE	3 h MAE	3 h SS (%)	24 h RMSE	24 h MAE	24 h SS (%)	72 h RMSE	72 h MAE	72 h SS (%)
2019	1.892 / 1.948	1.442 / 1.489	+2.9	2.208 / 2.230	1.693 / 1.710	+1.0	2.475 / 2.404	1.913 / 1.866	-3.0
2020	2.040 / 2.122	1.547 / 1.627	+3.9	2.396 / 2.461	1.831 / 1.889	+2.7	2.714 / 2.728	2.083 / 2.097	+0.5
2021	2.348 / 2.197	1.613 / 1.588	-6.9	2.545 / 2.457	1.820 / 1.797	-3.6	2.835 / 2.750	2.045 / 1.975	-3.1
2022	1.993 / 2.088	1.512 / 1.599	+4.5	2.336 / 2.397	1.797 / 1.868	+2.5	2.655 / 2.629	2.029 / 2.045	-1.0
2023	2.163 / 2.260	1.544 / 1.609	+4.3	2.406 / 2.481	1.745 / 1.824	+3.0	2.674 / 2.651	1.981 / 1.994	-0.9
2024	2.427 / 2.535	1.715 / 1.805	+4.3	2.762 / 2.870	2.042 / 2.129	+3.8	2.893 / 2.957	2.183 / 2.231	+2.2
Pooled	2.149 / 2.196	1.561 / 1.618	+2.1 [-0.8, +4.1]	2.445 / 2.486	1.820 / 1.867	+1.6 [+0.0, +3.0]	2.708 / 2.688	2.038 / 2.033	-0.8 [-2.6, +1.1]
Years HGB better			5 of 6			5 of 6			2 of 6

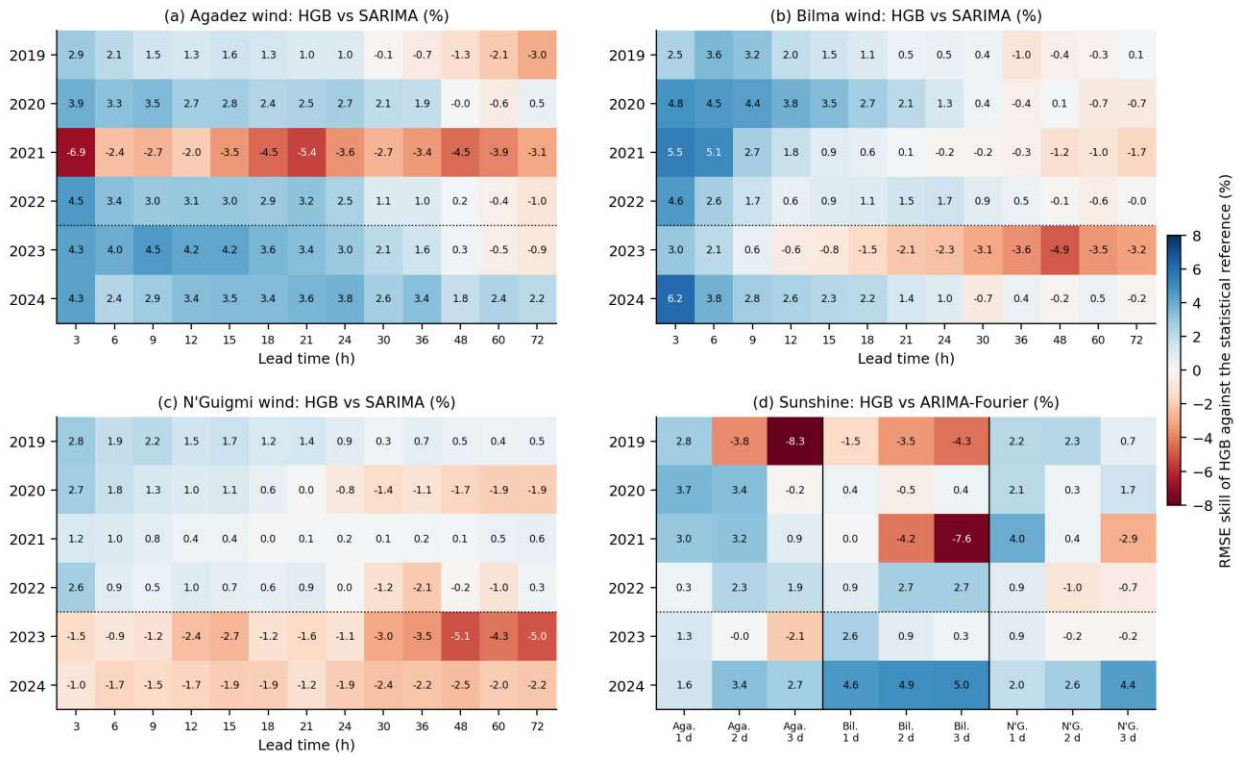
Table 8. Statistical significance of the RMSE difference between HGB and the statistical reference.

SARIMA for wind, ARIMA-Fourier for sunshine. SS: RMSE skill of HGB (%) with its 95 % moving-block bootstrap interval; p: Diebold-Mariano p-value on the daily mean squared-error differential (Bartlett HAC variance, Harvey-Leybourne-Newbold correction), Holm-adjusted value within station and target in brackets; Yrs: number of rolling test years (of six) in which HGB has the lower RMSE. Evidence, fixed / rolling: S+ significant gain (adjusted $p < 0.05$ and interval above zero), N+ numerical gain only, N- numerical loss only.

Station	Target, lead	Fixed 2023-2024 SS (%)	Fixed p (Holm)	Rolling 2019-2024 SS (%)	Rolling p (Holm)	Yrs	Evidence
Agadez	Wind 3 h	+4.3 [+2.4, +5.9]	<0.001 (<0.001)	+2.1 [-0.8, +4.1]	0.160 (1.000)	5	S+ / N+
Agadez	Wind 6 h	+3.0 [+1.5, +4.5]	<0.001 (0.001)	+2.0 [+0.8, +3.2]	0.001 (0.014)	5	S+ / S+
Agadez	Wind 12 h	+3.6 [+2.2, +5.3]	<0.001 (<0.001)	+2.2 [+0.8, +3.4]	0.002 (0.029)	5	S+ / S+
Agadez	Wind 24 h	+3.5 [+2.2, +5.0]	<0.001 (<0.001)	+1.6 [+0.0, +3.0]	0.065 (0.589)	5	S+ / N+
Agadez	Wind 36 h	+2.5 [+0.8, +4.3]	0.013 (0.053)	+0.7 [-0.7, +2.0]	0.355 (1.000)	4	N+ / N+
Agadez	Wind 48 h	+1.2 [-0.7, +3.2]	0.266 (0.797)	-0.5 [-2.3, +1.0]	0.486 (1.000)	3	N+ / N-
Agadez	Wind 72 h	+0.6 [-1.8, +3.3]	0.672 (0.845)	-0.8 [-2.6, +1.1]	0.392 (1.000)	2	N+ / N-
Agadez	Sunshine 1 d	-0.3 [-2.4, +1.8]	0.768 (1.000)	+2.1 [+0.8, +3.4]	0.001 (0.008)	6	N- / S+
Agadez	Sunshine 2 d	-0.8 [-3.0, +1.4]	0.531 (1.000)	+1.4 [-0.6, +3.2]	0.114 (0.685)	4	N- / N+
Agadez	Sunshine 3 d	-2.9 [-5.5, -0.3]	0.035 (0.246)	-1.0 [-3.4, +1.2]	0.330 (1.000)	3	N- / N-
Bilma	Wind 3 h	+5.2 [+2.0, +8.1]	0.007 (0.082)	+4.6 [+3.4, +5.8]	<0.001 (<0.001)	6	N+ / S+

Bilma	Wind 6 h	+3.4 [+1.4, +5.0]	<0.001 (0.003)	+3.8 [+2.8, +4.7]	<0.001 (<0.001)	6	S+ / S+
Bilma	Wind 12 h	+1.8 [-0.5, +3.5]	0.059 (0.592)	+1.9 [+0.9, +2.9]	<0.001 (0.003)	5	N+ / S+
Bilma	Wind 24 h	-0.4 [-3.2, +1.5]	0.646 (1.000)	+0.5 [-0.4, +1.3]	0.260 (0.717)	4	N- / N+
Bilma	Wind 36 h	-0.8 [-4.0, +1.4]	0.458 (1.000)	-0.6 [-1.6, +0.4]	0.239 (0.717)	2	N- / N-
Bilma	Wind 48 h	-1.7 [-4.8, +0.7]	0.121 (1.000)	-0.8 [-1.7, -0.0]	0.089 (0.624)	1	N- / N-
Bilma	Wind 72 h	-1.2 [-3.6, +0.6]	0.224 (1.000)	-0.7 [-1.6, +0.1]	0.091 (0.624)	1	N- / N-
Bilma	Sunshine 1 d	+2.5 [+0.4, +4.2]	0.017 (0.116)	+1.5 [-0.3, +3.1]	0.076 (0.534)	5	N+ / N+
Bilma	Sunshine 2 d	+0.8 [-2.0, +2.5]	0.513 (1.000)	+0.3 [-2.1, +2.1]	0.715 (1.000)	3	N+ / N+
Bilma	Sunshine 3 d	-0.2 [-3.3, +2.0]	0.903 (1.000)	-0.4 [-3.2, +1.7]	0.679 (1.000)	4	N- / N-
N'Guigmi	Wind 3 h	-1.2 [-4.8, +2.1]	0.344 (1.000)	+0.9 [-0.9, +2.4]	0.322 (1.000)	4	N- / N+
N'Guigmi	Wind 6 h	-1.5 [-5.0, +1.9]	0.285 (1.000)	+0.2 [-1.4, +1.8]	0.915 (1.000)	4	N- / N+
N'Guigmi	Wind 12 h	-2.4 [-6.8, +1.6]	0.130 (1.000)	-0.3 [-2.3, +1.4]	0.610 (1.000)	4	N- / N-
N'Guigmi	Wind 24 h	-1.6 [-5.1, +1.7]	0.241 (1.000)	-0.6 [-2.3, +0.9]	0.343 (1.000)	3	N- / N-
N'Guigmi	Wind 36 h	-2.9 [-8.7, +1.8]	0.149 (1.000)	-1.6 [-4.2, +0.3]	0.091 (1.000)	2	N- / N-
N'Guigmi	Wind 48 h	-3.9 [-10.6, +1.0]	0.084 (1.000)	-1.7 [-4.6, +0.5]	0.105 (1.000)	2	N- / N-
N'Guigmi	Wind 72 h	-3.8 [-10.0, +1.0]	0.091 (1.000)	-1.5 [-4.3, +0.7]	0.150 (1.000)	3	N- / N-
N'Guigmi	Sunshine 1 d	-0.0 [-1.7, +1.7]	0.965 (1.000)	+2.0 [+0.8, +3.4]	0.005 (0.034)	6	N- / S+
N'Guigmi	Sunshine 2 d	-0.5 [-2.2, +0.9]	0.525 (1.000)	+0.8 [-0.3, +1.8]	0.144 (0.864)	4	N- / N+
N'Guigmi	Sunshine 3 d	-0.3 [-1.5, +0.8]	0.683 (1.000)	+0.7 [-0.6, +1.9]	0.251 (1.000)	3	N- / N+

676



677

678 **Figure 15.** Rolling-origin skill of HGB against the statistical reference, by test year (2019-2024) and lead time. (a-c) 3-hourly
679 wind against SARIMA(2,0,1)(1,1,1)_8 at Agadez, Bilma and N'Guigmi. (d) Daily sunshine against ARIMA with annual
680 Fourier terms, leads of 1-3 days. For each year Y the models and all training-dependent features were re-estimated on data up to
681 31 December of Y - 1. Blue: HGB better; red: HGB worse. The dotted line separates the four earlier folds from the two years of
682 the fixed-origin test.

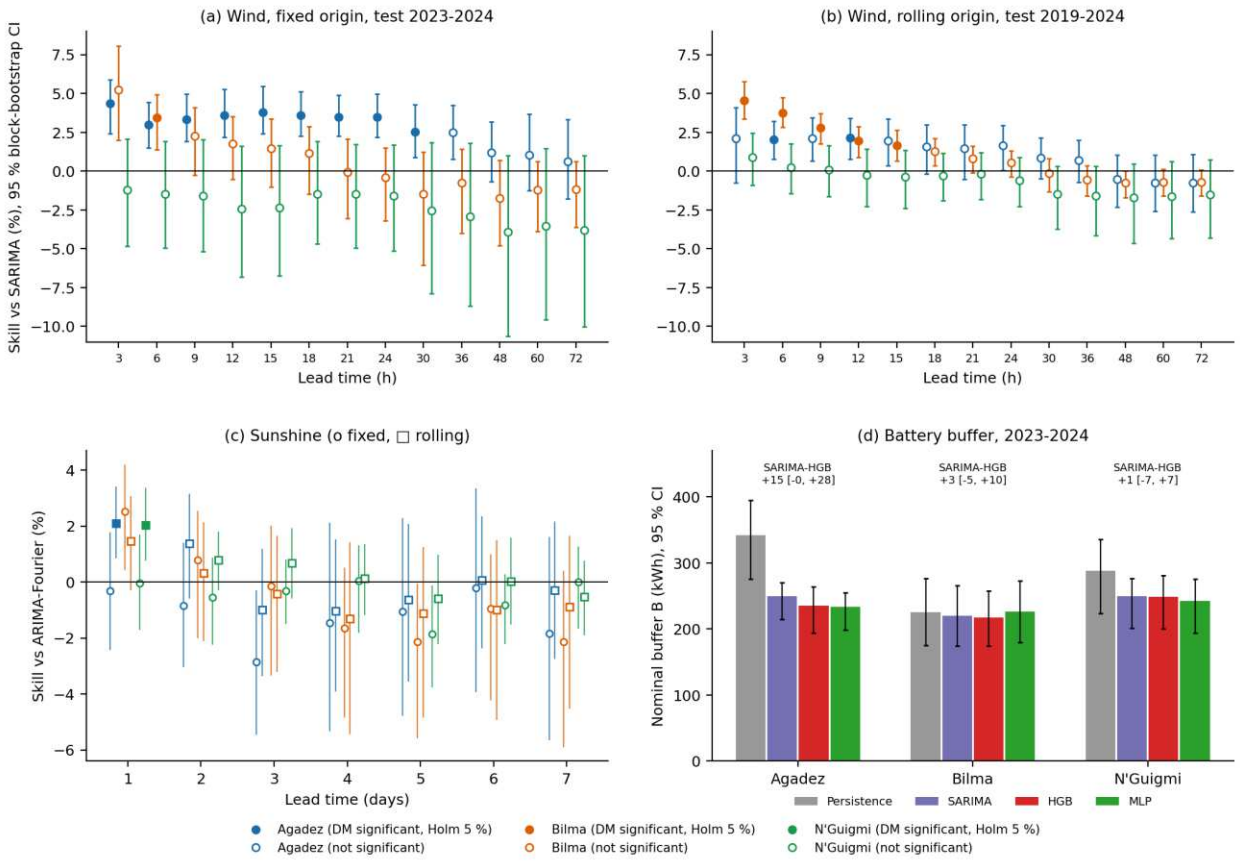


Figure 16. Uncertainty and significance of the model differences. (a) Wind skill of HGB against SARIMA in the fixed-origin test (2023-2024) and (b) pooled over the rolling-origin test years (2019-2024), with 95 % moving-block bootstrap intervals (28-day blocks, 2,000 resamples); filled markers: Diebold-Mariano test significant at 5 % after Holm adjustment within station. (c) Same for sunshine (circles: fixed origin; squares: rolling origin). (d) Nominal battery buffer (Eq. 22) by forecast model for 2023-2024 with 95 % block-bootstrap intervals, and the SARIMA-minus-HGB difference with its interval (kWh).

4.12 Stability of the conformal calibration

Split-conformal recalibration assumes that the conformity scores of the calibration year are exchangeable with those of the test years. Table 9 and Figure 17 test this assumption. At Agadez it holds for 2023 but not for 2024. The 2022 and 2023 scores are statistically indistinguishable (Kolmogorov-Smirnov $p = 0.51$ at 3 h and 0.15 at 24 h), and the recalibrated 80 % interval covers 0.78-0.79 of the 2023 observations. In 2024 the score distribution shifts ($D = 0.067$-0.087, $p < 0.001$): the 80th percentile of the 2024 scores is 0.58 m/s at 3 h, three times the correction of 0.19 m/s estimated on 2022, and coverage falls to 0.74. The 90 % interval shows the

same drop (0.89 to 0.86 at 3 h). A sequential update that recalibrates on 2023 before forecasting 2024 does not repair it (0.76 and 0.73 at 3 and 24 h), because the 2023 scores, whose 80th percentile was 0.29 m/s at 3 h, still understate the 2024 errors. The annual rolling scheme gives the broader picture (Figure 17b-c). Across 2020-2024 the recalibrated Agadez wind coverage ranges from 0.73 to 0.81, against 0.69-0.75 for the raw intervals, and Agadez sunshine from 0.74 to 0.85, against 0.63-0.67 raw. Recalibration thus removes most of the systematic under-coverage but leaves a year-to-year spread of up to eight percentage points. Coverage is not uniform within the year either. For sunshine, the recalibrated intervals over-cover in the dry season and under-cover in the monsoon months at all three stations (0.82-0.85 against 0.76-0.77), while for N'Guigmi wind the opposite happens (0.78 in the dry season and 0.85 in the wet season at 3 h). Bilma wind is the most stable case (0.79-0.82 in every split of the 2023-2024 test), because its raw intervals were already calibrated and the correction is close to zero. Its Kolmogorov-Smirnov tests are nonetheless significant, which indicates that a score distribution can change shape without moving its 80th percentile.

These results qualify the recommendation of Section 4.9. Conformal recalibration is worth applying, since it removed most of the under-coverage where it existed and left well-calibrated cases essentially unchanged, but its finite-sample guarantee is a marginal one over exchangeable data and does not survive a change of regime. The Agadez record carries several signs of such change: the borderline 2018-2019 inhomogeneity and the uneven frequency of zero-sunshine days (Section 2), a reporting schedule for the evening and night slots that varies from year to year, and the November 2021 surge. A reserve rule based on these intervals should therefore carry a margin for the observed year-to-year spread, or the correction should be updated online with a scheme built for distribution shift [52]; both options remain to be tested on measured plant data.

Table 9. Conformal recalibration under non-stationarity, nominal 80 % interval. $\hat{q}$: conformal correction estimated on 2022 (m/s or h/day); coverage after recalibration by test year and season (dry: October-May; wet: June-September); KS p: Kolmogorov-Smirnov p-value comparing 2022 conformity scores with those of 2023 and 2024; Seq. 2024: coverage in 2024 after recalibration on 2023; Rolling: range of annual coverage 2020-2024 (fit to Y - 2, calibrate on Y - 1, test on Y). Wind on the de-discretized scale.

Station	Case	$\hat{q}$ (2022)	2023	2024	Dry	Wet	KS p 2023	KS p 2024	Seq. 2024	Rolling 2020-24
Agadez	Wind 3 h	0.19	0.78	0.74	0.77	0.75	0.508	<0.001	0.76	0.76-0.81
Agadez	Wind 24 h	0.28	0.79	0.74	0.77	0.76	0.146	<0.001	0.73	0.73-0.81
Agadez	Sunshine 1 d	0.38	0.85	0.79	0.85	0.77	0.370	0.062	0.78	0.74-0.85
Bilma	Wind 3 h	-0.01	0.80	0.79	0.80	0.80	<0.001	<0.001	0.79	0.79-0.83
Bilma	Wind 24 h	0.00	0.82	0.80	0.81	0.80	<0.001	<0.001	0.78	0.78-0.85
Bilma	Sunshine 1 d	0.12	0.81	0.79	0.82	0.76	0.099	0.037	0.82	0.76-0.83
N'Guigmi	Wind 3 h	0.03	0.80	0.81	0.78	0.85	0.591	0.469	0.83	0.78-0.82
N'Guigmi	Wind 24 h	0.01	0.79	0.80	0.78	0.84	0.151	0.328	0.82	0.74-0.82
N'Guigmi	Sunshine 1 d	0.32	0.82	0.82	0.84	0.77	0.437	0.480	0.77	0.74-0.82

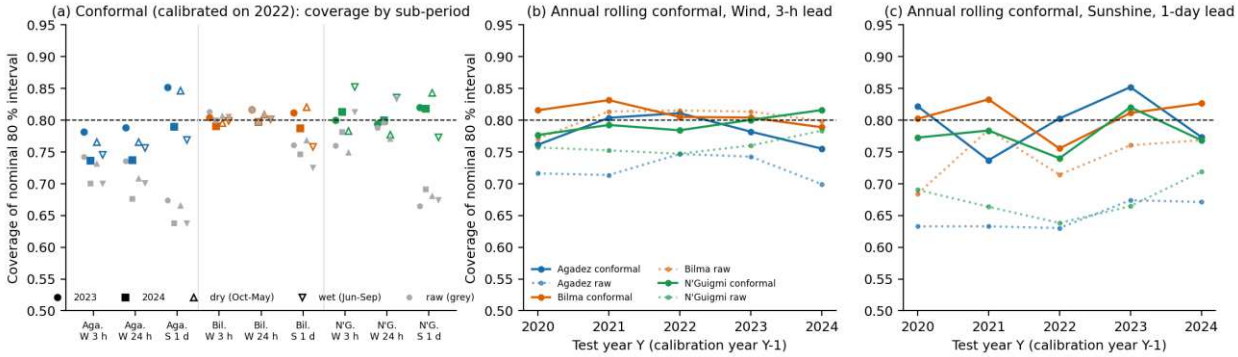


Figure 17. Stability of the split-conformal recalibration under non-stationarity. (a) Coverage of the nominal 80 % interval after recalibration on 2022 (coloured) and before it (grey), by test year (2023, 2024) and by season (dry: October-May; wet: June-September), for wind at 3-h (W 3 h) and 24-h (W 24 h) leads and sunshine at 1-day lead (S 1 d). (b-c) Annual rolling scheme for Y = 2020-2024: quantile models fitted up to Y - 2, calibrated on Y - 1 and tested on Y, with features rebuilt at each cutoff; solid: recalibrated, dotted: raw. Dashed line: nominal 0.80.

4.13 Propagating wind quantiles through the power curve

Table 10 and Figure 18 compare the point-forecast energy of Section 4.10 with the expectation of Eq. 24. At Agadez the mean forecast 10-m speed of the day-ahead slots barely differs between the point forecast and the mean of the predictive distribution (5.03 and 4.96 m/s), yet the expected turbine output is 16.7 kWh per slot against 13.2 kWh for the output at the mean speed. This 26 % gap is the convexity (Jensen) effect of the cubic part of the power curve: speeds in the upper part of the predictive distribution add more energy than the lower

part removes (Figure 18a). Over the 06-21 h window the expected wind energy is 84.7 kWh/day, against 70.1 kWh/day from the point forecast and 96.0 kWh/day computed from the observations. Propagating the quantiles therefore removes more than half of the systematic under-forecast (bias -11.3 against -25.9 kWh/day) while leaving the MAE unchanged (48.7 kWh/day). At Bilma and N'Guigmi the effect is proportionally much larger, because most point forecasts fall below the hub-height cut-in speed and give no output at all (73 % and 83 % of slots). The point method then forecasts 2.4 and 1.7 kWh/day, against 11.4 and 11.0 kWh/day from the observations. The quantile expectation gives 13.3 kWh/day at Bilma, close to unbiased (+1.9 kWh/day), and 4.5 kWh/day at N'Guigmi, where the upper tail of the predictive distribution is still too thin. The expectation is not the best forecast for every purpose: at Bilma, where the modelled output is zero on most days, its MAE is higher than that of the point method (14.4 against 10.8 kWh/day), as one would expect for a mean rather than a median forecast.

The distribution also changes the reserve. Because the point method under-forecasts wind energy, it also under-forecasts the days on which production falls short of the forecast. Recomputing the shortfall of Eq. 21 with the quantile expectation instead of the point forecast raises the CVaR95 buffer from 235 to 248 kWh at Agadez, from 218 to 237 kWh at Bilma and from 248 to 254 kWh at N'Guigmi, that is by 2-9 %. Part of the apparent advantage of the smaller buffers in Table 6 is therefore a by-product of a biased energy forecast rather than a sign of lower risk. The day-specific reserve of Eq. 25 uses the whole predictive distribution instead of a fixed buffer, and its behaviour depends on the dependence assumed between slots and on calibration (Figure 18d). At Bilma and N'Guigmi it averages 78-101 kWh, well below the static in-sample 95th percentile of the shortfall (123 and 151 kWh), but it is exceeded on 8.8-9.9 % of days instead of 5 %, which mirrors the under-coverage of the raw sunshine intervals (Section 4.9). At Agadez the opposite happens: the reserve averages 200 kWh with independent slots and 319 kWh with comonotonic slots, above the static 137 kWh, and it is exceeded on only 2.6 % and 0.1 % of days, because the lower sunshine quantile at Agadez is pinned near zero by the zero-sunshine days. A distribution-based reserve is thus feasible with the models of this study, but it is only as reliable as the calibration of its quantiles, and these figures should be read as an indicative comparison, to be validated against measured plant output.

Table 10. Day-ahead wind and total energy from point forecasts and from propagated quantiles.

Test period 2023-2024 (forecasts issued at 18:00 on D - 1, 06-21 h wind window, illustrative 100 kWp + 30 kW plant). Observed-based: energy modelled from the observed reports (Section 3.8). Convexity gap: E[P(V)] - P(E[V]) of the same predictive distribution. Buffer: nominal buffer of Eq. 22 with the shortfall computed from each forecast. Dynamic reserve: Eq. 25 with independent / comonotonic slots, and the share of days on which the realised shortfall exceeds it (nominal 5 %).

Quantity	Agadez	Bilma	N'Guigmi
Complete wind days / joint PV-wind days	695 / 691	696 / 694	727 / 693
Mean slot speed, point / predictive mean (m/s)	5.03 / 4.96	1.99 / 1.97	1.52 / 1.57
Wind energy, observed-based (kWh/day)	96.0	11.4	11.0
Wind energy, point forecast (kWh/day)	70.1	2.4	1.7
Wind energy, median quantile (kWh/day)	68.7	0.7	1.3
Wind energy, expected from quantiles (kWh/day)	84.7	13.3	4.5
Convexity gap (kWh/day)	17.2	11.6	3.2
Bias, point / expected (kWh/day)	-25.9 / -11.3	-9.0 / +1.9	-9.3 / -6.5
MAE, point / expected (kWh/day)	48.7 / 48.7	10.8 / 14.4	10.2 / 10.2
80 % coverage of daily wind energy, indep. / comon.	0.55 / 0.89	0.86 / 0.91	0.74 / 0.78
Buffer B, point / expected (kWh)	235 / 248	218 / 237	248 / 254
Static VaR95 of shortfall, in-sample (kWh)	137	123	151
Dynamic reserve, mean, indep. / comon. (kWh)	200 / 319	78 / 85	97 / 101
Days exceeding the reserve, indep. / comon. (%)	2.6 / 0.1	9.9 / 8.8	9.1 / 8.8

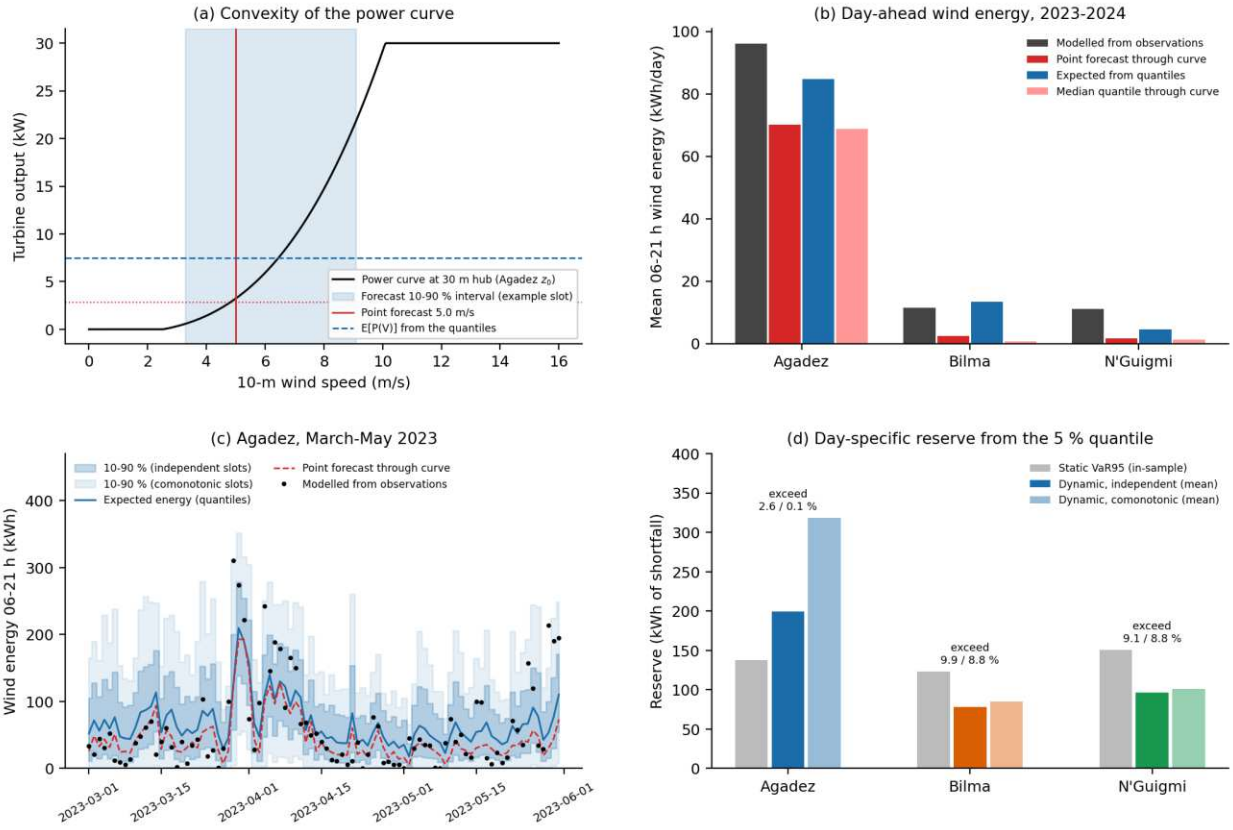


Figure 18. Propagation of wind-speed quantile forecasts through the turbine power curve, day-ahead forecasts for 2023-2024. (a) Power curve at 30-m hub height for the Agadez roughness length, with the 10-90 % interval of one representative forecast slot, its point forecast, the output at that point forecast (dotted) and the expected output from the full predictive distribution (dashed). (b) Mean 06-21 h wind energy per day: modelled from observations, from the point forecast, from the median quantile and expected from the quantiles (Eq. 24). (c) Agadez, March-May 2023: daily wind energy with 10-90 % bands for independent and comonotonic slots. (d) Day-specific reserve R_D (Eq. 25) averaged over the test period, compared with the static in-sample 95th percentile of the shortfall; labels give the share of days on which the realized shortfall exceeded R_D (independent / comonotonic).

5. Discussion

The first lesson is methodological. Once the models were trained on real data and scored against a properly specified SARIMA, the value of machine learning turned out to depend on the site, the lead time and the year. The only gains that are both statistically significant and stable across years are short-range wind gains (up to

15 h at Bilma, and up to about a day at Agadez in five years out of six) and a small one-day sunshine gain at two stations; elsewhere the differences are a few percent in either direction and within sampling noise. A study that compared its network only with other networks, or only with raw persistence, would have reported 20-50 % skill at every station and missed this completely. Beating raw persistence here is almost free, because persistence is badly hurt by the diurnal cycle and by the Agadez night gaps. A single two-year test would also have misled us, in both directions: it made N'Guigmi look like a site where SARIMA is better and hid the 2021 failure of the tree model at Agadez. The persistence-plus-SARIMA protocol, evaluated over several test years with explicit uncertainty, is therefore the transferable contribution of this article, more than any model choice.

The second lesson concerns the physics-derived features. The heat indices and wind-regime labels derived from established station physics are physically well motivated, but with only station variables available they add little beyond what the lagged target and calendar already encode. The one significant degradation (Agadez sunshine on strong-Harmattan days) is a reminder that additional features also add variance when the training set is small. In practice, the next gain for Sahelian forecasting is more likely to come from new information than from new transformations of the same six station variables. The obvious candidates are satellite aerosol or cloud products and the numerical weather prediction fields an operational forecaster would actually use.

The third lesson is operational. The day-ahead shortfall of a PV-dominated plant is governed by sunshine error, and for sunshine no model does materially better than a well-specified statistical reference. At Bilma and N'Guigmi the CVaR95-based buffer is therefore almost the same whichever non-naive forecast is used (218-250 kWh). At Agadez, where the turbine delivers a real share of the energy (a daytime capacity factor of 21 % in this plant configuration, against 2-3 % at the two other sites), better wind forecasts reduce the buffer, from 250 kWh with SARIMA to 235 kWh with HGB, and from 342 kWh with persistence, although only the reduction relative to persistence is clearly larger than its sampling uncertainty (Section 4.11). For the mini-grid sizing problem this means that forecast quality should enter as a site-specific adjustment to battery capacity, not as a blanket reduction. This is consistent with the broader finding that the dynamic performance

of PV-wind-battery mini-grids in this climate is highly site- and configuration-dependent, as shown in a companion climate and environmental performance assessment of such a system at Tafaghatt, Niger [54].

A fourth lesson concerns what the models cannot do. The 2021 Agadez episode exposes the main structural weakness of tree ensembles for this problem: they cannot forecast values outside the range of their training targets, so a rare surge, or a burst of erroneous reports, penalises them far more than a linear state-space model. A hybrid in which the tree model corrects the SARIMA forecast instead of replacing it would keep the linear model's ability to follow such departures and deserves testing. The same record argues for treating calibration as something to monitor rather than a property established once, since a conformal correction estimated on one year was adequate the next year and not the year after. Finally, the energy translation matters as much as the forecast. Passing a point forecast through a convex power curve biased the day-ahead wind energy low by 27 % at Agadez and by 79-85 % at the two low-wind stations, and that bias flowed into a buffer that looked smaller than it should.

6. Limitations

The architecture reported here is a gradient-boosting ensemble and a 64-32 perceptron trained with scikit-learn, rather than the convolutional-recurrent attention network that motivated this study. For roughly 3,600 daily or 29,000 three-hourly observations per station, a model of this size is proportionate. The learning curves show the 64-32 network overfitting within 4 to 21 epochs, so a network with tens of thousands more weights would need a much larger training set or strong pretraining to be justified. Pretraining on reanalysis such as ERA5, ERA5-Land or NASA POWER [55,56] was not part of this study; its benefit therefore remains untested here and should not be inferred from this article.

The wind speed unit is not documented in the station file. It is treated as m/s, and the large difference in mean speed between Agadez (4.8) and the other two stations (1.1-1.8), together with the very different calm fractions, may partly reflect instrument exposure or reporting practice rather than climate alone. Skill scores are ratios and are not affected by a common unit error. The energy-based results of Section 4.10 are affected,

since a knots-to-m/s error would change turbine output through the cubic power curve. Agadez wind skill is daytime skill only. The energy translation relies on assumptions carried in Section 3.8 (default Angström-Prescott coefficients without local calibration, a generic turbine curve, $PR_0 = 0.85$, no soiling) and on an illustrative plant size. The buffers should therefore be read as a comparison between forecast models, not as a design value. The buffers of Table 6 use point forecasts. Section 4.13 suggests that this biases the wind energy low, and the dependence between slots and between PV and wind, which a day-specific reserve needs, is bracketed there by two extreme assumptions rather than modelled. The zero-sunshine days at Agadez are kept as physical, but their uneven frequency across years may partly reflect recording practice, and they widen the lower tail of every Agadez sunshine forecast. The regime labels are themselves model outputs and their uncertainty is not propagated into the forecasts. The CRPS is approximated on a 19-level quantile grid. Finally, the rolling-origin evaluation samples six test years for the point forecasts, but its first folds train on only four or five years, the probabilistic and energy analyses still rest mainly on 2023-2024, and the significance tests are corrected for multiple leads within a station but not across stations.

7. Conclusion

On the measured 2015-2024 record of Agadez, Bilma and N'Guigmi, gradient boosting and a compact neural network forecast daily sunshine and 3-hourly wind far better than persistence. They are not generally better than a carefully specified SARIMA. Gradient boosting gives statistically significant gains only for short-range wind (up to 30 h at Agadez in 2023-2024 and up to 15 h at Bilma over six rolling test years) and a small one-day sunshine gain at Agadez and N'Guigmi; elsewhere the differences lie within sampling noise, and at Agadez the advantage reversed in one year out of six. The compact network is weaker than boosting almost everywhere. Heat and wind-regime features derived from established station physics do not deliver the regime-concentrated gains that were hypothesised. Cross-site transfer helps Agadez by a few percent when local data are scarce. Quantile forecasts need conformal recalibration at Agadez before they can be trusted, and the correction has to be monitored, because its coverage drifted from one test year to the next. At the level that matters for the mini-grid, the CVaR95 of the day-ahead shortfall, a better forecast lowers the

recommended battery buffer by about a third relative to persistence at Agadez. Relative to SARIMA the saving is about 6 % at Agadez and within sampling noise elsewhere, and point forecasts passed through the power curve understate the buffer by 2-9 %. That modest, site-dependent return is the honest answer to how much of the planned capacity an operator can count on the next day.

List of abbreviations

HGB: Histogram gradient-boosting regressor; MLP: Multilayer perceptron; SARIMA: Seasonal autoregressive integrated moving average; ARIMA: Autoregressive integrated moving average; RMSE: Root mean squared error; MAE: Mean absolute error; SS: Skill score; PV: Photovoltaic; TMAX, TMIN: Daily maximum and minimum temperature; UMAX, UMIN: Daily maximum and minimum relative humidity; WBGT: Wet-bulb globe temperature; FAO-56: Food and Agriculture Organization Irrigation and Drainage Paper 56; CRPS, CRPSS: Continuous ranked probability score and its skill score; PICP: Prediction interval coverage probability; CVaR95: Conditional value at risk at the 95th percentile; DoD: Depth of discharge; AIC: Akaike information criterion; DM test: Diebold-Mariano test; HAC: Heteroskedasticity- and autocorrelation-consistent (variance estimator); SNHT: Standard normal homogeneity test; LSTM: Long short-term memory (neural network); CNN: Convolutional neural network; CQR: Conformalized quantile regression; ERA5: Fifth generation ECMWF atmospheric reanalysis; NASA POWER: NASA Prediction of Worldwide Energy Resources; UTC: Coordinated universal time.

Declarations

Availability of data and materials

All results in this article were computed from the station file "Données-météo-Agadez-Bilma-Nguigmi-2015-2024.xlsx" and nothing else. The analysis scripts (quality control, feature construction, statistical references, point, rolling-origin and quantile forecasts, conformal recalibration, energy translation and figures) and the

derived per-sample test predictions supporting this study are available from the corresponding author upon reasonable request.

Competing interests

The author declares that he has no competing interests.

Funding

This research received no external funding from any public, commercial, or not-for-profit funding agency.

Authors' contributions

MGM conceived and designed the study, carried out the data processing, modelling, and statistical analysis, and wrote and approved the final manuscript.

Acknowledgements

The author thanks the National Meteorological Service (DMN) of Niger for the underlying synoptic station records.

Appendix A. Data provenance

Table A1 lists, for this study only, the period retained, the observations used, the values removed or imputed, the variables reconstructed, the external parameters and the status of independent validation.

Table A1. Provenance of the series used in this study. Counts are values present after quality control (Agadez / Bilma / N'Guigmi).

Item	Treatment in this study	Count (Agadez / Bilma / N'Guigmi)
Period and splits	1 January 2015 - 31 December 2024. Training 2015-2022 (inner split 2015-2020 / 2021-2022), fixed test 2023-2024, rolling test years 2019-2024.	3,653 days and 29,224 3-h slots per station
Sunshine duration n	Used as reported; no gap filling; zero-sunshine days kept (Section	3,332 / 3,585 / 3,525

	2).	days
10-m wind speed and direction (3-hourly)	Used as reported, whole units treated as m/s; gaps left missing on the regular grid.	18,511 / 26,114 / 28,531 slots
TMAX / TMIN	Used as reported.	3,649 / 3,619 / 3,619 and 3,648 / 3,619 / 3,619 days
UMAX / UMIN	Used as reported.	3,608 / 3,618 / 3,619 and 3,606 / 3,617 / 3,619 days
Station pressure	Used as reported in the models; training-period monthly mean substituted only in the air-density term of Eq. 20.	3,650 / 3,557 / 2,824 days
Values removed by quality control	Sunshine above the astronomical day length (Agadez 5 March 2022, Bilma 31 July 2015); UMIN > UMAX, both values set to missing (Agadez 9 June and 27 September 2024, Bilma 27 May 2024). No TMIN > TMAX case, no pressure outside 880-1015 hPa, no direction code outside 0-36 and no speed outside 0-60. Rows for non-existent calendar dates in the monthly source layout (e.g. 30 February) discarded before quality control.	3 / 2 / 0 events
Values imputed	None in the record. HGB and SARIMA handle gaps natively; the MLP uses training-period median imputation with missingness indicators.	-
Reconstructed or derived variables	Regular 3-h grid; daily wind aggregates; direction unit vectors; wet-bulb temperature (Eq. 2), WBGT approximation (Eq. 3), near-dawn vapour pressure; hot-day and heat-wave flags; k-means regimes; climatologies; de-discretised wind for probabilistic forecasts; global irradiation (Eq. 18); PV and wind energy (Eqs. 19-20).	-
Training-dependent quantities	Estimated on training years only; re-estimated for every rolling-origin fold and conformal variant.	-
External parameters (assumed)	FAO-56 astronomy and Angström-Prescott coefficients 0.25 / 0.50; $PR_0 = 0.85$, $\gamma = -0.004\ K^{-1}$, NOCT = 45 °C; $z_0 = 0.03 / 0.01 / 0.05$ m; generic turbine with $V_{ci} = 3$, $V_r = 12$, $V_{co} = 25$ m/s and 30-m hub; battery depth of discharge 0.8 and round-trip efficiency 0.92; plant of 100 kWp + 30 kW (illustrative).	-
Independent validation	Not available: the file contains no irradiance, aerosol, visibility or power measurement, and the wind unit is not documented. All energy figures are modelled.	-

References

[1] Moumouni Guero M, Prodjinonto V (2026) Technical design and optimal sizing of a photovoltaic-wind hybrid system for rural electrification: case study of Tafaghatt, Niger. Curr J Appl Sci Technol 45(9):75-86. doi:10.9734/cjast/2026/v45i94749

[2] Ladjal B, Nadour M, Bechouat M, Sedraoui M, Kahal M (2025) Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria. Sci Rep 15:15404. doi:10.1038/s41598-025-94239-z

[3] Obiora CN, Ali A (2021) Hourly photovoltaic power forecasting using CNN-LSTM hybrid model. In: Proceedings of the 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS). IEEE. doi:10.1109/ITMS52826.2021.9615335

[4] Zidoum H, Ibrahim H (2026) A cross-location evaluation of temporal fusion transformers and LSTM for long-term solar PV forecasting in arid environments. Energy Inform 9:73. doi:10.1186/s42162-026-00672-3

[5] Hussan U, Wang H, Peng J, Jiang H, Rasheed H (2026) Transformer-based renewable energy forecasting: a comprehensive review. Renew Sustain Energy Rev 226:116356. doi:10.1016/j.rser.2025.116356

[6] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30 (NeurIPS 2017)

[7] Wang G, Zhou Y, Yan Y, Zhou Z, Yang Z, Dai L, Huang J (2026) Probabilistic photovoltaic power forecasting with reliable uncertainty quantification via multi-scale temporal-spatial attention and conformalized quantile regression. Sustainability 18(2):739. doi:10.3390/su18020739

[8] Amnuaypongsa W, Wangdee W, Songsiri J (2024) Probabilistic solar power forecasting using multi-objective quantile regression. In: 2024 18th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). doi:10.1109/PMAPS61648.2024.10667174

[9] Idogho C, et al (2025) Machine learning-based solar photovoltaic power forecasting for Nigerian regions. Energy Sci Eng. doi:10.1002/ese3.70013

[10] Ojo GO, Alausa OA, Omotola OE (2026) A geospatially-enhanced machine learning framework for solar radiation forecasting in Nigeria. FUDMA J Sci 10(4):290-298. doi:10.33003/fjs-2026-1004-4776

[11] Ul Haq I, Kumar A, Singh Rathore P (2025) Machine learning approaches for wind power forecasting: a comprehensive review. Discov Appl Sci 7:1139

[12] Ayyavu S, Sayeed MS, Abdul Razak SF (2025) A multi-step day-ahead wind power forecasting based on VMD-LSTM-EFG-ABC technique. Energy Inform 8:111. doi:10.1186/s42162-025-00568-8

[13] Zhang Y, Liu Y, Huang K-E, Zhu H, Wang M (2026) Dust-induced losses in solar photovoltaic potential and their long-term changes over North Africa. Earths Future 14:e2025EF006914. doi:10.1029/2025EF006914

[14] Bonkaney A, Madougou S, Adamou R (2017) Impacts of cloud cover and dust on the performance of photovoltaic module in Niamey. J Renew Energy 2017:9107502. doi:10.1155/2017/9107502

[15] Weiss K, Khoshgoftaar TM, Wang DD (2016) A survey of transfer learning. J Big Data 3:9. doi:10.1186/s40537-016-0043-6

[16] Verdone A, Panella M, De Santis E, Rizzi A (2025) A review of solar and wind energy forecasting: from single-site to multi-site paradigm. Appl Energy 392:126016

[17] Weisser D, Hashimoto-Cullen C, Guedj B (2026) Cross-domain offshore wind power forecasting: transfer learning through meteorological clusters. arXiv:2601.19674

[18] Wang Y, Bi Y, Guo Y, Liu X, Sun W, Yu Y, Yang J (2025) A wind and solar power prediction method based on temporal convolutional network-attention-long short-term memory transfer learning and sensitive meteorological features. Appl Sci 15(3):1636. doi:10.3390/app15031636

[19] Depoortere J, Driesen J, Suykens J, Kazmi HS (2025) SolNet: open-source deep learning models for photovoltaic power forecasting across the globe. Int J Forecast 41(3):1223-1236

[20] Tajjour S, Chandel SS, Malik H, Garcia Marquez FP, Alotaibi MA (2025) Daily power generation forecasting for a grid-connected solar power plant using transfer learning technique. Appl Intell 55:383

[21] Schreiber J, Sick B (2022) Model selection, adaptation, and combination for transfer learning in wind and photovoltaic power forecasts. arXiv:2204.13293

[22] Sarmas E, Dimitropoulos N, Marinakis V, Mylona Z, Doukas H (2022) Transfer learning strategies for solar power forecasting under data scarcity. Sci Rep 12:14643

[23] Mohamed MG, Vincent P, Jean-Louis F (2022) Correlation of meteorological parameters to characterize wind sites: a case study of N'Guigmi, Niger. Glob NEST J 24(2):337-343. doi:10.30955/gnj.004155

[24] Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome

[25] Ben Taieb S, Bontempi G, Atiya AF, Sorjamaa A (2012) A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Syst Appl 39(8):7067-7083. doi:10.1016/j.eswa.2012.01.039

[26] Stull R (2011) Wet-bulb temperature from relative humidity and air temperature. J Appl Meteorol Climatol 50(11):2267-2269. doi:10.1175/JAMC-D-11-0143.1

[27] Australian Bureau of Meteorology (n.d.) Thermal comfort observations: about the approximation to the WBGT. http://www.bom.gov.au/info/thermal_stress/

[28] Kong Q, Huber M (2022) Explicit calculations of wet-bulb globe temperature compared with approximations and why it matters for labor productivity. Earths Future 10:e2021EF002334. doi:10.1029/2021EF002334

[29] Chomette O, Legrand M, Marticorena B (1999) Determination of the wind speed threshold for the emission of desert dust using satellite remote sensing in the thermal infrared. J Geophys Res Atmos 104(D24):31207-31215. doi:10.1029/1999JD900756

[30] Sesay MM, Ngunyi A, Imboga H (2026) An explainable hybrid transformer-BiLSTM framework for multivariate household energy demand forecasting with weather integration. Energy Inform 9:33. doi:10.1186/s42162-026-00635-8

[31] Seabold S, Perktold J (2010) Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference, pp 92-96

[32] Durbin J, Koopman SJ (2012) Time series analysis by state space methods, 2nd edn. Oxford University Press, Oxford

[33] Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735-1780

[34] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825-2830

[35] Semmelmann L, Henni S, Weinhardt C (2022) Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data. Energy Inform 5(Suppl 1):24. doi:10.1186/s42162-022-00212-9

[36] Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189-1232. doi:10.1214/aos/1013203451

[37] Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on typical tabular data? Adv Neural Inf Process Syst 35 (NeurIPS 2022), Datasets and Benchmarks Track

[38] Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations (ICLR 2015)

[39] Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33-50. doi:10.2307/1913643

[40] Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. Biometrics 65(4):1254-1261. doi:10.1111/j.1541-0420.2009.01191.x

[41] Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359-378. doi:10.1198/016214506000001437

[42] Romano Y, Patterson E, Candes EJ (2019) Conformalized quantile regression. Adv Neural Inf Process Syst 32 (NeurIPS 2019)

[43] Beichter M, Phipps K, Frysztacki MM, Mikut R, Hagenmeyer V, Ludwig N (2022) Net load forecasting using different aggregation levels. Energy Inform 5(Suppl 1):19. doi:10.1186/s42162-022-00213-8

[44] Breiman L (2001) Random forests. Mach Learn 45:5-32. doi:10.1023/A:1010933404324

[45] Kunsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17(3):1217-1241. doi:10.1214/aos/1176347265

[46] Sandia National Laboratories, PV Performance Modeling Collaborative (n.d.) Sandia module temperature model. https://pvpmc.sandia.gov/modeling-guide/2-dc-module-iv/module-temperature/sandia-module-temperature-model/

[47] Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2(3):21-41

[48] Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. Int J Forecast 16(4):437-450. doi:10.1016/S0169-2070(00)00065-0

[49] Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13(3):253-263. doi:10.1080/07350015.1995.10524599

[50] Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13(2):281-291. doi:10.1016/S0169-2070(96)00719-4

[51] Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6(2):65-70

[52] Gibbs I, Candes E (2021) Adaptive conformal inference under distribution shift. Adv Neural Inf Process Syst 34 (NeurIPS 2021)

[53] Swaroopa Rani B, Jatoth C, Venu S (2026) Explainable AI-driven energy forecasting: a DGMR-based feature extraction and EGST-Net prediction framework for transparent decision-making. Energy Inform 9:49. doi:10.1186/s42162-026-00656-3

[54] Moumouni Guero M, Prodjinonto V (2026) Climate, environmental and dynamic performance assessment of a PV-wind-battery mini-grid in Tafaghatt, Niger. Int J Environ Clim Change 16(9):197-213. doi:10.9734/ijecc/2026/v16i95641

[55] Copernicus Climate Data Store (n.d.) ERA5-Land hourly data documentation. https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=documentation

[56] NASA POWER Project (n.d.) Data sources and methodology, frequently asked questions. NASA Langley Research Center. https://power.larc.nasa.gov/docs/faqs/data/