Solar Resource Exposure to Curtailment: Evidence from Constrained Photovoltaic Plants in Brazil

Felipe De Souza Oliveira[1]
Rafael Morais de Souza[1]
Victor Eduardo de Mello Valério[2]
Douglas Sad Silveira[1]

[1] *Department of Economics, Universidade Federal de Juiz de Fora (UFJF), Brazil*
[2] *Department of Production Engineering, Universidade Federal de Itajubá (UNIFEI), Brazil*

Resumo

Este artigo investiga o *curtailment* fotovoltaico no Brasil. Combinando dados operacionais e meteorológicos, mapeamos a concentração espacial das restrições e avaliamos a previsibilidade da irradiação em usinas congestionadas. Utilizando modelos de aprendizado de máquina e técnicas de interpretabilidade por permutação (individual e agrupada), demonstramos que os cortes são altamente concentrados e a irradiação local segue padrões previsíveis de curto prazo. Essa previsibilidade deriva majoritariamente do ambiente de irradiação e defasagens do recurso, com baixo impacto de variáveis atmosféricas amplas. Os resultados confirmam que modelos preditivos localizados são cruciais para monitorar congestionamentos, precificar riscos, guiar investimentos em armazenamento e otimizar a expansão elétrica nacional.

Abstract

This paper investigates photovoltaic curtailment in Brazil. By combining operational and meteorological data, we map the spatial concentration of these restrictions and evaluate the predictability of the solar resource at congested plants. Using machine learning models and permutation-based interpretability techniques (individual and grouped), we demonstrate that curtailment is highly concentrated and that local irradiation follows predictable short-term patterns. This predictability is primarily driven by the radiation environment and lagged resource conditions, with limited impact from broader atmospheric variables. The results confirm that localized predictive models are crucial for monitoring grid congestion, pricing risks, guiding energy storage investments, and optimizing national grid expansion.

Palavras-chave: *Curtailment* solar; Geração fotovoltaica; Exposição do recurso solar; Aprendizado de máquina.
Keywords: Solar curtailment; Photovoltaic generation; Solar-resource exposure; Machine learning.
Área de Avaliação: Economia Regional
Classificação JEL: C53, L94, Q42, Q48

Acknowledgements: This research project received support from FAPEMIG, CNPQ and CAPES. Felipe de Souza Oliveira thanks FAPEMIG (Grant RED-00161-22) and CAPES (Grant 909693/2025-26) for financial support. Douglas Silveira thanks CNPQ (Grant 169328/2023-6) for financial support.

1 Introduction

Electricity systems require a continuous balance between supply and demand. This requirement becomes harder to satisfy as variable renewable energy sources increase the need for flexibility, balancing, transmission coordination, storage, and demand-side adjustment (Bird et al. 2016; Newbery et al. 2018; Joos and Staffell 2018). Brazil offers a relevant case for this discussion. The country combines a historically renewable electricity matrix, a large hydroelectric base, and rapid growth in photovoltaic generation (Paim et al. 2019; Losekann and Hallack 2018; Brandão et al. 2024). This combination supports the decarbonization of the power sector, but it also creates new integration challenges. Solar generation grows in locations and hours that do not always match the structure of demand or the available capacity of the transmission network (Silva et al. 2026; Operador Nacional do Sistema Elétrico 2025b).

Curtailment gives a visible operational expression to this mismatch. When the system cannot absorb, transmit, or safely accommodate available renewable output, the system operator restricts generation to preserve reliability and balance (Bird et al. 2016; Yasuda et al. 2022). The international literature treats this problem as part of a broader flexibility challenge associated with high shares of solar and wind generation (Newbery et al. 2018; Psarros et al. 2025). One important mechanism behind this challenge is the so-called duck curve, in which high photovoltaic output depresses net load during daylight hours and produces steep ramping requirements later in the day (California ISO 2016; Hou et al. 2019; Olczak et al. 2021). In Brazil, recent operational evidence indicates that these pressures now affect the National Interconnected System, especially during periods of high photovoltaic output, reduced net demand, hydropower availability, and constrained transmission capacity (Operador Nacional do Sistema Elétrico 2025b, 2025a; Silva et al. 2026).

Recent work on Brazil emphasizes the system-level nature of this problem. Silva et al. (2026) show that the rapid expansion of micro and mini distributed generation reshapes the aggregate net-load curve, deepens midday valleys, increases evening ramping requirements, and relates positively to energy-imbalance curtailment in centralized renewable plants. This evidence clarifies how distributed photovoltaic generation changes the operational behavior of the Brazilian system. However, curtailment also has a local and project-level dimension. Restrictions do not affect all plants equally. They concentrate in specific regions, corridors, and grid nodes, where generation assets face different levels of exposure to lost output and revenue risk (Vieira et al. 2025; Silva et al. 2025).

This project-level dimension matters because curtailment affects the economics of photovoltaic investment. O'Shaughnessy, Cruce, and Xu (2021) argue that rising curtailment risk changes the allocation of risk between generators and electricity buyers and can influence contract design, project value, and investment incentives. Related work also emphasizes that curtailment risk interacts with broader renewable-investment risks, including policy uncertainty, market-design choices, and the ability of contractual arrangements to protect revenue streams (Darling et al. 2011; Gatzert and Vogl 2016; Bartlett 2019). This perspective highlights an important point for empirical analysis: curtailment risk depends not only on the existence of grid restrictions, but also on the renewable resource available when those restrictions occur. A plant exposed to frequent grid constraints faces a more severe economic problem when local irradiation remains high and predictable during constrained periods. Therefore, measuring and forecasting the solar resource at constrained nodes can improve the assessment of curtailment exposure.

This paper studies that issue in the Brazilian centralized photovoltaic sector. We ask three related questions. First, where does photovoltaic curtailment concentrate across centralized plants in Brazil? Second, at the most affected locations, can short-run solar irradiation dynamics be predicted with useful accuracy? Third, which information blocks sustain this predictability? These questions complement the system-level perspective in Silva et al. (2026) and the contractual-risk perspective in O'Shaughnessy, Cruce, and Xu (2021). Rather than treating curtailment only as an aggregate flexibility outcome or as an abstract financial risk, we focus on the local solar-resource conditions faced by the plants that experience the strongest grid-imposed restrictions.

To answer these questions, we combine spatial curtailment analysis with localized solar-irradiation forecasting. We first use operational curtailment records from the Brazilian system operator to identify the photovoltaic plants and regions most exposed to grid-imposed restrictions. We then construct plant-level meteorological datasets using NASA POWER hourly information at the selected plant coordinates. NASA POWER has become a useful source for applied renewable-energy and meteorological analysis because it provides spatially consistent atmospheric and solar variables for locations where high-quality ground observations may be limited (Sparks 2018).

The predictive exercise aims to capture short-run changes in solar irradiation and compares a naive daily-persistence benchmark with four standard machine-learning models: Lasso, Ridge, Random Forest, and XGBoost. These models are well suited to time-series prediction problems with many lagged covariates, nonlinear relationships, and potentially correlated predictors (Parmezan, Souza, and Batista 2019; Masini, Medeiros, and Mendes 2023; Brown et al. 2025). The regularized linear models provide parsimonious and stable specifications in a high-dimensional lag structure, while the tree-based models allow nonlinearities, threshold effects, and interactions among meteorological predictors. We evaluate the models in a rolling-window design that mimics a real-time forecasting environment. We also use individual and grouped permutation importance to identify the variables and information blocks that contribute most to predictive performance (Altmann et al. 2010; Molnar 2020; Au et al. 2022; Titz, Pütz, and Witthaut 2024). This interpretability step is central to the paper because the forecasting exercise does not function as a prediction contest. Instead, it helps characterize the information structure behind solar-resource availability in constrained areas.

The results provide three main findings. First, photovoltaic curtailment in Brazil displays strong spatial and plant-level concentration. A limited set of centralized plants accounts for a large share of total curtailed solar generation, with especially relevant restrictions in Minas Gerais and in a broader corridor that reaches parts of the Northeast. This pattern reinforces the importance of moving beyond national aggregation when studying curtailment exposure in Brazil, a limitation that recent system-level work also recognizes (Silva et al. 2026). Second, irradiation at the most constrained plants follows predictable short-run dynamics. Across linear and nonlinear specifications, the machine-learning models improve on the naive benchmark and reveal useful patterns in the evolution of local solar-resource availability. Third, the variable-importance results show that predictive content concentrates in a small number of interpretable blocks. Radiation-environment variables and lagged target values account for most of this content, while broader atmospheric variables play more limited roles.

These findings inform the policy debate on curtailment management. They show that the solar resource exposed to curtailment in Brazil's most constrained photovoltaic nodes does not behave as a purely erratic process. Instead, recent radiation conditions and short-run persistence explain a substantial part of its variation. This evidence can support localized monitoring of constrained plants and improve discussions about curtailment-risk assessment, storage siting, and operational planning. It can also contribute to broader debates on transmission reinforcement, flexibility procurement, demand response, and market design, all of which the literature identifies as relevant instruments for integrating high shares of renewable generation (Denholm et al. 2016; Golden and Paulos 2015; Newbery et al. 2018; Psarros et al. 2025). At the same time, forecasting does not remove the structural causes of curtailment. It provides information that can help operators, investors, and policymakers assess exposure while the system develops the transmission capacity, storage resources, demand response, and market arrangements required for deeper renewable integration.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the curtailment records, plant selection, and meteorological data. Section 4 presents the predictive design and permutation-importance procedures. Section 5 reports the curtailment, forecasting, and variable-importance results. Section 6 concludes.

2 Literature Review

Our paper relates to renewable curtailment, system flexibility, investment risk, and solar-resource forecasting. While variable renewable energy supports decarbonization, it demands enhanced grid flexibility and market adaptation (Bird et al. 2016; Newbery et al. 2018; Joos and Staffell 2018).Curtailment occurs when the system cannot accommodate available generation, revealing infrastructure and operational limits (Bird et al. 2016; Yasuda et al. 2022; Psarros et al. 2025).This is exemplified by the "duck curve" in high-penetration solar systems, where midday generation depresses net load and exacerbates evening ramping requirements (California ISO 2016; Hou et al. 2019; Olczak et al. 2021).

Brazil's historically hydro-based system, now facing rapid photovoltaic expansion, offers a highly relevant setting.[1] Distributed generation reshapes the aggregate net-load curve and exacerbates energy-imbalance curtailment in centralized plants (Silva et al. 2026), raising questions about cost allocation (Stringer et al. 2021). Beyond system-level pressures, curtailment possesses a local dimension, concentrating in specific transmission corridors and grid nodes (Vieira et al. 2025; Silva et al. 2025).We build on this by moving from aggregate analyses to plant-level solar-resource conditions in the most constrained locations.

A second literature strand examines curtailment as an economic and contractual risk. By weakening the link between available output and delivered energy, curtailment alters revenue expectations, risk allocation, and project economics (O'Shaughnessy, Cruce, and Xu 2021; Darling et al. 2011; Gatzert and Vogl 2016; Bartlett 2019). Furthermore, large-scale resource assessments often underrepresent system-integration costs (Heinrichs et al. 2026).Consequently, curtailment risk depends not just on restriction frequency, but on the system's capacity to absorb the solar resource at specific locations.

The third strand develops solar forecasting methods for grid balancing and real-time management (Hua et al. 2015; Liu et al. 2022; Xiao et al. 2023).While the literature offers diverse time-series, machine-learning, and hybrid approaches (Li et al. 2016; Marchesoni-Acland and Alonso-Suárez 2020; Limouni et al. 2023),we utilize these methods differently: to evaluate whether the solar resource exposed to curtailment contains systematic, short-run information for planning.

This objective makes model interpretability paramount.Permutation importance provides a model-agnostic evaluation of individual predictors (Altmann et al. 2010), while grouped permutation importance assesses broader, correlated information blocks (Molnar 2020; Au et al. 2022). This dual approach identifies which information sets drive operational outcomes (Titz, Pütz, and Witthaut 2024; Brown et al. 2025).

Our contribution bridges these strands.We complement aggregate studies in Brazil by focusing on the most constrained centralized plants. By connecting curtailment risk to localized resource availability, and applying interpretable machine-learning methods to assess both individual and grouped variable predictability, we position forecasting as an empirical tool for curtailment-exposure analysis rather than an end in itself.

3 Data

This section describes the two datasets used in the empirical analysis. The first dataset contains operational curtailment records from the Brazilian system operator. We use it to identify where photovoltaic restrictions concentrate and to select the plants exposed to the strongest grid-imposed constraints. The second dataset contains localized meteorological series from NASA POWER. We use it to characterize and predict short-run solar-resource availability at the selected plant locations.

1. Salgueiro et al. (2026) show that distributed PV adoption in Minas Gerais follows spatial and socioeconomic patterns, reinforcing the importance of local heterogeneity in solar-energy outcomes.

3.1 Curtailment records and plant selection

We obtain curtailment records from the public data portal of the Operador Nacional do Sistema Elétrico (ONS).[2] The sample covers the monthly releases from April 2024 to December 2025 and contains observations at 30-minute frequency for 87 centralized photovoltaic plants connected to the Sistema Interligado Nacional (SIN). The dataset includes plants classified by ONS as Type I, Type II-B, and Type II-C. These categories comprise centrally monitored generation assets whose operation can face dispatch restrictions due to system-level constraints.

The ONS records constrained generation through the *constrained-off* framework. We identify a curtailment event when the limited-generation variable contains a valid numerical value. This rule excludes missing entries and keeps the observations in which the operator reports an effective restriction. We then combine the curtailment records with plant-level metadata, including location, submarket, and the official reason for the restriction. These variables allow us to study curtailment by region, by time of day, by operational reason, and by plant.

The curtailment data support the first step of the empirical analysis. We use them to document the spatial distribution of restrictions and to identify the plants most exposed to curtailment. Figure 1 summarizes the geographic distribution of curtailed photovoltaic generation. The pattern shows strong spatial heterogeneity. Minas Gerais concentrates the largest number of events and the largest restricted volumes in the sample. A broader restriction corridor also reaches parts of the Northeast, including Bahia, Piauí, and Ceará. This pattern matches the broader diagnosis that solar capacity has expanded rapidly in high-irradiation regions while transmission reinforcement and system flexibility have not always adjusted at the same pace (Operador Nacional do Sistema Elétrico 2025b; Silva et al. 2026).

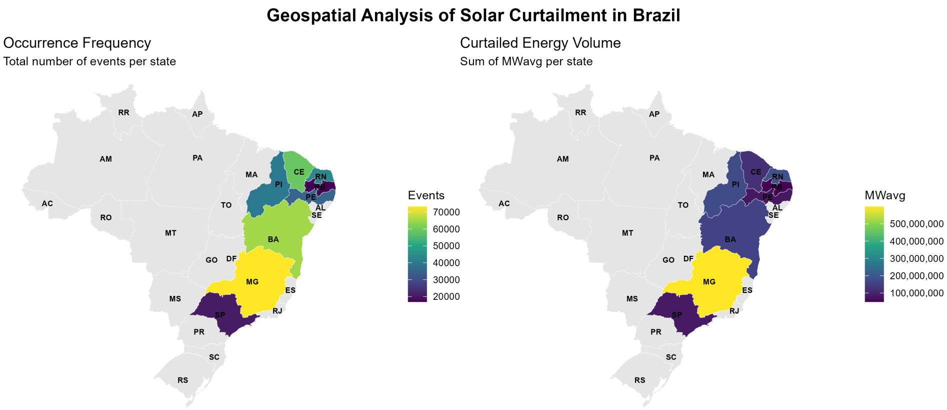


Figure 1: Curtailment spatial distribution.

The curtailment records also reveal a strong intraday pattern. Figure 2 reports restricted generation by hour and by operational reason. Curtailed volumes peak around the middle of the day, when solar irradiation and photovoltaic generation potential are highest. During the main daylight interval, energy-balance restrictions account for most of the curtailed volume. In contrast, reliability restrictions become more relevant during ramping periods, when solar output changes more rapidly and the operator must preserve system security.

2. https://dados.ons.org.br/dataset/

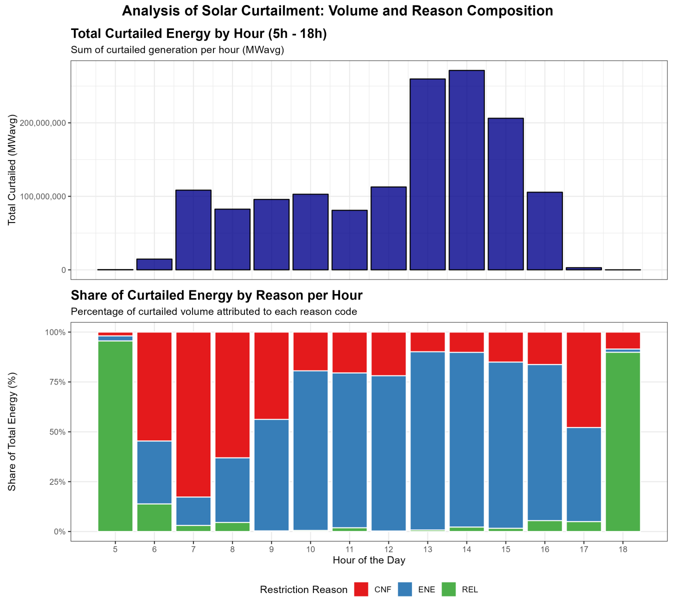


Figure 2: Curtailment amount and restriction reason by time of day.

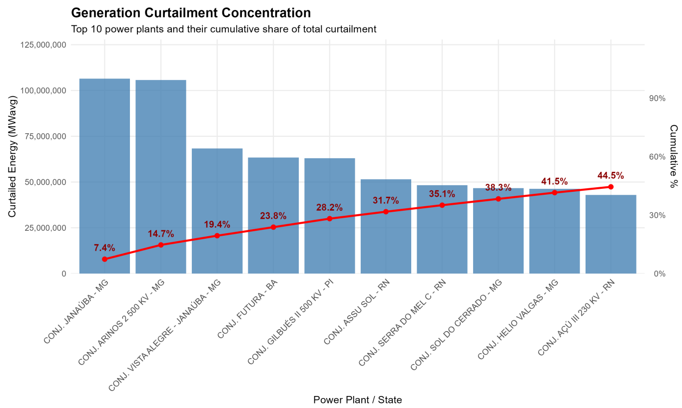


Figure 3: Top 10 photovoltaic plants by curtailment concentration.

Finally, the curtailment data show that restrictions concentrate strongly at the plant level. Figure 3 ranks the ten most affected photovoltaic plants. These plants account for 44.5% of total photovoltaic curtailment in the sample, although they represent only a small fraction of the assets in the dataset. The ranking highlights the Janaúba, Arinos 2, and Vista Alegre complexes in Minas Gerais, as well as Futura in Bahia and Gilbués II in Piauí. This concentration motivates the second step of the empirical analysis: we focus the localized solar-resource exercise on the plants that face the strongest curtailment exposure.

We construct localized hourly meteorological time series for each constrained plant using the NASA POWER project from January 1, 2023, to December 31, 2024. This spatial dataset provides reliable solar and weather variables where ground-station data are limited (Sparks 2018).

Table 1 details the raw variables extracted and their roles. Following the solar-forecasting literature (Succetti et al. 2020; Mohamad Radzi et al. 2023), we use all-sky irradiance as our target and retain a consistent predictor set—including clear-sky irradiance, temperature, moisture, pressure, and wind components—across all machine-learning models. Other variables remain solely for diagnostic purposes.

To capture physical and temporal structures, we construct a solar-intensity index (the ratio of all-sky to clear-sky irradiance) alongside deterministic hour and season indicators. Furthermore, because curtailment concentrates during daylight hours, retaining full hourly series introduces long nighttime zero-value se-

Table 1: Raw NASA POWER variables and role in the predictive exercise

Variable	NASA POWER parameter	Unit	Role
All-sky irradiance	ALLSKY_SFC_SW_DWN	Wh/m^2	Target series
Clear-sky irradiance	CLRSKY_SFC_SW_DWN	Wh/m^2	Predictor
Air temperature	T2M	°C	Predictor
Dew point	T2MDEW	°C	Predictor
Relative humidity	RH2M	%	Predictor
Surface pressure	PS	kPa	Predictor
Precipitation	PRECTOTCORR	mm	Predictor
Eastward wind at 50 m	U50M	m/s	Predictor
Northward wind at 50 m	V50M	m/s	Predictor
Wind speed	WS10M; WS50M	m/s	Raw diagnostics
Wind direction	WD10M; WD50M	degrees (°)	Raw diagnostics
All-sky clearness index	ALLSKY_KT	Unitless	Raw diagnostics

quences that can induce data sparsity and model underfitting (Syntetos and Boylan 2005; Perumean-Chaney et al. 2013; Petropoulos et al. 2022). To preserve the intraday cycle while mitigating this mechanical bias, we downsample the data to six evenly spaced daily observations (01:00, 05:00, 09:00, 13:00, 17:00, and 21:00).

We focus on short-run changes in solar-resource availability. The target is the first difference of the log-transformed all-sky irradiance: $\Delta \log(1 + \text{ALLSKY}_{i,t})$, where i indexes the plant and t the filtered observation. This stabilizes variance and shifts the focus from levels to short-run changes (Hamilton 2020).

Continuous predictors are transformed according to their support: we apply $\log(1+x)$ before differencing to skewed, nonnegative variables (e.g., precipitation), and take simple first differences for the rest (e.g., temperature, wind). We then standardize all continuous differenced predictors by twice their within-location standard deviation to ensure comparable magnitudes and improve regularization (Gelman and Pardoe 2007; Hastie, Tibshirani, and Friedman 2009).

Finally, to simulate a real-time forecasting environment, the models exclude contemporaneous information. We construct the information set using only six lags of the target and meteorological predictors, which perfectly captures the preceding 24-hour cycle.

4 Methodology

This section describes the empirical methods used to evaluate short-run solar-resource predictability at constrained photovoltaic plants.

Subsection 4.1 presents the rolling-window design and defines how the predictive exercise uses past information to forecast short-run changes in solar irradiation. Subsection 4.2 introduces the daily-persistence benchmark and the machine-learning models used in the exercise. Subsection 4.3 defines the performance metrics used to compare predictive accuracy. Subsection 4.4 presents the individual and grouped permutation-importance procedures used to identify the information blocks that sustain predictive performance. The models therefore serve as empirical devices for recovering interpretable information patterns in solar-resource availability, not as proposed forecasting algorithms. Figure 4 summarizes the full sequence of the empirical strategy.

4.1 Forecasting design

We estimate the predictive exercise separately for each selected plant location. Let $y_{i,t}$ denote the transformed solar-irradiation target at location i and time t, and let $\mathbf{X}_{i,t}$ denote the vector of lagged predictors.

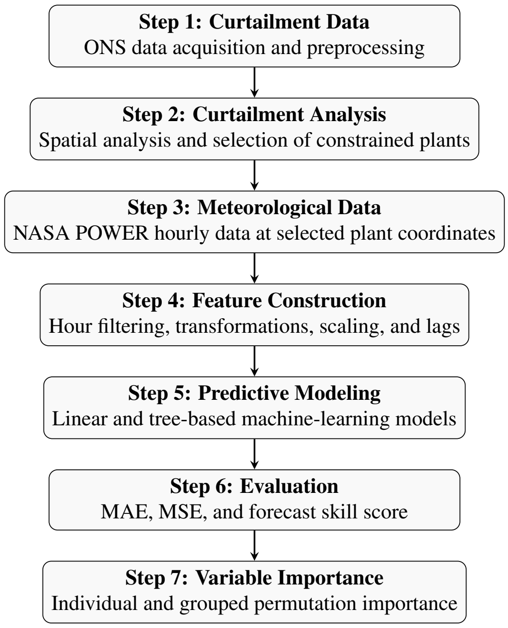


Figure 4: Methodological framework.

The forecasting problem asks whether information observed before t helps predict short-run changes in solar-resource availability at t. We use a rolling-window design to preserve the sequential structure of the data. In each rolling iteration, the model uses a fixed historical window to estimate the forecasting rule and then produces a one-step-ahead prediction. Since the filtered sequence contains one observation every four hours, one step ahead corresponds to a four-hour-ahead forecast.

We use two training windows. The 12-month window serves as the main specification because it allows the models to capture the annual solar cycle. The 3-month window serves as a shorter-memory robustness exercise. After each prediction, the window advances by one observation, and the model updates the training sample before producing the next forecast. This design gives the exercise a real-time interpretation because the model only uses information available before the target realization.

4.2 Benchmark and predictive models

We compare the machine-learning models with a naive daily-persistence benchmark. The benchmark sets the forecast for a selected hour equal to the value observed at the same selected hour on the previous day. Since the restricted sequence contains six observations per day, the naive forecast is

$$\widehat{y}_{i,t}^{Naive} = y_{i,t-6}. \tag{1}$$

This benchmark captures daily repetition in solar-resource dynamics without using additional meteorological predictors. Improvements over this benchmark indicate that the models extract useful information from the lagged radiation, meteorological, and temporal variables.

We then estimate two regularized linear models, Lasso and Ridge.[3] For each location, the linear specification takes the form

$$y_{i,t} = \beta_0 + \mathbf{X}_{i,t}'\beta + \varepsilon_{i,t}. \tag{2}$$

Regularization helps in this setting because the predictor set contains many lagged variables and correlated

3. We use these models because they provide complementary ways to test and interpret the information content of the predictor set.

meteorological covariates. Lasso solves

$$\widehat{\beta}^{Lasso} = \arg\min_{\beta} \left[\frac{1}{N} \sum_{s=1}^{N} \left(y_s - \beta_0 - \mathbf{X}_s' \beta \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right], \tag{3}$$

where $\lambda \geq 0$ controls the penalty. The ℓ_1 penalty shrinks coefficients and can set some coefficients exactly equal to zero, which allows Lasso to perform variable selection (Tibshirani 1996). Ridge solves

$$\widehat{\beta}^{Ridge} = \arg\min_{\beta} \left[\frac{1}{N} \sum_{s=1}^{N} \left(y_s - \beta_0 - \mathbf{X}_s' \beta \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right]. \tag{4}$$

The ℓ_2 penalty shrinks coefficients toward zero but usually does not produce sparse solutions (Hoerl and Kennard 1970). This feature makes Ridge useful when several correlated predictors jointly contribute to the target (Hastie, Tibshirani, and Friedman 2009).

We also estimate two tree-based nonlinear models, Random Forest and XGBoost. Random Forest averages predictions from many regression trees trained on bootstrap samples and random subsets of predictors (Breiman 2001). This structure allows the model to capture nonlinearities and interactions while reducing the instability of a single tree. XGBoost implements gradient tree boosting (Friedman 2001; Chen and Guestrin 2016). It builds an additive sequence of trees, where each new tree targets the residual errors left by the previous ensemble. These nonlinear specifications complement the regularized linear models because they allow threshold effects and interactions among radiation, meteorological, and temporal predictors.

4.3 Evaluation metrics

We evaluate predictive performance with two error metrics and one benchmark-relative measure. Let y_t denote the observed target in the validation sample and let $\widehat{y}_t$ denote the corresponding forecast. The Mean Absolute Error is

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |y_t - \widehat{y}_t|. \tag{5}$$

The MAE measures the average absolute prediction error and treats all deviations linearly. The Mean Squared Error is

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (y_t - \widehat{y}_t)^2. \tag{6}$$

The MSE penalizes larger errors more strongly because it squares the prediction residuals. Reporting both metrics helps us compare average absolute performance and sensitivity to larger forecast errors. We also report a forecast skill score relative to the naive benchmark:

$$SS = \left(1 - \frac{MSE_{Model}}{MSE_{Naive}} \right) \times 100. \tag{7}$$

A positive value indicates that the model reduces MSE relative to the daily-persistence benchmark. A value equal to zero indicates that the model matches the benchmark, while a negative value indicates weaker performance than the benchmark.

4.4 Variable importance measures

The forecasting exercise also aims to identify the information driving predictive performance. We use permutation importance (PI) to measure a predictor's contribution by comparing the median absolute prediction error before and after randomly disrupting it (Altmann et al. 2010). For a fitted model m, loss function $L(\cdot)$, and original data $\mathscr{D}$, the importance of permuting predictor j is:

$$PI_j = L\left(m, \mathscr{D}^{\pi(j)}\right) - L(m, \mathscr{D}). \tag{8}$$

However, individual PI can understate the role of highly correlated meteorological variables. To address this, we also compute Grouped Permutation Importance (GPI) by permuting predefined variable blocks simultaneously (Au et al. 2022; Plagwitz et al. 2022).

We define these groups based on physical interpretation: the radiation environment (clear-sky irradiance and solar-intensity lags), lagged target, temperature and moisture, wind (eastward and northward components), surface pressure, precipitation, and deterministic time structure (hour and season). These groups connect predictive performance to interpretable sources, revealing which information blocks explain solar-resource predictability at constrained plants, rather than identifying the structural causes of curtailment itself.

5 Results

This section reports the empirical results. Subsection 5.1 presents the spatial and plant-level concentration of photovoltaic curtailment and identifies the constrained plants used in the predictive exercise. Subsection 5.2 evaluates the 12-month rolling-window predictive performance across plants and models.Subsection 5.3 discusses grouped permutation importance to identify the main information blocks that sustain predictability. Subsection 5.4 then examines individual permutation importance to show which variables within the dominant blocks contribute most to the forecasts.

5.1 Spatial and Plant-Level Curtailment Concentration

Photovoltaic curtailment in Brazil concentrates in specific regions and hours. The spatial evidence points to Minas Gerais as the main concentration of restricted photovoltaic generation, while the restriction corridor also reaches parts of the Northeast, especially Bahia and Piauí. The intraday evidence reinforces this pattern by showing that curtailed volumes rise during daylight hours and peak around periods of high solar availability. Together, these patterns motivate a localized analysis of curtailment exposure rather than an exclusively national-level assessment. The plant-level ranking sharpens this point: the ten most affected photovoltaic plants account for 44.5% of total photovoltaic curtailment in the sample. This concentration indicates that a relatively small group of centralized plants absorbs a large share of the restriction burden. It also provides the empirical basis for selecting the locations used in the predictive exercise.

Figure 5 reports the locations selected for the predictive analysis. The selected set includes the Janaúba and Vista Alegre area in Minas Gerais, Arinos 2 in Minas Gerais, Futura in Bahia, and Gilbués II in Piauí. Together, these plants represent the main curtailment-exposure nodes identified in the plant-level ranking. Since Vista Alegre lies close to Janaúba and shares the same local meteorological environment, we use the Janaúba meteorological location to represent both complexes in the forecasting exercise.[4]

This selection links the curtailment evidence to the predictive exercise. The next subsection evaluates whether solar-resource availability at these constrained locations follows predictable short-run dynamics under the 12-month rolling-window design.

4. Vista Alegre is located in the same municipality as Janaúba and lies less than 20 km from the Janaúba complex. Given this spatial proximity, the localized NASA POWER series for Janaúba provides a suitable meteorological representation for both plants.

Figure 5: Locations selected for the predictive exercise.

5.2 Forecast Performance

Table 2 reports the predictive performance of the 12-month rolling-window specification. The table compares the daily-persistence benchmark with the linear and nonlinear machine-learning models. The error metrics refer to the transformed target, defined as the first difference of the log-transformed all-sky irradiance series.

The results show that all machine-learning models improve on the naive benchmark in every location. This pattern indicates that solar-resource availability at constrained photovoltaic nodes contains predictable short-run structure beyond daily persistence. The magnitude of the gains varies across models and plants, but the overall result is stable: the lagged radiation, meteorological, and temporal information used by the models improves the prediction of short-run irradiation changes.

Table 2: Forecast performance by location, 12-month rolling window

Location	Model	MAE	MSE	Skill Score (MSE)
	Naive	0.1449	0.0905	Baseline
	Lasso	0.1208	0.0491	45.75%
Arinos	Ridge	0.1092	**0.0353**	**60.99%**
	Random Forest	0.1048	0.0438	51.60%
	XGBoost	**0.1033**	0.0399	55.91%
	Naive	0.1226	0.0453	Baseline
	Lasso	0.1041	0.0290	35.98%
Futura	Ridge	**0.0936**	**0.0213**	**52.98%**
	Random Forest	0.0956	0.0269	40.62%
	XGBoost	0.1002	0.0283	37.53%
	Naive	0.1297	0.0741	Baseline
	Lasso	0.1116	0.0417	43.72%
Gilbués II	Ridge	0.1027	**0.0312**	**57.89%**
	Random Forest	**0.0921**	0.0330	55.47%
	XGBoost	0.0996	0.0349	52.90%
	Naive	0.1300	0.0667	Baseline
	Lasso	0.1171	0.0436	34.63%
Janaúba	Ridge	0.1062	**0.0314**	**52.92%**
	Random Forest	**0.1049**	0.0392	41.23%
	XGBoost	0.1067	0.0373	44.08%

Ridge provides the most consistent performance when we evaluate models by MSE. Its skill score ranges from 52.92% in Janaúba to 60.99% in Arinos, and it records the lowest MSE in all four locations. This result suggests that the predictive signal is not purely sparse. Instead, several correlated predictors appear to contribute jointly to the forecast, which is consistent with the structure of the meteorological feature set. Lasso also improves on the naive benchmark in all locations, although its gains are smaller than those obtained with Ridge. The tree-based models also perform well. Random Forest and XGBoost improve on the naive benchmark across all 12-month comparisons and, in several cases, deliver the lowest MAE. XGBoost records the lowest MAE in Arinos, while Random Forest records the lowest MAE in Gilbués and Janaúba. These results reinforce the interpretation that nonlinear specifications capture useful patterns in local solar-resource availability, even though Ridge remains the most stable specification under the squared-error criterion.

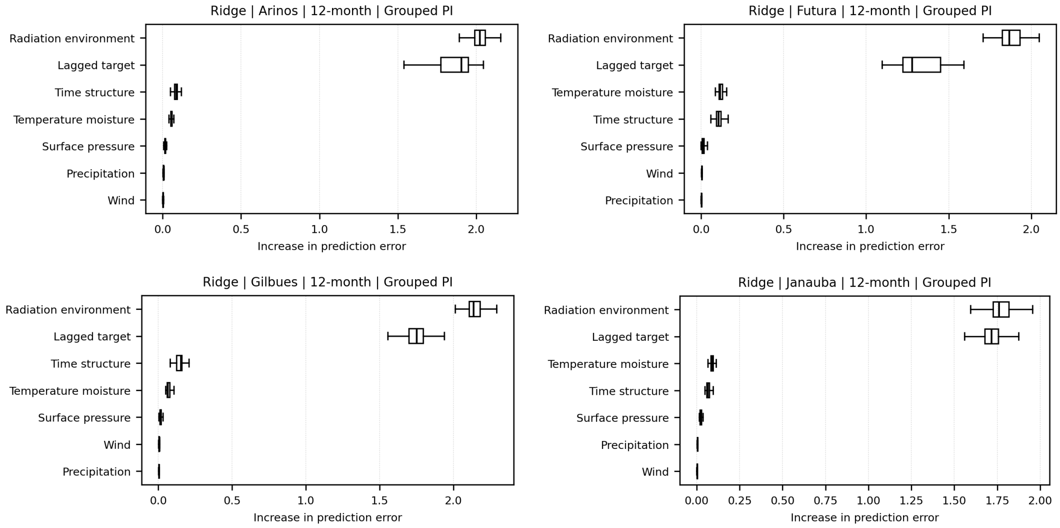
Overall, the performance results support the main premise of the predictive exercise. The objective is not to identify a single dominant algorithm, but to verify whether the exposed solar resource at constrained plants contains predictable short-run structure. The evidence in Table 2 supports this interpretation across both regularized linear and tree-based specifications. This motivates the permutation-importance analysis that follows, where we examine which information blocks explain the predictive gains.

5.3 Grouped Permutation Importance

We next examine grouped permutation importance to identify the information blocks that sustain predictive performance. Figures 6 and 7 report the results for Ridge and Random Forest under the 12-month rolling-window specification. We focus on these two models because they provide complementary perspectives: Ridge summarizes the regularized linear structure, while Random Forest allows nonlinearities and interactions among predictors.

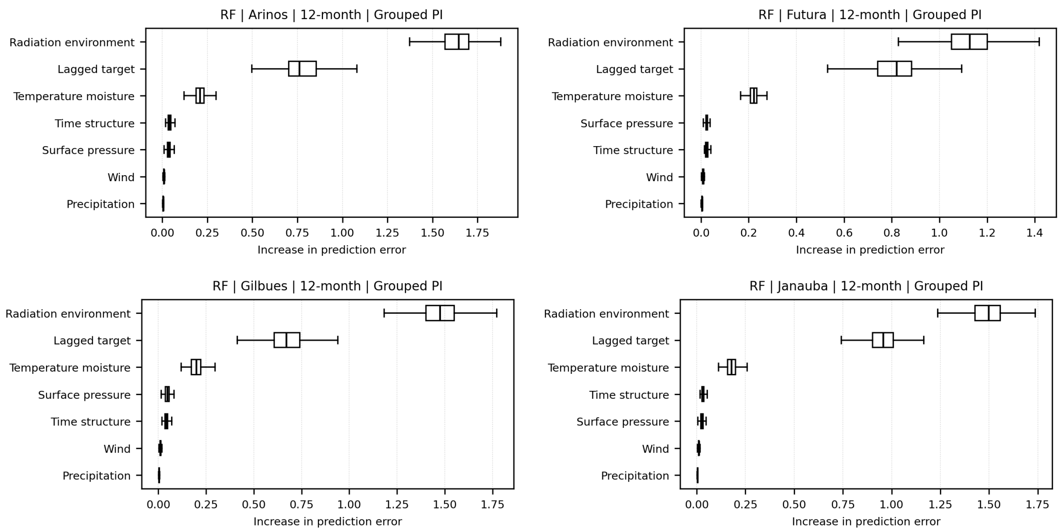
The Ridge results in Figure 6 show a stable hierarchy across locations. The *radiation environment* group ranks first in Arinos, Futura, Gilbués II, and Janaúba. Since this group includes clear-sky irradiance and solar-intensity lags, its dominance indicates that the model relies heavily on the recent local radiation benchmark and on deviations from clear-sky conditions. The *lagged target* group ranks second in all locations, showing that short-run persistence in the transformed irradiation series also carries substantial predictive content. The remaining groups play a much smaller role in the Ridge specification. Time structure and temperature-moisture variables contribute modestly, while surface pressure, precipitation, and wind remain close to zero across locations. This pattern suggests that the regularized linear model does not obtain its predictive gains from a diffuse set of atmospheric variables. Instead, it relies mainly on two interpretable information blocks: recent radiation conditions and the recent history of the solar-resource target.

Figure 6: Grouped permutation importance for Ridge



The Random Forest results in Figure 7 reinforce the same substantive conclusion. The *radiation environment* group remains the most important block in all four locations, and the *lagged target* group consistently ranks second. This consistency matters because Random Forest allows nonlinear relationships and interactions among predictors. The fact that both Ridge and Random Forest point to the same two leading groups suggests that the main predictive structure does not depend on a specific functional form. Random Forest assigns a somewhat larger role to the *temperature-moisture* group than Ridge does, especially in Futura, Gilbués II, and Janaúba. This result suggests that the nonlinear specification captures additional interactions between irradiation dynamics and local atmospheric conditions. Even so, this information remains secondary to the radiation-environment and lagged-target groups. Surface pressure and time structure contribute modestly, while wind and precipitation remain close to zero across locations.

Figure 7: Grouped permutation importance for Random Forest



These results show that solar-resource predictability at constrained photovoltaic nodes comes mainly from recent radiation conditions and short-run persistence in the target series. Other meteorological variables add information, especially under nonlinear specifications, but they do not drive the forecasts. Grouped permutation importance therefore helps identify the broader information blocks behind predictive performance without overinterpreting individual variables that may be highly correlated.

5.4 Individual Permutation Importance

We now examine individual permutation importance to identify which variables account for the patterns observed in the grouped analysis. This step complements the grouped permutation results, but it requires a more cautious interpretation. Since several meteorological predictors are correlated across lags and variable groups, individual permutation scores may distribute predictive content across related variables. For this reason, we use the individual results mainly to unpack the dominant information blocks identified in Subsection 5.3.

Figure 8: Top 10 individual permutation importance for Ridge

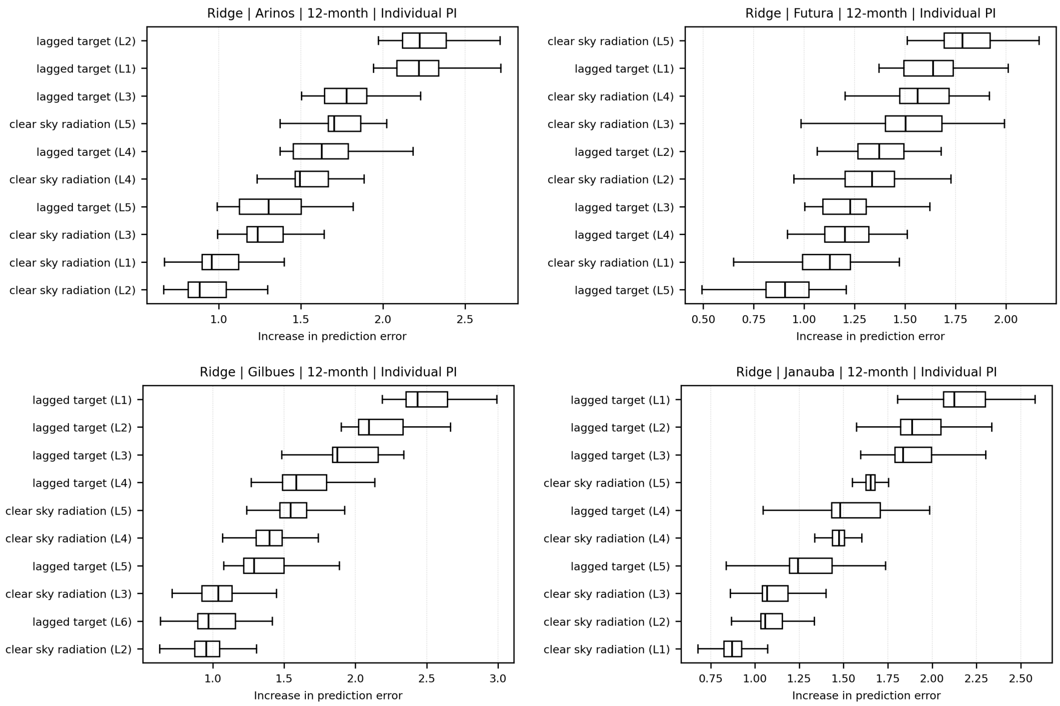


Figure 8 reports the top ten individual predictors for Ridge. The results show that the two dominant groups in the grouped analysis, *radiation environment* and *lagged target*, are also the main sources of individual predictive content. Across locations, the top variables consist almost entirely of lagged target values and clear-sky radiation lags. This pattern reinforces the interpretation that Ridge relies on short-run persistence in the transformed irradiation series and on the local radiation benchmark.

The relative ordering differs across locations, but the substantive message remains stable. In Janaúba and Gilbués II, the first lags of the target variable appear among the most important predictors, indicating a strong role for recent persistence. In Futura, clear-sky radiation lags occupy several of the highest positions, suggesting a stronger role for the deterministic radiation environment. Arinos combines both patterns, with leading target lags and clear-sky radiation lags appearing among the top predictors. Thus, Ridge spreads predictive content across multiple lags within the same two information blocks rather than relying on a broad set of unrelated meteorological variables.

Figure 9: Top 10 individual permutation importance for Random Forest

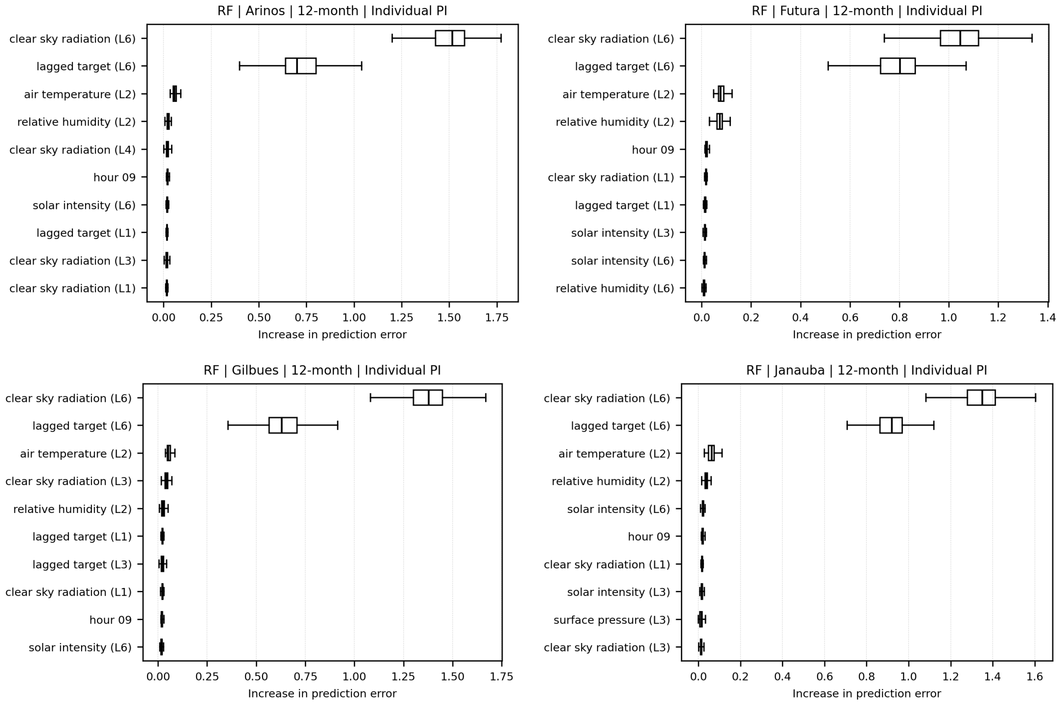


Figure 9 reports the corresponding results for Random Forest. The nonlinear specification produces a more concentrated individual-importance pattern. In all locations, the most important predictor is a clear-sky radiation lag, especially the sixth lag, which corresponds to the same selected hour on the previous day. The lagged target at the sixth lag also ranks among the most important predictors. This result is consistent with the daily-persistence structure of solar irradiation and with the grouped finding that radiation-environment and lagged-target variables dominate predictive performance.

Random Forest also gives modest importance to some temperature and humidity variables, especially second-lag air temperature and relative humidity. This result is consistent with the grouped permutation evidence, where the *temperature-moisture* block plays a somewhat larger role for Random Forest than for Ridge. Still, these variables remain far below the leading radiation and target lags. Other variables, such as hour indicators, solar-intensity lags, surface pressure, wind, and precipitation, enter the top ten only occasionally and with small importance values.

Taken together, the individual permutation results reinforce the grouped analysis. The models do not rely on a diffuse collection of meteorological predictors. Instead, the individual variables with the highest importance generally come from the same two information blocks: clear-sky radiation and lagged target values. Ridge distributes importance across several recent lags, while Random Forest concentrates importance more sharply on the previous-day timing embedded in the sixth lag. This distinction reflects differences in model structure, but it does not change the main conclusion: solar-resource predictability at constrained photovoltaic plants comes primarily from the recent radiation environment and from persistence in local irradiation dynamics.

6 Conclusions

Curtailment has become an important challenge for photovoltaic integration in power systems with growing shares of variable renewable generation. The problem does not only concern aggregate flexibility. It also creates plant-level exposure when photovoltaic assets remain able to generate but face grid-imposed restrictions. This paper studies that exposure in Brazil by combining operational curtailment records from

ONS with localized solar-resource information for the centralized photovoltaic plants most affected by restrictions. The curtailment evidence shows strong spatial and plant-level concentration, which motivates a localized analysis rather than an exclusively national-level assessment.

The predictive results show that short-run solar-resource availability at these constrained nodes contains useful structure. Across linear and nonlinear machine-learning models, the forecasts generally improve on the naive daily-persistence benchmark. The permutation-importance results show that this predictive content concentrates mainly in two interpretable blocks: the radiation environment and lagged target values. Broader atmospheric variables, such as wind, precipitation, and surface pressure, play a more limited role. These findings indicate that local solar-resource predictability at constrained nodes depends primarily on recent radiation conditions and short-run persistence.

The results contribute to the renewable-integration debate by showing that the resource exposed to curtailment can be characterized with localized meteorological information and interpretable predictive tools. This evidence can support monitoring of constrained photovoltaic nodes, curtailment-exposure assessment, storage-siting discussions, and operational planning. It can also help system planners and renewable-energy investors identify where grid restrictions coincide with predictable solar-resource availability. At the same time, forecasting does not remove the structural causes of curtailment. Transmission constraints, limited flexibility resources, demand-side rigidity, and market-design choices remain central elements of the broader integration problem.

The analysis also has limitations. We study the predictability of the solar resource exposed to curtailment, but we do not model the full operational and economic chain that links resource availability, dispatch restrictions, compensation rules, and investment decisions. Future work could build on this evidence by combining localized resource forecasts with curtailment-event models, revenue data, or storage-operation frameworks. These extensions would help evaluate how predictable solar-resource exposure translates into economic losses, mitigation opportunities, and policy design in systems with rising photovoltaic penetration.

References

Altmann, André, Laura Toloşi, Oliver Sander, and Thomas Lengauer. 2010. "Permutation importance: A corrected feature importance measure." *Bioinformatics* 26 (10): 1340–1347.

Au, Quang, Julia Herbinger, Clemens Stachl, Bernd Bischl, and Giuseppe Casalicchio. 2022. "Grouped feature importance and combined features effect plot." *Data Mining and Knowledge Discovery* 36:1401–1450.

Bartlett, Jay. 2019. *Reducing risk in merchant wind and solar projects through financial hedges.* Technical report. Resources for the Future.

Bird, Lori, Debra Lew, Michael Milligan, E. Maria Carlini, Ana Estanqueiro, Damian Flynn, Emilio Gomez-Lazaro, et al. 2016. "Wind and solar energy curtailment: A review of international experience." *Renewable and Sustainable Energy Reviews* 65:577–586.

Brandão, S. Q., Erik Eduardo Rego, R. V. Pillar, and R. N. F. de Carvalho. 2024. "Hydropower enhancing the future of variable renewable energy integration: A regional analysis of capacity availability in Brazil." *Energies* 17 (13).

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Brown, David P, Daniel O Cajueiro, Andrew Eckert, and Douglas Silveira. 2025. "Evaluating the role of information disclosure on bidding behavior in wholesale electricity markets." *Energy Economics* 146:108505.

California ISO. 2016. *What the duck curve tells us about managing a green grid.* Technical report. Fast Facts technical brief. Folsom, CA: California Independent System Operator. https://www.caiso.com/documents/flexibleresourceshelprenewables_fastfacts.pdf.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794.

Darling, Seth B., Fengqi You, Thomas Veselka, and Alfonso Velosa. 2011. "Assumptions and the levelized cost of energy for photovoltaics." *Energy & Environmental Science* 4 (9): 3133–3139.

Denholm, Paul, Josh Novacheck, Jennie Jorgenson, and Matthew O'Connell. 2016. "Impact of flexibility options on grid economic carrying capacity of solar and wind: Three case studies."

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics,* 1189–1232.

Gatzert, Nadine, and Nicole Vogl. 2016. "Evaluating investments in renewable energy under policy risks." *Energy Policy* 95:238–252.

Gelman, Andrew, and Iain Pardoe. 2007. "Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components." *Sociological Methodology* 37 (1): 23–51.

Golden, Ryan, and Bentham Paulos. 2015. "Curtailment of renewable energy in California and beyond." *The Electricity Journal* 28 (6): 36–50.

Hamilton, James D. 2020. *Time Series Analysis.* Princeton University Press.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Springer.

Heinrichs, Heidi Ursula, Jann M. Weinand, Sebastian Kebrich, Juan Camilo Gómez Trillos, Shuying Chen, Maxmilian Hoffmann, Tsamara Tsani, Ruihong Chen, Wenxuan Hu, Johannes Schmidt, et al. 2026. "Large-scale resource assessments for solar photovoltaics: A review of potential definitions, methodologies and future research needs." *Renewable Energy,* 125080.

Hoerl, Arthur E, and Robert W Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55–67.

Hou, Qian, Ning Zhang, Ershun Du, Ming Miao, Fei Peng, and Chongqing Kang. 2019. "Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China." *Applied Energy* 242:205–215.

Hua, Xie, Yang Le, Wang Jian, and Vassilios Agelidis. 2015. "Short-term power forecasting for photovoltaic generation based on wavelet neural network and residual correction of Markov chain." In *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC),* 1–5. IEEE.

Joos, Michael, and Iain Staffell. 2018. "Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany." *Renewable and Sustainable Energy Reviews* 86:45–65. https://doi.org/10.1016/j.rser.2018.01.009.

Li, Jiaming, John K. Ward, Jingnan Tong, Lyle Collins, and Glenn Platt. 2016. "Machine learning for solar irradiance forecasting of photovoltaic system." *Renewable Energy* 90:542–553.

Limouni, Tariq, Reda Yaagoubi, Khalid Bouziane, Khalid Guissi, and El Houssain Baali. 2023. "Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model." *Renewable Energy* 205:1010–1024.

Liu, Caicheng, Ming Li, Yunjun Yu, Ziyang Wu, Hai Gong, and Feier Cheng. 2022. "A review of multitemporal and multispatial scales photovoltaic forecasting methods." *IEEE access* 10:35073–35093.

Losekann, Luciano, and Michelle Carvalho Metanias Hallack. 2018. "Novas energias renováveis no Brasil: desafios e oportunidades." In *Desafios da Nação: Artigos de Apoio,* edited by João Alberto De Negri, Bruno César Araújo, and Ricardo Bacelette, 2:631–654. Brasília: Instituto de Pesquisa Econômica Aplicada.

Marchesoni-Acland, Franco, and Rodrigo Alonso-Suárez. 2020. "Intra-day solar irradiation forecast using RLS filters and satellite images." *Renewable Energy* 161:1140–1154.

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. 2023. "Machine learning advances for time series forecasting." *Journal of Economic Surveys* 37 (1): 76–111.

Mohamad Radzi, Putri Nor Liyana, Muhammad Naveed Akhter, Saad Mekhilef, and Noraisyah Mohamed Shah. 2023. "Review on the application of photovoltaic forecasting using machine learning for very short-to long-term forecasting." *Sustainability* 15 (4): 2942.

Molnar, Christoph. 2020. *Interpretable Machine Learning.* Lulu.com.

Newbery, David, Michael G. Pollitt, Robert A. Ritz, and Wadim Strielkowski. 2018. "Market design for a high-renewables European electricity system." *Renewable and Sustainable Energy Reviews* 91:695–707.

O'Shaughnessy, Eric, Jesse Cruce, and Kaifeng Xu. 2021. "Rethinking solar PV contracts in a world of increasing curtailment risk." *Energy Economics* 98:105264.

Olczak, Paweł, Paweł Jaśko, Dominik Kryzia, Dorota Matuszewska, Mariusz I. Fyk, and Artur Dyczko. 2021. "Analyses of duck curve phenomena potential in Polish PV prosumer households' installations." *Energy Reports* 7:4609–4622.

Operador Nacional do Sistema Elétrico. 2025a. *Critérios para Gestão de Excedentes Energéticos.* Technical report NT-ONS DOP 0022/2025. Rio de Janeiro: Operador Nacional do Sistema Elétrico.

———. 2025b. *Diagnóstico e Perspectiva da Evolução dos Cortes de Geração no Brasil.* Technical report RT DGL-ONS 0189/2025. Rio de Janeiro: Operador Nacional do Sistema Elétrico.

Paim, Maria-Antonia, Arthur R. Dalmarco, Chih-Hai Yang, Pablo Salas, Sigrid Lindner, Jean-Francois Mercure, José Baltazar Salgueirinho Osório de Andrade Guerra, Cristiane Derani, Tatiana Bruce da Silva, and Jorge E. Viñuales. 2019. "Evaluating regulatory strategies for mitigating hydrological risk in Brazil through diversification of its electricity mix." *Energy Policy* 128:393–401.

Parmezan, Antonio R. S., Vinicius M. A. Souza, and Gustavo E. A. P. A. Batista. 2019. "Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model." *Information Sciences* 484:302–337.

Perumean-Chaney, Suzanne E., Charity Morgan, David McDowall, and Inmaculada Aban. 2013. "Zero-inflated and Overdispersed: What's One to Do?" *Journal of Statistical Computation and Simulation* 83 (9): 1671–1683.

Petropoulos, Fotios, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K. Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J. Bessa, Jakub Bijak, John E. Boylan, et al. 2022. "Forecasting: Theory and Practice." *International Journal of Forecasting* 38 (3): 705–871.

Plagwitz, Lucas, Alexander Brenner, Michael Fujarski, and Julian Varghese. 2022. "Supporting AI-Explainability by Analyzing Feature Subsets in a Machine Learning Model." In *MIE,* 109–113.

Psarros, G. N., P. A. Dratsas, P. P. Chinaris, and S. A. Papathanassiou. 2025. "Assessing the implications of RES technology mix on curtailments, storage requirements and system economics." *Applied Energy* 381:125159.

Salgueiro, Guilherme Perobelli, Fernando Salgueiro Perobelli, Edson Paulo Domingues, and Victor Eduardo de Melo Valério. 2026. "Spatial and socioeconomic factors of distributed photovoltaic generation in Minas Gerais." *Energy Economics* 154:109121.

Silva, Walquiria N., I. B. Renolphi, Giovani G. T. Vieira, Luís F. N. Lourenço, Maurício B. C. Salles, and R. M. Monaro. 2025. "Wind and solar curtailment in Brazil: Quantifying grid constraints in the energy transition." In *2025 16th IEEE International Conference on Industry Applications,* 1–8.

Silva, Walquiria N., Giovani G. T. T. Vieira, Erik Eduardo Rego, Lucas F. C. Simone, Luís F. N. Lourenço, and Maurício B. C. Salles. 2026. "Curtailment pressures and grid flexibility in Brazil: Lessons from the fast-rising distributed generation." *Utilities Policy* 99:102148.

Sparks, Adam H. 2018. "nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R." *Journal of Open Source Software* 3 (30): 1035. https://doi.org/10.21105/joss.01035. https://doi.org/10.21105/joss.01035.

Stringer, Naomi, Navid Haghdadi, Anna Bruce, and Iain MacGill. 2021. "Fair consumer outcomes in the balance: Data driven analysis of distributed PV curtailment." *Renewable Energy* 173:972–986.

Succetti, Federico, Antonello Rosato, Rodolfo Araneo, and Massimo Panella. 2020. "Deep neural networks for multivariate prediction of photovoltaic power time series." *IEEE Access* 8:211490–211505.

Syntetos, Aris A., and John E. Boylan. 2005. "The Accuracy of Intermittent Demand Estimates." *International Journal of Forecasting* 21 (2): 303–314.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–288.

Titz, Maximilian, Sebastian Pütz, and Dirk Witthaut. 2024. “Identifying drivers and mitigators for congestion and redispatch in the German electric power system with explainable AI.” *Applied Energy* 356:122351.

Vieira, Giovani T. T., Walquiria N. Silva, Maurício B. C. Salles, and Luís F. N. Lourenço. 2025. “Assessment of constrained-off in renewable energy integration in Brazil.” In *2025 International Conference on Clean Electrical Power,* 121–127.

Xiao, Zenan, Xiaoqiao Huang, Jun Liu, Chengli Li, and Yonghang Tai. 2023. “A novel method based on time series ensemble model for hourly photovoltaic power prediction.” *Energy* 276:127542.

Yasuda, Yoh, Lori Bird, E. Maria Carlini, Poul Børre Eriksen, Ana Estanqueiro, Damian Flynn, Daniel Fraile, et al. 2022. “C-E (curtailment–energy share) map: An objective and quantitative measure to evaluate wind and solar curtailment.” *Renewable and Sustainable Energy Reviews* 160:112212.