

Article

A Photovoltaic Soiling Assessment Method Based on Irradiance Temporal Features and Power Spatial Consistency

Xiaoshi Xu ¹, Jifeng Song ^{2,*} , Hongyang Quan ¹, Shuai Zhang ¹, Jinguo He ¹ and Lianglin Zou ¹ ¹ School of New Energy, North China Electric Power University, Beijing 102206, China² Institute of Energy Power Innovation, North China Electric Power University, Beijing 102206, China

* Correspondence: songjifeng@ncepu.edu.cn

Abstract

Operational data from photovoltaic (PV) plants are widely used for soiling analysis, but their reliability is often reduced by weather disturbances, irradiance fluctuations, equipment heterogeneity, and measurement noise. This study proposes a baseline-referenced four-step filtering framework (S1–S5) to extract more physically consistent samples from inverter-level operational data. The framework combines irradiance-condition screening, irradiance and power temporal-trend constraints, and power–irradiance coupling consistency, and evaluates the retained samples using metrics of overall dispersion, intraday smoothness, interday continuity, spatial consistency, and sample retention. Based on 5 min data from the Lvhua PV plant in Shanghai, 280 parameter combinations were scanned under unified R^2 and coupling-correlation (CC) thresholds. The results show that the first two filtering steps provide the dominant reduction in overall variability, improving the dispersion metric by approximately 97% and 89%, respectively, while the later steps further improve device-response and coupling consistency. The recommended configuration, W7_R75 with CC = 0.95, retains 8.57% of valid samples and achieves a balanced improvement in data quality and sample support. Stricter CC thresholds (0.97–0.98) provide additional smoothness and spatial-consistency gains but reduce retention to about 5–7%. The proposed framework provides a reproducible data-screening basis for PV soiling assessment and operational data-quality control.

Keywords: photovoltaic soiling; operational data screening; irradiance temporal features; power spatial consistency



Academic Editors: José Vicente Muñoz Díez, Luís Miguel Nieto Nieto, Angel G. Gonzalez-Rodriguez and Slawomir Gulkowski

Received: 14 May 2026

Revised: 11 June 2026

Accepted: 29 June 2026

Published: 30 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In recent years, the global energy system has been accelerating its transition toward low-carbon and renewable energy [1–4]. Renewable energy technologies such as solar, wind, hydropower, and biomass are playing an increasingly important role in replacing conventional fossil fuels. Among them, photovoltaic (PV) power generation has become one of the fastest-growing renewable energy technologies because of its wide resource availability, simple system structure, and short construction cycle [5–7]. By the end of 2025, the global cumulative installed PV capacity had reached approximately 2800 GW and was still growing rapidly, with new installations in 2026 projected to reach 500–667 GW. As installed capacity continues to expand, the focus of the industry is gradually shifting from deployment scale to operational efficiency and operation-and-maintenance (O&M) cost control. Consequently, data quality and O&M strategies during plant operation have become increasingly important [8–10].

During PV plant operation, dust accumulation on module surfaces is widely recognized as one of the major factors affecting system performance. Large ground-mounted PV plants are often built in regions with favorable irradiance conditions, which are frequently associated with arid climates and dusty environments, making soiling on module surfaces more evident [3,11–13]. Dust deposited on the module surface reduces effective light transmittance [14,15], thereby lowering power output and causing generation losses; under some environmental conditions, the daily energy loss can even reach a relatively high level [16]. Therefore, key O&M decisions in PV plants, such as whether modules should be cleaned, when cleaning should be performed, and how cleaning frequency should be determined, essentially depend on reliable assessment of the soiling state of the modules.

A variety of methods have been developed for PV soiling assessment from different perspectives [17–20]. Early studies mainly focused on the overall impact of soiling on PV system performance [21–23]. Mani and Pillai systematically summarized the mechanisms by which dust deposition affects module transmittance and power output, and pointed out that soiling can cause substantial generation losses in arid and desert environments [12]. Subsequent review studies by Sarver et al. and Costa et al. further examined regional differences in soiling characteristics, deposition mechanisms, and cleaning strategies, showing that soiling has become an important factor affecting the long-term operational performance of PV plants [13]. In terms of specific assessment methods, one group of studies relies on direct measurements or comparative experiments, such as reference modules, transmittance measurements, or power comparisons before and after cleaning. Gostein et al. investigated soiling losses in large PV plants through field measurement systems and proposed a long-term monitoring approach based on reference modules [24]. Boyle et al. analyzed the effect of natural soiling on module performance through field experiments. These methods have clear physical meaning, but they usually require additional sensors or dedicated experimental conditions, which makes large-scale deployment costly [25].

With the widespread use of PV monitoring systems, another line of research has turned to soiling assessment based on plant operational data. Deceglie et al. proposed a soiling extraction method based on the relationship between power and irradiance, and used long-term operational data to analyze soiling evolution [26]. Micheli et al. further modeled PV soiling and analyzed its spatial distribution using time-series analysis and statistical approaches [11]. Compared with direct measurement methods, these data-driven approaches are more practical for engineering applications, because they enable plant-level soiling assessment without requiring additional hardware installation [27].

However, under real plant operating conditions, soiling assessment based on operational data still faces substantial challenges. The power output of a PV system is influenced not only by soiling, but also by cloud shading, irradiance fluctuations, sensor errors, and equipment operating conditions, all of which can disrupt the originally stable relationship between power and irradiance [28–31]. From a physical perspective, under ideal clear-sky conditions, solar irradiance is mainly governed by the solar elevation angle and usually follows a smooth, approximately unimodal intraday pattern [32]. Under such conditions, the coupling between PV power output and irradiance remains relatively stable, allowing the slow performance degradation caused by soiling to be identified from operational data [33,34]. In contrast, under cloudy or rapidly changing weather conditions, irradiance may fluctuate sharply at short time scales and produce spatially inconsistent shading effects across the plant, leading to pronounced differences in the power responses of different sub-arrays or inverters [11]. Such high-frequency variations induced by weather disturbances often mask the slow performance changes caused by soiling [19], thereby reducing the reliability of soiling assessment. Therefore, when operational data are used for soiling

analysis, identifying and screening time periods with relatively stable irradiance conditions is a key step toward improving assessment accuracy.

Although a variety of soiling assessment methods have been reported, clear limitations remain in the treatment of real operational data. On the one hand, most existing studies have focused on estimating or predicting soiling losses, whereas the preprocessing of operational data itself has received relatively limited systematic attention [35–40]. In many cases, data screening relies only on simple thresholds or single-step rules, which are often insufficient to handle complex weather disturbances and changing operating conditions [31]. On the other hand, existing methods generally lack a systematic mechanism to balance data-quality improvement against sample retention, and the evaluation often relies on a single metric, which provides an incomplete picture of screening performance. Overly strict screening can reduce data noise, but may also substantially shrink the usable sample set; in contrast, loose screening retains more data, but may introduce more abnormal samples and thus reduce the stability of the assessment [41]. In addition, for inverter-level operational data, the responses of different inverters may differ markedly because of local weather disturbances or variations in equipment condition, which further increases the complexity of data processing [19]. Therefore, it remains an open problem to develop a data-screening procedure that jointly considers physical consistency, data quality, and sample support, and to systematically evaluate how different parameter settings affect the screening results [42].

Against this background, this study takes the data processing procedure itself as the main research focus and proposes a four-step filtering method referenced to the raw-data baseline (S1). In addition, the effects of different filtering parameters on data quality and sample support are systematically investigated. The main contributions of this study are as follows:

- (1) A four-step data filtering framework referenced to the raw-data baseline (S1) is proposed to progressively remove unstable samples in raw operational data that are affected by weather disturbances and abnormal responses;
- (2) A joint evaluation system is established with overall dispersion, intraday smoothness, interday continuity, spatial consistency, and sample support as its core components, so as to quantify the effects of different filtering steps and parameter settings on data quality from multiple perspectives;
- (3) Through a systematic scan of window size, unified R^2 threshold, and coupling-correlation (CC) threshold, the trade-off between data-quality improvement and sample retention under different filtering intensities is analyzed, providing a reproducible basis for parameter selection;
- (4) Based on inverter-level operational data from a real photovoltaic plant, stage-wise experimental validation and 3D mechanism visualization are provided to reveal the role of the proposed method in improving data availability, reliability, and physical consistency.

It should be noted that clear-sky detection and irradiance-based data-quality-control methods are highly useful for identifying periods with stable irradiance conditions, but they do not necessarily guarantee that the PV power response during those periods is also physically consistent. In inverter-level operational data, even under apparently clear or smooth irradiance conditions, local shading, inverter anomalies, communication errors, or power-irradiance mismatch may still remain. Therefore, the key distinction of the proposed framework is that it does not stop at identifying stable irradiance periods; it further evaluates the temporal consistency of inverter power and the local coupling between power and irradiance. This extension from irradiance-condition screening to

power-response consistency screening is essential for obtaining reliable SLR samples from real plant operational data.

The remainder of this paper is organized as follows. Section 2 presents the methodology, including the study site, data and preprocessing, the definition of soiling loss ratio (SLR), the four-step filtering procedure referenced to the raw-data baseline (S1), the evaluation metrics, and the parameter-scanning strategy. Section 3 presents the parameter-scanning results and experimental analysis, and discusses the roles of different filtering stages as well as the benefit–cost trade-off of Step 5. Section 4 concludes the paper.

2. Methodology

2.1. Data Source and Preprocessing

This study is based on inverter-level operational data from the Lvhua PV plant in Lvhua Town, Chongming District, Shanghai (31.62° N, 121.40° E; altitude ~4 m). The data have a temporal resolution of 5 min, which allows short-term variations during plant operation to be captured. The study period spans from 6 February 2025 to 3 December 2025, covering approximately 10 months of operational data. The site belongs to a humid subtropical monsoon climate zone, characterized by hot, humid summers and mild winters, with frequent cloud cover that poses typical challenges for data-driven soiling assessment. The raw data mainly include plane-of-array irradiance *G*, module back-surface temperature *T*, inverter output power *P*, and the rated capacity *C* of each inverter. The definitions, physical meanings, and data sources of these variables are summarized in Table 1.

Table 1. The main input variables used in this study and their physical meanings.

Variable	Symbol	Unit	Resolution	Source
Plane-of-array irradiance	<i>G</i>	W/m ²	5 min	On-site irradiance monitoring
Module back-surface temperature	<i>T</i>	°C	5 min	On-site temperature monitoring
Inverter output power	<i>P</i>	kW	5 min	SCADA/inverter monitoring
Rated inverter capacity	<i>C</i>	kW	Static parameter	Equipment inventory

The studied plant consists of 11 PV arrays and 176 inverters, which makes it suitable for characterizing both the spatial differences among devices under real operating conditions and the temporal variability of plant operation. The use of inverter-level data allows the responses of different devices to be compared under the same meteorological conditions, and further enables analysis of how the proposed filtering method improves data quality from a spatial perspective.

To ensure comparability in the subsequent filtering results, the raw operational data were preprocessed in a unified manner. The preprocessing mainly involved three simple steps. First, the time indices of all data streams were standardized, and only daytime records (06:00–18:00) were retained to exclude the influence of nighttime non-generation data. Second, the power, module temperature, and irradiance data were aligned to a common timestamp so that all variables at a given time corresponded to the same observation conditions. Finally, the aligned data were merged into a unified master table and cached for repeated use in the subsequent filtering experiments under different parameter combinations, thereby improving computational efficiency. These preprocessing steps provide a consistent input for the subsequent baseline-referenced filtering analysis.

2.2. SLR Definition

To evaluate how data filtering affects the stability of soiling estimation, SLR is used as the core evaluation metric in this study. It is defined as follows [19]:

$$SLR = 1 - \frac{P_{i,actual}}{P_{i,theory}} \quad (1)$$

In the above equation, $P_{i,actual}$ is the measured power of the soiled PV group at each time, and $P_{i,theory}$ is the corresponding theoretical power, which is calculated as follows:

$$P_{i,theory} = C \times \frac{G}{1000} \times \alpha_T \times \eta_s \quad (2)$$

In the above equation, C is the maximum rated power of the string, G is the irradiance of the plane of the array, α_T is the temperature correction factor, and η_s is the integrated system efficiency coefficient. The temperature correction factor α_T is calculated as follows:

$$\alpha_T = 1 + \gamma(T_{cell} - 25) \quad (3)$$

γ is the power temperature coefficient. For the PV modules used in the studied plant, γ is set to $-0.300\%/^{\circ}\text{C}$. T_{cell} is the PV module temperature.

To avoid relying on a fixed empirical value, η_s is determined in this study using a clean-day fitting approach. Specifically, clear-sky data following continuous rainfall events are selected for calibration. Intense or continuous rainfall events can usually remove a substantial portion of the soiling accumulated on the module surface, allowing the modules to approach a clean state. Under such condition, η_s is calibrated by comparing the measured power with the theoretical power. The corresponding equation is given as follows:

$$\eta_s = \frac{P_{actual}}{C \times \frac{G}{1000} \times \alpha_T} \quad (4)$$

The calculated plant-level η_s is 0.86. It should be noted that this value represents an average efficiency factor at the plant level and cannot fully capture the actual efficiency differences among all modules or inverters within the plant. However, since all SLR calculations in this study use the same η_s , and the analysis mainly focuses on relative variations across time and across strings, the systematic bias introduced by η_s is largely canceled out in ratio- and difference-based calculations. Its influence on the data-stability metrics considered in this study is therefore limited.

2.3. Baseline-Referenced Four-Step Filtering Procedure

To improve the reliability of operational data for soiling analysis, this study proposes a multi-stage filtering method referenced to the raw-data baseline, in which raw operational data are progressively screened in a stage-wise manner. The method constrains the data from multiple perspectives, including irradiance conditions, temporal trends, and inter-device consistency, so as to gradually remove abnormal samples caused by weather disturbances, equipment anomalies, or measurement errors. As shown in Figure 1, the overall procedure can be divided into five stages, S1–S5, where S1 denotes the raw-data baseline and S2–S5 correspond to the four actual filtering steps.

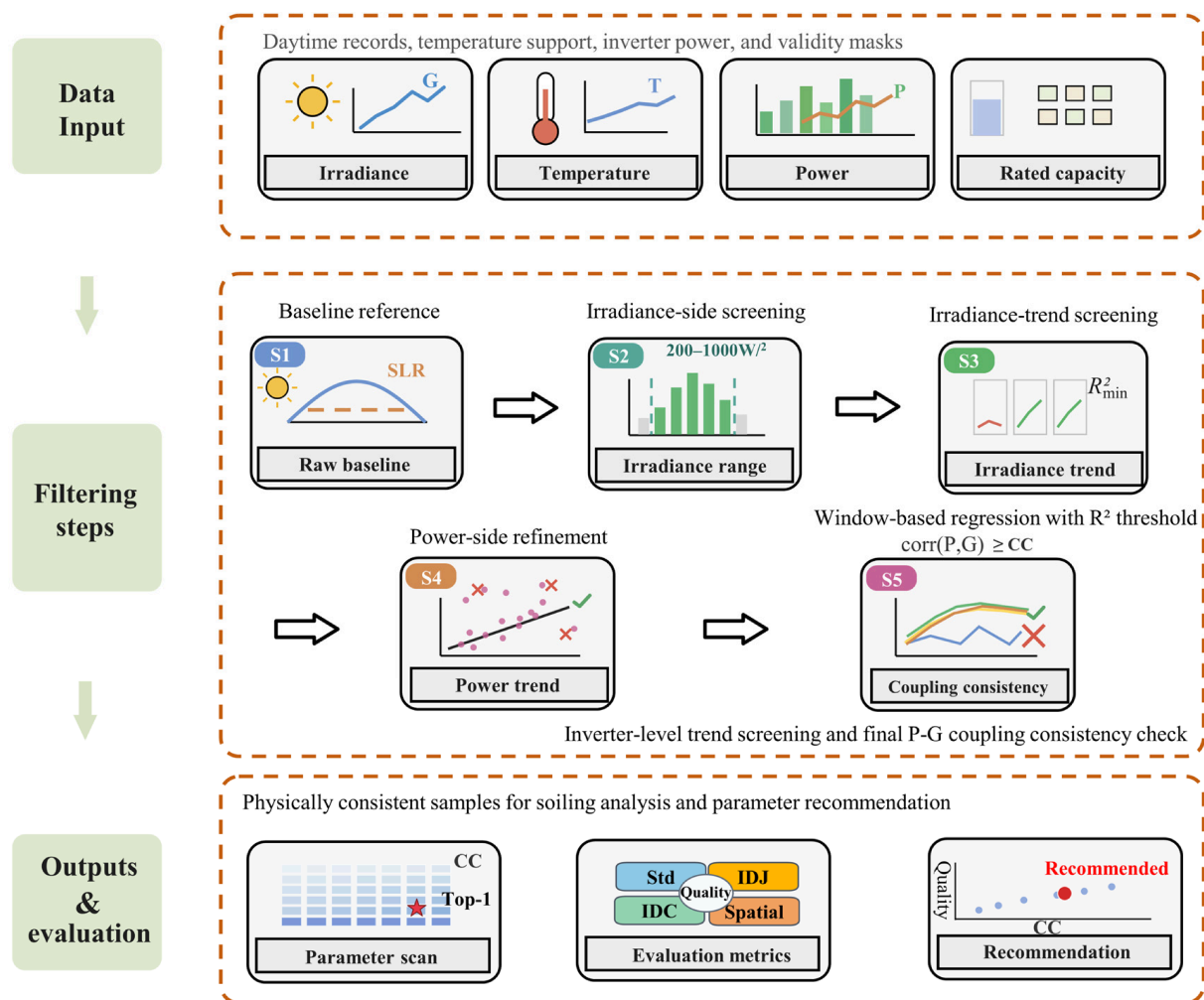


Figure 1. Baseline-referenced four-step filtering framework.

S1 Raw baseline

The SLR values are first calculated from the preprocessed data. This stage is used to establish the raw-data baseline and serves as the reference for calculating sample retention in the subsequent filtering steps.

S2 Irradiance-range filtering

Under low-irradiance conditions, module power output is more susceptible to incidence angle effects and spectral variation [43], and the relationship between power and irradiance may deviate substantially from its normal behavior. Extremely high irradiance values, on the other hand, may contain short-term extreme irradiance events such as cloud enhancement [44], which can introduce additional disturbances into the subsequent SLR calculation. Therefore, only irradiance data within the range of 200–1000 W/m² are retained in this study.

S3 Irradiance-trend regression filtering

Under clear-sky conditions, irradiance typically follows a unimodal pattern at the daily scale and exhibits a clear monotonic trend within short time windows [31,32]. Therefore, the trend characteristics of the irradiance time series can be used to identify stable weather conditions. Specifically, for each natural day, a sliding-window linear regression is applied

to the irradiance sequence. With a window length of w , the regression model is defined as follows:

$$G_k = a_G x_k + b_G \quad (5)$$

where a_G is the irradiance variation slope, b_G is the regression intercept, and x_k is the sample index within the window.

The goodness of fit is evaluated by the coefficient of determination R^2 :

$$R_G^2 = 1 - \frac{\sum (G_k - \overline{G_k})^2}{\sum (G_k - \overline{G_k})^2} \quad (6)$$

In addition, a directional constraint on the diurnal variation is imposed:

$$\text{sign}(a_G) = \begin{cases} +1, & h < 12 \\ -1, & h \geq 12 \end{cases} \quad (7)$$

where h denotes the hour value of the center time within the window. For the flat-top irradiance region around solar noon, windows with small slopes are directly accepted to avoid over-filtering stable samples.

S4 Power-trend filtering

Even after irradiance-trend screening, abnormal power responses may still remain for many inverters, for example due to local equipment faults, local shading, or communication disturbances. Therefore, the same sliding-window trend analysis used in Step 3 is further applied to the power sequence of each inverter. The screening criteria are the same as those in Step 3, namely the goodness-of-fit threshold and the directional constraint on diurnal variation. Unlike Step 3, however, this step is performed separately for each inverter. When an inverter does not satisfy the criteria at a given time, only the corresponding power point is marked as missing, while the entire timestamp is retained.

S5 Irradiance–power coupling filtering

After irradiance-trend screening and power-trend screening, it is still necessary to ensure that the power response remains sufficiently coupled with irradiance variation at the local scale. Therefore, an additional irradiance–power coupling-consistency constraint is introduced. This step is mainly used to remove device-level data with abnormal power responses.

Within each sliding window, the correlation coefficient between power and irradiance is calculated as follows:

$$\rho_i = \text{corr}(P_i, G_i) \quad (8)$$

The following condition must also be satisfied:

$$\rho_i \geq \rho_{\min} \quad (9)$$

where $\rho_{\min}$ is the threshold for irradiance–power correlation.

2.4. Evaluation Metrics

To evaluate the effect of data filtering on SLR stability and sample support, five metrics were used in this study: Std [26] (the standard deviation of all valid SLR samples), Intra-Day_Jump (IDJ, an intraday local fluctuation metric), Interday Continuity (IDC, an interday continuity metric), Spatial_Std [45] (a cross-inverter dispersion metric), and Retention_Rate

(RR, the sample retention rate relative to the S1 baseline). The definitions of these metrics are given below.

2.4.1. Std: Statistical Stability Metric

Std is used to quantify the overall dispersion of all valid SLR samples after filtering. A lower Std indicates that the retained SLR samples are more concentrated and less affected by short-term disturbances. It is calculated as:

$$Std = \sqrt{\frac{1}{M-1} \sum_{(i,t) \in S} (SLR_{i,t} - \overline{SLR})^2} \quad (10)$$

And:

$$\overline{SLR} = \frac{1}{M} \sum_{(i,t) \in S} SLR_{i,t} \quad (11)$$

where S is the set of all valid samples, and M is the total number of samples.

2.4.2. IntraDay_Jump (IDJ)

To characterize the short-term fluctuation of the filtered SLR at the intraday scale, IDJ is defined based on the system-average SLR series. Here, the system-average SLR refers to the time series obtained by averaging the SLR values of all valid inverters at each timestamp. This metric is used to quantify the local jump amplitude between adjacent valid time points, and a smaller value indicates a smoother series. It is defined as follows:

$$IDJ = \frac{1}{D} \sum_{d=1}^D \text{median} \left(\frac{\bar{S}_{d,k} - \bar{S}_{d,k-1}}{\Delta t_{d,k}} \right) \quad (12)$$

where $\bar{S}_{d,k}$ denotes the system-average SLR at the k -th valid time point on day d , $\Delta t_{d,k}$ denotes the time interval between two adjacent valid samples, and D is the number of valid days. To reduce the influence of local outliers, the median of the absolute first-order difference per unit time is first calculated for each day, and then a simple average is taken over all valid days to obtain the overall IDJ.

2.4.3. Interday Continuity (IDC): Temporal Consistency Metric

IDC is used to quantify the continuity of daily mean SLR values between adjacent valid days. Since soiling usually evolves gradually in the absence of rainfall or cleaning events, a lower IDC indicates better interday consistency:

$$TD = \frac{1}{D-1} \sum_{d=2}^D |SLR_d - SLR_{d-1}| \quad (13)$$

where SLR_d is the daily mean soiling loss ratio on day d , and D is the number of valid days.

2.4.4. Spatial_Std: Spatial-Consistency Metric

Spatial_Std is used to measure the dispersion of SLR estimates among different inverters across the plant at the same timestamp, and reflects consistency in the spatial dimension. Under ideal conditions, the arrays corresponding to different inverters within the same plant are subject to similar meteorological conditions, and their soiling variation trends are therefore expected to be highly synchronized. Accordingly, excessively large differences in SLR among inverters at the same timestamp usually indicate the presence of factors such as local cloud shading, local equipment anomalies, communication problems, or data-matching errors.

In this study, the standard deviation of SLR across all inverters is calculated at each timestamp, and the resulting spatial standard deviations are then averaged over all valid timestamps to obtain the spatial-consistency metric. Its definition is given as follows:

$$S_Std = \frac{1}{T} \sum_{t=1}^T \sqrt{\frac{1}{M} \sum_{i=1}^M (SLR_{i,t} - \overline{SLR}_t)^2} \quad (14)$$

where T is the number of valid timestamps, and M denotes the number of valid inverters at timestamp i .

2.4.5. Retention_Rate (RR): Sample Retention Metric

When evaluating a data-screening method, it is not sufficient to consider only the improvement in data quality; the amount of valid samples retained after filtering must also be taken into account. If the screening criteria are overly strict, data fluctuations and abnormal samples may be significantly reduced, but the number of usable samples may also become too small, thereby weakening the statistical reliability of subsequent soiling analysis. Therefore, a sample retention metric is further introduced in this study to measure the effect of different filtering parameter combinations on sample quantity.

RR is the percentage of valid samples retained under a given parameter combination relative to the number of baseline samples in S1. Its definition is given as follows:

$$RR = \frac{M}{M_{raw}} \times 100\% \quad (15)$$

where M is the number of valid samples retained under the current filtering condition, and M_{raw} is the number of valid SLR samples in S1.

2.4.6. Composite Score (CS): Composite Scoring Metric

To avoid judging parameter performance based on a single metric, a composite evaluation metric, CS, is constructed in this study. Unlike a simple formulation based on “sample RR/penalty term,” the proposed method first treats sample support as a threshold constraint. When RR is lower than 3%, the corresponding parameter combination is considered to lack sufficient data support, and no CS is calculated. When RR is equal to or greater than 3%, a sample-support factor is introduced as follows:

$$Support = \min\left(\frac{RR}{10\%}, 1\right) \quad (16)$$

This definition means that once RR reaches 10%, the benefit of additional sample support becomes saturated and no longer increases linearly with sample size, thereby preventing an excessive number of samples from dominating the final score.

On this basis, the quality metrics are further processed using robust normalization and converted into dimensionless utility values. For any lower-is-better quality metric, the utility function is defined as follows:

$$u(x) = u_{min} + (1 - u_{min}) \cdot clip\left(\frac{q_{0.9} - x}{q_{0.9} - q_{0.1}}, 0, 1\right) \quad (17)$$

where x denotes the value of the metric under a given parameter combination; $q_{0.1}$ and $q_{0.9}$ denote the 10th and 90th percentiles of that metric among all parameter combinations, respectively; $clip(\cdot)$ denotes a clipping function that constrains the result to the interval $[0, 1]$; and $u_{min} = 0.05$ is the minimum utility value, introduced to avoid the direct collapse of the CS to zero when one metric performs extremely poorly.

The final CS is defined as follows:

$$CS = 100 \cdot Support^{0.15} \cdot u(Std)^{0.25} \cdot u(IDC)^{0.25} \cdot u(IDJ)^{0.25} \cdot u(Spatial_Std)^{0.25} \quad (18)$$

A higher CS value indicates that the corresponding parameter combination achieves a better overall balance between data quality and sample support.

The weighting scheme and the 3% RR threshold were selected to reflect the methodological objective of improving data reliability while maintaining minimum sample support. RR was assigned a lower weight than each individual quality metric to prevent sample quantity from dominating the CS; otherwise, parameter combinations with high retention but limited quality improvement could be favored. The four quality metrics, Std, IDJ, IDC, and Spatial_Std, were assigned equal weights because they represent complementary dimensions of data quality, including statistical stability, intraday smoothness, interday continuity, and cross-inverter spatial consistency.

2.5. Parameter Scanning Strategy

To systematically analyze the effects of filtering parameters on data quality and sample retention, a combined scanning strategy based on a unified R^2 threshold and a CC threshold is adopted in this study. The trend window size is set as follows, corresponding to time scales of approximately 25–120 min. The irradiance fitting threshold is set as follows:

$$W \in \{5, 7, 9, 11, 13, 15, 20, 24\}$$

The irradiance fitting threshold is set as follows:

$$R^2_{min} \in \{0.75, 0.80, 0.85, 0.90, 0.95\}$$

Accordingly, a total of $8 \times 5 = 40$ trend-parameter combinations are obtained. On this basis, the CC threshold is further scanned as follows:

$$\rho_{min} \in \{0.80, 0.85, 0.90, 0.95, 0.97, 0.98, 0.99\}$$

In total, 280 parameter settings are obtained. The evaluation metrics are calculated for each setting, based on which the trade-off between data-quality improvement and sample retention under different filtering intensities is analyzed.

3. Results and Discussion

3.1. Variation in Optimal Parameters Under Different Coupling Thresholds

To evaluate the effect of parameter selection under the unified-threshold framework on the filtering results, a systematic scan was conducted over 280 parameter combinations formed by eight window sizes, five trend thresholds, and seven coupling thresholds. Under each fixed CC value, all parameter combinations were ranked using the CS, and the highest-scoring combination was taken as the optimal parameter configuration for that CC condition. For convenience, different parameter combinations are denoted by symbols; for example, W7_R90 represents a window size of 7 and a trend threshold of 0.90.

As shown in Figure 2 and Table 2, the optimal parameter combination exhibits a clear stage-wise variation with CC. When $CC = 0.80$ – 0.90 , the optimal combination remains W15_R90. Under this relatively loose coupling constraint, the final RR is about 6.25%, with $Std \approx 0.167$, $IDJ \approx 0.0058$, $Spatial_Std \approx 0.040$, and $IDC \approx 0.036$. This indicates that when the final coupling constraint is weak, the framework tends to rely on a longer trend window and a stricter unified R^2 threshold to ensure sample quality.

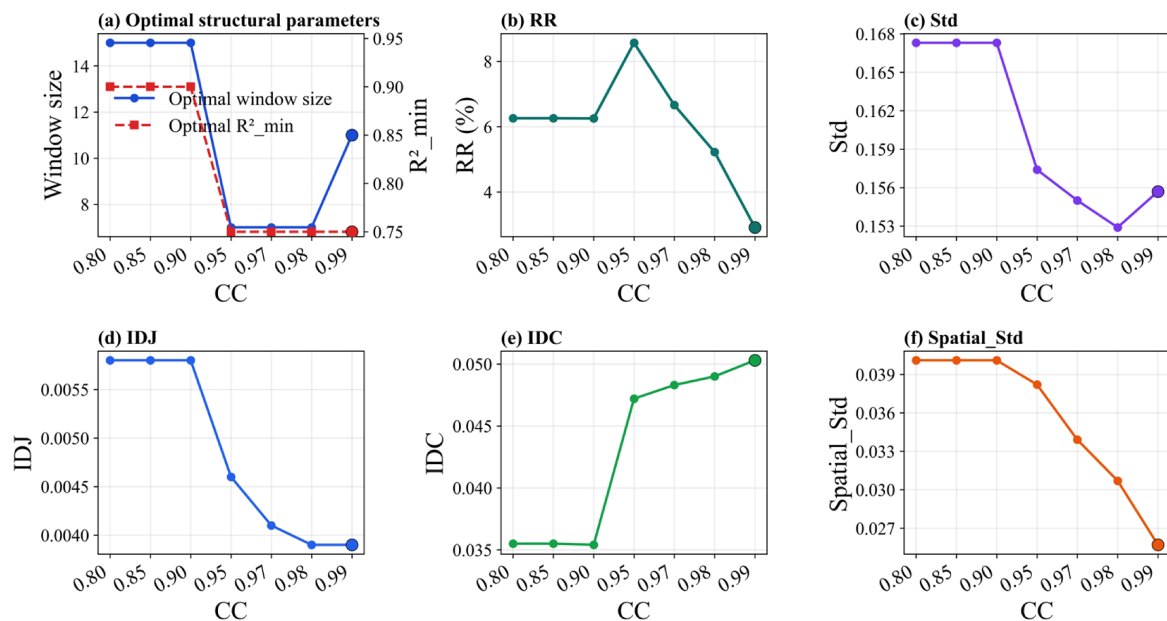


Figure 2. Optimal parameter combinations and corresponding key metrics under different coupling thresholds. (a) Optimal structural parameters under different CC conditions; (b) final RR; (c) Std; (d) IDJ; (e) IDC; and (f) Spatial_Std. Each point corresponds to the parameter combination with the highest CS under the current CC condition.

Table 2. Top-1 parameter combinations and corresponding final metrics under different CC values.

CC	Top-1 Parameter	RR (%)	Std	IDJ	IDC	Spatial_Std	CS
0.80	W15_R90	6.26	0.167	0.0058	0.036	0.040	61.35
0.85	W15_R90	6.26	0.167	0.0058	0.036	0.040	60.54
0.90	W15_R90	6.25	0.167	0.0058	0.035	0.040	58.75
0.95	W7_R75	8.57	0.157	0.0046	0.047	0.038	58.42
0.97	W7_R75	6.66	0.155	0.0041	0.048	0.033	56.60
0.98	W7_R75	5.22	0.153	0.0039	0.049	0.030	50.34
0.99	W11_R75	2.92	0.156	0.0039	0.050	0.026	/

When CC increases from 0.95 to 0.98, the selected optimal setting shifts to W7_R75, indicating that a shorter window with a looser trend threshold becomes preferable under stricter coupling constraints. Because the comparison is based on the optimal parameter combination under each CC condition, RR does not necessarily decrease monotonically with CC; the relatively higher RR at CC = 0.95 mainly reflects the shift from W15_R90 to W7_R75 rather than a relaxation of the coupling constraint.

Within this range, increasing CC mainly improves local smoothness and spatial consistency while reducing sample support. At CC = 0.95, the framework retains 8.57% of valid samples and achieves Std = 0.157, IDJ = 0.0046, and Spatial_Std = 0.038. Raising CC to 0.98 further improves IDJ and Spatial_Std to about 0.0039 and 0.030, but RR decreases to 5.22% and IDC slightly increases. Therefore, CC = 0.95 provides a more balanced operating point, whereas CC = 0.97–0.98 may be selected when higher sample purity is prioritized.

When CC = 0.99, no parameter combination satisfies the minimum sample-support threshold required by the current scoring criterion. Under this threshold, the best raw case retains only about 2.92% of the samples, indicating that the filtering becomes overly strict and is unfavorable for practical engineering applications.

Overall, the range of 0.80–0.90 does not make full use of the final coupling-consistency constraint, whereas 0.99 leads to excessive sample loss. By comparison, 0.95–0.98 better reflects an effective trade-off between data-quality improvement and sample support.

3.2. Stage-Wise Filtering Effects Referenced to the Raw Baseline

Figure 3 shows that $S1 \rightarrow S2$ and $S2 \rightarrow S3$ provide the dominant improvements in the early filtering process. $S1 \rightarrow S2$ sharply reduces the overall dispersion, with Std decreasing by about 97%, indicating that irradiance-range filtering effectively removes low-irradiance edge samples that are prone to unstable SLR estimates. $S2 \rightarrow S3$ further reduces Std by about 89% and improves interday and spatial consistency, showing that irradiance-trend filtering mainly suppresses weather-induced instability at the daily and plant-wide scales. In contrast, its direct effect on short-term local jumps is relatively limited, as reflected by the smaller improvement in IDJ.

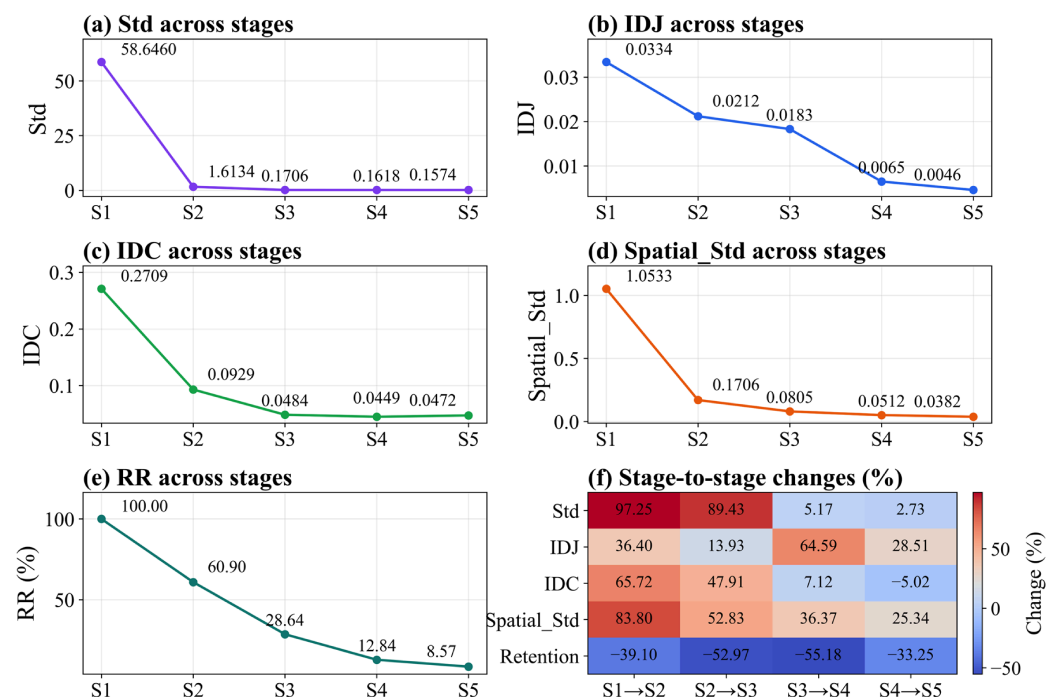


Figure 3. Metric variations in the optimal parameter combinations across stages S1–S5.

Compared with the first two steps, $S3 \rightarrow S4$ plays a greater role in local refinement. Although the further improvement in Std at this stage is limited, only about 3.96–5.17%, the improvements in IDJ and Spatial_Std are much more pronounced. In particular, IDJ improves by about 51.71–72.65% during $S3 \rightarrow S4$, while Spatial_Std improves by about 31.19–51.02%. This indicates that power-trend filtering does not continue to substantially compress the overall distribution width; rather, it mainly removes inverter-level abnormal responses that still exhibit local power drops or trend inconsistency even under relatively stable irradiance conditions.

Figure 4 provides a visual interpretation of how different disturbances are removed during the filtering process. In the representative case on 3 August 2025, S2 mainly removes low-irradiance edge periods, while S3 further excludes irradiance-unstable segments associated with cloudy disturbances. The enlarged views in Figure 4f,g show the additional role of the later stages: S4 removes inverter-level local power-drop anomalies, whereas S5 further eliminates residual power–irradiance mismatch segments. These visual patterns support the functional distinction between the irradiance-stability filters in S2–S3 and the local power-response refinement in S4–S5. To verify that this behavior is not limited to this

single-day example, additional S1–S5 visualizations for representative days from different seasons and for a continuous one-week period are provided in the Supplementary Material, where all cases were processed using the same filtering parameters.

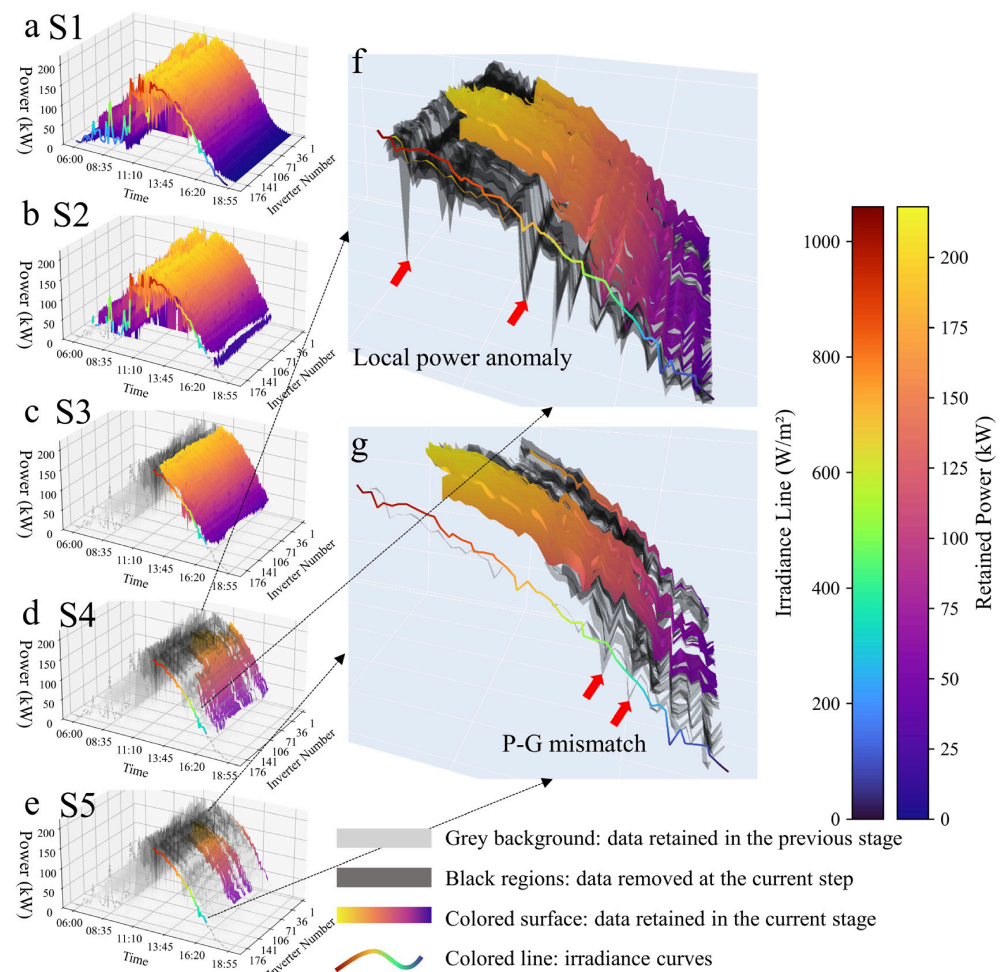


Figure 4. Stage-wise 3D visualization of S1–S5 and local mechanism illustration of S4/S5. (a) S1 raw-data baseline before filtering; (b) S2 retained data after irradiance-range filtering; (c) S3 retained data after irradiance-trend filtering; (d) S4 retained data after power-trend filtering; (e) S5 retained data after power-irradiance coupling filtering; (f) enlarged local view of inverter-level local power-drop anomalies removed during S4; (g) enlarged local view of residual power-irradiance mismatch segments removed during S5. The grey background indicates data retained in the previous stage, dark regions indicate data removed in the current stage, the colored surface indicates data retained in the current stage, and the colored line represents the irradiance curve.

In summary, the stage-wise results indicate that the proposed data-screening framework has a relatively clear internal functional division: S2 and S3 are responsible for the main overall noise reduction and the removal of weather-unstable samples, whereas S4 and S5 further improve local temporal consistency and spatial consistency after the data have entered a relatively stable irradiance regime.

3.3. Quality Gain and Sample Cost of S5

After irradiance-trend screening and power-trend screening, S5 further imposes an irradiance–power coupling-consistency constraint. Figure 5 selects IDJ and Spatial_Std as representative indicators of the gain brought by S5, reflecting improvements in local temporal smoothness and spatial consistency, respectively.

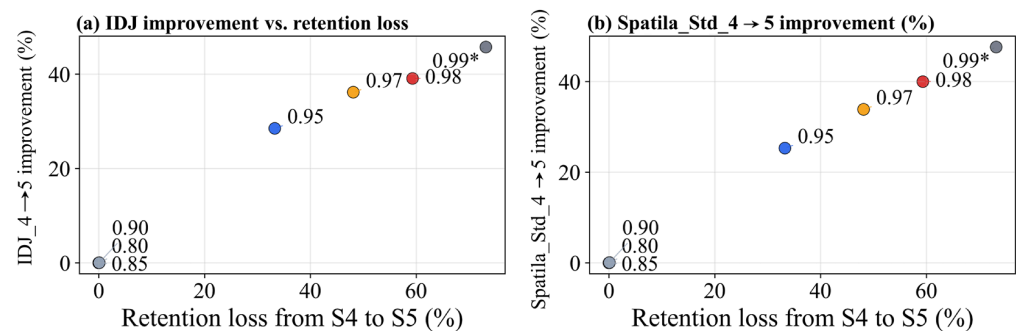


Figure 5. Benefit–cost trade-off of S4→S5. Each colored dot represents the best parameter combination under a fixed CC, and the number next to each dot indicates the corresponding CC value. Grey dots denote relatively loose or invalid reference cases, while blue, orange, and red dots highlight CC = 0.95, 0.97, and 0.98, respectively. The asterisk (*) indicates that the corresponding case does not satisfy the minimum sample-support criterion ($RR < 3\%$) and is shown only as a reference rather than a valid recommended solution.

When $CC = 0.80$ – 0.90 , S5 contributes almost no additional improvement to the final results. Taking the optimal combination W15_R90 as an example, the improvements in IDJ and Spatial_Std from S4 to S5 are both close to zero, and the change in RR is less than 0.1%. This indicates that under weak coupling constraints, the outputs of S4 and S5 are nearly identical, and the final coupling-consistency check is essentially ineffective within this threshold range.

When CC increases to 0.95 and above, the role of S5 becomes much more evident. Taking the optimal combination W7_R75 at $CC = 0.95$ as an example, S4→S5 improves Std by about 2.73%, IDJ by about 28.51%, and Spatial_Std by about 25.34%, while reducing RR by about 33.25% relative to S4. When CC increases to 0.97, the corresponding improvements increase to about 4.20%, 36.15%, and 33.86%, respectively, while RR decreases by about 48.10%. At 0.98, the improvements in Std, IDJ, and Spatial_Std from S4 to S5 reach about 5.47%, 39.10%, and 40.00%, respectively, whereas RR decreases by about 59.31%.

Therefore, Step 5 is better understood as a final sample-purification step. Its main role is to remove residual power–irradiance mismatch samples that remain after Step 4, but this gain is accompanied by a clear cost in sample support.

3.4. Recommended Parameter Range

Considering the balance between data quality and sample support, a CC of about 0.95 can be regarded as a preferable engineering compromise under the current framework. Under this threshold, the optimal combination W7_R75 still retains a final RR of about 8.57%, while already achieving clear improvements in key metrics such as Std, IDJ, and Spatial_Std. Compared with lower thresholds, this setting enables the final coupling-consistency screening to play a more effective role; compared with stricter thresholds, it retains a higher level of valid sample support.

If the application places greater emphasis on sample purity and can accept a lower RR, $CC = 0.97$ or 0.98 may also be considered. Under these two settings, the retained samples show further improvement in local smoothness and spatial consistency, but RR decreases to about 6.66% and 5.22%, respectively. By contrast, $CC = 0.99$ is not suitable as a recommended threshold for routine engineering use because the sample loss is too large.

Based on the above results, W7_R75 with $CC = 0.95$ is recommended as the primary parameter setting for subsequent soiling analysis and routine engineering applications. The detailed parameter recommendation is given in Table 3.

Table 3. A comparison of the recommended and alternative parameter combinations.

Scheme Type	CC	Parameter Combination	RR (%)	Std	IDC	IDJ	Spatial_Std
Recommended	Loose	W15_R90	6.26	0.167	0.036	0.0058	0.040
	0.95	W7_R75	8.57	0.157	0.047	0.0046	0.038
	Strict	W7_R75	5.22	0.153	0.049	0.0039	0.030
	Overly strict	W11_R75	2.92	0.156	0.050	0.0039	0.025

3.5. Comparison with Existing Methods

To further evaluate the proposed framework, we compared the recommended configuration (W7_R75, CC = 0.95) with several existing clear-sky and irradiance-based data-screening methods, including the CSI threshold method, Reno and Hansen (2016) [32], Jordan and Hansen (2023) [31], and Bright et al. [42]. All methods were applied to the same dataset and evaluated using the same metrics, including Std, IDJ, IDC, Spatial_Std, RR, and CS.

As shown in Figure 6, the proposed S5 framework achieves the best overall performance. Compared with the existing benchmark methods, S5 provides lower Std, IDJ, IDC, and Spatial_Std while maintaining a higher retention ratio. This indicates that the proposed framework not only removes unstable irradiance conditions, but also further excludes local power anomalies and power-irradiance mismatch samples through the power-trend and coupling-consistency constraints. Therefore, the proposed method provides a more balanced trade-off between data-quality improvement and sample retention than conventional clear-sky or irradiance-only screening methods.

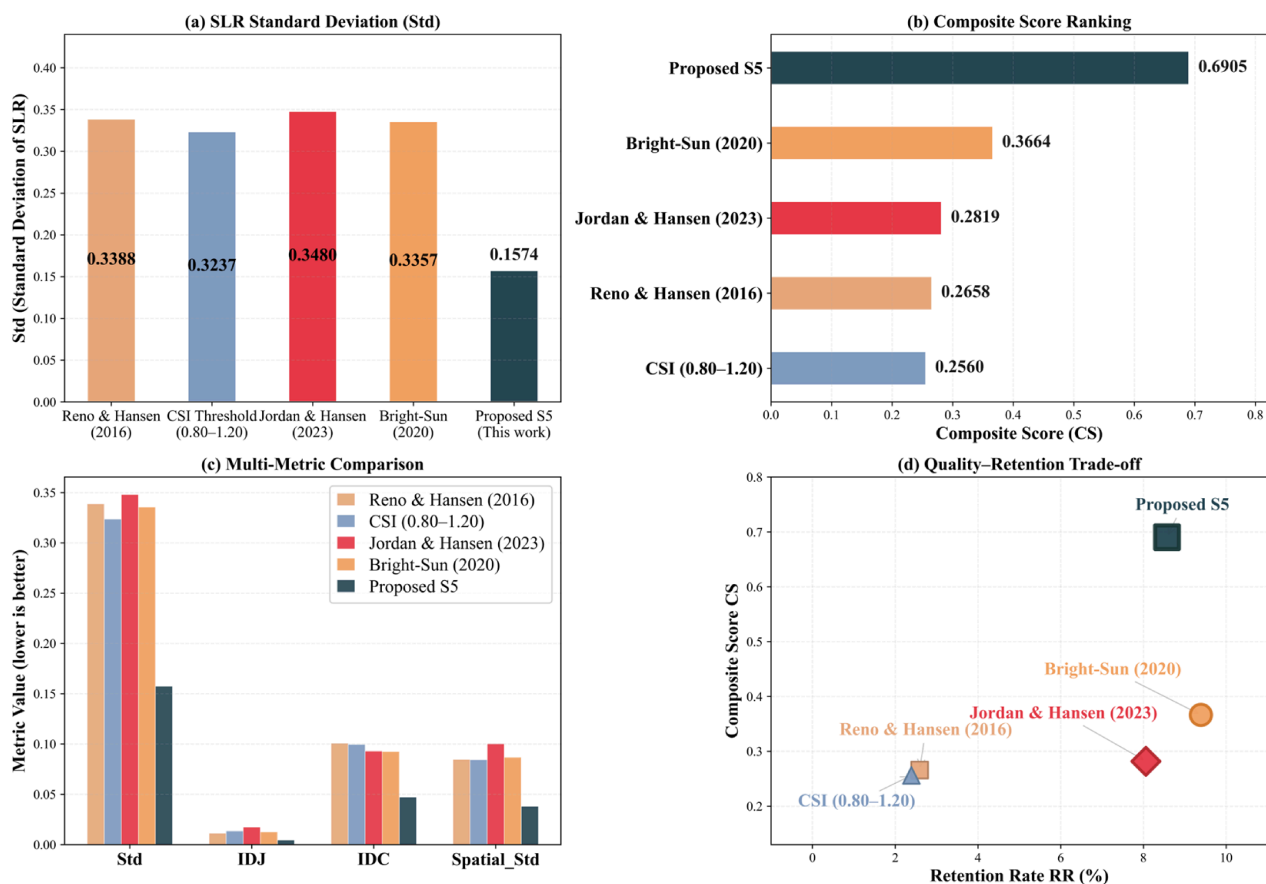


Figure 6. Comparison of data-screening methods [31,32,42]: (a) SLR standard deviation; (b) Composite Score ranking; (c) multi-metric grouped comparison; (d) Quality–Retention trade-off.

4. Conclusions

This study proposes a five-stage filtering framework (S1–S5) referenced to the raw-data baseline (S1). The method aims, through simple and efficient data screening, to reduce as much as possible the influence of non-soiling factors on photovoltaic soiling assessment, rather than directly providing an absolute ground truth of soiling severity. Its goal is to improve the data quality and physical reliability of the samples used for soiling analysis, thereby providing a more dependable basis for subsequent soiling trend estimation and cleaning-strategy optimization. The framework sequentially applies four physical constraints—irradiance range, irradiance temporal trend, power temporal trend, and irradiance–power coupling consistency—to progressively screen operational data from multiple perspectives. Meanwhile, a multi-metric evaluation system encompassing overall Std, IDJ, IDC, Spatial_Std, and RR was established to systematically quantify the contribution of each filtering stage.

Systematic analysis based on inverter-level operational data from a real PV plant shows that the baseline-referenced four-step filtering framework can significantly improve data stability. Specifically, irradiance-range filtering and irradiance-trend filtering account for the main overall noise suppression, whereas power-trend filtering and coupling-consistency filtering mainly serve to further enhance sample consistency in both the temporal and spatial dimensions. The mechanism visualization results further indicate that Step 4 mainly removes local power-drop anomalies, while Step 5 mainly removes residual power–irradiance mismatch segments, reflecting a progressive increase in data reliability from coarse screening to fine refinement. Among the 280 parameter combinations, CC = 0.95 with window size W7 and R^2 threshold R75 achieves a favorable engineering trade-off between data-quality improvement and sample retention.

The current work still has some limitations: the data validation is confined to a single PV plant at Lvhu, Shanghai (belonging to a subtropical monsoon climate zone). Future work should include cross-site verification under different climatic conditions (e.g., Mediterranean, hot desert, tropical monsoon, temperate oceanic climates) and different system configurations to examine the generalizability of the proposed method. In addition, subsequent research should incorporate independent direct soiling measurements (e.g., reference module pairs, transmissometer readings) or documented cleaning events with pre- and post-cleaning or rainfall comparisons, so as to directly quantify the improvement in soiling estimation accuracy brought by the proposed filtering framework.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/en19133102/s1>, Figure S1: Multi-season S1–S5 3D visualizations of the proposed filtering framework for representative dates in different seasons and irradiance conditions, including winter (19 February 2025), spring (29 May 2025), summer (17 June 2025), autumn (18 September 2025), and late autumn/early winter (13 November 2025), using the same filtering parameters; Figure S2: Continuous one-week S1–S5 3D visualizations from 19 August 2025 to 25 August 2025 using the same filtering parameters.

Author Contributions: Conceptualization, X.X., J.S., H.Q., S.Z., J.H. and L.Z.; methodology, X.X. and J.S.; validation, H.Q.; formal analysis, X.X. and J.H.; investigation, X.X., J.S., H.Q., S.Z. and L.Z.; data curation, S.Z., J.H. and L.Z.; writing—original draft, X.X. and H.Q.; writing—review and editing, X.X., J.S., H.Q., S.Z., J.H. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Project “Vice President of Science and Technology” of Changping District, Beijing (202502007023) and partly supported by the Fundamental Research Funds for the Central Universities (2024JC007).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to restrictions related to research confidentiality and data management requirements.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

PV	Photovoltaic
SLR	Soiling loss ratio
P_{actual}	Measured inverter output power
P_{theory}	Theoretical clean-condition power
G	Irradiance
C	Rated inverter/string capacity
α_T	Temperature correction factor
γ	Power temperature coefficient
η_s	Integrated system efficiency coefficient
MPPT	Maximum Power Point Tracking
S1	Raw-data baseline
S2	Irradiance-range filtering
S3	Irradiance-trend filtering
S4	Power-trend filtering
S5	Irradiance–power coupling filtering
W	Sliding-window length
R ²	Coefficient of determination
CC	Coupling correlation
Std	Standard deviation of valid SLR samples
IDJ	IntraDay_Jump
IDC	Interday continuity
Spatial_Std	Spatial standard deviation
RR	Retention rate
CS	Composite score

References

1. Zaihidee, F.M.; Mekhilef, S.; Seyedmahmoudian, M.; Horan, B. Dust as an unalterable deteriorative factor affecting PV panel's efficiency: Why and how. *Renew. Sustain. Energy Rev.* **2016**, *65*, 1267–1278. [\[CrossRef\]](#)
2. He, B.; Lu, H.; Zheng, C.; Wang, Y. Characteristics and cleaning methods of dust deposition on solar photovoltaic modules—A review. *Energy* **2023**, *263*, 126083. [\[CrossRef\]](#)
3. Ilse, K.; Micheli, L.; Figgis, B.W.; Lange, K.; Daßler, D.; Hanifi, H.; Wolfertstetter, F.; Naumann, V.; Hagendorf, C.; Gottschalg, R.; et al. Techno-Economic Assessment of Soiling Losses and Mitigation Strategies for Solar Power Generation. *Joule* **2019**, *3*, 2303–2321. [\[CrossRef\]](#)
4. Rusănescu, C.O.; Rusănescu, M.; Istrate, I.A.; Constantin, G.A.; Begea, M. The Effect of Dust Deposition on the Performance of Photovoltaic Panels. *Energies* **2023**, *16*, 6794. [\[CrossRef\]](#)
5. Choi, J.; Son, S.-W.; Lee, S.; Park, S. Advancing global solar photovoltaic power forecasting with sub-seasonal climate outlooks. *Renew. Energy* **2024**, *237*, 121803. [\[CrossRef\]](#)
6. Li, M.; Geng, Y.; Zhou, S.; Sarkis, J. Clean energy transitions and health. *Heliyon* **2023**, *9*, e21250. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Xiao, H.; Song, F.; Zheng, X.; Chen, J. Community-based energy revolution: An evaluation of China's photovoltaic poverty alleviation Program's economic and social benefits. *Energy Policy* **2023**, *177*, 113555. [\[CrossRef\]](#)
8. Oviedo Hernandez, G.; Godinho Ariolli, D.M.; Enriquez Paez, P.S.; Chiantore, P.V. Trends and innovations in photovoltaic operations and maintenance. *Prog. Energy* **2022**, *4*, 042002. [\[CrossRef\]](#)
9. Hernández-Callejo, L.; Gallardo-Saavedra, S.; Alonso-Gómez, V. A review of photovoltaic systems: Design, operation and maintenance. *Sol. Energy* **2019**, *188*, 426–440. [\[CrossRef\]](#)
10. Cipriani, G.; D'Amico, A.; Guarino, S.; Manno, D.; Traverso, M.; Di Dio, V. Convolutional Neural Network for Dust and Hotspot Classification in PV Modules. *Energies* **2020**, *13*, 6357. [\[CrossRef\]](#)

11. Micheli, L.; Fernández, E.F.; Aguilera, J.T.; Almonacid, F. Economics of seasonal photovoltaic soiling and cleaning optimization scenarios. *Energy* **2021**, *215*, 119018. [[CrossRef](#)]
12. Mani, M.; Pillai, R. Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations. *Renew. Sustain. Energy Rev.* **2010**, *14*, 3124–3131. [[CrossRef](#)]
13. Sarver, T.; Al-Qaraghuli, A.; Kazmerski, L.L. A comprehensive review of the impact of dust on the use of solar energy: History, investigations, results, literature, and mitigation approaches. *Renew. Sustain. Energy Rev.* **2013**, *22*, 698–733. [[CrossRef](#)]
14. Costa, S.C.S.; Diniz, A.S.A.C.; Kazmerski, L.L. Dust and soiling issues and impacts relating to solar energy systems: Literature review update for 2012–2015. *Renew. Sustain. Energy Rev.* **2016**, *63*, 33–61. [[CrossRef](#)]
15. Sayyah, A.; Horenstein, M.N.; Mazumder, M.K. Energy yield loss caused by dust deposition on photovoltaic panels. *Sol. Energy* **2014**, *107*, 576–604. [[CrossRef](#)]
16. Dida, M.; Boughali, S.; Bechki, D.; Bouguettaia, H. Output power loss of crystalline silicon photovoltaic modules due to dust accumulation in Saharan environment. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109787. [[CrossRef](#)]
17. Jiao, X.; Li, X.; Yang, Y.; Xiao, W. Novel and comprehensive approach for power loss estimation of soiled photovoltaic modules. *Sol. Energy* **2024**, *268*, 112283. [[CrossRef](#)]
18. Chen, L.; Fan, S.; Sun, S.; Cao, S.; Sun, T.; Liu, P.; Gao, H.; Zhang, Y.; Ding, W. A detection model for dust deposition on photovoltaic (PV) panels based on light transmittance estimation. *Energy* **2025**, *322*, 135284. [[CrossRef](#)]
19. Cao, X.; Tang, Z.; Xu, X.; Kuang, B.; Zou, W.; Shangguan, X.; Zhu, H.; Song, J. The Impact of Cloudy Weather on the Calculation Accuracy of the Soiling Loss Ratio in Photovoltaic Systems. *Energies* **2026**, *19*, 471. [[CrossRef](#)]
20. Huang, M.; Yang, R.; Tang, G.H.; Pu, J.H.; Sun, Q.; Du, M. Quantifying the effects of dust characteristics on the performance of radiative cooling PV systems. *Appl. Energy* **2025**, *377*, 124672. [[CrossRef](#)]
21. Gholami, A.; Khazaei, I.; Eslami, S.; Zandi, M.; Akrami, E. Experimental investigation of dust deposition effects on photo-voltaic output performance. *Sol. Energy* **2018**, *159*, 346–352. [[CrossRef](#)]
22. Nath, A.; Pal, A.; Ilango, G.S. Impact of inverter DC to AC ratio on soiling losses and cleaning intervals in large solar PV plants. *Sol. Energy* **2025**, *300*, 113825. [[CrossRef](#)]
23. Younis, A.; Cotfas, P.A.; Cotfas, D.T. Systematic indoor experimental practices for simulating and investigating dust deposition effects on photovoltaic surfaces: A review. *Energy Strategy Rev.* **2024**, *51*, 101310. [[CrossRef](#)]
24. Gostein, M.; Caron, J.R.; Littmann, B. IEEE-Measuring Soiling Losses at Utility-scale PV Power Plants. In *Proceedings of the IEEE Photovoltaic Specialists Conference*; IEEE: New York, NY, USA, 2014.
25. Boyle, L.; Flinchbaugh, H.; Hannigan, M.P. Natural soiling of photovoltaic cover plates and the impact on transmission. *Renew. Energy* **2015**, *77*, 166–173. [[CrossRef](#)]
26. Deceglie, M.G.; Micheli, L.; Muller, M. Quantifying Soiling Loss Directly From PV Yield. *IEEE J. Photovolt.* **2018**, *8*, 547–551. [[CrossRef](#)]
27. Agbogla, J.; Sekyere, C.K.K.; Forson, F.K.; Opoku, R.; Baah, B. Soiling estimation methods in solar photovoltaic systems: Review, challenges and future directions. *Results Eng.* **2025**, *26*, 104810. [[CrossRef](#)]
28. Jamil, I.; Lucheng, H.; Habib, S.; Aurangzeb, M.; Ali, A.; Ahmed, E.M. Performance Ratio Analysis Based on Energy Production for Large-Scale Solar Plant. *IEEE Access* **2022**, *10*, 5715–5735. [[CrossRef](#)]
29. Kazem, H.A.; Al-Waeli, A.H.A.; Chaichan, M.T.; Sopian, K.; Gholami, A.; Alnaser, W.E. Dust and cleaning impact on the performance of photovoltaic: An outdoor experimental study. *Energy Sources Part A Recovery Util. Environ. Eff.* **2023**, *45*, 3107–3124. [[CrossRef](#)]
30. Jiang, H.; Lu, L.; Sun, K. Experimental investigation of the impact of airborne dust deposition on the performance of solar photovoltaic (PV) modules. *Atmos. Environ.* **2011**, *45*, 4299–4304. [[CrossRef](#)]
31. Jordan, D.C.; Hansen, C. Clear-sky detection for PV degradation analysis using multiple regression. *Renew. Energy* **2023**, *209*, 393–400. [[CrossRef](#)]
32. Reno, M.J.; Hansen, C.W. Identification of periods of clear sky irradiance in time series of GHI measurements. *Renew. Energy* **2016**, *90*, 520–531. [[CrossRef](#)]
33. Jordan, D.C.; Kurtz, S.R.; VanSant, K.; Newmiller, J. Compendium of photovoltaic degradation rates. *Prog. Photovolt. Res. Appl.* **2016**, *24*, 978–989. [[CrossRef](#)]
34. Jordan, D.C.; Kurtz, S.R. Photovoltaic Degradation Rates—An Analytical Review. *Prog. Photovolt. Res. Appl.* **2011**, *21*, 12–29. [[CrossRef](#)]
35. Micheli, L.; Deceglie, M.G.; Muller, M. Mapping Photovoltaic Soiling Using Spatial Interpolation Techniques. *IEEE J. Photovolt.* **2019**, *9*, 272–277. [[CrossRef](#)]
36. You, S.; Lim, Y.J.; Dai, Y.; Wang, C.-H. On the temporal modelling of solar photovoltaic soiling: Energy and economic impacts in seven cities. *Appl. Energy* **2018**, *228*, 1136–1146. [[CrossRef](#)]
37. Coello, M.; Boyle, L. Simple Model for Predicting Time Series Soiling of Photovoltaic Panels. *IEEE J. Photovolt.* **2019**, *9*, 1382–1387. [[CrossRef](#)]

38. Javed, W.; Guo, B.; Figgis, B. Modeling of photovoltaic soiling loss as a function of environmental variables. *Sol. Energy* **2017**, *157*, 397–407. [[CrossRef](#)]
39. Lindig, S.; Louwen, A.; Moser, D.; Topic, M. Outdoor PV System Monitoring—Input Data Quality, Data Imputation and Filtering Approaches. *Energies* **2020**, *13*, 5099. [[CrossRef](#)]
40. Chermite, C.; Douiri, M.R. Harnessing hybrid intelligence: Four vector metaheuristic and differential evolution for optimized photovoltaic parameter extraction. *Comput. Electr. Eng.* **2025**, *123*, 110276. [[CrossRef](#)]
41. Hu, A.; Sun, Q.; Liu, H.; Zhou, N.; Tan, Z.a.; Zhu, H. A Novel Photovoltaic Array Outlier Cleaning Algorithm Based on Sliding Standard Deviation Mutation. *Energies* **2019**, *12*, 4316. [[CrossRef](#)]
42. Bright, J.M.; Sun, X.; Gueymard, C.A.; Acord, B.; Wang, P.; Engerer, N.A. Bright-Sun: A globally applicable 1-min irradiance clear-sky detection model. *Renew. Sustain. Energy Rev.* **2020**, *121*, 109706. [[CrossRef](#)]
43. Wolfertstetter, F.; Esquelli, A.; Wilbert, S.; Hanrieder, N.; Blum, N.; Korevaar, M.; Bergmans, T.; Zarzalejo, L.; Polo, J.; Alami-Merrouni, A.; et al. Incidence angle and diffuse radiation adaptation of soiling ratio measurements of indirect optical soiling sensors. *J. Renew. Sustain. Energy* **2021**, *13*, 033703. [[CrossRef](#)]
44. Tzoumanikas, P.; Nikitidou, E.; Bais, A.F.; Kazantzidis, A. The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system. *Renew. Energy* **2016**, *95*, 314–322. [[CrossRef](#)]
45. Leloux, J.; Narvarte, L.; Trebosc, D. Review of the performance of residential PV systems in France. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1369–1376. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.