Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260362
Published: June 9, 2026

Before the Drop: Transfer Entropy as an Early-Warning Signal for Photovoltaic System Disturbances

Md Tohidul Islam, Congyu Wu, Yuxin Wang and Yong Wang
School of Systems Science and Industrial Engineering
Binghamton University, State University of New York
Binghamton, NY 13902, USA
mislam33@binghamton.edu, congyu.wu@binghamton.edu, yuxinw@binghamton.edu, yongwang@binghamton.edu

Abstract

Photovoltaic (PV) monitoring is commonly performed through individual system indicators such as Performance Ratio (PR). Although useful, these indicators often identify faults only after measurable performance degradation has already occurred. Drone-based visual and thermal inspection can provide physical diagnosis, but it is periodic, not continuous. We propose an information-theoretic framework for early-warning monitoring of co-located PV systems using Transfer Entropy (TE). Instead of treating each PV system as an isolated unit, the proposed approach models a solar facility as a directed network of interacting time series. Three years of 5-minute resolution data from the Desert Knowledge Australia Solar Centre (DKASC) in Alice Springs were analyzed which covers 12 co-located PV systems and 43,044 clean daytime observations from 2019 to 2021. TE was compared with Pearson correlation and Mutual Information (MI) as symmetric baseline measures. Results show that all 132 directed system pairs exhibit significant TE, with values ranging from 0.029 to 0.139 bits and a 4.7× spread in information-flow strength. The TE network reveals identifiable source-sink behavior, with M1_A, M7_A, and M2_C acting as dominant sources and M4_A and M5_A acting as dominant sinks. Fault-adjacent periods show significantly higher TE than normal periods, with mean TE increasing from 0.0772 to 0.0945 bits (approximately 22%). Welch's t-test and the Kolmogorov–Smirnov (KS) test confirm statistical significance. Early-warning case studies further suggest that TE may rise before visible PR degradation, supporting its potential for risk-aware PV monitoring, predictive maintenance, and inspection planning.

Keywords
Photovoltaic System Monitoring, Transfer Entropy, Fault Detection, Predictive Maintenance, Information Theory

1. Introduction

Solar PV energy has become an important component of the global transition toward sustainable power generation. As PV installations scale in size and complexity, early fault detection is essential for maintaining energy yield, reducing downtime, and minimizing maintenance costs. PV systems are subject to a wide range of performance-degrading faults, including soiling, shading, module cracking, inverter failures, metering errors, and communication outages. Prior work has shown that degradation and fault behavior can directly affect long-term system reliability and energy output (Jordan and Kurtz 2013; Pillai and Rajasekar 2018; Triki-Lahiani et al. 2018).

Conventional PV monitoring relies on system-level performance indicators such as Performance Ratio (PR), which compares measured production with expected output under available irradiance. Although PR is widely used in monitoring standards and performance assessment (Khalid et al. 2016; Walker and Desai 2021), it is fundamentally reactive: it signals a problem only after measurable performance degradation has already occurred. Drone-based visual and thermal inspection can provide more direct physical diagnosis by identifying hotspots, cracks, soiling, and module-level defects. Recent studies show that UAV-based inspection and aerial infrared thermography are increasingly important for PV plant maintenance (de Oliveira et al. 2022; Michail et al. 2024; Mellit and Kalogirou

2025). However, drone inspection is periodic rather than continuous, requiring scheduling, equipment, and suitable operating conditions. Between inspection cycles, the most continuously available diagnostic source is the time-series data already collected by monitoring infrastructure.

In a co-located PV facility, multiple systems operate under nearly identical irradiance and weather conditions. Their power-output time series may therefore contain information not only about individual performance, but also about how systems interact during normal and abnormal operation. This creates an opportunity to study the solar facility as a network of interacting time series rather than a collection of isolated units. If the past behavior of one system helps predict the future behavior of another, there may be directional information flow between them. During faults, outages, or system stress, this information-flow structure may change before standard PR monitoring clearly deteriorates.

To quantify this behavior, we use TE, a model-free information-theoretic measure of directed temporal dependence introduced by Schreiber (2000). TE measures how much the past of one time series reduces uncertainty about the future of another, after accounting for the target series' own past. This distinguishes TE from Pearson correlation and Mutual Information (MI), both of which are symmetric and cannot identify information sources, sinks, or directionality in a co-located network. TE is interpreted here as evidence of directional information flow or predictive directional dependence, not as proof of physical causality. This is consistent with prior work connecting TE to predictive causality (Barnett et al. 2009).

This paper proposes and evaluates a TE-based monitoring framework using three years of real-world data from the DKASC in Alice Springs, Australia. The dataset includes documented operational events such as outages, metering failures, and remediation activities (Desert Knowledge Australia Solar Centre 2026), making it well-suited for testing whether information-theoretic signals can reveal directed system structure, change during fault-adjacent periods, and provide early-warning evidence before visible PR degradation appears. The contribution is an applied early-warning framework that can complement PR monitoring and drone-based inspection to support more proactive, risk-aware PV maintenance.

1.1 Objectives

The main objective of this study is to develop and evaluate an information-theoretic early-warning framework for fault-related behavior in co-located PV systems. The specific objectives are:

Objective 1. To model co-located PV systems as a directed information-flow network using TE and identify source and sink behavior among systems.
Objective 2. To compare TE with Pearson correlation and MI to evaluate whether TE provides directional information beyond symmetric dependence caused by shared irradiance and weather conditions.
Objective 3. To examine whether TE values differ significantly between normal operating periods and fault-adjacent periods associated with documented DKASC operational events.
Objective 4. To evaluate whether TE shows early-warning behavior before visible degradation appears in Performance Ratio during major documented disturbance events.

2. Literature Review

The research literature relevant to this study spans three primary areas: PV system monitoring and fault detection, information-theoretic methods for directed time-series analysis, and network-based approaches to monitoring complex co-located systems. This section reviews key contributions in each area and identifies the gap addressed by this paper.

2.1 PV Monitoring and Fault Detection

PV systems are subject to a broad range of degradation mechanisms and fault types that can reduce energy production and increase maintenance costs over time. Jordan and Kurtz (2013) conducted a comprehensive analytical review of PV degradation rates across technologies, climates, and monitoring data sources, establishing that the median annual degradation rate is approximately 0.5%, with higher rates observed under certain field conditions and technologies. This foundational study demonstrated that cumulative performance losses over system lifetimes are economically significant and necessitate effective monitoring strategies.

Pillai and Rajasekar (2018) provided a systematic review of protection challenges and fault diagnosis methods for PV systems, categorizing faults into electrical, environmental, and shading types and evaluating hardware-level and software-level detection approaches. Triki-Lahiani et al. (2018) surveyed fault detection and monitoring system architectures, covering model-based, signal-processing, and data-driven strategies, and noted that no single approach comprehensively handles all fault categories. A recurring limitation in both reviews is that existing methods predominantly analyze individual systems in isolation, without capturing interaction effects among co-located PV arrays that operate under shared environmental conditions.

2.2 Performance Ratio and Conventional Monitoring

Khalid et al. (2016) analyzed Performance Ratio across a broad portfolio of grid-connected PV installations worldwide, examining how environmental factors, technology type, system age, and operational practices influence PR variability, and providing empirical benchmarks for interpreting PR values in field conditions. Walker and Desai (2021) evaluated 75 federal PV installations in the United States and found that while PR-based monitoring is informative and standardized, it is fundamentally reactive: performance problems are flagged only after measurable production losses have already occurred. This reactive limitation creates a monitoring gap between the onset of abnormal system behavior and the point at which standard indicators clearly deteriorate which motivates the development of complementary early-warning signals.

2.3 UAV-Based Inspection and Thermal Imaging

Unmanned aerial vehicle-based visual and thermal inspection has become an important tool for PV plant diagnosis and predictive maintenance. de Oliveira et al. (2022) reviewed automatic PV plant inspection using aerial infrared thermography, showing that thermal imaging can detect hotspots, degraded bypass diodes, soiled cell strings, and module-level defects that remain invisible in PR trends and electrical monitoring data. Michail et al. (2024) provided an updated comprehensive review of UAV-based diagnostic approaches for photovoltaic plants, covering multispectral imaging, machine learning-assisted defect classification, and autonomous flight path planning. Mellit and Kalogirou (2025) reviewed the integration of embedded artificial intelligence with infrared thermographic imaging for predictive maintenance and identified the need for integrated monitoring systems that bridge continuous electrical monitoring and periodic physical inspection. A shared limitation across all these approaches is their periodic rather than continuously operating nature: between inspection cycles, early deterioration of system-level behavior remains undetected.

2.4 Information Theory and Transfer Entropy

Shannon (1948) established the mathematical foundations of information theory by introducing entropy as a quantitative measure of uncertainty and information in probabilistic systems. Cover and Thomas (2006) provide a comprehensive treatment of entropy, conditional entropy, mutual information, and related quantities that underpin the measures used in this study. These tools are model-free and can capture nonlinear dependencies, making them well-suited for analyzing complex real-world time series.

Schreiber (2000) introduced Transfer Entropy as a model-free, asymmetric measure of directed information transfer between dynamical systems. TE measures how much the past of a source system reduces uncertainty about the future state of a target system, conditioned on the target's own past. This directional property distinguishes TE from Pearson correlation and mutual information, which are symmetric and cannot identify which system drives and which responds. Barnett et al. (2009) showed that for Gaussian processes, TE is mathematically equivalent to Granger causality, providing a formal theoretical basis for interpreting TE as evidence of directional temporal dependence. Vicente et al. (2011) analyzed TE as a practical measure of effective connectivity in complex dynamical systems and demonstrated its ability to detect directed interactions in settings where causal structure is not known in advance. Lizier (2014) developed an open-source information-theoretic toolkit for complex systems research that has become a standard resource for practical TE computation, confirming the tractability of TE-based analysis for applied monitoring problems.

2.5 Network-Based Analysis of Complex Systems

Boccaletti et al. (2006) provided a comprehensive review of complex network theory, including centrality metrics, community structure, and dynamic processes on networks. Network-based methods have been applied across domains including infrastructure resilience, power system monitoring, and environmental sensor networks to reveal structural properties and vulnerability patterns that are invisible when individual components are analyzed in isolation. Treating a co-located PV facility as a directed information-flow network, where each system is a node and each directed edge represents TE-quantified information flow, represents a useful and relatively underexplored application of these network concepts to solar energy monitoring.

2.6 Research Gap and Contribution

The reviewed literature establishes that conventional PR monitoring is reactive, drone-based inspection is periodic, and existing data-driven monitoring methods largely treat PV systems as independent units. TE provides a rigorous, model-free framework for quantifying directional temporal dependence, but its application to co-located PV system monitoring using continuously available metering data has not been systematically explored. This paper addresses this gap by developing and validating a TE-based directed network monitoring framework using three years of real-world DKASC data, with the goal of supporting proactive, risk-aware PV inspection and maintenance planning.

3. Methods

This study develops an information-theoretic monitoring framework for analyzing directional relationships among co-located PV systems. The method is designed to determine whether time-series relationships among PV systems can reveal abnormal behavior before conventional performance indicators clearly degrade.

Figure 1 summarizes the overall workflow. Raw PV time-series data and operational event records are first converted into clean PR signals. The cleaned PR series are then discretized into symbolic states and analyzed through two parallel tracks. The first track uses TE and Markov entropy rate to quantify directed information flow and temporal complexity. The second track uses Pearson correlation and MI as symmetric baseline measures.

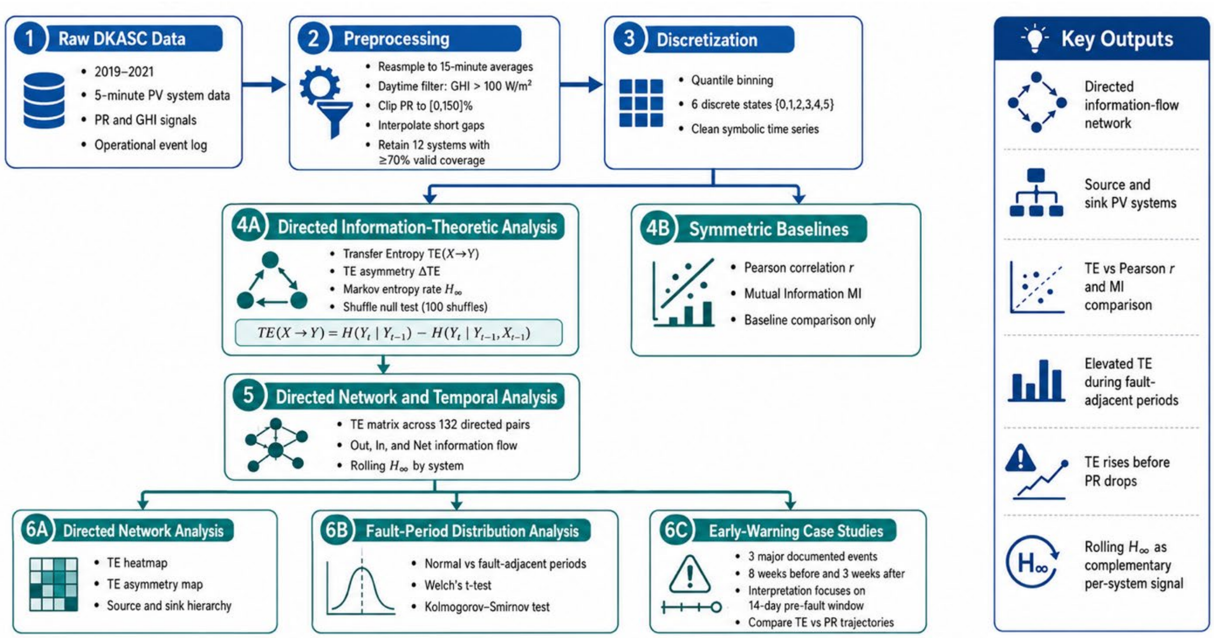

Figure 1. Methodology framework for information-theoretic monitoring of co-located PV systems. Raw DKASC time-series data are preprocessed into clean PR signals, discretized into symbolic states, analyzed using directed TE and symmetric baselines, and evaluated through network analysis, fault-adjacent distribution analysis, and early-warning case studies.

3.1 Data Preprocessing and Performance Ratio Signals

The raw DKASC time-series data were resampled from 5-minute to 15-minute averages to reduce high-frequency noise while preserving system-level temporal behavior. Only daytime observations were retained using a GHI threshold of GHI > 100 W/m². PR values were clipped to [0, 150]% to reduce extreme sensor artifacts. Short missing gaps of up to four consecutive 15-minute steps were linearly interpolated, and rows with remaining missing values were removed. Only PV systems with at least 70% valid daytime coverage over the full three-year period were retained. The final dataset contained 12 PV systems and 43,044 clean daytime observations arranged as a multivariate time-series matrix.

3.2 Discretization of PV System States

Information-theoretic measures require probability distributions over discrete states. Each system's continuous PR time series was discretized into six equal-frequency (quantile-based) bins, producing an integer-valued symbolic sequence $X_t \in \{0, 1, 2, 3, 4, 5\}$. Quantile binning was selected because PV distributions may be skewed during faults or sensor artifacts; equal-frequency bins improve the stability of empirical probability estimates.

3.3 Information-Theoretic Measures

This study uses Shannon entropy, conditional entropy, Mutual Information, Transfer Entropy, TE asymmetry, and Markov entropy rate. Shannon (1948) and Cover and Thomas (2006) provide the mathematical foundations. Table 1 summarizes the measures used.

Table 1. Information-theoretic measures used in this study.

Measure	Formula	Role in This Study
Shannon Entropy	$H(X) = -\Sigma_x\, p(x) \log_2 p(x)$	Measures uncertainty of a single PV system state
Conditional Entropy	$H(Y\|X) = H(X,Y) - H(X)$	Building block for TE computation
Mutual Information	$MI(X;Y) = H(Y) - H(Y\|X)$	Symmetric baseline for shared dependence
Transfer Entropy	$TE(X\rightarrow Y) = H(Y_t\|Y_{t-1}) - H(Y_t\|Y_{t-1}, X_{t-1})$	Main measure of directed temporal information flow
TE Asymmetry	$A_{TE}(X,Y) = TE(X\rightarrow Y) - TE(Y\rightarrow X)$	Measures net directionality between two systems
Markov Entropy Rate	$H\infty = -\Sigma_i\, \pi_i\, \Sigma_j\, T_{ij} \log_2(T_{ij})$	Per-system temporal complexity measure

For two PV systems X and Y, Transfer Entropy at lag $\tau = 1$ step (15 minutes) was computed as:

$$TE(X\rightarrow Y) = H(Y_t \mid Y_{t-1}) - H(Y_t \mid Y_{t-1}, X_{t-1}) \tag{1}$$

Equation (1) asks whether the recent past of system X reduces uncertainty about the next state of system Y, after accounting for Y's own recent past. The directional asymmetry between two systems was computed as:

$$A_{TE}(X, Y) = TE(X\rightarrow Y) - TE(Y\rightarrow X) \tag{2}$$

Equation (2) compares the two directions of information flow for the same system pair. A positive $A_{TE}(X, Y)$ indicates stronger information flow from X to Y; a negative value indicates stronger flow from Y to X. TE is interpreted as directional information flow or predictive directional dependence, not as proof of physical causality (Barnett et al. 2009).

3.4 Directed Network Construction

After computing TE for every ordered pair, the facility was represented as a directed weighted network where each PV system is a node and each directed edge X→Y is weighted by the corresponding TE value. For each system i, outgoing flow, incoming flow, and net information flow were computed as:

$$Out_i = \Sigma_{j\neq i}\, TE(i\rightarrow j), \quad In_i = \Sigma_{j\neq i}\, TE(j\rightarrow i), \quad Net_i = Out_i - In_i \tag{3}$$

Equation (3) summarizes whether system i mainly sends or receives predictive information. A positive Net_i indicates information-source behavior; a negative Net_i indicates information-sink behavior.

3.5 Statistical Significance Testing

TE significance was assessed using a shuffle-based null test. For each directed pair X→Y, the source series X was randomly shuffled 100 times and TE was recomputed to generate a null distribution. Shuffling destroys temporal ordering while preserving the marginal distribution. Observed TE was considered significant at $\alpha = 0.05$ if it exceeded the 95th percentile of the null distribution. This test was applied to all 132 directed pairs.

3.6 Fault-Adjacent Distribution Analysis

To test whether TE changes during documented disturbances, each timestamp was labeled fault-adjacent (within ±7 days of a documented DKASC event), normal (at least 14 days from any event), or buffer (excluded). Pairwise TE was computed separately for normal and fault-adjacent periods. Welch's t-test evaluated mean differences; the Kolmogorov–Smirnov test evaluated distributional differences. Using both tests provides stronger validation: the t-test assesses mean shifts while the KS test assesses full distributional changes.

3.7 Early-Warning Case Study Analysis

To evaluate early-warning potential, three major documented DKASC events were selected: the January 2021 major data outage, the November 2019 system-wide outage, and the August 2021 network and UPS outage. For each event, weekly mean TE and weekly mean PR were computed over a window spanning 8 weeks before and 3 weeks after the event and normalized to [0, 1] for visual comparison. The interpretation focused on whether TE rose during the 14-day pre-event window while PR remained relatively stable.

3.8 Markov Entropy Rate

Rolling Markov entropy rate was computed for each individual PV system using 14-day windows advanced by one day. A first-order Markov transition matrix was estimated from discretized PR states within each window. The entropy rate was computed as:

$$H\infty = -\Sigma_i \, \pi_i \, \Sigma_j \, T_{ij} \log_2(T_{ij}) \tag{4}$$

where π_i is the stationary probability of state i and T_{ij} is the transition probability from state i to state j. Equation (4) measures the average uncertainty of a system's state transitions over time. Higher entropy rate indicates less predictable behavior and lower entropy rate indicates more regular behavior. Unlike TE, which measures cross-system information flow, Markov entropy rate provides a per-system temporal complexity signal that can support abnormal behavior detection.

3.9 Computational Implementation

All analyses were implemented in Python using pandas, NumPy, SciPy, Matplotlib, and Seaborn. The full TE computation for 132 directed system pairs with 100 shuffle tests required approximately 15–20 minutes on a standard laptop CPU.

4. Data Collection

This study uses publicly available data from the DKASC, located in Alice Springs, Northern Territory, Australia, at approximately 23°46′S, 133°52′E, at an elevation of about 558 m (Desert Knowledge Australia Solar Centre 2026). The site is in a semi-arid region with an average global horizontal irradiance of approximately 6.17 kWh/m²/day, providing a high-irradiance test environment. The DKASC facility hosts PV systems based on diverse solar cell technologies, including monocrystalline silicon, polycrystalline silicon, amorphous silicon, cadmium telluride, and heterojunction modules. All systems are co-located at the same physical site, meaning they experience nearly identical irradiance, temperature, and weather conditions. This shared environmental setting makes it possible to examine cross-system directional dependence beyond common weather-driven responses.

For this study, three full calendar years of data spanning 2019 to 2021 were used. The original 5-minute resolution data were downloaded from the public DKASC portal under a Creative Commons license. After resampling to 15-minute averages, applying the GHI daytime filter, clipping PR to [0, 150]%, interpolating short gaps, and filtering by coverage, the final dataset retained 12 PV systems and 43,044 clean daytime observations.

In addition to the time-series power measurements, the DKASC provides operational notes documenting outages, metering failures, equipment issues, site modifications, and remediation activities. A total of 21 documented operational events occurring between November 2019 and December 2021 were identified and used as reference anchors for the fault-adjacent analysis. Table 2 summarizes the dataset parameters.

Table 2. Summary of the DKASC dataset used in this study.

Parameter	Value
Facility	DKASC, Alice Springs, Northern Territory, Australia
Coordinates	23°46′S, 133°52′E; elevation ≈558 m
Years covered	2019, 2020, and 2021
Raw data resolution	5-minute averaged readings
Analysis resolution	15-minute averages after resampling
Total raw rows	309,120 rows (after removing boundary duplicates)
Daytime rows used	43,044 observations (GHI > 100 W/m² filter applied)
PV systems analyzed	12 systems with ≥70% valid daytime coverage
Documented fault events	21 operational events, November 2019–December 2021
Data access	Publicly available from the DKASC portal under a Creative Commons license

5. Results and Discussion
5.1 Numerical Results
This section presents the primary quantitative findings from the TE analysis. This covers the directed information-flow network structure, the comparison between TE and symmetric baseline measures, and the fault-sensitivity analysis.

5.1.1 Directed Information-Flow Network
TE was computed for all 132 ordered directed pairs among the 12 selected PV systems. All 132 pairs were statistically significant under the shuffle-based null test at $\alpha = 0.05$. Because all systems share the same irradiance forcing, this saturation is expected. The analytical focus is therefore on TE magnitude and directional asymmetry. TE values ranged from 0.029 to 0.139 bits, corresponding to a 4.7× spread across directed pairs. The mean TE across all pairs was 0.079 bits, with a mean directional asymmetry ($|A_{TE}|$) of 0.027 bits. Table 3 summarizes the source-sink classification of each system.

Table 3. Per-system Transfer Entropy network metrics. Positive net flow indicates source behavior; negative net flow indicates sink behavior.

System	Out [bits]	In [bits]	Net [bits]	Role
M1_A	1.018	0.686	+0.332	Primary source
M7_A	1.004	0.714	+0.290	Source
M2_C	0.890	0.673	+0.218	Source
M5_B	0.956	0.824	+0.132	Weak source
M5_C	0.941	0.831	+0.110	Weak source
M3_A	0.834	0.861	−0.028	Near-neutral
M4_C	0.915	0.967	−0.052	Near-neutral
M6_B	0.878	0.967	−0.090	Weak sink
M6_A	0.884	0.973	−0.090	Weak sink
M4_B	0.716	0.877	−0.161	Sink
M5_A	0.769	1.094	−0.325	Primary sink

M4_A	0.609	0.946	−0.337	Primary sink

The results reveal that M1_A, M7_A, and M2_C are dominant information sources, with net flow values of +0.332, +0.290, and +0.218 bits, respectively. M4_A and M5_A are the dominant sinks with net flow values of −0.337 and −0.325 bits. The largest single-pair asymmetry was 0.073 bits between M1_A and M5_A. This source-sink structure indicates that the DKASC facility operates as a structured directed information-flow network, not a set of independent systems.

5.1.2 Comparison with Symmetric Baseline Measures

Table 4 compares TE against Pearson correlation and MI, demonstrating the additional directional information provided by TE.

Table 4. Comparison of Transfer Entropy against symmetric baseline measures across 132 directed system pairs.

Property	**Pearson r**	**Mutual Information**	**Transfer Entropy**
Symmetric ($X \rightarrow Y = Y \rightarrow X$)	Yes	Yes	No
Captures directionality	No	No	Yes
Model-free	No	Yes	Yes
Conditions on temporal lag	No	No	Yes
Mean value across pairs	0.857	1.075 bits	0.079 bits
Mean directional asymmetry	0.000	0.000	0.027 bits
Identifies sources and sinks	No	No	Yes
Fault-period shift detected	N/A	N/A	Yes, $p < 0.001$

High Pearson correlation (mean 0.857) and MI (mean 1.075 bits) confirm strong shared dependence, as expected from co-located systems. However, both measures are symmetric and cannot identify information sources or sinks. TE reveals a mean asymmetry of 0.027 bits, approximately 34% of mean TE magnitude. Critically, the relationship between |Pearson r| and $|A_{TE}|$ across the 132 directed pairs is weak ($r = 0.224$, $p = 0.071$) that suggests the directional structure captured by TE is not an artifact of shared irradiance-driven co-movement.

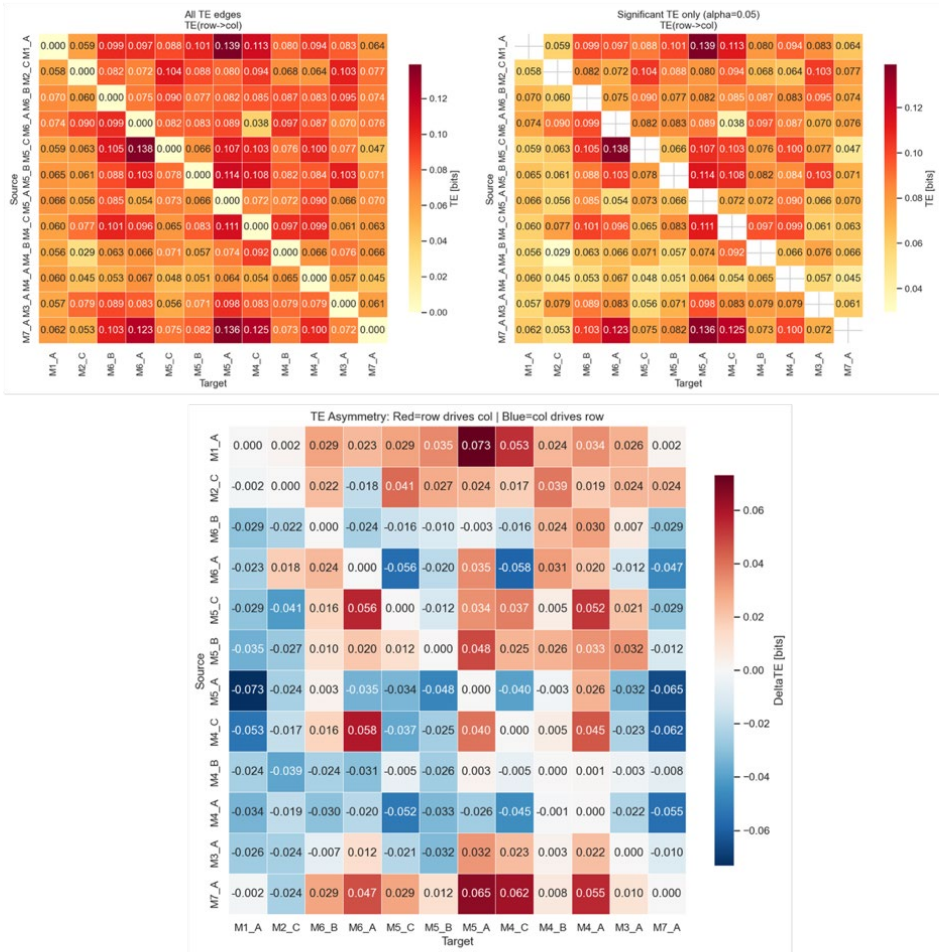

Figure 2. Directed information-flow structure: (a) TE matrix (TE row→column); (b) TE asymmetry matrix.

5.2 Graphical Results

Figure 2 shows the Transfer Entropy matrix (TE values for all 132 directed pairs) and the TE asymmetry matrix. The TE matrix reveals a heterogeneous distribution of information-flow strengths across the facility. The asymmetry matrix confirms that information flow is not bidirectionally balanced, with clear patterns of dominance in the directions predicted by the source-sink hierarchy in Table 3.

Figure 3 provides the baseline comparison among Pearson correlation, Mutual Information, and Transfer Entropy. The scatter plots confirm that TE captures a distinct directional component beyond shared statistical dependence. The TE asymmetry distribution shows a non-trivial positive mean of 0.027 bits, indicating systematic directionality in the PV network that is absent from symmetric measures.

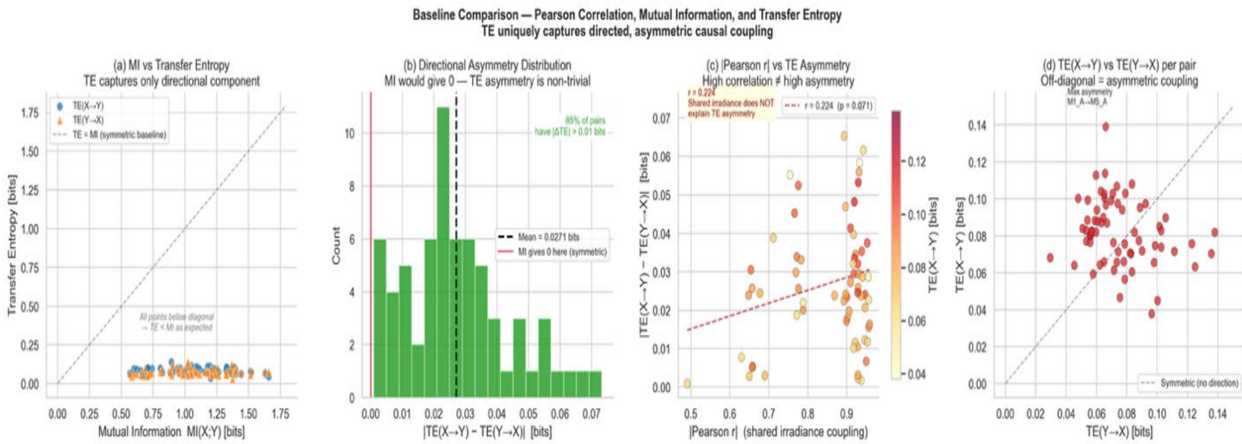

Figure 3. Baseline comparison between Pearson correlation, MI, and TE.

Figure 4 shows the significance framing results. The left panel plots absolute Pearson correlation against absolute TE asymmetry across system pairs that shows a weak and marginally insignificant relationship ($r = 0.224$, $p = 0.071$). This suggests that TE asymmetry is not strongly explained by shared irradiance-driven correlation alone. The right panel shows the non-uniform TE magnitude distribution with a 4.7× spread from 0.029 to 0.139 bits.

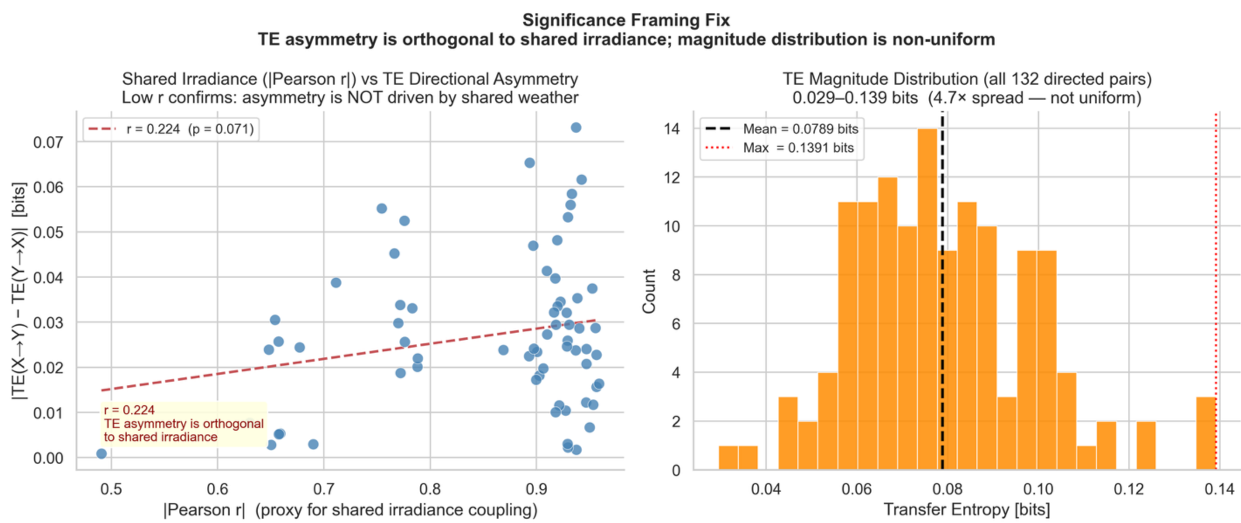

Figure 4. Significance framing: TE asymmetry is weakly related to shared irradiance-driven correlation (left); TE magnitude distribution is non-uniform across directed system pairs (right).

Figure 5 shows the Transfer Entropy distributions during normal and fault-adjacent periods. The fault-adjacent distribution is visibly shifted rightward relative to the normal distribution. Mean TE increased from 0.0772 bits (normal) to 0.0945 bits (fault-adjacent), a 22% increase. The Welch's t-test yielded $t = 6.35$ and $p = 9.27 \times 10^{-10}$; the KS test yielded $D = 0.326$ and $p = 1.35 \times 10^{-6}$. Both tests confirm that fault-adjacent periods are associated with statistically significant elevation of TE across the PV network.

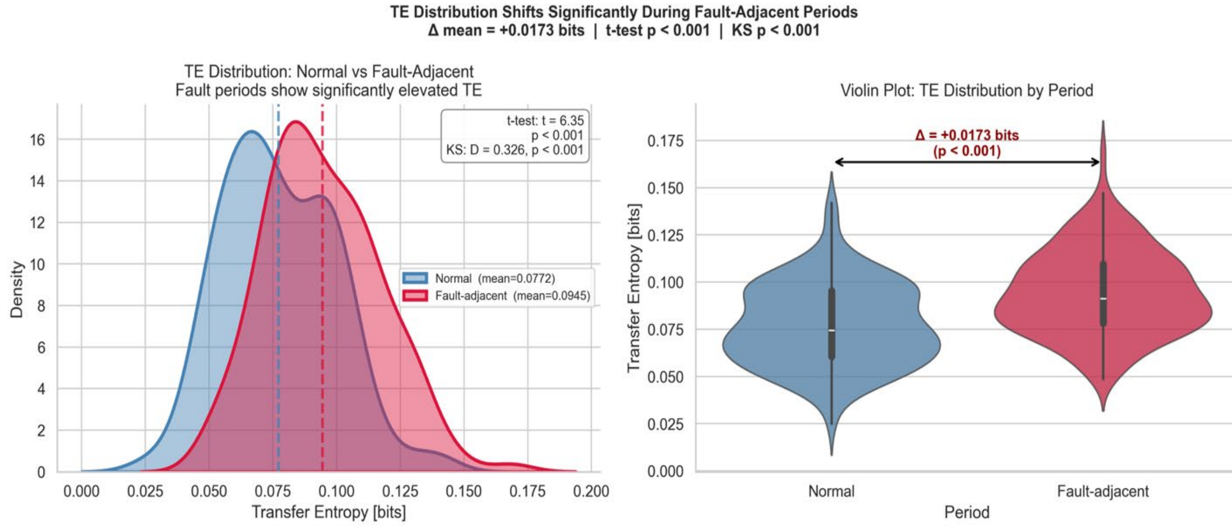

Figure 5. Transfer Entropy distributions during normal and fault-adjacent periods (density plot and violin plot).

Figure 6 presents the early-warning case studies for three major events. For all three events, normalized TE rises during the 14-day pre-fault window while normalized PR remains relatively stable, with PR degradation appearing at or after the documented event date. The pattern is most clearly visible for the January 2021 major data outage, where TE is distinctly elevated in the pre-fault window and PR drops sharply only at the event. These results support the interpretation that TE captures emerging system instability before PR-based monitoring visibly deteriorates.

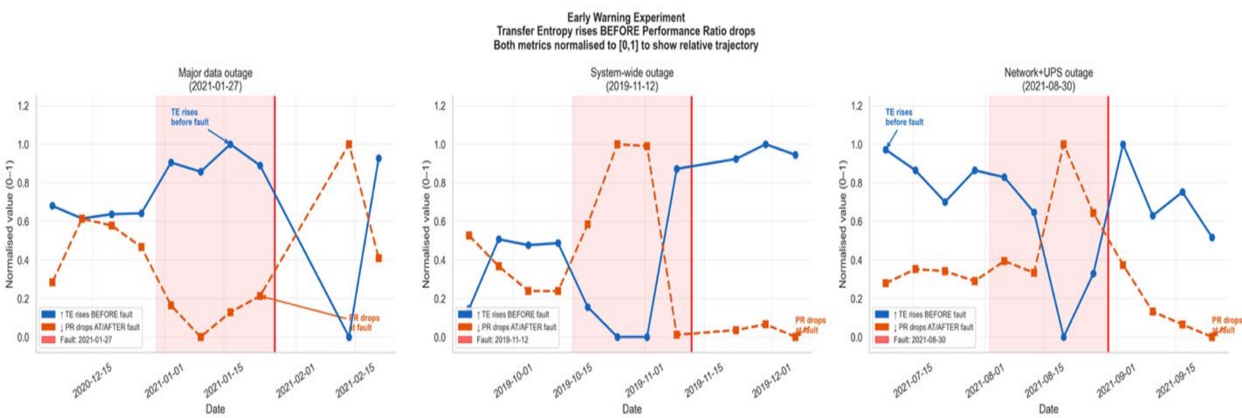

Figure 6. Early-warning case studies for three major documented events. TE often rises before visible PR degradation (both normalized to [0,1]).

5.3 Proposed Improvements

The proposed improvement is to add a continuous information-theoretic monitoring layer to existing PV maintenance workflows.

First, the study used a fixed 15-minute lag for TE computation. In future, we plan to evaluate sensitivity to lag choice by testing 30-minute, 60-minute, and multi-step lags to determine whether different temporal scales reveal additional or more stable directional structure. Second, the six-bin equal-frequency discretization should be compared against alternative bin counts (four and eight bins) to assess robustness of the source-sink classification and fault-sensitivity findings.

Third, the framework can be made more operationally actionable by combining cross-system TE with per-system rolling Markov entropy rate. A practical inspection-prioritization rule would be:

> *Flag a PV system for inspection when its incoming TE from source systems exceeds its recent rolling baseline AND its rolling Markov entropy rate is elevated relative to its seasonal pattern.*

This dual-signal rule requires two independent information-theoretic indicators to be simultaneously elevated, reducing false alarms. Such a rule would not replace drone-based physical inspection but would provide a continuous early-warning layer to prioritize when and where inspection resources are deployed. Fourth, more granular per-system fault labels should be developed using inverter logs and string-level monitoring to enable finer fault attribution. Fifth, validation on geographically and technologically diverse PV facilities is needed to test whether the source-sink structure and fault-sensitivity findings generalize beyond the DKASC site.

5.4 Validation

This section validates that all four research objectives stated in Section 1.1 have been met.

Objective 1 (Directed network structure). TE was computed for all 132 directed pairs. All pairs were statistically significant under the shuffle-based null test at $\alpha = 0.05$. TE values ranged from 0.029 to 0.139 bits. A clear source-sink hierarchy was identified (Table 3), with M1_A, M7_A, and M2_C as primary sources and M4_A and M5_A as primary sinks. Objective 1 is met.

Objective 2 (Comparison with symmetric measures). TE was compared against Pearson r and MI (Table 4). While high symmetric dependence is confirmed, TE reveals a mean asymmetry of 0.027 bits that is only weakly associated with shared correlation structure ($r = 0.224$, $p = 0.071$). TE provides directional information unavailable from symmetric measures. Objective 2 is met.

Objective 3 (Fault sensitivity). Mean TE increased by 22% during fault-adjacent periods (from 0.0772 to 0.0945 bits). This shift was confirmed by Welch's t-test ($t = 6.35$, $p < 0.001$) and the KS test ($D = 0.326$, $p < 0.001$). Documented disturbances are associated with statistically significant elevation of directional information flow. Objective 3 is met.

Objective 4 (Early-warning behavior). For three major documented events, normalized TE rose during the 14-day pre-fault window while normalized PR remained relatively stable. PR degradation appeared at or after each documented event. These results support the early-warning potential of TE-based monitoring, treated as case-study evidence. Objective 4 is met.

6. Conclusion

This paper presents a system-level information-theoretic framework for early-warning monitoring of co-located photovoltaic systems. By treating the solar facility as a directed network of interacting time series and using TE to quantify directional information flow, the framework reveals system-level behavior invisible to conventional performance monitoring.

Using three years of DKASC Alice Springs data, the framework was applied to 12 co-located PV systems and 43,044 clean daytime observations. TE values across 132 directed pairs ranged from 0.029 to 0.139 bits, corresponding to a 4.7× spread in information-flow strength. A clear source-sink hierarchy was identified, with systems M1_A, M7_A, and M2_C as primary sources and M4_A and M5_A as primary sinks. This demonstrate co-located PV systems contribute differently to the facility's temporal dynamics.

TE captured directional structure not available from symmetric measures. While Pearson correlation (mean 0.857) and MI (mean 1.075 bits) confirmed strong shared dependence, TE revealed a mean directional asymmetry of 0.027 bits that was only weakly related to the shared correlation structure ($r = 0.224$, $p = 0.071$). This suggests that TE captures directional behavior beyond weather-driven co-movement.

The fault-period analysis provided the strongest statistical validation. Mean TE increased from 0.0772 to 0.0945 bits during fault-adjacent periods, a 22% increase confirmed by both Welch's t-test ($t = 6.35$, $p < 0.001$) and the KS test ($D = 0.326$, $p < 0.001$), establishing that documented disturbances are associated with measurable changes in directional information flow across the PV network. Early-warning case studies further suggest that TE may detect emerging system instability before Performance Ratio visibly deteriorates.

All four research objectives were met. The unique contribution of this paper is the application of Transfer Entropy-based directed network analysis to co-located PV system monitoring, demonstrating that continuously available metering data can support interpretable early-warning signals beyond what conventional performance indicators provide. Combined with drone-based physical inspection, TE-based monitoring can help identify when a solar facility is entering an abnormal operating state and which systems may deserve closer investigation.

Future priorities include validation across geographically diverse PV facilities with varying climates, technologies, and operational condition, alongside sensitivity analysis for discretization choices, integration of electrical topology data for finer fault attribution, and prospective testing of the dual-signal inspection-prioritization rule in operational settings.

References

Barnett, L., Barrett, A. B. and Seth, A. K., Granger causality and transfer entropy are equivalent for Gaussian variables, *Physical Review Letters*, vol. 103, p. 238701, 2009.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D., Complex networks: Structure and dynamics, *Physics Reports*, vol. 424, pp. 175–308, 2006.

Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, 2nd Edition, John Wiley & Sons, Hoboken, NJ, 2006.

de Oliveira, A. K. V., Aghaei, M. and Rüther, R., Automatic inspection of photovoltaic power plants using aerial infrared thermography: A review, *Energies*, vol. 15, no. 6, 2022.

Desert Knowledge Australia Solar Centre, DKASC Alice Springs data download and notes on data, Available: https://dkasolarcentre.com.au/locations/alice-springs, Accessed on April 8, 2026.

Jordan, D. C. and Kurtz, S. R., Photovoltaic degradation rates—an analytical review, *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 1, pp. 12–29, 2013.

Khalid, A. M., Mitra, I., Warmuth, W. and Schacht, V., Performance ratio – crucial parameter for grid connected PV plants – comparative analysis and factors affecting its value, *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 1139–1158, 2016.

Lizier, J. T., JIDT: An information-theoretic toolkit for studying the dynamics of complex systems, *Frontiers in Robotics and AI*, vol. 1, p. 11, 2014.

Mellit, A. and Kalogirou, S., Recent advances in the application of infrared thermographic imaging and embedded artificial intelligence for fault diagnosis and predictive maintenance of photovoltaic plants: Challenges and future directions, *Renewable and Sustainable Energy Reviews*, vol. 223, p. 116057, 2025.

Michail, A., Livera, A., Tziolis, G., Carus Candás, J. L., Fernandez, A., Antuña Yudego, E., Fernández Martínez, D., Antonopoulos, A., Tripolitsiotis, A., Partsinevelos, P., Koutroulis, E. and Georghiou, G. E., A comprehensive review of unmanned aerial vehicle-based approaches to support photovoltaic plant diagnosis, *Heliyon*, vol. 10, no. 1, p. e23983, 2024.

Pillai, D. S. and Rajasekar, N., A comprehensive review on protection challenges and fault diagnosis in PV systems, *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 18–40, 2018.

Schreiber, T., Measuring information transfer, *Physical Review Letters*, vol. 85, pp. 461–464, 2000.

Shannon, C. E., A mathematical theory of communication, *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

Triki-Lahiani, A., Bennani-Ben Abdelghani, A. and Slama-Belkhodja, I., Fault detection and monitoring systems for photovoltaic installations: A review, *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2680–2692, 2018.

Vicente, R., Wibral, M., Lindner, M. and Pipa, G., Transfer entropy—a model-free measure of effective connectivity for the neurosciences, *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.

Walker, A. and Desai, J., Understanding solar photovoltaic system performance: An assessment of 75 federal photovoltaic systems, U.S. Department of Energy, Energy Efficiency and Renewable Energy, 2021.