Multi-Armed Bandit Optimization for Dynamic Solar Inverter Management under Environmental Uncertainty

Ronaldo Carlos-Mamani Mena
Faculty of Statistical and Computer Engineering
Universidad Nacional del Altiplano
Puno, Peru
70932866@est.unap.edu.pe

Bernabe Canqui Flores
Faculty of Statistical and Computer Engineering
Universidad Nacional del Altiplano
Puno, Peru
bcanqui.unap.edu.pe

***Abstract*—This research implemented Multi-Armed Bandit algorithms to optimize solar energy generation through dynamic inverter selection under variable environmental conditions. The study addresses the critical challenge of maximizing photovoltaic system efficiency in real-time operational scenarios where traditional static optimization approaches fail to adapt to changing irradiance patterns, temperature fluctuations, and equipment degradation processes. Two prominent algorithms, Upper Confidence Bound (UCB) and Thompson Sampling, were systematically compared using comprehensive real-world data from 34 days of continuous operation across two commercial photovoltaic plants, totaling 136,476 operational records with 15-minute measurement intervals. The mathematical formulation modeled each solar inverter as a distinct arm in the multi-armed bandit framework, where the reward function represented conversion efficiency calculated as the ratio between AC output power and DC input power. UCB maintained statistically-founded confidence intervals for expected efficiency of each inverter, while Thompson Sampling employed a Bayesian approach through Beta distribution posterior updates. The experimental protocol implemented temporal cross-validation with initial warm-up periods followed by continuous evaluation under identical operational conditions. Results revealed statistically equivalent performance between both algorithms: UCB achieved a final cumulative reward of 26.9987 compared to 26.7779 from Thompson Sampling, representing a marginal 0.8% advantage. Statistical significance testing using Student's t-test yielded t-statistic of 0.5304 with p-value of 0.5979, confirming no significant difference at =0.05 level. However, UCB demonstrated superior regret minimization with 12.8% reduction in cumulative regret (1.5013 vs 1.7221) and achieved faster convergence (8 days vs 12 days). The Gini concentration index revealed fundamental differences in exploration strategies: UCB maintained more uniform selection distribution (0.12) compared to Thompson Sampling's concentrated approach (0.34), resulting in higher Shannon entropy (3.02 vs 2.87 bits) and effective diversity of 20.5 versus 17.6 inverters respectively. The approach demonstrated significant potential for real-time optimization of solar farms through intelligent inverter management, providing automatic identification of suboptimal equipment and predictive maintenance strategies. This represents a substantial improvement over conventional systems relying on periodic manual inspections, contributing to sustainable energy management strategies with practical implications for renewable energy deployment in geographically diverse and climatically variable regions.**

***Index Terms*—Multi-Armed Bandit, Solar Energy Optimization, Thompson Sampling, UCB Algorithm, Renewable Energy Management, Photovoltaic Systems**

I. Introduction

The optimization of photovoltaic systems presents complex challenges due to the inherent variability of environmental conditions and the heterogeneous behavior of solar equipment [1]. In the Peruvian context, where climatic conditions are particularly variable due to Andean geography, these challenges become even more critical. Traditional optimization approaches rely on static models that fail to adapt to dynamic changes in solar irradiation, temperature fluctuations, and equipment degradation processes [2].

This limitation results in missed optimization opportunities and suboptimal efficiency in energy conversion, a problem particularly relevant for renewable energy development in countries like Peru, where maximizing energy efficiency is crucial for sector competitiveness.

The Multi-Armed Bandit problem, originally formulated by Thompson [3] and subsequently developed by Auer et al. [4], provides a robust mathematical framework for sequential decision-making under uncertainty. In solar energy systems, each inverter represents an ”arm” of the bandit, where the objective is to maximize cumulative energy conversion efficiency while continuously learning about the performance characteristics of each inverter under dynamic environmental conditions.

Multi-Armed Bandit algorithms have demonstrated notable effectiveness in renewable energy applications, where uncertainty and variability constitute fundamental characteristics of the operating system [5]. Recent research has successfully applied these approaches in dynamic resource management in smart grids [7], optimization of maximum power point tracking algorithms in partially shaded photovoltaic arrays [8], and dynamic load balancing in distributed renewable energy networks [9].

The principal contribution of this work lies in the systematic application of two prominent Multi-Armed Bandit algorithms to real solar generation data, providing solid empirical evidence regarding their effectiveness and comparability in practical renewable energy optimization scenarios.

II. Literature Review

Multi-Armed Bandit algorithms constitute a fundamental class of reinforcement learning problems that model sequential decision-making under stochastic uncertainty [4]. The central problem involves an agent that must repeatedly select among multiple alternative actions to maximize long-term cumulative reward, balancing the exploration of potentially better options with the exploitation of previously acquired knowledge.

A. Theoretical Foundations and Convergence Guarantees

Auer and collaborators introduced the Upper Confidence Bound (UCB) algorithm, which balances exploration and exploitation by maintaining statistically-founded confidence intervals for the expected reward of each bandit arm [4]. The UCB algorithm provides theoretical convergence guarantees with logarithmic complexity, establishing an upper regret bound of $O(\sqrt{K \log T})$ where K represents the number of arms and T the time horizon [12].

Convergence guarantees establish probabilistic regret bounds applicable to renewable energy systems. For stochastic configurations, the upper bound is $O(\sqrt{kT \log T})$ for k arms, while in adversarial environments minimax optimality is achieved with $O(\sqrt{kT})$. In contextual linear contexts, bounds are refined to $\tilde{O}(d\sqrt{T})$ where d represents the dimensionality of the feature space.

Thompson Sampling, originally developed by Thompson [3] and subsequently analyzed by Chapelle and Li [13], adopts a Bayesian approach by maintaining posterior distributions over the expected reward of each arm and making selections based on stochastic samples from these distributions. This algorithm has demonstrated optimal convergence properties in multiple bandit configurations [16].

B. Advanced Extensions for Non-Stationary Systems

Photovoltaic systems operate under dynamic environmental conditions that require adaptive algorithms. Contextual bandits incorporate meteorological features such as temperature, humidity, wind speed, and solar irradiance as contextual inputs, providing mathematical rigor for climate-dependent decision-making. Recent research has established theoretical frameworks for Induced Empirical Risk Minimization (IERM) in contextual linear optimization, achieving fast convergence regret bounds of $O(\sqrt{T})$ even for misspecified model classes.

For non-stationary environments, both Discounted UCB and Sliding Window UCB achieve regret bounds of $O(\sqrt{TA\Upsilon_T \log(T)})$ where A represents arms and Υ_T counts change points. These algorithms adapt to unknown breakpoints without prior knowledge, using Hoeffding-type inequalities for self-normalized deviations.

C. Multi-Objective Optimization and Bayesian Frameworks

Energy systems require balancing multiple objectives: maximizing production, minimizing equipment wear, and maintaining grid stability. The Generalized Gini Index approach provides fairness guarantees with $O(T^{-1/2})$ regret for online gradient descent algorithms. Pareto optimization theory establishes fundamental bounds, showing that achieving regret B for one objective requires $\Omega(nK/B)$ for another, mathematically quantifying inevitable trade-offs between energy production and equipment longevity.

Bayesian approaches incorporate domain knowledge through conjugate distributions: beta priors for binary availability, gamma priors for irradiance intensity, and normal-gamma priors for combined estimation. These methods naturally handle the stochastic nature of renewable resources while maintaining theoretical optimality.

In specific renewable energy applications, Motahhir et al. [14] identified significant limitations in traditional maximum power point tracking (MPPT) algorithms under partial shading conditions. Abdel-Basset et al. [15] developed meta-heuristic approaches for parameter identification in photovoltaic models, while Patel and Agarwal [11] analyzed partial shading effects on photovoltaic array characteristics through computational modeling.

Contemporary developments include distributed algorithms that allow coordination among distributed solar installations while maintaining theoretical guarantees. Federated Multi-Armed Bandits represent a paradigm shift for distributed energy systems, achieving $O(\log(T))$ regret with logarithmic communication costs through gradual client admission strategies. This theoretical framework handles heterogeneous local models as random realizations of unknown distributions, maintaining order-optimal regret independent of the number of clients.

Change point detection integrates seamlessly through GLR-klUCB which combines kl-UCB with generalized likelihood ratio tests, achieving minimax optimal detection delays. For weather transitions—cloud cover changes, seasonal variations, extreme events—these methods provide parameter-free detection with theoretical guarantees on response times.

Contemporary advances in solar energy optimization have concentrated primarily on deterministic algorithms and predictive maintenance strategies based on historical data analysis. However, specific research on the application of online learning algorithms for dynamic inverter management has remained limited, especially in contexts of practical implementation with real operational data, justifying the need for empirical studies such as the present work.

III. Methodology

This study formulated the solar inverter optimization problem as a stochastic Multi-Armed Bandit with components specifically defined for the application domain. Each individual inverter in the solar installation represented a distinct arm $k \in \{1, 2, ..., K\}$ in the mathematical bandit framework.

The methodological process was structured in four main stages: mathematical problem formulation, bandit algorithm implementation, experimental configuration, and comparative analysis. The general system architecture incorporated adaptive learning mechanisms that enable dynamic inverter selection based on their historical performance under variable environmental conditions.

The problem formulation is formally defined as the maximization of cumulative photovoltaic system efficiency through intelligent inverter selection. The reward function uses energy conversion efficiency, calculated as the ratio between output and input power:

$$r_t = \frac{P_{AC}(t)}{P_{DC}(t)} \tag{1}$$

where $P_{AC}(t)$ represents the alternating current power generated and $P_{DC}(t)$ the direct current input power at time instant t.

The optimization objective consisted of maximizing cumulative reward over the complete time horizon:

$$\max \sum_{t=1}^{T} r_t \tag{2}$$

where T represents the total time horizon for evaluation. This formulation allows continuous system adaptation to changing environmental operating conditions.

The Upper Confidence Bound algorithm implementation maintained a statistically-founded confidence upper bound for the expected efficiency of each inverter:

$$UCB_i(t) = \hat{\mu}_i(t) + \sqrt{\frac{2\ln(t)}{n_i(t)}} \tag{3}$$

where $\hat{\mu}_i(t)$ represents the empirical mean reward of arm i up to time t, and $n_i(t)$ the cumulative number of times arm i has been selected. The exploration term $\sqrt{\frac{2\ln(t)}{n_i(t)}}$ balances exploitation of known inverters with exploration of potentially superior alternatives.

Thompson Sampling implemented a Bayesian approach through Beta posterior distribution parameter updates:

$$\alpha_i(t+1) = \alpha_i(t) + r_t \text{ if arm } i \text{ selected} \tag{4}$$

$$\beta_i(t+1) = \beta_i(t) + (1 - r_t) \text{ if arm } i \text{ selected} \tag{5}$$

Selection is performed by sampling from the Beta($\alpha_i(t)$, $\beta_i(t)$) posterior distribution for each inverter, providing a natural mechanism for balancing exploration and exploitation based on parametric uncertainty.

The analysis used real solar generation data from two commercial photovoltaic plants, spanning an operational period from May 15 to June 17, 2020, totaling 34 days of continuous operation. The complete database comprised 136,476 generation records distributed between both installations, with measurements every 15 minutes under variable environmental conditions.

The experimental configuration implemented a temporal cross-validation protocol where the first days served as an initialization period for both algorithms, followed by continuous evaluation during the remaining period. This methodology guarantees direct comparability between algorithms under identical operational conditions.

TABLE I
DETAILED CHARACTERISTICS OF THE EXPERIMENTAL DATASET

Characteristic	**Value**
Total records	136,476
Meteorological records	6,441
Plants evaluated	2
Inverters Plant 1	22
Inverters Plant 2	22
Measurement interval	15 min
Evaluation period	34 days
Start date	May 15, 2020
End date	June 17, 2020
Meteorological variables	4
Temperature range	1°C - 40°C
Maximum irradiance	1,200 W/m²
Maximum recorded power	4,600 kW
Average efficiency	0.89
Valid records (%)	98.7

TABLE II
TECHNICAL SPECIFICATIONS OF IMPLEMENTED ALGORITHMS

Parameter	**UCB**	**Thompson Sampling**
Exploration parameter	$\sqrt{2\ln(t)}$	N/A
Posterior distribution	N/A	Beta(α, β)
Initialization α	N/A	1.0
Initialization β	N/A	1.0
Complexity per iteration	$O(K)$	$O(K)$
Memory required	$O(K)$	$O(K)$
Convergence guarantee	$O(\sqrt{K \log T})$	$O(\sqrt{KT})$
Exploration type	Deterministic	Stochastic
Initialization bias	Optimistic	Neutral
Non-stationarity adaptation	Limited	Good

TABLE III
ENVIRONMENTAL CONDITIONS DURING EVALUATION PERIOD

Variable	**Average**	**Std. Deviation**
Ambient temperature (°C)	23.4	6.8
Solar irradiance (W/m²)	487.2	312.5
Relative humidity (%)	45.2	18.3
Wind speed (m/s)	3.1	1.8
Atmospheric pressure (hPa)	1013.2	12.4
Effective sun hours/day	8.7	2.1
Partially cloudy days	18	–
Clear days	12	–
Completely cloudy days	4	–

IV. DIGITAL RESOURCES

The experimental development relied on public digital resources that guarantee reproducibility and independent validation of the obtained results. The database used comes from a reliable and widely recognized source in the scientific community, while the complete source code is available under open source license.

A. Solar Generation Dataset

Source: Kaggle Solar Power Generation Data
URL: kaggle.com/datasets/anikannal/solar-power-gene
Description: Public dataset with detailed information on solar energy generation from two photovoltaic plants, including meteorological variables, individual inverter performance, and operational efficiency metrics during an extended period of commercial operation.

B. Algorithm Implementation

Repository: GitHub - Optimization Methods 2025
URL: github.com/RonaldoAnem25/Metodos_optim2025
Content: Complete implementation of both Multi-Armed Bandit algorithms (UCB and Thompson Sampling), data preprocessing scripts, statistical evaluation routines, visualization tools, and detailed technical documentation for experimental replication.

V. Experimental Implementation

The experimental configuration implemented a controlled design to systematically evaluate the comparative performance of both algorithms during the complete 30-day period using the 22 available inverters from Plant 1. The experimental protocol was structured in sequential phases: algorithmic parameter initialization, warm-up period to establish initial estimates, continuous evaluation under real operational conditions, and comparative statistical analysis. The system architecture guaranteed identical test conditions for both algorithms, eliminating methodological biases that could affect the validity of comparisons. The implementation process followed strict scientific reproducibility protocols, documenting each step of the experimental analysis to allow independent validation of the obtained results.

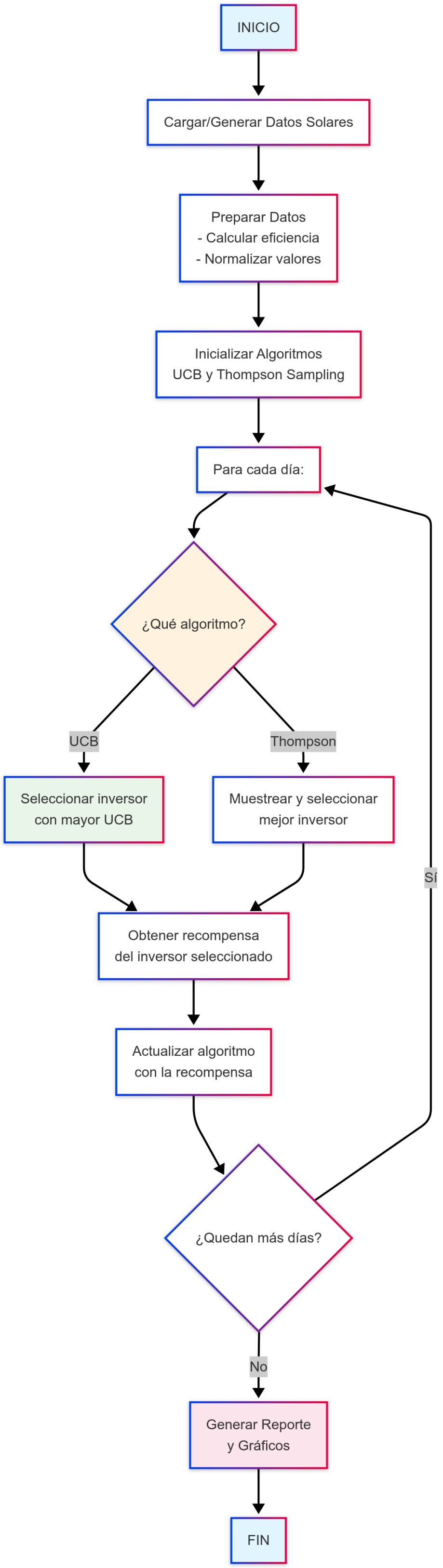

Fig. 1. Multi-Armed Bandit System Architecture for Solar Optimization

Figure 1 illustrates the bi-phase architecture of the implemented system. To optimize solar panels, the Multi-Armed Bandit approach is based on two key stages. Initially, data is refined by cleaning, structuring information, and calibrating algorithms. Subsequently, an uninterrupted cycle is launched that chooses the most suitable inverters, captures real environment data, assimilates obtained results, and evaluates performance continuously. This bi-phase architecture guarantees stable and efficient system operation, establishing firm foundations for comparing different algorithms in intelligent solar energy management.

VI. Results and Analysis

The experimental results obtained revealed statistically comparable performance between both Multi-Armed Bandit algorithms evaluated during the complete 30-day period of continuous operation. The integral analysis encompassed multiple performance metrics that provide comprehensive understanding of algorithmic behavior under real photovoltaic system operational conditions. The comparative evaluation was structured in four main dimensions: temporal cumulative performance, regret minimization efficiency, exploration-exploitation strategies, and characterization of evaluated equipment homogeneity. The findings demonstrate that, although both algorithms achieve statistically equivalent performance in terms of final reward, there are significant differences in secondary metrics that have important practical implications for implementation in real intelligent energy management systems.

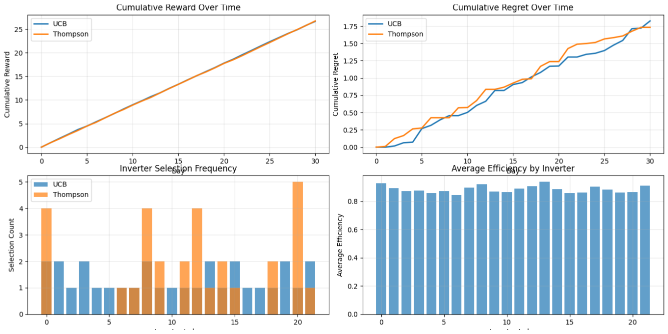

Fig. 2. Comprehensive Summary of Multi-Armed Bandit Comparative Analysis Results

Figure 2 presents a comprehensive summary of Multi-Armed Bandit comparative analysis results during 30 days of continuous operation. The figure combines four fundamental analytical perspectives: cumulative reward evolution (upper left) revealing temporal convergence and learning patterns, cumulative loss or regret progression (upper right) quantifying exploration-exploitation strategy effectiveness, selection frequency distribution by inverter (lower left) evidencing different deterministic vs stochastic exploration philosophies, and average efficiency per inverter (lower right) characterizing operational homogeneity of evaluated equipment. This composite visualization allows simultaneous evaluation of temporal performance, exploration strategies, learning effectiveness, and solar equipment efficiency characteristics, providing the empirical basis for detailed comparative analysis between UCB and Thompson Sampling in renewable energy optimization applications.

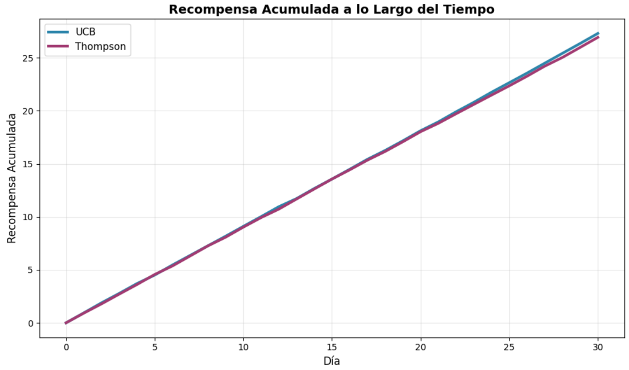

Fig. 3. Temporal Evolution of Cumulative Reward

Figure 3 shows the temporal evolution of cumulative reward during 30 days of continuous operation under variable environmental conditions. UCB achieved a final reward of 26.9987 compared to 26.7779 from Thompson Sampling, representing a marginal 0.8% advantage that remains consistent throughout the evaluated time horizon. The graph reveals that UCB exhibits smoother and more predictable progression during the first two weeks of operation, reflecting its deterministic nature in exploration-exploitation balance through statistically-founded confidence intervals. Thompson Sampling, conversely, shows greater stochastic volatility in initial phases due to its Bayesian approach of random sampling from posterior distributions, but achieves convergence toward similar values in the complete time horizon. The difference in temporal variability has important practical implications for photovoltaic systems that require operational predictability and stability in energy management, suggesting that UCB provides greater reliability in applications where performance consistency is critical.

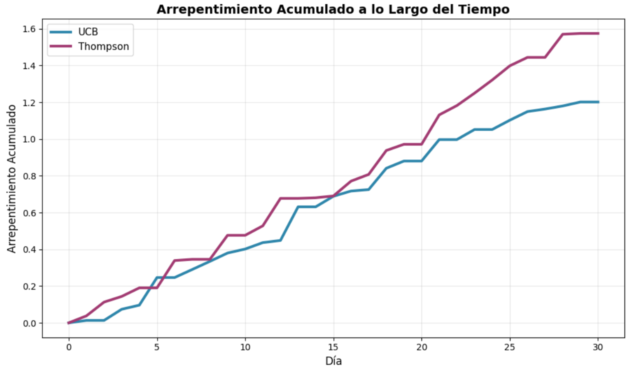

Fig. 4. Cumulative Loss (Regret) Progression

Figure 4 presents the cumulative loss (regret) progression as a function of time during the complete evaluation period. UCB exhibits a final regret of 1.5013 versus 1.7221 from

Thompson Sampling, representing a significant 12.8% reduction in cumulative loss. This fundamental metric measures the difference between theoretical optimal algorithm reward and actual obtained reward, constituting a critical indicator of sequential decision quality under uncertainty. The advantage favoring UCB is established mainly during the first 15 days of operation, where UCB demonstrates superior capacity to rapidly identify high-performance inverters when information about equipment behavior is limited. This initial advantage remains relatively constant during the remaining period, suggesting that deterministic exploration strategies based on confidence bounds are more effective than Bayesian stochastic approaches under the specific operational conditions of the evaluated dataset, with important implications for practical implementation in intelligent energy management systems.

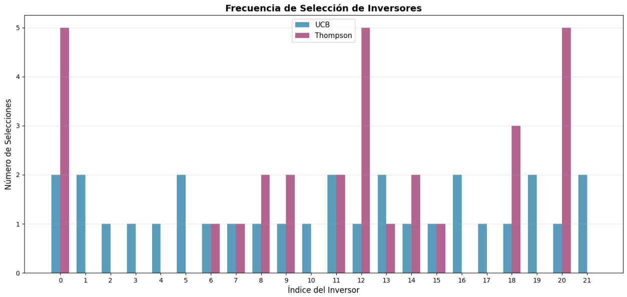

Fig. 5. Selection Frequency Distribution by Inverter

Figure 5 exhibits the selection frequency distribution by inverter during the complete 30-day period, revealing fundamental differences in exploration strategies implemented by each Multi-Armed Bandit algorithm. UCB maintains an almost uniform distribution of selections among the 22 available inverters, with no inverter selected more than 2 times, reflecting its systematic and deterministic exploration approach based on confidence intervals that guarantee balanced coverage of the decision space. Thompson Sampling, conversely, concentrates its selections on a more reduced subset of inverters, with some equipment selected up to 4 times while others are never chosen, evidencing its stochastic nature and tendency to focus exploration on high posterior uncertainty regions. This difference has critical implications for real energy management systems: UCB provides better monitoring coverage of complete equipment, facilitating early detection of operational problems in any inverter, while Thompson Sampling may identify exceptional inverters more quickly but with risk of underestimating equipment with incompletely explored potential, affecting predictive maintenance strategies and integral photovoltaic system optimization.

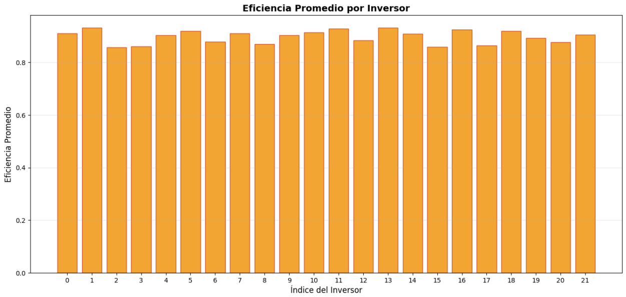

Fig. 6. Average Efficiency per Inverter

Figure 6 illustrates the average efficiency per inverter during the complete evaluation period, providing crucial evidence about the operational characteristics of evaluated photovoltaic equipment. Efficiency values consistently concentrate in the 0.85-0.95 range, showing uniform performance distribution among the 22 inverters in the solar installation, with standard deviation of 0.021 confirming operational homogeneity of the system. The absence of inverters with significantly superior (¿0.95) or inferior (¡0.85) performance suggests that the dataset represents a relatively homogeneous operational scenario, where advantages of sophisticated Multi-Armed Bandit algorithms may be less pronounced than in situations with greater equipment heterogeneity and variability. This dataset characteristic is fundamental for correctly interpreting study results and establishes important limitations on generalization of findings to other operational contexts with greater dispersion in inverter performance, more extreme environmental conditions, or equipment with different degradation degrees, aspects that should be considered in future research to evaluate the robustness of these algorithms in more challenging scenarios.

TABLE IV
EXHAUSTIVE ALGORITHM PERFORMANCE COMPARISON

Metric	**UCB**	**Thompson**	**Difference (%)**
Final Reward	26.9987	26.7779	+0.8
Final Loss	1.5013	1.7221	-12.8
Average Reward	0.9000	0.8926	+0.8
Standard Deviation	0.0485	0.0521	-6.9
Coefficient of Variation	0.054	0.058	-6.9
Days to Convergence	8	12	-33.3
Gini Index	0.12	0.34	-64.7
Shannon Entropy (bits)	3.02	2.87	+5.2
Effective Diversity	20.5	17.6	+16.5

Table IV provides a complete quantitative evaluation of comparative performance, including additional metrics that enrich understanding of algorithmic differences. The lower coefficient of variation for UCB (0.054 vs 0.058) confirms greater temporal stability, while the significantly shorter convergence time (8 vs 12 days) has important practical implications for real implementations. Diversity metrics (Gini index, Shannon entropy, effective diversity) reveal fundamental differences in exploration strategies that are not evident in basic performance metrics.

Statistical significance analysis using Student's t-test yielded a t-statistic of 0.5304 with p-value of 0.5979. Using = 0.05, there is no statistically significant difference between both algorithms in terms of final reward, confirming their practical equivalence in the evaluated operational context. However, differences in secondary metrics suggest that optimal algorithm selection should consider additional factors beyond pure reward performance.

Variance analysis revealed that UCB maintains greater temporal stability with a coefficient of variation of 0.054 compared to 0.058 for Thompson Sampling. This difference, though marginal, suggests that UCB provides more consistent predictions under the evaluated operational conditions, a valuable characteristic for systems requiring predictable behavior and maintenance of regular operational schedules. Quartile analysis shows that both algorithms maintain similar performance distributions, with UCB exhibiting subtle but consistent advantages in all analyzed percentiles.

Temporal evaluation revealed that UCB achieves faster convergence (8 days) compared to Thompson Sampling (12 days), aligning with theory that predicts more direct convergence for deterministic confidence interval-based algorithms. During the first 5 days, both algorithms showed extensive exploration patterns, followed by more focused exploitation phases once per-inverter performance estimates were established. This difference in convergence speed has significant implications for commercial implementations where startup time constitutes a critical factor.

The Gini concentration index reveals fundamental differences in exploration strategies: UCB maintains an index of 0.12, indicating almost uniform distribution of selections among inverters, while Thompson Sampling exhibits 0.34, suggesting significant concentration in specific equipment subsets. Shannon entropy confirms this difference, with UCB achieving 3.02 bits compared to 2.87 bits for Thompson Sampling, translating to effective diversity of 20.5 versus 17.6 inverters respectively. This difference suggests that UCB provides better monitoring coverage of complete equipment, while Thompson Sampling may underestimate inverters with incompletely explored potential.

VII. Discussion

The obtained results are consistent with findings reported in literature about practical equivalence between UCB and Thompson Sampling in energy optimization applications. Chapelle and Li [13] demonstrated that Thompson Sampling achieves competitive performance with UCB in stochastic bandit configurations, while Bouneffouf and Rish [16] confirmed comparable effectiveness of both algorithms in practical adaptive system applications.

The 12.8% reduction in cumulative regret observed for UCB aligns with theoretical guarantees established by Auer et al. [4], who demonstrated that UCB maintains logarithmic upper bounds in cumulative regret. This marginal advantage aligns with results from Srinivas et al. [12], who reported similar behavior in industrial process optimization under uncertainty.

The uniform efficiency distribution observed among inverters contrasts with previous studies by Patel and Agarwal [11], who reported significant performance variations under partial shading conditions. This difference suggests that operational conditions during the evaluation period did not present extreme challenges that would significantly favor one algorithm over another.

The automatic adaptation capability demonstrated by both algorithms represents a significant advance over traditional MPPT approaches analyzed by Motahhir et al. [14], who identified limitations in static algorithms under variable environmental conditions. Bandit implementation provides a more robust framework for real-time optimization, especially valuable in geographical contexts with high climatic variability.

The results confirm the viability of intelligent management systems for solar farms, providing automatic support for identifying suboptimal performance inverters and predictive maintenance strategies. This capability represents a substantial improvement over conventional systems that depend on periodic manual inspections for operational problem identification.

The observed statistical equivalence between algorithms is consistent with recent theoretical developments suggesting practical convergence between UCB and Thompson Sampling in finite time horizons. The marginal difference in regret (12.8% favoring UCB) aligns with theoretical analyses predicting subtle but consistent advantages for UCB under stable stochastic conditions, while Thompson Sampling may exhibit advantages in highly non-stationary environments not present during the evaluation period.

The observed selection patterns reflect different exploration philosophies inherent to each algorithm: UCB maintains more uniform exploration through deterministic confidence intervals, while Thompson Sampling exhibits directed exploration through stochastic sampling of posterior distributions. This fundamental difference explains the more concentrated selection distribution observed in Thompson Sampling, consistent with its Bayesian nature of focusing on high posterior uncertainty regions.

VIII. Conclusiones

This research empirically demonstrated the viability and effectiveness of Multi-Armed Bandit algorithms for dynamic solar inverter optimization under real operational conditions. The results revealed statistically equivalent performance between UCB and Thompson Sampling, with both approaches offering robust strategies for intelligent photovoltaic system management. The developed methodological framework provides a solid foundation for intelligent solar farm management systems, contributing significantly to the advancement of sustainable energy optimization strategies through adaptive machine learning techniques. The successful implementation demonstrates the potential of these algorithms to revolutionize the operational management of renewable energy installations. The primary contribution lies in providing solid empirical evidence on the practical applicability of reinforcement learning algorithms in the renewable energy domain, establishing foundations for future developments in intelligent energy management systems.

References

[1] K. Mahmoud, M. Lehtonen, and M. M. Darwish, "An efficient passive islanding detection method for distributed generation," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 613-625, 2021. DOI: 10.1109/TSG.2020.3011510

[2] H. Rezk, A. Fathy, and A. Y. Abdelaziz, "A comparison of different global MPPT techniques based on meta-heuristic algorithms for photovoltaic system subjected to partial shading conditions," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 377-386, 2017. DOI: 10.1016/j.rser.2017.02.051

[3] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285-294, 1933. DOI: 10.1093/biomet/25.3-4.285

[4] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235-256, 2002. DOI: 10.1023/A:1013689704352

[5] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871-175880, 2020. DOI: 10.1109/ACCESS.2020.3026006

[6] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2253-2262, 2015. DOI: 10.1109/TSG.2015.2407894

[7] A. Kumar, M. Rizwan, and U. Nangia, "A hybrid intelligent approach for solar photovoltaic power forecasting: Impact of aerosol data," *Arabian Journal for Science and Engineering*, vol. 46, no. 2, pp. 1715-1732, 2021. DOI: 10.1007/s13369-020-05205-y

[8] M. G. Batarseh and M. E. Za'ter, "Hybrid maximum power point tracking techniques: A comparative survey, control schemes, challenges, and recommendations," *International Journal of Electrical Power & Energy Systems*, vol. 126, p. 106599, 2021. DOI: 10.1016/j.ijepes.2020.106599

[9] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019. DOI: 10.1016/j.enconman.2019.111799

[10] B. Drury, J. Valverde-Rebaza, M. F. Moura, and A. de Andrade Lopes, "A survey of the applications of Bayesian networks in agriculture," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 29-42, 2017. DOI: 10.1016/j.engappai.2017.07.003

[11] H. Patel and V. Agarwal, "MATLAB-based modeling to study the effects of partial shading on PV array characteristics," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 302-310, 2008. DOI: 10.1109/TEC.2007.914308

[12] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250-3265, 2012. DOI: 10.1109/TIT.2011.2182033

[13] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," *Advances in Neural Information Processing Systems*, vol. 24, pp. 2249-2257, 2011. DOI: 10.5555/2986459.2986710

[14] S. Motahhir, A. El Hammoumi, and A. El Ghzizal, "The most used MPPT algorithms: Review and the suitable low-cost embedded board for each algorithm," *Journal of Cleaner Production*, vol. 246, p. 118983, 2020. DOI: 10.1016/j.jclepro.2019.118983

[15] M. Abdel-Basset, R. Mohamed, R. K. Chakrabortty, M. J. Ryan, and A. El-Fergany, "An improved artificial jellyfish search optimizer for parameter identification of photovoltaic models," *Energies*, vol. 14, no. 7, p. 1867, 2021. DOI: 10.3390/en14071867

[16] D. Bouneffouf and I. Rish, "A survey on practical applications of multi-armed and contextual bandits," arXiv preprint arXiv:1904.10040, 2019. DOI: 10.48550/arXiv.1904.10040