Causal Inference Testing in MVL Signals

Evelyn Tadlock
Division of Mathematics
Texas Woman's University
Denton, USA
0009-0000-9792-088X

Allison Reed
Division of Biology
Texas Woman's University
Denton, USA
0009-0000-0742-8366

Micah Thornton
Division of Mathematics
Texas Woman's University
Denton, USA
0000-0002-9093-045X

***Abstract*—We extend causal inference models pioneered by Donald Rubin and Judea Pearl in the context of multi-valued logic signals with arbitrary alphabet size. Here, "causal relationships" refers to statistical treatment–outcome relationships in the context of MVL sequences and not to physiological causation. Our approach was developed and tested using patterned sequences with embedded causal relationships and naturally occurring sequences such as discretized power levels from a bicycle ride. Our framework was also tested on synthetic sequences from true and pseudo-random number generators. Our results show we can discriminate between some purely random sequences and those with embedded causal dependencies.**

***Index Terms*—Multiple-Valued Logic, Causal Inference, Propensity Score, State-Transition Graphs, Signal Analysis**

I. Introduction

Causal inference attempts to analyze whether and how a single event or treatment causally affects another. The most notable frameworks for causal inference are the Rubin Causal Model (RCM) and Pearl's structural model. Both frameworks have the power to quantify cause-and-effect relationships in observational data. However, they have been primarily developed for binary-valued or single-valued systems, where there is treatment versus control in biomedical research or binary node states in Pearl's directed acyclic graphs (DAGs) [1].

Some contemporary electronic devices use multi-valued logic instead. GPS-enabled cycling devices convert continuous measurements (power, heart rate) into k-ary zones, thus generating MVL signals that require appropriate causal inference frameworks [2], [3]. An athlete is typically able to view this data through a piece of software that takes these continuous measurements and discretizes them for analysis. For example, power output is commonly categorized into seven different discrete zones based on Functional Threshold Power [4], [5] and thus they become multi-valued systems that need MVL frameworks instead of simple binary ones.

In this work, we extend both the Rubin and Pearl models to multi-valued logic (MVL) signals, enabling a direct analysis of causality in k-ary data streams. Our methodologies adapt propensity score matching (PSM) and the directed graphical model to an MVL setting, then introduce a statistical testing procedure for detecting non-random dependencies within the MVL sequences. Our proposed approach is validated both on created data, where causal structure is inserted and controlled, and on real-world k-ary signals derived from discretized interval-based bicycle power output.

Although the validation of the proposed MVL causal inference framework has been carried out by using discretized cycling power signals, the model is not limited to the specific domain. Healthcare monitoring can be a source of multi-valued physiologic representations where continuous measurements are converted into symbolic bands, such as heart rate zones, glucose ranges, EEG encodings, and sleep stage classifications. Treatment outcome dependencies, which are temporally structured, may thus be examined from the perspective of MVL causal signal methodologies.

II. Background & Prior Works

A. Rubin Causal Model

The Rubin Causal Model [6] estimates potential treatment outcomes by comparing similar units. It uses the propensity score, which was introduced by Rosenbaum and Rubin [7], to summarize these similarities and support the estimation of causal effects. A full framework for this approach is provided in Imbens and Rubin [8].

B. Pearl's Structural Model

Pearl's causal framework [9], [10] represents causal relationships with directed acyclic graphs (DAGs), where nodes denote variables and edges indicate causal influence [11]. Pearl's model operates on three levels, which are association, intervention, and counterfactuals. The model also uses do-calculus to identify causation from observed data [11], [12].

C. Continuous vs. Multi-Valued Discrete

Alternative approaches have been suggested to extend causal inference frameworks to non-binary treatments. In their 2004 study, Hirano and Imbens [13] suggested that the propensity score method be generalized to situations where treatments are continuous and the treatment T is in the continuous interval $[t_0, t_1]$. For the generalized propensity score, $r(t, x) = f_{T|X}(t \mid x)$ is the conditional density of the treatment level t and covariates x. Dose-response functions, $\mu(t) = \mathbb{E}[Y(t)]$, are the average potential outcomes for a given treatment t. Other previous causal inference research have also focused on scenarios of multiple treatment values. Imbens (2000) [14] referred to propensity scores as the core instrumental variables in dose-response estimation, and Imai and van Dyk (2004) [15] extended the concepts to general treatment regimes. Han et al. in their study of causal analysis with multidimensional

treatments point out that treatments can have different values whether categorical or continuous, and the classical binary causal estimators are not sufficient.

Although prior research does not mention MVL and state-transition structures, they are a backdrop for extending causal frameworks to more complex multi-level signals. These treatment frameworks are mostly concentrated on the continuous treatment, or static categorical case, and lack the temporal aspect and state transitions of a multi-valued logical system.

Prior extensions of the Rubin model to multi-valued treatments look at unit-level exposure variation only, and do not consider temporally indexed symbolic sequences. On the other hand, MVL signals consist of ordered states, transition dependencies, and block-defined treatment–outcome partitions, and thus need sequence-specific decomposition and multinomial treatment encoding which go beyond traditional generalized propensity score formulations. Generalized propensity score approaches operate at the unit-level exposure and do not account for temporally indexed symbolic sequences or block-defined combinatorial treatment partitions.

Our method is a step ahead in this respect by deriving the multi-level treatment case to k-ary MVL signals where each level is a discretized state for transition graphs.

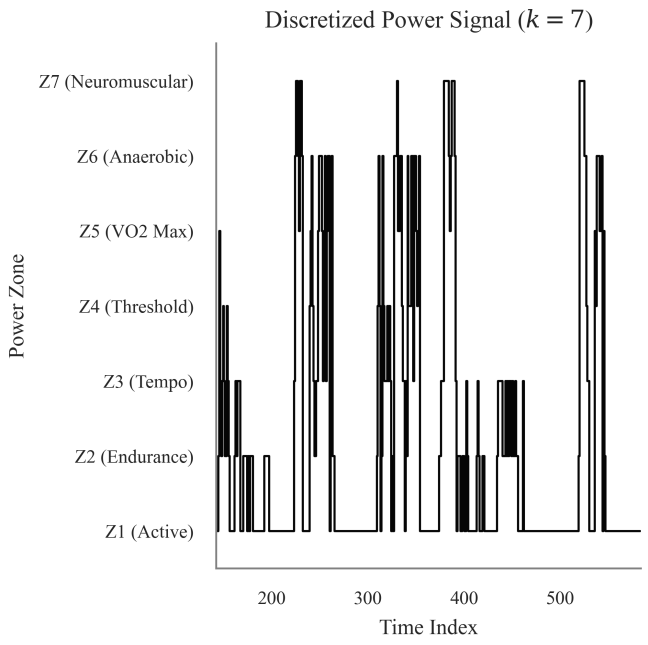

Fig. 1. Discretized cycling power signal into seven physiological zones (Z1–Z7).

III. Methodology and Approach

A. Rubin Causal Model

1) Multiple-Valued Signal Representation: The methodology for the Rubin Causal model (RCM) utilizes multi-valued logic (MVL) signals of alphabet length k, where each of the signal positions has a discrete value from the set $\{0, 1, \ldots, k-1\}$. Random number generators are utilized to produce k-ary symbols for the synthetic sequences of random numbers, and an equal-width discretization step is also performed to transform continuous measurements to MVL sequences for the real-world analysis. For a continuous signal of size N the range of observed values is divided into k bins of equal width, and each measurement is then allocated its bin index in the set $\{0, 1, \ldots, k-1\}$. Here, discretizations preserve temporal order and relative magnitude; creating a symbolic form suitable for causal inference as shown in Fig. 1.

Extending the Rubin model to MVL signals raises issues such as the expansion of the combinatorial treatment, sparsity of the propensity score, and sensitivity to block size and discretization. We limited the study to treatment classes that were observed and we tested the strength of our results across different discretization schemes to deal with these issues. Restricting analysis to observed treatment classes prevents exponential growth of empty or sparse treatment categories and stabilizes propensity estimation.

2) Signal Decomposition: To extend RCM to signal causality detection, the treatment (t_i), characteristics i, and the outcome (o_i) positions are defined in each block. A multi-valued signal of alphabet size k and length N is split into N/n substrings of size n. Of the n elements in each substring, we designate a portion of them for the indication of the treatment t_i, a portion of them for the characteristics c_{ij}, where j indexes the covariate position within each block, and another portion of them for the outcome o_i. The number of positions which are designated as treatment positions is n_t, the number designated as characteristics is n_c, and the number designated as outputs is n_o.

3) Discretization and Covariate Encoding: Multi-valued signals and categorical treatment allocations are handled by the covariate positions of the feature and are encoded as integer MVL symbols in $\{0, 1, \ldots, k-1\}$. If there are no covariates present, then a constant feature is used to ensure convergence of the model. Any continuous or categorical data values were discretized into $0, 1 \ldots, k-1$. Any missing data is assigned to a default bin, although in our experiments we did not have any missing data.

4) Propensity Score Estimation: Following the splitting of the string into the constituent substrings and the assignment of each substring to one of k^{n_t} possible treatments, which were restricted to those observed, a propensity model, utilizing a single multinomial logistic regression, estimates the probability of each treatment class. In contrast to fitting k^{n_t} separate binary logistic regressions, a single multinomial logistic regression model is used in which the treatment $t \in 1, 2, \ldots, k^{n_t}$ functions as a multi-class outcome. The model returns the probability that a substring belongs to each treatment type when regressed on the n_c characteristic positions. Thus, for every one of the k^{n_t} treatment types implied by the selected treatment positions in each substring, the multinomial model provides the estimated propensity score, with the characteristic positions included as the independent factor variables. For treatment type $t \in 1, 2, \ldots, k^{n_t}$, let t_k denote the t^{th} treatment

category, and the model yields the probability $\Pr(t = t_k)$ used for matching.

5) Greedy Matching Algorithm: Blocks are assigned to non-control treatment groups and then matched to control blocks that have similar propensity scores. The control group blocks are defined as those with the most prevalent treatment class in general, but could be relaxed to include all observations with a different treatment type. Matching is then performed greedily, without replacement, to create pairs with similar covariate distributions but different treatment assignments. Because the multinomial logistic regression produces a full vector of treatment probabilities for each substring, the matching is performed using the propensity score corresponding to the observed treatment category of each block (though in general euclidean distance could also be used). Each block from a non-control treatment group is matched to the control block whose multinomial propensity score for that treatment is closest, thereby ensuring comparable likelihoods of receiving that treatment under the observed characteristics.

6) Outcome Comparison via Hamming Distance: Hamming distance is generalized to compute each pair by counting the number of differing positions in the outcome vectors. The mean distance of all matched pairs is compared to the expected mean under null hypothesis of random assignment given by $n_o(1 - 1/k)$.

7) Statistical Inference: A z-score statistical test quantifies the deviation of the observed mean of mismatches from its null expectation along with a binomial-based standard error.

8) Sequence Generation and Controls: The RCM methodology was evaluated both with real data and with data generated artificially. A real data sequence was generated with a number of pseudo-random generators as Linear Feedback Shift Register (LFSR), Mersenne Twister, NumPy's PCG64, os.urandom, and secrets. A patterned deterministic generator was used as a positive control while Knuth's MMIX linear congruential generator was used as a classic reference to measure model stability to a wider variety of signals [16]. Knuth's was utilized to introduce a reference point for the rest of the generators, which then provided the basis for the assessment of the models stability when dealing with a wide variety of signals.

The real-world data consisted of cycling power measurements in watts, which were initially stored in a CSV file. The data originated from a Flexible and Interoperable Data File (.fit), which is a binary format used by Garmin devices that are used for performance metrics. The power data was converted into MVL signals and the continuous power was discretized into 7 levels of k, each corresponding to intensity zones that are the typical standard for analyzing cycling power output [5], as seen previously in Fig 1. The representative sample mode was used, with each row in the CSV file providing the treatment positions (power at time t) and the outcome positions (power at time t) forming the discrete blocks used for the causal analysis. This analysis evaluates causal dependence within the MVL-discretized signal structure, but does not imply physiological causation of power output.

9) Evaluation and Testing: The evaluation metrics consisted of the mean Hamming distance between the matched outcome vectors, the expected null value under the random assignment, the z-score, and the associated p-value. The methodology was applied to both random and real-world signals via a programmed command-line interface.

B. Pearl's Causal Model

The Pearl causal model adapts many features of the Rubin model's methodology and also extends the multi-valued signal representation. It preserves the same discretizations structure and block decompositions. However, it is also distinguished by the representation of causal dependence via state-transition graphs and the normalization of outcomes by treatment groups.

1) Multiple-Valued Signal Representation: Following in the same fashion as with the Rubin Model, each signal is represented as a sequence over an alphabet of size k and every position takes on a discrete value from $0, 1 \ldots, k-1$. Data that is continuous is discretized into k bins of equal width, which preserves the temporal order and relative magnitude, which is consistent with MVL representation and enables the direct comparison of causal relationships between the two frameworks.

2) Signal Decomposition: As done in the Rubin model, a multi-valued signal of length N is sub-sequenced into blocks of size n. Each block is then divided into their designated treatment and outcome positions. The variables are discretized into k bins, which ensures every block entry has been encoded as an integer in $0, 1..., k-1$. The encoding permits uniform state-transition graph building to be performed for the outcome analysis.

3) Discretization and Covariate Encoding: Our PSM extension implements the same discretization conventions that were used in the RCM by representing all signals as MVL symbols in $\{0, 1, \ldots, k-1\}$. The same data is used, however, no additional covariate encoding was required because the methodology relies only on treatment and outcome positions within each block and the number of states defined by the state-bit length. This allows for full compatibility within the graph-construction and normalization steps. Any missing values in the original sequence are also assigned to default MVL bins to preserve the uniformity of the blocks.

4) Outcome Normalization by Treatment Group: Different treatment positions within each block are considered different groups. The data for each group that is defined by a unique treatment tuple has the outcomes normalized in regard to each groups anchor treatment value. Each result is then substituted with $(y - t_{\text{anchor}}) \bmod k$. where t_{anchor} is the treatment value for that group. The normalization in this instance highlights the differences in the attributes of the treatment to their absolute outcome values. This operation also aligns with Pearl's concept of intervention, where the effects of treatment variables on results are isolated by keeping the structural mechanism constant [10].

5) State-Transition Graph Construction: The order of normalized outcomes for every treatment group is employed

to each state-transition directed graph. States are defined as state bits of size m over the normalized outcome sequence, and edges are transitions from one state to the following. The resulting graph is a treatment temporal or structural dependency representation.

6) Jaccard Similarity Evaluation: To know the causal effect between the treatment grouped state transition graphs, the group comparisons were first obtained. In these comparisons, two types of Jaccard similarities were applied. The first type measures the absolute intersection type of edge sets over the edge sets of the treatment groups and is defined as $J_{\text{all}} = |\cap_{g=1}^{G} E_g| / |\cup_{g=1}^{G} E_g|$, which gives the strict intersection of all edge sets. The second type measures the average similarity over the complete pairs of groups and is given by $J_{\text{pair}} = \frac{1}{\binom{G}{2}} \sum_{i<j} |E_i \cap E_j| / |E_i \cup E_j|$, which represents the average similarity across all treatment group pairs. Both metrics determine how much transition structure is uniformly independent from a set of MVL sequences.

Jaccard similarity was chosen because it can be easily interpreted in the case of sparse transition graph comparisons. Other metrics, such as graph edit distance or spectral similarity, need weighted normalization or a probabilistic model, whereas Jaccard simply reflects the degree to which the transition structures in different treatment groups overlap.

The pairwise measure here is more sensitive since complete agreement across groups can be restrictive in multi group settings. In our analysis, the pairwise measure is the primary measure of the degrees of randomness, since partial agreement can reveal causal relationships even when complete overlap is absent [17]. Causal relationships can be revealed from the transition graphs, hence the presence of state transition structure can suggest the presence of causal dependence. The strict all groups measure provides a complementary perspective by capturing absolute agreement in scenarios where it exists.

The collection of metrics is in alignment with the practices in comparative causal network analysis [18]. In this line of analysis, similarity indices capture the shared edges of the graphs. In the case of multi-group scenarios where full structural overlap is unlikely, pairwise averaging tends to work particularly well, which is supported by findings in community detection [19] and other varieties of network comparison [20], [21]. Hence, this dual metric design heightens sensitivity to the presence of a partial causal structure in the elastic MVL systems.

7) Statistical Inference: The Jaccard similarity score is evaluated against the real-world signal data, that were read in from CSV files and synthetic signals that were generated by the aforementioned different random and pseudo-random bit generators. In both instances, the observed value is then compared against a null distribution devised using simulations with the same organization. A z-score and an empirical p-value are also computed to measure how distant the observed Jaccard similarity is from the null expectation.

8) Sequence Generation and Controls: Control sequences synthetically generated are derived from random and pseudo-random number bit generators (Mersenne Twister, PCG64, `os.urandom`, `secrets`, LFSR, and MMIX LCG), as used in the Rubin model. Control pattern-dependent sequences once again also functioned as a positive control. The Pearl causal test can thus be used to determine the distinction between independent randomness and structured causal dependencies.

9) Evaluation and Testing: The testing process compares the Jaccard similarity score calculated on actual signal data to that calculated on synthetic control sequences. Both cases are tested with the same configuration of subsequence length, treatment and outcome positions, and state bit size. The empirical results include a calculated z-score and p-value that quantify to what degree the observed Jaccard similarity departs from the null expectation under randomness. This permits testing of whether the observed structure represents a statistically significant causal dependency rather than random fluctuation.

C. Statistical Framework

If the null hypothesis is true and the treatment positions do not affect the outcome positions, the MVL results behave as independent uniform symbols. Under this assumption, Rubin's test expects an average mismatch rate $P(Y \neq Y') = 1 - \frac{1}{k}$ and an expected Hamming distance $\mathbb{E}[D_H] = m(1 - \frac{1}{k})$, and evaluates the observed mismatch proportion $\hat{p} = D_H/m$ using a normal approximation with variance $\text{Var}(\hat{p}) = ((1 - 1/k)(1/k))/(mn)$.

For Pearl's model, independent MVL sequences produce state-transition graphs with expected Jaccard similarity $\mathbb{E}[J] \approx 1/k^M$, which approaches zero as either the alphabet size k or state-bit length M increases. The observed statistics are compared against these null expectations using empirical distributions generated from MT19937, PCG64, `os.urandom`, `secrets`, LFSR, and Knuth's MMIX LCG. Patterned deterministic sequences serve as positive controls, confirming that both methodologies correctly detect non-random MVL structure.

IV. Results

We first investigated whether causal effects in random n-ary sequences produced by classical generators were detected. We tested sequences from a Linear Feedback Shift Register (LFSR), Mersenne Twister, NumPy PCG64, `os.urandom`, `secrets`, and Knuth's MMIX linear congruential generator, together with a patterned deterministic sequence as a positive control. The real dataset consisted of discretized cycling power (watts) from a CSV file.

A. Rubin's Causal Model Results

When applied to the discretized power signal using Functional Threshold Power (FTP)-based seven-level quantization ($k = 7$), the Rubin MVL model detected a statistically significant treatment–outcome relationship. Our model also evaluated the effect of treatment power signals (rows 42–2042) and outcome signals (rows 3042–6042). A total of 2001 paired samples were evaluated with 7-zone (FTP-based) binning ($k = 7$). For each paired sample, Hamming distances

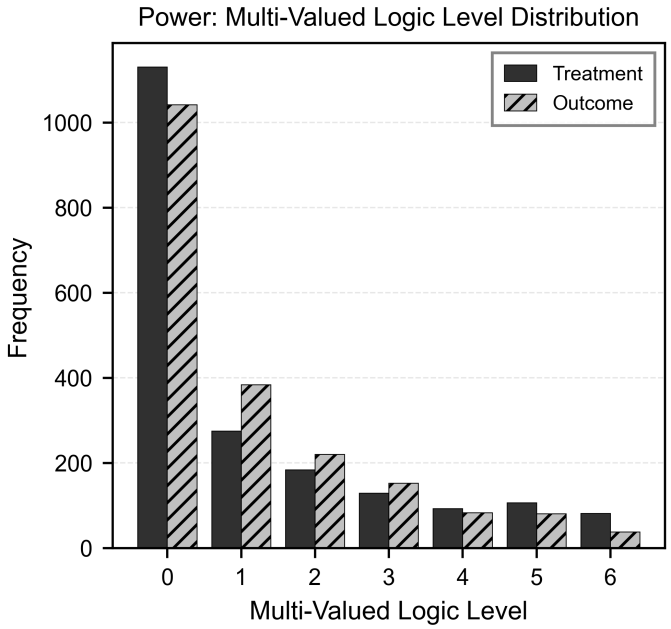

Fig. 2. Distribution of multi-valued logic levels in treatment and outcome segments for the discretized cycling power signal.

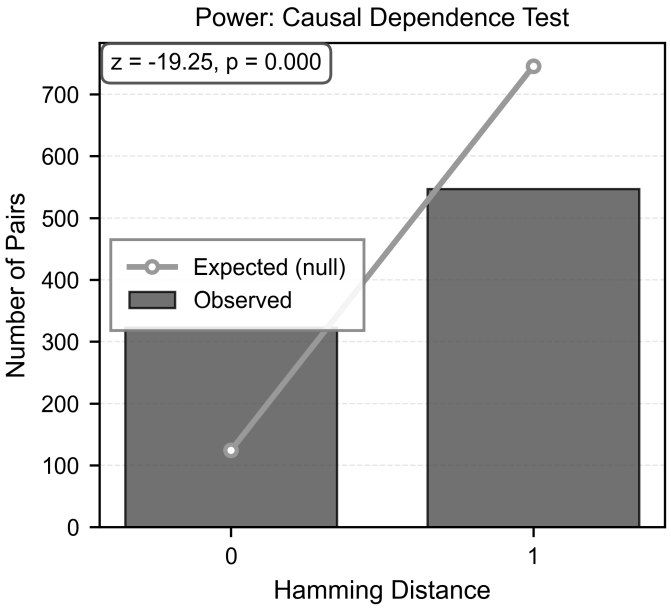

Fig. 3. Observed vs. null Hamming distance distribution for the Rubin MVL causal inference test.

were computed, and the average distance was 0.629. This average distance was significantly lower than what the null hypothesis predicted (i.e., there is no treatment and outcome segment dependence): $z = -19.25$, $p < 10^{-6}$. Therefore, this indicates the presence of causal dependence. The distribution of logic levels and the observed vs. null Hamming distances for the causal inference test are shown in Figs. 2 and 3.

The detected dependence reflects structured persistence in exertion state transitions. Treatment intervals condition subsequent outcome distributions, indicating non-random temporal progression within discretized physiologic workload states.

To test for robustness, we ran different tests of power zone bins and of treatment/outcome time windows to check how the causal inference test would still be valid under these different changes. In Table I, the Z-scores for four representative runs are reported: three runs using 7-zone (FTP-based) binning at different watt thresholds and one run using 3-level equal-width bins. The causal dependence that was observed in all these cases was very significant ($p < 0.001$).

TABLE I
RUBIN MVL: CAUSAL DEPENDENCE TESTS

Run	Time Windows	Discretization	Z-score
1	42–1442 vs 2042–3442	7-zone (100W)	-8.50
2	442–1642 vs 1742–2642	7-zone (125W)	-8.31
3	42–2042 vs 3042–6042	7-zone (175W)	-19.25
4	42–3042 vs 4042–6042	3-level bins	-11.74

All p-values < 0.001

In order to maintain reproducibility, a fixed seed, 8675309, was used in all RNG tests. We ran Rubin's MVL causal dependence test on 7-level ($k = 7$) signals of length $N = 100{,}000$ for various random, pseudo-random, and patterned controls. The treatment positions were [2, 8, 10]; outcome positions were [5, 11, 13]; and subsequence size was 16. These findings are summarized in Table II and visualized in Fig. 4.

TABLE II
RUBIN MVL: RNG RANDOMNESS COMPARISON

Generator	Avg Dist	Z-score	P-value	Interp.
MT	2.643	0.624	0.533	Random
PCG64	2.438	-1.250	0.211	Random
urandom	2.586	0.131	0.896	Random
secrets	2.774	1.863	0.063	Random
LFSR	2.549	-0.499	0.618	Random
MMIX LCG	2.500	-0.667	0.505	Random
Patterned	3.000	21.131	<0.001	Non-random

(Seed = 8675309, $k = 7$, subseqsize=16).

B. Pearl's Causal Model Results

The same treatment, outcome positions, and subsequence length were used for the Pearl model, which evaluates causality via differences in state-transition graphs. We again performed multiple tests using different treatment/outcome windows; the results for each run are summarized in Table III. The real-world power data again showed a significant causal effect, as shown in Table III.

TABLE III
PEARL MVL: CAUSAL DEPENDENCE TEST

Run	Time Windows	Discretization	Z-score
1	42–1442 vs 2042–3442	7-zone (100W)	-32.32
2	442–1642 vs 1742–2642	7-zone (125W)	-21.571
3	42–2042 vs 3042–6042	7-zone (175W)	-10.19
4	42–3042 vs 4042–6042	3-level bins	0.004

All p-values < 0.01

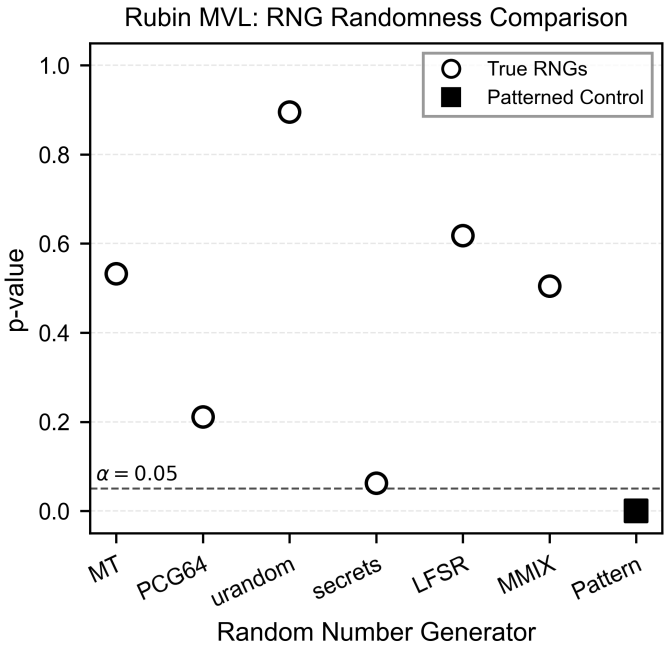

Fig. 4. Rubin MVL: p-values for the causal dependence test applied to various random number generators ($k = 7$, subseqsize=16, seed=8675309). True RNGs yield high p-values, whereas the patterned control is detected as non-random.

The Jaccard similarity for treatment-normalized outcome graphs was 0.119 ($z = -32.32$, $p = 0.0099$). The state-transition structure for the power data is visualized in Fig. 5.

As shown in Table IV all random and pseudo-random controls yielded $J = 0.000$ and $p = 1.000$ (no structure), while the patterned control produced $J = 1.000$ and $p = 0.010$.

TABLE IV
PEARL MVL: RNG RANDOMNESS COMPARISON

Generator	Jaccard	Avg Edges	Z-score	P-val	Interp.
MT	0.0757	48.85	1.4320	0.139	Random
PCG64	0.0758	48.81	1.877	0.045	Non-random
urandom	0.0755	48.68	0.068	0.452	Random
secrets	0.0756	48.69	0.934	0.390	Random
LFSR	0.0096	3.14	-249.539	0.0099	Non-random
MMIX LCG	0.0752	48.53	-0.359	0.713	Random
Patterned	1.000	1.000	3509.47	0.0099	Non-random

(Seed = 8675309, $k = 7$, subseqsize=16).

These results are illustrated in Fig. 6, where the edge-count distributions for each generator reveal that random and pseudo-random signals form sparse, diffuse graphs, whereas the patterned control produces dense, structured transitions characteristic of causal dependency.

V. CONCLUSIONS & FUTURE WORKS

A. Conclusions

This work adapts two foundational causal inference frameworks from statistical analysis to the domain of multiple-valued logical signals. The Rubin and Pearl causal models, traditionally applied in observational studies and epidemiology, have been extended to detect causal relationships in

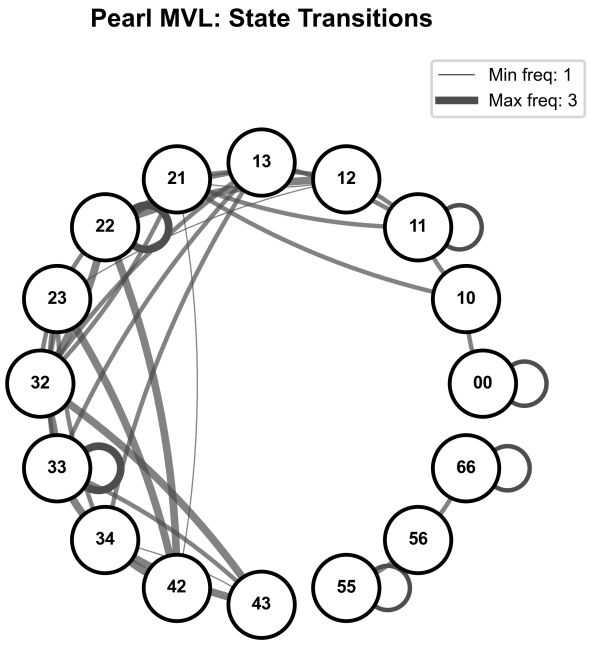

Fig. 5. Pearl MVL: State-transition graph for discretized cycling power data (FTP-based 7 zones, 2 state bits). Edge thickness reflects transition frequency.

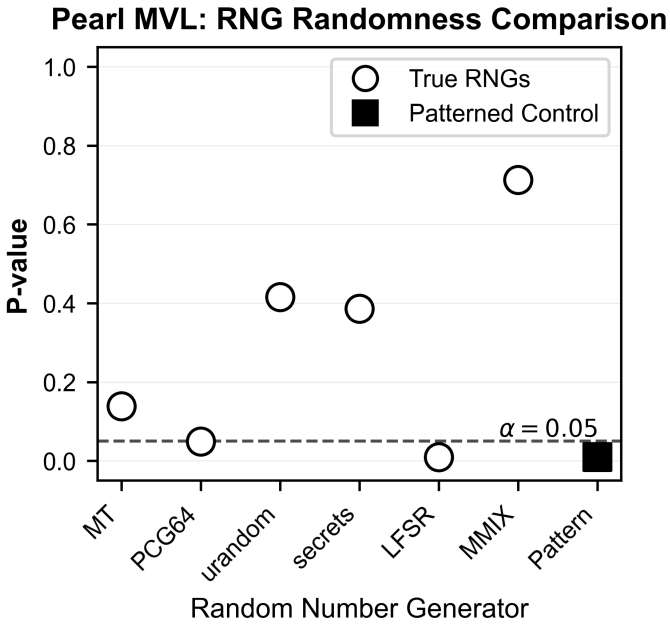

Fig. 6. Pearl MVL: p-values for the causal dependence test applied to various random number generators ($k = 7$, subseqsize=16, seed=8675309, state bits=2). Patterned controls are detected as non-random.

multi-valued signals. The methodology performed in this study demonstrates that both propensity score matching and graph-based causal inference can be effectively applied beyond their original functional contexts.

The results from the empirical testing have validated the approach across both synthetic and real-world data. Using propensity score matching and Hamming distance comparisons the Rubin model was able to successfully distinguish between truly random sequences and those embedded with causal structure. When applied to discretized cycling power data, the model detected a statistically significant treatment-outcome relationship, which indicated non-random causal dependence between power interval signals. The Pearl model, which employed state-transition graphs and Jaccard similarity metrics to identify the causal patterns, also correctly classified random and pseudo-random bit generators having random or non-random sequences. The real-world cycling power data was detected as having meaningful causal structure. Both methodologies proved robust across different signal characteristics and alphabet sizes.

Both methodologies proved to be quite robust across different signal characteristics and can be extended to different alphabet sizes and multi-valued systems. The corresponding nature of both models allows for flexibility in the analysis of causal inference in signal analysis. The Rubin model focuses on covariate-matched outcome comparisons and the Pearl examines structural transition patterns. Both of these techniques have broad applicability in domains where discrete temporal sequences may arise. Applications that could benefit from these methodologies can include digital systems analysis, physiological monitoring, and performance analytics. In physiologic monitoring contexts, discretized signals such as heart rate zones or glucose bands may be analyzed using MVL causal signal frameworks to detect structured treatment–response dependencies.

B. Future Works

There are several promising directions that can emerge from this work, which warrant further investigations.

1) Multi-Signal Causal Analysis: The most natural extension would involve the analysis of causal relationships across multiple concurrent signals. For example, in cycling, one could examine the joint causal effects of heart rate and power interval output signals, rather than treating them as isolated sequences.

2) Temporal Lag Analysis: The current methodology assumes simultaneous treatment–outcome relationship within blocks. In future work, there needs to be an incorporation of temporal lag structures. This would involve allowing the treatment positions at time t to influence the outcomes at $t + \Delta t$. Such a functionality would be relevant in the analysis of delayed physiological responses or any cascading effects in complex systems. For example, this could be applied to caloric burn effects after long cycling rides.

REFERENCES

[1] J. Pearl and D. Mackenzie, *The book of why: the new science of cause and effect*. Basic books, 2018.

[2] H. Charvátová, A. Procházka, and O. Vyšata, "Motion assessment for accelerometric and heart rate cycling data analysis," *Sensors*, vol. 20, no. 5, p. 1523, 2020.

[3] H. Charvátová, A. Procházka, S. Vaseghi, O. Vyšata, and M. Vališ, "Gps-based analysis of physical activities using positioning and heart rate cycling data," *Signal, Image and Video Processing*, vol. 11, no. 2, pp. 251–258, 2017.

[4] H. Allen and A. Coggan, *Training and Racing with a Power Meter*. Boulder, CO, USA: VeloPress, 2 ed., 2012.

[5] A. Coggan, "Power Training Levels.." https://www.trainingpeaks.com/blog/power-training-levels/, 2019. Describes the 7-zone power training model based on physiological responses.

[6] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[7] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[8] G. W. Imbens and D. B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. USA: Cambridge University Press, 2015.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[10] J. Pearl, *Causality*. Cambridge University Press, 2 ed., 2009.

[11] J. Pearl, "Causal diagrams for empirical research," *Biometrika*, vol. 82, no. 4, pp. 669–688, 1995.

[12] I. Shpitser and J. Pearl, "Identification of joint interventional distributions in recursive semi-markovian causal models," in *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI)*, pp. 1219–1226, 2006.

[13] K. Hirano and G. W. Imbens, "The propensity score with continuous treatments," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (W. A. Shewhart, S. S. Wilks, A. Gelman, and X.-L. Meng, eds.), Wiley, 2004.

[14] G. W. Imbens, "The role of the propensity score in estimating dose–response functions," *Biometrika*, vol. 87, no. 3, pp. 706–710, 2000.

[15] K. Imai and D. A. van Dyk, "Causal inference with general treatment regimes: Generalizing the propensity score," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 854–866, 2004.

[16] D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Reading, MA, USA: Addison-Wesley, 3 ed., 1997.

[17] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002.

[18] A. Defilippo, F. M. Giorgi, P. Veltri, *et al.*, "Understanding complex systems through differential causal networks," *Scientific Reports*, vol. 14, p. 27431, 2024.

[19] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, no. 3, p. 033015, 2009.

[20] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, 06 2009.

[21] S. Shvydun, "Models of similarity in complex networks," *PeerJ Computer Science*, vol. 9, p. e1371, 2023.