

Article

Dependence-Aware Lag-Resolved Correlation Analysis in Multi-Sensor Stochastic Systems

Andrew Graham Ryan 

Independent Researcher, Warrington WA5 8TA, UK; andrewryan63@gmail.com

Abstract

Exploratory lag-resolved correlation analysis is widely used when potential dependence between concurrently recorded signals may be delayed, transient, or weak. In autocorrelated time series, however, scanning across candidate lags creates a multiple-comparison problem, while temporal dependence invalidates naive permutation or parametric correlation tests. This paper presents a practical dependence-aware scan-level inference procedure for exploratory lag-resolved correlation analysis under temporal autocorrelation. The procedure combines baseline standardisation, lag-resolved Pearson correlation, dependence-preserving surrogate construction (here implemented using block permutation), and max-statistic correction so that inference is performed on the largest absolute correlation observed across the scanned lag domain rather than on post hoc selected lags. The contribution is integrative rather than metric-driven: established components are assembled into a fixed, auditable inference pipeline that preserves within-channel temporal structure while controlling familywise error across exploratory lag scans. The procedure is intended for multi-sensor stochastic systems in which weak synchrony must be distinguished from artefacts of temporal dependence and analytical flexibility. Synthetic null simulations illustrate false-positive inflation under naive lag scanning, characterise calibration through a block length sensitivity analysis, and compare complete inference pipelines, demonstrating that, under the dependence regimes examined here, valid exploratory lag inference requires scan-level multiplicity control in addition to dependence-preserving surrogate generation.

Keywords: lag-resolved correlation; permutation testing; time series analysis; multi-sensor systems; dependence-aware inference; multiple comparisons

1. Introduction

Lag-resolved correlation analysis is a routine tool in multi-sensor time series work [1,2]. When signals are recorded concurrently but potential coupling is delayed, intermittent, or weak, scanning across temporal offsets is often the natural first step. This situation arises across a wide range of applications: distributed sensor arrays, physiological recordings, environmental monitoring, and financial time series, where contemporaneous correlation is rarely sufficient.

At first glance, the task appears straightforward: compute correlations across a range of lags and inspect the result. In practice, it is not. Once temporal dependence and baseline drift enter the picture, the statistical assumptions underpinning standard correlation tests can fail. Autocorrelation reduces the effective number of independent observations; lag scanning multiplies the number of implicit tests. Taken together, these effects can inflate false-positive rates in ways that are easy to overlook and difficult to correct after the fact [3].



Academic Editor: Frank Werner

Received: 21 May 2026

Revised: 9 July 2026

Accepted: 19 July 2026

Published: 22 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

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

A common response is to employ permutation-based inference. Properly constructed, permutation tests avoid distributional assumptions and can adapt naturally to complex data. However, naive permutation—simple shuffling or random reordering—destroys precisely the dependence structure that defines the problem. The resulting null distributions may be tractable, but they are not appropriate for autocorrelated signals. More careful surrogate constructions exist, including block-based resampling [4,5] and frequency domain methods [6], yet these approaches are often applied to fixed hypotheses or predefined lags. They do not, by themselves, address the additional multiplicity introduced when the analyst searches across an exploratory lag domain.

The present paper addresses this practical inferential gap by specifying a dependence-aware scan-level inference procedure for exploratory lag-resolved correlation analysis under temporal autocorrelation. The procedure integrates baseline standardisation, lag-resolved Pearson correlation, dependence-preserving block permutation surrogates, and max-statistic correction within a fixed analysis pipeline. Its contribution is integrative rather than metric-driven: it does not introduce a new dependence measure, but assembles established statistical components into an auditable procedure that treats lag selection as part of the inferential problem rather than as an informal preprocessing step.

The aim is deliberately modest. The procedure is intended to constrain analytical flexibility in settings where multiple baselines, lag ranges, dependence structures, or null constructions might otherwise be explored post hoc. By fixing the baseline window, lag domain, surrogate construction, block length, number of permutations, and decision rule before inference, the analysis produces corrected pairwise evidence for lag-resolved synchrony under a specified dependence-preserving null model. The resulting detections are interpreted operationally: they indicate structure exceeding null expectations under the chosen procedure, not causality, directionality, or mechanism.

The remainder of the paper is organised as follows. Section 2 reviews related work on dependence-preserving surrogates, multiple testing correction, and applied lag-resolved analysis. Section 3 introduces the core components of the proposed procedure and defines the deviation metrics and lag-resolved statistics. Section 4 describes the construction of dependence-preserving permutation null models and the associated max-statistic correction procedure. Section 5 presents the fixed algorithmic pipeline and implementation details, including a formal statement of the procedure. Section 6 presents synthetic null simulations demonstrating false-positive inflation under naive lag scanning, examines the calibration properties of the proposed procedure through a block length sensitivity analysis, and compares the proposed inference framework with established dependence-preserving surrogate pipelines. Limitations and scope are discussed in Section 7, followed by concluding remarks in Section 8.

2. Related Work

The present procedure sits at the intersection of three established areas of statistical practice: dependence-preserving surrogate construction for autocorrelated signals, multiple testing correction for exploratory inference, and lag-resolved cross-correlation analysis in applied multi-sensor systems. Each of these areas is well-developed individually. However, published work typically emphasises either dependence-preserving surrogate construction for fixed hypotheses or multiplicity correction within other search domains, while comparatively little attention has been given to integrating both within a pre-specified inference procedure for exploratory lag-resolved scanning. The present work addresses this practical methodological integration.

2.1. Dependence-Preserving Surrogates and Block Resampling

Resampling procedures designed for temporally dependent observations have long been used to avoid the unrealistic independence assumptions underlying naive permutation tests. Early work on block bootstrap methods demonstrated that resampling contiguous segments of a time series can preserve short-range dependence structure while supporting nonparametric inference under weak stationarity assumptions [4,5]. Related approaches, including the stationary bootstrap of Politis and Romano [7], extended this logic by introducing random block lengths to improve asymptotic behaviour under certain dependence conditions.

Frequency domain surrogate procedures have also been widely used in nonlinear time series analysis. Theiler et al. [6] introduced phase-randomised surrogates that preserve the power spectrum of the original process while disrupting temporal alignment structure. Variants and extensions of these methods continue to be used in neuroscience, climate analysis, and physiological signal processing [8,9].

More recently, Sheng and Shou [10] introduced the truncated time-shift test, a non-parametric procedure for testing dependence between stationary autocorrelated time series. Their work illustrates continued methodological activity in dependence-preserving testing for correlated signals. The present procedure addresses a complementary problem: controlling familywise error across an exploratory lag domain rather than testing dependence at a single fixed alignment.

These methods address an important inferential problem: preserving within-series temporal structure under null hypothesis generation. However, they are most commonly applied to fixed hypotheses or predetermined lag structures. When exploratory lag scanning is performed across a broad lag domain, dependence-preserving surrogates alone do not address the multiplicity inflation arising from repeated lag-wise comparisons.

2.2. Multiple Testing and Scan-Level Inference

The problem of multiplicity correction under exploratory analysis has likewise received extensive attention. Classical familywise error rate (FWER) procedures based on resampling were formalised by Westfall and Young [2], while max-statistic approaches became particularly influential in neuroimaging applications involving large-scale correlated testing [1]. In these settings, inference is based on the extremal statistic observed across a search domain rather than on individual post hoc selected comparisons.

In practical lag-resolved correlation analysis, it remains common to evaluate correlations across multiple candidate lags without formally incorporating the lag-selection process into the inferential correction procedure. Many standard implementations either assume approximate independence, rely on parametric dependence models, or focus on fixed hypotheses rather than exploratory lag-domain searches.

The conditions under which naive permutation tests fail to control error under temporal dependence have recently been formalised. Romano and Tirlea [11] show that standard permutation tests are generally not asymptotically level- α when the null hypothesis concerns zero autocorrelation rather than independence, describing a failure mechanism complementary to the naive parametric threshold inflation demonstrated in Section 6.2.

The present work adopts the max-statistic principle specifically because lag scanning itself constitutes an exploratory search process. Rather than treating the selected lag as fixed after inspection, the procedure performs inference on the largest absolute correlation observed anywhere within the scanned lag domain under a dependence-preserving null model.

2.3. Lag-Resolved Correlation in Applied Multi-Sensor Systems

Lag-resolved correlation analysis is routinely used in applied multi-sensor settings where coupling may be delayed, transient, or weak. Examples include functional connectivity analysis in neuroscience [1], environmental and climate monitoring [12], and distributed sensing applications involving autocorrelated multichannel observations. In such applications, exploratory lag scanning is often performed operationally to identify candidate temporal alignments between channels.

Despite widespread use, inferential treatment varies substantially across domains. Some analyses rely on uncorrected lag-wise correlations; others apply parametric thresholds without accounting for temporal dependence; still others employ surrogate procedures without formal multiplicity correction across the lag domain. As a result, apparent synchrony can reflect a combination of autocorrelation, baseline drift, and exploratory flexibility rather than reproducible cross-channel structure [3].

Recent applied work confirms that this problem remains practically significant in modern autocorrelated time series settings. Lun et al. [13] demonstrate that autocorrelation in environmental time series can substantially inflate the apparent significance of cross-correlations under standard testing procedures, motivating dependence-aware inference strategies for lag-resolved analysis.

The present procedure addresses this gap by integrating dependence-preserving block permutation surrogates with max-statistic correction in a single, fixed inference pipeline specifically designed for exploratory lag scanning.

3. Procedure Overview

The procedure introduced here is designed to detect weak, lag-resolved correlation structure between multiple stochastic signals while explicitly accounting for temporal dependence and exploratory lag selection. It is formulated as a practical scan-level inference procedure rather than a domain-specific model. The emphasis is on how statistical decisions are constrained—what is fixed *a priori*, what is allowed to vary, and how inference is ultimately performed.

At a high level, the procedure comprises four components: baseline standardisation of individual channels; construction of deviation metrics; evaluation of lag-resolved correlation statistics; and statistical inference using dependence-preserving permutation null models with max-statistic correction. Each component is specified explicitly, and none relies on assumptions about the physical origin of the underlying signals.

The procedure operates on discretely sampled time series $X_i(t_k)$, where $i = 1, \dots, N$ indexes the channel or sensor and $k = 1, \dots, T$ indexes discrete sampling positions. The quantity t_k denotes the time associated with sample index k . Throughout, signals are treated as stochastic processes characterised empirically by their observed dependence structure.

3.1. Baseline Standardisation

To facilitate comparison across channels and to remove static offsets and standardise signal scale, each time series is first transformed into a standardised deviation signal. A baseline distribution is estimated for each channel using a predefined reference interval (baseline window) W_i or calibration period. The choice of W_i is fixed *a priori* and is kept constant throughout the analysis.

For each channel i , the baseline mean μ_i and standard deviation σ_i are estimated empirically. The raw signal $X_i(t_k)$ is then mapped to a standardised deviation signal $E_i(t_k)$ via

$$E_i(t_k) = \frac{X_i(t_k) - \mu_i}{\sigma_i}. \quad (1)$$

This transformation yields dimensionless signals with comparable scale, allowing subsequent cross-channel correlation analysis without introducing channel-specific weighting.

3.2. Lag-Resolved Correlation Statistic

Lag-resolved correlation is evaluated between pairs of standardised deviation signals $E_i(t_k)$ and $E_j(t_{k+\tau})$, where τ denotes an integer lag expressed in sampling units. For each channel pair (i, j) , correlation statistics are computed across a predefined symmetric lag range $\tau \in [-\tau_{\max}, \tau_{\max}]$.

In the present implementation, Pearson correlation is used as the dependence metric, yielding

$$C_{ij}(\tau) = \text{corr}_k(E_i(t_k), E_j(t_{k+\tau})), \quad (2)$$

where $\text{corr}_k(\cdot, \cdot)$ denotes the sample Pearson correlation evaluated over the overlapping sample indices k for which both $E_i(t_k)$ and $E_j(t_{k+\tau})$ are defined. Pearson correlation is therefore computed using the $T - |\tau|$ paired observations available at lag τ . The procedure itself does not depend on this specific choice. Alternative dependence measures may in principle be substituted within the same inferential framework, although such extensions are not developed or evaluated in the present study.

3.3. Exploratory Lag Scanning and Test Statistic

Exploratory lag scanning introduces an implicit multiple-comparison problem, as correlation statistics are evaluated across many candidate lag values. To summarise lag-resolved structure for inference, the procedure records the maximum absolute correlation observed across the scanned lag range,

$$r_{\max,ij} = \max_{\tau} |C_{ij}(\tau)|. \quad (3)$$

The associated lag τ^* at which this maximum occurs is also recorded. The statistic $r_{\max,ij}$ serves as the primary test statistic for assessing whether observed cross-channel structure exceeds expectations under appropriate null models. The lag τ^* is treated strictly as a descriptive property of the detected structure rather than as evidence of causality, directionality, or temporal precedence.

3.4. Inference Strategy Overview

Statistical inference is performed by comparing the observed test statistic $r_{\max,ij}$ to a null distribution generated via dependence-preserving permutation procedures. These null models are constructed to maintain key empirical properties of the individual observed signals, especially their marginal distributions and within-channel autocorrelation structure, while disrupting cross-channel alignment so that observed synchrony remains the object of inference. Max-statistic correction is applied across lag scans to control familywise error under exploratory testing. The detailed construction of permutation null models and the associated correction procedures are described in Section 4.

4. Dependence-Preserving Permutation Null Models

Permutation-based inference provides a natural alternative to parametric testing when distributional assumptions are unreliable or difficult to justify [14]. In the context of lag-resolved correlation analysis, however, permutation procedures must be constructed with particular care. Naive resampling schemes—such as random shuffling of observations—destroy temporal dependence and yield null distributions that substantially underestimate the probability of observing large correlation values under realistic noise conditions.

The present procedure therefore employs permutation null models that explicitly preserve the empirical dependence structure of the observed signals. The goal is not to approximate an idealised notion of independence, but rather to generate surrogate data that reflect what would be expected in the absence of cross-channel coupling, *conditional on the observed temporal structure*.

4.1. Principles of Dependence Preservation

Let $\{E_i(t_k)\}$ denote the standardised deviation signal for channel i . Each $E_i(t_k)$ typically exhibits temporal autocorrelation, nonstationarity, and finite-sample structure that cannot be ignored without distorting inference. A valid permutation scheme must therefore satisfy two constraints:

1. The marginal distribution of each channel must be preserved.
2. The temporal dependence structure within each channel must be preserved to a degree sufficient to maintain realistic autocorrelation properties.

Importantly, preservation of dependence is required only *within* channels. Any synchrony or alignment *between* channels is treated as the object of inference and must be disrupted by the permutation procedure.

The procedure therefore assumes that the deviation signals are approximately weakly stationary over the analysis interval after baseline standardisation, so that within-channel dependence may be represented meaningfully by the chosen surrogate construction.

4.2. Block-Based Permutation Strategy

In the present implementation, dependence preservation is achieved using a block-based permutation scheme. Each deviation signal $E_i(t_k)$ is partitioned into contiguous blocks of fixed length L , chosen to substantially exceed the dominant autocorrelation scale of the signal. The blocks are then permuted randomly, while the internal structure of each block is left unchanged.

In practice, the block length L should be selected with reference to the dominant autocorrelation scale of the signal. A useful starting point is the smallest autocorrelation lag ℓ at which the empirical autocorrelation function $\rho(\ell)$ decays below a modest threshold (for example, $|\rho(\ell)| < 0.2$). This should be regarded only as an initial reference scale rather than a sufficient choice of block length. As demonstrated empirically in Section 6.3, satisfactory calibration under strong temporal dependence may require block lengths that substantially exceed the dominant autocorrelation scale. In the AR(1) simulations examined here, reliable calibration was achieved only once L substantially exceeded that scale. The objective is therefore not to estimate an optimal block length, but to select a conservatively large value that adequately preserves within-channel dependence while retaining a sufficient number of permutable blocks for reliable null calibration.

In multichannel applications, different channels may exhibit different autocorrelation scales. The empirical findings of Section 6.3 motivate a conservative practical approach to block length selection. Where inference is performed independently for each channel pair, the selected block length may be referenced to the more slowly decaying channel within that pair. Alternatively, where a common block length is preferred across an entire multichannel system, a single value referenced to the largest estimated autocorrelation scale provides a conservative choice that prioritises preservation of within-channel temporal dependence, although at the cost of reduced permutation diversity.

Although the optimal value of L remains data-dependent, the simulation results presented here indicate that underestimating block length poses a greater practical risk to calibration than moderate overestimation. Excessively large block lengths are not without cost, however, since they reduce the number of independently permutable blocks and

consequently the resolution of the empirical null distribution. Block length should therefore be chosen to balance adequate preservation of within-channel dependence against sufficient permutation diversity for reliable statistical inference.

A known limitation of fixed block permutation schemes is that boundary discontinuities can arise at block junctions, introducing artificial structure at the seams between concatenated blocks. In practice, this effect is modest when L substantially exceeds the dominant autocorrelation scale, since boundary artefacts remain localised relative to the preserved within-block dependence. Alternative surrogate strategies address this limitation in different ways: the stationary bootstrap [7] uses geometrically distributed block lengths to smooth boundary transitions, whereas phase randomisation surrogates [6,8] operate in the frequency domain and avoid block boundaries at the cost of assuming approximately linear Gaussian structure. The fixed block scheme is retained here for its operational simplicity, parameter transparency, and broad applicability to non-Gaussian signals.

Formally, let

$$B_{i,m} = \{E_i(t_{(m-1)L+1}), \dots, E_i(t_{mL})\}, \quad (4)$$

where $m = 1, \dots, \lfloor T/L \rfloor$ indexes the contiguous complete blocks of channel i . If T is not an exact multiple of L , the remaining observations form a final shorter block that is appended unchanged after permutation of the complete blocks, thereby preserving the temporal ordering of the remaining observations.

A permuted surrogate signal $\tilde{E}_i(t_k)$ is constructed by randomly reordering the set of complete blocks $\{B_{i,m}\}$ and concatenating them in the new order. The block length L is fixed *a priori* and applied consistently across all permutations and channel pairs. Sensitivity to the choice of L is examined empirically in Section 6.3.

4.3. Lag-Resolved Null Distribution Construction

For each permutation replicate $p = 1, \dots, P$, surrogate signals $\tilde{E}_i^{(p)}(t_k)$ are generated independently for each channel. The lag-resolved correlation statistic is then computed for each channel pair, yielding

$$\tilde{C}_{ij}^{(p)}(\tau) = \text{corr}_k(\tilde{E}_i^{(p)}(t_k), \tilde{E}_j^{(p)}(t_{k+\tau})), \quad (5)$$

where $\text{corr}_k(\cdot, \cdot)$ denotes the sample Pearson correlation evaluated over the overlapping sample indices for which both surrogate signals are defined.

As in Equation (2), correlations are computed using the $T - |\tau|$ paired observations available at lag τ . The normalisation is therefore applied consistently to both the observed and surrogate signals.

4.4. Maximum Statistic Null Distribution

For each permutation replicate, the maximum absolute lag-resolved correlation is extracted across the entire scanned lag range,

$$\tilde{r}_{\max,ij}^{(p)} = \max_{\tau} |\tilde{C}_{ij}^{(p)}(\tau)|. \quad (6)$$

The collection $\{\tilde{r}_{\max,ij}^{(1)}, \dots, \tilde{r}_{\max,ij}^{(P)}\}$ therefore forms the empirical null distribution of the maximum statistic under the null hypothesis of no between-channel dependence.

4.5. Max-Statistic Correction and Significance Assessment

Use of the maximum statistic directly implements scan-level multiplicity correction across the lag scan. Rather than testing each lag independently, inference is based on the probability of observing a correlation peak at least as large as the observed maximum

anywhere within the scanned lag range under the null model. The corrected p -value is computed as

$$p_{\text{corr},ij} = \frac{1 + \sum_{p=1}^P \mathbf{1}(\tilde{r}_{\text{max},ij}^{(p)} \geq r_{\text{max},ij})}{P + 1}, \quad (7)$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. This is the standard add-one Monte Carlo estimator for permutation p -values, ensuring a strictly positive corrected p -value following the standard Monte Carlo convention. Under the dependence-preserving surrogate construction described above, the resulting significance assessment accounts simultaneously for temporal dependence and the multiplicity introduced by exploratory lag scanning. Its empirical calibration is evaluated in Section 6.

4.6. Interpretation Under the Procedure

Rejection of the null hypothesis indicates that the observed lag-resolved correlation structure exceeds what is expected under dependence-preserving noise alone. The procedure deliberately refrains from attributing causal or mechanistic interpretation to the detected structure. The role of the null model is strictly operational: to define what constitutes an unexpected degree of synchrony given the empirical properties of the data.

The inferential framework described here may be extended to alternative surrogate schemes, including frequency domain phase randomisation or model-based resampling, provided that the resulting null distributions preserve temporal dependence and are paired with an appropriate scan-level max-statistic correction.

Table 1 summarises the principal inferential properties of commonly used approaches to lag-resolved correlation analysis and places the proposed framework within this methodological context. The parametric and surrogate-based rows of this schematic comparison are instantiated empirically in Section 6.4.

Table 1. Comparison of inference properties across common approaches to lag-resolved correlation analysis. The “Exploratory lag scanning” column indicates whether the approach is commonly applied in exploratory lag-scanning settings, not whether it provides valid scan-level inference. Entries for block bootstrap and surrogate methods reflect standard practice; these methods can in principle be extended to address multiplicity and exploratory scanning, but this integration is not their typical application.

Approach	Dependence Preserved	Multiplicity Controlled	Exploratory Lag Scanning
Parametric single-lag tests	×	×	×
Naive permutation testing	×	×	✓
Block bootstrap/surrogates	✓	Typically omitted	Typically omitted
Ad hoc lag selection	×	×	✓
Proposed procedure	✓	✓	✓

Note: ✓ indicates that the property is present; × indicates that it is absent. Bold type identifies the proposed procedure.

5. Algorithmic Pipeline

The procedure is implemented as a fixed, sequential pipeline designed to constrain analytical flexibility and minimise researcher degrees of freedom [3]. Each stage is specified *a priori*, and intermediate outputs are treated as inputs to subsequent steps without post hoc adjustment. Figure 1 provides a schematic overview of the full analysis procedure.

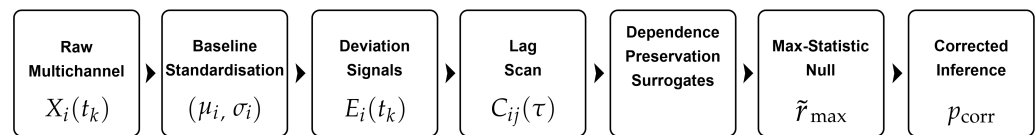


Figure 1. Schematic overview of the proposed dependence-aware lag-resolved inference pipeline. Raw multichannel signals are standardised using empirically calibrated baselines to form deviation signals $E_i(t_k)$. Lag-resolved correlations $C_{ij}(\tau)$ are evaluated over a scanned lag domain. Dependence-preserving surrogate signals are generated to construct an empirical max-statistic null distribution for the peak correlation magnitude, yielding a scan-level corrected p -value p_{corr} across lags. Empirical familywise error calibration is evaluated in Section 6.

5.1. Input Specification

The pipeline operates on a collection of discretely sampled time series $\{X_i(t_k)\}_{i=1}^N$ acquired concurrently from N sensors or channels. Signals are assumed to share a common sampling rate and time base. Missing data, if present, must be handled prior to analysis using a fixed preprocessing rule (e.g., interpolation or exclusion), as the procedure itself does not incorporate data imputation.

All analysis parameters—including baseline windows, lag range $[-\tau_{\max}, \tau_{\max}]$, block length L , and number of permutations P —are specified before any correlation statistics are computed.

5.2. Baseline Calibration and Deviation Transformation

Each input signal is mapped to a standardised deviation representation using the transformation defined in Equation (1). For channel i , a baseline interval is selected and used to estimate the baseline mean μ_i and standard deviation σ_i . These quantities define a fixed affine transformation applied uniformly to the full signal, normalising scale across channels and improving comparability by removing static offsets while preserving the temporal structure of fluctuations.

5.3. Lag-Resolved Correlation Evaluation

For each ordered channel pair (i, j) , lag-resolved correlation statistics are computed over the predefined lag range using the standardised deviation signals $E_i(t_k)$ and $E_j(t_{k+\tau})$. No smoothing or interpolation is applied across lags; each lag is treated as a discrete candidate alignment.

The output is a lag profile for each channel pair, together with the maximum absolute correlation $r_{\max,ij}$ and its associated lag τ^* .

5.4. Null Model Generation

Dependence-preserving permutation surrogates are generated independently for each channel using the block-based strategy described in Section 4. For each permutation replicate, the full lag-resolved correlation analysis is repeated, yielding a surrogate maximum statistic $\tilde{r}_{\max,ij}^{(p)}$.

This stage dominates computational cost; however, the procedure is embarrassingly parallel and can be accelerated using standard parallelisation strategies when large numbers of permutations are required.

For N channels, P permutations, series length T , and a scanned lag range of size $2\tau_{\max} + 1$, overall complexity scales as $\mathcal{O}(N^2 \cdot P \cdot \tau_{\max} \cdot T)$. In practice, $P \geq 1999$ supports a minimum attainable p -value of approximately 5×10^{-4} under the $(P + 1)$ convention.

5.5. Statistical Decision Rule

The observed test statistic $r_{\max,ij}$ is compared against the empirical null distribution $\{\tilde{r}_{\max,ij}^{(p)}\}$. Statistical significance is assessed using the max-statistic corrected p -value defined in Equation (7). No additional lag-wise correction is required, as multiplicity across the scanned lag domain is already incorporated into the max-statistic procedure.

The present procedure is designed to control familywise error across lags within an individual channel pair, with its empirical calibration assessed in Section 6. In analyses involving N channels, $N(N-1)/2$ pairwise comparisons may be performed, each yielding one corrected pairwise p -value. If global familywise error rate or false discovery rate control across all channel pairs is required, secondary correction procedures such as Bonferroni adjustment or Benjamini–Hochberg procedures may be applied to the set of corrected pairwise p -values.

The present paper addresses within-pair lag multiplicity specifically; cross-pair multiplicity control is a natural extension but lies outside the current scope.

If $p_{\text{corr},ij} < \alpha$, the result is classified as a detection of lag-resolved synchrony exceeding dependence-preserving null expectations.

5.6. Outputs and Reproducibility

The pipeline produces, for each channel pair, the observed lag-resolved correlation profile, the peak statistic $r_{\max,ij}$, the associated lag τ^* , and the corrected p -value p_{corr} .

Because all parameters are fixed in advance and the null model is defined operationally, the procedure is fully reproducible given the original data and random seeds used for permutation.

5.7. Formal Procedure Summary

Table 2 summarises the full inferential pipeline.

Table 2. Formal summary of the dependence-aware lag-resolved inference procedure.

Step	Procedure
1	Fix all analysis parameters ($\mathcal{W}_i, \tau_{\max}, L, P, \alpha$) before inspecting lag-resolved statistics.
2	Estimate baseline mean μ_i and standard deviation σ_i for each channel from the predefined baseline window $\mathcal{W}_i$.
3	Transform raw signals into standardised deviation signals using Equation (1).
4	Compute lag-resolved correlations $C_{ij}(\tau)$ for all channel pairs over $\tau \in [-\tau_{\max}, \tau_{\max}]$.
5	Record the maximum absolute correlation $r_{\max,ij}$ and associated lag τ^* for each channel pair.
6	Partition each deviation signal into contiguous blocks of length L .
7	Generate P dependence-preserving surrogate signals by random permutation of block order.
8	For each surrogate replicate, recompute lag-resolved correlations and the surrogate maximum statistic $\tilde{r}_{\max,ij}^{(p)}$.
9	Construct the empirical max-statistic null distribution from $\{\tilde{r}_{\max,ij}^{(p)}\}_{p=1}^P$.
10	Compute corrected max-statistic p -values $p_{\text{corr},ij}$ using Equation (7).
11	Report lag-resolved synchrony only when $p_{\text{corr},ij} < \alpha$.

6. Synthetic Illustration and Calibration Analysis

To demonstrate the operation of the proposed procedure and assess its calibration properties, we apply the full pipeline under a strict null hypothesis of no cross-channel coupling. The goal is to establish that naive lag scanning inflates false-positive rates under

temporal dependence, to characterise how calibration properties of the proposed procedure depend on the choice of block length relative to the dominant autocorrelation scale, and to compare the proposed scan-level inference framework with established dependence-preserving surrogate pipelines. In particular, the analysis examines conditions under which familywise error control is reliably achieved under varying dependence regimes.

6.1. Simulation Design

For the initial null illustration, each channel was generated independently as a stationary first-order autoregressive (AR(1)) process [15],

$$X_t^{(i)} = \varphi X_{t-1}^{(i)} + \varepsilon_t^{(i)}, \quad \varepsilon_t^{(i)} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1 - \varphi^2), \quad (8)$$

where $i \in \{1, 2\}$ indexes the simulated channels. The innovations were generated independently for each channel, thereby satisfying the strict null hypothesis of no between-channel coupling.

The autoregressive coefficient was fixed at $\varphi = 0.7$ and the series length at $T = 2000$. This configuration produces appreciable within-channel temporal dependence (the ACF decays below 0.2 by approximately lag 5) while maintaining a strict between-channel null. Each simulated series was standardised using Equation (1), with the baseline window W_i taken as the full series; because the simulated processes are stationary with zero mean and unit variance, this step is numerically near-inert but is retained so that the pipeline executed corresponds exactly to Table 2. AR(1) processes were selected because they provide the simplest stationary dependence model in which autocorrelation can be varied systematically while preserving a strict between-channel null, thereby isolating the inferential behaviour of the proposed procedure independently of higher-order process structure. Controlled synthetic simulations were selected as the primary validation instrument because the true null hypothesis is known exactly by construction. In contrast, real observational datasets rarely provide independent verification that no cross-channel dependence is present, making empirical false-positive rates difficult to interpret as evidence of calibration. Synthetic null validation is therefore a methodological requirement for evaluating the empirical calibration of false-positive control under known conditions, rather than a substitute for subsequent application to real data. For each simulated pair, the lag-resolved correlation $C_{12}(\tau)$ was evaluated over the symmetric lag domain $\tau \in [-40, 40]$. The maximum absolute correlation was defined as

$$r_{\max,12} = \max_{\tau} |C_{12}(\tau)| \quad (9)$$

and was recorded for each replicate. A total of 200 Monte Carlo replicates were performed. For each replicate, $P = 1999$ dependence-preserving block permutation surrogates were generated using a reference block length of $L = 120$ to produce the corrected p -value. The nominal significance level was $\alpha = 0.05$.

6.2. False-Positive Inflation Under Naive Lag Scanning

Under a naive analysis, $r_{\max,12}$ was compared against the nominal two-sided single-lag Pearson correlation threshold at $\alpha = 0.05$ without correction for multiplicity across the lag scan. This produced severe inflation of the empirical familywise false-positive rate: the naive threshold was exceeded in all replicates, yielding an empirical FWER of 1.00 (Table 3).

In contrast, the proposed procedure with dependence-preserving block permutation surrogates and max-statistic correction produced corrected p -values that were well-calibrated under the null. At the reference block length $L = 120$, the empirical familywise

error rate (FWER) was 0.045 at $\alpha = 0.05$ —near the nominal level, as expected of a properly calibrated max-statistic procedure.

Table 3. Empirical familywise error rate (FWER) at nominal $\alpha = 0.05$ across block lengths L under two AR(1) dependence regimes. Results are based on 200 Monte Carlo replicates with $T = 2000$, $\tau_{\max} = 40$, and $P = 1999$. Monte Carlo standard errors are approximately 0.015–0.020 for estimates near the nominal level, increasing to approximately 0.029 for the largest corrected FWER estimate.

φ	L	$L/\text{ACF Decay Scale}$	Corrected FWER	Naive FWER
0.7	20	$\sim 4\times$	0.090	1.000
0.7	40	$\sim 8\times$	0.070	1.000
0.7	80	$\sim 16\times$	0.080	1.000
0.7	120	$\sim 24\times$	0.045	1.000
0.7	200	$\sim 40\times$	0.055	1.000
0.95	20	$\sim 0.6\times$	0.205	1.000
0.95	40	$\sim 1.3\times$	0.120	1.000
0.95	80	$\sim 2.5\times$	0.085	0.995
0.95	120	$\sim 3.8\times$	0.065	1.000
0.95	200	$\sim 6.3\times$	0.040	1.000

Figure 2 shows a representative empirical null distribution of $r_{\max,12}$ obtained from block permuted surrogates, together with the naive single-lag threshold and an example observed $r_{\max,12}$ from a null replicate. The naive threshold lies well below the bulk of the empirical null distribution, demonstrating its inadequacy for scan-level inference under temporal dependence.

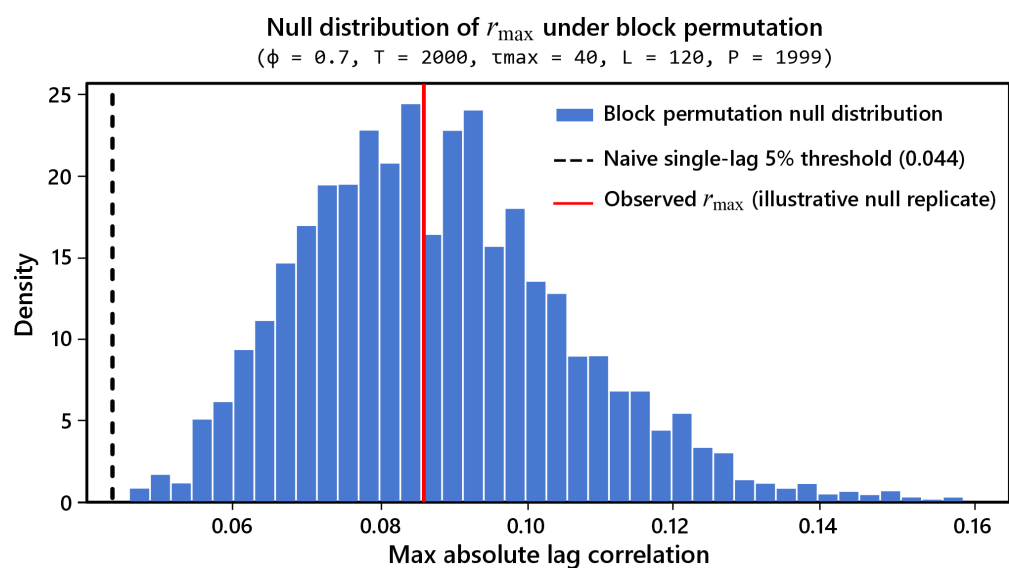


Figure 2. Simulation under a strict null hypothesis (no cross-channel coupling) illustrating false-positive inflation in exploratory lag scanning ($\varphi = 0.7$, $T = 2000$, $\tau_{\max} = 40$). The histogram shows the empirical null distribution of $r_{\max,12}$ generated using dependence-preserving block permutations ($L = 120$, $P = 1999$). The dashed line indicates the nominal two-sided single-lag parametric 5% threshold, which fails to control familywise error across the lag scan. The solid line marks an example observed $r_{\max,12}$ from a null replicate. The separation between the naive threshold and the empirical null distribution illustrates why scan-level correction is required for valid exploratory lag inference under temporal dependence.

6.3. Block Length Sensitivity Analysis

To assess sensitivity of the calibration properties to the choice of block length L , the simulation above was repeated across $L \in \{20, 40, 80, 120, 200\}$ under two autocor-

relation regimes: a moderately dependent AR(1) process with $\varphi = 0.7$ and a strongly dependent AR(1) process with $\varphi = 0.95$. In both cases, $T = 2000$, $\tau_{\max} = 40$, and $P = 1999$ were held fixed. Each (φ, L) combination was evaluated as an independent Monte Carlo experiment with newly generated simulated series and surrogate replicates. Table 3 reports the empirical FWER of the corrected procedure and the naive approach at each block length.

For the moderately dependent process ($\varphi = 0.7$), empirical FWER was highest at $L = 20$ and thereafter fluctuated within the expected Monte Carlo variability, with the larger block lengths achieving values close to the nominal significance level. At $L = 20$ (approximately $4 \times$ the decay scale), corrected FWER was 0.090. For larger block lengths, corrected FWER estimates ranged from 0.045 to 0.080, with no systematic monotonic pattern observed once L substantially exceeded the autocorrelation scale.

The strongly dependent process ($\varphi = 0.95$) produced a more severe version of the same effect. When L was smaller than or only marginally larger than the dominant autocorrelation scale (ACF below 0.2 by lag 32), substantial inflation occurred: corrected FWER reached 0.205 at $L = 20$ and 0.120 at $L = 40$. As L increased further beyond the autocorrelation scale, empirical FWER decreased progressively to 0.085 at $L = 80$, 0.065 at $L = 120$, and 0.040 at $L = 200$.

These results provide empirical support for the block length heuristic described in Section 4.2. In the present simulation setting, the procedure becomes reliably calibrated only when block length substantially exceeds the dominant autocorrelation scale; block lengths that do not substantially exceed the dominant autocorrelation scale can produce appreciable inflation even in otherwise well-behaved AR(1) settings. By contrast, the naive approach yields FWER near unity across all scenarios, confirming that temporal dependence and exploratory lag scanning together generate severe false-positive inflation unless explicitly controlled. The empirical calibration reported here is established under the weak stationarity assumptions described in Section 4. Real observational data may exhibit trends, seasonal structure, regime changes, or other departures from these assumptions that can affect the validity of any dependence-preserving resampling procedure, including block permutation. The present results therefore establish calibration under controlled simulation conditions rather than under arbitrary real-world dynamics. Application to observational data remains an important direction for future work, but such studies address practical performance rather than the calibration of false-positive control, which requires a known null hypothesis.

6.4. Comparison with Existing Dependence-Preserving Inference Pipelines

The block permutation null model described in Section 4 is one of several dependence-preserving surrogate constructions available for autocorrelated time series. Two widely used alternatives are phase-randomised surrogates [6] and the stationary bootstrap [7]. A natural question is whether these established approaches alone resolve the false-positive inflation demonstrated in Section 6.2, or whether scan-level max-statistic correction performs the essential inferential role.

To address this question, we compare complete inference pipelines rather than surrogate constructions in isolation. Six inference pipelines were evaluated under the same strict null model used in Section 6.1, comprising two independent stationary AR(1) processes with identical autoregressive coefficient φ . All pipelines used the same lag-resolved Pearson statistic, the same lag family $\tau \in [-40, 40]$, and the same decision threshold $\alpha = 0.05$. The pipelines differed only in the construction of the null distribution and in whether multiplicity across the exploratory lag scan was incorporated into the inferential procedure. The comparison comprised: (i) a naive single-lag parametric threshold; (ii) phase-randomised surrogates evaluated using a single-lag surrogate threshold; (iii) sta-

tionary bootstrap surrogates evaluated using a single-lag surrogate threshold; (iv) the proposed block permutation pipeline with scan-level max-statistic correction, with significance assessed using the corrected p -value of Equation (7); (v) phase randomisation combined with the same max-statistic correction; and (vi) the stationary bootstrap combined with the same max-statistic correction. The single-lag evaluations therefore represent these dependence-preserving surrogate procedures applied without scan-level multiplicity correction, whereas pipelines (iv)–(vi) evaluate the same surrogate constructions once scan-level correction is incorporated. The max-statistic pipelines are included to distinguish the contribution of dependence-preserving surrogate construction from that of scan-level multiplicity control.

All pipelines were evaluated using 200 Monte Carlo replicates with series length $T = 2000$, surrogate sample size $P = 1999$, and nominal significance level $\alpha = 0.05$. For block permutation, a fixed block length of $L = 120$ was used, whereas the stationary bootstrap employed a geometric block length distribution with expected block length $L = 120$.

Table 4 reports the empirical familywise error rate (FWER) for each pipeline, with Figure 3 providing the corresponding graphical summary. Three principal observations emerge.

Table 4. Empirical familywise error rate (FWER) at nominal $\alpha = 0.05$ for six inference pipelines under the strict null, at two AR(1) dependence regimes. All pipelines share the same lag-resolved Pearson statistic and lag family $\tau \in [-40, 40]$, differing only in their null construction and in whether scan-level max-statistic correction is applied. Pipelines (ii) and (iii) employ the single-lag surrogate threshold defined in Section 6.4, whereas pipelines (iv)–(vi) incorporate scan-level max-statistic correction. Results are based on 200 Monte Carlo replicates with $T = 2000$, $P = 1999$ and $L = 120$.

Pipeline	Dependence Preserved	Scan-Level Correction	Empirical FWER	
			$\varphi = 0.70$	$\varphi = 0.95$
Naive single-lag threshold	×	×	1.000	0.995
Phase randomisation (single-lag)	✓	×	0.785	0.275
Stationary bootstrap (single-lag)	✓	×	0.780	0.290
Block permutation + max (proposed)	✓	✓	0.055	0.050
Phase randomisation + max	✓	✓	0.065	0.040
Stationary bootstrap + max	✓	✓	0.060	0.045

Note: ✓ indicates that the property is present; × indicates that it is absent. Bold type identifies the proposed block permutation + max pipeline and its corresponding empirical FWER values.

First, dependence preservation alone is insufficient to provide valid exploratory lag inference. Phase randomisation and stationary bootstrap surrogates, when evaluated using single-lag thresholds without scan-level correction, yielded empirical FWER between 0.275 and 0.785. Here, the single-lag surrogate threshold is defined as the $(1 - \alpha)$ quantile of the pooled per-lag surrogate correlation magnitudes across all surrogates and lag values, rather than the distribution of scan maxima. Pooling across surrogate replicates and lag values provides a stable estimate of the marginal single-lag null distribution while deliberately omitting scan-level multiplicity correction, thereby isolating the effect of lag selection on familywise error. Although substantially lower than the near-unity false-positive rate of the naive parametric pipeline, these values remain far above the nominal significance level. Preserving temporal dependence therefore mitigates false-positive inflation but does not by itself control familywise error across an exploratory lag scan.

Second, pairing the same two surrogate constructions with scan-level max-statistic correction restored nominal calibration. The proposed block permutation pipeline achieved empirical FWER of 0.055 and 0.050 for $\varphi = 0.7$ and $\varphi = 0.95$, respectively. These values arise from an independent Monte Carlo comparison experiment and are consistent with

the corresponding $L = 120$ results in Table 3 to within Monte Carlo sampling error (approximately 0.015 at 200 replicates). Phase randomisation combined with max-statistic correction yielded 0.065 and 0.040, and the stationary bootstrap combined with max-statistic correction yielded 0.060 and 0.045. Thus, within the surrogate methods examined here, scan-level maximum statistic correction is the component consistently associated with nominal familywise error control. These results support the central methodological claim of this work: preserving temporal dependence is necessary but not sufficient for valid exploratory lag inference, whereas embedding dependence-preserving surrogates within a scan-level maximum statistic framework restores nominal familywise error control.

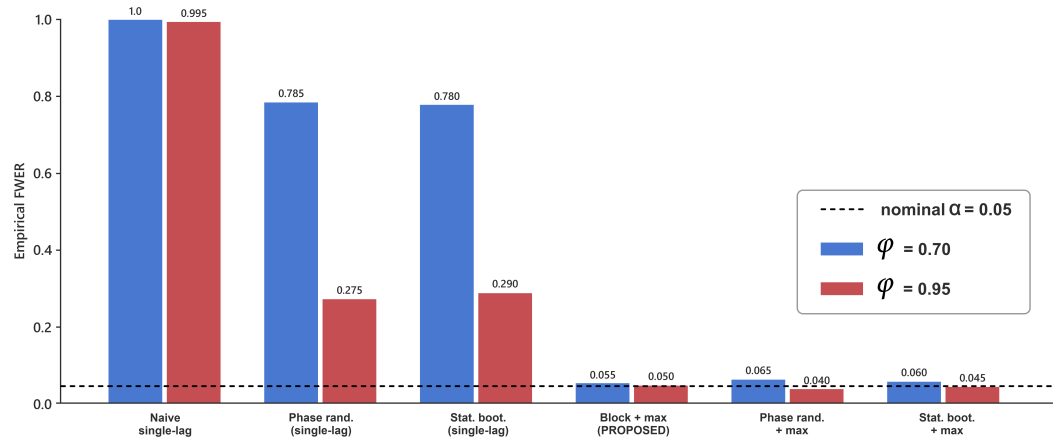


Figure 3. Empirical familywise error rate (FWER) under the strict null for six inference pipelines evaluated at two AR(1) dependence regimes ($\phi = 0.70$ and $\phi = 0.95$). Dependence-preserving surrogate methods applied using single-lag thresholds substantially reduce false-positive inflation relative to the naive parametric baseline but remain far above the nominal significance level. Pairing the same surrogate constructions with scan-level max-statistic correction restores nominal calibration irrespective of the surrogate mechanism. The proposed block permutation pipeline exhibits the same calibrated behaviour, supporting the conclusion that scan-level multiplicity control, rather than the specific surrogate construction, is the critical component of valid exploratory lag inference.

Third, the uncorrected dependence-preserving pipelines exhibited lower FWER at $\phi = 0.95$ than at $\phi = 0.7$. This behaviour is consistent with stronger autocorrelation broadening the surrogate null distribution while simultaneously reducing the effective number of independent lag comparisons. Despite this reduction, all uncorrected pipelines remained substantially above the nominal level, whereas all max-statistic pipelines remained well-calibrated.

All calibrated pipelines lie within Monte Carlo sampling error of the nominal significance level and are not meaningfully distinguishable from one another. Accordingly, the comparison concerns calibration under the strict null rather than relative superiority of individual surrogate methods. Table 4 therefore provides empirical validation of the conceptual comparison introduced in Table 1 for the parametric and surrogate-based approaches considered here.

7. Limitations and Scope

The procedure presented here adopts a calibration-oriented approach to exploratory lag inference. Its primary objective is not to maximise detection sensitivity, but to provide reliable statistical inference under realistic temporal dependence while controlling familywise error across exploratory lag scans. This design choice introduces several limitations that should be made explicit.

First, statistical detection of lag-resolved synchrony does not establish causality, directionality, or mechanism. The procedure is agnostic with respect to the physical or

informational origin of observed structure. A statistically significant result indicates that the detected alignment exceeds expectations under a specified dependence-preserving null model; it does not imply direct interaction or temporal precedence between channels.

Second, statistical power depends on data length and dependence structure. Finite sample sizes and strong autocorrelation broaden null distributions, reducing sensitivity to weak effects. Because the primary objective of the present study is to establish reliable familywise error control under exploratory lag scanning, the simulation study is intentionally focused on calibration under the strict null; comparative power under specific alternative dependence structures is left for future investigation. The present study therefore does not characterise the power of the proposed procedure against meaningful alternatives, and the trade-off between conservative false-positive control and detection sensitivity remains to be quantified. Such a trade-off is inherent when controlling familywise error under exploratory lag scanning. Non-detections should therefore be interpreted as inconclusive rather than as evidence of absence.

Third, the present implementation employs linear correlation as the dependence metric. While this choice facilitates interpretability and computational efficiency, it may fail to capture nonlinear or higher-order dependencies. The procedure itself does not preclude alternative statistics, but any substitution must be paired with null models that preserve the relevant dependence structure.

Fourth, the block-based permutation strategy introduces a tuning parameter—the block length L . Although L is fixed *a priori*, different choices may affect sensitivity to specific timescales. Moderate overestimation of L is generally preferable to underestimation in order to preserve conservative null behaviour; the sensitivity analysis in Section 6.3 demonstrates that the procedure remains well-calibrated across a range of block lengths when L substantially exceeds the dominant autocorrelation scale. The boundary discontinuity limitation of fixed block schemes is discussed in Section 4.2.

Fifth, the present procedure controls familywise error across exploratory lag scans within an individual channel pair. In analyses involving many sensors, additional multiplicity arises across the set of pairwise comparisons. Although secondary procedures such as Bonferroni adjustment or Benjamini–Hochberg procedures may be applied to the resulting corrected pairwise p -values, the present work focuses specifically on within-pair lag multiplicity rather than global cross-pair error control.

The procedure is not intended for settings in which relevant lags are known *a priori*. In such cases—for example, when physical propagation delays are well-characterised—targeted testing at specific lags is likely to be more powerful than exploratory lag scanning with familywise error correction. The present approach is most appropriate when candidate lags are uncertain, transient, or weakly constrained by prior knowledge.

Finally, the procedure assumes concurrent sampling and a shared time base across channels. Extensions to irregularly sampled data or asynchronous sensor arrays would require additional preprocessing or modification of the lag-resolved statistics.

Within these constraints, the scope of the procedure is deliberately broad. It is applicable to multi-sensor stochastic systems in which weak or transient synchrony must be distinguished from correlated noise under exploratory analysis, and may be adapted to a wide range of settings provided that dependence-aware null models can be constructed consistently.

8. Conclusions

This paper has introduced a dependence-aware procedure for lag-resolved correlation analysis in multi-sensor stochastic systems. By integrating baseline standardisation, lag-resolved correlation statistics, dependence-preserving permutation null models, and max-

statistic correction, the procedure addresses a common but often underappreciated source of false-positive structure arising from exploratory lag scanning under temporal dependence.

The contribution is practical and inferential rather than metric-driven. The procedure constrains analytical flexibility by fixing inference decisions *a priori* and pairing exploratory alignment with conservative statistical control. The comparative evaluation against phase-randomised and stationary bootstrap surrogate pipelines further demonstrates that preserving temporal dependence alone is insufficient for valid exploratory lag inference. Rather, nominal familywise error is consistently restored when dependence-preserving surrogate constructions are embedded within a scan-level max-statistic framework. In doing so, the proposed procedure provides a disciplined means of distinguishing reproducible lag-resolved synchrony from artefacts of correlated noise without relying on unrealistic independence assumptions or post hoc lag selection.

Because the procedure is formulated independently of any particular signal origin or coupling mechanism, it is intended for application across a broad range of multi-sensor contexts, subject to the assumptions and limitations discussed throughout the manuscript. Its utility lies not in maximising detection rates, but in ensuring that detected structure remains statistically defensible under realistic dependence conditions.

The present implementation employs linear correlation and regularly sampled signals for clarity and computational simplicity. However, the inferential structure itself is not tied to a specific dependence metric or sampling scheme. In principle, nonlinear or rank-based measures, information-theoretic statistics, and irregularly sampled data could be accommodated within the same inferential framework, provided that corresponding dependence-preserving null models and scan-level correction procedures are developed consistently. These extensions have neither been developed nor empirically evaluated in the present study and therefore remain directions for future methodological research.

Future work may extend the procedure to additional surrogate constructions, adaptive block length selection, cross-pair multiplicity correction, power comparisons under controlled alternatives, or asynchronous sampling settings, while preserving the core principle established here: valid exploratory lag-resolved inference requires explicit control of both temporal dependence and multiplicity across the lag search domain, a conclusion that is consistent across the dependence-preserving surrogate mechanisms examined in the present study.

Funding: This research received no external funding.

Data Availability Statement: The Python simulation code, parameter configurations, fixed random seeds, and reproducibility materials used to generate the synthetic null simulations, block length sensitivity analyses, and pipeline comparison experiments reported in Section 6 are publicly available at Zenodo: <https://doi.org/10.5281/zenodo.20311242>. The illustrative data used in this study are synthetically generated for methodological demonstration purposes.

Acknowledgments: During the preparation of this manuscript, the author used OpenAI ChatGPT for language refinement, formatting assistance, and code-debugging support. The author reviewed and verified all outputs and takes full responsibility for the content of this publication. All scientific decisions, methodological choices, simulations, statistical analyses, interpretation, and final editorial review were performed and verified by the author.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Nichols, T.E.; Holmes, A.P. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.* **2002**, *15*, 1–25. [CrossRef] [PubMed]
2. Westfall, P.H.; Young, S.S. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*; Wiley: New York, NY, USA, 1993.

3. Simmons, J.P.; Nelson, L.D.; Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **2011**, *22*, 1359–1366. [[CrossRef](#)] [[PubMed](#)]
4. Künsch, H.R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **1989**, *17*, 1217–1241. [[CrossRef](#)]
5. Lahiri, S.N. *Resampling Methods for Dependent Data*; Springer: New York, NY, USA, 2003.
6. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Farmer, J.D. Testing for nonlinearity in time series: The method of surrogate data. *Phys. D* **1992**, *58*, 77–94. [[CrossRef](#)]
7. Politis, D.N.; Romano, J.P. The stationary bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 1303–1313. [[CrossRef](#)]
8. Schreiber, T.; Schmitz, A. Surrogate time series. *Phys. D* **2000**, *142*, 346–382. [[CrossRef](#)]
9. Lancaster, G.; Iatsenko, D.; Pidde, A.; Ticcinelli, V.; Stefanovska, A. Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* **2018**, *748*, 1–60. [[CrossRef](#)]
10. Sheng, X.; Shou, W. A rigorous and versatile statistical test for correlations between stationary time series. *PLoS Biol.* **2024**, *22*, e3002758. [[CrossRef](#)]
11. Romano, J.P.; Tirlea, M.A. Permutation testing for dependence in time series. *J. Time Ser. Anal.* **2022**, *43*, 781–807. [[CrossRef](#)]
12. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Academic Press: Oxford, UK, 2011.
13. Lun, D.; Fischer, S.; Viglione, A.; Blöschl, G. Significance testing of rank cross-correlations between autocorrelated time series with short-range dependence. *J. Appl. Stat.* **2022**, *49*, 2934–2950. [[CrossRef](#)] [[PubMed](#)]
14. Good, P.I. *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 3rd ed.; Springer: New York, NY, USA, 2005.
15. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*, 2nd ed.; Springer: New York, NY, USA, 1991.

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