



Representation-dependent topology of visibility and recurrence networks under adaptive decomposition and coordinate compression

Agustín M. de los Riscos^{a,b}, Julio E. Sandubete^{b,c}*, Ana Lazcano^b

^a Faculty of Statistical Studies, Universidad Complutense de Madrid, Avda. Complutense, s/n, Madrid, 28040, Spain

^b Faculty of Law, Business and Government, Universidad Francisco de Vitoria, Carretera de Pozuelo a Majadahonda km. 1.800, Pozuelo de Alarcón, 28223, Madrid, Spain

^c CRIA-BDHS Research Group, Higher Polytechnic School of Technology and Science, Universidad Camilo José Cela, C/ Castillo de Alarcón, 49. Urb. Villafranca del Castillo, Villanueva de la Cañada, 28692, Madrid, Spain

ARTICLE INFO

Dataset link: <https://github.com/agusriscos/GraphEMD>

Keywords:

Time-series networks
Visibility graphs
Recurrence networks
Adaptive signal decomposition
Multiscale representation
Econophysics

ABSTRACT

Time-series-to-network mappings are widely used in statistical physics and econophysics, often treating graph topology as intrinsic to observed dynamics. We test how strongly inferred topology depends on representation choices in multiscale financial signals. Four adaptive decomposition methods are evaluated on controlled non-stationary signals. Cross-mode coupling of stochastic ensemble methods is assessed over 30 Monte Carlo replications, while known-drift recovery is evaluated on a prespecified reference realization. No method is universally superior, so the operational choice is completed on the empirical panel. Five financial series are decomposed, their oscillatory modes compressed by fast independent component analysis (FastICA), and native and compressed representations mapped to natural visibility (NVG), horizontal visibility (HVG), and recurrence networks. HVG edge counts retain their order of magnitude after compression, but mean clustering increases and the native scale-ordered gradient is lost. Dense NVGs on slow native modes arise from natural-visibility geometry and disappear after FastICA recombination and truncation; native NVGs exceed FastICA NVGs by 0.706 on the mean log10 edge-count scale. Recurrence networks remain sparse and fragmented under local-distance calibration, with stronger fragmentation after compression; fixed recurrence-rate and Theiler-window checks show this result is protocol-dependent. Multi-seed and full-rank FastICA controls distinguish truncation loss from invertible linear recombination. Thus, these mappings provide complementary rather than interchangeable descriptions, and their descriptors must be interpreted jointly with the decomposition, coordinate representation, and encoding rule.

1. Introduction

Within statistical-physics approaches to complex systems, time-series-to-network mappings provide relational representations of dynamical data, in which temporal observations or reconstructed states become nodes and geometric or phase-space criteria determine the edges. Visibility graphs and recurrence networks thereby make connectivity, clustering, path structure, and degree heterogeneity available as network descriptors of the encoded dynamics [1–6]. These descriptors are commonly interpreted as

* Corresponding author at: Faculty of Law, Business and Government, Universidad Francisco de Vitoria, Carretera de Pozuelo a Majadahonda km. 1.800, Pozuelo de Alarcón, 28223, Madrid, Spain.

E-mail addresses: ariscos@ucm.es (A.M.d.l. Riscos), je.sandubete@ufv.es (J.E. Sandubete), ana.lazcano@ufv.es (A. Lazcano).

<https://doi.org/10.1016/j.physa.2026.132124>

Received 1 August 2026; Received in revised form 18 September 2026

Available online 9 October 2026

0378-4371/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

properties of the observed dynamics. However, every inferred network is constructed from a particular representation of the signal. For a non-stationary multiscale process, separating temporal scales or subsequently recombining and reducing the resulting channels may alter the input supplied to the network operator and, consequently, the topology obtained from it. The relevant question is therefore not only what dynamical information a time-series network encodes, but also which of its topological properties remain stable when the representation preceding the encoding is changed.

Financial price series provide a demanding test bed for this question because persistent long-horizon drift, medium-term cyclical structure, intermittent fluctuations, and short-horizon volatility coexist within the same realization, with relative contributions that vary over time, across market regimes, and across instruments [7,8]. Descriptors computed directly on the observed path consequently aggregate information from several temporal scales. Fixed-basis spectral representations can separate some of this structure, but their prescribed basis or resolution may be restrictive when oscillations evolve in frequency or occur only locally. Empirical mode decomposition (EMD) and its ensemble extensions instead separate oscillatory intrinsic mode functions (IMFs) from a slow residual without imposing spectral bands a priori [9–11], whereas variational mode decomposition (VMD) obtains band-limited components through a variational formulation [12]. The outputs of these adaptive operators depend jointly on the method, its calibration, and the observed signal. Their modes therefore define alternative representations of the same process rather than neutral preprocessing products.

Visibility and recurrence mappings extract different relational aspects from such representations. Natural and horizontal visibility graphs encode geometric ordering in the time–amplitude plane, whereas recurrence networks encode proximity between states in a reconstructed phase space [1–4]. In financial applications, these operators have mainly been applied to raw prices, returns, or aggregate volatility series to characterize market states, asymmetries, and episodes of instability [13–15]. A smaller group of studies has applied visibility graphs to IMF channels extracted by EMD [16,17]. In those settings, however, the decomposition is typically adopted before the network analysis, the number of encoded channels follows directly from the sifting output, and one network operator is selected without auditing whether alternative representations or encoders preserve the same structural conclusions. The literature therefore provides limited evidence on whether graph differences attributed to the dynamics are stable with respect to scale separation and coordinate representation, or are partly induced by those upstream choices.

This study addresses that question within a deliberately bounded design. The four decomposition methods are benchmarked to select an operational multiscale representation; the study does not claim to compare the graph topologies generated by all four methods. The topological audit is instead conducted within the retained decomposition by comparing its native oscillatory modes with a reduced set of FastICA coordinates and by mapping both representations through natural visibility, horizontal visibility, and recurrence encoders. The analysis asks three related questions: whether replacing the native mode panel with a reduced linear coordinate representation preserves its network structure; which descriptors respond more or less strongly to that representation change; and whether the three encoding rules provide equivalent or complementary accounts of the same oscillatory dynamics. Financial series serve as the empirical test bed, while the broader complex-systems question concerns the dependence of inferred topology on the representation and encoding rule from which it is constructed.

The experimental design separates the stages of this dependence. EMD, EEMD, CEEMDAN, and VMD are first evaluated on controlled signals containing close-frequency beating, chirps, localized bursts, noisy superpositions, and a multi-component stress test with slow drift. For the stochastic ensemble methods, cross-mode linear coupling is treated as an operational diagnostic related to mode mixing and is reported over thirty Monte Carlo replications, with multiple-comparison control for the pairwise correlation tests [18]. The selection is then refined on a heterogeneous panel comprising a global equity index, three sector exchange-traded funds, and a COMEX gold-futures series. Before decomposition, the series are characterized in terms of stationarity patterns, residual dependence beyond a simple linear specification, volatility clustering, rolling price-level dispersion, and local oscillatory behaviour. The retained oscillatory panels are subsequently compressed with FastICA [19,20], and the native and compressed channels are mapped in parallel to natural visibility, horizontal visibility, and recurrence networks. This design makes it possible to assess the effect of linear recombination and dimensionality reduction against a native-mode baseline while holding the encoding definitions and empirical sample fixed. Fig. 1 in Section 3 summarizes the processing chain.

The objective is structural characterization rather than prediction. No forecasting model is estimated because a downstream error measure would aggregate the effects of decomposition, compression, and network encoding without identifying which stage produced them. Each stage is therefore examined through diagnostics appropriate to its role: cross-mode linear coupling and slow-component allocation for decomposition; numerical decorrelation and compression diagnostics for FastICA; and connectivity, density, path, and clustering structure for the resulting networks.

The contributions of the paper are the following:

1. An empirical characterization of representation-dependent topology in multiscale time-series networks. Comparing native oscillatory modes with their FastICA-compressed coordinates shows that replacing the native panel with a reduced linear representation does not affect all network constructions uniformly: HVG edge counts retain the same order of magnitude after compression, whereas HVG clustering becomes higher and less dispersed; NVGs lose the extreme edge counts of the slowest native modes; and recurrence networks remain highly fragmented under the common channel-specific calibration, with the reduced panel more fragmented than the native one. Formal channel-level contrasts and symmetric native/reduced summaries are reported in Sections 4.4 and 4.5.
2. Evidence that natural visibility, horizontal visibility, and recurrence networks provide complementary rather than interchangeable descriptions of the same multiscale dynamics. Their connectivity, density, clustering, and path structure respond differently to the scale content and coordinate representation of the encoded channel. In particular, HVG clustering on native

modes decreases with IMF extraction order, but FastICA scores are not ordered by scale, so that ranking cannot be recovered from the reduced index Z_j . The selected encoder is therefore part of the substantive interpretation rather than a neutral final step (Section 4.5).

3. A controlled and replicated benchmark that supports the selection of an operational decomposition without claiming universal superiority. EMD, EEMD, CEEMDAN, and VMD are exposed to the same catalogue of nonlinear and non-stationary signal pathologies, while the stochastic ensemble methods are evaluated over thirty Monte Carlo replications. Cross-mode linear coupling is reported as an operational diagnostic related to mode mixing rather than as an exhaustive measure of scale separation (Section 4.2).
4. A stagewise audit design that separates decomposition behaviour, coordinate compression, and network encoding. The retained settings are treated as operational choices rather than statistically optimal or universally valid solutions, and preservation of multiscale structure is assessed directly through the topology induced by each representation rather than inferred solely from reconstruction or forecasting accuracy (Sections Section 4.3–4.5).

The remainder of this paper is organized as follows. Section 2 reviews time-series-to-network mappings, adaptive multiscale representations, and coordinate compression, with particular attention to the dependence of graph structure on the encoded signal. Section 3 describes the empirical signal characterization, the operational decomposition, the FastICA compression, and the three network encoders. Section 4 reports the synthetic benchmark, the empirical decomposition, the compression results, and the topological audit. Section 5 discusses the implications for the representation dependence and comparability of time-series networks, states the limitations of the analysis, and concludes.

2. Related work

2.1. Time-series networks and representation dependence

Transforming a time series into a network is an established approach in complex-systems research. The transformation replaces the sequential description of the observed process with a relational representation in which nodes correspond to observations, states, or temporal patterns, and edges encode geometric, ordinal, transitional, or phase-space relations [1,5,6,21]. The resulting network can be characterized through degree, density, clustering, path structure, and connectedness, thereby providing descriptors of dynamical organization that complement conventional time-domain and frequency-domain statistics. Reviews of the field distinguish visibility, transition, and proximity mappings; the present study focuses on visibility and recurrence-based constructions [5,6].

The natural visibility graph (NVG) connects observations that have unobstructed line of sight in the time–amplitude plane [1]. Its topology is sensitive to convexity, amplitude variation, and long-range geometric structure. In particular, smooth or slowly varying segments can create many long-range visibility relations, so dense NVGs on the lowest-frequency modes of an adaptive decomposition are an expected consequence of that line-of-sight geometry rather than an anomalous empirical finding. The horizontal visibility graph (HVG) applies a stricter ordinal criterion and produces a sparser network for which analytical results are available for several stochastic and deterministic processes [2]. At the level of a single channel, both constructions possess relevant invariance properties. NVGs are invariant under positive affine transformations of amplitude, whereas HVGs depend only on amplitude ordering and are therefore invariant under strictly increasing pointwise transformations [1,2]. These properties do not imply invariance under adaptive scale separation or multichannel linear compression. Such operations construct new temporal channels and can alter the amplitude ordering, waveform geometry, and scale content supplied to the network mapping.

Recurrence networks encode a different aspect of the dynamics. A trajectory is reconstructed in phase space, and two states are connected when their distance is below a prescribed threshold [3,4]. Their network descriptors can therefore be related to the local geometry and recurrence structure of the reconstructed dynamics. This interpretation is conditional on the reconstruction and network-construction choices. Embedding dimension, delay, distance norm, temporal exclusion, and recurrence threshold can materially affect connectivity, clustering, and component structure, and inappropriate choices may generate topological artifacts [6,22,23]. Consequently, the calibration of a recurrence network is part of its substantive interpretation rather than a neutral implementation detail.

Previous work has introduced multiscale network constructions by coarse-graining a series before encoding or by modifying the visibility rule to admit penetrable or scale-dependent connections [24,25]. These approaches study how network structure changes with the resolution imposed by the graph construction. The present setting is different: scale is first represented through an adaptive decomposition, after which the resulting channels are either encoded directly or compressed into a reduced coordinate representation. The question is therefore not only whether graph topology changes across temporal scales, but whether conclusions obtained from those scales are preserved after the representation itself is recombined and reduced.

Financial applications provide an established test domain for these constructions. Visibility networks have been used to characterize foreign exchange and equity series, temporal market states, directional asymmetries, and periods of instability [13–15,26]. Most studies encode raw prices, returns, or aggregate volatility. A smaller set of contributions applies visibility mappings to components obtained from empirical mode decomposition [16,17]. These studies show that high-frequency and low-frequency components can generate different degree, density, and clustering patterns. They do not, however, directly audit whether the structural conclusions obtained from the native decomposed channels are preserved after those channels are compressed or whether alternative encoders respond similarly to the same representation change.

2.2. Adaptive multiscale representations of non-stationary signals

Financial markets are a conventional empirical domain in econophysics because their price and return series exhibit non-stationarity, heavy tails, volatility clustering, and scale-dependent fluctuations [7,27]. Price levels commonly combine persistent drift, medium-term oscillatory behaviour, intermittent events, and short-horizon fluctuations within the same realization [8]. They therefore provide a heterogeneous multiscale setting in which a network constructed from the raw path may aggregate dynamical structures operating at substantially different characteristic scales.

Fixed-basis spectral and wavelet representations remain useful for many non-stationary signals, but their predefined bases or resolutions may be restrictive when oscillations evolve in frequency or are localized in time. Empirical mode decomposition (EMD) instead represents a signal as a finite set of intrinsic mode functions (IMFs) and a residual obtained through an iterative sifting procedure [9]. The IMFs are ordered approximately from faster to slower local oscillations, while the final residual contains the component that cannot be further decomposed under the adopted stopping rule. In financial applications, this residual is commonly interpreted as a slow trend component, although the economic interpretation of individual modes cannot be inferred solely from their extraction order.

A principal limitation of EMD is mode mixing, whereby oscillations of different scales appear within the same IMF or a single oscillatory process is distributed across several IMFs [28]. Analytical results for two-tone signals show that the separation obtained by EMD depends on the relative amplitudes and frequencies of the components [28]. Intermittency, localized bursts, noise, and drift can further complicate the allocation of signal content across modes. Pairwise linear correlations between extracted modes can identify one form of residual cross-mode coupling, but they do not provide an exhaustive measure of mode mixing because shared nonlinear, spectral, or time-localized structure can remain even when linear correlation is small.

Ensemble empirical mode decomposition (EEMD) reduces some forms of mode mixing by averaging decompositions obtained after adding independent noise realizations [10]. Its output is stochastic and can retain a non-zero reconstruction discrepancy. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) modifies this construction by introducing adaptive noise at successive stages and enforcing complete reconstruction up to numerical precision [11]. Variational mode decomposition (VMD) follows a different principle, obtaining a prescribed number of compact band-limited modes from a variational optimization problem [12]. The slowest VMD mode is not equivalent by construction to the residual generated by an EMD-family sifting procedure. The methods therefore optimize different representation properties, and reconstruction accuracy alone cannot determine which allocation of slow and oscillatory content is most appropriate for a subsequent network analysis.

Applied studies commonly evaluate decomposition methods through the performance of a downstream forecasting model. CEEMDAN and VMD components have been supplied to recurrent, support-vector, and hybrid predictors, either through one predictor per mode or after grouping selected components [29–31]. Such experiments determine whether a complete prediction pipeline is useful, but forecast error does not isolate the behaviour of the decomposition itself. It also does not show how the selected representation affects a network subsequently inferred from its channels. For the present purpose, comparison among EMD, EEMD, CEEMDAN, and VMD is therefore an operational stage used to obtain a defensible multiscale representation, rather than the principal contribution or a search for a universally superior decomposition method.

2.3. Coordinate compression of decomposed signals

An adaptive decomposition commonly produces several oscillatory channels. Encoding each channel separately preserves the native representation but also makes the resulting graph panel dependent on the number of modes returned by the decomposition. Combining or compressing those channels before network construction changes not only the computational size of the panel but also the time series supplied to the encoder. Dimensionality reduction is therefore a representation choice with potential structural consequences, rather than merely an implementation device.

Principal component analysis (PCA) and independent component analysis (ICA) provide two standard linear approaches to this problem. PCA constructs orthogonal directions ordered by explained variance and produces uncorrelated scores for centred data [32]. ICA assumes a linear mixing model and estimates a demixing transformation using higher-order distributional information, commonly by maximizing a contrast of non-Gaussianity after centring and whitening [19,20,33]. Whitening imposes zero sample cross-correlation on the retained coordinates, so a near-diagonal correlation matrix is an expected property of the transformation and not empirical evidence of statistical independence. Independence concerns the joint distribution and would require additional diagnostics beyond pairwise correlation.

A distinction is also required between coordinate transformation and compression. When all p dimensions are retained, the fitted ICA mapping can be interpreted as a change of coordinates within the whitened subspace. When only $k < p$ coordinates are retained, as in the present study, the mapping also performs dimensionality reduction and is not invertible. Differences between networks constructed from the native modes and from the retained ICA coordinates can therefore reflect linear recombination, changed amplitude ordering and waveform geometry, and discarded dimensions. The compressed representation is treated here as an operational reduced coordinate system, not as a lossless re-expression of the original IMF panel.

ICA has been applied to synchronized financial returns to identify latent shock structures [34], but applications to panels generated by adaptive decomposition are less common. Existing studies have combined complementary ensemble EMD with ICA or EMD with PCA before neural forecasting [35,36]. By contrast, many decomposition-based forecasting architectures preserve the complete mode panel and assign separate predictors to individual components [30,37]. In both cases, compression is assessed mainly through downstream predictive accuracy. Limited attention has been paid to whether the compressed coordinates induce the same connectivity, clustering, and path structure as the native oscillatory channels when both are mapped to networks.

2.4. Research gap and study positioning

The preceding literature identifies several ways in which representation enters a time-series network analysis. Visibility mappings depend on the geometric or ordinal organization of the encoded channel, while recurrence networks additionally depend on phase-space reconstruction and threshold calibration. Adaptive decomposition determines which temporal structures are assigned to each channel, and dimensionality reduction constructs further channels by recombining and truncating the decomposed representation. Nevertheless, these stages are usually studied separately. Decomposition is commonly evaluated through reconstruction or forecast error, compression through explained variance or downstream performance, and network encoding through descriptors calculated from a representation chosen in advance.

The first gap concerns the operational selection of the multiscale representation. EMD, EEMD, CEEMDAN, and VMD differ in stochasticity, reconstruction, mode cardinality, and treatment of slow signal content. Comparisons based on a single realization are insufficient for stochastic ensemble decompositions, while reconstruction error does not reveal how signal content has been allocated across modes. A controlled benchmark over known non-stationary signal structures can support the choice of a representation, provided that cross-mode linear correlation is interpreted as a limited diagnostic of coupling rather than a complete measurement of mode mixing. Such a benchmark informs the subsequent analysis but does not establish a universally optimal decomposition.

The central gap concerns the preservation of network structure after compression. Existing multiscale visibility studies show that networks vary with the scale or resolution of their input, and IMF-based applications show that different native modes generate different graph statistics. It remains less clear whether those scale-dependent conclusions persist when the native mode panel is replaced by a reduced linear representation. In particular, there is limited direct evidence on which descriptors remain comparatively stable, which are strongly altered, and whether natural visibility, horizontal visibility, and recurrence mappings respond in the same way to the same representation change.

A related issue concerns comparability across encoders. Visibility and recurrence networks describe different relations and possess different construction dependencies. Agreement among their descriptors cannot be assumed, while recurrence-network outcomes must be interpreted conditionally on the adopted embedding and threshold procedure. A parallel audit of the three mappings is therefore necessary to distinguish a regularity that is preserved across encoders from one induced by a particular network construction.

The present study is positioned at the intersection of these gaps. The decomposition benchmark is used only to select an operational representation. The main analysis then compares networks inferred from the native oscillatory modes with networks inferred from a reduced FastICA representation, while applying natural visibility, horizontal visibility, and recurrence mappings in parallel. The study does not propose a new decomposition, a new ICA algorithm, or a new network encoder, and it does not claim universal invariance or optimality. Its contribution is a controlled assessment of how representation and encoding choices condition the network topology inferred from multiscale signals.

3. Experimental design

3.1. Study architecture and comparison logic

The experimental design separates the construction of a multiscale representation from the comparison of the networks inferred from it. Each financial series is processed independently through four linked stages. First, the empirical signal is characterized in terms of stationarity, residual temporal dependence, volatility clustering, rolling price-level dispersion, and local oscillatory behaviour. These diagnostics describe the signal supplied to the decomposition and are not treated as necessary or sufficient conditions for empirical mode decomposition. Second, a controlled synthetic benchmark and an empirical comparison are used to select an operational multiscale representation. Third, the native oscillatory panel is compressed with FastICA into a reduced set of decorrelated coordinates. Fourth, both the native oscillatory modes and the reduced coordinates are mapped in parallel through natural visibility, horizontal visibility, and recurrence-network constructions.

No joint model is fitted across instruments. Each asset traverses the processing chain independently, so the resulting graph panels describe asset-specific representations rather than cross-sectional interactions. The comparison has two complementary dimensions. The between-representation comparison contrasts the graphs obtained from the native oscillatory modes with those obtained from the reduced FastICA representation under the same network mapping. The within-representation comparison contrasts the three network mappings applied to a given representation. The design is therefore a structural audit of representation and encoding choices rather than a forecasting or predictive-model-selection exercise.

The input data, source code, configuration files, and analysis scripts are available in the public GraphEMD repository at <https://github.com/agusriscos/GraphEMD>. Graph objects were stored using PyTorch Geometric [38]. Fig. 1 summarizes the complete experimental design.

3.2. Empirical signal characterization

Before decomposition, each instrument is characterized through three groups of diagnostics. Their purpose is to document the statistical and scale-related properties of the empirical signals, not to define a formal admissibility gate or to select instruments *ex post*. All procedures are applied independently to each series. Inferential results are reported at the nominal significance level $\alpha = 0.05$.

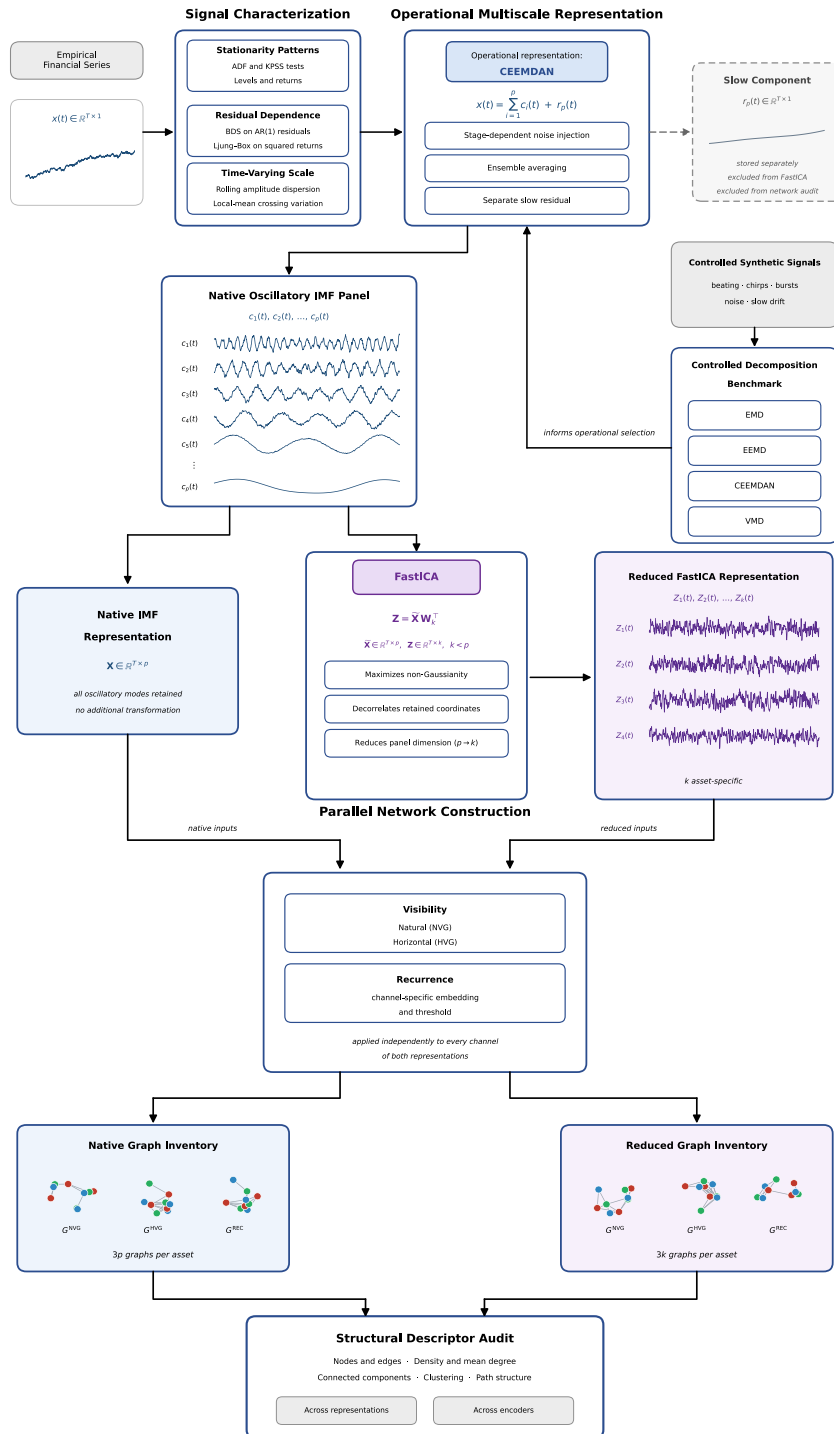


Fig. 1. Experimental design for auditing representation-dependent topology. Empirical signal characterization precedes the operational selection of a multiscale decomposition, informed by a controlled synthetic benchmark over EMD, EEMD, CEEMDAN, and VMD. The retained decomposition separates an oscillatory IMF panel from a slow component. The slow component is stored separately and excluded from the network audit. The oscillatory panel follows two parallel branches: direct encoding of the native IMFs and FastICA compression into a reduced coordinate representation. Natural visibility, horizontal visibility, and recurrence networks are constructed from both representations, and their graph descriptors are compared across representations and encoding rules.

Stationarity patterns. Closing prices and daily percentage returns are examined with the augmented Dickey–Fuller (ADF) test [39], with lag order selected by the Akaike information criterion, and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [40]. For both closing-price levels and daily returns, the implemented KPSS specification includes a deterministic trend and uses automatic bandwidth selection.

The joint pattern of failure to reject a unit root and rejection of trend stationarity in price levels, together with rejection of a unit root and failure to reject the implemented trend-stationarity null in daily returns, is interpreted as evidence consistent with integrated price levels and stationary short-horizon increments. These outcomes characterize the empirical panel but are not treated as mathematical requirements for adaptive decomposition.

Residual dependence and volatility clustering. The BDS statistic [41] is calculated for the residuals of an AR(1) specification fitted to closing prices, over embedding dimensions $m = 3, 4, 5$. Because the underlying price levels are non-stationary, the BDS outcomes are used only as descriptive evidence that a simple linear AR(1) specification does not exhaust the dependence present in the observed path. They are not interpreted as standalone evidence of deterministic chaos or as identification of a particular nonlinear data-generating mechanism.

Volatility dependence is characterized separately through the Ljung–Box statistic applied to squared daily percentage returns at lag 10 [42]. Rejection of the no-autocorrelation null for r_t^2 is reported as evidence consistent with volatility clustering of ARCH type [43,44].

Time-varying amplitude and oscillatory behaviour. Local amplitude variation is described using the standard deviation of closing-price levels within rolling windows of 252 trading days. The temporal heterogeneity of this dispersion measure is summarized by its coefficient of variation. Because it is computed from price levels rather than returns, this quantity is interpreted as a descriptive proxy for local amplitude variability and not as a conventional financial-volatility estimator.

Oscillatory variation is described through crossings of the local window mean. Within each 252-trading-day window, the closing-price observations are centred by the mean of that window. A crossing is recorded whenever the sign of the centred series changes between two consecutive observations. The mean and coefficient of variation of the resulting window-wise crossing counts summarize variation in local oscillatory behaviour. They are not interpreted as estimates of instantaneous physical or spectral frequency.

The coefficients of variation are used descriptively. No universal threshold is imposed as a necessary or sufficient condition for adaptive decomposition. The empirical outcomes are reported in Section 4.1.

3.3. Operational multiscale representation

Let $x(t)$ denote one financial series observed at $t = 1, \dots, T$. The purpose of this stage is to construct an operational multiscale representation that separates oscillatory content from the slowest component before the network audit. Four candidate methods are considered: empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and variational mode decomposition (VMD) [9–12].

The comparison supports the choice of the representation used in the subsequent network analysis. It is not intended to establish a universal ranking of decomposition methods, and the graph topologies generated by all four methods are not compared.

EMD-family representation. For the EMD family, the observed signal is represented as

$$x(t) = \sum_{i=1}^p c_i(t) + r_p(t), \quad (1)$$

where $c_i(t)$ is the i th intrinsic mode function and $r_p(t)$ is the final residual under the adopted stopping rule. The IMFs are ordered approximately from higher to lower local frequency. The residual is treated operationally as the slow component, but no claim is made that it corresponds uniquely to an economic trend or to a particular generating mechanism.

EMD extracts the components through iterative sifting [9,28]. EEMD averages decompositions generated after adding independent noise realizations [10]. Its finite ensemble can leave a non-zero reconstruction discrepancy, and its output depends on the noise realizations. CEEMDAN introduces stage-dependent noise and a uniquely defined residual at each extraction stage, producing a complete decomposition up to the numerical tolerance of the implementation [11]. VMD follows a different principle and estimates a prescribed number of compact, band-limited modes through a variational optimization problem [12]. Under the adopted VMD configuration, its zero-frequency mode is treated as the slow-component analogue, although it is not mathematically equivalent to the residual generated by EMD-family sifting.

Controlled synthetic benchmark. The candidate methods are first evaluated on signals whose composition is known by construction. The benchmark comprises close-frequency beating, a linear chirp, a localized high-frequency burst on a slow carrier, a noisy superposition of close tones and a chirp, and a multi-component stress signal combining spectral proximity, intermittency, noise, and slow drift. The signal definitions and results are presented in Section 4.2.

Cross-mode linear coupling is summarized over the first $\min\{8, p\}$ oscillatory components through

$$M_{\text{corr}} = \text{median}_{i < j} |\text{corr}(c_i, c_j)|. \quad (2)$$

This statistic is treated as an operational diagnostic of residual linear coupling associated with one manifestation of mode mixing. It is not interpreted as a complete measure of spectral, nonlinear, or time-localized overlap between components.

EEMD and CEEMDAN are stochastic because their output depends on the ensemble noise realizations. For the cross-mode-coupling analysis defined in Eq. (2), each stochastic decomposition is repeated over $R = 30$ independent seeds, and the resulting distributions of M_{corr} are summarized through percentile intervals. The percentage of pairwise correlations for which a zero-correlation null is rejected after Bonferroni correction [18] is reported as a complementary diagnostic. Because the extracted modes retain temporal dependence, these rejection percentages are interpreted descriptively and not as exact finite-sample inference for independent observations. This $R = 30$ replication design applies to the cross-mode-coupling diagnostics; the known-drift selection check described below is evaluated on the prespecified seed-42 reference realization.

Numerical reconstruction error is reported separately. A small normalized reconstruction error indicates closure of the decomposition but does not show how the signal has been allocated across slow and oscillatory components. Reconstruction and cross-mode coupling are therefore treated as complementary rather than interchangeable diagnostics.

For the multi-component stress signal, whose known ascending slow drift is $d(t) = 8t/T$, slow-component allocation is evaluated explicitly on the prespecified seed-42 reference realization. Let $\mathcal{E}$ denote the set of eligible methods. A method belongs to $\mathcal{E}$ when $\rho_S(s_{\text{slow}}, t) > 0$ and $\text{FracIncr}(s_{\text{slow}}) \geq 0.65$. Among the eligible methods, the operational selection score is

$$J_B(m) = \frac{1}{2} \log_{10} \widetilde{\mathcal{R}_{\Delta^2, m}} + \frac{1}{2} (1 - \widetilde{\text{FracIncr}_m}), \quad (3)$$

where the tilde denotes min-max normalization over the eligible set $\mathcal{E}$; that is, $\widetilde{q}_m = (q_m - \min_{j \in \mathcal{E}} q_j) / (\max_{j \in \mathcal{E}} q_j - \min_{j \in \mathcal{E}} q_j)$. The method with the lowest J_B is retained by this synthetic slow-component check. The eligibility threshold of 0.65 and the equal weighting of the two terms are operational design choices rather than estimated universal parameters. This known-drift check is based on the reference realization and is distinct from the $R = 30$ cross-mode-coupling analysis.

Empirical calibration and operational selection. The candidate methods are subsequently compared on the empirical panel using method-specific calibration. This asymmetry reflects differences in their construction and does not define a common optimization problem.

For CEEMDAN, candidate configurations span noise scales $\varepsilon \in [0.03, 0.12]$, ensemble sizes between 100 and 220 trials, and maximum IMF counts between 9 and 14. Configurations with $\text{NRMSE} < 10^{-10}$ form the primary candidate set. MSCI World additionally retains the prespecified reference configuration ($\varepsilon = 0.05$, $N_{\text{trials}} = 100$, $\text{max_imf} = 14$). Although this configuration lies outside the primary machine-precision set, its NRMSE of 6.8×10^{-7} remains negligible relative to the signal norm and is reported explicitly. Among the retained candidates, the operational setting is selected by prioritizing the concentration of long-horizon variation in a smooth slow component, using temporal linearity, the fraction of positive first differences, and roughness as descriptive criteria. Ties are resolved through lower linear-fit error and weaker cross-mode linear coupling.

EEMD is evaluated under a common empirical configuration across assets, with noise width 0.05, 100 trials, sifting threshold 0.25, and a maximum of 14 modes. VMD is evaluated over candidate mode counts $K \in [6, 12]$ and bandwidth penalties $\alpha \in [10^3, 5 \times 10^3]$, with the DC constraint enabled. Its retained configuration minimizes mean absolute cross-mode correlation within the evaluated grid.

Classical EMD is retained in the synthetic benchmark as a deterministic reference but is not carried to the empirical calibration. This is a scope decision and does not imply that EMD lacks a residual or slow component. CEEMDAN is ultimately retained as the operational representation for the network audit according to the empirical results reported in Section 4.3.

For the retained representation, the complete oscillatory panel

$$\mathbf{C} = \{c_1, \dots, c_p\} \quad (4)$$

defines the native representation. The residual r_p is stored separately for reconstruction and traceability but is excluded from the FastICA fit and from both graph inventories. The subsequent comparison therefore concerns alternative representations of the same oscillatory panel.

3.4. Reduced FastICA representation

Let

$$\mathbf{X} = \begin{bmatrix} c_1(1) & \cdots & c_p(1) \\ \vdots & & \vdots \\ c_1(T) & \cdots & c_p(T) \end{bmatrix} \in \mathbb{R}^{T \times p} \quad (5)$$

denote the native oscillatory panel, with one IMF per column and observations aligned by time. The slow component is not included in $\mathbf{X}$. Each column is standardized before fitting FastICA [19,20,33].

FastICA is used in its parallel form with unit-variance whitening (scikit-learn FastICA; convergence tolerance 10^{-4} ; maximum of 3000 iterations). The operational network audit retains the fit obtained with random seed 42. Robustness is assessed by repeating the fit over twenty initializations (seeds 42–61) at neighbouring ranks $k = 3, 4, 5$ and at full rank $k = p$, as reported in Section 4.4. For a retained dimensionality $k < p$, the fitted transformation produces

$$\mathbf{Z} = \widetilde{\mathbf{X}} \mathbf{W}_k^\top, \quad \mathbf{W}_k \in \mathbb{R}^{k \times p}, \quad \mathbf{Z} \in \mathbb{R}^{T \times k}, \quad (6)$$

where $\widetilde{\mathbf{X}}$ is the standardized IMF matrix and $\mathbf{W}_k$ is the retained unmixing matrix. Because $k < p$, the operation combines a linear recombination of the original modes with dimensionality reduction and is not invertible. The retained coordinates are therefore not treated as a lossless re-expression of the native IMF panel.

Whitening imposes zero sample cross-correlation among the estimated coordinates by construction. The near-zero correlations reported in Section 4.4 are consequently numerical checks of the fitted solution and not empirical evidence of statistical independence. Independence concerns the complete joint distribution and is not established in this study. The outputs $Z_1, \dots, Z_k$ are therefore described as decorrelated coordinates.

Candidate dimensions are evaluated over

$$k \in \{2, \dots, p-1\}. \quad (7)$$

For each converged candidate, the recorded diagnostics include the maximum absolute off-diagonal correlation among the retained coordinates, the mean column-wise reconstruction coefficient of determination, and the relative Frobenius reconstruction error

$$E_F(k) = \frac{\|\mathbf{X} - \hat{\mathbf{X}}_k\|_F}{\|\mathbf{X}\|_F}, \quad (8)$$

where $\hat{\mathbf{X}}_k$ is obtained through the inverse FastICA mapping. Reconstruction is reported to quantify the information discarded by the reduction; it is not the sole objective of the compression stage.

Because whitening makes the off-diagonal sample correlations numerically negligible for every converged candidate, cross-correlation is used only as a convergence diagnostic and does not discriminate among candidate dimensions. Compact candidates are identified by requiring mean column-wise reconstruction R^2 to reach at least 72% of the maximum value over the candidate grid. Relative Frobenius error is then used as a secondary reconstruction diagnostic, and parsimony determines the operational dimensionality among candidates with comparable reconstruction. For XLP, these diagnostics do not identify a unique preferred value, and $k = 4$ is retained for consistency with the MSCI World reference setting.

The operational dimensions used throughout the subsequent network analysis are $k = 4$ for MSCI World and XLP, and $k = 5$ for XLE, XLV, and COMEX gold futures. These values are fixed design settings rather than estimates of a uniquely identified latent dimension. The reconstruction diagnostics and the asset-specific selection rationale are reported in Section 4.4, and all topological findings are conditional on the retained dimensions.

A full-rank control with $k = p$ retains every oscillatory coordinate and therefore isolates the invertible linear recombination from dimensionality reduction. Reconstruction diagnostics at $k = p$ quantify whether residual error at the operational $k < p$ is attributable to truncation rather than to the ICA recombination itself. The multi-seed experiment assesses stability of reconstruction and numerical decorrelation across initializations; it does not rebuild the full graph inventories for every seed, and seed-wise invariance of topological descriptors is therefore not claimed.

The component labels $Z_1, \dots, Z_k$ are internal to each FastICA fit. Their order and sign are not identified, and equal indices do not imply one-to-one correspondence between different assets.

3.5. Network construction and structural audit

For each asset, the three network mappings are applied to two representations: the p native oscillatory IMFs and the k retained FastICA coordinates. The resulting inventory contains $3p$ native graphs and $3k$ compressed-representation graphs per instrument. The slow component is excluded from both inventories by design.

The native and compressed panels have different cardinalities, and their channels are not matched one to one. The comparison is therefore conducted through graph-level descriptor distributions and encoder-specific inventories rather than through direct correspondence between individual native and compressed channels.

All reported graph statistics are calculated after converting the stored edge lists into simple undirected graphs. Weighted visibility constructions are not considered because they would introduce edge weighting as an additional experimental factor [45].

Natural visibility graphs. For a channel $\{y_t\}_{t=1}^T$, the natural visibility graph (NVG) has node set $V = \{1, \dots, T\}$. For $i < j$, an edge exists when

$$A_{ij}^{\text{NVG}} = \mathbb{1} \left[y_\ell < y_i + \frac{t_\ell - t_i}{t_j - t_i} (y_j - y_i) \quad \forall \ell \in \{i+1, \dots, j-1\} \right]. \quad (9)$$

Thus, two observations are connected when no intermediate observation obstructs the line segment joining them in the time–amplitude plane [1]. The construction preserves the complete sample length, so $|V| = T$.

Horizontal visibility graphs. The horizontal visibility graph (HVG) uses the same node set and defines

$$A_{ij}^{\text{HVG}} = \mathbb{1} [y_\ell < \min(y_i, y_j) \quad \forall \ell \in \{i+1, \dots, j-1\}]. \quad (10)$$

HVG connectivity depends on relative amplitude ordering and generally produces a sparser graph than NVG [2]. As with NVG, the construction preserves the complete sample length.

For implementation, the visibility relations are generated in increasing temporal direction. Before graph descriptors are calculated, the stored relations are converted into a simple undirected graph, so each unordered node pair contributes a single edge to the reported edge count.

Recurrence networks. Recurrence networks are constructed from a channel-specific delay embedding [3,4]. For delay τ and embedding dimension d , the reconstructed state is

$$\mathbf{u}_i = [y_i, y_{i+\tau}, \dots, y_{i+(d-1)\tau}]^T, \quad i = 1, \dots, T - (d-1)\tau. \quad (11)$$

The delay is selected from the first local minimum of the time-delayed mutual information [46]. Mutual information is estimated after discretizing the channel using $\lfloor \sqrt{T} \rfloor$ equally spaced cut points over its observed range. Delays from 1 to 50 are evaluated. When no local minimum is found, the first delay at which the discrete mutual-information slope becomes positive is retained. If no such change occurs, $\tau = 2$ is used as the fallback value.

The embedding dimension is selected with the false-nearest-neighbour criterion [47]. Candidate dimensions up to $d = 10$ are evaluated. For every embedded state, its nearest non-self neighbour is identified in dimension d . The neighbour is classified as false when

$$\frac{|y_{i+d\tau} - y_{j+d\tau}|}{\|\mathbf{u}_i^{(d)} - \mathbf{u}_j^{(d)}\|_2} > 10. \quad (12)$$

The first dimension for which the proportion of false neighbours falls below 0.10 is retained. When no candidate satisfies this condition, the dimension with the smallest observed false-neighbour proportion is selected. For the false-nearest-neighbour search used to select d , no temporal-exclusion or Theiler window is imposed beyond removal of the state itself.

Two embedded states are connected when their Euclidean distance does not exceed a channel-specific radius:

$$A_{ij}^{\text{REC}} = \mathbb{1} \left[\|\mathbf{u}_i - \mathbf{u}_j\|_2 \leq \epsilon_{\text{rec}} \right], \quad i \neq j. \quad (13)$$

The radius is estimated independently for each channel. With random seed 42, up to 1000 embedded states are sampled without replacement. For every sampled state, distances to its 99 nearest non-self neighbours are calculated and pooled; when fewer neighbours are available, all available non-self neighbours are used. The radius ϵ_{rec} is set to the tenth percentile of this pooled local-distance distribution and is then applied to all embedded states of the channel.

No additional channel standardization is applied specifically before the recurrence-distance calculation. Because the radius is recalibrated independently for every channel, its numerical magnitude adapts to the scale of that channel. Nevertheless, the resulting density, connectivity, and component structure remain conditional on the embedding procedure, local neighbourhood size, sampled states, selected percentile, and Theiler-window choice.

The main between-representation comparison uses the channel-specific local-distance radius above with no Theiler window ($W = 0$). This construction does not impose a common absolute radius or a fixed global recurrence rate. To assess calibration dependence, the same channels are also rebuilt under a fixed recurrence rate $\text{RR} = 5\%$ with $W = 0$, and under $\text{RR} = 5\%$ with Theiler windows $W \in \{0, \tau, 2\tau, 5\tau\}$, while retaining each channel's delay τ and embedding dimension d . The sensitivity analysis covers both native CEEMDAN and reduced FastICA panels for all five assets (ten panels; seventy-four channels including residues). The primary native-versus-reduced audit therefore evaluates a common calibration protocol; the RR and Theiler checks quantify how strongly fragmentation depends on threshold and temporal exclusion [22,23].

Structural descriptors. Each simple undirected graph is characterized by its number of nodes $|V|$, number of edges $|E|$, edge density, number of connected components, mean degree, diameter estimate, and local clustering-coefficient distribution.

Edge density is calculated as

$$\rho_G = \frac{2|E|}{|V|(|V| - 1)}, \quad (14)$$

and mean degree as

$$\bar{k}_G = \frac{2|E|}{|V|}. \quad (15)$$

For connected graphs, the reported diameter is estimated through a double-sweep breadth-first-search procedure: a first traversal identifies a distant node, and a second traversal from that node provides the maximum observed shortest-path distance. The value is therefore an operational diameter estimate rather than an exact all-pairs diameter. Diameter is not reported for disconnected graphs.

The mean, median, and standard deviation of the node-wise local clustering coefficients are calculated using the available values. Clustering coefficients are unavailable for nine computationally demanding native NVGs because these calculations were omitted from the original graph inventory. These values are reported as *n.a.*, excluded from clustering summaries and contrasts, and never interpreted as zero.

Comparison logic. Two comparisons are performed. The within-representation comparison evaluates how NVG, HVG, and recurrence constructions describe the same channel differently. The between-representation comparison contrasts the distributions of graph descriptors obtained from the native IMF panel and from the reduced FastICA panel under each encoder.

Formal native-versus-reduced contrasts treat each oscillatory IMF or FastICA coordinate as the analysis unit and exclude the slow residue. For central descriptors we report the difference of means (native minus reduced), Cohen's d with pooled standard deviation, a 95% bootstrap confidence interval based on 5000 resamples, and a two-sided permutation p -value based on 10,000

permutations. Because several channels come from the same asset, these analyses provide operational support for the observed descriptor distributions rather than independent population trials.

For visibility graphs, whose construction introduces no estimated embedding parameters, differences primarily reflect the changed input representation. For recurrence networks, differences reflect both the changed representation and the channel-specific recalibration of delay, embedding dimension, and distance radius.

Aggregate edge counts across an entire graph inventory are reported only as descriptive workload and connectivity summaries. They are not treated as direct measures of topological preservation because the native and compressed representations contain different numbers of channels. The results of the structural audit are presented in Section 4.5.

4. Results

4.1. Data and signal characterization

The empirical panel comprises five daily closing-price series: the MSCI World global equity index, three U.S. sector exchange-traded funds (XLE, energy; XLP, consumer staples; XLV, health care), and the COMEX gold futures series identified by the Yahoo Finance ticker GC=F. The original series cover the period from January 12, 2012 to April 20, 2026. After calendar alignment, the empirical analysis is restricted to the 3584 dates jointly available for all five series. Daily percentage returns are used for the return-based diagnostics, whereas decomposition, compression, and network construction operate on the aligned closing-price levels.

The panel is deliberately heterogeneous. MSCI World provides a broad global equity trajectory; XLE contributes cyclical and commodity-linked sector variation; XLP represents a comparatively defensive consumption sector; XLV provides health-care-specific dynamics; and COMEX gold futures provide a non-equity commodity series. The purpose of this composition is not to infer cross-asset relations, since every series is processed independently, but to examine whether the representation and network effects remain visible across signals with different trend, amplitude, and oscillatory characteristics.

Table 1 reports the signal-characterization diagnostics defined in Section 3.2. For all five series, ADF fails to reject a unit root in closing-price levels while KPSS rejects the implemented trend-stationarity null. On daily returns, ADF rejects a unit root and KPSS fails to reject the implemented trend-stationarity null. This joint pattern is consistent with integrated price levels and stationary short-horizon increments.

The BDS statistic rejects its independence null for the residuals of the AR(1) specification at embedding dimensions $m = 3, 4, 5$ on every asset. Given that the regressions are fitted to non-stationary price levels, these results are interpreted descriptively as evidence that a simple AR(1) specification does not exhaust temporal dependence; they are not treated as standalone proof of deterministic chaos or of a particular nonlinear mechanism. The Ljung–Box statistic applied to squared returns at lag 10 also rejects the absence of autocorrelation for all assets, consistently with volatility clustering.

Rolling price-level dispersion is temporally heterogeneous, with coefficients of variation ranging from 0.378 for XLP to 0.952 for COMEX gold futures. The mean number of crossings of the local 252-day window mean ranges from 11.04 to 13.95. These statistics describe changes in local amplitude and oscillatory behaviour and are not used as pass–fail criteria for decomposition.

4.2. Controlled decomposition benchmark

The four candidate methods are first compared on signals whose composition is known by construction. Five families are generated ($T = 2500$, sampling rate 500 Hz, duration 5 s), each representing a different challenge for adaptive decomposition: close-frequency tones that produce amplitude beating; a linear chirp with continuously evolving instantaneous frequency; a localized high-frequency burst on a low-frequency carrier; a noisy superposition of close tones and a chirp; and a multi-component stress signal combining spectral proximity, ascending and descending chirps, localized oscillations, Gaussian noise, and a slow linear drift. Fig. 2 shows the resulting waveforms.

Each family is decomposed with EMD, EEMD, CEEMDAN, and VMD [9–12]. Method-specific settings for EEMD, CEEMDAN, and VMD are calibrated on the multi-component stress signal, while EMD enters as a deterministic reference under the shared mode cap.

Cross-mode linear coupling is evaluated over the first $\min\{8, N_{\text{IMF}}\}$ oscillatory components, excluding the slow component. The first summary is the median absolute Pearson correlation between mode pairs. The second is $\pi_{0.05}^{\text{Bonf}}$, the percentage of pairwise zero-correlation tests rejected after Bonferroni correction. As discussed in Section 3.3, this percentage is used as a descriptive complement because the temporal dependence of the extracted modes limits a conventional interpretation of the individual correlation tests.

EEMD and CEEMDAN are evaluated over $R = 30$ independent noise seeds. Their reported values are therefore accompanied by 95% percentile intervals across replications. EMD and VMD are deterministic under the implemented settings and are evaluated once per signal family. When the corresponding method-specific remainder is included, all four implementations reconstruct the synthetic signals to numerical precision. Closure therefore does not discriminate among the methods in this benchmark.

No method dominates every signal family. VMD produces the lowest median coupling on the close-tone, noisy-superposition, and multi-component stress signals. On the stress signal, its median is 0.004 and $\pi_{0.05}^{\text{Bonf}} = 7\%$. CEEMDAN and EEMD both attain a replication median of 0.008 on that family, with intervals [0.006, 0.009] and [0.006, 0.013], respectively. EMD produces a median of 0.010 and a rejection percentage of 14%.

The stochastic variation is most visible for CEEMDAN on the close-tone and chirp signals. On close tones, its median absolute correlation ranges up to 0.048 and its corrected rejection percentage up to 38% across seeds. On the chirp, the corresponding

Table 1

Empirical signal characterization by asset (independent analysis per series, $T = 3584$). Test statistics are shown with p -values in parentheses. BDS is applied to residuals from an AR(1) model fitted to closing prices. The Ljung–Box statistic is applied to squared daily returns at lag 10. Price-dispersion CV is the coefficient of variation of the rolling 252-day standard deviation of closing-price levels. Mean crossings denotes the average number of crossings of the local window mean within rolling 252-day windows.

Diagnostic	MSCI	XLE	XLP	XLV	GC=F
Stationarity: closing prices					
ADF	1.66 ($p = 0.998$)	−0.08 ($p = 0.952$)	−0.55 ($p = 0.882$)	−0.61 ($p = 0.870$)	2.96 ($p = 1.000$)
KPSS	1.33 ($p \leq 0.010$)	1.61 ($p \leq 0.010$)	0.87 ($p \leq 0.010$)	0.70 ($p \leq 0.010$)	1.40 ($p \leq 0.010$)
Stationarity: daily returns (%)					
ADF	−19.73 ($p < 0.001$)	−11.68 ($p < 0.001$)	−18.89 ($p < 0.001$)	−16.47 ($p < 0.001$)	−62.65 ($p < 0.001$)
KPSS	0.03 ($p \geq 0.100$)	0.05 ($p \geq 0.100$)	0.02 ($p \geq 0.100$)	0.03 ($p \geq 0.100$)	0.05 ($p \geq 0.100$)
Residual dependence: BDS on AR(1) residuals					
$m = 3$	21.27 ($p < 0.001$)	17.28 ($p < 0.001$)	22.71 ($p < 0.001$)	21.53 ($p < 0.001$)	24.85 ($p < 0.001$)
$m = 4$	25.28 ($p < 0.001$)	20.75 ($p < 0.001$)	25.54 ($p < 0.001$)	24.93 ($p < 0.001$)	26.37 ($p < 0.001$)
$m = 5$	28.35 ($p < 0.001$)	23.42 ($p < 0.001$)	27.84 ($p < 0.001$)	27.72 ($p < 0.001$)	27.44 ($p < 0.001$)
Volatility clustering: Ljung–Box on r_T^2					
Ljung–Box	3236.82 ($p < 0.001$)	2267.29 ($p < 0.001$)	5072.60 ($p < 0.001$)	4816.87 ($p < 0.001$)	318.03 ($p < 0.001$)
Time-varying amplitude and oscillatory behaviour					
Price-dispersion CV (252d)	0.597	0.462	0.378	0.427	0.952
Mean local-mean crossings	12.04	12.67	13.95	13.31	11.04

Table 2

Controlled benchmark of cross-mode linear coupling. Each cell reports the median absolute pairwise correlation among the first $\min\{8, N_{\text{IMF}}\}$ oscillatory components. Bracketed ranges are 95% replication intervals for EEMD and CEEMDAN over $R = 30$ noise seeds. Parenthetical values report $\pi_{0.05}^{\text{Bonf}}$, the percentage of pairwise zero-correlation tests rejected after Bonferroni correction; corresponding replication intervals are again shown for the stochastic methods. Lower values indicate weaker linear coupling, but not necessarily complete spectral or nonlinear separation.

Family	EMD	EEMD	CEEMDAN	VMD
Close tones	0.038 (40%)	0.012 [0.009, 0.019] (14% [11%, 19%])	0.019 [0.012, 0.048] (25% [14%, 38%])	0.002 (13%)
Linear chirp	0.009 (33%)	0.009 [0.006, 0.015] (25% [21%, 30%])	0.022 [0.010, 0.089] (40% [27%, 55%])	0.038 (33%)
Burst on carrier	0.057 (67%)	0.010 [0.007, 0.015] (25% [18%, 29%])	0.008 [0.004, 0.011] (20% [13%, 29%])	0.070 (60%)
Close tones + chirp + noise	0.028 (33%)	0.012 [0.008, 0.019] (21% [18%, 29%])	0.011 [0.009, 0.014] (19% [14%, 25%])	0.004 (27%)
Multi-component stress	0.010 (14%)	0.008 [0.006, 0.013] (25% [25%, 29%])	0.008 [0.006, 0.009] (24% [19%, 24%])	0.004 (7%)

correlation interval extends to 0.089. These ranges show why a single stochastic decomposition would not characterize the method reliably in those regimes.

Averaged over the five signal families, EEMD has the lowest mean replication median of absolute cross-mode correlation, approximately 0.010. VMD, CEEMDAN, and EEMD nevertheless exchange relative positions across the individual challenges. The benchmark therefore supports method-specific characterization rather than a universal ordering. In particular, Table 2 does not select CEEMDAN through mode mixing: on the multi-component stress signal, VMD attains the lowest median absolute cross-mode correlation (0.004), while EEMD and CEEMDAN both obtain 0.008.

Cross-mode coupling alone does not determine how the slow part of the signal is allocated. On the multi-component stress signal, the slow-component energy fraction

$$\eta_{\text{slow}} = \frac{\|s_{\text{slow}}\|_2^2}{\|x\|_2^2} \quad (16)$$

is approximately 0.4% for the retained VMD DC mode and 75.5% for the CEEMDAN residual. These values do not by themselves establish that one allocation is intrinsically correct. They show that the methods place the slow and oscillatory content in substantially different representations, even when reconstruction is numerically equivalent.

The stress signal also contains a known linear drift $d(t) = 8t/T$ over $T = 5$ s, which provides an operational check of slow-component allocation. This check is evaluated on the prespecified seed-42 reference realization using the eligibility rule and score J_B defined in Eq. (3). CEEMDAN and EMD satisfy the eligibility conditions. CEEMDAN attains both lower roughness (2.78×10^{-6}

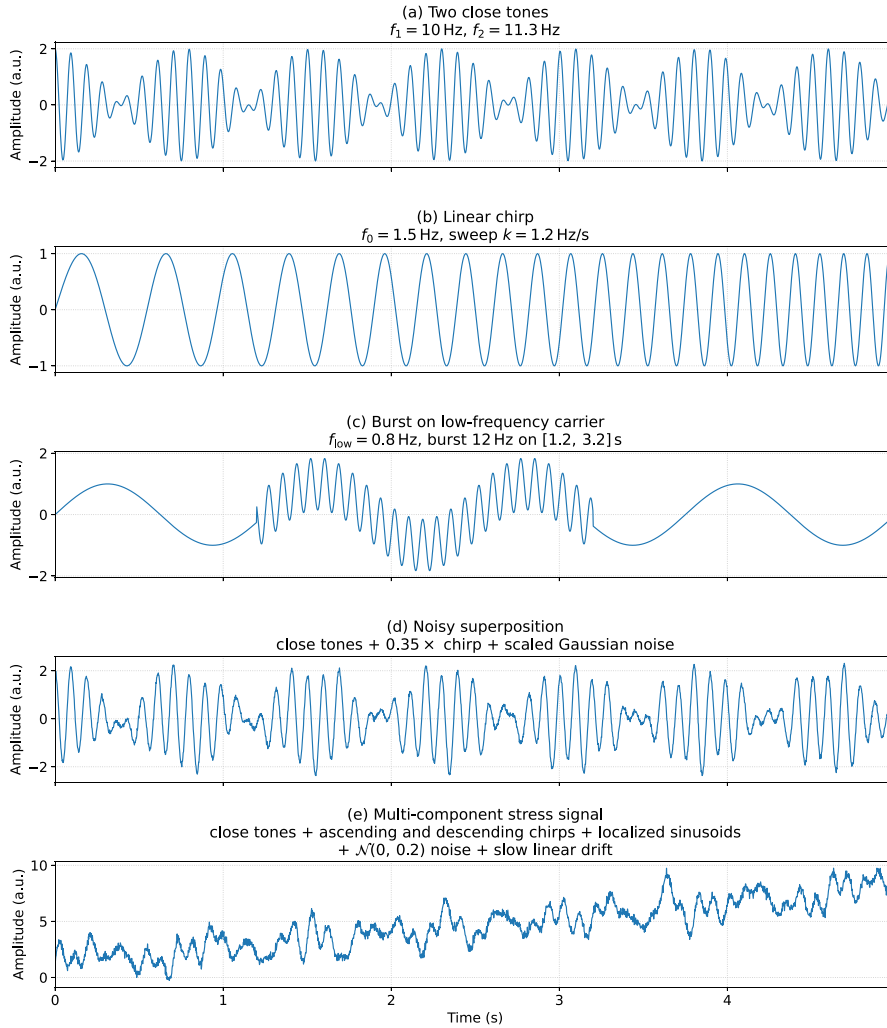


Fig. 2. Controlled signals used in the decomposition benchmark ($T = 2500$, 500 Hz, 5 s). (a) Close tones at 10 Hz and 11.3 Hz, producing amplitude beating. (b) Linear chirp with $f_0 = 1.5$ Hz and sweep rate 1.2 Hz/s. (c) A 12 Hz localized burst superimposed on a 0.8 Hz carrier over $t \in [1.2, 3.2]$ s. (d) Noisy superposition of close tones and a scaled chirp. (e) Multi-component stress signal combining close tones, ascending and descending chirps, localized oscillations, Gaussian noise, and a slow linear drift.

versus 3.34×10^{-6}) and a larger fraction of positive increments (0.745 versus 0.709), and therefore obtains the lower selection score. VMD has a positive rank correlation with time but fails the increment criterion (0.581), whereas EEMD fails both the direction and increment criteria ($\rho_S = -0.977$; fraction = 0.501). The 0.65 eligibility threshold and the equal weighting in J_B are operational design choices, not estimated universal parameters. Because this known-drift check uses a single reference realization, it is interpreted as an operational consistency check rather than a stochastic robustness result. The separate cross-mode-coupling results for EEMD and CEEMDAN remain based on $R = 30$ independent seeds. The synthetic result therefore supports the suitability of CEEMDAN for separating a smooth slow component, but does not by itself establish universal superiority. The operational choice is completed on the empirical panel using the signed diagnostics in Table 3.

4.3. Operational decomposition on the empirical panel

CEEMDAN, EEMD, and VMD are applied independently to each empirical series using the method-specific calibration protocols defined in Section 3.3. Classical EMD is not carried forward beyond the synthetic benchmark. This is a scope decision: EMD does produce a residual, but the empirical comparison is restricted to the two ensemble candidates and the variational candidate selected for operational calibration.

Table 3

Properties of the signed operational slow component by asset and method. Roughness values are reported in scientific notation. NRMSE entries below 10^{-15} are written as $< 10^{-15}$; all other reconstruction errors use scientific notation with a common precision. NRMSE is at or below 10^{-6} for every method and asset, so the principal differences concern the allocation of signal content rather than numerical closure.

Asset	Method	R_{lin}^2	$\mathcal{R}_{\Delta^2}$	Frac. incr.	ρ_S	NRMSE
MSCI World	CEEMDAN	0.949	1.28×10^{-6}	0.992	1.000	6.8×10^{-7}
	EEMD	0.955	4.83×10^{-2}	0.516	1.000	$< 10^{-15}$
	VMD	0.930	1.11×10^{-4}	0.932	1.000	$< 10^{-15}$
XLE	CEEMDAN	0.489	1.96×10^{-6}	0.568	0.547	$< 10^{-15}$
	EEMD	0.943	1.40×10^{-1}	0.513	0.995	$< 10^{-15}$
	VMD	0.465	2.10×10^{-4}	0.690	0.531	$< 10^{-15}$
XLP	CEEMDAN	0.993	9.11×10^{-7}	1.000	1.000	$< 10^{-15}$
	EEMD	0.988	1.01×10^{-1}	0.514	0.999	$< 10^{-15}$
	VMD	0.984	1.34×10^{-4}	0.918	0.999	$< 10^{-15}$
XLV	CEEMDAN	0.982	1.04×10^{-6}	0.977	1.000	$< 10^{-15}$
	EEMD	0.974	2.13×10^{-1}	0.510	0.993	$< 10^{-15}$
	VMD	0.976	9.42×10^{-5}	0.821	0.994	$< 10^{-15}$
COMEX gold futures ($GC=F$)						
	CEEMDAN	0.709	1.16×10^{-6}	0.666	0.753	$< 10^{-15}$
	EEMD	0.670	8.09×10^{-2}	0.505	0.720	$< 10^{-15}$
	VMD	0.502	2.73×10^{-4}	0.604	0.760	$< 10^{-15}$

The comparison concerns both numerical closure and the allocation of the signal between oscillatory and slow components. Reconstruction fidelity is measured by

$$\text{NRMSE} = \frac{\sqrt{T^{-1} \sum_{t=1}^T (\hat{x}_t - x_t)^2}}{\sqrt{T^{-1} \sum_{t=1}^T x_t^2}} = \frac{\|\hat{x} - x\|_2}{\|x\|_2}. \quad (17)$$

The signed slow component is defined numerically as the explicit CEEMDAN residual, the implicit EEMD remainder $x - \sum_i c_i$, and the VMD DC mode. It is characterized by four descriptors: R_{lin}^2 , the coefficient of determination of a linear fit against time; $\mathcal{R}_{\Delta^2}$, the roughness based on the normalized second difference; the fraction of positive first differences; and the Spearman correlation with time, ρ_S .

Fig. 3 compares the original series with the aggregate oscillatory components obtained by each decomposition method. The lower row reports the signed method-specific slow component $s_m(t) = x(t) - o_m(t)$, where $o_m(t)$ denotes the sum of the oscillatory modes under method m . No absolute value is applied. Retaining the sign makes the negative or oscillatory EEMD remainder visually distinguishable from the CEEMDAN and VMD slow components and aligns the figure with the signed monotonicity and roughness diagnostics in Table 3.

Numerical closure is non-discriminating. Every operational pipeline reconstructs the corresponding series with NRMSE at or below 10^{-6} , and most values are at numerical precision. The methods differ substantially, however, in the roughness and directional consistency of the slow component.

According to the signed numerical slow components summarized in Table 3, CEEMDAN produces the lowest roughness for every asset, with $\mathcal{R}_{\Delta^2}$ between approximately 9×10^{-7} and 2×10^{-6} . VMD produces values near 10^{-4} , while EEMD produces values between approximately 5×10^{-2} and 2×10^{-1} . The difference spans two to five orders of magnitude despite near-equivalent reconstruction.

The signed EEMD remainder reported in Table 3 has a fraction of positive first differences close to 0.51 for every asset. It therefore rises and falls in almost equal proportion, even when its long-run trajectory is highly correlated with time. This directional behaviour is now visible in the signed slow-component panels of Fig. 3, which distinguish the oscillatory EEMD remainder from the smoother CEEMDAN and VMD slow components. Linear fit and rank correlation capture the overall direction of the component, whereas roughness and first-difference balance capture its local oscillation.

The signed diagnostics in Table 3 indicate that CEEMDAN produces a comparatively smooth and strongly monotone slow component for MSCI World, XLP, and XLV. The distinction is less direct for XLE and COMEX gold futures, whose slow representations are less linear under every method. On XLE, EEMD has a larger R_{lin}^2 than CEEMDAN, and VMD has a larger fraction of positive increments. CEEMDAN is retained there because its roughness is approximately two orders of magnitude below VMD and several orders below EEMD. For COMEX gold futures, CEEMDAN also has the lowest roughness and the highest fraction of positive increments, although VMD has a slightly larger Spearman correlation with time.

The empirical comparison therefore does not identify a universally optimal decomposition. Under the adopted method-specific calibration and the criterion required for the subsequent representation audit, CEEMDAN provides the most consistently smooth separation between oscillatory and slow content over the five-asset panel. It is consequently retained as the operational representation for the remaining stages.

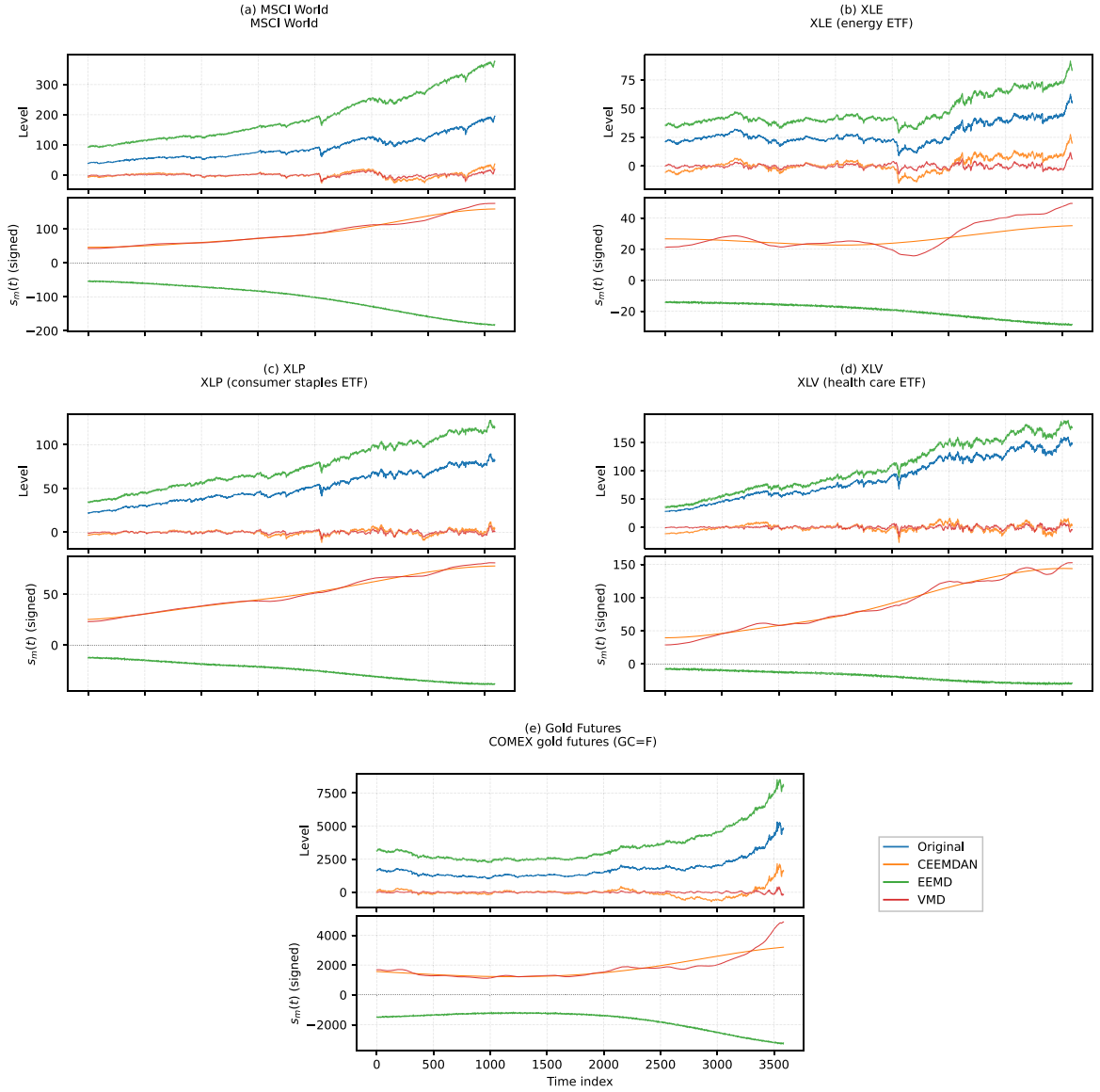


Fig. 3. Operational decomposition of the empirical panel. Each asset is shown with two aligned rows. The upper row compares the original closing-price series with the aggregate oscillatory components obtained from CEEMDAN, EEMD, and VMD. For CEEMDAN and EEMD, the aggregate is the sum of the extracted oscillatory IMFs; for VMD, it is the sum of the retained non-DC modes. The lower row shows the signed method-specific slow component $s_m(t) = x(t) - o_m(t)$ (CEEMDAN: explicit residual; EEMD: implicit remainder after summing oscillatory IMFs; VMD with $DC = 1$: DC mode). Signs are retained. Vertical scales differ across assets and are not intended for cross-asset amplitude comparison.

4.4. Reduced FastICA representation

The retained CEEMDAN decompositions produce between eight and nine oscillatory modes per asset, excluding the slow residual. FastICA reduces each native oscillatory panel to an asset-specific set of decorrelated coordinates. The operational dimensionalities are $k = 4$ for MSCI World and XLP and $k = 5$ for XLE, XLV, and COMEX gold futures, as specified in Section 3.4.

The operational network inventory retains the seed-42 fit. Robustness is assessed by repeating FastICA for all five assets with twenty initializations (seeds 42–61) at neighbouring ranks $k = 3, 4, 5$ and at full rank $k = p$ ($p = 8$ for MSCI World, XLE, XLP, and XLV; $p = 9$ for COMEX gold futures), giving 400 fits in total (Tables 4 and 5). All 300 reduced-rank runs converge. Two of the 100 full-rank runs, both for XLV, reach the iteration limit; their returned reconstructions are nevertheless numerically exact ($R^2 = 1.000$). For every fixed asset–rank combination, mean column-wise R^2 and the relative Frobenius error are stable across seeds to machine precision, and the maximum absolute off-diagonal score correlation remains at numerical round-off. At the operational ranks, for example, R^2 equals 0.636 for MSCI World ($k = 4$), 0.734 for XLE ($k = 5$), 0.618 for XLP ($k = 4$), 0.743 for XLV ($k = 5$),

Table 4

FastICA robustness across random initializations. For each asset and rank k , entries are mean $\pm$ sd over N seeds. $k = p$ is the full-rank control (linear recombination without dimension reduction).

Asset	k	R^2 mean	Rel. Frobenius	max corr Z	Conv. frac.
MSCI World	3	$0.508 \pm 4.93 \times 10^{-17}$	$0.662 \pm 5.67 \times 10^{-17}$	1.53×10^{-15}	1.00
MSCI World	4*	$0.636 \pm 4.56 \times 10^{-17}$	$0.589 \pm 4.56 \times 10^{-17}$	1.77×10^{-15}	1.00
MSCI World	5	$0.759 \pm 4.56 \times 10^{-17}$	$0.567 \pm 3.42 \times 10^{-17}$	2.34×10^{-15}	1.00
MSCI World	$p = 8$	1.000 ± 0.000	$6.90 \times 10^{-16} \pm 1.77 \times 10^{-16}$	2.86×10^{-15}	1.00
XLE	3	$0.485 \pm 2.61 \times 10^{-17}$	$0.897 \pm 1.04 \times 10^{-16}$	7.56×10^{-16}	1.00
XLE	4	0.614 ± 0.000	$0.685 \pm 1.13 \times 10^{-16}$	1.05×10^{-15}	1.00
XLE	5*	$0.734 \pm 4.56 \times 10^{-17}$	$0.247 \pm 3.40 \times 10^{-17}$	2.16×10^{-15}	1.00
XLE	$p = 8$	1.000 ± 0.000	$8.17 \times 10^{-16} \pm 1.91 \times 10^{-16}$	2.68×10^{-15}	1.00
XLP	3	$0.496 \pm 3.26 \times 10^{-17}$	$0.767 \pm 6.52 \times 10^{-17}$	9.46×10^{-16}	1.00
XLP	4*	$0.618 \pm 3.42 \times 10^{-17}$	$0.754 \pm 6.52 \times 10^{-17}$	1.32×10^{-15}	1.00
XLP	5	0.734 ± 0.000	$0.488 \pm 5.43 \times 10^{-17}$	1.97×10^{-15}	1.00
XLP	$p = 8$	1.000 ± 0.000	$8.45 \times 10^{-16} \pm 1.87 \times 10^{-16}$	3.23×10^{-15}	1.00
XLV	3	0.495 ± 0.000	0.729 ± 0.000	9.18×10^{-16}	1.00
XLV	4	$0.630 \pm 5.43 \times 10^{-17}$	0.628 ± 0.000	6.11×10^{-15}	1.00
XLV	5*	$0.743 \pm 4.56 \times 10^{-17}$	$0.503 \pm 4.93 \times 10^{-17}$	8.40×10^{-14}	1.00
XLV	$p = 8$	1.000 ± 0.000	$1.70 \times 10^{-15} \pm 1.62 \times 10^{-16}$	1.32×10^{-14}	0.90
Gold Futures	3	$0.501 \pm 6.71 \times 10^{-17}$	0.533 ± 0.000	1.08×10^{-15}	1.00
Gold Futures	4	$0.614 \pm 4.93 \times 10^{-17}$	$0.456 \pm 1.71 \times 10^{-17}$	2.14×10^{-15}	1.00
Gold Futures	5*	0.725 ± 0.000	$0.452 \pm 2.03 \times 10^{-17}$	1.74×10^{-15}	1.00
Gold Futures	$p = 9$	1.000 ± 0.000	$7.87 \times 10^{-16} \pm 1.82 \times 10^{-16}$	6.13×10^{-15}	1.00

* operational k used in the manuscript for that asset. Off-diagonal Pearson correlations among FastICA scores are at numerical roundoff for every seed in this run.

Table 5

Transformation versus truncation: operational $k < p$ compared with the full-rank control $k = p$. Values are mean across seeds. At $k = p$ the relative Frobenius error is at numerical roundoff ($\sim 10^{-15}$), so residual error at k^* is attributable to truncation rather than to the ICA transform itself.

Asset	p	k^*	$R^2(k^*)$	$R^2(k = p)$	Frob(k^*)	Frob($k = p$)
MSCI World	8	4	0.636	1.000	0.589	6.90×10^{-16}
XLE	8	5	0.734	1.000	0.247	8.17×10^{-16}
XLP	8	4	0.618	1.000	0.754	8.45×10^{-16}
XLV	8	5	0.743	1.000	0.503	1.70×10^{-15}
Gold Futures	9	5	0.725	1.000	0.452	7.87×10^{-16}

and 0.725 for gold futures ($k = 5$). Mean R^2 increases monotonically from $k = 3$ to $k = 5$ for every asset. At $k = p$, $R^2 = 1.000$ and the relative Frobenius error is of order 10^{-15} for every asset, showing that the reconstruction gap at $k < p$ results from truncating coordinates rather than from the invertible ICA recombination itself. Component order and sign remain unidentified. The multi-seed experiment establishes robustness of reconstruction and numerical decorrelation; it does not establish seed-wise invariance of every graph descriptor, because the full graph inventories were not rebuilt for each initialization.

Mean column-wise reconstruction R^2 at the operational ranks therefore ranges from 0.62 for XLP to 0.74 for XLV, and relative Frobenius error from 0.25 for XLE to 0.75 for XLP. These diagnostics quantify the information discarded by reducing p native modes to $k < p$ coordinates. They are not interpreted as direct measures of topological preservation.

Fig. 4 shows the retained coordinates. The panels contain both high-frequency fluctuations and slower oscillatory structures, but the coordinate index has no common meaning across assets. FastICA component order and sign are not identified; consequently, Z_j for one asset is not assumed to correspond to Z_j for another. The subsequent network audit therefore compares descriptor distributions within each representation and encoder rather than matching components by numerical index.

4.5. Representation-dependent network topology

Natural visibility, horizontal visibility, and recurrence networks are constructed from both the native oscillatory IMFs and the retained FastICA coordinates. The slow residual is excluded from both graph inventories, so the comparison concerns two representations of the oscillatory content.

The native panel contains 41 oscillatory IMFs across the five assets and therefore generates 123 graphs. The reduced panel contains 23 FastICA coordinates and generates 69 graphs. The inventory ratio is

$$q_{\text{graph}} = \frac{69}{123} \approx 0.56. \quad (18)$$

Total node count decreases in nearly the same proportion, from approximately 4.40×10^5 to 2.47×10^5 , because every visibility graph contains the full aligned sample and recurrence graphs lose only the final embedding observations.

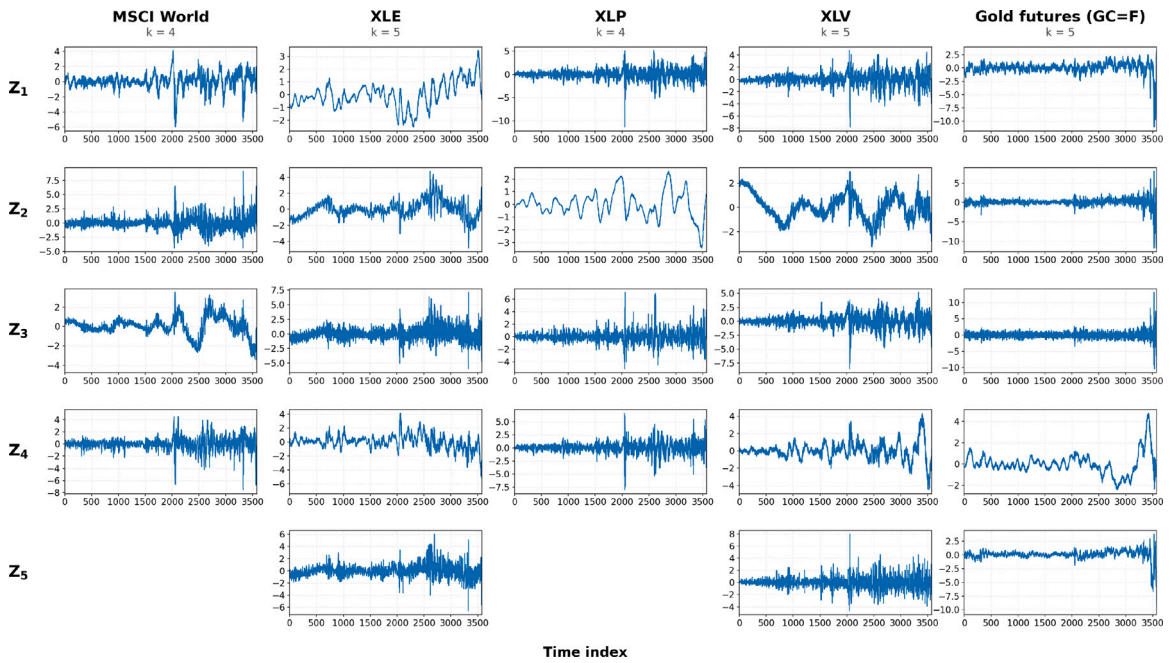


Fig. 4. Reduced FastICA representations obtained from the oscillatory CEEMDAN panels. Columns correspond to MSCI World, XLE, XLP, XLV, and COMEX gold futures (GC=F); rows show the retained coordinates $Z_1, \dots, Z_5$. The operational dimensionality k is displayed for each asset, and cells beyond the retained dimension are left empty. Coordinate indices are internal to each asset-specific fit: their order and sign are not identified and do not establish correspondence between assets. Vertical scales are displayed independently to emphasize temporal structure and are not used for cross-panel amplitude comparison. The CEEMDAN slow component is excluded from the FastICA representation and from the network audit.

Table 6

Mean connectivity statistics on the *native* oscillatory IMF panel by asset and network construction (residue excluded). Edge counts and recurrence component counts are averaged over the p oscillatory IMFs. Columns match Table 10 (reduced FastICA panel) with p in place of k . Recurrence $\tilde{N}_{cc}$ values come from the original graph inventory. Panel means under the rebuilt local-distance protocol used for RR/Theiler sensitivity (Table 12) can differ slightly because ϵ_{rec} is re-estimated from a finite sample (e.g. MSCI World: 853 versus 871.0).

Asset	p	HVG $ E $	NVG $ E $	REC $ E $	$\tilde{D}_{HVG}$	$\tilde{N}_{cc,REC}$
MSCI World	8	6927	238 728	17 233	302	853
XLE	8	6845	379 350	17 218	380	798
XLP	8	6906	243 966	16 648	338	836
XLV	8	6866	208 743	16 950	329	901
Gold futures	9	6965	476 992	17 493	333	610

Total edge count decreases more sharply, from approximately 1.38×10^7 in the native inventory to 9.7×10^5 in the reduced inventory, so that the raw inventory ratio $q_E \approx 0.07$ (Eq. (19)). This ratio combines inventories with different numbers of channels and is dominated by a small number of very dense low-frequency native NVGs. It is retained only as a descriptive inventory statistic and is not used to support structural preservation, compression quality, or computational savings; observed execution and memory costs remain unmeasured.

$$q_E = \frac{E_{\text{tot, reduced}}}{E_{\text{tot, native}}} \approx 0.07. \quad (19)$$

Symmetric documentation is provided by native-panel connectivity and clustering tables parallel to the reduced-coordinate summaries (Tables 6–7), together with per-mode and summary values for the two lowest-frequency NVGs in each asset (Tables 8 and 9). Formal channel-level contrasts appear in Table 11.

Table 10 summarizes connectivity on the reduced coordinates. Visibility graphs preserve the complete aligned sample length. Recurrence networks are shorter by the $(d-1)\tau$ observations consumed by delay embedding, approximately thirty observations on average.

Horizontal visibility produces the most homogeneous connectivity across the reduced panel. Mean edge counts range from approximately 7.1×10^3 to 7.2×10^3 , corresponding to densities near 1.1×10^{-3} and mean degrees close to four. Every HVG instance

Table 7

Mean local clustering coefficients on the *native* oscillatory IMF panel by asset and network construction (residue excluded). Each value is averaged over the p IMFs. Columns match Table 14.

Asset	p	$\bar{C}_{\text{HVG}}$	$\bar{C}_{\text{NVG}}$	$\bar{C}_{\text{REC}}$
MSCI World	8	0.53	0.64	0.31
XLE	8	0.51	0.68	0.34
XLP	8	0.52	0.67	0.34
XLV	8	0.52	0.68	0.32
Gold futures	9	0.53	0.68	0.38

Table 8

Native-panel low-frequency NVG edge counts by mode. Low-frequency modes are the two highest-index oscillatory IMFs (and IMF₉ when present). Asset-level mean, median and range are computed over those modes only. An entry of *n.a.* in the $\bar{C}$ column indicates that the clustering coefficient was not computed in the original graph inventory because of computational cost. These missing clustering values are excluded from clustering summaries and contrasts and are never treated as zero.

Asset	Mode	$ E $	Density	$\bar{C}$
MSCI World	IMF_7	417 517	6.49×10^{-02}	0.483
MSCI World	IMF_8	1162 388	1.81×10^{-01}	0.607
XLE	IMF_7	682 912	1.06×10^{-01}	<i>n.a.</i>
XLE	IMF_8	1 949 471	3.03×10^{-01}	<i>n.a.</i>
XLP	IMF_7	436 054	6.78×10^{-02}	<i>n.a.</i>
XLP	IMF_8	1 137 231	1.77×10^{-01}	<i>n.a.</i>
XLV	IMF_7	501 317	7.79×10^{-02}	<i>n.a.</i>
XLV	IMF_8	817 660	1.27×10^{-01}	<i>n.a.</i>
Gold futures	IMF_8	1 252 601	1.95×10^{-01}	<i>n.a.</i>
Gold futures	IMF_9	2 235 125	3.48×10^{-01}	<i>n.a.</i>

Table 9

Native low-frequency NVG edge-count summaries (mean, median, range) over the two highest-index oscillatory IMFs per asset (IMF₈–IMF₉ for gold). These modes drive the extreme native NVG edge totals contrasted with Table 10.

Asset	Modes	Mean $ E $	Median $ E $	Min $ E $	Max $ E $
MSCI World	IMF ₇ , IMF ₈	789 952	789 952	417 517	1162388
XLE	IMF ₇ , IMF ₈	1316 192	1316 192	682 912	1949471
XLP	IMF ₇ , IMF ₈	786 642	786 642	436 054	1137231
XLV	IMF ₇ , IMF ₈	659 488	659 488	501 317	817 660
Gold futures	IMF ₈ , IMF ₉	1743 863	1743 863	1252 601	2235125

Table 10

Mean connectivity statistics on the retained FastICA coordinates by asset and network construction. Edge counts and recurrence component counts are averaged over the k coordinates. Recurrence parameters and thresholds are recalibrated independently for every channel according to Section 3.5; the networks therefore do not share a common absolute radius or recurrence rate. Reported diameters apply to HVG only in this summary because all recurrence networks are disconnected.

Asset	k	HVG $ E $	NVG $ E $	REC $ E $	$\bar{D}_{\text{HVG}}$	$\bar{N}_{\text{cc,REC}}$
MSCI World	4	7161	14 362	16 918	40	1428
XLE	5	7110	20 746	17 256	73	897
XLP	4	7089	27 845	16 207	96	1140
XLV	5	7141	13 917	16 630	44	1408
Gold futures	5	7124	15 326	17 338	49	798

is connected. These values remain of the same order as the native oscillatory mean, about 6.9×10^3 edges (native asset means 6845–6965; Table 6). The native channel mean is slightly lower than the FastICA mean (difference -221.9 , 95% CI $[-340.4, -116.9]$, $d = -0.764$, $p = 0.0036$; Table 11), as is mean degree (difference -0.122 , $p = 0.0047$). Aggregate HVG connectivity is therefore similar in scale after compression, but not invariant or unchanged.

NVG edge counts change much more under the same representation shift. Asset-level means on the reduced coordinates range from approximately 1.4×10^4 for XLV to 2.8×10^4 for XLP, whereas native means range from 2.1×10^5 to 4.8×10^5 (Table 6). At channel level, the native-minus-reduced difference in $\log_{10} |E|$ is 0.706 (95% CI $[0.476, 0.946]$, $d = 1.158$, $p = 0.0002$). Among the

Table 11

Formal contrasts between native IMF and FastICA-reduced graph descriptors (oscillatory channels only; residue excluded). Difference = mean(native)–mean(reduced). 95% bootstrap CIs ($N = 5000$); two-sided permutation p ($N = 10,000$). Cohen's d uses the pooled standard deviation.

Contrast	n_{nat}	n_{ica}	Diff	95% CI	d	p_{perm}
NVG $\log_{10} E $	41	23	0.706	[0.476, 0.946]	1.158	0.0002
HVG $ E $	41	23	−221.9	[−340.4, −116.9]	−0.764	0.0036
HVG clustering	41	23	−0.109	[−0.136, −0.081]	−1.514	0.0001
NVG clustering	32	23	−0.064	[−0.095, −0.033]	−1.006	0.0008
REC components	41	23	−325.9	[−558.8, −93.8]	−0.664	0.0130
REC $ E $	41	23	221.4	[−97.7, 541.4]	0.357	0.1783
HVG mean degree	41	23	−0.122	[−0.184, −0.064]	−0.753	0.0047
NVG mean degree	41	23	164.8	[84.3, 257.9]	0.696	0.0065

Table 12

Recurrence fragmentation across assets and representations (oscillatory modes only; residue excluded). Entries are mean $\tilde{N}_{\text{cc}}$ under protocols A/B/C (C uses $W = \tau$). Under A the realized RR is also reported.

Asset	Rep.	RR _A	$\tilde{N}_{\text{cc}}^A$	$\tilde{N}_{\text{cc}}^B$	$\tilde{N}_{\text{cc}}^C$
MSCI World	FastICA	0.0027	1427.8	316.0	387.5
MSCI World	native	0.0027	871.0	142.5	478.2
Gold Futures	FastICA	0.0028	798.2	165.8	316.0
Gold Futures	native	0.0028	610.2	117.7	402.3
XLE	FastICA	0.0028	896.6	117.4	284.4
XLE	native	0.0027	797.6	124.8	377.1
XLP	FastICA	0.0026	1139.5	258.8	346.8
XLP	native	0.0026	836.4	143.2	470.6
XLV	FastICA	0.0026	1408.0	282.6	369.4
XLV	native	0.0027	901.4	172.0	368.2

two slowest oscillatory IMFs per asset, individual NVG edge counts range from 417,517 to 2,235,125 (Tables 8 and 9). Dense NVGs on those smooth, slowly varying modes are consistent with the established line-of-sight geometry of natural visibility (Section 2.1); the empirical finding is the disappearance of that extreme tail after scales are recombined and truncated by FastICA.

Recurrence edge counts on the reduced coordinates lie between approximately 1.6×10^4 and 1.7×10^4 , but their connectivity differs qualitatively from both visibility constructions. Under the adopted local-distance calibration, recurrence networks constructed from the reduced coordinates split into hundreds or more than one thousand connected components. Their diameters are therefore undefined. Mean degree remains near nine or ten, consistently with the radius derived from the tenth percentile of the pooled distances to the 99 nearest non-self neighbours.

Both representations are highly fragmented under this rule. The reduced panel has more connected components (native-minus-reduced difference −325.9, 95% CI [−558.8, −93.8], $d = -0.664$, $p = 0.013$), whereas recurrence edge counts are compatible with no difference between representations (difference 221.4, 95% CI [−97.7, 541.4], $p = 0.178$; Table 11). Delay, embedding dimension, and radius are estimated separately for each channel, and the realized recurrence rate is only 0.24%–0.32%. Setting a global rate of $RR = 5\%$ reduces mean component counts in every panel; for MSCI World they fall from 871.0 to 142.5 in the native representation and from 1427.8 to 316.0 after FastICA (Table 12). Under this fixed-rate construction with $W = 0$, three native MSCI modes become connected (Table 13). In a paired analysis of the nine MSCI native channels (eight oscillatory IMFs and the residue), the mean A-minus-B reduction is 647.6 components (95% bootstrap CI [411.0, 867.7]; sign-flip permutation $p = 0.0074$). Applying a Theiler window $W = \tau$ raises the oscillatory-panel means to 478.2 (native) and 387.5 (FastICA). Extending the window to $W = 2\tau$ and $W = 5\tau$ leaves mid- and low-frequency modes strongly fragmented: for MSCI World native at $RR = 5\%$, IMF₆ changes from 1 component at $W = 0$ to 606, 600, and 593 at $W = \tau$, 2τ , and 5τ , while IMF₈ changes from 1 to 639, 680, and 881 (Table 13). Fast modes change little. Fragmentation is therefore strongly conditional on threshold and temporal exclusion; it does not disappear completely, because excluding temporally adjacent states restores hundreds of components in several mid- and low-frequency modes.

Table 14 shows a similarly encoder-dependent pattern. HVG clustering on the reduced coordinates is concentrated in a narrow range, with asset means between 0.62 and 0.64. Channel-level values lie approximately between 0.59 and 0.66.

Across the 41 native oscillatory IMFs, mean HVG clustering decreases with extraction order (Spearman $\rho = -0.97$, 95% bootstrap CI [−0.98, −0.95]) and averages 0.520 (high values near 0.66 on the fastest modes, down to approximately 0.33 on the slowest; Table 7). After FastICA, coordinates mix scales without a fast-to-slow ordering, so the native gradient cannot be recovered from the Z_j index. The native-minus-reduced difference in mean clustering is −0.109 ($d = -1.514$, $p = 0.0001$; Table 11), and the coefficient of variation falls from 0.170 on native IMFs to 0.031 after FastICA. Preservation of the order of magnitude of HVG edge count should therefore not be confused with preservation of the scale-ordered clustering interpretation.

NVGs have the highest mean local clustering on the retained coordinates, between 0.70 and 0.75 across assets. Median node-level coefficients are approximately 0.76–0.80, except for one XLP coordinate with a lower value near 0.57. Recurrence networks have lower and more heterogeneous mean clustering, between 0.23 and 0.34 at asset level and approximately 0.21–0.59 across individual

Table 13

Theiler-window sensitivity at fixed $RR = 5\%$ (MSCI World, native CEEM-DAN). Columns are $W = m\tau$. Entries are N_{cc} .

Mode	$W = 0$	$W = \tau$	$W = 2\tau$	$W = 5\tau$
IMF ₁	198	201	201	205
IMF ₂	130	130	130	132
IMF ₃	439	442	441	441
IMF ₄	357	580	580	572
IMF ₅	13	531	531	547
IMF ₆	1	606	600	593
IMF ₇	1	697	665	625
IMF ₈	1	639	680	881
Residue	1	1	1	831

Table 14

Mean local clustering coefficients on the retained FastICA coordinates by asset and network construction. Each value is averaged over the k coordinates retained for the corresponding asset.

Asset	k	$\hat{C}_{HVG}$	$\hat{C}_{NVG}$	$\hat{C}_{REC}$
MSCI World	4	0.63	0.75	0.23
XLE	5	0.62	0.72	0.34
XLP	4	0.62	0.70	0.32
XLV	5	0.63	0.75	0.23
Gold futures	5	0.64	0.74	0.32

channels. Their low median node-level coefficients are consistent with the large number of low-degree vertices in the fragmented recurrence graphs.

The combined audit yields two main results. First, replacing the native IMF panel with a reduced linear representation does not affect all network constructions in the same way (Table 11). HVG edge counts retain their order of magnitude, with a modest shift in the mean; HVG clustering rises and tightens, and the native ranking by IMF index cannot be read on Z_j ; reduced-coordinate NVGs never reach the edge counts of the slowest native IMFs; and recurrence graphs remain highly fragmented, with more components after FastICA under the original rule, while recurrence edge counts do not separate the representations.

Second, the three mappings are not interchangeable. Applied to the same coordinates, HVG produces sparse and connected graphs with comparatively homogeneous descriptors; NVG produces denser graphs with high local clustering; and the recurrence construction produces fragmented networks whose topology is conditional on phase-space reconstruction and threshold calibration. The inferred network structure must therefore be interpreted jointly with both the representation supplied to the mapping and the encoding rule used to construct the graph.

5. Discussion and conclusions

5.1. Representation dependence of the inferred topology

The central question of this study is whether the topology inferred from a multiscale time series can be interpreted independently of the representation and network construction from which it is obtained. The decomposition benchmark supports the selection of an operational multiscale representation, but the main comparison concerns the networks constructed from the native oscillatory modes and from their reduced FastICA representation. The results show that representation and encoding choices contribute jointly to the observed graph structure. Consequently, the resulting topology should not be treated as an intrinsic property of the original financial series alone.

FastICA replaces the native IMF panel with a reduced linear representation containing four or five coordinates per asset. Because $k < p$, this operation combines linear recombination with dimensionality reduction and cannot be interpreted as a lossless change of coordinates. Mean reconstruction R^2 ranges from 0.62 to 0.74, confirming that part of the native panel is discarded. These reconstruction values, however, do not predict how the network descriptors will change.

The network audit shows that the effect of compression depends on the encoding rule. After reduction, HVG edge counts and degrees remain close to the native oscillatory values, and every reduced-coordinate HVG is connected. There is nevertheless a small rise in mean $|E|$ on the reduced panel (native-minus-reduced difference -221.9 ; the confidence interval excludes zero; Table 11). This statement concerns averages rather than node-for-node graph identity, and it does not claim that the two samples are equal.

NVG edge counts react much more strongly to the same change of representation. The slowest native IMFs can produce NVGs with close to 10^6 edges; the FastICA coordinates do not. Native channels sit 0.706 decades above the reduced ones in $\log_{10} |E|$ (95% CI [0.476, 0.946], $d = 1.158$). Dense low-frequency NVGs are consistent with the established line-of-sight geometry of natural visibility (Section 2.1); the empirical finding is the removal of that extreme tail after scale recombination and truncation. NVG depends on the shape of the series in the time–amplitude plane, so mixing scales inside one coordinate is sufficient to change the graph.

Recurrence graphs behave differently again. With the per-channel calibration, native and reduced series both split into hundreds or more than a thousand components. The FastICA side has the larger mean component count (Table 11). Delay, embedding dimension, and radius are fitted separately on each channel, and the realized recurrence rate is approximately 0.3%. A fixed RR of 5% reduces that splitting on all five assets, native and FastICA; Theiler windows $W = m\tau$ restore components on slower modes (Table 12). Recurrence topology is therefore protocol-dependent rather than an intrinsic property of the modes or of FastICA alone.

Taken together, these findings show that representation reduction does not act uniformly across graph constructions. An HVG descriptor that barely moves can still change substantially under NVG or recurrence. Whether a representation is “preserved” depends on which encoder and which statistic are examined. Preservation must therefore be assessed with respect to a specified encoder and descriptor rather than inferred from reconstruction accuracy alone.

5.2. Complementarity of the network encodings

The three mappings provide structurally different descriptions of the same channels. Horizontal visibility produces sparse, connected graphs with mean degree close to four and comparatively homogeneous clustering. Natural visibility produces denser graphs with higher local clustering and greater sensitivity to low-frequency waveform geometry. Recurrence networks encode phase-space proximity under the adopted embedding and threshold calibration, but are highly fragmented and do not support global path summaries.

A further distinction appears in horizontal-visibility clustering. Across the 41 native oscillatory IMFs, mean HVG clustering decreases with extraction order (Spearman $\rho = -0.97$, 95% bootstrap CI $[-0.98, -0.95]$), from approximately 0.66 on the fastest modes to approximately 0.33 on the slowest, with panel mean 0.520. After FastICA, asset-level means lie between 0.62 and 0.64 and the channel mean is 0.628; the native-minus-reduced difference is -0.109 ($d = -1.514$, $p = 0.0001$). Because FastICA coordinates are mixtures without a fast-to-slow ordering, the native scale gradient cannot be recovered from the Z_j index. This pattern is consistent with the native HVG descriptors retaining information associated with the characteristic scale of the encoded mode, whereas each reduced coordinate combines contributions from several native scales. It should not be interpreted as a universal one-to-one relationship between HVG clustering and physical frequency, because IMF ordering and the resulting descriptors are both signal- and decomposition-dependent.

The three encoders are therefore complementary rather than interchangeable. Agreement between them cannot be assumed, and disagreement does not necessarily indicate an error. Each construction emphasizes a different relation: ordinal visibility, geometric visibility, or reconstructed-state proximity. A substantive interpretation of a time-series network must identify which relation was encoded and which upstream representation supplied the channel.

5.3. Role of the operational decomposition

The synthetic benchmark does not identify a universally superior decomposition method. EMD, EEMD, CEEMDAN, and VMD exchange relative positions across close-frequency tones, chirps, intermittent bursts, noisy superpositions, and the multi-component stress signal. Replication of EEMD and CEEMDAN also shows that a single stochastic realization can provide an incomplete description of their behaviour.

Numerical closure is similarly non-discriminating on the empirical panel. CEEMDAN, EEMD, and VMD reconstruct the five series with NRMSE at or below 10^{-6} , but allocate slow and oscillatory content differently. The roughness of the signed operational slow component differs by several orders of magnitude. Under the adopted method-specific calibration, CEEMDAN produces the lowest roughness for every asset and the most consistently monotone slow component across the panel. EEMD produces a much rougher remainder whose fraction of positive first differences remains close to one half, while VMD occupies an intermediate position.

These outcomes support CEEMDAN as the operational representation used in the network audit, but they do not establish intrinsic superiority. The synthetic benchmark characterizes method behaviour and checks the slow-component criterion on a known drift; the empirical panel determines the retained representation. The methods are calibrated through different search spaces and objectives, and classical EMD is not carried to the empirical comparison. The result is therefore a selection within the implemented experimental protocol, not a general ranking of adaptive decomposition algorithms.

5.4. Limitations

Remarking that some limitations bound the interpretation of the results.

First, FastICA is non-convex and component order and sign remain unidentified. The multi-seed experiment (seeds 42–61) and neighbouring ranks $k = 3, 4, 5$, together with the full-rank control $k = p$, establish that reconstruction R^2 and numerical decorrelation are stable across initializations and that residual error at the operational $k < p$ is due to truncation rather than to the invertible ICA recombination (Tables 4 and 5). The selected dimensionalities remain operational rather than uniquely identified optima. Graph inventories were not rebuilt for every seed, so seed-wise topological invariance is not claimed.

Second, formal channel-level contrasts between native and reduced descriptor distributions are now reported (Table 11), including Cohen’s d , bootstrap confidence intervals, and permutation p -values. Because several channels come from the same asset, these analyses provide operational support for the observed distributions rather than independent population trials or proof of generality beyond the panel. Nine computationally demanding native NVGs have unavailable clustering coefficients because these

calculations were omitted from the original graph inventory. These missing values are excluded from clustering summaries and contrasts and prevent a complete clustering comparison for the affected native graphs; they are never treated as zero.

Third, recurrence-network topology remains conditional on reconstruction and threshold choices. Mutual information and false nearest neighbours select τ and d , and the primary audit uses the local-distance radius with $W = 0$. Fixed-rate (RR = 5%) and Theiler-window ($W \in \{0, \tau, 2\tau, 5\tau\}$) checks show that fragmentation is strongly protocol-dependent (Tables 12 and 13). Sensitivity to percentile, neighbourhood size, sampling seed, maximum lag, or alternative embedding rules beyond those checks is not claimed.

Fourth, the empirical panel still contains five instruments over one common sampling interval. The new analyses were performed on all five series, but the market universe has not been enlarged, and statistical representativeness across asset classes, frequencies, windows, or regimes is not claimed. The BDS results remain descriptive because the statistic is applied to residuals from AR(1) models fitted to non-stationary price levels. Rolling price-level dispersion and local-mean crossings remain operational descriptors rather than formal admissibility conditions for decomposition.

Fifth, the decomposition comparison uses method-specific calibration. The candidate spaces, objectives, and degree of asset-specific tuning differ between CEEMDAN, EEMD, and VMD. This design is appropriate for selecting an operational representation but does not constitute a symmetric algorithmic benchmark. Classical EMD is evaluated only on the controlled signals. The synthetic benchmark characterizes method behaviour and checks the slow-component criterion on a known drift; the empirical panel determines the representation retained for the network audit.

Sixth, all networks are undirected and unweighted. Weighted visibility constructions could preserve amplitude information not represented in the current adjacency matrices [45], but they would introduce an additional design factor. Directed, multilayer, and temporal network constructions remain outside the present scope.

Finally, the aggregate edge-count ratio $q_E \approx 0.07$ is retained only as a descriptive inventory statistic. It is dominated by the dense NVGs of the lowest-frequency native modes and combines inventories with different numbers of graphs. No execution-time benchmark, memory analysis, or distributed implementation is reported.

5.5. Future work

The multi-seed reconstruction diagnostics and the $k = p$ control now separate truncation from linear recombination at the signal level. The most immediate remaining extension is therefore graph-level multi-seed validation: rebuilding the full NVG, HVG, and recurrence inventories under alternative FastICA initializations to test whether the reported topological contrasts are seed-invariant. Comparing truncated ICA with PCA and related compressors, and complementing decorrelation diagnostics with higher-order dependence measures, would further clarify the compression stage without presuming statistical independence.

A second direction is broader external replication. Additional asset classes, sampling frequencies, and non-overlapping historical windows would permit a formal assessment of whether the observed representation effects persist across regimes. A rolling-window design could determine whether descriptor stability changes during crises or transitions. Such an extension should retain chronological alignment while distinguishing within-asset representation effects from genuine cross-asset relations.

A third direction is further sensitivity analysis of network construction beyond the RR and Theiler checks already reported. This includes alternative distance norms, embedding rules, neighbourhood sizes, and percentile thresholds for recurrence networks, and weighted or directed visibility variants for amplitude and temporal-asymmetry information.

The resulting graph inventories could subsequently be supplied to classification, regime-detection, or forecasting models. Such downstream experiments would address a question not tested in the present study: whether differences in graph structure translate into differences in predictive or discriminative information. No such benefit is assumed from the current topological results alone.

5.6. Concluding remarks

This study shows that the topology inferred from a multiscale financial time series depends jointly on representation, encoder, and calibration, with the magnitude and direction of that dependence quantified where the data permit. Replacing native CEEMDAN modes with a reduced FastICA representation preserves the order of magnitude of HVG edge counts while raising mean HVG clustering and destroying the native scale-ordered clustering gradient; it removes the extreme natural-visibility edge growth associated with the lowest-frequency native modes, an effect expected from NVG geometry; and, under the local-distance recurrence rule, it increases connected-component counts, although that fragmentation is strongly conditional on recurrence rate and Theiler-window choices.

The contribution is not a new decomposition method, dimensionality-reduction algorithm, or network encoder. It is a controlled audit of how established operations interact. The results indicate that decomposition, compression, and network construction cannot be interpreted as independent preprocessing choices. Graph descriptors should therefore be reported together with the scale representation, coordinate system, and encoding parameters from which they were obtained.

Financial series provide the empirical test bed, but the methodological implication is broader. From a statistical-physics perspective, network observables derived from non-stationary signals are properties of the complete representation pipeline rather than of the observed trajectory alone. Accordingly, their interpretation requires explicit consideration of the scale decomposition, coordinate representation, and network mapping through which they were constructed.

CRediT authorship contribution statement

Agustín M. de los Riscos: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Julio E. Sandubete:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Ana Lazcano:** Writing – review & editing, Supervision, Resources, Methodology, Data curation, Conceptualization.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the authors used ChatGPT (OpenAI) and Claude (Anthropic) to assist with language editing and document formatting. After using these tools and services, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This work forms part of the research project Responsible AI: Distributed and Efficient Management of Temporal Dependence in Large Time Series Datasets, funded by Universidad Francisco de Vitoria, Madrid, under its 2025 Call for Research Projects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data files, source code, configuration files, and scripts used to generate the analyses, tables, and figures are publicly available at <https://github.com/agusriscos/GraphEMD>.

References

- [1] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuño, From time series to complex networks: The visibility graph, *Proc. Natl. Acad. Sci.* 105 (13) (2008) 4972–4975, <http://dx.doi.org/10.1073/pnas.0709247105>.
- [2] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: Exact results for random time series, *Phys. Rev. E* 80 (4) (2009) 046103.
- [3] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, *Europhys. Lett.* 4 (9) (1987) 973–977, <http://dx.doi.org/10.1209/0295-5075/4/9/004>.
- [4] R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Complex network analysis of recurrences, *New J. Phys.* 12 (3) (2010) 033025, <http://dx.doi.org/10.1088/1367-2630/12/3/033025>.
- [5] V.F. Silva, M.E. Silva, P. Ribeiro, F. Silva, Time series analysis via network science: concepts and algorithms, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 11 (6) (2021) e1404, <http://dx.doi.org/10.1002/widm.1404>.
- [6] Y. Zou, R.V. Donner, N. Marwan, J.F. Donges, J. Kurths, Complex network approaches to nonlinear time series analysis, *Phys. Rep.* 787 (2019) 1–97, <http://dx.doi.org/10.1016/j.physrep.2018.12.004>.
- [7] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, *Quant. Finance* 1 (2) (2001) 223–236.
- [8] N. Nava, T. Di Matteo, T. Aste, Financial time series forecasting using empirical mode decomposition and support vector regression, *Risks* 6 (1) (2018) 7, <http://dx.doi.org/10.3390/risks6010007>.
- [9] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 454 (1971) (1998) 903–995, <http://dx.doi.org/10.1098/rspa.1998.0193>.
- [10] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv. Adapt. Data Anal.* 1 (01) (2009) 1–41.
- [11] M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2011, pp. 4144–4147, <http://dx.doi.org/10.1109/ICASSP.2011.5947265>.
- [12] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.* 62 (3) (2014) 531–544.
- [13] S. Mutua, C. Gu, H. Yang, Visibility graph based time series analysis, *PLoS One* 10 (11) (2015) e0143015, <http://dx.doi.org/10.1371/journal.pone.0143015>.
- [14] L. Yu, H. Yao, J. Wang, Analysis of stock price motion asymmetry via visibility-graph algorithm, *Front. Phys.* 8 (2020) 539521, <http://dx.doi.org/10.3389/fphy.2020.539521>.
- [15] M. Serafino, A. Gabrielli, G. Caldarelli, G. Cimini, Statistical validation of financial time series via visibility graph, 2017, arXiv preprint [arXiv:1710.10980](https://arxiv.org/abs/1710.10980).
- [16] B. Zhang, J. Wang, W. Fang, Volatility behavior of visibility graph EMD financial time series from ising interacting system, *Phys. A* 432 (2015) 301–314, <http://dx.doi.org/10.1016/j.physa.2015.03.057>.
- [17] G. Liu, F. Xiao, C.-T. Lin, Z. Cao, A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation, *IEEE Trans. Fuzzy Syst.* 28 (11) (2020) 2677–2690, <http://dx.doi.org/10.1109/TFUZZ.2020.2972823>.
- [18] O.J. Dunn, Multiple comparisons among means, *J. Amer. Statist. Assoc.* 56 (293) (1961) 52–64, <http://dx.doi.org/10.1080/01621459.1961.10482090>.
- [19] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. Neural Netw.* 10 (3) (1999) 626–634, <http://dx.doi.org/10.1109/72.761722>.
- [20] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Netw.* 13 (4–5) (2000) 411–430, [http://dx.doi.org/10.1016/S0893-6080\(00\)00026-5](http://dx.doi.org/10.1016/S0893-6080(00)00026-5).
- [21] Y. Yang, H. Yang, Complex network-based time series analysis, *Phys. A* 387 (5–6) (2008) 1381–1386, <http://dx.doi.org/10.1016/j.physa.2007.10.055>.
- [22] R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Ambiguities in recurrence-based complex network representations of time series, *Phys. Rev. E* 81 (1) (2010) 015101, <http://dx.doi.org/10.1103/PhysRevE.81.015101>.

- [23] J.F. Donges, J. Heitzig, R.V. Donner, J. Kurths, Analytical framework for recurrence network analysis of time series, *Phys. Rev. E* 85 (4) (2012) 046105, <http://dx.doi.org/10.1103/PhysRevE.85.046105>.
- [24] Z.-K. Gao, Q. Cai, Y.-X. Yang, W.-D. Dang, S.-S. Zhang, Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series, *Sci. Rep.* 6 (2016) 35622, <http://dx.doi.org/10.1038/srep35622>.
- [25] M. Han, Q. Fan, G. Ling, Multiscale online-horizontal-visibility-graph correlation analysis of financial market, *Phys. A* 607 (2022) 128195, <http://dx.doi.org/10.1016/j.physa.2022.128195>.
- [26] Y. Yang, J. Wang, H. Yang, J. Mang, Visibility graph approach to exchange rate series, *Phys. A* 388 (20) (2009) 4431–4437, <http://dx.doi.org/10.1016/j.physa.2009.06.051>.
- [27] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Econophysics: Financial time series from a statistical physics point of view, *Phys. A* 279 (1–4) (2000) 443–456, [http://dx.doi.org/10.1016/S0378-4371\(00\)00010-8](http://dx.doi.org/10.1016/S0378-4371(00)00010-8).
- [28] G. Rilling, P. Flandrin, One or two frequencies? The empirical mode decomposition answers, *IEEE Trans. Signal Process.* 56 (1) (2008) 85–95.
- [29] S. Lahmiri, Modeling and predicting historical volatility in exchange rate markets, *Phys. A* 471 (2017) 387–395, <http://dx.doi.org/10.1016/j.physa.2016.12.061>.
- [30] Y. Lin, Y. Yan, J. Xu, Y. Liao, F. Ma, Forecasting stock index price using the CEEMDAN-LSTM model, *North Am. J. Econ. Financ.* 57 (2021) 101421, <http://dx.doi.org/10.1016/j.najef.2021.101421>.
- [31] A. Bareket, B. Pärvi, Enhancing stock market forecasting using recursive time series decomposition techniques, *J. Mach. Intell. Data Sci.* 4 (2023) 43–51, <http://dx.doi.org/10.11159/jmids.2023.006>, URL <https://jmids.avestia.com/2023/006.html>.
- [32] I.T. Jolliffe, *Principal Component Analysis*, second ed., Springer, 2002, <http://dx.doi.org/10.1007/b98835>.
- [33] P. Comon, Independent component analysis, a new concept? *Signal Process.* 36 (3) (1994) 287–314, [http://dx.doi.org/10.1016/0165-1684\(94\)90029-9](http://dx.doi.org/10.1016/0165-1684(94)90029-9).
- [34] A.D. Back, A.S. Weigend, A first application of independent component analysis to extracting structure from stock returns, *Int. J. Neural Syst.* 8 (4) (1997) 473–484, <http://dx.doi.org/10.1142/S0129065797000458>.
- [35] Y. Chen, P. Zhao, Z. Zhang, J. Bai, Y. Guo, A stock price forecasting model integrating complementary ensemble empirical mode decomposition and independent component analysis, *Int. J. Comput. Intell. Syst.* 15 (1) (2022) 75, <http://dx.doi.org/10.1007/s44196-022-00140-2>.
- [36] C. Zhang, H. Pan, Y. Ma, X. Huang, Analysis of Asia Pacific stock markets with a novel multiscale model, *Phys. A* 534 (2019) 120939, <http://dx.doi.org/10.1016/j.physa.2019.04.175>.
- [37] J. Cao, Z. Li, J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Phys. A* 519 (2019) 127–139, <http://dx.doi.org/10.1016/j.physa.2018.11.061>.
- [38] M. Fey, J.E. Lenssen, Fast graph representation learning with PyTorch Geometric, in: *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019, pp. 1–9, <http://dx.doi.org/10.48550/arXiv.1903.02428>, URL <https://arxiv.org/abs/1903.02428>, arXiv:1903.02428.
- [39] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.* 74 (366) (1979) 427–431.
- [40] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econometrics* 54 (1–3) (1992) 159–178, [http://dx.doi.org/10.1016/0304-4076\(92\)90104-Y](http://dx.doi.org/10.1016/0304-4076(92)90104-Y).
- [41] W.A. Brock, W.D. Dechert, J.A. Scheinkman, B. LeBaron, A test for independence based on the correlation dimension, *Econometric Rev.* 15 (3) (1996) 197–235, <http://dx.doi.org/10.1080/07474939608800353>.
- [42] G.M. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (2) (1978) 297–303.
- [43] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50 (4) (1982) 987–1007.
- [44] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* 31 (3) (1986) 307–327.
- [45] P. Xu, R. Zhang, Y. Deng, A novel visibility graph transformation of time series into weighted networks, *Chaos Solitons Fractals* 117 (2018) 201–208, <http://dx.doi.org/10.1016/j.chaos.2018.07.039>.
- [46] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A* 33 (2) (1986) 1134–1140, <http://dx.doi.org/10.1103/PhysRevA.33.1134>.
- [47] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phys. Rev. A* 45 (6) (1992) 3403–3411, <http://dx.doi.org/10.1103/PhysRevA.45.3403>.