This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2026.3676810

Detecting Stable Cross-Impact Patterns in Bivariate Time Series

Gennady Andrienko, Natalia Andrienko, Maram Akila, Bahavathy Kathirgamanathan, and Miguel Ponce-de-Leon

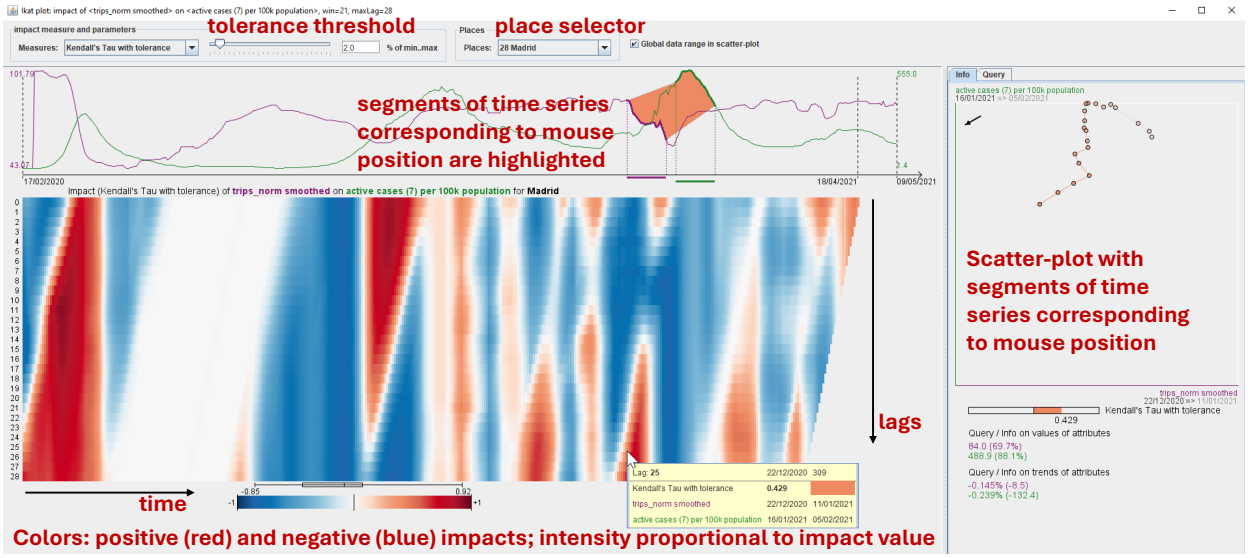


Fig. 1: Visualization of impact values. The control panel, located at the top, allows for selecting time series, impact measures to apply, and setting parameters such as the tolerance threshold. The central section features a plot of the base and response time series, aligned with the impact matrix. This matrix displays impact values along the timeline and across various time lags in the form of an Ikat plot. The plot is named after a Southeast Asian dyeing technique used to pattern textiles [51]. For any selected cell in the Ikat plot, the corresponding time series fragments are highlighted in bold on the timelines above the plot and presented on the scatter plot to the right, with a directed arrow indicating the progression from the first to the last point.

Abstract—This paper presents a visual analytics workflow for detecting stable cross-impact patterns in time series pairs. A sliding window technique computes multiple impact measures, including a novel Kendall's tau variant that tolerates minor fluctuations. Evaluating these measures across various time lags reveals dynamic relationships between time series. An interactive Ikat plot facilitates the exploration of impact distributions, helping identify intervals where specific cross-impacts remain stable (e.g., trends in one series followed by similar or opposite trends in another after a lag). These intervals are extracted as events, whose temporal (and, when applicable, spatial) distributions can be analyzed to uncover broader patterns across multiple time series pairs and over extended time spans. This includes identifying co-occurring cross-impacts and variations in cross-impact presence or type across different periods and data subsets. Experiments on real-world datasets demonstrate the framework's ability to isolate robust patterns, providing a scalable and interpretable approach to analyzing complex temporal dynamics.

Index Terms—Visual analytics, time series, impact

1 Introduction

In many real-world scenarios, dependencies between time series often manifest with varying lags due to economic, physical, or behavioral delays. For instance, gold and crude oil prices exhibit dynamic relationships influenced by inflation expectations, geopolitical events, and macroeconomic policies, where oil often responds with a lag to gold price shocks. Similarly, interest rates and stock market indices demonstrate lead-lag effects, as monetary policy changes influence equity markets over different time horizons. In meteorology, temperature and electricity consumption show seasonal dependencies, with energy demand lagging behind temperature shifts. In epidemics, lagged correlations have been found between the intensity of population mobility and the rate of new infections [11]. Identifying lagged relationships and assessing their strength is crucial for forecasting, risk management, and decision-making in various domains [65].

Detecting and understanding cross-impacts between time series in both temporal and spatial contexts is a fundamental challenge in data analysis. Identifying these relationships requires not only the computation of statistical impact measures over time but also careful consideration of different time lags. However, short-term fluctuations, periodic variations, and irregular trends complicate this process, necessitating robust methods for preprocessing and analysis.

We address these challenges by introducing a systematic approach that computes multiple impact measures over a sliding window with varying time lags. In addition to traditional measures, such as Pearson correlation, rank correlation, and Kendall's tau, we have introduced

- *Gennady and Natalia Andrienko are with Fraunhofer Institute IAIS and with City, University of London. E-mail: {gennady|natalia}.andrienko@iais.fraunhofer.de.*
- *Maram Akila and Bahavathy Kathirgamanathan are with Fraunhofer Institute IAIS.*
- *Miguel Ponce-de-Leon is with Barcelona Supercomputing Center.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

a novel variant of Kendall's tau that is less sensitive to minor temporal deviations in value sequences and thus more robust against small fluctuations.

These computations form the highly important first step of our analytical workflow for detecting, extracting, and analyzing stable cross-impact patterns in datasets containing multiple bivariate time series (Fig. 2). The success of the whole process depends on finding an appropriate combination of options for pattern extraction. To facilitate informed selection and testing of impact measures and parameter settings, we developed an interactive visual interface including the novel Ikat plot, which is controlled by query widgets.

In the initial phase of the workflow, analysts use a few example time series pairs to explore how detected patterns vary depending on parameter settings. Once suitable settings are identified, i.e., those that yield patterns aligned with the analysis objectives, they are applied to the full dataset to detect similar patterns across all pairs. The resulting patterns are then extracted and represented as "impact events", which undergo further analysis as outlined in Fig. 2. In this way, the Ikat plot acts as a tool for hypothesis generation and parameter sensitivity analysis, enabling analysts to visually assess the stability and characteristics of cross-impacts before proceeding to automated event extraction.

We wish to clarify that this work presents a conceptual workflow design and a proof of concept, rather than a polished, ready-to-use software system. Our contribution includes a prototype implementation and a case study conducted by the authors, among whom are epidemiology experts. The aim is to lay the groundwork for the future development of robust, user-friendly analytical tools. The intended audience for such tools comprises time-series experts that seek to uncover complex lagged relationships between dynamic phenomena existing in various application domains such as epidemiology, economics, and environmental science.

Given the novelty of the Ikat plot design and its key role in the entire workflow, we evaluated the plot's interpretability and representational power through a two-stage study: first, with four time-series professionals to validate pattern recognition in static representations, and subsequently with five visualization experts to refine interactive features. The expert feedback suggested that the concept is promising for identifying interpretable cross-impacts, but still considerable work is needed for developing a fully polished software tool and evaluating its usability and effectiveness with a broader range of users.

Our main contributions thus include

- A novel, robust variant of Kendall's tau tailored for time series analysis that tolerates minor temporal fluctuations.
- The Ikat plot, an interactive visualization supporting parameter exploration and pattern detection in the initial workflow stage.
- An event extraction methodology applicable to large sets of time series pairs.
- A comprehensive event extraction and analysis workflow (Fig. 2) to identify significant cross-impact patterns.
- A proof-of-concept case study on epidemiological data, demonstrating the utility of the approach.

The remainder of this paper is organized as follows. In Section 2, we review the related work. Section 3 introduces the key part of the methodology, including the computation of impact measures, the design of the interactive visualization for pattern discovery, and the event extraction process. Section 4 describes our two-stage evaluation of the Ikat plot design. Section 5 presents the entire workflow within a case study. Finally, Section 6 critically discusses the approach and outlines directions for future research.

2 Related Work

This section highlights key themes in time series analysis, addressing fundamental properties of time that shape forecasting models and anomaly detection techniques. We distinguish between univariate and multivariate time series, where the latter enable the study of interdependencies between variables.

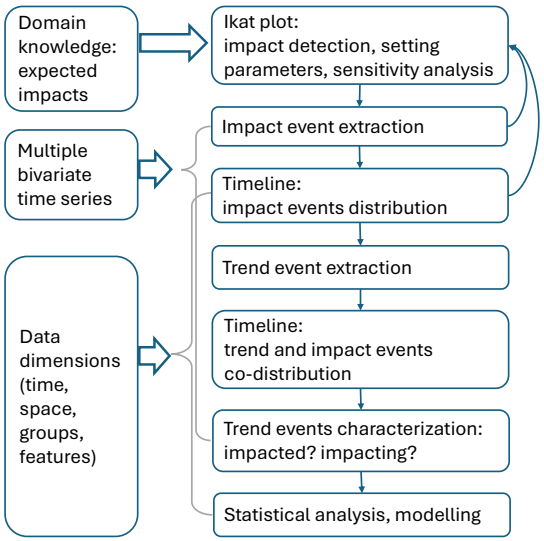


Fig. 2: General workflow for detecting and analyzing impact events from multiple bivariate time series.

2.1 Time Series Analysis Fundamentals

Fundamental properties of time and time-referenced data include temporal dependency and temporal cycles. Temporal dependency refers to the influence of past values on future observations, forming the basis for forecasting models such as autoregressive methods and recurrent neural networks. This dependency can be short-term (e.g., stock price momentum) or long-term (e.g., climate trends). Temporal cycles capture recurring patterns over fixed or variable intervals, driven by external factors like seasons, daily routines, or economic activities. Examples include annual temperature fluctuations, rush-hour traffic, and holiday-driven sales. Understanding these properties is key to time series modeling, anomaly detection, and predictive analytics.

Time series data vary in the number of series and attributes. A univariate time series tracks a *single attribute* over time, while a multivariate time series records *multiple attributes* per time step. This differs from multiple univariate time series, which track the same attribute across different entities or locations. This distinction is crucial, as univariate methods analyze individual trends, whereas multivariate techniques capture interdependencies among variables.

Time series analysis serves diverse fields. In statistics, econometrics, and meteorology, it focuses on forecasting; in epidemiology the focus is to infer causal relationships among relevant features [11, 65]; in engineering and signal processing, it aids signal detection. It also supports data mining, machine learning, and pattern recognition tasks like clustering, classification, and anomaly detection. Analytical techniques include autocorrelation analysis, seasonality detection, trend decomposition, and Fourier-based frequency analysis, mostly applied to univariate series.

2.2 Time Series Visualization

Various visualization techniques for time series are covered in classical books [13, 17, 57] and cartographic literature [34, 60], while [3] reviews modern information visualization approaches.

Analytical tools include TimeSearcher [28] for selecting time intervals, TimeSearcher2 [15] for motif detection, and LoCoMotif [58] for identifying time-warped motifs. TimePool [22] supports various queries over multiple time series and visually presents times when query conditions are satisfied. ChronoDesk [39] allows flexible grouping and arrangement of multiple time series, thus supporting summarization and comparison of them. RelaQ [37] enables relation-driven queries over multiple time series, supporting complex queries composed of primitives like correlation, similarity, and causality with lag modificators. The primitives are calculated either over fixed windows or using dynamic time warping [33, 48].

Limitation of existing approaches	How our method addresses it
Impact detection methods often fail to handle minor fluctuations or noise, leading to misidentification of significant trends.	Our tolerance-based Kendall's tau is robust against minor fluctuations by treating them as ties, focusing on substantial patterns.
Models struggle with non-linear, time-varying dependencies and variable lags, often requiring strict assumptions like stationarity.	Our approach captures dynamic, lagged relationships without strict data assumptions.
Sophisticated motif discovery and query methods enable discovery of patterns, but do not provide sufficient overview of their distributions.	We combine interactive query with Ikat plot visualization that provides a overall level overview of patterns distribution across time and lags.
Visualization and query tools like RelaQ and CRPs are tailored for synchronization analysis and do not reveal interpretable, lagged impact patterns across continuous time and lag dimensions.	Our interactive Ikat plot visualizes impact values over time and lags, enabling easy exploration of stable, lagged cross-impacts.
Lack of comprehensive, visual workflows to extract and analyze meaningful cross-impact events from raw time series data.	We provide a full visual analytics workflow for impact detection, parameter tuning, and event extraction for subsequent analysis.

Table 1: Summary of limitations in existing approaches and our contributions

Time series pattern mining, pioneered by E. Keogh [66], introduced matrix profiles to uncover structures in univariate time series, such as stock prices or temperature records. Additionally, paper [59] and later [55] propose assessing time series similarity, encoding it with similarity-preserving colors, and mapping these colors onto a calendar view to highlight recurring patterns. OutVis [24] integrates a timeline view with dynamic queries on a parallel coordinates plot representing time series features, enabling users to detect and interpret outliers effectively.

Within the univariate paradigm, growth matrix visualization [30] identifies varying-length intervals showing trends like sustained increases or decreases. A triangular model [46] was later developed to facilitate reasoning over temporal intervals.

For multivariate time series, methods include anomaly detection [56] considering anomalies in their temporal contexts, assessing time interval similarity based on multivariate data distributions [5], identifying periods of stable multivariate dynamics [7], and detecting episodes and characterizing them by common topics [10].

A special case of multivariate time series is univariate time series recorded at different spatial locations, also known as spatial time series. GeoChron [20] divides time into fixed-length intervals, computes correlations between these intervals across spatially adjusted locations, and extracts communities from the resulting spatial relationship network. The evolution of these communities is then visualized using a storyline-based approach, enabling users to trace spatially coherent temporal patterns over time. In similar problem settings, VolumeSTCube [21] captures the evolution of the spatial distributions of temporal patterns in synchronized spatial time series using a continuous variant of the space–time cube.

2.3 Analysis of Bivariate Time Series

A special case of multivariate time series is bivariate data sets, which consist of two attributes. For example, [50] constructs 2D trajectories from weekly pairs of financial indicators and applies self-organizing maps to group them based on shape similarity. This approach was further developed in Congnostics [41], which extends analysis beyond simple scatter plots of time series. The system computes various relationship measures characterizing the shape of trajectories, which are then used for clustering and UMAP embedding of time series pairs.

Cross Recurrence Plots (CRPs) [38] offer another analytical approach for bivariate time series by visualizing the recurrence of similar states between two time-dependent signals. By plotting a two-dimensional matrix where each point indicates a recurrence between the two systems, CRPs reveal synchronization patterns, shared dynamics, and temporal alignment.

Machine learning methods for bivariate time series face significant challenges in capturing the complexities of temporal lags, asynchrony, and nonlinear dependencies [12]. These methods often assume fixed lag relationships, which can cause them to miss dependencies that occur across varying delays. Another issue is that dependencies may be time-varying, requiring explicit modeling of dynamic lag structures. Approaches like Recurrent Neural Networks theoretically support long-term temporal modeling, but in practice they often struggle beyond a few lag steps due to vanishing gradients and overfitting.

LFPeers [16] analyzes multiple multivariate time series of epidemiological indicators and related attributes in two phases: a search phase to find similar countries and an exploration phase to analyze the outcomes of their implemented measures. This approach supports epidemiologists in learning from the pandemic by comparing effective and ineffective strategies. The system aims to improve understanding of cause-and-effect relationships in the pandemic and offers a general concept for temporal and multimodal similarity search and exploration

COPE (Co-Occurrence Pattern Exploration) [36] is a visual exploration framework designed to discover temporal relationship patterns between event locations in spatial time series. Using a novel approach to detecting and exploring co-occurrence patterns by enabling flexible event extraction and interactive visual analysis, the system enables analyst to specify relationship of interest through predicates and examples, and then explore distribution of co-occurrences in space and time. The approach allows to take into account possible lags, but does not support systematic exploration of events in respect to varying lags.

The development of visual causality analysis has evolved from static causal discovery to more practical and time-aware approaches. Early work of J.Wang and K.Mueller [61] introduced interactive tools that allowed users to refine algorithmically derived causal networks. This was extended in [62], which addressed real-world challenges like Simpson's Paradox and introduced model scoring for better refinement. More recently, DOMINO [63] expanded causality analysis to time-dependent data, integrating logic-based causality to identify causal effects across time windows. Unlike traditional methods like Granger causality [25], which primarily tests whether one time series can predict another, DOMINO provides an interactive framework that allows users to visually explore and validate temporal causal relationships, enhancing interpretability and usability in real-world applications.

2.4 Concluding remarks

In contrast to causality-oriented techniques, which seek to mathematically prove directed causal influence, our approach focuses on the **interpretive exploration of temporal associations**. By utilizing a robust Kendall's τ with tolerance, we prioritize the visual identification of stable cross-impact patterns, allowing experts to involve domain knowledge to infer potential causality in a way that is more robust to noise and easier to validate than purely algorithmic causal discovery.

With this interpretive focus, when time series have geographic references, the analysis necessitates a transformation from raw sequences into discrete, meaningful units. This aligns with established frameworks for spatio-temporal data [4], specifically the **extraction of events** from time series to facilitate analysis within spatial, temporal, and situational contexts [6]. Given the inherent complexity of both the data and the underlying impact measures, our approach adopts the human-centered machine learning paradigm [9], emphasizing explainable methods that keep the analyst in the loop.

However, despite the abundance of methods for bivariate and spatio-temporal analysis, existing tools often fall short in supporting the systematic detection of stable patterns that evolve over **variable time lags**. Traditional machine learning techniques frequently struggle with non-linear, time-varying dependencies, while specialized causality frameworks often prioritize predictive influence over the discovery of interpretable lagged impact intervals. Our proposed method addresses this gap by combining robust statistical measures with a **sliding window strategy** to capture dynamic, lag-dependent relationships. Supported by an interactive visual interface, our approach bridges data-driven detection with visual reasoning, enabling the analysis of complex, asynchronous dynamics across both temporal and spatial dimensions. Table 1 summarizes the limitations of existing approaches and provides the final motivation for our proposed workflow.

3 Detection and Extraction of Impact Patterns

This section details the **initial phase** of the proposed workflow (Fig. 2), focusing on the computation of cross-impacts and the visual calibration of extraction parameters. This stage is critical, as the success of the subsequent analysis depends on identifying the optimal settings for pattern detection.

We begin by describing methods for quantifying the impact between pairs of time series, including our robust variant of Kendall's τ. Next, we explain how to apply these measures across different segments of the time series using varying lags to account for temporal delays. Then, we introduce the Ikat plot, a novel interactive visualization designed to provide a comprehensive overview of impact distributions. By enabling analysts to explore how different parameters influence pattern visibility, the Ikat plot serves as the primary tool for finding the settings that will later be applied to the entire dataset. The subsequent analysis process is presented through a case study in Section 5.

3.1 Kendall's Tau with Tolerance

Let's consider a simple example of two monotonically increasing time series $X = [1.0, 2.5, 2.9, 4.2]$ and $Y = [1.1, 2.0, 3.2, 3.9]$. For these data, the Pearson correlation is equal to 0.961, and Spearman rank correlation is 1, indicating perfect monotonicity. Pearson correlation measures linear relationships, and Spearman rank correlation assesses monotonicity based on rank differences. While Pearson correlation can be sensitive to outliers and Spearman correlation may not capture certain nonlinear patterns, Kendall's Tau [31] offers a more robust alternative by evaluating the concordance of paired observations. It compares the number of concordant and discordant pairs, where concordant pairs maintain the same relative order in both variables, while discordant pairs do not.

Kendall's Tau is particularly useful in time series analysis, as it is more resistant to noise, non-stationarity, and lagged dependencies, which are common issues in real-world data. In contrast, more complex measures, such as Granger causality [25], impose strict assumptions like stationarity, making them harder to apply and interpret. Formally, given a time series of length T, if c is the number of concordant pairs and d is the number of discordant pairs, Kendall's Tau is computed as:

$$\tau_n = \frac{c-d}{c+d} = \frac{2(c-d)}{T(T-1)}$$

The value of Kendall's Tau ranges from -1 (perfect disagreement) to +1 (perfect agreement), with 0 indicating no clear relationship. For the given toy time series, all 6 pairs are concordant (Table 2, column "Standard result"), and therefore Kendall's Tau is equal to 1.

3.1.1 Introducing tolerance

Kendall's Tau was originally designed for rankings, assuming that all distinct values differ significantly. However, when applied to continuous values, even small, often random fluctuations can be treated as distinct ranks, potentially distorting the results. To address this sensitivity to minor fluctuations, we introduce tolerance thresholds ε_x and ε_y for the X and Y series respectively. Differences smaller than these thresholds are treated as ties, effectively ignoring fluctuations that may be due to noise rather than meaningful trends.

The original Kendall's Tau algorithm and our extension are detailed in Algorithm 1, with modifications highlighted in gray. For example, in Table 2, pairs (1,2) and (2,3) have at least one difference below the tolerance threshold (0.8), causing them to be treated as ties in the modified calculation. This yields a final value of $4/6 = 0.667$ (Table 2, column "Modified result").

The tolerance thresholds ε_x and ε_y can be specified in two ways: as absolute values (e.g., $\varepsilon_x = 14.293$) or as percentages of the series value range. For percentage-based specification, the thresholds are computed as $\varepsilon = (max - min) * percentage/100$, where *min* and *max* represent the minimum and maximum values of the respective time series. To ensure robust threshold calculation in the presence of outliers, the system optionally uses percentile-based range estimation (e.g., 5th to 95th percentiles) instead of absolute min-max values, preventing extreme values from distorting the tolerance calibration.

Algorithm 1 Compute Kendall's Tau with Tolerance

Require: Two input arrays X and Y of equal length T
Require: Tolerance parameters ε_x and ε_y
Ensure: Kendall's Tau coefficient with tolerance

```
if n ≤ 1 then
    return 0
end if
concordant ← 0, discordant ← 0                          ▷ Initialization
for i ← 0 to T − 2 do
    for j ← i + 1 to T − 1 do
        diff_x ← X[i] − X[j]
        diff_y ← Y[i] − Y[j]
        tied_x ← |diff_x| < ε_x
        tied_y ← |diff_y| < ε_y
        if tied_x and tied_y then                       ▷ Both are within tolerance
            continue                                    ▷ Ignore the tied pair
        end if
        sign_x ← sgn(diff_x) if not tied_x else 0
        sign_y ← sgn(diff_y) if not tied_y else 0
        if sign_x = sign_y then
            concordant ← concordant + 1
        else if sign_x ≠ 0 or sign_y ≠ 0 then
            discordant ← discordant + 1
        end if
    end for
end for
totalPairs ← T(T−1)/2                                   ▷ Compute total number of possible pairs
τ ← (concordant − discordant)/totalPairs                ▷ Compute Kendall's Tau
return τ
```

$$\sigma_\tau = \sqrt{\frac{2(2T+5)}{9T(T-1)}} \tag{1}$$

In statistics, the significance thresholds for Kendall's Tau were determined using the asymptotic variance of the statistic for large T under the null hypothesis of independence. In this asymptotic case the distribution for τ, under the null hypothesis, converges to a normal distribution around $\tau = 0$ with a variance given by Equation (1). Using, for example, a significance level of $p = 0.05$ (two-tailed), values of τ above or below approximately $\pm 1.96\sigma_\tau$ are statistically significant.

The introduction of tolerance thresholds ε_x and ε_y in our modified Kendall's Tau affects the calculation by effectively increasing the number of 'tied' pairs. When differences between data points fall within these tolerances, they are treated as ties rather than strict increases or decreases. This increased presence of ties in the data has a known statistical effect on Kendall's Tau: it leads to a larger standard error (σ_τ) and, consequently, requires larger absolute values of τ to reach statistical

Pair (i, j)	(x_i, x_j)	(y_i, y_j)	$x_i - x_j$	$y_i - y_j$	Standard Sign (s_1, s_2)	Standard Result	Below Tolerance?	Modified Sign (s_1, s_2)	Modified Result
(0,1)	(1.0, 2.5)	(1.1, 2.0)	-1.5	-0.9	(-1, -1)	Concordant	No	(-1, -1)	Concordant
(0,2)	(1.0, 2.9)	(1.1, 3.2)	-1.9	-2.1	(-1, -1)	Concordant	No	(-1, -1)	Concordant
(0,3)	(1.0, 4.2)	(1.1, 3.9)	-3.2	-2.8	(-1, -1)	Concordant	No	(-1, -1)	Concordant
(1,2)	(2.5, 2.9)	(2.0, 3.2)	-0.4	-1.2	(-1, -1)	Concordant	Yes (x within tol.)	(0, -1)	Ignored (Tie)
(1,3)	(2.5, 4.2)	(2.0, 3.9)	-1.7	-1.9	(-1, -1)	Concordant	No	(-1, -1)	Concordant
(2,3)	(2.9, 4.2)	(3.2, 3.9)	-1.3	-0.7	(-1, -1)	Concordant	Yes (y within tol.)	(-1, 0)	Ignored (Tie)

Table 2: Pairwise comparisons for standard and tolerance-based Kendall's Tau, assuming example time series $X = [1.0, 2.5, 2.9, 4.2]$, $Y = [1.1, 2.0, 3.2, 3.9]$, and tolerance thresholds $\varepsilon_x = \varepsilon_y = 0.8$.

significance. In essence, with tolerance, the measure becomes more conservative, requiring stronger evidence of a monotonic relationship to be deemed significant [32, 52, 53]. While a precise analytical formula for the variance of Kendall's Tau with arbitrary tolerances can be complex and is beyond the scope of this paper, our experimental evaluation empirically accounts for this effect by observing the behavior of the metric and adjusting our interpretation of significance thresholds accordingly based on the specific tolerance values used.

Based on Equation 1, Figure 3 illustrates the critical values for $|\tau|$ across a range of time series lengths (T) from approximately 5 to 50, and for different significance levels ($p = 0.01, 0.05$). These thresholds represent the minimum absolute Kendall's Tau value required for a correlation to be considered statistically significant at the respective p-value, assuming no ties in the data.

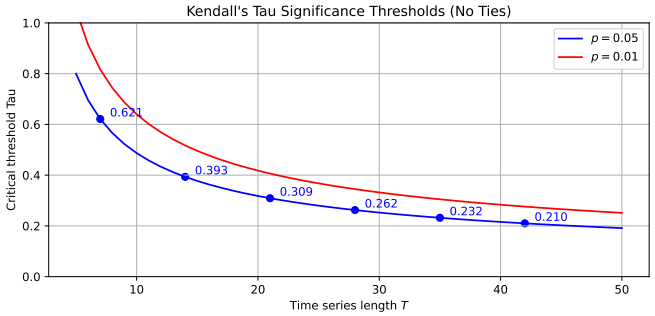


Fig. 3: Critical thresholds of Kendall's Tau for $p = 0.05$ and $p = 0.01$ without tolerance. Marked points correspond to selected T values of $7, 14, 21, 28, 35, 42$ on the $p = 0.05$ line.

It is important to note that the critical values presented here serve as general indicators of desired statistical significance at given p under assumptions such as approximate normality of Kendall's Tau distribution. In practical applications involving real data, the distribution of τ values may deviate from theoretical assumptions due to factors like non-normality, autocorrelation, or complex dependence structures. Therefore, analysts need to assess empirically the distribution of τ in their specific context and adjust significance thresholds accordingly. Interactive exploration of different tolerance values, as demonstrated in our implementation, helps identify appropriate thresholds that balance noise reduction with pattern sensitivity.

3.1.2 Parameters of Kendall's Tau with tolerance

Generally, we analyze a pair of time series of length T, denoted as $X = \{x_0, x_1, \ldots, x_{T-1}\}$ and $Y = \{y_0, y_1, \ldots, y_{T-1}\}$, using a time window of length T^W and considering time lags l in the range $0, 1, \ldots, L$.

The selection of the time window T^W should be guided by domain knowledge and respect relevant temporal cycles inherent in the data. For instance, many human activities exhibit weekly patterns due to work schedules, business hours, and social routines, which are reflected in corresponding time series data. Financial markets demonstrate similar cyclicality with trading patterns that vary between weekdays and weekends, while epidemiological data may show weekly reporting cycles due to administrative practices. Therefore, choosing T^W to align

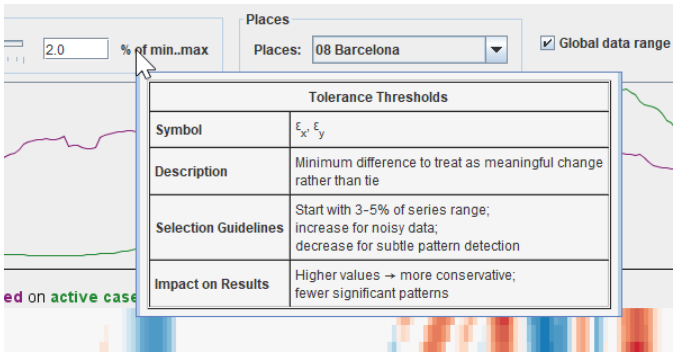


Fig. 4: An example of a tooltip-based presentation of parameter guidance directly in the user interface, aligned with Table 3.

with these natural periodicities - such as multiples of 7 days for weekly cycles - helps ensure that the computed impact measures capture meaningful relationships rather than artifacts of temporal misalignment.

The maximum lag L should reflect the anticipated range of influence delays as estimated by domain experts, representing the longest reasonable time period over which one series might affect another. In many applications, L should be aligned with the length of relevant temporal cycles; for example, in economic data with weekly patterns, setting L to multiples of 7 days ensures that lagged relationships are captured across complete cycles. This domain-informed selection prevents both computational overhead from examining irrelevantly long lags and the risk of missing meaningful delayed impacts that occur within expected timeframes.

Table 3 summarizes the key parameters and provides guidance for setting these parameters. The same information is presented in the user interface for guiding parameters tuning, see an example in Fig. 4.

After setting appropriate parameter values, we compute the matrix

$$\tau = \{\tau_{tl}\} = \{\tau([x_t, x_{t+1}, \ldots, x_{t+T^W-1}], [y_{t+l}, y_{t+l+1}, \ldots, y_{t+l+T^W-1}])\}$$

where l varies within the interval $[0, L]$ and t ranges from 0 to $T - T^W - L$. This is achieved by shifting Y back in time by l time steps, applying a sliding window of length T^W over both X and shifted Y, and computing the τ value for the time series segments contained in the window. We refer to X as the **base** time series and Y as the **response** time series, emphasizing that Y is expected to react to earlier changes in X.

The computation of impact measures over a sliding window for T time steps and L lags is $O(T \times L \times f(W))$, where $f(W)$ is the cost of computing the impact measure over a window of size W. For our robust Kendall's τ, this is $O(T \times L \times W^2)$. Given that W is typically small (e.g., 7 days) and L is constrained by domain logic, the approach scales linearly with the length of the time series, making it highly efficient for multi-year daily data.

In the following subsection, we present the method with an illustrative example using daily time series of two financial indicators: crude oil futures *"CL=F"* and gold futures *"GC=F"* over the years 2023–2024. These data were extracted from `finance.yahoo.com` using the `yfinance` Python library.

Parameter	Symbol	Description	Selection Guidelines	Impact on Results
Tolerance thresholds	ε_x, ε_y	Minimum difference to treat as meaningful change rather than tie	Start with 3–5% of series range; Increase for noisy data; Decrease for subtle pattern detection	Higher values → more conservative, fewer significant patterns
Time window length	T^W	Number of consecutive time points used for τ calculation	Align with natural cycles (e.g., 7, 14 days); Balance smoothing vs. temporal resolution; Consider domain-specific periodicities	Larger T^W → smoother patterns, less sensitivity to short-term fluctuations
Maximum lag	L	Maximum time delay considered between base and response series	Set based on expected influence delays; Consider domain knowledge; Align with temporal cycles	Larger L → increased computational cost, potential for spurious long-term patterns
Significance threshold	$\lvert\tau\rvert$	Minimum absolute tau value for statistical significance	Use Figure 3 as starting point; Adjust upward when tolerance creates many ties; Validate empirically with data	Lower thresholds → more detected patterns, higher false positive risk

Table 3: Summary of Key Parameters for Kendall's Tau with Tolerance

3.2 Interactive Visualization of Impact Distribution

Following the computation of impact measures, analysts need to identify stable and meaningful patterns within a large search space. Because the specific time lags and the periods of occurrence are not known a priori, the analysis requires an exploratory environment capable of navigating a two-dimensional distribution (absolute time × time lag). This leads to four logically derived desiderata for the interactive visualization:

- **D1**. Since cross-impacts may manifest with different lags and last for varying periods at different points in time, the display must provide a 2D overview of the distribution of τ over the timeline and across different lags enabling detection of patterns and assessing their stability and depth.
- **D2**. The 2D impact distribution must be visually linked to the original base and response time series. This enables experts to verify that a computed pattern corresponds to observable behavior in the raw data.
- **D3**. Meaningful insights often depend on specific conditions (e.g., "high price of gold followed by a rise of oil"). The visualization must therefore enable the identification of periods based on a combination of properties: value ranges, direction of change, and the resulting impact strength.
- **D4**. The ultimate goal is to generalize findings from representative pairs to the entire dataset. Therefore, the visualization must support the extraction of "impact events" from multiple time series pairs transforming identified visual patterns into standalone, structured data items that will be used in the subsequent analysis across the full dataset.

3.2.1 Ikat plot

In our illustrative example, we consider gold as the base time series and crude oil as the response time series. These time series are shown in the top line plot in Fig. 5. Below it are two images of the ***Ikat plot*** representing the $T \times L$ matrices of τ values computed over the same T^W using the original Kendall's Tau and our tolerance-adjusted variant, with $\varepsilon_x = \varepsilon_y$ set to 3% of the variables' value ranges.

The Ikat plot displays the distribution of impact patterns over time and across various lags (**D1**). Negative impact values are shown in shades of blue, positive in shades of red, and impact absence in white. Introducing tolerance reduces the intensity of colors in regions of minor changes or fluctuations in the time series, as small variations are treated as ties when computing τ. This results in lower impact values, replacing saturated red and blue by white or near-white shades.

The results of impact calculation depend on the chosen time window T^W. To determine the optimal window for specific data and analysis objectives, multiple Ikat plots need to be visually compared. Figure 5 demonstrates the influence of T^W by presenting plots for $T^W = 7, 14, 21, 28$ chosen according to the weekly cycles of business activities. For $T^W = 7$, the plot exhibits small, scattered patterns, likely caused by fluctuations in the unsmoothed daily time series. As T^W increases, color variations become smoother across the matrix, revealing patterns at different scales.

All variants of the Ikat plot reveal an interplay between two basic geometric patterns: vertical and diagonal stripes. Vertical stripes appear when a short trend in the base time series corresponds to a prolonged

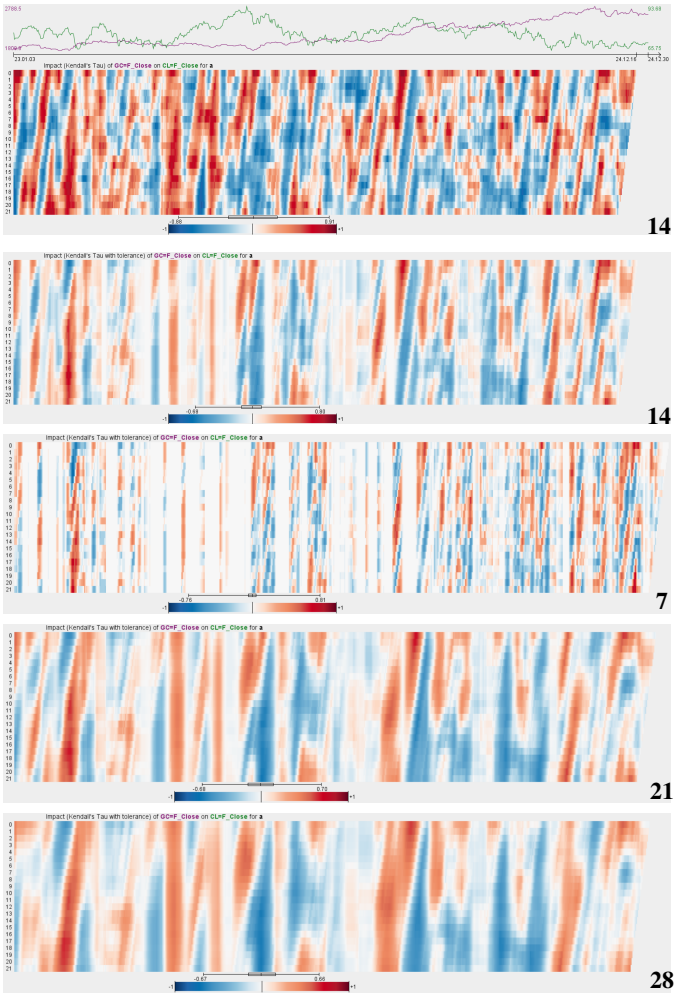


Fig. 5: The line plot at the top shows time series of gold and crude oil futures prices. Below it, five Ikat plots represent the dynamics of impact values over time (horizontal axis) and across lags from 0 to 21 days (vertical axis). The top Ikat plot displays the original Kendall's Tau values, and the remaining four show Kendall's Tau with a 3% tolerance. The numbers in the lower right corners are the sliding window lengths in days.

trend in the response time series, resulting in sustained high impact values across an extended range of time lags (i.e., along the vertical axis). Diagonal stripes emerge when the response time series exhibits a short-lived trend. In this case, the impact weakens as the time lag increases, causing the depth of the effect along the Y-axis to diminish with each step along the X-axis. Vertical interruptions within diagonal patterns occur when the base time series has a flat segment or undergoes a trend reversal. A typology of patterns that can be observed in the Ikat plot is presented in Table 4.

Figure 1 presents the user interface designed to support the required functions. The Ikat plot provides an overview of τ dynamics (**D1**),

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2026.3676810

while visual linkage to time periods (**D2**) and corresponding fragments of the base and response time series is supported through coordination with the time graph display (top) and the scatter plot (right). Figure 6 illustrates how the UI explains the computed impact for a matrix cell under the cursor, displaying all relevant details in a popup window and in the scatterplot area.

Fig. 6: Fragment of the Ikat plot in Fig. 5, demonstrating the impact on November 03, 2024, with $lag = 3$ days. The corresponding segments of the base and response time series are highlighted in bold in the time graph above, also marked by bars below the time line. A sequence of points corresponding to the time window is shown in the scatter plot to the right, with gold on the horizontal axis and crude oil on the vertical axis. An arrow shows direction of points in the time window.

3.2.2 Query

To enable identification of periods characterized by specific properties (**D3**), we support several kinds of queries:

- **Time** queries restrict the time interval and/or time lags.
- **Level** queries restrict the matrix to time intervals when values of base or response time series fall within the specified range.
- **Trend** queries restrict the matrix to time intervals when trends of base or response time series fall within the specified range. Trend is computed as difference between median values in last and first thirds of the time window:

  $$median(\texttt{idx} + \lceil 2/3\, T^W \rceil) - median(\texttt{idx})$$

  where $median(\text{idx})$ calculates the median of the time series in a window of length $\lfloor T^W/3 \rfloor$ starting from index `idx`.
- **Impact** queries restrict the matrix to cells with impact values within the specified range.

For level and trend queries, values are linearly normalized to $[0,1]$ using the *min* and *max* values of the entire time series, similar to how the tolerance threshold is set. Conditions are specified through dynamic query sliders, possibly using "Not" modifiers. Several types of conditions can be combined. For example, Fig. 7 displays the result of combining one level condition ($gold \geq 50\%$), one trend condition ($trend(gold) \geq 1\%$), and one impact condition $|\text{impact}| \geq 0.3$. Figure 8 explains how different types of query conditions affect the selection of regions in the τ matrix.

Filtered-out cells do not disappear completely from the plot. Instead, they are shown as narrow colored stripes, contrasting with the fully covered cells that satisfy the query. This design choice ensures that the context of filtered-out data remains visible while maintaining a clear distinction between selected and non-selected cells. Unlike alternatives such as opacity or blur, narrow stripes preserve the original color encoding and do not distort the perceived data magnitude [64]. Adjusting opacity can unintentionally change the perceived intensity of color, which is already mapped to quantitative values in our heatmaps, potentially leading to misinterpretation. Similarly, blur introduces spatial

Fig. 7: Ikat plot (Figs. 5,6) after applying a dynamic query [2], as shown on the right. Cells of the τ matrix that do not meet the query conditions are visually de-emphasized.

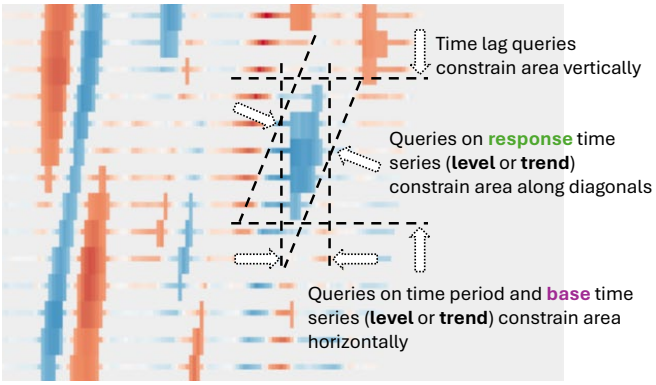


Fig. 8: Different kinds of queries constrain results in different ways.

uncertainty and can obscure boundaries, making it harder to discern fine-grained patterns [40].

Moreover, reducing the occupied area through stripes provides a discrete, pre-attentive visual cue [13, 14] that scales well in dense and sparse regions. This approach avoids the visual blending and overplotting issues that opacity-based techniques can create, particularly in datasets with high data density. By maintaining both the color and shape distinction, the stripes offer a more robust contextual view of the filtered-out data. This is especially useful when investigating query sensitivity, as analysts can still perceive the distribution and density of filtered-out cells relative to the selected cells. Previous work on visualization design emphasizes the importance of preserving context without overpowering the primary data and maintaining perceptual separability between encodings [18, 40]. Our choice of narrow stripes directly supports these principles by allowing users to analyze patterns and gaps in the data without introducing perceptual ambiguity.

3.3 Event Extraction

As stated in the introduction, the visual query interface centered on the Ikat plot is used at the initial stage of the analysis workflow (Fig. 2). This enables the analyst to select parameter settings that reveal the kind of impact patterns aligned with the analysis goals. The effects of these parameter settings are tested on several pairs of time series selected from a large dataset. Once the analyst identifies the settings capable of capturing the patterns sought, these settings are used by an automatic method to extract *impact events* from the entire dataset (**D4**).

An impact event is a standalone temporal object defined by the temporal attributes t_1, t_2, l_1, l_2, where $[t_1, t_2]$ denotes the interval in the base time series that exerts an impact, and $[l_1, l_2]$ specifies the corresponding interval in the response time series, expressed as time lags relative to the start time t_1. Geometrically, these four attributes define a rectangular region within the matrix where columns represent time steps in the base time series and rows represent time lags (as in

Pattern Type	Visual Representation	Inferred Meaning	Example
Sustained Impact	Vertical stripes across multiple lags.	A specific event or trend in the base time series is followed by a prolonged, consistent trend in the response series.	A sustained marketing campaign leads to a prolonged increase in sales over many weeks.
Transient Impact	Diagonal stripes, often fading with increasing lag.	A short-lived trend in the base series leads to a short-lived trend in the response series. The impact is strongest at a specific lag.	A sudden weather event (e.g., a cold snap) causes a spike in energy consumption that lasts only for a few days.
Lag-Specific Impact	Horizontal bands at a particular lag value.	The relationship between the time series is stable and consistent over a long period, but only for a very specific time delay.	A factory production increase consistently leads to a downstream delivery increase at a fixed lag of 5 days.
Episodic Impact	Localized "hotspots" or small, isolated colored patches.	A transient, short-lived impact that is only significant at a specific time step and for a narrow range of lags.	A single, one-time policy announcement leads to a very strong but short-lived impact on stock market volatility.
Stable Trend-Driven Impact	Large, contiguous areas of uniform color.	A strong, stable, and persistent relationship between the two time series across a wide range of time steps and lags.	A fundamental economic relationship where prices consistently rise and fall together over a multi-year period.
Periods of "Noise"	White or near-white regions, interrupting other patterns.	Indicates a lack of significant impact or a period where the relationship is too weak to be considered significant.	A period of economic stability where fluctuations in one time series have no significant impact on the other.

Table 4: Typology of Cross-Impact Patterns Observable on the Ikat Plot

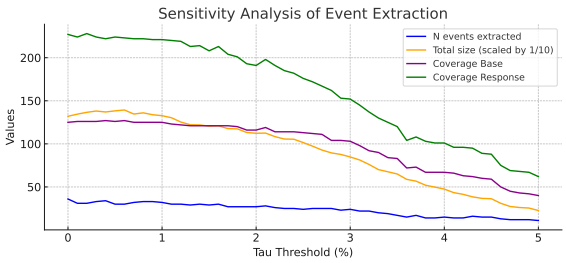


Fig. 9: Sensitivity of the event extraction procedure to the tolerance.

Ikat plot, which is a visual representation of such a matrix). Algorithms for event extraction are detailed in Appendix B.

Extracted events are described by their temporal attributes such as event duration (the horizontal extent) and minimal depth (the range of lags covered), as well as further features, including the base and response trend parameters and the value of the chosen impact measure.

3.4 Parameter sensitivity analysis

As parameter settings for event extraction are typically identified using selected pairs of time series, the analyst may be unsure whether these settings will generalize to all remaining data, or whether small variations in parameter values might lead to significant differences in event extraction results. To address this, an automatic procedure can be run for each parameter, iteratively varying its value v within a range $[v-\eta, v+\eta]$, using an increment δ (where η and δ are specified by the analyst). At each step i, the procedure performs event extraction using the current parameter value v_i, recording the number of extracted events and aggregate statistics of event features. The resulting sequences of numeric indicators can be visualized to assess the effects of parameter variation.

Typically, as thresholds (e.g., tolerance, impact strength, or trend inclination) increase, the counts of extracted events and their lengths and depths decrease. A line graph visualization (as in Fig. 9) can reveal bend points where the rate of change increases. Such bend points indicate critical parameter values that merit further interactive examination using the Ikat plot. Patterns visible in the Ikat plot then guide the analyst in selecting the most appropriate critical value, based on domain knowledge and the specific task requirements.

Figure 9 demonstrates the sensitivity analysis procedure. For the time series used in our running example, we systematically varied the tolerance thresholds ε_x and ε_y within the intervals suggested by Fig. 3, applying the same value to both, assuming similar behavior in the base and response time series. The line graphs in Fig. 9 represent trends in the number of extracted events, total event size, and coverage metrics as tolerance thresholds increase. The lines suggest that the Ikat plot should be used to explore and compare patterns at threshold values of 2% and 3% for making the final choice.

3.5 Concluding remarks

Before evaluating the interpretability of the Ikat plot, we briefly compare our methodology to established approaches for analyzing relationships between time-series pairs—specifically Cross-Recurrence Plots (CRP) and Dynamic Time Warping (DTW). While these methods are widely adopted, the Ikat plot approach differs in several key aspects. Specifically, the Ikat plot reveals impact relationships rather than general similarity, explicitly considers time lags, is less sensitive to noise, and focuses on detection of stable interpretable patterns. A more detailed comparison is provided in Appendix C.

4 Evaluation of Ikat Plot Interpretability

The proposed workflow (Fig. 2) is designed for time-series experts analyzing complex, multi-lag relationships. While the Ikat plot is only one component of this workflow, it performs a critical role: informing the parameter settings for event extraction. Since the Ikat plot is a novel design, unlike familiar timeline-based bar charts (Figs. 13 and 14), it was important to test how users can perceive and interpret its unique visual patterns. To assess its representational power and interpretability, we conducted a two-stage evaluation:

- **Time series expert assessment**: We recruited four professionals in time-series data analysis. Due to their limited availability, this study focused exclusively on the static visual representation.
- **Follow-up evaluation**: We later engaged five experts in information visualization and human-computer interaction. They provided detailed feedback on the visualization and interaction techniques, based on which we refined the tool's visual responses.

4.1 Evaluation of pattern discovery

In the first stage, evaluation materials were distributed to participants via email. These included a concise description of the Ikat plot's purpose, information content, and visual encoding. We provided three screenshots generated from COVID-19 and population mobility data for Spanish provinces [42]: Barcelona, Madrid, and Seville (referenced by identifiers 08, 28, and 41; Fig. 19). Population mobility (normalized daily trips) served as the base variable, and COVID-19 incidence

(normalized daily cases) as the response variable. All plots used fixed parameters and a predefined query: "*increase in base variable followed by increase in response variable after a time lag of at least 7 days*." The experts were asked to identify and interpret the patterns exhibited. Their responses were then validated against a curated chronology of Spain's COVID-19 policies and epidemiological milestones (www.andalucia.com/health/coronavirus/chronology).

Prior to the expert study, we conducted an exploratory experiment using five Large Language Models (LLMs) to test the clarity of our explanatory materials and tasks, which could be judged from the closeness of the LLMs' outputs to what we expected from the human participants. The use of LLMs to simulate various aspects of human behavior has been explored in prior research (e.g., [1]). Their ability to detect and verbalize graphical patterns [19, 67] makes them suitable for a preliminary, machine-based sanity check of our task description and stimuli. However, we do not regard LLM results as a substitute for human evaluation of interpretability..

In this experiment, GPT-4.1, GPT-4o, o4-mini, Claude 3 Opus, and Gemini 1.5 Pro (accessed via FhGenie platform fhgenie.fraunhofer.de) consistently recognized pattern types, such as sustained, trend-driven, episodic, and transient, and assessed impact strengths and durations. The models converged on location-level characterizations, emphasizing episodic patterns for Barcelona, a strong sustained vertical band (7–12 day lags) for Madrid, and a mixture of impacts for Seville. Some models noted interpretation challenges due to visual overlaps. We emphasize that these LLM results are used only as an auxiliary check that the patterns we expected are in principle detectable from the plots under our prompts; they are not used as evidence that the visualization is effective or interpretable for human analysts.

In the subsequent expert evaluation, participants identified patterns and estimated lags based solely on three static images. Despite the inability to test parameter sensitivity or determine precise event dates, the experts' professional backgrounds enabled them to extract meaningful insights.

All four experts identified recurring positive impacts of the mobility on the disease incidence concentrated in 2020–early 2021, typically with 7–14 day lags, occasionally extending up to 21 days. They consistently characterized Location 28 as showing a strong, sustained pattern, Location 08 as episodic, and Location 41 as a mixture of overlapping structures. While they differed in strength estimates and emphasized different typologies (e.g., vertical sustained bands vs. rhomboid "stable trend-driven" forms), their summaries aligned on where and when the major signals occurred. Detailed per-expert findings are provided in Appendix E.

Agreement was strongest on lag ranges, temporal clustering (spring and summer 2020, late 2020/early 2021), and coexistence of multiple pattern types across locations. Divergences were related to strength calibration from static color scales, typology boundaries, and short-lag omissions. We address these through typology clarification, lag-sensitivity checks, improved temporal annotations, and robustness controls (see Appendix E for a full discussion).

The fact that experts extracted meaningful insights from static images suggests that the Ikat visualization is interpretable for time-series specialists under our study conditions. Remaining divergences in strength estimation and event dating underscore the added value of interactive implementations. Overall, these preliminary results indicate that Ikat plots can serve as a useful basis for exploring lagged cross-impacts in time series, but more extensive and comparative studies are needed to fully assess their effectiveness.

4.2 Evaluation of interactive visualization

In the second stage, the evaluation took the form of an online seminar. After introducing the purposes and content of the Ikat plot, we demonstrated its interactive operations and explained their effects. Along the session, we were answering emerging questions and noting immediate feedback. In particular, the participants appreciated the dual-axis design (absolute and relative time), although it was initially somewhat hard to understand. The relationships between the upper and lower parts of the plot were found non-obvious. Despite this, the overall reception was very positive. Two participants expressed their wish to provide more detailed feedback after contemplation.

During a follow-up meeting, participants provided concrete advice for improving visual clarity and user interaction, in particular, for more expressive visual linking between the upper and lower parts of the plot. Based on this feedback, we enhanced the tool. For example, as shown in Figs. 1 and 6, hovering over the lower display now highlights the corresponding time intervals and links between the base and response time series in the upper part. The improvements are briefly described and illustrated in Appendix A.

Our evaluations have several important limitations. The interpretability study involved only four time-series experts and relied on static screenshots, so we could not observe how analysts would interact with the full system or quantify task performance. The second stage with five visualization and HCI experts was conducted as an informal online seminar and design discussion, rather than a controlled task-based study. Moreover, we did not include comparative baselines against alternative visual encodings or existing tools for lagged cross-impact analysis. Consequently, our findings should be viewed as preliminary, formative evidence about the promise of the Ikat plot, rather than a definitive demonstration of its effectiveness or superiority. There remains an obvious need for more comprehensive evaluation studies, which we plan to address in future work.

5 Case Study: Mobility and COVID-19 Impact Analysis

We demonstrate the complete analytical workflow (Fig. 2) by examining the interactions between population mobility and COVID-19 incidence across the provinces of Spain. Using publicly available datasets [42–45], we analyzed two daily time series per province from February 14, 2020, to May 9, 2021: (1) normalized daily internal trips and (2) reported active COVID-19 cases, both calculated per 100,000 residents. To mitigate reporting delays, weekend effects, and high-frequency fluctuations, we applied a **7-day moving average** (back-smoothing) to both series.

5.1 Defining and Extracting Impact Events

Impact events are defined as situations where changes in one time series (e.g., mobility) are followed by systematic changes in another (e.g., COVID-19 cases). We focus on the following types of relationships, as summarized schematically in Fig. 10:

- **Negative impacts**, where increasing disease incidence leads to decreasing mobility and vice versa.
- **Positive impacts**, where increasing or decreasing mobility contributes to a rise or reduction in disease incidence.

Impact values were computed using a **7-day sliding window** with **time lags ranging from 7 to 21 days**. This window length was chosen to ensure reliable impact scores while remaining sensitive to short-lived events. The lag range was informed by domain knowledge regarding COVID-19 incubation periods and the typical time required for mobility policies to respond to epidemiological trends.

To capture negative impacts of disease on mobility, the disease series was treated as the base and the mobility series as the response (Fig. 11A). Conversely, to capture positive impacts of mobility on disease spread, the mobility series was used as the base and the COVID-19 series as the response (Fig. 11B).

Before extracting impact events across the full Spanish dataset, we used the Ikat plot to identify suitable parameters for event extraction by considering representative provinces with distinct dynamics, including Madrid (Fig. 1), Barcelona (Fig. 11), and Seville. The resulting settings were then cross-validated against other provinces, such as Navarra, Valencia, and Zaragoza. We also performed automated parameter sensitivity analysis (Section 3.4). This exploratory phase allowed us to observe how varying thresholds influenced the stability and density of the visual patterns.

Based on this interactive refinement, we adopted the following settings to ensure a balance between sensitivity to subtle trends and robustness against noise:

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2026.3676810

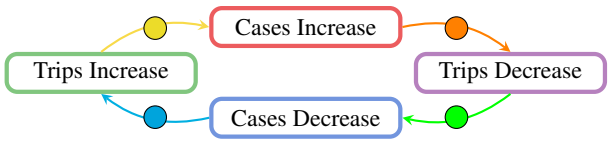
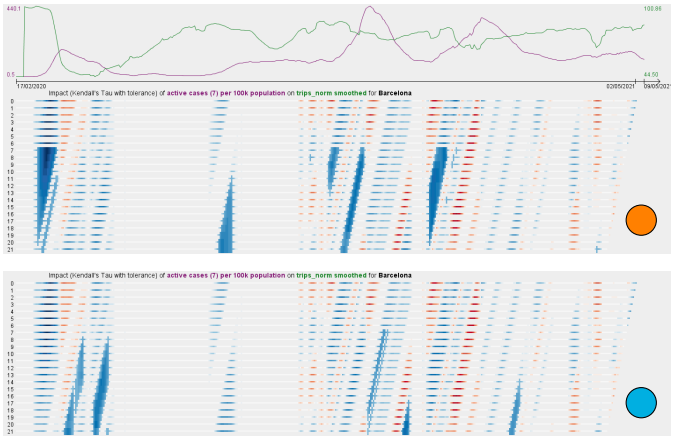


Fig. 10: Expected impacts between trends in time series of cases and trips. Nodes represent trend events, and arrows - impact events. Colors correspond to legends in Figs. 11, 13, 14, 18.



(A) Detection of negative impacts of the disease trends on population mobility.

(B) Detection of positive impacts of the mobility trends on the pandemic development.

Fig. 11: Example Ikat plot screenshots used to select and test event extraction parameters in the case study. The colored circle in the bottom-right corner of each image indicates the event type, as defined in Fig. 10.

- **Tolerance threshold** ($\varepsilon = 2\%$): We set the tolerance at 2% of the value range. This level was identified as effective for filtering out minor fluctuations while preserving meaningful signal.
- **Trend thresholds** ($\pm 3\%$): To focus on intervals of clear progression, we restricted the analysis to periods where the difference between the first and last thirds of the 7-day window is at least 3%. This ensured that extracted events were based on sustained trends rather than transient noise, as illustrated in Fig. 12.
- **Impact strength** (τ): We established a threshold of $|\tau| \geq 0.3$. Visual inspection of the Ikat plots suggested that values below this magnitude often represented ambiguous or unstable relationships, whereas ± 0.3 consistently captured prominent cross-impact patterns.

Specifically, the impact value was constrained according to four distinct event typologies, each represented by a specific color coding in the subsequent analysis:

- **Cases increase** $\rightarrow$ **Trips decrease** (impact ≤ -0.3)
- **Cases decrease** $\rightarrow$ **Trips increase** (impact ≤ -0.3)
- **Trips increase** $\rightarrow$ **Cases increase** (impact ≥ 0.3)
- **Trips decrease** $\rightarrow$ **Cases decrease** (impact ≥ 0.3)

Fig. 12: Choosing the trend settings for event extraction by example of Barcelona. Top: No trend constraints. Bottom: Trend thresholds are set to $\pm 3\%$, so that impacts of minor changes and fluctuations are ignored.

By using the Ikat plot to validate these settings on a subset of the data, we ensured that the automated event extraction subsequently applied to all provinces was grounded in visually verified, stable relationships.

In addition to the impact events, we also extracted **trend events** from each time series using the same ***sliding window*** *of 7 days* and $\pm 3\%$ ***trend thresholds*** as for the impact event extraction. The extracted event types include:

- **Case trends:** intervals of increasing or decreasing case counts.
- **Mobility trends:** intervals of increasing or decreasing numbers of trips.

Finally, for all event types, additional constraints were applied to enforce a ***minimum event duration*** and sufficient ***lag range***, ensuring statistical robustness.

5.2 Visualization of Events

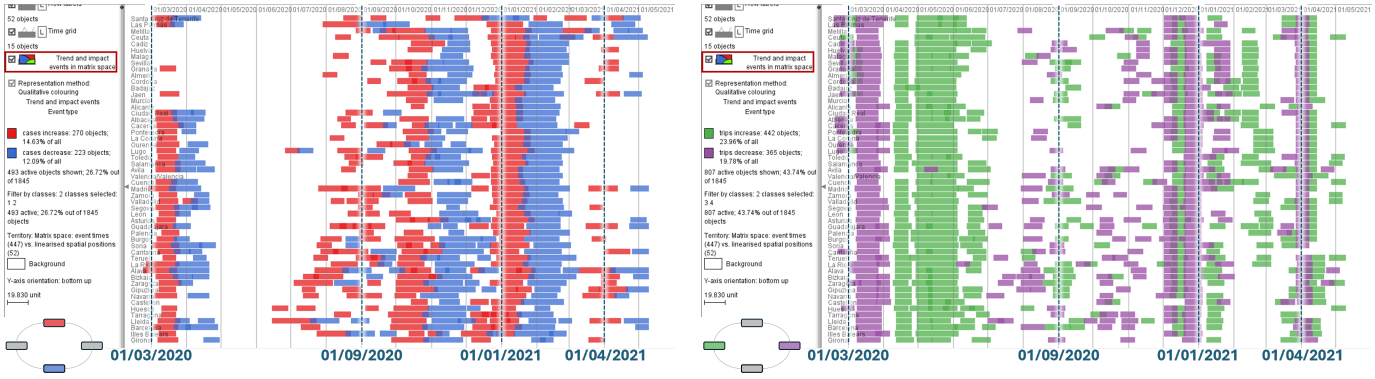
Following event extraction, the next phase of the workflow (Fig. 2) focuses on analyzing event distributions across time and other relevant dimensions of the data, specifically, the entities associated with each time-series pair (in our case study, these are the Spanish provinces). While the proposed methodology does not *prescribe* a specific visualization technique for this exploration, the task inherently requires a display with at least two dimensions, one of which must represent time to facilitate temporal localization of the events end exhibit temporal relationships between them.

As our data have geographical references, one potentially suitable visualization technique is the space–time cube (STC) [23, 27, 35]. An example can be seen in Appendix D (Fig. 18). An STC offers an integrated view of temporal and spatial distributions of events, but effective interpretation requires interactive rotation and navigation to reduce occlusion.

Instead, we adopted the more effective timeline chart [47]. In this 2D display, one dimension (usually horizontal) represents time and the other is used for arranging time-referenced information, such as events, according to a certain purpose-driven principle. In our case, the positions of the second (vertical) dimension correspond to the provinces. Each event is shown as a horizontal bar spanning its start and end times, with bar length proportional to event duration (e.g., Fig. 13).

To preserve spatial context within such a 2D chart, we ordered the provinces along the vertical axis using Sammon's mapping [49] applied to their geographic centroids. This nonlinear dimensionality reduction projects 2D geographic distances into a 1D order. Consequently, neighboring provinces tend to appear near one another on the vertical axis, allowing for the detection of spatially correlated temporal patterns. Bar colors consistently encode event types as defined in Fig. 10.

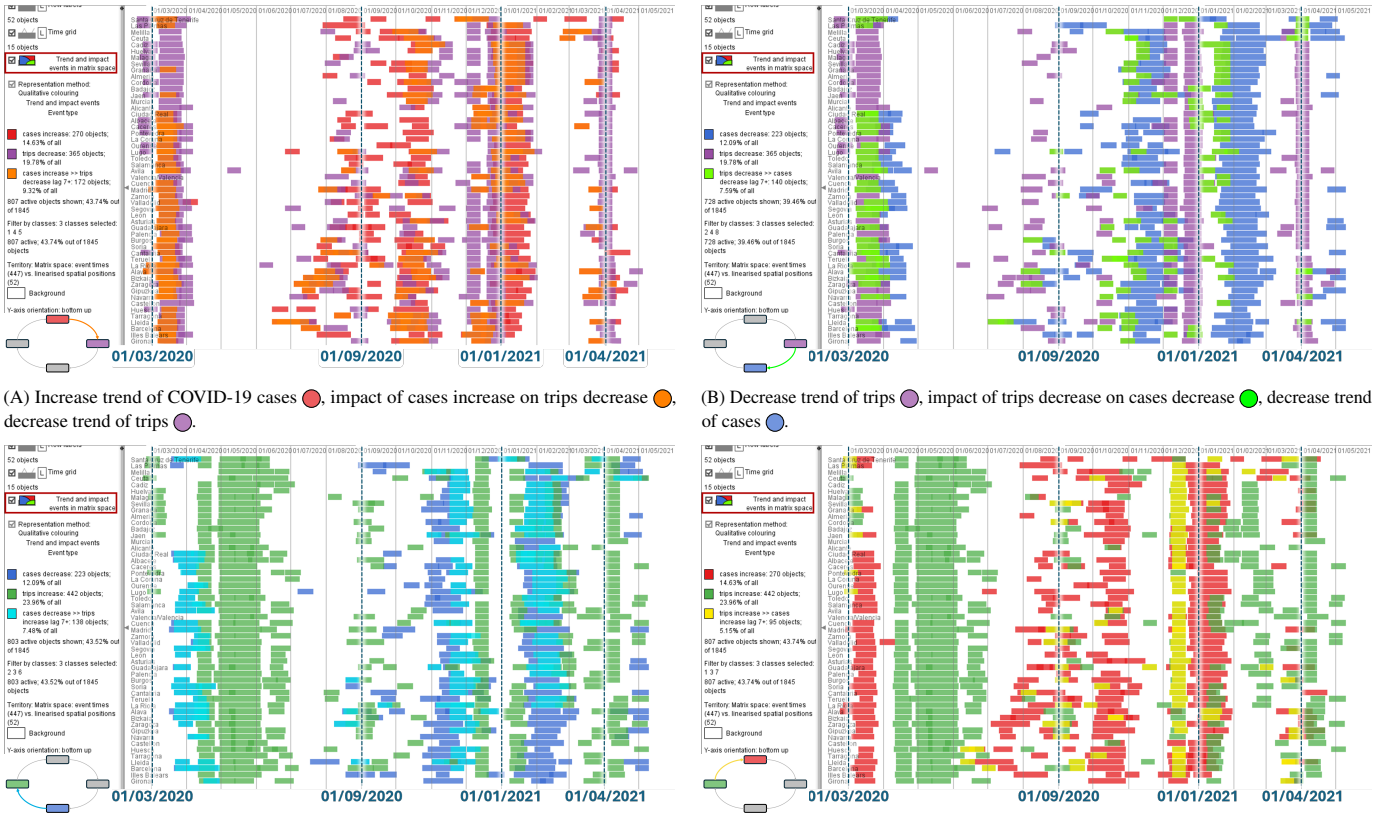
The initial workflow phase yields eight distinct event types (four trend types and four impact types; Fig. 10). Visualizing all types simultaneously is challenging due to inherent temporal overlaps, because impact events include trend events as their integral parts. For instance, a mobility increase may belong to both a preceding "recovery" impact (*Cases decrease* $\rightarrow$ *Trips increase*) and a subsequent "spread"



(A) COVID-19 trend events, with red for increase and blue for decrease.

(B) Mobility trend events with green for increase and purple for decrease.

Fig. 13: Visualization of the distributions of the extracted trend events along the timeline (horizontal dimension) and over the set of provinces (vertical dimension).



(A) Increase trend of COVID-19 cases, impact of cases increase on trips decrease, decrease trend of trips.

(B) Decrease trend of trips, impact of trips decrease on cases decrease, decrease trend of cases.

(C) Decrease trend of cases, impact of cases decrease on trips increase, increase trend of trips.

(D) Increase trend of trips, impact of trips increase on cases increase, increase trend of cases.

Fig. 14: Visualization of trend and impact event distributions corresponding to expected transitions between trends (Fig. 10). Impact events always partly or fully cover corresponding trend events. For example, an impact event of type "cases increase » trips decrease" in panel A is visually represented as a bar drawn over the corresponding trend events "cases increase" and "trips decrease".

impact (*Trips increase* → *Cases increase*). To de-clutter the display, we utilize interactive filtering to focus on specific combinations of event types, as shown in Figs. 13 and 14.

To aid temporal orientation and interpretation, the presented screenshots are additionally annotated with vertical lines marking several anchor dates (March 1, 2020; September 1, 2020; January 1, 2021; April 1, 2021). These markers provide temporal reference points that help approximate when particular patterns occurred.

Figure 13A shows the distribution of COVID-19 trends: red for increases and blue for decreases. The trends form three major epidemic "waves" that are broadly synchronized across provinces, interspersed with periods of more heterogeneous regional behavior.

Figure 13B shows the distribution of mobility trends: green for increases and purple for decreases. Similar to the epidemic curves, mobility exhibits several nationwide phases of reduction and recovery, often coinciding with policy interventions such as lockdowns or easing of restrictions [54], but also periods where provinces diverge, likely reflecting region-specific measures or behavioral responses.

Figure 14 provides combined views of impact and trend events. These visualizations reveal how specific shifts in mobility or disease incidence translated into measurable lagged impacts, and highlight periods where trends emerged without generating subsequent cross-impact effects.

5.3 Exploration of trends and impacts

Figure 14 presents four screenshots of the event timeline chart, showing different combinations of extracted trend and impact events. The visualizations highlight relationships between increases or decreases in COVID-19 cases and mobility allowing us to assess whether the COVID-19 or mobility trend events had significant impacts or occurred without measurable effects.

Overlaps between trend and impact events. Impact events are derived from trend events and always at least partially overlap with them in time. Each impact event (e.g., cases increase → trips decrease) corresponds to a period where a significant trend in one time series (e.g., an increase of cases) is followed by a trend in another series (e.g., a decrease in trips) after a time lag. Since impact events inherently link two trend events, the bar of the impact event is drawn over the respective trend bars.

By analyzing the combined displays of trend and impact events (Fig. 14), we observe several key patterns:

1. Mobility Responses to Pandemic Dynamics

- **Strong initial response to rising cases:** In March and in autumn 2020, increasing COVID-19 incidence was typically followed by decreasing mobility, reflecting imposed restrictions (lockdown has started on March 15) and voluntary behavioral changes. Later in the pandemic (e.g., summer 2020, spring of 2021), increases in cases often did not result in mobility reductions (Fig. 14A).
- **Recovery of mobility after case declines:** When cases decreased, mobility generally increased, likely due to relaxation of restrictions or behavioral changes (Fig. 14C).
- **Seasonal influences:** A notable drop in mobility occurred across all 52 provinces before and during Christmas 2020; in 35 of them the drop was not related to case increases. Similarly, the trips decreased in all provinces around Easter 2021 (beginning of April). The display also shows a synchronous decrease of mobility in 44 provinces in the beginning of December 2020. This trend appeared due to mobility restrictions imposed during several consecutive holidays [29].
- **Regional differences:** The provinces of Andalusia, located in southern Spain and positioned in the upper part of the timeline chart, were less affected by the first wave of the pandemic than other regions. This is evident from the absence of COVID-19 increase trends in the upper left area of Fig. 13A). Nevertheless, mobility decreased nationwide during this period (Fig. 13B). Andalusia also experienced lower pandemic impact in the summer of 2020.

2. Impact of Mobility Changes on COVID-19 Incidence

- **Mobility restrictions were most effective early in the pandemic:** The strongest impact of reduced mobility on declining cases was seen in March-April 2020 and again in October-November 2020. However, in later periods, many decreases in mobility did not result in a corresponding reduction in COVID-19 cases. Overall, only 137 out of 365 (37.5%) mobility decrease events led to a decline in cases, but these accounted for 46% (103/223) of all case decreases, indicating that mobility reductions played a role in disease control but were not the sole factor.
- **Limited impact of mobility increases on case surges:** While some mobility increases were followed by a rise in cases (Fig. 14D), this effect diminished over time. Between April and June 2020, widespread mobility increases had little measurable effect on COVID-19 incidence. In 2021, this impact further weakened, suggesting moderating factors like vaccination and immunity. Only 89 out of 442 (20%) mobility increase events resulted in increased cases. Conversely, only 30% (70/230) of all case increases (excluding the first wave) were preceded by rising mobility within a 1–3 week time frame.

3. Asymmetric Influence Between Mobility and Cases. The effect of rising COVID-19 cases on reducing mobility is stronger and more consistent than the reverse. This asymmetry suggests that factors like restrictions and public awareness influence mobility more reliably than mobility influences virus transmission.

4. Variability in Impact Duration. Some impact events are short-lived, while others persist for weeks. This variation may be linked to region-specific factors such as population density, policy changes, and mobility habits.

5. Spatio-Temporal Differences in Responses. Some provinces exhibit strong, repeated correlations between mobility and COVID-19 trends, while others show inconsistent relationships. The presence of trend events without corresponding impacts in certain provinces suggests that interventions or external factors significantly alter mobility-disease interactions. There are also prolonged time periods when trend events (e.g., increasing cases or mobility) do not lead to measurable impacts in most provinces. These cases likely reflect external influences such as vaccination rates, seasonal variations, or local interventions (e.g., mask mandates).

Concluding Remark. The visualization allowed us to detect stable patterns, regional differences, and periods where expected impacts do not occur, highlighting the complex interplay between mobility and pandemic trends. We complement this rough visual assessment with a more rigorous analysis of the trend-impact relationships by quantifying the proportions of trend events linked to impact events, either as results of prior influences or as contributors to subsequent changes.

5.4 Quantitative analysis and modeling

To obtain quantitative data for the analysis, we applied the **event context extraction** method [8]. For each trend event, we defined two temporal neighborhoods:

- **Preceding context** (3 weeks before event start) – to check if a prior impact event influenced the trend.
- **Following context** (3 weeks before event end) – to check if the trend contributed to a subsequent impact event.

Each trend event $\mathbf{X\ trend_X}$ (where X is mobility or disease cases, and $trend_X$ is increase or decrease) was characterized as:

- **Impacted** if an impact event $\mathbf{Y\ trend_Y} \rightarrow \mathbf{X\ trend_X}$ existed within the 3 weeks before its start.
- **Had impact** if an impact event $\mathbf{X\ trend_X} \rightarrow \mathbf{Y\ trend_Y}$ existed within the 3 weeks before its end.

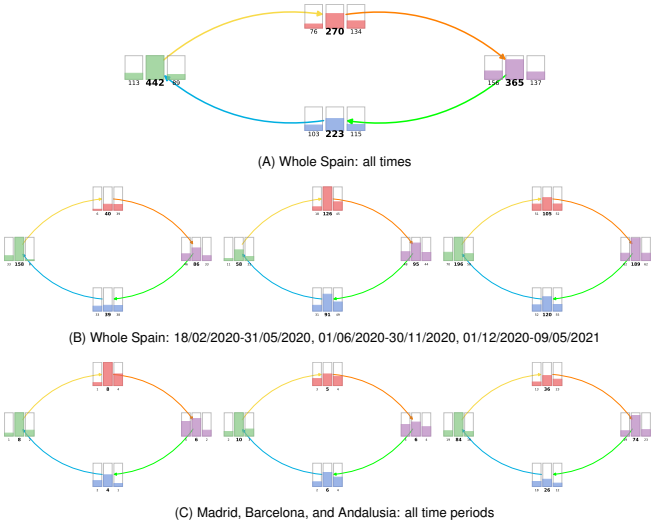
Based on these definitions, we generated two binary attributes: **"Was impacted"** (*yes*/*no*) and **"Had impact"** (*yes*/*no*). To visualize the results, we used impact transition graphs (Fig. 15), whose layout is consistent with the graph of expected impacts shown in Fig. 10, with nodes representing trend events and arrows signifying impact events.

Each graph displays trend-impact relationships through colored bar charts corresponding to different trend event types, encoded by colors. Bar heights are proportional to event counts, with the middle bars indicating the total number of events, the left bars representing impacted events, and the right bars showing impacting events. Different event subsets can be selected using spatial, temporal, and other filters, and the display updates to reflect the corresponding statistics.

The top graph in Fig. 15 shows all events nationwide over the entire study period. The three graphs in the middle row depict trend-impact relationships during the three pandemic waves, while the bottom row presents event subsets for Madrid, Barcelona, and Andalusia. Each graph uses its own scale for the bar charts, with the tallest bar representing the maximum value within the subset and the remaining bars scaled proportionally. This supports the comparison of the proportions across the graphs.

To obtain further insights into the event properties, modeling can be used. Specifically, we are interested which features of an event are predictive for whether it will have an impact on subsequent events. We choose to build an inherently interpretable [26] logistic regression model, as its feature weights describe direct dependencies between the target variable ("having impact") and the input features. In contrast to time-series analysis, features are extracted from the entire event durations and not fixed time windows to keep correspondence with the visually analyzed events.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2026.3676810



(A) Whole Spain: all times

(B) Whole Spain: 18/02/2020-31/05/2020, 01/06/2020-30/11/2020, 01/12/2020-09/05/2021

(C) Madrid, Barcelona, and Andalusia: all time periods

Fig. 15: Impact transition graphs showing trend-impact relationships. Colored bar charts represent different trend event types, with bar heights proportional to event counts: total (middle bars), impacted (left), and impacting (right). The top graph summarizes all events nationwide, while the three graphs in the middle row depict relationships during the three pandemic waves. The bottom row presents subsets for Madrid, Barcelona, and Andalusia.

	Mobility Event				Covid Event			
	Wave 1	*Wave 2*	*Wave 3*	*All*	*Wave 1*	*Wave 2*	*Wave 3*	*All*
Model Accuracy	0.920	0.700	0.710	0.700		0.730	0.720	0.620
Average Mobility	1.056	0.288	0.678	1.339	failed to build a model: all trends have impact	0.069	0.618	0.080
Trend type (increase/decrease)	0.428	0.023	0.370	0.225		0.181	0.912	0.026
Extreme event value	0.373	0.320	0.041	0.050		0.341	0.029	0.141
Average Cases	1.406	0.455	0.638	1.395		0.469	0.527	0.589
X coordinate province	0.981	0.100	0.214	0.291		0.086	0.296	0.015
Y coordinate province	0.157	0.361	0.120	0.050		0.096	0.175	0.072
Event Start Case value	0.783	0.105	0.216	0.360		0.489	0.734	0.270
Event End Case value	2.543	0.806	0.361	0.365		0.204	0.150	0.093
Event Start Mobility value	1.555	0.331	0.273	0.130		0.123	0.135	0.034
Event End Mobility Value	0.031	0.062	0.590	1.043		0.320	0.373	0.799
Amplitude of Covid values	0.378	0.571	0.482	0.409		0.073	0.580	0.200
Amplitude of Mobility Values	0.152	0.274	0.177	0.023		0.361	0.458	0.658

Fig. 16: Feature weights in logistic regression models predicting the expected impacts from mobility (left) and case trends (right).

We model mobility and disease event impacts separately. Given the substantial differences between pandemic waves, we create per-wave models alongside global ones. Results on the balanced model accuracies can be found in Fig. 16, ranging from 70% to 92%. For the first COVID wave, a model for the pandemic event impacts cannot be trained as all events had impact. Figure 16 also shows feature importance as absolute model coefficients, revealing, e.g., a spatial distribution for the first wave mobility event impacts. Notably, the importance of features varied significantly between pandemic waves, suggesting that seasonality and evolving mitigation strategies played a decisive role, while other factors were subordinate. The importance scores of the features can guide the queries described in Sec. 3.2.

5.5 Insights gained

The analysis of trend and impact events provides numerous insights into the dynamics of mobility and COVID-19 incidence. We identified periods with stable cross-impacts, where mobility and COVID-19 cases exhibited consistent dependencies. The strongest linkages existed in early 2020, with mobility reductions playing a key role in controlling disease spread (Fig. 15, middle left). Over time, the influence of mobility on transmission weakened (Fig. 15, middle center and right), likely due to changing public behavior (e.g., wearing masks and social distancing), improved interventions, and vaccination. The persistence of impact events varied across regions (Fig. 15 bottom), indicating heterogeneity in local responses to pandemic conditions. Additional analysis can be done by combining epidemiological data with the dynamics of mobility between provinces, assessing risk of disease transmission, and studying its dynamic impact in space and time.

6 Discussion and Conclusion

A key contribution of our work is a general workflow for detecting and extracting impact events from multiple bivariate time series, followed by visualization and comprehensive analysis of event distributions and relationships across multiple data dimensions (Fig. 2). By leveraging a modified Kendall's tau measure with tolerance, an interactive Ikat plot visualization, and event extraction techniques, we provided an approach for exploring dynamic relationships between time series at varying time lags. This section reflects upon various aspects of our approach and outlines possible future research directions.

6.1 Impact Measures and Parameter Sensitivity

Instead of relying on traditionally applied Granger causality, which poses strong data requirements and is difficult to interpret, we propose using an easy-to-understand measure that captures stable dependencies more intuitively. The introduction of tolerance thresholds in Kendall's tau helped mitigate the effect of minor fluctuations, leading to a more robust detection of significant impact patterns. Our study demonstrated that impact measures are sensitive to parameter choices, including:

- The length of the sliding time window: Shorter windows captured more transient dependencies, while longer windows smoothed fluctuations but risked obscuring fine-grained patterns.
- The threshold for Kendall's tau with tolerance: An appropriate cutoff should be determined by analyzing the distribution of values in areas surrounding potential events. This ensures that detected dependencies are not artifacts of local fluctuations.

The computational complexity of our approach remains manageable, as it scales linearly with respect to the length of time series and quadratically with the time window size. The flexibility of the approach can be enhanced by expressing tolerance in terms of standard deviation or quartile-based measures for reducing influence of outliers.

6.2 Event Extraction, Visualization and Interpretation

Our event extraction method transforms abstract statistical dependencies into interpretable events with temporal and, when relevant, spatial references. Once extracted, these events can be analyzed as a standalone dataset. Understanding larger-scale patterns of event occurrence and relationships requires examining the distribution of these events over time and across relevant data dimensions. For this purpose, we primarily use timeline charts, where one dimension represents time and events are visualized as bars positioned along the time axis. The remaining display dimension represents the set of time series from which the events were extracted, i.e., the locations or entities described by the time series. A meaningful linear arrangement of these items along the dedicated display dimension is crucial for detecting variation patterns among groups of entities or regions in space.

Our case study provides an example of achieving a linear ordering by applying a dimensionality reduction method to the geographic positions of the provinces. This approach tends to group geographically neighboring provinces together in the plot. However, since neither local nor global relationships are strictly preserved, it is essential to validate the detected patterns. To do this, we used geographic maps, but due to space constraints, we omit the details.

The ordering can also be derived by applying dimensionality reduction to other attributes present in the original data, which may be particularly relevant for non-spatial data. A viable approach, always available, is to arrange the original time series based on their similarity. When additional attributes are available, such as overall characteristics of patients in medical studies or products in sales analyses, these attributes can also be utilized to obtain meaningful arrangements. Such arrangements enable analysts to determine whether the distribution of events is related to item similarities.

Since a one-dimensional arrangement cannot fully represent similarity relationships without introducing distortions, it is essential to have tools that help verify whether display neighbors are indeed similar and whether similar items are not separated in the display. One

possible approach is to highlight the most similar items for the item under the cursor, allowing analysts to quickly assess the validity of the arrangement and identify potential inconsistencies.

6.3 Limitations and Generalizability

The metrics we use for impact detection, including the Kendal's Tau, implicitly assume regular temporal sampling of the data and identical temporal resolution across all time series. When this assumption is not fulfilled, or when some values are missing, the data need to be preprocessed by applying resampling and/or value permutation. Presence of measurement errors in the data can also distort the results of impact event extraction and downstream analysis; however, this is a common problem necessitating data cleaning before using any method or analytical workflow.

The proposed approach for impact detection and event extraction is general by design and can be applied to bivariate time series in any domain. However, the following analysis applied to the extracted events may be specific for the application domain and depend on the analysis goals. Our case study should be considered as an example demonstrating a possible purpose of extracting impact events from a large set of bivariate time series, visualization techniques applicable for exploration of the impact distribution and variation, and a possibility of using the events for impact modeling. While these ideas are transferrable to other application domains, other visualization and/or modeling methods may be more suitable.

6.4 Future Research Directions

Several potential extensions of our approach can be explored:

- **Explicit Encoding of Changes**: The Ikat plot always shows the parts of the data satisfying the current query settings and immediately updates in response to any change. The previous state of the display is not visible anymore, and it is hard to understand what has changed. The visualization design could be enhanced by techniques highlighting the changes. For instance, upon user's request, the elements that have changed could start blinking or alternating between the old and new appearance.
- **Full Technique Evaluation**: An obvious limitation of our evaluation study with time series experts was the use of static images, which could affect the interpretability of the visualization. A study with participants using the interactive visual interface is necessary for a comprehensive assessment of the analytical potential of the Ikat plot technique.
- **Comprehensive Threshold Sensitivity Assessment**: Our implementation enables interactive visual exploration of the sensitivity of event detection results to the Kendall's tau tolerance threshold. It also helps to obtain a graph showing how the number of detected events and their numerical characteristics change as the threshold varies within a specified range. However, to support deeper analysis and well-grounded decisions during event extraction, it would be beneficial to include computation and visual representation of threshold influence in space and time.
- **Enhanced Multivariate Analysis**: It would be exciting to find approaches to extend the method from bivariate time series to multidimensional data representing more complex systems, such as chains or networks of impacts.

Our study highlights the importance of combining computational analysis with interactive visual exploration to uncover stable cross-impact patterns in time series data. Future work will focus on refining the methodology and extending its applicability to more complex impact patterns, diverse domains, and real-time scenarios.

Acknowledgments

This work was supported by Federal Ministry of Education and Research of Germany and the state of NRW as part of the *Lamarr Institute for Machine Learning and Artificial Intelligence*, by EU in project *CrexData* (grant agreement no. 101092749), and *MePreCiSa* project of the UNICO I+D Cloud program (REGAGE22e00052915462).

References

[1] G. V. Aher, R. I. Arriaga, and A. T. Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *International conference on machine learning*, pp. 337–371. PMLR, 2023. doi: 10.48550/arXiv.2208.10264 9

[2] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic queries for information exploration: an implementation and evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, 8 pages, p. 619–626. Association for Computing Machinery, New York, NY, USA, 1992. doi: 10.1145/142750.143054 7

[3] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, second edition ed., 2023. doi: 10.1007/978-1-4471-7527-8 2

[4] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer Science & Business Media, 2013. doi: 10.1007/978-3-642-37583-5 3

[5] G. Andrienko, N. Andrienko, S. Bremm, T. Schreck, T. Von Landesberger, P. Bak, and D. Keim. Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum*, 29(3):913–922, 2010. doi: 10.1111/j.1467-8659.2009.01664.x 3

[6] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Politz. Identifying place histories from activity traces with an eye to parameter impact. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):675–688, 2012. doi: 10.1109/TVCG.2011.153 3

[7] N. Andrienko and G. Andrienko. It's about time: Analytical time periodization. *Computer Graphics Forum*, 42(6):e14845, 2023. doi: 10.1111/cgf.14845 3

[8] N. Andrienko and G. Andrienko. Exploring Relationships between Events in Context . In *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2024. doi: 10.2312/eurova.20241114 12

[9] N. Andrienko, G. Andrienko, L. Adilova, and S. Wrobel. Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications*, 42(1):123–133, 2022. doi: 10.1109/MCG.2021.3130314 3

[10] N. Andrienko, G. Andrienko, and G. Shirato. Episodes and topics in multivariate temporal data. *Computer Graphics Forum*, 42(6):e14926, 2023. doi: 10.1111/cgf.14926 3

[11] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner. Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(11):1247–1254, 2020. doi: 10.1016/S1473-3099(20)30553-3 1, 2

[12] S. Bennett, M. Cucuringu, and G. Reinert. Lead–lag detection and network clustering for multivariate time series with an application to the us equity market. *Machine Learning*, 111(12):4497–4538, Dec 2022. doi: 10.1007/s10994-022-06250-4 3

[13] J. Bertin. *Sémiologie Graphique: Les Diagrammes, les Réseaux, les Cartes*. Mouton, 1967. 2, 7

[14] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, 1983. 7

[15] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman. Interactive pattern search in time series. In *Visualization and Data Analysis 2005*, vol. 5669, pp. 175 – 186. International Society for Optics and Photonics, SPIE, 2005. doi: 10.1117/12.587537 2

[16] J. Burmeister, J. Bernard, and J. Kohlhammer. LFPeers: Temporal Similarity Search in Covid-19 Data. In K. Vrotsou and J. Bernard, eds., *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2021. doi: 10.2312/eurova.20211098 3

[17] W. S. Cleveland. *The Elements of Graphing Data*. Wadsworth Advanced Books and Software, 1985. 2

[18] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: 10.1080/01621459.1984.10478080 7

[19] X. Dai, H. Qu, Y. Shen, B. Zhang, Q. Wen, W. Fan, D. Li, J. Tang, and C. Shan. How do large language models understand graph patterns? A benchmark for graph pattern comprehension, 2025. doi: 10.48550/arXiv.2410.05298 9

[20] Z. Deng, S. Chen, T. Schreck, D. Deng, T. Tang, M. Xu, D. Weng, and Y. Wu. Visualizing large-scale spatial time series with geochron. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1194–1204, 2024. doi: 10.1109/TVCG.2023.3327162 3

[21] Z. Deng, J. Huang, C. Ruan, J. Li, S. Gao, and Y. Cai. Volume-based space-time cube for large-scale continuous spatial time series. *IEEE*

Transactions on Visualization and Computer Graphics, pp. 1–15, 2025. doi: 10.1109/TVCG.2025.3537115 3

[22] T. Feng, Y. Hu, J. Yang, T. Polk, Y. Zhao, S. Liu, and Z. Yang. Timepool: Visually answer "which and when" questions on univariate time series. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 201–205, 2023. doi: 10.1109/VIS54172.2023.00049 2

[23] P. Gatalsky, N. Andrienko, and G. Andrienko. Interactive analysis of event data using space-time cube. In *Proceedings of IV' 04*, pp. 145–152, 2004. doi: 10.1109/IV.2004.1320137 10

[24] J. Gonzalez and T. Dang. Outviz: Visualizing the outliers of multivariate time series. In *Proceedings of the 12th International Conference on Advances in Information Technology*, IAIT '21, article no. 37, 5 pages. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3468784.3471606 3

[25] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. doi: 10.2307/1912791 3, 4

[26] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Compututins Surveys*, 51(5), article no. 93, 42 pages, Aug. 2018. doi: 10.1145/3236009 12

[27] T. Hägerstrand. What about people in regional science. *Transport Sociology: Social aspects of transport planning*, pp. 143–158, 1970. doi: 10.1111/j.1435-5597.1970.tb01464.x 10

[28] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004. doi: 10.1057/palgrave.ivs.9500061 2

[29] M. V. Isabel Valdés. Madrid will close its borders for 10 days in early December — english.elpais.com. https://english.elpais.com/spanish_news/2020-11-20/madrid-will-close-its-borders-for-10-days-in-early-december.html. [Accessed 25-02-2025]. 12

[30] D. A. Keim, T. Nietzschmann, N. Schelwies, J. Schneidewind, T. Schreck, and H. Ziegler. A Spectral Visualization System for Analyzing Financial Time Series Data . In *EUROVIS - Eurographics /IEEE VGTC Symposium on Visualization*. The Eurographics Association, 2006. doi: /10.2312/VisSym/EuroVis06/195-202 3

[31] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 06 1938. doi: 10.1093/biomet/30.1-2.81 4

[32] M. G. Kendall. *The Advanced Theory of Statistics, Vol. 2: Inference and Relationship*. Charles Griffin & Company, 1945. 5

[33] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, Mar 2005. doi: 10.1007/s10115-004-0154-9 2

[34] M. Kraak. *Mapping time : illustrated by Minard's map of Napoleon's Russian campaign of 1812*. ESRI, United States, 2014. 2

[35] M.-J. Kraak. The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference*, pp. 1988–1996, August 2003. 10

[36] J. Li, S. Chen, K. Zhang, G. Andrienko, and N. Andrienko. COPE: Interactive Exploration of Co-Occurrence Patterns in Spatial Time Series . *IEEE Transactions on Visualization & Computer Graphics*, 25(08):2554–2567, Aug. 2019. doi: 10.1109/TVCG.2018.2851227 3

[37] S. Liu, Y. Tian, Z. Deng, W. Cui, H. Zhang, D. Weng, and Y. Wu. Relation-driven query of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 31(8):4210–4225, 2025. doi: 10.1109/TVCG.2024.3397554 2

[38] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438(5):237–329, 2007. doi: 10.1016/j.physrep.2006.11.001 3

[39] L. Meng, S. Liu, K. Yang, J. Xu, Z. Deng, D. Weng, and Y. Wu. Chronodeck: A visual analytics approach for hierarchical time series analysis. *IEEE Transactions on Visualization and Computer Graphics*, 31(12):10488–10502, 2025. doi: 10.1109/TVCG.2025.3602273 2

[40] T. Munzner. *Visualization analysis and design*. CRC press, 2014. 7

[41] B. D. Q. Nguyen, R. Hewett, and T. Dang. Congnostics: Visual Features for Doubly Time Series Plots. In *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2020. doi: 10.2312/eurova.20201086 3

[42] M. Ponce-de Leon, J. Del Valle, J. M. Fernandez, M. Bernardo, D. Cirillo, J. Sanchez-Valle, M. Smith, S. Capella-Gutierrez, T. Gullón, and A. Valencia. Covid-19 flow-maps an open geographic information system on covid-19 and human mobility for spain. *Scientific Data*, 8(1):310, 2021. doi: 10.1038/s41597-021-01093-5 8, 9

[43] M. Ponce-de Leon, J. del Valle, J. M. Fernández, M. Bernardo, D. Cirillo, J. Sanchez-Valle, M. Smith, S. Capella-Gutierrez, T. Gullón, and A. Valencia. COVID19 Flow-Maps daily cases reports, 2021. doi: 10.5281/zenodo.5217386 9

[44] M. Ponce-de Leon, J. del Valle, J. M. Fernández, M. Bernardo, D. Cirillo, J. Sanchez-Valle, M. Smith, S. Capella-Gutierrez, T. Gullón, and A. Valencia. COVID19 Flow-Maps daily-mobility for Spain, 2021. doi: 10.5281/zenodo.5539411 9

[45] M. Ponce-de Leon, J. del Valle, J. M. Fernández, M. Bernardo, D. Cirillo, J. Sanchez-Valle, M. Smith, S. Capella-Gutierrez, T. Gullón, and A. Valencia. COVID19 Flow-Maps population data, 2021. doi: 10.5281/zenodo.5226351 9

[46] Y. Qiang, M. Delafontaine, M. Versichele, P. D. Maeyer, and N. V. de Weghe. Interactive analysis of time intervals in a two-dimensional space. *Information Visualization*, 11(4):255–272, 2012. doi: 10.1177/1473871612436775 3

[47] D. Rosenberg and A. Grafton. *Cartographies of Time: A History of the Timeline*. Princeton Architectural Press, 2013. 10

[48] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, 1978. doi: 10.1109/TASSP.1978.1163055 2

[49] J. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18(5):401–409, 1969. doi: 10.1109/T-C.1969.222678 10

[50] T. Schreck, J. Bernard, T. von Landesberger, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive kohonen maps. *Information Visualization*, 8(1):14–29, 2009. doi: 10.1057/ivs.2008.29 3

[51] Z. Shafie and K. Ghazizadeh. A comparative study between the colors and patterns of ikats in iran and uzbekistan. *Journal of Art and Civilization of the Orient*, 10(35):29–44, 2022. doi: 10.22034/jaco.2022.325930.1230 1

[52] D. J. Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC Press, 6th ed., 2020. doi: 10.1201/9780429186196 5

[53] S. Siegel and N. J. Castellan. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 2nd ed., 1988. 5

[54] M. Smith, M. Ponce-de Leon, and A. Valencia. Evaluating the policy of closing bars and restaurants in cataluña and its effects on mobility and covid19 incidence. *Scientific Reports*, 12(1):9132, 2022. doi: 10.1038/s41598-022-11531-y 11

[55] M. Steiger, J. Bernard, S. Mittelstädt, H. Lücke-Tieke, D. Keim, T. May, and J. Kohlhammer. Visual analysis of time-series similarities for anomaly detection in sensor networks. *Computer Graphics Forum*, 33(3):401–410, 2014. doi: 10.1111/cgf.12396 3

[56] J. Suschnigg, B. Mutlu, G. Koutroulis, H. Hussain, and T. Schreck. Mandala—visual exploration of anomalies in industrial multivariate time series data. *Computer Graphics Forum*, 44(1):e70000, 2025. doi: 10.1111/cgf.70000 3

[57] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977. 2

[58] D. Van Wesenbeeck, A. Yurtman, W. Meert, and H. Blockeel. Locomotif: discovering time-warped motifs in time series. *Data Mining and Knowledge Discovery*, 38(4):2276–2305, Jul 2024. doi: 10.1007/s10618-024-01032-z 2

[59] J. Van Wijk and E. Van Selow. Cluster and calendar based visualization of time series data. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)*, pp. 4–9, 1999. doi: 10.1109/INFVIS.1999.801851 3

[60] I. Vasiliev. Mapping time. *Cartographica*, 34(2):1–51, 1997. doi: 10.3138/D357-234G-2M62-4373 2

[61] J. Wang and K. Mueller. The visual causality analyst: An interactive interface for causal reasoning. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):230–239, 2016. doi: 10.1109/TVCG.2015.2467931 3

[62] J. Wang and K. Mueller. Visual causality analysis made practical. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 151–161, 2017. doi: 10.1109/VAST.2017.8585647 3

[63] J. Wang and K. Mueller. Domino: Visual causal reasoning with time-dependent phenomena. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5342–5356, 2023. doi: 10.1109/TVCG.2022.3207929 3

[64] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019. 7

[65] W. Xi, T. Pei, Q. Liu, C. Song, Y. Liu, X. Chen, J. Ma, and Z. Zhang.

Quantifying the time-lag effects of human mobility on the covid-19 transmission: a multi-city study in china. *IEEE Access*, 8:216752–216761, 2020. doi: 10.1109/ACCESS.2020.3038995 1, 2

[66] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh. Matrix Profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 1317–1322, 2016. doi: 10.1109/ICDM.2016.0179 3

[67] Y. Zhang, H. Wang, S. Feng, Z. Tan, X. Han, T. He, and Y. Tsvetkov. Can LLM graph reasoning generalize beyond pattern memorization?, 2024. doi: 10.48550/arXiv.2406.15992 9

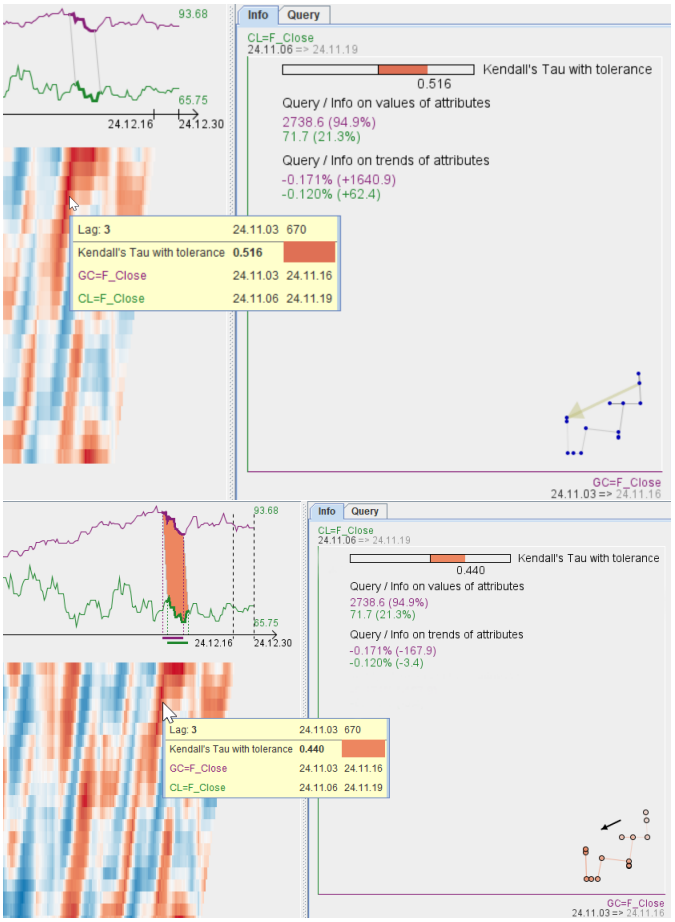


Fig. 17: Figure 6 before (top) and after (bottom) improvement.

A Improvements of Visualization

In response to consultations with several information visualization experts, the visual design of the Ikat plot tool has been revised. Figure 17 illustrates the visualization before (top) and after (bottom) these revisions; a complete updated screenshot is also available in the revised Fig. 1.

Specifically, the active intervals of the base and response variables are now explicitly represented by colored bars below the time axis. To better communicate the relationship between these variables, the area between the affected portions of the two time series is filled with a color indicating the impact value. Additionally, the visual clarity of the scatter plot has been improved to facilitate more precise recognition of the two variables co-evolution patterns.

B Algorithms for Event Extraction

To find candidate regions for constructing events, the method computes, for each matrix cell, a binary value $b[i][j]$ that is *true* if cell (i, j) satisfies all query conditions. The resulting boolean matrix b is supplied to Algorithm 2, which recursively calls Algorithm 3, to identify contiguous regions containing *true* values. The function *isValidEvent* checks whether the detected regions satisfy additional query constraints specifying the minimal event duration (the horizontal extent), minimal depth (the range of lags covered), and numeric characteristics of the event's shape. If so, these regions are used to construct events.

C Conceptual Comparison with Existing Approaches

In this section, we briefly compare our Ikat plot-based methodology to established approaches for analyzing relationships between time series pairs, particularly Cross-Recurrence Plots (CRP) and Dynamic Time Warping (DTW). While these methods have been widely adopted, the Ikat plot approach differs in several essential conceptual ways:

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2026.3676810

Algorithm 2 Extract Events

Input: Boolean matrix b
Output: List of events *events*
$events \leftarrow []$ ▷ Initialize an empty list of events
Initialize a 2D boolean array *visited* with the same dimensions as b
for each row r in b **do**
  for each column c in $b[r]$ **do**
    if $b[r][c] =$ true **and** $visited[r][c] =$ false **then**
      $event \leftarrow$ new Event() ▷ Create a new Event object
      dfs(b, visited, r, c, event) ▷ Depth-First Search
      if isValidEvent$(event, b)$ **then**
        $events \leftarrow events \cup \{event\}$ ▷ Add event to list
      end if
    end if
  end for
end for
return *events*

Algorithm 3 Depth-First Search for Event Exploration

Input: Boolean matrix b, Visited matrix *visited*, row, column, Event object *event*
Output: Updated Event object
if row is out of bounds **or** col is out of bounds **or** visited[row][col] = true **or** b[row][col] = false **then return**
end if
$visited[row][col] =$ true ▷ Mark the current cell as visited
$event.coordinates.add($new Point$(col, row))$ ▷ Add the current cell to the event
▷ Explore neighbors (Up, Down, Left, Right)
dfs(b, visited, row - 1, col, event) ▷ Up
dfs(b, visited, row + 1, col, event) ▷ Down
dfs(b, visited, row, col - 1, event) ▷ Left
dfs(b, visited, row, col + 1, event) ▷ Right

- **From Similarity to Impact.** CRP and DTW focus on quantifying similarity or alignment between time series, typically capturing periods where both series evolve in similar ways. In contrast, the Ikat plot is designed to reveal *impact* relationships, which may be positive (e.g., increases in one series lead to increases in another) or negative (e.g., increases in one series lead to decreases in the other). This allows for the detection of asymmetric and potentially causal relationships, rather than just co-movement.
- **Explicit Time Lag Consideration.** While CRP and DTW can capture temporal misalignments, they do not directly expose or analyze the range of possible time lags at which impacts may occur. The Ikat plot explicitly maps impact strength across a user-defined range of lags, enabling the discovery of lead-lag effects and delayed responses.
- **Reduced Sensitivity to Noise.** Standard similarity-based methods can be overly sensitive to short-term fluctuations and noise, which may obscure meaningful patterns. Our use of a tolerance-based Kendall's tau measure within the Ikat plot framework treats minor fluctuations as ties, focusing attention on substantial, stable impact patterns.
- **Interpretability of Patterns.** The visual encoding in CRP and DTW often emphasizes global structure or synchronization, but may lack direct interpretability for complex lagged relationships. The Ikat plot, by contrast, enables analysts to recognize and interpret local, lag-specific, and temporally stable cross-impacts through intuitive geometric patterns (e.g., vertical, diagonal stripes, or hotspots).
- **Local Patterns versus Global Similarity.** DTW is optimized to find a single global alignment path that minimizes total distance between two time series, potentially overlooking local but meaningful impact episodes. The Ikat plot supports the identification and extraction of such localized events, which may only occur during specific intervals or at selected lags.
- **Event Extraction for Downstream Analysis.** Unlike CRP and DTW, which primarily produce similarity scores or recurrence matrices, the Ikat-based workflow includes explicit event extraction. This enables downstream quantitative and visual analysis of impact events across multiple time series pairs, supporting further modeling and hypothesis generation.

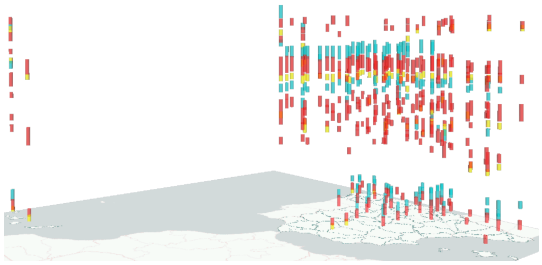


Fig. 18: Impact events **Cases decrease → Trips increase** 🔵, **Trips increase → Cases increase** 🟡, and trend events **Cases increase** 🔴 are shown in space-time cube.

D Example of Geographic Visualization of Extracted Events

Figure 18 is an example of possible geographic visualization of events. Alternative representations include maps with colored dots representing individual events accompanied by interactive time filter controls, map with diagrams showing counts of events of different type over time periods etc.

E Evaluation of Ikat Plot Interpretability

The goal of this evaluation is to assess the representational power and interpretability of the Ikat plot for time-series experts. Our approach comprised: (1) a preliminary exploratory study using five Large Language Models (LLMs), (2) a study with four human experts in time-series analysis, and (3) an alignment of these findings with a curated chronology of Spain's COVID-19 policy and epidemiological milestones.

Each agent (human or LLM) received a description of the analytical framework (Section 3) and Ikat plot visualizations for three Spanish regions (Locations 08, 28, and 41; Fig. 19), using a predefined query: "*increase in trips followed by increase in cases after a time lag of* ≥ 7 *days*."

E.1 Exploratory LLM-Based Check of Task and Stimuli

We utilized five LLMs (GPT-4.1, GPT-4o, o4-mini, Claude 3 Opus, and Gemini 1.5 Pro) to conduct a rapid, reproducible "stress test" of the visualization's clarity. The models consistently recognized key patterns, including sustained vertical bands, episodic events, and lag-specific structures.

The models largely converged on location-level characterizations: identifying Location 28 as having the most stable sustained patterns, Location 08 as having clear episodic structures, and Location 41 as exhibiting high complexity with overlapping events. This convergence suggests that our prompts and explanatory text were clear enough for the models to recover the intended pattern descriptions. However, we do not treat this as evidence about human interpretability; our conclusions about visualization effectiveness for human analysts are based solely on the expert study reported in Sections 4.1 and E.2.

E.2 Human Expert Evaluation

For human expert evaluation, we engaged four professionals in time series analysis, referred to as E1–E4. Given their tight work schedules and very limited availability, we designed a streamlined assessment process. Rather than requiring them to learn and interact with our full implementation, which would require substantial time investment,

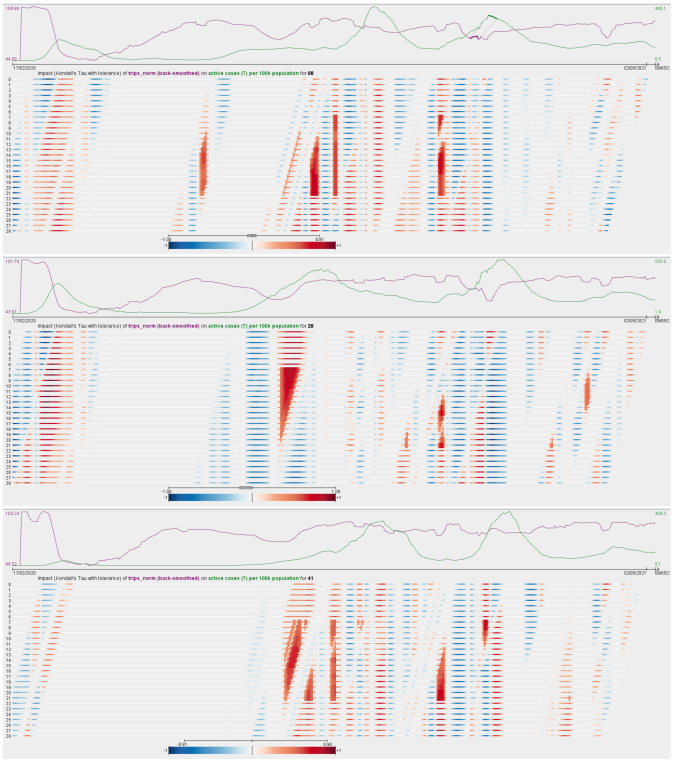


Fig. 19: kat plots used in the evaluation: Locations 08 (Barcelona), 28 (Madrid), and 41 (Seville).

we provided static Ikat plot images (Fig. 19) along with structured evaluation forms. Detailed instructions are attached as a separate PDF file. Experts were asked to identify patterns, estimate lags and impact strengths, and provide interpretations based solely on static images. While this lacks the depth of interactive analysis, their expertise enabled them to extract meaningful insights. This approach allowed us to gather valuable feedback within tight time constraints.

Below we summarize the experts' findings by location.

Location 08 (Barcelona):

- E1: Multiple impact events including sustained (lags 7, 10–11 days; strengths up to strong, $\tau \approx 0.8$; persistence 10–14 days) and transient (lag ≈10 days, medium). Significant impacts limited to lags < 21 days. Visual cues (solid vs. narrow stripes; red/blue polarity) aid detection.
- E2: Patterns described as "stable trend-driven rhomboids" with lags ≈ 12–22 days, moderate strength, concentrated in 2020 transitions (summer, back-to-school, Christmas). Absence of strong 2022 events consistent with vaccination.
- E3: Five major events (transient, lag-specific, sustained), delays 7–11 days, persistence 10–14 days. One sustained event shows a gap (likely Christmas mobility dip). Later events exhibit slightly shorter delays (possible faster reporting or variant effects).
- E4: Five major sustained events and one episodic. Contrasted diagonal vs. flat bounds (selection cut-offs and onset timing). Warned of overemphasis on high-frequency mobility fluctuations relative to slower case trends.

Consensus: Repeated mobility → case associations in 2020–early 2021 with typical lags of 7–14 days, some longer-lag structure (12–22 days). E4 highlighted methodological cautions on typology clarity, lag cut-offs, and level-shift effects.

Location 28 (Madrid):

- E1: Mix of stable trend-driven, sustained, and episodic events; lags 7–20 days; strengths weak-to-strong (up to $\tau \approx 0.7$). White gaps and alternating colors indicate ambiguity.
- E2: Identified summer triangle (lags ≈ 7–22) and a late-year pattern (lags ≈14–21). Noted that Madrid's Non-Pharmaceutical Interventions (NPIs) were often lighter or differently timed than elsewhere.
- E3: One dominant, very strong sustained event with ≈7-day delay (second wave), plus smaller episodic patches. Strength decays with longer lags, consistent with incubation/reporting delays.
- E4: Six events including sustained, episodic, and stable trend-driven types. Main vertical/area-spanning event plus medium-strength sustained patterns linked to oscillations/lockdowns. Causal interpretation uncertain; precise dating difficult without interactivity.

Consensus: A major sustained pattern during the late-2020/early-2021 wave, with additional shorter events (summer, late-year). Lags mainly 7–14 days, extending up to ≈21. White/ambiguous intervals indicate weak coupling.

Location 41 (Seville):

- E1: Multiple events (stable trend-driven, lag-specific, episodic); lags 7–21 days. Two clusters precede large infection increases; noted gaps and alternating colors.
- E2: No form provided.
- E3: Several sustained/transient events (second and third waves). One later event with ≈16-day lag (possible secondary transmission). Holiday-adjacent sustained event (Christmas) with ≈7-day delay.
- E4: Nine events including one major stable trend-driven diagonal band, multiple medium sustained links, and late events on high mobility plateaus. Interpretation challenging; sensitive to parameter settings. Conjectured that some late mobility variations may not affect epidemic dynamics.

Consensus: Mixed sustained/episodic structure with typical lags 7–14 days and longer-lag features (≈16–21 days). Interpretation complexity is higher due to overlapping patterns and mobility-level effects.

E.3 Alignment with Spain's COVID-19 Chronology

Expert observations showed a high degree of alignment with historical milestones (www.andalucia.com/health/coronavirus/chronology). Key matches include:

- **Spring/Summer 2020:** Reopening and de-escalation periods corresponded with 7–14 day lag clusters.
- **Winter 2020–21:** Holiday mobility relaxations were clearly visible as strong sustained clusters, particularly in Madrid (Location 28).
- **Late 2021–22:** Vaccination and reduced case reporting correlated with the observed sparsity of strong impact events.

E.4 Cross-Expert Convergence and Divergences

Convergence

- Lags: Predominantly 7–14 days; longer-lag structure (up to ≈21–22 days) noted by multiple experts.
- Temporal clustering: Consistently identified in spring 2020 reopening, summer 2020, and late 2020/early 2021.
- Typology: Sustained (vertical bands), stable trend-driven (rhomboids/triangles), episodic, transient, and lag-specific patterns coexist.
- Mixed intervals: White gaps/alternating colors reflect weak or ambiguous associations.

Divergences and Caveats

- **Strength calibration:** E1 reported higher peaks ($\tau \approx 0.8$ for 08) compared to E2's "not very high" ratings for similar areas, likely due to static-image color-scale perception.
- **Typology boundaries:** E4 called for clearer criteria/examples distinguishing sustained vs. stable trend-driven vs. episodic/transient patterns, and emphasized the influence of selection cut-offs.
- **Lag policy:** E4 cautioned that the ≥ 7 day minimum lag may hide contemporaneous/shorter-lag effects; suggested sensitivity analyses.
- **Confounding:** E4 highlighted artifacts from ordering/level shifts (events on high mobility plateaus may not causally drive cases), recommending detrending/controls.
- **Event dating:** Several experts noted difficulty in precise temporal localization with static images, underlining the limitations of our assessment setup.

These results confirm that while the Ikat plot is sufficiently interpretable in static form, interactivity is essential for detailed sensitivity analysis.