Descubrimiento causal sobre vectores en series temporales no estacionarias

Joaquín Mateos[0009−0002−8189−3236], Antonio R. Moya[0000−0002−3657−4749], and Sebastián Ventura[0000−0003−4216−6378]*

Departamento de Ciencias de la Computación e Inteligencia Artificial, Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DasCI), Universidad de Córdoba, Córdoba, 14071, España
sventura@uco.es

Resumen El descubrimiento causal en series temporales multivariantes es fundamental para comprender los mecanismos de sistemas complejos. El estado del arte en dicha tarea presenta grandes dificultades a la hora de tratar con variables agrupadas de alta dimensionalidad y comportamientos no estacionarios, debido a las suposiciones realizadas en su desarrollo. Para superar estas limitaciones, se propone aplicar un descubrimiento causal en dos fases: en primer lugar se emplea un iVAE [9] para encontrar, para cada grupo, una representación latente identificable y de menor dimensionalidad, apoyándose precisamente en la no-estacionariedad de las series. A continuación, sobre esta representación, se adapta el algoritmo PCMCI [21], perteneciente al estado del arte para descubrimiento causal sobre series temporales, para su aplicación a nodos multivariantes y series no estacionarias. La experimentación en escenarios sintéticos con varios entornos heterogéneos demuestra que, en el caso de que el conjunto de series temporales esté afectado por cambios de régimen y no haya conocimiento explícito de ellos (como suele ser en entornos reales), nuestra propuesta mantiene un comportamiento robusto tanto en términos de F1 como de SHD, superando a los algoritmos en el estado del arte.

Keywords: *Descubrimiento causal · Grupos de series temporales · Identificabilidad.*

1. Introducción

La inferencia causal en series temporales se ha convertido en una herramienta fundamental para comprender los mecanismos subyacentes en sistemas complejos, con aplicaciones en diversos ámbitos, desde climatología [11,16] hasta economía [8]. Un paso fundamental, tanto en entornos científicos como industriales, es la construcción de un grafo que designe las relaciones entre distintas variables de estudio. En muchos casos no es posible disponer de expertos que puedan definir estas relaciones, e incluso en el caso de sí tenerlos, la designación del conjunto de todas las variables que afectan a cada dato de estudio puede suponer un problema nada trivial, especialmente en contextos de alta dimensionalidad. Por ello,

* Autor de correspondencia.

existe una rama creciente en la última década, llamada Descubrimiento Causal sobre series temporales (ts-CD, por sus siglas en inglés), que explora el uso de una variedad de métodos estadísticos y computacionales para encontrar el grafo que mejor represente las relaciones observadas en un conjunto de datos [2,5].

Dentro del ts-CD, el caso de alta dimensionalidad entraña especiales problemas debido a la posible creación de relaciones espurias y dificultad de búsqueda de relaciones intrincadas en los datos, por parte de los distintos algoritmos. Por ello es de especial interés el estudio de algoritmos que no se centren en encontrar estas relaciones variable a variable, sino que, a partir de conjuntos de variables, típicamente bien definidas en un entorno real — por ejemplo, sensores en ciertos sistemas de un vehículo, o datos climatológicas en áreas bien diferenciadas — sean capaces de encontrar las relaciones causales entre los grupos.

El resto del artículo está estructurado de la siguiente forma: en la Sección 2 se da un breve resumen de los trabajos previos realizados sobre esta temática y los problemas específicos que enfrentan. En la Sección 3 se expone nuestra propuesta para solucionar estos problemas, y en la Sección 4 se explican los experimentos realizados junto a sus resultados. Finalmente, en la Sección 5 se resumen y concluyen las aportaciones realizadas en este artículo.

2. Antecedentes

Los fundamentos teóricos acerca del descubrimiento causal sobre grupos de variables fueron establecidos recientemente en [25]. Aquí se propusieron, además, los acercamientos más sencillos al descubrimiento causal en este contexto, como el ***Micro-Level***. Este método, en primer lugar, utiliza un algoritmo de descubrimiento causal directamente sobre las variables observadas, $\mathcal{D} = \{X_t^1, ..., X_t^M\}$, construyendo así un grafo sobre ellas, y aplica su definición de grafo a nivel de grupos, $\boldsymbol{X}_t^1, ..., \boldsymbol{X}_t^N \in 2^{\mathcal{D}}$. Según esta definición, existe una relación causal entre dos grupos cuando esta existe al menos entre dos variables del grupo, en la misma dirección:

$$\boldsymbol{X}_{t-\tau}^i \to \boldsymbol{X}_t^j \iff \exists X_{t-\tau}^{i'} \in \boldsymbol{X}_{t-\tau}^i,\ X_t^{j'} \in \boldsymbol{X}_t^j \quad \text{t.q.}\ \ X_{t-\tau}^{i'} \to X_t^{j'} \tag{1}$$

Sin embargo, los mismos autores advierten que esta definición puede pecar de permisiva, provocando que cada fallo a la hora de crear una sola arista a nivel de nodo cree también una arista a nivel grupal, llevando a tener muchos falsos positivos. Así, proponen el uso de una reducción de dimensionalidad dentro de cada grupo, $h : \boldsymbol{X}_t^i \mapsto \boldsymbol{Z}_t^i$, con $\dim(\text{soporte}(\boldsymbol{Z}_t^i)) \leq \dim(\text{soporte}(\boldsymbol{X}_t^i))$, para evitar esta explosión de aristas, al tener que realizar menos comprobaciones reales de aristas.

Un problema evidente que surge al decidir reducir esta dimensionalidad, como se explica en [15], es la elección de la dimensionalidad objetivo, que debe elegirse con anterioridad a reducciones de dimensionalidad estándar como el Análisis de Componentes Principales (PCA). Por ello proponen el empleo de una Agregación Adaptativa (**Adag**), que elija la dimensión de la representación como la menor dimensión que alcance un umbral de cierta métrica. Tras diversos resultados teóricos y empíricos concluyen que la métrica cuya maximización más

información permite preservar para algoritmos de descubrimiento causal basados en restricciones es la *métrica de independencia para consistencia de agregación*: $c_{ind}(\boldsymbol{Z}_t, \boldsymbol{X}_t) = \frac{|\mathfrak{C}|}{|\mathfrak{C}|+|\mathfrak{I}|}$, donde $\mathfrak{C}$ denota el conjunto de independencias entre grupos que pueden encontrarse en el espacio de dimensionalidad reducida, $\boldsymbol{Z}_t$ y también se cumplen sobre el espacio original, $\boldsymbol{X}_t$; mientras que en $\mathfrak{I}$ están las independencias que pueden encontrarse sobre $\boldsymbol{Z}_t$ pero son falsas al estudiarlas sobre los grupos de $\boldsymbol{X}_t$.

De forma independiente a esta línea, en la literatura se han diseñado dos algoritmos basados en redes neuronales. **GroupRESIT** [4] asume modelos de ruido aditivo no lineal (GANM). Se extrae el orden causal comprobando la independencia de los residuos en redes profundas, para refinar este orden en un grafo definitivo mediante una regularización MURGS [24], eliminando así relaciones espurias entre grupos. Otra propuesta basada en redes neuronales es **gCDMI** [1], el cual primero utiliza la metodología de DeepAR [22] para capturar relaciones no lineales entre series temporales multivariantes. A continuación infiere una causalidad $\boldsymbol{X}^i_{t^-} \to \boldsymbol{X}^j_t$ realizando intervenciones basadas en reemplazar la causa $\boldsymbol{X}^i_t$ por *knockoffs*, y confirma los enlaces causales cuando estos residuos tienen una alteración significativa.

Tras estudiar los métodos actuales para el descubrimiento causal sobre grupos se han detectado dos problemas no tratados, de especial importancia en el contexto de series temporales:

Generación de una representación identificable: como se demuestra en el Teorema 1 de [14], cuando se utilizan distribuciones *a priori* incondicionales sobre el espacio latente, $p(\boldsymbol{z})$, como ocurre en los *Autoencoders* Variacionales (VAEs) [10], las variables latentes obtenidas pueden resultar completamente entrelazadas, incluso cuando existe una representación original no entrelazada. Por otro lado, un *autoencoder* no variacional, como se explica en [9], tampoco proporciona una representación adecuada para la inferencia causal debido a su falta de identificabilidad[1]. Cuando un método de reducción de dimensionalidad no es identificable, las transformaciones arbitrarias admisibles en el espacio latente pueden violar fácilmente la suposición de fidelidad causal [17] al ocultar o cancelar dependencias subyacentes. Por ejemplo, si tenemos dos grupos con variables de misma varianza, $\boldsymbol{X}^1_t = \{X^1_t, X^2_t\}$, $\boldsymbol{X}^2_t = \{X^3_t\}$, con $\boldsymbol{X}^1_{t-1} \to \boldsymbol{X}^2_\tau$, con una relación $X^3_t = X^1_{t-1} + X^2_{t-1} + \varepsilon$, entonces una reducción de dimensionalidad podría darnos las variables $Z^1_t = X^1_t - X^2_t$; $Z^2_t = X^3_t$, en cuyo caso sería $Z^1_{t-1} \perp\!\!\!\perp Z^2_t$, y ningún algoritmo de descubrimiento causal podría encontrar la relación $Z^1_{t-1} \to Z^2_t$.

[1] Según la definición clásica de identificabilidad, Definición 5.2 de [12], un modelo estadístico definido por una familia de distribuciones $\mathcal{P}_\Theta = \{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$, es identificable respecto de una relación de equivalencia $\sim$ cuando si $p_{\boldsymbol{\theta}}(\boldsymbol{x}) = p_{\boldsymbol{\theta}'}(\boldsymbol{x})$ en casi todo punto $\boldsymbol{x}$, entonces $\boldsymbol{\theta} \sim \boldsymbol{\theta}'$.

En el Apéndice D.2 de [9] se demuestra que los modelos clásicos, con cualquier tipo de *prior* incondicional $p_{\boldsymbol{\theta}}(\boldsymbol{z})$, son inidentificables, tanto considerando como equivalentes las permutaciones de las variables, $\sim_P$, como considerando equivalentes las transformaciones lineales invertibles, $\sim_A$.

Series temporales no estacionarias: los métodos tradicionales de descubrimiento causal asumen estacionariedad de las series temporales. Sin embargo, en la mayoría de los sistemas complejos reales, las dinámicas cambian constantemente. Romper esta suposición supone un problema crítico a nivel teórico y, como se verá en la Sección 4, también a nivel práctico.

Un enfoque particularmente útil para modelar esta no estacionariedad es el estudio de los cambios de régimen, es decir, se asume que las distribuciones de las variables del sistema pasan por diferentes estados o "entornos" a lo largo del tiempo (por ejemplo, transiciones entre distintas estaciones del año o cambios en el entorno por el que transita un vehículo).

Trabajos como los basados en el descubrimiento causal a partir de datos no estacionarios [7] o la predicción causal invariante [19,20] demuestran que los verdaderos mecanismos causales tienden a permanecer invariantes a través de los distintos regímenes. Por ello, formalizar estos cambios de régimen mediante una variable de entorno indicadora permite aislar la no estacionariedad y, como se realiza en nuestra propuesta, aprovechar esta señal adicional para garantizar la identificabilidad de representaciones latentes complejas, mejorando así el descubrimiento causal incluso bajo el supuesto de que los cambios de regímenes no son plenamente conocidos. En la Figura 1 se muestra un ejemplo de estos cambios de régimen, basados en cambios de media y desviación típica, dentro de series temporales causalmente relacionadas.

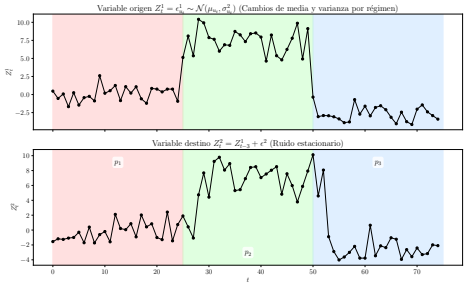


Figura 1. Ejemplo de cambios de régimen sobre dos variables causalmente relacionadas, $Z_{t-3}^1 \rightarrow Z_t^2$. Se aprecia cómo el cambio de régimen en Z_t^1 deriva en un cambio de régimen de Z_{t+3}^2, a pesar de que el ruido de esta segunda variable sea estacionario.

3. Propuesta

Sea $\boldsymbol{X}_t = \{\boldsymbol{X}_t^1, ..., \boldsymbol{X}_t^N\} \subset 2^{\mathcal{D}}$ una serie temporal multivariante dividida en N grupos disjuntos, $\boldsymbol{X}_t^i = \{X_t^{m_1}, ..., X_t^{m_{M_i}}\}$.

Sea $\boldsymbol{u}_t$ una variable no estacionaria de fondo en el instante t. En nuestro caso, $\boldsymbol{u}_t$ se modela como un vector *one-hot* categórico $\boldsymbol{u}_t \in \{0,1\}^{n_{regimes}}$, que representa distintos regímenes, o una división por bloques igualmente espaciados de un índice temporal continuo.

El objetivo del algoritmo, cuyo esquema general se ilustra en la Figura 2, es reconstruir el grafo causal que representa las dependencias temporales entre los grupos de variables $\boldsymbol{X}_t^i$. Para ello, el proceso consta de dos fases principales: en primer lugar, se aísla la influencia de la no estacionariedad y la interdependencia de las series temporales mediante su proyección en un espacio latente identificable. A continuación, se lleva a cabo el descubrimiento causal aplicando pruebas de independencia condicional directamente sobre este espacio latente.

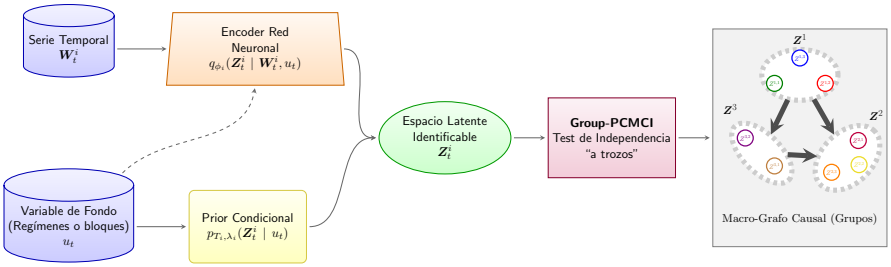


Figura 2. Esquema general del marco algorítmico propuesto. Las series temporales observadas ($\boldsymbol{X}_t$) y la variable de fondo ($\boldsymbol{u}_t$) actúan como entrada del iVAE (compuesto por el *encoder* y el *prior*). Este modelo aísla la no estacionariedad y proyecta los datos en un espacio latente identificable ($\boldsymbol{Z}_t$), sobre el que se estudian las independencias mediante el algoritmo Group-PCMCI para inferir el grafo causal subyacente.

3.1. Identificabilidad del espacio latente mediante iVAE

Para cada grupo $\boldsymbol{X}_t^i$, el algoritmo extrae una representación latente de baja dimensionalidad, $\boldsymbol{Z}_t^i$, utilizando un *Autoencoder* Variacional identificable (iVAE) [9]. Como explican los autores, este tipo de VAE solventa el problema de la inidentificabilidad introducido en la Sección 2. Para ello, emplea nuestra variable de fondo no estacionaria, $\boldsymbol{u}_t$, de tal forma que el *prior* del VAE deje de ser incondicional y pase a ser $p_{\boldsymbol{T}_i,\boldsymbol{\lambda}_i}(\boldsymbol{Z}_t^i \mid \boldsymbol{u}_t)$. Este condicionamiento rompe las ambigüedades del espacio latente; de forma intuitiva, la variabilidad de $\boldsymbol{u}_t$ restringe el conjunto de soluciones válidas para el modelo, forzando al *decoder* a aprender un único mecanismo de generación que asocia a estas distribuciones no estacionarias los datos que se observan en cada régimen.

Formalmente, este modelo define una distribución conjunta: $p_{\boldsymbol{\theta}_i}(\boldsymbol{X}_t^i, \boldsymbol{Z}_t^i \mid \boldsymbol{u}_t) = p_{\boldsymbol{f}_i}(\boldsymbol{X}_t^i \mid \boldsymbol{Z}_t^i) p_{\boldsymbol{T}_i,\boldsymbol{\lambda}_i}(\boldsymbol{Z}_t^i \mid \boldsymbol{u}_t)$. Esta formulación divide la influencia de la componente no estacionaria en dos partes. Por un lado, un *prior* condicional $p_{\boldsymbol{T}_i,\boldsymbol{\lambda}_i}$, que asume una distribución perteneciente a la familia exponencial, definida por sus estadísticos suficientes $\boldsymbol{T}_i$ (que en nuestro caso asumen una distribución gaussiana), y sus parámetros naturales, $\boldsymbol{\lambda}_i$, fuertemente dependientes de la información de fondo $\boldsymbol{u}_t$. Por otro, un modelo generador $p_{\boldsymbol{f}_i}$, que asocia al espacio latente las observaciones mediante un *decoder* $\boldsymbol{f}_i$, sometido a un ruido aditivo independiente.

Dado que el *posterior* real es intratable, el modelo se entrena maximizando la Cota Inferior de la Evidencia (ELBO) a través de una aproximación variacional $q_{\boldsymbol{\phi}_i}(\boldsymbol{Z}_t^i \mid \boldsymbol{X}_t^i, \boldsymbol{u}_t)$. A nivel computacional, el modelo se implementa mediante redes neuronales: una red neuronal sencilla, como un perceptrón multicapa, procesa $\boldsymbol{u}_t$ para predecir los parámetros $\boldsymbol{\lambda}_i$, mientras que $q_{\boldsymbol{\phi}_i}$ y $\boldsymbol{f}_i$ conforman las arquitecturas del *encoder* y *decoder*, respectivamente.

Es precisamente la variabilidad temporal inyectada por $\boldsymbol{u}_t$ en los parámetros del *prior* lo que proporciona el anclaje necesario para garantizar teóricamente la recuperabilidad de las verdaderas fuentes latentes (identificabilidad $\sim_A$ y $\sim_P$). Finalmente, para la posterior fase de descubrimiento causal se utiliza la media de

estas distribuciones latentes inferidas, ya que es el estadístico que concentra la señal principal de la representación, quedando el resto de parámetros destinados a modelar la incertidumbre [14].

3.2. Descubrimiento causal sobre el espacio generado (Group-PCMCI)

Una vez inferidas las secuencias latentes identificables $\boldsymbol{Z}_t^i$ asociadas a cada grupo, procedemos a modelar sus relaciones causales temporales. Para resolver problemas fundamentales del descubrimiento causal en series temporales —la autocorrelación en alta dimensionalidad (que produce un incremento de falsos positivos) y las influencias indirectas— adaptamos el enfoque PCMCI [21]. Este acercamiento opera en dos grandes fases, sobre el espacio latente:

1. $\mathbf{PC}_1$ **(Selección adaptativa de condiciones)**: para cada nodo objetivo $\boldsymbol{Z}_t^j$ se realiza, en primer lugar, el clásico enfoque de descubrimiento de Peter-Clark (PC) [23]. Se inicializa su conjunto preliminar de padres $\widehat{Pa}(\boldsymbol{Z}_t^j)$ con el grafo temporal pasado completamente conectado (hasta un retraso $\tau_{\text{máx}}$). A continuación, se eliminan iterativamente los ancestros candidatos que sean condicionalmente independientes respecto a combinaciones de variables de tamaño p (comenzando en $p = 0$ e iterando hasta $p_{\text{máx}}$): $\boldsymbol{Z}_{t-\tau}^i \perp\!\!\!\perp \boldsymbol{Z}_t^j \mid \mathcal{S}$, con $\mathcal{S} \subseteq \widehat{Pa}(\boldsymbol{Z}_t^j) \setminus \{\boldsymbol{Z}_{t-\tau}^i\}$, $|\mathcal{S}| = p$. Si el p-valor de independencia supera nuestro nivel de significancia estipulado (por ejemplo, $\alpha_{pc} = 0.1$), se rechaza la conexión, permitiendo converger de forma ágil hacia un conjunto reducido y muy relevante de posibles variables conductoras.
2. **Independencia Condicional Momentánea (MCI)**: una vez purgado el grafo para retener únicamente la mínima dimensionalidad descriptiva de posibles padres, mitigamos la falsa causalidad espuria producida por factores como la autocorrelación. La prueba definitiva de causalidad de $\boldsymbol{Z}_{t-\tau}^i \to \boldsymbol{Z}_t^j$ se condiciona combinando el entorno parental del nodo destino $\boldsymbol{Z}_t^j$ con los retardos correspondientes al nodo origen de causalidad $\boldsymbol{Z}_{t-\tau}^i$: $\boldsymbol{Z}_{t-\tau}^i \perp\!\!\!\perp \boldsymbol{Z}_t^j \mid \left(\widehat{Pa}(\boldsymbol{Z}_t^j) \setminus \{\boldsymbol{Z}_{t-\tau}^i\}\right) \cup \widehat{Pa}(\boldsymbol{Z}_{t-\tau}^i)$. Esto fuerza un control causal riguroso para asegurar que un nodo está añadiendo nueva información momentánea (*Momentary Information Transfer*), no presente ni en su propio pasado dinámico ni en otras rutas conjuntas.

3.3. Test de Independencia Condicional

Los tests de independencia tradicionales, sobre toda la serie temporal, requieren asumir estacionariedad. Para evitarlo, utilizamos los cambios de régimen (marcados por la variable de fondo $\boldsymbol{u}_t$ o mediante subseries equidistantes [9]) para aislar la dinámica temporal.

Nuestro objetivo es evaluar la independencia condicional global $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{W}$ a lo largo de toda la serie, siendo $\boldsymbol{X}$ el potencial nodo origen (por ejemplo, una variable latente en el pasado $\boldsymbol{Z}_{t-\tau}^i$), $\boldsymbol{Y}$ el potencial nodo destino a explicar ($\boldsymbol{Z}_t^j$), y $\boldsymbol{W}$ es el conjunto condicionante (los posibles padres y variables de contexto). Para ello se aplica un test de independencia dentro de cada régimen, para obtener una serie de p-valores p_r, como se indica en la Figura 1, y se agregan estos mediante el Test de Combinación de Cauchy (CCT) [13].

Para cada régimen r, se modela el efecto condicionante de $\boldsymbol{W}$ mediante el cálculo de los residuos de $\boldsymbol{X}$ y $\boldsymbol{Y}$ respecto a $\boldsymbol{W}$: $\boldsymbol{\xi}_r^{\boldsymbol{X}}$ y $\boldsymbol{\xi}_r^{\boldsymbol{Y}}$. A continuación, sobre estos residuos, se aplica el test *Max-Corr* [15]; se calcula la matriz de correlación de Pearson por pares, y se extrae el estadístico por régimen $\rho_r = \max_{j,k} |\rho_{j,k}|$. Se evalúa la significancia a partir de este estadístico mediante una distribución t de Student, con ajuste de grados de libertad paramétricos: $df_r = \max\{1, n_r - 2 - df_z\}$, donde n_r es el número de muestras en el régimen y df_z es el grado de libertad de la regresión escogida — la dimensionalidad de $\boldsymbol{W}$ en una regresión por mínimos cuadrados, o la traza de la matriz de Hat [6] para regresiones regularizadas como la Ridge — para el cálculo de los residuos.

Este proceso produce un p-valor local p_r, que debe acotarse mediante la corrección de Bonferroni, aplicada sobre el espacio multivariante de $\boldsymbol{X}$ y $\boldsymbol{Y}$. Tras ello, se agregan estas evidencias locales mediante el CCT. Para ello, se le asigna a cada régimen un peso $w_r = n_r / \sum_{i=1}^{R} n_i$, proporcional al tamaño de cada muestra. Finalmente, el estadístico global se define transformando los p-valores locales mediante la función tangente: $T = \sum_{r=1}^{R} w_r \tan(\pi(0.5 - p_r))$, el cual, bajo la hipótesis nula de independencia condicional global, y asumiendo regularidad en las colas de distribuciones bivariadas subyacentes, sigue de forma asintótica una distribución de Cauchy estándar, independientemente de las estructuras de correlación temporal entre los estadísticos locales.

El orden de complejidad del *pipeline* está acotado por sus dos fases. La inferencia (iVAE) presenta una complejidad de $\mathcal{O}(E \cdot T \cdot W)$ (E: épocas, T: longitud de serie, W: parámetros, que absorbe las variables observadas intra-grupo). Por su parte, el descubrimiento (Group-PCMCI) está acotado por $\mathcal{O}(N^2 \cdot \tau_{\max} \cdot (N\tau_{\max})^{p_{\max}} \cdot T)$. Al aislar la búsqueda al espacio latente reducido (N grupos) y limitar el condicionamiento ($p_{\max} = 3$), se evita la explosión combinatoria, garantizando la escalabilidad.

4. Experimentación

4.1. Definición de los experimentos

Para hacer pruebas con los algoritmos previamente explicados se ha seguido una metodología popular en la literatura de descubrimiento causal sobre series temporales [3,21], consistente en una generación aleatoria de grafos, construcción de modelos estructurales a partir de estos grafos, muestreo de los modelos estructurales y ejecución de algoritmos de descubrimiento causal sobre el muestreo. Así, en primer lugar se genera aleatoriamente un grafo real $\mathcal{G}^{real}$, el cual asocia a cada grupo $\boldsymbol{X}_t^i$ un conjunto de padres $Pa(\boldsymbol{X}_t^i)$. Este, para mantener las suposiciones de algunos algoritmos como gCDMI, contiene únicamente nodos retardados, con aristas de la forma, $\boldsymbol{X}_{t-\tau}^j \to \boldsymbol{X}_t^i$, siendo $\tau > 0$. De esta forma, el modelo estructural causal [17] queda definido por $\boldsymbol{X}_t^i = f^i\left(Pa(\boldsymbol{X}_t^i)\right) + \boldsymbol{\epsilon}_t^i$, donde $\boldsymbol{\epsilon}_t^i$ es un ruido cuya distribución suele considerarse de componentes independientes e idénticamente distribuidas, y en nuestro caso será una combinación de distribuciones normales y Weibull. Las funciones f^i estarán construidas de la forma: $f^i\left(Pa(\boldsymbol{X}_t^i)\right) = \sum_{\boldsymbol{X}_{t-\tau}^j \in Pa(\boldsymbol{X}_t^i)} \alpha_\tau^{j\to i} f_\tau^{j\to i}(\boldsymbol{X}_{t-\tau}^j)$, siendo $f_\tau^{j\to i} : \mathbb{R}^{\dim(\text{soporte}(\boldsymbol{X}_t^j))} \to \mathbb{R}^{\dim(\text{soporte}(\boldsymbol{X}_t^i))}$, en nuestro caso, una combinación

lineal de funciones univariantes lineales, cuadráticas, sinusoidales y sigmoidales, además de las funciones bivariantes del producto, una supresión racional, el máximo y el mínimo. El valor $\alpha_\tau^{j\to i}$ indica la fuerza de la dependencia, que, en nuestro caso, se ha tomado como un valor aleatorio en $\{0.3, 0.5\}$.

Para evaluar la robustez de los algoritmos ante dinámicas no estacionarias, generamos un escenario compuesto por múltiples entornos causales heterogéneos [18], introduciendo un ruido dependiente del régimen $\boldsymbol{\epsilon}_t^i \sim \mathcal{D}(\boldsymbol{\mu}_{\boldsymbol{u}_t}^i, \boldsymbol{\Sigma}_{\boldsymbol{u}_t}^i)$.

4.2. Adaptación de métodos e hiperparámetros

En el caso de nuestra propuesta, se toma un conjunto condicionante de tamaño máximo $p_{\text{máx}} = 3$; relativamente bajo debido a la naturaleza multivariante del problema. Además, se hacen dos versiones del algoritmo: una en la que la variable de fondo $\boldsymbol{u}_t$ identifica exactamente los regímenes considerados, y otra en la que, en ausencia de esta información, se divide la serie temporal de entrada en 10 intervalos equidistantes.

Por otro lado, dado que los algoritmos del estado del arte para descubrimiento causal sobre grupos no están intrínsecamente preparados para un escenario no estacionario, se ha implementado para cada uno de ellos una adaptación que sea capaz de aprovechar el conocimiento de los cambios de régimen incluido en la variable de fondo $\boldsymbol{u}_t$. Adaptaremos métodos basados en restricciones y en modelos predictivos.

Métodos basados en restricciones; Para los algoritmos que fundamentan su descubrimiento en pruebas de independencia condicional, tales como el enfoque **Micro-level** [25]; **PCA+PCMCI**, aplicando PCMCI sobre la primera componente principal [25]; el uso de **Adag** para generar el *embedding* con una dimensionalidad específica [15], y la aplicación del ya explicado **Group-PCMCI** directamente sobre el espacio observado, se ha implementado la misma estrategia de evaluación propuesta en nuestro marco de trabajo, adaptando el flujo de estos algoritmos para calcular los p-valores de independencia condicional de manera aislada dentro de cada régimen temporal (definido por $\boldsymbol{u}_t$) y, a continuación, agregar dicha evidencia local utilizando el método CCT.

Métodos basados en modelos predictivos (*Deep Learning*) En los algoritmos basados en redes neuronales, la información no estacionaria se integra directamente en el modelado predictivo. En **GroupRESIT**, la variable $\boldsymbol{u}_t$ se incorpora en los perceptrones regresores concatenando el vector de características observadas con el del régimen ($\boldsymbol{u}_t$), tanto en la Fase I (extracción de residuos para calcular el orden causal) como en la Fase II (selección del modelo final). Por su parte, en **gCDMI** se utiliza $\boldsymbol{u}_t$ como covariables dinámicas en DeepAR , actuando como variables de contexto externas para que la red recurrente condicione sus predicciones según el régimen vigente en cada instante.

4.3. Resultados

Los experimentos realizados han tomado 10 grafos aleatoriamente generados, cada uno con 50 variables distintas agrupadas en 8 grupos, un retraso máximo de $\tau_{\text{máx}} = 3$, y una fracción del 20 % de las posibles aristas inter-grupo posibles. A partir de estos grafos se han generado los modelos estructurales causales co-

rrespondientes, y un muestreo de las variables de estos modelos ha permitido generar conjuntos de datos con 2 000 instantes de tiempo cada uno. Además se dejan disponible en el material suplementario[2]más métricas sobre este mismo experimento, otro experimento equivalente estudiando cómo afecta un aumento en la fracción de variables afectadas por el cambio de régimen, con resultados parecidos a los aquí obtenidos, y un análisis de sensibilidad en el que se comprueba que el número de bloques para realizar la división de las series, en el caso de asumir ausencia de conocimiento sobre los regímenes, no tiene una influencia significativa en los resultados.

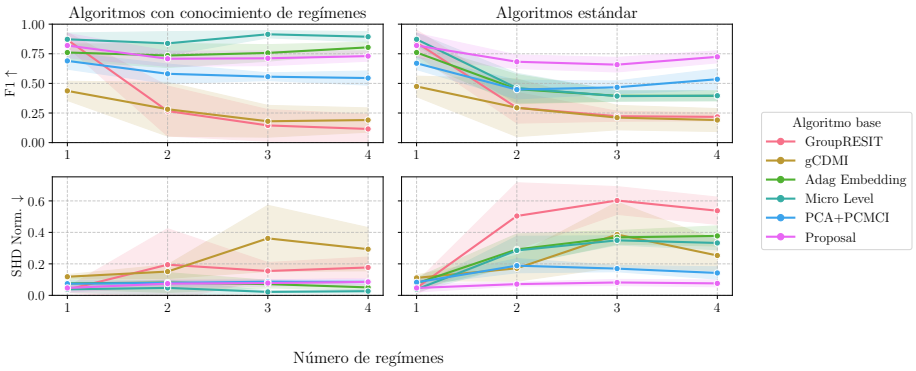


Figura 3. Gráfico de las medias y desviaciones típicas de los *F1 Score* ↑ obtenidos en los experimentos, para un número de regímenes en aumento.

Para comprobar la influencia del cambio de régimen se han realizado estos mismos experimentos, con todas las variables afectadas por los regímenes, para un número de regímenes creciente, y se han mostrado los resultados en la Figura 3, en términos tanto de F1, comparando las aristas de la clase de Markov del grafo predicho con las del real, como de SHD[3]. Se aprecia cómo los algoritmos, al menos los basados en restricciones, asumen un empeoramiento mucho menor que los estándar, haya más o menos variables afectadas por cambios de régimen, cuando se les incorpora conocimiento acerca de los regímenes. Esto probable sea debido a que la comprobación de la independencia condicional de forma independiente en cada régimen para su posterior acumulación con el método

[2] El material suplementario está disponible en la siguiente web: `https://www.uco.es/kdis/descubrimiento-causal-sobre-vectores-en-series-temporales-no-estacionarias/`. Además, por los resultados allí presentados, queda claro que el decrecimiento del F1 en los algoritmos sin conocimiento de regímenes se debe a un decrecimiento radical de la precisión; esto cobra mucho sentido cuando se considera que los regímenes no observados actúan entre las variables observadas como factores de confusión: $X_t^i \leftarrow \boldsymbol{u}_t \rightarrow X_t^j$.

[3] La *Structural Hamming Distance* (SHD) mide la distancia entre las clases de Markov del grafo predicho con el grafo real, en términos del número de aristas que no comparten, tanto porque las tenga el predicho y no el real, como viceversa. Su normalización divide por el número total de aristas posibles, $N^2 \cdot \tau_{\text{máx}}$.

CCT permite que los algoritmos sean relativamente agnósticos a los cambios. Además, nuestro propuesta es capaz, cuando todas las variables son modificadas, de obtener el mayor F1 ante la ausencia de conocimiento real sobre los regímenes, lo cual suele ser el escenario en contextos reales.

En la Figura 4 se muestran, en particular, los resultados obtenidos en el caso en el que todas las variables son afectadas por los cambios de régimen, en términos tanto de F1 como de SHD normalizada. Aquí se aprecia como, en general, los métodos utilizados para introducir conocimiento sobre el régimen son funcionales, pues se obtienen, en general, mejores resultados entre los métodos que tienen conocimiento acerca de los regímenes — esta mejora es especialmente apreciable en los métodos basados en restricciones, pues su propio diseño permite dividir correctamente el intervalo temporal en distintos regímenes.

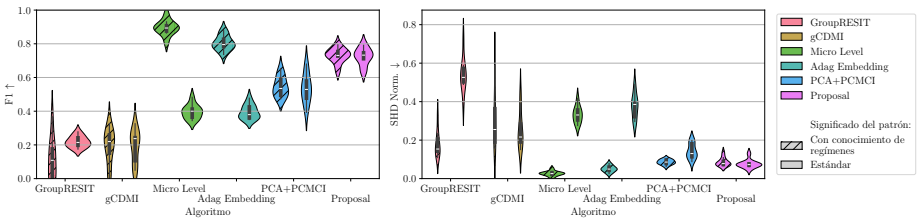


Figura 4. Comparación de *F1 Score* ↑ y SHD ↓, en forma de *violinplot*, obtenido con los distintos algoritmos de descubrimiento causal en un escenario con todas las variables afectadas por los cambios de régimen.

Por otro lado, se aprecia que nuestra propuesta, a pesar de no tener grandes diferencias con otros métodos a la hora de utilizar la información sobre conocimiento de fondo, cuando este conocimiento no está disponible, es capaz de obtener resultados relativamente equivalentes a los obtenidos cuando tiene esta información, lo cual le da una clara ventaja respecto al resto de métodos en este escenario más realista. Esta mejora podría deberse a la capacidad que tiene el método propuesto de generar representaciones latentes que incorporen la información más relevante acerca de cada grupo de variables, incluso ante cambios no conocidos en las distribuciones de los datos.

5. Conclusiones

En este artículo se ha presentado un marco algorítmico para resolver el problema del descubrimiento causal sobre grupos de series temporales cuando no se cumplen los clásicos supuestos de estacionariedad. A diferencia de los métodos tradicionales, que asumen distribuciones sin cambios estadísticos a lo largo del tiempo, nuestra propuesta aprovecha estos cambios, mediante un iVAE [9], para construir un espacio latente identificable. Este espacio latente reduce la dimensionalidad de cada grupo, permitiendo así la ejecución de una adaptación realizada al algoritmo PCMCI [21] (Group-PCMCI), sobre este espacio latente, de forma independiente en subconjuntos temporales de la serie completa, para recuperar un grafo compatible con las dinámicas causales y lidiar con la autocorrelación.

La experimentación realizada sobre datos sintéticos demuestra empíricamente la superioridad de este enfoque frente a métodos del estado del arte (como

GroupRESIT [4], gCDMI [1] o combinaciones de PCA y PCMCI [15,21,25]). Si bien los algoritmos existentes mejoran significativamente cuando son adaptados para recibir información acerca de los regímenes, nuestra propuesta tiene una clara ventaja bajo un supuesto más general, en el que se asumen regímenes no conocidos.

Sin embargo, dado que los tests de independencia se están realizando sobre un espacio latente, y no es trivial la extrapolación del descubrimiento causal a variables observadas, [15], las relaciones inferidas a partir de estos patrones deben interpretarse como evidencias de asociaciones compatibles con un modelo causal, y no como una identificación causal garantizada.

En definitiva, la combinación de representaciones latentes identificables y pruebas de independencia grupales genera una herramienta fiable para descubrir relaciones *macro* en sistemas complejos del mundo real (como entornos climáticos, industriales o logísticos), donde las distribuciones de los datos fluctúan de forma natural.

Trabajo futuro Es necesario extender la validación empírica a contextos reales. En la literatura se han realizado pruebas de la efectividad del descubrimiento causal en series temporales no estacionarias [7] y en grupos de ellas [4], de forma independiente, por lo que es interesante estudiar los casos de intersección entre estos conceptos, en los que hay grupos de series temporales con estadísticas fluctuantes en el tiempo. Además, un estudio de ablación podría permitir aislar el efecto individual de la generación de representación latente y de la fase de descubrimiento en el rendimiento final del modelo.

Por otro lado, también sería interesante extender el marco teórico de los iVAE [9] para estudiar si sería posible garantizar suficiencia causal sobre el espacio latente sin realizar la suposición de estacionaridad. Esto podría permitir extender los resultados acerca de la *aggregation faithfulness* introducida en [15] a series temporales no estacionarias. De esta forma se podría asegurar teóricamente que, bajo ciertos supuestos, el algoritmo es capaz de recuperar el grafo causal subyacente en el límite de número de datos y capacidad de cómputo infinitos.

Agradecimientos Esta investigación ha sido financiada parcialmente por el proyecto PID2023-148396NB-I00 del Ministerio de Ciencia e Innovación y por el proyecto ProyExcel-0069 de la Consejería de Universidad, Investigación e Innovación de la Junta de Andalucía.

Referencias

1. Ahmad, W., Denzler, J., Shadaydeh, M.: Group interventions on deep networks for causal discovery in subsystems (2025)
2. Assaad, C.K., Devijver, E., Gaussier, E.: Survey and evaluation of causal discovery methods for time series. Journal of Artificial Intelligence Research **73**, 767–819 (2022)
3. Ferdous, M.H., Hasan, U., Gani, M.O.: Cdans: Temporal causal discovery from autocorrelated and non-stationary time series data. In: Proceedings of the 8th Machine Learning for Healthcare Conference. vol. 219, pp. 186–207. PMLR (2023)
4. Göbler, K., Windisch, T., Drton, M.: Nonlinear causal discovery for grouped data. In: Proceedings of the Forty-First Conference on Uncertainty in Artificial Intelligence. JMLR.org (2025). https://doi.org/10.48550/arXiv.2506.05120

5. Hasan, U., Hossain, E., Gani, M.O.: A survey on causal discovery methods for iid and time series data. arXiv preprint arXiv:2303.15027 (2023)
6. Hastie, T., Tibshirani, R., Friedman, J., et al.: The elements of statistical learning (2009). https://doi.org/10.1007/978-0-387-84858-7
7. Huang, B., Zhang, K., Zhang, J., Ramsey, J., Sanchez-Romero, R., Glymour, C., Schölkopf, B.: Causal discovery from heterogeneous/nonstationary data with independent changes (2020)
8. Huynh, V.N., Kreinovich, V., Sriboonchitta, S.: Causal inference in econometrics. Springer (2016)
9. Khemakhem, I., Kingma, D., Monti, R., Hyvarinen, A.: Variational autoencoders and nonlinear ica: A unifying framework. In: International conference on artificial intelligence and statistics. pp. 2207–2217. PMLR (2020)
10. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2022)
11. Kretschmer, M., Coumou, D., Donges, J.F., Runge, J.: Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. Journal of climate **29**(11), 4069–4081 (2016)
12. Lehmann, E.L., Casella, G.: Theory of Point Estimation. Springer, second edn. (1998)
13. Liu, Y., Xie, J.: Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. Journal of the American Statistical Association **115**(529), 393–402 (2020)
14. Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., Bachem, O.: Challenging common assumptions in the unsupervised learning of disentangled representations. In: Proceedings of the 36th International Conference on Machine Learning. vol. 97, pp. 4114–4124. PMLR (2019)
15. Ninad, U., Wahl, J., Gerhardus, A., Runge, J.: Causal discovery on vector-valued variables and consistency-guided aggregation (2025)
16. Nowack, P., Runge, J., Eyring, V., Haigh, J.D.: Causal networks for climate model evaluation and constrained projections. Nature communications **11**(1), 1415 (2020)
17. Pearl, J.: Causality. Cambridge university press (2009)
18. Peters, J., Bühlmann, P., Meinshausen, N.: Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society Series B: Statistical Methodology **78**(5), 947–1012 (2016)
19. Peters, J., Bühlmann, P., Meinshausen, N.: Causal inference using invariant prediction: identification and confidence intervals (2015)
20. Pfister, N., Bühlmann, P., Peters, J.: Invariant causal prediction for sequential data (2018)
21. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances **5**(11), eaau4996 (2019)
22. Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: Deepar: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting **36**(3), 1181–1191 (2020)
23. Spirtes, P., Glymour, C.N., Scheines, R.: Causation, prediction, and search. MIT press (2000)
24. Teyssier, M., Koller, D.: Ordering-based search: a simple and effective algorithm for learning bayesian networks. In: Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence. p. 584–590. AUAI Press (2005)
25. Wahl, J., Ninad, U., Runge, J.: Foundations of causal discovery on groups of variables (2024). https://doi.org/10.48550/arXiv.2306.07047