The Annals of Statistics
2024, Vol. 52, No. 1, 103–130
https://doi.org/10.1214/23-AOS2325

CHARACTERIZATION OF CAUSAL ANCESTRAL GRAPHS FOR TIME SERIES WITH LATENT CONFOUNDERS

BY ANDREAS GERHARDUS[a]

Institute of Data Science, German Aerospace Center, [a]*andreas.gerhardus@dlr.de*

In this paper, we introduce a novel class of graphical models for representing time-lag specific causal relationships and independencies of multivariate time series with unobserved confounders. We completely characterize these graphs and show that they constitute proper subsets of the currently employed model classes. As we show, from the novel graphs one can thus draw stronger causal inferences—without additional assumptions. We further introduce a graphical representation of Markov equivalence classes of the novel graphs. This graphical representation contains more causal knowledge than what current state-of-the-art causal discovery algorithms learn.

1. Introduction. In recent decades, causal graphical models have become a standard tool for reasoning about causal relationships, for example, Pearl (2009), Spirtes, Glymour and Scheines (2000), Koller and Friedman (2009). The most basic and popular class of models are directed acyclic graphs (DAGs). In their interpretation as causal Bayesian networks, these graphs specify interventional distributions and causal effects in terms of the observational distribution, for example, Spirtes, Glymour and Scheines (1993), Pearl (1995), Pearl (2000). DAGs can only model acyclic causal relationships among variables that are not subject to latent confounding, that is, such that there are no unobserved common causes of observed variables. The latter assumption is known as *causal sufficiency* and intuitively means that all variables relevant for describing the system's causal relationships are modeled explicitly. If causal sufficiency cannot be asserted, as is often the case, then one approach is to instead work with maximal ancestral graphs (MAGs); see Richardson and Spirtes (2002), Zhang (2008a). This larger class of graphs retains a well-defined causal interpretation in presence of latent confounding.

MAGs can even represent selection variables, that is, unobserved variables that determine which sample points belong to the observed population. In this paper, we rule out selection variables by assumption. It is then sufficient to work with a subclass of MAGs that, following Mooij and Claassen (2020), are called *directed maximal ancestral graphs (DMAGs)*. Assuming the absence of selection variables is common both in the literature on causal effect estimation and causal discovery, for example, Zhang (2006), Perković et al. (2017) and Entner and Hoyer (2010), Malinsky and Spirtes (2018), Gerhardus and Runge (2020). As an advantage, DMAGs convey significantly stronger inferences about the presence of causal ancestral relationships than MAGs. Moreover, for time series there is exactly one sample point per time step, and hence potential selection bias would at least not go unnoticed.

To use any of these model classes for causal reasoning, one needs to already know the system's causal structure in form of the respective graph. If this knowledge is not available and experiments are infeasible, then one must rely on observational causal discovery, for example, Spirtes, Glymour and Scheines (2000), Peters, Janzing and Schölkopf (2017), which refers to learning causal relationships from observational data under suitable enabling

Received December 2022; revised July 2023.
MSC2020 subject classifications. Primary 62A09, 62D20, 62M10; secondary 68T30, 68T37.
Key words and phrases. Causal graph, ancestral graph, time series, latent variable, causal inference, causal discovery.

assumptions. So-called independence-based methods, also called constraint-based methods, attempt to learn the causal graph from independencies in the observed probability distribution. In general, learning the graph from independencies is an underdetermined problem since distinct graphs may describe the same set of independencies. This nonuniqueness is known as *Markov equivalence*. Without more assumptions, it is then only possible to learn those features of the causal graph that it shares with all its Markov equivalent graphs. These shared features can in turn be represented by certain graphs, which for the case of MAGs are *partial ancestral graphs (PAGs)*; see Ali, Richardson and Spirtes (2009), Zhang (2008b). There are sound and complete causal discovery algorithms for learning PAGs, for example, the FCI algorithm; see Spirtes, Meek and Richardson (1995), Spirtes, Glymour and Scheines (2000), Zhang (2008b). Here, *sound* refers to correctness of the method and *complete* to it learning all shared features. The refinement of PAGs obtained by restricting from MAGs to DMAGs are called *directed partial ancestral graphs (DPAGs)* in Mooij and Claassen (2020).

The causal graphical model framework outlined above does not inherently rely on temporal information, and the nontemporal setting so far is its major domain of application. However, dynamical systems and time series data are ubiquitous and of great interest to science and beyond. In this setting, Granger causality (see Granger (1969)) is a widely-used framework for causal analyses. This framework employs a predictive notion of causality, according to which a time series X has a causal influence on time series Y if the past of X helps in predicting the present of Y given that the pasts of all time series other than X are already known. Granger causality has two central limitations: First, it requires the absence of latent confounders, that is, unobserved time series that are a common cause of two observed time series. Second, it cannot in general deal with contemporaneous causal influences, that is, causal influences on time scales below the sampling interval. For an in-depth discussion of these limitations, see, for example, Peters, Janzing and Schölkopf ((2017), Chapter 10).

Since the causal graphical model framework is not subject to these two limitations, in recent years there has been a growing interest in adapting it to the time series setting. Generally, there are three ways to do this. The first approach, for example, Eichler and Didelez (2007), Eichler (2010), Eichler and Didelez (2010) and Didelez (2008), Mogensen and Hansen (2020), uses a graph in which there is one vertex per component time series. The edges then summarize the causal influences at all time lags, thus giving a conveniently compressed graphical representation of the causal relationships. However, the information about time lags of individual cause-and-effect relationships is lost. The second approach uses graphs with one vertex per component time series and time step, thus resolving the time lags. There are various causal discovery methods that implement this approach, for example, Chu and Glymour (2008), Hyvärinen et al. (2010), Entner and Hoyer (2010), Malinsky and Spirtes (2018), Runge (2020), Pamfil et al. (2020), Gerhardus and Runge (2020), and application works from diverse domains, for example, Kretschmer et al. (2016), Huckins et al. (2020), Saetia, Yoshimura and Koike (2021). By resolving time lags, it becomes possible to obtain a data-driven process understanding and to study the effect of interventions on particular time steps of variables. However, learning a time-resolved graph is statistically more challenging than learning a time-collapsed graph and one might need to compromise on the number of resolved time steps. Assaad, Devijver and Gaussier (2022) propose a third, intermediate approach with two vertices per component time series (one for the present time step and one for the entire past).

We follow the second approach. In this case, the temporal information inherent to time series restricts the connectivity pattern (i.e., absence and presence of edges, edge orientations) of the resulting time-resolved graphs. Namely, since we here consider graphical models in which directed edges signify causal influences (DAGs, DMAGs and DPAGs), the directed edges must not point backwards in time. In addition, we assume time invariant causal relationships. This invariance, known as *causal stationarity*, implies that the graph's edges are repetitive in time. For DAGs that represent time series *without* latent confounders, which we call *time series DAGs (ts-DAGs)*, these are the only restrictions on the connectivity pattern.

For DMAGs that represent time series *with* latent confounders, the corresponding restrictions on the connectivity pattern have, however, not yet been worked out. Although there are works on independence-based time series causal discovery with latent confounding (see Entner and Hoyer (2010), Malinsky and Spirtes (2018), Gerhardus and Runge (2020)), no characterization of the associated class of graphical models has been given. This is the conceptual gap that we close in the present work, that is, we completely characterize which DMAGs are obtained by marginalizing ts-DAGs, and hence can serve as causal graphical model for causally stationary time series with latent confounders. We call the novel graphs defined by this characterization *time series DMAGs (ts-DMAGs)* and show that these novel graphs constitute a strictly smaller model class than the previously considered model classes. We further show that, without imposing additional assumptions, one can draw stronger causal inferences from ts-DMAGs than from the previously considered graphs. We also introduce *time series DPAGs (ts-DPAGs)* as representations of Markov equivalence classes of ts-DMAGs. Time series DPAGs are more informative than the graphs learned by current latent time series causal discovery algorithms. As a remark, since contemporaneous causal interactions are allowed without restrictions other than acyclicity, the time series case considered here formally subsumes, and hence is more general than the (acyclic) nontemporal case.

The structure of this paper is as follows: In Section 2, we summarize basic graphical concepts and introduce our notation. In Section 3, we first specify the considered type of causally stationary time series processes. We then introduce ts-DMAGs, a class of causal graphical models for representing the causal relationships and independencies among only the observed variables of such processes at finitely many regularly spaced observed time steps. In Section 4, we analyze ts-DMAGs and first derive several properties that they necessarily have. With Theorems 1 and 2, we then completely characterize ts-DMAGs by a single necessary and sufficient condition. We further show that ts-DMAGs are a strict subset of the classes of graphical models that have previously been considered in the literature (see Section 4.8). For this reason, and as we demonstrate with examples, one can draw stronger causal inferences from ts-DMAGs than from the previously considered graphs. We further introduce the concept of *stationarification* in order to illuminate various discussions. In Section 5, we put these developments to use in the context of causal discovery by defining ts-DPAGs as representations of the Markov equivalence classes of ts-DMAGs. We show that these graphs contain more causal information than the output of current causal discovery algorithms. Moreover, we point out an incorrect claim in the literature that, as we argue, has misguided recent developments (see the discussion below Theorem 3). We also present an algorithm that learns ts-DPAGs from data. We give further theoretical results and all proofs in the Supplementary Material (Gerhardus (2023)).

2. Basic graphical concepts and notation. Our notation and terminology is a mixture of those used in Maathuis and Colombo (2015), Perković et al. (2017) and Mooij and Claassen (2020) as well as some idiosyncratic notation.

A *graph* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consists of a set of vertices $\mathbf{V}$ together with a set of edges $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$. The vertices $i, j \in \mathbf{V}$ are *adjacent* if $(i, j) \in \mathbf{E}$ or $(j, i) \in \mathbf{E}$. We then say that there is an *edge between i and j* and that *i is an adjacency of j*, and similarly for i and j interchanged.

Throughout this paper, we only consider directed partial mixed graphs. These are graphs that satisfy three conditions: First, there is at most one edge between any pair of vertices. Second, no vertex is adjacent to itself. Third, there are at most four types of edges: *directed edges* ($\rightarrow$), *bidirected edges* ($\leftrightarrow$), *partially directed edges* ($\circ\!\!\rightarrow$) and *nondirected edges* ($\circ\!\!-\!\!\circ$).

The third condition is formalized by a decomposition of $\mathbf{E}$ as $\mathbf{E} = \mathbf{E}_{\rightarrow} \dot\cup \mathbf{E}_{\leftrightarrow} \dot\cup \mathbf{E}_{\circ\!\rightarrow} \dot\cup \mathbf{E}_{\circ\!-\!\circ}$ that specifies the *edge types* (also called *edge orientations*). This decomposition is considered part of the specification of a concrete graph. A *directed mixed graph* is a partial mixed graph without partially directed and nondirected edges, and a *directed graph* is a directed mixed graph without bidirected edges. The *skeleton* of a graph is the object obtained when disregarding the information about the decomposition of $\mathbf{E}$ into $\mathbf{E}_{\rightarrow} \dot\cup \mathbf{E}_{\leftrightarrow} \dot\cup \mathbf{E}_{\circ\!\rightarrow} \dot\cup \mathbf{E}_{\circ\!-\!\circ}$.

Given directed partial mixed graphs $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$, we say that $\mathcal{G}'$ *is a* subgraph *of* $\mathcal{G}$ and that $\mathcal{G}$ *is a* supergraph *of* $\mathcal{G}'$, denoted as $\mathcal{G}' \subseteq \mathcal{G}$ or $\mathcal{G} \supseteq \mathcal{G}'$, if $\mathbf{V}' \subseteq \mathbf{V}$ and $(i, j) \in \mathbf{E}'_{\bullet}$ with $\bullet \in \{\rightarrow, \leftrightarrow, \circ\!\rightarrow, \circ\!-\!\circ\}$ implies $(i, j) \in \mathbf{E}_{\bullet}$. Given a directed partial mixed graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, its induced subgraph *on* $\mathbf{V}' \subseteq \mathbf{V}$ is the graph $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ such that $(i, j) \in \mathbf{E}'_{\bullet}$ with $\bullet \in \{\rightarrow, \leftrightarrow, \circ\!\rightarrow, \circ\!-\!\circ\}$ if and only if $i, j \in \mathbf{V}'$ and $(i, j) \in \mathbf{E}_{\bullet}$.

We denote a directed edge $(i, j) \in \mathbf{E}_{\rightarrow}$ as $i \rightarrow j$ or $j \leftarrow i$ and say $i \rightarrow j$ *($j \leftarrow i$) is in* $\mathcal{G}$ if $(i, j) \in \mathbf{E}_{\rightarrow}$; similarly for the other edge types. We view edges as composite objects of the symbols at their ends—the *edge marks*—which are *tails*, *heads* or *circles*. For example, $i \circ\!\rightarrow j$ has a circle mark at i and a head mark at j, and $i \rightarrow j$ has a tail mark at i. Tails and heads are *noncircle marks* and *unambiguous* orientations. Circle marks are *ambiguous* orientations. The symbol '$*$' is a wildcard for all three marks. For example, $*\!\!\rightarrow$ may be $\rightarrow$, $\leftrightarrow$ or $\circ\!\rightarrow$.

A *walk* in $\mathcal{G}$ is an ordered sequence $\pi = (i_1, i_2, \dots, i_n)$ of vertices such that i_k and i_{k+1} are adjacent in $\mathcal{G}$ for all $k = 1, \dots, n-1$. The integer $n \geq 1$ is the *length of* π and a vertex in this sequence is said to *be on* π. A *path* is a walk on which every vertex occurs at most once. For a path $\pi = (i_1, i_2, \dots, i_n)$, the vertices i_1 and i_n are the *end-point vertices of* π; all other vertices on π are the *non-end-point vertices of* π. We refer to π as a *path between* i_1 *and* i_n and graphically represent it by $i_1 *\!\!-\!\!* i_2 *\!\!-\!\!* \dots *\!\!-\!\!* i_n$ where $i_k *\!\!-\!\!* i_{k+1}$ is the unique edge between i_k and i_{k+1}. Such a graphical representation can also specify a path. We say that π *is out of* i_1 if $i_1 \rightarrow i_2$ in $\mathcal{G}$ and that π *is into* i_1 if $i_1 \leftarrow\!\!* i_2$ in $\mathcal{G}$; similarly for the other end-point vertex. For $1 \leq a < b \leq n$, we write $\pi(i_a, i_b)$ for the path $(i_a, i_{a+1}, \dots i_b)$ and $\pi(i_b, i_a)$ for the path $(i_b, i_{b-1}, \dots i_a)$. Both of these are *subpaths of* π. Given walks $\pi_1 = (i_1, i_2, \dots, i_n)$ and $\pi_2 = (j_1, j_2, \dots, j_m)$ with $i_n = j_1$, we write $\pi_1 \oplus \pi_2$ for the walk $(i_1, i_2, \dots, i_n, j_2, \dots, j_m)$. A vertex i_k on path π is a *collider on* π if it is a non-end-point vertex of π and $\pi(i_{k-1}, i_{k+1})$ is $i_{k-1} *\!\!\rightarrow i_k \leftarrow\!\!* i_{k+1}$, else it is a *noncollider on* π. If the vertices i and k are nonadjacent, then the path $i *\!\!-\!\!* j *\!\!-\!\!* k$ is an *unshielded triple* and the path $i *\!\!\rightarrow j \leftarrow\!\!* k$ an *unshielded collider*. A *path* of length $n = 1$ is called *trivial*. The path $\pi = (i_1, i_2, \dots, i_n)$ is a *directed path* if $i_k \rightarrow i_{k+1}$ in $\mathcal{G}$ for all $1 \leq k \leq n-1$ or $i_k \leftarrow i_{k+1}$ in $\mathcal{G}$ for all $1 \leq k \leq n-1$. In the former case, we speak of a *directed path from* i_1 *to* i_n, in the latter case of a *directed path from* i_n *to* i_1. The previous definitions equally apply to walks.

If the edge $i \rightarrow j$ is in $\mathcal{G}$, then i is a *parent* of j and j is a *child* of i. The vertex i is an *ancestor* of j and j is a *descendant* of i if $i = j$ or if there is a directed path from i to j. The set of parents and ancestors of a vertex i in $\mathcal{G}$ are respectively denoted as $\mathrm{pa}(i, \mathcal{G})$ and $\mathrm{an}(i, \mathcal{G})$. We say *vertex* i *is an ancestor of a set* $\mathbf{S}$ *of vertices* and $\mathbf{S}$ *is a descendant of* i if at least one element of $\mathbf{S}$ is a descendant of i. Similarly, *vertex* i *is a descendant of a set* $\mathbf{S}$ *of vertices* and $\mathbf{S}$ *is an ancestor of* i if at least one element of $\mathbf{S}$ is an ancestor of i.

A directed partial mixed graph $\mathcal{G}$ has a *directed cycle* if there are distinct vertices i and j with $i \in \mathrm{an}(j, \mathcal{G})$ and $j \in \mathrm{an}(i, \mathcal{G})$. A *directed acyclic graph (DAG)* $\mathcal{D}$ is a directed graph without directed cycles. A directed partial mixed graph $\mathcal{G}$ has an *almost directed cycle* if the edge $i \leftrightarrow j$ is in $\mathcal{G}$ and $i \in \mathrm{an}(j, \mathcal{G})$. A *directed ancestral graph* is a directed mixed graph without directed cycles and almost directed cycles. An *inducing path between* i *and* j is a path π between i and j such that all non-end-point vertices of π are colliders on π and ancestors of i or j. A *directed maximal ancestral graph (DMAG)* $\mathcal{M}$ is a directed ancestral graph that has no inducing paths between nonadjacent vertices. Every DAG is a DMAG. *Directed partial ancestral graphs (DPAGs)* $\mathcal{P}$ are directed partial mixed graphs that represent Markov equivalence classes of DMAGs; see Definition 5.2 in Section 5.1 for a formal definition.

3. A class of causal graphical models for time series with latent confounders. In this section, we first formally specify the considered type of time series processes; see Section 3.1. We then explain how, if there are no unobserved variables, certain DAGs with an infinite number of vertices (Definition 3.4) can model these processes as causal Bayesian networks; see Sections 3.2 and 3.3. Importantly, Definition 3.6 in Section 3.4 introduces so-called *time series DMAGs (ts-DMAGs)*. These graphs are projections of the infinite DAGs and represent the causal relationships and independencies among only a subset of observed variables at a finite number of regularly sampled or regularly subsampled observed time steps. Time series DMAGs constitute the novel class of causal graphical models which is the central topic of this paper.

3.1. *Structural vector autoregressive processes.* We consider multivariate time series $\{\mathbf{V}_t\}_{t\in\mathbb{Z}}$, where $\mathbf{V}_t = (V_t^1, \dots, V_t^{n_V})$ with the component time series $V^i = \{V_t^i\}_{t\in\mathbb{Z}}$ for $1 \le i \le n_V$, that are generated by an *acyclic structural vector autoregressive process with contemporaneous influences*, for example, Malinsky and Spirtes (2018). That is to say, for all $t \in \mathbb{Z}$ (time index) and $1 \le i \le n_V$ (variable index) the value of V_t^i is determined as

$$V_t^i := f^i\big(PA_t^i, \epsilon_t^i\big), \tag{1}$$

where f^i is a measurable function that depends on all its arguments, the random variables ϵ_t^i (so-called "noise" variables) are jointly independent (with respect to both indices) and have a distribution that may depend on i but not on t, and $PA_t^i \subseteq \{V_{t-\tau}^k | 1 \le k \le n_V, 0 \le \tau \le p_{\text{ts}}\} \setminus \{V_t^i\}$. Here, the *order* p_{ts} of the process is the smallest integer for which the set inclusion in the previous sentence holds (for all i and t). We demand that $0 \le p_{\text{ts}} < \infty$.

We allow contemporaneous causal influences (i.e., $V_{t-\tau}^k \in PA_t^i$ with $\tau = 0$). Further, for all $\Delta t \in \mathbb{Z}$ we assume the sets PA_t^i and $PA_{t-\Delta t}^i$ to be consistent in the sense that $V_{t-\tau}^k \in PA_t^i$ if and only if $V_{t-\tau-\Delta t}^k \in PA_{t-\Delta t}^i$. Acyclicity means the system of equations is recursive. The attribute *structural* asserts that equation (1) is a *structural causal model (SCM)*, for example, Bollen (1989), Pearl (2009), Peters, Janzing and Schölkopf (2017), which we indicate by the ":=" symbol. Because of this causal interpretation, we refer to the variables PA_t^i as *causal parents* of V_t^i and to the consistency of PA_t^i and $PA_{t-\Delta t}^i$ as *causal stationarity*. The restriction of PA_t^i to variables $V_{t-\tau}^k$ with $\tau \ge 0$ ensures that there is no causal influence backward in time.

3.2. *Time series DAGs.* The causal parentships specified by an SCM are graphically represented by the SCM's *causal graph*, for example, Spirtes, Glymour and Scheines (2000), Pearl (2009), Peters, Janzing and Schölkopf (2017). The causal graph is a directed graph with one vertex per variable, typically excluding the noise variables, and directed edges from each variable to all variables of which it is a causal parent. The same construction applies to structural processes as in equation (1). However, as we capture by the below three notions, the resulting "temporal causal graphs" carry more structure than their nontemporal counterparts.

First, the random variable V_t^i corresponds to a particular time step t of a particular component time series V^i. This correspondence is captured by the following notion.

DEFINITION 3.1 (Time series structure). A graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ has a *time series structure* if $\mathbf{V} = \mathbf{I} \times \mathbf{T}$, where $\mathbf{I} = \{1, 2, \dots, n\}$ with $n \ge 1$ is the *variable index set* and $\mathbf{T} = \{t \in \mathbb{Z} | t_s \le t \le t_e\}$ with $t_s \in \mathbb{Z} \cup \{-\infty\}$ and $t_e \in \mathbb{Z} \cup \{+\infty\}$ and $t_s \le t_e$ is the *time index set*.

We say that a vertex $(i, t) \in \mathbf{V}$ is *at time* t, and if $t_a \le t \le t_b$ to be *in the time window* $[t_a, t_b]$. We further say $(i, t_i) \in \mathbf{V}$ is *before* $(j, t_j) \in \mathbf{V}$ and $(j, t_j) \in \mathbf{V}$ is *after* $(i, t_i) \in \mathbf{V}$ if

$t_i < t_j$. An edge $((i, t_i), (j, t_j)) \in \mathbf{E}$ has *length* or *lag* $|t_i - t_j|$. We call edges of length zero *contemporaneous* and call all other edges *lagged*.

Second, below equation (1) we explicitly restricted the causal parents PA_t^i to only contain vertices that are before or at time t. This restriction is captured by the following notion.

DEFINITION 3.2 (Time order). A directed partial mixed graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with time series structure is *time ordered* if $((i, t_i), (j, t_j)) \in \mathbf{E}_{\rightarrow}$ implies $t_i \leq t_j$.

In a time ordered graph $\mathcal{G}$, the ancestral relationship $(i, t_i) \in \mathrm{an}((j, t_j), \mathcal{G})$ implies $t_i \leq t_j$. This fact shows that also indirect causal influences are correctly restricted to not go backwards in time as soon as this restriction is imposed on direct causal influences.

Third, the property of causal stationarity (see Section 3.1) restricts the edges to be repetitive in time. This restriction is captured by the following notion.

DEFINITION 3.3 (Repeating edges). A directed partial mixed graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with time series structure has *repeating edges* if the following holds: If $((i, t_i), (j, t_j)) \in \mathbf{E}_{\bullet}$ with $\bullet \in \{\rightarrow, \leftrightarrow, {\circ\!\!\rightarrow}, {\circ\!\!-\!\!\circ}\}$ and $(i, t_i + \Delta t), (j, t_j + \Delta t) \in \mathbf{V}$, then $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}_{\bullet}$.

REMARK (on Definition 3.3). Section 3 is concerned with DAGs and DMAGs only. In these graphs, there are by definition no edges of the types ${\circ\!\!\rightarrow}$ or ${\circ\!\!-\!\!\circ}$. However, in Section 5 we will apply the concept of repeating edges also to DPAGs. Since these graphs (DPAGs) can contain edges ${\circ\!\!\rightarrow}$ or ${\circ\!\!-\!\!\circ}$, we already here formulate Definition 3.3 in sufficient generality.

By combining the three notions introduced in Definitions 3.1, 3.2 and 3.3, we define the following class of graphical models, which plays an important role throughout the paper.

DEFINITION 3.4 (Time series DAG). A *time series DAG* (*ts-DAG*) is a DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ with time series structure $\mathbf{V} = \mathbf{I} \times \mathbf{T}$ with $\mathbf{T} = \mathbb{Z}$ that is time ordered and has repeating edges.

Due to time order and repeating edges, a ts-DAG $\mathcal{D}$ is fully specified by its variable index set together with its edges that point to a vertex at time t. Hence, if the longest edge of $\mathcal{D}$ is of finite length $p_{\mathcal{D}} < \infty$ then one unambiguously specifies $\mathcal{D}$ by drawing all vertices within the time window $[t - p_{\mathcal{D}}, t]$ and the edges between these; see Figure 1 for an example. In slight abuse of notation, we sometimes denote vertices by the random variable that they represent.

3.3. *Time series DAGs as causal graphs for structural vector autoregressive processes.* Since the structural process in equation (1) is acyclic by assumption, that is, since the system of equations is recursive, its causal graph is acyclic (hence the terminology). In combination with the discussions in the previous subsection, we thus get the following result.

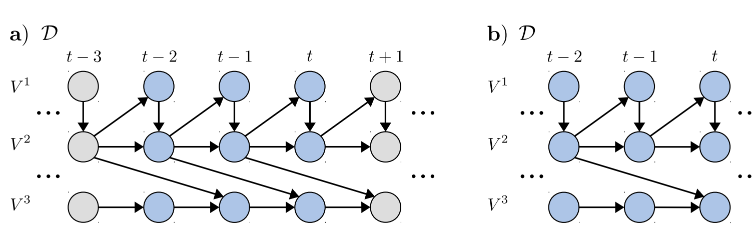

FIG. 1. *Two illustrations of the same ts-DAG $\mathcal{D}$ that represents a structural process as in equation (1) of order $p_{\mathcal{D}} = p_{\mathrm{ts}} = 2$ with three component time series V^1, V^2 and V^3. Given the implicit assertion that there is no edge of length larger than those depicted, the ts-DAG is uniquely specified by showing a segment of $p_{\mathrm{ts}} + 1$ successive time steps. The horizontal dots indicate that the structure is repeated into the infinite past and infinite future.*

LEMMA 3.5. *The causal graph of an acyclic and causally stationary structural vector autoregressive process as in equation* (1) *is a ts-DAG.*

This observation has been made before, for example, in Runge et al. (2012) and Peters, Janzing and Schölkopf (2013) where ts-DAGs have respectively been called *time series graphs* and *full time graphs*. We note that, because of time order, the assumption of acyclicity restricts only the ts-DAG's contemporaneous edges.

In the non-time series setting, an acyclic SCM defines a unique distribution over the SCM's variables (pushforward of the noise distribution by the structural assignments). According to the *causal Markov condition* (see Spirtes, Glymour and Scheines (2000)), the SCM's causal graph is a Bayesian network for this so-called *entailed distribution* (Pearl (2009)), which in turn implies that d-separations (see Pearl (1988), denoted by "$\perp\!\!\!\perp$") in the causal graph imply the corresponding independencies in the distribution (Verma and Pearl (1990), Geiger, Verma and Pearl (1990)). The *causal faithfulness condition* (see Spirtes, Glymour and Scheines (2000)) assumes the reverse implication, that is, that all independencies imply the corresponding d-separations. Then d-separations and independencies are in one-to-one correspondence.

Although acyclic, the time series setting specified by equation (1) is more complicated: Since time is indexed by $t \in \mathbb{Z}$ (as opposed to, e.g., $t \in \mathbb{N}$), there is no initial "starting" distribution that can be pushforwarded to explicitly define a unique entailed distribution. Instead, we need to ask whether equation (1) *implicitly* defines a distribution; and if yes, how many. Following the terminology in Bongers, Blom and Mooij (2018), this question asks for *solutions* to equation (1), that is, for stochastic processes which satisfy equation (1) almost surely. The existence of such solutions as well as their uniqueness (up to almost sure equality) and properties are nontrivial and not considered here. Rather, for the purpose of this paper we *assume* that equation (1) is solved by a (not necessarily unique) strictly stationary stochastic process whose finite-dimensional distributions satisfy the causal Markov and causal faithfulness condition with respect to its ts-DAG. This assumption is common in the literature, cf. Entner and Hoyer (2010), Malinsky and Spirtes (2018), Gerhardus and Runge (2020), and is here only needed for the connection to causality. The results of the present paper are technically about marginalizing the independence (i.e., d-separation) models of ts-DAGs and remain valid also without that additional assumption. The issue of solving equation (1) is an important aspect to consider in future work.

3.4. *Time series DMAGs.* In most real-world scenarios, unobserved common causes cannot be excluded. As mentioned in Section 1 for the non-time series setting, directed maximal ancestral graphs (DMAGs) are often used for causal modeling in the presence of unobserved variables. This use of DMAGs as causal graphical models was pioneered in Richardson and Spirtes (2002), which defines a *marginalization/projection* procedure that from a DAG $\mathcal{G}$ over vertices $\mathbf{V}$, of which only a subset $\mathbf{O} \subseteq \mathbf{V}$ is observed, constructs a DMAG $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ over the observed variables $\mathbf{O}$ only (see also Zhang (2008a)). The projection of $\mathcal{D}$ to $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ has two properties: First, both graphs have the same ancestral relationships among vertices in $\mathbf{O}$. Second, d-separations in $\mathcal{D}$ among vertices in $\mathbf{O}$ are in one-to-one correspondence to the similar concept of m-separation in $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ (also denoted by "$\perp\!\!\!\perp$"). These two properties ensure that if $\mathcal{D}$ is a causal graph then also $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ carries causal meaning and can be used for causal reasoning as explained in Zhang (2008a).

Below, we generalize the construction of such "causal" DMAGs to the time series setting. To begin, we first note that for time series there are two types of unobserved variables:

- *Unobservable variables*: Some component time series $L^1, \dots, L^{n_L}$ with $0 \le n_L < n_V$ may be unobserved entirely by the experimental setup. We call these L^i *unobservable* and call the other component time series $O^1, \dots, O^{n_O}$ with $n_O = n_V - n_l \le \infty$ *observable*. The variable index set $\mathbf{I}$ of the ts-DAG $\mathcal{D}$ accordingly decomposes as $\mathbf{I} = \mathbf{I_O} \dot\cup \mathbf{I_L}$. This first type of unobserved variables is similar to the case of unobserved variables in the non-time series setting.

- *Temporally unobserved variables*: In addition, throughout the paper we will treat only a finite number of time steps $\mathbf{T_O}$ as observed. This construction is specific to the time series setting and means that at times $\mathbf{T} \setminus \mathbf{T_O}$ also the observable time series are treated as unobserved. The rational for doing so is that in practice only finitely many observations are available, and hence one can only reason about DMAGs of finite temporal extension.

Throughout the paper, we restrict the set $\mathbf{T_O}$ of observed time steps to take one of the following two forms:

- *Regular sampling*: All time steps within a time interval $[t-p, t]$ for some nonnegative integer $p < \infty$ and a reference time step t are observed, that is, $\mathbf{T_O} = \{t - \tau \,|\, 0 \le \tau \le p\}$.
- *Regular subsampling*: Every nth time step, for $n \ge 2$ an integer, within $[t-p, t]$ with $p < \infty$ is observed, that is, $\mathbf{T_O} = \{t - \tau \,|\, 0 \le \tau \le p,\ \tau \mod n = 0\}$.

The time window length p is not restricted relative to the order p_{ts} of the data-generating process, that is, we allow all of $p < p_{\text{ts}}$ and $p = p_{\text{ts}}$ and $p > p_{\text{ts}}$. The reference time step t is arbitrary since the ts-DAG $\mathcal{D}$ has repeating edges. We are led to the following definition.

DEFINITION 3.6 (Time series DMAG). Let $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ be a ts-DAG with variable index set $\mathbf{I}$, let $\mathbf{I_O} \subseteq \mathbf{I}$, and let $\mathbf{T_O} \subsetneq \mathbb{Z}$ be regularly sampled or regularly subsampled. The *time series DMAG implied by $\mathcal{D}$ over $\mathbf{O} = \mathbf{I_O} \times \mathbf{T_O}$*, denoted as $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ or $\mathcal{M}_{\mathbf{I_O}\times\mathbf{T_O}}(\mathcal{D})$ and also referred to as a *ts-DMAG*, is the DMAG on the vertex set $\mathbf{O}$ that is obtained by applying the MAG latent projection defined in Zhang ((2008a), pp. 1442–1443) to $\mathcal{D}$ with $\mathbf{L} = \mathbf{V} \setminus \mathbf{O}$ being the set of latent vertices.

Figure 2 illustrates the construction of ts-DMAGs as projections of ts-DAGs. We stress that *all* vertices prior to the observed time window (i.e., before time $t - p$) are treated as unobserved, even if they are observable, and hence would be observed for a larger value of p.

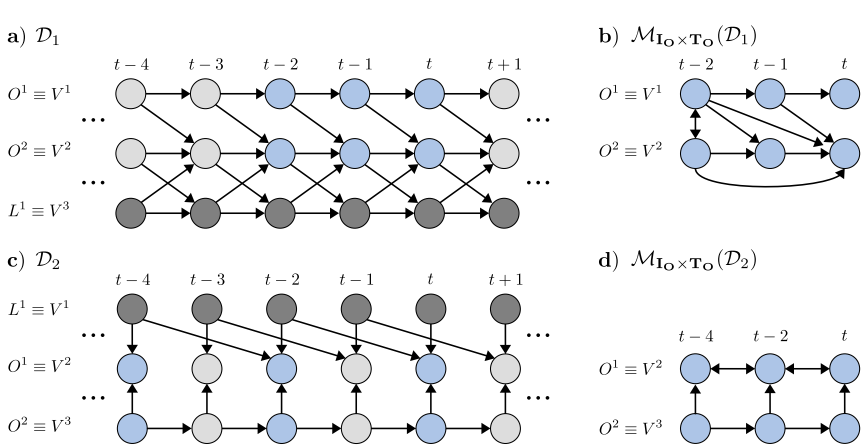

FIG. 2. *The ts-DAG $\mathcal{D}_1$ in part* (*a*) *implies the ts-DMAG $\mathcal{M}_{\mathbf{I_O}\times\mathbf{T_O}}(\mathcal{D}_1)$ in part* (*b*) *for $\mathbf{I_O} = \{1, 2\}$ and $\mathbf{T_O} = \{t-2, t-1, t\}$* (*regular sampling*). *The ts-DAG $\mathcal{D}_2$ in part* (*c*) *implies the ts-DMAG $\mathcal{M}_{\mathbf{I_O}\times\mathbf{T_O}}(\mathcal{D}_2)$ in part* (*d*) *for $\mathbf{I_O} = \{2, 3\}$ and $\mathbf{T_O} = \{t-4, t-2, t\}$* (*regular subsampling*). *Color coding*: *Observed vertices are light blue*, *unobservable vertices are dark gray*, *temporally unobserved observable vertices are light gray*.

REMARK (on Defintion 3.6). The time series DMAG $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ is defined as the MAG latent projection of an *infinite* object, namely of the ts-DAG $\mathcal{D}$. An implementation of this projection in a procedure that always terminates in finite time is possible but nontrivial. Such a procedure is discussed in Gerhardus et al. (2023). For the present paper, however, this procedure is not needed because all theoretical results and examples either do not require the explicit construction of ts-DMAGs or one can carry out the required projections by hand.

Time series DMAGs are the central objects of interest in this paper and a significant part of the paper deals with deriving their properties. We will see that the repeating edges property of ts-DAGs $\mathcal{D}$ plays an essential role in this regard. As a first step, the following lemma notes which of the defining properties of ts-DAGs carry over to ts-DMAGs.

LEMMA 3.7. *Let $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ be a ts-DMAG. Then*:

1. *$\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ has a time series structure.*
2. *$\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ is time ordered.*
3. *There are cases in which $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ does not have repeating edges.*

While according to part 1 of Lemma 3.7 every ts-DMAG is a DMAG with time series structure, part 2 implies that the reverse is not true. Namely, DMAGs with time series structure that are not time ordered cannot be ts-DMAGs. We thus see that ts-DMAGs are a proper subclass of DMAGs with time series structure. The following example shows that ts-DMAGs $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$ do not in general have repeating edges.

EXAMPLE 3.8. The ts-DMAG in part (b) of Figure 2 does not have repeating edges because there is the edge $O^1_{t-2} {\leftrightarrow} O^2_{t-2}$ although O^1_{t-1} and O^2_{t-1} (and O^1_t and O^2_t) are nonadjacent.

Despite this fact, the repeating edges property of the ts-DAG $\mathcal{D}$ strongly restricts the connectivity pattern of the ts-DMAG $\mathcal{M}_{\mathbf{O}}(\mathcal{D})$. We will work out these restrictions in Section 4.

4. Characterization of ts-DMAGs. The main goal of this section is to characterize the space of ts-DMAGs, that is, to find conditions that specify exactly which DMAGs with time series structure are ts-DMAGs. Theorem 1 in Section 4.6 achieves this goal by providing a single condition that is both necessary and sufficient. The theorem uses the notion of *canonical ts-DAGs*; see Definition 4.11 in Section 4.5. In Section 4.4, we introduce *stationarified ts-DMAGs* and, more generally, the concept of *stationarification*. This concept simplifies the definition of canonical ts-DAGs and is useful to describe the output of two recent time series causal discovery algorithms (see Section 5.5). In Section 4.8, we show that ts-DMAGs constitute a strict subset of the classes of graphical models that have so far been used in the literature for describing time-lag specific causal relationships and independencies in time series with latent confounders. Section 4.3 discusses several properties that ts-DMAGs necessarily have, but which can also be obeyed by DMAGs that are not ts-DMAGs. These properties are useful for the discussions in Sections 4.8 and 5. In Section 4.2, we show that regular sampling and regular subsampling are equivalent from a graphical point of view. Section 4.7 gives a characterization of the space of stationarified ts-DMAGs. At first, however, we spell out the motivation for the analysis.

4.1. *Motivation.* When using a class of graphs to represent causal knowledge, it is desirable to know which graphs belong to this class and which do not. Otherwise, it is impossible to fully characterize which causal claims a given graph of that class conveys. Another, a posteriori motivation has been mentioned in the previous paragraph: In Section 4.8, we will see that ts-DMAGs are a strict subset of the previously employed model classes. Thus, when using ts-DMAGs as targets of inference in causal discovery or to reason about causal effects, it is, respectively, possible to learn more qualitative causal relationships (see Sections 5.4 and 5.5 for an in-depth discussion) and to identify more causal effects (see Example 5.8) from data without having imposed any additional assumption or restriction.

4.2. *Equivalence of regular subsampling and regular sampling.* In Section 3.4, we restricted the set of observed time steps $\mathbf{T_O}$ to regular sampling or regular subsampling. While different at first sight, these two cases are equivalent in the following sense.

LEMMA 4.1. *Let $\mathcal{D}$ be a ts-DAG and $1 \leq n_{steps} \in \mathbb{N}$. For $1 \leq n \in \mathbb{Z}$ define the set $\mathbf{T}^n_\mathbf{O} = \{t - m \cdot n | 0 \leq m \leq n_{steps} - 1\}$. Then, with equality up to relabeling vertices*:

1. *For every $n > 1$, there is a ts-DAG $\mathcal{D}'$ such that $\mathcal{M}_{\mathbf{I_O} \times \mathbf{T}^n_\mathbf{O}}(\mathcal{D}) = \mathcal{M}_{\mathbf{I_O} \times \mathbf{T}^1_\mathbf{O}}(\mathcal{D}')$.*
2. *For every $n > 1$, there is a ts-DAG $\mathcal{D}'$ such that $\mathcal{M}_{\mathbf{I_O} \times \mathbf{T}^1_\mathbf{O}}(\mathcal{D}) = \mathcal{M}_{\mathbf{I_O} \times \mathbf{T}^n_\mathbf{O}}(\mathcal{D}')$.*

Lemma 4.1 implies the following: Every property that ts-DMAGs necessarily have in case of regular sampling is also necessarily obeyed in case of regular subsampling (part 1) and vice versa (part 2). Moreover, every set of additional properties that, when imposed on a DMAG $\mathcal{M}$ with time series structure, is sufficient for $\mathcal{M}$ to be a ts-DMAG in case of regular sampling is also sufficient in case of regular subsampling (part 2) and vice versa (part 1).

Due to this equivalence, we from here on restrict to regular sampling, without losing generality, and write $\mathcal{M}^p(\mathcal{D})$ for $\mathcal{M}_\mathbf{O}(\mathcal{D})$ where $\mathbf{O} = \mathbf{I_O} \times \mathbf{T_O}$ and $\mathbf{T_O} = \{t - \tau | 0 \leq \tau \leq p\}$.

4.3. *Properties of ts-DMAGs.* In this subsection, we discuss several properties that ts-DMAGs $\mathcal{M}^p(\mathcal{D})$ necessarily have. These properties are such that a certain graphical property persists when the involved vertices are shifted in time. We use the following definitions.

DEFINITION 4.2 (Time-shift persistent graphical properties). A partial mixed graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with time series structure has…

1. … *repeating adjacencies* if the following holds: If $((i, t_i), (j, t_j)) \in \mathbf{E}$ and $(i, t_i + \Delta t), (j, t_j + \Delta t) \in \mathbf{V}$, then $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}$.
2. … *past-repeating adjacencies* if the following holds: If $((i, t_i), (j, t_j)) \in \mathbf{E}$ and $(i, t_i + \Delta t), (j, t_j + \Delta t) \in \mathbf{V}$ with $\Delta t < 0$, then $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}$.
3. … *repeating orientations* if the following holds: If $((i, t_i), (j, t_j)) \in \mathbf{E}_\bullet$ with $\bullet \in \{\rightarrow, \leftrightarrow, \circ\!\!\rightarrow, \circ\!\!-\!\!\circ\}$ and $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}$, then $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}_\bullet$.

A DMAG $\mathcal{M} = (\mathbf{V}, \mathbf{E})$ with time series structure has

4. … *repeating ancestral relationships* if the following holds: If $(i, t_i) \in \mathrm{an}((j, t_j), \mathcal{M})$ and $(i, t_i + \Delta t), (j, t_j + \Delta t) \in \mathbf{V}$, then $(i, t_i + \Delta t) \in \mathrm{an}((j, t_j + \Delta t), \mathcal{M})$.
5. … *repeating separating sets* if the following holds: If $(i, t_i) \perp\!\!\!\perp (j, t_j) | \mathbf{S}$ and $\{(i, t_i + \Delta t), (j, t_j + \Delta t)\} \cup \mathbf{S}_{\Delta t} \subseteq \mathbf{V}$, where $\mathbf{S}_{\Delta t}$ is obtained by shifting every vertex in $\mathbf{S}$ by Δt time steps, then $(i, t_i + \Delta t) \perp\!\!\!\perp (j, t_j + \Delta t) | \mathbf{S}_{\Delta t}$.

REMARK (on Definition 4.2). Section 4 is concerned with DAGs and DMAGs only. However, in Section 5 we will apply the concepts of repeating adjacencies, past-repeating adjacencies and repeating orientations also to DPAGs (which are a special case of partial mixed graphs). Hence, we already here formulate the definition in sufficient generality.

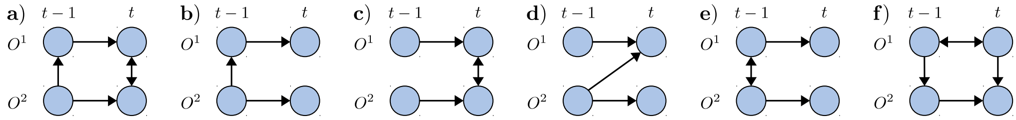

FIG. 3. *Examples of time ordered DMAGs with time series structure for illustrating the properties from Definition 4.2 and the repeating edges property from Definition 3.3. In each case, we state which of these properties apply.* a) *Repeating adjacencies, repeating separating sets, past-repeating adjacencies.* b) *Repeating orientations, repeating separating sets, past-repeating adjacencies.* c) *Repeating orientations, repeating ancestral relationships.* d) *All but repeating separating sets.* e) *All but repeating repeating edges and repeating adjacencies.* f) *All.*

Figure 3 illustrates the five properties introduced by Definition 4.2 as well as their distinctions. Below we will make frequent use of the implications expressed by the following lemma.

LEMMA 4.3.

1. *Repeating edges is equivalent to the combination of repeating adjacencies and repeating orientations.*
2. *Repeating adjacencies implies past-repeating adjacencies.*
3. *Repeating ancestral relationships implies repeating orientations.*
4. *In graphs with time index set* $\mathbf{T} = \mathbb{Z}$, *repeating edges implies repeating ancestral relationships and repeating separating sets.*

These implications further show that the combination of repeating adjacencies and repeating ancestral relationships implies repeating edges. Importantly, repeating orientations does *not* imply repeating ancestral relationships; see part (b) of Figure 3 for an example.

Since ts-DAGs have repeating edges, according to Lemma 4.3 they in fact also have all five properties given in Definition 4.2. How about ts-DMAGs? While these in general *do not* inherit repeating edges from the underlying ts-DAG (see part 3 of Lemma 3.7), the following lemma shows that ts-DMAGs do feature some of the weaker time-shift persistent properties.

LEMMA 4.4.

1. *Time series DMAGs* $\mathcal{M}^p(\mathcal{D})$ *have repeating ancestral relationships.*
2. *Time series DMAGs* $\mathcal{M}^p(\mathcal{D})$ *have repeating orientations.*
3. *Time series DMAGs* $\mathcal{M}^p(\mathcal{D})$ *have repeating separating sets.*
4. *Time series DMAGs* $\mathcal{M}^p(\mathcal{D})$ *have past-repeating adjacencies.*
5. *There are cases in which a ts-DMAG* $\mathcal{M}^p(\mathcal{D})$ *does not have repeating adjacencies.*

The ts-DMAGs in parts (b) and (d) of Figure 2 indeed satisfy the properties asserted by parts 1 through 4 of Lemma 4.4. Moreover, part 5 of Lemma 4.4 clarifies why ts-DMAGs may fail to have repeating edges: They do not necessarily have repeating adjacencies but only the weaker property of past-repeating adjacencies. The following example illustrates this fact.

EXAMPLE 4.5. Consider the ts-DAG $\mathcal{D}_1$ in part (a) of Figure 2. In this graph, the d-separation $O^1_{t+\Delta t} \perp\!\!\!\perp O^2_{t+\Delta t} | O^1_{t+\Delta t-1}$ holds for all $\Delta t \in \mathbb{Z}$. Hence, the vertices O^1_t and O^2_t (and, similarly, O^1_{t-1} and O^2_{t-1}) are nonadjacent in the ts-DMAG $\mathcal{M}^2(\mathcal{D}_1)$ in part (b) of the figure. However, since O^1_{t-3} is temporally unobserved and the d-separation $O^1_{t+\Delta t} \perp\!\!\!\perp O^2_{t+\Delta t} | \mathbf{S}$ requires that $O^1_{t+\Delta t-1} \in \mathbf{S}$, the vertices O^1_{t-2} and O^1_{t-2} are adjacent in $\mathcal{M}^2(\mathcal{D}_1)$.

That ts-DMAGs have repeating orientations and repeating separating sets has already been found and used in Entner and Hoyer (2010).

4.4. *Stationarified ts-DMAGs*. Example 4.5 shows that in a ts-DMAG $\mathcal{M}^p(\mathcal{D})$ there may be an edge $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ with $\Delta t < 0$ even if the vertices (i, t_i) and (j, t_j) are nonadjacent in $\mathcal{M}^p(\mathcal{D})$. This is the case even though one then knows that $(i, t_i + \Delta t)$ and $(j, t_j + \Delta t)$ *can* be d-separated in underlying ts-DAG $\mathcal{D}$, just not by a set of vertices that is within the observed time window. One might thus view such an edge $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ in $\mathcal{M}^p(\mathcal{D})$ as an artifact of the chosen time window, and hence prefer to manually remove the edge by subjecting the ts-DMAG to the following operation.

DEFINITION 4.6 (Stationarification). Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a directed partial mixed graph with time series structure. The *stationarification* of $\mathcal{G}$, denoted as stat($\mathcal{G}$), is the graph stat$(\mathcal{G}) = (\mathbf{V}', \mathbf{E}')$ defined as follows:

1. It has the same set of vertices as $\mathcal{G}$, that is, $\mathbf{V}' = \mathbf{V}$.
2. There is an edge $((i, t_i), (j, t_j)) \in \mathbf{E}'_\bullet$ with $\bullet \in \{\rightarrow, \leftrightarrow, \circ\!\!\rightarrow, \circ\!\!-\!\!\circ\}$ if and only if $((i, t_i + \Delta t), (j, t_j + \Delta t)) \in \mathbf{E}_\bullet$ in $\mathcal{G}$ for all Δt with $(i, t_i + \Delta t), (j, t_j + \Delta t) \in \mathbf{V}$.

REMARK (on Definition 4.6). Section 4 is concerned with DAGs and DMAGs only. In these graphs, there are by definition no edges of the types $\circ\!\!\rightarrow$ or $\circ\!\!-\!\!\circ$. However, in Section 5 we will apply the concept of sationarification also to DPAGs. Since these graphs (DPAGs) can contain edges $\circ\!\!\rightarrow$ or $\circ\!\!-\!\!\circ$, we already here formulate the definition in sufficient generality.

To see that the process of stationarification indeed achieves what it is supposed to do, consider the ts-DMAG $\mathcal{M}_1$ in part (a) of Figure 4. In this graph, there is the edge $O^1_{t-2} \leftrightarrow O^2_{t-2} \in \mathbf{E}_\leftrightarrow$ while the vertices O^1_{t-1} and O^2_{t-1} (and, similarly, O^1_t and O^2_t) are nonadjacent. According to part 2 of Definition 4.6 (note the "for all Δt"), the vertices O^1_{t-2} and O^2_{t-2} are therefore nonadjacent in the stationarification stat($\mathcal{M}_1$) of $\mathcal{M}_1$ as shown in part (b) of Figure 4.

Stationarification removes an edge $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ also if (i, t_i) and (j, t_j) are adjacent but if the edges $(i, t_i){*\!\!-\!\!*}(j, t_j)$ and $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ have different orientations (note the "$\bullet$" subscripts on $\mathbf{E}'_\bullet$ and $\mathbf{E}_\bullet$ in part 2 of Definition 4.6). This effect, illustrated by parts (c) and (d) of Figure 4, ensures that stat($\mathcal{G}$) is the unique largest subgraph of $\mathcal{G}$ with repeating edges. For graphs with repeating orientations (as, e.g., ts-DMAGs), this effect does not occur and stationarification only concerns adjacencies (as, e.g., in parts (a) and (b) of Figure 4).

Since ts-DMAGs $\mathcal{M}^p(\mathcal{D})$ have repeating orientations and past-repeating adjacencies, their stationarifications stat($\mathcal{M}^p(\mathcal{D})$) can be characterized with the following simpler condition.

LEMMA 4.7. *The stationarification* stat($\mathcal{M}^p(\mathcal{D})$) *of a ts-DMAG* $\mathcal{M}^p(\mathcal{D})$ *is the unique subgraph of* $\mathcal{M}^p(\mathcal{D})$ *in which the vertices* $(i, t_j - \tau)$ *and* (j, t_j) *with* $\tau \geq 0$ *are adjacent if and only if the vertices* $(i, t - \tau)$ *and* (j, t) *are adjacent in* $\mathcal{M}^p(\mathcal{D})$.

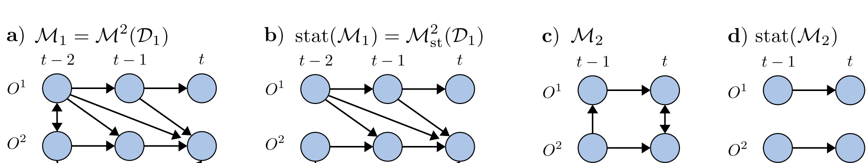

FIG. 4. *A ts-DMAG* $\mathcal{M}_1 = \mathcal{M}^2(\mathcal{D}_1)$ (*the same as in part* (*b*) *of Figure* 2) *and a DMAG with time series structure* $\mathcal{M}_2$ (*the same as in part* (*a*) *of Figure* 3) *together with their stationarifications. Note that although* $\mathcal{M}_2$ *has repeating adjacencies its contemporaneous edges are not in* stat($\mathcal{M}_2$) *because these edges do not have the same orientation.*

Because the stationarification stat($\mathcal{G}$) is a subgraph of $\mathcal{G}$, a time series structure and time order naturally carry over from $\mathcal{G}$ to stat($\mathcal{G}$). Moreover, we can prove the following.

LEMMA 4.8. *The stationarification* stat($\mathcal{M}^p(\mathcal{D})$) *of a ts-DMAG* $\mathcal{M}^p(\mathcal{D})$ *is a DMAG.*

We thus refer to stat($\mathcal{M}^p(\mathcal{D})$) as a *stationarified ts-DMAG* and abbreviate stat($\mathcal{M}^p(\mathcal{D})$) as $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$. However, as the following example shows, a stationarified ts-DMAG $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ may not be the MAG latent projection of any ts-DAG, that is, may not be a ts-DMAG.

EXAMPLE 4.9. The stationarified ts-DMAG $\mathcal{M}^2_{\mathrm{st}}(\mathcal{D}_1)$ in part (b) of Figure 4 implies the d-separation $O^1_{t-2} \perp\!\!\!\perp O^2_{t-2}$ and the d-connections $O^1_{t-1} \not\perp\!\!\!\perp O^2_{t-1}$ and $O^1_t \not\perp\!\!\!\perp O^2_t$. The graph $\mathcal{M}^2_{\mathrm{st}}(\mathcal{D}_1)$ does thus not have repeating separating sets and can, by means of Lemma 4.4, not be a ts-DMAG. Also, note that in the underlying ts-DAG $\mathcal{D}_1$, shown in part (a) of Figure 2, the d-connection $O^1_{t-2} \not\perp\!\!\!\perp O^2_{t-2}$ holds. From this observation, we learn that $(i, t_i) \perp\!\!\!\perp (j, t_j) | \mathbf{S}$ in $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ does not necessarily imply $(i, t_i) \perp\!\!\!\perp (j, t_j) | \mathbf{S}$ in $\mathcal{D}$.

The vertices $(i, t - \tau_i)$ and $(j, t - \tau_j)$ with $0 \le \tau_j \le \tau_i \le p$ are adjacent in a stationarified ts-DMAG $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ if and only if they cannot be d-separated by any set of observable vertices within $[t - p - \tau_j, t]$ in the underlying ts-DAG $\mathcal{D}$ (instead of $[t - p, t]$, which is what a ts-DMAG would assert). The orientation of edges, however, retains the standard meaning: Tail and head marks, respectively, convey (non)ancestorship according to the ts-DAG $\mathcal{D}$. The following lemma says that stationarification does not change ancestral relationships.

LEMMA 4.10. *The ts-DMAG* $\mathcal{M}^p(\mathcal{D})$ *and its stationarification* $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ *agree on ancestral relationships, that is,* $(i, t_i) \in \mathrm{an}((j, t_j), \mathcal{M}^p(\mathcal{D}))$ *if and only if* $(i, t_i) \in \mathrm{an}((j, t_j), \mathcal{M}^p_{\mathrm{st}}(\mathcal{D}))$.

Since $\mathcal{M}^p(\mathcal{D})$ and $\mathcal{D}$ by construction of the MAG latent projection agree on ancestral relationships, Lemma 4.10 implies that also the stationarified ts-DMAG $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ agrees with the ancestral relationships of $\mathcal{D}$. Thus, $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ has repeating ancestral relationships.

In summary, edges in the ts-DMAG $\mathcal{M}^p(\mathcal{D})$ that are not also in $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ are due to marginalizing over observable vertices before $t - p$. Such edges disappear when p is sufficiently increased; see also Gerhardus ((2023), Section B.8). However, as we will show in Section 5.3, these additional edges play a useful role in causal discovery. In Section 5.5, we will further use the concept of stationarification to describe the SVAR-FCI causal discovery algorithm from Malinsky and Spirtes (2018) and the LPCMCI causal discovery algorithm from Gerhardus and Runge (2020).

4.5. *Canonical ts-DAGs.* In the current subsection, we return to the goal of characterizing the space of ts-DMAGs. To this end, we first recall the concept of *canonical DAGs*.

DEFINITION 4.11 (Canonical DAG. From Section 6.1 of Richardson and Spirtes (2002), specialized to the case of *directed* ancestral graphs). Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a directed ancestral graph. The canonical DAG $\mathcal{D}_c(\mathcal{G})$ of $\mathcal{G}$ is the graph $\mathcal{D}_c(\mathcal{G}) = (\mathbf{V}^{\mathrm{ca}}, \mathbf{E}^{\mathrm{ca}})$ defined as follows:

1. Its vertex set is $\mathbf{V}^{\mathrm{ca}} = \mathbf{V} \cup \mathbf{L}$, where $\mathbf{L} = \{l_{ij} | (i, j) \in \mathbf{E}_{\leftrightarrow}\}$.
2. Its edge set $\mathbf{E}^{\mathrm{ca}} = \mathbf{E}^{\mathrm{ca}}_{\rightarrow}$ consists of the edges:
   - $i \rightarrow j$ for all $(i, j) \in \mathbf{E}_{\rightarrow}$ and
   - $l_{ij} \rightarrow i$ for all $l_{ij} \in \mathbf{L}$ and
   - $l_{ij} \rightarrow j$ for all $l_{ij} \in \mathbf{L}$.

Intuitively, the canonical DAG $\mathcal{D}_c(\mathcal{G})$ of a directed ancestral graph $\mathcal{G}$ is obtained by replacing each bidirected edge $i \leftrightarrow j$ in $\mathcal{G}$ with $i \leftarrow l_{ij} \rightarrow j$ where l_{ij} is an additionally inserted, unobserved vertex. The canonical DAG $\mathcal{D}_c(\mathcal{G})$ is a DAG and has the convenient property that there are no edges pointing into unobserved vertices, and hence that there are also no edges between two unobserved vertices. Despite this simple structure of unobserved vertices, the following result shows that canonical DAGs are expressive enough to generate all DMAGs.

LEMMA 4.12 (Theorem 6.4 in Richardson and Spirtes (2002), specialized to *directed* ancestral graphs). *If $\mathcal{M}$ is a DMAG over vertex set $\mathbf{O}$, then the MAG latent projection $\mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$ of the canonical DAG $\mathcal{D}_c(\mathcal{M})$ of $\mathcal{M}$ equals $\mathcal{M}$, that is, $\mathcal{M} = \mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$.*

Lemma 4.12 means that every DMAG is the MAG latent projection of some DAG. Moreoever, the condition $\mathcal{M} = \mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$ yields a characterization of DMAGs in the sense that a directed ancestral graph $\mathcal{G}$ is a DMAG if and only if it meets the condition $\mathcal{G} = \mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{G}))$. Because DMAGs are already characterized by definition,[1] the alternative characterization by the condition $\mathcal{G} = \mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{G}))$ is of limited use in this case.

For ts-DMAGs, however, there is no definitional characterization. In addition, because not every DMAG with time series structure is a ts-DMAG (see the explanation below Lemma 3.7), characterizing ts-DMAGs is a nontrivial task. In the remaining parts of the current subsection and Section 4.6, we show that ts-DMAGs can be characterized by an appropriate generalization of the condition $\mathcal{G} = \mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{G}))$. The first step of such a generalization is to find an appropriate generalization of canonical DAGs.

The generalization of canonical DAGs to the time series setting is nontrivial for the following reason. Consider an edge $(i, t_i){*\!\!-\!\!*}(j, t_j)$ in a DMAG $\mathcal{M}$ with time series structure that is not in the DMAG's stationarification $\mathrm{stat}(\mathcal{M})$. If, depending on the orientation of the edge $(i, t_i){*\!\!-\!\!*}(j, t_j)$ in $\mathcal{M}$, either $(i, t_i) \rightarrow (j, t_j)$ or $(i, t_i) \leftarrow (j, t_j)$ or $(i, t_i) \leftarrow (l_{ij}, t_{ij}) \rightarrow (j, t_j)$ with (l_{ij}, t_{ij}) unobserved were included in a "canonical ts-DAG" $\mathcal{D}_c(\mathcal{M})$, then the repeating edges property of ts-DAGs would require the same structure to be present at all other time steps, too. Hence, in $\mathcal{D}_c(\mathcal{M})$ there would be $(i, t_i + \Delta t) \rightarrow (j, t_j + \Delta t)$ or $(i, t_i + \Delta t) \leftarrow (j, t_j + \Delta t)$ or $(i, t_i + \Delta t) \leftarrow (l_{ij}, t_{ij} + \Delta t) \rightarrow (j, t_j + \Delta t)$ for all $\Delta t \in \mathbb{Z}$. Consequently, in the MAG latent projection $\mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$ of $\mathcal{D}_c(\mathcal{M})$ there would be an edge $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ of the same type for all Δt. But then also in the stationarification $\mathrm{stat}(\mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M})))$ of $\mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$ there would be the edge $(i, t_i + \Delta t){*\!\!-\!\!*}(j, t_j + \Delta t)$ for all Δt. Hence, $\mathcal{M}$ could not equal $\mathcal{M}_{\mathbf{O}}(\mathcal{D}_c(\mathcal{M}))$.

Given these considerations, the canonical ts-DAG $\mathcal{D}_c(\mathcal{M})$ of a ts-DMAG $\mathcal{M}$ should instead only take into account the edges in the stationarification $\mathrm{stat}(\mathcal{M})$ of $\mathcal{M}$. We are thus lead to the following definition, which for use further below is not restricted to ts-DMAGs but more generally applies to acyclic directed mixed graphs.

DEFINITION 4.13 (Canonical ts-DAG). Let $\mathcal{G}$ be an acyclic directed mixed graph with time series structure and let $\mathbf{V} = \mathbf{I} \times \mathbf{T}$ with $\mathbf{T} = \{t - \tau | 0 \leq \tau \leq p\}$ be its set of vertices. Denote with $\mathbf{E}^{\mathrm{stat}}$ the set of edges of $\mathrm{stat}(\mathcal{G})$. The canonical ts-DAG associated to $\mathcal{G}$, denoted as $\mathcal{D}_c(\mathcal{G})$, is the graph $\mathcal{D}_c(\mathcal{G}) = (\mathbf{V}^{\mathrm{ca}}, \mathbf{E}^{\mathrm{ca}})$ defined as follows:

1. Its set of vertices is $\mathbf{V}^{\mathrm{ca}} = (\mathbf{I} \cup \mathbf{J}) \times \mathbb{Z}$, where $\mathbf{J} = \{(i, j, \tau) | ((i, t - \tau), (j, t)) \in \mathbf{E}^{\mathrm{stat}}_{\leftrightarrow}\}$. The variable index set is $\mathbf{I} \cup \mathbf{J}$ and the time index set is $\mathbb{Z}$.
2. Its set of edges $\mathbf{E}^{\mathrm{ca}} = \mathbf{E}^{\mathrm{ca}}_{\rightarrow}$ are, for all $\Delta t \in \mathbb{Z}$,
   - $(i, t - \tau + \Delta t) \rightarrow (j, t + \Delta t)$ for all $((i, t - \tau), (j, t)) \in \mathbf{E}^{\mathrm{stat}}_{\rightarrow}$ and

[1] As directed ancestral graphs without inducing paths between nonadjacent vertices, see Section 2.

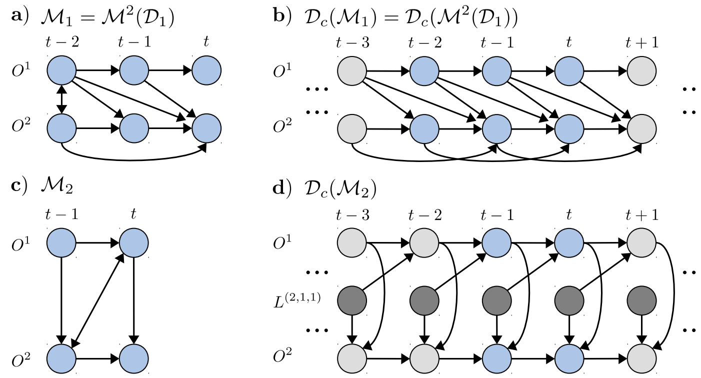

FIG. 5. *A ts-DMAG $\mathcal{M}_1 = \mathcal{M}^2(\mathcal{D}_1)$ (the same as in part (b) of Figure 2 and part (a) of Figure 4) and a DMAG with time series structure $\mathcal{M}_2$ together with their canonical ts-DAGs. In $\mathcal{D}_c(\mathcal{M}_1)$, there is no unobservable time series because in $\mathcal{M}_1$ there is no bidirected edge that is repetitive in time, and hence there is no bidirected edge in* $\mathrm{stat}(\mathcal{M}_1)$. *The unobservable time series $L^{(2,1,1)}$ in $\mathcal{D}_c(\mathcal{M}_2)$ in the notation of Definition 4.13 corresponds to* $(2, 1, 1) \in \mathbf{J}$ *and results from the edge $O^2_{t-1} \leftrightarrow O^1_t$ in* $\mathrm{stat}(\mathcal{M}_2) = \mathcal{M}_2$.

- $((i, j, \tau), t + \Delta t) \rightarrow (i, t + \Delta t)$ for all $(i, j, \tau) \in \mathbf{J}$ and
- $((i, j, \tau), t - \tau + \Delta t) \rightarrow (j, t + \Delta t)$ for all $(i, j, \tau) \in \mathbf{J}$.

Figure 5 illustrates canonical ts-DAGs. Intuitively, the canonical ts-DAG $\mathcal{D}_c(\mathcal{G})$ of $\mathcal{G}$ is obtained in three steps: First, replace $\mathcal{G}$ by its stationarification $\mathrm{stat}(\mathcal{G})$. Second, in $\mathrm{stat}(\mathcal{G})$ replace every bidirected edge $(i, t_j - \tau) \leftrightarrow (j, t_j)$ with $(i, t_j - \tau) \leftarrow ((i, j, \tau), t_j - \tau) \rightarrow (j, t_j)$ where $((i, j, \tau), t_j - \tau)$ is an additionally inserted, unobserved vertex. Third, repeat this structure into the infinite past and future according to the repeating edges property. This intuition identifies the vertices $((i, j, \tau), s)$ with $(i, j, \tau) \in \mathbf{J}$ and $s \in \mathbb{Z}$ as analogs of the unobserved vertices $l_{ij} \in \mathbf{L}$ in standard canonical DAGs (see Definition 4.11 above) and, in addition, means that the time series indexed by $\mathbf{J}$ are treated as unobservable. The key difference between standard canonical DAGs and canonical ts-DAGs is the first of the three steps, that is, the application of stationarification. A similarity is that also in canonical ts-DAGs there are no edges into unobservable vertices, and hence no edges between two unobservable vertices.

Canonical ts-DAGs are indeed ts-DAGs and, by means of the following result, yield the desired generalization of Lemma 4.12.

LEMMA 4.14. *Let $\mathcal{D}$ be a ts-DAG with variable index set* $\mathbf{I}$. *Let* $\mathbf{I_O} \subseteq \mathbf{I}$ *and* $\mathbf{T_O} = \{t - \tau \,|\, 0 \leq \tau \leq p\}$ *with* $p \geq 0$. *Then* $\mathcal{M}_\mathbf{O}(\mathcal{D}) = \mathcal{M}_\mathbf{O}(\mathcal{D}_c(\mathcal{M}_\mathbf{O}(\mathcal{D})))$ *with* $\mathbf{O} = \mathbf{I_O} \times \mathbf{T_O}$.

REMARK (on Lemma 4.14). The lemma involves two different MAG latent projections: First, the projection of the ts-DAG $\mathcal{D}$ to the ts-DMAG $\mathcal{M}_\mathbf{O}(\mathcal{D})$. Second, the projection of the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}_\mathbf{O}(\mathcal{D}))$ of $\mathcal{M}_\mathbf{O}(\mathcal{D})$ to $\mathcal{M}_\mathbf{O}(\mathcal{D}_c(\mathcal{M}_\mathbf{O}(\mathcal{D})))$. In the first projection, the time series indexed by $\mathbf{I} \setminus \mathbf{I_O}$ are unobservable. In the second projection, the time series indexed by the set $\mathbf{J}$ are unobservable. In both projections, all vertices before $t - p$ and after t are temporally unobserved. However, since the set of observed variables is the same in both projections (namely $\mathbf{O}$), no confusion arises when writing $\mathcal{M}^p(\mathcal{D}) = \mathcal{M}^p(\mathcal{D}_c(\mathcal{M}^p(\mathcal{D})))$ instead of $\mathcal{M}_\mathbf{O}(\mathcal{D}) = \mathcal{M}_\mathbf{O}(\mathcal{D}_c(\mathcal{M}_\mathbf{O}(\mathcal{D})))$. From here on, we adopt this notation.

Lemma 4.14 says that the composition of creating the canonical ts-DAG and then projecting back to the original vertices is the identity operation on the space of ts-DMAGs; see Figure 6. This result is far from obvious for two reasons: First, if an edge $(i, t_i){*\!-\!*}(j, t_j)$ in a ts-DMAG $\mathcal{M}^p(\mathcal{D})$ is not in the stationarified ts-DMAG $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ then in the canonical ts-DAG

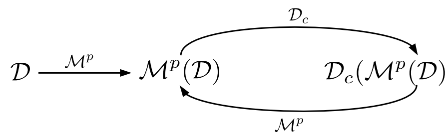

FIG. 6. *Conceptual illustration of Lemma 4.14.*

$\mathcal{D}_c(\mathcal{M}^p(\mathcal{D}))$ there is neither $(i, t_i) \rightarrow (j, t_j)$ nor $(i, t_i) \leftarrow (j, t_j)$ nor $(i, t_i) \leftarrow (l_{ij}, t_{ij}) \rightarrow (j, t_j)$ with (l_{ij}, t_{ij}) unobservable. Hence, the edge $(i, t_i) {*\!\!-\!\!*} (j, t_j)$ needs to appear in the MAG latent projection $\mathcal{M}^p(\mathcal{D}_c(\mathcal{M}^p(\mathcal{D})))$ of $\mathcal{D}_c(\mathcal{M}^p(\mathcal{D}))$ in a nontrivial way, namely because of marginalizing over the temporally unobserved vertices. Second, this marginalization over the vertices before $t - p$ must not create superfluous edges.

EXAMPLE 4.15. The example in Figure 7 illustrates Lemma 4.14. This example also shows that the original ts-DAG $\mathcal{D}$ and the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}^p(\mathcal{D}))$ need not be equal.

We stress that Lemma 4.14 holds for *arbitrary* ts-DAGs $\mathcal{D}$. In particular, in $\mathcal{D}$ there may be what in Malinsky and Spirtes (2018) is referred to as "auto-lag confounders," namely unobservable autocorrelated component time series L, that is, $L_{t-\tau} \rightarrow L_t$ with L unobservable.

4.6. *A necessary and sufficient condition that characterizes ts-DMAGs.* Lemma 4.14 readily implies the following characterization of ts-DMAGs as a subclass of DMAGs with time series structure by a single necessary and sufficient condition.

THEOREM 1. *Let $\mathcal{M}$ be a DMAG with time series structure and time index set $\mathbf{T} = \{t - \tau \,|\, 0 \leq t \leq p\}$. Then $\mathcal{M}$ is a ts-DMAG, that is, there is a ts-DAG $\mathcal{D}$ such that $\mathcal{M} = \mathcal{M}^p(\mathcal{D})$ if and only if the MAG latent projection $\mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$ of the canonical ts-DAG $\mathcal{D}_c(\mathcal{M})$ of $\mathcal{M}$ equals $\mathcal{M}$, that is, if and only if $\mathcal{M} = \mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$.*

Theorem 1 is one of the central results of this paper. The following four examples are included for its illustration.

EXAMPLE 4.16. The DMAG $\mathcal{M}_2$ in part (c) of Figure 5 is a ts-DMAG. This conclusion follows because the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}_2)$ in part (d) of the figure projects to $\mathcal{M}_2$.

EXAMPLE 4.17. One may use Theorem 1 to confirm that none of the four DMAGs in parts (a)–(d) of Figure 3 is a ts-DMAG. In these cases, this conclusion also follows because each of these four graphs violates at least one of the necessary conditions in Lemmas 3.7 and 4.4.

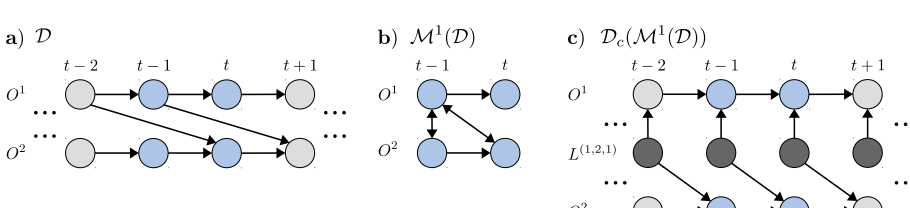

FIG. 7. *A ts-DAG $\mathcal{D}$ together with the ts-DMAG $\mathcal{M}^1(\mathcal{D})$ and the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}^1(\mathcal{D}))$ of the ts-DMAG. Marginalizing $\mathcal{D}_c(\mathcal{M}^1(\mathcal{D}))$ to the observed vertices gives back $\mathcal{M}^1(\mathcal{D})$. Same color coding as in Figure 2.*

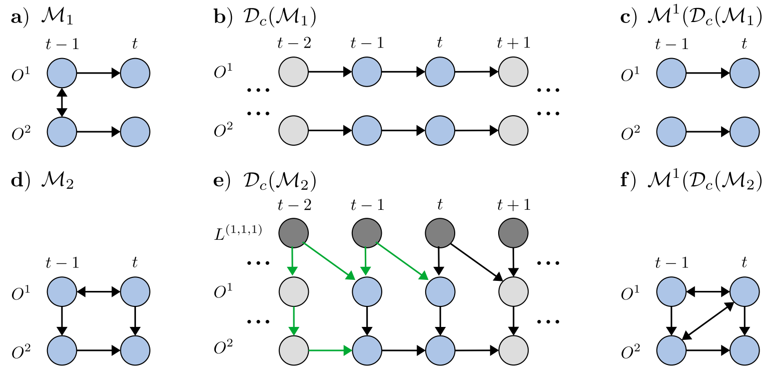

FIG. 8. *Two examples of DMAGs with time series structure, $\mathcal{M}_1$ (the same as in part part (e) of Figure 3) and $\mathcal{M}_2$ (the same as in part (f) of Figure 3), that are not ts-DMAGs although they obey all necessary conditions in Lemmas 3.7 and 4.4. See also the discussions in Examples 4.18 and 4.19.*

EXAMPLE 4.18. The DMAG $\mathcal{M}_1$ in part (a) of Figure 8 is not a ts-DMAG because its canonical ts-DAG $\mathcal{D}_c(\mathcal{M}_1)$ in part (b) projects to the ts-DMAG $\mathcal{M}^1(\mathcal{D}_c(\mathcal{M}_1))$ in part (c), which is a proper subgraph of $\mathcal{M}_1$. This example also demonstrates that the equality $\text{stat}(\mathcal{M}) = \mathcal{M}^p_{\text{st}}(\mathcal{D}_c(\mathcal{M}))$ is not sufficient for $\mathcal{M}$ to be a ts-DMAG.

EXAMPLE 4.19. The DMAG $\mathcal{M}_2$ in part (d) of Figure 8 is not a ts-DMAG since its canonical ts-DAG $\mathcal{D}_c(\mathcal{M}_2)$ in part (e) projects to the ts-DMAG $\mathcal{M}^1(\mathcal{D}_c(\mathcal{M}_2))$ in part (f), which is a proper supergraph of $\mathcal{M}_2$. The edge $O^2_{t-1} {\leftrightarrow} O^1_t$ in $\mathcal{M}^1(\mathcal{D}_c(\mathcal{M}_2))$ is due to the green colored inducing path $O^2_{t-1} {\leftarrow} O^2_{t-2} {\leftarrow} O^1_{t-2} {\leftarrow} L^{(1,1,1)}_{t-2} {\rightarrow} O^1_{t-1} {\leftarrow} L^{(1,1,1)}_{t-1} {\rightarrow} O^1_t$ in $\mathcal{D}_c(\mathcal{M}_2)$.

Importantly, the DMAGs considered in Examples 4.18 and 4.19 obey all necessary conditions given in Lemmas 3.7 and 4.4. The condition $\mathcal{M} = \mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$ is thus strictly stronger than even the combination of all these necessary conditions. This observation clearly demonstrates the significance and nontriviality of Theorem 1.

As an alternative to Theorem 1, we also characterize ts-DMAGs as a subclass of directed mixed graphs with time series structure.

THEOREM 2. *Let $\mathcal{G}$ be a directed mixed graph with time series structure and time index set $\mathbf{T} = \{t - \tau | 0 \leq t \leq p\}$. Then $\mathcal{G}$ is a ts-DMAG, that is, there is a ts-DAG $\mathcal{D}$ such that $\mathcal{G} = \mathcal{M}^p(\mathcal{D})$ if and only if $\mathcal{G}$ is acyclic and $\mathcal{G} = \mathcal{M}^p(\mathcal{D}_c(\mathcal{G}))$.*

Theorem 2 is even stronger than Theorem 1 because Theorem 2 does not require the graph $\mathcal{G}$ to be ancestral and/or maximal. Acyclicity, however, is needed because the definition of canonical ts-DAGs $\mathcal{D}_c(\mathcal{G})$ requires $\mathcal{G}$ to be acyclic, as does the notion of d-separation.

4.7. *Implications for stationarified ts-DMAGs.* A ts-DMAG $\mathcal{M}^p(\mathcal{D})$ by definition uniquely determines its stationarification $\mathcal{M}^p_{\text{st}}(\mathcal{D})$. How about the opposite? That is, can a ts-DMAG $\mathcal{M}^p(\mathcal{D})$ be uniquely determined from its stationarification $\mathcal{M}^p_{\text{st}}(\mathcal{D})$? At first it seems perfectly conceivable that different ts-DMAGs have the same stationarification, which would make it impossible to uniquely determine $\mathcal{M}^p_{\text{st}}(\mathcal{D})$ from $\mathcal{M}^p(\mathcal{D})$. However, as a corollary to the observation $\mathcal{D}_c(\mathcal{G}) = \mathcal{D}_c(\text{stat}(\mathcal{G}))$ and Lemma 4.14 we get the following result.

LEMMA 4.20. *Let $\mathcal{D}$ be a ts-DAG. Then the ts-DMAG $\mathcal{M}^p(\mathcal{D})$ equals the MAG latent projection $\mathcal{M}^p(\mathcal{D}_c(\mathcal{M}^p_{\text{st}}(\mathcal{D})))$ of the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}^p_{\text{st}}(\mathcal{D}))$ of the stationarification $\mathcal{M}^p_{\text{st}}(\mathcal{D}) = \text{stat}(\mathcal{M}^p(\mathcal{D}))$ of $\mathcal{M}^p(\mathcal{D})$, that is, $\mathcal{M}^p(\mathcal{D}) = \mathcal{M}^p(\mathcal{D}_c(\mathcal{M}^p_{\text{st}}(\mathcal{D})))$.*

According to Lemma 4.20, one can always uniquely determine $\mathcal{M}^p(\mathcal{D})$ from $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$. A ts-DMAG $\mathcal{M}^p(\mathcal{D})$ and its stationarification $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ thus carry the exact same information about the underlying ts-DAG $\mathcal{D}$. In this sense, $\mathcal{M}^p(\mathcal{D})$ and $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ are, if interpreted in the correct way, equivalent descriptions.

Lastly, we also arrive at two characterizations of stationarified ts-DMAGs.

LEMMA 4.21. *Let $\mathcal{M}$ be a DMAG with time series structure and time index set* $\mathbf{T} = \{t - \tau | 0 \leq t \leq p\}$. *Then $\mathcal{M}$ is a stationarified ts-DMAG, that is, there is a ts-DAG $\mathcal{D}$ such that $\mathcal{M} = \mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ if and only if $\mathcal{M} = \mathcal{M}^p_{\mathrm{st}}(\mathcal{D}_c(\mathcal{M}))$.*

LEMMA 4.22. *Let $\mathcal{G}$ be a directed mixed graph with time series structure and time index set* $\mathbf{T} = \{t - \tau | 0 \leq t \leq p\}$. *Then $\mathcal{G}$ is a stationarified ts-DMAG, that is, there is a ts-DAG $\mathcal{D}$ such that $\mathcal{G} = \mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ if and only if $\mathcal{G}$ is acyclic and $\mathcal{G} = \mathcal{M}^p_{\mathrm{st}}(\mathcal{D}_c(\mathcal{G}))$.*

4.8. *Comparison with previously considered model classes.* The author is aware of two distinct classes of graphical models based on DMAGs that have so far been used to represent time-lag specific causal relationships in time series with latent confounders. Here, we show that both these model classes are strictly larger than the class of ts-DMAGs.

The first previously used model class, employed by the tsFCI algorithm from Entner and Hoyer (2010), are DMAGs with time series structure that are time ordered and have repeating orientations as well as past-repeating adjacencies. Lemmas 3.7 and 4.4 show that ts-DMAGs fall into this model class. The reverse, however, is not true: The graphs in part (b) of Figure 3 and parts (a) and (d) of Figure 8 fall into the model class used by tsFCI but are not ts-DMAGs.

The second previously used model class, employed by the SVAR-FCI algorithm from Malinsky and Spirtes (2018) and LPCMCI from Gerhardus and Runge (2020), are DMAGs with time series structure that are time ordered and have repeating edges. From Lemma 3.7 and Definition 4.6, we see that each ts-DMAG $\mathcal{M}^p(\mathcal{D})$ is associated to a graph in this model class, namely to the stationarified ts-DMAG $\mathcal{M}^p_{\mathrm{st}}(\mathcal{D}) = \mathrm{stat}(\mathcal{M}^p(\mathcal{D}))$. Lemma 4.20 further implies that the mapping $\iota : \mathcal{M}^p(\mathcal{D}) \mapsto \mathcal{M}^p_{\mathrm{st}}(\mathcal{D})$ is injective. Conversely, not all graphs in the model class used by SVAR-FCI and LPCMCI are stationarified ts-DMAGs: The graph in part (d) of Figure 8 is an example.

5. Markov equivalence classes of ts-DMAGs and causal discovery. This section discusses the implications of the concepts and results of Section 4 for causal discovery. To this end, Definition 5.7 in Section 5.4 introduces *time series DPAGs (ts-DPAGs)* as graphs that represent Markov equivalence classes of ts-DMAGs. Time series DPAGs are refinements of DPAGs obtained by incorporating our background knowledge about the data generating process—namely that the data are generated by a process as in equation (1) and that the observed time steps are regularly (sub)sampled. We further introduce several alternative refinements of DPAGs (see Sections 5.1 and 5.2), concretely DPAGs which represent Markov equivalence classes of stationarified ts-DMAGs and DPAGs which incorporate only some of the necessary properties of ts-DMAGs as background knowledge. As we show, these alternative DPAGs carry less information about the underlying ts-DAG than ts-DPAGs do. Using the introduced terminology, in Section 5.5 we discuss and compare three algorithms for independence-based causal discovery in time series with latent confounders and show that none of them learns ts-DPAGs. That is, all of these algorithms are conceptually suboptimal as they fail to learn causal properties of the underlying ts-DAG that in principle can be learned. As opposed to that, Algorithm 1 in Section 5.6 does learn ts-DPAGs and in this sense is *complete*. Another important result is Theorem 3 in Section 5.3, according to which DPAGs based on stationarified DMAGs carry less causal information than DPAGs based on nonstationarified DMAGs. Theorem 3 corrects an erroneous claim that has appeared in the literature;

see the explanation below Theorem 3 in Section 5.3 and the discussion of the SVAR-FCI algorithm in Section 5.5 for more details.

5.1. *Background knowledge and DPAGs*. Markov equivalent DMAGs by definition have the same m-separations, and thus cannot be distinguished by statistical independencies. They might, however, be distinguished if additional assumptions are made. One type of such assumptions is *background knowledge*, that is, the assertion that DMAGs with certain properties can be excluded as these are in conflict with a priori knowledge about the system of study.

DEFINITION 5.1 (Background knowledge, cf. Mooij and Claassen (2020)). A *background knowledge* $\mathcal{A}$ is a Boolean function on the set of all DMAGs. If $\mathcal{A}(\mathcal{M}) = 1$, then $\mathcal{M}$ is said to be *consistent* with $\mathcal{A}$, else it is said to be *inconsistent* with $\mathcal{A}$.

Combining Definition 2 in Mooij and Claassen (2020) with the definition of PAGs in Andrews, Spirtes and Cooper (2020), we refine DPAGs by background knowledge as follows.

DEFINITION 5.2 (DPAGs refined by background knowledge). Let $\mathcal{M}$ be a DMAG, let $[\mathcal{M}]$ be its Markov equivalence class, and for a background knowledge $\mathcal{A}$ let $[\mathcal{M}]_\mathcal{A}$ be the subset of $[\mathcal{M}]$ that is consistent with $\mathcal{A}$, that is, $[\mathcal{M}]_\mathcal{A} = \{\mathcal{M} \in [\mathcal{M}] | \mathcal{A}(\mathcal{M}) = 1\}$. Then:

1. A directed partial mixed graph $\mathcal{P}$ is a *DPAG for* $\mathcal{M}$ if:
   - $\mathcal{P}$ has the same skeleton (i.e., the same set of adjacencies) as $\mathcal{M}$ and
   - every noncircle mark in $\mathcal{P}$ is also in $\mathcal{M}$.
2. A DPAG $\mathcal{P}$ for $\mathcal{M}$ is called *maximally informative (m.i.) with respect to* $[\mathcal{M}]' \subseteq [\mathcal{M}]$ if:
   - every noncircle mark in $\mathcal{P}$ is in every element of $[\mathcal{M}]'$ and
   - for every circle mark in $\mathcal{P}$ there are $\mathcal{M}_1, \mathcal{M}_2 \in [\mathcal{M}]'$ such that in $\mathcal{M}_1$ there is a tail mark and in $\mathcal{M}_2 \in [\mathcal{M}]'$ there is a head mark instead of the circle mark.
3. The *maximally informative (m.i.) DPAG with respect to* $\mathcal{A}$, denoted as $\mathcal{P}(\mathcal{M}, \mathcal{A})$, is the m.i. DPAG of $\mathcal{M}$ with respect to $[\mathcal{M}]_\mathcal{A}$.
4. The *conventional m.i. DPAG for* $\mathcal{M}$ is the m.i. DPAG $\mathcal{P}(\mathcal{M}) = \mathcal{P}(\mathcal{M}, \mathcal{A}_\varnothing)$, where $\mathcal{A}_\varnothing$ is the "empty" background knowledge for which $\mathcal{A}_\varnothing = 1$ constant.

To compare different background knowledges and the accordingly refined DPAGs, we employ the following terminology.

DEFINITION 5.3 (Stronger/weaker background knowledge, more/less informative DPAG). Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be background knowledges, and let $\mathcal{P}_1$ and $\mathcal{P}_2$ be DPAGs for $\mathcal{M}$. We say:

- $\mathcal{A}_1$ is *stronger* than $\mathcal{A}_2$ and $\mathcal{A}_2$ is *weaker* than $\mathcal{A}_1$ if $\mathcal{A}_1(\mathcal{M}) = 1$ implies $\mathcal{A}_2(\mathcal{M}) = 1$.
- $\mathcal{P}_1$ is *more informative* than $\mathcal{P}_2$ and $\mathcal{P}_2$ is *less informative* than $\mathcal{P}_1$ if every circle mark in $\mathcal{P}_1$ is also in $\mathcal{P}_2$.

It follows that $\mathcal{P}(\mathcal{M}, \mathcal{A}_1)$ is more informative than $\mathcal{P}(\mathcal{M}, \mathcal{A}_2)$ if $\mathcal{A}_1$ is stronger than $\mathcal{A}_2$. By construction, $\mathcal{P}(\mathcal{M}, \mathcal{A})$ is the most informative DPAG for $\mathcal{M}$ that can be learned from statistical independencies together with the background knowledge $\mathcal{A}$.

5.2. *Considered background knowledge.* In the below discussions, we are interested in the following background knowledge.

DEFINITION 5.4 (Specific background knowledge). The *background knowledge of*:

- ... *an underlying ts-DAG* is the background knowledge $\mathcal{A}_{\mathcal{D}}$ for which $\mathcal{A}_{\mathcal{D}}(\mathcal{M}) = 1$ if and only if $\mathcal{M}$ is a ts-DMAG, that is, $\mathcal{A}_{\mathcal{D}}(\mathcal{M}) = 1$ if and only if there is a ts-DAG $\mathcal{D}$ with $\mathcal{M} = \mathcal{M}^p(\mathcal{D})$.
- ... *an underlying ts-DAG for stationarifications* is the background knowledge $\mathcal{A}_{\mathcal{D}}^{\text{stat}}$ for which $\mathcal{A}_{\mathcal{D}}^{\text{stat}}(\mathcal{M}) = 1$ if and only if $\mathcal{M}$ is a stationarified ts-DMAG, that is, $\mathcal{A}_{\mathcal{D}}^{\text{stat}}(\mathcal{M}) = 1$ if and only if there is a ts-DAG $\mathcal{D}$ with $\mathcal{M} = \mathcal{M}_{\text{st}}^p(\mathcal{D})$.
- ... *time order and repeating ancestral relationships* is the background knowledge $\mathcal{A}_{\text{ta}}$ for which $\mathcal{A}_{\text{ta}}(\mathcal{M}) = 1$ if and only if $\mathcal{M}$ is time ordered and has repeating ancestral relationships.
- ... *time order and repeating orientations* is the background knowledge $\mathcal{A}_{\text{to}}$ for which $\mathcal{A}_{\text{to}}(\mathcal{M}) = 1$ if and only if $\mathcal{M}$ is time ordered and has repeating orientations.

The first background knowledge $\mathcal{A}_{\mathcal{D}}$ is as much background knowledge as is available in the time series setting defined in Section 3.1. In Section 5.4, we will use $\mathcal{A}_{\mathcal{D}}$ to define ts-DPAGs. The second background knowledge $\mathcal{A}_{\mathcal{D}}^{\text{stat}}$ is the equivalent background knowledge when working with stationarified ts-DMAGs $\mathcal{M}_{\text{st}}^p(\mathcal{D})$ instead of ts-DMAGs $\mathcal{M}^p(\mathcal{D})$. We will use $\mathcal{A}_{\mathcal{D}}^{\text{stat}}$ to compare causal discovery based on $\mathcal{M}^p(\mathcal{D})$ with causal discovery based on $\mathcal{M}_{\text{st}}^p(\mathcal{D})$. Given that a ts-DMAG $\mathcal{M}^p(\mathcal{D})$ and its stationarification $\mathcal{M}_{\text{st}}^p(\mathcal{D})$ are in one-to-one correspondence (see Section 4.7), one might also expect the corresponding DPAGs to carry the same information. Interestingly, as we will show in Section 5.3, this expectation is incorrect. The third and fourth background knowledge $\mathcal{A}_{\text{ta}}$ and $\mathcal{A}_{\text{to}}$ equally apply to both standard and stationarified ts-DMAGs. They are included for comparison with existing causal discovery algorithms.

The four specified background knowledges compare as follows: Since both ts-DMAGs $\mathcal{M}^p(\mathcal{D})$ and stationarified ts-DMAGs $\mathcal{M}_{\text{st}}^p(\mathcal{D})$ are time ordered and have repeating ancestral relationships, both $\mathcal{A}_{\mathcal{D}}$ and $\mathcal{A}_{\mathcal{D}}^{\text{stat}}$ are stronger than $\mathcal{A}_{\text{ta}}$. Since repeating ancestral relationships imply repeating orientations, $\mathcal{A}_{\text{ta}}$ is stronger than $\mathcal{A}_{\text{to}}$. For stationarified ts-DMAGs $\mathcal{M}_{\text{st}}^p(\mathcal{D})$, however, $\mathcal{A}_{\text{ta}}$ and $\mathcal{A}_{\text{to}}$ are equivalent (as follows from Lemma 4.3). In our notation, this equivalence is expressed as $\mathcal{P}(\mathcal{M}_{\text{st}}^p(\mathcal{D}), \mathcal{A}_{\text{ta}}) = \mathcal{P}(\mathcal{M}_{\text{st}}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$.

5.3. *DPAGs of ts-DMAGs $\mathcal{M}^p(\mathcal{D})$ carry more information than DPAGs of stationarified ts-DMAGs $\mathcal{M}_{\text{st}}^p(\mathcal{D})$.* In this subsection, we show that when working with the background knowledges specified in Definition 5.4, DPAGs of ts-DMAGs can never carry less but may carry more information about the underlying ts-DAG than DPAGs of stationarified ts-DMAGs. This is so despite the fact that, as explained in Section 4.7, a ts-DMAG and its stationarification are in one-to-one correspondence. Toward proving the claim, we first note the following.

LEMMA 5.5. *Let $\mathcal{D}$ be a ts-DAG and let $\mathcal{A} \in \{\mathcal{A}_{\mathcal{D}}, \mathcal{A}_{\text{ta}}, \mathcal{A}_{\text{to}}\}$. Then the graph* $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$ *is a DPAG for $\mathcal{M}_{\text{st}}^p(\mathcal{D})$.*

In particular, both DPAGs $\mathcal{P}(\mathcal{M}_{\text{st}}^p(\mathcal{D}), \mathcal{A}^{\text{stat}})$ and $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$ have the same adjacencies. Moreover, it is well-defined to ask whether one of the two DPAGs is more informative than the other. The following result answers this question.

THEOREM 3. *Let $\mathcal{D}$ be a ts-DAG and let $(\mathcal{A}, \mathcal{A}^{\text{stat}})$ either be $(\mathcal{A}_{\text{to}}, \mathcal{A}_{\text{to}})$ or $(\mathcal{A}_{\text{ta}}, \mathcal{A}_{\text{ta}})$ or $(\mathcal{A}_{\mathcal{D}}, \mathcal{A}_{\mathcal{D}}^{\text{stat}})$. Then*:

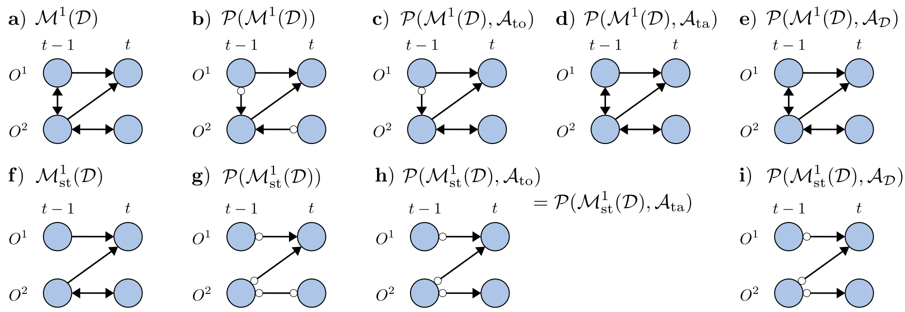

FIG. 9. *An example for illustrating Theorem 3; see also the discussion in Example 5.6.*

1. *Every noncircle mark (head or tail) in* $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}^{\text{stat}})$ *is also in* $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$.
2. *Every noncircle mark in* $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}^{\text{stat}})$ *is also in* $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A})$.
3. *There are cases in which a noncircle mark that is in* $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$ *is not also in* $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}^{\text{stat}})$.
4. *There are cases in which a noncircle mark that is in* $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A})$ *is not also in* $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}^{\text{stat}})$, *even regarding adjacencies that are shared by both graphs.*

Theorem 3 contradicts the opposite claim in Malinsky and Spirtes (2018) according to which more unambiguous edge orientations (heads or tails) may be inferred if, as licensed by the assumption of causal stationarity, the property of repeating adjacencies is enforced in causal discovery; see Section 5.5 for more details. The following example illustrates Theorem 3.

EXAMPLE 5.6. Parts (a) and (b) of Figure 9 respectively show a ts-DMAG $\mathcal{M}^p(\mathcal{D})$ and its conventional m.i. DPAG $\mathcal{P}(\mathcal{M}^p(\mathcal{D}))$. To derive $\mathcal{P}(\mathcal{M}^p(\mathcal{D}))$ one may, for example, apply the FCI orientation rules (see Zhang (2008b)) to the skeleton of $\mathcal{M}^p(\mathcal{D})$. Part (c) of the same figure shows $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$, where the head at O^2_t on $O^2_{t-1} {\leftrightarrow} O^2_t$ follows by time order. Repeating orientations does not help in orienting the last remaining circle mark on $O^1_{t-1} {\circ\!\!\rightarrow} O^2_{t-1}$ because O^1_t and O^2_t are nonadjacent. The stronger background knowledge $\mathcal{A}_{\text{ta}}$ is, however, sufficient to do so: Vertex O^1_{t-1} cannot be an ancestor of O^2_{t-1} because O^1_t is not an ancestor of O^2_t, which in turn follows because there is no possibly directed path from O^1_t to O^2_t; see Zhang ((2006), p. 81f). We hence get the DPAG $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}})$ shown in part (d). Since there are no circle marks left, $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}})$ here equals the DPAG $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\mathcal{D}})$ in part (e).[2] The graphs $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}}))$, $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}}))$ and $\text{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\mathcal{D}}))$ are respectively obtained by removing the edge between O^1_{t-1} and O^2_{t-1} from the graphs in parts (c), (d) and (e). The stationarified ts-DMAG $\mathcal{M}^p_{\text{st}}(\mathcal{D})$ and its conventional m.i. DPAG $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}))$ are shown in parts (f) and (g). Part (h) shows $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\text{to}})$, where there is a head mark at O^2_t on $O^2_{t-1} {\circ\!\!\rightarrow} O^2_t$ due to time order. As explained in Section 5.2, the equality $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\text{to}}) = \mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\text{ta}})$ always holds. With the characterization of stationarified ts-DMAGs in Lemma 4.21 (or Lemma 4.22) we can further show that in this example the graph in (h) equals $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\mathcal{D}})$ in part (i). Note that the ts-DMAG $\mathcal{M}^1(\mathcal{D})$ in part (a) is indeed a ts-DMAG. For example, its canonical ts-DAG $\mathcal{D}_c(\mathcal{M}^1(\mathcal{D}))$ projects to $\mathcal{M}^1(\mathcal{D})$.

[2] In general, the DPAGs $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\mathcal{D}})$ and $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}})$ are not equal and can contain circle marks.

Theorem 3 and Example 5.6 show that $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A})$ and $\mathrm{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$ have more unambiguous edge marks than $\mathcal{P}(\mathcal{M}^p_{\mathrm{st}}(\mathcal{D}), \mathcal{A}^{\mathrm{stat}})$. It thus is *conceptually* advantageous to work with DPAGs $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A})$ of ts-DMAGs—or with their stationarifications $\mathrm{stat}(\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}))$, if one prefers graphs with repeating edges—rather than with DPAGs $\mathcal{P}(\mathcal{M}^p_{\mathrm{st}}(\mathcal{D}), \mathcal{A}^{\mathrm{stat}})$ of stationarified ts-DMAGs. One might argue, though, that the additional ambiguous orientations (i.e., circle marks) which $\mathcal{P}(\mathcal{M}^p_{\mathrm{st}}(\mathcal{D}), \mathcal{A}^{\mathrm{stat}})$ has as compared to $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A})$ might turn into unambiguous orientations (i.e., head or tail marks) in $\mathcal{P}(\mathcal{M}^{\tilde{p}}_{\mathrm{st}}(\mathcal{D}), \mathcal{A}^{\mathrm{stat}})$ for an increased length $\tilde{p} > p$ of the observed time window.[3] However, increasing p to $\tilde{p}$ also increases the number of observed vertices, and thus yields a higher-dimensional causal discovery problem. Having more observed vertices typically hurts finite-sample performance of causal discovery; see, for example, the simulation studies in Gerhardus and Runge (2020). On the other hand, algorithms that work with stationarified ts-DMAGs rather than ts-DMAGs may scale more favorably with the length p of the observed time window because they remove the edges $O^i_{t-\Delta t-\tau} {*\!\!-\!\!*} O^j_{t-\Delta t}$ for all Δt as soon as the edge $O^i_{t-\tau} {*\!\!-\!\!*} O^j_t$ is removed and, therefore, typically make fewer independence tests. From a *practical* perspective, there thus is a trade-off between working with ts-DMAGs versus working with stationarified ts-DMAGs, which calls for empirical evaluation in future work.

5.4. *Time series DPAGs*. In Section 5.3, we showed that DPAGs of ts-DMAGs always carry more information about the underlying ts-DAG than DPAGs of stationarified ts-DMAGs. Because of this fact, we choose to define *time series DPAGs* as the former type of DPAGs.

DEFINITION 5.7 (Time series DPAG). Let $\mathcal{D}$ be a ts-DAG with variable index set $\mathbf{I}$, let $\mathbf{I_O} \subseteq \mathbf{I}$ and let $\mathbf{T_O} \subsetneq \mathbb{Z}$ be regularly sampled or regularly subsampled. The *time series DPAG implied by $\mathcal{D}$ over $\mathbf{O} = \mathbf{I_O} \times \mathbf{T_O}$*, denoted as $\mathcal{P}_{\mathbf{O}}(\mathcal{D})$ or $\mathcal{P}_{\mathbf{I_O}\times\mathbf{T_O}}(\mathcal{D})$ and also referred to as a *ts-DPAG*, is the m.i. DPAG $\mathcal{P}(\mathcal{M}_{\mathbf{O}}(\mathcal{D}), \mathcal{A}_{\mathcal{D}})$.

REMARK (on Definition 5.7). The equivalence of regular sampling and regular subsampling (see Section 4.2) carries over to ts-DPAGs. We hence restrict to regular sampling without loss of generality and write $\mathcal{P}^p(\mathcal{D})$ for $\mathcal{P}_{\mathbf{O}}(\mathcal{D})$ with $\mathbf{O} = \mathbf{I_O} \times \mathbf{T_O}$ and $\mathbf{T_O} = \{t - \tau | 0 \leq \tau \leq p\}$.

The following example discusses a case in which the use of the strongest background knowledge $\mathcal{A}_{\mathcal{D}}$ leads to strictly more unambiguous edge orientations than $\mathcal{A}_{\mathrm{ta}}$. We thus cannot replace $\mathcal{A}_{\mathcal{D}}$ with $\mathcal{A}_{\mathrm{ta}}$ in the definition of ts-DPAGs without losing information.

EXAMPLE 5.8. The ts-DMAG $\mathcal{M}^1(\mathcal{D})$ in part (a) of Figure 10 gives rise to $\mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\mathrm{ta}})$ in part (b) with a circle mark at O^1_t on $O^1_{t-1} {\circ\!\!\rightarrow} O^1_t$. According to the stronger background knowledge $\mathcal{A}_{\mathcal{D}}$, one can orient this edge as $O^1_{t-1} \rightarrow O^1_t$ because the opposite hypothesis gives the graph in part (d) in Figure 8, which by means of Theorem 1 was shown to not be a ts-DMAG; see Example 4.19. Thus, from the ts-DPAG $\mathcal{P}^1(\mathcal{D})$ we can conclude that O^1_{t-1} has a causal influence on O^1_t whereas from $\mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\mathrm{ta}})$ we can only conclude that this causal influence might but also might not exist.

Furthermore (see Gerhardus (2023), Lemma B.8), in the ts-DMAG $\mathcal{M}^1(\mathcal{D})$ the pair (O^1_{t-1}, O^1_t) cannot suffer from latent confounding.[4] Thus, the causal effect of O^1_{t-1} on O^1_t

[3]There are examples with this property, but it is unknown to the author whether this property is a general fact.

[4]Interestingly, we can draw this conclusion although $O^1_{t-1} \rightarrow O^1_t$ is not *visible*, thereby suggesting that the notion of visibility from Zhang (2008a) needs refinement for ts-DMAGs and ts-DPAGs.

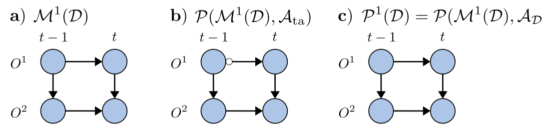

FIG. 10. *A case in which $\mathcal{P}^p(\mathcal{D})$ has a noncircle mark that is not in $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}})$.*

is identifiable and can be estimated from observations by adjusting for the empty set. Importantly, if we interpret $\mathcal{M}^1(\mathcal{D})$ not as a ts-DMAG but as a "standard" DMAG, then the causal effect of O^1_{t-1} on O^1_t would be unidentifiable as follows from Lemma 10 in Zhang (2008a).

Example 5.8 clearly demonstrates the importance of our characterization of ts-DMAGs for the tasks of causal discovery and causal inference. Moreover, the following result shows that ts-DPAGs are complete with respect to ancestral relationships.

LEMMA 5.9. *If in a ts-DPAG $\mathcal{P}^p(\mathcal{D})$ there is an edge $(i, t_i){\circ\!\!-\!\!*}\,(j, t_j)$, then there are ts-DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ such that both ts-DMAGs $\mathcal{M}^p(\mathcal{D}_1)$ and $\mathcal{M}^p(\mathcal{D}_2)$ are Markov equivalent to the ts-DMAG $\mathcal{M}^p(\mathcal{D})$ and $(i, t_i) \in \text{an}((j, t_j), \mathcal{D}_1)$ and $(i, t_i) \notin \text{an}((j, t_j), \mathcal{D}_2)$.*

5.5. *Existing causal discovery algorithms do not learn ts-DPAGs.* To the best of the authors' knowledge, so far there is no causal discovery algorithm that learns ts-DPAGs $\mathcal{P}^p(\mathcal{D})$. Hence, all existing causal discovery algorithms fail to learn some causal relationships that can be learned. This failure also applies to the independence-based algorithms tsFCI from Entner and Hoyer (2010), SVAR-FCI from Malinsky and Spirtes (2018) and LPCMCI from Gerhardus and Runge (2020).[5] Below, we discuss and compare these three algorithms *conceptually* (but also note the *practical* considerations discussed at the end of Section 5.3).

The *tsFCI* algorithm from Entner and Hoyer (2010) refines the well-known FCI algorithm (see Spirtes, Meek and Richardson (1995), Spirtes, Glymour and Scheines (2000), Zhang (2008b)), to structural processes as in equation (1). To this end, see the blue colored instructions in parts 2(a) and 2(b) of Algorithm 1 in Entner and Hoyer (2010), tsFCI imposes time order from the start and enforces repeating orientations at all steps. In addition, see the blue colored instructions in parts 1(b) and 1(c) of Algorithm 1 in Entner and Hoyer (2010), tsFCI excludes future vertices from conditioning sets and uses repeating separating sets to avoid unnecessary independence tests (these latter two modifications are, however, only relevant computationally and statistically but not conceptually). Importantly, Entner and Hoyer (2010) introduces two variants of the algorithm. The first variant, which we call *tsFCI^l* (with "l" for "lagged"), assumes that in the data-generating ts-DAG there are no contemporaneous edges, and hence orients all contemporaneous edges in the DPAG as bidirected. This first variant is as specified by Algorithm 1 in Entner and Hoyer (2010). However, in Section 6 of Entner and Hoyer (2010) (see, in particular, their footnote 3) the authors explain the minor modifications that have to be done when not making the additional assumption of no contemporaneous causation. Moreover, there they also show an application of the resulting more general variant. We refer to this second variant as *tsFCI^{l+c}* (with "$l+c$" for "lagged plus contemporaneous"). To summarize, in our terminology tsFCI^{l+c} attempts to learn the DPAG $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$ of the ts-DMAG $\mathcal{M}^p(\mathcal{D})$. Since $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$ may contain circle marks that are not in the ts-DPAG $\mathcal{P}^p(\mathcal{D}) = \mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\mathcal{D}})$ (see Examples 5.6 and 5.8), tsFCI^{l+c} does not learn all ancestral relationships that can be learned when using the available background knowledge

[5]The same is true for the score-based and hybrid algorithms from Gao and Tian (2010) and Malinsky and Spirtes (2018).

$\mathcal{A}_{\mathcal{D}}$. From Example 5.6, we even conclude that tsFCI^{l+c} learns fewer orientations as can be learned with the weaker background knowledge $\mathcal{A}_{\text{ta}}$.

As compared to tsFCI^{l+c}, the more recent *SVAR-FCI* algorithm from Malinsky and Spirtes (2018) enforces repeating adjacencies by removing the edges $O^i_{t-\Delta t-\tau} {*\!\!-\!\!*} O^j_{t-\Delta t}$ for all Δt as soon as the edge $O^i_{t-\tau} {*\!\!-\!\!*} O^j_t$ is removed—even in cases where there is no associated separating set in the observed time window.[6] This modification is achieved by the respective second lines in the "then" clauses in steps 5 and 11 of Algorithm 3.1 in Malinsky and Spirtes (2018). Consequently, SVAR-FCI finds a skeleton which has repeating adjacencies, that is, SVAR-FCI finds the skeleton of the stationarified ts-DMAG $\mathcal{M}^p_{\text{st}}(\mathcal{D})$ rather than the skeleton of the ts-DMAG $\mathcal{M}^p(\mathcal{D})$. On the skeleton of $\mathcal{M}^p_{\text{st}}(\mathcal{D})$, the algorithm then applies the FCI orientation rules, augmented with the background knowledge of time order and repeating orientations. In our terminology, SVAR-FCI hence attempts to learn the DPAG $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\text{to}})$ of the stationarified ts-DMAG $\mathcal{M}^p_{\text{st}}(\mathcal{D})$. Now recall Theorem 3, which says that all unambiguous edge orientations in this DPAG $\mathcal{P}(\mathcal{M}^p_{\text{st}}(\mathcal{D}), \mathcal{A}_{\text{to}})$ are also in $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$—the one learned by tsFCI^{l+c}—while there are cases in which the opposite is not true. Thus, if ground-truth knowledge of (conditional) independencies is given, SVAR-FCI can never learn more unambiguous edge orientations than tsFCI^{l+c} while there are cases in which it learns strictly fewer. The additional edge removals thus actually have the opposite effect of what was intended in Malinsky and Spirtes (2018). Moreover, also SVAR-FCI fails to learn all identifiable ancestral relationships of the underlying ts-DAG.

EXAMPLE 5.10. Assume ground-truth knowledge about (conditional) independencies. When applied to the ts-DMAG in part (a) of Figure 9, tsFCI^{l+c} returns the graph in part (c) whereas SVAR-FCI returns the graph in part (h) with strictly fewer unambiguous edge marks. This difference is relevant: From tsFCI^{l+c}'s output, we can conclude that O^2_{t-1} generically has a causal effect on O^1_t whereas from SVAR-FCI's output we can only conclude that O^2_{t-1} might but also might not generically have a causal effect on O^1_t. As another difference, tsFCI^{l+c} gives the edge $O^1_{t-1} {\circ\!\!\rightarrow} O^2_{t-1}$ whereas SVAR-FCI gives that O^1_{t-1} and O^2_{t-1} are nonadjacent. At first, one might think that SVAR-FCI is at the advantage in this regard, because the absence of an edge between O^1_{t-1} and O^2_{t-1} correctly conveys that the pair (O^1_{t-1}, O^2_{t-1}) is not confounded by unobservable variables. Note, however, that tsFCI^{l+c} conveys the same conclusion by means of having learned that O^1_t and O^2_t are nonadjacent (cf. last paragraph in Section 4.4). In fact, see Theorem 3, one can always post-process the output of tsFCI^{l+c} by stationarification stat$(\cdot)$ to obtain a graph that, compared to the graph learned by SVAR-FCI, has the same adjacencies and the same or more unambiguous edge orientations. In the current example, this post-processing step amounts to removing the edge $O^1_{t-1} {\circ\!\!\rightarrow} O^2_{t-1}$ from the graph in part (c) of Figure 9.

The *LPCMCI* algorithm from Gerhardus and Runge (2020) applies several modifications to SVAR-FCI that significantly improve the finite-sample performance. The infinite sample properties are unchanged, however. Thus, also LPCMCI in general learns fewer orientations than tsFCI^{l+c} and fails to learn all ancestral relationships that can be learned.

5.6. *ts-DPAGs can be learned from data.* In this subsection, we show that ts-DPAGs can, at least in principle, be learned from data. In fact, using the characterization of ts-DMAGs by Theorem 1, we can immediately write down Algorithm 1 for this purpose.

[6]To clarify, Malinsky and Spirtes (2018) do not mention tsFCI^{l+c} but refer to tsFCI^l when writing "tsFCI."

Algorithm 1 Learning ts-DPAGs

1: Apply any causal discovery algorithm on the time window $[t-p, t]$ that determines a DPAG $\mathcal{P}^*$ for $\mathcal{M}^p(\mathcal{D})$ which is at least as informative as the conventional m.i. DPAG $\mathcal{P}(\mathcal{M}^p(\mathcal{D}))$. The tsFCI^{l+c} algorithm meets this requirement, whereas SVAR-FCI and LPCMCI do not meet this requirement.

2: Let $\mathbf{M}^*$ be the set of all DMAGs that are represented by and are Markov equivalent to the DPAG $\mathcal{P}^*$. This step can, for example, be done with the Zhang MAG listing algorithm described in Malinsky and Spirtes (2016).

3: Let $\mathbf{M}$ be the set of all DMAGs $\mathcal{M} \in \mathbf{M}^*$ for which $\mathcal{M} = \mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$. This step can be executed as follows: For each $\mathcal{M} \in \mathbf{M}^*$, first construct the canonical ts-DAG $\mathcal{D}_c(\mathcal{M})$ by applying Definition 4.13 and, second, apply the MAG latent projection to determine the ts-DMAG $\mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$ and, third, check for equality of the graphs $\mathcal{M}$ and $\mathcal{M}^p(\mathcal{D}_c(\mathcal{M}))$.

4: Let $\mathcal{P}$ be the m.i. DPAG with respect to the set $\mathbf{M}$. Note that parts 1 and 2 of Definition 5.2 specify how $\mathcal{P}$ is determined from $\mathbf{M}$.

5: **return** ts-DPAG $\mathcal{P} = \mathcal{P}^p(\mathcal{D})$

Practically, however, finding the set of candidate DMAGs $\mathbf{M}^*$ in step 2 is expected to become computationally infeasible for large graphs $\mathcal{P}^*$. This expectation is based on the empirical finding in Malinsky and Spirtes (2016) according to which the Zhang MAG listing algorithm (there used not for causal discovery but for causal effect estimation and in a nontemporal setting) becomes too slow for graphs with about 15 to 20 vertices. On the contrary, when using tsFCI^{l+c} in step 1, the DPAG $\mathcal{P}^*$ already incorporates the background knowledge $\mathcal{A}_{\text{to}}$ of time order and repeating orientations. Hence, $\mathcal{P}^*$ will tend to have fewer circle marks than a typical DPAG in the nontemporal setting. Therefore, Algorithm 1 might be feasible for yet larger graphs. Nevertheless, it would be desirable to instead derive orientation rules that impose the background knowledge $\mathcal{A}_{\mathcal{D}}$ directly on $\mathcal{P}^*$, and thus entirely circumvent the need to determine the set $\mathbf{M}^*$. Moreover, recalling from the remark on Definition 3.6, an implementation of the projection procedure required for step 3 is possible but nontrivial; see Gerhardus et al. (2023). The following example illustrates Algorithm 1.

EXAMPLE 5.11. Consider the graph $\mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\text{ta}})$ in part (b) of Figure 10, which in this example equals $\mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\text{to}})$. Given ground-truth knowledge of (conditional) independencies, this graph is the output of tsFCI^{l+c} (and of SVAR-FCI and LPCMCI) on any ts-DAG $\mathcal{D}$ that projects to $\mathcal{M}^1(\mathcal{D})$ in part (a) of the figure. Such $\mathcal{D}$ exists, for example, the canonical ts-DAG $\mathcal{D}_c(\mathcal{M}^1(\mathcal{D}))$ of $\mathcal{M}^1(\mathcal{D})$. There is exactly one circle mark in $\mathcal{P}^* = \mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\text{ta}})$, namely on $O^1_{t-1} {\circ\!\!\rightarrow} O^1_t$. This circle mark can be oriented either as a tail ($O^1_{t-1} \rightarrow O^1_t$), giving rise to a DMAG $\mathcal{M}_1$, or as a head ($O^1_{t-1} \leftrightarrow O^1_t$), giving rise to a DMAG $\mathcal{M}_2$. Both of these candidates are represented by and Markov equivalent to $\mathcal{P}^*$, hence $\mathbf{M}^* = \{\mathcal{M}_1, \mathcal{M}_2\}$. Moving to step 3, the first candidate $\mathcal{M}_1$ passes the check $\mathcal{M}_1 = \mathcal{M}^1(\mathcal{D}_c(\mathcal{M}_1))$, whereas (see Example 5.8) $\mathcal{M}_2 \neq \mathcal{M}^1(\mathcal{D}_c(\mathcal{M}_2))$ for the second candidate $\mathcal{M}_2$. Thus, $\mathbf{M} = \{\mathcal{M}_1\}$. Since there is only a single element $\mathcal{M}_1$ in $\mathbf{M}$, this DMAG $\mathcal{M}_1$ according to parts 1 and 2 of Definition 5.2 equals the ts-DPAG $\mathcal{P}^1(\mathcal{D})$. Noting that $\mathcal{P}^1(\mathcal{D})$ (learned by Algorithm 1) has an additional unambiguous edge mark as compared to $\mathcal{P}(\mathcal{M}^1(\mathcal{D}), \mathcal{A}_{\text{to}})$ (learned by tsFCI^{l+c}, SVAR-FCI and LPCMCI), we see that Algorithm 1 is indeed more informative than the existing algorithms.

REMARK (on Example 5.11). The example has two nongeneric properties. First, see (Gerhardus (2023), Figure C and Example B.15), in general there can be circle marks in the ts-DPAG $\mathcal{P}^p(\mathcal{D})$. Second, the ts-DMAG $\mathcal{M}^p(\mathcal{D}) = \mathcal{M}^1(\mathcal{D})$ here has repeating edges, and thus equals $\mathcal{M}^p_{\text{st}}(\mathcal{D}) = \mathcal{M}^1_{\text{st}}(\mathcal{D})$. Only if the equality $\mathcal{M}^p(\mathcal{D}) = \mathcal{M}^p_{\text{st}}(\mathcal{D})$ holds, then also SVAR-FCI and LPCMCI learn the DPAG $\mathcal{P}^* = \mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{to}})$, which then also necessarily equals $\mathcal{P}(\mathcal{M}^p(\mathcal{D}), \mathcal{A}_{\text{ta}})$. In general, however, $\mathcal{M}^p(\mathcal{D})$ and $\mathcal{M}^p_{\text{st}}(\mathcal{D})$ are not equal and neither SVAR-FCI nor LPCMCI may be used for step 1 of Algorithm 1; see Figure 9 and Example 5.6.

6. Discussion. In this paper, we developed and analyzed ts-DMAGs, a class of graphical models for representing time-lag specific causal relationships and independencies among finitely many regularly (sub)sampled time steps of causally stationary multivariate time series with unobserved components. As a central result, Theorems 1 and 2 completely characterize ts-DMAGs. Examples demonstrated that ts-DMAGs constitute a strictly smaller class of graphical models than the graphs that have previously been employed in the literature; see Section 4.8 for details. At the same time, using ts-DMAGs does not require additional assumptions or restrictions on the data-generating process. From ts-DMAGs, one can thus draw stronger causal inferences than from the previously employed model classes, both in causal discovery and causal effect estimation. In addition, we defined ts-DPAGs as representations of Markov equivalence classes of ts-DMAGs. Time series DPAGs contain as much information about the ancestral relationships as can in principle be learned from observational data under the standard assumptions of independence-based causal discovery. We then showed that current time series causal discovery algorithms do not learn ts-DPAGs, that is, they fail to learn some causal relationships that can be learned. As opposed to that, Algorithm 1 does learn ts-DPAGs. With Theorem 3, we corrected the incorrect claim from the literature that causal discovery on stationarified DMAGs gives more unambiguous edge orientations than causal discovery on nonstationarified DMAGs—in fact, the opposite is true. We envision that these results will be used to improve time series causal inference methods that resolve time lags, which in turn can have applications in diverse scientific and technical domains.

The results presented here point to various directions of future research. First, it would be valuable to consider the causal discovery problem in more detail. In particular, it is desirable to develop orientation rules that impose the background knowledge of an underlying ts-DAG $\mathcal{A}_{\mathcal{D}}$ without the need for listing all DMAGs consistent with $\mathcal{A}_{\mathcal{D}}$. Second, it remains open to characterize the causal inferences that can be drawn from ts-DMAGs and ts-DPAGs. As shown by Example 5.8, deriving such a characterization is a nontrivial task that goes beyond the corresponding task in the nontemporal setting. Third, one may analyze the additional restrictions and causal inferences that follow when, as opposed to this work, assumptions on the connectivity pattern of the underlying ts-DAG are imposed. Lastly, it is desirable to generalize our results to cases with cyclic causal relationships and selection variables.

Acknowledgments. I thank Jakob Runge for helpful discussions and suggestions. I thank Tom Hochsprung and Wiebke Günther for careful proofreading and suggestions on how to make the paper more accessible. I thank two anonymous reviewers and two anonymous associate editors for suggestions and questions that helped me to improve the paper.

SUPPLEMENTARY MATERIAL

Supplement to a Characterization of causal ancestral graphs for time series with latent confounders (DOI: 10.1214/23-AOS2325SUPP; .pdf). This Supplementary Material contains: First, a glossary of abbreviations and frequently used symbols. Second, theoretical results that were omitted from the main text due to space constraints. Third, proofs of all theoretical results presented in the main text together with various auxiliary results that are used in these proofs.

REFERENCES

ALI, R. A., RICHARDSON, T. S. and SPIRTES, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

ANDREWS, B., SPIRTES, P. and COOPER, G. F. (2020). On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 4002–4011. PMLR.

ASSAAD, C. K., DEVIJVER, E. and GAUSSIER, E. (2022). Discovery of extended summary graphs in time series. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 96–106. PMLR.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0996025 https://doi.org/10.1002/9781118619179

BONGERS, S., BLOM, T. and MOOIJ, J. M. (2018). Causal modeling of dynamical systems. Preprint. Available at arXiv:1803.08784.

CHU, T. and GLYMOUR, C. (2008). Search for additive nonlinear time series causal models. *J. Mach. Learn. Res.* **9** 967–991. MR2417260

DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x

EICHLER, M. (2010). Graphical Gaussian modelling of multivariate time series with latent variables. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterington, eds.). *Proceedings of Machine Learning Research* **9** 193–200. PMLR.

EICHLER, M. and DIDELEZ, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* (*UAI*2007) (R. Parr and L. van der Gaag, eds.) 109–116. AUAI Press.

EICHLER, M. and DIDELEZ, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3

ENTNER, D. and HOYER, P. O. (2010). On causal discovery from time series data using FCI. In *Proceedings of the 5th European Workshop on Probabilistic Graphical Models* (P. Myllymäki, T. Roos and T. Jaakkola, eds.) 121–128. Helsinki Institute for Information Technology HIIT, Helsinki, Finland.

GAO, W. and TIAN, Z. (2010). Latent ancestral graph of structure vector autoregressive models. *J. Syst. Eng. Electron.* **21** 233–238.

GEIGER, D., VERMA, T. and PEARL, J. (1990). Identifying independence in Bayesian networks *Networks* **20** 507–534. MR1064736 https://doi.org/10.1002/net.3230200504

GERHARDUS, A. (2024). Supplement to "Characterization of causal ancestral graphs for time series with latent confounders." https://doi.org/10.1214/23-AOS2325SUPP

GERHARDUS, A. and RUNGE, J. (2020). High-recall causal discovery for autocorrelated time series with latent confounders. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 12615–12625. Curran Associates, Red Hook, NY.

GERHARDUS, A., WAHL, J., FALTENBACHER, S., NINAD, U. and RUNGE, J. (2023). Projecting infinite time series graphs to finite marginal graphs using number theory. Preprint. Available at arXiv:2310.05526.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

HUCKINS, J. F., DASILVA, A. W., HEDLUND, E. L., MURPHY, E. I., ROGERS, C., WANG, W., OBUCHI, M., HOLTZHEIMER, P. E., WAGNER, D. D. et al. (2020). Causal factors of anxiety and depression in college students: Longitudinal ecological momentary assessment and causal analysis using Peter and Clark momentary conditional independence. *JMIR Mental Health* **7** e16684.

HYVÄRINEN, A., ZHANG, K., SHIMIZU, S. and HOYER, P. O. (2010). Estimation of a structural vector autoregression model using non-Gaussianity. *J. Mach. Learn. Res.* **11** 1709–1731. MR2653353

KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models*: *Principles and Techniques*. MIT Press, Cambridge, MA.

KRETSCHMER, M., COUMOU, D., DONGES, J. F. and RUNGE, J. (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *J. Climate* **29** 4069–4081.

MAATHUIS, M. H. and COLOMBO, D. (2015). A generalized back-door criterion. *Ann. Statist.* **43** 1060–1088. MR3346697 https://doi.org/10.1214/14-AOS1295

MALINSKY, D. and SPIRTES, P. (2016). Estimating causal effects with ancestral graph Markov models. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models* (A. Antonucci, G. Corani and C. P. Campos, eds.). *Proceedings of Machine Learning Research* **52** 299–309. PMLR.

MALINSKY, D. and SPIRTES, P. (2018). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of* 2018 *ACM SIGKDD Workshop on Causal Discovery* (T. D. Le, K. Zhang, E. Kıcıman, A. Hyvärinen and L. Liu, eds.). *Proceedings of Machine Learning Research* **92** 23–47. PMLR.

MOGENSEN, S. W. and HANSEN, N. R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Statist.* **48** 539–559. MR4065173 https://doi.org/10.1214/19-AOS1821

MOOIJ, J. M. and CLAASSEN, T. (2020). Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence* (*UAI*) (J. Peters and D. Sontag, eds.). *Proceedings of Machine Learning Research* **124** 1159–1168. PMLR.

PAMFIL, R., SRIWATTANAWORACHAI, N., DESAI, S., PILGERSTORFER, P., GEORGATZIS, K., BEAUMONT, P. and ARAGAM, B. (2020). DYNOTEARS: Structure learning from time-series data. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 1595–1605. PMLR.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*. *The Morgan Kaufmann Series in Representation and Reasoning*. Morgan Kaufmann, San Francisco, CA. MR0965765

PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82** 669–688. MR1380809 https://doi.org/10.1093/biomet/82.4.669

PEARL, J. (2000). *Causality*: *Models*, *Reasoning*, *and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

PERKOVIĆ, E., TEXTOR, J., KALISCH, M. and MAATHUIS, M. H. (2017). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.* **18** 1–62. MR3827108

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2013). Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26**. Curran Associates, Red Hook, NY.

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference*: *Foundations and Learning Algorithms*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3822088

RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015

RUNGE, J. (2020). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence* (*UAI*) (J. Peters and D. Sontag, eds.). *Proceedings of Machine Learning Research* **124** 1388–1397. PMLR.

RUNGE, J., HEITZIG, J., PETOUKHOV, V. and KURTHS, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108** 258701. https://doi.org/10.1103/PhysRevLett.108.258701

SAETIA, S., YOSHIMURA, N. and KOIKE, Y. (2021). Constructing brain connectivity model using causal network reconstruction approach. *Front. Neuroinform.* **15** 619557. https://doi.org/10.3389/fninf.2021.619557

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (1993). *Causation*, *Prediction*, *and Search*. *Lecture Notes in Statistics* **81**. Springer, New York. MR1227558 https://doi.org/10.1007/978-1-4612-2748-9

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. (With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper, and Thomas Richardson.) MR1815675

SPIRTES, P., MEEK, C. and RICHARDSON, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI*1995) (P. Besnard and S. Hanks, eds.) 499–506. Morgan Kaufmann, San Francisco, CA.

VERMA, T. and PEARL, J. (1990). Causal Networks: Semantics and Expressiveness. In *Uncertainty in Artificial Intelligence* (R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, eds.). *Machine Intelligence and Pattern Recognition* **9** 69-76. North-Holland. https://doi.org/10.1016/B978-0-444-88650-7.50011-1

ZHANG, J. (2006). Causal inference and reasoning in causally insufficient systems, Ph.D. thesis, Department of Philosophy, Carnegie Mellon University.

ZHANG, J. (2008a). Causal reasoning with ancestral graphs. *J. Mach. Learn. Res.* **9** 1437–1474. MR2426048

ZHANG, J. (2008b). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif. Intell.* **172** 1873–1896. MR2459793 https://doi.org/10.1016/j.artint.2008.08.001