The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Spatio-Temporal Hierarchical Causal Models

Xintong Li[1], **Haoran Zhang**[1], **Xiao Zhou**[2,3,4*]
[1]School of Statistics, Renmin University of China Beijing, China
[2]Gaoling School of Artificial Intelligence, Renmin University of China Beijing, China
[3]Beijing Key Laboratory of Research on Large Models and Intelligent Governance
[4]Engineering Research Center of Next-Generation Intelligent Search and Recommendation, MOE
{lixintong, zhanghaoran1015, xiaozhou}@ruc.edu.cn

Abstract

The abundance of fine-grained spatio-temporal data, such as traffic sensor networks, offers vast opportunities for scientific discovery. However, inferring causal relationships from such observational data remains challenging, particularly due to unobserved confounders that are specific to units (e.g., geographical locations) yet influence outcomes over time. Most existing methods for spatio-temporal causal inference assume that all confounders are observed, an assumption that is often violated in practice. In this paper, we introduce *Spatio-Temporal Hierarchical Causal Models* (**ST-HCMs**), a novel graphical framework that extends hierarchical causal modeling to the spatio-temporal domain. At the core of our approach is the *Spatio-Temporal Collapse Theorem*, which shows that a complex ST-HCM converges to a simpler flat causal model as the amount of subunit data increases. This theoretical result enables a general procedure for *causal identification*, allowing ST-HCMs to recover causal effects even in the presence of unobserved, time-invariant unit-level confounders, a scenario where standard non-hierarchical models fail. We validate the effectiveness of our framework on both synthetic and real-world datasets, demonstrating its potential for robust causal inference in complex dynamic systems.

Code — https://github.com/CAMELLIAxt/ST-HCMs

Introduction

Fine-grained spatio-temporal data are becoming ubiquitous, generated by sources ranging from satellite imagery in environmental science to sensor networks in urban planning (Wang et al. 2018; Xu and Zhou 2024; Ali et al. 2024; Dong et al. 2024; Li, Li, and Zhou 2025). Such data inherently possess a hierarchical structure, offering unprecedented opportunities to uncover causal mechanisms in complex dynamic systems. However, reliable causal inference from observational data remains challenging (Hernán and Robins 2010; Reich et al. 2021), obscuring mechanistic insights and impeding accurate policy intervention predictions in complex spatio-temporal systems like urban operations (Zhang and Ning 2023; Dong, Zhang, and Zhang 2025; Xie et al. 2025a).

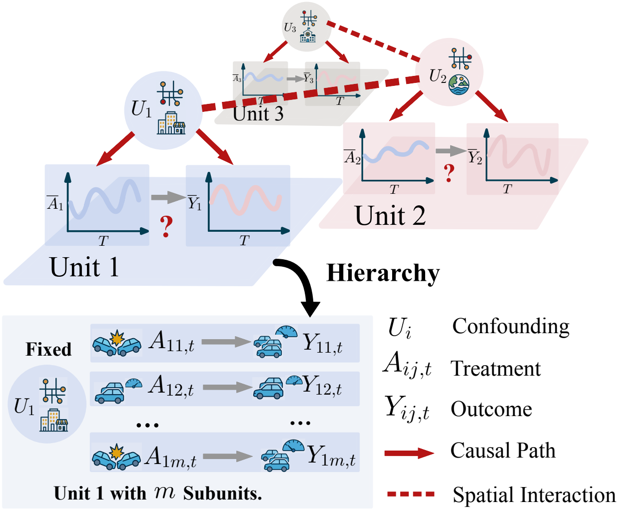


Figure 1: An example of hierarchical causality in traffic.

Consider the urban traffic system in Figure 1, where a city is divided into several regions (units), and within each region, numerous sensors monitor traffic flow (outcome) on specific road segments (subunits). A traffic accident (treatment) can disrupt the flow, but a region's unique characteristics, such as its geographical layout and road network structure, serve as an **unobserved confounder**. This confounder affects both the likelihood of accidents and typical traffic patterns, making it difficult to disentangle the true causal effect of an accident from the region's inherent characteristics (Weinstein and Blei 2024; Clarke and Polselli 2025). The problem is further compounded by temporal dynamics and spatial spillovers from neighboring regions, which introduces additional layers of complexity (Robins 1986; Aronow and Samii 2017; Reich et al. 2021).

This hierarchical structure offers a powerful, yet often underutilized, path forward. Standard analytical methods often bypass this structure, instead relying on the strong "no unmeasured confounding" assumption (Hernán and Robins 2006; Ali, Faruque, and Wang 2024; Oprescu et al. 2025), which is untenable for complex systems like the one in Figure 1. While instrumental variables (IV) provide a theoretical alternative (Martinussen et al. 2017; Cheng et al. 2024),

---

[*] Corresponding author.

finding a valid instrument that is both relevant and exogenous is notoriously difficult in practice. Methods like fixed-effect (FE) models do leverage within-unit variation (Angrist and Pischke 2009; Jensen, Burroni, and Rattigan 2020), but they typically impose restrictive parametric assumptions, such as linearity, limiting their applicability.

In contrast, our approach builds on the key insight illustrated in Figure 1: we treat each unit as a local natural experiment, enabling non-parametric estimation of region-specific causal effects by analyzing the relationship between treatments and subunit-level outcomes. This paper extends this powerful principle to spatio-temporal settings, where such effects are confounded not only by static heterogeneity but also by dynamic evolution and spatial interactions.

While the idea of leveraging hierarchical structure for causal inference has been explored in static settings (Weinstein and Blei 2024), its extension to the spatio-temporal domain presents non-trivial. The introduction of time and space creates two fundamental challenges: (1) **dynamic confounding**, where the history of the system acts as a time-varying confounder for future treatments and outcomes, and (2) **spatial confounding**, where the states of neighboring units interfere with each other's causal processes. To address these challenges, we introduce **Spatio-Temporal Hierarchical Causal Models (ST-HCMs)**, a novel graphical framework that explicitly models these dependencies. The core of our solution is a **Spatio-Temporal Collapse Theorem**, which provides a theoretical guarantee that the complex, dynamic ST-HCM mathematically converges to a simpler, equivalent representation. This theorem is the key that enables causal identification, allowing us to disentangle effects from the web of spatio-temporal confounding.

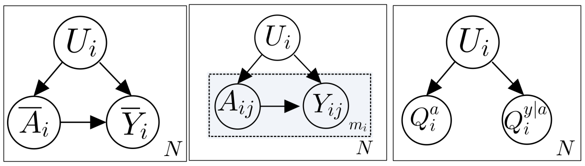
The contributions of this paper are as follows:

- **Conceptually,** we propose the first graphical framework (ST-HCMs) for causal modeling of nested spatio-temporal data, providing a formal formulation to reason about causality within spatio-temporal systems.
- **Theoretically,** we prove a Spatio-Temporal Collapse Theorem (Theorem 2) that guarantees a complex ST-HCM converges to a simpler, equivalent model, which in turn enables causal identification (Theorem 3,4) despite unobserved confounders.
- **Experimentally,** we validate our framework through extensive simulations, showing that it robustly recovers causal effects despite unobserved confounders and spatial spillovers. On a real-world dataset, we further demonstrate that it substantially reduces the biases inherent to traditional non-hierarchical models.

Preliminaries

Notation Consider a system composed of N units, indexed by $i \in \{1, \ldots, N\}$, observed over T discrete time steps, indexed by $t \in \{1, \ldots, T\}$. Each unit i contains m_i subunits, indexed by $j \in \{1, \ldots, m_i\}$. For simplicity, we assume $m_i = m$ for all i. Let V be a finite index set for endogenous variable types. This set is partitioned into subunit-level variables S and unit-level variables $U = V \setminus S$.

We define the key variables as follows:



(a) Aggregated (b) HCM (c) Collapsed

Figure 2: Hierarchy and Collapsing in HCMs.

- $X^v_{ij,t}$: The value of a subunit-level variable $v \in S$ for subunit j in unit i at time t.
- $X^w_{i,t}$: The value of a unit-level variable $w \in U$ for unit i at time t.
- $\boldsymbol{X}_{i,<t}$: The full history of all variables associated with unit i prior to time t.
- $\mathcal{N}(i)$: The set of spatial neighbors of unit i.
- $\mathcal{X}$: A generic state space. $\mathcal{P}(\mathcal{X})$ denotes the space of all probability measures on $\mathcal{X}$.
- U_i: A time-invariant, unobserved random variable representing the static latent properties of unit i.
- $\gamma^v_{i,t}, \epsilon^v_{ij,t}$: Time-varying exogenous noise variables at the unit and subunit levels, respectively.

Causal Graphical Models and HCMs A Causal Graphical Model (CGM) represents causal relationships as a Directed Acyclic Graph (DAG), where nodes are variables and directed edges denote causal effects. Interventions are denoted by the do-operator (Pearl 2009).

The static Hierarchical Causal Model (HCM) framework (Weinstein and Blei 2024) is designed to overcome confounding in nested data. As illustrated in Figure 2a, in a standard aggregated view, an unobserved confounder U_i biases the estimated effect of treatment $\bar{A}_i$ on outcome $\bar{Y}_i$. The key insight of HCM is to leverage the fine-grained subunit data within each unit, as depicted in Figure 2b. This hierarchical structure enables a powerful identification strategy known as collapsing. The procedure transforms the HCM into an equivalent, flat causal graph (Figure 2c, where the subunit-level causal mechanisms are succinctly captured by new random variables, termed Q-variables (e.g., $Q^a_i, Q^{y|a}_i$). However, the static HCM is limited to cross-sectional settings and does not account for temporal dynamics or spatial dependencies, which are essential in spatio-temporal domains. Our work extends this collapsing principle to the dynamic spatio-temporal domain.

Spatio-Temporal Hierarchical Causal Models

In this section, , we introduce the proposed Spatio-Temporal Hierarchical Causal Model (ST-HCM). To build intuition, we first extend the static HCM to the temporal setting (T-HCM), then incorporate spatial dependencies to obtain the ST-HCM and establish its convergence properties.

A Temporal Extension: T-HCMs

Definition 1 (Temporal Hierarchical Structural Causal Model (T-HSCM)). *A T-HSCM, $\mathcal{M}_{t-hscm}$, is defined by:*

(1) a summary causal graph $\mathcal{G}_T$ whose unrolling over time yields a DAG; (2) a set of static exogenous variables $U_i \sim P(U)$; (3) sets of time-varying exogenous noises $\{\gamma^v_{i,t}\}, \{\epsilon^v_{ij,t}\}$; and (4) a set of deterministic mechanism functions $\{f^v\}_{v \in V}$. The value of each endogenous variable is generated as follows:

- *For a subunit-level variable $v \in S$:*

$$x^v_{ij,t} = f^v(U_i, pa_U(x^v_{i,t}), pa_S(x^v_{ij,t}), \boldsymbol{X}_{i,<t}, \gamma^v_{i,t}, \epsilon^v_{ij,t}).$$

- *For a unit-level variable $w \in U$:*

$$x^w_{i,t} = f^w(U_i, pa_U(x^w_{i,t}), \\ \{\{x^{v'}_{ij,t}\}_{v' \in pa_S(w,t)}\}^m_{j=1}, \boldsymbol{X}_{i,<t}, \gamma^w_{i,t}),$$

  where the mechanism f^w is permutation-invariant with respect to the subunit index j.

By integrating out the exogenous noises from the structural model, we obtain the T-HCGM. This graphical model is described by stochastic mechanisms, centered on the dynamic evolution of Q-variables, which represent the conditional distributions governing subunit-level phenomena.

Definition 2 (Temporal Hierarchical Causal Graphical Model (T-HCGM))**.** *A T-HCGM, $\mathcal{M}_{t-hcgm}$, corresponding to a T-HSCM, is characterized by a two-stage generative process at each time step t:*

1. ***Q-variable Evolution:*** *A vector of Q-variables $\boldsymbol{Q}_{i,t}$ is generated for each unit i, conditional on its macro-history: $\boldsymbol{Q}_{i,t} \sim p(\boldsymbol{q}_t | U_i, \boldsymbol{S}_{i,<t})$ where $\boldsymbol{S}_{i,<t} = (X^U_{i,<t}, \boldsymbol{Q}_{i,<t})$ is the macro-history.*
2. ***Variable Generation:*** *Endogenous variables are drawn conditional on the current Q-variables and their history.*

To prove the convergence of this model to a simpler, collapsed representation, we require the following assumptions.

Assumption 1 (Mechanism Convergence)**.** *For any $w \in U$, its conditional distribution's generating mechanism is continuous with respect to the empirical measure of its subunit-level parent histories. Let $\mathcal{P}_{i,<t}$ be the true probability measure on the subunit history space, determined by the macro-history $Hist^{col}_{i,<t} = (U_i, \boldsymbol{S}_{i,<t})$. Let $\hat{\mathcal{P}}_{i,m,<t}$ be the empirical measure from m subunits. We assume:*

$$\lim_{m \to \infty} \mathbb{E}_{\hat{\mathcal{P}}_{i,m,<t}}[\mathrm{KL}(p(X^w_{i,t} | \boldsymbol{S}_{i,<t}, \mathcal{P}_{i,<t}) \\ || \, p(X^w_{i,t} | \boldsymbol{S}_{i,<t}, \hat{\mathcal{P}}_{i,m,<t}))] = 0. \tag{1}$$

This assumption formalizes the idea that with enough subunit data, the uncertainty from sampling subunit histories becomes negligible for determining unit-level outcomes .

Assumption 2 (Regularity of State Spaces)**.** *The state spaces for all variables are Polish spaces. Furthermore, the class of functions defining subunit histories forms a Glivenko-Cantelli class, ensuring the uniform convergence of empirical measures.*

This is a standard technical condition for ensuring the regularity of the underlying probability spaces (Hernán and Robins 2006; Ali, Faruque, and Wang 2024).

Theorem 1 (Convergence of T-HCMs)**.** *Under Assumptions 1 and 2, as the number of subunits $m \to \infty$, the joint distribution of the macro-state history from a T-HCGM, marginalized over subunit variables ($P^{marg}_{m,T}$), converges in Kullback-Leibler divergence to the distribution of its corresponding Dynamic Collapsed Model ($P_{col,T}$). For any finite time T:*

$$\lim_{m \to \infty} \mathrm{KL}(P_{col,T} \, || \, P^{marg}_{m,T}) = 0. \tag{2}$$

Proof Sketch. The proof proceeds by mathematical induction on the time step t, using the chain rule: $\mathrm{KL}(P_t || Q_t) = \mathrm{KL}(P_{t-1} || Q_{t-1}) + \mathbb{E}_{P_{t-1}}[\mathrm{KL}(P(S_t | H_{t-1}) || Q(S_t | H_{t-1}))]$. The inductive step shows that the expected KL-divergence of the one-step transition kernels converges to zero. We decompose this inner KL term into two parts: one for the Q-variable evolution and one for the unit-level variable generation. The KL-divergence for the Q-variable part is identically zero by definition, as the collapsed model preserves this mechanism exactly. The KL-divergence for the unit-level variable part converges to zero due to a continuous mapping argument: Assumption 2 ensures the convergence of the input, and Assumption 1 ensures the continuity of the function. Thus, the output distributions converge. ☐

The Framework: ST-HCMs

We now incorporate spatial dependencies to construct the full Spatio-Temporal Hierarchical Causal Model. This introduces new challenges, primarily the symmetric, contemporaneous interactions between units, which can create cycles in the causal graph. We address this by imposing a causal ordering on spatial interactions.

Definition 3 (Spatio-Temporal Hierarchical Causal Graphical Model (ST-HCGM))**.** *An ST-HCGM, $\mathcal{M}_{st-hcgm}$, extends the T-HCGM by allowing the generating mechanisms to depend on the states of spatial neighbors $\mathcal{N}(i)$.*

- *The Q-variable evolution for unit i now depends on its own macro-history and that of its neighbors:*

$$\boldsymbol{Q}_{i,t} \sim p(\boldsymbol{q}_t | U_i, \boldsymbol{S}_{i,<t}, \{\boldsymbol{S}_{k,<t}\}_{k \in \mathcal{N}(i)}). \tag{3}$$

- *The generation of unit-level variables $X^w_{i,t}$ similarly depends on the history of unit i and its neighbors $\mathcal{N}(i)$.*

To ensure the model is well-defined, we introduce the following assumptions regarding its spatial structure.

Assumption 3 (Spatial Markov Property (S1))**.** *The state of any unit i at time t is conditionally independent of all non-neighboring units, given the history of unit i and its designated neighbors $\mathcal{N}(i)$.*

Assumption 4 (Spatial Causal Ordering (S2))**.** *There exists a known, global ordering $<$ on the set of all units $\{1, \ldots, N\}$. For any two neighbors $i, j \in \mathcal{N}(i)$, contemporaneous causal influence is unidirectional: unit j can influence unit i at time t only if $j < i$.*

Assumption 5 (Spatial Homogeneity (S3))**.** *All units share the same set of mechanism functions $\{f^v\}$. The functional form of spatial influence is invariant across space, depending only on the relative relationship between units, not their absolute identities .*

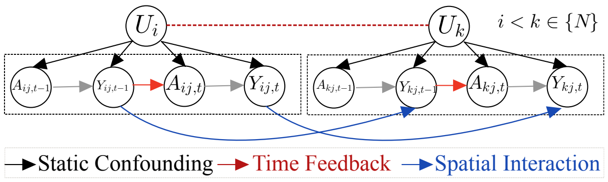


Figure 3: The challenge of identifiability.

Theorem 2 (Convergence of ST-HCMs). *Under Assumptions 1, 2, and 3-5, as the number of subunits $m \to \infty$, the joint distribution of the macro-state history from an ST-HCGM, marginalized over subunit variables ($P^{marg}_{m,T,N}$), converges in Kullback-Leibler divergence to the distribution of its corresponding Spatio-Temporal Dynamic Collapsed Model ($P_{col,T,N}$). For any finite N and T:*

$$\lim_{m\to\infty} \mathrm{KL}(P_{col,T,N} \,||\, P^{marg}_{m,T,N}) = 0. \tag{4}$$

Proof Sketch. We construct a single "Super-Unit" model, $\mathcal{M}_{super}$, whose state at time t is the concatenated vector of states of all N spatial units, i.e., $\boldsymbol{S}_{super,t} = (\boldsymbol{S}_{1,t}, \ldots, \boldsymbol{S}_{N,t})$. Assumption 4 guarantees that the internal dependencies within this Super-Unit's state vector at any time t are acyclic. Then $\mathcal{M}_{super}$ is a valid (though high-dimensional) T-HCM satisfying all premises of Theorem 1. Since the collapse for $\mathcal{M}_{super}$ is mathematically equivalent to the collapse process for the original ST-HCM. the convergence guaranteed by Theorem 1 for $\mathcal{M}_{super}$ directly implies the convergence for the ST-HCM. □

Identification and Estimation

Having established definition and convergence properties for ST-HCMs, we now address its primary purpose: enabling causal inference from spatio-temporal data. This section tackles identifiability, determining when causal effects can be uniquely computed from observational data. We develop sufficient conditions for identifiability in ST-HCMs.

The Identifiability Problem in Dynamic HCMs

The fundamental challenge of causal inference in dynamic hierarchical systems is threefold (Figure 3). First, we face unobserved unit-level confounding, represented by variable U_i. Second, the system's history acts as a confounder for current treatments and future outcomes. Third, neighboring units' states can confound effects, creating spatial dependencies that traditional methods cannot address. Our goal is to identify the full post-intervention distribution $P(Y_{k,t}|\,\mathrm{do}(A_{i,t'} = a^*))$, which describes the outcome at a location k and time t under an intervention on a treatment at location i and time t'. Thanks to the convergence theorems (Theorems 1 and 2), this problem is equivalent to identifying the effect in the corresponding Dynamic Collapsed Model (DCM). Our identifiability results are derived by analyzing the graphical structure of this DCM.

Identifiability via Sequential Adjustability

Our first identification strategy relies on the principle of adjusting for a sufficiently rich set of observed covariates to block all backdoor paths between treatment and outcome. We begin by stating the result for the purely temporal case, which forms the foundation for the more general theorem.

Proposition 1 (Identifiability in T-HCMs via Adjustability). *In a T-HCGM satisfying the premises of Theorem 1, the causal effect of a subunit-level treatment $A_{ij,t'}$ on a future outcome $Y_{ij,t}$ ($t \geq t'$) is identifiable if, in its corresponding DCM, all backdoor paths from the treatment node $Q^A_{i,t'}$ to the outcome node $Q^Y_{i,t}$ are blocked by conditioning on the observable macro-history $\boldsymbol{S}_{i,<t'}$.*

This principle naturally extends to the full spatio-temporal setting, where the adjustment set must also include the history of spatial neighbors.

Theorem 3 (Identifiability in ST-HCMs via Adjustability). *In an ST-HCGM satisfying the premises of Theorem 2, the causal effect $P(Y_{k,t}|\,\mathrm{do}(A_{i,t'} = a^*))$ is identifiable if, in its ST-DCM, all backdoor paths from the treatment $Q^A_{i,t'}$ to the outcome $Q^Y_{k,t}$ are blocked by conditioning on the observable spatio-temporal history. This history comprises the macro-states of unit i and its neighbors $\mathcal{N}(i)$ prior to time t'.*

Identifiability via Dynamic Instruments

When a sufficiently rich history is not observable, we can achieve identification by leveraging an exogenous source of variation—an instrumental variable.

Proposition 2 (Identifiability in T-HCMs via Instruments). *In a T-HCGM, the causal effect of $A_{ij,t'}$ on $Y_{ij,t}$ is identifiable if there exists a valid subunit-level instrumental variable $Z_{ij,t'}$. A valid instrument must (1) be a cause of $A_{ij,t'}$ (relevance) and (2) have no unblocked backdoor path to $Y_{ij,t}$ in the DCM that does not pass through $A_{ij,t'}$.*

This result also generalizes directly to the spatio-temporal framework, requiring a correspondingly instrument.

Theorem 4 (Identifiability in ST-HCMs via Instruments). *In an ST-HCGM, the causal effect $P(Y_{k,t}|\,\mathrm{do}(A_{i,t'} = a^*))$ is identifiable if there exists a valid spatio-temporal subunit-level instrumental variable $Z_{ij,t'}$. A valid instrument must satisfy the exclusion restriction with respect to the entire system: its node $Q^Z_{i,t'}$ in the ST-DCM must be d-separated from all static confounders $\{U_k\}$ and the complete, unobserved history of the entire spatio-temporal system.*

ATE with Time-Invariant Spatial Confounding

To demonstrate our framework's practical power, we focus on a classic problem in real-world applications: **estimating average treatment effects with unobserved, time-invariant spatial confounding.** This scenario is insightful since traditional spatio-temporal methods often fail, while our hierarchical structure provides a unique solution.

Problem Formulation We consider an ST-HCM where the primary challenge is an unobserved, time-invariant, unit-level confounder, U_i. This confounder affects the treatment assignment, outcomes, and potentially the dynamics of neighboring units through spatial spillover, creating complex confounding across both time and space.

Our goal is to identify and estimate the **Average Treatment Effect (ATE)** of a global, static treatment policy. Specifically, we are interested in the effect on the outcome at a future time horizon T, resulting from setting the subunit-level treatment A to a constant value a^* for all units i and all time steps $t \leq T$. The formal estimand is:

$$\text{ATE}_T = \mathbb{E}[Y_T|\text{do}(A_t = 1)] - \mathbb{E}[Y_T|\text{do}(A_t = 0)], \qquad (5)$$

where $\mathbb{E}[Y_T|\text{do}(A_t = a^*)]$ denotes the expected global average outcome at time T under the specified policy.

Identification Strategy The identification of this ATE is made possible by our theoretical framework, specifically by applying Theorem 3. The strategy involves three conceptual steps: (1) collapsing the ST-HCM to its equivalent ST-DCM, as guaranteed by Theorem 2; (2) applying a conditional version of sequential adjustment on the ST-DCM; and (3) leveraging the hierarchical structure to average out the unobserved confounder U_i.

Assuming that the dynamic and spatial confounding paths are blockable by the observable history (comprising the history of unit i and its neighbors $\mathcal{N}(i)$), Theorem 3 ensures that the ATE is identifiable. The identification formula can be expressed as an expectation over the distribution of the unobserved confounders $\mathbf{U} = (U_1, ..., U_N)$:

$$\mathbb{E}[Y_T|\text{do}(A_t = a^*)] = \mathbb{E}_{\mathbf{U}}\left[\mathbb{E}_{\text{G-Comp}}[Y_T|\text{do}(A_t = a^*), \mathbf{U}]\right].$$

The inner expectation, $\mathbb{E}_{\text{G-Comp}}[\cdot]$, represents the conditional causal effect for units with fixed static properties $\mathbf{U}$, which can be computed via a recursive procedure analogous to Robins' G-Computation formula. The outer expectation, $\mathbb{E}_{\mathbf{U}}[\cdot]$, is then computed by averaging these conditional effects over the population of units.

Estimation Algorithm We translate the identification strategy into a practical, two-stage estimation algorithm.

Stage 1: Learning Conditional Dynamics. For each unit i, we estimate a unit-specific model, $\mathcal{M}_i$, that captures the full spatio-temporal dynamics conditional on its latent static properties U_i. This model is trained on the subunit-level panel data centered at unit i, including the histories of its neighbors. The choice of model for $\mathcal{M}_i$ is flexible. Any method capable of capturing the relevant temporal and spatial dependencies can be employed, provided it can yield an estimate of the conditional expectation of the outcome. As our core contribution lies in the causal identification framework itself, rather than a specific predictive model, we prioritize methods that are well-established and interpretable. Suitable candidates range from parametric models, such as Linear Mixed-Effects Models (LMM) (Faraway 2006), to non-parametric approaches like Gradient Boosting Machines (GBM) (Natekin and Knoll 2013).

Stage 2: Conditional G-Computation via Simulation. With the trained models $\{\mathcal{M}_i\}_{i=1}^N$, we simulate the effect of the intervention for each unit and then aggregate the results. The procedure is detailed in Algorithm 1. For each unit i, we recursively simulate the system's evolution under a fixed treatment policy a^*. At each time step t, the model $\mathcal{M}_i$ predicts the outcome $\hat{y}_{i,t}(a^*)$ based on the simulated history of unit i and the observed history of its neighbors. The final outcome at time T, $\hat{\mu}_i(T; a^*)$, represents an estimate of the conditional causal effect for that unit. Finally, we average these estimates across all units to obtain the ATE.

Algorithm 1: ATE Estimation in ST-HCM

```
**Input:** Spatio-temporal data $\mathcal{D}$, time horizon $T$.
**Output:** Estimated Average Treatment Effect $\widehat{\text{ATE}}_T$.
**Stage 1: Learning Conditional Dynamics**
Train a set of unit-specific models $\{\mathcal{M}_i\}_{i=1}^N$ from $\mathcal{D}$.
**Stage 2: Simulation and Aggregation**
Initialize result sets $\mathcal{R}_0 \leftarrow \emptyset$, $\mathcal{R}_1 \leftarrow \emptyset$.
for $a^* \in \{0, 1\}$ do
    for $i = 1$ to $N$ do
        Initialize simulation history $\mathcal{H}_i$.
        for $t = 1$ to $T$ do
            Get neighbors' history $\mathcal{H}_{\mathcal{N}(i),<t}$ from $\mathcal{D}$.
            Predict outcome $\hat{y}_{i,t}^{(a^*)} \leftarrow \mathcal{M}_i(\text{do}(A_t = a^*), \mathcal{H}_i, \mathcal{H}_{\mathcal{N}(i),<t})$.
            Update $\mathcal{H}_i \leftarrow \mathcal{H}_i \cup \{(a^*, \hat{y}_{i,t}^{(a^*)})\}$.
        end for
        Add final outcome $\hat{y}_{i,T}^{(a^*)}$ to result set $\mathcal{R}_{a^*}$.
    end for
end for
**Final Calculation**
$\hat{\mathbb{E}}[Y_T|\text{do}(A_t = 1)] \leftarrow \text{mean}(\mathcal{R}_1)$.
$\hat{\mathbb{E}}[Y_T|\text{do}(A_t = 0)] \leftarrow \text{mean}(\mathcal{R}_0)$.
$\widehat{\text{ATE}}_T \leftarrow \hat{\mathbb{E}}[Y_T|\text{do}(A_t = 1)] - \hat{\mathbb{E}}[Y_T|\text{do}(A_t = 0)]$.
return $\widehat{\text{ATE}}_T$
```

Experiments

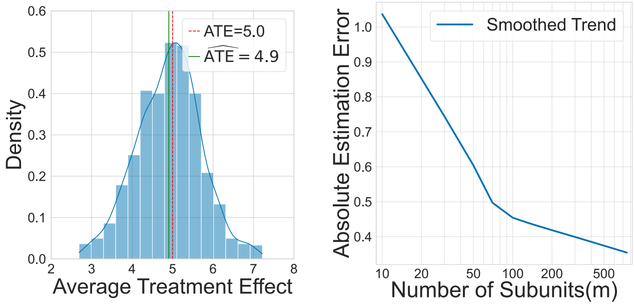
We conduct experiments on both a synthetic setting and a real-world case to evaluate ST-HCMs. We design 4 research questions to guide the evaluation:

- **RQ1 (Correctness)** Can our method recover known causal effects in an ideal synthetic setting?
- **RQ2 (Superiority)** How does our framework perform compared to baselines that neglect either the hierarchical structure or spatio-temporal dependencies of the data?
- **RQ3 (Robustness)** How robust is our method to violations of its key assumptions?
- **RQ4 (Applicability)** How sensitive are causal estimates to underlying structural assumptions in a real-world system, and how does ST-HCM address this challenge?

We implement our estimator with Linear Mixed Models (LMM) and Gradient Boosting Machines (GBM) models to serve as the unit-specific conditional mechanism $\mathcal{M}_i$.

Simulation Experiments

Correctness To verify the fundamental correctness of our proposed estimation procedure, we first assess its ability to produce unbiased and consistent estimates of the Average Treatment Effect (ATE). We generate data from a canonical spatio-temporal hierarchical process with a known true ATE.



(a) Unbiasedness (b) Consistency

Figure 4: Validation of ST-HCM estimator.

Figure 4 demonstrates the statistical properties of our ST-HCM estimator. The left panel shows the distribution of ATE estimates over multiple independent trials. The mean of this distribution (4.91) is nearly identical to the true ATE (5.0), demonstrating that our estimator is **unbiased**. The right panel shows that the absolute estimation error converges towards zero as the number of subunits per unit (m) increases. This confirms the **consistency** of our estimator, a key property guaranteed by our Collapse Theorem 2. These results provide strong evidence that our framework is correctly specified and theoretically sound (**answering RQ1**).

Superiority over Baselines Having established the correctness of our framework, we now conduct a rigorous stress test to evaluate its superiority over several strong baselines in the presence of unobserved confounding and spatial spillovers. To isolate the core causal identification capabilities of each method, we instantiate our framework and the hierarchical baselines using Linear Mixed Models (LMMs). This creates a challenging but well-understood linear setting where performance differences can be clearly attributed to the structural assumptions of each model. We compare our ST-HCM (LMM) against two primary classes of baselines:

1. **Aggregated Models**: A standard spatio-temporal LMM applied to data aggregated at the unit-time level, which ignores the hierarchical structure.
2. **Non-Spatial Models**: A Temporal HCM that models the hierarchy but ignores spatial dependencies, equivalent to a panel data model with fixed effects.

Table 1 presents the ATE absolute error for these models across a wide grid of confounding strengths (γ) and spatial spillover strengths (ρ). The results reveal two clear patterns.

First, hierarchy is essential for handling confounding. Across all levels of spatial spillover (all row groups), the error of the Aggregated model increases dramatically with the confounding strength. In contrast, both hierarchical models (T-HCM and ST-HCM) maintain low error, demonstrating their robustness to the unobserved unit-level confounder U_i. For instance, at $\rho = 2.0$ and $\gamma = 4.0$, the Aggregated model's error is 7.94, whereas the hierarchical models' errors are an order of magnitude smaller.

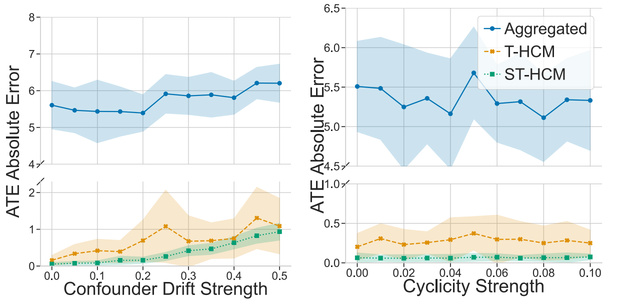
Second, spatial modeling is crucial under spillover. While the T-HCM performs well when spillover is mild ($\rho = 0.0, 0.5$), its error systematically increases as ρ grows. Our proposed ST-HCM, which explicitly models these spatial dependencies, consistently maintains the lowest error across nearly all settings. At $\gamma = 4.0$, as ρ increases from 0.0 to 2.0, the error of the T-HCM explodes from 0.11 to 2.35, while our ST-HCM's error remains stable at approximately 0.11. These findings indicate that our ST-HCM framework, by correctly modeling both the hierarchical structure and spatio-temporal dependencies, provides substantially more accurate causal estimates than methods that ignore either dimension of the data's complexity (**answering RQ2**).

		Confounding Strength (γ)				
ρ	**Model**	**0.0**	**1.0**	**2.0**	**3.0**	**4.0**
0.0	Aggregated	0.173	1.909	2.886	3.009	3.086
	T-HCM	**0.054**	**0.070**	**0.084**	0.126	**0.114**
	ST-HCM	0.057	0.073	0.086	**0.118**	0.117
1.0	Aggregated	0.189	2.261	3.624	4.013	4.260
	T-HCM	0.132	0.108	0.161	0.251	0.224
	ST-HCM	**0.056**	**0.066**	**0.076**	**0.127**	**0.121**
1.5	Aggregated	0.211	3.420	5.051	5.950	6.734
	T-HCM	0.174	0.135	0.335	0.507	0.606
	ST-HCM	**0.058**	**0.066**	**0.064**	**0.123**	**0.113**
2.0	Aggregated	0.228	3.962	5.620	6.819	7.942
	T-HCM	0.329	0.271	0.894	1.320	2.354
	ST-HCM	**0.056**	**0.072**	**0.070**	**0.124**	**0.110**

Table 1: Mean ATE Absolute Error for different levels of confounding strength (γ) and spatial spillover (ρ). Both T-HCM (no spatial modeling) and ST-HCM (full model) are proposed hierarchical estimators. T-HCM is optimal when $\rho = 0$, while ST-HCM excels when $\rho > 0$, demonstrating the importance of correctly specifying spatial dependencies.

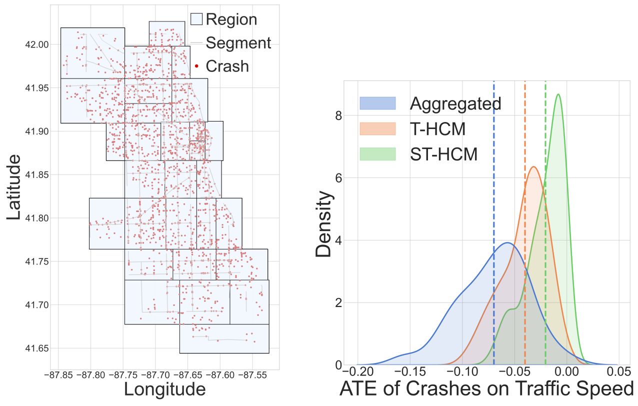
Robustness We assess the robustness of our ST-HCM (LMM) estimator by systematically violating two key assumptions: the time-invariance of the unobserved confounder and the causal ordering of spatial effects (Assumption 4). Our simulations are based on a challenging linear setting ($N = 16, m = 50, T = 8$) with fixed baseline confounding ($\gamma = 2.0$) and spatial spillover ($\rho = 1.5$).

As illustrated in Figure 5, our proposed model demonstrates considerable robustness. When the time-invariance assumption is relaxed by introducing a temporal drift to the confounder (Figure 5a), the error of our ST-HCM (LMM) increases mildly but remains significantly lower than that of the non-spatial T-HCM baseline. More impressively, when the spatial ordering assumption is violated by introducing simultaneous feedback loops (Figure 5b), the performance of our ST-HCM (LMM) remains nearly unaffected, maintaining a consistently low error and substantially outperforming both the T-HCM and Aggregated baselines across all violation intensities. This suggests our framework is a robust tool for practical applications where these assumptions may not strictly hold (**answering RQ3**).



(a) Time-Variant Confounding (b) Violation of Spatial Order

Figure 5: Robustness analysis of estimators.



(a) Chicago Layout (b) Estimated ATE Distributions

Figure 6: Real-world application.

Real-World Application: Urban Traffic Dynamics

We demonstrate the applicability of our framework in real-world using a Chicago traffic data set[1]. Our analysis uses traffic speed data from 1,025 road segments (subunits) across 29 traffic regions (units) over a one-week period, comprising over 1.2 million observations and 1943 treatment events (Figure 6a).

To capture the complex non-linearities inherent in traffic systems, we employ Gradient Boosting Machines (GBM) as the implementation for the unit-specific conditional mechanism $\mathcal{M}_i$. We compare our full ST-HCM against a T-HCM that ignores spatial effects, and an Aggregated model that ignores the hierarchy entirely.

Results summarized in Figure 6b reveal that causal estimates are sensitive to structural assumptions. Progressively incorporating the hierarchical and then the spatial structure systematically attenuates the estimated ATE. Notably, this process also sharpens the posterior distribution, demonstrating that a more correctly specified model yields not only a less biased but also a more precise estimate by disentangling the treatment effect from unobserved confounders and spatial spillovers (**answering RQ4**).

Related Work

Hierarchical causal models The hierarchical causal model framework (Weinstein and Blei 2024), exploits subunit variation to address unobserved unit-level confounding, generalizing fixed-effects and related econometric methods (Gelman and Hill 2007; Wooldridge 2010). However, they are designed for static i.i.d. settings and cannot capture temporal or spatial dependencies. Recent advances such as DML-panel methods with correlated random effects and dynamic multivariate panel models suggest possible extensions toward temporal settings (Clarke and Polselli 2025; Tikka and Helske 2024). Our ST-HCMs extend HCMs by explicitly modeling temporal dynamics and spatial interactions while still addressing latent unit-level confounders.

Causal inference for spatio-temporal data Existing approaches fall into two categories: (1) causal discovery methods that learn dynamic causal graphs from observational data (Pamfil et al. 2020; Zhao et al. 2023; Gong et al. 2024; Wang et al. 2025), and (2) causal effect estimators that adapt frameworks like G-computation, marginal structural models, or propensity scores to dynamic or spatial settings, with recent extensions using recurrent or transformer architectures (Bica, Alaa, and van der Schaar 2020; Melnychuk, Frauen, and Feuerriegel 2022; Bhattacharya and Sen 2024; Li et al. 2025; Oprescu et al. 2025). However, most assume no latent unit-level confounders, which is unrealistic in domains like environment or transportation. In contrast, ST-HCMs address this gap by structurally incorporating unobserved, static confounders, enabling more credible causal inference.

Panel data and IV-based causal inference Panel methods such as fixed effects, difference-in-differences, and synthetic control estimate effects via within-unit variation or counterfactuals (Wooldridge 2010; Abadie, Diamond, and Hainmueller 2010), but their i.i.d. and linearity assumptions limit use in spatio-temporal contexts (Millimet and Bellemare 2023; Arkhangelsky and Imbens 2024). IV-based approaches mitigate confounding, for example through DeepIV and semiparametric IVs (Hartford et al. 2017; Xie et al. 2025b; Cui et al. 2025); and the recent Time-dependent Instrumental Factor Model (TIFM) addresses time-varying confounders with learned substitutes (Cheng et al. 2024). Nevertheless, IV methods rely on valid instruments, which are frequently difficult to obtain. The proposed ST-HCMs offer a flexible solution by integrating hierarchical modeling with graphical and spatio-temporal reasoning.

Conclusion

In this paper, we introduced Spatio-Temporal Hierarchical Causal Models (ST-HCMs), a novel framework enabling reliable causal inference from nested spatio-temporal data. Our theoretical foundation rests on a Spatio-Temporal Collapse Theorem, guaranteeing that hierarchical models converge to simpler, tractable counterparts, providing rigorous causal identification. Extensive experiments validate this framework, demonstrating theoretical soundness, practical advantages over baselines that ignore hierarchy or spatial effects, and robustness to assumption violations.

---

[1] https://data.cityofchicago.org/

Acknowledgements

This work was supported by Public Computing Cloud, Renmin University of China and the fund for building world-class universities (disciplines) of Renmin University of China.

References

Abadie, A.; Diamond, A.; and Hainmueller, J. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, 105(490): 493–505.

Ali, S.; Faruque, O.; and Wang, J. 2024. Estimating direct and indirect causal effects of spatiotemporal interventions in presence of spatial interference. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 213–230. Springer.

Ali, S.; Hasan, U.; Li, X.; Faruque, O.; Sampath, A.; Huang, Y.; Gani, M. O.; and Wang, J. 2024. Causality for Earth Science – A Review on Time-series and Spatiotemporal Causality Methods. arXiv:2404.05746.

Angrist, J. D.; and Pischke, J.-S. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arkhangelsky, D.; and Imbens, G. 2024. Causal models for longitudinal and panel data: a survey. *The Econometrics Journal*, 27(3): C1–C61.

Aronow, P. M.; and Samii, C. 2017. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4): 1912 – 1947.

Bhattacharya, S.; and Sen, S. 2024. Causal effect estimation under network interference with mean-field methods. arXiv:2407.19613.

Bica, I.; Alaa, A. M.; and van der Schaar, M. 2020. Time Series Deconfounder: Estimating Treatment Effects over Time in the Presence of Hidden Confounders. arXiv:1902.00450.

Cheng, D.; Xu, Z.; Li, J.; Liu, L.; Liu, J.; Gao, W.; and Le, T. D. 2024. Instrumental variable estimation for causal inference in longitudinal data with time-dependent latent confounders. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 11480–11488.

Clarke, P. S.; and Polselli, A. 2025. Double machine learning for static panel models with fixed effects. *The Econometrics Journal*, utaf011.

Cui, E. H.; Lu, X.; Zhou, J.; Zhou, H.; and Li, G. 2025. A Semiparametric Bayesian Method for Instrumental Variable Analysis with Partly Interval-Censored Time-to-Event Outcome. arXiv:2501.14837.

Dong, P.; Wang, X.-L.; Bose, I.; Ng, K. K. H.; Zhang, X.; and Zhang, X. 2024. Causally-Aware Spatio-Temporal Multi-Graph Convolution Network for Accurate and Reliable Traffic Prediction. arXiv:2408.13293.

Dong, P.; Zhang, X.; and Zhang, X. 2025. A data-driven approach for spatio-temporal causal analysis in large-scale urban traffic networks. *Transportation Research Part E: Logistics and Transportation Review*, 202: 104244.

Faraway, J. J. 2006. *Extending Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton, FL: Chapman & Hall/CRC. ISBN 1-584-88424-X.

Gelman, A.; and Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gong, C.; Zhang, C.; Yao, D.; Bi, J.; Li, W.; and Xu, Y. 2024. Causal discovery from temporal data: An overview and new perspectives. *ACM Computing Surveys*, 57(4): 1–38.

Hartford, J.; Lewis, G.; Leyton-Brown, K.; and Taddy, M. 2017. Deep IV: A Flexible Approach for Counterfactual Prediction. In Precup, D.; and Teh, Y. W., eds., *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 1414–1423. PMLR.

Hernán, M. A.; and Robins, J. M. 2006. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7): 578–586.

Hernán, M. A.; and Robins, J. M. 2010. Causal inference.

Jensen, D.; Burroni, J.; and Rattigan, M. 2020. Object Conditioning for Causal Inference. In Adams, R. P.; and Gogate, V., eds., *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, 1072–1082. PMLR.

Li, J.; Li, X.; and Zhou, X. 2025. FAP-CD: Fairness-Driven Age-Friendly Community Planning via Conditional Diffusion Generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27): 28168–28176.

Li, S.; Li, H.; Li, X.; Xu, Y.; Lin, Z.; and Jiang, H. 2025. Causal Intervention Is What Large Language Models Need for Spatio-Temporal Forecasting. *IEEE Transactions on Cybernetics*, 55(8): 3825–3837.

Martinussen, T.; Vansteelandt, S.; Tchetgen Tchetgen, E. J.; and Zucker, D. M. 2017. Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics*, 73(4): 1140–1149.

Melnychuk, V.; Frauen, D.; and Feuerriegel, S. 2022. Causal Transformer for Estimating Counterfactual Outcomes. arXiv:2204.07258.

Millimet, D.; and Bellemare, M. F. 2023. Fixed effects and causal inference.

Natekin, A.; and Knoll, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, Volume 7 - 2013.

Oprescu, M.; Park, D. K.; Luo, X.; Yoo, S.; and Kallus, N. 2025. GST-UNet: Spatiotemporal Causal Inference with Time-Varying Confounders. arXiv:2502.05295.

Pamfil, R.; Sriwattanaworachai, N.; Desai, S.; Pilgerstorfer, P.; Beaumont, P.; Georgatzis, K.; and Aragam, B. 2020. DYNOTEARS: Structure Learning from Time-Series Data. arXiv:2002.00498.

Pearl, J. 2009. Causal inference in statistics: An overview. *Statistics Surveys*, 3(none): 96 – 146.

Reich, B. J.; Yang, S.; Guan, Y.; Giffin, A. B.; Miller, M. J.; and Rappold, A. 2021. A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*, 89(3): 605–634.

Robins, J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9): 1393–1512.

Tikka, S.; and Helske, J. 2024. dynamite: An R Package for Dynamic Multivariate Panel Models. arXiv:2302.01607.

Wang, K.; Varambally, S.; Watson-Parris, D.; Ma, Y.-A.; and Yu, R. 2025. Discovering Latent Causal Graphs from Spatiotemporal Data. arXiv:2411.05331.

Wang, Y.; Zhou, X.; Mascolo, C.; Noulas, A.; Xie, X.; and Liu, Q. 2018. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data.

Weinstein, E. N.; and Blei, D. M. 2024. Hierarchical Causal Models. arXiv:2401.05330.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Xie, Z.; Hong, Z.; Lyu, W.; Wang, H.; Wang, G.; and Zhang, D. 2025a. CoAlign: Uncertainty Calibration of LLM for Geospatial Repartition. In Rehm, G.; and Li, Y., eds., *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, 254–263. Vienna, Austria: Association for Computational Linguistics. ISBN 979-8-89176-288-6.

Xie, Z.; Lyu, W.; Song, Y.; Wang, H.; Yang, G.; Liu, Y.; He, T.; Zhang, D.; and Wang, G. 2025b. Scalable Area Difficulty Assessment with Knowledge-enhanced AI for Nationwide Logistics Systems. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, 2713–2724. New York, NY, USA: Association for Computing Machinery. ISBN 9798400712456.

Xu, Z.; and Zhou, X. 2024. CGAP: Urban Region Representation Learning with Coarsened Graph Attention Pooling. arXiv:2407.02074.

Zhang, W.; and Ning, K. 2023. Spatiotemporal heterogeneities in the causal effects of mobility intervention policies during the COVID-19 outbreak: A spatially interrupted time-series (SITS) analysis. *Annals of the American Association of Geographers*, 113(5): 1112–1134.

Zhao, W.; Zhang, S.; Wang, B.; and Zhou, B. 2023. Spatiotemporal causal graph attention network for traffic flow prediction in intelligent transportation systems. *PeerJ Computer Science*, 9: e1484.