Toward Causal Inference for Spatio-Temporal Data
Conflict and Forest Loss in Colombia
Christiansen, Rune; Baumann, Matthias; Kuemmerle, Tobias; Mahecha, Miguel D.; Peters, Jonas Martin

Published in:
Journal of the American Statistical Association

DOI:
10.1080/01621459.2021.2013241

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Toward Causal Inference for Spatio-Temporal Data: Conflict and Forest Loss in Colombia

Rune Christiansen, Matthias Baumann, Tobias Kuemmerle, Miguel D. Mahecha & Jonas Peters


To link to this article: https://doi.org/10.1080/01621459.2021.2013241
Toward Causal Inference for Spatio-Temporal Data: Conflict and Forest Loss in Colombia

Rune Christiansen, Matthias Baumann, Tobias Kuemmerle, Miguel D. Mahecha, and Jonas Peters

1. Introduction

1.1. Spatio-Temporal Data Analysis

In principle, all data are spatio-temporal data: Any observation of any phenomenon occurs at a particular point in space and time. If information on the spatio-temporal origin of data is available, this information can be exploited for statistical modeling in various ways; this is the study of spatio-temporal statistics (e.g., Cressie and Wikle 2015; Wikle, Zammit-Mangion, and Cressie 2019). Spatio-temporal statistical models find their application in many environmental and sustainability sciences, and have been used, for example, for the analysis of biological growth patterns (Chaplain, Singh, and McLachlan 1999), to model meteorological fields (Bertolacci et al. 2019), or to assess the development of land-use change (Liu, Huang, and Zhang 2017) and sea level rise (Zammit-Mangion et al. 2015). They are used in epidemiology for prevalence mapping of infectious diseases (Giorgi et al. 2018), and play a key role in socio-economic research, for example, for the modeling of housing prices (Holly, Pesaran, and Yamagata 2010), or for election forecasting (Pavia, Larraz, and Montero 2008). In almost all of these domains, the abundance of spatio-temporal data has increased rapidly over the last decades. Several advances aim to improve the accessibility of such datasets, for example, via “data cube” approaches (e.g., Nativi, Mazzetti, and Craglia 2017; Mahecha et al. 2020).

Most spatio-temporal statistical models are models for the observational distribution, that is, they model processes that are passively observed. By allowing for spatio-temporal trends and dependence structures, such models can be accurate descriptions of complex processes, and have proven to be effective tools for spatio-temporal prediction, inference and forecasting (e.g., Wikle, Zammit-Mangion, and Cressie 2019). However, to answer interventional questions such as “How does a certain policy change affect land-use patterns?”, we require a model for the intervention distribution, that is, for data generated under a change in the data generating mechanism—we require a causal model for the data-generating process.

1.2. Causality

Causal models can be used to quantify causal relations between random variables, (e.g., by analyzing the change in expected value of $Y$ when intervening on $X$). However, existing causal models do not apply well to our setting, since they are either restricted to independent and identically distributed (iid) or time series data, or they rely on various assumptions which cannot be expected to hold in our application (see Appendix A for a review of existing work). In this article, we introduce a novel class of causal models for multivariate spatio-temporal stochastic processes. A spatio-temporal dataset may then be viewed as a single realization from such a model, observed at discrete points in space and time. The full causal structure
among all variables of a spatio-temporal process can hardly be fully specified. In practice, however, a full causal specification may also not be necessary: we are often interested in quantifying only certain causal relationships, while being indifferent to other parts of the causal structure. The causal models introduced in this article are well adapted to such settings. They allow us to model a causal influence of a vector of covariates on a target variable while leaving other parts of the causal structure unspecified. In particular, the models accommodate largely unspecified autocorrelation patterns in the response variable, which are a common phenomena in spatio-temporal data.

The introduced framework allows us to formally talk about causality in a spatio-temporal context and can be used to construct well-defined targets of inference. As an example, we define the intervention effect ("causal effect") of $X$ on $Y$. We show that this effect can be estimated from observational spatio-temporal data and introduce a corresponding estimator. We prove consistency and verify this finding by a simulation experiment. We furthermore construct a nonparametric hypothesis test for the causal effect being zero. Our methods do not rely on any distributional assumptions on the data-generating process. They furthermore allow for the influence of arbitrarily many latent confounders even not varying across time. In principle, our method also allows to analyze problems where temporal and spatial dimensions are interchanged, meaning that confounders may vary in time but remain static across space.

Our work has been motivated by the following application.

1.3. Conflict and Forest Loss in Colombia

Tropical forests are rich in biodiversity (Kreft and Jetz 2007), store large amounts of carbon (Avitabile et al. 2016), play an important role in climate-regulation, and provide livelihoods to millions of people (Lambin and P. Meyfroidt 2011). Yet, tropical forests continue to be under pressure due to agricultural expansion (Angelsen and Kaimowitz 1999), mining (Sonter et al. 2017), timber harvest (Pearson, Brown, and Casarim 2014), or urban expansion (DeFries et al. 2010). A problem that is still only partly understood is the interaction between forest loss and armed conflicts (Baumann and Pradhan 2009), which may be unobserved. Failing to account for confounding effects (e.g., Sánchez-Cuervo and Aide 2013; Castro-Nunez et al. 2017). We believe that a purely data-driven approach can also be a useful addition to this debate.

In our analysis, we use a dataset containing the following variables.

- $X_{t}^{s}$: binary conflict indicator for location $s$ at year $t$.
- $Y_{t}^{s}$: absolute forest loss in location $s$ from year $t - 1$ to year $t$, measured in km².
- $W_{t}^{s}$: distance from location $s$ to the closest road, measured in km.

Data are annually aggregated, covering the years from 2000 to 2018, and spatially explicit at a 10km × 10km-resolution. We provide a detailed description of the data processing in Section 4.

A summary of the dataset can be seen in Figure 1. Visually, there is a strong positive dependence between the occurrence of a conflict and the loss of forest canopy. This observation is supported by simple summary statistics: the average forest loss across measurements classified as conflict events significantly exceeds that from non-conflict events by almost 50% (Figure 1, left), conforming previous findings (e.g., Landholm, Pradhan, and Kropp 2019). When seeking a causal explanation for the observed data, however, we regard such an analysis as flawed in two ways. First, both conflicts and forest loss predominantly occur in areas with high accessibility (Figure 1, right), indicating that the potential causal effect of $X$ on $Y$ is confounded by $W$. In fact, we expect the existence of several other confounders (e.g., population density, market infrastructure, etc.), many of which may be unobserved. Failing to account for confounding variables leads to biased estimates of the causal effect. Second, strong spatial dependencies in $X$ and $Y$ reduce the effective sample size, and a standard $t$-test thus exaggerates the significance of the observed difference in sample averages. To test hypotheses about $X$ and $Y$, we need statistical tests which are adapted to the spatio-temporal nature of data.

1.4. Contributions and Structure of the Article

Apart from the case study, this article contains three main theoretical contributions: the definition of a causal model for spatio-temporal data, a method for estimating causal effects (both Section 2), and a hypothesis test for the overall existence of such effects (Section 3). Section 4 applies our methodology to the above example. All data used for our analysis are publicly available. A description of how it can be obtained, along with an implementation of our method and reproducing scripts for all our figures and results, can be found at github.com/runesen/spatio_temporal_causality. All proofs are contained in Appendix H.
cumulative forest loss 2000–2018
distance to closest road

Figure 1. Temporally aggregated summary of the dataset described in Section 1.3. Conflicts are predictive of exceedances in forest loss (left), but this dependence is partly induced by a common dependence on accessibility, which we measure by the mean distance to a road (right). Failing to account for this variable and other confounders biases our estimate of the causal influence of conflict on forest loss.

2. Quantifying Causal Effects for Spatio-Temporal Data

A spatio-temporal dataset may be viewed as a single realization of a spatio-temporal stochastic process, observed at discrete points in space-time. In this section, we provide a formal framework to quantify causal relations among the components of a multivariate spatio-temporal process, and show how to estimate causal effects from observational data.

2.1. Causal Models for Spatio-Temporal Processes

Throughout this section, let \((\Omega, A, P)\) be some background probability space. A \(p\)-dimensional spatio-temporal process \(Z\) is a random variable taking values in the sample space \(Z_p\) of all \((B(\mathbb{R}^2 \times \mathbb{N}), B(\mathbb{R}^p))\)-measurable functions, where \(B(\cdot)\) denotes the Borel \(\sigma\)-algebra. We equip \(Z_p\) with the \(\sigma\)-algebra \(\mathcal{F}_p\), defined as the smallest \(\sigma\)-algebra such that for all \(B \in B(\mathbb{R}^p)\), the mapping \(Z_p \ni z \mapsto \int_B z(x)dx\) is \((\mathcal{F}_p, B(\mathbb{R}))\)-measurable. The induced probability measure \(P\) on the measurable space \((Z_p, \mathcal{F}_p)\), for every \(F \in \mathcal{F}_p\) defined by \(P(F) := P(Z^{-1}(F))\), is said to be the distribution of \(Z\). Throughout this article, we use the notation \(Z^s_t\) to denote the random vector obtained from marginalizing \(Z\) at location \(s\) and time point \(t\). We use \(Z_i\) for the time series \((Z^s_i)_{s \in \mathbb{N}}\), \(Z^s\) for the spatial process \((Z^s_j)_{j \in \mathbb{R}^2}\), and \(Z^{(s)}\) for the spatio-temporal process corresponding to the coordinates in \(S \subseteq \{1, \ldots, p\}\). We call a spatio-temporal process weakly stationary if the marginal distribution of \(Z^s_i\) is the same for all \((s, t) \in \mathbb{R}^2 \times \mathbb{N}\), and time-invariant if \(P(Z^1 = Z^2 = \cdots) = 1\).

Multivariate spatio-temporal processes are used for the joint modeling of different phenomena, each of which corresponds to a coordinate process. Let us consider a decomposition of these coordinate processes into disjoint ‘bundles’. We are interested in specifying causal relations among these bundles while leaving the causal structure among variables within each bundle unspecified. Similarly to a graphical model (Lauritzen 1996), our approach relies on a factorization of the joint distribution of \(Z\) into a number of components, each of which models the conditional distribution for one bundle given several others. This approach induces a graphical relation among the different bundles. We will equip these relations with a causal interpretation by additionally specifying the distribution of \(Z\) under certain interventions on the data-generating process.

Definition 1 (Causal graphical models for spatio-temporal processes). A causal graphical model for a \(p\)-dimensional spatio-temporal process \(Z\) is a triplet \((\mathcal{S}, \mathcal{G}, \mathcal{P})\) consisting of:

- a family \(\mathcal{S} = (S^k)_{k=1}^k\) of nonempty, disjoint sets \(S_1, \ldots, S_k \subseteq \{1, \ldots, p\}\) with \(\bigcup_{j=1}^k S_j = \{1, \ldots, p\}\),
- a directed acyclic graph \(\mathcal{G}\) with vertices \(S_1, \ldots, S_k\), and
- a family \(\mathcal{P} = (\mathcal{P}^j)^{k}_{j=1}\) of collections \(\mathcal{P}^j = \{P^j_x\}_{x \in Z_{\mathcal{P}A_j}}\) of distributions on \((Z_{|S_j}, \mathcal{F}_{|S_j})\), where for every \(j\), \(PA_j := \bigcup_{i:S_i \rightarrow S_j \in \mathcal{G}} S_i\). Whenever \(PA_j = \emptyset\), \(\mathcal{P}^j\) consists only of a single distribution which we denote by \(\mathcal{P}^j\).

Since \(\mathcal{G}\) is acyclic, we can without loss of generality assume that \(S_1, \ldots, S_k\) are indexed such that \(S_i \not\rightarrow S_j\) in \(\mathcal{G}\) whenever \(i > j\). The above components induce a unique joint distribution \(\mathcal{P}\) over \(Z\). For every \(F = \times^k_{j=1} F_j\), it is defined by

\[
\mathcal{P}(F) = \int_{F_1} \cdots \int_{F_k} \mathcal{P}^n(\mathcal{P}A_k) (dz^{(S_k)}) \cdots \mathcal{P}^1 (dz^{(S_1)}).
\]
We call \( \mathbb{P} \) the observational distribution. For each \( j \in \{1, \ldots, k\} \), the conditional distribution of \( \mathbf{Z}(\mathbf{S}^j) \) given \( \mathbf{Z}(\mathbf{PA}^j) \) as induced by \( \mathbb{P} \) equals \( \mathbb{P}^{(j)} \). We define an intervention on \( \mathbf{Z}(\mathbf{S}^j) \) as replacing \( \mathbb{P}^{(j)} \) by another model \( \hat{\mathbb{P}}^{(j)} \). This operation results in a new graphical model \((S, G, \hat{\mathbb{P}})\) for \( Z \) which induces, via (1), a new distribution \( \hat{\mathbb{P}} \), the interventional distribution.

Assume that we perform an intervention on \( \mathbf{Z}(\mathbf{S}^i) \). By definition, the resulting interventional distribution differs from the observational distribution only in the way in which \( \mathbf{Z}(\mathbf{S}^i) \) depends on \( \mathbf{Z}(\mathbf{PA}^i) \), while all other conditional distributions \( \mathbf{Z}(\mathbf{S}^j) | \mathbf{Z}(\mathbf{PA}^i), j \neq i \), remain the same. This property is analogous to the modularity property of structural causal models (Haavelmo 1944; Pearl 2009; Peters, Janzing, and Schölkopf 2017) and justifies a causal interpretation of the conditionals in \( \hat{\mathbb{P}} \). We refer to the graph \( G \) as the causal structure of \( Z \), and sometimes write \( Z = [\mathbf{Z}(\mathbf{S}^i) | \mathbf{Z}(\mathbf{PA}^i)] \cdots [\mathbf{Z}(\mathbf{S}^j)] \) to emphasize this structure.

### 2.2. Latent Spatial Confounder Model

Motivated by the example introduced in Section 1.3, we are particularly interested in scenarios where a target variable \( Y \) is causally influenced by a vector of covariates \( X \), and where \((X, Y)\) are additionally affected by some latent variables \( H \). In general, inferring causal effects under arbitrary influences of latent confounders is impossible, and we therefore need to impose additional restrictions on the variables in \( H \). Here we make the fundamental assumption that they do not vary across time (alternatively, one can assume that the hidden variables are invariant over space, see Appendix B.3).

**Definition 2 (Latent spatial confounder model).** Consider a spatio-temporal process \((X, Y, H) = (X_t^i, Y_t^j, H_t^k)_{(i,j,k)\in\mathbb{R}^2\times\mathbb{N}}\) over a real-valued response \( Y \), a vector of covariates \( X \in \mathbb{R}^d \) and a vector of latent variables \( H \in \mathbb{R}^h \). We call a causal graphical model over \((X, Y, H)\) with causal structure \([\mathbf{Y} | \mathbf{X}, \mathbf{H}]|\mathbf{H} | \mathbf{H}[\mathbf{H}]\) a latent spatial confounder model (LSCM) if both of the following conditions hold true for the observational distribution.

- The latent process \( H \) is weakly stationary and time-invariant.
- There exists a measurable function \( f : \mathbb{R}^{d+h+1} \to \mathbb{R} \) and an iid sequence \( \epsilon^1, \epsilon^2, \ldots \) of weakly stationary error processes independent of \((X, H)\), such that

\[
Y_t^j = f(X_t^i, H_t^k, \epsilon_t^j) \quad \text{for all } (s, t) \in \mathbb{R}^2 \times \mathbb{N}. \tag{2}
\]

We require that for all \( x \in \mathbb{R}^d, f(x, H_t^1, \epsilon_0^j) \) has finite expectation.

Throughout this section, we assume that \((X, Y, H)\) come from an LSCM. The above definition says that for every \( s, t \), \( Y_t^j \) depends on \((X, H)\) only via \((X_t^i, H_t^k)\), and that this dependence remains the same for all points in space-time. Together with the weak stationarity of \( H \) and \( \epsilon \), this assumption ensures that the average causal effect of \( X_s^j \) on \( Y_t^j \) (which we introduce below) remains the same for all \( s, t \). Our model class imposes no restrictions on the marginal distribution of \( X \). The spatial dependence structure of the error process \( \epsilon \) must have the same marginal distributions everywhere, but is otherwise unspecified.

### 2.3. Estimation of the Average Causal Effect

#### 2.3.1. Definition and Consistency

In practice, we only observe the process \((X, Y)\) at a finite number of points in space and time. We assume that at every temporal instance, we observe the process at the same spatial locations \(s_1, \ldots, s_n \in \mathbb{R}^2\) (these locations need not lie on a regular grid). To simplify notation, we furthermore assume that the observed time points are \( t = 1, 2, \ldots, m \), that is, we have

\[1\] We are grateful to Steffen Lauritzen for emphasizing this viewpoint.
access to a dataset \((X^m_n, Y^m_n) = (X^m_t, Y^m_t)_{(t,s) \in \{s_1, \ldots, s_k\} \times \{1, \ldots, m\}}\). The proposed method is based on the following key idea: for every \(s \in \{s_1, \ldots, s_k\}\), we observe several time instances \((X^m_t, Y^m_t)\), \(t \in \{1, \ldots, m\}\), all with the same conditionals \(Y^m_t \mid (X^m_t, H^m_t)\). Since \(H\) is time-invariant, we can, for every \(s\), estimate \(f_{Y \mid (X,H)}(\cdot, h)_s\) for the (unobserved) realization \(h_s\) of \(H^m_t\) using the data \((X^m_t, Y^m_t)\), \(t \in \{1, \ldots, m\}\). The expectation in (4) is then approximated by averaging estimates obtained from different spatial locations. This idea is visualized in Figure 2. More formally, our method requires as input a model class for the regressions \(f_{Y \mid (X,H)}(\cdot, h)\), \(h \in \mathbb{R}^d\), alongside with a suitable estimator \(\hat{f}_{Y \mid X} = (\hat{f}_{Y \mid X}^{m})_{m \in \mathbb{N}}\), and returns

\[
\hat{f}_{\text{AVE}(X \to Y)}(X^m_n, Y^m_n)(x) := \frac{1}{n} \sum_{i=1}^{n} \hat{f}_{Y \mid X}^{m}(X^m_i, Y^m_i)(x), \tag{5}
\]

an estimator of the average causal effect (3) within the given model class. In Section 3, we furthermore provide a statistical test for the overall existence of a causal effect. Our approach may be seen as summarizing the output of a spatially varying regression model (e.g., Gelfand et al. 2003) that is allowed to change arbitrarily from one location to the other (within the model class dictated by \(\hat{f}_{Y \mid X}\)). By permitting such flexibility, our method does not rely on observing data from a continuous or spatially connected domain, and accommodates complex influences of the latent variables. An implementation can be found in our code package, see Section 1.4.

To prove consistency of our estimator, we let the number of observable points in space-time increase. Let therefore \((s_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}^2\) be a sequence of spatial coordinates, and consider the array of data \((X^m_n, Y^m_n)_{n \geq 1}\), where for every \(n, m \in \mathbb{N}\), \((X^m_n, Y^m_n) = (X^m_t, Y^m_t)_{(t,s) \in \{s_1, \ldots, s_k\} \times \{1, \ldots, m\}}\). We want to prove that the corresponding sequence of estimators (5) consistently estimates (3). To obtain such a result, we need two central assumptions.

**Assumption 1 (LLN for the latent process).** For all measurable functions \(\varphi : \mathbb{R}^d \to \mathbb{R}\) with \(\mathbb{E}[|\varphi(H^m_1)|] < \infty\) it holds that
\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(H^m_i) \to \mathbb{E}[\varphi(H^m_1)] \text{ in probability as } n \to \infty.
\]

Assumption 1 ensures that, for an increasing number of spatial locations at which data are observed, the spatial average in (5) approximates the expectation in (3). The assumption is satisfied, for example, for a stationary Gaussian process \(H^1\) that is sampled regularly across an increasing spatial domain, see Proposition 1 in Appendix D. As the number of observed time points tends to infinity, we furthermore require the estimators \(\hat{f}_{Y \mid X}^m\) to converge to the integrand in (3), at least in some area \(\mathcal{X} \subseteq \mathbb{R}^d\).

**Assumption 2 (Consistent estimators of the conditional expectations).** There exists \(\mathcal{X} \subseteq \mathbb{R}^d\) such that for all \(x \in \mathcal{X}\) and \(s \in \mathbb{R}^d\), it holds that \(\hat{f}_{Y \mid X}^m(X^m_s, Y^m_s)(x) \to f_{Y \mid (X,H)}(x, H^m_s)\) as \(m \to \infty\), in probability under \(\mathbb{P}(\cdot \mid H = h)\). It follows from the dominated convergence theorem that this assumption implies Assumption 2. Under Assumptions 1 and 2, we obtain the following consistency result.

**Theorem 1 (Consistent estimator of the average causal effect).** Let \((X, Y, H)\) come from an LSCM as defined in Definition 2. Let \((s_n)_{n \in \mathbb{N}}\) be a sequence of spatial coordinates such that the marginalized process \((H^m_n)_{m \in \mathbb{N}}\) satisfies Assumption 1, and assume that for all \(x \in \mathcal{X}\), \(s \in \mathbb{R}^2\) and almost all \(h\), \(\hat{f}_{Y \mid X}^m(X^m_n, Y^m_n)(x) \to f_{Y \mid (X,H)}(x, H^m_s)\) as \(m \to \infty\), in probability under \(\mathbb{P}(\cdot \mid H = h)\). Let furthermore \(\hat{f}_{Y \mid X} = (\hat{f}_{Y \mid X}^m)_{m \in \mathbb{N}}\) be an estimator satisfying Assumption 2. Then we have the following consistency result. For all \(x \in \mathcal{X}\), \(\delta > 0\) and \(\alpha > 0\), there exists \(N \in \mathbb{N}\) such that for all \(n \geq N\) we can find \(M_n \in \mathbb{N}\) such that for all \(m \geq M_n\) we have that
\[
\mathbb{P}\left(\left|\hat{f}_{\text{AVE}(X \to Y)}(X^m_n, Y^m_n)(x) - f_{\text{AVE}(X \to Y)}(x)\right| > \delta\right) \leq \alpha. \tag{6}
\]

Apart from the LSCM structure, the above result does not rely on any particular distributional properties of the data. Assumptions 1 and 2 do not impose strong restrictions on the data-generating process and hold true for several model classes,
including linear and nonlinear models. In Appendix D, we provide sufficient conditions under which these assumptions are true.

### 2.3.2. An Example LSCM

To illustrate the consistency result in Theorem 1, we now consider a simple example with one covariate ($d = 1$) and two hidden variables ($\ell = 2$).

**Example 1 (Latent Gaussian process and a linear average causal effect).** Let $\zeta, \psi, \xi^1, \xi^2 \in \mathbb{N}$, be independent versions of a univariate stationary spatial Gaussian process with mean 0 and covariance function $u \mapsto \exp(-\frac{1}{2}|u|^2)$. For notational simplicity, let $\tilde{H}$ and $\hat{H}$ denote the respective first and second coordinate process of $H$. We define a marginal distribution over $\hat{H}$ and conditional distributions $X | \tilde{H}$ and $Y | (X, \tilde{H})$ by specifying that for all $(s, t) \in \mathbb{R}^2 \times \mathbb{N},$

$$H^s_t = (\tilde{H}^s_t, \hat{H}^s_t) = (\zeta, 1 + \frac{1}{2} \psi + \frac{\sqrt{2}}{2} \psi_1),$$

$$X^s_t = \exp(-|s|^2/(1000) + (0.2 + 0.1 \cdot \sin(2\pi t/100)) \cdot \hat{H}^s_t \cdot \hat{H}^s_t + 0.5 \cdot \xi^1_t),$$

$$Y^s_t = (1.5 + \tilde{H}^s_t \cdot \tilde{H}^s_t) \cdot X^s_t + (\tilde{H}^s_t)^2 + |\hat{H}^s_t| \cdot \epsilon^1_t.$$

Interventions on $X, Y, \text{ or } H$ are defined as in Definition 1. In this LSCM, the average causal effect $f_{\text{AVE}}(X \rightarrow Y)$ is the linear function $x \mapsto \beta_0 + \beta_1 x$, with $\beta_0 := \mathbb{E}(\tilde{H}^2) = 1$ and $\beta_1 := 1.5 + \mathbb{E}(\tilde{H}^2)$. We define a spatial sampling scheme $(s^i)_{i \in \mathbb{N}}$ for every $j \in \mathbb{N}$ and $k \in \{1, \ldots, 25\}$ by $s_{25(j-1)+k} = (j, k)$. Given a sample $(X^m, Y^m) = (X^i, Y^i)_{(i,k)\in\{1,\ldots,25\}\times\{1,\ldots,25\}}$ from $(X, Y)$, we construct an estimator of $f_{\text{AVE}}(X \rightarrow Y)$ by

$$\hat{f}^m_{\text{AVE}}(X \rightarrow Y)(X^m, Y^m) = \frac{1}{n} \sum_{i=1}^{n} (1 \cdot x) \hat{\beta}^m_{\text{OLS}}(X^m_{s^i}, Y^m_{s^i}),$$

where $\hat{\beta}^m_{\text{OLS}}(X^m_{s^i}, Y^m_{s^i}) \in \mathbb{R}^2$ is the OLS estimator for the linear regression at spatial location $s^i$, that is of $Y^m_{s^i} = (Y^1_{s^i}, \ldots, Y^n_{s^i})$ on $X^m_{s^i} = (X^1_{s^i}, \ldots, X^m_{s^i})$ (we assume that the regression includes an intercept term). It follows by Propositions 1 and 2 (see in particular Example 4 and Remark 6 in Appendix E) that Assumptions 1 and 2 are satisfied. Hence, (7) is a consistent estimator of $f_{\text{AVE}}(X \rightarrow Y)$.

Figure 3 illustrates our procedure using a numerical experiment based on Example 1. The example shows that we can estimate causal effects even under complex influences of the latent process $H$. To construct the estimator $\hat{f}^m_{\text{AVE}}(X \rightarrow Y)$, we have used the influence of $(X, H)$ on $Y$ linear in $X$. However, we do not assume knowledge of the particular functional dependence of $Y$ on $H$; we obtain consistency under any influence of the form $Y^m_{s^i} = f_1(H^s_t) \cdot X^s_t + f_2(H^s_t, \epsilon^1_t)$, see Proposition 2.

### 2.3.3. Extensions

Our methodology can be extended in several directions. In Appendix B, we provide details on how to model the influence of observed (time- and space-varying) confounders and temporally lagged casual effects of $X$ on $Y$. We furthermore argue that our method also applies in situations where the roles of time and space are interchanged, that is, when the hidden confounders are time-varying, but remain static across space. The extension which includes observed confounders is furthermore explored empirically in Appendix F.1.

### 3. Testing for the Existence of Causal Effects

The previous section has been concerned with the quantification and estimation of the causal effect of $X$ on $Y$. In this section, we introduce hypothesis tests for this effect being zero. We consider the null hypothesis

$$H_0 : (X, Y) \text{ come from an LSCM with a function } f \text{ that is constant w.r.t. } X^1,$$

strictly speaking, Example 4 and Remark 6 in Appendix E show that (L1)–(L3) are satisfied for bounded basis functions. We are confident that the same holds true in the current example.

In this example, a standard regression approach overestimates the causal effect. Similarly, one can construct examples where the causal effect would be underestimated.
which formalizes the assumption of “no causal effect of X on Y” within the LSCM framework. We construct a hypothesis test for $H_0$ using data resampling. Our approach acknowledges the existence of spatial dependence in the data without modeling it explicitly. It thus does not rely on distributional assumptions apart from the LSCM structure.

For the construction of a resampling test, we closely follow the setup presented in Pfister et al. (2018). We require a data permutation scheme which, under the null hypothesis, leaves the distribution of the data unaffected. In particular, it must preserve the dependence between the distribution of the data unaffected. In particular, it must preserve the dependence between $X_n^m$, $Y_n^m$ be the observed data. For every $(x, y) \in \mathbb{R}^{(d+1) \times m}$ and every permutation $\sigma$ of the elements in $\{1, \ldots, m\}$, let $\sigma(x, y) \in \mathbb{R}^{(d+1) \times m}$ denote the permuted array with entries $(\sigma(x, y))^i_j = (x^i_j, y^i_j)^{\sigma(i)}$. We then have the following exchangeability property.

**Proposition 3 (Exchangeability).** For any permutation $\sigma$ of the elements in $\{1, \ldots, m\}$, we have that, under $H_0$, $\sigma(X_n^m, Y_n^m)$ is equal in distribution to $(X_n^m, Y_n^m)$.

Proposition 3 is the cornerstone for the construction of a valid resampling test. Under the null hypothesis, we can compute pseudo-replications of the observed sample $(X_n^m, Y_n^m)$ using the permutation scheme described above. Given any test statistic $\hat{T}: \mathbb{R}^{(d+1) \times m} \rightarrow \mathbb{R}$, we obtain a $p$-value for $H_0$ by comparing the value of $\hat{T}$ calculated on the original dataset with the empirical null distribution of $\hat{T}$ obtained from the resampled datasets. The choice of $\hat{T}$ determines the power of the test. More formally, let $M := m^d$ and let $\sigma_1, \ldots, \sigma_M$ be all permutations of the elements in $\{1, \ldots, m\}$. By Proposition 3, each $\sigma_j$ yields a new dataset with the same distribution as $(X_n^m, Y_n^m)$. Let $B \in \mathbb{N}$ and let $k_1, \ldots, k_B$ be independent, uniform draws from $\{1, \ldots, M\}$. For every $(x, y)$, we define

$$p_T(x, y) := \frac{1 + \left| \{b \in \{1, \ldots, B\} : \hat{T}(\sigma_{k_b}(x, y)) \geq \hat{T}(x, y) \right|}{1 + B}$$

and construct for every $\alpha \in (0, 1)$ a test $\varphi_T^\alpha: \mathbb{R}^{(d+1) \times m} \rightarrow \{0, 1\}$ of $H_0$ defined by $\varphi_T^\alpha = 1 \Leftrightarrow p_T \leq \alpha$. The following level guarantee for $\varphi_T^\alpha$ follows directly from (Pfister et al. 2018, prop. B.4).

**Corollary 1 (Level guarantee of resampling test).** Assume that for all $k, \ell \in \{1, \ldots, B\}$, $k \neq \ell$, it holds that, under $H_0$, $P(\hat{T}(\sigma_k(X_n^m, Y_n^m)) = \hat{T}(\sigma_\ell(X_n^m, Y_n^m))) = 0$. Then, for every $\alpha \in (0, 1)$, the test $\varphi_T^\alpha$ has correct level $\alpha$.

Corollary 1 ensures valid test level for a large class of test statistics. The particular choice of test statistic should depend on the alternative hypothesis that we seek to have power against. Within the LSCM model class, it makes sense to quantify deviations from the null hypothesis using functionals of the average causal effect, that is, $T = \psi(f_{\text{AVE}(X \rightarrow Y)})$ for some suitable function $\psi$. As a test statistic, we then use the plug-in estimator $\hat{T}(X_n^m, Y_n^m) = \psi(f_{\text{AVE}(X \rightarrow Y)}^n(X_n^m, Y_n^m))$. An implementation of the above testing procedure is contained in our code package, see Section 1.4. A power-analysis can be found in Appendix F.3.

4. Conflict and Forest Loss in Colombia

We return to the problem of conflict (X) and forest loss (Y). As discussed in Section 1.3, there may exist several confounders of the (potential) causal effect of X on Y. We believe that many of these vary at a relatively low temporal frequency (e.g., population density, market infrastructure), justifying the application of the methods developed above.

4.1. Data Description and Preprocessing

Our analysis is based on two main datasets: (i) a remote sensing-based forest loss dataset for the period 2000–2018, which identifies annual forest loss at a spatial resolution of 30m × 30m using Landsat satellites (Hansen et al. 2013). Here, forest loss is defined as complete canopy removal. (ii) Spatially explicit information on conflict events from 2000 to 2018, based on the Georeferenced Event Dataset (GED) from the Uppsala Conflict Data Program (UCDP) (Croicu and Sundberg 2015). In this dataset, a conflict event is defined as “an incident where armed force was used by an organized actor against another organized actor, or against civilians, resulting in at least one direct death at a specific location and a specific date”. Such events were identified through global newswire reporting, global monitoring of local news, and other secondary sources such as reports by non-governmental organizations (for information on the data collection as well as control for quality and consistency of the data, please refer to Croicu and Sundberg 2015). We homogenized these datasets through aggregation to a spatial resolution of $10\,\text{km} \times 10\,\text{km}$ by averaging the annual forest loss within each grid, and by counting all conflict events occurring in the same year and within the same grid. As a proxy for accessibility, we additionally calculated, for each spatial grid, the average Euclidean distance to the closest road segment, using spatial data from https://diva-assis.org containing all primary and secondary roads in Colombia. We regard this variable as relatively constant throughout the considered time-span.

4.2. Quantifying the Causal Influence of Conflict on Forest Loss

We assume that $(X, Y)$ come from an LSCM as defined in Definition 2. Since $X^t_i$ is binary, we can characterize the causal influence of X on Y by $T := f_{\text{AVE}(X \rightarrow Y)}(1) - f_{\text{AVE}(X \rightarrow Y)}(0)$, that is, the difference in expected forest loss $E[Y|^t_i]$ under the respective interventions enforcing conflict ($X^t_i := 1$) and peace ($X^t_i := 0$). Positive values of $T$ correspond to an augmenting effect of conflict on forest loss and negative values correspond to a reducing effect. Our goal is to estimate $T$, and to test the hypothesis $H_0 : T = 0$ (no causal effect of X on Y). To construct an estimator of the average causal effect of the form (5), we require estimators of the conditional expectations $x \mapsto \psi(f_{\text{AVE}(X \rightarrow Y)}(x))$ for some suitable function $\psi$. As a test statistic, we then use the plug-in estimator $\hat{T}(X_n^m, Y_n^m) = \psi(f_{\text{AVE}(X \rightarrow Y)}^n(X_n^m, Y_n^m))$. An implementation of the above testing procedure is contained in our code package, see Section 1.4. A power-analysis can be found in Appendix F.3.
\textbf{Figure 4.} Testing for a causal influence of conflict (X) on forest loss (Y) using our method (right) and two alternative approaches (left and middle) which are based on different and arguably less realistic assumptions on the causal structure. The process W corresponds to the mean distance to a road, and H represents unobserved time-invariant confounders. Each of the above models gives rise to a different expression for the test statistic $T = \hat{f}_{\text{AVE}}(X \rightarrow Y)(1) - \hat{f}_{\text{AVE}}(X \rightarrow Y)(0)$ (indicated by red vertical bars), see Sections 4.2 and 4.3. The gray histograms illustrate the empirical null distributions of $T$ under the respective null hypotheses obtained from 999 resampled datasets. The results show that our testing procedure—we obtain a valid test level for any test statistic $T$ even if the above assumptions on the underlying LSCM, the above estimator is approximately unbiased. Note, however, that even if the above estimator is biased, this does not affect the level guarantee of our testing procedure—we obtain a valid test level for any test statistic $T$ satisfying the assumptions of Corollary 1.

\subsection*{4.3. Alternative Assumptions on the Causal Structure}

To emphasize the relevance of the assumed causal structure, we compare our method with two alternative approaches based on different assumptions about the grund truth: Model 1 assumes no confounders of (X, Y) and Model 2 assumes that the only confounder is the observed process W (mean distance to a road). In reality, we expect the existence of several confounders in addition to W, for example, population density or market infrastructure. Even though none of the models may be a precise description of the data-generating mechanism, we therefore regard both Models 1 and 2 as less realistic than the LSCM. In both models we can, similarly to Definition 3, define the average causal effect of X on Y. Under Model 1, $\hat{f}_{\text{AVE}}(X \rightarrow Y)$ coincides with the conditional expectation of $Y_i$ given $X_i$, which can be estimated simply using sample averages (as is done in Figure 1 left). Under Model 2, $\hat{f}_{\text{AVE}}(X \rightarrow Y)$ can, analogously to Proposition 2, be computed by adjusting for the confounder W. For each $x \in \{0,1\}$, we obtain an estimate $\hat{f}_{\text{AVE}}(X \rightarrow Y)(x)$ by calculating sample averages of Y across different subsets $(s, t) \in \mathbb{R}^2 \times \mathbb{N}: X_i = x, W_j \in W_j, j \in J$ (we here construct these by considering 100 equidistant quantiles of W), and subsequently averaging over the resulting values. In both models, we furthermore test the hypothesis of no causal effect of X on Y using approaches similar to the ones presented in Section 3. Under the LSCM assumption, we have constructed a permutation scheme that permutes the values of Y along the time axis, to preserve the dependence between Y and H, see Proposition 3. Similarly, we construct a permutation scheme for Model 2 by permuting observations of Y corresponding to similar values of the confounder W (i.e., values within the same quantile range). Under the null hypothesis corresponding to Model 1, X and Y are (conditionally) independent, and we therefore permute the values of Y completely at random. Strictly speaking, the permutation schemes for Models 1 and 2 require additional exchangeability assumptions on Y in order to yield valid tests. In Appendix G, we repeat the analysis for Model 1 using a spatial block-permutation to account for the spatial dependence in Y, and obtain similar results.
The results of applying our method and the two alternative approaches to the entire study region are depicted in Figure 4. Under Model 1, there is an enhancing, highly significant causal effect of conflict on forest loss ($\hat{T} = 0.073, P = 0.002$). When adjusting for accessibility (quantified by $W$, Model 2), the size of the estimated causal effect shrinks, and becomes insignificant ($\hat{T} = 0.049, P = 0.168$). (Note that we have considered other confounders, too, yet obtained similar results. For example, when adjusting for population density, which we consider as moderately temporally varying, we obtain $\hat{T} = 0.038$ and $P = 0.214$.) When applying the methodology proposed in this article, that is, adjusting for all time-invariant confounders, the estimated effect swaps sign ($\hat{T} = -0.018, P = 0.578$), but is insignificant. We also tested for spatial spill-over effects by modeling causal effects that are spatially lagged by up to one or two pixels, respectively. In both cases, the effect was negative and insignificant (results not shown here). One reason for the above non-finding could be the time delay between the proposed cause (conflict) and effect (forest loss). To account for this potential issue, we also test for a causal effect of $X$ on $Y$ that is temporally lagged by one year, that is, we use an estimator similar to (8), where we compare the average forest loss succeeding conflict events with the average forest loss succeeding non-conflict events. Again, the estimated influence of $X$ on $Y$ is negative and insignificant ($\hat{T} = -0.0293, P = 0.354$). Additionally, we perform alternative versions of the last two tests to account for potential autocorrelation in the response variable, by adopting a temporal block-permutation scheme. In both cases, the test is insignificant, see Appendix G.

The above analysis provides an estimate for the average causal effect, see Equation (3), which, in particular, averages over space. Given that Colombia is a country with high ambiental and socio-economic heterogeneity, we furthermore conduct an analysis at the department level (see Figure 5). In fact, there is considerable spatial variation in the estimated causal effects, with significant positive as well as negative effects (Figure 5 middle). This variation may be seen as evidence for an interaction effect between conflict and the assumed hidden founders. In most departments, the estimated causal effect is negative (although mostly insignificant), meaning that conflict tends to decrease forest loss. The strongest positive and significant causal influence is identified in the La Guajira department ($\hat{T} = 0.398, P = 0.047$). Although this region is commonly associated with semi-arid to very dry conditions, most conflicts occurred in the South-Western areas, at the beginning of Caribbean tropical forests (see Figure 1). In fact, these zones have also been also identified by Negret et al. (2019) as having been strongly affected by deforestation pressure in the wake of conflict. Interestingly, the neighboring Magdalena department shows the opposite effect ($\hat{T} = -0.218, P = 0.004$). The positive effect in the department of Huila ($\hat{T} = 0.095, P = 0.023$) is again in line with the findings by Negret et al. (2019) (based on a visual inspection of their attribution maps). Out of the 8 departments that are mostly controlled by FARC (Figure 5 right), 6 have a negative test statistic, meaning that conflict reduces forest loss. This can be explained in part by the internal governance of this group, where forest cover was a strategic advantage for both their own protection as well as for cocaine production. Overall, of course, the peace-induced acceleration of forest loss has to be discussed with caution, and should not be interpreted reversely as if conflict per se is a measure of environmental protection (Clerici et al. 2020).

### 4.5. Interpretation in Light of the Colombian Peace Process

In late 2012, negotiations that later would be known as the “Colombian peace process” started between the then president of Colombia and the strongest group of rebels, the FARC, lasting until 2016. Peace was declared by both parties upon a revised agreement in October 2016, and became effective in 2017. The agreement was signed only by the FARC, while other guerilla groups remain active.
negotiations marked a steadily decreasing number of conflicts (Figure 6, left). Since this decrease is the result of governmental intervention, rather than a natural resolution of local tensions, the peace process provides an opportunity to verify the intervention effects estimated in Section 4.4. As can be seen in Figure 6 (right), Colombia experienced a steep increase in the total forest loss in the final phase of the peace negotiations. Although there may be several other factors which have contributed to this development, we observe that these results align with our previous finding of an overall negative causal effect of conflict on forest loss ($\hat{T} < 0$).

5. Conclusions and Future Work

In this article, we have investigated the relationship between conflict and forest loss in Colombia based on yearly measurements from 2000 to 2018. Our methodology allows for a causal rather than a purely predictive analysis. We find that conflicts are predictive of exceedances in forest loss, but we find no significant evidence of a causal relation when analyzing this problem on country level: once all (time-invariant) confounders are adjusted for, there is a negative but insignificant correlation between conflict and forest loss. Our analysis on department level suggests that this nonfinding could be due to locally varying effects of opposite directionality, which would approximately cancel out in our final estimate. In most departments, we find negative (mostly insignificant) effects, although we also identify a few departments where conflict seems to have increased forest loss. It would be interesting to furthermore explore these differences in causal dependence, for example, by adjusting for socio-political factors. The overall negative influence of conflict on forest loss estimated by our method is in line with the observation that during the final phase of the peace process, which stopped many of the existing conflicts, the total forest loss in Colombia has increased. However, these results should be interpreted with caution. Overall, we find that, once all time-invariant confounders are adjusted for, conflicts have only weak explanatory power for predicting forest loss, and the potential causal effect is therefore likely to be small, compared to other drivers of forest loss. Furthermore, our results rely on the assumption that all latent confounders are time-invariant, which, in practice, can only be expected to hold approximately. In Appendix E.1, we explore how our estimator is affected by temporal variation in the latent confounders.

To analyze the relation between conflict and forest loss, we have introduced a novel causal framework for spatio-temporal data. From a methodological perspective, this article contains three main contributions: the definition of a class of causal models for multivariate spatio-temporal stochastic processes, a procedure for estimating causal effects within this model class, and a nonparametric hypothesis test for the overall existence of such effects. Our method allows for the influence of arbitrarily many latent confounders, as long as these confounders do not vary across time (or space). We prove asymptotic consistency of our estimator, and verify this finding empirically using simulated data. Our results hold under weak assumptions on the data generating process, and do not rely on any particular distributional properties of the data. We prove sufficient conditions under which these rather general assumptions hold true. The proposed testing procedure is based on data resampling and provably obtains valid level in finite samples.

Supplementary Material

The supplementary material is structured as follows. Appendix A contains a review of existing causal methodology, Appendix B explores several extensions of our methods, and Appendix C proposes an alternative definition of the average causal effect, our inferential target. In Appendix D, we provide sufficient conditions for Assumptions A1 and A2, and Appendix E provides two examples of data generating processes which satisfy these assumptions. Appendix F contains several simulation experiments which investigate the numerical performance of our methods, Appendix G shows results of the case study where we use adapted versions of our testing procedure, and Appendix H contains all proofs.

Acknowledgments

This work contributes to the Global Land Programme, and was supported by Carlsberg Foundation (JP) and a research grant (18968) from VILLUM FONDEN (RC, JP). MDM thanks ESA for funding the Earth System Data Lab. We are grateful to Steffen Lauritzen, Niels Richard Hansen and Niklas Pfister for valuable comments on the methodology, and to Lina Estupinan-Suarez and Alejandro Salazar for helpful discussions on the case study. We thank two anonymous referees and the associate editor for many helpful comments.

ORCID

Rune Christiansen http://orcid.org/0000-0002-8202-1176
Tobias Kuemmerle http://orcid.org/0000-0002-9775-142X
Jonas Peters https://orcid.org/0000-0002-1487-7511
Miguel D. Mahecha https://orcid.org/0000-0003-3031-613X
Matthias Baumann https://orcid.org/0000-0003-2375-3622