Scalable Sensitivity and Uncertainty Analyses for Causal-Effect Estimates of Continuous-Valued Interventions

Anonymous Author(s)
Affiliation
Address
email

Abstract

Estimating the effects of continuous-valued interventions from observational data is a critically important task for climate science, healthcare, and economics. Recent work focuses on designing neural network architectures and regularization functions to allow for scalable estimation of average and individual-level dose-response curves from high-dimensional, large-sample data. Such methodologies assume ignorability (observation of all confounding variables) and positivity (observation of all treatment levels for every covariate value describing a set of units), assumptions problematic in the continuous treatment regime. Scalable sensitivity and uncertainty analyses to understand the ignorance induced in causal estimates when these assumptions are relaxed are less studied. Here, we develop a continuous treatment-effect marginal sensitivity model (CMSM) and derive bounds that agree with the observed data and a researcher-defined level of hidden confounding. We introduce a scalable algorithm and uncertainty-aware deep models to derive and estimate these bounds for high-dimensional, large-sample observational data. We work in concert with climate scientists interested in the climatological impacts of human emissions on cloud properties using satellite observations from the past 15 years. This problem is known to be complicated by many unobserved confounders.

1 Introduction

Understanding the causal effect of a continuous variable (termed "treatment") on individual units and subgroups is crucial across many fields. In economics, we might like to know the effect of price on demand from different customer demographics. In healthcare, we might like to know the effect of medication dosage on health outcomes for patients of various ages and comorbidities. And in climate science, we might like to know the effects of anthropogenic emissions on cloud formation and lifetimes under variable atmospheric conditions. In many cases, these effects must be estimated from observational data as experiments are often costly, unethical, or otherwise impossible to conduct.

Estimating causal effects from observational data can only be done under certain conditions, some of which are not testable from data. The most prominent are the common assumptions that all confounders between treatment and outcome are measured ("no hidden confounders"), and any level of treatment could occur for any observable covariate vector ("positivity"). These assumptions and their possible violations introduce uncertainty when estimating treatment effects. Estimating this uncertainty is crucial for decision-making and scientific understanding. For example, understanding how unmeasured confounding can change estimates about the impact of emissions on cloud properties can help to modify global warming projection models to account for the uncertainty it induces.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

We present a novel marginal sensitivity model for continuous treatment effects. This model is used to develop a method that gives the user a corresponding interval representing the "ignorance region" of the possible treatment outcomes per covariate and treatment level [D'A19] for a specified level of violation of the no-hidden confounding assumption. We adapt prior work [Tan06, KMZ19, JMGS21] to the technical challenge presented by continuous treatments. Specifically, we modify the existing model to work with propensity score densities instead of propensity score probabilities (see Section 3 below) and propose a method to relate ignorability violations to the unexplained range of outcomes. Further, we derive bootstrapped uncertainty intervals for the estimated ignorance regions and show how to efficiently compute the intervals, thus providing a method for quantifying the uncertainty presented by finite data and possible violations of the positivity assumption. We validate our methods on synthetic data and provide an application on real-world satellite observations of the effects of anthropogenic emissions on cloud properties. For this application, we develop a new neural network architecture for estimating continuous treatment effects that can take into account spatiotemporal covariates. We find that the model accurately captures known patterns of cloud deepening in response to anthropogenic emission loading with realistic intervals of uncertainty due to unmodeled confounders in the satellite data.

2 Problem Setting

Let the random variable $\mathbf{X} \in \mathcal{X}$ model observable covariates. For clarity, we will assume that $\mathcal{X}$ is a d-dimensional continuous space: $\mathcal{X} \subseteq \mathbb{R}^d$, but this does not preclude more diverse spaces. Instances of $\mathbf{X}$ are denoted by $\mathbf{x}$. The observable continuous treatment variable is modeled as the random variable $\mathrm{T} \in \mathcal{T} \subseteq \mathbb{R}$. Instances of T are denoted by t. Let the random variable $\mathrm{Y} \in \mathcal{Y} \subseteq \mathbb{R}$ model the observable continuous outcome variable. Instances of Y are denoted by y. Using the Neyman-Rubin potential outcomes framework [Ney23, Rub74, Sek08], we model the potential outcome of a treatment level t by the random variable $\mathrm{Y_t} \in \mathcal{Y}$. Instances of $\mathrm{Y_t}$ are denoted by $\mathrm{y_t}$. We assume that the observational data, $\mathcal{D}_n$, consists of n realizations of the random variables, $\mathcal{D}_n = \{(\mathbf{x}_i, \mathrm{t}_i, \mathrm{y}_i)\}_{i=1}^n$. We let the observed outcome be the potential outcome of the assigned treatment level, $\mathrm{y}_i = \mathrm{y}_{\mathrm{t}_i}$, thus assuming non-interference and consistency [Rub80]. Moreover, we assume that the tuple $(\mathbf{x}_i, \mathrm{t}_i, \mathrm{y}_i)$ are i.i.d. samples from the joint distribution $P(\mathbf{X}, \mathrm{T}, \mathrm{Y_T})$, where $\mathrm{Y_T} = \{\mathrm{Y_t} : \mathrm{t} \in \mathcal{T}\}$.

We are interested in the **conditional average potential outcome (CAPO)** function, $\mu(\mathbf{x}, \mathrm{t})$, and the **average potential outcome (APO)** — or dose-response function — $\mu(\mathrm{t})$, for continuous valued treatments. These functions are defined by the expectations:

$$\mu(\mathbf{x}, \mathrm{t}) \coloneqq \mathbb{E}\left[\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x}\right] \quad (1) \qquad \mu(\mathrm{t}) \coloneqq \mathbb{E}\left[\mu(\mathbf{X}, \mathrm{t})\right]. \quad (2)$$

Under the assumptions of ignorability, $\mathrm{Y_T} \perp\!\!\!\perp \mathrm{T} \mid \mathbf{X}$, and positivity, $p(\mathrm{t} \mid \mathbf{X} = \mathbf{x}) > 0 : \forall \mathrm{t} \in \mathcal{T}, \forall \mathbf{x} \in \mathcal{X}$ — jointly known as *strong ignorability* [RR83] — the CAPO and APO are identifiable from the observational distribution $P(\mathbf{X}, \mathrm{T}, \mathrm{Y_T})$ as:

$$\widetilde{\mu}(\mathbf{x}, \mathrm{t}) = \mathbb{E}\left[\mathrm{Y} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x}\right] \quad (3) \qquad \widetilde{\mu}(\mathrm{t}) = \mathbb{E}\left[\widetilde{\mu}(\mathbf{X}, \mathrm{t})\right]. \quad (4)$$

In practice, however, these assumptions rarely hold. For example, there will almost always be unobserved confounding variables, thus violating the ignorability (also known as unconfoundedness or exogeneity) assumption, $\mathrm{Y_T} \not\perp\!\!\!\perp \mathrm{T} \mid \mathbf{X}$. Moreover, due to both the finite sample of observed data, $\mathcal{D}$, and also the continuity of treatment T, there will most certainly be values, $\mathrm{T} = \mathrm{t}$, that are unobserved for a given covariate measurement, $\mathbf{X} = \mathbf{x}$, leading to violations or near violations of the positivity assumption (also known as overlap).

3 Methods

We propose the continuous marginal sensitivity model (CMSM) as a new marginal sensitivity model (MSM [Tan06]) for continuous treatment variables. The set of conditional distributions of the potential outcomes given the observed treatment assigned, $\{P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x}) : \mathrm{t} \in \mathcal{T}\}$, are identifiable from data, $\mathcal{D}$. But, the set of marginal distributions of the potential outcomes, $\{P(\mathrm{Y_t} \mid, \mathbf{X} = \mathbf{x}) : \mathrm{t} \in \mathcal{T}\}$, each given as a continuous mixture,

$$P(\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x}) = \int_{\mathcal{T}} p(\mathrm{t}' \mid \mathbf{x}) P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}', \mathbf{X} = \mathbf{x}) d\mathrm{t}',$$

are not. This is due to the general unidentifiability of the component distributions, $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}', \mathbf{X} = \mathbf{x})$, where $\mathrm{Y_t}$ cannot be observed for units under treatment level $\mathrm{T} = \mathrm{t}'$ for $\mathrm{t}' \neq \mathrm{t}$: the well-known "fundamental problem of causal inference" [Hol86]. Yet, under the ignorability assumption, the factual $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$ and counterfactual $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}', \mathbf{X} = \mathbf{x})$ are equal for all $\mathrm{t}' \in \mathcal{T}$. Thus, $P(\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x})$ and $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$ are identical, and any divergence between them is indicative of hidden confounding. But, such divergence is not observable in practice.

The CMSM supposes a degree of divergence between the unidentifiable $P(\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x})$ and the identifiable $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$ by assuming that the rate of change of $P(\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x})$ with respect to $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$ is bounded by some value greater than or equal to 1. The Radon-Nikodym derivative formulates the divergence, $\lambda(\mathrm{y_t}; \mathbf{x}, \mathrm{t}) = \frac{dP(\mathrm{Y_t} \mid \mathbf{X}=\mathbf{x})}{dP(\mathrm{Y_t} \mid \mathrm{T}=\mathrm{t}, \mathbf{X}=\mathbf{x})}$, under the assumption that $P(\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x})$ is absolutely continuous with respect to $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$, $\forall \mathrm{t} \in \mathcal{T}$.

Proposition 1. *Under the additional assumption that $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$ is equivalent to the Lebesgue measure, the Radon-Nikodym derivative above is equal to the ratio between the unidentifiable "complete" propensity density for treatment $p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x})$ and the identifiable "nominal" propensity density for treatment $p(\mathrm{t} \mid \mathbf{x})$,*

$$\lambda(\mathrm{y_t}; \mathbf{x}, \mathrm{t}) = \frac{p(\mathrm{t} \mid \mathbf{x})}{p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x})}, \tag{5}$$

Proof (Appendix A.3) and an analysis of this proposition are given in Appendix A.

The value $\lambda(\mathrm{y_t}; \mathbf{x}, \mathrm{t})$ cannot be identified from the observational data alone; the merit of the CMSM is that enables a domain expert to express their belief in what is a plausible degree hidden confounding through the parameter $\Lambda \geq 1$. Where, $\Lambda^{-1} \leq p(\mathrm{t} \mid \mathbf{x})/p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x}) \leq \Lambda$, reflects a hypothesis that the "complete", unidentifiable propensity density for subjects with covariates $\mathbf{X} = \mathbf{x}$ can be different from the observable "nominal" propensity density by at most a factor of Λ. These inequalities allow us to bound inverse of the unidentifiable complete propensity density in terms of functions of the inverse of the identifiable nominal propensity density:

$$\frac{1}{\Lambda p(\mathrm{t} \mid \mathbf{x})} \leq \frac{1}{p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x})} \leq \frac{\Lambda}{p(\mathrm{t} \mid \mathbf{x})}. \tag{6}$$

Remark. Note that the CMSM is defined in terms of a *density ratio*, $p(\mathrm{t} \mid \mathbf{x})/p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x})$, whereas the MSM for binary-valued treatments is defined in terms of an *odds ratio*, $\frac{P(\mathrm{t}|\mathbf{x})}{(1-P(\mathrm{t}|\mathbf{x}))} / \frac{P(\mathrm{t}|\mathrm{y_t},\mathbf{x})}{(1-P(\mathrm{t}|\mathrm{y_t},\mathbf{x}))}$. Importantly, naively substituting densities into the MSM for binary-treatments would violate the condition that $\lambda > 0$ as the densities $p(\mathrm{t} \mid \mathbf{x})$ or $p(\mathrm{t} \mid \mathrm{y_t}, \mathbf{x})$ can each be greater than one, which would result in a negative $1 - p(\mathrm{t} \mid \cdot)$. The odds ratio is familiar to practitioners. The density ratio is less so. We offer a transformation of the sensitivity analysis parameter Λ in terms of the unexplained range of the outcome later.

3.1 Continuous Treatment Effect Bounds Without Ignorability

The CAPO and APO (dose-response) functions cannot be point identified from observational data without ignorability. Under the CMSM with a given Λ, we can only identify a set of CAPO and APO functions jointly consistent with the observational data $\mathcal{D}$ and the continuous marginal sensitivity model. All of the functions in this set are possible from the point of view of the observational data alone. So to cover the range of all possible functional values, we seek an interval function that maps covariate values, $\mathbf{X} = \mathbf{x}$, to the upper and lower bounds of this set for every treatment value, t.

For $\mathrm{t} \in \mathcal{T}$ and $\mathbf{x} \in \mathcal{X}$, let $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x})$ denote the density of the distribution $P(\mathrm{Y_t} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x})$. As a reminder, this distribution is identifiable from observational data, but without further assumptions the CAPO, $\mu(\mathbf{x}, \mathrm{t}) = \mathbb{E}[\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x}]$, is not. We can express the CAPO in terms of its identifiable and unidentifiable components as

$$\begin{aligned} \mu(\mathbf{x}, \mathrm{t}) &= \frac{\int_{\mathcal{Y}} \mathrm{y_t} \frac{p(\mathrm{y_t}|\mathrm{t},\mathbf{x})}{p(\mathrm{t}|\mathrm{y_t},\mathbf{x})} d\mathrm{y_t}}{\int_{\mathcal{Y}} \frac{p(\mathrm{y_t}|\mathrm{t},\mathbf{x})}{p(\mathrm{t}|\mathrm{y_t},\mathbf{x})} d\mathrm{y_t}} = \widetilde{\mu}(\mathbf{x}, \mathrm{t}) + \frac{\int_{\mathcal{Y}} w(\mathrm{y}, \mathbf{x})(\mathrm{y} - \widetilde{\mu}(\mathbf{x}, \mathrm{t}))p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}) d\mathrm{y}}{(\Lambda^2 - 1)^{-1} + \int_{\mathcal{Y}} w(\mathrm{y}, \mathbf{x}) p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}) d\mathrm{y}}, \\ &\equiv \mu(w(\mathrm{y}, \mathbf{x}); \mathbf{x}, \mathrm{t}, \Lambda) \end{aligned} \tag{7}$$

where, by a one-to-one change of variables, $\frac{1}{p(\mathrm{t}|\mathrm{y_t},\mathbf{x})} = \frac{1}{\Lambda p(\mathrm{t}|\mathbf{x})} + w(\mathrm{y},\mathbf{x})(\frac{\Lambda}{p(\mathrm{t}|\mathbf{x})} - \frac{1}{\Lambda p(\mathrm{t}|\mathbf{x})})$ with $w : \mathcal{Y} \times \mathcal{X} \to [0,1]$. Both [KMZ19] and later [JMGS21] provide analogous expressions for the CAPO in the discrete treatment regime under the MSM, and we provide our derivation in Lemma 1.

The uncertainty set that includes all possible values of $w(\mathrm{y},\mathbf{x})$ that agree with the CMSM, *i.e.*, the set of functions that violate ignorability by no more than Λ, can now be expressed as $\mathcal{W} = \{w : w(\mathrm{y},\mathbf{x}) \in [0,1] \quad \forall \mathrm{y} \in \mathcal{Y}, \forall \mathbf{x} \in \mathcal{X}\}$.

With this set of functions, we can now define the CAPO and APO bounds under the CMSM. The CAPO lower, $\underline{\mu}(\mathbf{x},\mathrm{t};\Lambda)$, and upper, $\overline{\mu}(\mathbf{x},\mathrm{t};\Lambda)$, bounds under the CMSM with parameter Λ are:

$$\begin{aligned}\underline{\mu}(\mathbf{x},\mathrm{t};\Lambda) &\coloneqq \inf_{w\in\mathcal{W}} \mu(w(\mathrm{y},\mathbf{x});\mathbf{x},\mathrm{t},\Lambda) \\ &= \inf_{w\in\mathcal{W}^H_{\text{ni}}} \mu(w(\mathrm{y});\mathbf{x},\mathrm{t},\Lambda)\end{aligned} \tag{8}$$

$$\begin{aligned}\overline{\mu}(\mathbf{x},\mathrm{t};\Lambda) &\coloneqq \sup_{w\in\mathcal{W}} \mu(w(\mathrm{y},\mathbf{x});\mathbf{x},\mathrm{t},\Lambda) \\ &= \sup_{w\in\mathcal{W}^H_{\text{nd}}} \mu(w(\mathrm{y});\mathbf{x},\mathrm{t},\Lambda)\end{aligned} \tag{9}$$

Where the sets $\mathcal{W}^H_{\text{ni}} = \{w : w(y) = H(\mathrm{y}_H - \mathrm{y})\}_{\mathrm{y}_H\in\mathcal{Y}}$, and $\mathcal{W}^H_{\text{nd}} = \{w : w(y) = H(\mathrm{y} - \mathrm{y}_H)\}_{\mathrm{y}_H\in\mathcal{Y}}$, and $H(\cdot)$ is the Heaviside step function. Lemma 2 in appendix D proves the equivalence in eq. (9) for bounded Y. The equivalence in eq. (8) can be proved analogously.

The APO lower, $\underline{\mu}(\mathrm{t};\Lambda)$, and upper, $\overline{\mu}(\mathrm{t};\Lambda)$, bounds under the CMSM with parameter Λ are:

$$\underline{\mu}(\mathrm{t};\Lambda) \coloneqq \mathbb{E}\left[\underline{\mu}(\mathbf{X},\mathrm{t};\Lambda)\right] \tag{10}$$

$$\overline{\mu}(\mathrm{t};\Lambda) \coloneqq \mathbb{E}\left[\overline{\mu}(\mathbf{X},\mathrm{t};\Lambda)\right] \tag{11}$$

Remark. It is worth pausing here and breaking down Equation (7) to get an intuitive sense of how the specification of Λ in the CMSM affects the bounds on the causal estimands. When $\Lambda \to 1$, then the $(\Lambda^2-1)^{-1}$ term (and thus the denominator) in Equation (7) tends to infinity. As a result, the CAPO under Λ converges to the empirical estimate of the CAPO — $\mu(w(\mathrm{y});\mathbf{x},\mathrm{t},\Lambda\to 1) \to \widetilde{\mu}(\mathbf{x},\mathrm{t})$ — as expected. Thus, the supremum and infimum in Equations (8) and (9) become independent of w, and the ignorance intervals concentrate on point estimates. Next, consider complete relaxation of the ignorability assumption, $\Lambda \to \infty$. Then, the $(\Lambda^2-1)^{-1}$ term tends to zero, and we are left with,

$$\mu(w;\cdot,\Lambda\to\infty) \to \widetilde{\mu}(\mathbf{x},\mathrm{t}) + \frac{\int_{\mathcal{Y}} w(\mathrm{y})(\mathrm{y}-\widetilde{\mu}(\mathbf{x},\mathrm{t}))p(\mathrm{y}\mid\mathrm{t},\mathbf{x})d\mathrm{y}}{\int_{\mathcal{Y}} w(\mathrm{y})p(\mathrm{y}\mid\mathrm{t},\mathbf{x})d\mathrm{y},} = \widetilde{\mu}(\mathbf{x},\mathrm{t}) + \mathop{\mathbb{E}}_{p(w(\mathrm{y})|\mathbf{x},\mathrm{t})}[\mathrm{Y}-\widetilde{\mu}(\mathbf{x},\mathrm{t})],$$

where, $p(w(\mathrm{y})\mid\mathbf{x},\mathrm{t}) \equiv \frac{w(\mathrm{y})p(\mathrm{y}|\mathrm{t},\mathbf{x})}{\int_{\mathcal{Y}} w(\mathrm{y}')p(\mathrm{y}'|\mathrm{t},\mathbf{x})d\mathrm{y}'}$, a distribution over Y given $\mathbf{X}=x$ and $\mathrm{T}=\mathrm{t}$. Thus, when we *relax* the ignorability assumption entirely, the CAPO can be anywhere in the range of Y.

The parameter Λ relates to the proportion of unexplained range in Y assumed to come from unobserved confounders after observing $\mathbf{x}$ and t. When a user sets Λ to 1, they assume that the entire unexplained range of Y comes from unknown mechanisms independent of T. As the user increases Λ, they attribute some of the unexplained range of Y to mechanisms causally connected to T. For bounded $\mathrm{Y_t}$, this proportion can be calculated as:

$$\rho(\mathbf{x},\mathrm{t};\Lambda) \coloneqq \frac{\overline{\mu}(\mathbf{x},\mathrm{t};\Lambda) - \underline{\mu}(\mathbf{x},\mathrm{t};\Lambda)}{\overline{\mu}(\mathbf{x},\mathrm{t};\Lambda\to\infty) - \underline{\mu}(\mathbf{x},\mathrm{t};\Lambda\to\infty)} = \frac{\overline{\mu}(\mathbf{x},\mathrm{t};\Lambda) - \underline{\mu}(\mathbf{x},\mathrm{t};\Lambda)}{\mathrm{y_{max}} - \mathrm{y_{min}} \mid \mathbf{X}=x,\mathrm{T}=\mathrm{t}}.$$

The user can sweep over a set of Λ values and report the bounds corresponding to a ρ value they deem tolerable (e.g., $\rho = 0.5$ yields bounds for the assumption that half the unexplained range in Y is due to unobserved confounders).

For another way to interpret Λ, in Appendix A.3.1 we Λ can be presented as a bound on the Kullback–Leibler divergence between the nominal and complete propensity scores through the relationship: $|\log(\Lambda)| \geq D_{\mathrm{KL}}\left(P(\mathrm{Y_t}\mid\mathrm{T}=\mathrm{t},\mathbf{X}=\mathbf{x})||P(\mathrm{Y_t}\mid\mathbf{X}=\mathbf{x})\right)$.

3.2 Semi-Parametric Interval Estimator

Following [JMGS21], we develop a semi-parametric estimator of the bounds in eqs. (8) to (11). Under assumption Λ, the bounds on the expected potential outcome over $\mu(w(\mathrm{y});\mathbf{x},\mathrm{t},\Lambda)$ are completely defined in terms of identifiable quantities: namely, the conditional density of the outcome given the assigned treatment and measured covariates, $p(\mathrm{y}\mid\mathrm{t},\mathbf{x})$; and the conditional expected outcome $\widetilde{\mu}(\mathbf{x},\mathrm{t})$. Thus, we define a density estimator, $p(\mathrm{y}\mid\mathrm{t},\mathbf{x},\boldsymbol{\theta})$, and estimator, $\mu(\mathbf{x},\mathrm{t};\boldsymbol{\theta})$, parameterized

by instances $\boldsymbol{\theta}$ of the random variable $\boldsymbol{\Theta}$. The choice of density estimator is ultimately up to the user and will depend on the scale of the problem examined and the distribution of the outcome variable Y. In Section 3.5, we will outline how to define appropriate density estimators for high-dimensional, large-sample, continuous-valued treatment problems.

Next, we need an estimator of the integrals in $\mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta})$, eq. (7). We use Monte-Carlo (MC) integration to estimate the expectation of arbitrary functions $h(y)$ with respect to the parametric density estimate $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta})$: $I(h(\mathrm{y})) \coloneqq \frac{1}{m}\sum_{i=1}^{m} h(\mathrm{y}_i), \quad \mathrm{y}_i \sim p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta})$. We outline how the Gauss-Hermite quadrature rule is an alternate estimator of these expectations in Appendix C. The integral estimators allow for the semi-parametric estimators for the CAPO and APO bounds under the CMSM to be defined.

The semi-parametric CAPO bound estimators under the CMSM with sensitivity parameter Λ are:

$$\underline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \inf_{w \in \mathcal{W}_{\mathrm{ni}}^{H}} \mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}) \quad (12) \qquad \overline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \sup_{w \in \mathcal{W}_{\mathrm{nd}}^{H}} \mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}) \quad (13)$$

where,

$$\mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}) \equiv \widetilde{\mu}(\mathbf{x}, \mathrm{t}; \boldsymbol{\theta}) + \frac{I\left(w(\mathrm{y})(\mathrm{y} - \widetilde{\mu}(\mathbf{x}, \mathrm{t}; \boldsymbol{\theta}))\right)}{(\Lambda^2 - 1)^{-1} + I\left(w(\mathrm{y})\right)}.$$

The semi-parametric APO bound estimators under the CMSM with sensitivity parameter Λ are:

$$\underline{\mu}(\mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \mathbb{E}\left[\underline{\mu}(\mathbf{X}, \mathrm{t}; \Lambda, \boldsymbol{\theta})\right] \quad (14) \qquad \overline{\mu}(\mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \mathbb{E}\left[\overline{\mu}(\mathbf{X}, \mathrm{t}; \Lambda, \boldsymbol{\theta})\right] \quad (15)$$

Theorem 1. *In the limit of data ($n \to \infty$) and MC samples ($m \to \infty$), for observed $(\mathbf{X} = \mathbf{x}, \mathrm{T} = \mathrm{t}) \in \mathcal{D}_n$, we assume that $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta})$ converges in measure to $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x})$, $\widetilde{\mu}(\mathbf{x}, \mathrm{t}; \boldsymbol{\theta})$ is a consistent estimator of $\widetilde{\mu}(\mathbf{x}, \mathrm{t})$, and $p(\mathrm{t} \mid \mathrm{y}_{\mathrm{t}}, \mathbf{x})$ is bounded away from 0 uniformly for all $\mathrm{y}_{\mathrm{t}} \in \mathcal{Y}$. Then, $\underline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \xrightarrow{p} \underline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda)$ and $\overline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \xrightarrow{p} \overline{\mu}(\mathbf{x}, \mathrm{t}; \Lambda)$. Proof in Appendix E.*

3.3 Solving for w

We are interested in a scalable algorithm to compute the intervals on the CAPO function, eqs. (12) and (13), and the APO (dose-response) function, eqs. (14) and (15). The need for scalability stems not only from dataset size. The intervals also need to be evaluated for arbitrarily many values of the continuous treatment variable, t, and the sensitivity parameter Λ. The bounds on the CAPO function can be calculated independently for each instance $\mathbf{x}$, and the limits on the APO are an expectation over the CAPO function bounds.

The upper and lower bounds of the CAPO function under treatment, t, and sensitivity parameter, Λ, can be estimated for any observed covariate value, $\mathbf{x}$, as

$$\widehat{\underline{\mu}}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \mu(H(\underline{\mathrm{y}} - \mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}),$$

and

$$\widehat{\overline{\mu}}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \mu(H(\mathrm{y} - \overline{\mathrm{y}}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}),$$

where $\underline{\mathrm{y}}$ and $\overline{\mathrm{y}}$ are found using Algorithm 1. See Algorithm 2 and Appendix F for alternative approaches.

Algorithm 1 Grid Search Interval Optimizer

Require: $\mathbf{x}$ is an instance of $\mathbf{X}$, t is a treatment level to evaluate, Λ is a belief in the amount of hidden confounding, $\boldsymbol{\theta}$ are optimized model parameters, $\widehat{\mathcal{Y}}$ is a set of unique values $\{\mathrm{y} \sim p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta})\}$.

```
function GRIDSEARCH(x, t, Λ, θ, Ŷ)
    μ̄ ← −∞, ȳ ← 0
    μ̲ ← ∞, y̲ ← 0
    for y_H ∈ Ŷ do
        κ̄ ← μ(H(y − y_H); x, t, Λ, θ)
        κ̲ ← μ(H(y_H − y); x, t, Λ, θ)
        if κ̄ > μ̄ then
            μ̄ ← κ̄, ȳ ← y_H
        if κ̲ < μ̲ then
            μ̲ ← κ̲, y̲ ← y_H
    return y̲, ȳ
```

The upper and lower bounds for the APO (dose-response) function under treatment $\mathrm{T} = \mathrm{t}$ and sensitivity parameter Λ can be estimated over any set of observed covariates $\mathcal{D}_{\mathbf{x}} = \{\mathbf{x}_i\}_{i=1}^{n}$, as

$$\widehat{\underline{\mu}}(\mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \widehat{\underline{\mu}}(\mathbf{x}_i, \mathrm{t}; \Lambda, \boldsymbol{\theta}), \qquad \widehat{\overline{\mu}}(\mathrm{t}; \Lambda, \boldsymbol{\theta}) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \widehat{\overline{\mu}}(\mathbf{x}_i, \mathrm{t}; \Lambda, \boldsymbol{\theta}), \qquad \mathbf{x}_i \in \mathcal{D}_{\mathbf{x}}.$$

3.4 Uncertainty about the Continuous Treatment Effect Interval

Following [ZSB19], [DG21], and [CCN+21], we construct $(1 - \alpha)$ statistical confidence intervals for the upper and lower bounds under the CMSM using the percentile bootstrap estimator. Let $P_{\mathcal{D}}$

be the empirical distribution of the observed data sample, $\mathcal{D} = \{\mathbf{x}_i, \mathrm{t}_i, \mathrm{y}_i\}_{i=1}^n = \{\mathbf{S}_i\}_{i=1}^n$. Let $\widehat{P}_{\mathcal{D}} = \{\widehat{\mathcal{D}}_j\}_{k=1}^{n_b}$ be the bootstrap distribution over n_b datasets, $\widehat{\mathcal{D}}_k = \{\widehat{\mathbf{S}}_i\}_{i=1}^n$, sampled with replacement from the empirical distribution, $P_{\mathcal{D}}$. Let Q_α be the α-quantile of $\mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta})$ in the bootstrap resampling distribution: $Q_\alpha := \inf_{\mu^*} \left\{ \widehat{P}_{\mathcal{D}}(\mu(w(\mathrm{y}); \mathbf{x}, \mathrm{t}, \Lambda, \boldsymbol{\theta}) \leq \mu^*) \geq \alpha \right\}$. Finally, let $\boldsymbol{\theta}_k$ be the parameters of the model of the k-th bootstrap sample of the data. Then, the bootstrap confidence interval of the upper and lower bounds of the CAPO function under the CMSM is given by: $\mathrm{CI}_b\left(\mu(\mathbf{x}, \mathrm{t}; \Lambda, \alpha)\right) := \left[\underline{\mu}_b(\mathbf{x}, \mathrm{t}; \Lambda, \alpha), \overline{\mu}_b(\mathbf{x}, \mathrm{t}; \Lambda, \alpha)\right]$, where,

$$\underline{\mu}_b(\mathbf{x}, \mathrm{t}; \Lambda, \alpha) = Q_{\alpha/2}\left(\left\{\widehat{\underline{\mu}}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}_k)\right\}_{k=1}^b\right), \ \overline{\mu}_b(\mathbf{x}, \mathrm{t}; \Lambda, \alpha) = Q_{1-\alpha/2}\left(\left\{\widehat{\overline{\mu}}(\mathbf{x}, \mathrm{t}; \Lambda, \boldsymbol{\theta}_k)\right\}_{k=1}^b\right).$$

Furthermore, the bootstrap confidence interval of the upper and lower bounds of the APO (dose-response) function under the CMSM are given by: $\mathrm{CI}_b\left(\mu(\mathrm{t}; \Lambda, \alpha)\right) := \left[\underline{\mu}_b(\mathrm{t}; \Lambda, \alpha), \overline{\mu}_b(\mathrm{t}; \Lambda, \alpha)\right]$, where,

$$\underline{\mu}_b(\mathrm{t}; \Lambda, \alpha) = Q_{\alpha/2}\left(\left\{\widehat{\underline{\mu}}(\mathrm{t}; \Lambda, \boldsymbol{\theta}_k)\right\}_{k=1}^b\right), \qquad \overline{\mu}_b(\mathrm{t}; \Lambda, \alpha) = Q_{1-\alpha/2}\left(\left\{\widehat{\overline{\mu}}(\mathrm{t}; \Lambda, \boldsymbol{\theta}_k)\right\}_{k=1}^b\right).$$

3.5 Scalable Continuous Treatment Effect Estimation

Following [SJS17], [SLB+20], and [NYLN21], we propose using neural-network architectures with two basic components: a feature extractor, $\boldsymbol{\phi}(\mathbf{x}; \boldsymbol{\theta})$ ($\boldsymbol{\phi}$, for short) and a conditional outcome prediction block $f(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta})$. The feature extractor design will be problem and data specific. In Section 5, we look at using both a simple feed-forward neural network, and also a transformer [VSP+17]. For the conditional outcome block, we depart from more complex structures ([SLB+20, NYLN21]) and simply focus on a residual [HZRS16], feed-forward, S-learner [KSBY19] structure. For the final piece of the puzzle, we follow [JMGS21] and propose a n_{y} component Gaussian mixture density:

$$p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta}) = \sum_{j=1}^{n_{\mathrm{y}}} \widetilde{\pi}_j(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta}) \mathcal{N}\left(\mathrm{y} \mid \widetilde{\mu}_j(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta}), \widetilde{\sigma}_j^2(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta})\right),$$

and $\widetilde{\mu}(\mathbf{x}, \mathrm{t}; \boldsymbol{\theta}) = \sum_{j=1}^{n_{\mathrm{y}}} \widetilde{\pi}_j(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta}) \widetilde{\mu}_j(\boldsymbol{\phi}, \mathrm{t}; \boldsymbol{\theta})$ [Bis94]. Models are optimized by minimizing the log-likelihood of $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}, \boldsymbol{\theta})$.

4 Related Works

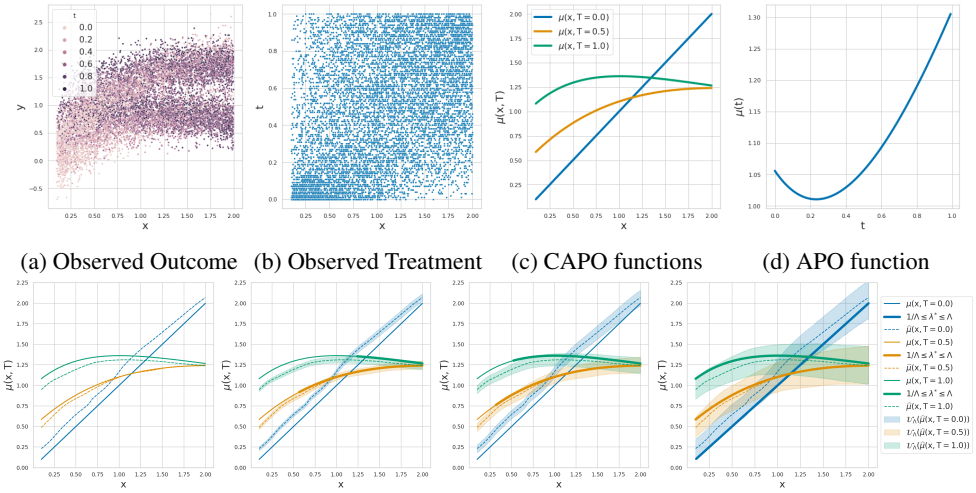
Scalable Continuous Treatment Effect Estimation. Using neural networks to provide scalable solutions for estimating the effects of continuous-valued interventions has received significant attention in recent years. [BJvdS20] provide a Generative Adversarial Network (GAN) approach. The dose-response network (DRNet) [SLB+20] provides a more direct adaptation of the TarNet [SJS17] architecture for continuous treatments. The varying coefficient network VCNet [NYLN21] generalizes the DRNet approach and provides a formal result for incorporating the target regularization technique presented by [SBV19]. The RieszNet [CCN+21] provides an alternative approach for targeted regularization. Adaptation of each method is straightforward for use in our sensitivity analysis framework by replacing the outcome prediction head of the model with a suitable density estimator.

Sensitivity and Uncertainty Analyses for Continuous Treatment Effects. The prior literature for continuous-valued treatments has focused largely on parametric methods assuming linear treatment/outcome, hidden-confounder/treatment, and hidden-confounder/outcome relationships [CHH16, DHCH16, MSDH16, Ost19, CH20a, CH20b]. In addition to linearity, these parametric methods need to assume the structure and distribution of the unobserved confounding variable(s). [CKC+19] allows for sensitivity analysis for arbitrary structural causal models under the linearity assumption. The MSM relaxes both the distributional and linearity assumptions, as does our CMSM extension. A two-parameter sensitivity model based on Riesz-Frechet representations of the target functionals, here the APO and CAPO, is proposed by [CCN+21] as a way to incorporate confidence intervals and sensitivity bounds. In contrast, we use the theoretical background of the marginal

sensitivity model to derive a one-parameter sensitivity model. [DBSC21] purport to quantify the bias induced by unobserved confounding in the effects of continuous-valued interventions, but they do not present a formal sensitivity analysis. Simultaneously and independently of this work, [MVSG] are deriving a sensitivity model that bounds the partial derivative of the log density ratio between complete and nominal propensity densities.

5 Experiments

Here we empirically validate our method. First, we consider a synthetic structural causal model (SCM) to demonstrate the validity of our method. Next, we show the scalability of our methods by applying them to a real-world climate-science-inspired problem. Implementation details (appendix H), datasets (appendix G), and code are provided in the supplementary materials.



(a) Observed Outcome (b) Observed Treatment (c) CAPO functions (d) APO function

(e) Λ=1.0 (f) Λ=1.1 (g) Λ=1.2 (h) Λ=1.6

Figure 1: Figures 1a to 1d: Synthetic data and ground truth functions. Figures 1e to 1h Causal uncertainty under hypothesized Λ values. Solid lines are ground truth; thick solid lines where the true λ^* is within the range of hypothesized Λ, thin solid lines otherwise. The dotted lines are the estimated CAPO. Shaded regions are estimated CMSM intervals.

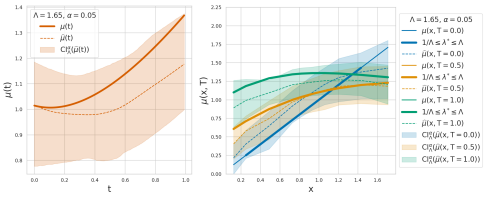
5.1 Synthetic

Figure 1 presents the synthetic dataset (additional details about the SCM are given in Appendix G.1). Figure 1a plots the observed outcomes, y, against the observed confounding covariate, x. Each datapoint is colored by the magnitude of the observed treatment, t. The binary unobserved confounder, u, induces a bi-modal distribution in the outcome variable, y, at each measured value, x. Figure 1b plots the assigned treatment, t, against the observed confounding covariate, x. We can see that the coverage of observed treatments, t, varies for each value of x. For example, there is uniform coverage at $\mathrm{X} = 1$, but low coverage for high treatment values at $\mathrm{X} = 0.1$, and low coverage for low treatment values at $\mathrm{X} = 2.0$. Figure 1c plots the true CAPO function over the domain of observed confounding variable, X, for several values of treatment ($\mathrm{T} = 0.0$, $\mathrm{T} = 0.5$, and $\mathrm{T} = 1.0$). For lower magnitude treatments, t, the CAPO function becomes more linear, and for higher values, we see more effect heterogeneity and attenuation of the effect size as seen from the slope of the CAPO curve for $\mathrm{T} = 0.5$ and $\mathrm{T} = 1.0$. Figure 1d plots the the APO function over the domain of the treatment variable T.

Causal Uncertainty We want to show that in the limit of large samples (we set n to $100k$), the bounds on the CAPO and APO functions under the CMSM include the ground truth when the CMSM is correctly specified. That is, when $1/\Lambda \leq \lambda^*(\mathrm{t}, \mathrm{x}, \mathrm{u}) \leq \Lambda$, for user specified parameter Λ, the estimated intervals should cover the true CAPO or APO. This is somewhat challenging to demonstrate as the true density ratio $\lambda^*(\mathrm{t}, \mathrm{x}, \mathrm{u})$ (eq. (50)), varies with t, x, and u. Figures 1e to 1h work towards communicating this. In Figure 1e, we see that each predicted CAPO function (dashed lines) is biased away from the true CAPO functions (solid lines). We use thick solid lines to indicate cases where $1/\Lambda \leq \lambda^*(\mathrm{t}, \mathrm{x}, \mathrm{u}) \leq \Lambda$, and thin solid lines otherwise. Therefore thick solid lines indicate areas

where we expect the causal intervals to cover the true functions. Under the erroneous assumption of ignorability ($\Lambda = 1$), the CMSM bounds have no width. In Figure 1f, we see that as we relax our ignorability assumption ($\Lambda = 1.1$) the intervals (shaded regions) start to grow. Note the thicker orange line: this indicates that for observed data described by $\mathrm{X} > 0.5$ and $\mathrm{T} = 0.5$, the actual density ratio is in the bounds of the CMSM with parameter $\Lambda = 0.5$. We see that our predicted bounds cover the actual CAPO function for these values. We see our bounds grow again in Figure 1g when we increase Λ to 1.2. We see that more data points have λ^* values that lie in the CMSM range and that our bounds cover the actual CAPO function for these values. In Figure 1h we again increase the parameter of the CMSM. We see that the bounds grow again, and cover the true CAPO functions for all of the data that satisfy $1/\Lambda \leq \lambda^*(\mathrm{t}, \mathrm{x}, \mathrm{u}) \leq \Lambda$.

Statistical Uncertainty Now we relax the infinite data assumption and set $n = 1000$. This decrease in data will increase the estimator error for the CAPO and APO functions. So the estimated functions will not only be biased due to hidden confounding, but they may also be erroneous due to finite sample variance. We show this in Figure 2b where the blue dashed line deviates from the actual blue solid line as $\mathbf{x}$ increases beyond 1.0. However, Figure 2b shows that under the correct CMSM, the uncertainty aware confidence intervals, section 3.4, cover the actual CAPO functions for the range of treatments considered. Figure 2a demonstrates that this holds for the APO function as well.

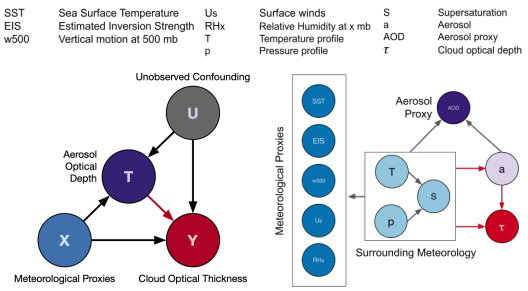


(a) APO Function (b) CAPO Functions

Figure 2: Statistical and causal uncertainty, α is statistical significance level for the bootstrap. see Figure 1 for other details.

5.2 Estimating Aerosol-Cloud-Climate Effects from Satellite Data

Background The development of the model above, and the inclusion of treatment as a continuous variable with multiple, unknown confounders, is motivated by a real-life use case for a prime topic in climate science. Aerosol-cloud interactions (ACI) occur when anthropogenic emissions in the form of aerosol enter a cloud and act as cloud condensation nuclei (CCN). An increase in the number of CCN results in a shift in the cloud droplets to smaller sizes which increases the brightness of a cloud and delays precipitation, increasing the cloud's lifetime, extent, and possibly thickness [Two77, Alb89, TCGB17]. However, the magnitude and sign of these effects are heavily dependent on the environmental conditions surrounding the cloud [DL20]. Clouds remain the largest source of uncertainty in our future climate projections [MDZP+21]; it is pivotal to understand how human emissions may be altering their ability to cool. Our current climate models fail to accurately emulate ACI, leading to uncertainty bounds that could offset global warming completely or double the effects of rising CO_2 [BRA+13].



(a) (b)

Figure 3: Causal diagrams. Figure 3a, a simplified causal diagram representing what we are reporting within; aerosol optical depth (AOD, regarded as the treatment T) modulates cloud optical depth (τ, Y), which itself is affected by hidden confounders (U) and the meteorological proxies (X). Figure 3b, an expanded causal diagram of ACI. The aerosol (a) and aerosol proxy (AOD), the true confounders (light blue), their proxies (dark blue), and the cloud optical depth (red).

Defining the Causal Relationships Clouds are integral to multiple components of the climate system, as they produce precipitation, reflect incoming sunlight, and can trap outgoing heat [SF09]. Unfortunately, their interconnectedness often leads to hidden sources of confounding when trying to address how anthropogenic emissions alter cloud properties.

Ideally, we would like to understand the effect of aerosols (a) on the cloud optical thickness, denoted τ. However, this is currently impossible. Aerosols come in varying concentrations, chemical

compositions, and sizes [SGW+16] and we cannot measure these variables directly. Therefore, we use aerosol optical depth (AOD) as a continuous, 1-dimensional proxy for aerosols. Figure 3b accounts for the known fact that AOD is an imperfect proxy impacted by its surrounding meteorological environment [CNP+17]. The meteorological environment is also a confounder that impacts cloud thickness τ and aerosol concentration a. Additionally, we depend on simulations of the current environment in the form of reanalysis to serve as its proxy.

Here we report AOD as a continuous treatment and the environmental variables as covariates. However, aerosol is the actual treatment, and AOD is only a confounded, imperfect proxy (Figure 3a). This model cannot accurately capture all causal effects and uncertainty due to known and unknown confounding variables. We use this simplified model as a test-bed for the methods developed within this paper and as a demonstration that they can scale to the underlying problem. Future work will tackle the more challenging and realistic causal model shown in Figure 3b, noting that the treatment of interest a is multi-dimensional and cannot be measured directly.

Model We use daily observed $1^\circ \times 1^\circ$ means of clouds, aerosol, and the environment from sources shown in Table 1 of Appendix G. To model the spatial correlations between the covariates on a given day, we use multi-headed attention [VSP+17] to define a transformer-based feature extractor. Modeling the spatial dependencies between meteorological variables is motivated by confounding that may be latent in the relationships between neighboring variables. These dependencies are unobserved from the perspective of a single location. This architectural change respects both the assumed causal graph (fig. 3a) and some of the underlying physical causal structure. We see in Figure 4 (Left) that modeling *context* with the transformer architecture significantly increases the predictive accuracy of the model when compared to a simple feed-forward neural network (*no context*).

Discussion & Results The results for the APO of cloud optical depth (τ) as the "treatment", AOD, increases are shown in Figure 4. As the assumed strength of confounding increases ($\Lambda > 1$), the range of uncertainty in the treatment outcome increases. Within this range of confounding, the modeled outcomes agree with two conflicting hypotheses. First, that aerosol acts to invigorate the cloud, inducing a large response that would follow a maximum curve within this uncertainty range [CS11, DL21]. And second, that aerosol has little impact on cloud depth, and the actual response is a minimal, flat line [GGS+19]. We further find the reported dose-response curves in agreement with multiple estimates of aerosol-cloud interactions using satellite observations [BTG02, MSJ+07, TCQB19]. The upper bound for $\log \Lambda = .2$ agrees with measurements of the in-cloud environment and aerosol-cloud interactions from aircraft-mounted sensors [PZ13], this may indicate the need for additional control variables when using satellite data.

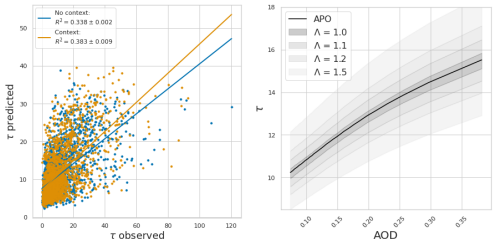


Figure 4: Left: The values of the observed, true τ against the modeled τ. Right: The curve for continuous treatment outcome of our aerosol proxy (AOD) on cloud optical depth (τ). The darkest shaded region (Λ = 1) represents the uncertainty in the treatment outcome from the ensemble due to finite data. As the strength of confounders increases ($\Lambda > 1.0$), the range of uncertainty in the treatment outcome increases.

The resolution of the satellite observations ($1^\circ \times 1^\circ$ daily means) could be averaging various cloud types and obscuring the signal. Future work will investigate how higher resolution (20km $\times$ 20km) data with constraints on cloud type may resolve some confounding influences. However, even our more detailed causal model (Figure 3b) cannot account for all confounders; we expected, and have seen, imperfections in our model of this complex effect. The model's results require further expert validation to interpret the outcomes and uncertainty.

Societal Impact Geoengineering of clouds by aerosol seeding could offset some amount of warming due to climate change, but also have disastrous global impacts on weather patterns [DGL+22]. Given the uncertainties involved in understanding aerosol-cloud interactions, it is paramount that policy makers are presented with projected outcomes if a proposals assumptions are wrong or relaxed.

References

[Alb89] Bruce A Albrecht. Aerosols, cloud microphysics, and fractional cloudiness. *Science*, 245(4923):1227–1230, 1989.

[BAC+15] Michael G Bosilovich, Santha Akella, Lawrence Coy, Richard Cullather, Clara Draper, Ronald Gelaro, Robin Kovach, Qing Liu, Andrea Molod, Peter Norris, et al. Merra-2: Initial evaluation of the climate. 2015.

[Bis94] Christopher M. Bishop. Mixture density networks. 1994.

[BJvdS20] Ioana Bica, James Jordon, and Mihaela van der Schaar. Estimating the effects of continuous-valued interventions using generative adversarial networks. *arXiv preprint arXiv:2002.12326*, 2020.

[BP06] Bryan A Baum and Steven Platnick. Introduction to modis cloud products. In *Earth science satellite remote sensing*, pages 74–91. Springer, 2006.

[BRA+13] Olivier Boucher, David Randall, Paulo Artaxo, Christopher Bretherton, Gragam Feingold, Piers Forster, V-M Kerminen, Yutaka Kondo, Hong Liao, Ulrike Lohmann, et al. Clouds and aerosols. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 571–657. Cambridge University Press, 2013.

[BTG02] Francois-Marie Bréon, Didier Tanré, and Sylvia Generoso. Aerosol effect on cloud droplet size monitored from satellite. *Science*, 295(5556):834–838, 2002.

[CCN+21] Victor Chernozhukov, Carlos Cinelli, Whitney Newey, Amit Sharma, and Vasilis Syrgkanis. Omitted variable bias in machine learned causal models. *arXiv preprint arXiv:2112.13398*, 2021.

[CH20a] Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):39–67, 2020.

[CH20b] Carlos Cinelli and Chad Hazlett. An omitted variable bias framework for sensitivity analysis of instrumental variables. *Work. Pap*, 2020.

[CHH16] Nicole Bohme Carnegie, Masataka Harada, and Jennifer L Hill. Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *Journal of Research on Educational Effectiveness*, 9(3):395–420, 2016.

[CKC+19] Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl, and Elias Bareinboim. Sensitivity analysis of linear structural causal models. In *International conference on machine learning*, pages 1252–1261. PMLR, 2019.

[CNP+17] Matthew W Christensen, David Neubauer, Caroline A Poulsen, Gareth E Thomas, Gregory R McGarragh, Adam C Povey, Simon R Proud, and Roy G Grainger. Unveiling aerosol–cloud interactions–part 1: Cloud contamination in satellite products enhances the aerosol indirect forcing estimate. *Atmospheric Chemistry and Physics*, 17(21):13151–13164, 2017.

[CS11] Matthew W Christensen and Graeme L Stephens. Microphysical and macrophysical responses of marine stratocumulus polluted by underlying ships: Evidence of cloud deepening. *Journal of Geophysical Research: Atmospheres*, 116(D3), 2011.

[D'A19] Alexander D'Amour. On multi-cause approaches to causal inference with unobserved counfounding: Two cautionary failure cases and a promising alternative. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3478–3486. PMLR, 2019.

[DBSC21] Gianluca Detommaso, Michael Brückner, Philip Schulz, and Victor Chernozhukov. Causal bias quantification for continuous treatment. *arXiv preprint arXiv:2106.09762*, 2021.

[DG21] Jacob Dorn and Kevin Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *arXiv preprint arXiv:2102.04543*, 2021.

[DGL+22] Michael S Diamond, Andrew Gettelman, Matthew D Lebsock, Allison McComiskey, Lynn M Russell, Robert Wood, and Graham Feingold. Opinion: To assess marine cloud brightening's technical feasibility, we need to know what to study—and when to stop. *Proceedings of the National Academy of Sciences*, 119(4), 2022.

[DHCH16] Vincent Dorie, Masataka Harada, Nicole Bohme Carnegie, and Jennifer Hill. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in medicine*, 35(20):3453–3470, 2016.

[DL20] Alyson Douglas and Tristan L'Ecuyer. Quantifying cloud adjustments and the radiative forcing due to aerosol–cloud interactions in satellite observations of warm marine clouds. *Atmospheric Chemistry and Physics*, 20(10):6225–6241, 2020.

[DL21] Alyson Douglas and Tristan L'Ecuyer. Global evidence of aerosol-induced invigoration in marine cumulus clouds. *Atmospheric Chemistry and Physics*, 21(19):15103–15114, 2021.

[GGS+19] Edward Gryspeerdt, Tom Goren, Odran Sourdeval, Johannes Quaas, Johannes Mülmenstädt, Sudhakar Dipu, Claudia Unglaub, Andrew Gettelman, and Matthew Christensen. Constraining the aerosol influence on cloud liquid water path. *Atmospheric Chemistry and Physics*, 19(8):5331–5347, 2019.

[GMS+17] Ronald Gelaro, Will McCarty, Max J Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A Randles, Anton Darmenov, Michael G Bosilovich, Rolf Reichle, et al. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of climate*, 30(14):5419–5454, 2017.

[HI04] Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.

[Hol86] Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[JMGS21] Andrew Jesson, Sören Mindermann, Yarin Gal, and Uri Shalit. Quantifying ignorance in individual-level causal-effect estimates under hidden confounding. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4829–4838. PMLR, 18–24 Jul 2021.

[KMZ19] Nathan Kallus, Xiaojie Mao, and Angela Zhou. Interval estimation of individual-level causal effects under unobserved confounding. In *The 22nd international conference on artificial intelligence and statistics*, pages 2281–2290. PMLR, 2019.

[KSBY19] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

[MDZP+21] V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, , and B. Zhou. Ipcc, 2021: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. 2021.

[MSDH16] Joel A Middleton, Marc A Scott, Ronli Diakow, and Jennifer L Hill. Bias amplification and bias unmasking. *Political Analysis*, 24(3):307–323, 2016.

[MSJ+07] Gunnar Myhre, Frode Stordal, M Johnsrud, YJ Kaufman, D Rosenfeld, Trude Storelvmo, Jon Egill Kristjansson, Terje Koren Berntsen, A Myhre, and Ivar SA Isaksen. Aerosol-cloud interaction inferred from modis satellite data and global aerosol models. *Atmospheric Chemistry and Physics*, 7(12):3081–3101, 2007.

[MVSG] Myrl G Marmarelis, Greg Ver Steeg, and Aram Galstyan. Bounding the effects of continuous treatments for hidden confounders.

[Ney23] Jerzy Neyman. edited and translated by dorota m. dabrowska and terrence p. speed (1990). on the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1923.

[NYLN21] Lizhen Nie, Mao Ye, Qiang Liu, and Dan Nicolae. Vcnet and functional targeted regularization for learning causal effects of continuous treatments. *arXiv preprint arXiv:2103.07861*, 2021.

[Ost19] Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204, 2019.

[PZ13] D Painemal and P Zuidema. The first aerosol indirect effect quantified through airborne remote sensing during vocals-rex. *Atmospheric Chemistry and Physics*, 13(2):917–931, 2013.

[RHW86] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[RR83] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[Rub74] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[Rub80] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

[Rud16] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[SBV19] Claudia Shi, David M Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 2507–2517, 2019.

[Sek08] Jasjeet S Sekhon. The neyman-rubin model of causal inference and estimation via matching methods. *The Oxford handbook of political methodology*, 2:1–32, 2008.

[SF09] Bjorn Stevens and Graham Feingold. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature*, 461(7264):607–613, 2009.

[SGW+16] Nick AJ Schutgens, Edward Gryspeerdt, Natalie Weigum, Svetlana Tsyro, Daisuke Goto, Michael Schulz, and Philip Stier. Will a perfect model agree with perfect observations? the impact of spatial sampling. *Atmospheric Chemistry and Physics*, 16(10):6335–6353, 2016.

[SJS17] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, pages 3076–3085. PMLR, 2017.

[SLB+20] Patrick Schwab, Lorenz Linhardt, Stefan Bauer, Joachim M Buhmann, and Walter Karlen. Learning counterfactual representations for estimating individual dose-response curves. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5612–5619, 2020.

[Tan06] Zhiqiang Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.

[TCGB17] Velle Toll, Matthew Christensen, Santiago Gassó, and Nicolas Bellouin. Volcano and ship tracks indicate excessive aerosol-induced cloud water increases in a climate model. *Geophysical research letters*, 44(24):12–492, 2017.

[TCQB19] Velle Toll, Matthew Christensen, Johannes Quaas, and Nicolas Bellouin. Weak average liquid-cloud-water response to anthropogenic aerosols. *Nature*, 572(7767):51–55, 2019.

[Two77] Sean Twomey. The influence of pollution on the shortwave albedo of clouds. *Journal of the atmospheric sciences*, 34(7):1149–1152, 1977.

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[ZSB19] Qingyuan Zhao, Dylan S Small, and Bhaswar B Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(4):735–761, 2019.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
      i. we claim to introduce a novel marginal sensitivity model for continuous valued treatment effect (the CMSM). See Section 3.
      ii. we claim to derive bounds for the CAPO and APO functions that agree with the CMSMS and observed data. See Section 3.1.
      iii. we claim to provide tractable estimators of the CAPO and APO bounds. See Sections 3.2 and 3.3.
      iv. we claim to provide bounds that account for finite-sample (statistical) uncertainty. See Section 3.4.
      v. we claim to provide a novel architecture for scalable estimation of the effects of continuous valued interventions. See Section 3.5.
      vi. we claim that the bounds cover the true ignorance interval for well specified Λ. See Figures 1e to 1h and Theorem 1.
      vii. we claim that this model scales to real-world, large-sample, high-dimensional data. See Section 5.2
   (b) Did you describe the limitations of your work? [Yes]
      i. We have discussed the major limitation of sensitivity analysis methods, that unobserved confounding is not identifiable from data alone. We have tried to be honest and transparent that our method provides users with a way to communicate the uncertainty induced when relaxing the ignorability assumption. We do not claim that lambda is in any way identifiable without further assumptions.

ii. In Section 5.2, we have clearly discussed the limitations of analyses of aerosol-cloud interactions using satellite data where we only see underlying causal mechanisms through proxy variables. We hope this paper serves as a stepping stone for work that specifically addresses those issues.

(c) Did you discuss any potential negative societal impacts of your work? [Yes]

(d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...

(a) Did you state the full set of assumptions of all theoretical results? [Yes] We have five theoretical results. Proposition 1, Equation (7), Equation (8), Equation (9), and Theorem 1. All assumptions are stated for each.

(b) Did you include complete proofs of all theoretical results? [Yes] The proof of Proposition 1 is given in Appendix A.3. The proof of Equation (7) is given in Lemma 1. The proofs for Equation (8) and Equation (9) are given in Lemma 2. The proof for Theorem 1 is given in Appendix E.

3. If you ran experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Code, data, and instructions are provided in the suppleemental material.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We specify these details in Appendix G and Appendix H as well as in the provided code.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Both random seeds and random bootstrapped sampling of the training data.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] this is outlined in Appendix H

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [Yes] We use existing satellite data and open source code libraries that we have cited.

(b) Did you mention the license of the assets? [Yes]

(c) Did you include any new assets either in the supplemental material or as a URL? [Yes] we provide a new synthetic dataset and code base

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]