# Quantifying Causes of Arctic Amplification via Deep Learning based Time-series Causal Inference

Sahara Ali[*§], Omar Faruque[*], Yiyi Huang[*], Md. Osman Gani[*§], Aneesh Subramanian[†§], Nicole-Jeanne Schlegel[‡], Jianwu Wang[*§]
[*]Department of Information Systems, University of Maryland, Baltimore County, Baltimore, MD, United States
[†]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO, United States
[‡]Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ, United States
[§]NSF HDR Institute for Harnessing Data and Model Revolution in the Polar Regions (iHARP), United States
Emails: [*]{sali9,omarf1,yhuang10,mogani,jianwu}@umbc.edu, [†]aneeshcs@colorado.edu, [‡]nicole.schlegel@noaa.gov

***Abstract*—The warming of the Arctic, also known as Arctic amplification, is led by several atmospheric and oceanic drivers. However, the details of its underlying thermodynamic causes are still unknown. Inferring the causal effects of atmospheric processes on sea ice melt using fixed treatment effect strategies leads to unrealistic counterfactual estimations. Such methods are also prone to bias due to time-varying confoundedness. Further, the complex non-linearity in Earth science data makes it infeasible to perform causal inference using existing marginal structural techniques. In order to tackle these challenges, we propose TCINet - Time-series Causal Inference Network to infer causation under continuous treatment using recurrent neural networks and a novel probabilistic balancing technique. More specifically, we propose a neural network based potential outcome model using the long-short-term-memory (LSTM) layers for time-delayed factual and counterfactual predictions with a custom weighted loss. To tackle the confounding bias, we experiment with multiple balancing strategies, namely TCINet with the inverse probability weighting (IPTW), TCINet with stabilized weights using Gaussian Mixture Model (GMMs) and TCINet without any balancing technique. Through experiments on synthetic and observational data, we show how our research can substantially improve the ability to quantify leading causes of Arctic sea ice melt, further paving paths for causal inference in observational Earth science.**

***Index Terms*—Causal Inference, Deep Learning, LSTM, Arctic Amplification**

## I. Introduction

In the last few decades, Earth and Atmospheric scientists have observed greater climate change near the polar regions as compared to the rest of the world [20]. In 2018, the observed mean sea ice extent (SIE) at Kara Sea during the summer months of June, July and August (JJA) reduced to half of what it was in 1979, i.e. from 1.25 $million\ km^2$ to just 0.5 $million\ km^2$. What we are observing can happen in response to a change in global climate forcing. Due to the melting of highly reflective sea ice and snow regions in the Arctic and Antarctic, there is an increased absorption of solar radiation which amplifies the warming. This phenomenon, also known as polar amplification is causing the melting of polar ice sheets, resulting in sea level rise, and the rate of carbon uptake in the polar regions [20]. In light of this phenomenon, the warming of Arctic sea ice is referred to **Arctic Amplification** [28]. Though, it has not been scientifically proven if the Arctic has warmed more than the rest of the hemisphere, studying the cause of thinning and retreat of the Arctic sea ice is a significant and substantial topic of atmospheric research [10].

In this paper, we dive deeper into the concept of time-series Causal Inference (CI) and present the challenges and opportunities for performing CI to study Arctic amplification under continuous treatment effect. Causal inference can be defined as the process of estimating the causal effects (influence) of one event, process, state or object (a cause) on another event, process, state or object (an effect). For estimation of causal effect, there are two main categories of techniques, potential outcome framework and do-calculus. The potential outcome framework relies on hypothetical interventions such that it defines the causal effect as the difference between the outcomes that would be observed with and without exposure to the intervention [32]. This technique is widely used in epidemiology where patients are randomly divided into treated and controlled groups and the effectiveness of treatment is inferred by observing patients condition with and without undergoing a treatment [33]. The treatment effect can be measured at individual, treated group, sub-treated group and entire population levels [25]. At the population level, the treatment effect measured is called Average Treatment Effect (ATE). In case of Earth science observational data, we consider ATE a more suitable metric as a causal estimand which quantifies the mean difference observed in potential outcomes $Y$ given the exposure to treatment $X$, i.e., $Y(X = 1)$ versus the inexposure to the treatment, i.e., $Y(X = 0)$. The established standard approach of performing time-varying causal inference in case of linear time-series data is through the use of marginal structural models [30], whereas, recent development in deep learning has paved paths for robust techniques for performing causal inference on non-linear observational and longitudinal data [17]. While existing deep learning models majorly handle independent and identically distributed (i.i.d) data [17], we see only a handful of techniques capable of joint representation learning of continuous treatment and covariates in time-varying setting [4], [5], [21]. We compare in Table I the capability of existing deep learning and machine learning

methods in fulfilling Earth science requirements.

In light of above background, this paper presents a deep learning based time-series causal inference method that overcomes the limitations of existing causal effect estimation approaches in answering important research questions pertaining to the climate change effects in the Arctic. We present TCINet, a deep learning based time-series causal inference model, for counterfactual prediction under time-delayed continuous treatment. Our major contributions can be summarized as follows:

- We propose a deep learning based time-series causal inference model suitable for both time-varying and time-invariant treatment effect estimation, which includes a new definition for average treatment effect estimation in case of time-delayed continuous treatment.
- We propose a novel probabilistic weighting technique to balance time-varying confoundedness by leveraging Gaussian Mixture Model (GMM).
- We perform extensive experiments evaluating our approach and compare it with the state-of-the-art (SOTA) approaches using synthetic time series data for fixed and continuous time-delayed treatments; further verifying our quantified causal effect results of thermodynamic processes on the Arctic sea ice melt with domain knowledge.

Moving forward, we will use the following terminologies: treatment variable (which is an identified cause), potential outcome (the variable identified as effect) and covariates (a set of variables that are either common cause of both treatment and outcome or are descendants of treatment variables identified in the causal graph).

### A. Formulating Causal Inference for Earth Science

Climate data is non-stationary with climatological trends and visible annual seasonality cycles, therefore, binary or fixed treatment effect estimation can be an unrealistic way of quantifying causation. Further, in the absence of ground truth, the exposure to a policy change or applying dynamic treatment regime cannot be observed. This leads to the inability to evaluate model's performance for counterfactual predictions in observational data [34], [38]. Existing techniques such as marginal structural models [30], time-series regression [16], matching methods [38] and deep learning based counterfactual predictions struggle in accurately inferring causation under time-delayed continuous intervention [17]. To fill this gap, we propose a deep learning based inference model, based on the potential outcomes framework [32] and extending the recurrent methods based counterfactual approach [21] to study the impact of time-delayed treatment in the presence of time-varying covariates.

More formally, given treatment $X_t$ at timestep $t$, our model infers the time-delayed outcome $Y_{t+l}$ at $l$ steps ahead in future, in the presence of a set of $M$ time-dependent covariates $Z_t$. We give a generic formulation of our problem as follows. $Y(\hat{X}_t)$ is the potential outcome, i.e., forecasted values under intervened treatment $\hat{X}$ at time $t$, and $Y(X_t)$ is the potential outcome under treatment $X$ at time $t$ without intervention (also called placebo effect), whereas $Z_t$ represents the covariates at time $t$, and $f$ represents our proposed deep learning based inference model. We utilize both factual and counterfactual predictions of $Y$ for all $N$ timesteps to estimate ***lagged average treatment effect (LATE)*** under continuous intervention:

$$Y_{t+l}(X = x_t) = f(Z_t, x_t) \tag{1}$$

$$Y_{t+l}(\hat{X} = \hat{x}_t) = f(Z_t, \hat{x}_t) \tag{2}$$

$$LATE(l) = \frac{1}{N}\sum_{t=1}^{N} E[Y_{t+l}(\hat{X}_t) - Y_{t+l}(X_t)] \tag{3}$$

For consistent causal effect estimation under time-varying treatment, our proposed model holds the standard identifiability conditions or causal assumptions of consistency, positivity and conditional exchangeability [25], [30]. Our implementation code can be accessed at the iHARP GitHub repository[1].

## II. Related Work

Though causality based study is a comparatively a new paradigm in Earth science, causal inference has been a widely studied topic for decades in statistics, economics, public policy and even healthcare [17], [25], [41].

*1) G-Methods for Time-Varying Causal Inference:* Estimating time-varying causal or treatment effects leads to the problem of time-varying confounding, that is the common influence a past treatment or covariate might have on the future treatments and the future outcome. Robin's g-methods have shown to provide promising results on reducing bias caused by time-varying treatment and covariates on the potential outcome [26]. G-methods provide metrics to overcome the problem of time-varying confounding through standardization, g-computation, and inverse probability of treatment weighted (IPTW) estimators [26]. The prediction models of these estimators are typically based on linear or logistic regression such as Causal-ARIMA [24], Time Based Regression (TBR) [16] and Marginal Structural Models (MSMs) [30]. One big limitation of these methods is that, in case of complex non-linearity in treatment or outcome variables, the methods will lead to inaccurate results.

*2) Deep Learning based Causal Inference:* Causal inference methods based on representation learning or deep learning techniques [3] learn the representation of input data by extracting features from the covariate space [17], where majority of the existing deep learning based methods are developed for i.i.d data [17]. In these deep learning based CI methods, a single neural network (also called meta learner) can be trained to make predictions for both treatment and control groups individually to estimate the average treatment effect (ATE). Existing meta-learners include S(ingle)-learner [18], and T-learner or multi-task learners [15], [36] that jointly predict outcome for treated and controlled groups. X-learner [18] or cross-group learners are a hybrid form of meta learners that overcome the problem of unbalanced data in treatment and controlled groups. U-learner [27] and R-learner utilizes

[1] github.com/iharp-institute/causality-for-arctic-amplification

Robinson transformation to develop a custom loss function for conditional ATE estimation [27]. SCIGAN is another causal inference method for estimating the effects of continuous-valued interventions that aim to learn the distribution of unobserved counterfactuals using Generative Adversarial Networks (GANs) [6]. The limitations of CI methods for i.i.d. data is that these methods perform poorly on sequential or time-series data with no capability to handle time lags or time-varying confounding effects, thereby leading to invalid causal effect estimation results. For time-series causal inference, researchers have proposed methodologies based on machine learning and deep learning models that can also tackle the problem of time-varying confounding [25]. Recurrent Marginal Structural Networks (R-MSN) [21] and Counterfactual Recurrent Network (CRN) [5] are some of the recent models that claim to estimate causal effects in the presence of time-varying confounders, however, contrary to the claim, these methods are healthcare-specific and cannot be generalized for other domain areas like Earth science because these models require on-hot encoded treatment flags with multivariate combined dosage. Talking about counterfactuals, the most recent model, Time Series Deconfounder - a multi-task method, leverages the assignment of multiple treatments over time to enable the estimation of treatment effects in the presence of multi-cause hidden confounders [4]. The Conditional Instrumental Variable (CIV) [39] method measures the causal effect $\beta$ from covariates to the target variable using instrument variables that have a relation with covariates and target but are independent of any hidden confounder. To yield a better estimation the instrument variables are conditioned for single or multiple previous time steps in CIV. Though deep representation learning methods are capable of automatically learning the intrinsic correlations and are also effective in accurate counterfactual estimation, they often lead to predictions with high variance or uncertainty estimates.

*3) Time-Varying Causal Inference for Earth Science:* From climate or atmospheric science perspective, causality remains a lesser tapped area [14], [34], [35] and climatologist still rely on dynamical modeling techniques where certain atmospheric variables are nudged or perturbed as initial conditions in the physical simulation models (also called Earth System Models) to evaluate the outcome of these interventions on target variables [12], [22], [40]. Applying deep learning techniques to infer causal effects of climate change offers a data-driven and cost-effective solution to the problem. Deep learning (DL) models can work more efficiently and effectively than current climate model simulators that are highly computationally expensive. Our work will build on top of deep learning based predictive models where we will extend them from fixed treatment to continuous treatment setting. Table I shows a holistic comparison of some of the time-series based causal inference methods and their capabilities to handles different causal inference scenarios.

## III. Datasets

To evaluate our model, we first generate synthetic data with time-delayed continuous treatment and time-varying covariates. We further provide details of the real world observational dataset pertaining to our research problem.

### A. Synthetic Data

Using gaussian white noise, we generate four non-linear time-series given in Equations 4 to 7, mimicking the non-linearity in dynamic climate models.

The corresponding true causal graph for these time-series is given in Figure 1. Here, we have taken $S3$ to be the treatment and $S4$ as the potential outcome. $S1$ and $S2$ will be considered as covariates where $S1$ is an observed time-varying confounder of both treatment and outcome. To generate counterfactuals, we intervene on $S3$, in two settings. First, we intervene on $S3$ as fixed treatment with binary values of $[0, 1]$ to generate counterfactual values of $S4$. Next, we intervene on $S3$ by increasing $S3$ by $10\%$ and generate corresponding $S4$ counterfactuals under continuous treatment.

$$S1_t = cos(\frac{t}{10}) + log(|S1_{t-6} - S1_{t-10}| + 1) + 0.1\varepsilon 1 \quad (4)$$

$$S2_t = 1.2e^{\frac{S1_{t-1}^2}{2}} + \varepsilon 2 \quad (5)$$

$$S3_t = -1.05e^{\frac{-S1_{t-1}^2}{2}} + \varepsilon 3 \quad (6)$$

$$S4_t = -1.15e^{\frac{-S1_{t-1}^2}{2}} + 1.35e^{\frac{-S3_{t-1}^2}{2}} + 0.28e^{\frac{-S4_{t-1}^2}{2}} + \varepsilon 4 \quad (7)$$

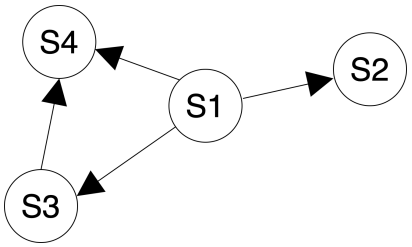


Fig. 1. Causal graph of non-linear synthetic data.

### B. Observational Arctic Data

We used observational sea-ice and reanalysis atmospheric and meteorological data which is available from 1979 till present. The reanalysis data is available with open access and can be obtained from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-5 global reanalysis product [2]. Whereas the sea ice concentration (SIC) values are obtained from Nimbus-7 SSMR and DMSP SSM/I-SSMIS passive microwave data version 1 [8] provided by the National Snow and Ice Data Center (NSIDC). The original data format is spatiotemporal from which we generated spatially averaged time-series combining sea ice extent values, oceanic and atmospheric variables. For this, daily gridded data over the regions of Barents Sea and Kara Sea, during 1979-2018, has been averaged using area-weighted method. The details of these variables are enlisted in Table II.

TABLE I
COMPARISON OF TCINET WITH EXISTING TIME-SERIES CAUSAL INFERENCE METHODS.

| Method | Binary/ fixed treatment | Continuous treatment | Time varying treatment | Time varying covariates | Applicable on Earth Science |
|---|---|---|---|---|---|
| Difference in Difference [19] | ✓ | ✗ | ✗ | ✗ | ✗ |
| Causal Impact [7] | ✓ | ✓ | ✓ | ✗ | ✓ |
| CIV [39] | ✓ | ✓ | ✓ | ✓ | ✓ |
| CRN [5] | ✓ | ✗ | ✓ | ✓ | ✗ |
| MSM [30] | ✓ | ✗ | ✓ | ✗ | ✗ |
| R-MSN [21] | ✓ | ✗ | ✓ | ✓ | ✗ |
| Time-series Deconfounder [4] | ✓ | ✗ | ✓ | ✓ | ✗ |
| TCINet (ours) | ✓ | ✓ | ✓ | ✓ | ✓ |

TABLE II
VARIABLES IN THE ARCTIC DATASET

| VARIABLE | RANGE | UNIT |
|---|---|---|
| SPECIFIC HUMIDITY | [0,0.1] | KG/KG |
| SHORTWAVE RADIATION | [0,1500] | $W/m^2$ |
| LONGWAVE RADIATION | [0,700] | $W/m^2$ |
| RAINFALL RATE | [0,800] | MM/DAY |
| SEA SURFACE TEMPERATURE | [200,350] | K |
| AIR TEMPERATURE | [200,350] | K |
| GREENLAND BLOCKING INDEX | [5000,5500] | M |
| SEA ICE EXTENT | [4, 13] | MILLION $Km^2$ |

## IV. METHODOLOGY

Following the same principle of meta-learning used in existing deep learning based causal inference approaches, we propose a time-varying causal inference model, called Time-series Causal Inference Network (TCINet), on top of our previous work on LSTM based sea-ice forecasting model [1]. The training and inference phases of our TCINet pipeline are illustrated in Figure 2. In the training phase, time-delayed treatments $X_{t-l}$ and time-varying covariates $Z_{t-l}$ are fed to our potential outcome model (see Subsection IV-B). To balance the bias due to time-varying covariates, we leverage Gaussian mixture modeling to compute stabilized weights (see Subsection IV-A). We also define a custom weighted loss to incorporate the balancing weights into our potential outcome model (see Subsection IV-C). In the inference phase, we perturb the treatment variable and feed it to the pretained outcome model to make factual and counterfactual predictions (see Subsection IV-D). We further explain how we estimate uncertainty during inference arguing on the feasibility of bootstrapping for time-series data.

### A. Balancing Time-varying Covariates

Balancing is a treatment adjustment strategy that aims to deconfound the treatment from outcome by forcing the treated and control covariate distributions as close as possible. When conducting observational studies, researchers often face the challenge that treatment assignment is not randomized, leading to potential confounding variables that can bias the treatment effect estimates. Inverse Probability of Treatment Weights (IPTW) [31] is a statistical technique used in causal inference to address confounding bias in observational studies. IPTW generates a pseudo-population in which treatments are independent of confounders. To calculate IPTW, we first need the predicted probabilities of the observed treatments given the covariates. This is also known as the propensity score, given by $prob(X|Z = z)$. When treatment and confounders are time-varying, these IPTW weights for time-fixed treatments need to be generalized. For a time-varying treatment $\bar{X}_t = (X_1, X_2, ..., X_t)$ and time-varying covariate $\bar{Z}_t = (Z_1, Z_2, ..., Z_t)$, the IP weights for every timestep $t$ are given by [13]:

$$IPTW(l) = \Pi_{t=1}^{l} \frac{1}{f(\bar{X}_t|\bar{Z}_t)} \tag{8}$$

Here, $l$ represents the lag or length of treatment sequence, $f(.)$ is the propensity score model, widely implemented using logistic regression following marginal structural modeling technique [30]. However, the propensity scores that are near 0 or 1 can yield extreme IPTW weights, leading to unstable estimates and inflated variances. To tackle this, [30] proposed the stabilized weights in which the IPTW is multiplied by the probability of receiving treatment, as given in Equation 9. Stabilized weights offer greater stability and reduce the variance in treatment effect estimation, which can improve the precision of the estimates. They are generally preferred in practice because of their improved numerical properties and stability.

$$SW(l) = \frac{\Pi_{t=1}^{l} f(X_t|\bar{X}_{t-1})}{\Pi_{t=1}^{l} f(X_t|\bar{X}_{t-1}, \bar{Z}_t)} \tag{9}$$

Here, $f(.)$ represents the probability density function (PDF) of treatment at every timestep given covariates and treatment history. In case of binary or discrete treatment, the PDF can be estimated using logistic regression or sigmoid function. However, in case of continuous treatment such as our case, this estimation becomes complex as it requires a parametric model to estimate the PDF at every stage $t$ [13].

We implement a Gaussian Mixture Model (GMM) [29] to estimate the probability density of treatment $X_t$ at every timestep $t$. The step-by-step implementation of GMM for calculating stabilized weights is given in Algorithm 1. We

refer to the conditional probability densities in Equation 9 as $X\_pdf$ and $XZ\_pdf$ in our algorithm. Whereas, the mean $\mu$, covariance $\Sigma$ and parameter $\alpha$, estimated as mixing coefficients, are all learned by the GMM model. First, we fit the GMM model on treatment history and covariates to learn these parameters. We then estimate the probability density of $X_t$ given these parameter values at every timestep $t$ using Equation 10.

$$f(X_t|\mu,\Sigma) = \left(\frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}}\right)\exp\left[-\frac{1}{2}(X_t-\mu)^T\Sigma^{-1}(X_t-\mu)\right] \quad (10)$$

**Algorithm 1:** Stabilized Weights for Continuous Treatment

**Data:** Treatment Data: $X$, Treatment History: $\bar{X}_{\text{hist}}$, Time-varying Covariates: $\bar{Z}$
**Result:** Stabilized Weight Estimates $SW$

**Function** PDF_calc($X$, $\bar{X}_{hist}$, $\bar{Z} = []$):
  // Concatenate the treatment history and covariates
  $\overline{XZ} \leftarrow concat(\bar{X}_{\text{hist}}, \bar{Z})$ ;
  $l \leftarrow$ length of sequence $XZ$;
  **for** $i \leftarrow 1$ **to** $l$ **do**
    $n_{\text{comp}} \leftarrow$ Number of components for GMM;
    // Create a GMM object
    $gmm \leftarrow$ GaussianMixture($n_{\text{comp}}$) ;
    // Fit the GMM model
    $gmm$.fit($\overline{XZ}_i$);
    // Extract model parameters:
    $(\alpha, \mu, \Sigma) \leftarrow (gmm.\text{weights}, gmm.\text{means}, gmm.\text{covariances})$;
    // Estimate PDF for every component
    **for** $j \leftarrow 1$ **to** $n_{comp}$ **do**
      $pdf_{\text{comp}}[j] \leftarrow \left(\frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma_j|}}\right) * \exp\left[-\frac{1}{2}(X_i-\mu_j)^T\Sigma_j^{-1}(X_i-\mu_j)\right]$;
    // Sum PDF over all components
    $pdf[i] \leftarrow \sum_{j=1}^{n_{\text{comp}}}(pdf_{\text{comp}}[j] \times \alpha[j])$;
  // Take product of PDFs over all sub-sequences
  $pdf_{\text{product}} \leftarrow \Pi_{i=1}^{l} pdf[i]$ **return** $pdf_{product}$
$X_pdf \leftarrow$PDF_calc($X$, $\bar{X}_{\text{hist}}$) ;
$XZ_pdf \leftarrow$PDF_calc($X$, $\bar{X}_{\text{hist}}$, $\bar{Z}$) ;
// Calculate stabilized weights at every timestep
**for** $k \leftarrow 1$ **to** $t_{timesteps}$ **do**
  $SW[k] \leftarrow \frac{X\_pdf[k]}{XZ\_pdf[k]}$;

### B. Potential Outcome Model (POM)

We develop an LSTM-based prediction model as our potential outcome model (POM), following the S-learner approach [17]. POM takes in input a 3D tensor of shape $N \times T \times F$. Here $N$ represents the mini-batch size, $T$ represents the time lag and $F$ comprises the covariates and the treatment variable at timestep $t$. The model comprises three LSTM layers with RELU activation, where first two many-to-many (also called seq2seq) layers are followed by a Dropout layer to cater uncertainty estimation. These seq2seq layers take in a sequence of input of length $l$ and learn the latent representations $\phi$ of treatment and covariates. The third LSTM layer is a many-to-one layer succeeded by three fully connected Dense layers with linear activation. The purpose of these layers is to combine the learned representations to jointly predict the potential outcome $Y_{t+l}$ at timestep $t+l$ where $l$ is the time-dependency or lag. The model is compiled using Adam optimizer using the early stopping technique.

For a joint input, the model will learn mixed representations of covariates and treatment. This will be problematic in causal effect estimation as we want to keep the covariates independent of the intervention on the treatment variable. This is where the balancing strategy comes into play. To debias the confoundedness, we use Gaussian Mixture Model (GMM), as discussed in Subsection IV-A, to get the stabilized weights $SW_t$ for time-varying treatment at each timestep $t$ given the confounders. To train POM to make weighted predictions for potential outcome $Y_{t+l}$, we implement a custom-weighted loss function.

### C. Custom Weighted Loss

We introduce the stabilized weights into the POM model using a custom weighted loss technique, to regularize the predictive model. The SW weights $SW_t$ calculated using GMM are inducted to the predictive model loss $L^{pred}$. This implies the final loss of the model will be a weighted average of prediction loss over observed data points $N$, as shown in Equation 11, where $L^{pred}$ is the mean squared error (MSE) loss.

$$L_{tot} = \frac{1}{N}\sum_{t=1}^{N} SW_t * L_t^{pred} \quad (11)$$

### D. Inference

Once the predictive model, namely POM, is trained on the observational data, the next step is to predict factual and counterfactual outcomes. We do this by perturbing the treatment variables at every timestep while retaining the observed values of time-varying covariates. The updated data is fed to the model to make counterfactual predictions while factual predictions are made without performing any nudging or intervention on the treatment variable. Once we have both predictions for all timesteps, we calculate LATE using Equation 3.

To gain confidence in the predicted counterfactual values, we analyze the predictive skill of the underlying deep learning

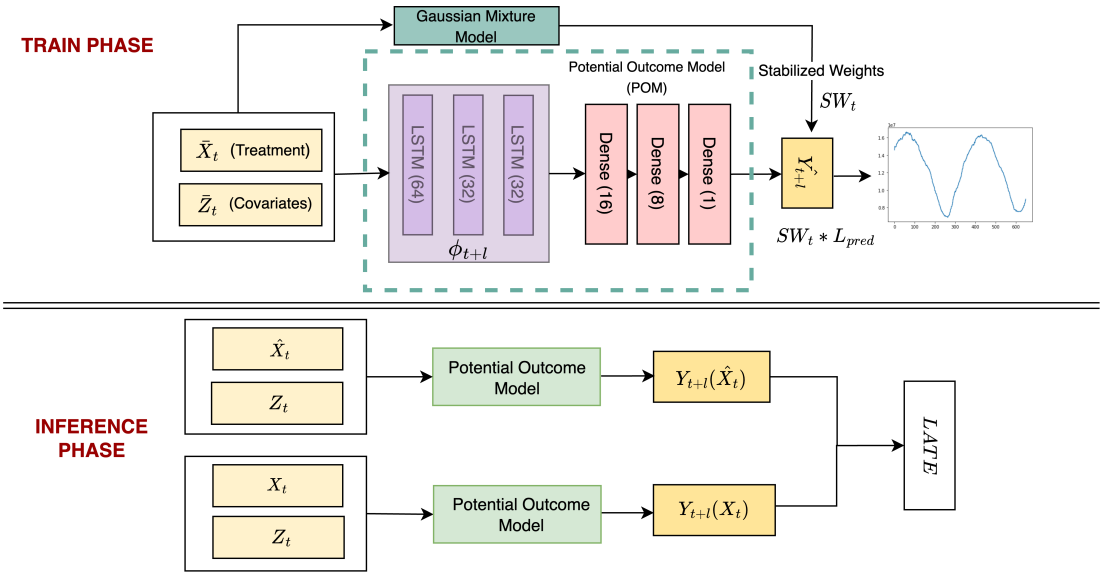


Fig. 2. Training and inference/test phase pipeline of our proposed TCINet model.

model and measure the model's epistemic uncertainty. Bootstrapping is a common approach used for quantifying model uncertainty in causal inference techniques [37], however, bootstrapping will lead to two potential problems in case of our data. First, bootstrapping requires random sampling of data for train and test split but sampling randomly from time-series data will corrupt the sequential pattern and lead to unrealistic results. Second, bootstrapping involves retraining the model every time a random number of samples are taken from the data. In case of TCINet, it will be computationally expensive to retrain the deep learning model *n* number of times as required by bootstrapping.

We therefore take an alternative approach, where we train the TCINet modules POM and GMM once and make predictions *n* times for each interventional scenario. We then calculate the mean and standard deviation of these predictions. The ATEs are recorded for observational data after making predictions for each case 50 times with a 95% confidence interval.

## V. Results & Analysis

In this section, we report our experimental setup and results obtained on synthetic and observational data, followed by a critical analysis of our findings using RMSE, LATE and PEHE scores.

### A. Evaluation Metrics

*1) Root Mean Square Error (RMSE):* We evaluate the performance of our predictive models using the Root Mean Square Error (RMSE) which can be only calculated for factual observational data but cannot be done for counterfactual predictions.

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{12}$$

*2) Precision in Estimated Heterogeneous Effects (PEHE):* This metric is commonly used in machine learning literature for calculating the average error across the predicted ATEs [9]. PEHE metric, measuring causal effect estimation skill, can only be calculated for synthetic data which has ground truth information.

$$\sqrt{PEHE} = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(ATE_i - \hat{ATE}_i)^2} \tag{13}$$

### B. Experimental Setup

We implement TCINet using Keras functional API with TensorFlow backend. The model has a total of $40,551$ trainable parameters. We compile the model using Adam optimizer with a $0.001$ learning rate and train it using early stopping technique. We train three variants of our model depending upon the underlying balancing strategies used in custom weighted loss: TCINet with SW weights using GMM which we refer to as TCINet-GMM, TCINet with IPTW weights using Logistic Regression model, which we refer to as TCINet-LR; and TCINet without any weighting using standard Mean Squared Error loss which we refer to as TCINet$^{-}$ in Table III.

### C. Results on Synthetic Data

We report our results on the three variants of the model; TCINet-GMM, TCINet-LR and TCINet and compare them with two state-of-the-art (SOTA) methods: time-varying CIV technique [39] and time-invariant Causal Impact [7] causal inference methods. We evaluate both the CIV and Causal Impact method using the synthetic dataset to measure the causal effect from the cause $S3$ to the target variable $S4$ in case of the fixed and continuous treatments explained in III-A. We report these results in Table III. Comparing the performance of three TCINet variants in Table III, we notice that all variants have marginal differences in RMSE scores, however, we see substantial differences in ATE estimation by

these models. This performance difference is also evident from the low PEHE scores for TCINet-GMM in Table III. Since CIV does not provide RMSE values on factual estimation, we compare its performance based on estimated ATE values. Though CIV is an easier model to implement, we notice that in both cases, i.e., fixed and continuous treatment effect estimation, CIV performs poorly as compared to TCINet variants and Causal Impact, which gives us more confidence in our model performance. It is important to note here that Causal Impact provides the second best performance in case of fixed and continuous treatments, however, inherently Causal Impact cannot work with time-varying covariates and is therefore not suitable for our case. Moving forward, we analyze the observational data using TCINet-GMM owing to its superior performance.

TABLE III
CAUSAL INFERENCE MODELS PERFORMANCE ON SYNTHETIC DATA UNDER FIXED AND CONTINUOUS TREATMENTS FOR ONE-STEP AHEAD PREDICTION (TRUE ATE = -0.0514)

| MODEL | TEST RMSE | ESTIMATED LATE | PEHE |
|---|---|---|---|
| FIXED TREATMENT | | | |
| **TCINET$^-$** | 1.079 | -0.040 | 1.132 |
| **TCINET-LR** | 1.142 | -0.037 | 1.227 |
| **TCINET-GMM** | 1.023 | -0.051 | 1.004 |
| **CIV [39]** | N/A | -0.219 | N/A |
| **CAUSAL IMPACT [7]** | N/A | -0.060 | 1.110 |
| CONTINUOUS TREATMENT | | | |
| **TCINET$^-$** | 1.026 | -0.036 | 1.221 |
| **TCINET-LR** | 1.000 | -0.049 | 1.143 |
| **TCINET-GMM** | 0.998 | -0.050 | 1.102 |
| **CIV [39]** | N/A | 0.515 | N/A |
| **CAUSAL IMPACT [7]** | N/A | -0.040 | 1.112 |

### D. TCINet for Arctic Amplification

After gaining confidence in the predictive skill of TCINet for synthetic data, we use the model to answer an important domain science question on the observational data as identified by Atmospheric scientists [11]: *How does increased Greenland Blocking (GBI) affect summertime regional Arctic sea ice melting given snowfall rate and solar radiation data?*

The Greenland block is a ridge of high pressure that sits near or over Greenland. It is the normalised area-weighted 500 hPa geopotential height over the region $60-80°N, 20-80°W$. To identify the regions of interest and time lag by which GBI affects sea ice extent, we performed lagged correlation test between daily GBI values and regional sea ice extent given by [23] for sixteen sub-regions. We conducted experiments for a lag of 0 to 30 days and found the highest correlation at day 8 between GBI and SIE in Barents Sea and Kara Sea (combined as BarKara Sea in our analysis).

To answer the domain science question, we first trained TCINet-GMM on forty years of our observational data. We then predict sea ice extent by perturbing the values of summertime (June, July, August) GBI to the following four values: 1) 40-year-averaged daily GBI, 2) double GBI annual trend, 3) triple GBI annual trend, 4) quadruple GBI annual trend.

In our efforts to quantify the effects of increasing GBI on declining sea ice, we first make predictions for summertime (June, July, August) sea ice for mean daily GBI values. We then perturb the GBI values by increasing them by a multiplicative factor of the daily recorded trend, i.e. 0.039. Our interpretation of ATE in case of observational sea ice data is that it reflects the average increase or decrease in sea ice extent under interventional treatment.

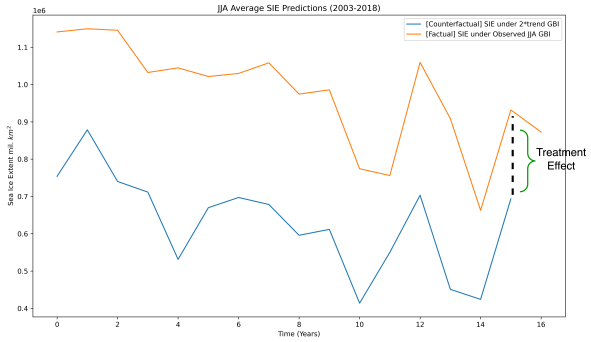


Fig. 3. Comparison of annual mean sea ice extent (SIE) predictions given observational data (factual) versus predictions under interventional GBI (counterfactual) between 2003-2018. Here, each data point represents summer (JJA) mean SIE predicted for that year.

As shown in Figure 3, we notice that increasing GBI leads to decrease in sea ice extent (blue line with counterfactual predictions). Quantitatively, our model predicts that the average daily sea ice extent value in JJA summer months would have decreased by 0.64, 0.65 and 0.69 million $km^2$ between 2003 to 2018, given the GBI was increased by 2, 3 and 4 times the daily trend. This aligns with the findings of [11] where summertime low clouds play an important role in driving sea ice melt by amplifying the adiabatic warming induced by a stronger anticyclonic circulation aloft.

## VI. DISCUSSION & FUTURE WORK

In this paper, we propose a deep learning based time-series inference method for time-varying causal inference under continuous treatment effects using stabilized weights. We introduce a probabilistic method of implementing stabilized weights through gaussian modeling. Through ablative study, we show how our proposed model balances confoundedness in case of time-delayed treatment. We presented one use-case of analyzing the causal relation between Greenland blocking and sea ice melt. Through experiments, we noticed our data-driven findings align with the literature on "increasing GBI leads to decreasing SIE". For our ongoing research, we will continue to analyze similar other use cases in the realm of Arctic Amplification, such as the effects of atmospheric processes on Arctic sea ice melt. We will further extend our work to spatiotemporal causal inference to explore the potential of neural networks in learning and answering important Earth

Science questions in the presence of temporal and spatial confounders.

## ACKNOWLEDGEMENT

This work is supported by NSF grants: CAREER: Big Data Climate Causality (OAC-1942714) and HDR Institute: HARP - Harnessing Data and Model Revolution in the Polar Regions (OAC-2118285).

## REFERENCES

[1] S. Ali, Y. Huang, X. Huang, and J. Wang. Sea ice forecasting using attention-based ensemble LSTM. Tackling Climate Change with Machine Learning Workshop at International Conference on Machine Learning (ICML). *arXiv:2108.00853*, 2021.

[2] B. Bell, H. Hersbach, A. Simmons, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, et al. The era5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741):4186–4227, 2021.

[3] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[4] I. Bica, A. Alaa, and M. Van Der Schaar. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International Conference on Machine Learning*, pages 884–895. PMLR, 2020.

[5] I. Bica, A. M. Alaa, J. Jordon, and M. van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. *arXiv preprint arXiv:2002.04083*, 2020.

[6] I. Bica, J. Jordon, and M. van der Schaar. Estimating the effects of continuous-valued interventions using generative adversarial networks. *Advances in Neural Information Processing Systems*, 33:16434–16445, 2020.

[7] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274, 2015.

[8] D. Cavalieri, C. Parkinson, P. Gloersen, and H. Zwally. Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. Technical report, NASA DAAC at the National Snow and Ice Data Center, 1996.

[9] J. L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

[10] M. M. Holland and C. M. Bitz. Polar amplification of climate change in coupled models. *Climate Dynamics*, 21(3):221–232, 2003.

[11] Y. Huang, Q. Ding, X. Dong, B. Xi, and I. Baxter. Summertime low clouds mediate the impact of the large-scale circulation on arctic sea ice. *Communications Earth & Environment*, 2(1):1–10, 2021.

[12] Y. Huang, X. Dong, B. Xi, and Y. Deng. A survey of the atmospheric physical processes key to the onset of arctic sea ice melt in spring. *Climate Dynamics*, 52(7):4907–4922, 2019.

[13] C. Huffman and E. van Gameren. Covariate balancing inverse probability weights for time-varying continuous interventions. *Journal of Causal Inference*, 6(2):20170002, 2018.

[14] C. T. Jerzak, F. Johansson, and A. Daoud. Integrating earth observation data into causal inference: challenges and opportunities. *arXiv preprint arXiv:2301.12985*, 2023.

[15] F. D. Johansson, U. Shalit, N. Kallus, and D. Sontag. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *The Journal of Machine Learning Research*, 23(1):7489–7538, 2022.

[16] J. Kerman, P. Wang, and J. Vaver. Estimating ad effectiveness using geo experiments in a time-based regression framework. Technical report, Google, Inc., 2017.

[17] B. Koch, T. Sainburg, P. Geraldo, S. Jiang, Y. Sun, and J. G. Foster. Deep learning of potential outcomes. *arXiv preprint arXiv:2110.04442*, 2021.

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

[19] M. Lechner. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3):165–224, 2011.

[20] S. Lee. A theory for polar amplification from a general circulation perspective. *Asia-Pacific Journal of Atmospheric Sciences*, 50(1):31–43, 2014.

[21] B. Lim. Forecasting treatment responses over time using recurrent marginal structural networks. *Advances in Neural Information Processing Systems*, 31, 2018.

[22] A. Marcovecchio, A. Behrangi, X. Dong, B. Xi, and Y. Huang. Precipitation influence on and response to early and late arctic sea ice melt onset during melt season. *International Journal of Climatology*, 42(1):81–96, 2022.

[23] W. N. Meier and J. S. Stewart. Arctic and Antarctic regional masks for sea ice and related data products, Version 1, 2023.

[24] F. Menchetti, F. Cipollini, and F. Mealli. Estimating the causal effect of an intervention in a time series setting: the C-ARIMA approach. *arXiv preprint arXiv:2103.06740*, 2021.

[25] R. Moraffah, P. Sheth, M. Karami, A. Bhattacharya, Q. Wang, A. Tahir, A. Raglin, and H. Liu. Causal inference for time series analysis: Problems, methods and evaluation. *Knowledge and Information Systems*, pages 1–45, 2021.

[26] A. I. Naimi, S. R. Cole, and E. H. Kennedy. An introduction to g methods. *International journal of epidemiology*, 46(2):756–762, 2017.

[27] X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

[28] M. Rantanen, A. Y. Karpechko, A. Lipponen, K. Nordling, O. Hyvärinen, K. Ruosteenoja, T. Vihma, and A. Laaksonen. The arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, 3(1):1–10, 2022.

[29] D. A. Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

[30] J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, pages 550–560, 2000.

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

[32] D. B. Rubin. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, 5(4):472–480, 1990.

[33] D. B. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[34] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):1–13, 2019.

[35] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. Causal inference for time series. *Nature Reviews Earth & Environment*, pages 1–19, 2023.

[36] U. Shalit, F. D. Johansson, and D. Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International conference on machine learning*, pages 3076–3085. PMLR, 2017.

[37] M. J. Smith, M. A. Mansournia, C. Maringe, P. N. Zivich, S. R. Cole, C. Leyrat, A. Belot, B. Rachet, and M. A. Luque-Fernandez. Introduction to computational causal inference using reproducible stata, r, and python code: A tutorial. *Statistics in medicine*, 41(2):407–432, 2022.

[38] E. A. Stuart, D. B. Rubin, and J. Osborne. Matching methods for causal inference: Designing observational studies. *Harvard University Department of Statistics mimeo*, 2004.

[39] N. Thams, R. Søndergaard, S. Weichwald, and J. Peters. Identifying causal effects using instrumental time series: Nuisance iv and correcting for the past. *arXiv preprint arXiv:2203.06056*, 2022.

[40] L. van Garderen, F. Feser, and T. G. Shepherd. A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline. *Natural Hazards and Earth System Sciences*, 21(1):171–186, 2021.

[41] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5):1–46, 2021.