AI Based Multi-Satellite Earth Remote Sensing and Causal Understanding of Earth Processes

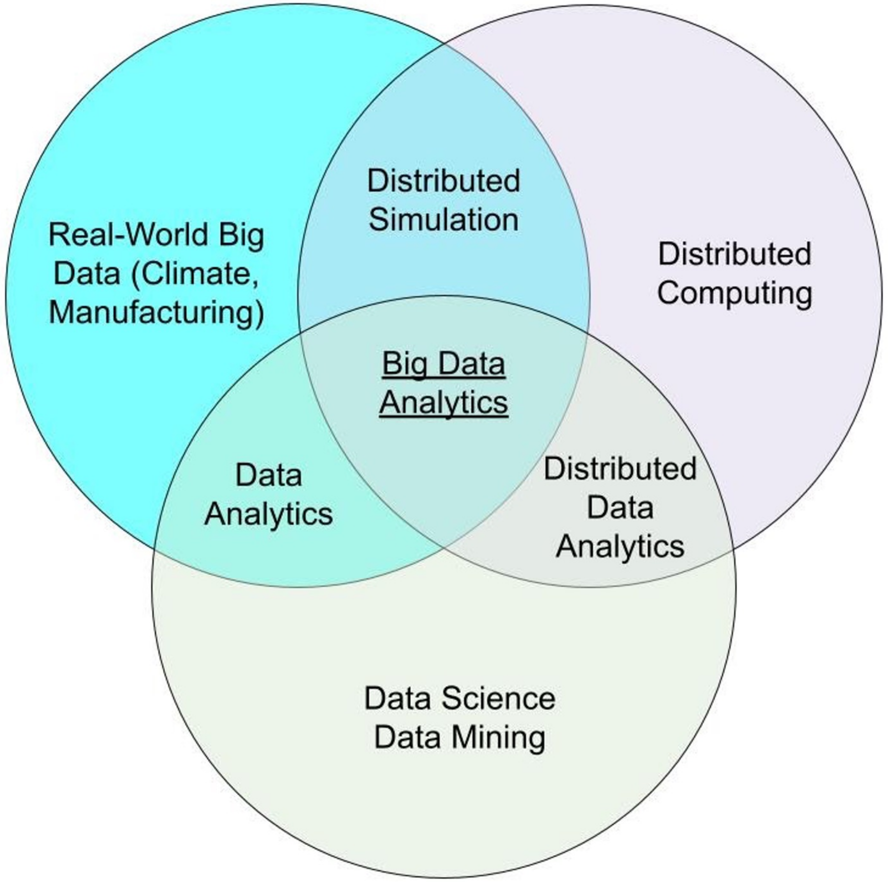
Jianwu Wang
Big Data Analytics Lab, UMBC (bdal.umbc.edu)
iHARP: NSF HDR Institute for Harnessing Data and Model Revolution in the Polar Regions (iharp.umbc.edu)

Outline

- My Research Overview
- Earth AI Challenges
  - Ever-increasing data volume
  - Data heterogeneity/variety
  - Method focus difference
- Study 1: Deep Multi-Sensor Domain Adaptation on Active and Passive Satellite Remote Sensing Data
  - Deals with data volume and data variety challenges
  - More at our papers at BigData2020 and SigSpatial2022
- Study 2: Quantifying Causes of Arctic Amplification via Deep Learning based Time-series Causal Inference
  - Bridge method focus difference between data science and Earth science
  - More at our paper at ICMLA 2023

My Research Portfolio

- Center on **big data analytics** with three connected and supporting areas
  - **Distributed computing**: key enabling computing infrastructure for big data
  - **Data science**: core research topic to learn patterns from big data
  - **Real-world big data**: research frontier to identify new research challenges and evaluate research contribution and impacts
- **Holistic** and **end-to-end** research
  - Identify new and important research challenges
  - Propose **integrative** solutions
  - Evaluate from many aspects: efficiency, effectiveness, helps to domain scientists, etc.



Recent Research Topics

Earth AI and Informatics

- Sea ice forecasting: [BigData2021], [IGARSS2022], [BDCAT2022]
- Multi-satellite cloud retrieval: [BigData2020], [SigSpatial2022]
- Dust detection: [BigData2020], [RemoteSensing2021]
- Ice bed topography: [ICMLA2023]
- Climate data clustering: [ECMLPKDD2023]
- Ocean eddy detection: [IGARSS2023]

Distributed Computing

- Reproducible big data analytics in the cloud: [TransCloudComputing2023]
- Edge-cloud for stream analytics: [FGCS2022]

Causal AI

- Deep learning based causal inference: [ICMLA2023]
- Causality benchmarking for sea ice: [FrontiersinBigData2021]
- Big data causality: [BigData2019], [SMDS2020]

Streaming Data Analytics

- Flood detection on edge: [ICMLA2023]
- Camouflaged object detection on edge: [SmartComp2023]
- Streaming data forecasting: [BigData2020]

Challenges in Earth AI

- Ever-increasing data volume
  - It is estimated available Earth data (simulation and observation) will increase from 100 PB in 2020 to 350 PB in 2030 [1]
- Data heterogeneity and variety
  - Data are generated from various sources (simulation, flight, satellites, etc.) with different resolution, spatial and temporal coverage
- Method focus difference: focuses of traditional data science are not applicable to climate science [2]
  - Accuracy vs. Physical plausibility and causality
- Because of these challenges, there are growing interests of applying AI and data science techniques in Earth science
  - NSF AI Institute and HDR Institute on Climate and Environment
  - AI/DS related programs in NASA ROSES programs
  - NOAA AI and Data strategy

A Review of Earth Artificial Intelligence, Computers & Geosciences, volume 159, 105034, DOI:10.1016/j.cageo.2022.105034, Elsevier, 2022

Outline

- My Research Overview
- Earth AI Challenges
  - Ever-increasing data volume
  - Data heterogeneity/variety
  - Method focus difference
- Study 1: Deep Multi-Sensor Domain Adaptation on Active and Passive Satellite Remote Sensing Data
  - Deals with data volume and data variety challenges
  - More at our papers at BigData2020 and SigSpatial2022
- Study 2: Quantifying Causes of Arctic Amplification via Deep Learning based Time-series Causal Inference
  - Bridges method focus difference between data science and Earth science
  - More at our paper at ICMLA 2023

Heterogeneous Active and Passive Sensing Data

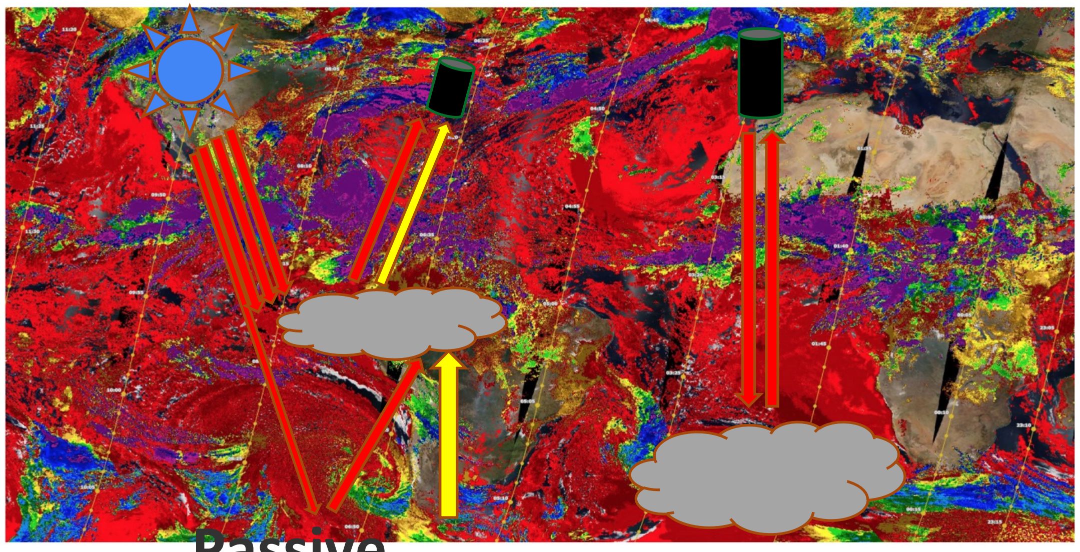


Fig. 1. An example showing the spatial coverage differences between VIIRS (global coverage) and CALIOP (yellow lines) data (Credits: NASA).

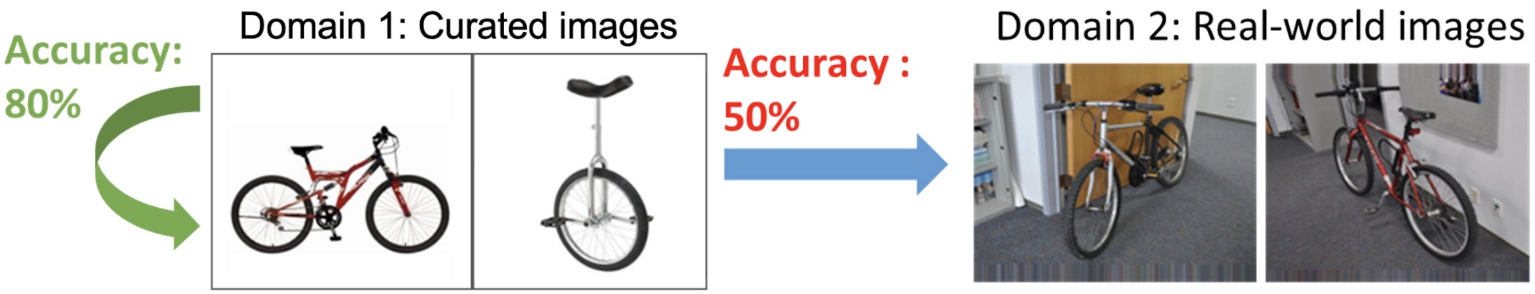
	Name
1	CALIOP_N_Clay_1km
2	CALIOP_N_Clay_5km
3	CALIOP_Liq_Fraction_1km
4	CALIOP_Liq_Fraction_5km
5	CALIOP_Ice_Fraction_1km
6	CALIOP_Ice_Fraction_5km
7	CALIOP_Clay_Top_Altitude
8	CALIOP_Clay_Base_Altitude
9	CALIOP_Clay_Top_Temperature
10	CALIOP_Clay_Base_Temperature
11	CALIOP_Clay_Optical_Depth_532
12	CALIOP_Clay_Opacity_Flag
13	CALIOP_Clay_Integrated_Attenuated_Backscatter_532
14	CALIOP_Clay_Integrated_Attenuated_Backscatter_1064
15	CALIOP_Clay_Final_Lidar_Ratio_532
16	CALIOP_Clay_Color_Ratio
17	CALIOP_Alay_Top_Altitude
18	CALIOP_Alay_Base_Altitude
19	CALIOP_Alay_Top_Temperature
20	CALIOP_Alay_Base_Temperature
21	CALIOP_Alay_Integrated_Attenuated_Backscatter_532
22	CALIOP_Alay_Integrated_Attenuated_Backscatter_1064
23	CALIOP_Alay_Color_Ratio
24	CALIOP_Alay_Optical_Depth_532
25	CALIOP_Alay_Aerosol_Type_Mode

	Name	Description
1	VIIRS_SZA	viirs solar zenith angle in degree
2	VIIRS_SAA	viirs solar azimuthal angle in degree
3	VIIRS_VZA	viirs viewing zenith angle in degree
4	VIIRS_VAA	viirs viewing azimuthal angle in degree
5	VIIRS_M1	Band wavelength range 0.402-0.422μm
6	VIIRS_M2	Band wavelength range 0.436-0.454μm
7	VIIRS_M3	Band wavelength range 0.478-0.488μm
8	VIIRS_M4	Band wavelength range 0.545-0.565μm
9	VIIRS_M5_B	Band wavelength range 0.662-0.682μm
10	VIIRS_M6	Band wavelength range 0.739-0.754μm
11	VIIRS_M7_G	Band wavelength range 0.846-0.885μm
12	VIIRS_M8	Band wavelength range 1.23-1.25μm
13	VIIRS_M9	Band wavelength range 1.371-1.386μm
14	VIIRS_M10_R	Band wavelength range 1.58-1.64μm
15	VIIRS_M11	Band wavelength range 2.23-2.28μm
16	VIIRS_M12	Band wavelength range 3.61-3.79μm
17	VIIRS_M13	Band wavelength range 3.97-4.13μm
18	VIIRS_M14	Band wavelength range 8.4-8.7μm
19	VIIRS_M15	Band wavelength range 10.26-11.26μm
20	VIIRS_M16	Band wavelength range 11.54-12.49μm

- Background: CALIOP active sensing data and VIIRS passive sensing data are heterogenous in spatial coverage, data features and data quality
- Challenge: How to leverage the high data quality of active sensors and the global spatial coverage of passive sensors so that we can retrieve high quality cloud properties globally?

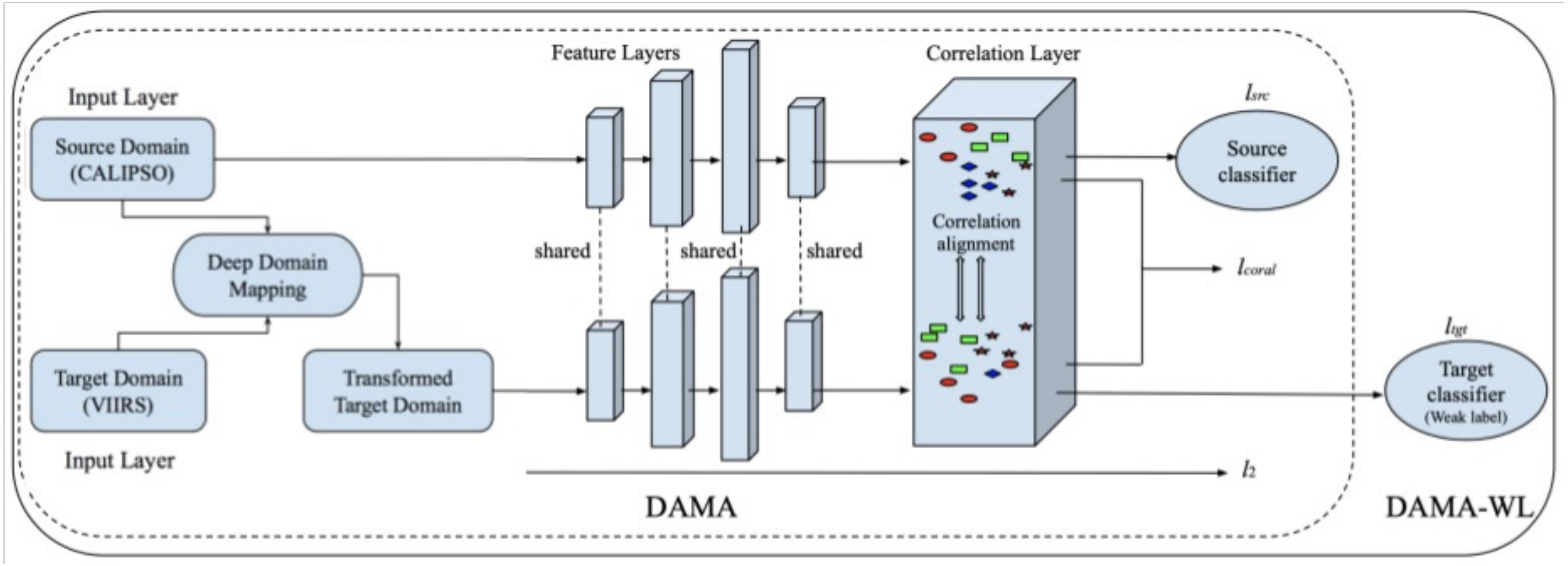
Domain Adaptation

- One type of transfer learning techniques to deal with distribution shift between two domains
- Most existing deep domain adaptation methods are for homogeneous domain adaptation in image classification



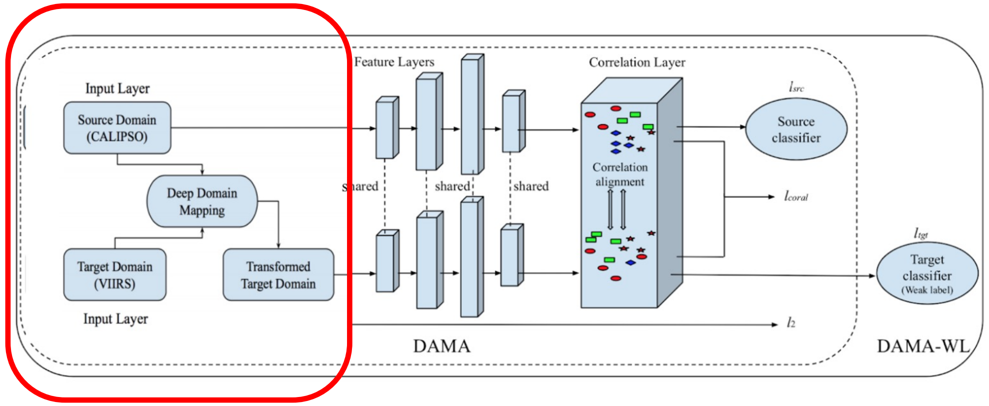
- Our work is the first deep domain adaptation approach in heterogeneous domains and remote sensing

Our Proposed Deep Domain Adaptation Model



- DAMA model: Domain mapping + feature extraction + correlation alignment + source classifier
- DAMA-WL model: Domain mapping + feature extraction + correlation alignment + source classifier + weak target classifier

Deep Domain Mapping

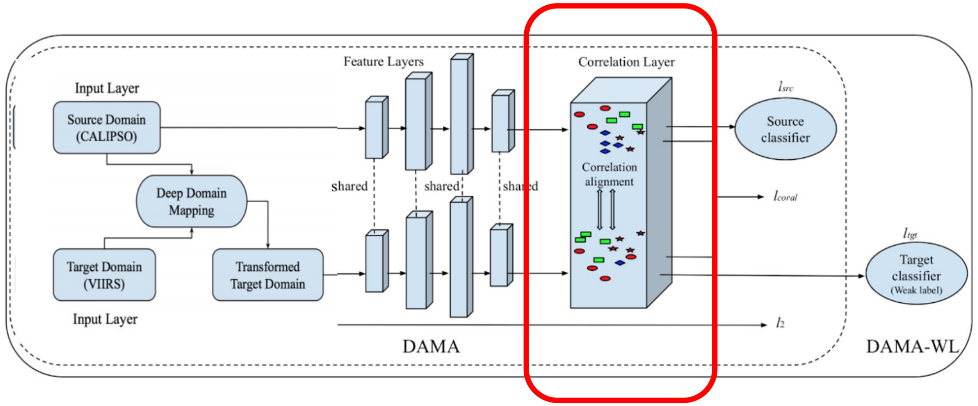


- Data collocation
  - Passive sensor data is collocated on active sensor's track (on-track)

- Transform the target domain into source domain feature space with L2 loss

$$l_2 = \frac{1}{n_t} \sum_{(i=1)}^{n_t} (DDM(u_i) - x_i)^2$$

Correlation Alignment



- **Correlation loss**
  - Measure the distance between the second order statistics (covariances) of the source and target data

$$l_{coral} = \frac{1}{4d^2}||C_s - C_t||_F^2$$

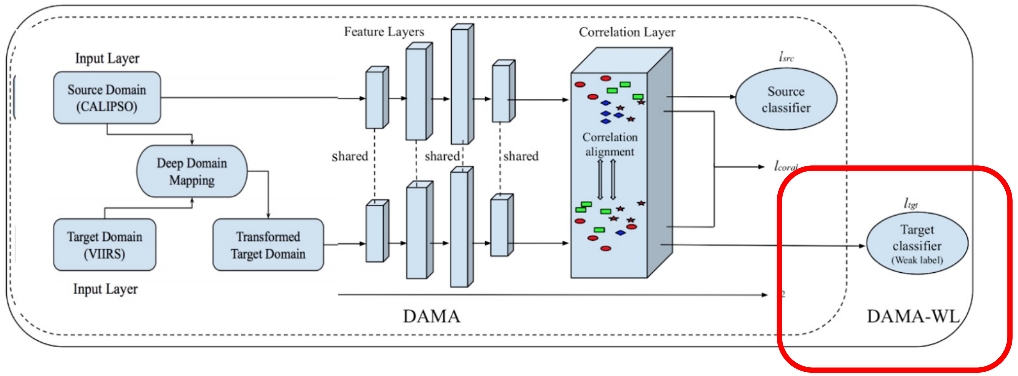
$$C_s = \frac{1}{n_s - 1}(D_s^T D_s - \frac{1}{n_s}(\mathbf{1}^T D_s)^T(\mathbf{1}^T D_s))$$

$$C_t = \frac{1}{n_t - 1}(D_t^T D_t - \frac{1}{n_t}(\mathbf{1}^T D_t)^T(\mathbf{1}^T D_t))$$

- **Combine correlation loss with source classification loss**

$$l = l_{src} + \sum_{(i=1)}^{t} \lambda_i l_{coral}$$

Domain Adaptation with Weak Supervision



- Incorporate the weak label information from the **target** domain
- **DAMA-WL**: weakly supervised learning

$$l^* = l_{src} + \sum_{(i=1)}^{t} \lambda_i l_{coral} + l_{tgt}$$

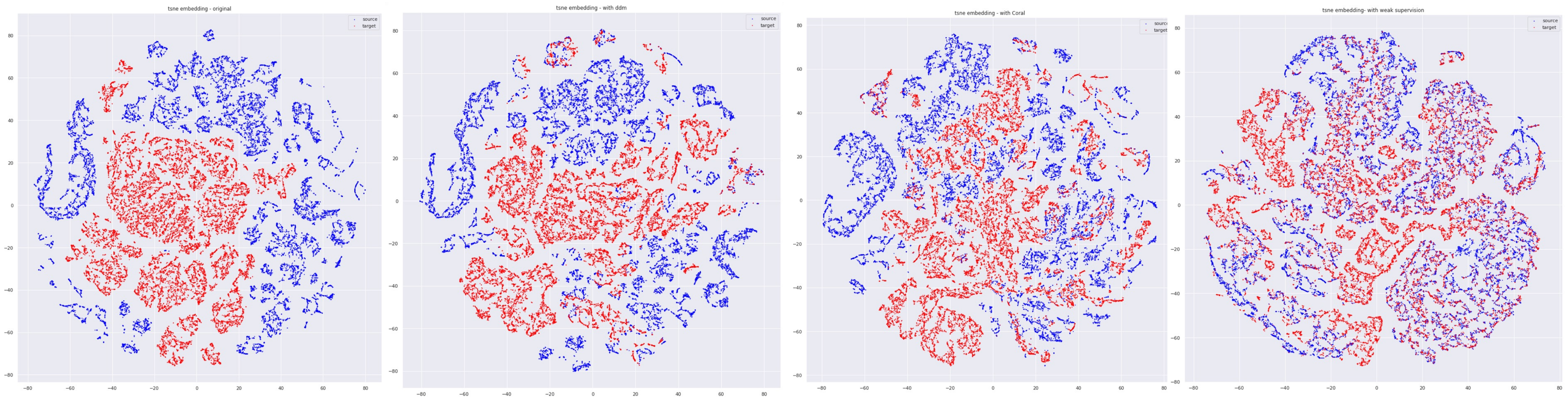
Experiments

TABLE II
ACCURACY ON PREDICTING THE CLOUD TYPES ON VIIRS (TARGET) DATASET WITH WEAK LABEL.

Models - Single Domain	**Label**	**Source**	**Target**	**Day-005**	**Day-013**	**Day-019**	**Day-024**	**Day-030**	**Jan. 2017**
Random Forest	CALIOP	VIIRS	VIIRS	0.957	0.947	0.934	0.933	0.917	0.939
Random Forest-WL	VIIRS	VIIRS	VIIRS	0.905	0.911	0.883	0.878	0.854	0.889
MLP-VIIRS	CALIOP	VIIRS	VIIRS	0.896	0.907	0.878	0.877	0.865	0.885
MLP-CALIOP	CALIOP	CALIOP	CALIOP	1.000	1.000	1.000	1.000	1.000	1.000
Models - Multiple Domains									
Domain Mapping Only	CALIOP	CALIOP	VIIRS	0.910	0.913	0.890	0.896	0.885	0.899
Correlation Align. Only	CALIOP	CALIOP	VIIRS	0.428	0.473	0.394	0.378	0.321	0.408
DAMA	CALIOP	CALIOP	VIIRS	0.956	0.948	0.934	0.936	0.926	0.941
DAMA-WL	CALIOP + VIIRS	CALIOP	VIIRS	**0.963**	**0.964**	**0.958**	**0.958**	**0.949**	**0.960**

- DAMA outperforms the domain adaptation baselines by ~5% to ~54%
- DAMA-WL brings additional ~2% accuracy improvement compared to the DAMA

Visualization of Learned Representations



- We use t-SNE techniques to visualize the learned representations
- The result shows data records in our source domain and target domain are more and more mixed along the pipeline: Original->DDM->Coral -> WL
  - Indicates the success of our domain adaptation approach

VDAM: VAE based Domain Adaptation for Cloud Property Retrieval from Multi-satellite Data

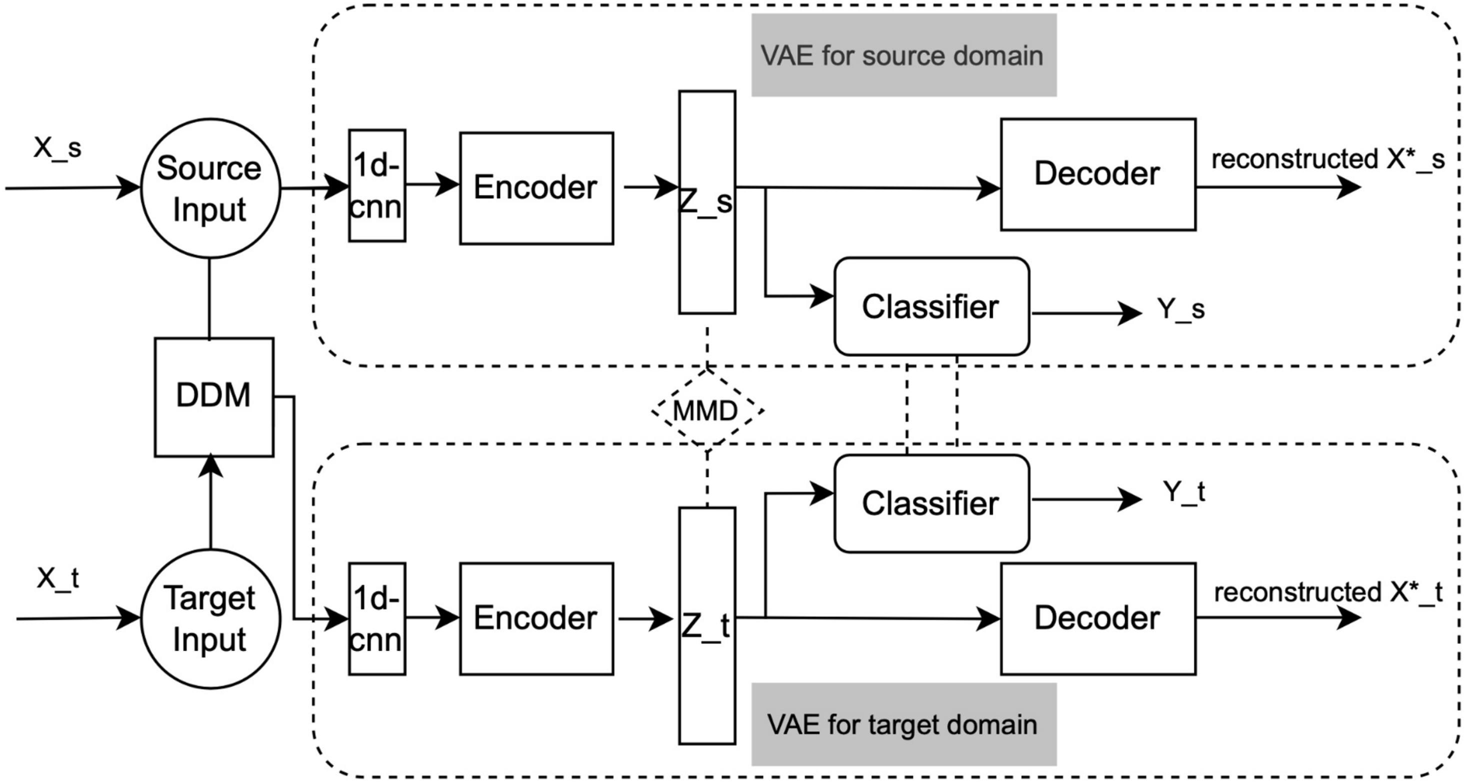
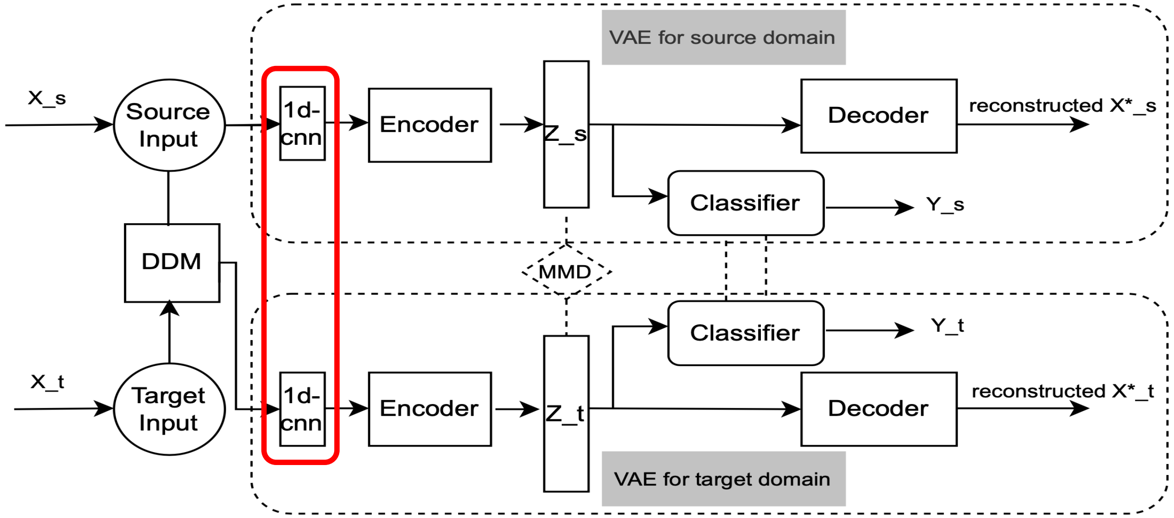


Figure 2: Network architecture of the proposed VAE based domain adaptation. For each domain, we construct a customized VAE model, which contains an encoder to extract latent features, a decoder for input data reconstruction, and a classifier for cloud property retrieval. The domain discrepancy between the source domain and target domain is minimized by a domain alignment technique (MMD).

Spatial Feature Representation Learning

- **Apply 1D-CNNs on the source branch and target branch**
  - 1-D data sequence follow the CALIOP orbiting track
  - Capture spatial dependency among the pixels sequence



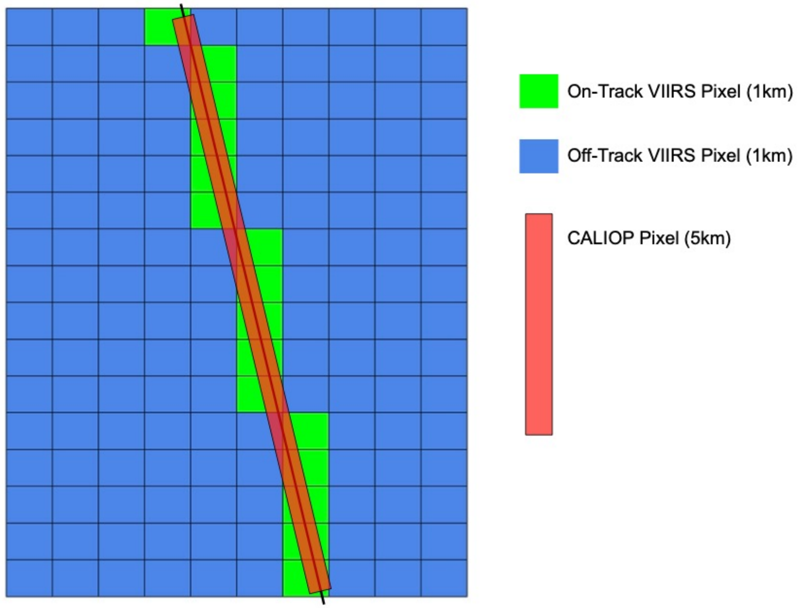
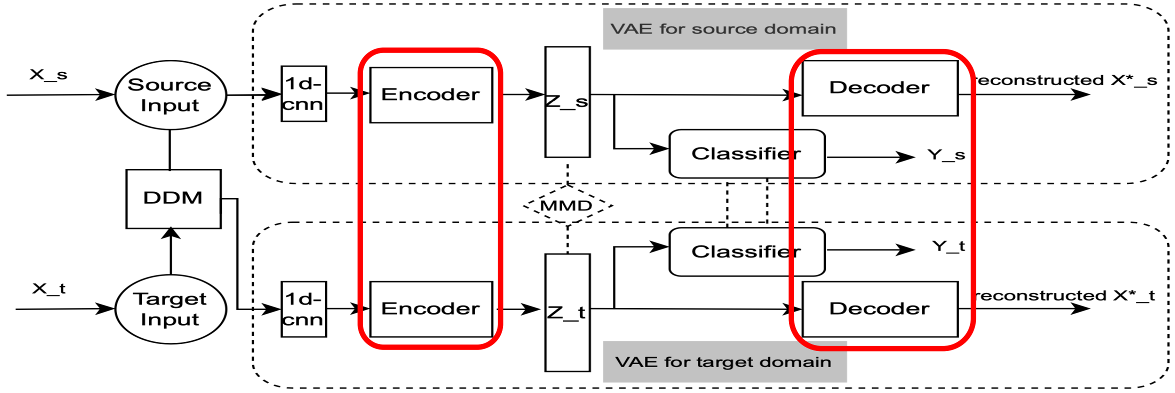


Figure 3: Illustration of on-track and off-track VIIRS pixels as well as the collocated CALIOP track and pixels. In our study, two 1D-CNN layers are applied on the overlapped VIIRS (green) and CALIOP (red) pixel sequences, respectively.

Variational Autoencoder Networks



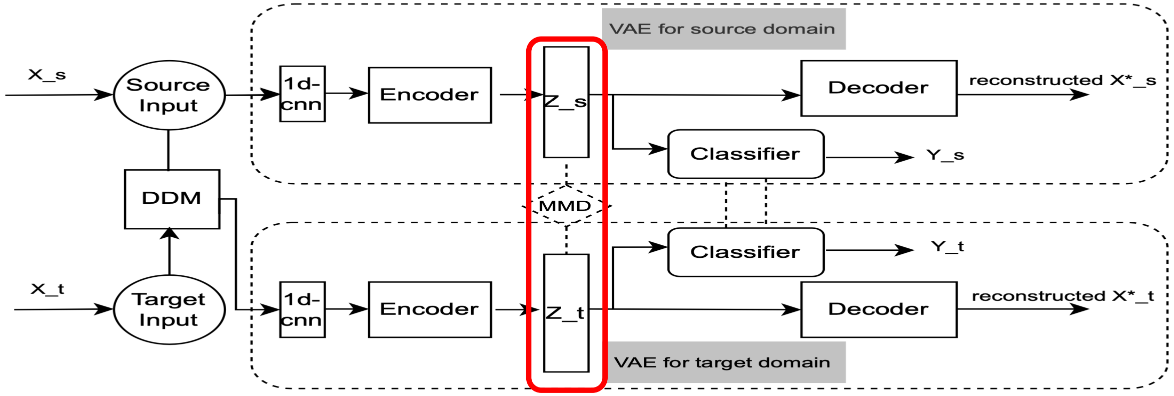
- Encoder maps the input data into a latent feature space, and approximates the posterior probability by a parameterized model
- Maximize the variational lower bound by optimizing the parameters θ and ϕ of the neural network
  - l_{KL} : minimize the KL divergence between approximate posterior distribution and true posterior distribution
  - l_R : maximize the expectation of the reconstructed data points sampled from the latent vector

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}^{(i)}) = -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z})\right]$$

- VAE for source domain (CALIOP) and VAE for target domain (VIIRS)

$$l_{VAE}^{s} = l_{KL}^{s} + l_{R}^{s} \text{ and } l_{VAE}^{t} = l_{KL}^{t} + l_{R}^{t}$$

MMD based Domain Alignment

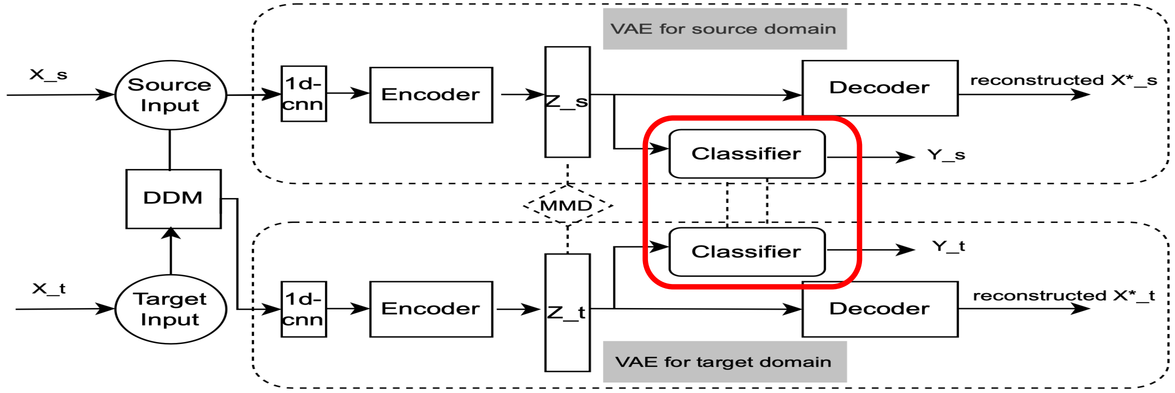


- **To minimize the distribution discrepancy between the source and target domain**
  - Add a feature adaptation layer to the auto encoder pairs of the source and target domain to measure the domain discrepancy loss
- **MMD (maximum mean discrepancy) loss**
  - Convert two sets of source and target domain features to a common reproducing kernel Hilbert space (RKHS)
  - Representing distances between distributions as distances between kernel embedding of distributions

$$\mathrm{MMD}(\mathbf{X}, \mathbf{Y}) = \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(x_i) - \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(y_i) \right\|_{\mathcal{H}},$$

where $\mathcal{H}$ is a universal RKHS, $|| * ||_{\mathcal{H}}$ is RKHS norm, and $\phi : X \rightarrow \mathcal{H}$.

Label Space Alignment



- **Fully connected classification layer for source domain (CALIOP)**
  - Standard cross entropy loss for strong source labels
- **Fully connected classification layer for target domain (VIIRS)**
  - Weighted cross entropy loss for weak target labels
  - Assign a higher weight when weak label from target domain differs to that of source domain
  - Encourage model to learn toward the more challenging area that the classifier is uncertain about

$$w_{s_i,t_i} = \begin{cases} 1.5 & \text{if label of } s_i \text{ differs to label of } t_i \\ 1.25 & \text{if label of } s_i \text{ is mixed cloud} \\ 1 & \text{if label of } s_i \text{ equals to label of } t_i \end{cases}$$

End-to-end Joint Training

- The model trained jointly in an end-to-end fashion in order to align the heterogeneous source and target domains and build the domain invariant classifier
- The joint loss is composed of 1) the loss of deep domain mapping, 2) the losses of VAE losses for source domain and target domain, 3) the loss of source classifier, 4) the loss of MMD based domain alignment and 5) the loss of target classifier with weak label

$$l = l_2 + l_{vae}^s + l_{vae}^t + l_{mmd} + l_C^s + l_C^t$$

Experiments

Table 1: Accuracy on predicting the cloud types on VIIRS (target) dataset.

Models - Single Domain	**Label**	**Source**	**Target**	**Jan. 2014**	**Jan. 2015**	**Jan. 2016**	**Jan. 2017**
Random Forest	CALIOP	VIIRS	VIIRS	0.815	0.823	0.821	0.828
Random Forest-WL	VIIRS	VIIRS	VIIRS	0.775	0.783	0.781	0.790
MLP-VIIRS	CALIOP	VIIRS	VIIRS	0.805	0.811	0.810	0.815
MLP-CALIOP	CALIOP	CALIOP	CALIOP	1.0	1.0	1.0	1.0
Models - Multiple Domains							
Auto Encoder model	CALIOP + VIIRS	CALIOP	VIIRS	0.821	0.833	0.830	0.836
Model without domain mapping	CALIOP + VIIRS	CALIOP	VIIRS	0.512	0.533	0.530	0.539
Model without 1d-CNN	CALIOP + VIIRS	CALIOP	VIIRS	0.855	0.863	0.861	0.866
DAMA-WL[12]	CALIOP + VIIRS	CALIOP	VIIRS	0.842	0.848	0.846	0.851
The proposed model	CALIOP + VIIRS	CALIOP	VIIRS	**0.868**	**0.872**	**0.871**	**0.878**

Ablation study

Conclusions from Study 1

- Utilizing data from multiple satellites jointly, we can achieve better information retrievals for targeted geophysics variables

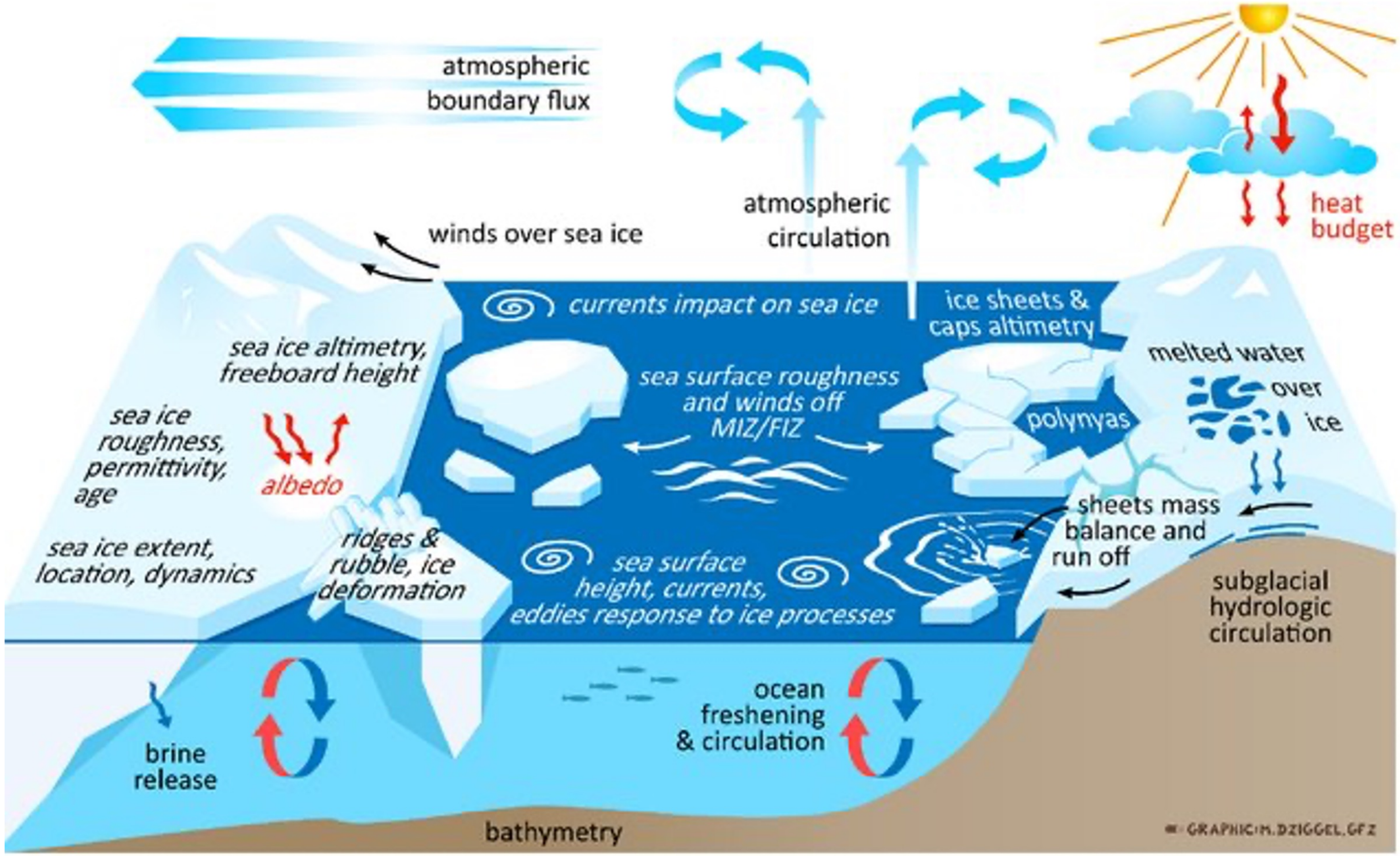
- We proposed deep domain adaptation methods with heterogeneous domain mapping and correlation alignment to employ both active and passive sensing data in cloud type detection

- Our VAE based deep domain adaptation model outperforms the first model (DAMA-WL) by capturing spatial feature on orbiting track, MMD based domain alignment and label space alignment (customized loss weights)

Outline

- My Research Overview
- Earth AI Challenges
  - Ever-increasing data volume
  - Data heterogeneity/variety
  - Method focus difference
- Study 1: Deep Multi-Sensor Domain Adaptation on Active and Passive Satellite Remote Sensing Data
  - Deals with data volume and data variety challenges
  - More at our papers at BigData2020 and SigSpatial2022
- Study 2: Quantifying Causes of Arctic Amplification via Deep Learning based Time-series Causal Inference
  - Bridges method focus difference between data science and Earth science
  - More at our paper at ICMLA 2023

Background: Dynamics Of Arctic Amplification



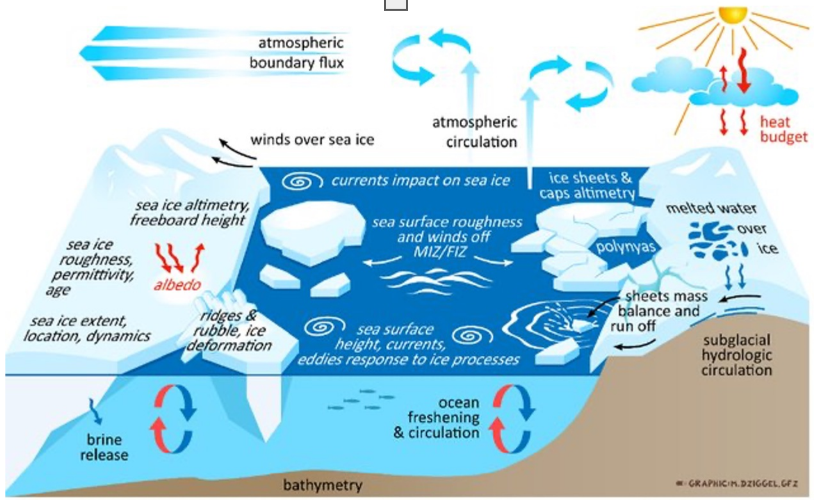
Sketch of different processes and interactions involving the cryosphere [3]

Pressing Questions

What is the cause / effect of ice melt?

How can I represent the complex dynamic interactions in data-driven way?

What are my regions of interest?



What datasets can help me in my task?

Do I look in the past or future of sea ice?

How will I validate my findings?

BACKGROUND ON CAUSAL INFERENCE

The process of inferring/quantifying the causal influence (***strength***) of one event, process, policy or treatment (*a* ***cause***) on another event, process, state or outcome (an ***effect***).

Common methods:

- Calculating average causal effect (ACE) via intervention, i.e., do-calculus [4]
- Calculating average treatment effect (ATE) via potential outcomes framework [5]

Problem Statement

Given the data Z_t (covariates) at timestep t, we want to forecast observed (factual) as well as counterfactual values of sea ice, i.e., potential outcome Y_{t+n}, at timestep $t + n$ by intervening/perturbing on certain atmospheric processes, i.e., time-varying treatment X_t

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

X_{hat} represents intervened treatment and X represents treatment without intervention or placebo effect

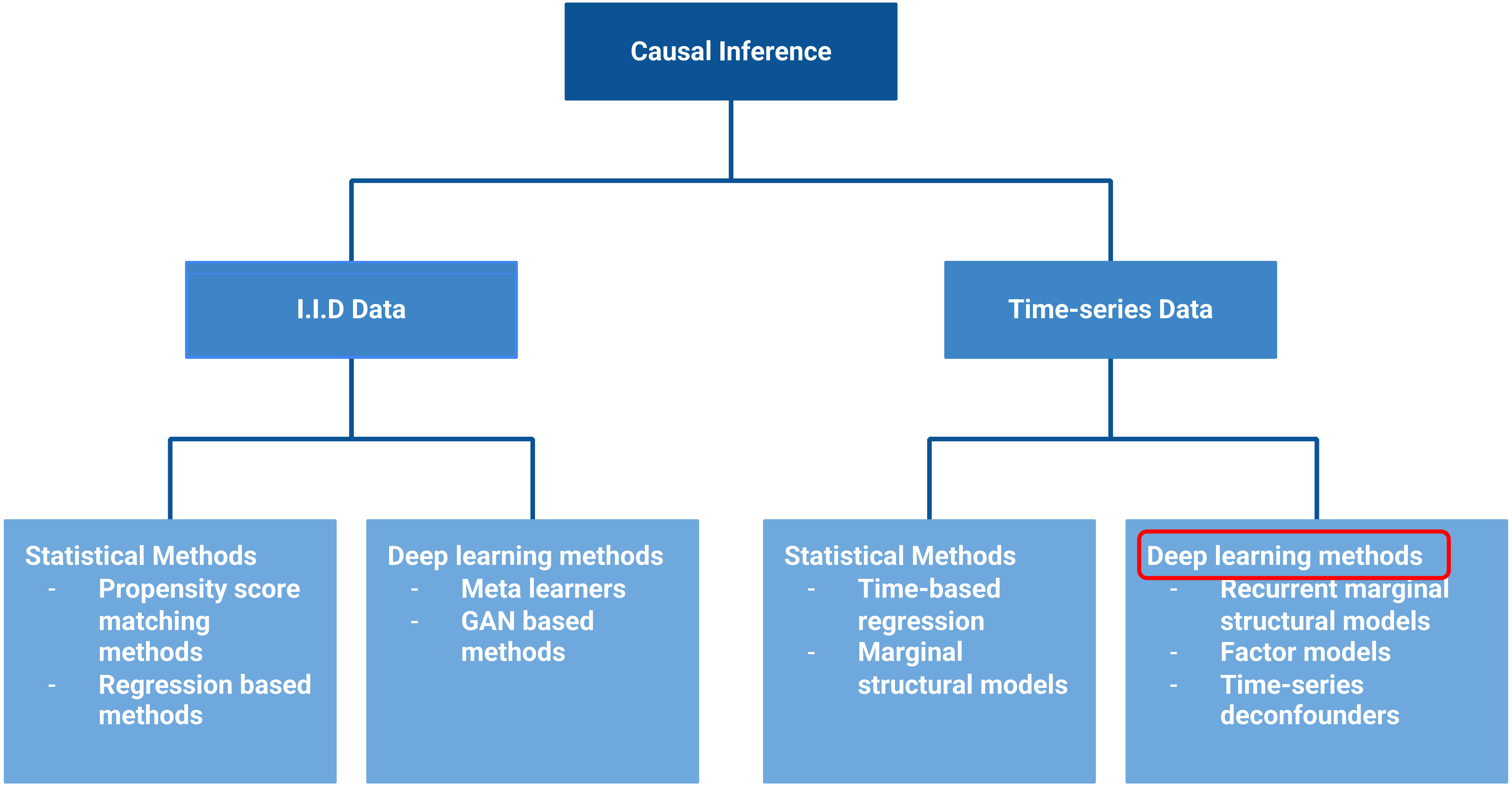
We want to estimate the lagged average treatment effect (LATE) of atmospheric process X, on the sea-ice variations, after a lag of l timesteps

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

Challenges and Proposed Solutions

CHALLENGE	PROPOSED SOLUTION
Predicting potential time-varying outcome with low predictive loss and high accuracy	Utilizing deep learning models for time-series data
Inability to evaluate the model’s performance for counterfactual predictions	Evaluating models on synthetic data
Tackling time-varying confounding effects	Balancing strategies: inverse probability of treatment weighting (IPTW), stabilized weighting (SW)

Related Work: Causal Inference Methods



Balancing Strategies – G-Methods

Generalized Propensity Score [6]

$$Prob(X_t | X_{t-1}, Z_t)$$

Inverse Probability of Treatment Weight [7]

$$IPTW = \Pi_{t=1}^{k} \frac{1}{f(\bar{X}|\bar{Z})}$$

leads to unstable estimates and inflated variance [5]

where, $\bar{X} = (X_1, X_2, ..., X_t)$ $\bar{Z} = (Z_1, Z_2, ..., Z_t)$,

Balancing Strategies (Cont')

Stabilized Weights for Binary Treatment [8]

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

can be estimated using logistic regression in case of binary/discrete treatment

where, $\bar{X} = (X_1, X_2, ..., X_t)$ $\bar{Z} = (Z_1, Z_2, ..., Z_t)$

What to do if treatment is continuous?

Balancing Strategies (2)

Stabilized Weights for Continuous Treatment

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

Here, f is the probability density function

where, $\bar{X} = (X_1, X_2, ..., X_t)$ $\bar{Z} = (Z_1, Z_2, ..., Z_t)$

Estimating the **probability density function (PDF)** mathematically [9]

$$f(X_t|\bar{X}_{t-1}) = \frac{1}{\sigma\sqrt{2\pi}} exp\{-\frac{1}{2\sigma^2}[X_t - (\bar{X}_{t-1}^T \alpha)]^2\}$$

$$f(X_t|\bar{X}_{t-1}, \bar{Z}_t) = \frac{1}{\sigma\sqrt{2\pi}} exp\{-\frac{1}{2\sigma^2}[X_t - (\bar{X}_{t-1}^T \alpha + \bar{Z}_t^T \beta)]^2\}$$

What We Propose

Calculating Stabilized Weights using **Probabilistic Modeling.**

We leverage Gaussian Mixture Model (**GMM**) for density estimation at every timestep $\boldsymbol{t}$

Learn the underlying distribution of treatment history and covariates using GMM to get its mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

Calculate the probability density of current treatment using the density estimation formula:

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

Calculating Stabilized Weights for Continuous Treatment

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

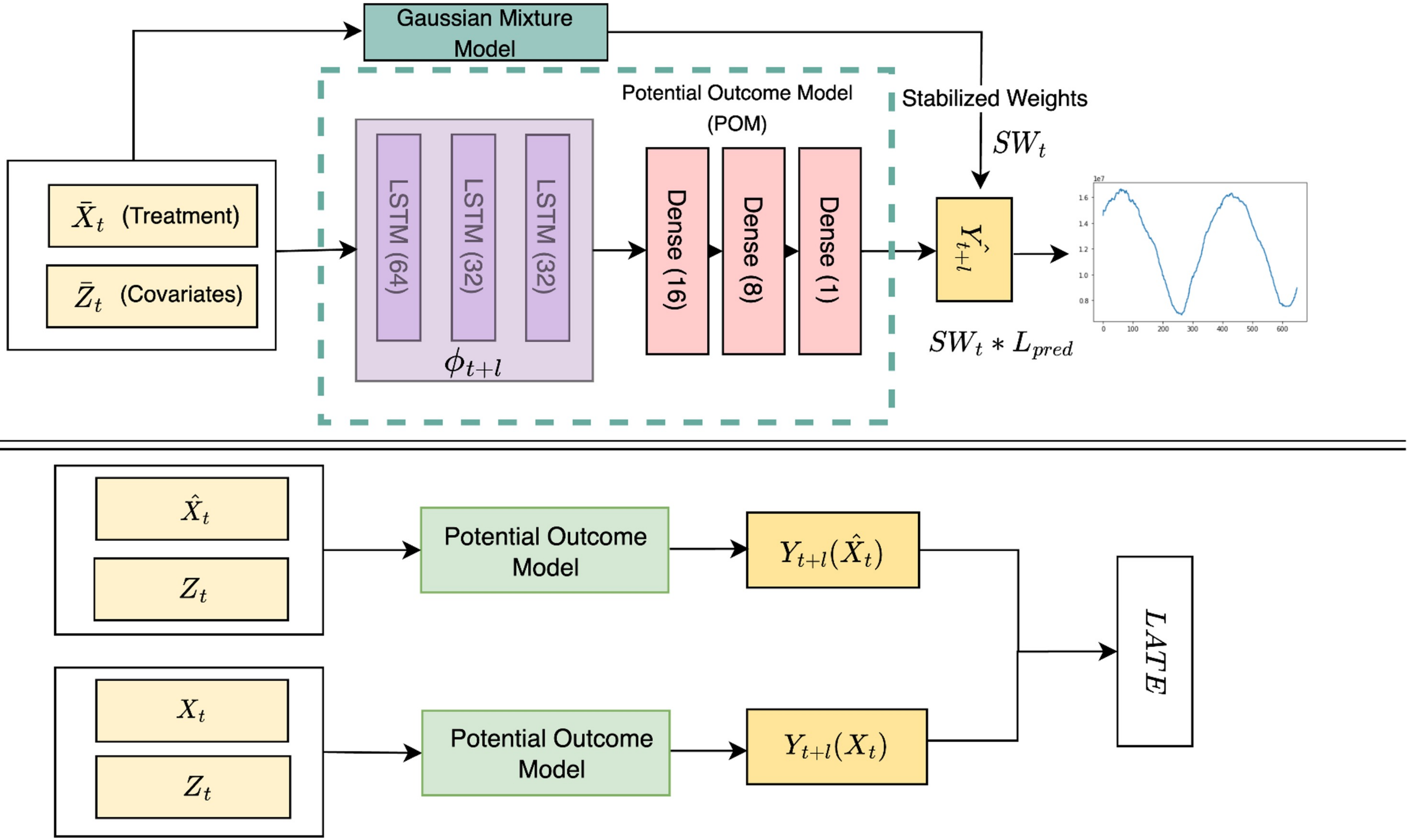
Function `PDF_calc(`X, $\bar{X}_{hist}$, $\bar{Z} = []$`):`
    `// Concatenate the treatment history and covariates`
    $\overline{XZ} \leftarrow concat(\bar{X}_{\text{hist}}, \bar{Z})$;
    $l \leftarrow$ length of sequence XZ;
    for $i \leftarrow 1$ **to** l **do**
        $n_{\text{comp}} \leftarrow$ Number of components for GMM;
        `// Create a GMM object`
        $gmm \leftarrow$ GaussianMixture(n_{comp}) ;
        `// Fit the GMM model`
        gmm.fit($\overline{XZ}_i$);
        `// Extract model parameters:`
        $(\alpha, \mu, \Sigma) \leftarrow$ (gmm.weights, gmm.means, gmm.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**
    $SW[k] \leftarrow \frac{X\_pdf[k]}{XZ\_pdf[k]}$;

TCINET (Time-Series Causal Inference Network)

TRAIN PHASE

INFERENCE PHASE



Dataset: Synthetic Data

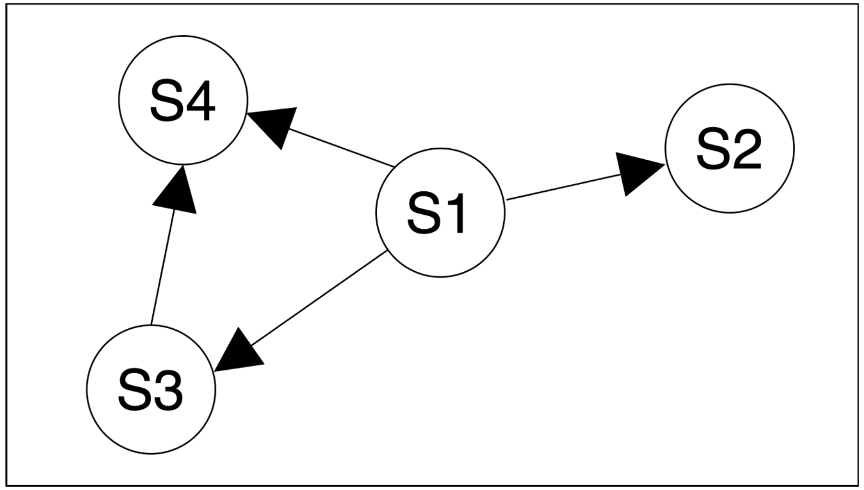
Non-linear time-series

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

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

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

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



(where, ε is Gaussian noise)

Dataset: Observational Data

- Time period: 1979 – 2018
  - Daily data: 14,610 temporal records
- Geolocation:
  - Barents Sea, Kara Sea
- Sources:
  - Nimbus-7 SSMR and DMSP SSM/I-SSMIS passive microwave data version.
  - ERA-5 global reanalysis product

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	[3 x 10^6, 14 x 10^6]	NSIDC

Evaluation Metrics

- Root Mean Square Error

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

- Precision Estimation of Heterogeneous Effects (PEHE)

$$\sqrt{PEHE} = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(ATE_i - A\hat{T}E_i)^2}$$

where, ATE is the Average Treatment Effect

Results – Synthetic Data

TABLE III
CAUSAL INFERENCE MODEL PERFORMANCE ON SYNTHETIC DATA UNDER FIXED TREATMENT FOR ONE-STEP AHEAD PREDICTION (TRUE ATE = -0.0514)

MODEL	TEST RMSE	ESTIMATED LATE	PEHE
TCINET†	1.079	-0.040	1.132
TCINET-LR	1.142	-0.037	1.227
TCINET-GMM	1.023	-0.051	1.004
CIV [40]	N/A	-0.219	N/A
CAUSAL IMPACT [7]	N/A	-0.060	1.110

TABLE IV
CAUSAL INFERENCE MODEL PERFORMANCE ON SYNTHETIC DATA UNDER CONTINUOUS TREATMENT FOR ONE-STEP AHEAD PREDICTION (TRUE ATE = -0.0514)

MODEL	TEST RMSE	ESTIMATED LATE	PEHE
TCINET†	1.026	-0.036	1.221
TCINET-LR	1.000	-0.049	1.143
TCINET-GMM	0.998	-0.050	1.102
CIV [40]	N/A	0.515	N/A
CAUSAL IMPACT [7]	N/A	-0.040	1.112

Results – Observational Data

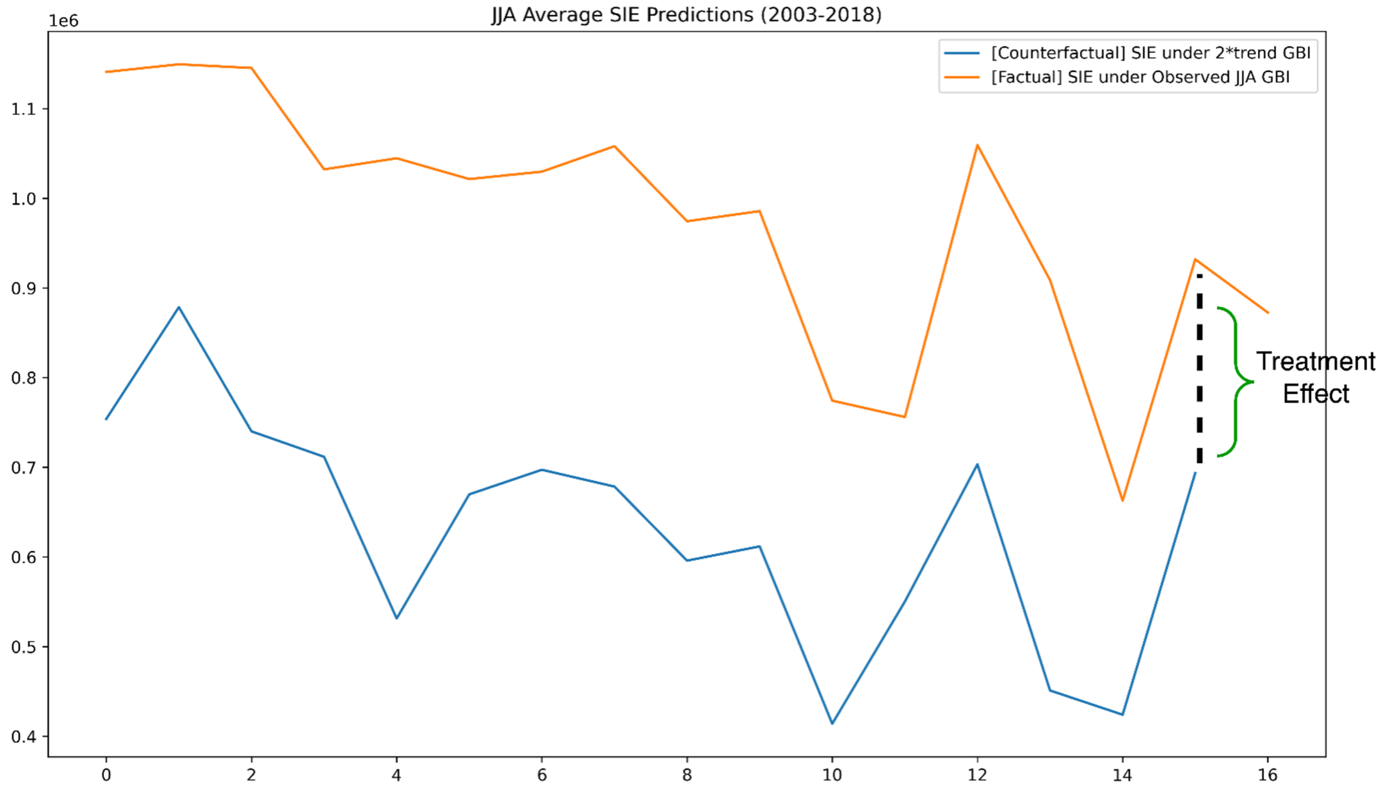


Fig: Annual mean sea ice extent (SIE) predictions under interventional GBI where each data point represents summer (JJA) mean SIE for that year.

Case study:

How does increased Greenland Blocking (GBI) affect summertime regional Arctic sea ice melting given snowfall rate and solar radiation data.

Treatments:

- Treatment 1: GBI✝ = 40 -year-averaged daily GBI
- Treatment 2: GBI + 2x GBI✝
- Treatment 3: GBI + 3x GBI✝
- Treatment 4: GBI + 4x GBI✝

TABLE V

CAUSAL EFFECT ESTIMATION BY TCINET-GMM (IN $million\ km^2$) ON OBSERVATIONAL ARCTIC DATA UNDER CONTINUOUS TREATMENTS.

TREATMENT	ESTIMATED LATE (TCINET-GMM)
40YR-AVG-GBI	-0.60 $million\ km^2$
GBI (2× TREND)	-0.64 $million\ km^2$
GBI (3× TREND)	-0.65 $million\ km^2$
GBI (4× TREND)	-0.69 $million\ km^2$

Conclusions from Study 2

- We propose a time-series based causal inference model for **continuous treatment effect estimation**
- We propose a novel probabilistic weighting technique to balance **time-varying confoundedness** by leveraging Gaussian Mixture Model (GMM)
- We compare model performance with state-of-the-art (SOTA) methods using **synthetic time-series** data for fixed and continuous time-delayed treatments
- We utilize the developed model to quantify the causal effects of thermodynamic processes on the **Arctic sea ice melt** and our results aligns with physics based understanding in [10]
- To the best of our knowledge, we are the first one to calculate **stabilized weights** for **continuous treatment effects estimation**

Acknowledgements

- Funding from NASA and NSF
  - Developing Passive Satellite Cloud Remote Sensing Algorithms Using Collocated Observations, Numerical Simulation and Deep Learning, NASA
  - CAREER: Big Data Climate Causality Analytics, NSF
  - NSF Harnessing Data Revolution (HDR) Institute: Harnessing Data and Model Revolution in the Polar Regions (iHARP), NSF

- My collaborators: Yiyi Huang, Sanjay Purushotham and Md Osman Gani (UMBC Information Systems), Chenxi Wang and Kerry Meyer (NASA GSFC), Zhibo Zhang (UMBC Physics), Aneesh Subramanian (University of Colorado Boulder, Atmospheric and Oceanic Sciences), Nicole-Jeanne Schlegel (NOAA Geophysical Fluid Dynamics Laboratory)

- My PhD students: Xin Huang, Sahara Ali, Omar Faruque

REFERENCES

1. Overpeck, et al., Climate Data Challenges in the 21st Century, *Science*, 2011
2. Faghmous and Kumar. A big data guide to understanding climate change: The case for theory-guided data science. *Big Data*, 2014
3. Holland, M. M. and Bitz, C. M. Polar amplification of climate change in coupled models. *Climate Dynamics*, 21(3-4):221–232, 2003.
4. Pearl, Judea. A probabilistic calculus of actions. *Uncertainty Proceedings 1994*. Morgan Kaufmann, 1994. 454-462.
5. Rubin, Donald B. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100.469 (2005): 322-331.
6. Hirano, Keisuke, and Guido W. Imbens. The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives 226164 (2004): 73-84.
7. Naimi, Ashley I., Stephen R. Cole, and Edward H. Kennedy. An introduction to g methods. International journal of epidemiology 46.2 (2017): 756-762.
8. Cole, Stephen R., and Miguel A. Hernán. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology* 168.6 (2008): 656-664.
9. Huffman, Curtis, and Edwin van Gameren. Covariate balancing inverse probability weights for time-varying continuous interventions. *Journal of Causal Inference* 6.2 (2018): 20170002.
10. 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.

THANK YOU!

KEY TERMINOLOGIES

Bias due to Time Varying Confounding

The common influence a past treatment T_t or covariate X_t might have on the future treatments T_{t+1} and the future outcome Y_{t+1}

Propensity Score

The probability of a unit being assigned to a particular treatment given a set of observed covariates.

Balancing strategies to reduce bias

Methods to reduce bias caused by time-varying treatment and covariates on the potential outcome, such as g-methods, propensity score matching, etc.

CAUSAL ASSUMPTIONS

- **Consistency**
  - The potential outcome for the treated subject $Y_{T=1}$ is considered equal to the observed outcome Y.
- **Positivity**
  - The probability of receiving treatment given some covariates X is always greater than zero. That is,

$$\Pr(T = t|X = x) > 0 \text{ where } \Pr(X = x) \neq 0$$

- **Conditional Exchangeability**
  - The conditional probability of receiving treatment depends only on the covariates X, that is, Y_T and treatment T are are statistically independent, given every possible value of X.
- **Stable Unit Treatment Value Assumption (SUTVA)**
  - The potential outcome Y_i on one unit i is not affected by the treatment effect on other units and there is no hidden variations of treatment.