UNIVERSITY OF MICHIGAN

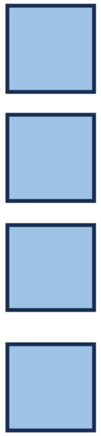
LSA STATISTICS
UNIVERSITY OF MICHIGAN

Statistical Methods for Spatio-Temporal Tensor Data

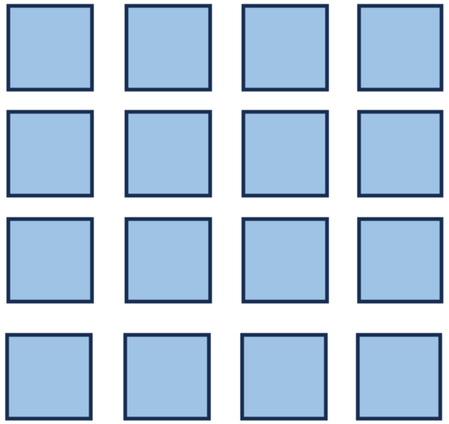
Hu Sun
Department of Statistics, University of Michigan

Doctoral Dissertation Defense
July 15, 2024
Advisor & Committee Chair: Prof. Yang Chen
Committee Members: Prof. Jian Kang, Prof. Kean Ming Tan, Prof. Ward B. Manchester

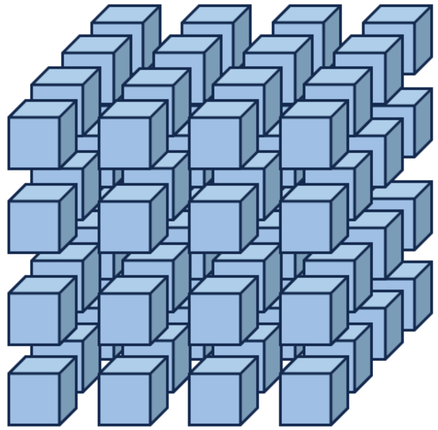
Introduction to Tensor



order-1 tensor $\mathbf{x}$
(vector)



order-2 tensor $\mathbf{X}$
(matrix)



order-3 tensor $\mathcal{X}$
(tensor)

Examples of Tensor Data

Examples of Tensor Data

rating tensor $\mathcal{X}$ in recommender system

order-3 tensor: $\mathcal{X}$ (user, item, context)

Examples of Tensor Data

rating tensor $\mathcal{X}$ in recommender system

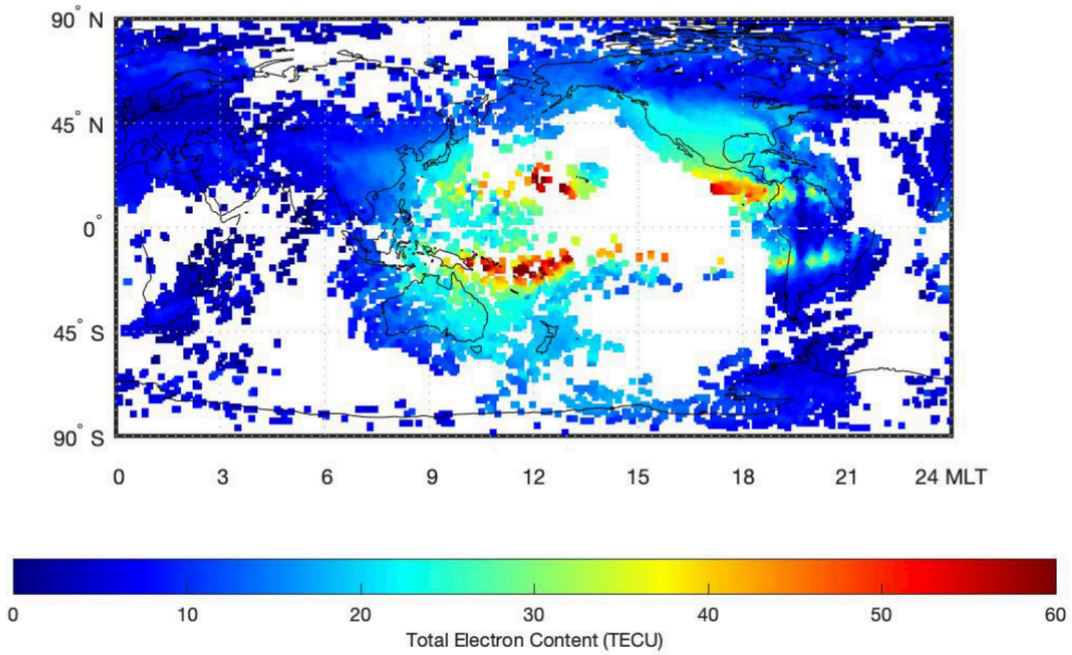
order-3 tensor: $\mathcal{X}$ (user, item, context)

functional MRI 3-D brain imaging tensor $\mathcal{Y}$

order-3 tensor: $\mathcal{Y}\ (x, y, z)$

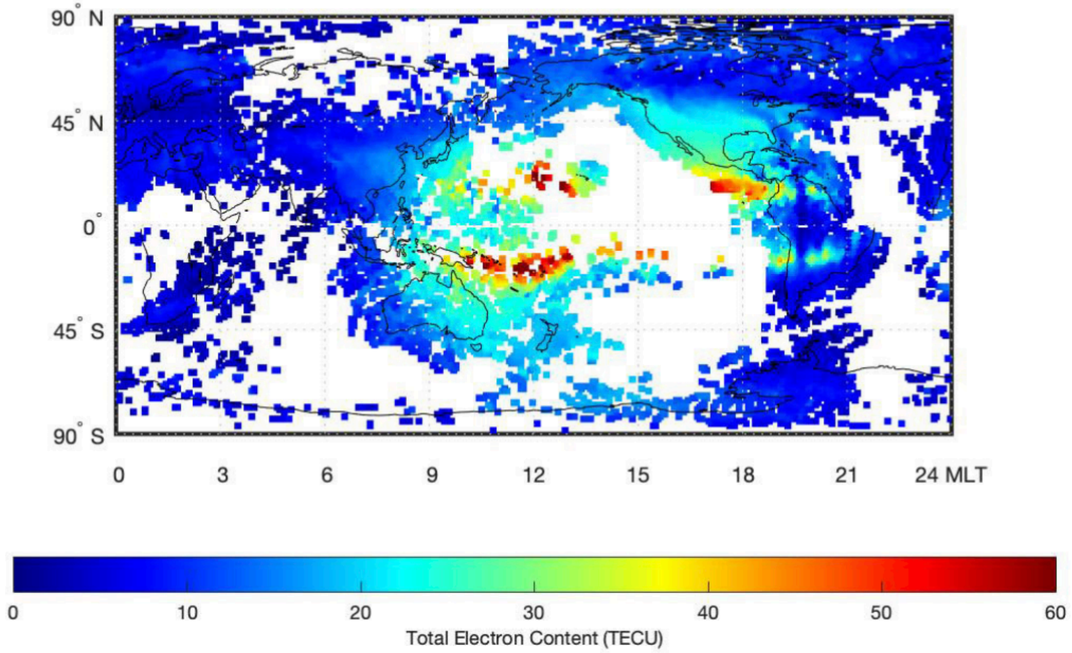
Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)



Spatio-Temporal Tensor

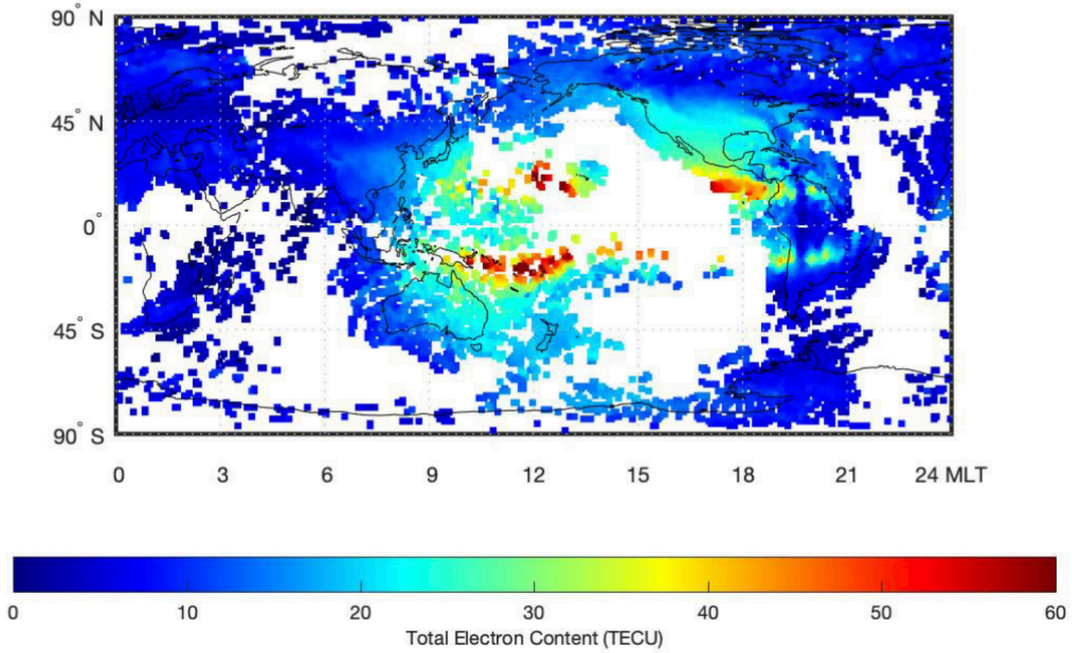
Motivating Dataset: Global Total Electron Content (TEC)



- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)

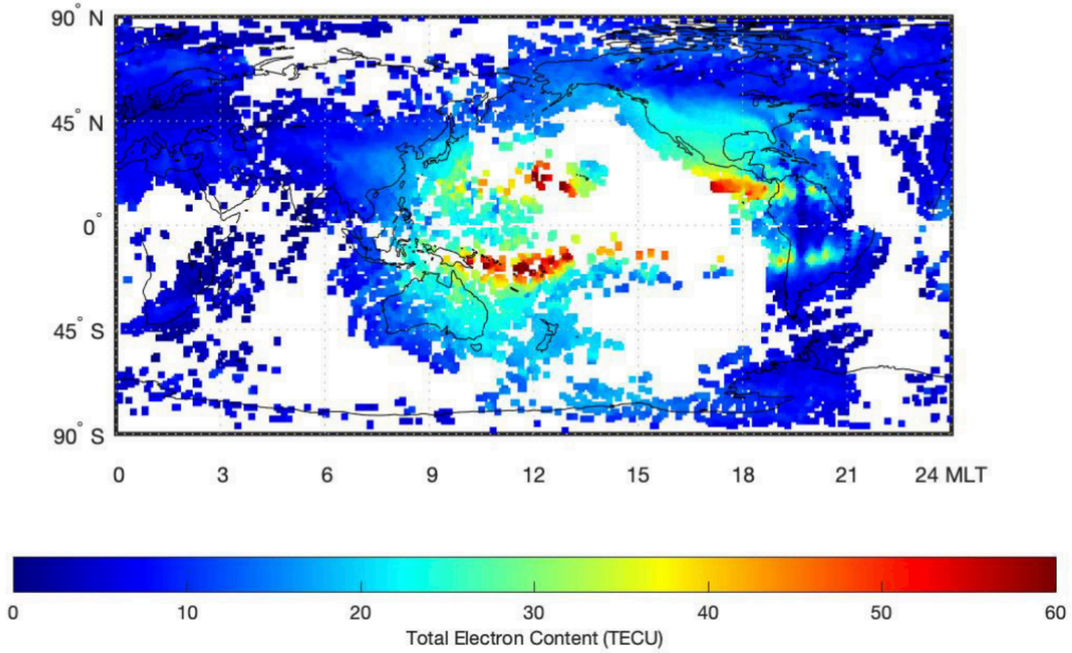


- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Challenges for such spatio-temporal tensors:

Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)



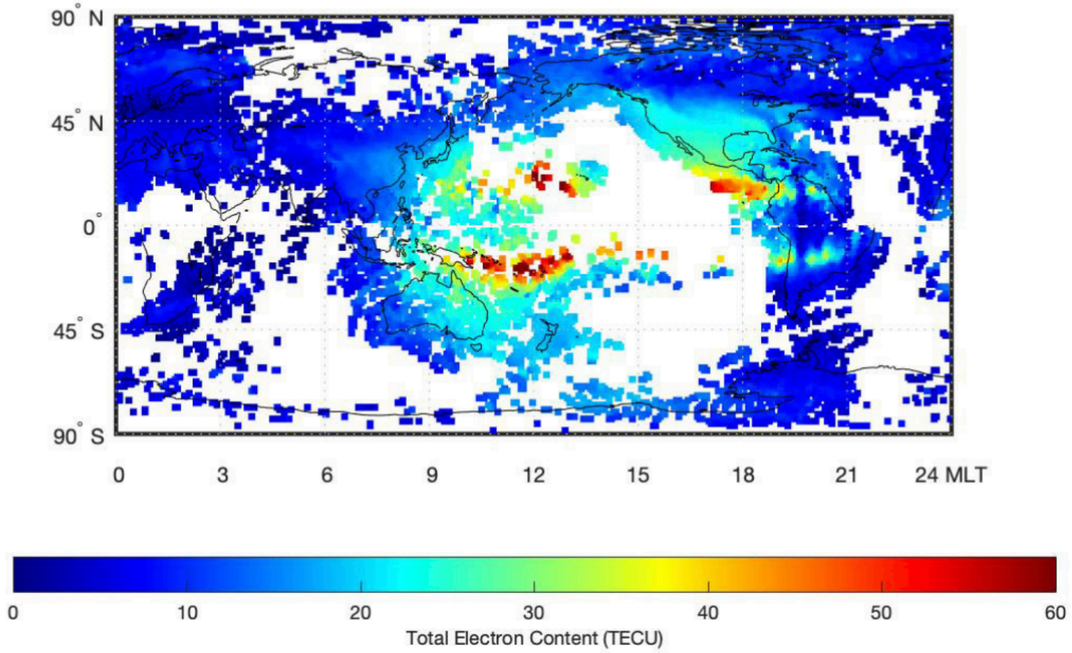
- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Challenges for such spatio-temporal tensors:

- multi-way spatio-temporal dependency

Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)



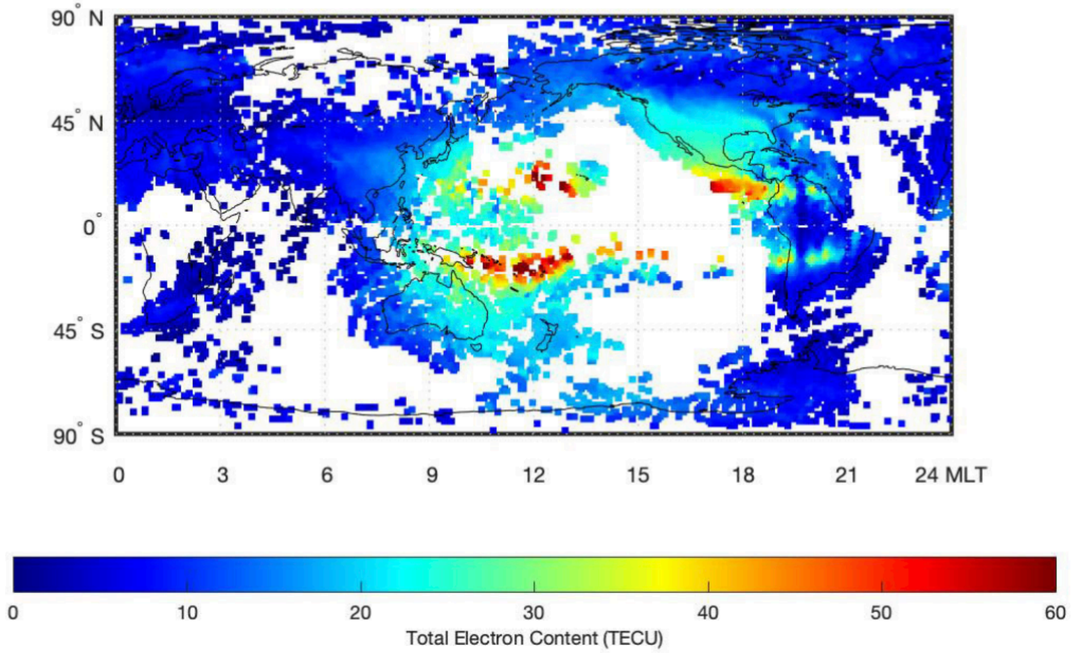
- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Challenges for such spatio-temporal tensors:

- multi-way spatio-temporal dependency
- dependent data missingness

Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)



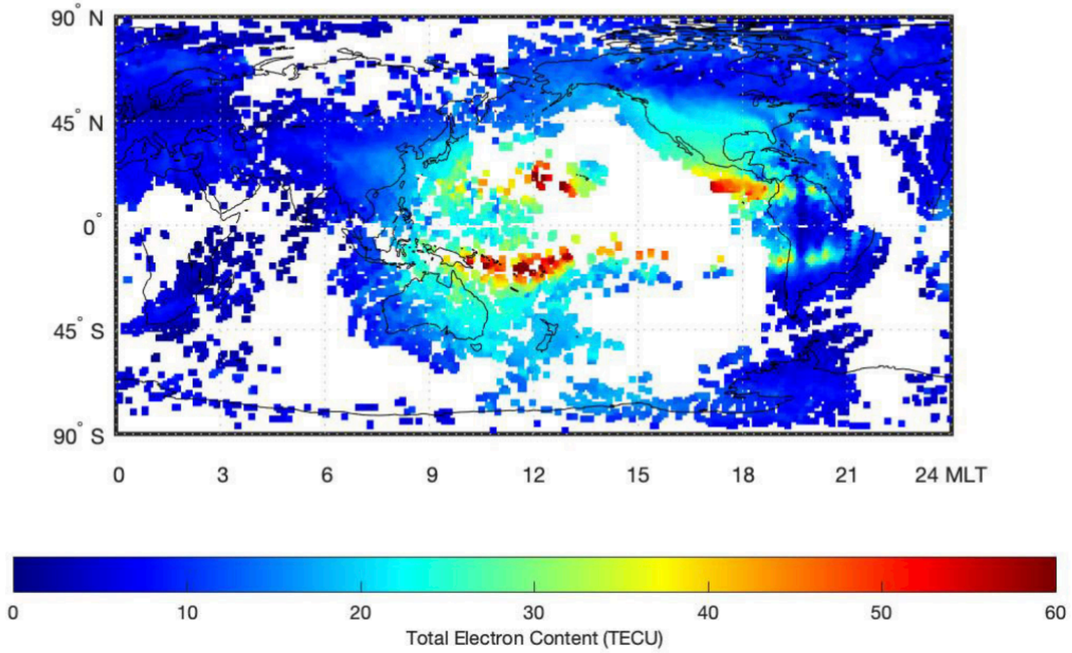
- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Challenges for such spatio-temporal tensors:

- multi-way spatio-temporal dependency
- dependent data missingness
- multiple data modalities

Spatio-Temporal Tensor

Motivating Dataset: Global Total Electron Content (TEC)



- Monitoring the global TEC is critical since TEC affect satellite communication and satellite navigation (e.g. GPS positioning).

Challenges for such spatio-temporal tensors:

- multi-way spatio-temporal dependency
- dependent data missingness
- multiple data modalities
- high data dimensionality

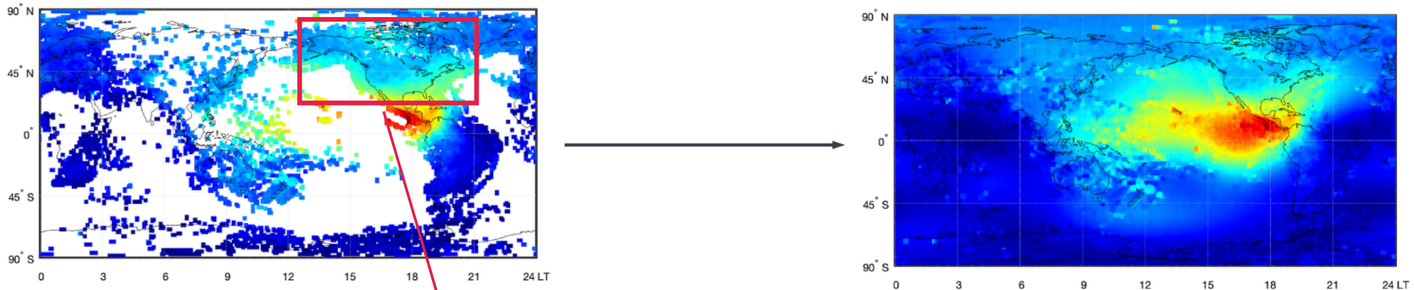
Dissertation Summary

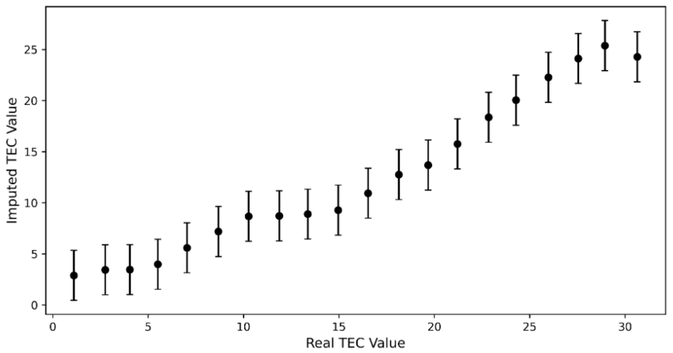
Dissertation Summary

I. Tensor Completion with Spatio-Temporal smoothing

Dissertation Summary

I. Tensor Completion with Spatio-Temporal smoothing

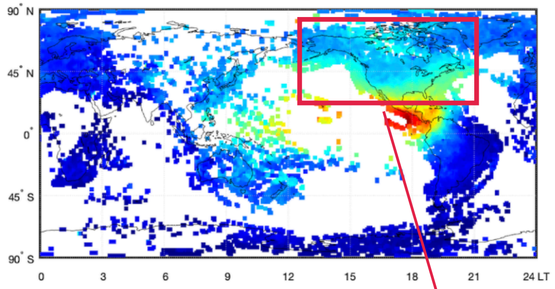


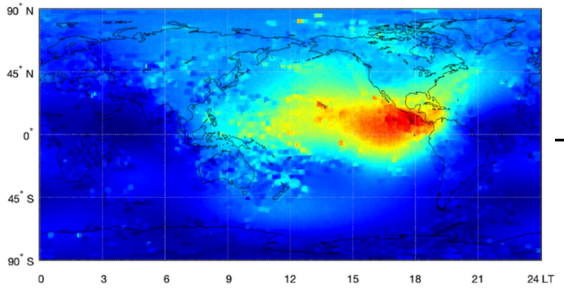


II. Uncertainty Quantification for Tensor Completion

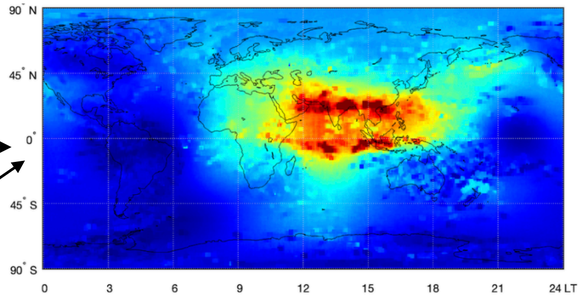
Dissertation Summary

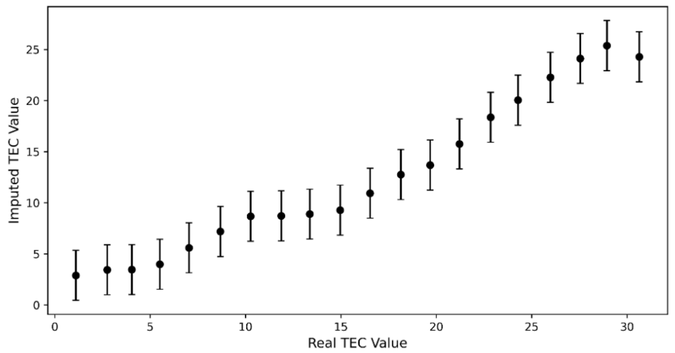
I. Tensor Completion with Spatio-Temporal smoothing

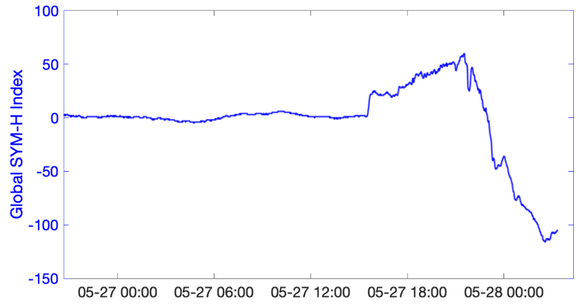




III. RKHS Method for Multi-modal Autoregression Model



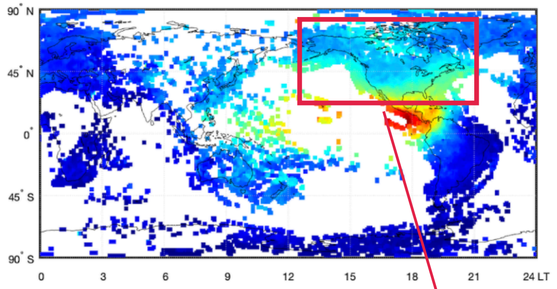


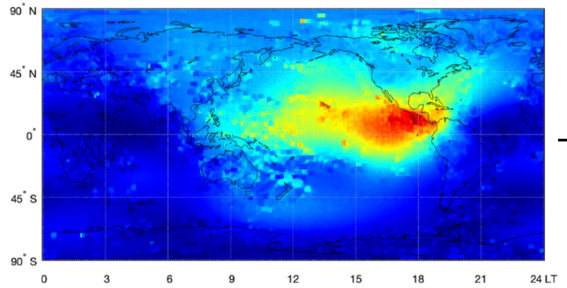


II. Uncertainty Quantification for Tensor Completion

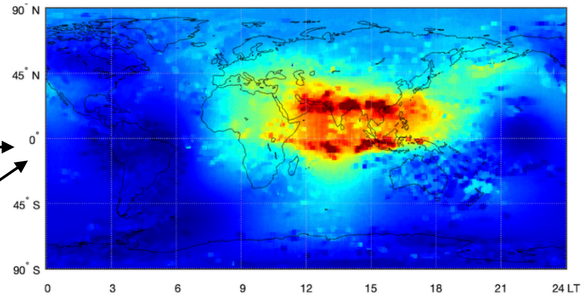
Dissertation Summary

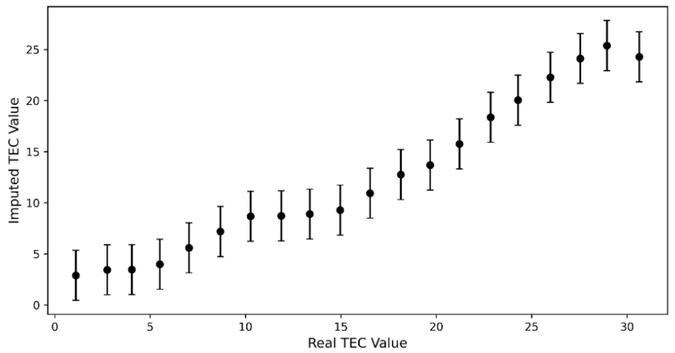
I. Tensor Completion with Spatio-Temporal smoothing



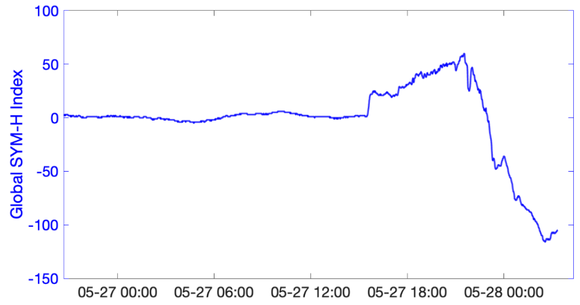


III. RKHS Method for Multi-modal Autoregression Model

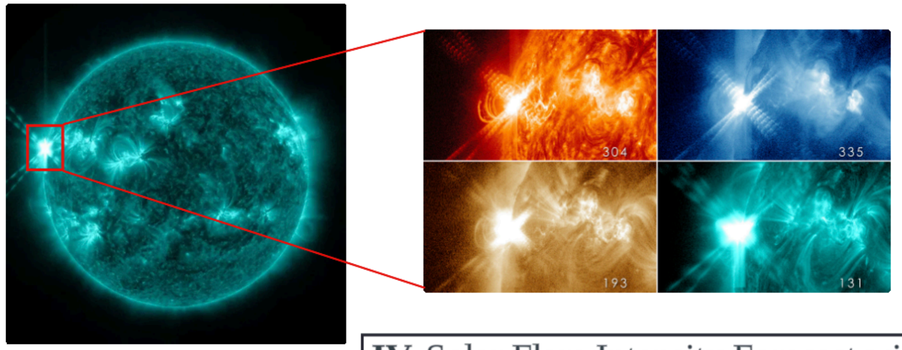




II. Uncertainty Quantification for Tensor Completion

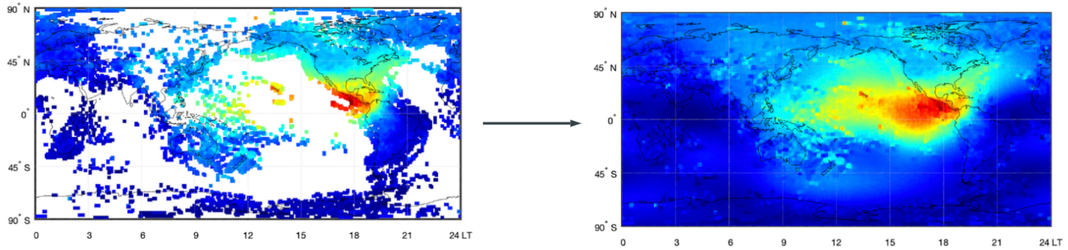


Solar Wind Parameters



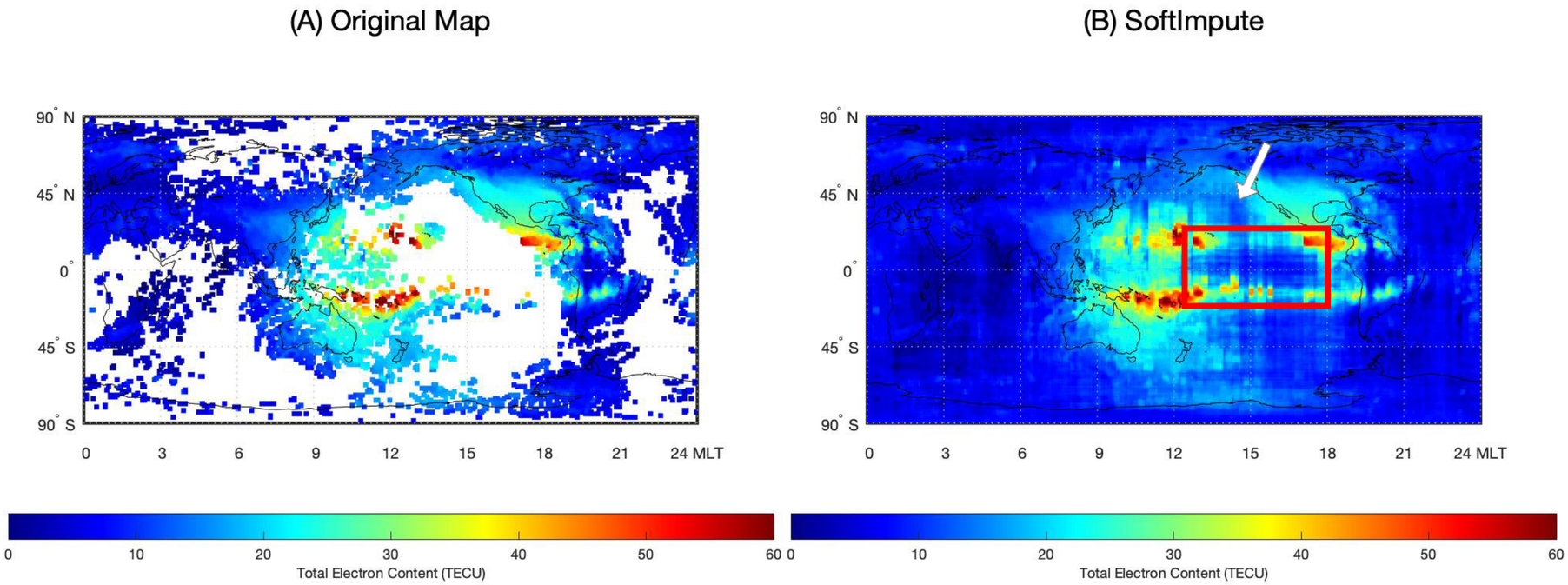
IV. Solar Flare Intensity Forecast with Scalar-on-Tensor Regression

I. Tensor Completion with Spatio-Temporal Smoothing

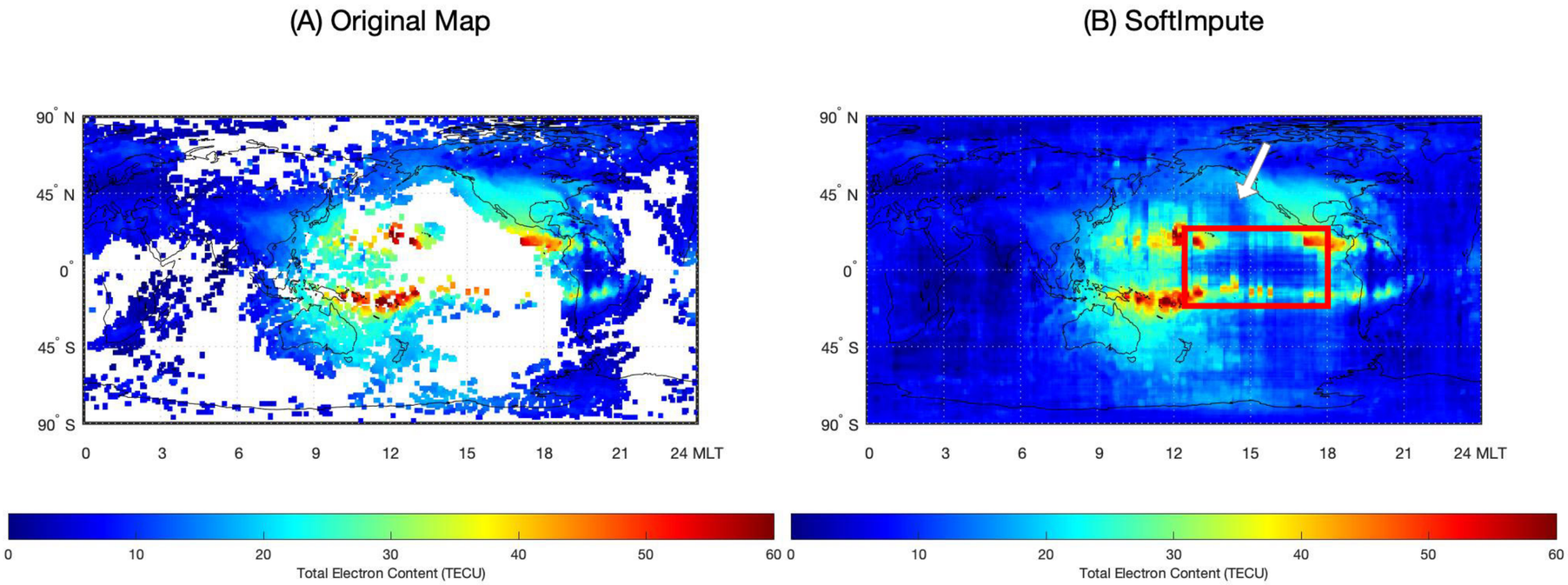


How to generate accurate imputations when data are missing in spatio-temporal patches?

Low-rank Completion



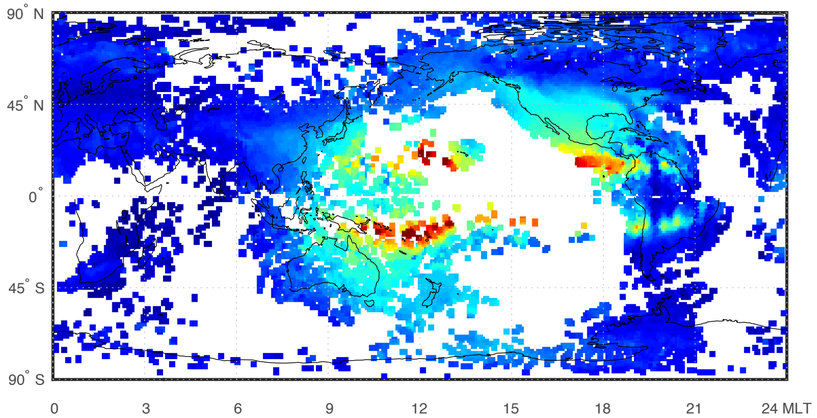
Low-rank Completion



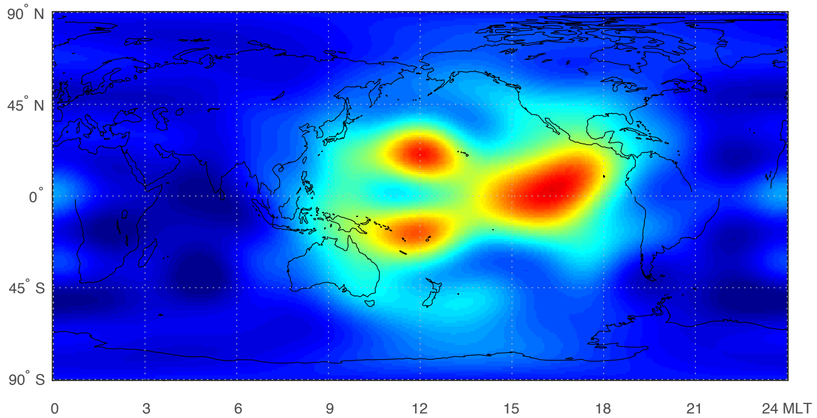
- Traditional matrix/tensor completion method [Hastie et al. (2015)] cannot impute regions with almost all entries missing (called the missing patches).

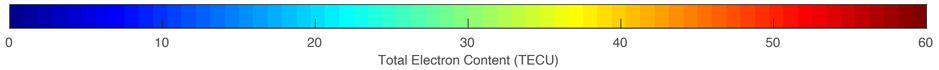
Spatio-Temporal Kernel Smoothing

(A) Madrigal TEC map

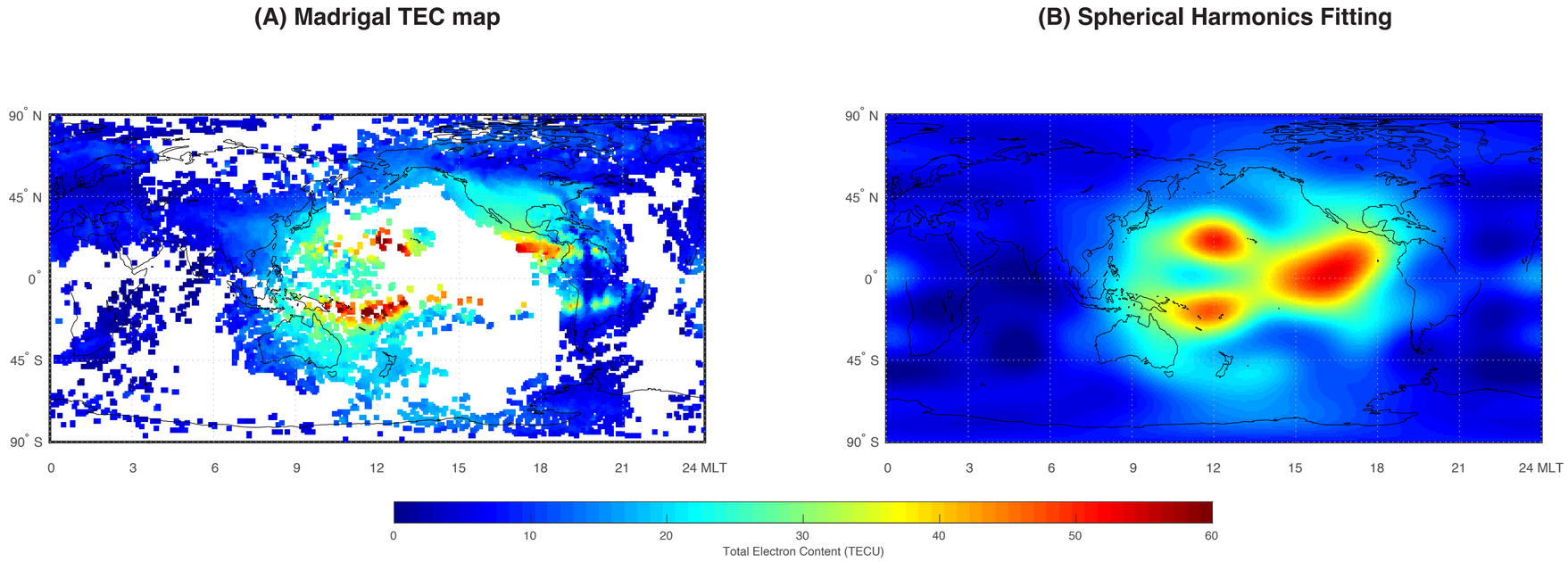


(B) Spherical Harmonics Fitting





Spatio-Temporal Kernel Smoothing



- Spatio-temporal kernel smoothing can impute the missing patches with more reasonable values, but tends to be overly-smooth.

Our Method

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$
- We aim at imputing $\mathbf{X}_1, \ldots, \mathbf{X}_T$ via solving the following non-convex optimization problem:

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$
- We aim at imputing $\mathbf{X}_1, \ldots, \mathbf{X}_T$ via solving the following non-convex optimization problem:

$$\min_{\mathbf{A}_t, \mathbf{B}_t, t \in [T]} \frac{1}{2} \sum_t \left\| \mathrm{P}_{\Omega_t} \left(\mathbf{X}_t - \mathbf{A}_t \mathbf{B}_t^\top \right) \right\|_{\mathrm{F}}^2 + \lambda_1 \sum_t \left(\left\| \mathbf{A}_t \right\|_{\mathrm{F}}^2 + \left\| \mathbf{B}_t \right\|_{\mathrm{F}}^2 \right)$$
$$+ \lambda_2 \sum_{t \geq 2} \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{A}_{t-1} \mathbf{B}_{t-1}^\top \right\|_{\mathrm{F}}^2 + \lambda_3 \sum_t \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{Y}_t \right\|_{\mathrm{F}}^2$$

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$
- We aim at imputing $\mathbf{X}_1, \ldots, \mathbf{X}_T$ via solving the following non-convex optimization problem:

reconstruction loss

$$\min_{\mathbf{A}_t, \mathbf{B}_t, t \in [T]} \frac{1}{2} \sum_t \left\| \mathrm{P}_{\Omega_t} \left(\mathbf{X}_t - \mathbf{A}_t \mathbf{B}_t^\top \right) \right\|_{\mathrm{F}}^2 + \lambda_1 \sum_t \left(\left\| \mathbf{A}_t \right\|_{\mathrm{F}}^2 + \left\| \mathbf{B}_t \right\|_{\mathrm{F}}^2 \right)$$

$$+ \lambda_2 \sum_{t \geq 2} \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{A}_{t-1} \mathbf{B}_{t-1}^\top \right\|_{\mathrm{F}}^2 + \lambda_3 \sum_t \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{Y}_t \right\|_{\mathrm{F}}^2$$

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$
- We aim at imputing $\mathbf{X}_1, \ldots, \mathbf{X}_T$ via solving the following non-convex optimization problem:

reconstruction loss

$$\min_{\mathbf{A}_t, \mathbf{B}_t, t \in [T]} \frac{1}{2} \sum_t \left\| \mathrm{P}_{\Omega_t} \left(\mathbf{X}_t - \mathbf{A}_t \mathbf{B}_t^\top \right) \right\|_{\mathrm{F}}^2 + \lambda_1 \sum_t \left(\left\| \mathbf{A}_t \right\|_{\mathrm{F}}^2 + \left\| \mathbf{B}_t \right\|_{\mathrm{F}}^2 \right)$$

$$+ \lambda_2 \sum_{t \geq 2} \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{A}_{t-1} \mathbf{B}_{t-1}^\top \right\|_{\mathrm{F}}^2 + \lambda_3 \sum_t \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{Y}_t \right\|_{\mathrm{F}}^2$$

temporal smoothing

Our Method

- 3-D spatio-temporal tensor (as a matrix time series): $\mathbf{X}_1, \ldots, \mathbf{X}_T \in \mathbb{R}^{M \times N}$
- We aim at imputing $\mathbf{X}_1, \ldots, \mathbf{X}_T$ via solving the following non-convex optimization problem:

reconstruction loss

$$\min_{\mathbf{A}_t, \mathbf{B}_t, t \in [T]} \frac{1}{2} \sum_t \left\| \mathrm{P}_{\Omega_t} \left(\mathbf{X}_t - \mathbf{A}_t \mathbf{B}_t^\top \right) \right\|_{\mathrm{F}}^2 + \lambda_1 \sum_t \left(\left\| \mathbf{A}_t \right\|_{\mathrm{F}}^2 + \left\| \mathbf{B}_t \right\|_{\mathrm{F}}^2 \right)$$

$$+ \lambda_2 \sum_{t \geq 2} \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{A}_{t-1} \mathbf{B}_{t-1}^\top \right\|_{\mathrm{F}}^2 + \lambda_3 \sum_t \left\| \mathbf{A}_t \mathbf{B}_t^\top - \mathbf{Y}_t \right\|_{\mathrm{F}}^2$$

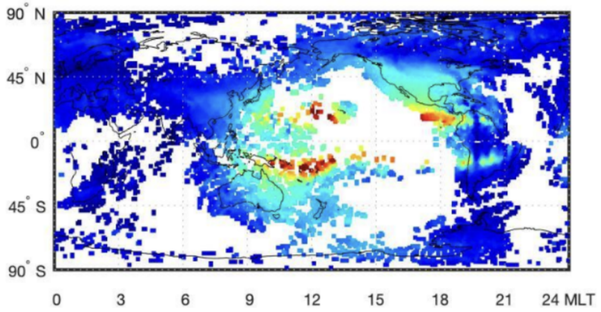
temporal smoothing

pre-imputed by kernel smoothing

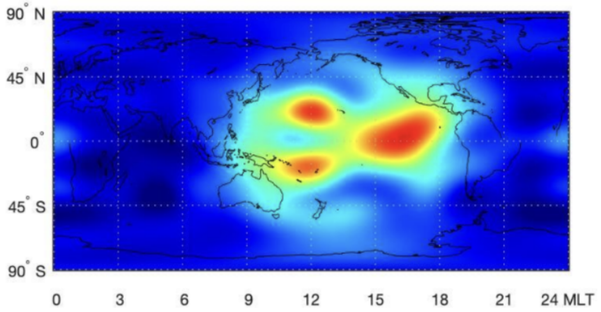
Results

Global Total Electron Content (TEC) Reconstruction

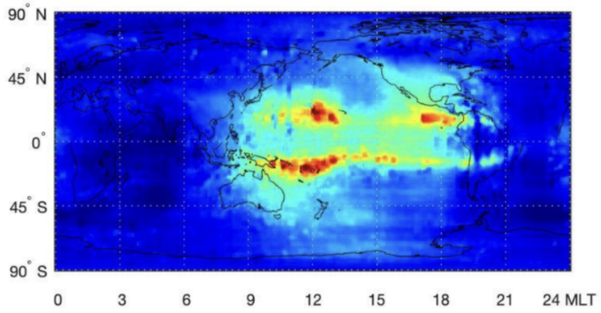
(A) Original Data



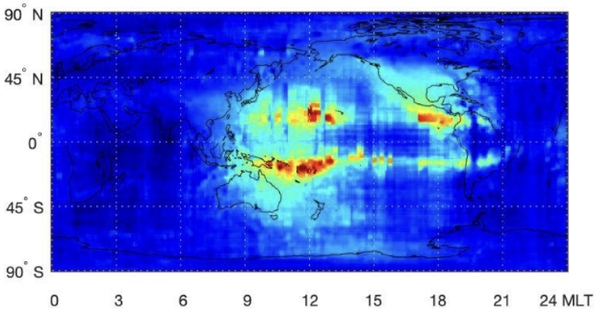
(B) Kernel Smoothing

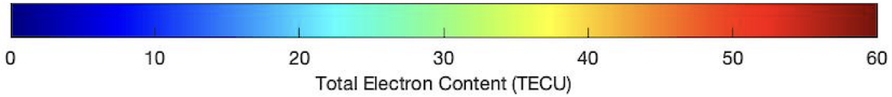


(C) Our Method

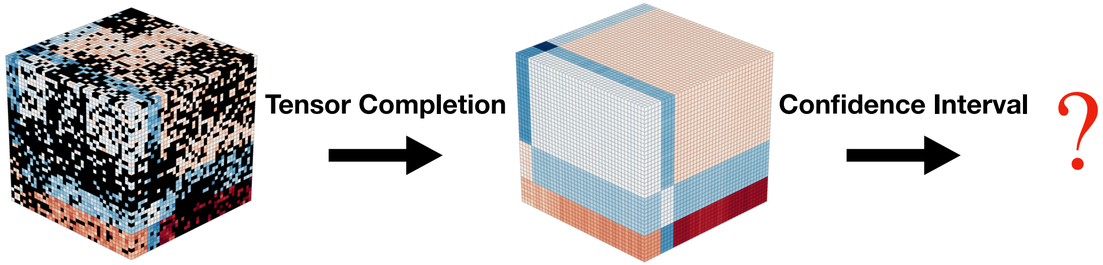


(D) Matrix Completion





II. Conformalized Tensor Completion with Riemannian Optimization

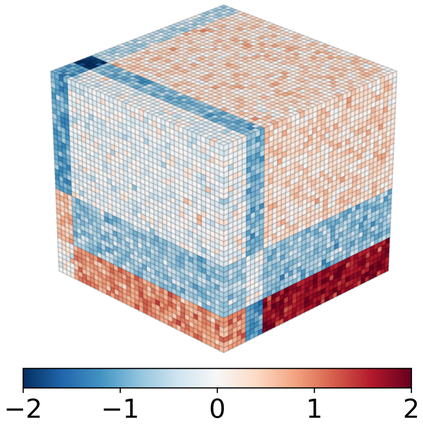


Research Question

Research Question

Noisy Data Tensor

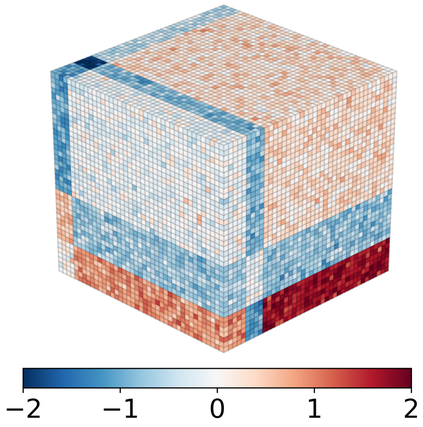
$\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$



Research Question

Noisy Data Tensor

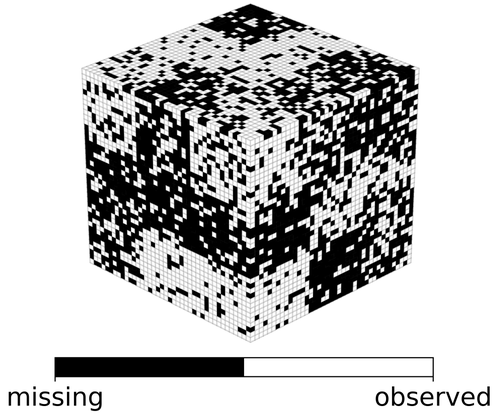
$\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$



+

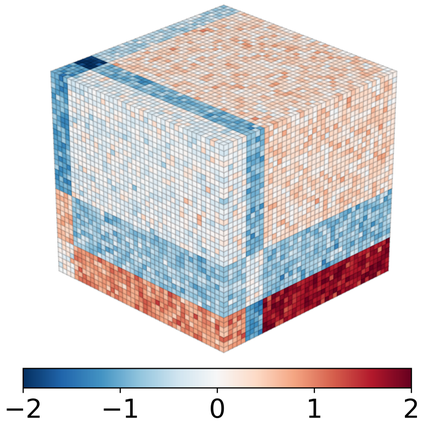
Missing Data

$\mathcal{W} \in \{-1,1\}^{d_1 \times \cdots \times d_K}$



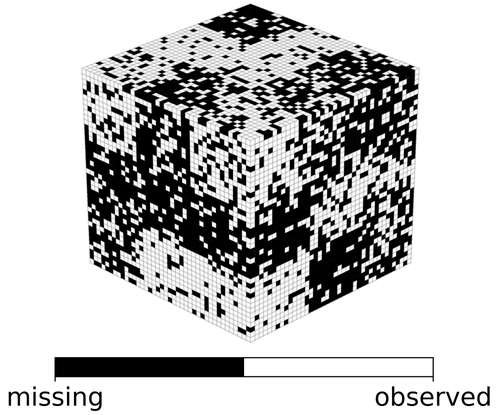
Research Question

Noisy Data Tensor
$\mathscr{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$



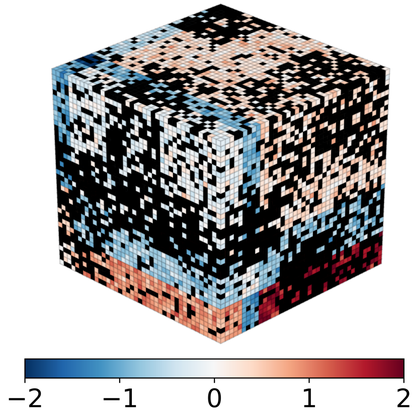
+

Missing Data
$\mathscr{W} \in \{-1,1\}^{d_1 \times \cdots \times d_K}$



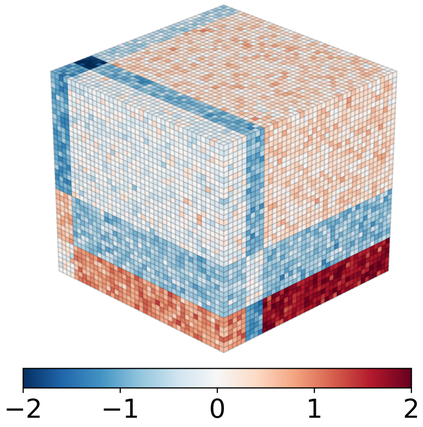
=

Partially-Observed Data Tensor $\mathscr{X}_{\mathbf{\Omega}}$
$\mathbf{\Omega} = \{s \in [d_1] \times \cdots [d_K] \,|\, [\mathscr{W}]_s = 1\}$



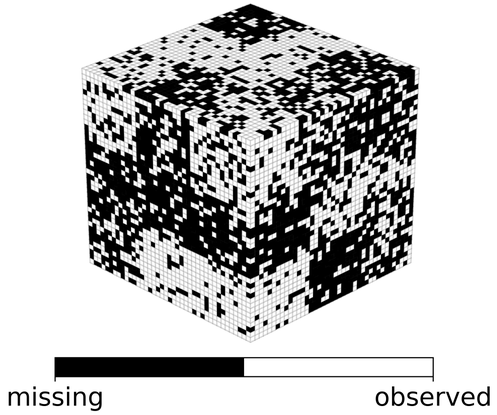
Research Question

Noisy Data Tensor
$\mathscr{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$



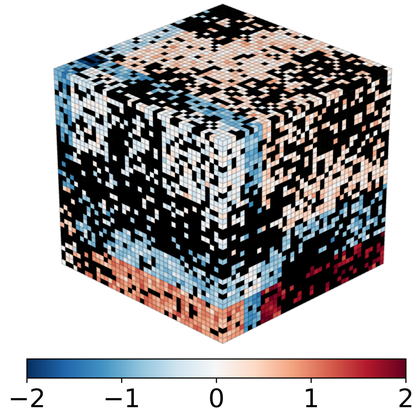
+

Missing Data
$\mathscr{W} \in \{-1,1\}^{d_1 \times \cdots \times d_K}$



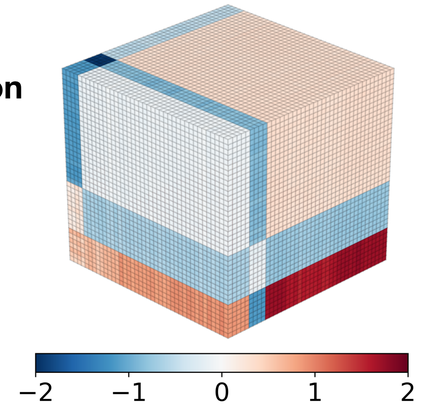
=

Partially-Observed Data Tensor $\mathscr{X}_{\mathbf{\Omega}}$
$\mathbf{\Omega} = \{s \in [d_1] \times \cdots [d_K] \,|\, [\mathscr{W}]_s = 1\}$



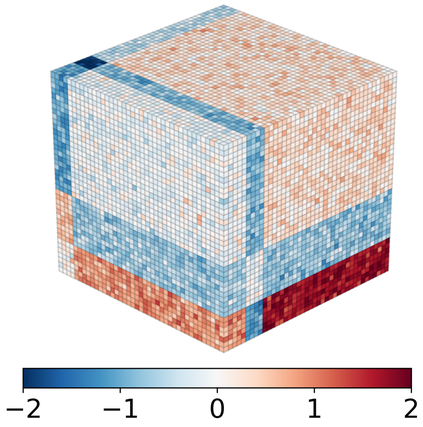
Tensor Completion

Completed Tensor $\widehat{\mathscr{X}}$



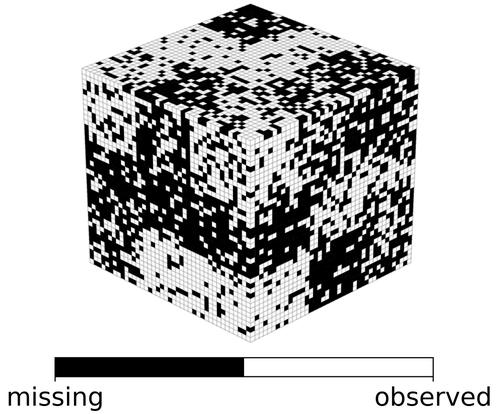
Research Question

Noisy Data Tensor
$\mathscr{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$



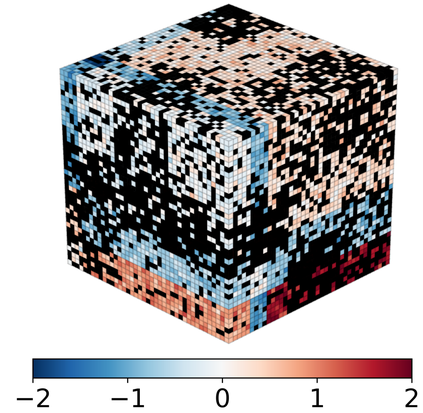
+

Missing Data
$\mathscr{W} \in \{-1,1\}^{d_1 \times \cdots \times d_K}$



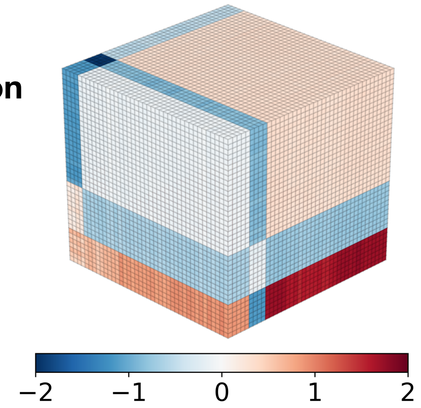
=

Partially-Observed Data Tensor $\mathscr{X}_{\boldsymbol{\Omega}}$
$\boldsymbol{\Omega} = \{s \in [d_1] \times \cdots [d_K] \,|\, [\mathscr{W}]_s = 1\}$



Tensor Completion

Completed Tensor $\widehat{\mathscr{X}}$



For any $s^* \in \boldsymbol{\Omega}^c = \{s \in [d_1] \times \cdots [d_K] \,|\, [\mathscr{W}]_s = -1\}$, how to construct a confidence region $\widehat{C}_{s^*}$ such that:

$$\mathrm{P}\left([\mathscr{X}]_{s^*} \in \widehat{C}_{s^*}\right) \geq 1 - \alpha,$$

with $1 - \alpha$ being a pre-specified level of confidence?

Literature Review

Two Types of Uncertainty Quantification

Literature Review

Two Types of Uncertainty Quantification

Suppose the data tensor $\mathcal{X}$ is generated by:

Literature Review

Two Types of Uncertainty Quantification

Suppose the data tensor $\mathcal{X}$ is generated by:

$$\mathcal{X} = \mathcal{X}^* + \mathcal{E},$$

Literature Review

Two Types of Uncertainty Quantification

Suppose the data tensor $\mathcal{X}$ is generated by:

$$\mathcal{X} = \mathcal{X}^* + \mathcal{E},$$

existing works construct confidence sets of $\widehat{\mathcal{X}}$ with two distinct goals:

Literature Review

Two Types of Uncertainty Quantification

Suppose the data tensor $\mathcal{X}$ is generated by:

$$\mathcal{X} = \mathcal{X}^* + \mathcal{E},$$

existing works construct confidence sets of $\widehat{\mathcal{X}}$ with two distinct goals:

- inference on true tensor: $\mathrm{P}\left([\mathcal{X}^*]_{s^*} \in \widehat{C}_{s^*}\right) \geq 1 - \alpha$

Literature Review

Two Types of Uncertainty Quantification

Suppose the data tensor $\mathcal{X}$ is generated by:

$$\mathcal{X} = \mathcal{X}^* + \mathcal{E},$$

existing works construct confidence sets of $\widehat{\mathcal{X}}$ with two distinct goals:

- inference on true tensor: $\mathrm{P}\left([\mathcal{X}^*]_{s^*} \in \widehat{C}_{s^*}\right) \geq 1 - \alpha$
- predictive UQ: $\mathrm{P}\left([\mathcal{X}]_{s^*} \in \widehat{C}_{s^*}\right) \geq 1 - \alpha$

Literature Review

Inference on True Tensor/Matrix

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]
  - relies on specific algorithm and can be extremely loose

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]
  - relies on specific algorithm and can be extremely loose
- Distributional analysis of $[\widehat{\mathcal{X}} - \mathcal{X}^*]_{s^*}$ [Chen et al. (2019), Xia and Yuan (2021), Cai et al. (2022), Farias et al. (2022)]

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]
  - relies on specific algorithm and can be extremely loose
- Distributional analysis of $[\widehat{\mathcal{X}} - \mathcal{X}^*]_{s^*}$ [Chen et al. (2019), Xia and Yuan (2021), Cai et al. (2022), Farias et al. (2022)]
  - relies on specific algorithm, noise distribution, and exact low-rank true tensor form

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]
  - relies on specific algorithm and can be extremely loose
- Distributional analysis of $[\widehat{\mathcal{X}} - \mathcal{X}^*]_{s^*}$ [Chen et al. (2019), Xia and Yuan (2021), Cai et al. (2022), Farias et al. (2022)]
  - relies on specific algorithm, noise distribution, and exact low-rank true tensor form
- Bayesian matrix completion [Mai and Alquiler (2015), Yuchi et al. (2023)]

Literature Review

Inference on True Tensor/Matrix

- Convert bound on $\|\widehat{\mathcal{X}} - \mathcal{X}^*\|_{\mathrm{F}}$ to confidence sets [Carpentier et al. (2018, 2019)]
  - relies on specific algorithm and can be extremely loose
- Distributional analysis of $[\widehat{\mathcal{X}} - \mathcal{X}^*]_{s^*}$ [Chen et al. (2019), Xia and Yuan (2021), Cai et al. (2022), Farias et al. (2022)]
  - relies on specific algorithm, noise distribution, and exact low-rank true tensor form
- Bayesian matrix completion [Mai and Alquiler (2015), Yuchi et al. (2023)]
- Assume $[\mathcal{W}]_s \overset{i.i.d.}{\sim} \mathrm{Bern}(p)$, i.e. data missing uniformly at random.

Literature Review

Predictive Inference of Matrix Completion

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees
- Conformal prediction based matrix completion

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees
- Conformal prediction based matrix completion
  - data is missing independently, i.e. $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(p_s)$ [Gui et al. (2023)]

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees
- Conformal prediction based matrix completion
  - data is missing independently, i.e. $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(p_s)$ [Gui et al. (2023)]
  - considers uncertainty of newly-arrived rows [Shao and Zhang (2023)]

Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees
- Conformal prediction based matrix completion
  - data is missing independently, i.e. $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(p_s)$ [Gui et al. (2023)]
  - considers uncertainty of newly-arrived rows [Shao and Zhang (2023)]
- Not directly transferrable from matrix to tensor setting.

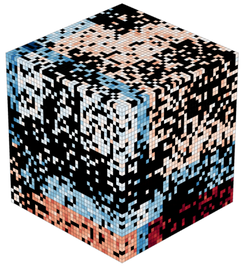
Literature Review

Predictive Inference of Matrix Completion

- Deep learning based matrix completion [Zeldes et al. (2017), Kasalicky et al. (2023)]
  - do not have coverage guarantees
- Conformal prediction based matrix completion
  - data is missing independently, i.e. $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(p_s)$ [Gui et al. (2023)]
  - considers uncertainty of newly-arrived rows [Shao and Zhang (2023)]
- Not directly transferrable from matrix to tensor setting.
- No previous work considers a missing mechanism under the spatio-temporal context.

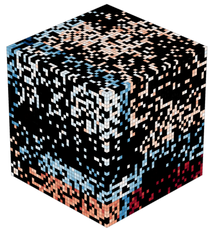
Classical Conformal Prediction (CP)

Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$

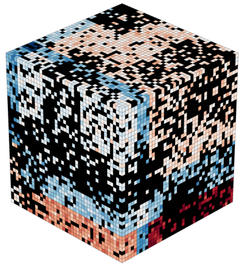


Classical Conformal Prediction (CP)

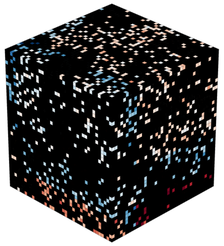
Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$



Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$

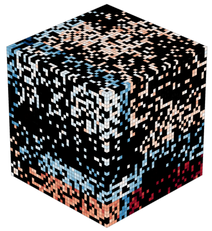


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$

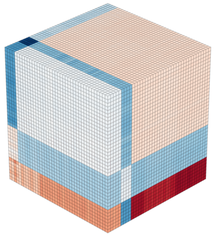


Classical Conformal Prediction (CP)

Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

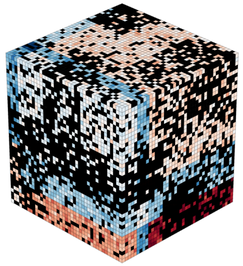


Completed Tensor $\widehat{\mathcal{X}}$

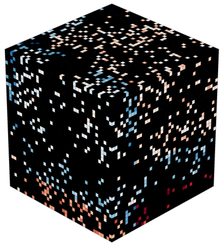


Tensor Completion

Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$

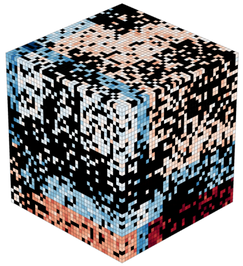


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$

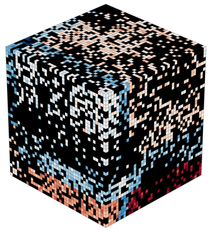


Classical Conformal Prediction (CP)

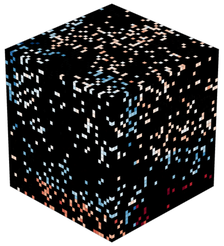
Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$



Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

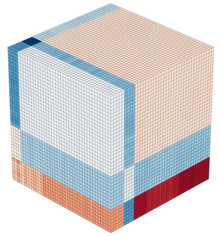


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$



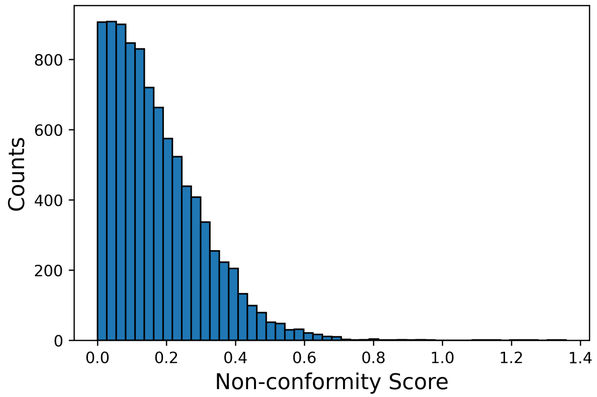
Tensor Completion

Completed Tensor $\widehat{\mathcal{X}}$



Non-conformity score

e.g. $\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s) = |[\mathcal{X}]_s - [\widehat{\mathcal{X}}]_s|$



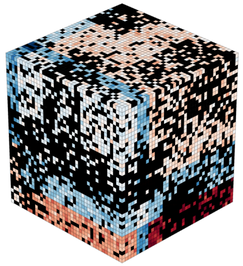
Classical Conformal Prediction (CP)

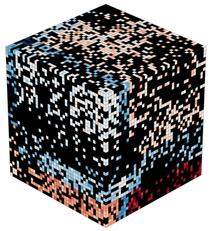
Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

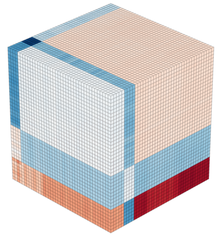
Completed Tensor $\widehat{\mathcal{X}}$

Tensor Completion

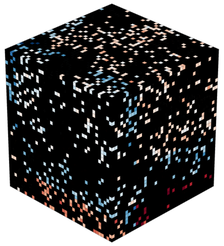
Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$





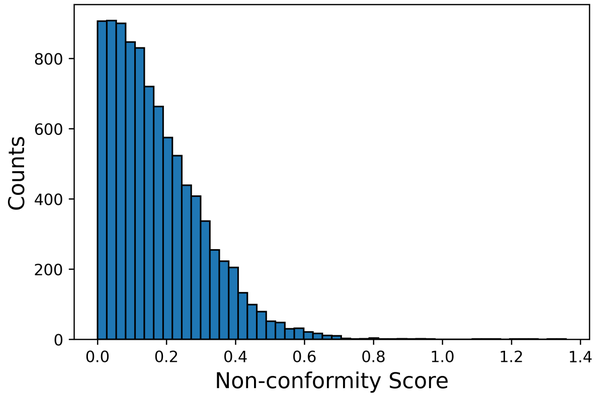


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$



Non-conformity score

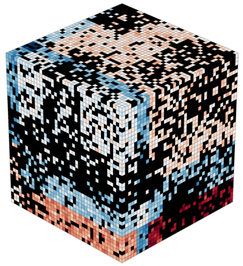
e.g. $\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s) = |[\mathcal{X}]_s - [\widehat{\mathcal{X}}]_s|$



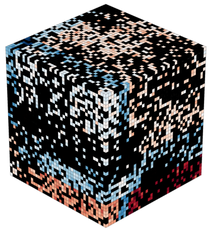
$$\widehat{C}_{1-\alpha,s*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathcal{X}}]_s) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s)} + \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{+\infty} \right) \right\}$$

Classical Conformal Prediction (CP)

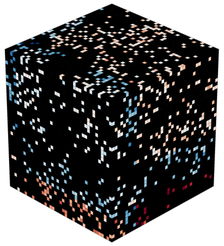
Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$



Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

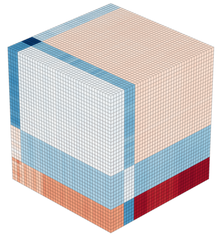


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$



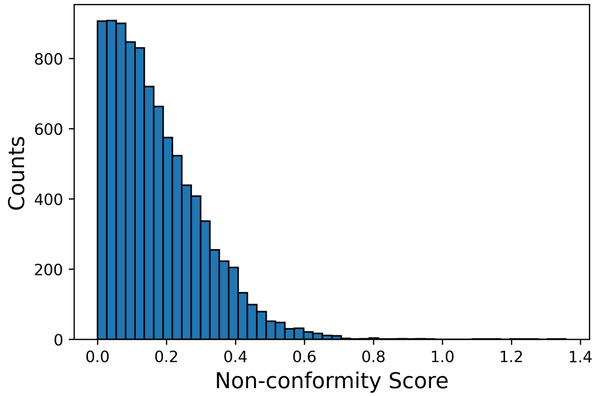
Tensor Completion

Completed Tensor $\widehat{\mathcal{X}}$



Non-conformity score

e.g. $\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s) = |[\mathcal{X}]_s - [\widehat{\mathcal{X}}]_s|$

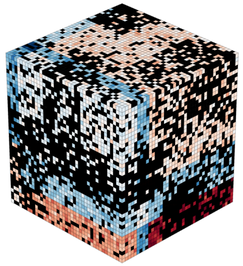


$$\widehat{C}_{1-\alpha,s^*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathcal{X}}]_s) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s)} + \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{+\infty} \right) \right\}$$

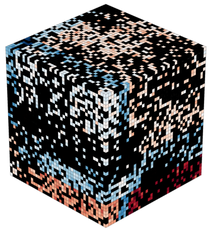
$(1-\alpha)$ level conformal interval at s^*

Classical Conformal Prediction (CP)

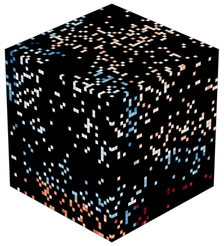
Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$



Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

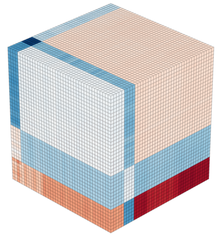


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$



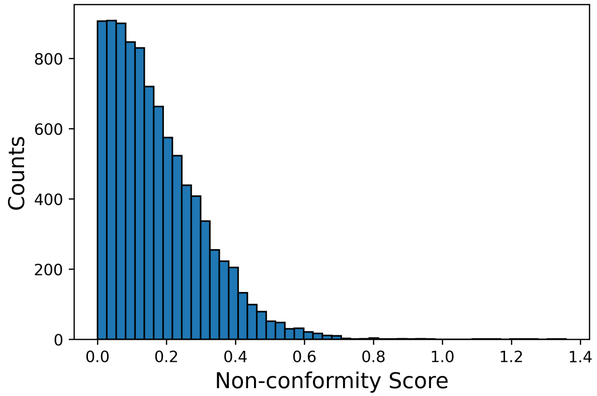
Tensor Completion

Completed Tensor $\widehat{\mathcal{X}}$



Non-conformity score

e.g. $\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s) = |[\mathcal{X}]_s - [\widehat{\mathcal{X}}]_s|$



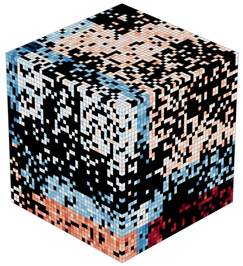
$$\widehat{C}_{1-\alpha,s*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathcal{X}}]_s) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s)} + \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{+\infty} \right) \right\}$$

$(1-\alpha)$ level conformal interval at $s*$

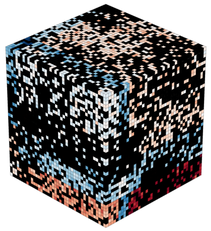
$(1-\alpha)$ quantile

Classical Conformal Prediction (CP)

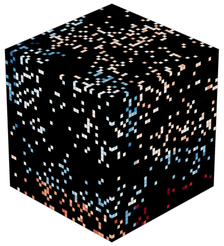
Data Tensor $\mathcal{X}_{\mathbf{\Omega}}$



Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

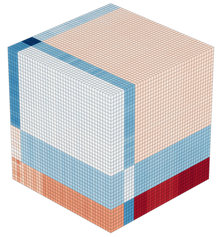


Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$



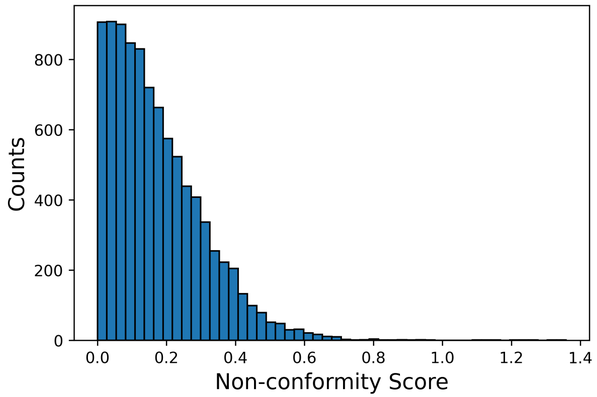
Tensor Completion

Completed Tensor $\widehat{\mathcal{X}}$



Non-conformity score

e.g. $\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s) = |[\mathcal{X}]_s - [\widehat{\mathcal{X}}]_s|$



$$\widehat{C}_{1-\alpha,s*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathcal{X}}]_s) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{\mathcal{S}([\mathcal{X}]_s, [\widehat{\mathcal{X}}]_s)} + \frac{1}{|\mathbf{\Omega}_{cal}| + 1} \delta_{+\infty} \right) \right\}$$

$(1-\alpha)$ level conformal interval at $s*$

$(1-\alpha)$ quantile

eCDF of the non-conformity score

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

- $[\mathscr{W}]_s \overset{i.i.d.}{\sim} \text{Bern}(p)$

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

- $[\mathscr{W}]_s \overset{i.i.d.}{\sim} \text{Bern}(p)$
- $s_1, \ldots, s_n$: indices of the n calibration entries

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

- $[\mathscr{W}]_s \overset{i.i.d.}{\sim} \text{Bern}(p)$
- $s_1, \ldots, s_n$: indices of the n calibration entries
- s_{n+1}: index of a test set entry

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

- $[\mathscr{W}]_s \overset{i.i.d.}{\sim} \text{Bern}(p)$
- $s_1, \ldots, s_n$: indices of the n calibration entries
- s_{n+1}: index of a test set entry
- $v_1, \ldots, v_n, v_{n+1}$: non-conformity scores

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

- $[\mathscr{W}]_s \overset{i.i.d.}{\sim} \text{Bern}(p)$
- $s_1, \ldots, s_n$: indices of the n calibration entries
- s_{n+1}: index of a test set entry
- $v_1, \ldots, v_n, v_{n+1}$: non-conformity scores
- V: the random variable of the non-conformity score at an unseen entry

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

$$
\begin{array}{ll}
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_1 \\
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_2 \\
\ldots\ldots & \ldots\ldots \\
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_{n+1}
\end{array}
$$

Validity of Classic Conformal Prediction

Tibshirani et al. (2019)

$$
\begin{array}{ll}
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_1 \\
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_2 \\
\ldots\ldots & \ldots\ldots \\
\left(s_1, s_2, \ldots, s_n, s_{n+1}\right) & V = v_{n+1}
\end{array}
$$

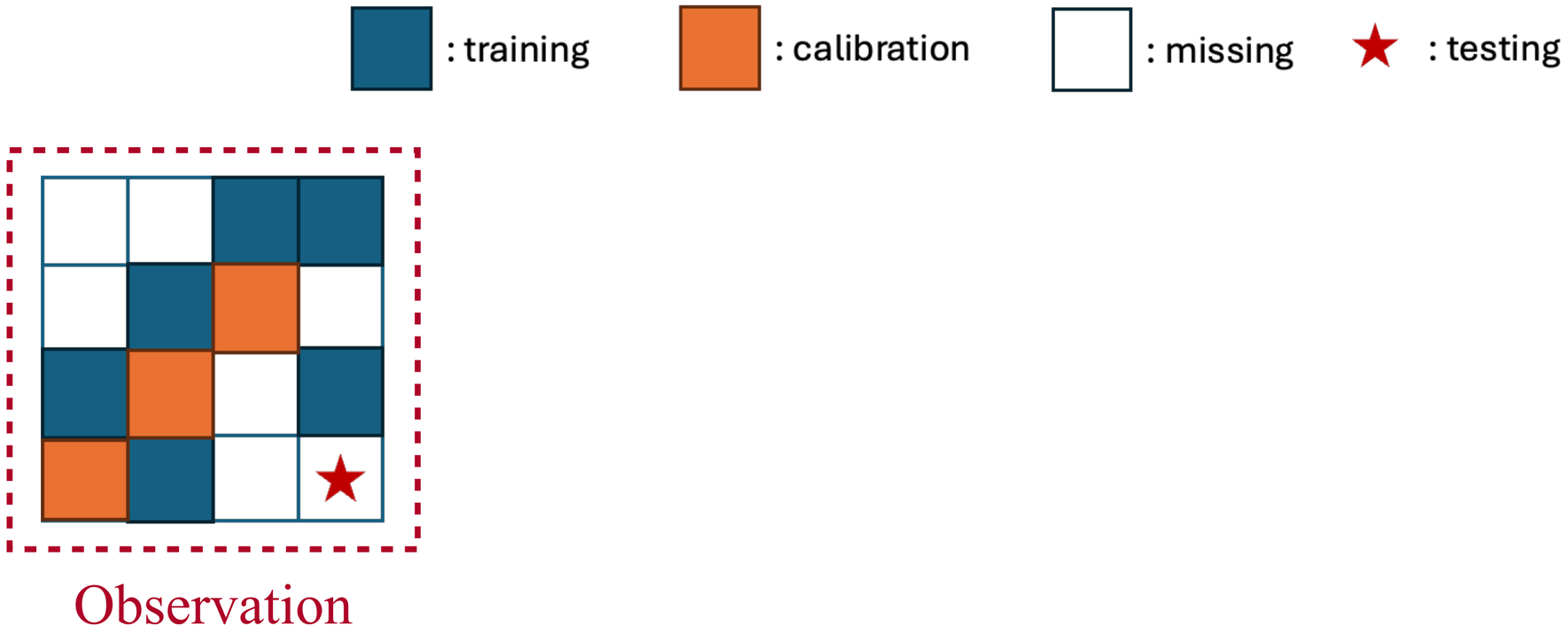
$$
\mathrm{P}(V = v_k) = \frac{\mathrm{P}([\mathscr{W}]_s = 1, \forall s \in \mathbf{\Omega}_{tr} \cup \{s_1, \ldots, s_{n+1}\} \backslash \{s_k\}, [\mathscr{W}]_s = -1, \text{o.w.})}{\sum_{l=1}^{n+1} \mathrm{P}([\mathscr{W}]_s = 1, \forall s \in \mathbf{\Omega}_{tr} \cup \{s_1, \ldots, s_{n+1}\} \backslash \{s_l\}, [\mathscr{W}]_s = -1, \text{o.w.})} = \frac{1}{n+1}
$$

Weighted Conformal Prediction

Leave-one-out Likelihood Weighting

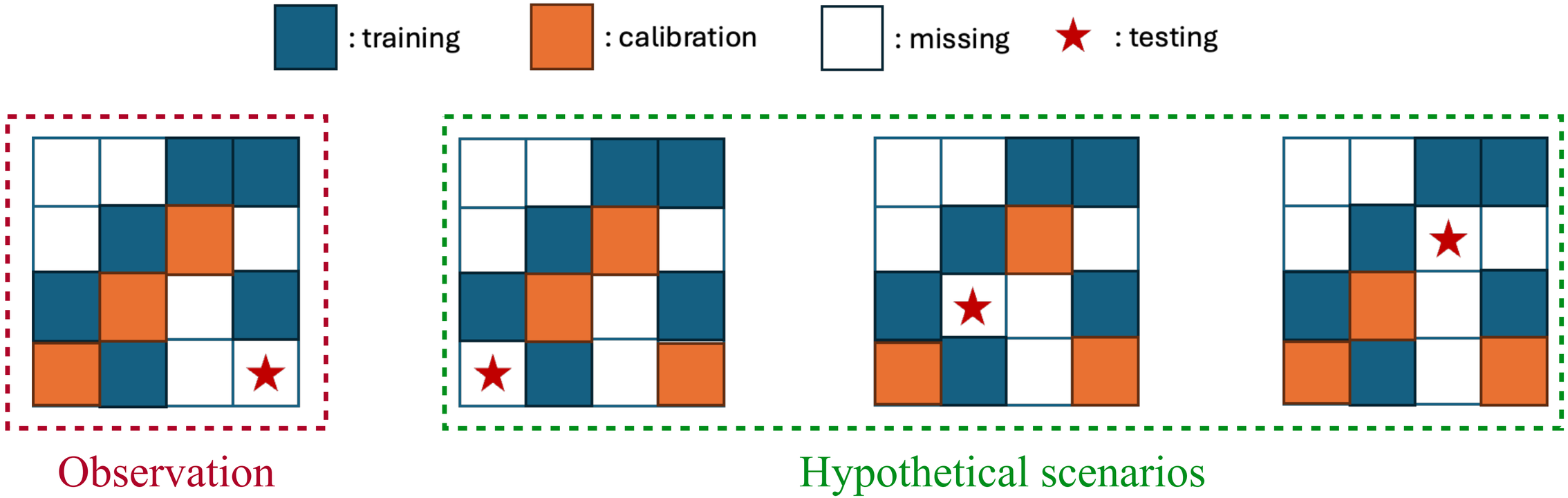
Weighted Conformal Prediction

Leave-one-out Likelihood Weighting



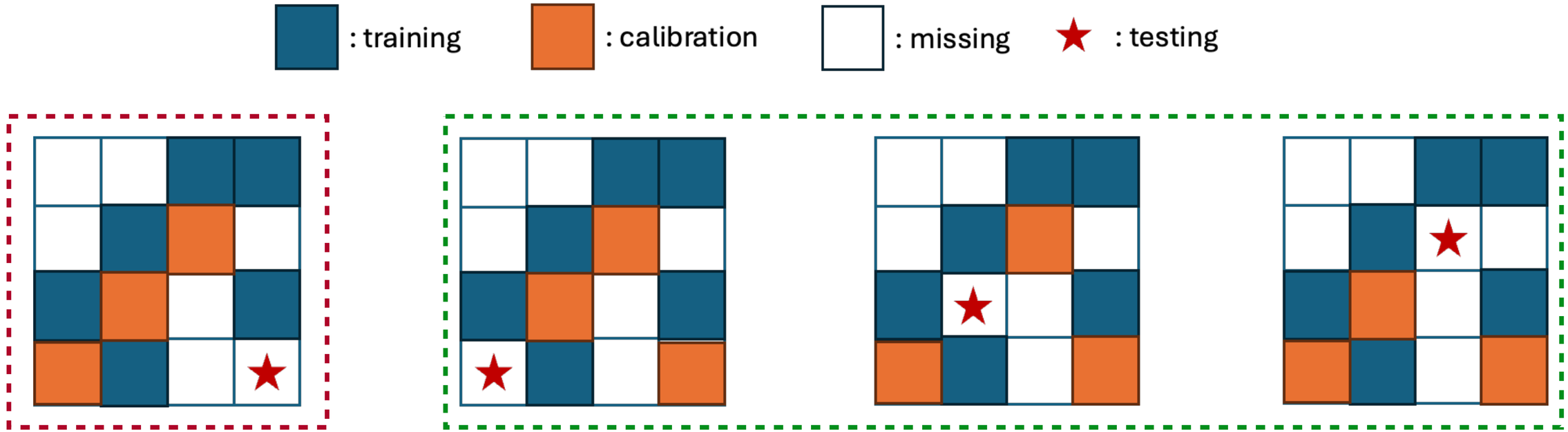
Weighted Conformal Prediction

Leave-one-out Likelihood Weighting



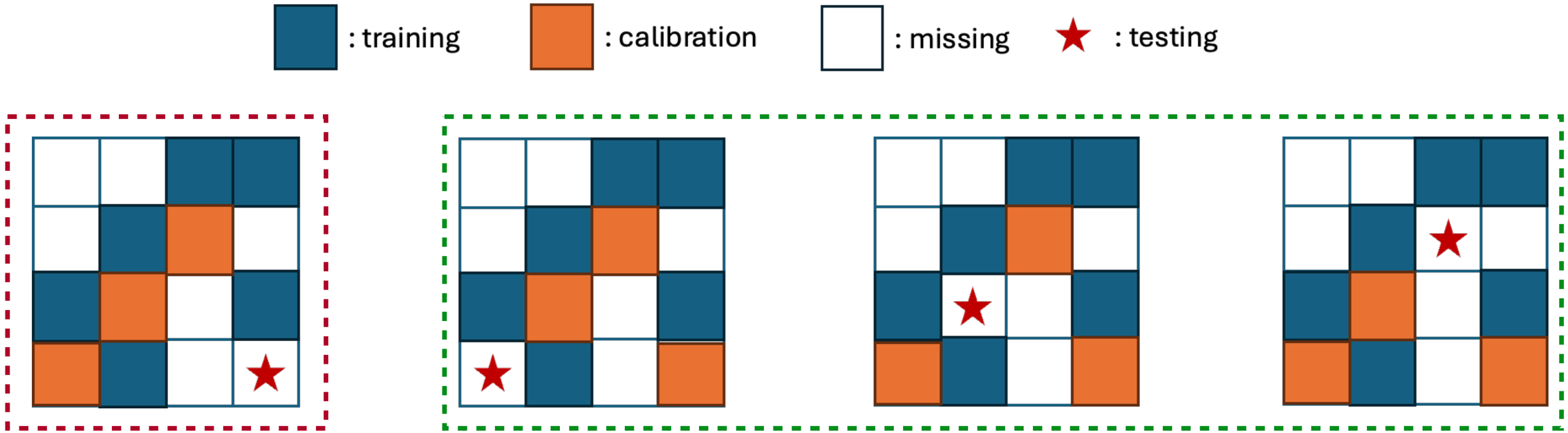
Weighted Conformal Prediction

Leave-one-out Likelihood Weighting



Weighted Conformal Prediction

Leave-one-out Likelihood Weighting



probability	p_1	p_2	p_3	p_4
weight of ★	$\frac{p_1}{p_1+p_2+p_3+p_4}$	$\frac{p_2}{p_1+p_2+p_3+p_4}$	$\frac{p_3}{p_1+p_2+p_3+p_4}$	$\frac{p_4}{p_1+p_2+p_3+p_4}$

Weighted Conformal Prediction

Leave-one-out Likelihood Weighting

Theorem 1

Let $\omega_k \propto \mathrm{P}([\mathscr{W}]_s = 1, \forall s \in \mathbf{\Omega}_{tr} \cup \{s_1, \ldots, s_{n_{cal}+1}\} \backslash \{s_k\}, [\mathscr{W}]_s = -1, \text{o.w.})$, then

$$\widehat{C}_{1-\alpha,s*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathscr{X}}]_{s*}) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \omega_s \cdot \delta_{\mathcal{S}([\mathscr{X}]_s, [\widehat{\mathscr{X}}]_s)} + \omega_{s*} \cdot \delta_{+\infty} \right) \right\}$$

is still valid, i.e. $\mathrm{P}([\mathscr{X}]_{s*} \in \widehat{C}_{1-\alpha,s*}) \geq 1 - \alpha$.

Weighted Conformal Prediction

Leave-one-out Likelihood Weighting

Theorem 1

Let $\omega_k \propto \mathrm{P}([\mathscr{W}]_s = 1, \forall s \in \mathbf{\Omega}_{tr} \cup \{s_1, \ldots, s_{n_{cal}+1}\} \backslash \{s_k\}, [\mathscr{W}]_s = -1, \text{o.w.})$, then

$$\widehat{C}_{1-\alpha,s*} = \left\{ x \in \mathbb{R} \,\middle|\, \mathcal{S}(x, [\widehat{\mathscr{X}}]_{s*}) \leq \hat{q}_{1-\alpha} \left(\sum_{s \in \mathbf{\Omega}_{cal}} \omega_s \cdot \delta_{\mathcal{S}([\mathscr{X}]_s, [\widehat{\mathscr{X}}]_s)} + \omega_{s*} \cdot \delta_{+\infty} \right) \right\}$$

is still valid, i.e. $\mathrm{P}([\mathscr{X}]_{s*} \in \widehat{C}_{1-\alpha,s*}) \geq 1 - \alpha$.

Now we need to estimate the likelihood $p(\mathscr{W})$ with a single sample.

Missing Propensity Model

Motivation

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:
  - model independent missingness with heterogeneous probabilities

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:
  - model independent missingness with heterogeneous probabilities

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad \mathscr{P} \in [0,1]^{d_1 \times \cdots \times d_K}$$

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:
  - model independent missingness with heterogeneous probabilities

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad \mathscr{P} \in [0,1]^{d_1 \times \cdots \times d_K}$$

  - model locally-dependent missingness

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:
  - model independent missingness with heterogeneous probabilities

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad \mathscr{P} \in [0,1]^{d_1 \times \cdots \times d_K}$$

  - model locally-dependent missingness

$$[\mathscr{W}]_{s_1} \perp [\mathscr{W}]_{s_2} \,|\, [\mathscr{W}]_j, j \in N(s_1)$$

Missing Propensity Model

Motivation

- We want to build a probabilistic model for a binary tensor $\mathscr{W}$ that can:
  - model independent missingness with heterogeneous probabilities

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad \mathscr{P} \in [0,1]^{d_1 \times \cdots \times d_K}$$

  - model locally-dependent missingness

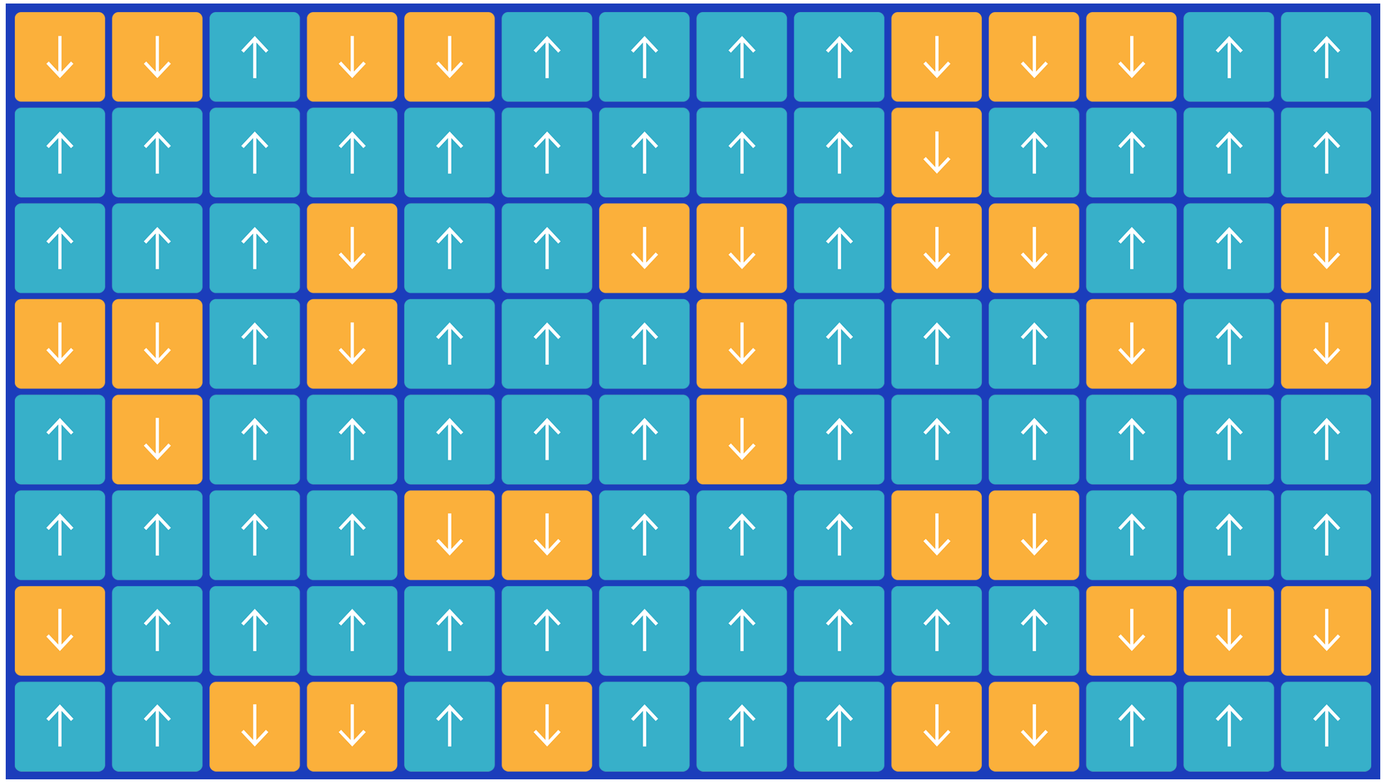
$$[\mathscr{W}]_{s_1} \perp [\mathscr{W}]_{s_2} \,|\, [\mathscr{W}]_j, j \in N(s_1)$$

where $N(s_1) = \{s \,|\, s \sim s_1\}$, and $s \sim s_1$ means s, s_1 are neighboring tensor entries.

Missing Propensity Model

Tensor Ising Model

- We adopt the idea from the Ising model [Cipra (1987)], which models the probability distribution of atomic spins on a lattice grid.

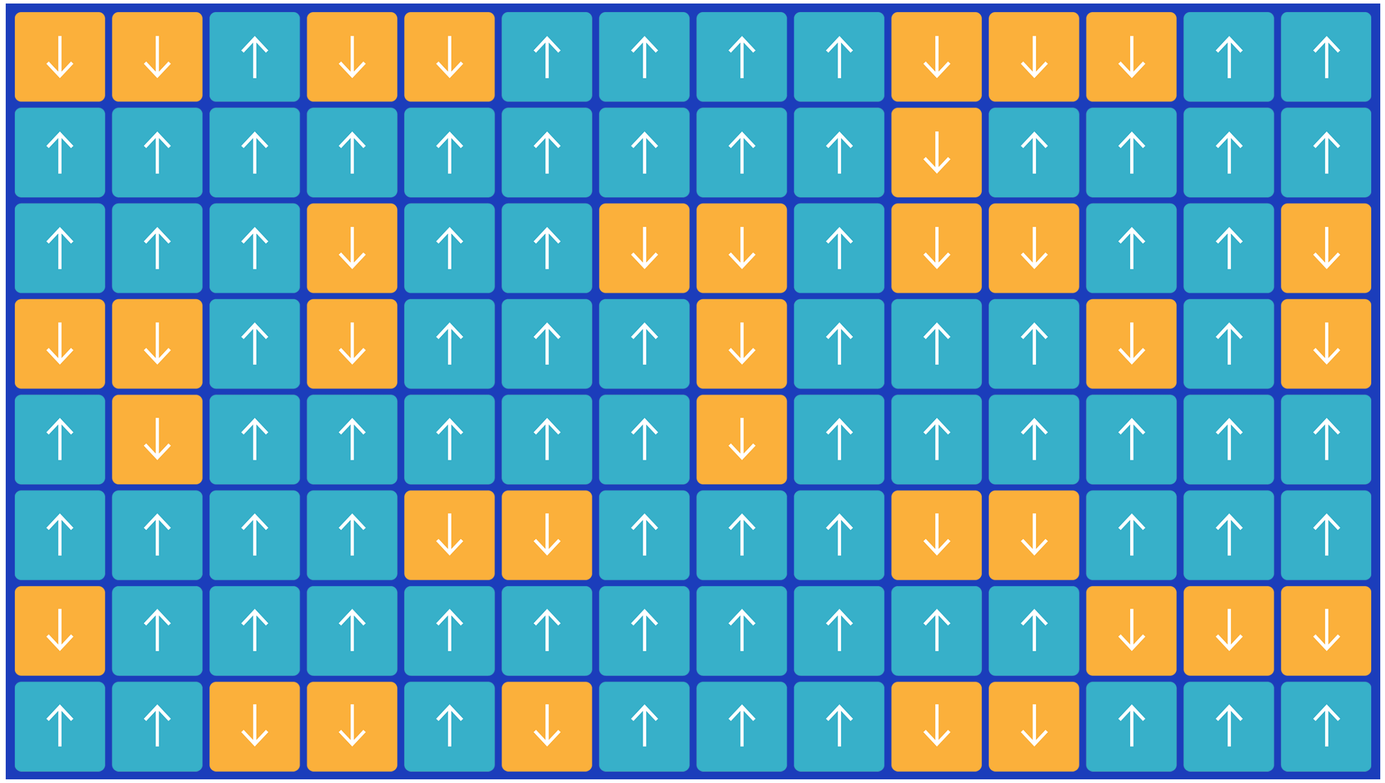


[source]

Missing Propensity Model

Tensor Ising Model

- We adopt the idea from the Ising model [Cipra (1987)], which models the probability distribution of atomic spins on a lattice grid.



[source]

Missing Propensity Model

Tensor Ising Model

- We model $p(\mathscr{W})$ via a Boltzmann distribution:

$$p(\mathscr{W}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g_{s_1 s_2} \cdot \overbrace{[\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2}}^{\text{"co-missingness"}} + \sum_{s_1} h_{s_1} \cdot [\mathscr{W}]_{s_1}\right]$$

Missing Propensity Model

Tensor Ising Model

- We model $p(\mathscr{W})$ via a Boltzmann distribution:

$$p(\mathscr{W}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g_{s_1 s_2} \cdot \overbrace{[\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2}}^{\text{"co-missingness"}} + \sum_{s_1} h_{s_1} \cdot [\mathscr{W}]_{s_1}\right]$$

$g_{s_1 s_2}$: interaction strength

Missing Propensity Model

Tensor Ising Model

- We model $p(\mathscr{W})$ via a Boltzmann distribution:

$$p(\mathscr{W}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g_{s_1 s_2} \cdot \overbrace{[\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2}}^{\text{"co-missingness"}} + \sum_{s_1} h_{s_1} \cdot [\mathscr{W}]_{s_1}\right]$$

$g_{s_1 s_2}$: interaction strength

h_{s_1}: individual missing propensity

Missing Propensity Model

Parameter Estimation Problem of Tensor Ising Model

Missing Propensity Model

Parameter Estimation Problem of Tensor Ising Model

- Given $\mathscr{W}$, we want to estimate g_{s_1,s_2} for arbitrary $s_1 \sim s_2$ and h_{s_1} for any s_1.

Missing Propensity Model

Parameter Estimation Problem of Tensor Ising Model

- Given $\mathscr{W}$, we want to estimate g_{s_1,s_2} for arbitrary $s_1 \sim s_2$ and h_{s_1} for any s_1.

- Existing works [Ravikumar, et al. (2010), Barber and Drton (2015), Liu et al. (2024)] on learning the graph structure of Ising model deal with **multiple samples** of **1-D or 2-D Ising model**.

Missing Propensity Model

Parameter Estimation Problem of Tensor Ising Model

- Given $\mathscr{W}$, we want to estimate g_{s_1,s_2} for arbitrary $s_1 \sim s_2$ and h_{s_1} for any s_1.

- Existing works [Ravikumar, et al. (2010), Barber and Drton (2015), Liu et al. (2024)] on learning the graph structure of Ising model deal with **multiple samples** of **1-D or 2-D Ising model**.

- In our work, we have **a single sample** but **an order-K Ising model**.

Missing Propensity Model

Latent Space Tensor Ising Model

Missing Propensity Model

Latent Space Tensor Ising Model

- We assume each tensor entry s has a latent feature $[\mathscr{B}]_s \in \mathbb{R}$, and:

Missing Propensity Model

Latent Space Tensor Ising Model

- We assume each tensor entry s has a latent feature $[\mathscr{B}]_s \in \mathbb{R}$, and:

$$p(\mathscr{W} \mid \mathscr{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathscr{B}]_{s_1}, [\mathscr{B}]_{s_2}) \cdot [\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2} + \sum_{s_1} h([\mathscr{B}]_{s_1}) \cdot [\mathscr{W}]_{s_1}\right],$$

Missing Propensity Model

Latent Space Tensor Ising Model

- We assume each tensor entry s has a latent feature $[\mathscr{B}]_s \in \mathbb{R}$, and:

$$p(\mathscr{W} \mid \mathscr{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathscr{B}]_{s_1}, [\mathscr{B}]_{s_2}) \cdot [\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2} + \sum_{s_1} h([\mathscr{B}]_{s_1}) \cdot [\mathscr{W}]_{s_1}\right],$$

where:

Missing Propensity Model

Latent Space Tensor Ising Model

- We assume each tensor entry s has a latent feature $[\mathscr{B}]_s \in \mathbb{R}$, and:

$$p(\mathscr{W} \mid \mathscr{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathscr{B}]_{s_1}, [\mathscr{B}]_{s_2}) \cdot [\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2} + \sum_{s_1} h([\mathscr{B}]_{s_1}) \cdot [\mathscr{W}]_{s_1}\right],$$

where:

- $g(\cdot, \cdot)$ is symmetric;

Missing Propensity Model

Latent Space Tensor Ising Model

- We assume each tensor entry s has a latent feature $[\mathscr{B}]_s \in \mathbb{R}$, and:

$$p(\mathscr{W} \mid \mathscr{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathscr{B}]_{s_1}, [\mathscr{B}]_{s_2}) \cdot [\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2} + \sum_{s_1} h([\mathscr{B}]_{s_1}) \cdot [\mathscr{W}]_{s_1}\right],$$

where:

- $g(\cdot, \cdot)$ is symmetric;
- $h(\cdot)$ is twice-continuously differentiable.

Example I: Bernoulli Model

Example I: Bernoulli Model

- If $g(\cdot, \cdot) = 0$, then:

Example I: Bernoulli Model

- If $g(\cdot, \cdot) = 0$, then:

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad [\mathscr{P}]_s = \frac{\exp[2h([\mathscr{B}]_s)]}{1 + \exp[2h([\mathscr{B}]_s)]}$$

Example I: Bernoulli Model

- If $g(\cdot\,,\cdot\,) = 0$, then:

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad [\mathscr{P}]_s = \frac{\exp[2h([\mathscr{B}]_s)]}{1 + \exp[2h([\mathscr{B}]_s)]}$$

- The conformal weight satisfies:

Example I: Bernoulli Model

- If $g(\cdot, \cdot) = 0$, then:

$$[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}([\mathscr{P}]_s), \quad [\mathscr{P}]_s = \frac{\exp[2h([\mathscr{B}]_s)]}{1 + \exp[2h([\mathscr{B}]_s)]}$$

- The conformal weight satisfies:

$$\omega_s \propto \frac{1 - [\mathscr{P}]_s}{[\mathscr{P}_s]} = \exp\left[-2h([\mathscr{B}]_s)\right]$$

Example II: Interaction-only Ising Model

Example II: Interaction-only Ising Model

- If $h(\cdot) = 0$,

Example II: Interaction-only Ising Model

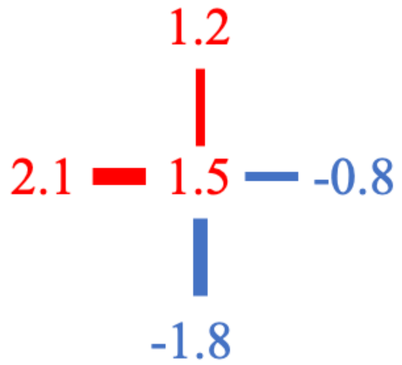
- If $h(\cdot) = 0$,

$$\widetilde{p}_s = \mathrm{P}([\mathscr{W}]_s = 1 \,|\, [\mathscr{W}]_{-s}, \mathscr{B}) = \frac{\exp\left[2 \sum_{j \in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j) \cdot [\mathscr{W}]_j\right]}{1 + \exp\left[2 \sum_{j \in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j) \cdot [\mathscr{W}]_j\right]}$$

Example II: Interaction-only Ising Model

- If $h(\cdot) = 0$,

$$\widetilde{p}_s = \mathrm{P}([\mathscr{W}]_s = 1 \,|\, [\mathscr{W}]_{-s}, \mathscr{B}) = \frac{\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}{1+\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}$$

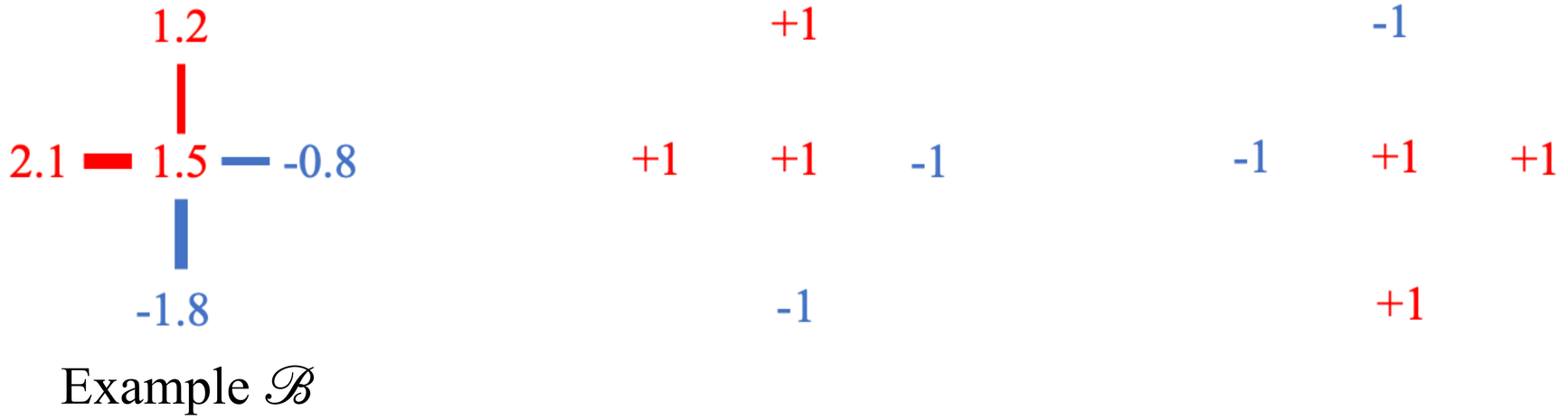


Example $\mathscr{B}$

Example II: Interaction-only Ising Model

- If $h(\cdot) = 0$,

$$\widetilde{p}_s = \mathrm{P}([\mathscr{W}]_s = 1 \,|\, [\mathscr{W}]_{-s}, \mathscr{B}) = \frac{\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}{1+\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}$$

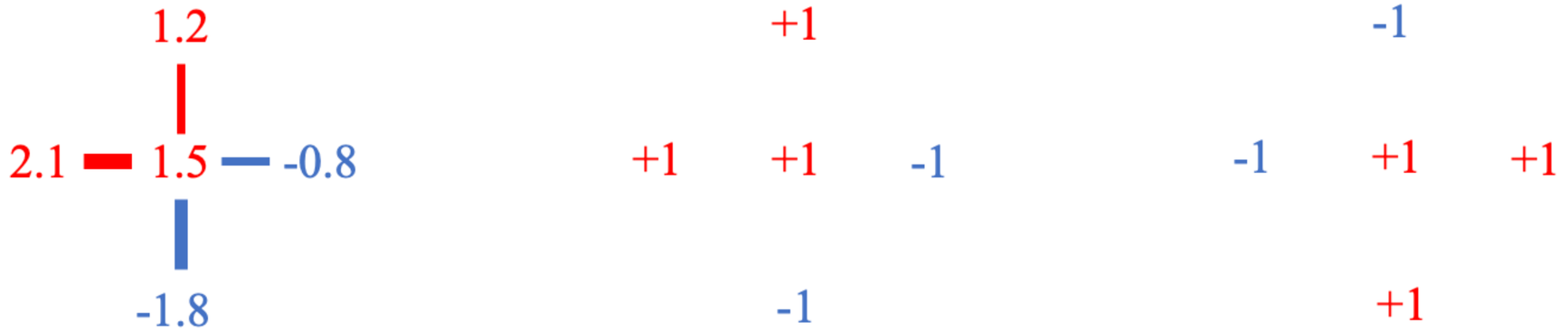


Example $\mathscr{B}$

Example II: Interaction-only Ising Model

- If $h(\cdot) = 0$,

$$\widetilde{p}_s = \mathrm{P}([\mathscr{W}]_s = 1 \,|\, [\mathscr{W}]_{-s}, \mathscr{B}) = \frac{\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}{1+\exp\left[2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}]_j)\cdot[\mathscr{W}]_j\right]}$$



Example $\mathscr{B}$ More likely scenario Less likely scenario

Example II: Interaction-only Ising Model

- The conformal weight (almost always) satisfies:

$$\omega_s \propto \frac{1-\widetilde{p}_s}{\widetilde{p}_s} = \exp\left[-2\sum_{j\in N(s)} g([\mathscr{B}]_s, [\mathscr{B}_j]) \cdot [\mathscr{W}]_j\right]$$

Exceptions are samples in $\mathbf{\Omega}_{cal} \cap N(s_{n+1})$, where $[\mathscr{W}]_{s_{n+1}} = -1$, but the leave-one-out procedure requires $[\mathscr{W}]_{s_{n+1}} = 1$.

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

- For theoretical reason [Barber et al. (2023)], we can only fit the Ising model with $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, defined as:

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

- For theoretical reason [Barber et al. (2023)], we can only fit the Ising model with $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, defined as:

$$[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s = 1, \quad \text{if and only if } s \in \mathbf{\Omega}_{tr}$$

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

- For theoretical reason [Barber et al. (2023)], we can only fit the Ising model with $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, defined as:

$$[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s = 1, \quad \text{if and only if } s \in \mathbf{\Omega}_{tr}$$

- The pseudo-likelihood of $\mathscr{W}_{\mathbf{\Omega}_{tr}}$ posits that $[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s \overset{ind.}{\sim} \text{Bern}\left(q \cdot \widetilde{p}_s\right)$, thus the MPLE is:

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

- For theoretical reason [Barber et al. (2023)], we can only fit the Ising model with $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, defined as:

$$[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s = 1, \quad \text{if and only if } s \in \mathbf{\Omega}_{tr}$$

- The pseudo-likelihood of $\mathscr{W}_{\mathbf{\Omega}_{tr}}$ posits that $[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s \overset{ind.}{\sim} \text{Bern}\left(q \cdot \widetilde{p}_s\right)$, thus the MPLE is:

$$\widehat{\mathscr{B}} = \underset{\mathscr{B} \in \mathbb{F}}{\text{argmin}} - \sum_{s:[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s=1} \log(q\widetilde{p}_s) - \sum_{s:[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s=-1} \log(1 - q\widetilde{p}_s)$$

Estimation Method

Maximum Pseudo-Likelihood Estimation (MPLE)

- For theoretical reason [Barber et al. (2023)], we can only fit the Ising model with $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, defined as:

$$[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s = 1, \quad \text{if and only if } s \in \mathbf{\Omega}_{tr}$$

- The pseudo-likelihood of $\mathscr{W}_{\mathbf{\Omega}_{tr}}$ posits that $[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s \overset{ind.}{\sim} \text{Bern}\left(q \cdot \widetilde{p}_s\right)$, thus the MPLE is:

$$\widehat{\mathscr{B}} = \underset{\mathscr{B} \in \mathbb{F}}{\text{argmin}} - \sum_{s:[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s=1} \log(q\widetilde{p}_s) - \sum_{s:[\mathscr{W}_{\mathbf{\Omega}_{tr}}]_s=-1} \log(1 - q\widetilde{p}_s)$$

where $q \in (0,1)$ is the probability of assigning data to the training set, and $\mathbb{F}$ is a feasible set of $\mathscr{B}$.

Estimation Method

Low-rank MPLE

Estimation Method

Low-rank MPLE

- Given a single sample $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, we need to impose a parsimonious structure over $\mathscr{B}$ for estimation.

Estimation Method

Low-rank MPLE

- Given a single sample $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, we need to impose a parsimonious structure over $\mathscr{B}$ for estimation.
- We set $\mathbb{F}$ to be the set of low tensor-train (TT) rank tensors [Oseledets (2011)]:

Estimation Method

Low-rank MPLE

- Given a single sample $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, we need to impose a parsimonious structure over $\mathscr{B}$ for estimation.
- We set $\mathbb{F}$ to be the set of low tensor-train (TT) rank tensors [Oseledets (2011)]:

$$\mathbb{F} = \{\mathscr{B} \in \mathbb{R}^{d_1 \times \cdots \times d_K} \,|\, [\mathscr{B}]_{i_1 \ldots i_K} = \prod_{k=1}^{K} [\mathscr{T}^k]_{:i_k:}, \mathscr{T}^k \in \mathbb{R}^{r_{k-1} \times d_k \times r_k}\},$$

Estimation Method

Low-rank MPLE

- Given a single sample $\mathscr{W}_{\mathbf{\Omega}_{tr}}$, we need to impose a parsimonious structure over $\mathscr{B}$ for estimation.
- We set $\mathbb{F}$ to be the set of low tensor-train (TT) rank tensors [Oseledets (2011)]:

$$\mathbb{F} = \{\mathscr{B} \in \mathbb{R}^{d_1 \times \cdots \times d_K} \,|\, [\mathscr{B}]_{i_1 \ldots i_K} = \prod_{k=1}^{K} [\mathscr{T}^k]_{:i_k:}, \mathscr{T}^k \in \mathbb{R}^{r_{k-1} \times d_k \times r_k}\},$$

with $r_0 = r_K = 1$, and $\mathbf{r} = (r_1, \ldots, r_{K-1})$ being the TT-rank.

Estimation Algorithm

Riemannian Gradient Descent (RGrad)

Estimation Algorithm

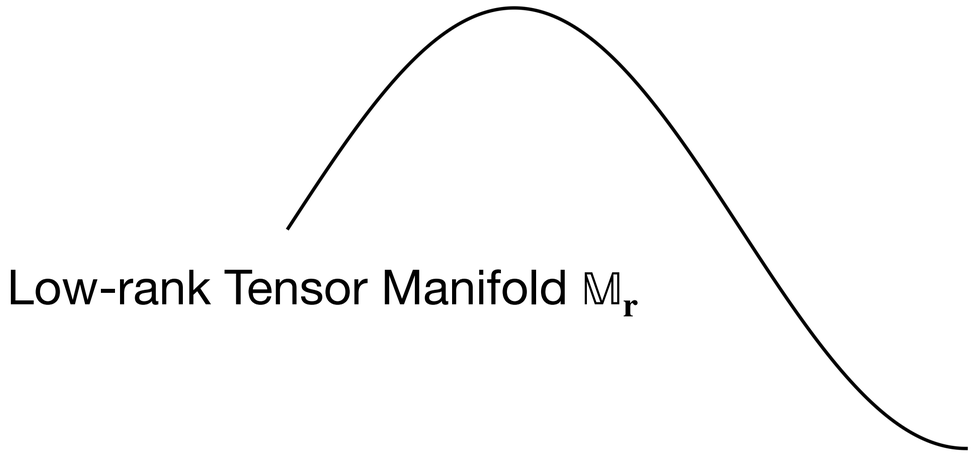
Riemannian Gradient Descent (RGrad)

- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]

Estimation Algorithm

Riemannian Gradient Descent (RGrad)

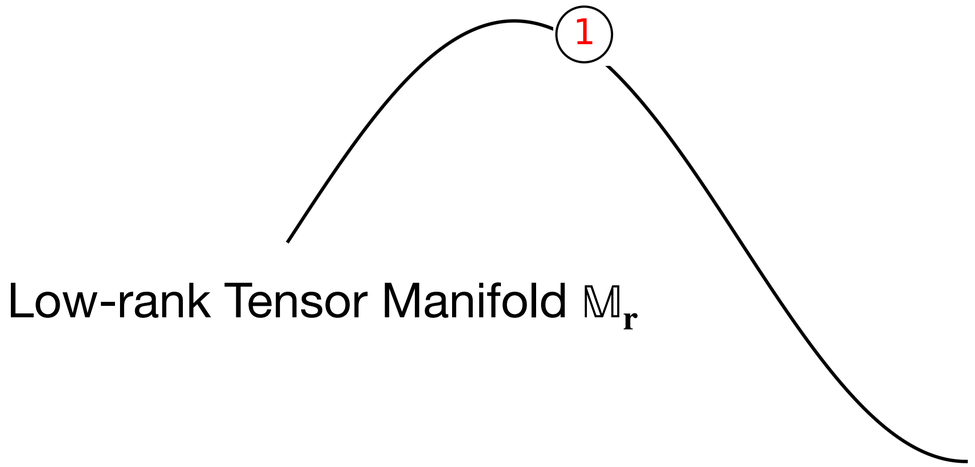
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

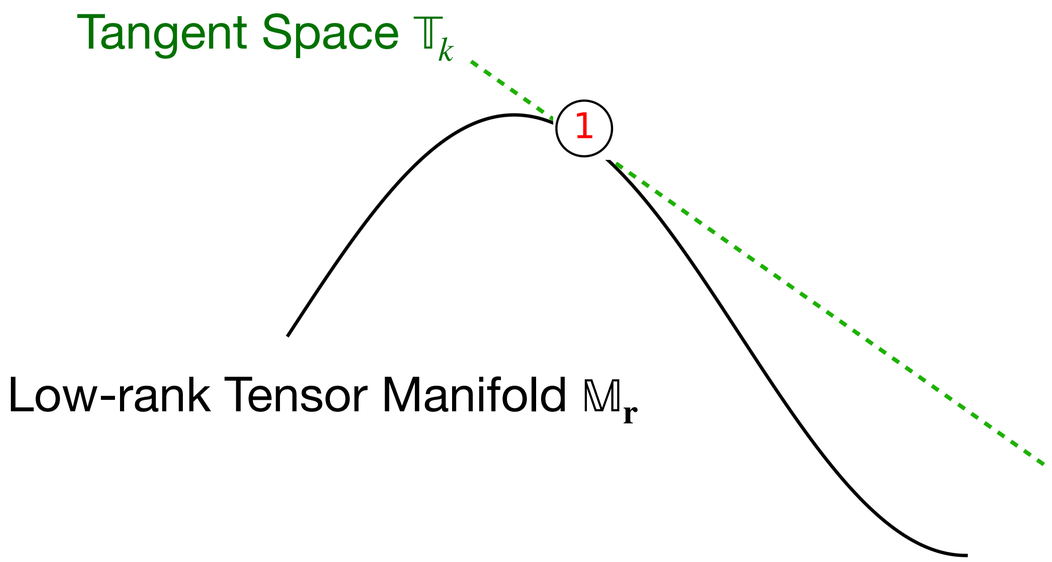
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

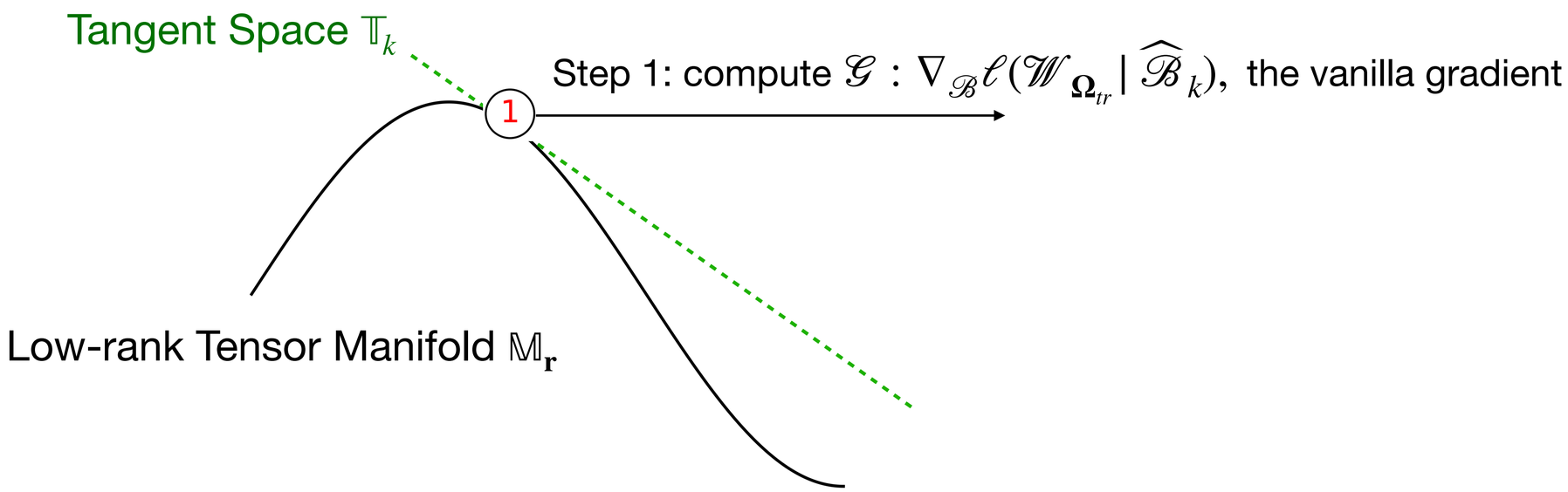
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

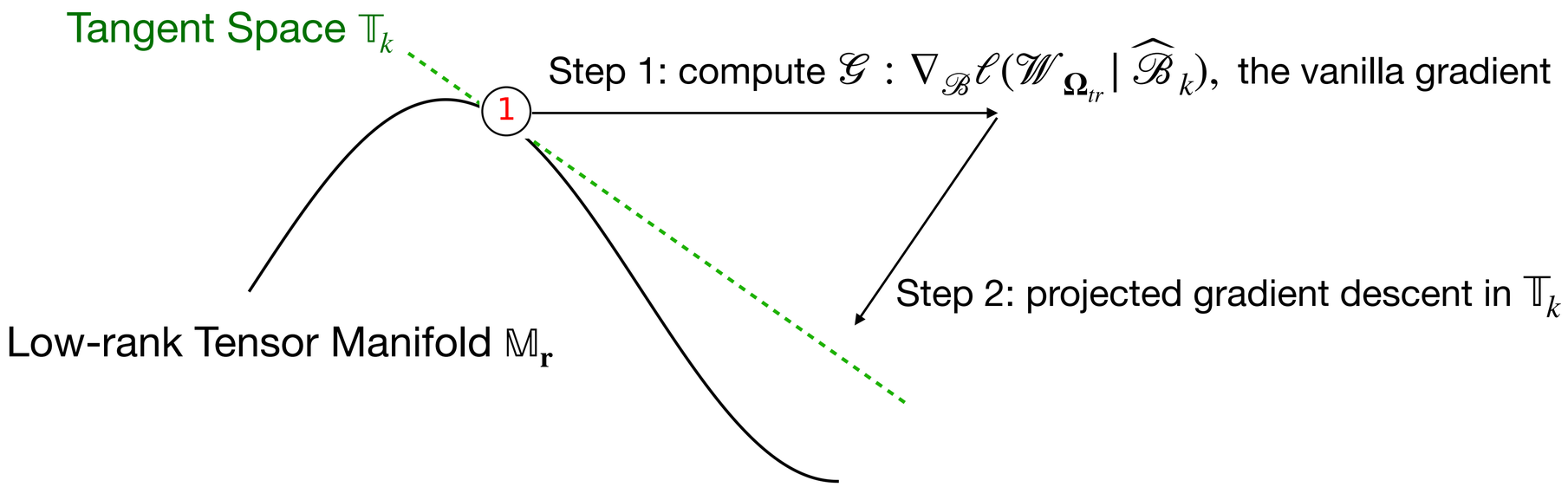
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

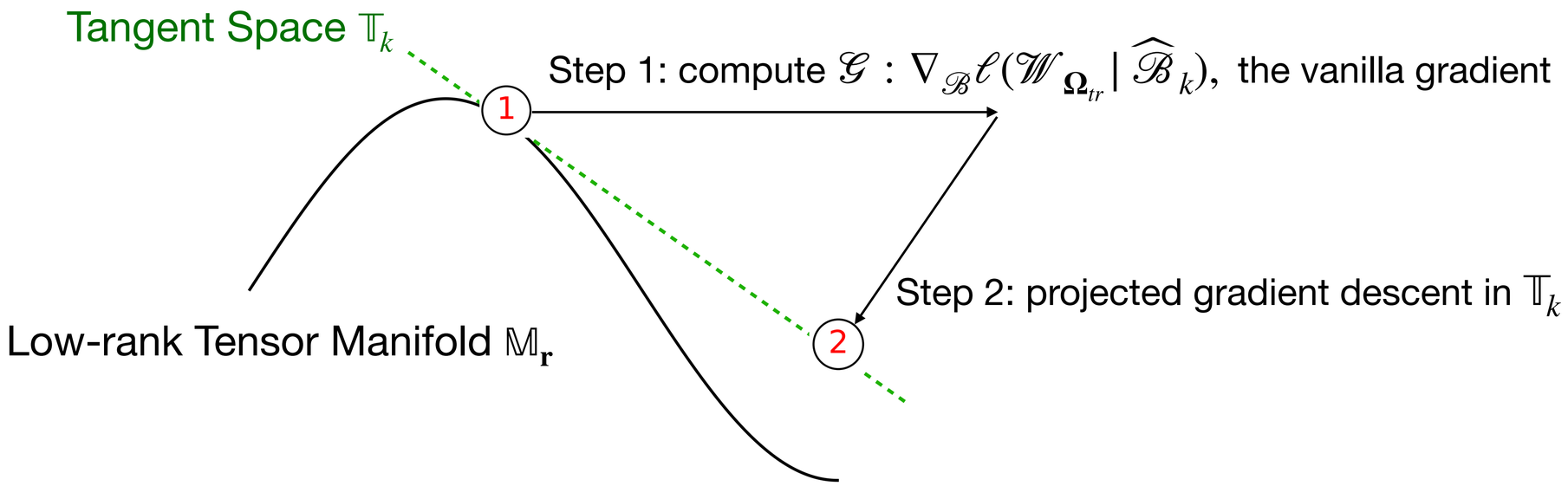
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

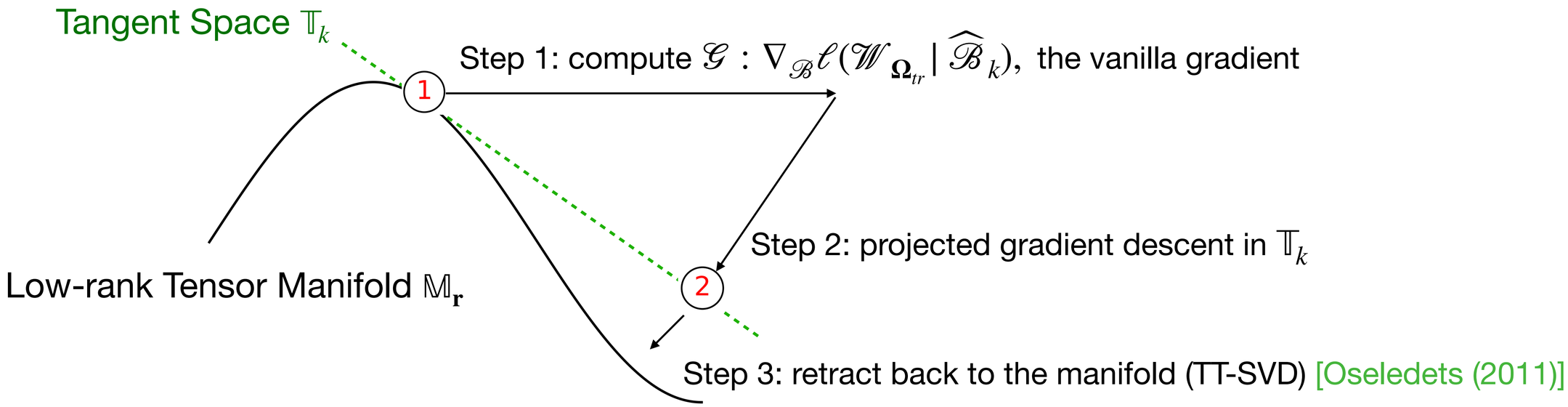
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]



Estimation Algorithm

Riemannian Gradient Descent (RGrad)

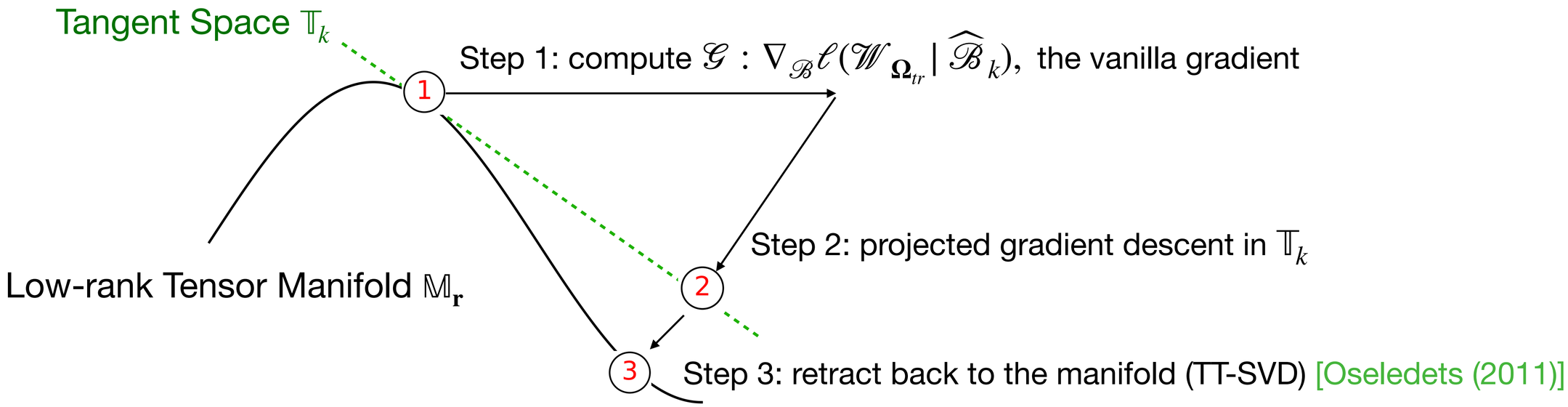
- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]

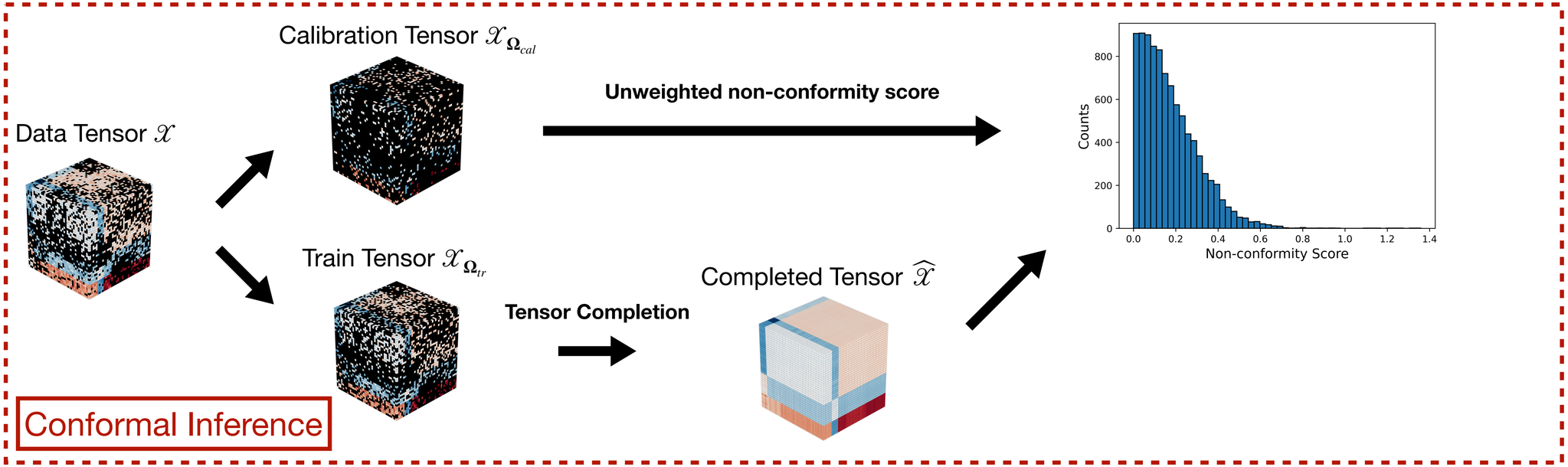


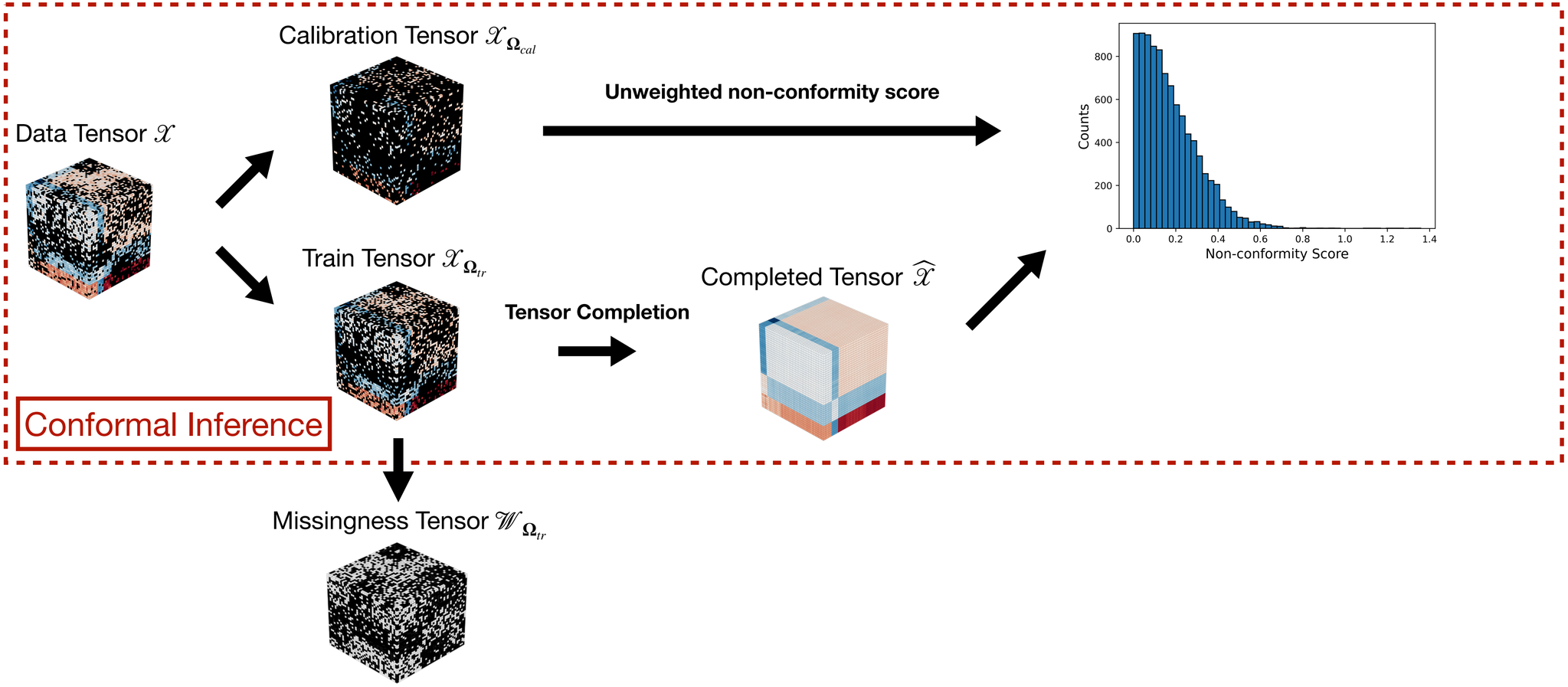
Estimation Algorithm

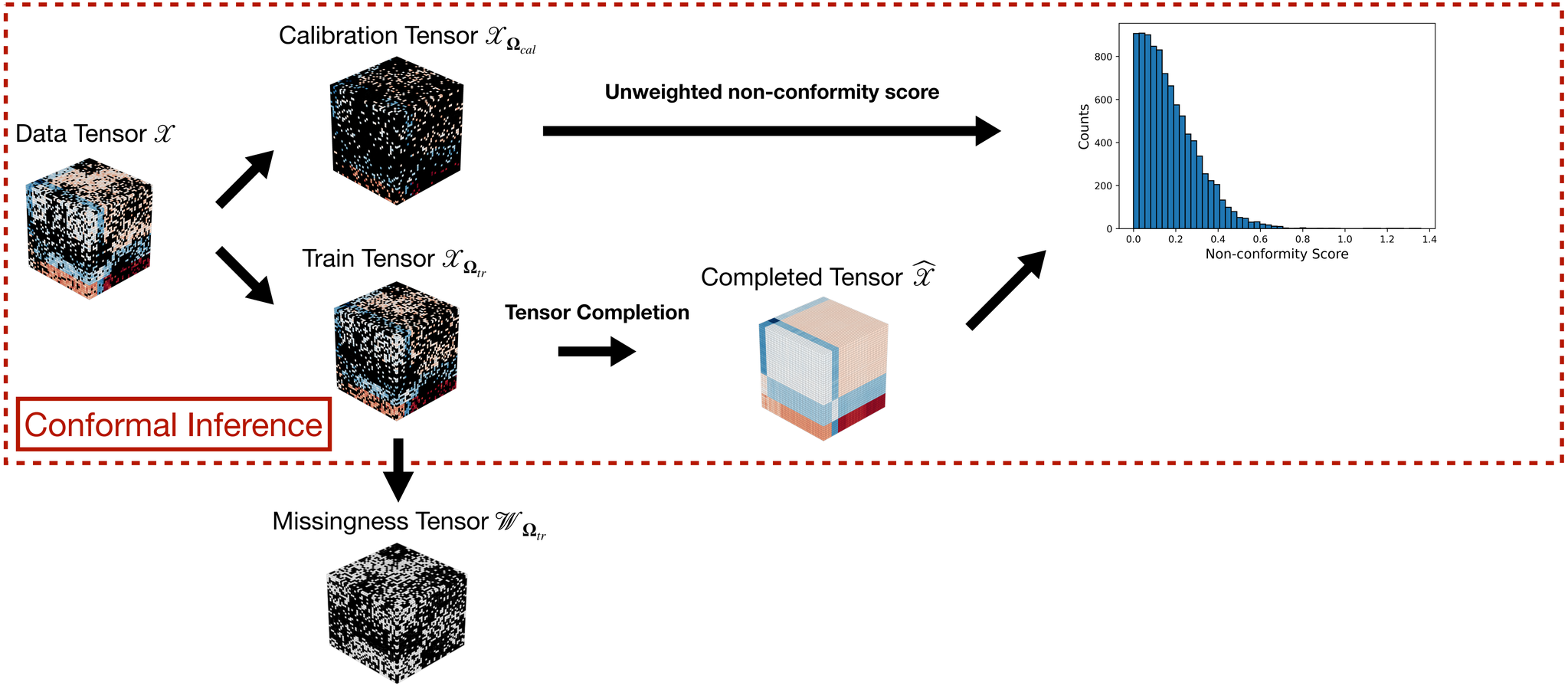
Riemannian Gradient Descent (RGrad)

- Low tensor-train rank tensors lie on a smooth manifold $\mathbb{M}_{\mathbf{r}}$ [Holtz et al. (2012)]




Calibration Tensor $\mathcal{X}_{\Omega_{cal}}$
Data Tensor $\mathcal{X}$
Train Tensor $\mathcal{X}_{\Omega_{tr}}$
Unweighted non-conformity score
Tensor Completion
Completed Tensor $\widehat{\mathcal{X}}$
Counts
Non-conformity Score
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Conformal Inference

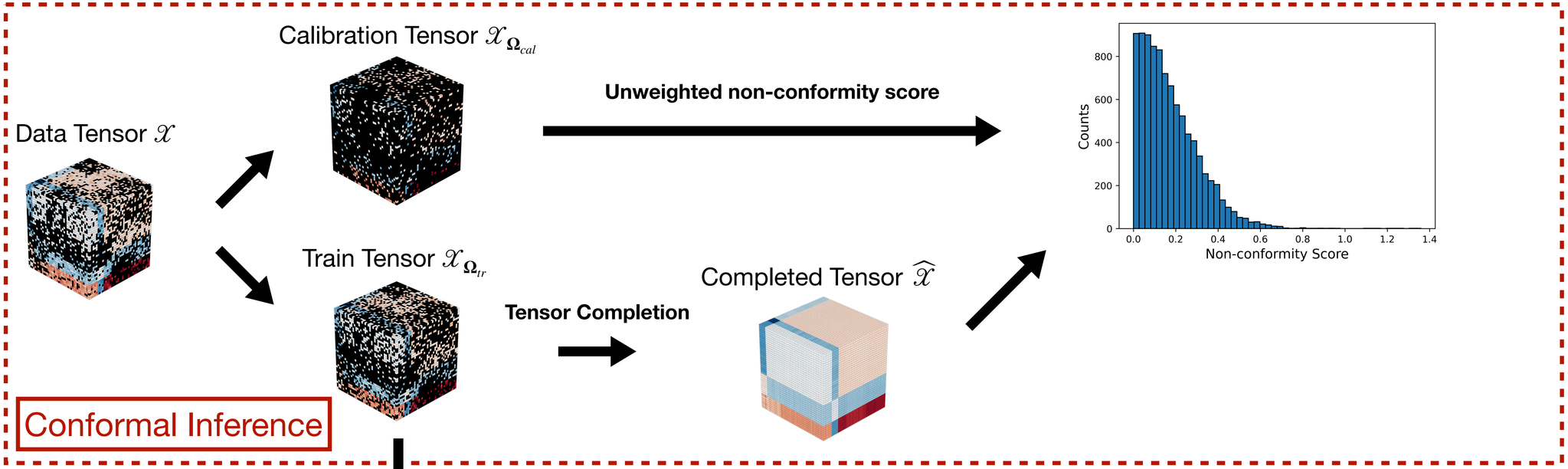

Calibration Tensor $\mathcal{X}_{\Omega_{cal}}$
Data Tensor $\mathcal{X}$
Train Tensor $\mathcal{X}_{\Omega_{tr}}$
Unweighted non-conformity score
Tensor Completion
Completed Tensor $\widehat{\mathcal{X}}$
Counts
Non-conformity Score
Conformal Inference
Missingness Tensor $\mathcal{W}_{\Omega_{tr}}$



Low-rank Tensor Ising Model

$$p(\mathcal{W} \mid \mathcal{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathcal{B}]_{s_1}, [\mathcal{B}]_{s_2})[\mathcal{W}]_{s_1}[\mathcal{W}]_{s_2}\right]$$

$$\times \exp\left[\sum_{s_1} h([\mathcal{B}]_{s_1})[\mathcal{W}]_{s_1}\right], \quad \mathcal{B} \in \mathbb{M}_{\mathbf{r}}$$

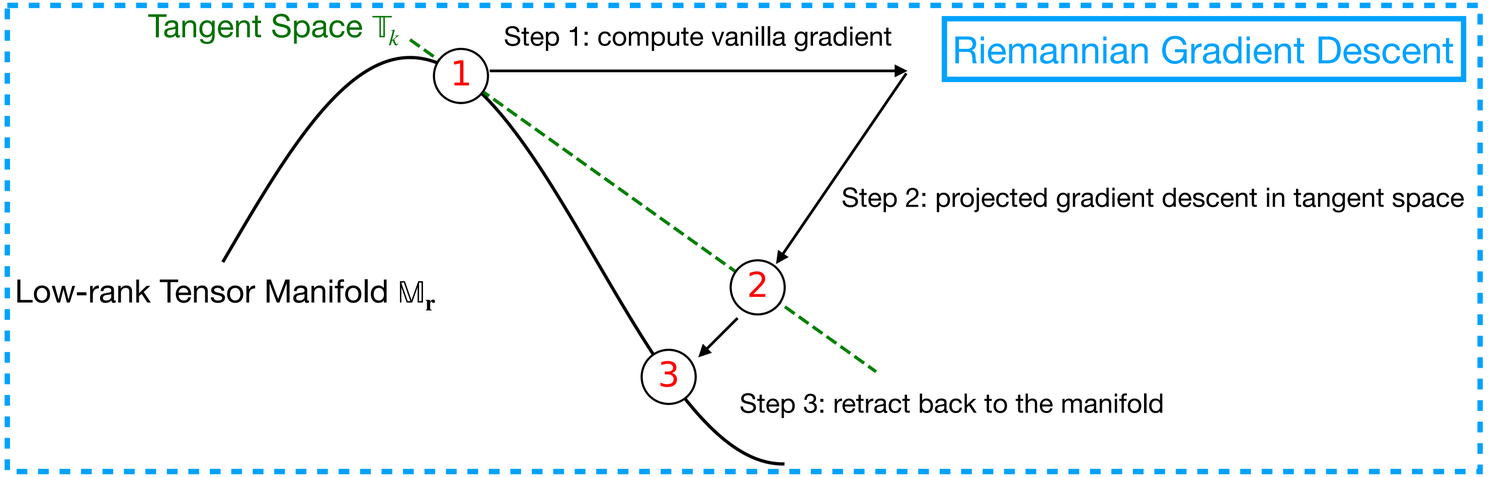


Low-rank Tensor Ising Model

$$p(\mathcal{W} \mid \mathcal{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathcal{B}]_{s_1}, [\mathcal{B}]_{s_2})[\mathcal{W}]_{s_1}[\mathcal{W}]_{s_2}\right]$$

$$\times \exp\left[\sum_{s_1} h([\mathcal{B}]_{s_1})[\mathcal{W}]_{s_1}\right], \quad \mathcal{B} \in \mathbb{M}_{\mathbf{r}}$$

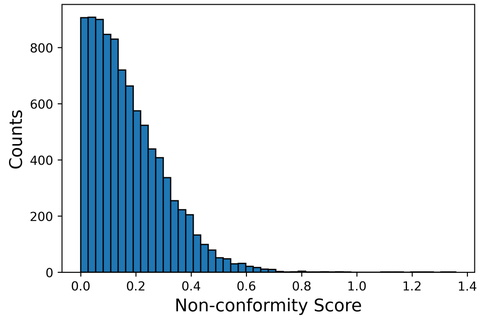
Riemannian Gradient Descent



Calibration Tensor $\mathcal{X}_{\mathbf{\Omega}_{cal}}$

Data Tensor $\mathcal{X}$

Unweighted non-conformity score



Train Tensor $\mathcal{X}_{\mathbf{\Omega}_{tr}}$

Tensor Completion

Completed Tensor $\widehat{\mathcal{X}}$

Conformal Inference

Missingness Tensor $\mathcal{W}_{\mathbf{\Omega}_{tr}}$

MPLE Estimator $\widehat{\mathcal{B}}$

Low-rank Tensor Ising Model

$$p(\mathcal{W} \mid \mathcal{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathcal{B}]_{s_1}, [\mathcal{B}]_{s_2})[\mathcal{W}]_{s_1}[\mathcal{W}]_{s_2}\right]$$

$$\times \exp\left[\sum_{s_1} h([\mathcal{B}]_{s_1})[\mathcal{W}]_{s_1}\right], \quad \mathcal{B} \in \mathbb{M}_{\mathbf{r}}$$

Riemannian Gradient Descent

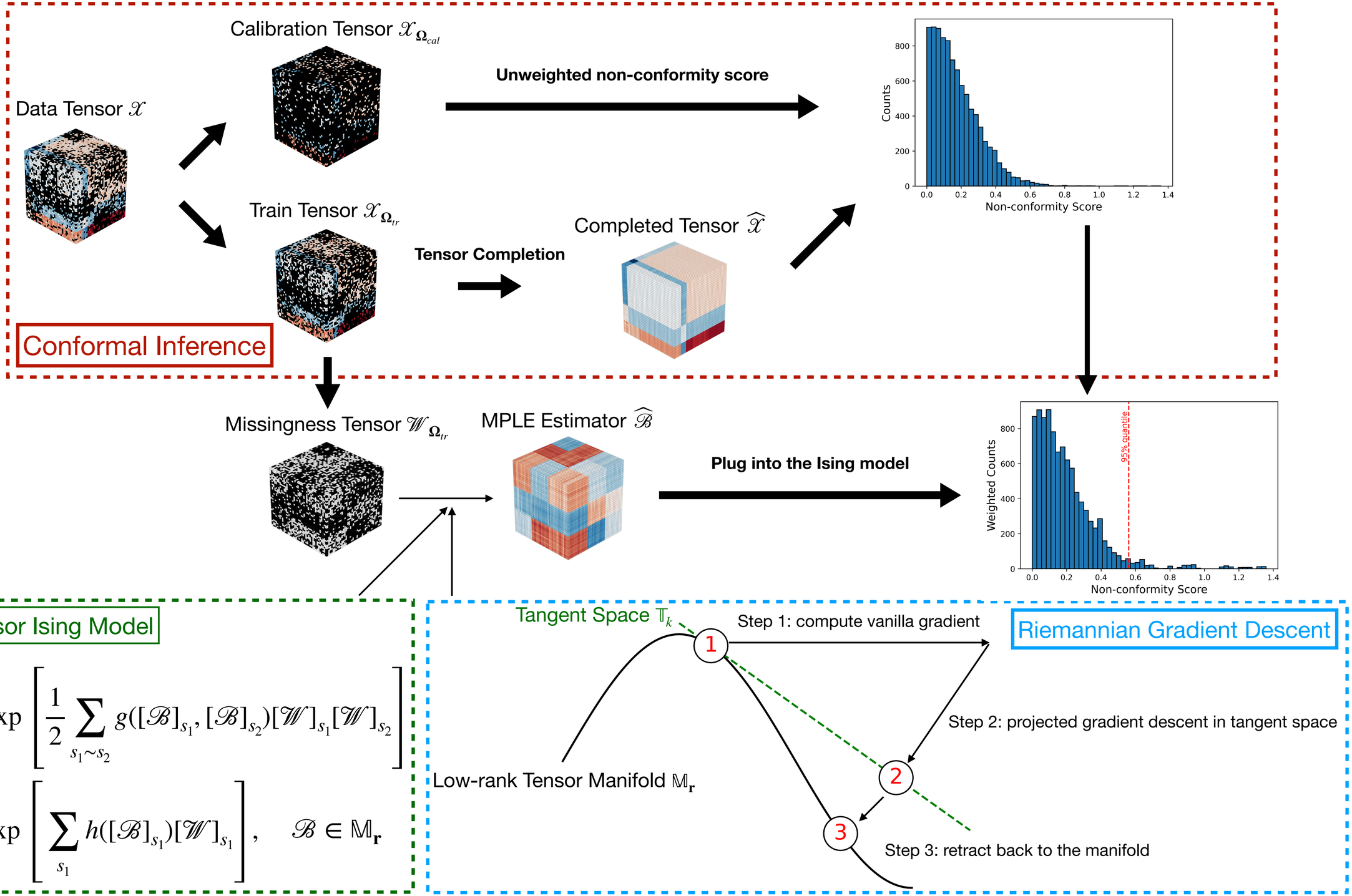
Tangent Space $\mathbb{T}_k$

Step 1: compute vanilla gradient

Step 2: projected gradient descent in tangent space

Step 3: retract back to the manifold

Low-rank Tensor Manifold $\mathbb{M}_{\mathbf{r}}$


Calibration Tensor $\mathcal{X}_{\Omega_{cal}}$
Data Tensor $\mathcal{X}$
Train Tensor $\mathcal{X}_{\Omega_{tr}}$
Unweighted non-conformity score
Counts
Non-conformity Score
Completed Tensor $\widehat{\mathcal{X}}$
Tensor Completion
Conformal Inference
Missingness Tensor $\mathscr{W}_{\Omega_{tr}}$
MPLE Estimator $\widehat{\mathscr{B}}$
Plug into the Ising model
Weighted Counts
95% quantile
Non-conformity Score
Low-rank Tensor Ising Model
$p(\mathscr{W} \mid \mathscr{B}) \propto \exp\left[\frac{1}{2}\sum_{s_1 \sim s_2} g([\mathscr{B}]_{s_1}, [\mathscr{B}]_{s_2})[\mathscr{W}]_{s_1}[\mathscr{W}]_{s_2}\right]$
$\times \exp\left[\sum_{s_1} h([\mathscr{B}]_{s_1})[\mathscr{W}]_{s_1}\right], \quad \mathscr{B} \in \mathbb{M}_{\mathbf{r}}$
Tangent Space $\mathbb{T}_k$
Step 1: compute vanilla gradient
Riemannian Gradient Descent
Step 2: projected gradient descent in tangent space
Low-rank Tensor Manifold $\mathbb{M}_{\mathbf{r}}$
Step 3: retract back to the manifold

Theoretical Results

MPLE error bound

Theoretical Results

MPLE error bound

Theorem 2 Assume that:

- $g(\cdot, \cdot) = 0$
- $h(\cdot)$ is twice-continuously differentiable and $h''(\cdot) \geq 0$
- $\|\mathscr{B}^*\|_\infty \leq \xi$

Theoretical Results

MPLE error bound

Theorem 2 Assume that:

- $g(\cdot, \cdot) = 0$
- $h(\cdot)$ is twice-continuously differentiable and $h''(\cdot) \geq 0$
- $\|\mathcal{B}^*\|_\infty \leq \xi$

$$\mathrm{P}\left[\frac{1}{\sqrt{d^*}}\|\widehat{\mathcal{B}} - \mathcal{B}^*\|_{\mathrm{F}} \leq 2C_{K,\xi}\sqrt{\frac{r^*\bar{d}}{d^*}}\right] \geq 1 - \exp\left(-C_1\bar{d}\log K\right),$$

where $r^*, \bar{d}, d^*$ are $\prod_k r_k, \sum_k d_k, \prod_k d_k$ for $\mathcal{B}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with tensor-train rank $\mathbf{r} = (r_1, \ldots, r_{K-1})$.

Theoretical Results

MPLE error bound

Theorem 2 Assume that:

- $g(\cdot, \cdot) = 0$
- $h(\cdot)$ is twice-continuously differentiable and $h''(\cdot) \geq 0$
- $\|\mathscr{B}^*\|_\infty \leq \xi$

$$\mathrm{P}\left[\frac{1}{\sqrt{d^*}}\|\widehat{\mathscr{B}} - \mathscr{B}^*\|_{\mathrm{F}} \leq 2C_{K,\xi}\sqrt{\frac{r^*\bar{d}}{d^*}}\right] \geq 1 - \exp\left(-C_1\bar{d}\log K\right),$$

RMSE of the MPLE

where $r^*, \bar{d}, d^*$ are $\prod_k r_k, \sum_k d_k, \prod_k d_k$ for $\mathscr{B}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with tensor-train rank $\mathbf{r} = (r_1, \ldots, r_{K-1})$.

Theoretical Results

MPLE error bound

Theorem 2 Assume that:

- $g(\cdot, \cdot) = 0$
- $h(\cdot)$ is twice-continuously differentiable and $h''(\cdot) \geq 0$
- $\|\mathscr{B}^*\|_\infty \leq \xi$

positive constants related to K and ξ

$$\mathrm{P}\left[\frac{1}{\sqrt{d^*}}\|\widehat{\mathscr{B}} - \mathscr{B}^*\|_{\mathrm{F}} \leq 2C_{K,\xi}\sqrt{\frac{r^*\bar{d}}{d^*}}\right] \geq 1 - \exp\left(-C_1\bar{d}\log K\right),$$

RMSE of the MPLE

where $r^*, \bar{d}, d^*$ are $\prod_k r_k, \sum_k d_k, \prod_k d_k$ for $\mathscr{B}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with tensor-train rank $\mathbf{r} = (r_1, \ldots, r_{K-1})$.

Theoretical Results

MPLE error bound

Theorem 2 Assume that:

- $g(\cdot, \cdot) = 0$
- $h(\cdot)$ is twice-continuously differentiable and $h''(\cdot) \geq 0$
- $\|\mathscr{B}^*\|_\infty \leq \xi$

positive constants related to K and ξ

$$\mathrm{P}\left[\frac{1}{\sqrt{d^*}}\|\widehat{\mathscr{B}} - \mathscr{B}^*\|_{\mathrm{F}} \leq 2C_{K,\xi}\sqrt{\frac{r^*\bar{d}}{d^*}}\right] \geq 1 - \exp\left(-C_1\bar{d}\log K\right),$$

RMSE of the MPLE

$\asymp (r/d)^{(K-1)/2}$, if $r_k \asymp r, d_k \asymp d, \forall k$

where $r^*, \bar{d}, d^*$ are $\prod_k r_k, \sum_k d_k, \prod_k d_k$ for $\mathscr{B}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with tensor-train rank $\mathbf{r} = (r_1, \ldots, r_{K-1})$.

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

$$\mathrm{E}\left[\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c}\mathbb{I}_{\{[\mathcal{X}]_s\in\widehat{C}_{1-\alpha,s}\}}\right] \geq 1-\alpha-\frac{2c_{K,\xi}}{(1-c)(1-q)}\sqrt{\frac{r^*\bar{d}}{d^*}} - \exp[-C_1\bar{d}\log K]-\exp\left[-\frac{c^2(1-q)d^*l_\xi}{2}\right]$$

where $q \in (0,1)$ is the train-calibration split ratio .

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

average coverage on $\boldsymbol{\Omega}^c$

$$\mathrm{E}\left[\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c} \mathbb{I}_{\{[\mathcal{X}]_s \in \widehat{C}_{1-\alpha,s}\}}\right] \geq 1-\alpha-\frac{2c_{K,\xi}}{(1-c)(1-q)}\sqrt{\frac{r^*\bar{d}}{d^*}}$$

$$-\exp[-C_1\bar{d}\log K]-\exp\left[-\frac{c^2(1-q)d^*l_\xi}{2}\right]$$

where $q \in (0,1)$ is the train-calibration split ratio .

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

average coverage on $\boldsymbol{\Omega}^c$ ↑ target coverage ↑

$$\mathrm{E}\left[\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c} \mathbb{I}_{\{[\mathcal{X}]_s \in \widehat{C}_{1-\alpha,s}\}}\right] \geq 1-\alpha-\frac{2c_{K,\xi}}{(1-c)(1-q)}\sqrt{\frac{r^*\bar{d}}{d^*}}$$

$$-\exp[-C_1\bar{d}\log K]-\exp\left[-\frac{c^2(1-q)d^*l_\xi}{2}\right]$$

where $q \in (0,1)$ is the train-calibration split ratio .

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

average coverage on $\boldsymbol{\Omega}^c$ | target coverage | MPLE estimation error

$$\mathrm{E}\left[\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c}\mathbb{I}_{\{[\mathcal{X}]_s\in\widehat{C}_{1-\alpha,s}\}}\right] \geq 1-\alpha-\frac{2c_{K,\xi}}{(1-c)(1-q)}\sqrt{\frac{r^*\bar{d}}{d^*}}$$

$$-\exp[-C_1\bar{d}\log K]-\exp\left[-\frac{c^2(1-q)d^*l_\xi}{2}\right]$$

where $q \in (0,1)$ is the train-calibration split ratio .

Theoretical Results

Coverage Guarantee

Theorem 3 Assume the same assumptions as Theorem 2, then for any $0 < c < 1$:

average coverage on $\boldsymbol{\Omega}^c$ | target coverage | MPLE estimation error

$$\mathrm{E}\left[\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c} \mathbb{I}_{\{[\mathcal{X}]_s \in \widehat{C}_{1-\alpha,s}\}}\right] \geq 1-\alpha-\frac{2c_{K,\xi}}{(1-c)(1-q)}\sqrt{\frac{r^*\bar{d}}{d^*}} - \exp[-C_1\bar{d}\log K] - \exp\left[-\frac{c^2(1-q)d^*l_\xi}{2}\right]$$

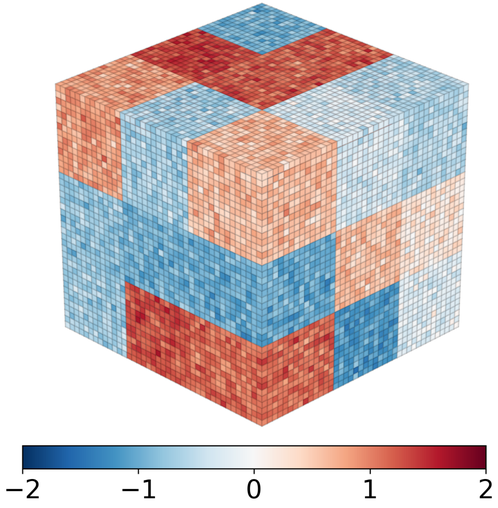
where $q \in (0,1)$ is the train-calibration split ratio.

sample size requirement for calibration

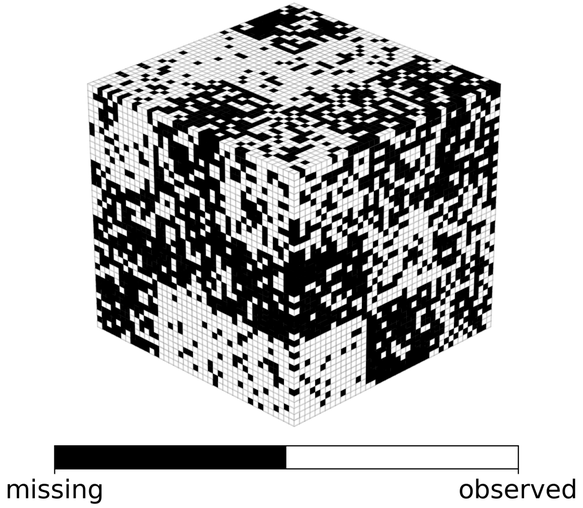
Numerical Experiments

Setup

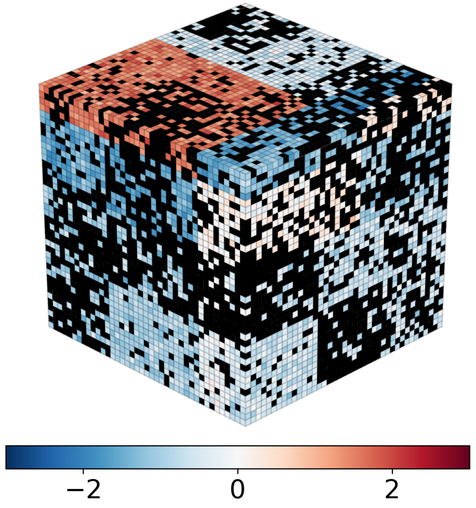
(a) Ground Truth $\mathcal{B}^*$



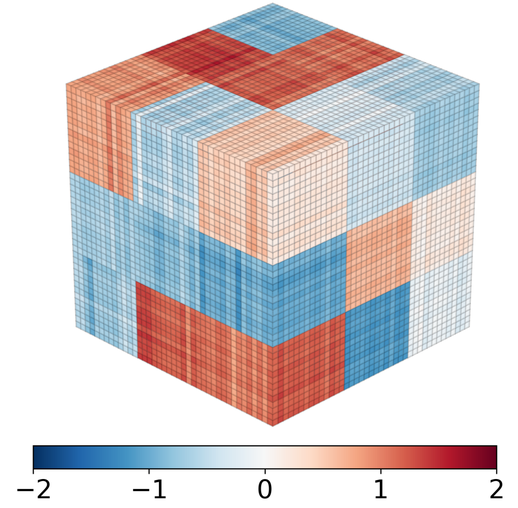
(b) Simulated Missingness $\mathcal{W}$



(c) Simulated Data Tensor $\mathcal{X}$



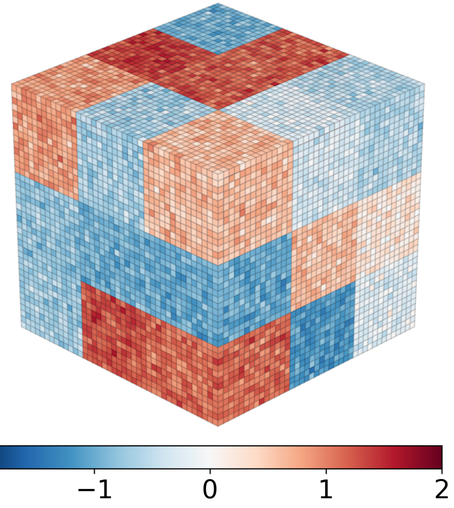
(d) Fitted $\widehat{\mathcal{B}}$ by RGrad



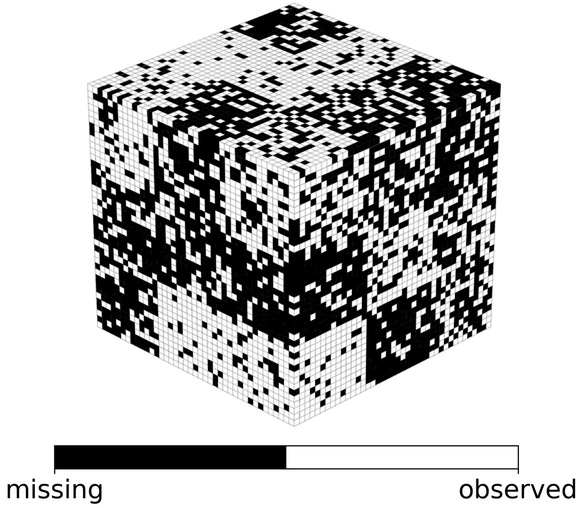
Numerical Experiments

Setup

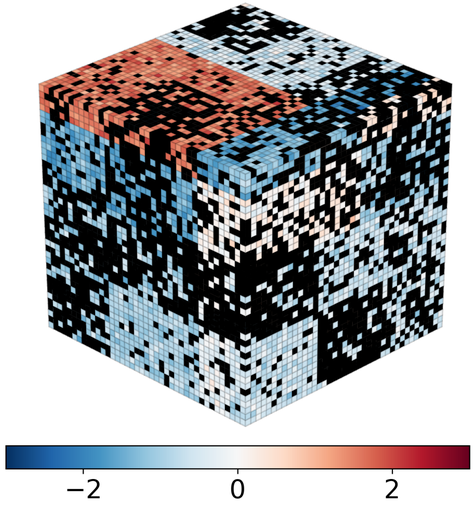
(a) Ground Truth $\mathcal{B}^*$



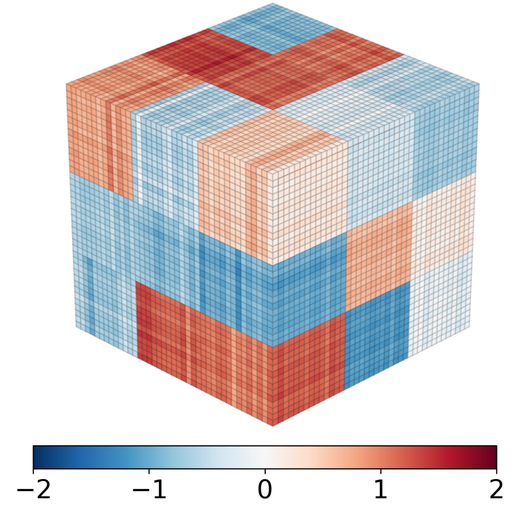
(b) Simulated Missingness $\mathcal{W}$



(c) Simulated Data Tensor $\mathcal{X}$



(d) Fitted $\widehat{\mathcal{B}}$ by RGrad



red: more likely to observe

blue: more likely to miss

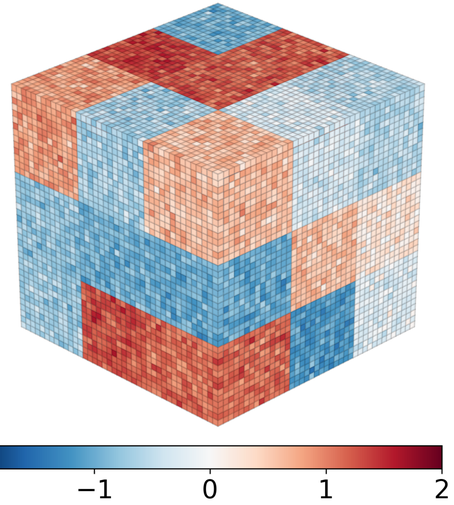
data tend to miss/observe together in blocks

Numerical Experiments

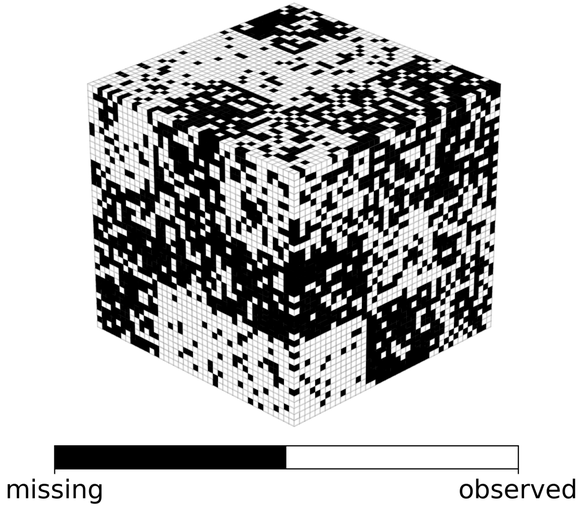
Setup

Simulated by MCMC with block-Gibbs sampler

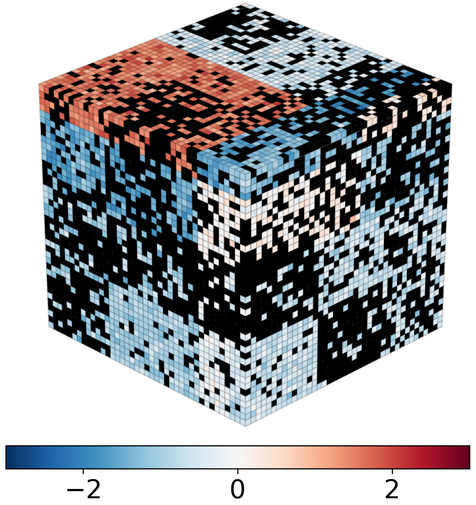
(a) Ground Truth $\mathcal{B}^*$



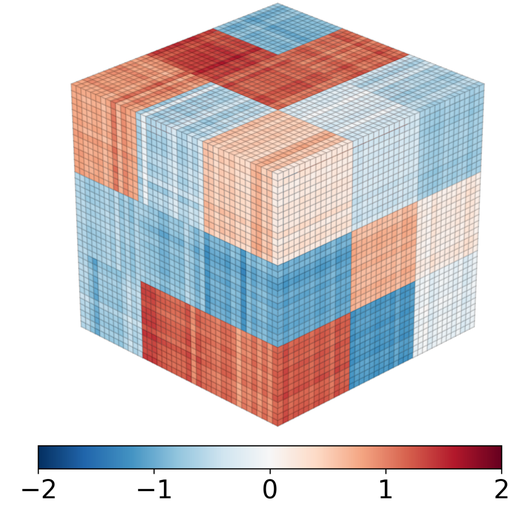
(b) Simulated Missingness $\mathcal{W}$



(c) Simulated Data Tensor $\mathcal{X}$



(d) Fitted $\widehat{\mathcal{B}}$ by RGrad



red: more likely to observe

blue: more likely to miss

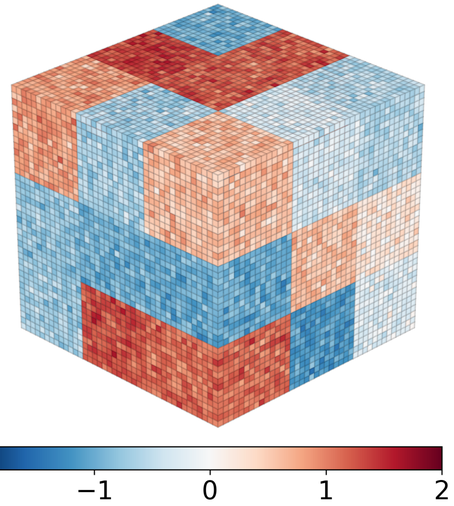
data tend to miss/observe together in blocks

Numerical Experiments

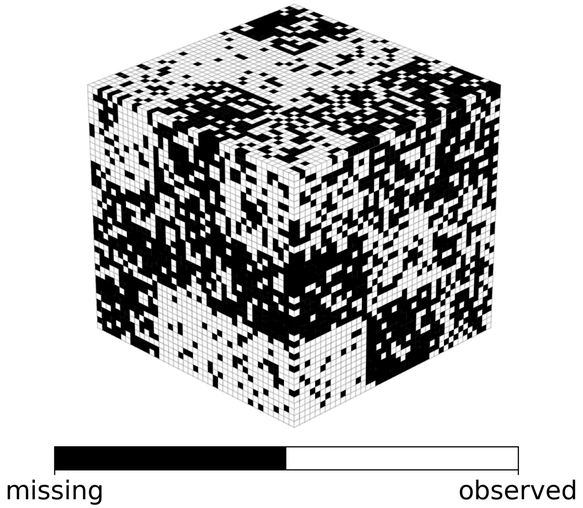
Setup

Simulated by MCMC with block-Gibbs sampler

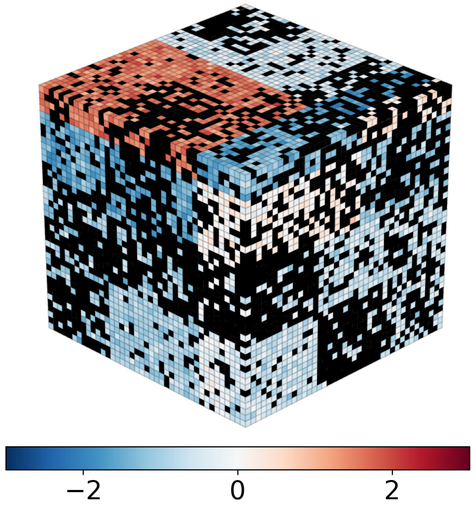
(a) Ground Truth $\mathcal{B}^*$



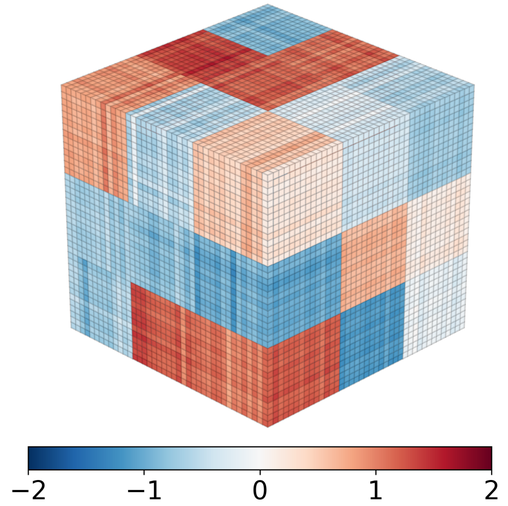
(b) Simulated Missingness $\mathcal{W}$



(c) Simulated Data Tensor $\mathcal{X}$



(d) Fitted $\widehat{\mathcal{B}}$ by RGrad



red: more likely to observe

blue: more likely to miss

data tend to miss/observe together in blocks

$\mathcal{X} = \mathcal{X}^* + \mathcal{E}, \mathcal{X}^*$ **is low rank**

Numerical Experiments

Setup

Numerical Experiments

Setup

- We consider two data missing patterns:

Numerical Experiments

Setup

- We consider two data missing patterns:
  - **Bernoulli**: $g(x, y) = 0, h(x) = x/2$, so $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(\frac{\exp([\mathscr{B}]_s)}{1 + \exp([\mathscr{B}]_s)})$

Numerical Experiments

Setup

- We consider two data missing patterns:
  - **Bernoulli**: $g(x, y) = 0, h(x) = x/2$, so $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(\frac{\exp([\mathscr{B}]_s)}{1 + \exp([\mathscr{B}]_s)})$
  - **Ising**: $g(x, y) = xy/15, h(x) = x/2$

Numerical Experiments

Setup

- We consider two data missing patterns:
  - **Bernoulli**: $g(x, y) = 0, h(x) = x/2$, so $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(\frac{\exp([\mathscr{B}]_s)}{1 + \exp([\mathscr{B}]_s)})$
  - **Ising**: $g(x, y) = xy/15, h(x) = x/2$
- We consider two types of noise distributions:

Numerical Experiments

Setup

- We consider two data missing patterns:
  - **Bernoulli**: $g(x,y)=0, h(x)=x/2$, so $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(\frac{\exp([\mathscr{B}]_s)}{1+\exp([\mathscr{B}]_s)})$
  - **Ising**: $g(x,y)=xy/15, h(x)=x/2$
- We consider two types of noise distributions:
  - **constant**: $[\mathscr{E}]_s \overset{i.i.d.}{\sim} N(0,\sigma^2)$

Numerical Experiments

Setup

- We consider two data missing patterns:
  - **Bernoulli**: $g(x,y)=0, h(x)=x/2$, so $[\mathscr{W}]_s \overset{ind.}{\sim} \text{Bern}(\frac{\exp([\mathscr{B}]_s)}{1+\exp([\mathscr{B}]_s)})$
  - **Ising**: $g(x,y)=xy/15, h(x)=x/2$
- We consider two types of noise distributions:
  - **constant**: $[\mathscr{E}]_s \overset{i.i.d.}{\sim} N(0,\sigma^2)$
  - **adversarial**: $[\mathscr{E}]_s \overset{ind.}{\sim} N(0,\sigma_s^2), \quad \sigma_s \propto [\mathscr{B}]_s^{-1}$ (higher uncertainty for missing entries)

Numerical Experiments

Evaluation Metrics & Benchmarks

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;
  - **Oracle:** assume that the missing propensity is known;

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;
  - **Oracle:** assume that the missing propensity is known;
  - **RGrad:** learn the missing propensity via tensor Ising model.

Numerical Experiments

Evaluation Metrics & Benchmarks

- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;
  - **Oracle:** assume that the missing propensity is known;
  - **RGrad:** learn the missing propensity via tensor Ising model.
- We evaluate each method by the average mis-coverage % in $\mathbb{Q}$:

Numerical Experiments

Evaluation Metrics & Benchmarks

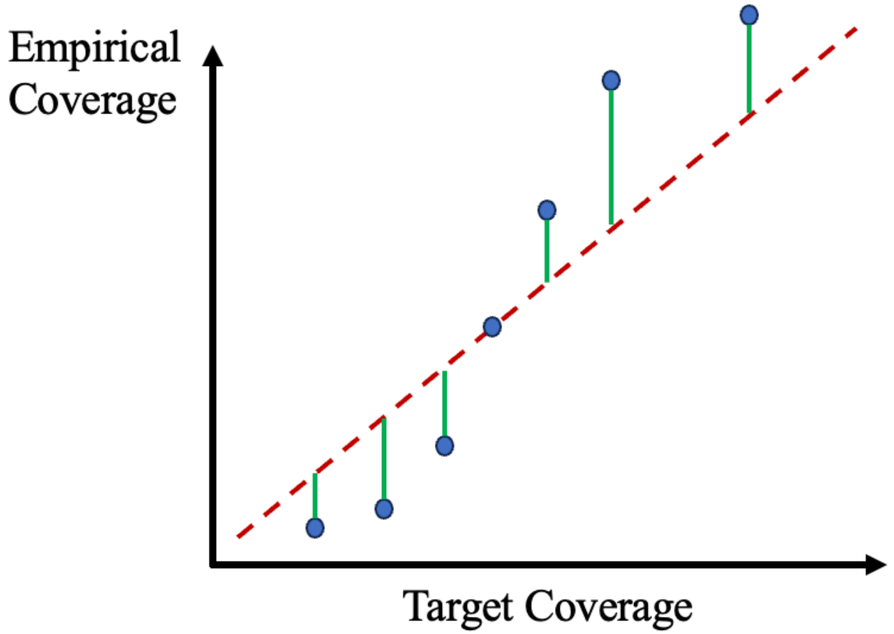
- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;
  - **Oracle:** assume that the missing propensity is known;
  - **RGrad:** learn the missing propensity via tensor Ising model.
- We evaluate each method by the average mis-coverage % in $\mathbb{Q}$:

$$\frac{1}{|\mathbb{Q}|}\sum_{\tau\in\mathbb{Q}}\left|\underbrace{\tau}_{\text{target coverage}} - \underbrace{\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c}\mathbb{I}_{\{[\mathcal{X}]_s\in\widehat{C}_{\tau,s}\}}}_{\text{empirical coverage at }\tau}\right|$$

Numerical Experiments

Evaluation Metrics & Benchmarks

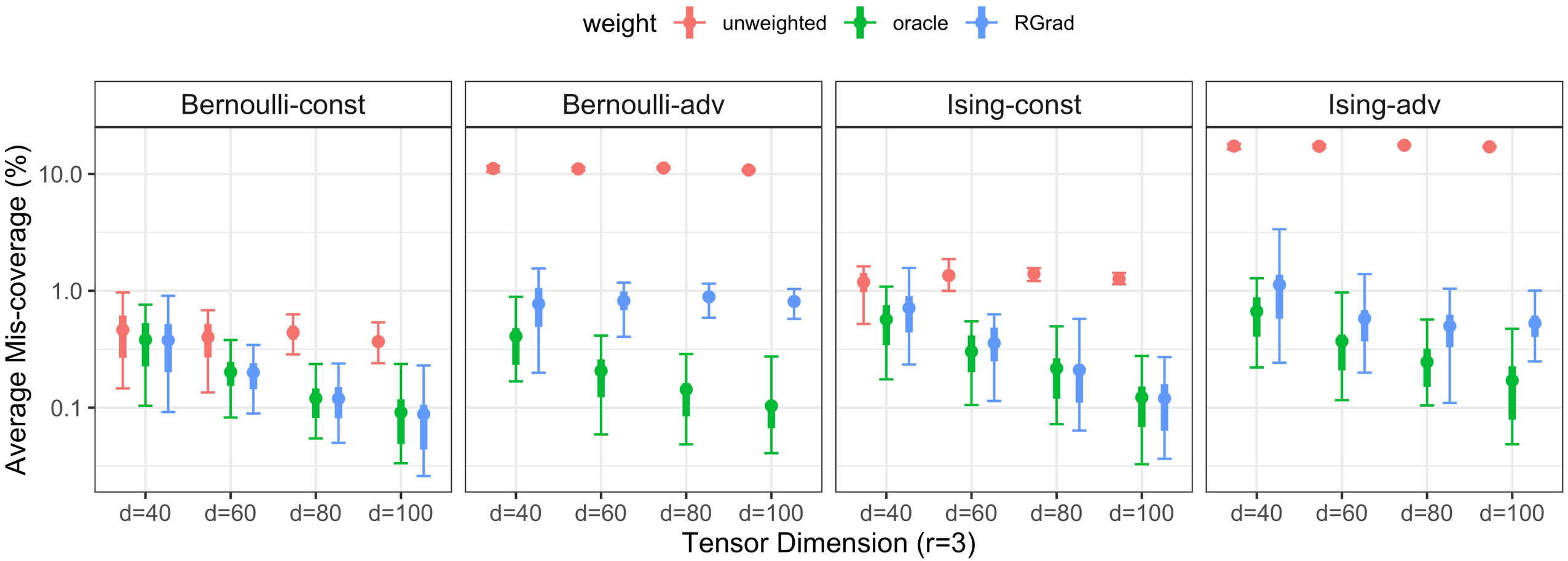
- We choose 20 coverage levels: $\mathbb{Q} = \{0.80, 0.81, \ldots, 0.99\}$
- We consider three versions of conformal prediction:
  - **Unweighted:** apply equal weight to all calibration entries;
  - **Oracle:** assume that the missing propensity is known;
  - **RGrad:** learn the missing propensity via tensor Ising model.
- We evaluate each method by the average mis-coverage % in $\mathbb{Q}$:



$$\frac{1}{|\mathbb{Q}|}\sum_{\tau\in\mathbb{Q}}\left|\underbrace{\tau}_{\text{target coverage}} - \underbrace{\frac{1}{|\boldsymbol{\Omega}^c|}\sum_{s\in\boldsymbol{\Omega}^c}\mathbb{I}_{\{[\mathcal{X}]_s\in\widehat{C}_{\tau,s}\}}}_{\text{empirical coverage at }\tau}\right|$$

Numerical Experiments

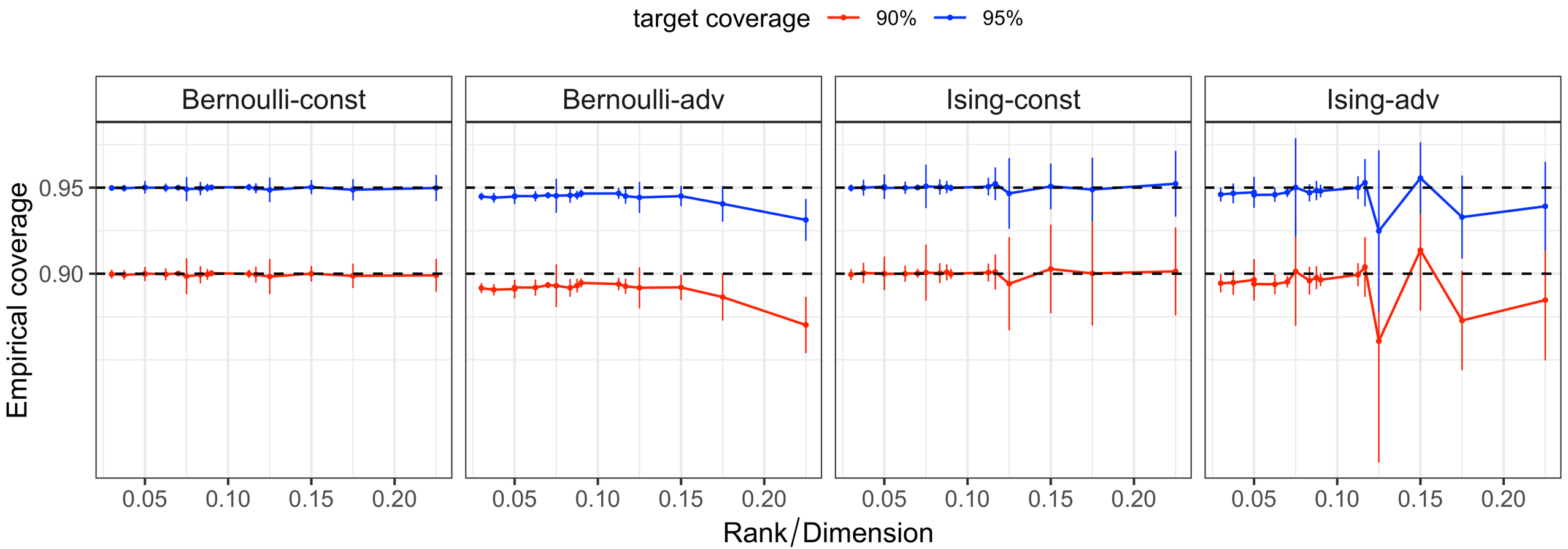
Conformal Inference Validation (rank($\mathcal{B}^*$) = (3,3))



const: $[\mathcal{E}]_s \overset{i.i.d.}{\sim} N(0,\sigma^2)$; adv: $[\mathcal{E}]_s \overset{ind.}{\sim} N(0,\sigma_s^2)$, $\sigma_s \propto [\mathcal{B}]_s^{-1}$ (higher uncertainty for missing entries)

Numerical Experiments

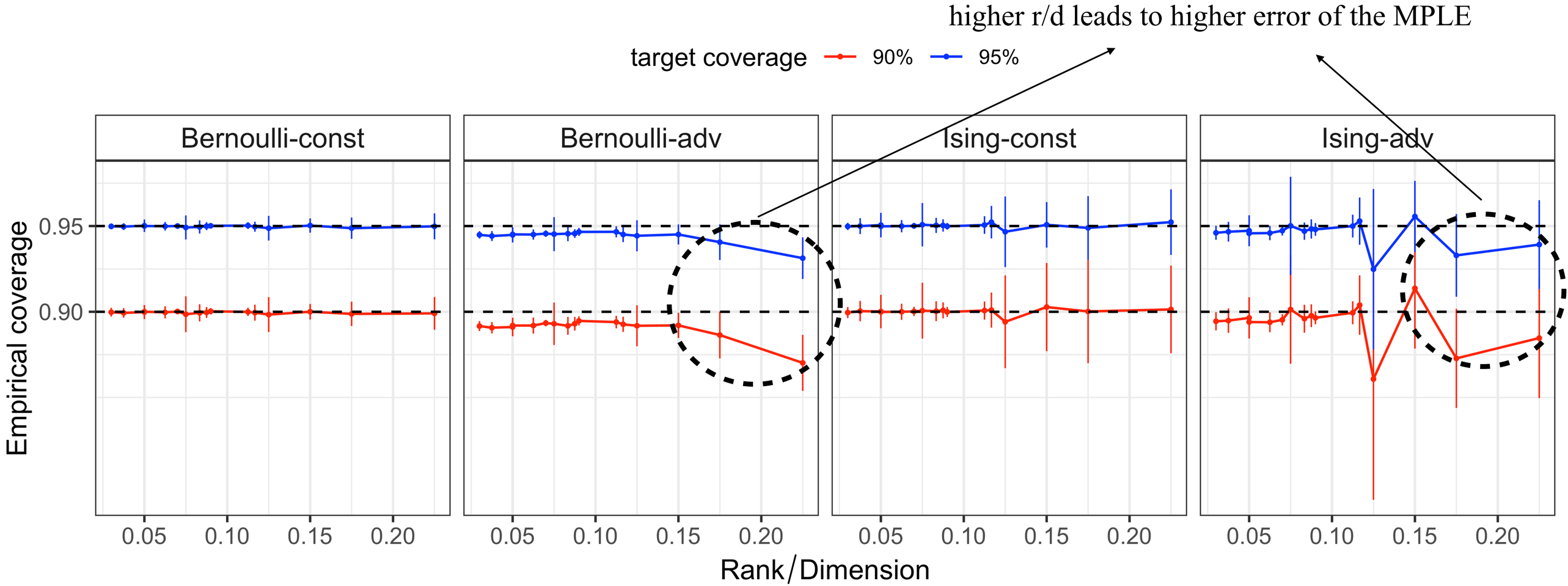
Conformal Inference Validation (90%, 95% CI)



x-axis: r/d of the tensor parameter $\mathcal{B}^*$

Numerical Experiments

Conformal Inference Validation (90%, 95% CI)

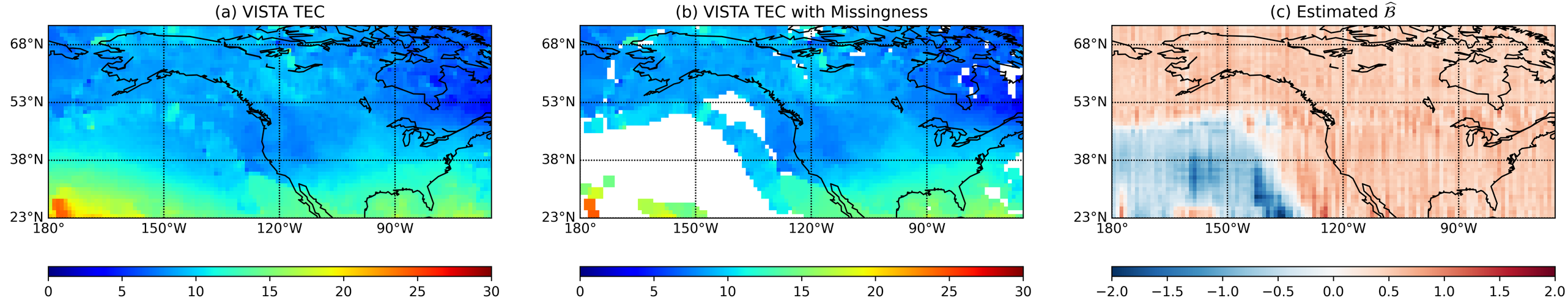


x-axis: r/d of the tensor parameter $\mathscr{B}^*$

Data Application

Regional TEC Reconstruction

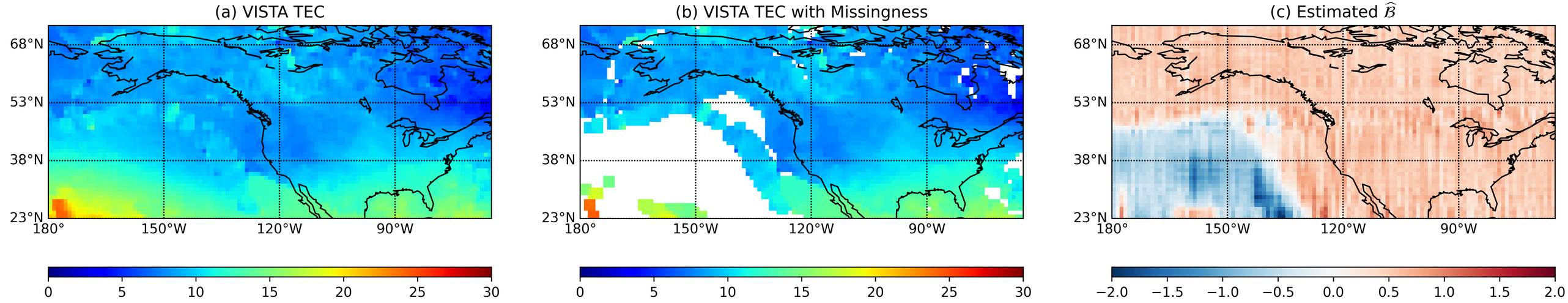
- We apply our method to reconstruct the total electron content (TEC) over the US and Canada for 15 days in Sept, 2017.



Data Application

Regional TEC Reconstruction

- We apply our method to reconstruct the total electron content (TEC) over the US and Canada for 15 days in Sept, 2017.

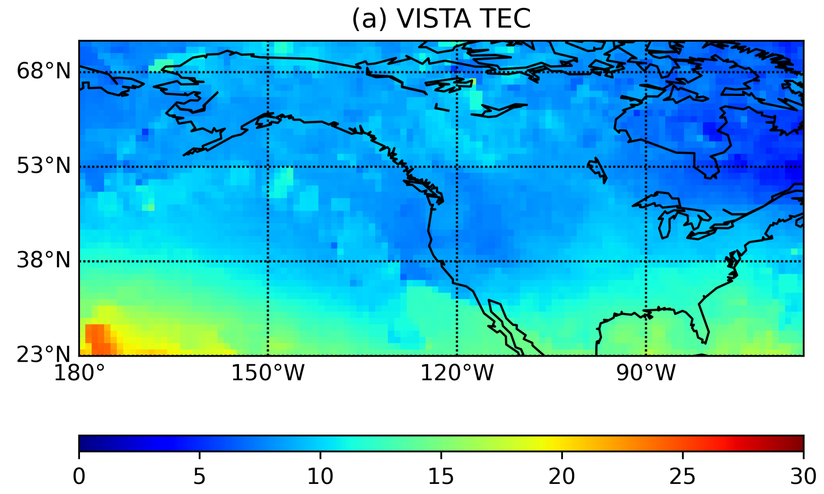


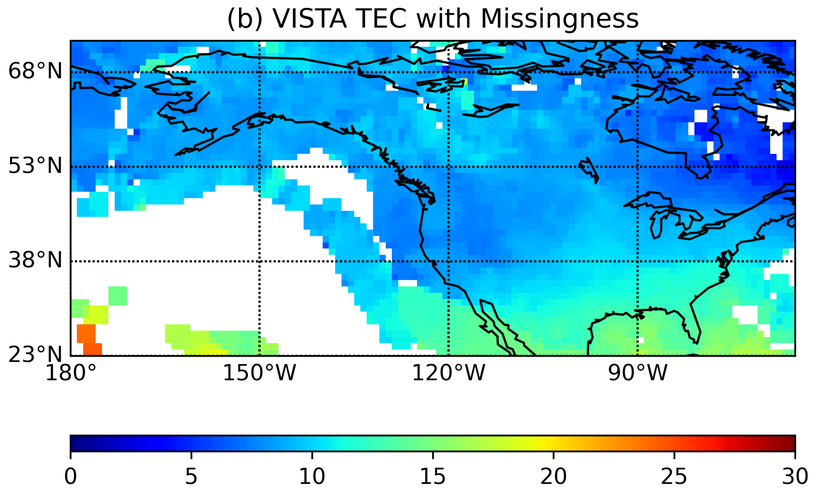
data without missingness

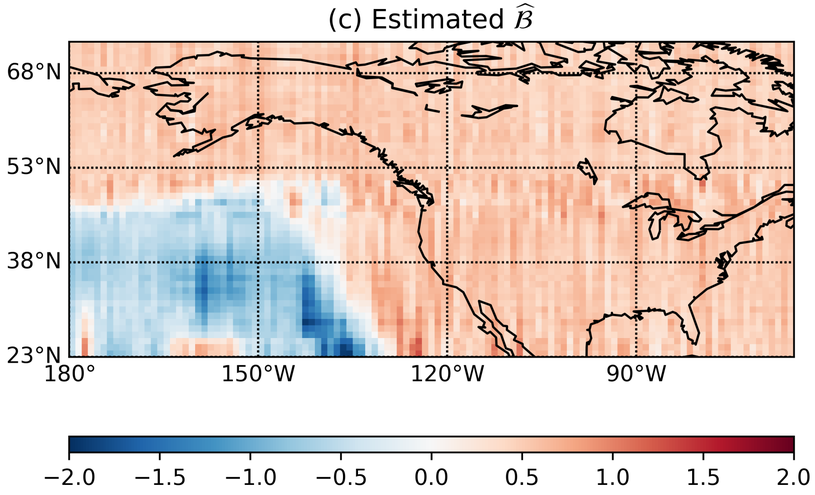
Data Application

Regional TEC Reconstruction

- We apply our method to reconstruct the total electron content (TEC) over the US and Canada for 15 days in Sept, 2017.







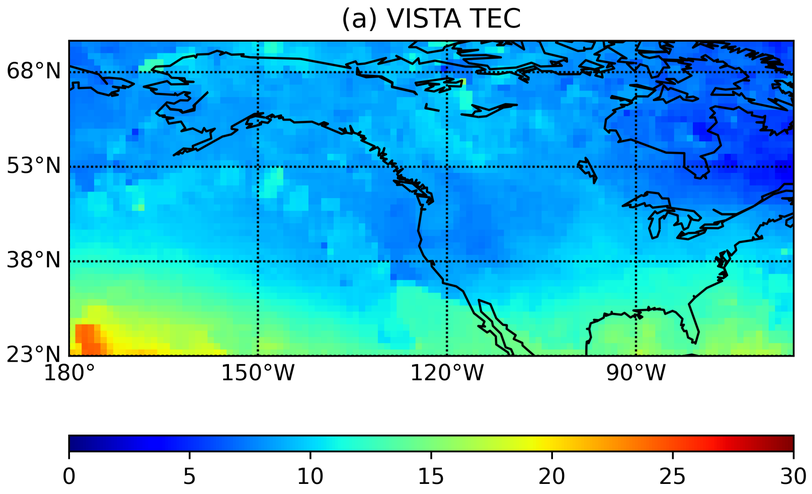
data without missingness

data with missingness

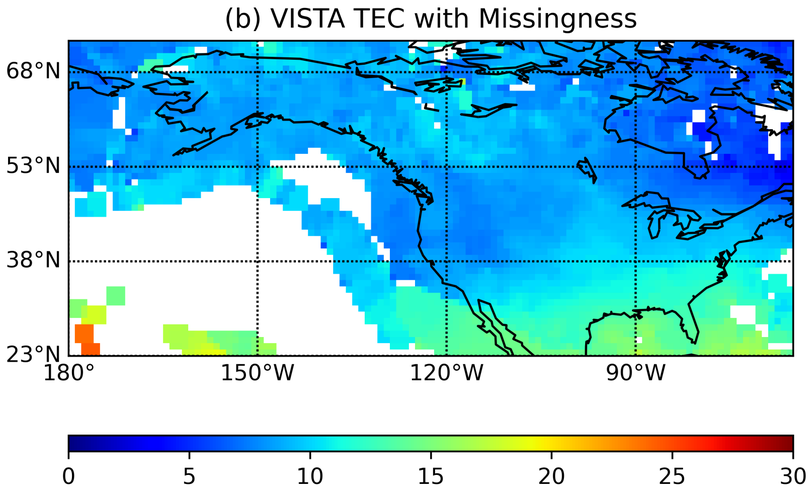
Data Application

Regional TEC Reconstruction

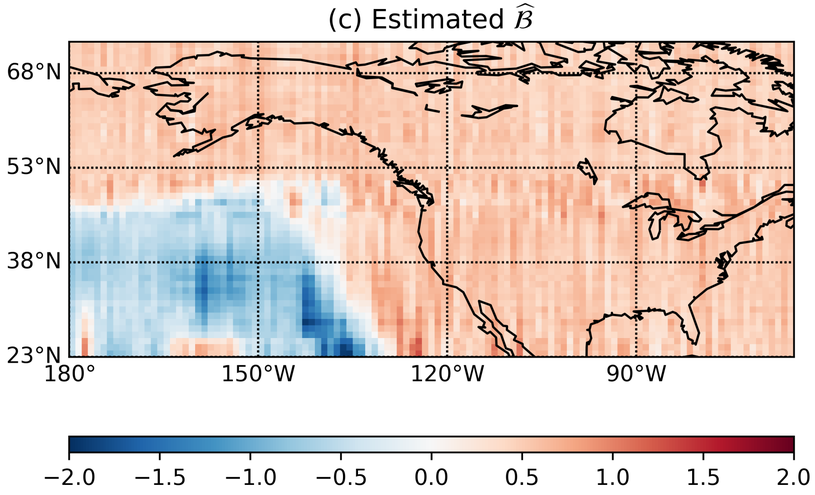
- We apply our method to reconstruct the total electron content (TEC) over the US and Canada for 15 days in Sept, 2017.



data without missingness



data with missingness



fitted $\widehat{\mathcal{B}}$

blue: more likely to miss

Data Application

Coverage Performance

- We **compare** our method under both Ising model and Bernoulli model with the unweighted conformal prediction:

Data Application

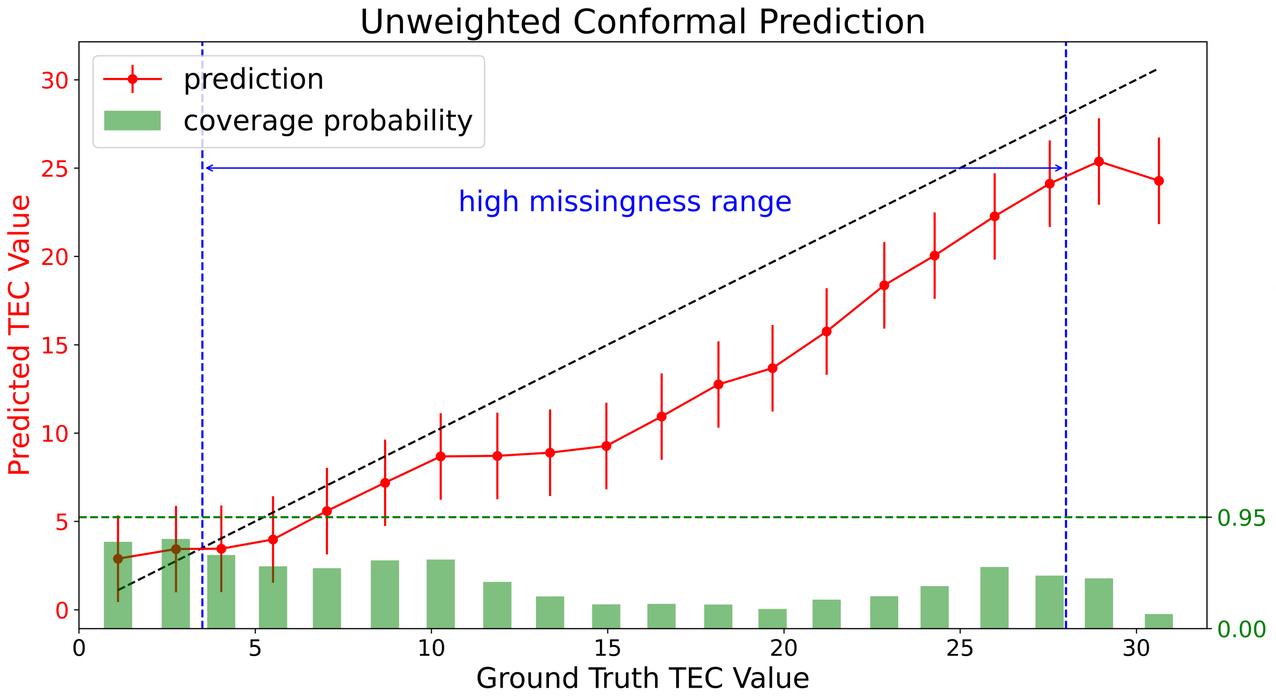
Coverage Performance

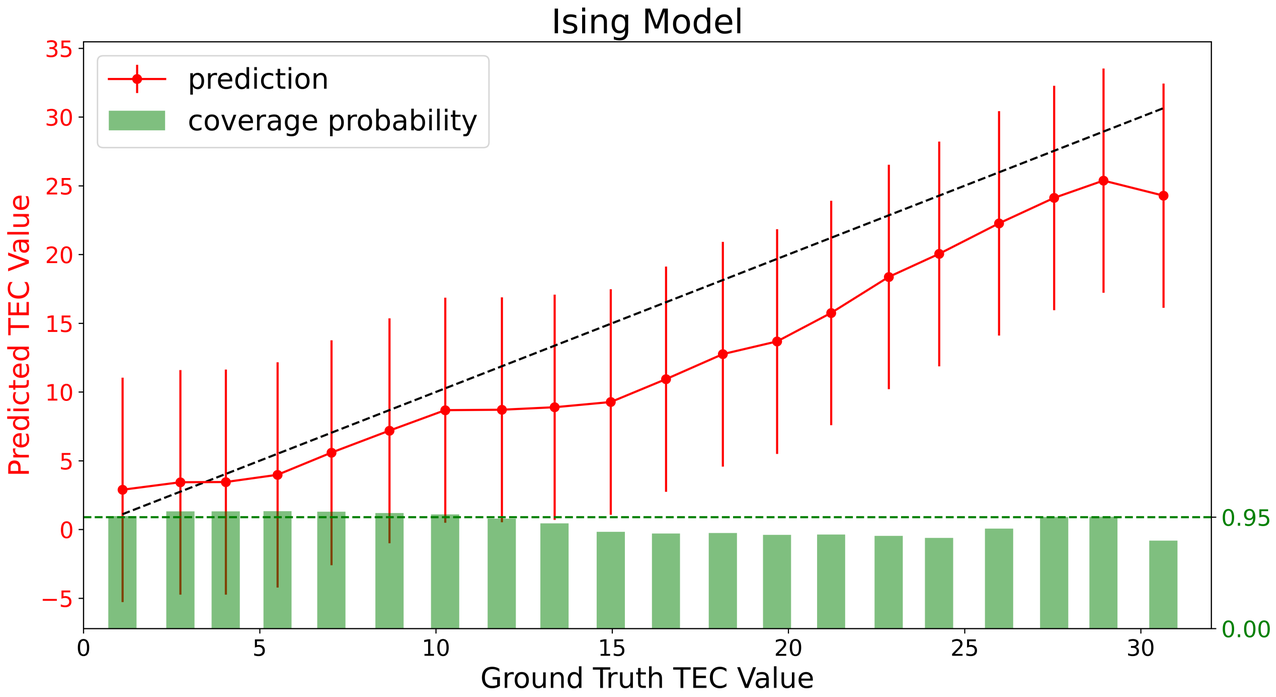
- We **compare** our method under both Ising model and Bernoulli model with the unweighted conformal prediction:

method	mis-coverage %	90% CI coverage %	95% CI coverage %
unweighted	42.1(6.49)	46.3(6.58)	52.3(7.23)
Bernoulli	23.1(5.34)	64.6(5.97)	76.8(5.03)
Ising	6.01(2.45)	90.0(6.06)	94.2(3.74)

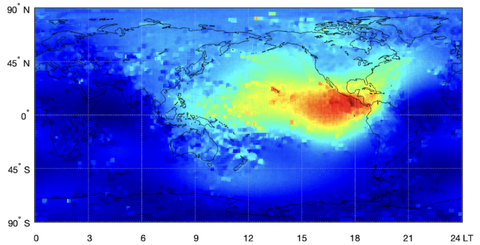
Data Application

A Single Day Example (Sept 6, 2017)

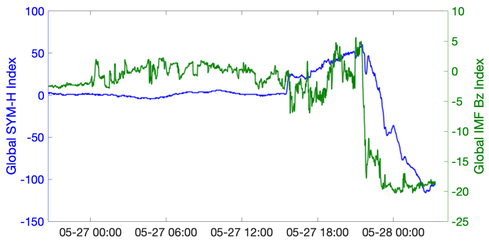




III. Matrix Autoregression with Auxiliary Data



+

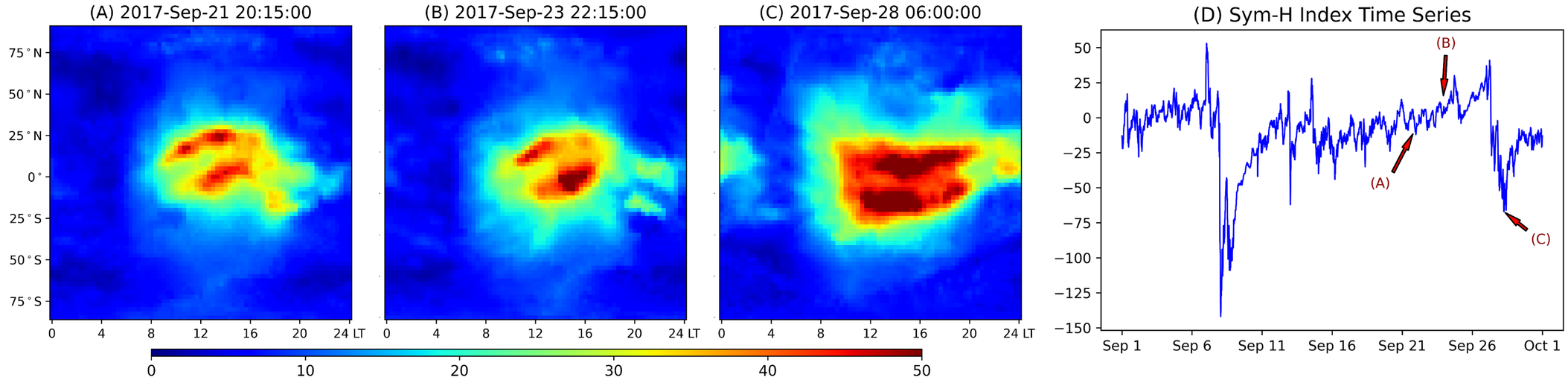


Spatial Matrix Time Series **Non-spatial Vector Time Series**

How to forecast future matrix data?

Motivating Example

Global TEC Forecast with Solar Wind Parameters



- Left 3 panels: Global TEC, measurement of the Earth's ionospheric activity
- Right panel: Sym-H index, measurement of the impact of solar eruption on Earth

Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

- Denote $\{\mathbf{X}_t\}_{t=1}^T$ as the matrix time series, and $\{\mathbf{z}_t\}_{t=1}^T$ as the vector time series, then our MARAC(P,Q) model is:

Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

- Denote $\{\mathbf{X}_t\}_{t=1}^T$ as the matrix time series, and $\{\mathbf{z}_t\}_{t=1}^T$ as the vector time series, then our MARAC(P,Q) model is:

$$\mathbf{X}_t = \underbrace{\sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top}_{\text{autoregressive part}} + \underbrace{\sum_{q=1}^{Q} \mathcal{G}_q \bar{\times} \mathbf{z}_{t-q}}_{\text{auxiliary part}} + \mathbf{E}_t, \quad \text{vec}\left(\mathbf{E}_t\right) \sim \mathbf{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\text{c}} \otimes \mathbf{\Sigma}_{\text{r}}\right)$$

Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

- Denote $\{\mathbf{X}_t\}_{t=1}^T$ as the matrix time series, and $\{\mathbf{z}_t\}_{t=1}^T$ as the vector time series, then our MARAC(P,Q) model is:

$$\mathbf{X}_t = \underbrace{\sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top}_{\text{autoregressive part}} + \underbrace{\sum_{q=1}^{Q} \mathcal{G}_q \bar{\times} \mathbf{z}_{t-q}}_{\text{auxiliary part}} + \mathbf{E}_t, \quad \text{vec}\left(\mathbf{E}_t\right) \sim \mathbf{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\text{c}} \otimes \mathbf{\Sigma}_{\text{r}}\right)$$

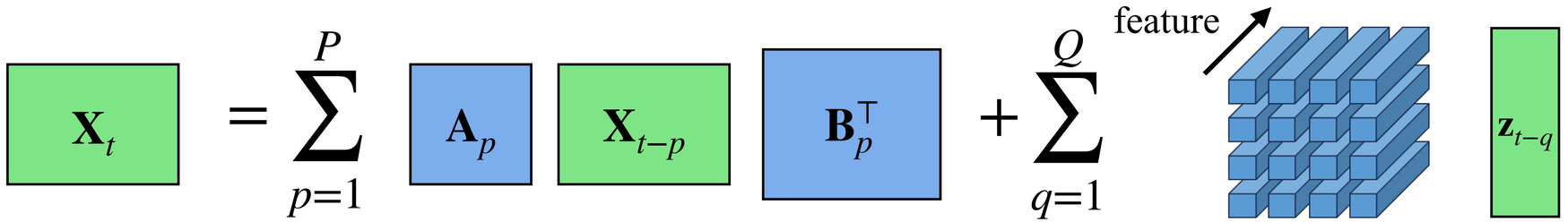
$$\mathbf{X}_t = \sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top$$

Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

- Denote $\{\mathbf{X}_t\}_{t=1}^T$ as the matrix time series, and $\{\mathbf{z}_t\}_{t=1}^T$ as the vector time series, then our MARAC(P,Q) model is:

$$\mathbf{X}_t = \underbrace{\sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top}_{\text{autoregressive part}} + \underbrace{\sum_{q=1}^{Q} \mathcal{G}_q \bar{\times} \mathbf{z}_{t-q}}_{\text{auxiliary part}} + \mathbf{E}_t, \quad \text{vec}\left(\mathbf{E}_t\right) \sim \mathbf{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\text{c}} \otimes \mathbf{\Sigma}_{\text{r}}\right)$$

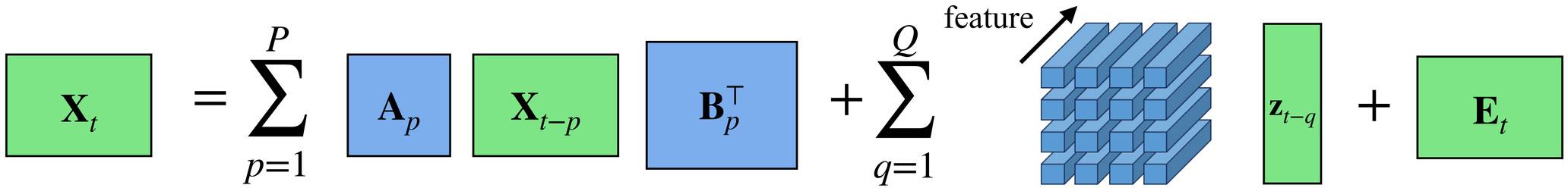


Our Model

Matrix AutoRegression with Auxiliary Covariates (MARAC)

- Denote $\{\mathbf{X}_t\}_{t=1}^T$ as the matrix time series, and $\{\mathbf{z}_t\}_{t=1}^T$ as the vector time series, then our MARAC(P,Q) model is:

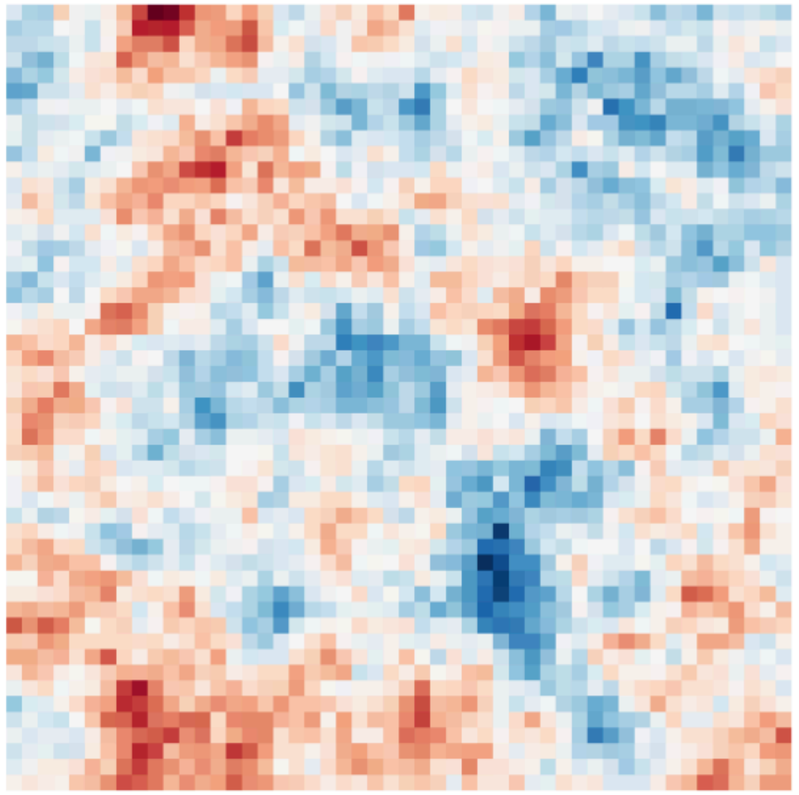
$$\mathbf{X}_t = \underbrace{\sum_{p=1}^{P} \mathbf{A}_p \mathbf{X}_{t-p} \mathbf{B}_p^\top}_{\text{autoregressive part}} + \underbrace{\sum_{q=1}^{Q} \mathcal{G}_q \bar{\times} \mathbf{z}_{t-q}}_{\text{auxiliary part}} + \mathbf{E}_t, \quad \text{vec}\left(\mathbf{E}_t\right) \sim \mathbf{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\text{c}} \otimes \mathbf{\Sigma}_{\text{r}}\right)$$



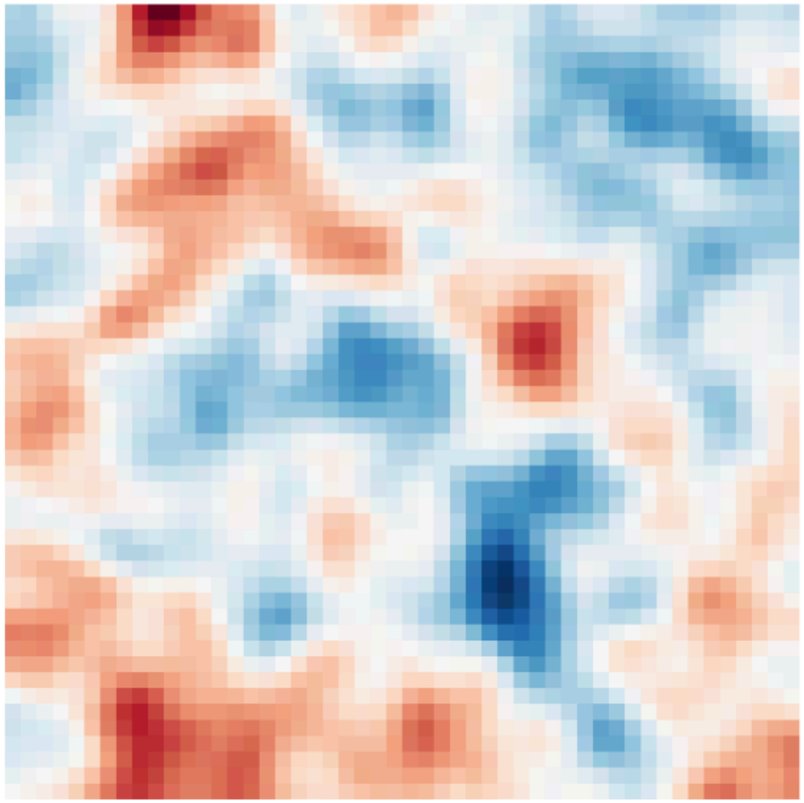
Model

Spatially-Smooth Regression Coefficient

Slice of Tensor Coefficient $\mathcal{G}_q(:,:,d)$

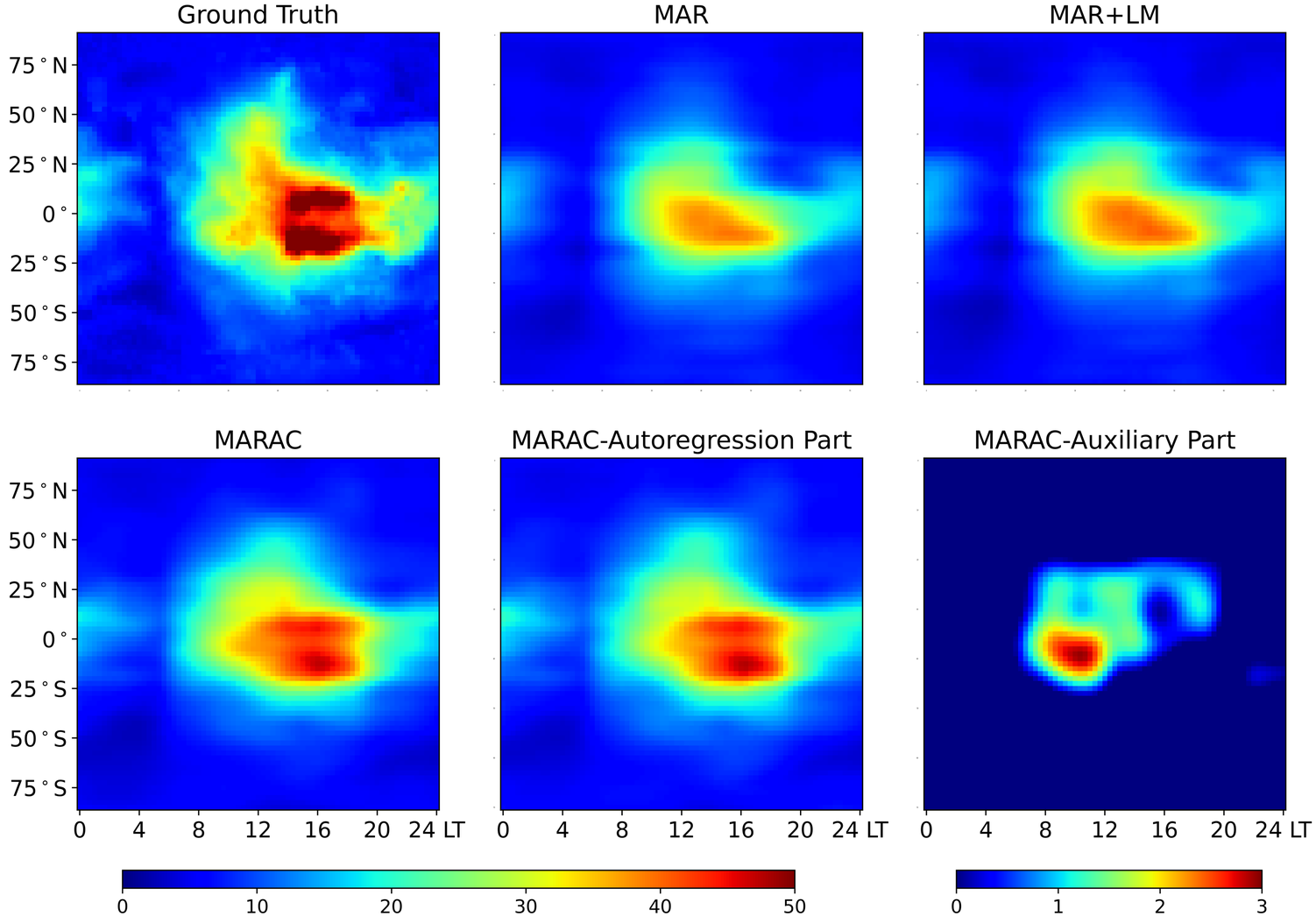


True Functional Parameter $g_{q,d}(\cdot)$



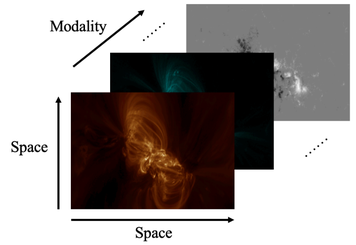
Real Data Application

Global Total Electron Content (TEC) Prediction



IV. Scalar-on-Tensor Gaussian Process with Contraction

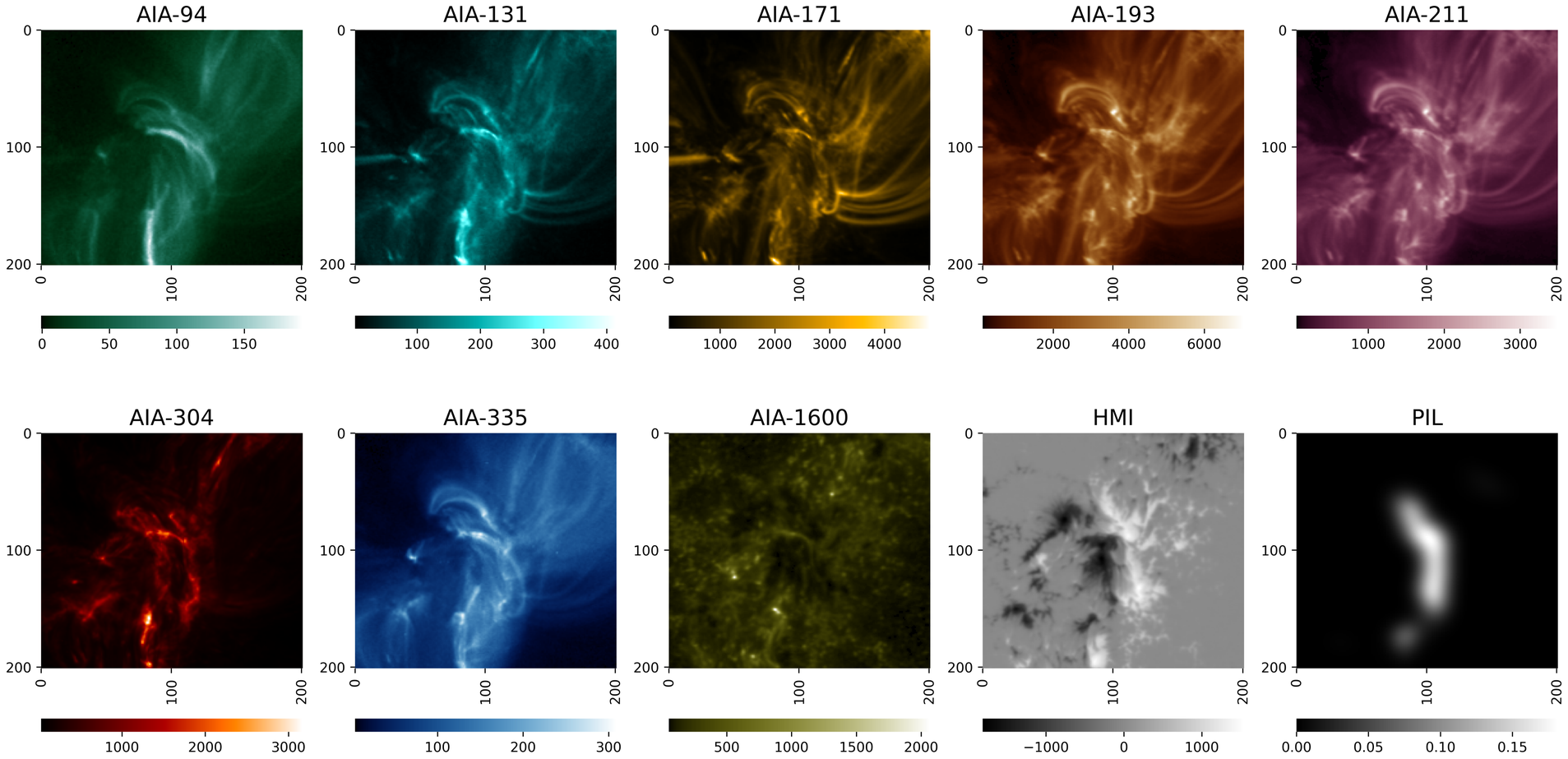
$$y = f\Big(\quad \Big) + \epsilon$$



scalar label **multi-channel imaging covariate**

Motivation

Covariate: multi-modal solar imaging data



Proposed Method

Tensor Gaussian Process with Contraction

Proposed Method

Tensor Gaussian Process with Contraction

Let $y \in \mathbb{R}$ be the scalar label and $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ be the tensor covariate, then we have:

Proposed Method

Tensor Gaussian Process with Contraction

Let $y \in \mathbb{R}$ be the scalar label and $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ be the tensor covariate, then we have:

$$y = g \circ h(\mathcal{X}) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

Proposed Method

Tensor Gaussian Process with Contraction

Let $y \in \mathbb{R}$ be the scalar label and $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ be the tensor covariate, then we have:

$$y = g \circ h(\mathcal{X}) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

$$\mathcal{Z}_{h \times w \times C} = h(\mathcal{X}) = \mathcal{X}_{H \times W \times C} \times_1 \mathbf{A}_{h \times H} \times_2 \mathbf{B}_{w \times W} \text{ (dimension reduction)}$$

Proposed Method

Tensor Gaussian Process with Contraction

Let $y \in \mathbb{R}$ be the scalar label and $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ be the tensor covariate, then we have:

$$y = g \circ h(\mathcal{X}) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

$$\mathcal{Z}_{h \times w \times C} = h(\mathcal{X}) = \mathcal{X}_{H \times W \times C} \times_1 \mathbf{A}_{h \times H} \times_2 \mathbf{B}_{w \times W} \quad \text{(dimension reduction)}$$

$$g(\cdot) \sim \mathbf{GP}\left(0, k(\cdot, \cdot)\right) \quad \text{(tensor Gaussian process)}$$

Proposed Method

Tensor Gaussian Process with Contraction

Let $y \in \mathbb{R}$ be the scalar label and $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$ be the tensor covariate, then we have:

$$y = g \circ h(\mathcal{X}) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

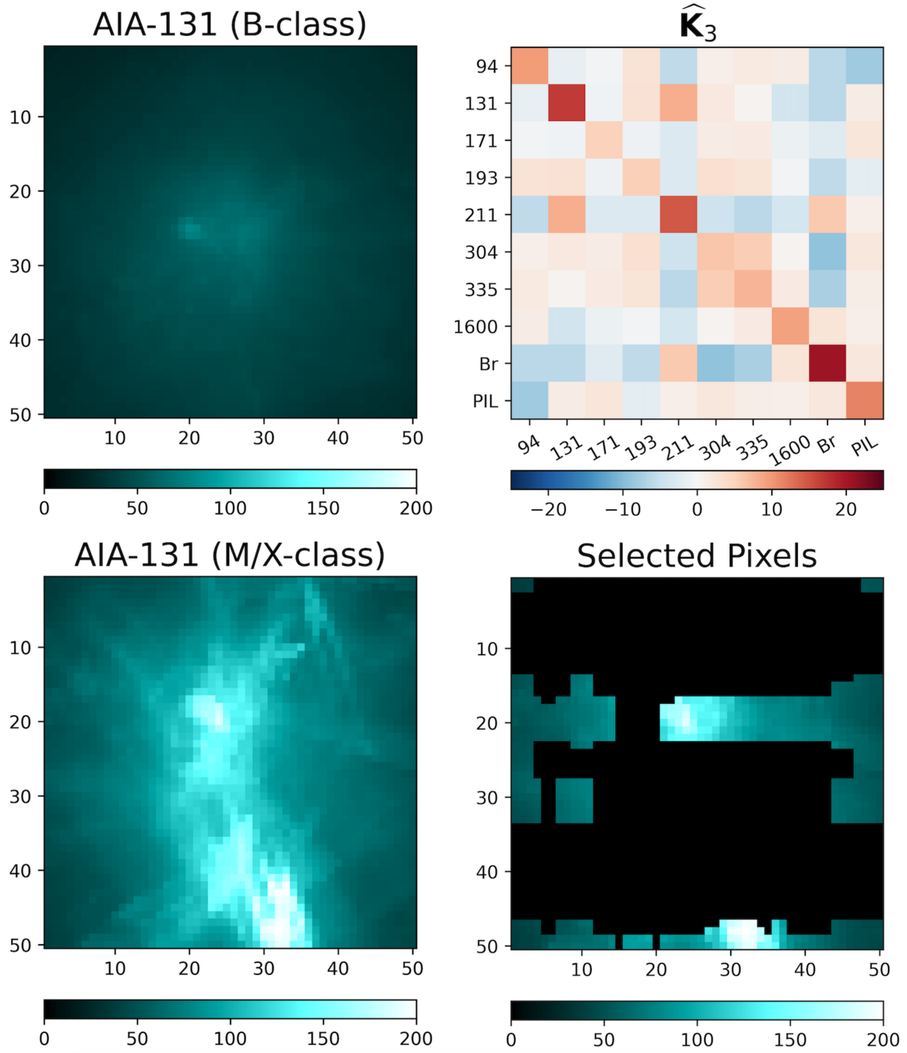
$$\mathcal{Z}_{h \times w \times C} = h(\mathcal{X}) = \mathcal{X}_{H \times W \times C} \times_1 \mathbf{A}_{h \times H} \times_2 \mathbf{B}_{w \times W} \quad \text{(dimension reduction)}$$

$$g(\cdot) \sim \mathbf{GP}\left(0, k(\cdot, \cdot)\right) \qquad \text{(tensor Gaussian process)}$$

$$k(\mathcal{Z}_1, \mathcal{Z}_2) = \text{vec}\left(\mathcal{Z}_1\right)^\top \left[\bigotimes_{j=1}^{3} \mathbf{K}_j\right] \text{vec}\left(\mathcal{Z}_2\right)$$

Real Data Application

Solar Flare Intensity Forecast



Modality pairwise interaction

Feature Selection

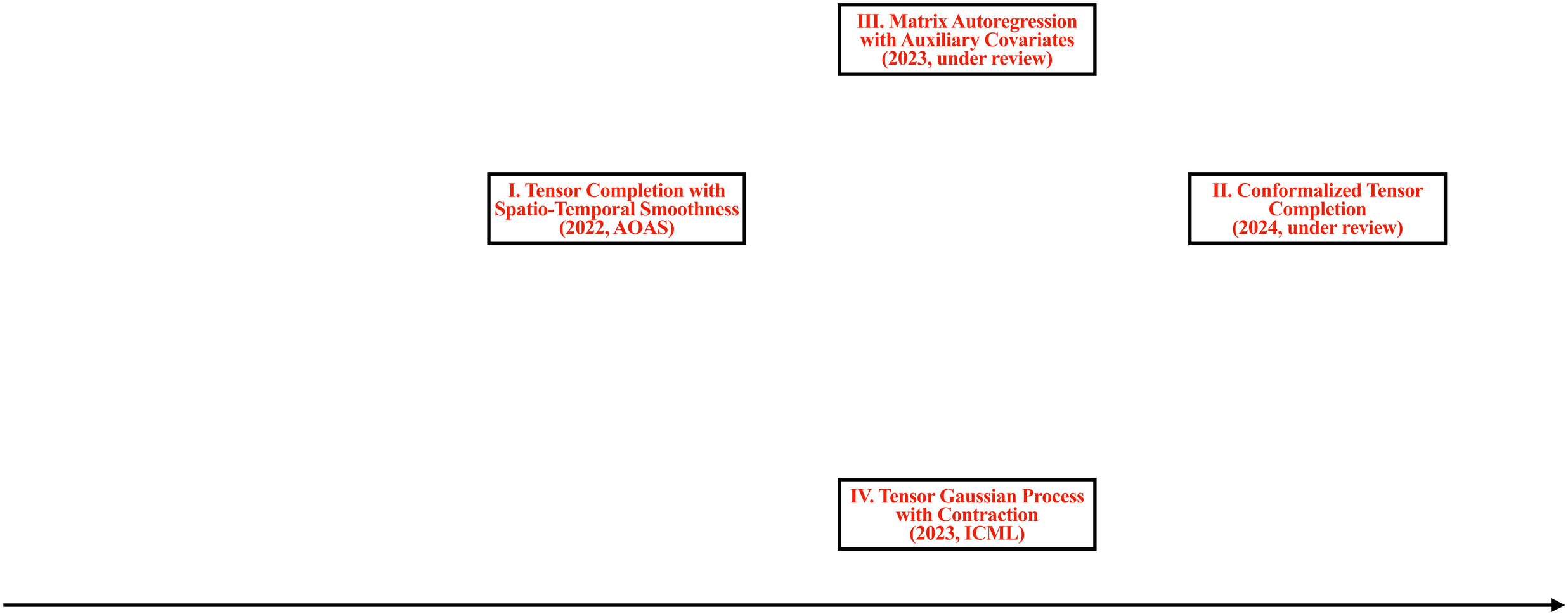
Final Remarks

Summary

Roadmaps of all works

Summary

Roadmaps of all works



Summary

Roadmaps of all works

III. Matrix Autoregression with Auxiliary Covariates (2023, under review)

I. Tensor Completion with Spatio-Temporal Smoothness (2022, AOAS)

II. Conformalized Tensor Completion (2024, under review)

Solar Flare Forecast with Deep Learning (2020, Space Weather)

Solar Flare Forecast with Spatial Statistics (2021, Space Weather)

IV. Tensor Gaussian Process with Contraction (2023, ICML)

Time

Summary

Roadmaps of all works

III. Matrix Autoregression with Auxiliary Covariates (2023, under review)

I. Tensor Completion with Spatio-Temporal Smoothness (2022, AOAS) → Open-Source Global TEC Database (2023, Scientific Data)

II. Conformalized Tensor Completion (2024, under review)

Solar Flare Forecast with Deep Learning (2020, Space Weather)

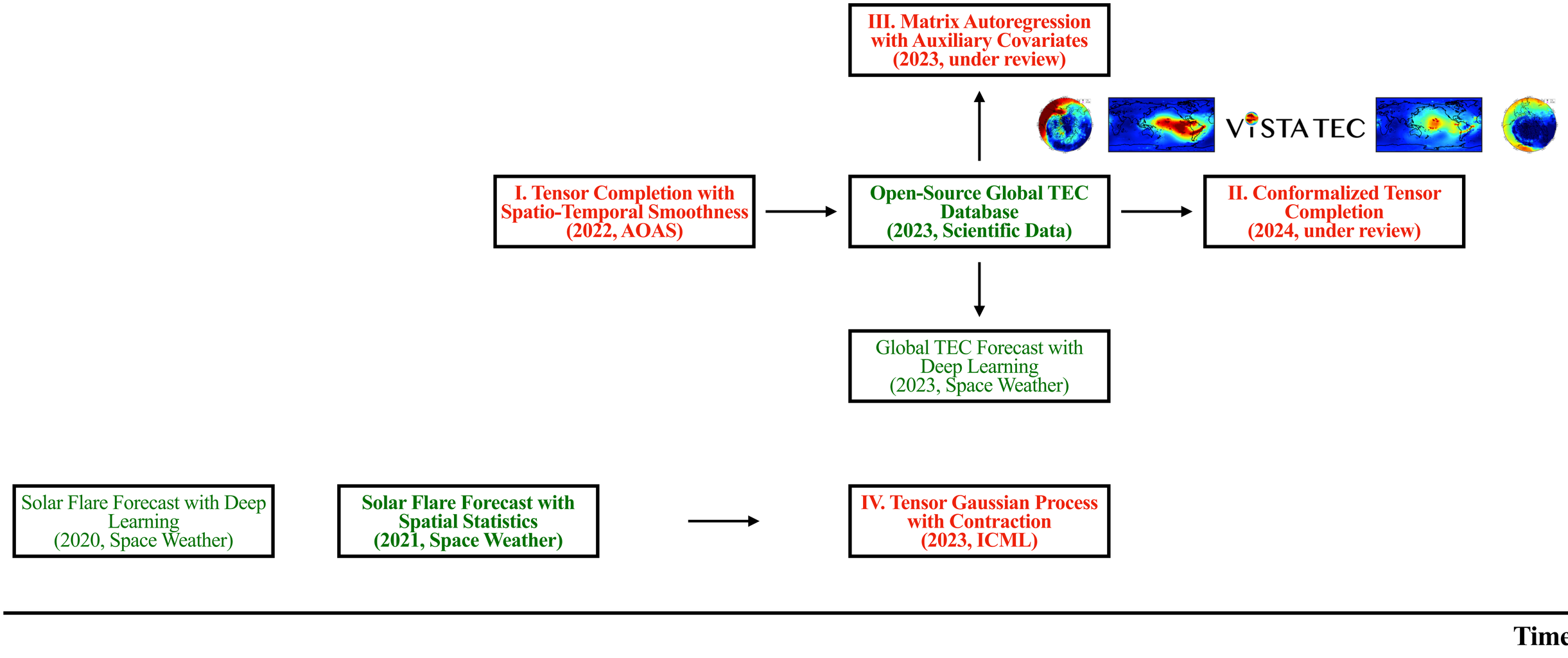
Solar Flare Forecast with Spatial Statistics (2021, Space Weather) →

IV. Tensor Gaussian Process with Contraction (2023, ICML)

Time

Summary

Roadmaps of all works



Future Directions

Future Directions

Theory

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

Methodology

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

Methodology

- Multi-modal tensor inference

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

Methodology

- Multi-modal tensor inference
- Online spatio-temporal tensor learning

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

Methodology

- Multi-modal tensor inference
- Online spatio-temporal tensor learning

Computation

Future Directions

Theory

- Concentration bounds of locally-dependent tensors

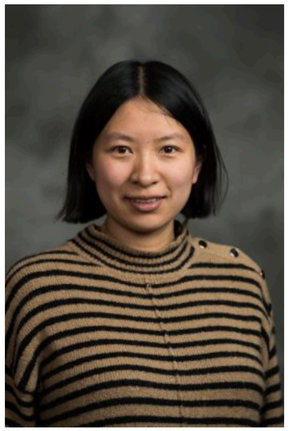
Methodology

- Multi-modal tensor inference
- Online spatio-temporal tensor learning

Computation

- Efficient hyper-parameter tuning

Acknowledgements




- **Statistics**: Prof. Yang Chen, Prof. Zuofeng Shang, Prof. Yuekai Sun, Dr. Yu Wang, Zhenbang Jiao, Yurui Chang
- **EECS**: Prof. Alfred Hero, Zeyu Sun
- **CLASP**: Prof. Ward Manchester, Prof. Shasha Zou, Prof. Tamas Gombosi, Dr. Xiantong Wang, Dr. Zihan Wang, Dr. Jiaen Ren
- **Outside of UM**: Dr. Meng Jin, Dr. Yang Liu, Dr. Anthea Coster, Dr. Monica Bobra

References I

1. Bi, X., Qu, A., & Shen, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Annals of Statistics*, *46*(6B), 3308-3333.
2. Wei, B., Peng, L., Guo, Y., Manatunga, A., & Stevens, J. (2023). Tensor response quantile regression with neuroimaging data. *Biometrics*, *79*(3), 1947-1958.
3. Chen, Y., Fan, J., Ma, C., & Yan, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, *116*(46), 22931-22937.
4. Gui, Y., Barber, R., & Ma, C. (2023). Conformalized matrix completion. *Advances in Neural Information Processing Systems*, *36*, 4820-4844.
5. Cai, C., Poor, H. V., & Chen, Y. (2022). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. *IEEE Transactions on Information Theory*, 69(1), 407-452.
6. Oseledets, I. V. (2011). Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5), 2295-2317.
7. Holtz, S., Rohwedder, T., & Schneider, R. (2012). On manifolds of tensors of fixed TT-rank. *Numerische Mathematik*, 120(4), 701-731.
8. Hastie, T., Mazumder, R., Lee, J. D., & Zadeh, R. (2015). Matrix completion and low-rank SVD via fast alternating least squares. *The Journal of Machine Learning Research*, *16*(1), 3367-3402.

References II

10. Chen, R., Xiao, H., & Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, *222*(1), 539-560.

11. Hsu, N. J., Huang, H. C., & Tsay, R. S. (2021). Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics*, *30*(4), 1143-1155.

12. Li, L., & Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, *112*(519), 1131-1146.

13. Carpentier, A., Eisert, J., Gross, D., & Nickl, R. (2019). Uncertainty quantification for matrix compressed sensing and quantum tomography problems. In *High Dimensional Probability VIII: The Oaxaca Volume* (pp. 385-430). Cham: Springer International Publishing.

14. Carpentier, A., Klopp, O., Löffler, M., & Nickl, R. (2018). Adaptive confidence sets for matrix completion. *Bernoulli*, *24*(4A), 2429-2460.

15. Farias, V., Li, A. A., & Peng, T. (2022, May). Uncertainty quantification for low-rank matrix completion with heterogeneous and sub-exponential noise. In *International Conference on Artificial Intelligence and Statistics* (pp. 1179-1189). PMLR.

16. Xia, D., & Yuan, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(1), 58-77.

References III

17. Mai, T. T., & Alquier, P. (2015). A Bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. *Electronic Journal of Statistics*, 9, 823-841.

18. Yuchi, H. S., Mak, S., & Xie, Y. (2023). Bayesian uncertainty quantification for low-rank matrix completion. *Bayesian Analysis*, 18(2), 491-518.

19. Kasalicky, P., Ledent, A., & Alves, R. (2023, September). Uncertainty-adjusted Inductive Matrix Completion with Graph Neural Networks. *In Proceedings of the 17th ACM Conference on Recommender Systems* (pp. 1169-1174).

20. Zeldes, Y., Theodorakis, S., Solodnik, E., Rotman, A., Chamiel, G., & Friedman, D. (2017). Deep density networks and uncertainty in recommender systems. *arXiv preprint* arXiv:1711.02487.

21. Shao, M., & Zhang, Y. (2023). Distribution-free matrix prediction under arbitrary missing pattern. *arXiv preprint* arXiv:2305.11640.

22. Tibshirani, R. J., Foygel Barber, R., Candes, E., & Ramdas, A. (2019). Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32.

23. Cipra, B. A. (1987). An introduction to the Ising model. *The American Mathematical Monthly*, 94(10), 937-959.

Reference IV

24. Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional Ising model selection using ℓ1-regularized logistic regression. *Annals of Statistics*, 38(1), 1287-1319.

25. Liu, T., Mukherjee, S., & Biswas, R. (2024). Tensor recovery in high-dimensional Ising models. *Journal of Multivariate Analysis*, 203, 105335.

26. Barber, R. F., & Drton, M. (2015). High-dimensional Ising model selection with Bayesian information criteria. *Electronic Journal of Statistics*, *9*, 567-607.

27. Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *The Annals of Statistics*, *51*(2), 816-845.