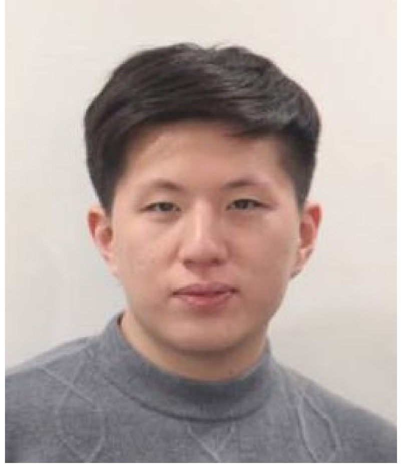
东南大學
SOUTHEAST UNIVERSITY

ICML
International Conference
On Machine Learning

A Difference Standardization Method for Mutual Transfer Learning



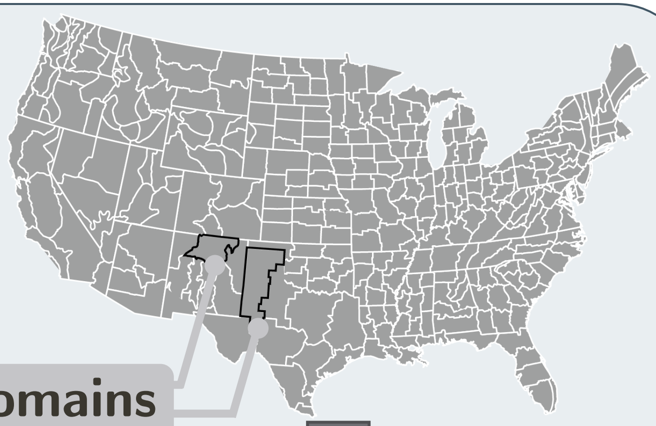
Haoqing Xu Meng Wang Beilun Wang*

Southeast University

Motivation: Learning from Multiple Data Domains in Big Data Applications

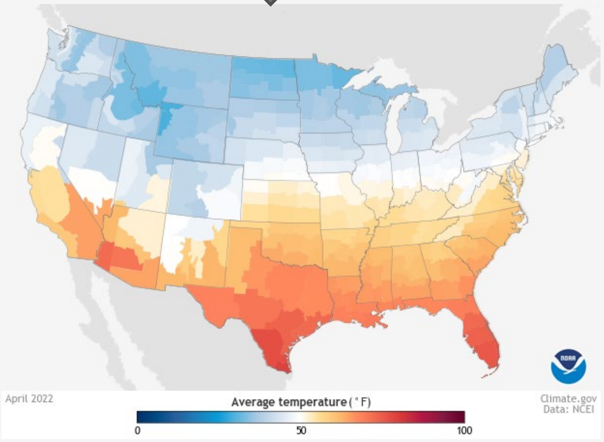
- Learning from **multiple data domains** is common:
  - Average Temperature Prediction
    - **Domain:** Climate Division
  - Gene Expression
    - **Domain:** Gene
  - ...
- Leverage abundant samples from **related domains**
  - **Improve** model generalizability
- #Domains M grows quickly

NOAA nClimDiv Database



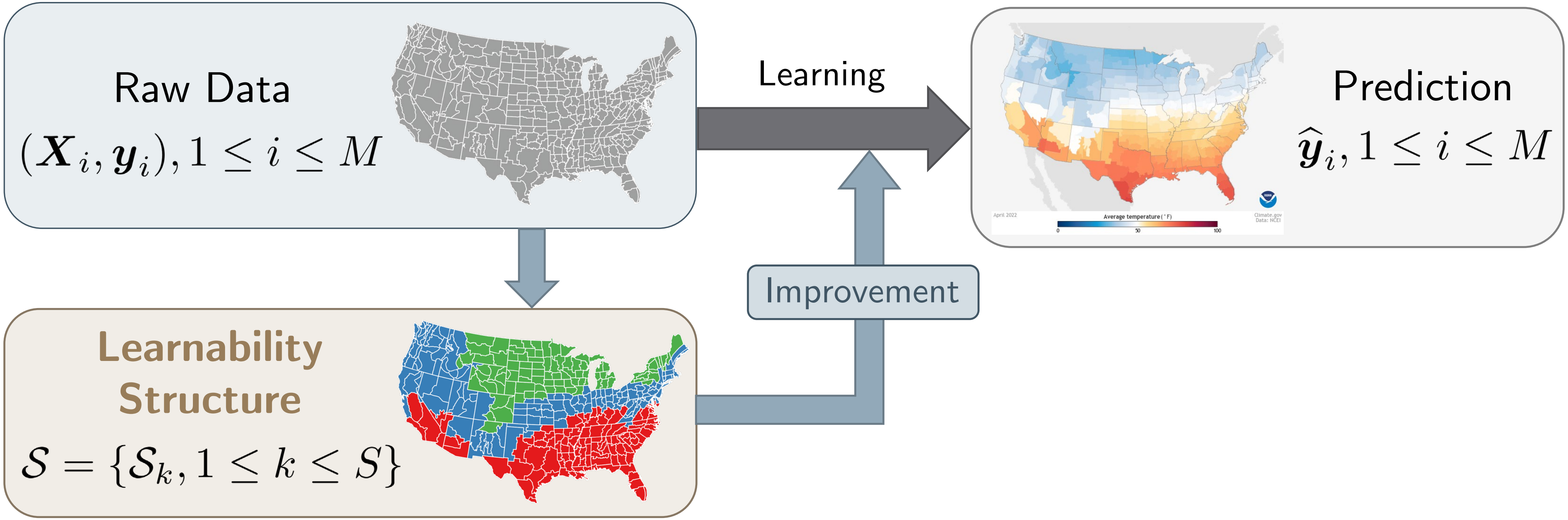
Data Domains

Predict Temperature from **Multiple Climate Divisions**



Motivation: Better Model via Mutual Transfer Learning with Learnability Structure

- Each data domain ➡ source/target
- **Similar** domains form subgroups $\mathcal{S}_k$ ➡ **Learnability Structure** $\mathcal{S}$



Cheng, C.-W., Qiao, X., and Cheng, G. Mutual transfer learning for massive data. In Proceedings of the 37th International Conference on Machine Learning, pp. 1800–1809, 2020.

Motivation: Learnability Structure obtained from Real Data

- Learnability Structure Recovery: Key to Mutual Transfer Learning

Artistic Styles Climatic Regions User Preferences

Challenge I: Mixed-Effects Heterogeneity

- Common Linear Model (M domains in total):

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}, 1 \leq i \leq M$$

  - Parameters are transferable among all the domains

- Linear **Mixed-Effects** Model:

Random Effects

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i(\boldsymbol{\theta}_i + \boldsymbol{u}_i) + \boldsymbol{\varepsilon}, 1 \leq i \leq M$$

Global Parameters

Heterogeneous Parameters

  - **Global** Parameters: Transferable among **all the domains**
  - **Heterogeneous** Parameters: Transferable **in one subgroup** $\boldsymbol{\theta}_i \equiv \boldsymbol{\alpha}_k, i \in \mathcal{S}_k$
  - Random Effects: Domain-Specific, cannot be transferred

Challenge I: Mixed-Effects Heterogeneity

- Linear Mixed-Effects Model:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i(\boldsymbol{\theta}_i + \boldsymbol{u}_i) + \boldsymbol{\varepsilon}, 1 \leq i \leq M$$

Random Effects ($\boldsymbol{u}_i$)

Global Parameters ($\boldsymbol{\beta}$)

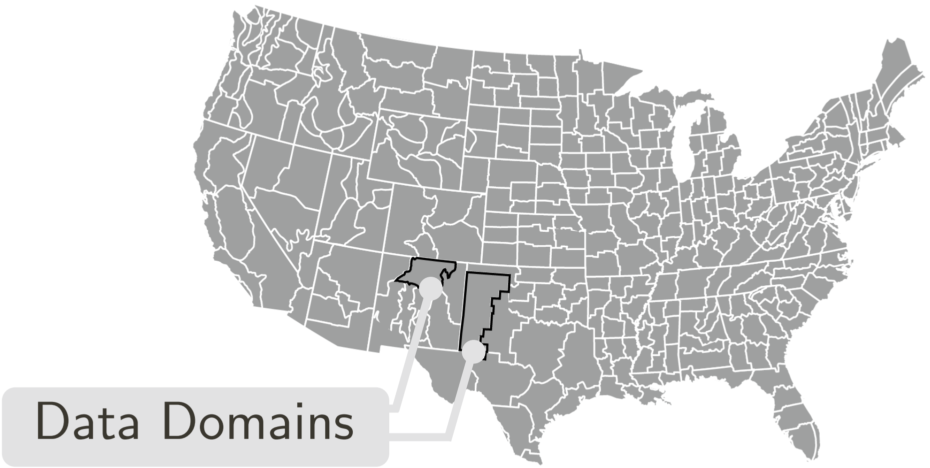
Heterogeneous Parameters ($\boldsymbol{\theta}_i$)

- However, previous methods have their limitations:

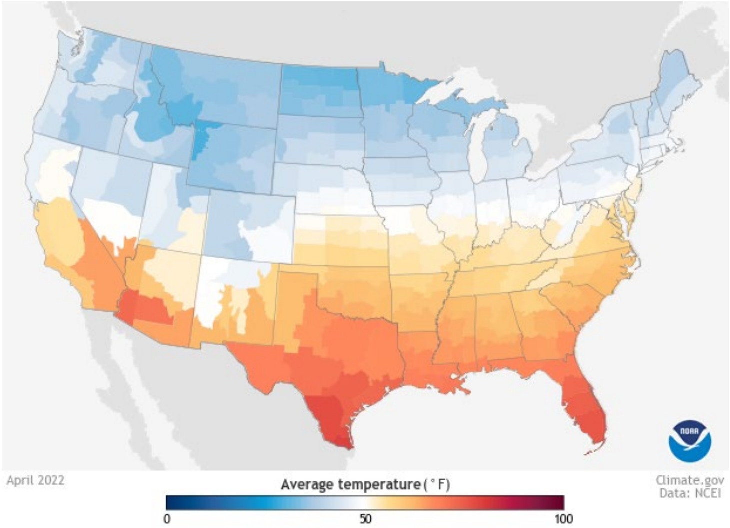
Previous Methods	**Learnability Structure Recovery**	**Mixed-Effects Model Learning**
Generalized Least Squares	✗	✓
Statistical Methods	✓	✗
Goal	✓	✓

Challenge II: Large-Scale Data Setting

- The number of domains M grows rapidly in big data applications
- Mutual transfer learning methods suffer from high time cost
  - CD Fusion[1]: Time Complexity $O[M^3(q^3 + n^2 q) + M^4 q^2]$
  - e.g., NOAA Dataset: 25.83 hr/iter ⊗



NOAA nClimDiv Database
$(M = \#\text{climate divisions} \geq 344)$



Monthly Average Temperature Prediction

[1](Cheng et al., 2020)

Proposed: A Difference Standardization Method for Mutual Transfer Learning (**DiffS**)

- Accurate Learnability Structure Recovery ✅
- Fast Estimation ✅

Raw Data

$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y})$

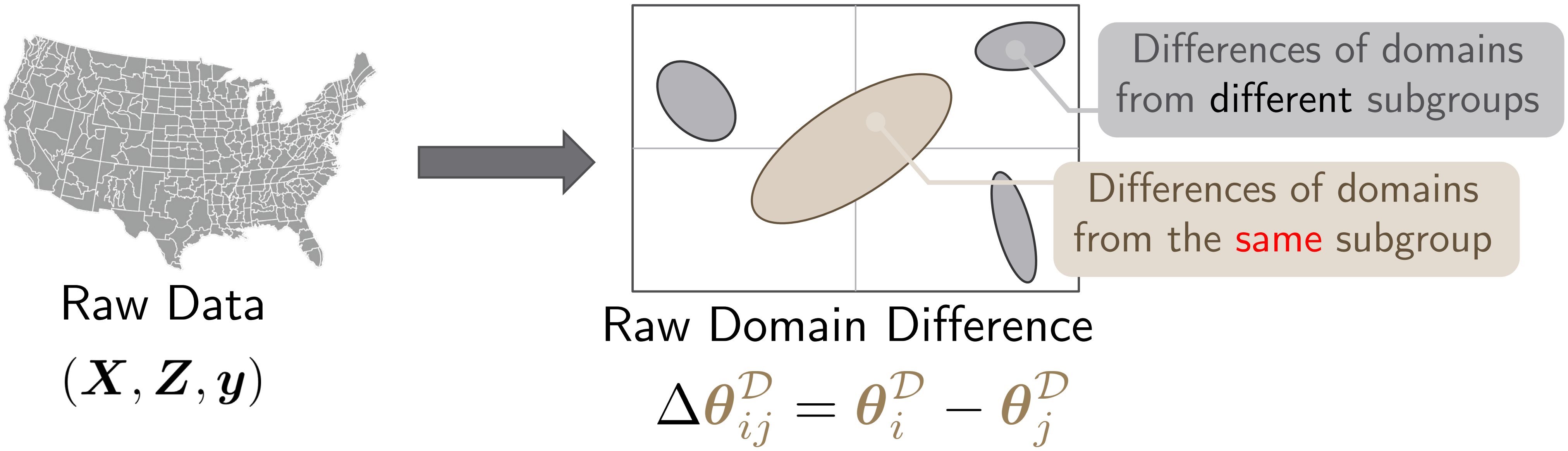
$$\min \mathcal{L}_{\mathrm{GLS}}(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$$

Domain-wise Generalized Least Squares (GLS) Estimate

$(\boldsymbol{\beta}_i^{\mathcal{D}}, \boldsymbol{\theta}_i^{\mathcal{D}})$

Proposed: A Difference Standardization Method for Mutual Transfer Learning (DiffS)

- Accurate Learnability Structure Recovery ✓
- Fast Estimation ✓

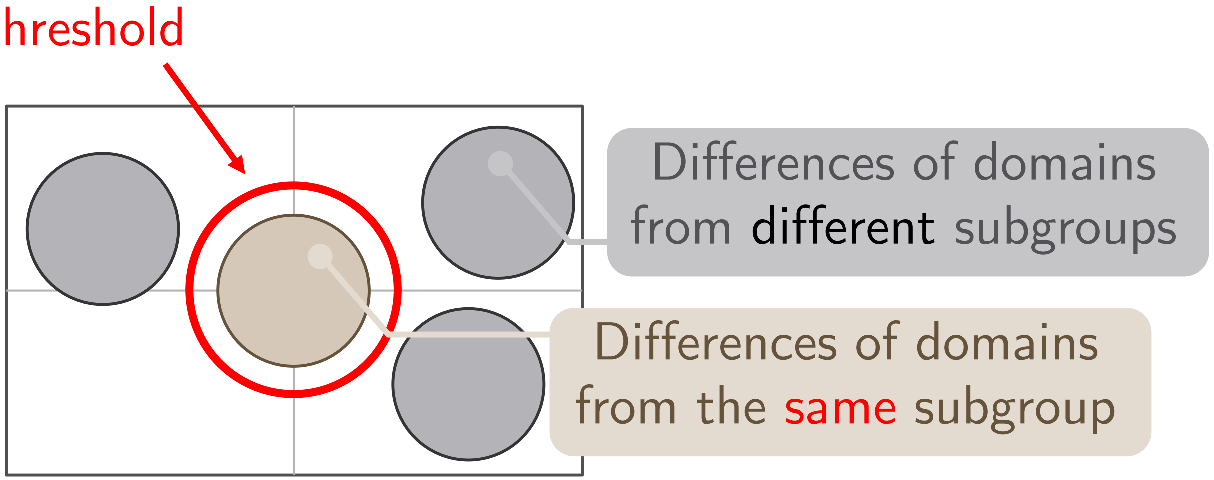


Proposed: A Difference Standardization Method for Mutual Transfer Learning (**DiffS**)

- Accurate Learnability Structure Recovery ✓
- Fast Estimation ✓

Raw Data

$(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y})$

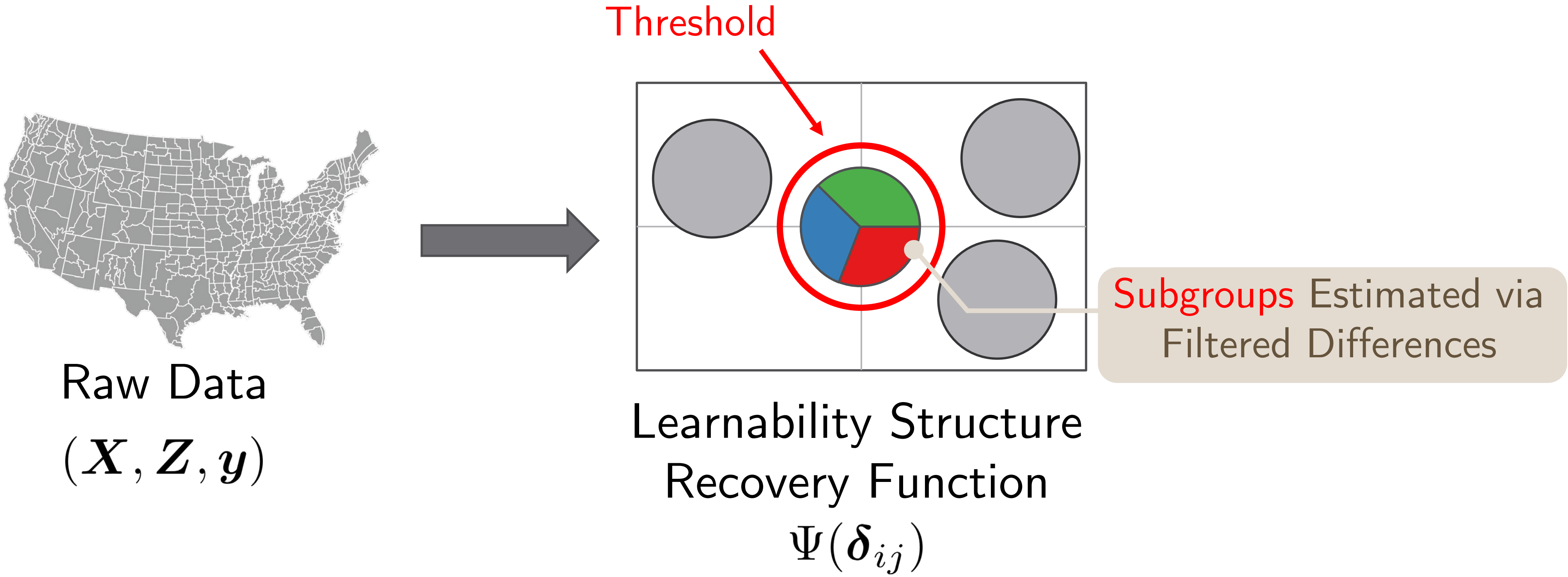


Standardized Domain Difference

$$\boldsymbol{\delta}_{ij} = \Sigma_{ij}^{-1/2}(\boldsymbol{\theta}_i^{\mathcal{D}} - \boldsymbol{\theta}_j^{\mathcal{D}})$$

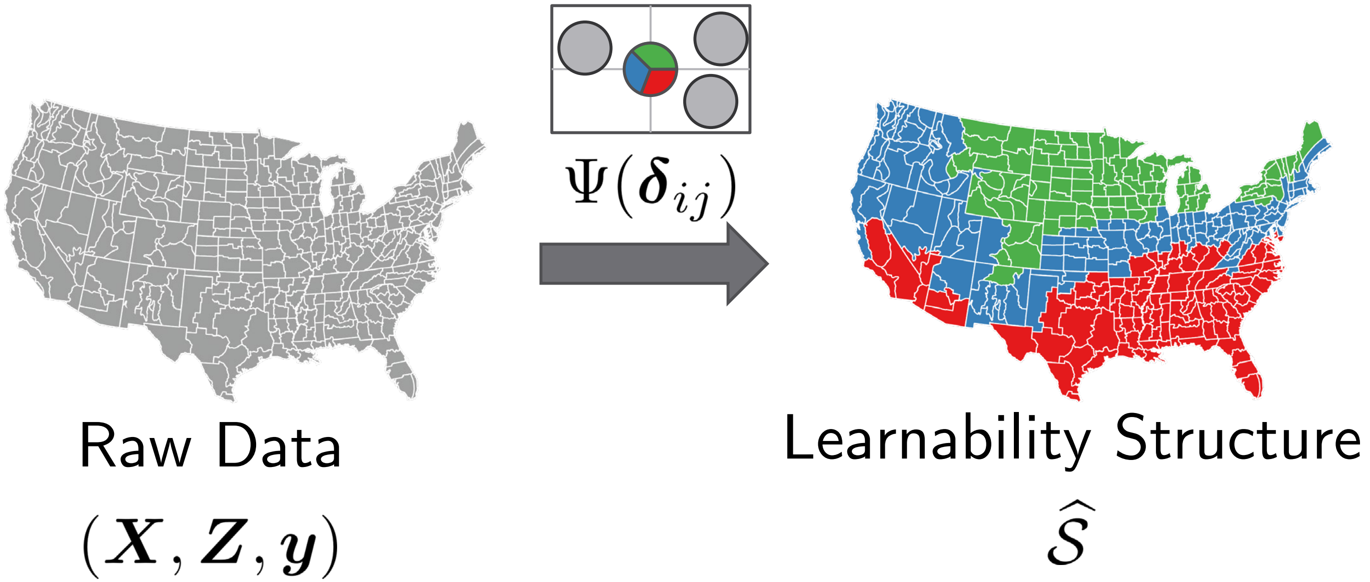
Proposed: A Difference Standardization Method for Mutual Transfer Learning (DiffS)

- Accurate Learnability Structure Recovery ✓
- Fast Estimation ✓



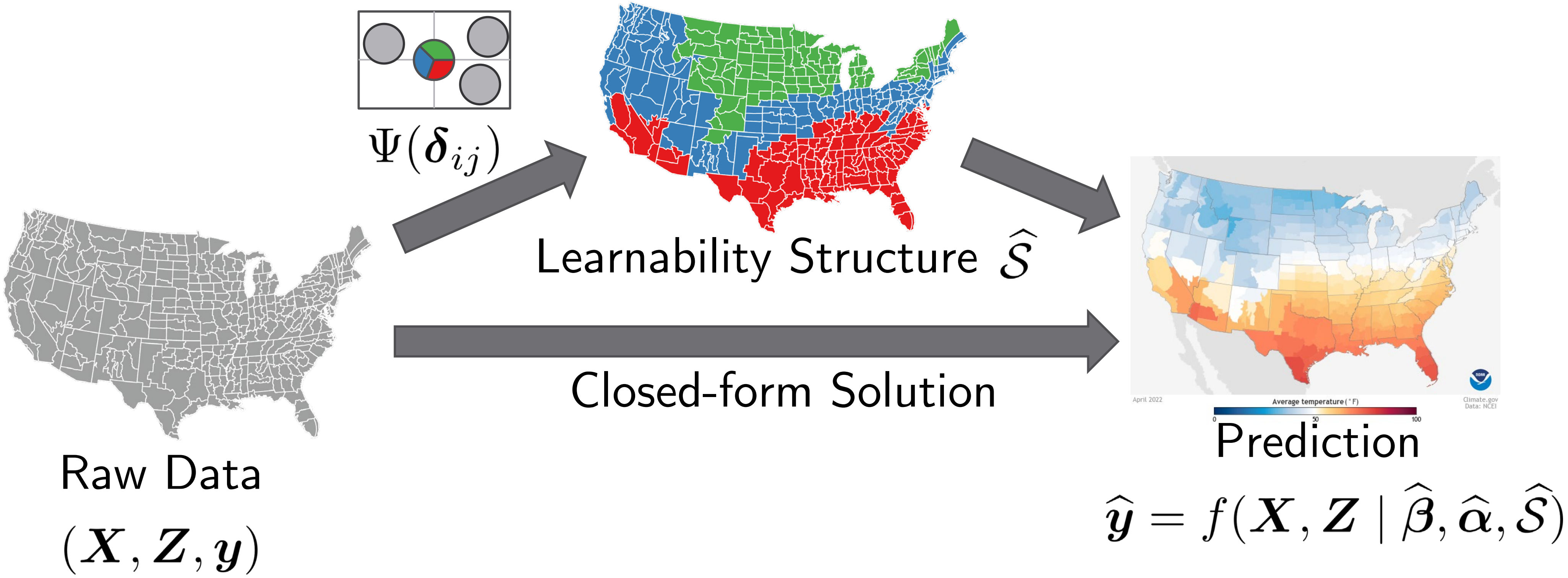
Proposed: A Difference Standardization Method for Mutual Transfer Learning (**DiffS**)

- Accurate Learnability Structure Recovery
- Fast Estimation



Proposed: A Difference Standardization Method for Mutual Transfer Learning (**DiffS**)

- Accurate Learnability Structure Recovery ✓
- Fast Estimation ✓



Solution to Challenge I: Learnability Structure Recovery with Standardized Domain Difference

- Proposed: **Standardized Domain Difference**

$$\boldsymbol{\delta}_{ij} = \Sigma_{ij}{}^{-1/2}(\boldsymbol{\theta}_i^{\mathcal{D}} - \boldsymbol{\theta}_j^{\mathcal{D}})$$

  - Σ_{ij}: calculated from raw data
  - $\boldsymbol{\theta}_i^{\mathcal{D}}$: domain-wise Generalized Least Squares estimate

- We found that: $\boldsymbol{\delta}_{ij} \sim \begin{cases} \mathcal{N}(\mathbf{0}, \boldsymbol{I}), & i, j \text{ in the same subgroup,} \\ \mathcal{N}\left(\Sigma_{ij}{}^{-1/2}\Delta\boldsymbol{\theta}_{ij}^{*}, \boldsymbol{I}\right), & \text{otherwise.} \end{cases}$

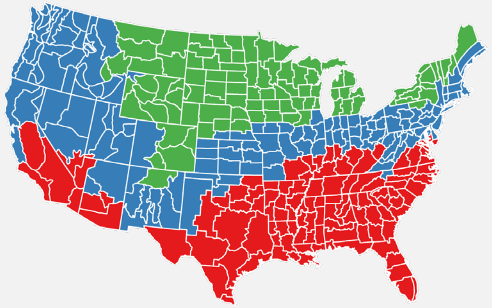
Raw Data $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{y})$

$\Delta\boldsymbol{\theta}_{ij}^{\mathcal{D}}$ →

$$\boldsymbol{\Delta} = \left(\|\boldsymbol{\delta}_{ij}\|^2\right)_{M\times M}$$

Standardized Domain Difference

→ $\Psi(\boldsymbol{\Delta})$



Learnability Structure $\mathcal{S}$

Solution to Challenge II: Closed-Form Solution based on DiffS Learnability Structure Estimate

- Objective of DiffS:

$$\min \ \mathcal{L}_{\text{DiffS}}(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{|\widehat{\mathcal{S}}|}) = \sum_{i=1}^{M} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{\theta}_i)^\top \boldsymbol{W}_i (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{\theta}_i),$$

$$\text{s.t.} \ \widehat{\mathcal{S}} = \Psi(\boldsymbol{\Delta}), \quad \boldsymbol{\theta}_i = \boldsymbol{\alpha}_s, \ \forall i \in \widehat{\mathcal{S}}_k, 1 \leq k \leq |\widehat{\mathcal{S}}|.$$

Learnability Recovery with **Standardized Domain Difference**

- **Closed-Form Solution:** $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}_1, \ldots, \widehat{\boldsymbol{\alpha}}_{|\widehat{\mathcal{S}}|}\right)^\top = [\boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{G}]^{-1} \boldsymbol{G}^\top \boldsymbol{W} \boldsymbol{y}$
  - $\boldsymbol{G}$ is constructed by $\boldsymbol{X}_i, \boldsymbol{Z}_i, \widehat{\mathcal{S}}$

Theoretical Results

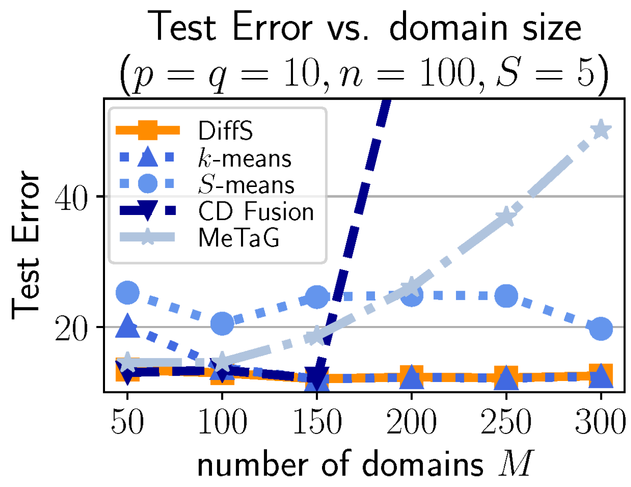
- **DiffS** is able to perfectly recover the learnability structure:

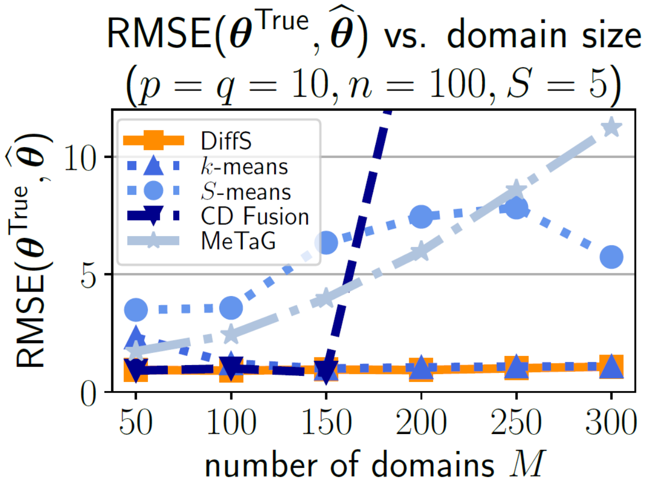
Theorem 4.3 (Learnability structure recovery guarantee). *Denoting $\mathcal{S}^*$ as the true learnability structure, supposing that the Assumption 4.1 is satisfied and learnability structure recovering is applied via Algorithm 2, thus $\widehat{\mathcal{S}} = \mathcal{S}^*$.*

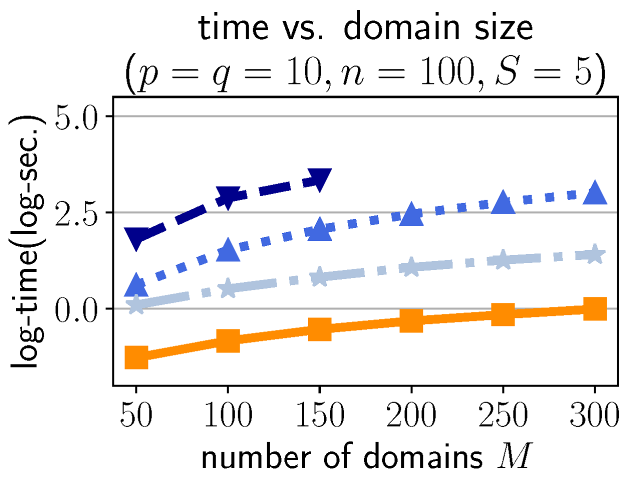
- **DiffS** achieves a significant complexity improvement:

CD Fusion (Baseline)	DiffS (Proposed)
$O[M^3(q^3 + n^2 q) + M^4 q^2]$	$O[M^2(n^2(p + Sq) + q^3)]$

Empirical Results on Synthetic Datasets





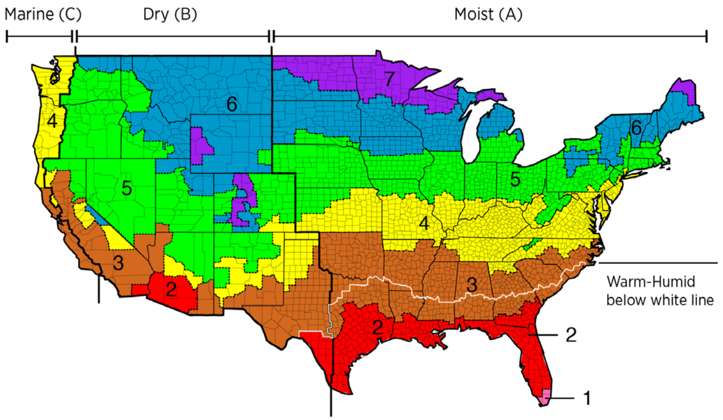


- **DiffS** has best prediction performance among baselines
- **DiffS** obtains best parameter estimation
- **DiffS** achieves the fastest estimation speed (~1000× CD Fusion)

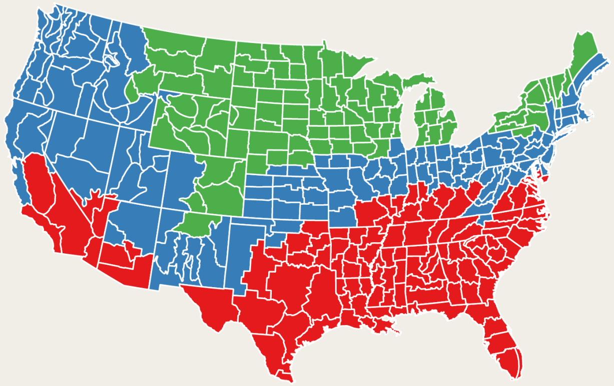
Empirical Results on Two Real-World Datasets

- Real-world tasks
  - NOAA nClimDiv Temperature Prediction
  - Microarray Gene Expression
- **DiffS** provides more reasonable results within an acceptable time.

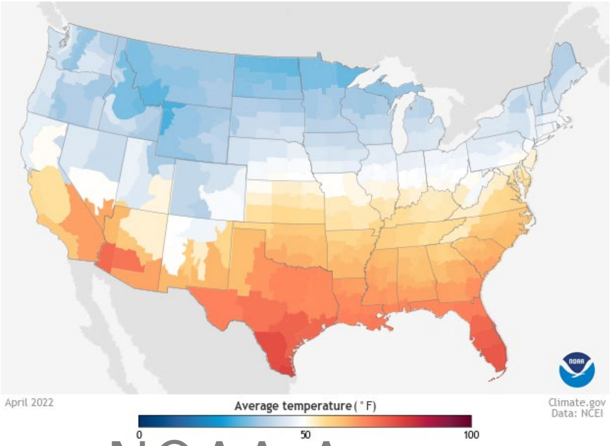
Method	NOAA nClimDiv Database	
	Timecost	Test Error
DiffS	**33.79**	$\mathbf{82.44 \pm 3.44}$
k-means	119.73	88.72 ± 8.91
MeTaG	32.49	$> 10^{200}$
	Microarray Data	
DiffS	**0.0169**	$\mathbf{0.93 \pm 0.14}$
k-means	0.2677	0.96 ± 0.13
MeTaG	0.2988	0.90 ± 0.14



Climate Zone from IECC



Learnability Structure estimated by **DiffS**



NOAA Average Temperature Outlook

Vose, R. S., Applequist, S., Squires, M., et al. Noaa's gridded climate divisional dataset (climdiv). NOAA National Climatic Data Center, 2014. doi: 10.7289/V5M32STR.

Thank you for your attention!

- **Haoqing Xu** (haoqing_xu@outlook.com)
- Meng Wang (meng.wang@seu.edu.cn)
- Beilun Wang* (beilun@seu.edu.cn)

- Spotlight Presentation
  - https://icml.cc/virtual/2022/spotlight/17526
- Poster: #17525
  - https://icml.cc/virtual/2022/poster/17525

東南大學
SOUTHEAST UNIVERSITY

ICML
International Conference On Machine Learning