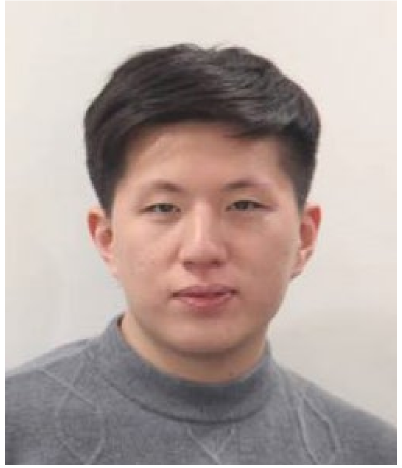
東南大學
SOUTHEAST UNIVERSITY

ICML
International Conference
On Machine Learning

Adaptive Group Personalization for Federated Mutual Transfer Learning

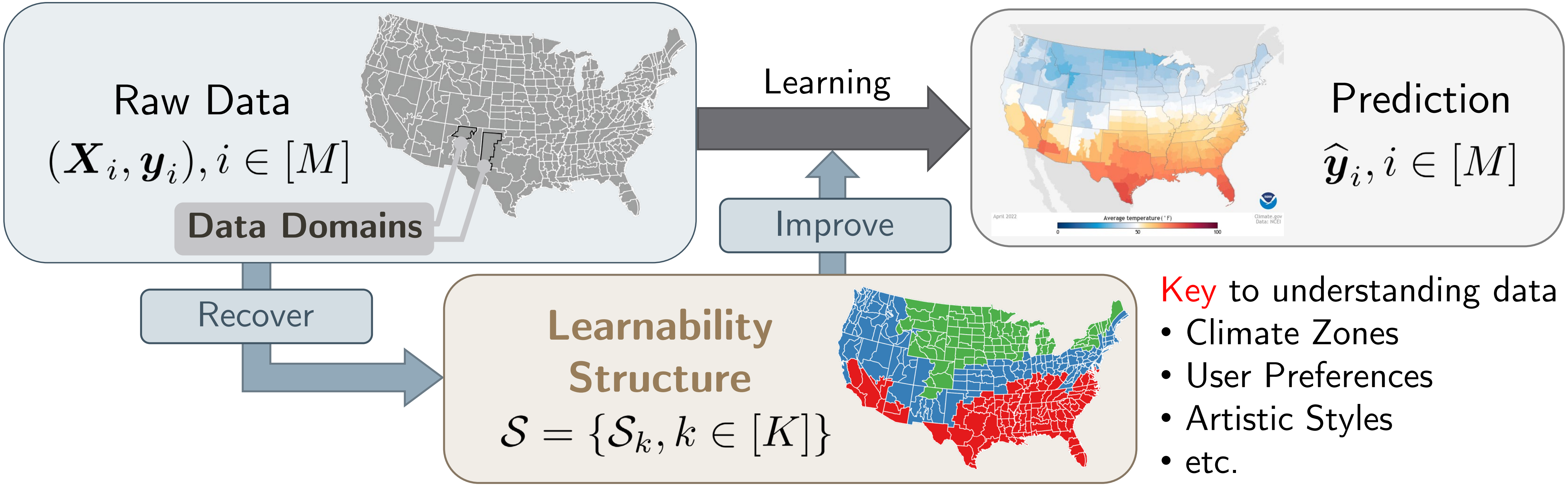



Haoqing Xu[1] Dian Shen[1] Meng Wang[2] Beilun Wang[1] *

[1]School of Computer Science and Engineering, Southeast University
[2]College of Design and Innovation, Tongji University

Motivation: Recovering Learnability Structure in Big Data Applications

- Mutual Transfer Learning[1]: Each data domain ⟶ source/target
- **Similar** domains form clusters $\mathcal{S}_k$ ⟶ **Learnability Structure** $\mathcal{S}$

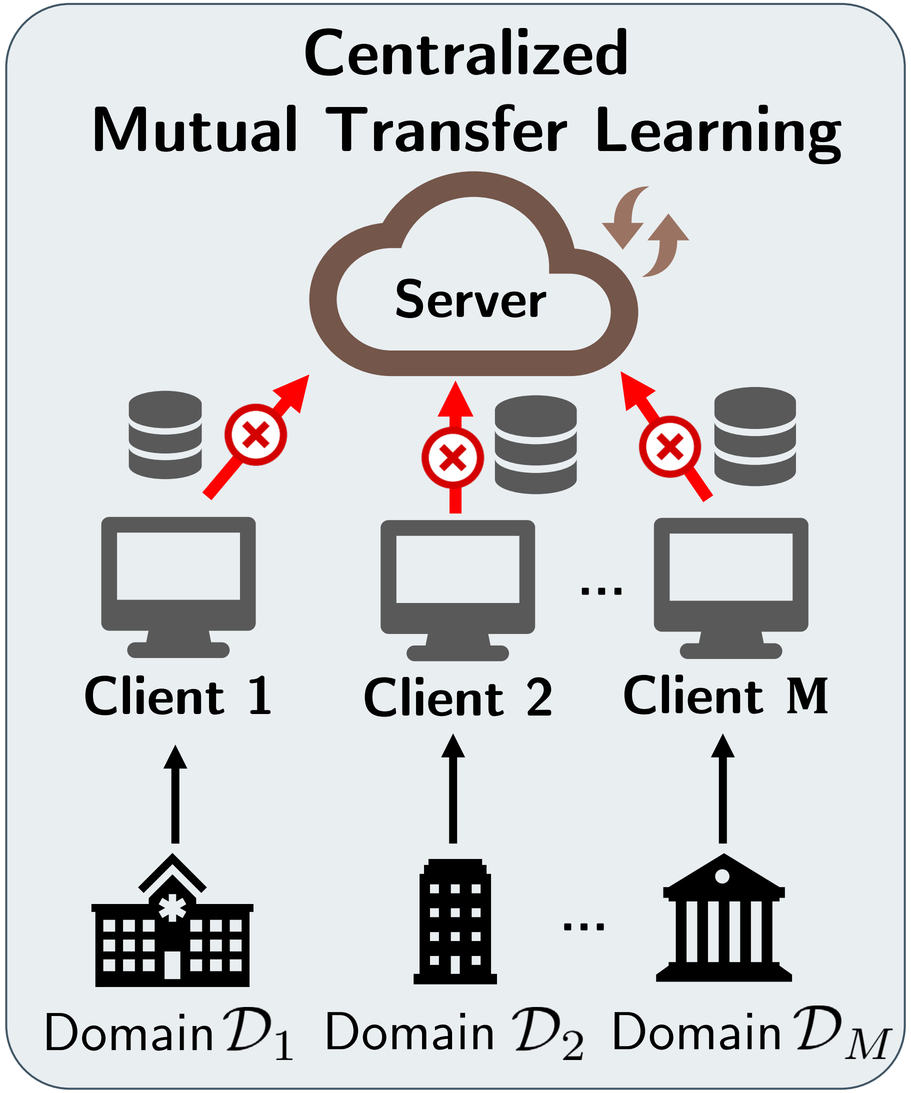


Key to understanding data
- Climate Zones
- User Preferences
- Artistic Styles
- etc.

[1] Cheng, Ching-Wei, Xingye Qiao, and Guang Cheng. "Mutual transfer learning for massive data." *International Conference on Machine Learning*. PMLR, 2020.

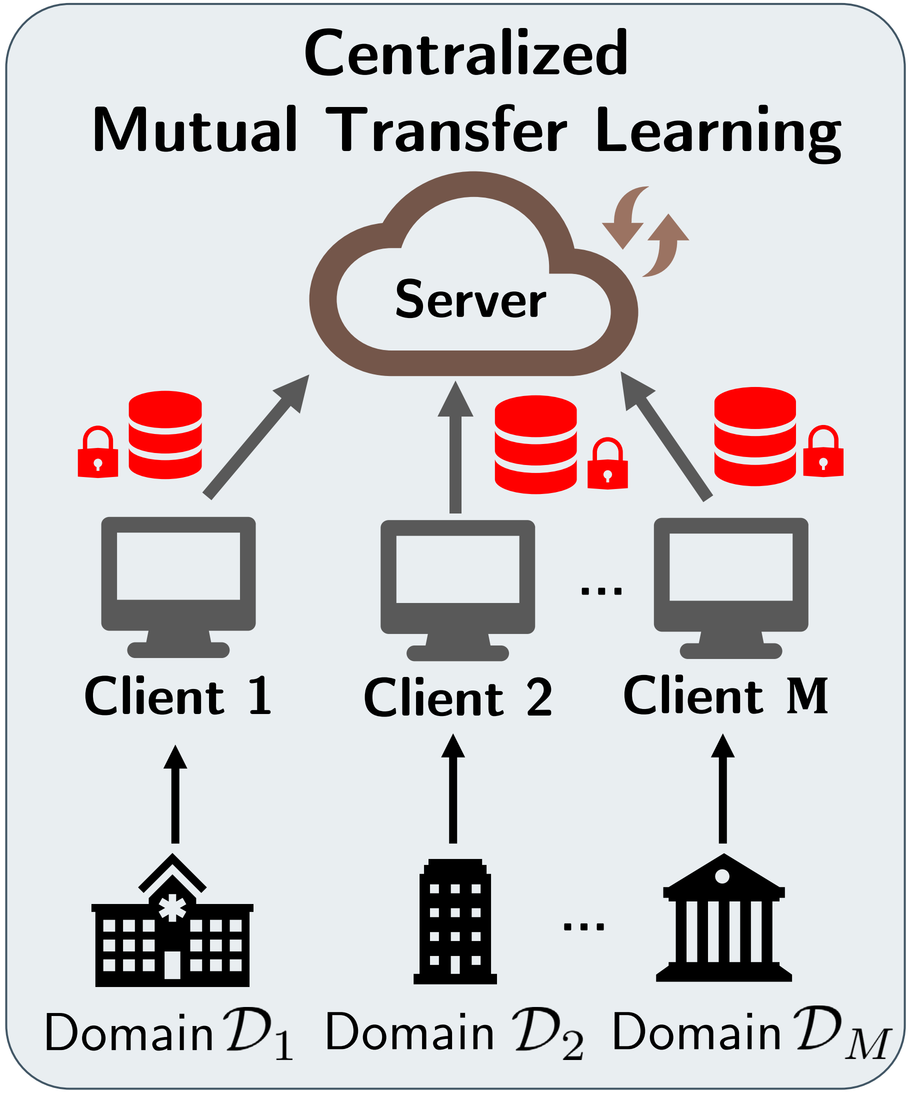
Motivation: Bottlenecks of Centralized Mutual Transfer Learning with Large-Scale data

- **Communication Overload**
  - TBs of data are collected by clients
  - Transferring to server is too costly



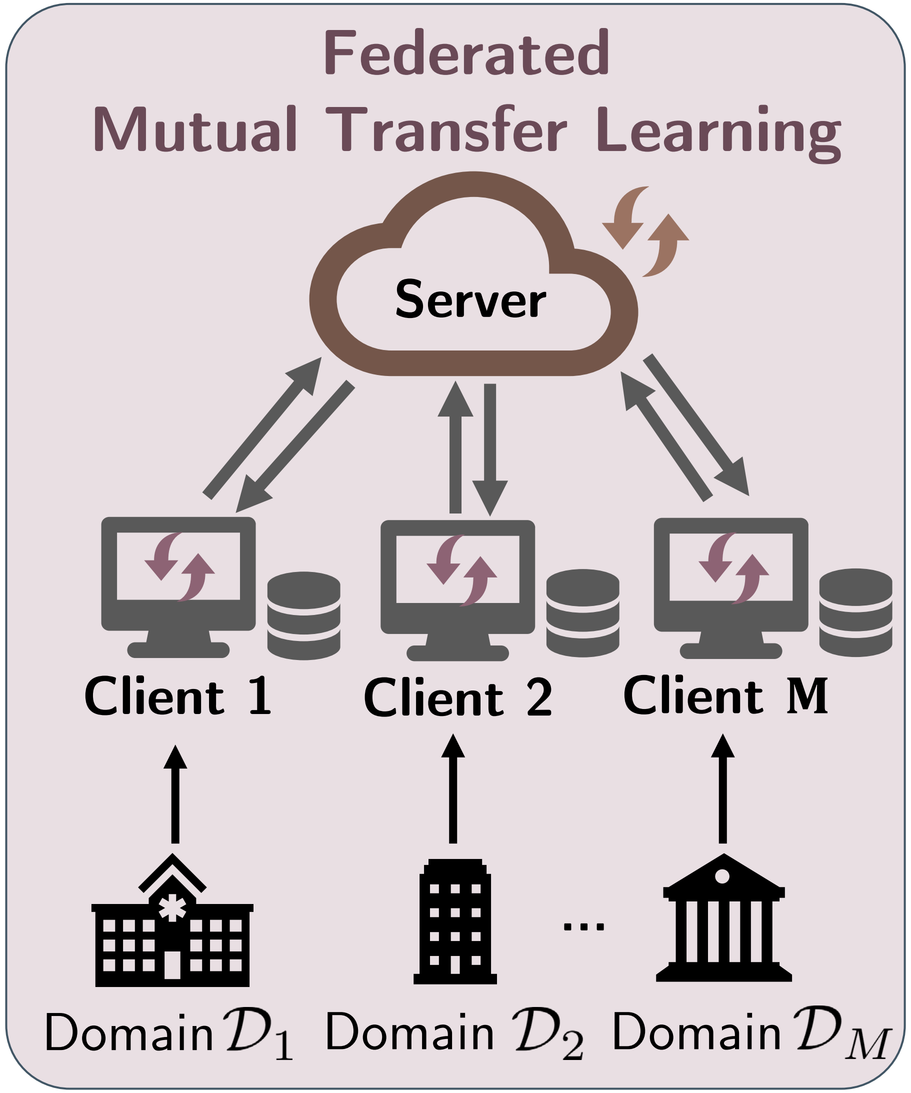
Motivation: Bottlenecks of Centralized Mutual Transfer Learning with Large-Scale data

- **Communication Overload**
  - TBs of data are collected by clients
  - Transferring to server is too costly
- **Privacy Concern**
  - Data contains sensitive information
  - Leakage may cause serious ethic issues



Motivation: Bottlenecks of Centralized Mutual Transfer Learning with Large-Scale data

- **Communication Overload**
  - TBs of data are collected by clients
  - Transferring to server is too costly
- **Privacy Concern**
  - Data contains sensitive information
  - Leakage may cause serious ethic issues
- **Federated Mutual Transfer Learning**
  - Transfer parameters only
  - Overcome the above problems simultaneously



Challenge I: Learnability 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 shared among all the domains

- **Mutual Transfer Learning:**

$$\boldsymbol{y}_i = \boldsymbol{X}_i \underbrace{\boldsymbol{\beta}}_{\text{Global Parameters}} + \boldsymbol{Z}_i(\underbrace{\boldsymbol{\alpha}_{k_i}}_{\text{Heterogeneous Parameters}} + \overbrace{\boldsymbol{u}_i}^{\text{Random Effects}}) + \boldsymbol{\varepsilon}, 1 \leq i \leq M$$

  - **Global** Parameters: Shared among **all the domains**
  - **Heterogeneous** Parameters: Shared **in one subgroup** $\boxed{\mathcal{D}_i \in \mathcal{S}_{k_i}}$
  - Random Effects: Domain-Specific, cannot be transferred

Challenge I: Learnability Heterogeneity

- **Mutual Transfer Learning:**

Random Effects

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i(\boldsymbol{\alpha}_{k_i} + \boldsymbol{u}_i) + \boldsymbol{\varepsilon}, 1 \le i \le M$$

Global Parameters

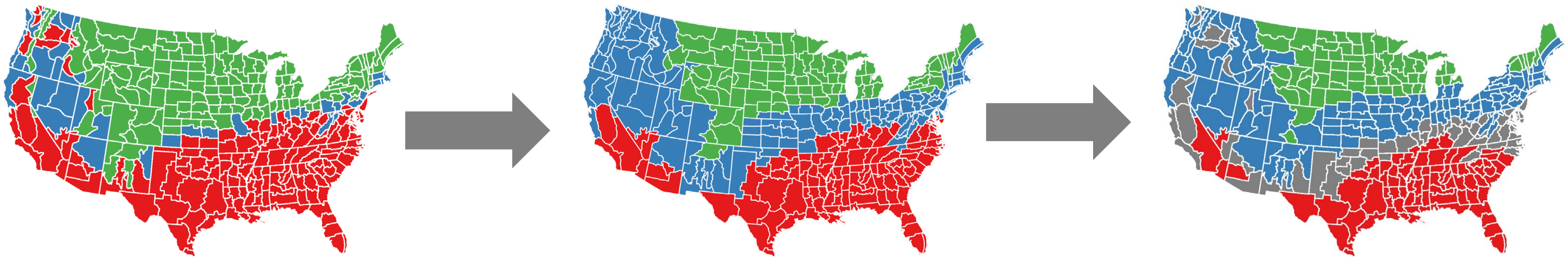
Heterogeneous Parameters

- However, previous methods cannot easily adapt to such task:

Previous Methods	Learnability Structure Recovery	Distributed Learning
Single-model FL	✗	✓
Centralized MTL	✓	✗
Goal	✓	✓

Challenge II: Concept Drift

- Multiple communication rounds → Long period of time
- Data distribution may change → **Concept drift** occurs
  - User preferences change due to new trends
  - Climate slightly changes due to human activities
  - etc…



$t = 1$ $t = 2$ $t = 3$

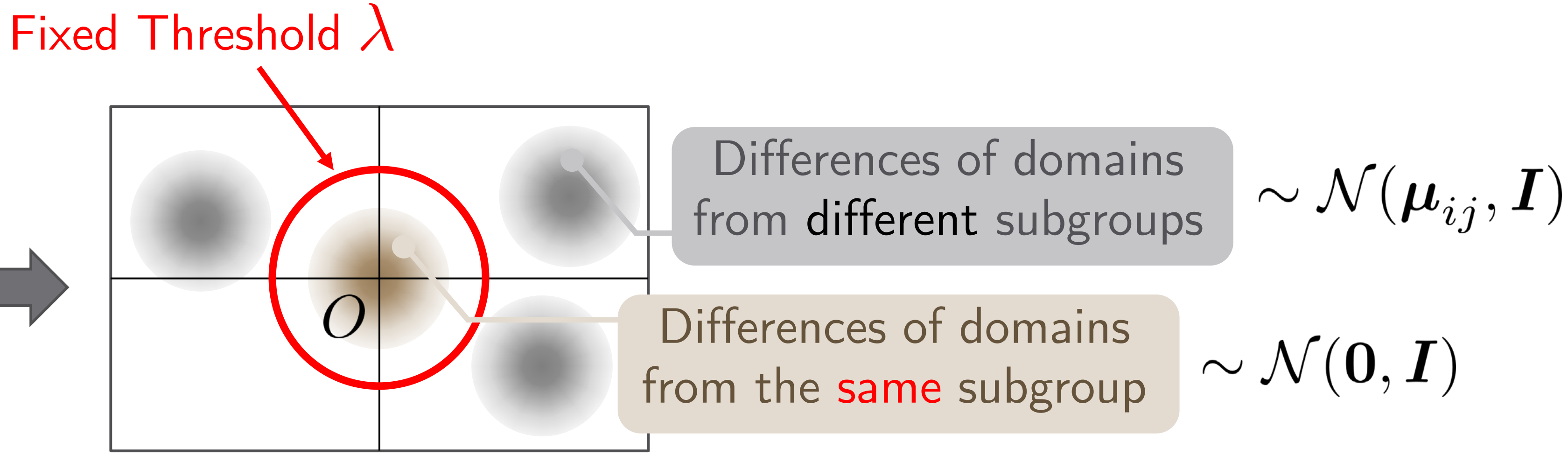
Learnability structure may change over time

Challenge II: Concept Drift

- Previous mutual transfer learning focus on stable data
  - DiffS[2] uses a fixed threshold: cannot well adapt to dynamic environment

Raw Data

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

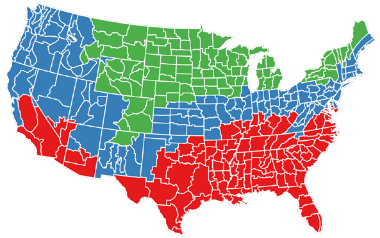


Standardized Domain Difference

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

where $\boldsymbol{\theta}_i^{\mathcal{D}} = \arg\min \ell_i(\boldsymbol{\beta}, \boldsymbol{\theta}_i)$

$\Psi(\|\boldsymbol{\delta}\|^2; \lambda)$



Learnability Structure

[2] Xu, Haoqing, Meng Wang, and Beilun Wang. "A difference standardization method for mutual transfer learning." *International Conference on Machine Learning*. PMLR, 2022.

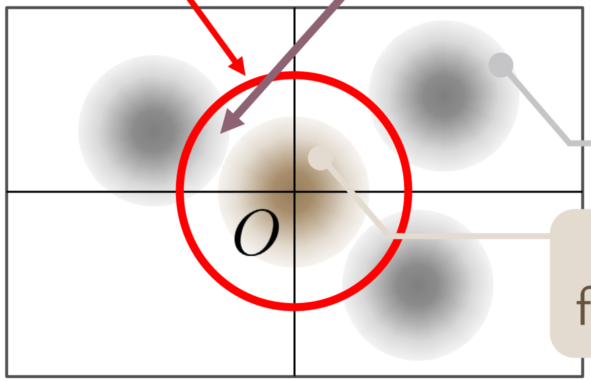
Challenge II: Concept Drift

- Previous mutual transfer learning focus on stable data
  - DiffS[2] uses a fixed threshold: cannot well adapt to dynamic environment

Raw Data

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

Fixed Threshold λ False Positives!



Differences of domains from different subgroups $\sim \mathcal{N}(\boldsymbol{\mu}_{ij}, \boldsymbol{I})$

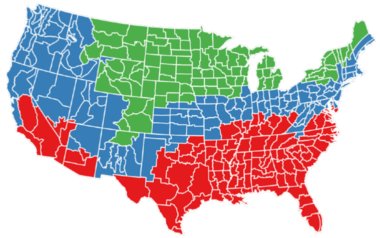
Differences of domains from the same subgroup $\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

Standardized Domain Difference

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

where $\boldsymbol{\theta}_i^{\mathcal{D}} = \arg\min \ell_i(\boldsymbol{\beta}, \boldsymbol{\theta}_i)$

$\Psi(\|\boldsymbol{\delta}\|^2; \lambda)$



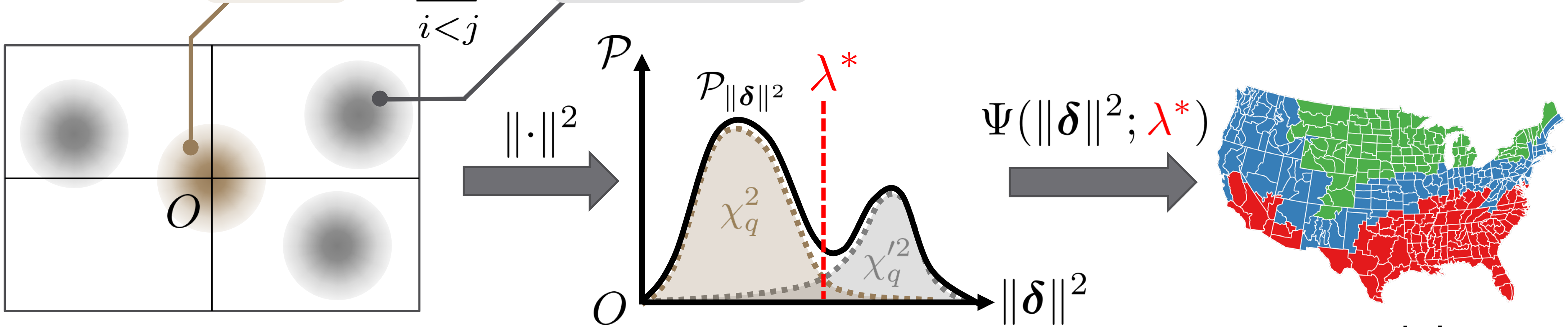
Learnability Structure

[2] Xu, Haoqing, Meng Wang, and Beilun Wang. "A difference standardization method for mutual transfer learning." *International Conference on Machine Learning*. PMLR, 2022.

Solution to Challenge II: Adaptive Threshold Correction under Concept Drift

- Adaptively correct the threshold with current distribution
- **Motivation:** Population $\mathcal{P}_{\|\boldsymbol{\delta}\|^2}(\lambda) = p_\chi \chi_q^2(\lambda) + \sum_{i<j} p_{ij} \chi_q'^2(\lambda; \boldsymbol{\mu}_{ij})$
- **Proposed:** Check the *optimal threshold condition*

$$p_\chi \chi_q^2(\lambda^*) = \sum_{i<j} p_{ij} \chi_q'^2(\lambda^*; \boldsymbol{\mu}_{ij}) \Rightarrow \mathcal{P}_{\|\boldsymbol{\delta}\|^2}(\lambda^*) = 2 p_\chi \chi_q^2(\lambda^*)$$



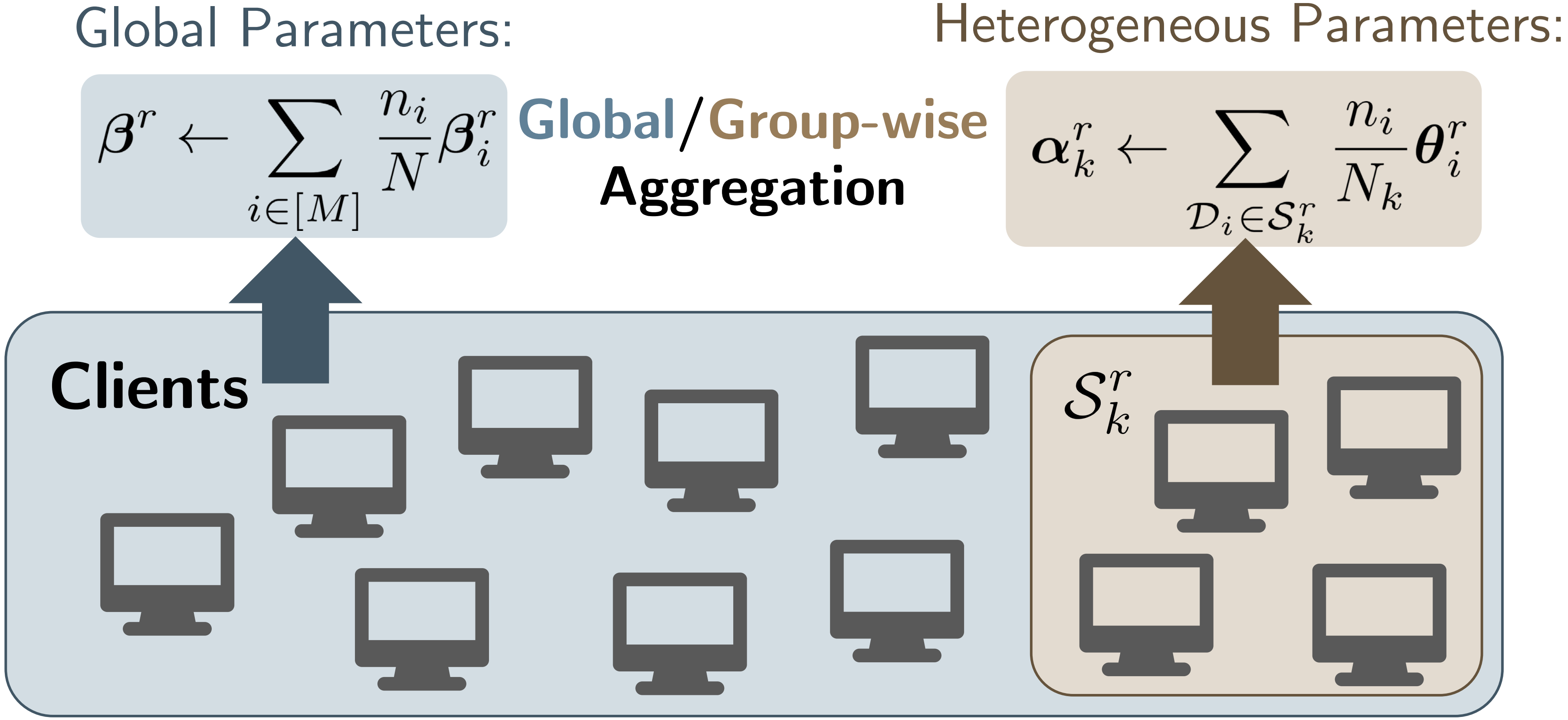
Standardized Domain Difference

Adaptive Threshold Correction $\Lambda(\|\boldsymbol{\delta}\|^2)$

Learnability Structure

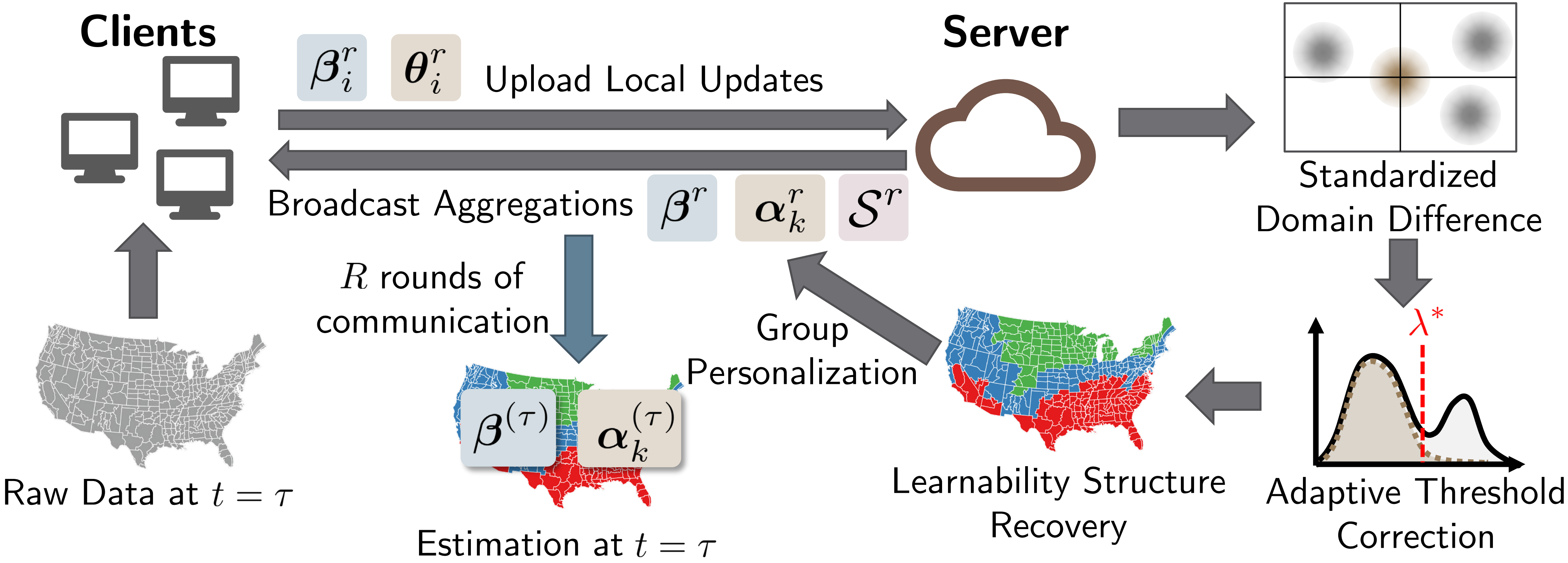
Solution to Challenge I: Group Personalization with Learnability Structure Recovery

- **Proposed:** Group Personalization based Mixed Aggregation
- With learnability structure recovered by $\mathcal{S}^r = \Psi\left(\|\boldsymbol{\delta}^r\|^2; \Lambda(\|\boldsymbol{\delta}^r\|)\right)$



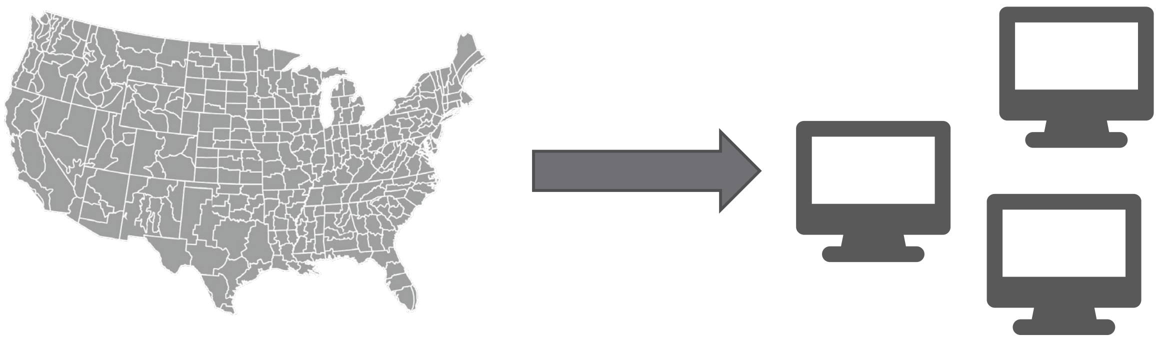
Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



Raw Data at $t = \tau$

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

Clients

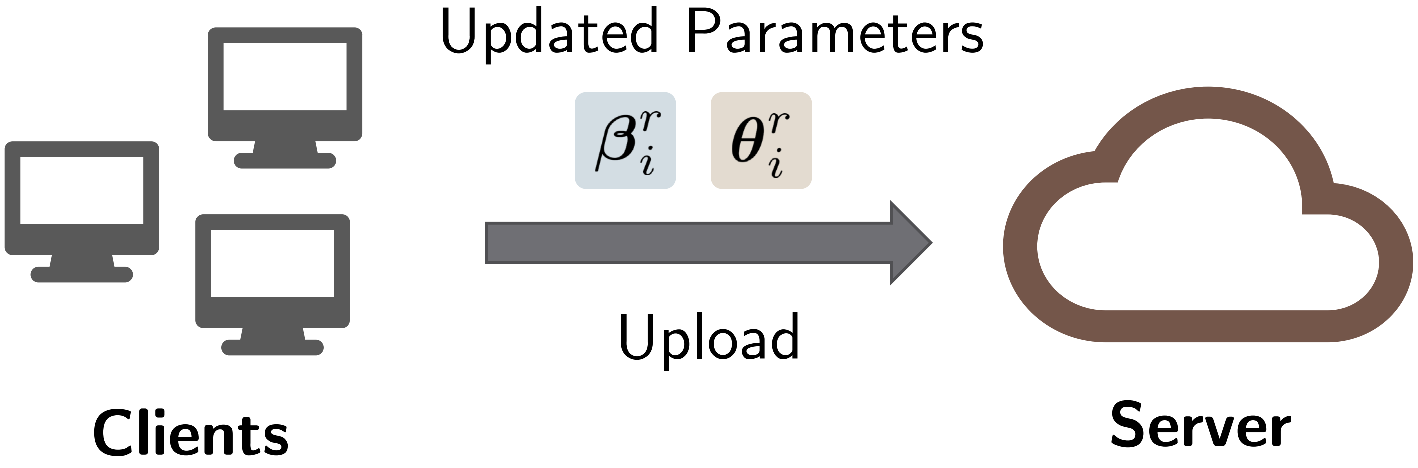
$$(\boldsymbol{\beta}_i^r, \boldsymbol{\theta}_i^r) = \arg\min \ell_i^{(\tau)}(\boldsymbol{\beta}, \boldsymbol{\theta}_i)$$

Client Local Updates with GLS Loss

$\boldsymbol{\beta}_i^r$ $\boldsymbol{\theta}_i^r$

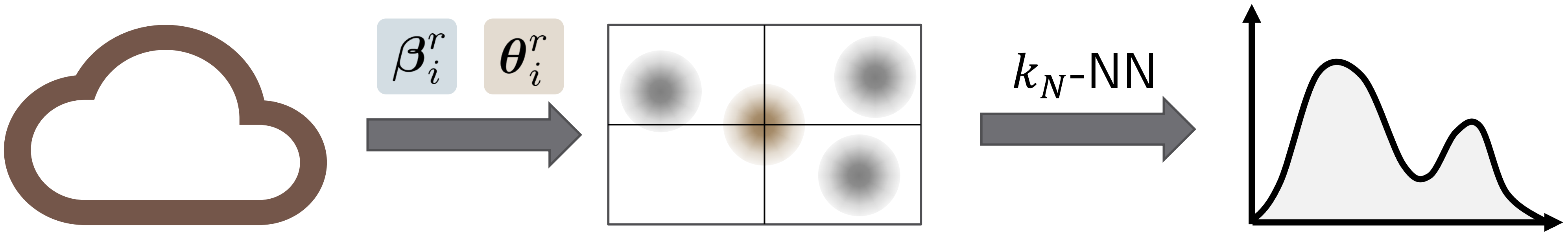
Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (**AdaGrP**)

- Accurate learnability structure recovery in Federated framework ✓
- Robustness against Concept Drift with Tuning-Free Solution ✓



Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (**AdaGrP**)

- Accurate learnability structure recovery in Federated framework ✓
- Robustness against Concept Drift with Tuning-Free Solution ✓



Server

Standardized Domain Difference

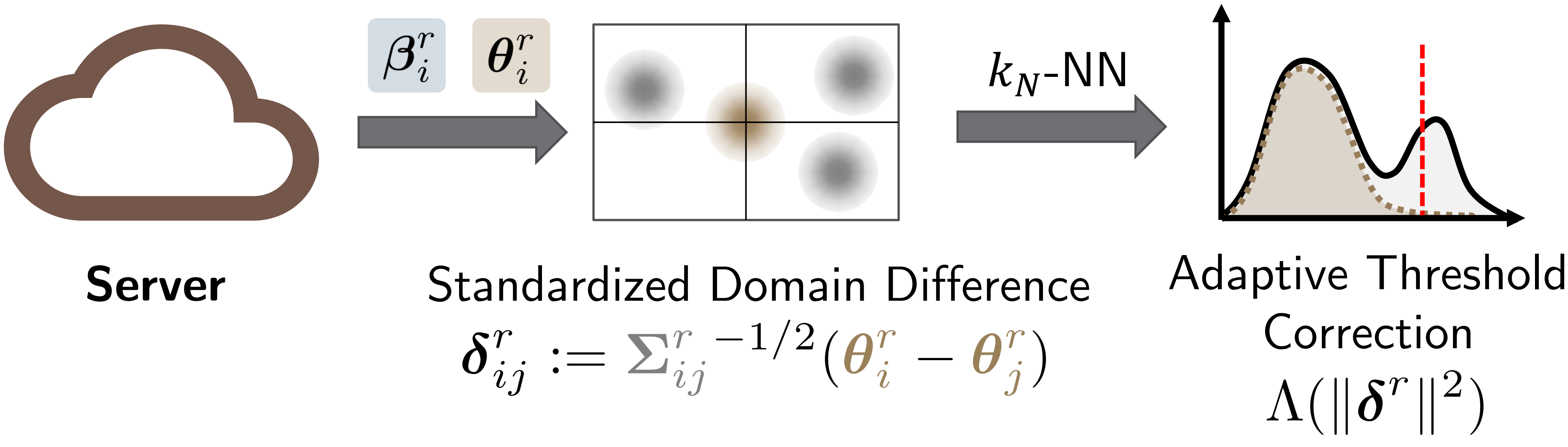
$$\boldsymbol{\delta}_{ij}^r := {\Sigma_{ij}^r}^{-1/2}(\boldsymbol{\theta}_i^r - \boldsymbol{\theta}_j^r)$$

Distance Density

$$\mathcal{P}_{\|\boldsymbol{\delta}^r\|^2}$$

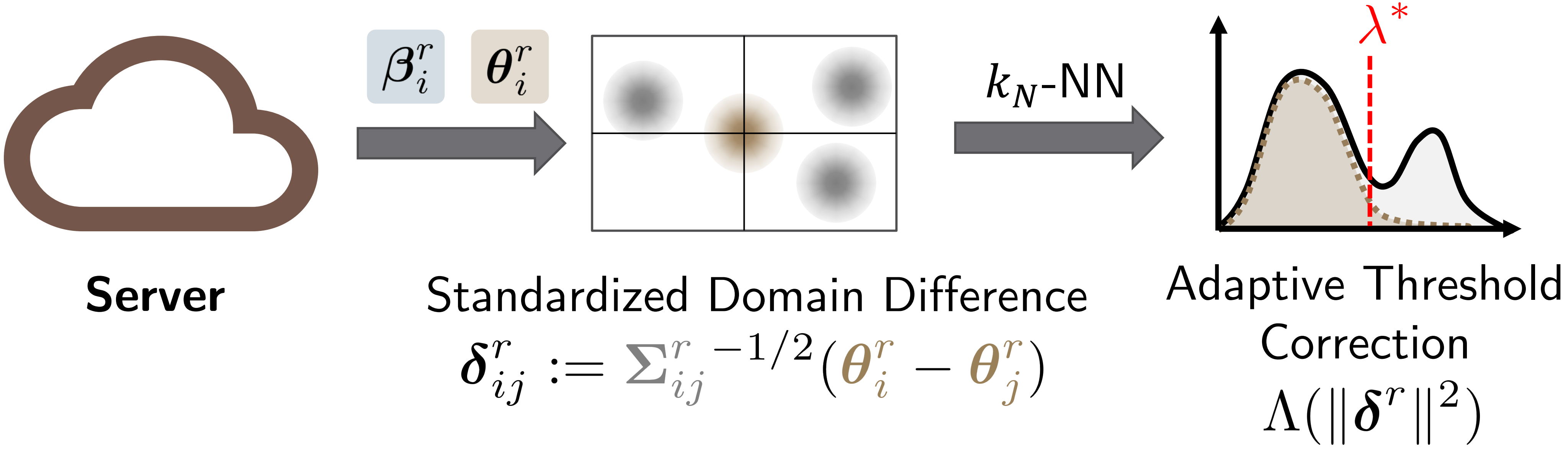
Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



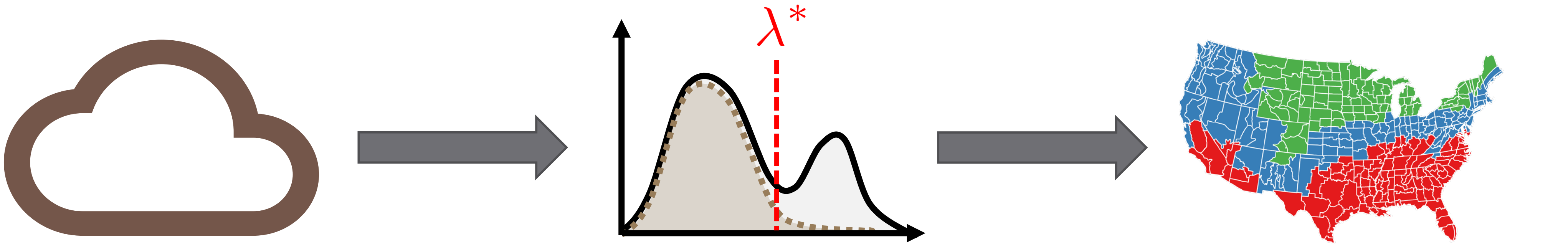
Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



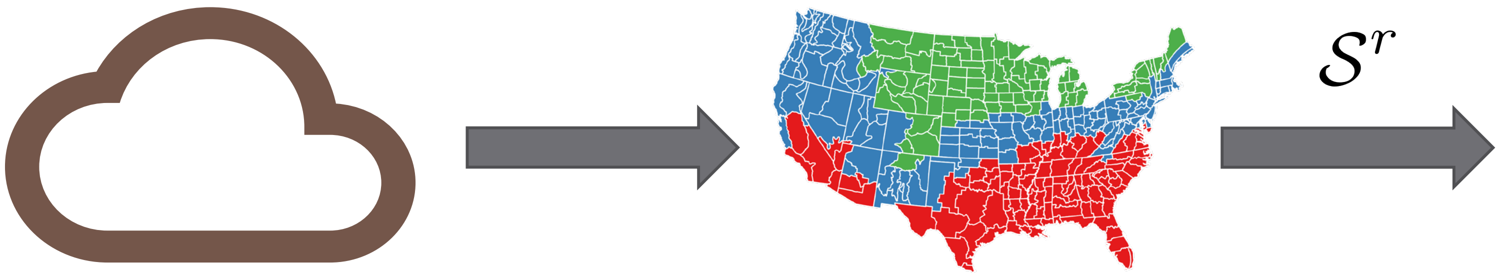
Server

Adaptive Threshold Correction
$\Lambda(\|\boldsymbol{\delta}^r\|^2)$

Learnability Structure Recovery
$\Psi(\boldsymbol{\delta}^r; \Lambda(\|\boldsymbol{\delta}^r\|))$

Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (**AdaGrP**)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution



Server

Learnability Structure Recovery

$$\Psi(\boldsymbol{\delta}^r; \Lambda(\|\boldsymbol{\delta}^r\|))$$

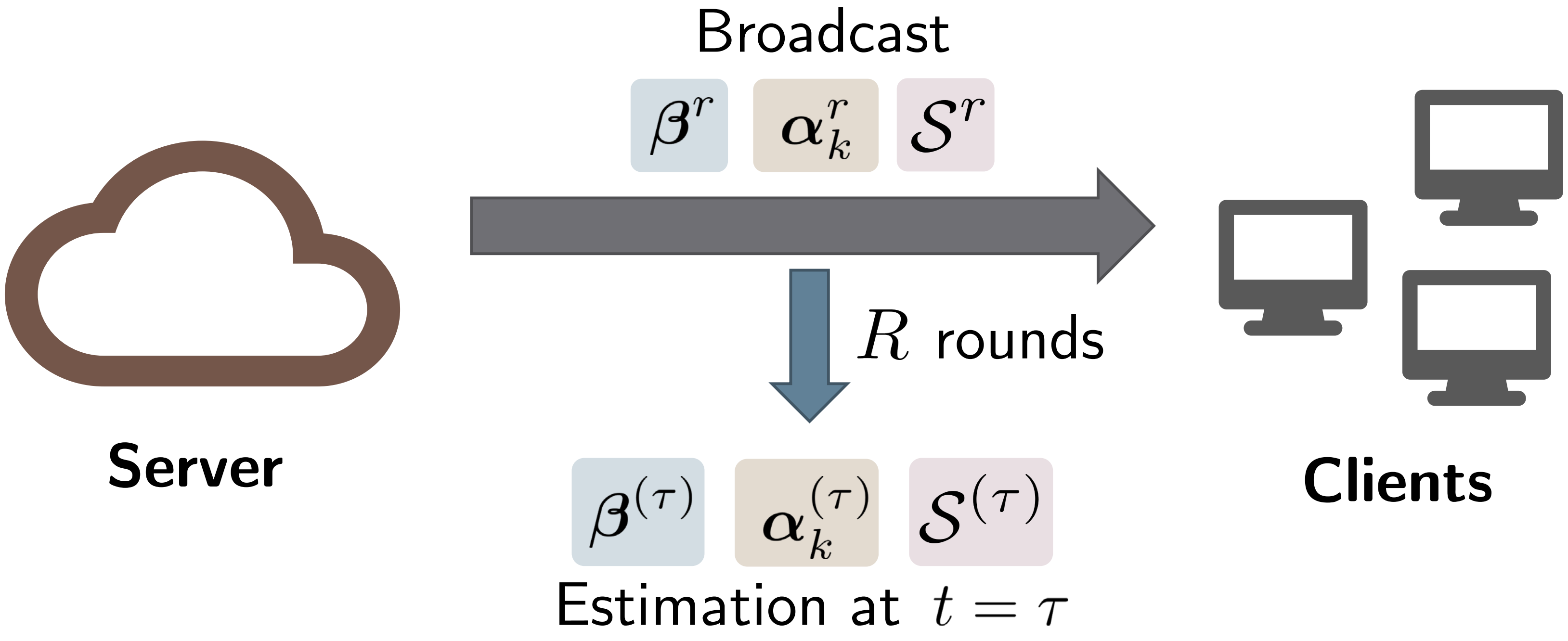
$$\boldsymbol{\beta}^r \leftarrow \sum_{i \in [M]} \frac{n_i}{N} \boldsymbol{\beta}_i^r$$

$$\boldsymbol{\alpha}_k^r \leftarrow \sum_{\mathcal{D}_i \in \mathcal{S}_k^r} \frac{n_i}{N_k} \boldsymbol{\theta}_i^r$$

Group Personalization

Proposed: Adaptive Group Personalization for Federated Mutual Transfer Learning (AdaGrP)

- Accurate learnability structure recovery in Federated framework
- Robustness against Concept Drift with Tuning-Free Solution

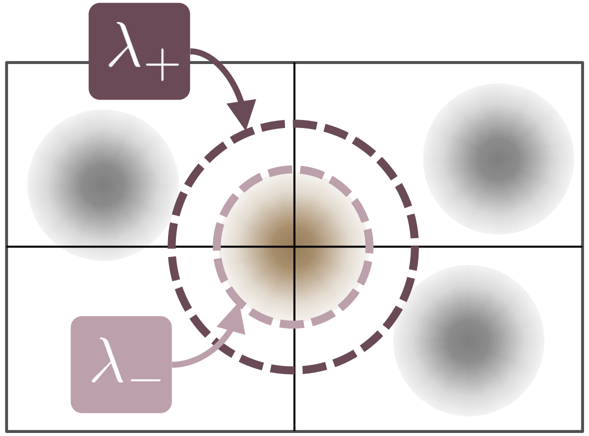


Theoretical Results

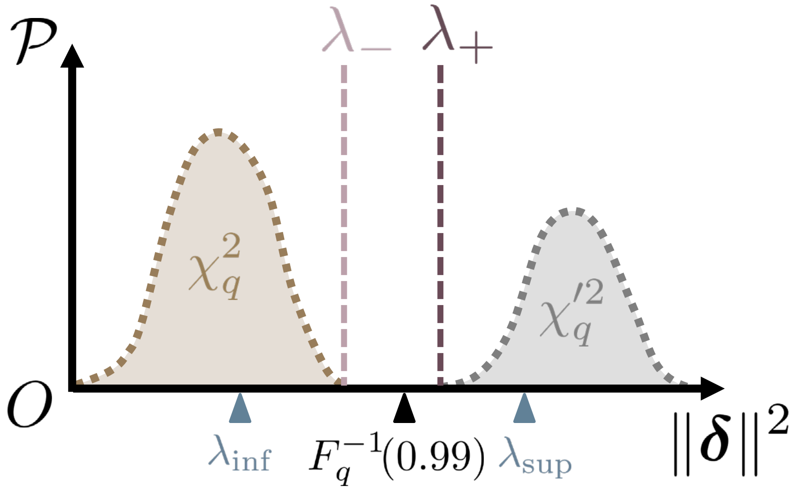
- **AdaGrP** is able to perfectly recover the learnability structure at every time step:

Theorem 4.5. *Assume Assumption 3.1 holds and* $\lambda_- < \lambda^* < \lambda_+$. *Under Condition 3, AdaGrP satisfies that* $\Psi(\boldsymbol{\theta}^r; \widetilde{\lambda}(\boldsymbol{\theta}^r)) = \Psi(\boldsymbol{\theta}^r; \lambda^*), \forall r \in [R]$. *With sufficient local updates that* $t > \frac{2\ln C_\lambda / C_M}{\ln(1-2\eta\omega)}$, $\Psi(\boldsymbol{\theta}^r; \widetilde{\lambda}(\boldsymbol{\theta}^r)) = \mathcal{S}^{(\tau)}, \forall r \in [R]$, *where* $C_\lambda = \min(\lambda_+ - \lambda^*, \lambda^* - \lambda_-)$.

- **AdaGrP** relaxes the condition of perfect recovery:



Standardized Domain Difference



DiffS (Baseline)

$\lambda_- \leq F_q^{-1}(0.99) \leq \lambda_+$

AdaGrP (Proposed)

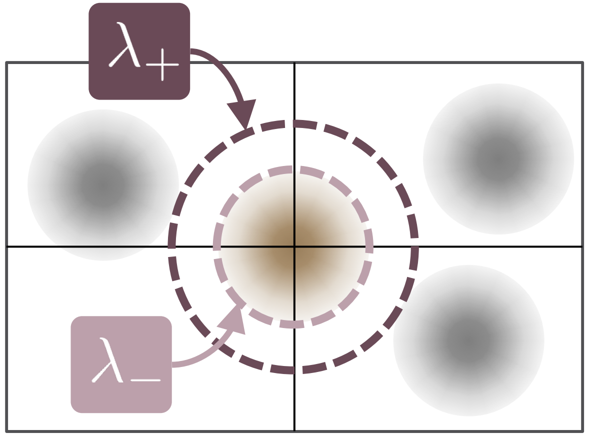
$(\lambda_-, \lambda_+) \cap (\lambda_{\text{inf}}, \lambda_{\text{sup}}) \neq \varnothing$

Theoretical Results

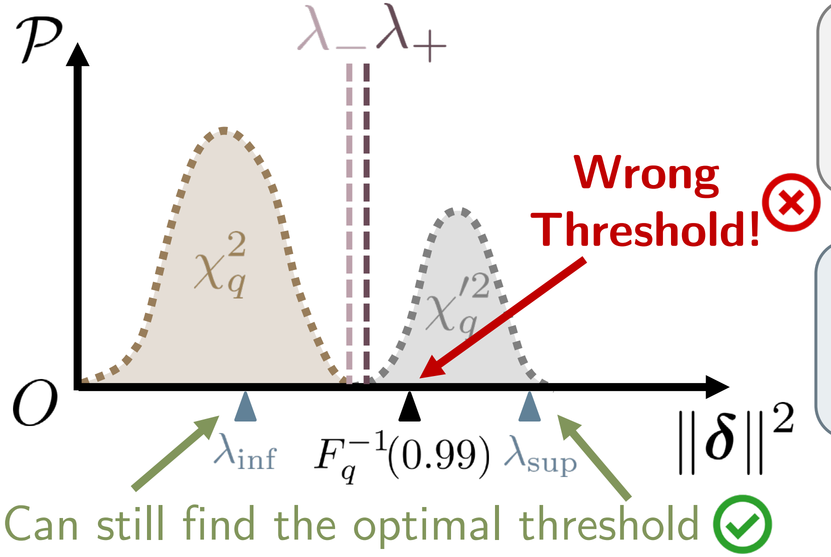
- **AdaGrP** is able to perfectly recover the learnability structure at every time step:

Theorem 4.5. *Assume Assumption 3.1 holds and $\lambda_- < \lambda^* < \lambda_+$. Under Condition 3, AdaGrP satisfies that $\Psi(\boldsymbol{\theta}^r; \widetilde{\lambda}(\boldsymbol{\theta}^r)) = \Psi(\boldsymbol{\theta}^r; \lambda^*), \forall r \in [R]$. With sufficient local updates that $t > \frac{2\ln C_\lambda / C_M}{\ln(1-2\eta\omega)}$, $\Psi(\boldsymbol{\theta}^r; \widetilde{\lambda}(\boldsymbol{\theta}^r)) = \mathcal{S}^{(\tau)}, \forall r \in [R]$, where $C_\lambda = \min(\lambda_+ - \lambda^*, \lambda^* - \lambda_-)$.*

- **AdaGrP** relaxes the condition of perfect recovery:



Standardized Domain Difference



DiffS (Baseline)

$\lambda_- \leq F_q^{-1}(0.99) \leq \lambda_+$

AdaGrP (Proposed)

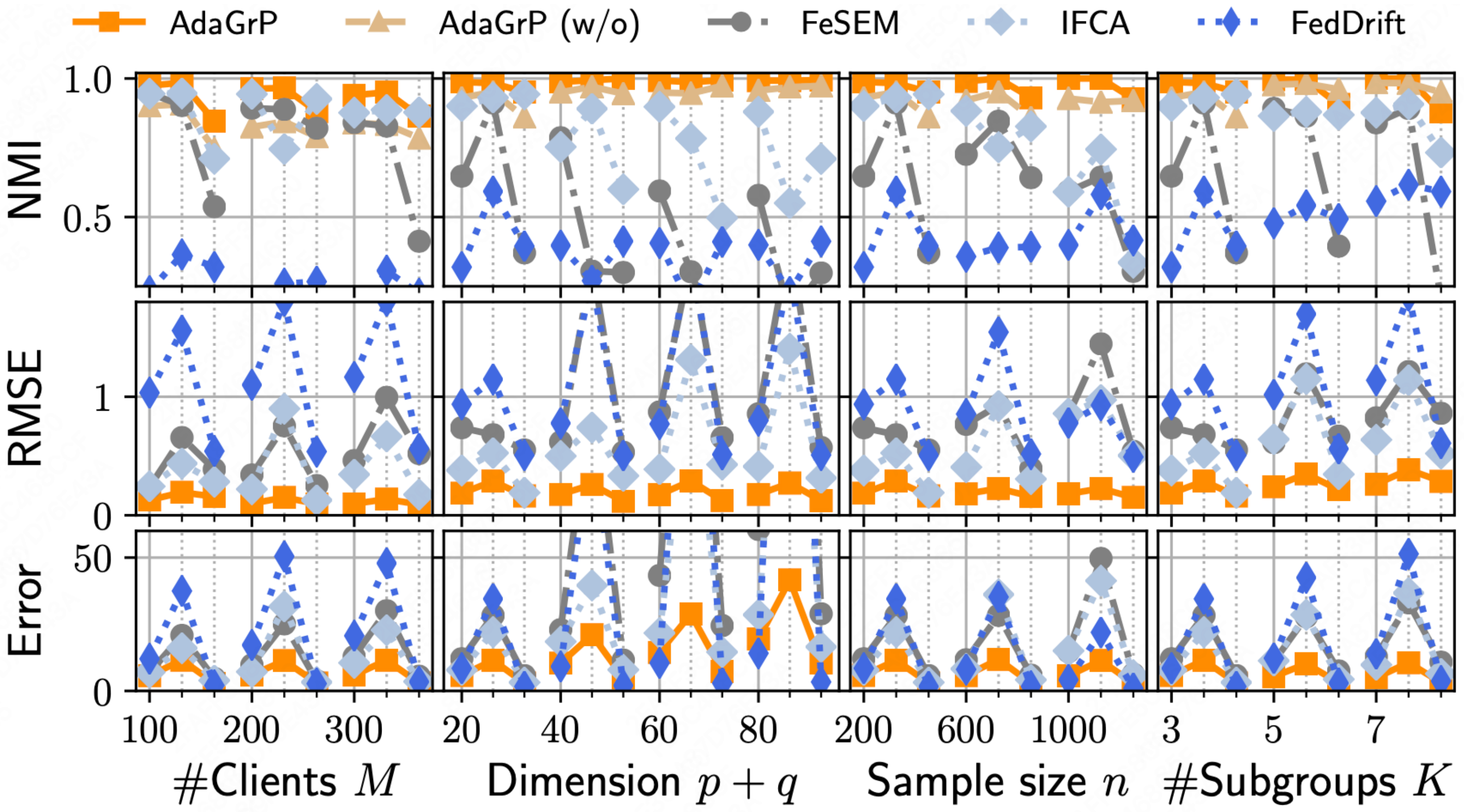
$(\lambda_-, \lambda_+) \cap (\lambda_{\inf}, \lambda_{\sup}) \neq \varnothing$

Results: Synthetic Data

Learnability Structure Recovery Performance

Parameter Estimation Error

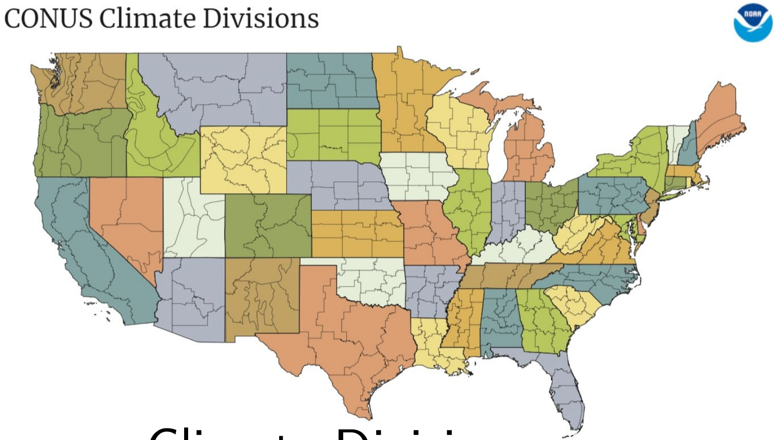
Prediction Error



- **AdaGrP** has the best estimation under concept drift environment
- **AdaGrP** achieves the most stable performance while tuning-free

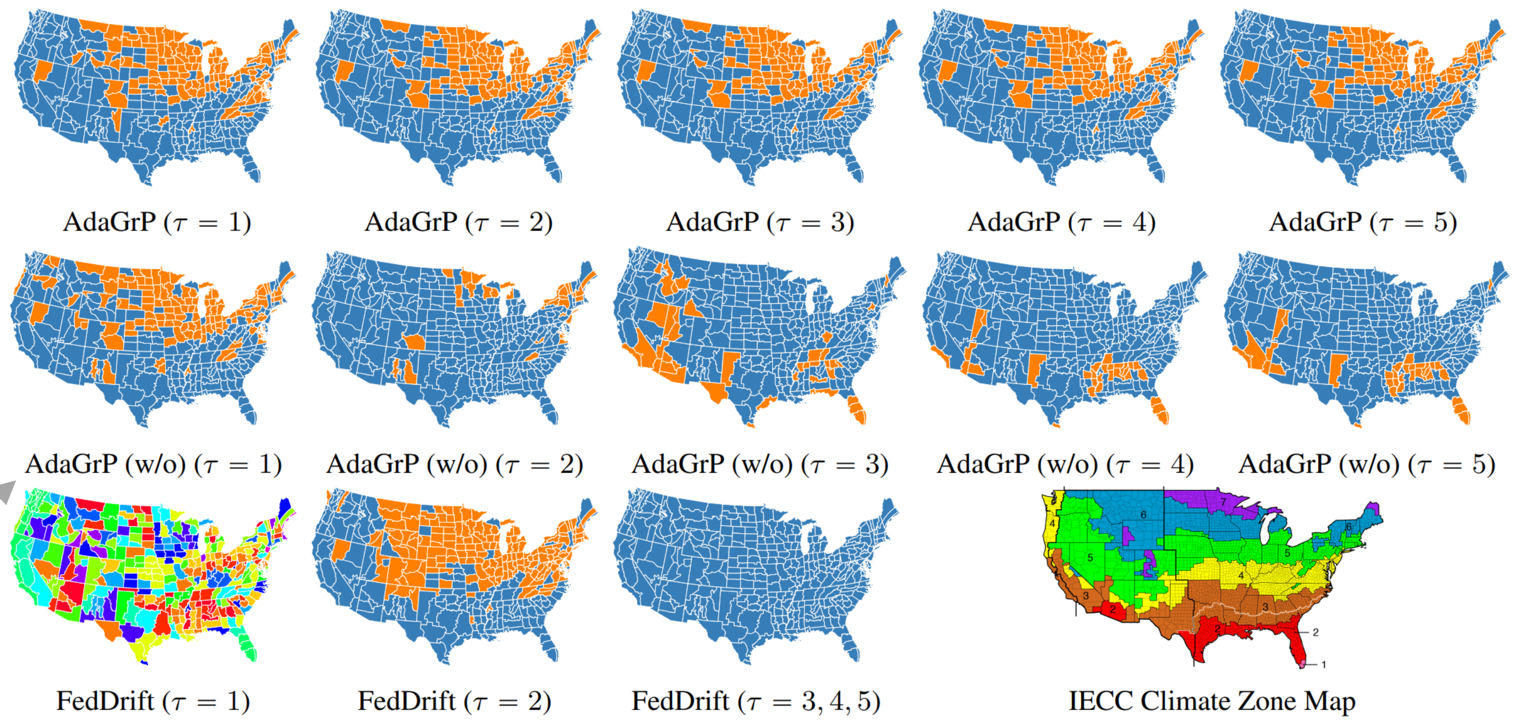
Results: NOAA nClimDiv Database[3]

- Real-world task: NOAA nClimDiv Temperature Prediction
  - Data from 344 domains, 25 years per time step, 5 time step in total
- **AdaGrP** estimates more properly based on IECC Climate Zone
  - FedDrift: Sensitive Hyper-parameters & Bad Learnability Recovery



Climate Divisions (Domains) on Map

AdaGrP (w/o): without Threshold Correction



[3] Vose, Russell S., et al. "NOAA Monthly US Climate Divisional Database (NClimDiv)." NOAA National Climatic Data Center, 2014. doi:10.7289/V5M32STR

Thank you for your attention!

- **Haoqing Xu** (xuhaoqing@seu.edu.com)
- Dian Shen
- Meng Wang
- Beilun Wang* (beilun@seu.edu.cn)

- Poster: Hall C 4-9 #2210 Wed 24 Jul 11:30 a.m. — 1 p.m.
  - icml.cc/virtual/2024/poster/34610

東南大學
SOUTHEAST UNIVERSITY

ICML
International Conference
On Machine Learning