# Spatiotemporal Predictive Modeling for Climate Science: Are we there yet?

Arindam Banerjee
*Dept of Computer Science & Engineering*
*University of Minnesota, Twin Cities*
banerjee@cs.umn.edu

Expeditions Workshop, August 4-5, 2015

# Spatiotemporal Predictive Modeling for Climate

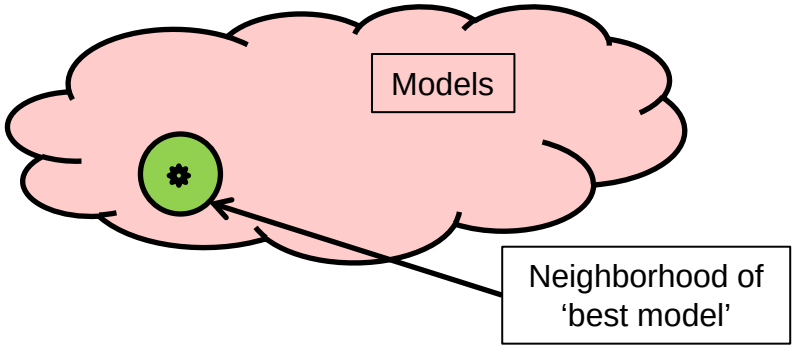
- Model phenomenon y using ‘predictors’ x
  - Example: y is total Indian summer monsoon rainfall (ISMR)
- ‘Predictors’ x: numerous spatial and temporal features
  - Highly correlated data, temporal lags, spatial teleconnections
- Mechanistic understanding: climate processes as predictors
  - Representation of processes: climate indices, spectral information
- Multiple mechanisms: Occam and Murphy
  - Different models for different phases of the climate system

# Small Sample Regime and Stability

- Modeling complex phenomenon with small samples
  - MPU*: High accuracy, limited stability
  - Stability needs to be a priority
  - Iterative Occam: refine stable, moderately accurate models


*minimum publishable unit

- ‘Proving’ stability: Uniform bounds for model class
  - How many samples are sufficient?
  - Explicit modeling assumptions



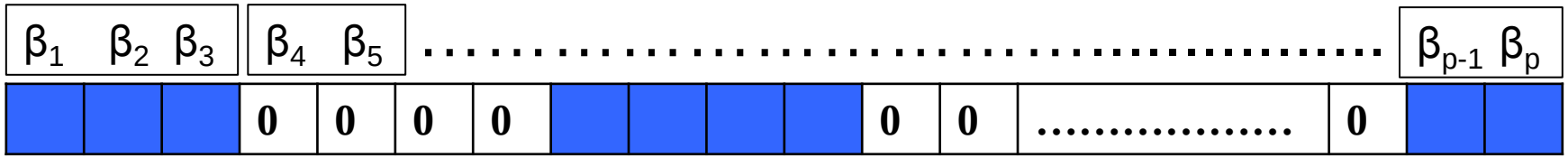

- ‘Testing’ stability: Robustness
  - Are the results stable?
  - Potentially more general scope

# Small Sample Regime: Bias, with Regularization

- General regularized regression (Banerjee et al., 2014)
  - E.g., sparsity, group-sparsity, hierarchy, ridge, low-rank

$$\min_{\beta \in R^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_n R(\beta) \right\}$$

$R(\beta)$ : bias structure of $\beta$ (e.g., based on domain knowledge)



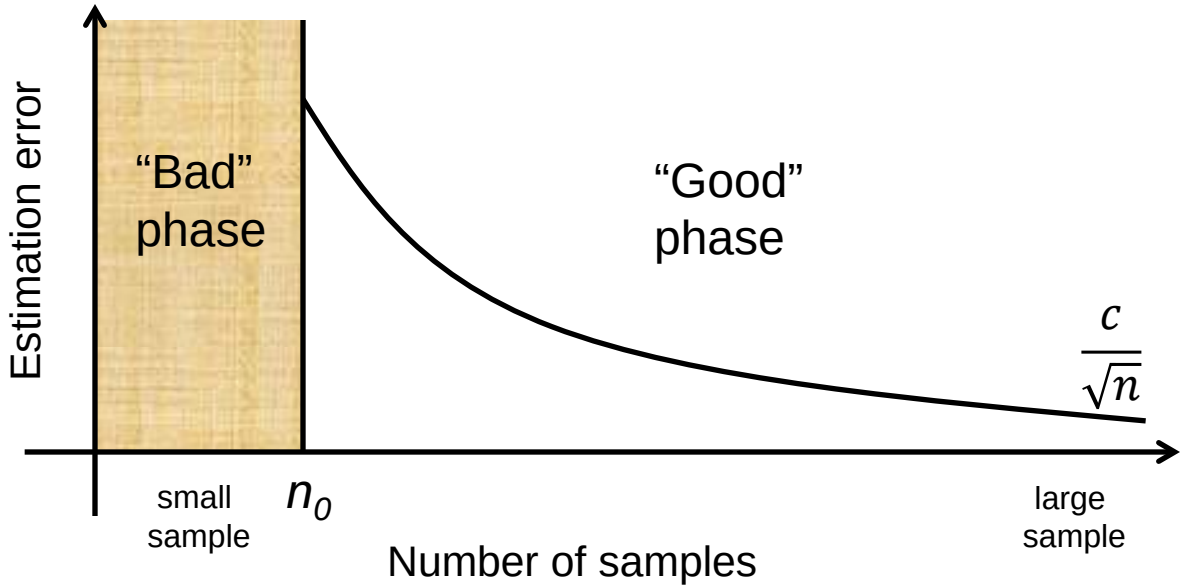

Group sparsity over linear regression coefficients

- Performance guarantees for regularized regression
  - How many samples are needed to accurately estimate β?
  - What is the rate of convergence?

# Why Should I Care?

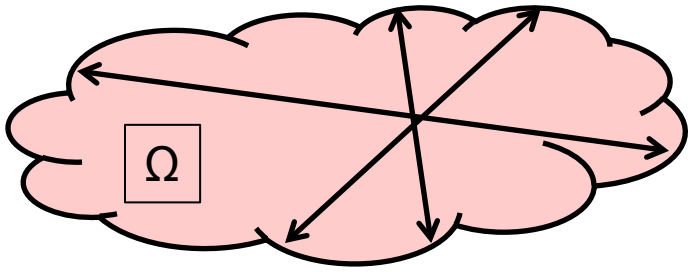
- Estimation in simple vs structured problems
  - Example: (a) mean of samples $\{x_i\}$ vs. (b) Lasso on samples $\{(y_i,x_i)\}$

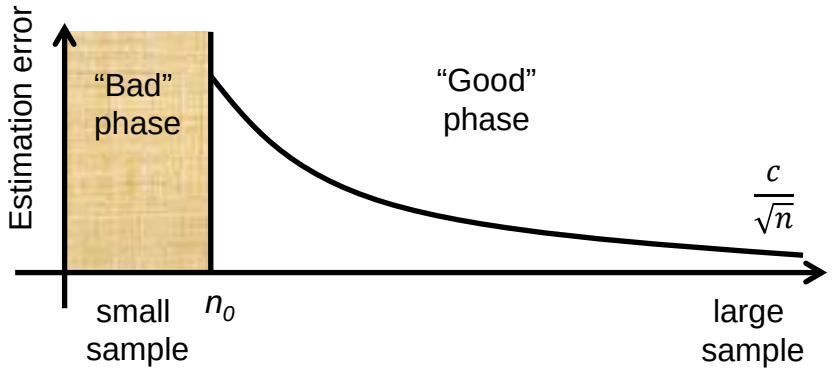
- Structured problem with $n$ samples
  - "Bad" phase: $n < n_0$, **<u>do not</u>** trust the estimate
  - "Good" phase: $n \geq n_0$, estimate is good, error decreases as $\frac{c}{\sqrt{n}}$



# Gaussian “width” of sets

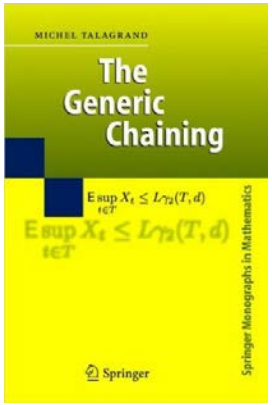
Gaussian “width” of sets




$$\min_{\beta \in R^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_n R(\beta) \right\}$$

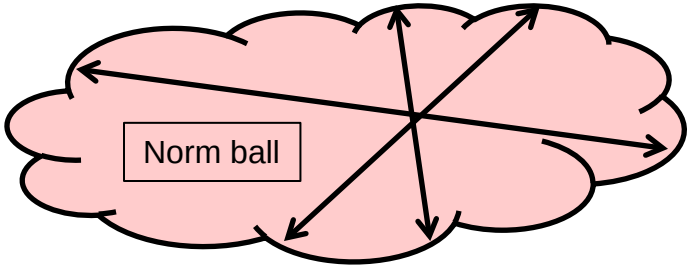
Supremum of (geometric) Gaussian process, indexed by $u \in \Omega$

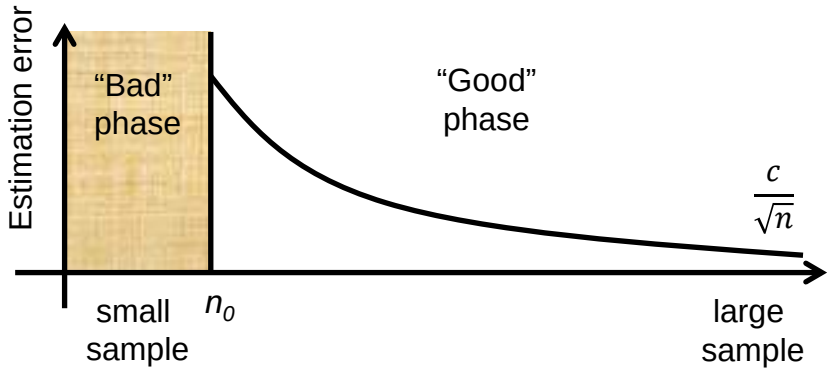



“What is the maximum level a certain river is likely to reach over the next 25 years? (Having experienced three times a few feet of water in my house, I feel a keen personal interest in this question.) …”

# Structured Models: Rate of convergence

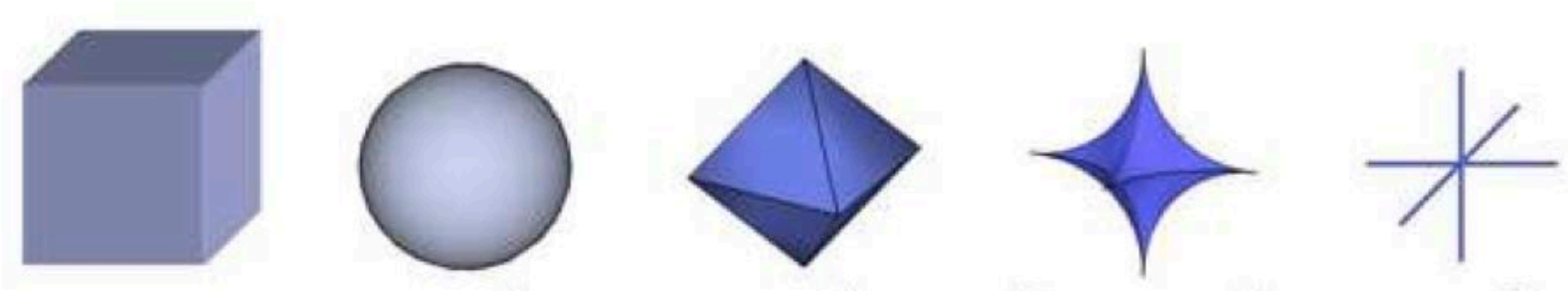
Gaussian width or "width" of sets




$$\min_{\beta \in R^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_n R(\beta) \right\}$$

*c* depends on "width" of norm ball, {*u* | *R(u)* ≤ *1*}



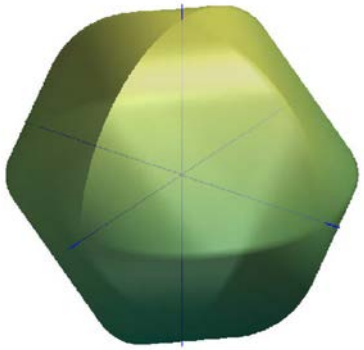
$L_\infty$ $L_2$ $L_1$ $L_q$, q ϵ(0,1) $L_0$

Banerjee et al., 2014

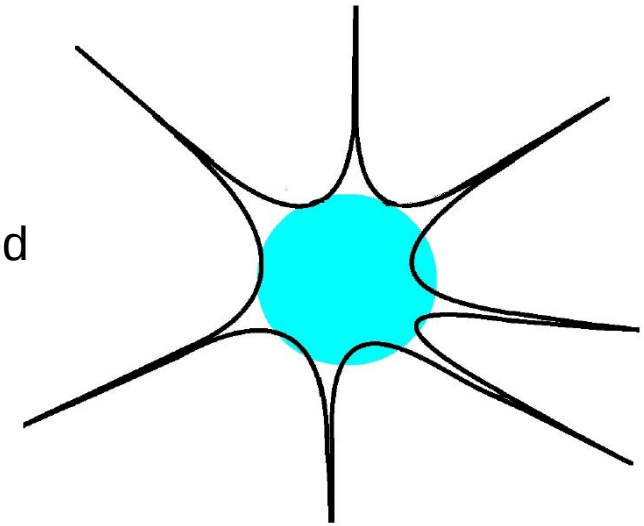
# Structured Convexity in High-Dimensions

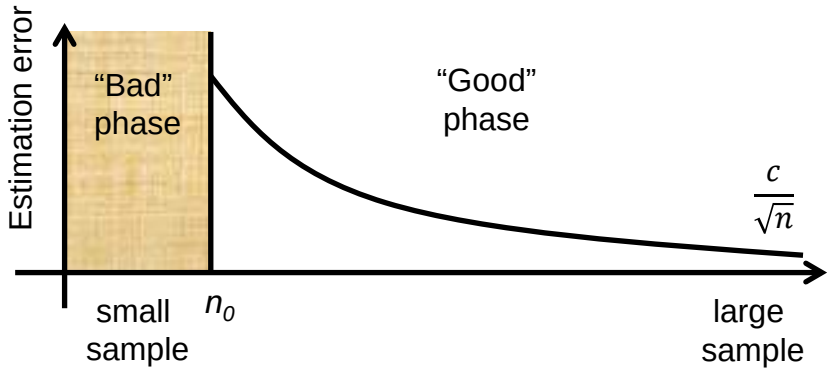
*c* depends on "width" of norm ball

Low-d



High-d





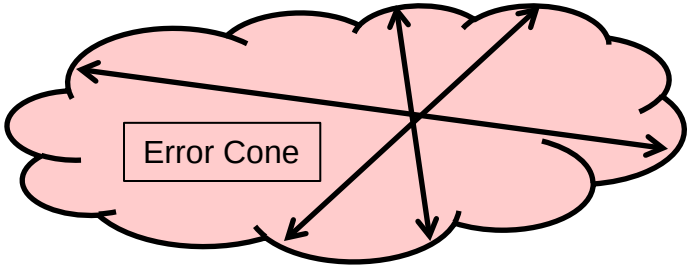

$$\min_{\beta \in R^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_n R(\beta) \right\}$$

- Gaussian widths
  - For $L_1$ ball: $O(\sqrt{\log p})$
  - For $L_2$ ball: $O(\sqrt{p})$

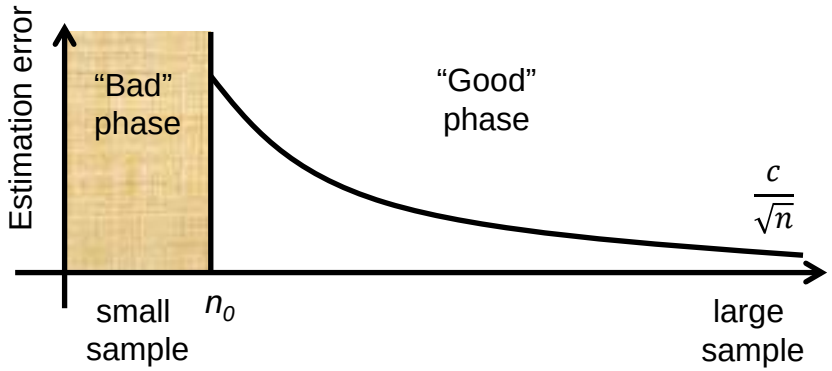
Milman, 1998, Vershynin, 2014

# Structured Models: Sample Complexity

Gaussian width or "width" of sets



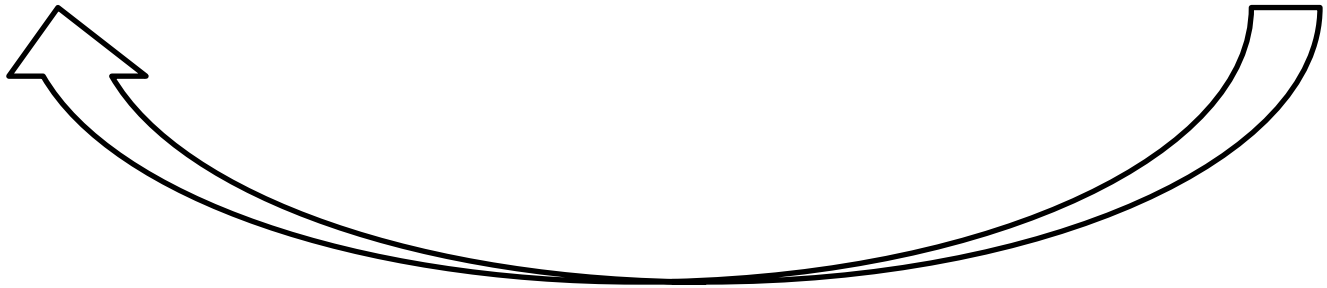

$n_0$ depends on "width" of error cone




$$\min_{\beta \in R^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda_n R(\beta) \right\}$$

Structured models have smaller error cones

Domain knowledge ⇒ Structured model ⇒ Stable model with less samples

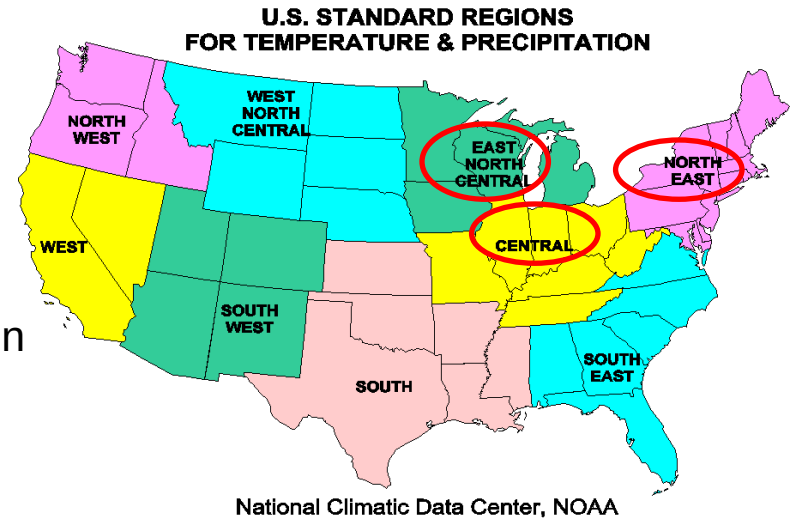


Bickel et al., 2009, Chandrashekaran et al., 2012, Negahban et al., 2012, Banerjee et al., 2014, Tropp, 2015

# Great Lakes Precipitation

Predictors:

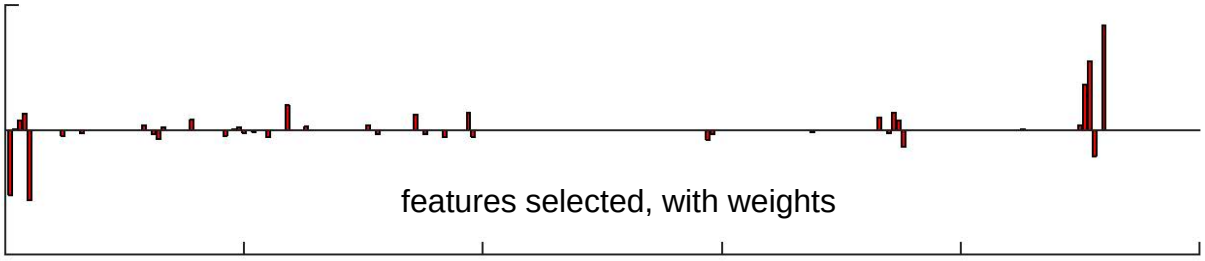
- Climate indices (long range), e.g., NAO monthly
- Atmospheric variables (station level, regional)



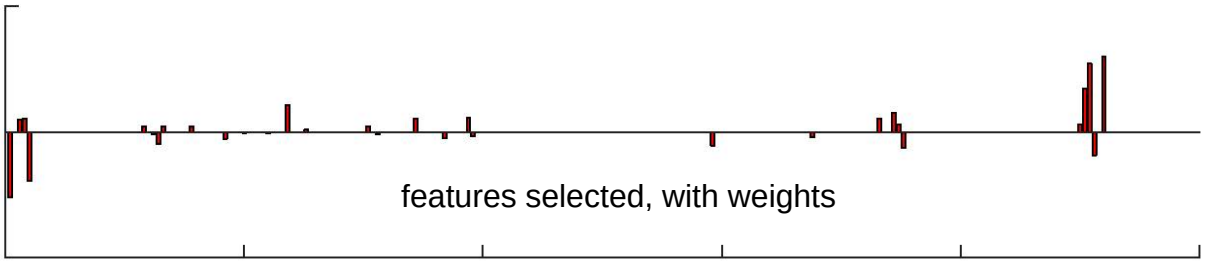

Stable sparse estimation: 2 levels of feature selection

- Lasso: $L_1$-norm regularized linear regression
- Stability test: Randomly permuting y, fixed x

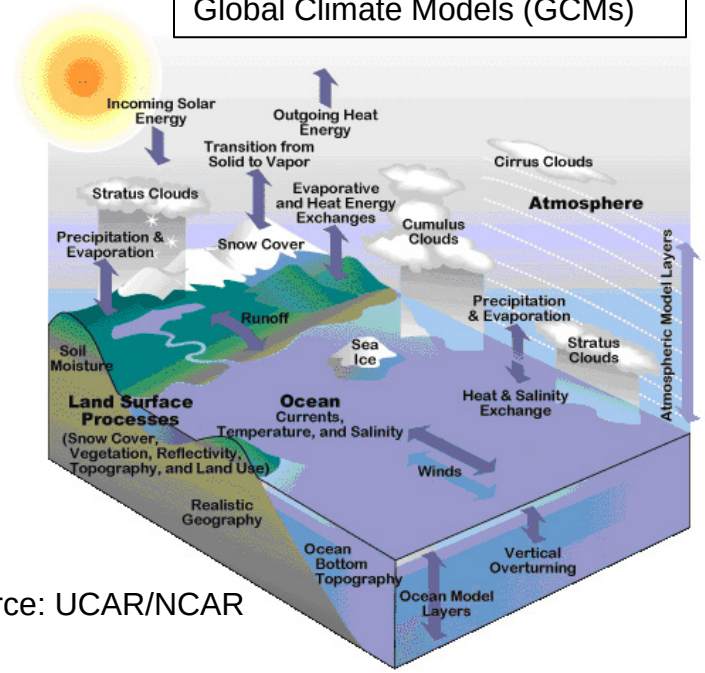
$\lambda = 1.0$




$\lambda = 10.0$

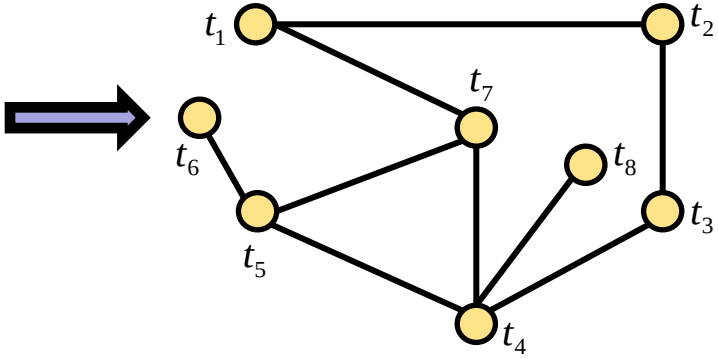


# Combing Global Climate Model Outputs




Source: UCAR/NCAR

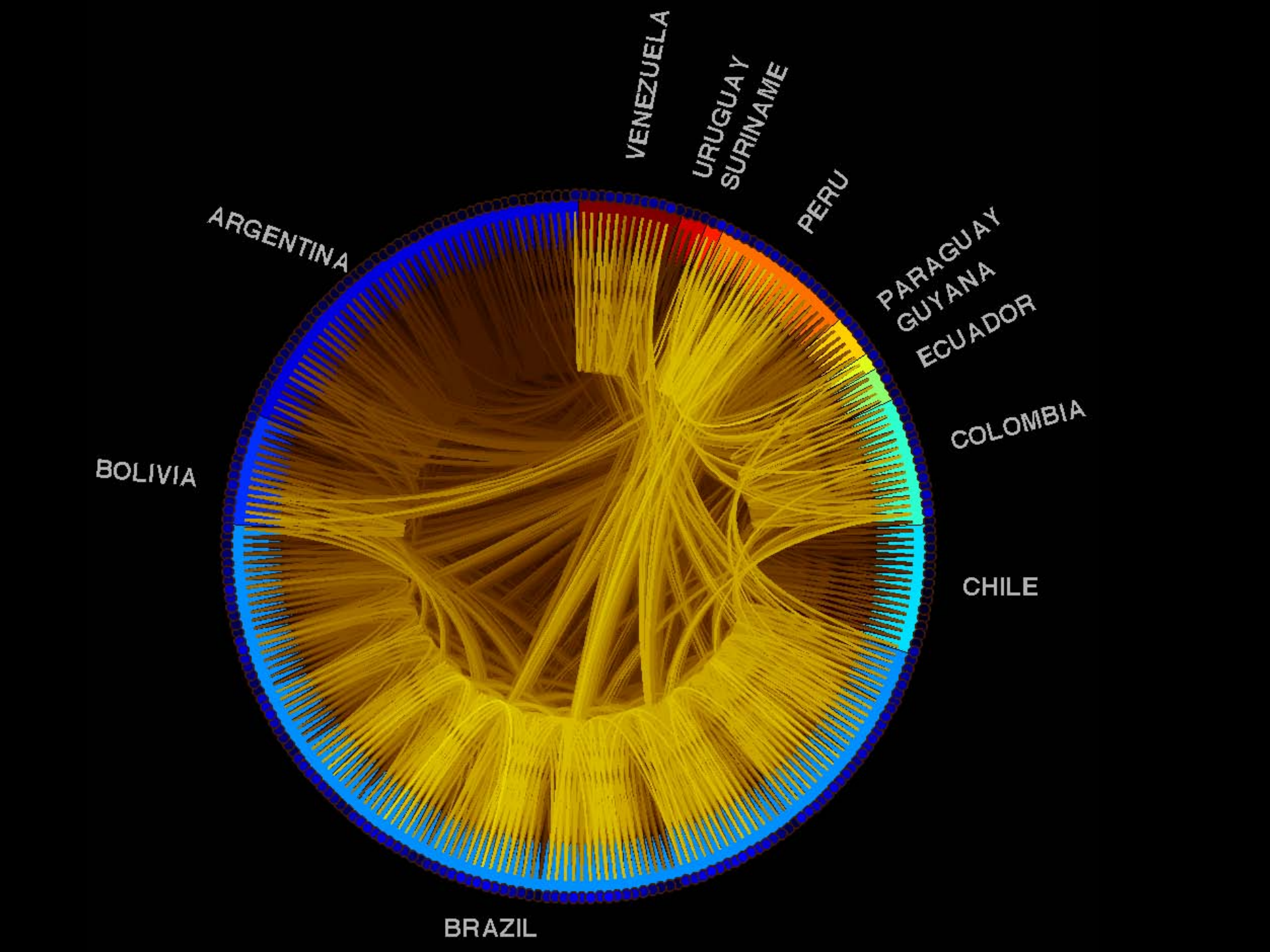
Tasks: South American regional temperature



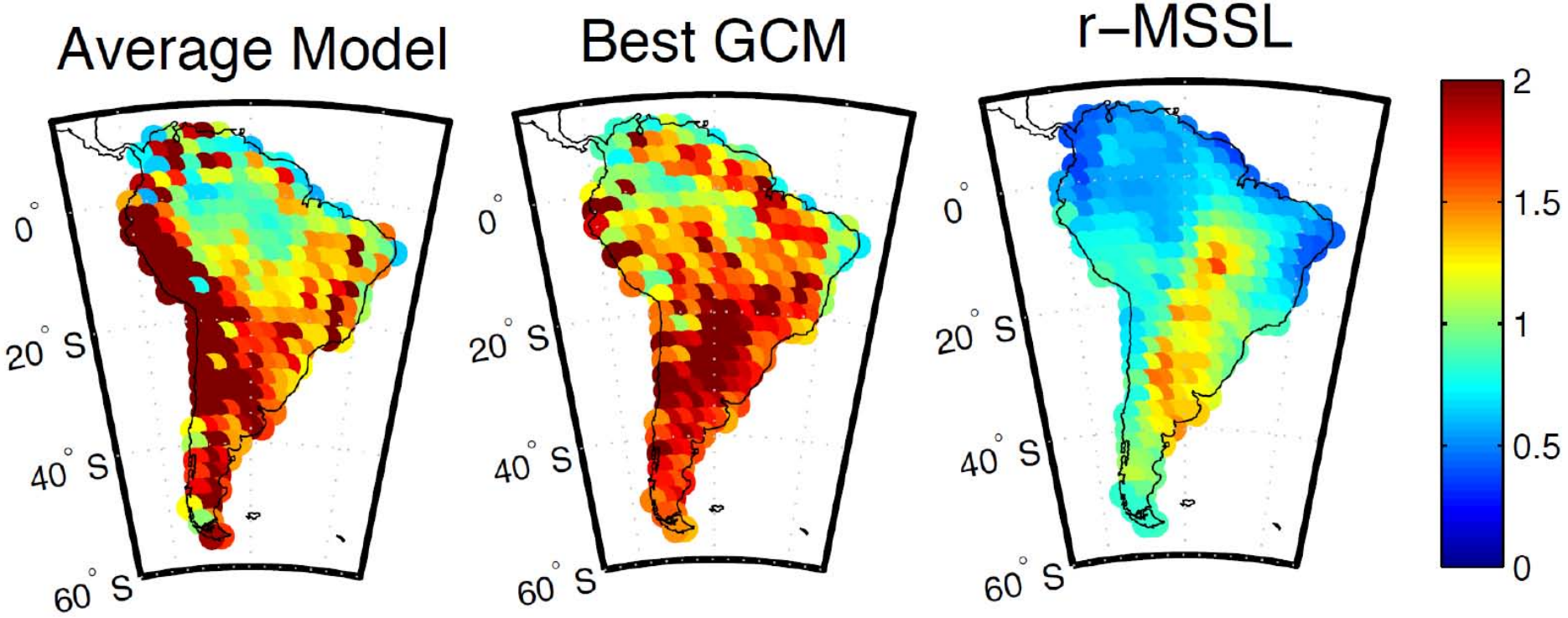

Dependency between different regions (tasks)

Combining GCM outputs as multi-task learning (Goncalves et al., 2014)

- Tasks: Climate model weights for a process, variable, or region
- Task based regularization
  - Model weights on related tasks should be similar
- Several alternatives as baselines
  - Model average (IPCC), 'Best' GCM, etc.


VENEZUELA
URUGUAY
SURINAME
PERU
PARAGUAY
GUYANA
ECUADOR
COLOMBIA
CHILE
BRAZIL
BOLIVIA
ARGENTINA

# RMSE Comparison: Multi-task vs Baselines




| Average Model | Best GCM | r–MSSL |
|---|---|---|
| **1.621 (±0.020)** | **1.410 (±0.037)** | **0.780 (±0.039)** |

# Latent Variable Models: Iterative Occam

- Predictive modeling, data $\{(x_i,y_i)\}$
  - Model with no latent variables, e.g., single linear regression (SLR)

$$\min_{\beta\in\mathbb{R}^p} \frac{1}{2n}\sum_{i=1}^{n}(y_i - \langle x_i,\beta\rangle)^2 + \lambda_n\|\beta\|_2^2$$

  - Model with latent variables, e.g., mixture of linear regression (MLR)

$$\min_{\substack{z_1,\ldots,z_n\in\{0,1\}\\ \beta_1,\beta_2\in\mathbb{R}^p}} \frac{1}{2n}\sum_{i=1}^{n}(y_i - \langle x_i,(z_i\beta_1 + \bar{z}_i\beta_2)\rangle)^2 + \lambda_n(\|\beta_1\|_2^2 + \|\beta_2\|_2^2)$$

- When will MLR succeed?
  - Multiple mechanisms responsible for y
  - At any given point, one mechanism is active/dominant

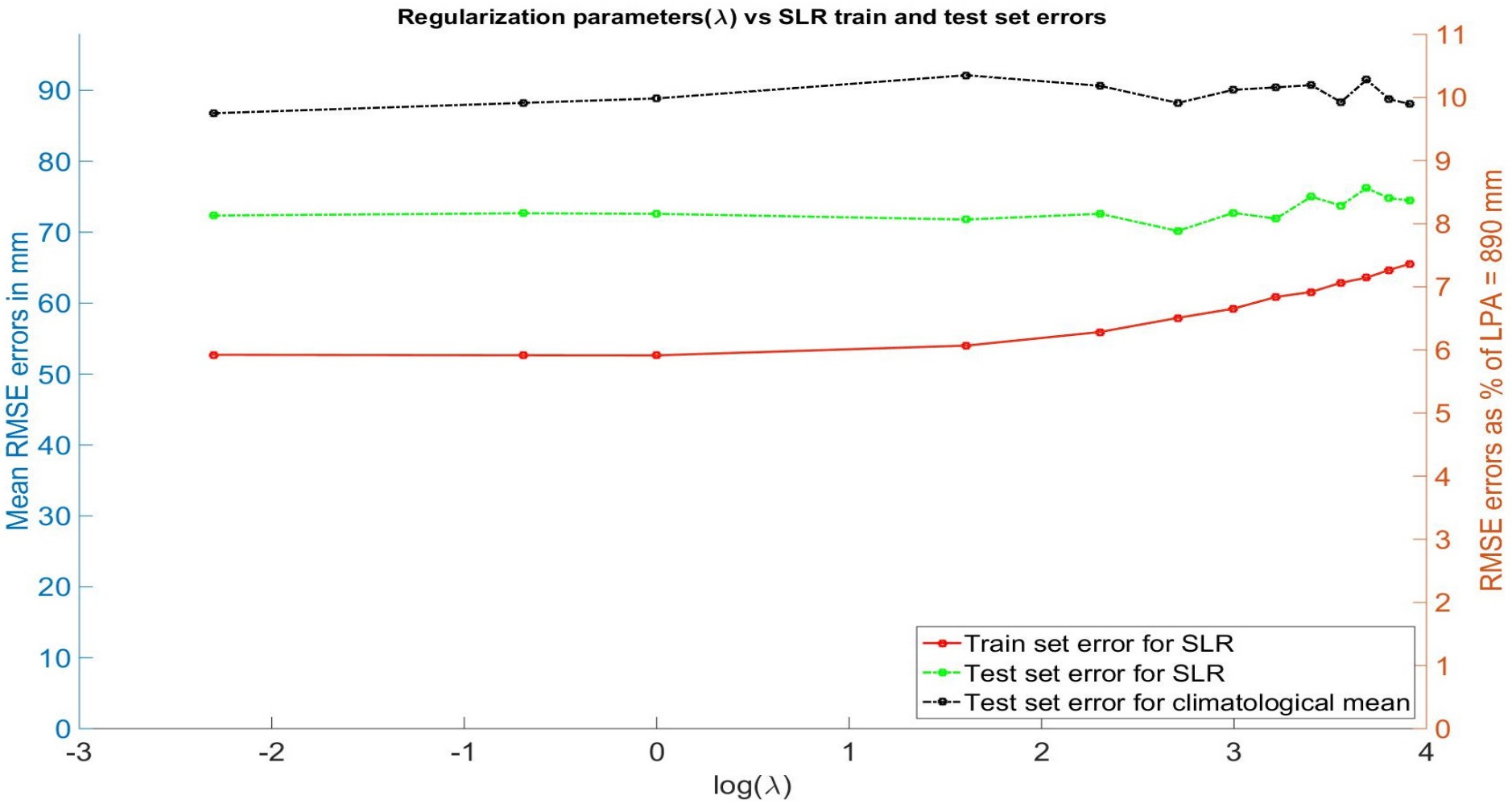
# Indian Summer Monsoon Rainfall (ISMR)

- Predicting total ISMR
  - Long period average (LPA): 890 mm, variation within 10% of LPA
  - Gowariker (1991), DelSole and Shukla (2002, 2006), Rajeevan et al. (2006), Saha et al. (2015)

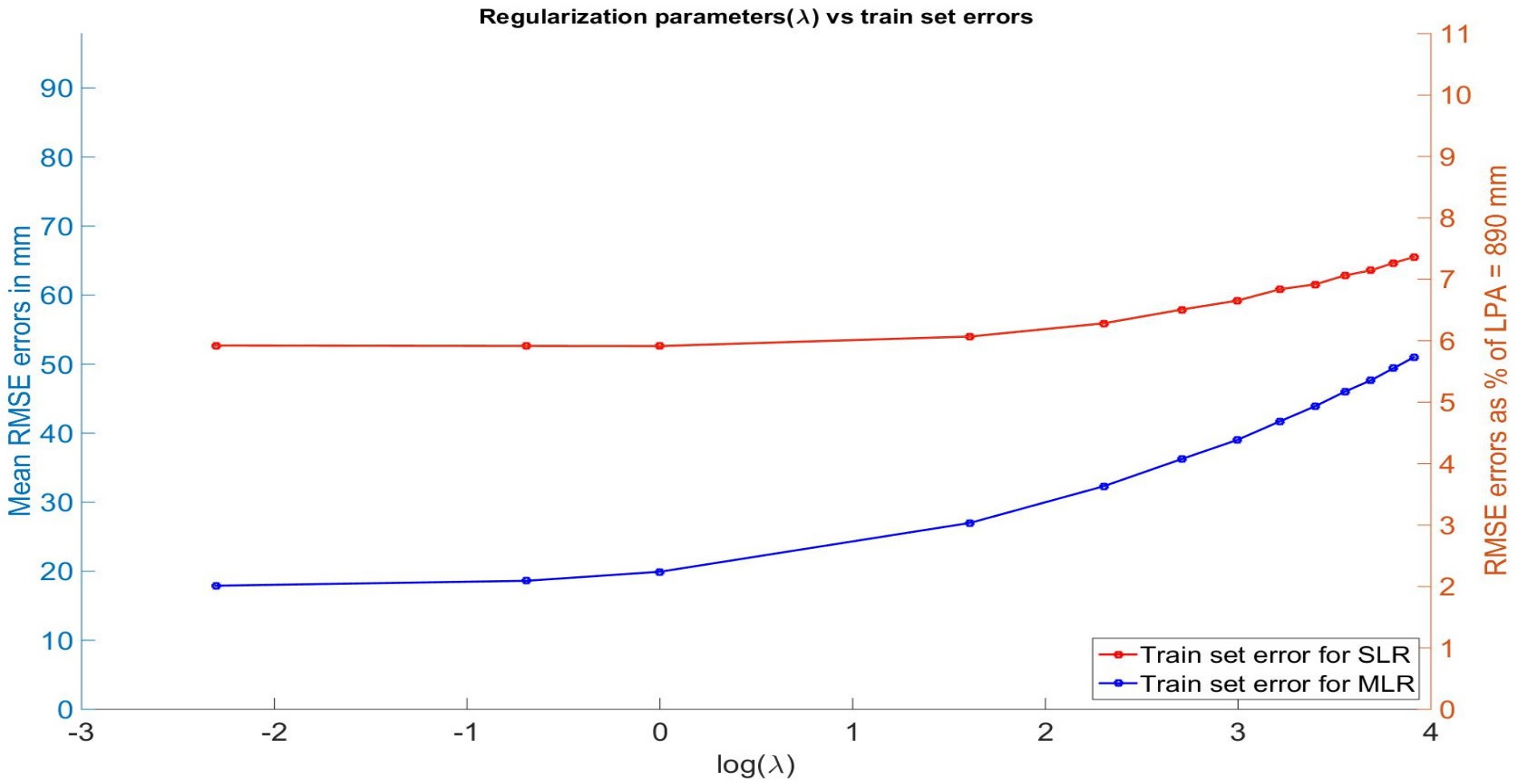
- Dataset: 66 years, 1948-2013, covariates from Saha et al. (2015)

- Methodology, repeated 200 times

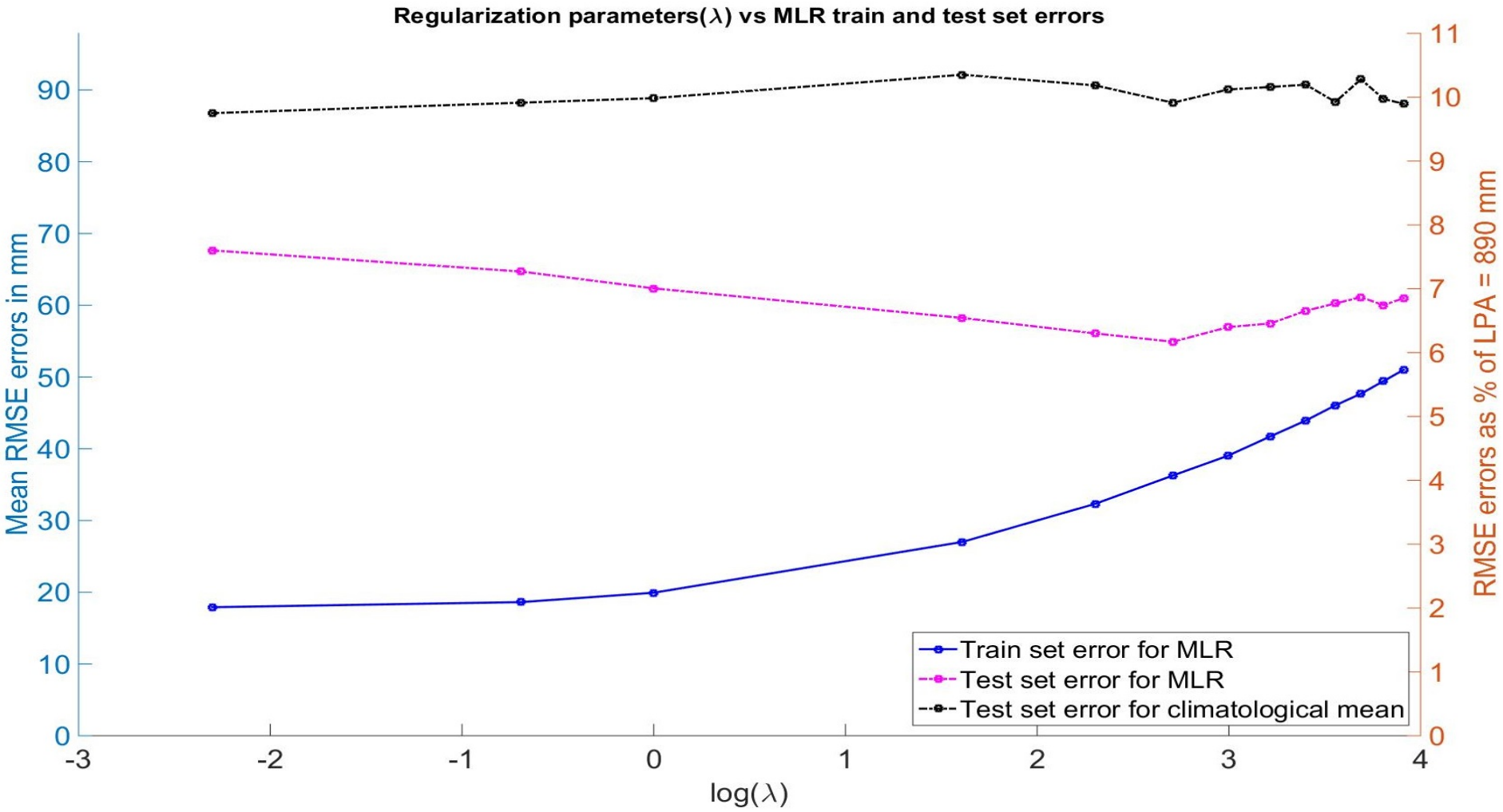
# ISMR Prediction: SLR, train vs test



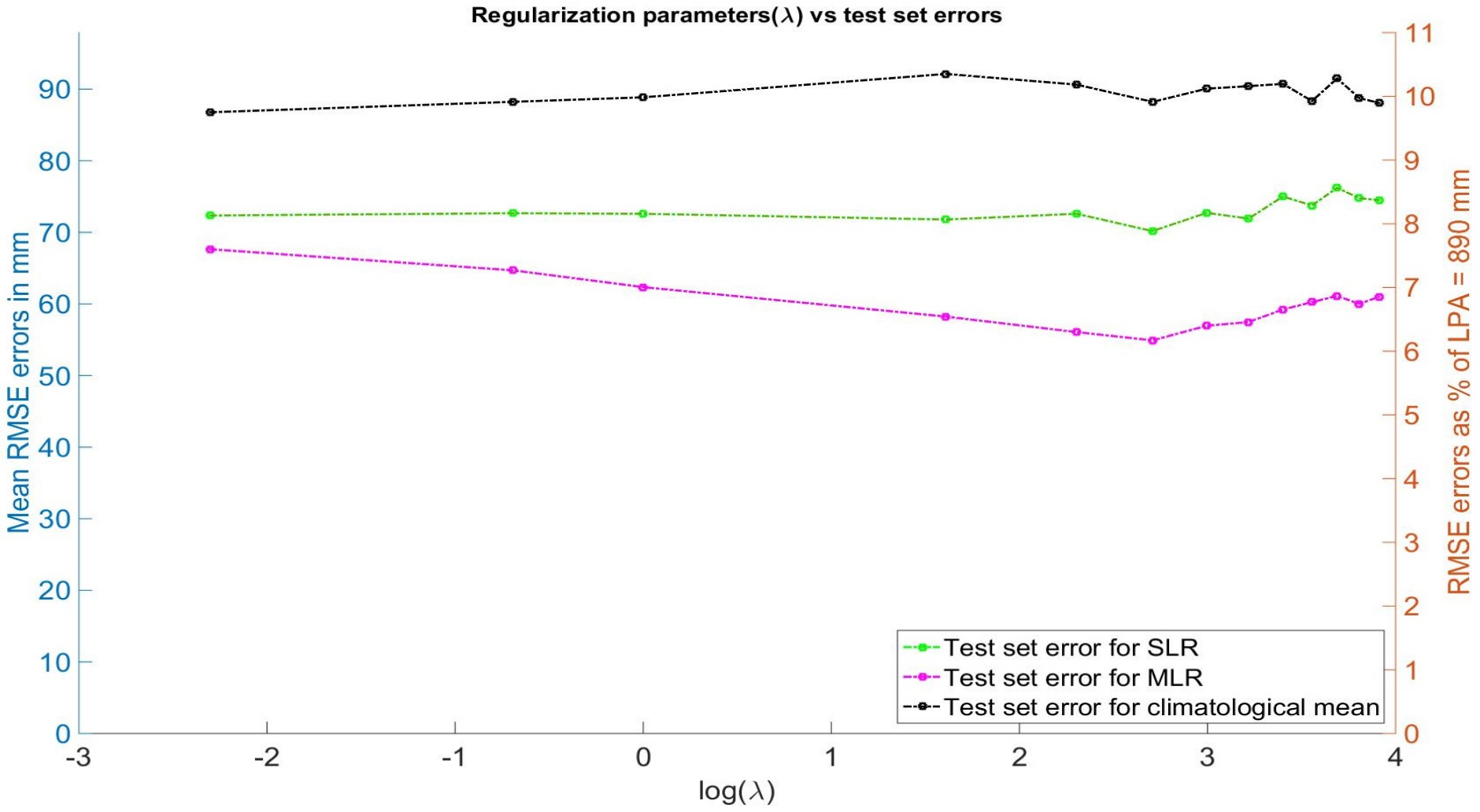
# ISMR Prediction: SLR vs MLR, train



# ISMR Prediction: MLR, train vs test



# ISMR Prediction: SLR vs MLR, test



# Are We There Yet?

- Identifying nonlinear multivariate dependencies
  - "Simple" nonlinearities: monotonic dependencies
  - "Complex" nonlinearities: combination of simple n… …es
- Complex processes and inter…
  - Oscillations and pha… …ependencies in each phase
  - Temporal l… …echanisms, long memory processes
  - S… …ects
  - …al variability
- Causal discovery
  - Mechanistic understanding, based on climate processes

Far from it … but we have a good start

# Conclusions

- Modeling complex phenomenon with small samples
  - Stability needs to be a priority
  - Iterative Occam: refine stable, moderately accurate models
- Domain knowledge helps build structured models
  - Structured models are stable with small samples
  - Identifies dominant factors, improves understanding
- Multiple mechanisms: Latent variable models
  - Different models for different phases of the climate system
- Better approaches for proving/testing stability

# Acknowledgements


- Students and Collaborators
  - Soumyadeep Chatterjee, Andre Goncalves, Vidyashankar Sivakumar
  - Puja Das, Farideh Fazayeli, Qiang Fu, Karthik Subbian, Huahua Wang
  - Tim DelSole and Claire Monteleoni
  - Vipin Kumar and Auroop Ganguly




# References


- A. Banerjee, S. Chen, F. Fazayeli, V. Sivakumar, Estimation with Norm Regularization, *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- S. Chatterjee, S. Chen, A. Banerjee, Generalized Dantzig Selector: Application to the k-support norm, *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- A. Goncalves, P. Das, S. Chatterjee, V. Sivakumar, F. Von Zuben, A. Banerjee, Multi-task Sparse Structure Learning, *Conference on Information and Knowledge Management (CIKM)*, 2014.
- H. Wang, F. Fazayeli, S. Chatterjee, and A. Banerjee, Gaussian Copula Precision Estimation with Missing Values. *Artificial Intelligence and Statistics (AISTATS)*, 2014.
- K. Subbian and A. Banerjee, Climate Multi-model Regression Using Spatial Smoothing, *SIAM Data Mining (SDM)*, 2013.
- S. Chatterjee, K. Steinhaeuser, A. Banerjee, S. Chatterjee, and A. Ganguly, Sparse Group Lasso: Consistency and Climate Applications, *SIAM Data Mining (SDM)*, 2012.