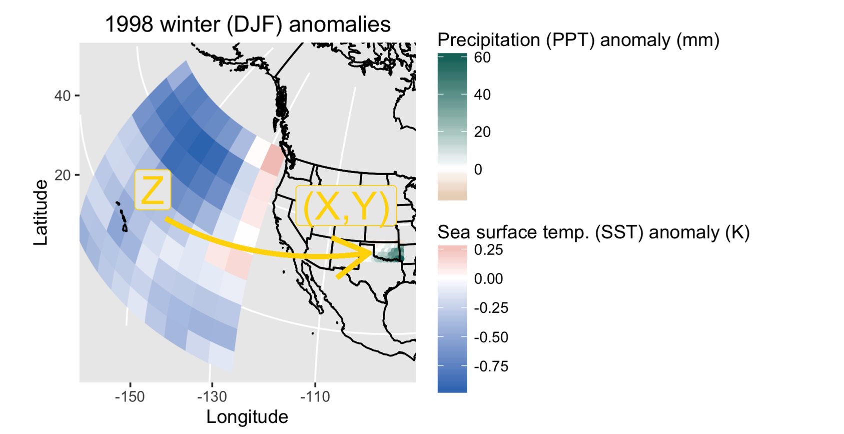
A Spatial-statistical approach to modeling teleconnections

Joshua Hewitt[1], Jennifer A. Hoeting[1], James Done[2], and Erin Towler[2]
[1]Department of Statistics, Colorado State University; [2]National Center for Atmospheric Research

Overview

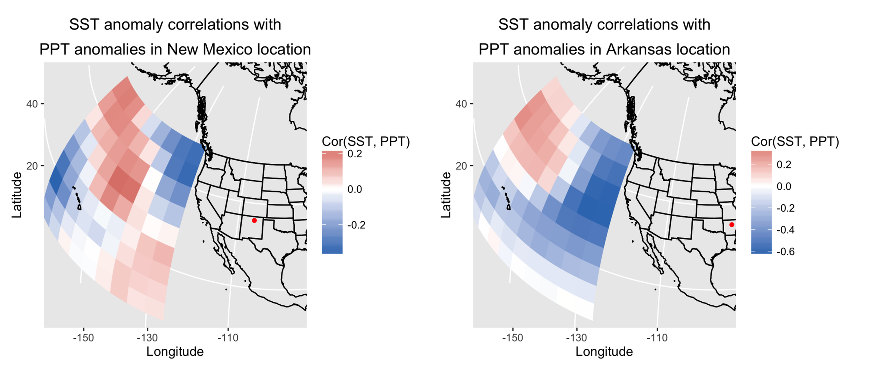
We develop a regional geostatistical model to study teleconnections—a climate phenomenon in which geographically distant areas influence regional climate patterns—at decadal time scales. Such a model could ultimately help regional planners use climate forecasts to study and prepare for local impacts of climate variability and change.



Teleconnections occur when remote covariates, like Pacific Ocean sea surface temperature anomalies, influence regional climate variables, like average Winter precipitation anomalies.

Regional teleconnections

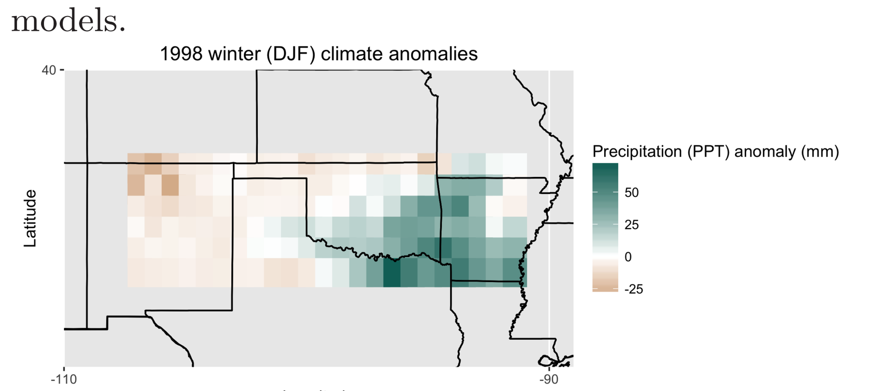
Data: (1981-2013) Average winter (DJF) land and sea surface temperature anomalies from ERA-Interim Reanalysis data; PRISM precipitation anomalies (total rain and melted snow).



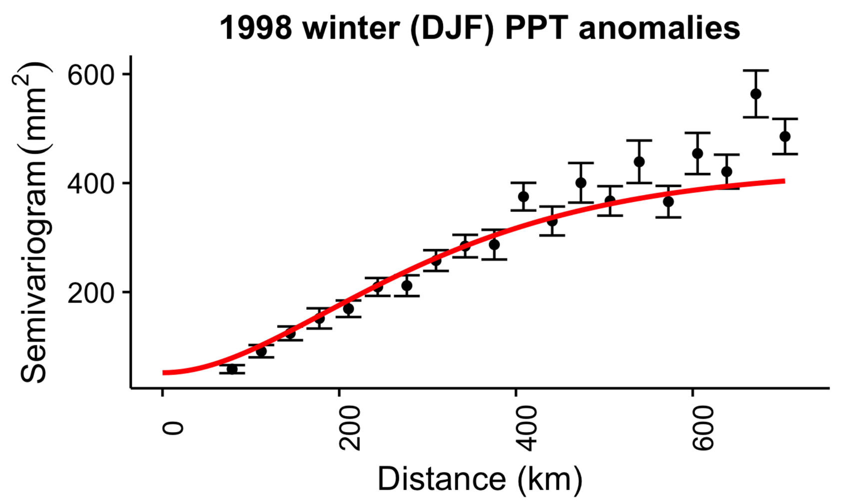
Pointwise correlation maps offer one way to study teleconnection effects. These maps highlight how teleconnection effects vary regionally. Teleconnection models should account for this behavior.

Spatial statistics

Spatial models extend linear regression techniques so we can use statistical models to estimate and test empirical relationships while accounting for spatial correlations in data. While CCA and EOF analyses estimate spatial patterns in data, they do not provide the same modeling flexibility and ability to test relationships as spatial models.



PPT anomalies display positive spatial correlations—a phenomena in which a variable's value at any location tends to be similar to its neighbors.



The semivariogram, $\gamma(d)$, quantifies spatial correlation by measuring a variable's variance as a function of distance, thereby quantifying how similar nearby locations tend to be. Our model uses the Matèrn correlation function with $\nu = 2$—a smoothly decaying spatial correlation function—to model this correlation in red.

$$\gamma(d) = \mathrm{Var}\left(Y(s_1) - Y(s_2)\right) \text{ for } \|s_1 - s_2\| = d$$

New spatial teleconnection modeling approach

We model teleconnection at each mainland location, s, and time, t, with a simple, flexible form

$$\underbrace{y(s,t)}_{\text{PPT anomaly}} = \underbrace{\sum_{i=0}^{p} \beta_i x_i(s,t)}_{\text{Local effects}} + \underbrace{\sum_{r=1}^{n_r} \alpha(r,s) z(r,t)}_{\text{Teleconnection effects}} + \underbrace{w(s,t)}_{\text{Random error}}$$

Data

Mainland:
- s — location
- $y(s,t)$ — PPT anomaly
- $x_1(s,t)$ — temp. anomaly
- $x_2(s,t)$ — longitude

Ocean:
- r — location
- $z(r,t)$ — SST anomaly

Effects

Mainland: β_i — regression coefficients; non-linear effects could be modeled with splines

Ocean: $\alpha(r,s)$ — teleconnection effect of ocean location, r, on mainland location, s

This model naturally induces a hierarchical Bayesian structure across timepoints, t

$$\begin{aligned}
\boldsymbol{Y}_t &\sim \mathcal{N}\left\{X_t\boldsymbol{\beta} + \left(I_{n_s} \otimes \boldsymbol{z}_t^T\right)\tilde{\boldsymbol{\alpha}},\ \Sigma\right\} \\
\tilde{\boldsymbol{\alpha}} &\sim \mathcal{N}\left\{\mathbf{0}_{n_s n_r \times 1},\ \Sigma \otimes R\right\} \\
\boldsymbol{\beta} &\sim \mathcal{N}\left\{\mathbf{0}_{p \times 1},\ \Lambda\right\} \\
\sigma^2 &\sim \text{Inv-Gamma}(k, \theta) \\
\rho &\sim \text{Uniform}(a, b)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma} &= \sigma_y^2\left(\Sigma(\rho_y) + \sigma_\varepsilon^2 I_{n_s}\right) \\
R &= \sigma_r^2 R(\rho_r) \\
\boldsymbol{z}_t &= \left[z(1,t) \dots z(r,t)\right]^T \\
\tilde{\boldsymbol{\alpha}} &= \left[\boldsymbol{\alpha}(s_1)^T \dots \boldsymbol{\alpha}(s_{n_s})^T\right]^T \\
\boldsymbol{\alpha}(s) &= \left[\alpha(r_1, s) \dots \alpha(r_{n_r}, s)\right]^T
\end{aligned}$$

Key implications:

1. This model enables inference about teleconnection effects (e.g., SST anomalies) while accounting for effects of local predictors (e.g., mainland temperature anomalies and location).
2. Modeled teleconnection effects, $\tilde{\boldsymbol{\alpha}}$, can capture large-scale correlations teleconnection predictors (e.g., SST anomalies).
3. Modeled teleconnection effects, $\tilde{\boldsymbol{\alpha}}$, vary smoothly across the mainland.

Estimation

- Estimate parameters with hybrid Gibbs algorithm applied to the model marginalized over spatial effects.
  - Conjugate updates computed for:
    - $\boldsymbol{\beta}$, σ_y^2
  - Random walk updates computed for:
    - $\log(\sigma_r^2)$, $\log(\sigma_\varepsilon^2)$, $\mathrm{logit}(\rho_y)$, $\mathrm{logit}(\rho_r)$
- Estimate teleconnection coefficients, $\tilde{\boldsymbol{\alpha}}$, using posterior parameter samples to draw (in parallel) composition samples of $\tilde{\boldsymbol{\alpha}}$.
- Marginal likelihood complexity:
  - Likelihood evaluation in $O(n_s^3 \vee n_r^3)$ by applying Kronecker product properties and Sherman-Morrison-Woodbury formula.

Marginal model: (let $Z = \left[\boldsymbol{z}_{t_1} \dots \boldsymbol{z}_{t_{n_t}}\right]$)

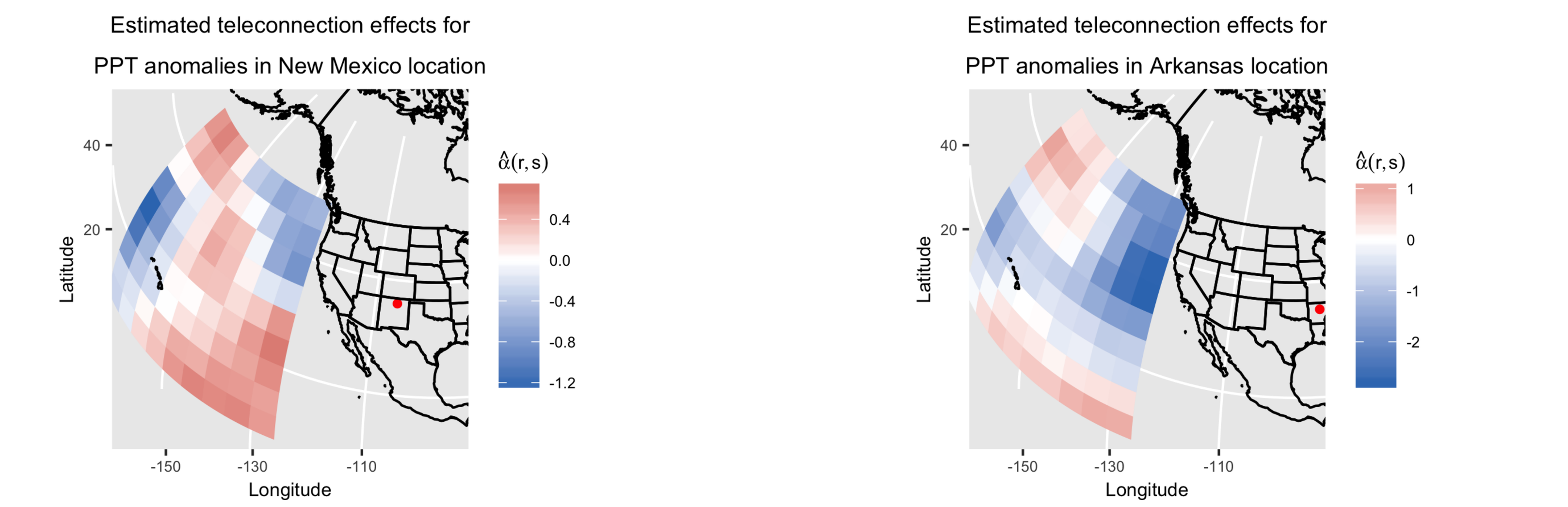
$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_{t_1} \\ \vdots \\ \boldsymbol{Y}_{t_{n_t}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} X_{t_1} & & \\ & \ddots & \\ & & X_{t_{n_t}} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta} \\ \vdots \\ \boldsymbol{\beta} \end{bmatrix},\ \left(I_{n_t} + Z^T R Z\right) \otimes \Sigma\right)$$

Composition-sample distribution:

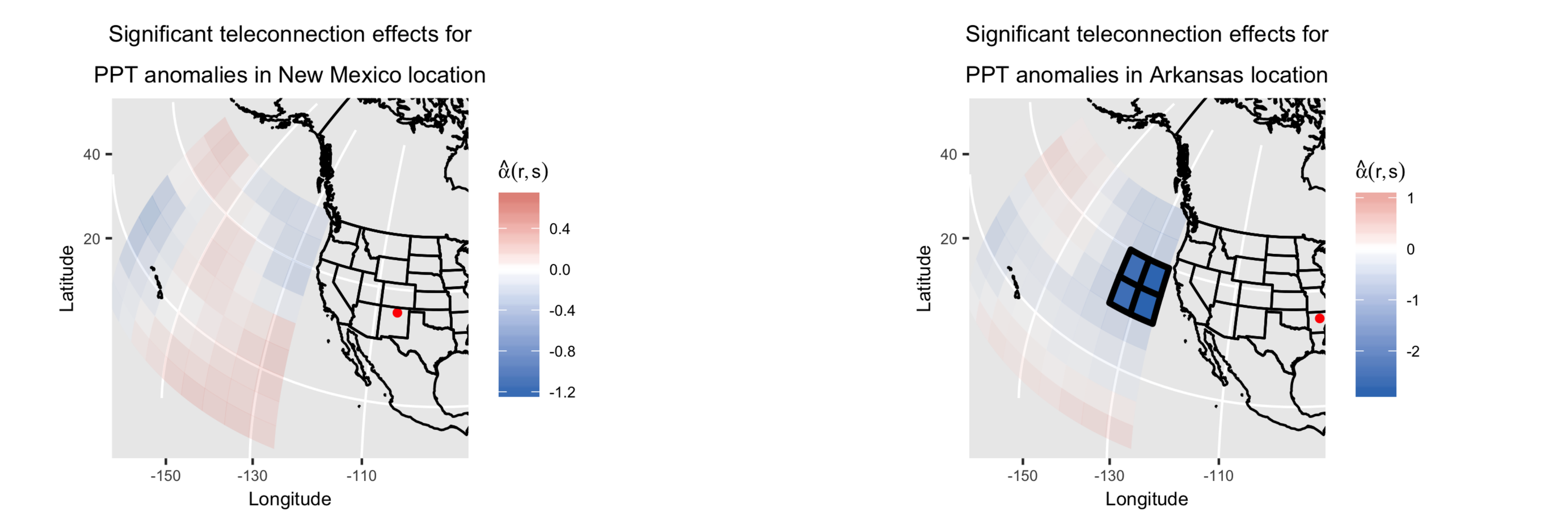
$$\tilde{\boldsymbol{\alpha}} | \boldsymbol{Y}, \cdot \sim \mathcal{N}\left(\sum_t \left((\boldsymbol{Y}_t - X_t\boldsymbol{\beta}) \otimes \left(R^{-1} + ZZ^T\right)^{-1} \boldsymbol{z}_t\right),\ \Sigma \otimes \left(R^{-1} + ZZ^T\right)^{-1}\right)$$

Results

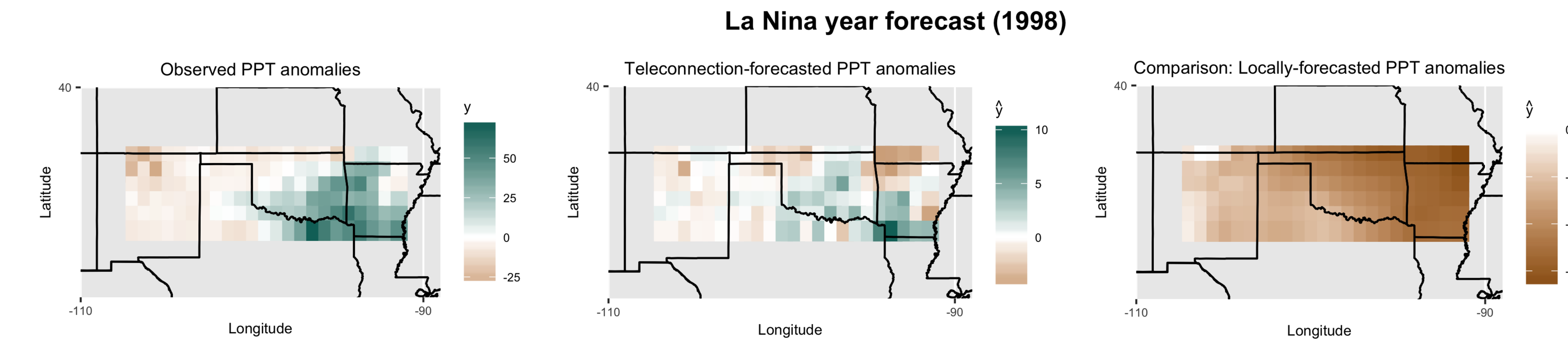
- Teleconnection effect estimates reproduce trends in observed pointwise correlation patterns

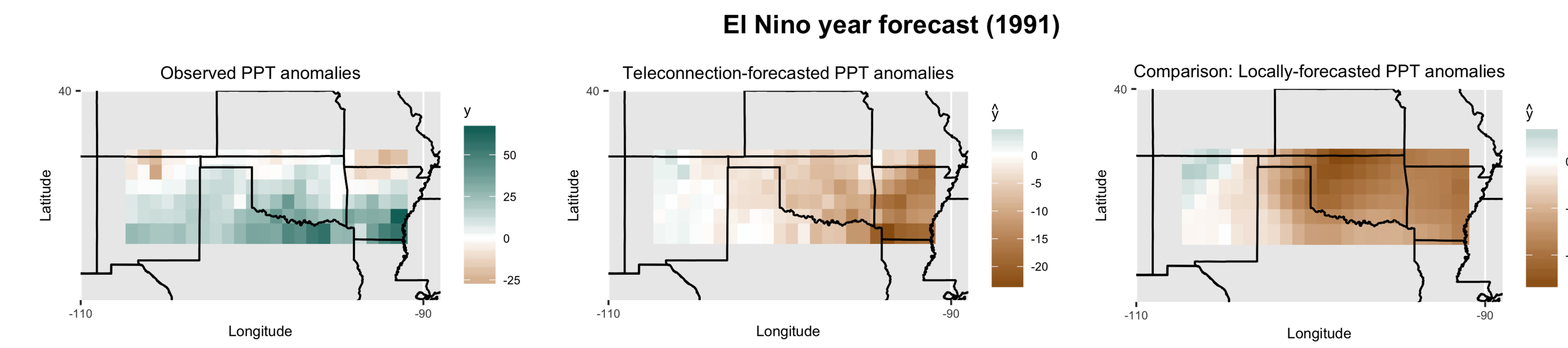


- Bayesian highest posterior density (HPD) intervals—a Bayesian version of confidence intervals—let us identify regions where non-spurious, scientifically meaningful teleconnection effects may exist
  - Following discussion in Towler, et. al. (2016), we search for novel ocean regions with significant teleconnections to the mainland region studied. For New Mexico, we found no regions with possible teleconnection. For Arkansas, we found a region off of California with a possible teleconnection (below right, highlighted in bold; significant based on a 70% HPD interval).



- PPT anomaly forecasts (below center) match observed patterns (below left) in La Niña years, but not in El Niño years. This discrepancy is likely an artifact of excluding the ENSO region from our analysis.
  - We compared the teleconnection model forecasts to a model that only includes the local predictor. The teleconnection model's root mean-squared prediction error (RMSPE) is 7% lower than the local forecast in the 1998 La Niña year. Posterior predictive loss is 10% lower.
  - The teleconnection model's RMSPE is generally high (La Niña RMSPE = 21 mm) since spatial models can greatly oversmooth predictions, thereby reducing prediction magnitudes.





Parameter estimates

	Posterior mean	90% HPD Interval
β_0	-3.54	(-35.2, 29.5)
β_1	-1.35	(-2.67, -0.00315)
β_2	-0.0294	(-0.353, 0.302)
σ_y^2	363	(319, 409)
σ_r^2	0.015	(0.0118, 0.0158)
ρ_y	82.1	(76.7, 86.8)
ρ_r	23.1	(7.74, 36.9)
σ_ε^2	0.0862	(0.057, 0.0742)

- 90% HPD Interval for β_1 suggests relationship of PPT with local temperature is weak. Future work will explore relationships with other local variables and their importance relative to remote variables.

Key contributions

- Geostatistical model that incorporates spatially disjoint predictors.
- Methodology for adding explanatory structure and statistical testing to teleconnection patterns.

Future work

- Reformulate model to predict anomaly categories (i.e., "High", "Medium", "Low") to increase skill.
- Refine initial, exploratory modeling to consider a larger Pacific region and additional variables to assess improvements to predictions.
- Investigate effect of spatial scale on predictive skill.

References

Banerjee, S., Carlin, B.P., and Gelfand, A.E., 2004, *Hierarchical Modeling and Analysis for Spatial Data*: Boca Raton, FL, Chapman and Hall/CRC.

Givens, G.H., and Hoeting, J., 2013, *Computational Statistics*: Hoboken, NJ, John Wiley & Sons, Inc.

Towler, E., PaiMazumder, D., and Holland, G., 2016, A framework for investigating large-scale patterns as an alternative to precipitation for downscaling to local drought: Climate Dynamics, p. 1-12.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. AGS - 1419558.

NSF research network on STATMOS through grant DMS-1106862.