NONPARAMETRIC TEST OF ASSUMPTIONS OF SPATIOTEMPORAL MODEL WITH VARYING FREQUENCIES

By:

Daniel David M. Pamplona*, Dr. Joseph Ryan G. Lansangan, Dr. Erniel B. Barrios

University of the Philippines

Background

Spatiotemporal Model

- a statistical representation of data sampled from specific locations over a period of time that characterizes their exogenous, spatial, and temporal dependencies (Cressie & Majure, 1997).

Malabanan and Barrios (2017) postulated a semiparametric spatiotemporal model, motivated by agricultural systems, with data measured at varying frequency, e.g.

- Amount of Rainfall (daily / weekly)
- Amount of Fertilizer Applied (quarterly)
- Corn Production (quarterly)

Spatiotemporal Model with Varying Frequencies (Malabanan & Barrios, 2017)

$$y_{it} = \sum_{k=1}^{K} f(X_{it_k}) + \beta Z_{it} + \gamma W_{it} + \varepsilon_{it}$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + a_{it}, |\rho| < 1, a_{it} \sim IID(0, \sigma_a^2), i = 1, \dots, n; t = 1, \dots, T; k = 1, \dots, K$$

Components

- Y_{it} - response in unit *i* at time *t*
- X_{it_k} - covariate measured at higher frequency
- $f(\cdot)$ - continuous function in X_{it_k}
- Z_{it} - covariate measured in the same frequency as response in spatial unit *i* at time *t*
- W_{it} - neighborhood system where spatial unit *i* belongs at time *t*
- ε_{it} - error terms for spatial unit *i* at time *t*

Estimation

Backfitting Algorithm
Spline Smoothing – nonparametric
Regression Model – parametric
AR(1) model – temporal parameter

Spatiotemporal Model with Varying Frequencies (Malabanan & Barrios, 2017)

$$y_{it} = \sum_{k=1}^{K} f(X_{it_k}) + \beta Z_{it} + \gamma W_{it} + \varepsilon_{it}$$

$$\varepsilon_{it} = \rho \varepsilon_{i,t-1} + a_{it}, |\rho| < 1, a_{it} \sim IID(0, \sigma_a^2), i = 1, \dots, n; t = 1, \dots, T; k = 1, \dots, K$$

Model Assumptions

- Constant nonparametric component effect across units and across time
- Constant parametric component (β) effect across units and across time
- Constant neighborhood variable (γ) effect across units and across time
- Constant temporal effect (ρ) across units

Testing of Constant Covariate Effect Across Spatial Units

Hypothesis

$H_0: \beta_i = \beta \quad \forall i$

$H_1: \beta_i \neq \beta \quad for\ some\ i$

Algorithm

1. Estimate $y_{it} = \sum_{k=1}^{K} f(x_{it_k}) + \beta z_{it} + \gamma w_{it} + \varepsilon_{it}$
2. Isolate residuals associated to covariate effect, $y_{it}^* = e_{it}^{\beta} = y_{it} - \left(\sum_{k=1}^{K} \widehat{f(X_{it_k})} + \hat{\gamma} W_{it}\right)$
3. In each spatial unit, generate k bootstrap samples of T pairs (z_{it}, y_{it}^*), and estimate
   $y_{it}^* = \beta_i z_{it} + \varepsilon_{it}, \quad i = 1, . . . , N, \quad t = 1, \ldots, T$
4. Compute $\hat{\sigma}_{\hat{\beta}_i}^* = \left[\frac{1}{k-1}\sum_{p=1}^{k}\left(\hat{\beta}_{ip}^* - \overline{\hat{\beta}_{ip}^*}\right)^2\right]^{1/2}, \quad \overline{\hat{\beta}_{ip}^*} = \frac{1}{k}\sum_{p=1}^{k}\hat{\beta}_{ip}^*$
5. Construct $(1-\alpha) * 100\%$ CI using
   $\overline{\hat{\beta}_{ip}^*} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\beta}_i}^*$
6. Test the hypothesis using the test statistic:
   $D = \frac{1}{\binom{N}{2}} \sum_{l=1}^{\binom{N}{2}} d_l$ where $d_l = \begin{cases} \min(|U_1 - L_2|, |L_1 - U_2|), & if\ CI\ pair\ does\ not\ overlap \\ 0, & otherwise \end{cases}$
7. Reject H_o if $D > D_{\alpha}^{null}$

III. Simulating Non-constant Covariate Effect

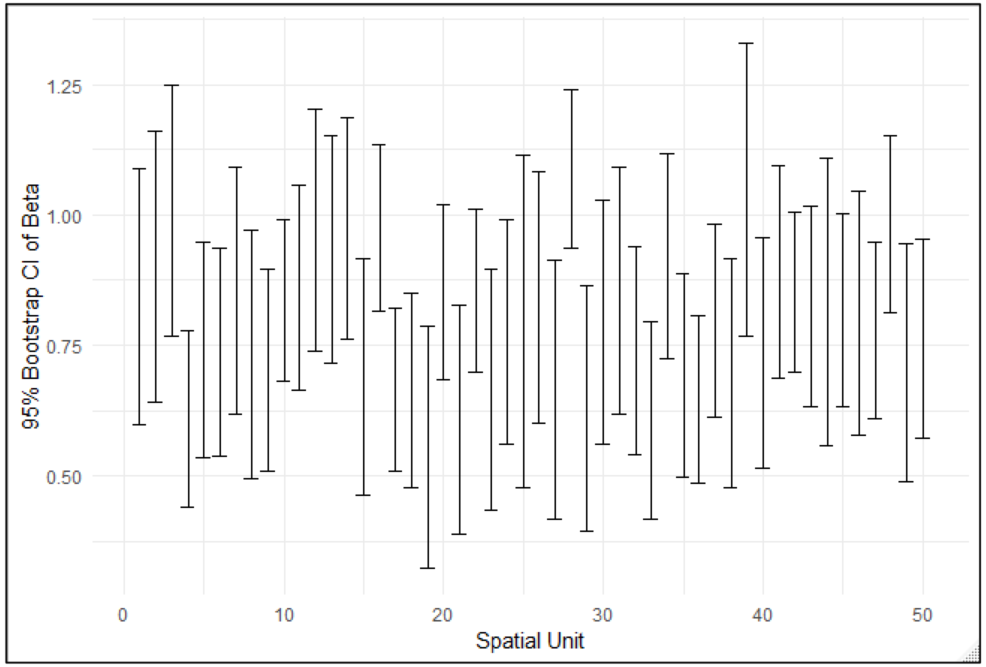
$$y_{it} = a\sum_{k=1}^{K} f(X_{it_k}) + (\beta + v_i\beta)Z_{it} + \gamma W_{it} + \varepsilon_{it}\ ,$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + ma_{it}$$

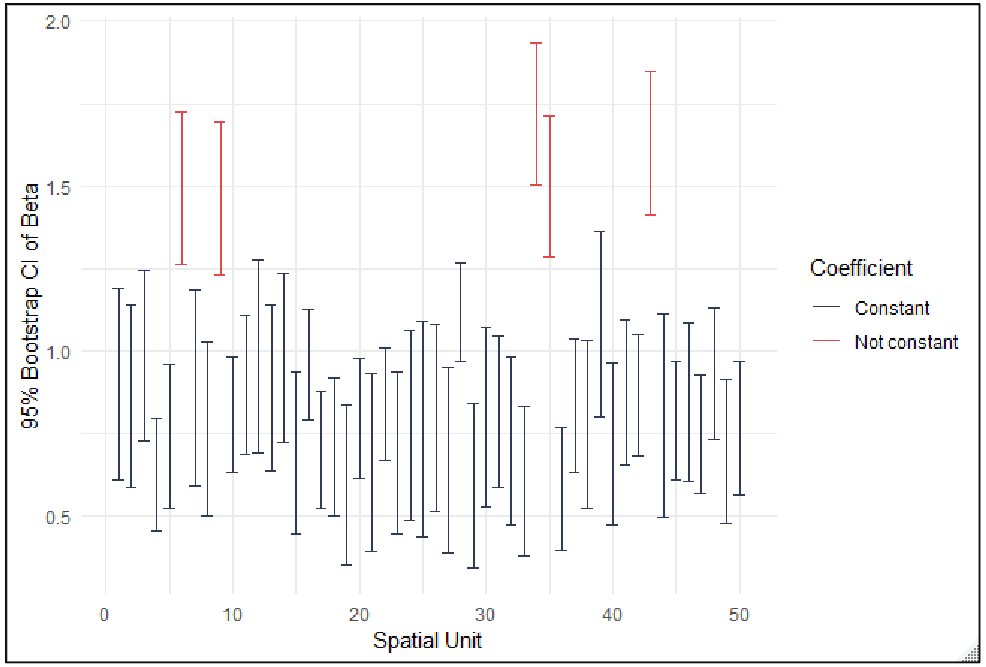
where v_i *is a nonzero constant for some spatial unit* i.

Characteristic	Settings
Percent of spatial units / time points with alternative parameter values	10% 20%
Percent difference from "true" parameter value in simulation	$v = 0.0$ $v = 0.3$ $v = 1.0$ $v = 2.0$

Illustration (Constant vs Non-constant Covariate Effect)



(a)



(b)

Sample Plot of 95% Bootstrap CI of β_i under (a) under H_0:$\beta_i = \beta = 0.79$; and (b) H_1: $\beta_i = \beta + v\beta$; $v = 1.0$ for 5 time points, $v = 0$ for 45 time points

Table 1. Average empirical size and power in testing constant covariate effect across time points for different simulation scenarios (m = 1)

Characteristics	Values	**Size under** $H_0: \beta_t = \beta$	**Power at** $H_1: \beta_t = \beta + v\beta$					
			5 time points			10 time points		
			$v = 0.3$	$v = 1$	$v = 2$	$v = 0.3$	$v = 1$	$v = 2$
Component Contributions (f(x)-z-w-e)	(30-30-30-10)	0.051	0.141	0.858	1.000	0.232	0.946	1.000
	(20-50-20-10)	0.043	0.392	0.995	1.000	0.528	1.000	1.000
	(20-20-50-10)	0.065	0.073	0.605	0.974	0.101	0.749	0.995
Error correlation	rho = 0.5	0.064	0.239	0.866	0.988	0.346	0.929	0.994
	rho = 0.9	0.058	0.094	0.572	0.872	0.129	0.696	0.917
Length of N and T	N=T	0.062	0.165	0.772	0.985	0.235	0.877	0.997
	N>T	0.047	0.263	0.893	1.000	0.372	0.938	1.000
	N<T	0.048	0.178	0.793	0.990	0.254	0.880	0.998
Nature of Covariate	no temporal correlation	0.064	0.230	0.879	0.995	0.320	0.946	0.998
	with temporal correlation	0.056	0.103	0.559	0.865	0.155	0.679	0.914

Color	(white)	(light yellow)	(light orange)	(green)
Proportion of Rejection	0.0-0.25	0.26-0.50	0.51-0.75	0.76-1.00

Table 2. Average empirical size and power in testing constant covariate effect across locations for different simulation scenarios (m = 1)

Characteristics	Values	**Size under** $H_0: \beta_i = \beta$	**Power at** $H_1: \beta_\mathrm{i} = \beta + v\beta$					
			5 locations			10 locations		
			$v = 0.3$	$v = 1$	$v = 2$	$v = 0.3$	$v = 1$	$v = 2$
Component Contributions (f(x)-z-w-e)	(30-30-30-10)	0.054	0.133	0.755	0.958	0.189	0.845	0.968
	(20-50-20-10)	0.051	0.302	0.921	0.982	0.433	0.963	0.985
	(20-20-50-10)	0.055	0.064	0.481	0.850	0.090	0.630	0.914
Error correlation	rho = 0.5	0.064	0.239	0.866	0.988	0.346	0.929	0.994
	rho = 0.9	0.058	0.094	0.572	0.872	0.129	0.696	0.917
Length of N and T	N=T	0.042	0.128	0.653	0.901	0.200	0.744	0.924
	N>T	0.041	0.134	0.697	0.928	0.193	0.800	0.956
	N<T	0.055	0.237	0.808	0.961	0.319	0.894	0.988
Nature of Covariate	no temporal correlation	0.056	0.237	0.878	0.999	0.341	0.946	1.000
	with temporal correlation	0.047	0.167	0.761	0.984	0.233	0.851	0.997

Color	White	Light yellow	Light orange	Green
Proportion of Rejection	0.0-0.25	0.26-0.50	0.51-0.75	0.76-1.00

Table 3. Average empirical size and power in testing constant covariate effect across time points for different simulation scenarios (m = 10)

Characteristics	Values	**Size under** $H_0: \beta_t = \beta$	**Power at** $H_1: \beta_t = \beta + v\beta$					
			5 time points			10 time points		
			$v = 0.3$	$v = 1$	$v = 2$	$v = 0.3$	$v = 1$	$v = 2$
Component Contributions (f(x)-z-w-e)	(30-30-30-10)	0.053	0.051	0.064	0.083	0.057	0.063	0.136
	(20-50-20-10)	0.055	0.051	0.069	0.212	0.054	0.095	0.311
	(20-20-50-10)	0.043	0.045	0.057	0.062	0.047	0.059	0.078
Error correlation	rho = 0.5	0.061	0.041	0.066	0.174	0.046	0.074	0.257
	rho = 0.9	0.052	0.057	0.061	0.064	0.060	0.070	0.093
Length of N and T	N=T	0.045	0.047	0.066	0.104	0.047	0.063	0.132
	N>T	0.062	0.038	0.057	0.145	0.048	0.071	0.227
	N<T	0.050	0.062	0.068	0.107	0.064	0.083	0.167
Nature of Covariate	no temporal correlation	0.050	0.053	0.069	0.151	0.049	0.087	0.234
	with temporal correlation	0.038	0.047	0.059	0.089	0.054	0.062	0.117

Color	(white)	(light yellow)	(light orange)	(light green)
Proportion of Rejection	0.0-0.25	0.26-0.50	0.51-0.75	0.76-1.00

Table 4. Average empirical size and power in testing constant covariate effect across locations for different simulation scenarios (m = 10)

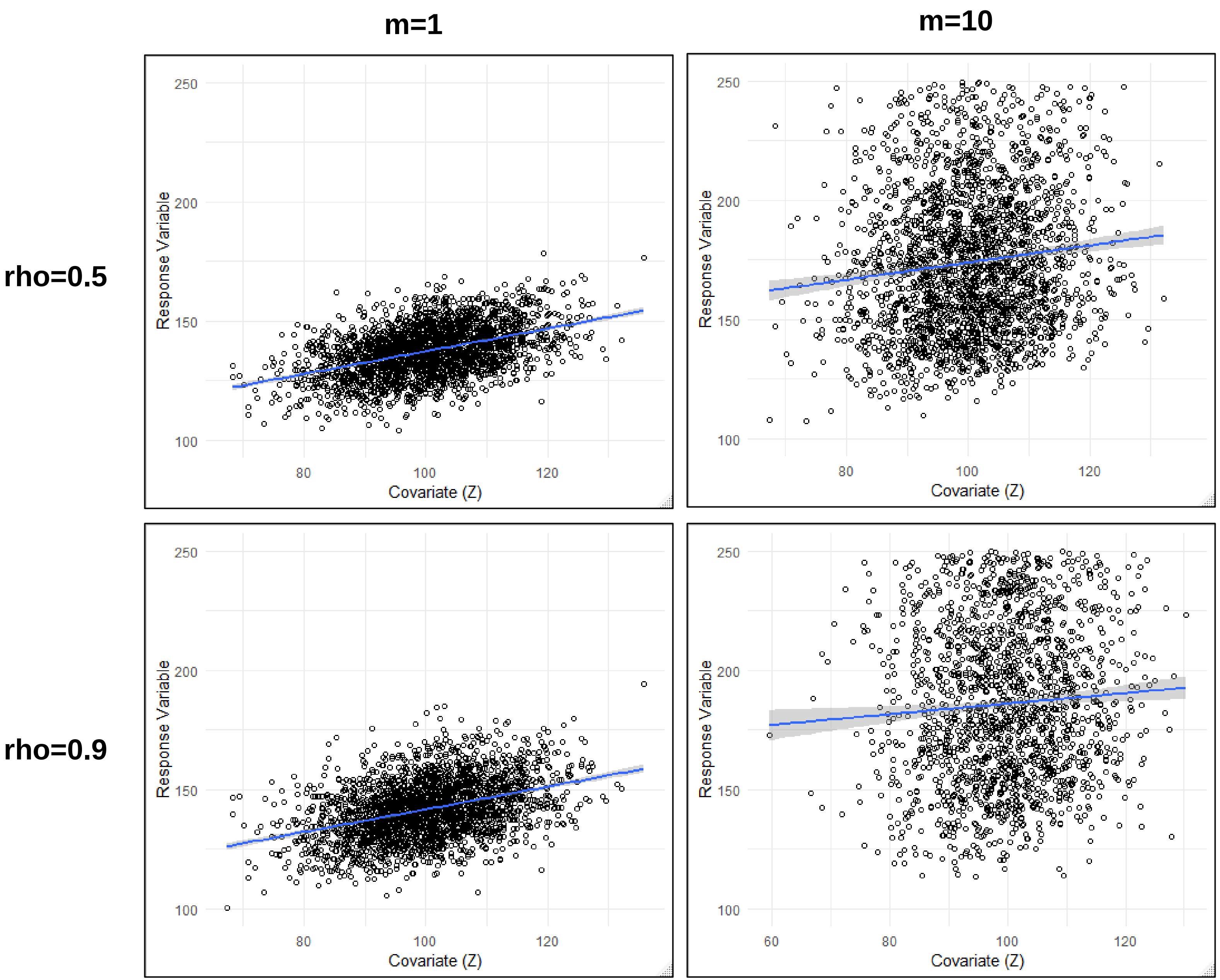
Characteristics	Values	**Size under** $H_0: \beta_i = \beta$	**Power at** $H_1: \beta_i = \beta + v\beta$					
			5 locations			10 locations		
			$v = 0.3$	$v = 1$	$v = 2$	$v = 0.3$	$v = 1$	$v = 2$
Component Contributions (f(x)-z-w-e)	(30-30-30-10)	0.044	0.055	0.06	0.096	0.052	0.067	0.149
	(20-50-20-10)	0.047	0.049	0.071	0.196	0.053	0.105	0.300
	(20-20-50-10)	0.048	0.047	0.06	0.067	0.049	0.052	0.078
Error correlation	rho = 0.5	0.042	0.063	0.083	0.168	0.055	0.093	0.260
	rho = 0.9	0.037	0.037	0.045	0.072	0.048	0.056	0.091
Length of N and T	N=T	0.046	0.045	0.056	0.099	0.050	0.067	0.151
	N>T	0.061	0.05	0.063	0.096	0.048	0.057	0.136
	N<T	0.056	0.056	0.073	0.164	0.056	0.100	0.239
Nature of Covariate	no temporal correlation	0.050	0.054	0.077	0.163	0.06	0.086	0.228
	with temporal correlation	0.048	0.044	0.05	0.075	0.045	0.059	0.122

Color	(white)	(light yellow)	(light orange)	(green)
Proportion of Rejection	0.0-0.25	0.26-0.50	0.51-0.75	0.76-1.00

Model Component Contribution to the Response at m=1 and at m=10

Target Model Component Contributions (f(x)-z-w-e) in %	Simulated Model Component Contribution **(m=1)**	Simulated Model Component Contribution **(m=10)**
(30-30-30-10)	(29.31–34.79–32.61–3.28)	(22.62–26.84–25.16–25.37)
(20-50-20-10)	(19.09–56.35–21.27–3.27)	(14.73–43.51–16.41–25.34)
(20-20-50-10)	(19.53–23.26–53.95–3.24)	(15.12–18.01–41.77–25.10)

Scatter Plot of Response and Covariate



Conclusions

1. The proposed procedure in testing for constant covariate effect across time points and spatial units is correctly sized under the different scenarios simulated.
2. Power of test follows an increasing pattern as frequency and degree of deviation from the true coefficient increase.
3. The test is sensitive under the presence of non-homogenous covariate effect that is at least twice the magnitude of the “true” coefficient, and when model errors are not too large.
4. The frequency of time points / locations with alternate coefficients only slightly improves power different.

References:

- Alonso, A., Peña, D., and Romo, J. (2001). Forecasting Time Series with Sieve Bootstrap. *Journal of Statistical Planning and Inference* 100 (2002) 1-11.
- Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli* 3(2), 123-148.
- Chernick, M.R. (1999). Bootstrap Methods: A Practitioner's Guide: New York: *John Wiley and Sons, Inc*.
- Cressie, N. and Majure, J. J. (1997). Spatio-temporal Statistical Modeling of Livestock Waste In Streams. *Journal of Agricultural, Biological, and Environmental Statistics* 2, 24-47.
- De Almeida, D.B., Ferreira K. R., de Oliveira A. G., Monteiro A. M. V., (2015). Temporal GIS and Spatiotemporal Data Sources. *Proceedings XVI GEOINFO*. Campos do Jordao, Brazil p. 1-13.
- Davison, A.C and Hinkley, D.V. (1997). Bootstrap Methods and their Application. Cambridge: *Cambridge University Press*.
- Efron, B. (1979). Bootstrap methods: Another Look at the Jackknife, *Annals of Statistics*, 7, 1-26.
- Efron, B. and Tibshirani, R. J. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy, *Statistical Science*, 1, 54-77.
- Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap, Chapman and Hall, New York.
- Guarte, J. and Barrios, E. (2012). Nonparametric Hypothesis Testing in a Spatial-Temporal Model: A Simulation Study. *Communications in Statistics – Simulation and Computation*. 42, 153-170.
- Hardle, W., Muller, M., Sperlich, S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models: An introduction. *Springer* 2004.

- Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap, Chapman and Hall, New York.
- Guarte, J. and Barrios, E. (2012). Nonparametric Hypothesis Testing in a Spatial-Temporal Model: A Simulation Study*. Communications in Statistics – Simulation and Computation*. 42, 153-170.
- Hardle, W., Muller, M., Sperlich, S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models: An introduction. *Springer* 2004.
- Hastie, T., James, G., Tibshirani, R. and Witten, D. (2013). An Introduction to Statistical Learning with Applications in R. *Springer Texts in Statistics*, p 187.
- Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics*. 17 1217-1241.
- Kreiss, J. and Frankie, J. (1992). Bootstrapping Stationary Autoregressive Moving Average Models. *J. Time Series Anal.* 13, 297-317
- Landagan, O. and Barrios, E. (2007). An Estimation Procedure for Spatial-Temporal Model. *Statistics and Probability Letter* 77, 401-406.
- Shao, J. and Tu, D. (1995). The Jacknife and Bootstrap. New York: Springer-Verlag, Inc.
- Squires, E. (2013). The Impact of Different Nitrogen Fertilizer Rates on Soil Characteristics, Plant Properties, and Economic Returns in a Southeastern Minnesota Cornfield. *Field Ecology*. St. Olaf College.
- Torre, D.M., Perez, G.J., Penology-Based Classification of Major Crops Areas in Central Luzon, Philippines from 2001-2013. University of the Philippines Diliman
- Zhu, J., Huang, H., and Wu, J. (2005). Modeling Spatial-temporal Binary Data using Markov Random Fields. *Journal of Agricultural, Biological, and Environmental Statistics* 10, 212-225.