SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

# WORKING PAPER SERIES

**Robust Estimation of a Dynamic Spatio-temporal Model with Structural Change**

by

***Stephen Jun V. Villejo***
***Erniel B. Barrios***
***Joseph Ryan G. Lansangan***

UPSS Working Paper No. 2016-01
February 2016

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

**Abstract**

We postulate a dynamic spatio-temporal model with constant covariate effect but with varying spatial effect over time and varying temporal effect across locations. To mitigate the effect of temporary structural change, the model can be estimated using the backfitting algorithm embedded with forward search algorithm and bootstrap. A simulation study is designed to evaluate structural optimality of the model with the estimation procedure. The fitted model exhibit superior predictive ability relative to the linear model. The proposed algorithm also consistently produced lower relative bias and standard errors for the spatial parameter estimates. While additional neighborhoods do not necessarily improve predictive ability of the model, it trims down relative bias on the parameter estimates, specially for spatial parameter. Location of the temporary structural change along with the degree of structural change contributes to relative bias of parameter estimates and in predictive ability of the model. The estimation procedure is able to produce parameter estimates that are robust to the occurrence of temporary structural change.

***Keywords:*** spatio-temporal model, backfitting algorithm, bootstrap, forward search, dynamic model

## 1. Introduction

Many data is generated from a compilation of those coming from different sources involving a dynamic interaction of spatial and temporal dependencies. While independence assumption can tremendously simplify modeling, it is easily violated in data sets indexed by space and time, consequently resulting to misleading understanding of the data generating process. Xu and Wikle (2007), observed that spatio-temporal modeling is essential yet complicated due to the dependence structure across space and time. Its application transcends many disciplines. According to Kyriakidis and Journel (1999), modeling of spatiotemporal distributions generated by processes evolving in both space and time is critical in many science and engineering fields. Statistical models built in the premise that units closely located in space and time are correlated are becoming more prominent in the literature. Cressie (1993) proposed to include time as an extra dimension in traditional spatial statistical techniques. Kyriakidis and Journel (1999) suggests viewing spatio-temporal processes from the multiple time series perspective. Cressie and Wikle (2002) combined spatial and time series techniques through a spatio-temporal dynamic model formulation. However, it is still of great necessity to improve the existing models and estimation methods which recognizes such idiosyncrasies.

Spatial and time interaction is evident in the context of agricultural production, (Landagan and Barrios, 2007). Similar crops are planted in the same location due to geographical, climatic, and soil conditions. Adjacent locations are most likely similar or resembles almost the same characteristics in terms of soil quality, availability and quality of irrigation systems, amount of rainfall, terrain and climate or weather condition. In addition to the spatial dependence, there could also exist dependence among values adjacent in time due to seasonal weather patterns. Yield of crops are affected by physical and geographical conditions, summary of indicator/quantitative variables that represent information within the area with similar conditions, and another term which accounts for production shocks. These components in the model will vary through space and time. Systems with such complexities which exhibit spatial and temporal dependence is best facilitated by a dynamic spatio-temporal models. This facilitates better understanding of the dynamic relationships in the system, paving the way for more accurate forecasts that can aid policy-making.

We postulate and characterize a dynamic spatio-temporal model and propose an estimation procedure that can withstand onset of temporary structural change. Following Hackl and Westlund (1989), an essential element of structural change is non-constancy of the parameters and could be evident in time-series or even in cross-sectional data. In other words, structural change happen when there is a change in the system and in the behavior of the random variables which occur at certain period of time or in particular neighborhoods. Occurrence of structural change is temporary since after a duration of time, the series will return to its original behavior. An estimation procedure that is robust to structural change is therefore necessary since the change will happen only temporarily. Moreover, viewing the system as dynamic rather than static is more realistic and universal. Occurrences and the behavior in a particular location may not be true in other locations. Hence, the goal of this paper to develop an estimation procedure for dynamic spatio-temporal processes that is robust to structural change.

## 2. Some Methods of Spatio-Temporal Modeling

Kyrikaidis and Journel (1999) reviewed numerous geostatistical space-time models, distinguished two perspectives in the model-building: one is calling for a single spatio-temporal random function model, and; the other calling for vectors of space random functions or vectors of time-series. Kyriakidis and Jounrel (1999) further noted that the practice of modelling a stochastic spatio-temporal model depends on the data set and the goal, the researcher has the flexibility in performing the modelling tasks to fit best the model for a phenomenon given a particular data at hand. Cressie (1993) proposed to include time as an extra dimension in the use of traditional spatial statistical techniques. A drawback of the method is that it ignores the fundamental differences between space and time since time is naturally-ordered while space is not. Another approach is to formulate the problem in a multivariate geostatistical perspective. This will again pose some limitation since it requires space-time covariance functions in which only quite little is known for such a class of functions. Kyriakidis and Journel (1999) suggest viewing spatio-temporal processes from the multiple time series perspective by associating a time series for each spatial location. However, a disadvantage of such approach is that it ignores the fundamental differences between space and time and it does not generally accounts for uncertainty in the observed data.

Another approach to modeling spatio-temporal processes is using a combination of spatial and time series techniques which is accomplished by a spatio-temporal dynamic model formulation (Cressie and Wikle, 2002). This procedure leads to high dimensionality of the state process making estimation problematic. One proposed solution is projecting the state-process onto some set of spectral basis functions to reduce dimensionality (Mardia etal 1998; Cressie and Wikle, 1999). Many solutions have been proposed (see, Xu and Wikle 2007), but it requires the model parameters to be explicitly known and this happens very seldom. Also, many parameters need to be estimated outright due to the high dimensionality and covariance matrices are therefore highly-parameterized causing difficulty in the estimation. Xu and Wikle (2007) proposed using a state-space framework where the unobserved state-process can be estimated via the Kalman filter. Since the model parameters are unknown, the standard approach is to use the expectation-maximization (EM) algorithm, Shumway and Stoffer (1982). In such cases when parameter vectors and matrices are of high dimension, Xu and Wikle (2007) proposed a spatio-temporal dynamic model formulation with parameter matrices restricted based on prior knowledge and/or common spatial models. The estimation is performed using the EM or general EM (GEM) algorithm. They noted that fully Bayesian (MCMC) approach is better when one has significant prior information about the dynamics of a complex model. Also, the EM or GEM approach can be used when the model is not so complex and when little is known about the dynamics of the model since the EM or GEM approach falls short in a high-dimensional parameter space due to convergence problems.

Landagan and Barrios (2007) proposed an estimation procedure of a spatial-temporal model using the backfitting algorithm infused with the Cochrane-Orcutt procedure. The idea is to alternately and iteratively estimate parameters of the covariate and temporal effects first then the spatial effect. The postulated spatial-temporal model is found to be superior to commonly used models like the ordinary linear model, linear model with autocorrelated errors and mixed linear model.

Dumanjug, et. al. (2010) introduced two bootstrap method into the algorithm of Landagan and Barrios (2007). The first method uses ordinary bootstrap and the second method uses time blocks of consecutive observations. The second method gave more efficient and reliable bootstrap estimates. Furthermore, the simulation study showed that the model gave very low prediction errors and are robust to the number of spatial units and time points.

Taking stimulus from disease prevalence and epidemics of infectious diseases, Bastero and Barrios (2011) postulated a spatio-temporal model with structural change, the change dynamics is represented by exponential functions that reasonable vanish after specific period of time. Occurrence of epidemics can alter the prevalence of diseases from normal, resulting to possible structural change.

**3. Dynamic Spatio-Temporal Model**

Barrios and Guarte (2013) developed a nonparametric test procedure based on the bootstrap to verify two model assumptions (constant temporal effect across space and constant spatial effect over time) postulated by Landagan and Barrios (2007). The procedure is able to verify the assumptions for reasonably sized data. Also, the test for constant spatial effect over time is robust to model misspecification. Suppose that the test for varying temporal effect across locations and varying spatial effect over time rejects the null hypotheses. This easily happens when spatial and temporal clustering is present as a result of poisson innovation of events over time and across space. This requires a dynamic model for a clearer abstraction of the spatio-temporal data generating process.

Consider the following model:

$$Y_{it} = \beta X_{it} + \delta_t W_{it} + \epsilon_{it} \qquad (1)$$

where,

$Y_{it}$= response variable from location i at time t

$X_{it}$ = covariate from location i at time t

$W_{it}$ = variable in the neighborhood system of location i at time t

$\epsilon_{it}$ = error term

Without loss of generality, assume that the error term is generated by an autoregressive process of order 1, given by

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + a_{it}, \quad |\rho_i| < 1, \quad a_{it} \sim IID(0, \sigma^2{}_a) \qquad (2)$$

The Equations (1) and (2) specifies the following:

i. constant covariate (β) effect across location and time

ii. varying temporal effect ($\rho_i$) across locations,

iii. varying spatial effect ($\delta_t$) across time.

The constant covariate effect assumes that characteristic of the covariates and their effect on the response variable will not significantly change over time and across space. As an example, if $X_{it}$ is area harvested at time t, this means that if area harvested is high at *t=0*, the area harvested will remain to be high at t = 1, 2, …, T. $W_{it}$ are variables which define the neighbourhood system. In Landagan and Barrios (2007), spatial effect $\delta_t$ is assumed to be constant over time while the temporal effect $\rho_i$ is assumed to be constant across locations. In this study, spatial parameter is assumed to vary across time while the temporal parameter is assumed to vary across locations. Furthermore, without loss of generality, we will consider only one covariate and one neighborhood variable. Geographic characteristics can be used in defining neighborhoods in an agricultural setup, see for example (Landagan and Barrios, 2007).

After accounting for spatial and temporal dependencies into the model, the phenomenon of interest may still exhibit structural change. In the agriculture example above, certain areas may be affected by natural calamities at specific times, the effect of area ($X_{it}$) on yield may deviate for some areas (at some time point) relative to other areas. The change could only be temporary since everything will return to the 'normal' conditions (pre-calamity) for affected areas in due time. In this case, one is interested in constructing models that are robust to these temporary change to gain better insights of the 'normal' times unaffected by temporary changes caused by certain events.

**Estimation Procedure**

The model summarized in Equations (1) and (2) are characterized by blocks of determinants that are related within each block but possibly unrelated across blocks (with minimal relations across blocks at the most). One can easily postulate that Equations (1) and

(2) is an additive model. Owing to the additivity of model in Equations (1) and (2), the backfitting algorithm should serve as the basic paradigm of estimation. Covariate effects, spatial effects, and temporal effects, can be estimated sequentially via the iterative processes involved in the backfitting algorithm. The effect of structural change that is localized only in certain components of the model can be mitigated through the forward search algorithm integrated into the backfitting algorithm. Furthermore, limited time series data for each spatial unit imposes constraints in estimating dynamic parameters. In this case, replicates in a time series bootstrap can facilitate estimation of dynamic parameters in a nearly overparameterized dynamic spatio-temporal model. Thus, to address structural change while estimating dynamic parameters, the forward search and the bootstrap methods are imbedded into the backfitting algorithm in estimating the model summarized in Equations (1) and (2).

The backfitting algorithm is used to estimate the parameters of the spatio-temporal model, i.e, parameters are estimated iteratively. Forward search algorithm and bootstrap method is incorporated in the backfitting algorithm to benefit from robustness of estimates that each method produces. The algorithm is given in the following steps:

**Step 1**: **Forward Search**

Using the realizations $\{y_{it}\}$, i=1, …, N, t=1, …, T, where i is the index for location and t is the index for time, $\beta$ and $\delta_t$ will be simultaneously estimated using ordinary least squares for each of the T time points. Forward search algorithm will be incorporated in each of the ordinary least squares estimation through the following steps. For each time point:

i. A subset of size n, n < N from N observations for each time point will be chosen. The n observations are ideal and outlier-free. The choice of the n observations in the initial subset corresponds to the smallest n residuals after performing ordinary least squares on the N units. The initial subset *n* is computed as 50% of *N.*

ii. Fit the model $Y_{it} = \beta X_{it} + \delta_t W_{it} + \epsilon_{it}$ using selected n observations, this leads to the parameter estimates $\hat{\beta}$ and $\widehat{\delta_t}$.

iii. Using the fitted model in ii, compute fitted response to N-n left-out observations, compute residuals.

iv. The observation corresponding to the smallest residual from the N-n residuals will be included in the subset of observations from (i).

v. Fit the model $Y_{it} = \beta X_{it} + \delta_t W_{it} + \epsilon_{it}$ on the n+1 observations in (iv).

vi. Iterate from ii, adding one observation at a time until all N locations have been included in the model or until the model is behaving wildly based on the Cook's D. An observation is said to be influential if the Cook's D value exceeds $\frac{4}{n}$ where n is the number of observations. The algorithm stops if the newly added observation yields a Cook's D exceeding the threshold.

The forward search is used to obtain robust estimates of the spatial and covariate parameters. Observations which behave wildly based on the postulated model will not be included in the estimation of the parameters.

**Step 2**. **Robust Estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\delta_t}$**

After obtaining the final subset from the forward search algorithm for each time point, the following is performed for each time point:

i. Without loss of generality, assuming that there are m units in the final subset after the forward search from Step 1, draw a simple random sample of size m with replacement from the subset.

ii. Fit the model $Y_{it} = \beta X_{it} + \delta_t W_{it} + \epsilon_{it}$. This gives a value of the parameter estimates computed using the bootstrap sample in (i).

iii. Repeat (i) and (ii) B times, yielding B $\hat{\beta}$ and B $\widehat{\delta_t}$.

iv. The Monte Carlo estimates for the mean and standard deviation of $\delta_t$ is computed. If we denote by $\hat{\beta}_{tk}$ the kth bootstrap estimate of β from time point $t$ and and $\hat{\delta}_{tk}$ as the kth bootstrap estimate of $\delta_t$, $k = 1, \dots, B$, then Monte Carlo estimates are computed from:

$$\hat{\delta}_t^{\,*} = \frac{1}{B}\sum_{k=1}^{B} \hat{\delta}_{tk} \text{ , S.E.}(\hat{\delta}_t^{\,*}) = \frac{1}{B-1}\sum_{k=1}^{B}(\hat{\delta}_{tk} - \hat{\delta}_t^{\,*})^2 \tag{3}$$

Since the covariate effect is constant, the Monte Carlo Estimate of β is computed as follows:

$$\hat{\beta}^* = \frac{1}{BT}\sum_{t=1}^{T}\sum_{k=1}^{B}\hat{\beta}_{tk} \tag{4}$$

The index *i* is for the time point while the index k is for the bootstrap replicate. The standard error of the estimate is then computed as follows:

$$S.E.(\hat{\beta}^*) = \frac{1}{BT-1}\sum_{t=1}^{T}\sum_{k=1}^{B}(\hat{\beta}_{tk} - \hat{\beta}^*)^2 \tag{5}$$

This step generates the set of parameter estimates $\{\hat{\delta}_1^{\,*}, \hat{\delta}_2^{\,*}, \dots, \hat{\delta}_t^{\,*}, \hat{\beta}^*\}$ with their corresponding standard errors. Covariate and spatial parameters are simultaneously estimated for each time point.

**Step 3**. **Modeling of the Residuals.**

For each $Y_{it}$, compute the residuals $e_{it} = Y_{it} - \hat{\beta}^{*}X_{it} - \widehat{\delta_t}^{*}W_{it}$. The residuals still contain information on the true error and the temporal parameter. For each location, an AR(1) model is estimated to get the temporal parameter estimates $\hat{\rho}_i's$.

**Step 4**. **Updating the Dependent Variable**.

In this step, we have the set of parameter estimates $\hat{\delta}_1{}^{*}, \hat{\delta}_2{}^{*}, \dots, \hat{\delta}_T{}^{*}, \hat{\beta}^{*}, \hat{\rho}_1{}^{*}, \hat{\rho}_2{}^{*}, \dots, \hat{\rho}_N{}^{*}$. A new dependent variable will be computed by adjusting for the temporal component, given by $Y_{it}{}^{new} = Y_{it}{}^{orig} - \hat{\rho}_i e_{i,t-1}$, where the initial value $e_{i,0} = 0$.

After *Step 4*, the algorithm iterates from *Step 1* using the new values of the dependent variable (less the temporal effect). Focus will be on the covariate-spatial effect when the ordinary least squares is once again in Step 1. In updating the estimates of the error term in *Step 3*, the original values of Y will be used. After updating the error terms and performing *Step 3*, the new dependent variable will once again be computed using the original values of the dependent variable and the updated estimates of the error terms from *Step 3.* The iteration converges when there are minimal changes in the values of the parameter estimates.

## 4. Simulation Study

The model along with the estimated procedure is evaluated using a simulated data in the balanced (N=T) and unbalanced (T<N, N>T) scenarios. The true model is given below:

$$Y_{it} = \beta X_{it} + \delta_t W_{it} + \epsilon_{it} \tag{6}$$

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + a_{it}, \quad |\rho_i| < 1, \quad a_{it} \sim IID(0, \sigma^2{}_a) \tag{7}$$

The covariate $X_{it}$ was generated from N(100, 10) while the spatial variable was generated from the Poisson distribution with varying parameters. Only one covariate and spatial variable were generated. Five and ten neighborhoods were generated. In case of 5 neighborhoods, means of the Poisson distribution are 10, 20, 30, 40 and 50. In case of 10 neighborhoods, the parameters are 10, 20, 30, …, 80, 90, 100. It should be noted that the elements in the neighborhoods are pre-determined since the neighborhood variables are usually geographic variables (Landagan and Barrios, 2007).

For the balanced data case, two settings are simulated: small data set, and large data set. For small data set, N=T=20 while for large data set, N=T=50. For the unbalanced case, the setting is either N=20 and T=50 or N=50 and T=20. Balanced and unbalanced cases in panel data models is investigated by Arellano and Bond (1991) in the context of a dynamic panel data. Panel data are special cases of spatio-temporal phenomenon since we have multiple observations of the same units over time. In panel data model, they postulated that T is small and N is large, both for the cases when T is the same for each unit and when T will vary from unit to unit. Model was estimated through generalized method of moments and the specification tests and tests of serial correlation with their asymptotic distributions gave negligible finite sample biases and smaller variances than with simpler instrumental variable estimators. Also, asymptotic counterparts give good approximations of the distribution of the serial correlation tests. Arellano and Bond (1991) noted that for the method works when N > T. On the other hand, Xu and Wikle (2007) noted that inferences in spatio-temporal statistical models would need a large number of spatial units, either N > T or N is large. In this study, the case when T > N and when N is small will be explored. This is very useful in application since many real-life data have more time points than spatial units.

Another setting is the number of predetermined neighborhoods: 5 or 10. The number of neighborhoods will show the performance of the estimation procedure when the population is divided to small number of neighborhood systems. For a fixed number of spatial units, a higher number of neighborhoods will divide the spatial units into more groupings. Arcenaeux and Nickerson (2009) stressed the importance of recognizing clustering of observations. There is a severe problem in the precision and biases of parameter estimates when ignoring clustering especially when it explains a large portion of the variance in the response variable. Efficiency is gained with the addition of clusters than with the addition of observations.

**Simulating Structural Change**

The simulation study aims to recreate the dynamics of agricultural production with structural change. The structural change is simulated by altering the parameter values on the affected time points and affected spatial units. The structural change is bounded by the following features: contamination in the recent period of the time series vs. contamination in the middle of the time series, structural change in all neighborhoods in the time points affected by the contamination versus structural change in 20% of the neighborhoods (if 5 neighborhoods, only 1 is affected; if 10 neighborhoods, 2 are affected), and the change in the parameter to be 10% vs. 20%.; and contamination in the covariate parameter only, or contamination in the spatial parameter only, or contamination in both parameters.

The scope of contamination will be facilitated by making all neighborhoods affected by the contamination versus having only a subset of neighborhoods contaminated. The settings where only 20% of the neighborhoods is contaminated may be viewed as the cases wherein the structural change occurred within specific locations. The presence of the contamination is either at the middle time points or at the recent time points of the series. The percentage of time points affected by the contamination is set to be 10%.

Information contributed by the predictors on the response variable is considered using three levels: equal share of information from the covariate and spatial variable, higher share for the covariate, and higher share for the spatial variable. A higher share of covariate is implemented by assuring that its percentage of contribution on the response variable is around 75% while the contribution of the spatial component is around 25%, and vice versa. When there is equal share of information on the covariate and spatial component, the percentage of information is approximately 50% for each.

Table 1 summarizes the simulation settings and scenarios. Furthermore, constant covariate parameter and variance of the error term are allowed to vary in the different scenarios to satisfy requirements on the share of information for each predictor. Since the spatial parameters varies across space and the temporal parameter varies over time, the parameter values are fixed for the different scenarios on the number of locations and time points.

**Table 1: Table on the different simulation settings**

| Settings | Scenarios |
|---|---|
| Number of Time Points | 20, 50 |
| Number of Spatial Units | 20, 50 |
| Number of Neighborhoods | 5, 10 |
| Presence of the Contamination | Recent Period of the Series vs. Middle Period of the Series |
| Scope of the Contamination | All neighborhoods vs. 20% of the neighborhoods |
| Degree of Contamination | 10% change in parameters vs. 20% Change vs. 0% Change which is a case of no contamination |
| Parameters Contaminated | Covariate only, Neighborhood variable only, Both |
| Share of Information on Predictors | Higher share for the covariate vs. Higher share for the neighborhood variable vs. equal share |

## 5. Results and Discussion

We present results of evaluation of the model/estimation procedures for the dynamic spatio-temporal model with or without structural change as well as the effect of number of clusters and the extent of contamination causing structural change. The proposed algorithm is also compared to the ordinary linear model (estimated through ordinary least squares) in terms of the MAPE, relative bias of the estimates, and standard error of the estimates for the different simulation scenarios.

### 5.1 Absence of Structural Change

In this section, predicted ability, relative bias, standard errors, etc., are presented both for the proposed method vis-a-vis ordinary linear model estimated using ordinary least squares.

***Predictive Ability***

We summarized in Table 2 MAPE and standard errors of the parameter estimates generated by the proposed algorithm and ordinary linear model, while Table 3 indicates the relative bias of the parameter estimates. Even without structural change, predictive ability of the postulated model (estimated based through proposed algorithm) is superior to ordinary linear model that fails to account some aspects of dependencies included in spatio-temporal model. Predictive ability of ordinary linear model improves when the covariate has a higher contribution to the variation of the dependent variable (share of information) relative to the spatial component. It's the exact opposite for the proposed algorithm where predictive ability is better when spatial component shares higher to the total variation in the dependent variable. In the proposed algorithm, data set with less time points than spatial units have higher predictive ability, similar is true for large data sets (more time points, more spatial units). Ordinary linear model exhibits better predictive ability when there are fewer neighborhoods. The proposed algorithm however, yield better predictive ability with more neighborhoods since the model aptly accounts for clustering that the data set may have exhibited.

**Table 2. Comparison of standard error and MAPE in the absence of structural change**

| | Proposed Algorithm | | | Ordinary Linear Model | | |
|---|---|---|---|---|---|---|
| **Scenarios** | $S.E.(\hat{\beta})$ | $S.E.(\hat{\delta})$ | **MAPE** | $S.E.(\hat{\beta})$ | $S.E.(\hat{\delta})$ | **MAPE** |
| **Overall** | 0.1273 | 0.1205 | 5.7851 | 0.2098 | 0.4442 | 41.3451 |
| **5 Neighborhoods** | 0.1002 | 0.1323 | 5.9079 | 0.1522 | 0.4570 | 40.4319 |
| **10 Neighborhoods** | 0.1544 | 0.1087 | 5.6623 | 0.2674 | 0.4314 | 42.2583 |
| **N=T=20** | 0.1885 | 0.1833 | 6.2434 | 0.3187 | 0.6757 | 34.2269 |
| **N=T=50** | 0.0707 | 0.0641 | 5.2293 | 0.1238 | 0.2615 | 48.7935 |
| **N=50, T=20** | 0.0682 | 0.0648 | 4.7617 | 0.1976 | 0.4175 | 40.5652 |
| **N=20,T=50** | 0.1819 | 0.1698 | 6.9060 | 0.1992 | 0.4222 | 41.7948 |
| **Equal Share** | 0.1109 | 0.1045 | 5.7712 | 0.2094 | 0.4432 | 40.5302 |
| **Higher share of** $\beta X$ | 0.2017 | 0.1925 | 6.0388 | 0.2112 | 0.4473 | 23.9098 |
| **Higher share of** $\delta W$ | 0.0694 | 0.0645 | 5.5453 | 0.2089 | 0.4421 | 59.5953 |

**Table 3. Comparison of relative bias in the absence of structural change**

| | Proposed Algorithm | | | Ordinary Linear Model | |
|---|---|---|---|---|---|
| **Scenario** | $\mathbf{Rel.\,bias}(\hat{\boldsymbol{\delta}})$ | $\mathbf{Rel.\,bias}(\hat{\boldsymbol{\beta}})$ | $\mathbf{Rel.\,bias}(\hat{\boldsymbol{\rho}})$ | $\mathbf{Rel.\,bias}(\hat{\boldsymbol{\delta}})$ | $\mathbf{Rel.\,bias}(\hat{\hat{\boldsymbol{\beta}}})$ |
| **Overall** | 3.3006 | 1.5491 | 37.9079 | 91.4870 | 1.5030 |
| **N=T=20** | 4.0097 | 2.2103 | 49.3289 | 78.2969 | 2.0999 |
| **N=T=50** | 2.2358 | 0.9583 | 27.1299 | 104.6778 | 0.9328 |
| **N=50, T=20** | 1.6082 | 1.5031 | 43.3382 | 78.4170 | 1.3773 |
| **N=20,T=50** | 5.3489 | 1.5249 | 31.8347 | 104.5562 | 1.6019 |
| **5 Neighborhoods** | 3.6287 | 1.5816 | 38.1204 | 91.4788 | 1.5644 |
| **10 Neighborhoods** | 2.9726 | 1.5167 | 37.6954 | 91.4951 | 1.4415 |
| **Equal Share** | 2.8473 | 1.3620 | 37.8866 | 91.4912 | 1.3274 |
| **Higher share of** $\boldsymbol{\beta X}$ | 5.3025 | 0.9729 | 37.9262 | 91.4628 | 0.9115 |
| **Higher share of** $\boldsymbol{\delta W}$ | 1.7521 | 2.3126 | 37.9109 | 91.5068 | 2.2700 |

***Relative Bias and Standard Errors***

The relative bias of the spatial parameter estimates from the proposed algorithm is consistently smaller compared to that of the ordinary linear model. However, it is the opposite for the covariate parameter. This is explained by the fact that ordinary least squares gives the best linear unbiased estimator (BLUE) on the assumption that the model is correctly specified and thus the relative bias from fitting a series of ordinary linear models per time point is smaller compared to that of the proposed algorithm. For the ordinary linear model, relative bias on the spatial parameter exceeds 100% when the length of the time series is long. Relative bias on the covariate parameter estimates is higher on the average when the spatial component has higher share of information than when the covariate has higher share of information. For the proposed algorithm, more neighborhoods gives smaller relative bias on the parameter estimates than with fewer neighborhoods. Moreover, relative bias on the parameter estimates are smaller when the data set is large or when there are more spatial units than time points. Relative bias of the estimate of the temporal parameter is high as expected since it is the last to be estimated in the backfitting algorithm. Nonetheless, the relative bias is smaller for longer time series compared to shorter time series.

The proposed algorithm also gave smaller standard error on the estimates compared to the ordinary linear model. More neighborhoods led to higher standard errors for the covariate parameter estimates and smaller standard errors for the spatial parameter estimates compared to the case when there are fewer neighborhoods. The standard errors are also smaller when there are more spatial units than time points or when the data set is large. In the proposed algorithm, the standard errors for parameter estimates are relatively small on the average when the spatial component has higher share of information and are relatively large when the covariate has higher share of information.

### 5.2 Presence of Structural Change

***Predictive Ability***

Table 4 and Table 5 shows the comparison of MAPE between the proposed algorithm and ordinary linear model for different simulation scenarios and the interaction between share of information and which parameters are contaminated on the MAPE, respectively. Predictive ability of the proposed algorithm are generally superior to ordinary linear model, specially if there are more spatial units than time points. When the data set is large and balanced, predictive ability is also better compared to the case of small data sets and unbalanced data set (more time points than spatial units).

**Table 4. Comparison of MAPE in the presence of structural change**

| Scenarios | Proposed Algorithm | Ordinary Linear Model |
|---|---|---|
| **Overall** | 8.0323 | 32.4043 |
| **5 Neighborhoods** | 8.0282 | 32.2846 |
| **10 Neighborhoods** | 8.0364 | 32.524 |
| **N=T=20** | 8.1435 | 28.1476 |
| **N=T=50** | 8.0337 | 36.698 |
| **N=50, T=20** | 7.7833 | 28.3237 |
| **T=50, N=20** | 8.1686 | 36.4478 |

**Table 5. Effect of 'share of information' and presence of contamination on MAPE**

| | Share of Information | | | |
|---|---|---|---|---|
| **Contamination** | **Equal** | **Higher in $\beta X$** | **Higher in $\delta W$** | **Overall** |
| $\boldsymbol{\beta}$ | 8.0913 | 8.4647 | 7.9156 | 8.1572 |
| $\boldsymbol{\delta}$ | 7.6639 | 7.6114 | 7.7898 | 7.6884 |
| **Both** | 8.1989 | 8.5200 | 8.0350 | 8.2513 |
| **Overall** | 7.9847 | 8.1987 | 7.9135 | 8.0323 |

The predictive ability does not vary over the number of neighborhoods. If there is higher share of information on the covariate, predictive ability is inferior than when there is equal share of information or higher information on the neighborhood variable. When the share of information from the covariate is higher, predictive ability is better when contamination is present only in the neighborhood parameter. When the spatial component has the higher share of information, predictive ability improves when contamination is present only in the neighborhood parameter. When contamination is present only in the covariate parameter, predictive ability generally suffers when the covariate has higher share of information but is better when more information is contained in the spatial component. When contamination is present only in the spatial parameter, predictive ability in inferior when the spatial component has higher share of information. And when contamination is present in both parameter, predictive ability in inferior when the covariate has higher share of information than otherwise, while better predictive when the spatial component has the higher share of information.

***Standard Error***

As can be seen from Table 5*,* standard error of the spatial parameter estimates are all consistently lower for the proposed algorithm than the ordinary linear model, the opposite is true for the covariate parameter estimates. For both of the proposed algorithm and ordinary linear model, the standard errors of the parameter estimates are relatively larger for small data sets and smaller for large data sets.

Regardless of the share of information between the explanatory variables, standard error is higher under the proposed algorithm when the covariate has higher information and smaller when the spatial component has higher share of information. The standard errors of the ordinary linear model generally differ with respect to the share of information. Moreover, the standard errors of the parameter estimates are relatively smaller when only the spatial parameter is contaminated than when the covariate parameter or both parameters are contaminated. The standard errors of the ordinary linear model does not vary according to which parameter is contaminated. Finally, when there are more neighborhoods, the standard error of the covariate parameter estimate increases while standard error of spatial parameters decreases.

**Table 6. Comparison of standard errors of the parameter estimates in the presence of structural change**

| | Proposed Algorithm | | Ordinary Linear Model | |
|---|---|---|---|---|
| **Scenarios** | $\mathbf{S.E.}\,(\hat{\boldsymbol{\beta}})$ | $\mathbf{S.E.}\,(\hat{\boldsymbol{\delta}})$ | $\mathbf{S.E.}\,(\hat{\boldsymbol{\beta}})$ | $\mathbf{S.E.}\,(\hat{\boldsymbol{\delta}})$ |
| **Overall** | 0.2532 | 0.2001 | 0.2145 | 0.4506 |
| **5 Neighborhoods** | 0.1861 | 0.2093 | 0.1562 | 0.4632 |
| **10 Neighborhoods** | 0.3203 | 0.1909 | 0.2728 | 0.438 |
| **N=T=20** | 0.3465 | 0.2901 | 0.3268 | 0.6839 |
| **N=T=50** | 0.1726 | 0.1266 | 0.1261 | 0.2661 |
| **N=50, T=20** | 0.2249 | 0.1548 | 0.2034 | 0.4243 |
| **T=50, N=20** | 0.2688 | 0.2287 | 0.2018 | 0.4279 |
| $\boldsymbol{\beta}$ **−contaminated** | 0.2655 | 0.2006 | 0.2123 | 0.4493 |
| $\boldsymbol{\delta}$ **− contaminated** | 0.2267 | 0.1989 | 0.2148 | 0.4512 |
| $\boldsymbol{\beta}, \boldsymbol{\delta}$ **− contaminated** | 0.2674 | 0.2007 | 0.2165 | 0.4512 |
| **Equal Share** | 0.2074 | 0.1682 | 0.2115 | 0.4475 |
| **Higher in** $\boldsymbol{\beta X}$ | 0.4339 | 0.3320 | 0.2194 | 0.4606 |
| **Higher in** $\boldsymbol{\delta W}$ | 0.1183 | 0.1001 | 0.2127 | 0.4435 |

***Share of Information, Presence of Contamination, Standard Error***

Table 7 illustrates the joint effect of share of information and which parameters are contaminated on the standard error of the covariate parameter estimates. When covariate has higher share of information, the standard error of the covariate parameter estimate is small on the average if the contamination is in the spatial parameter only compared to the case when covariate parameter or both parameters are contaminated. When the spatial component has higher share of information, the standard error for the covariate parameter estimate is slightly smaller on the average when the spatial parameter is contaminated. Regardless on which parameter is contaminated, the standard errors for both parameters are higher when the covariate has higher share of information. Standard errors are lower when the spatial component has higher share of information.

**Table 7. Effect of 'share of information' and presence of contamination on standard error of the covariate parameter estimates**

| | Share of Information | | | |
|---|---|---|---|---|
| **Contamination** | **Equal** | **Higher in $\beta X$** | **Higher in $\delta W$** | **Overall** |
| $\boldsymbol{\beta}$ | 0.2145 | 0.4644 | 0.1176 | 0.2655 |
| $\boldsymbol{\delta}$ | 0.1911 | 0.3734 | 0.1158 | 0.2267 |
| **Both** | 0.2167 | 0.4642 | 0.1214 | 0.2674 |
| **Overall** | 0.2074 | 0.4339 | 0.1183 | |

Table 8 shows the interaction between share of information and which parameters are contaminated on the standard error of the spatial parameter estimates. Standard error of spatial parameter estimates does not vary much according to which parameter/s is/are contaminated. In general, the standard errors for both parameters are higher when the covariate has higher share of information. Standard errors are also lower when the spatial component has higher share of information.

**Table 8. Effect of 'share of information' and presence of contamination on standard error of the spatial parameter estimates**

| | Share of Information | | | |
|---|---|---|---|---|
| **Contamination** | **Equal** | **Higher in $\beta X$** | **Higher in $\delta W$** | **Overall** |
| $\boldsymbol{\beta}$ | 0.1686 | 0.3333 | 0.0999 | 0.2006 |
| $\boldsymbol{\delta}$ | 0.1676 | 0.3289 | 0.1003 | 0.1989 |
| **Both** | 0.1684 | 0.3338 | 0.1001 | 0.2007 |
| **Overall** | 0.1682 | 0.3320 | 0.1001 | |

***Relative Bias***

In Table 9, relative bias of the parameter estimates from the proposed algorithm and the ordinary linear model are presented. Relative bias for the spatial parameter estimates are consistently lower under the proposed algorithm while the relative bias for the covariate parameter estimate is generally lower in ordinary linear model. The temporal parameter estimate has very high relative bias. Relative bias becomes smaller for longer time series. With more spatial units than time points or for large data sets, relative biases are smaller than for small data sets or for cases when there are more spatial units than time points. For cases when there are more neighborhoods, relative bias generally decreases.

**Table 9. Comparison of relative bias in the presence of structural change**

| | Proposed Algorithm | | | Ordinary Linear Model | |
|---|---|---|---|---|---|
| **Scenarios** | $\mathbf{Rel.\,bias(\hat{\delta})}$ | $\mathbf{Rel.\,bias(\hat{\beta})}$ | $\mathbf{Rel.\,bias(\hat{\rho})}$ | $\mathbf{Rel.\,bias(\hat{\delta})}$ | $\mathbf{Rel.\,bias(\hat{\beta})}$ |
| **Overall** | 6.3721 | 1.9782 | 38.5696 | 91.3683 | 1.8101 |
| **5 Neighborhoods** | 6.6552 | 2.0176 | 38.8454 | 91.3493 | 1.8747 |
| **10 Neighborhoods** | 6.0889 | 1.9388 | 38.2938 | 91.3873 | 1.7455 |
| **N=T=20** | 7.3751 | 2.6683 | 49.7352 | 78.1005 | 2.3797 |
| **N=T=50** | 5.2501 | 1.4241 | 28.2770 | 104.6108 | 1.3823 |
| **N=50, T=20** | 4.7551 | 1.8098 | 43.2634 | 78.2723 | 1.5454 |
| **T=20, N=50** | 8.1079 | 2.0106 | 33.0028 | 104.4897 | 1.9329 |
| $\boldsymbol{\beta}$ **−contaminated** | 6.1519 | 2.1378 | 38.5646 | 91.0759 | 1.9351 |
| $\boldsymbol{\delta}$ **−contaminated** | 6.3141 | 1.6034 | 38.2414 | 91.6790 | 1.5037 |
| $\boldsymbol{\beta}, \boldsymbol{\delta}$ **−contaminated** | 6.6502 | 2.1934 | 38.9027 | 91.3501 | 1.9914 |
| **Equal Share** | 5.3414 | 1.8142 | 38.3377 | 91.4403 | 1.6630 |
| **Higher in** $\boldsymbol{\beta X}$ | 10.4776 | 1.5536 | 39.3722 | 91.0797 | 1.3363 |
| **Higher in** $\boldsymbol{\delta W}$ | 3.2972 | 2.5668 | 37.9989 | 91.5850 | 2.4309 |

Relative biases are larger on the average when both parameters are contaminated. When the covariate has higher share of information, relative bias of the covariate parameter estimates is smaller while the relative bias of the spatial parameter estimate is larger. When the spatial component has higher share of information, the relative bias of the covariate parameter estimates is larger on the average while the relative bias of the spatial parameter estimates is smaller on the average.

***Share of Information, Contamination, and Relative Bias***

Tables 10 and 11 shows the interaction between share of information and which parameters are contaminated on the relative bias of the spatial parameter estimates and covariate parameter estimates, respectively.

Regardless of which parameter is contaminated, relative bias of the spatial parameter estimates are large when the covariate has higher share of information and are small when the spatial parameter component has higher share of information. When the spatial component has higher share of information, the relative bias of the spatial parameter estimates are bigger when the spatial parameter is contaminated compared to when the covariate parameter is contaminated. When the covariate has higher share of information, relative bias of the spatial parameter estimates are bigger when the covariate parameter is contaminated than when the spatial parameter is contaminated. When there is equal information on the covariate and spatial component, contamination on the spatial parameter estimates or on both parameters will give larger standard errors of the spatial parameter.

Regardless of which parameter is contaminated, higher information on the spatial component leads to large relative bias for the covariate parameter estimates. Furthermore, regardless of which component has higher share of information, there are on the average larger relative bias on the covariate parameter estimates when the covariate parameter is contaminated.

**Table 10. Effect of 'share of information' and presence of contamination on the relative bias of the spatial parameter estimates**

| | Share of Information | | | |
|---|---|---|---|---|
| **Contamination** | **Equal** | **Higher in $\boldsymbol{\beta X}$** | **Higher in $\boldsymbol{\delta W}$** | **Overall** |
| $\boldsymbol{\beta}$ | 5.0732 | 10.4930 | 2.8896 | 6.1519 |
| $\boldsymbol{\delta}$ | 5.3812 | 10.0802 | 3.4809 | 6.3141 |
| **Both** | 5.5698 | 10.8597 | 3.5210 | 6.6502 |

| | | | | |
|---|---|---|---|---|
| **Overall** | 5.3414 | 10.4776 | 3.2972 | |

**Table 11. Effect of 'share of information' and presence of contamination on the relative bias of the covariate parameter estimates**

| | Share of Information | | | |
|---|---|---|---|---|
| **Contamination** | **Equal** | **Higher in $\beta X$** | **Higher in $\delta W$** | **Overall** |
| $\boldsymbol{\beta}$ | 1.7607 | 1.5814 | 2.3587 | 1.9002 |
| $\boldsymbol{\delta}$ | 1.2797 | 0.9587 | 2.0373 | 1.4252 |
| **Both** | 1.7975 | 1.6028 | 2.4488 | 1.9497 |
| **Overall** | 1.6126 | 1.3810 | 2.2817 | |

***Contamination in All Neighborhoods vs 20% of Neighborhoods Affected***

In Table 12 we compare relative bias and standard errors of the parameter estimates and MAPE between contamination in all neighborhoods and contamination in only 20% of neighborhoods. When contamination is in the covariate parameter only, MAPE are higher on the average when structural change manifest in all neighborhoods. When contamination is in the spatial parameter only, the MAPE are generally higher when structural change is in 20% of the neighborhoods. Similar pattern can observed for their respective relative biases. When the covariate parameter is contaminated, the relative bias on the covariate parameter estimates is higher when all neighborhoods are contaminated. Similarly is true when the spatial parameter is contaminated, the relative bias of the spatial parameter is higher when all neighborhoods are contaminated then when only 20% of the neighborhoods are contaminated. Moreover, when only the covariate parameter is contaminated, covariate parameter estimate has higher standard error when all neighborhoods are contaminated but not the case for the spatial parameter estimate. When only the spatial parameter is contaminated, the spatial parameter has higher standard error when only 20% of the neighborhoods are contaminated, which is a similar pattern for the MAPE.

**Table 12. Comparison of relative bias and standard errors of estimates and MAPE (contamination in all neighborhoods vs. contamination only in 20% of neighborhoods)**

| | | Relative Bias | | | Standard Error | | MAPE |
|---|---|---|---|---|---|---|---|
| | | $\hat{\delta}$ | $\hat{\beta}$ | $\hat{\rho}$ | $\hat{\beta}$ | $\hat{\delta}$ | |
| **$\beta$ −contaminated** | *All NBs Cont.* | 0.0587 | 0.0235 | 0.3886 | 0.2790 | 0.1996 | 8.3105 |
| | *20% NBs Cont.* | 0.0644 | 0.0193 | 0.3827 | 0.2519 | 0.2015 | 8.0039 |
| **$\delta$ − contaminated** | *All NBs Cont.* | 0.0677 | 0.0157 | 0.3800 | 0.2231 | 0.1987 | 7.6575 |
| | *20% NBs Cont.* | 0.0586 | 0.0163 | 0.3848 | 0.2304 | 0.1991 | 7.7192 |
| **$\beta, \delta$ − contaminated** | **Higher in $\beta X$** | | | | | | |
| | *All NBs Cont.* | 0.1071 | 0.0205 | 0.4070 | 0.4949 | 0.3317 | 8.7509 |
| | *20% NBs Cont.* | 0.1101 | 0.0155 | 0.3922 | 0.4334 | 0.3358 | 8.2890 |
| | **Higher in $\delta W$** | | | | | | |
| | *All NBs Cont.* | 0.0407 | 0.0280 | 0.3748 | 0.1200 | 0.0995 | 7.9812 |
| | *20% NBs Cont.* | 0.0297 | 0.0271 | 0.3872 | 0.1229 | 0.1006 | 8.0887 |
| | **Equal** | | | | | | |
| | *All NBs Cont.* | 0.0595 | 0.0219 | 0.3847 | 0.2230 | 0.1681 | 8.2110 |
| | *20% NBs Cont.* | 0.0519 | 0.0185 | 0.3882 | 0.2103 | 0.1686 | 8.1868 |

When both parameters are contaminated and covariate has higher share of information, MAPE are higher when all neighborhoods are contaminated. However, when the spatial component has higher share of information, the MAPE are higher when only 20% of the neighborhoods are contaminated. When there is equal share of information on the explanatory variables, the MAPE are higher when all neighborhoods are contaminated. Covariate parameter estimate has higher relative bias when all neighborhoods are contaminated regardless of the share of information. For spatial parameter estimate, the relative bias is higher for the case of contamination in all neighborhoods when the spatial component has higher share of information or when there is equal share of information. The standard error of the spatial parameter estimate is consistently higher on the average when only 20% of neighborhoods are contaminated regardless of the share of information.

***Structural Change in the Middle Periods vs. Recent Periods***

From Table 13, predictive ability suffers when structural change occurred in the recent periods of the time series. However, standard error of the covariate parameter estimates are higher when structural change is in the middle periods while the standard error of the spatial parameter estimate is higher when structural change is in the recent time points of the series.

**Table 13. Effect of location of structural change on standard error of the parameter estimates and MAPE**

| Period of the Series | S.E. $(\hat{\beta})$ | S.E. $(\hat{\delta})$ | 4 | MAPE |
|---|---|---|---|---|
| **Recent** | 0.2486 | 0.2074 | | 8.3382 |
| **Middle** | 0.2578 | 0.1927 | | 7.7217 |

## 6. Conclusions

The dynamic spatio-temporal model estimated through the backfitting algorithm imbedded with the forward search algorithm and bootstrap is capable of representing process that exhibit dynamic interaction of space and time. The fitted model also exhibits robustness in the presence of temporary structural change.

The proposed algorithm works well when there are more spatial units than time points or when the data set is large. The relative bias and standard errors of the spatial parameter estimates are consistently lower for the proposed algorithm, but not for the covariate parameter estimates. The standard error of the covariate parameter estimates are lower when there is equal share of information between the covariate and spatial component or when the spatial component has higher share of information.

Although more neighborhoods will not have an apparent improvement in the predictive ability of the model, an advantage of having more neighborhoods is manifested in terms of smaller relative bias on the parameter estimates, becoming more prominent for spatial parameter estimates. Moreover, the standard error of the spatial parameter estimate decreases on the average, but increases for the covariate parameter estimates.

Contamination in all neighborhoods does not necessarily imply that the relative biases and MAPE are higher, than when contamination is limited to some neighborhoods. When only the covariate parameter is contaminated, the MAPE and relative bias of the covariate parameter estimate is higher when all neighborhoods are contaminated compared to cases where only some neighborhoods are contaminated. When only the spatial parameter is contaminated, MAPE are higher on the average when only some neighborhoods are contaminated.

Predictive ability of the model and the standard error of the parameter estimates deteriorates when structural change occurs in the recent periods in the series. The relative bias of the spatial parameter estimates are also larger when structural change occurs in the recent periods of the series.

## REFERENCES


Arellano, M., Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*. 58(2): 277-297.

Bastero, R., Barrios, E. (2011). Robust Estimation of a Spatiotemporal Model with Structural Change. *Communication in Statistics – Simulation and Computation*. 40: 448-468

Cressie, N (1993). **Statistics for Spatial Data**. A Wiley-Interscience Publication. Wiley, New York.

Cressie, N. and Wikle, C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*. 94(448): 1330-1340.

Cressie N., and Wikle, C., (2002) Space-time kalman filter. *Encyclopedia of Environmetrics*, Vol.4. Wiley, New York, pp. 2045-2049.

Dumanjug, C., Barrios, E., Lansangan J. (2010). Bootstrap procedures in spatiotemporal model. *Journal of Statistical Computation and Simulation*. 80:809-822.

Guarte, J., Barrios, E., (2013). Nonparametric Hypothesis Testing in a Spatial-Temporal Model: A Simulation Study. *Communications in Statistics – Simulation and Computation*. 42: 153-170.

Hackl, P., and Westlund A. H. (1989). Statistical Analysis of "Structural Change": An Annotated Bibliography. *Empirical Econom*ics. 14(2):167-192.

Kyriakidis, P., and Journel, A. (1999). Geostatistical space-time models: a review. *Mathematical Geology*. 31(6): 651-684.

Landagan, O., Barrios, E. (2007). An Estimation Procedure for Spatio-temporal Model. *Statistics and Probability Letters*. 77:401-406.

Mardia, K., Goodall, C., Redfern, E., Alonso, F. (1998). The kriged kalman filter. *Test*. 7:217-285 (with discussion).

Shumway, R. H., Stoffer, D. S. (2000). **Time Series Analysis and its Applications**. Springer, New York.

Xu, K. and Wikle, C. (2007). Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference*. 137:567-588.