
Compare and Evaluate the Performance of Gaussian Spatial Regression Models and Skew Gaussian Spatial Regression Based on Kernel Averaged Predictors

Somayeh Shahraki Dehsoukhteh

Department of Statistics, Faculty of Sciences, Zabol University, zabol, Iran

Email address:

Sshahraki7@gmail.com (S. S. Dehsoukhteh)

To cite this article:

Somayeh Shahraki Dehsoukhteh. Compare and Evaluate the Performance of Gaussian Spatial Regression Models and Skew Gaussian Spatial Regression Based on Kernel Averaged Predictors. *American Journal of Theoretical and Applied Statistics*. Vol. 4, No. 5, 2015, pp. 368-372. doi: 10.11648/j.ajtas.20150405.17

Abstract: In many problems in the field of spatial statistics, when modeling the trend functions, predictors or covariates are available and the goal is to build a regression model to describe the relationship between the response and predictors. Generally, in spatial regression models, the trend function is often linear and it is assumed that the response mean is a linear function of predictor values in the same location where the response variable is observed. But, in real applications, the neighboring predictors sometimes provide valuable information about the response variable particularly when the distance between the locations is small. Having considered this subject matter, Heaton and Gelfand [6] suggested using kernel averaged predictors for modeling trend functions in which neighboring predictor information are also used. The models proposed by Heaton and Gelfand seemed to be bound by data normality. So, in many more application problems, spatial response variables follow a skew distribution. Therefore, in this article, skew Gaussian spatial regression model is studied and the performance of the model is presented and evaluated in comparison with Gaussian spatial regression models based on kernel averaged predictors using simulation studies and real examples.

Keywords: Spatial Regression, Kernel, Skew Normal

1. Introduction

So far, statistic methods and different models have been presented for the analysis of spatial data. The basics of these models and methods can be seen in various books including [2,3,4,5,6,11]. In problems in the field of spatial statistics a response variable is measured in different locations in the area under study. Response variables are dependent in space so that close observations in space have much more correlation than those of the farther. It is assumed that for the continuous responses the residuals are normal. But in many application examples, spatial variables follow a skew distribution. One common method for analysis of such data skew normal distribution. Different generalization of this distribution are presented by [1,7]. Since using this modeling method has some difficulties, Zhang and El-shaarawi [12] analysed skew spatial data in another way while presenting a regression model. In this model, trend function is written as a linear function of the predictor values in the same location of response variable. But in real application, the neighboring

predictors sometimes provide valuable information about the response variable particularly when the distance between locations is small. So, in this situation, considering mean based only on the predictor variable value in the same location is not enough and it is also necessary to use neighboring information. Heaton and Gelfand [8,9] presented application method of neighboring information in spatial regression model with normal errors. In this article, the method they have proposed for skew Gaussian regression model is generalized. Then, using simulation and application examples, performance of this model is compared and evaluated to the model introduced by Heaton and Gelfand.

2. Spatial Regression Model Base on Kernel Averaged Predictors

The spatial regression model is presented as follows:

$$Y(s) = m(s) + \sigma_2 \omega_Y(s) + \sigma_0 \epsilon_Y(s)$$

Where $Y(s)$ is univariate response at location $s \in D \subseteq R^d$, $d \geq 1$, and $m(s)$ is a non-random function depending on s and used for modeling mean function. $\omega_Y(s)$ is a zero mean, unit variance GP with correlation function $\rho_Y(s)$ and $\epsilon_Y(s)$ is a zero mean Gaussian white noise process with unit variance. This process is considered to study measurement error in the model $\omega_Y(s)$ and $\epsilon_Y(s)$ are viewed as independent processes. $m(s)$ mean function is usually written as linear combination of predictor variables. Assume, we have a predictor variable $X(s)$. So,

$$m(s) = \beta_0 + \beta_1 X(s) \tag{1}$$

Where β_0, β_1 are regression coefficients. In (1) we saw that trend phrase is a function of predictor variable in the same location of s . But, as it was mentioned in the introduction section, we are going to apply information of neighboring location in the mean function structure. To achieve the aim, we use kernel averaged predictors on the whole area of study according to methods proposed by Heaton and Gelfand [8,9]. To show the method, assume $X(s)$ follows a Gaussian processes (GP) of the form,

$$X(s) = \mu_X(s) + \sigma_X \omega_X(s) \tag{2}$$

Where $\mu_X(s)$ is the mean surface location s and $\omega_X(s)$ is a zero mean, unit variance GP with correlation function $Corr(\omega_X(s), \omega_X(s')) = \rho_X(s, s' | \varphi_X)$, where φ_X denotes the parameter associated with ρ_X . Unobserved local covariate at s in incorporating information as $\tilde{X}(s)$ using a kernel function, i.e.

$$\tilde{X}(s) = \frac{1}{K(s|\xi)} \int_D K(s, u|\xi) X(u) du \tag{3}$$

Where $K(s, s'|\xi)$ is a kernel defining a weight on the distance between s , and s' with parameters ξ , $K(s, s'|\xi) \geq 0$ for all s, s' and $\int_D K(s, u|\xi) du = 1$. Because a valid GP was defined for $X(s)$, $\tilde{X}(s)$ is also a valid GP with mean

$$\mu_{\tilde{X}}(s) = \int_D K(s, u|\xi) \mu_X(u) du$$

And

$$\begin{aligned} cov(\tilde{X}(s), \tilde{X}(s')) &= \frac{\sigma_X^2}{K(s|\xi)K(s'|\xi)} \int_D \int_D K(s, u|\xi) K(s', u|\xi) \rho_X(u, v) dudv \\ &\equiv \sigma_{\tilde{X}}^2 \rho_{\tilde{X}}(s, s') \end{aligned}$$

Therefore, to account for effects of $\{X(s'): s' \in D\}$ on $Y(s)$, consider the linear model defined by

$$Y(s)|X(s), \tilde{X}(s) = \beta_0 + \beta_1 \tilde{X}(s) + \sigma_2 \omega_Y(s) + \sigma_0 \epsilon_Y(s) \tag{4}$$

3. Skew Gaussian Spatial Regression Based on Kernel Averaged Predictors

Assume that response variable is non-normal one, skew

Gaussian spatial regression model is presented as follows:

$$Y(s) = m(s) + \sigma_1 |\omega_{Y_1}(s)| + \sigma_2 \omega_{Y_2}(s) + \sigma_0 \epsilon_Y(s)$$

Where $Y(s)$, and $m(s)$ are defined according to section two. $\sigma_2 \geq 0$, $\sigma_0 \geq 0$, and σ_1 are true values. $\omega_{Y_1}(s)$ is also a stationary Gaussian random field with a zero mean, unit variance and correlation function $\rho_{Y_2}(h)$. Three processes of $\omega_{Y_1}(s)$, $\omega_{Y_2}(s)$ and $\epsilon_Y(s)$ are considered independent. Simply, it is illustrated that $\frac{Y(s)-m(s)}{\sqrt{\sigma_0^2 + \sigma_1^2 + \sigma_2^2}}$ has skew normal probability density as $2\varphi(x)\Phi(\alpha x)$ where $\alpha = \frac{\sigma_1}{\sqrt{\sigma_0^2 + \sigma_2^2}}$. As it has been shown α has a direct relationship with σ_1 , then σ_1 determines type of data skewness. i.e. $Y(s)$ distribution when $\sigma_1 > 0$, is skew right, $\sigma_1 = 0$, symmetric, and when $\sigma_1 < 0$, skew is left. The mean and correlation of random field $Y(\cdot)$ is also as follows:

$$\mu(s) = EY(s) = m(s) + \sigma_1 \sqrt{\frac{2}{\pi}}$$

$$\begin{aligned} C(h) = \frac{2\sigma_1^2}{\pi} &\left(\sqrt{1 - \rho_{Y_1}^2(h)} + \rho_{Y_1}(h) \arcsin(\rho_{Y_1}(h)) - 1 \right) \\ &+ \sigma_2^2 \rho_{Y_2}(h) + \sigma_0 I_{\{h=0\}} \end{aligned}$$

Therefore, to taking account the effects of neighboring covariates on response variable, skew Gaussian random field is as follows:

$$Y(s)|X(s), \tilde{X}(s) = \beta_0 + \beta_1 \tilde{X}(s) + \sigma_1 |\omega_{Y_1}(s)| + \sigma_2 \omega_{Y_2}(s) + \sigma_0 \epsilon_Y(s) \tag{5}$$

4. Comparison of Models Using Simulation Examples

Assume that sample size in $n \in \{50, 100\}$, data Gaussian model (4) with $\sigma_X^2 = 1$ and $\beta = (0, 1)^T$ and correlation function $K(s, s'|\xi) = I_{\{\|s-s'\| \leq r\}}$ where $\xi \in \{0, 0/1, 0/3, 0/5\}$ is simulated using R. Note that sampling plan is selected at random and exponential correlation function of

$$\rho_X(s, s'|\varphi_X) = \exp(-\varphi_X \|s - s'\|) \tag{6}$$

Is used with correlation parameter of $\varphi_X = 10$. We assume that location of observed responses and predictors were aligned with and confined to unit square in R^2 but D was taken to be all of R^2 so as to avoid difficulty in dealing with locations near boundary. Note that when $\xi = 0$ resulting in spatial regression model of point traditional predictor (PTP), i.e. $E(Y(s)|\theta, \tilde{X}) = \beta_0 + \beta_1 X(s)$. For each combination of (n, ξ) 20 data sets were simulated using an additional 25 values of $Y(s)$ left as a hold-out sample to determine predictive performance of the fitted models. Assume $B(\hat{\beta}_1)$, $MSE(\hat{\beta}_1)$, $CIC(\hat{\beta}_1)$ as the observed bias of the posteriori mean $\hat{\beta}_1$, mean square error of $\hat{\beta}_1$, and empirical %95 credible interval coverage for β_1 respectively. Furthermore,

we define $MSPE = \sum_{j=1}^{20} \sum_{i=1}^{25} (y^{(i)}(s_i) - \hat{y}^{(j)}(s_i))^2 / (25 \times 20)$ as the average predictive mean square error width across all of the 25 hold-out values where $y^{(i)}(s_i)$, i^{th} is left observed from j^{th} data set and $\hat{y}^{(j)}(s_i)$ is prediction value of it. To compare, two models were fitted to data. The first model was the kernel averaged predictor (KAP) given by (4) and the second one was a point traditional predictor (PTP). Discrete prior distributions for φ_X and ξ were used with mass at (5, 10, 15, 20) and (0, 0/1, 0/3, 0/5) respectively. Vague, but proper, conjugate prior distributions were assumed for the remaining parameters. Chains were run for an initial burn in period of 50000 draws and the following 5000 were retained as draws from the posterior distribution. Table 1 shows results of the model fitness. A sit is indicated in the table, and considering $B(\hat{\beta}_1)$, when $\xi = 0$, model bias of KAP is outstanding (for $n=50$, bias is 0/19, and for $n=100$ bias is 0/16) as it is expected estimation of PTP model is relatively bias (for $n=50$, bias is -0/03; and for $n=100$, it is -0/01). But when ξ increases, $B(\hat{\beta}_1)$ value also increases for PTP model. For instance, even for relatively small value of $\xi = 0/1$,

$B(\hat{\beta}_1)$ value under PTP for $n=50$ and $n=100$ equals to -0/19 and -0/28 respectively which are significant values. Furthermore, when ξ increases, $MSE(\hat{\beta}_1)$ value also increases for PTP. For example, for $\xi = 0/3$, $MSE(\hat{\beta}_1)$ value in PTP model for $n=50$ equals to 0/42 and 0/50 respectively; while this standard for KAP is 0/1 and 0/08 respectively. Besides that, taking account CIC, when true model is PTP, CIC value for KAP of both two sizes of $n=50$, and $n=100$ equals to 0/33 and 0/3 respectively which indicates that performance of the model is poor. (Note that for PTP these values equals to 0/90 and 0/97 respectively). But when $\xi > 0$, CIC value of KAP gradually becomes more when ξ increases and this indicates that the model performs well. (β_1 is estimated accurately). Having considered CIC value, the performance of PTP becomes weak quickly when ξ increases. Note that even for relatively small amount of $\xi = 0/1$, probable coverage of PTP for $n=50$ and $n=100$ equals to 0/17 and 0/05 respectively which are not suitable values. Having a general look at the table, we can say that PTP when $\xi > 0$ for different (n) values does not present such logical answers.

Table 1. Estimation performance comparison of the mean kernel averaged (KAP) and the point predictor models based on simulated data.

		50				100			
ξ	model	0	0/1	0/3	0/5	0	0/1	0/3	0/5
$B(\hat{\beta}_1)$	KAP	0/19	-0/03	-0/18	0/23	0/16	-0/06	-0/20	-0/26
	PTP	0/03	-0/19	-0/43	-0/56	-0/01	-0/28	-0/40	-0/48
$MSE(\hat{\beta}_1)$	KAP	0/06	0/08	0/10	0/12	0/06	0/04	0/08	0/10
	PTP	0/04	0/08	0/42	0/72	0/04	0/08	0/50	0/80
$CIC(\hat{\beta}_1)$	KAP	0/33	0/90	0/95	0/90	0/30	0/94	0/95	0/97
	PTP	0/90	0/17	0/06	0/09	0/97	0/05	0/01	0/05

Table 2. Predictive performance comparison of KAP and PTP based on simulated data.

		50				100			
ξ	model	0	0/1	0/3	0/5	0	0/1	0/3	0/5
MSPE	KAP	0/14	0/09	0/11	0/08	0/13	0/04	0/08	0/05
	PTP	0/05	0/12	0/19	0/20	0/03	0/19	0/23	0/24

Table 2 shows prediction results for the two models. As it is observed when $\xi > 0$ KAP model has great value regarding MSPE standard. For example, when $\xi = 0/1$, MSPE of KAP for $n=50$ and $n=100$ are 0/09 and 0/04 respectively while for PTP, they are 0/12 and 0/19 respectively which show weak performance of the model.

The important point resulting from simulation is that in correct use of PTP and KAP can result in unsuitable answers. Never the less, incorrect using of PTP instead use of KAP make the results significantly invalid (especially for high value of ξ). Therefore, it is necessary to pay much more attention to model selection in an application example.

5. Application Example

Air pollution refers to the existence of each kind of pollutant in the air, being either solid, liquid, gas or radioactive and non-radioactive radiation; so that the amount and length of their presence in the air endangers quality of life for human and other beings, and damages to ancient relics and assets. Considering researches that have been conducted in this field, carbon monoxide (CO) is one of the pollutants that causes

greatest damage to humans and animals. Carbon monoxide is a colorless and odorless gas, extremely poisonous which is produced by incomplete combustion of fossil fuels. In the process of organic material combustion, the amount of oxygen for combustion on is not enough, therefore, carbon monoxide is produced out of it. In fact, since this gas has negative effects on respiratory metabolism and on brain activities, so modeling and zoning values of co attracted a lot of attention to control and reduce it. Because air pollution is one of the major problems in Tehran metropolis, thus we consider this metropolis as a region for our study. It is necessary to mention that based on air quality control company, about one million and three hundred and fifty four thousand tons of carbon monoxide pollutant go into Tehran air annually. In this article we consider Co because of its great importance and its harmful effects. Data reviews show daily average amount of Co density per ppm related to the first of December from 2010 to January 23rd in 2011. And environment organization and quality control company for Tehran air have measured and recorded them in 37 stations of air pollutant measurement. Note that since some of the stations had technical problems, they did not

record any information. Therefore, from among 37 existing stations, data of only 16 station have been available. Since there is measurement and record error in data gathering, so it is logical to assume views as noised. One of the other effective factors influencing on air pollution including amount of co density is temperature. One major goal of this example is to study the amount of temperature effect on Co density. Notice that temperature data have been measured in 7 stations from among 16 stations so, we encounter with an misalignment problem. Assume $Y(s_i)$ and $X(s_i)$ show average amount of Co density and temperature amount in s ; location at 62 days respectively. To study data normality we use Shapiro-wilk test. Because p-value of test is nearly equals to 0/0086 and 0/1002 for response and predictive values; the hypothesis of data normality has not been proved, but there is no reason to reject this hypothesis for predictive variable. Therefore, four models are fitted to data: point traditional predictor with normal error (NPTP), kernel averaged predictor with normal error (NKAP), point traditional predictor with skew normal error (SNPTP) and kernel averaged predictor with skew normal error (SNKAP). It is necessary to mention that exponential correlation function was used for each one of the models. Doing MCMC algorithm and after studying required graphs, 20000 was determined as burn time. Then, 50000 samples were extracted from the posterior distribution. Out of each 10 obtained samples, one sample has been taken as the final one. This means that in the end 5000 samples were used for inferences. It is worthy of mentioning that with sensitivity analysis, it was determined that results of a posterior do not have much sensitivity than super parameter change. To choose better model from among selected models, there are different evaluation criteria; in this article we use Deviance Information Criterion and Cross-Validate Criterion to compare models. DIC for each of the models has been presented in table 3.

Table 3. DIC value of model compared.

Model	DIC
NPTP	1488/275
SNPTP	1471/036
NKAP	1465/117
SNKAP	1428/746

According to the table, it is indicated that DIC value for NKAP is fewer than NPTP, its value for SNKAP is fewer than SNPTP. Since DIC statistic shows deviance from true model, therefore, fewer value of this statistic indicates that fitted model is better. Taking account value of this statistic is the fewest value corresponding to SNKAP, we can claim that the model is better than the other models compared. Besides that, table 4 shows estimate point and %95 confidence interval for β_1 parameter.

Table 4. Point and Interval Estimation of β_1 parameter for different models.

Model	point estimation	%95 confidence interval
NPTP	0/01	(-0/04,0/06)
SNPTP	0/03	(-0/02,0/08)
NKAP	0/28	(0/24,0/32)
SNKAP	0/8	(0/11,0/5)

Based on this table we can conclude that the significant effect of temperature on Co amount for NKAP especially SNKAP. But we do not see such thing in the two other models. In other words, in KAP, using $X(s)$ instead of $\tilde{X}(s)$ reduces predictor effect on the response. MSE_{CV} for each model is presented in table 5.

Table 5. MSE_{CV} value of compared models.

Model	MSE_{CV}
NPTP	2/401
SNPTP	2/280
NKAP	1/849
SNKAP	1/327

According to this table, it is indicated that MSE_{CV} value of NKAP is fewer than that of NPTP; and its value for SNKAP is fewer than that of SNPTP. Having considered that MSE_{CV} value of SNKAP is the fewest value of all, it is claimed that this model is better than the other models compared.

6. Conclusions

In this article, we used kernel averaged predictors in modeling the trend function of spatial regression. Kernel is based upon weight between locations and is applied to describe the effect of covariate on response variable. The kernel was taken as parameter so that their function form was clear but dependent on unspecified parameters. Therefore, unobserved local covariate using intended kernel function in each location was defined in a manner to consider neighbor information. Important feature of this approach is to use neighbor information in the analysis and inference of the model without observing covariate variable. In application and simulation examples was shown that spatial regression model based on kernel averaged predictors has more effective performance than traditional spatial model, and it could display reasonable estimation of regression coefficient and suitable prediction. While taking account of skew normal distribution for error terms can get better results. Therefore, the following suggestions can be used for further research:

- We used exponential correlation function in application and simulation examples, while it is possible to we other correlation functions including matern [10] and compare their performance.
- We presented, skew Gaussian spatial using kernel averaged predictors while assuming predictor process as Gaussian. But, we can generalize this approach in a more general way so that both response and predictor processes are Gaussian.
- In all models, we considered one predictor variable. while this approach can be generalized in cases where we face some predictor variables.

Acknowledgments

We thank to zabol university to support this project.

References

- [1] Arellano-Valle, R. B. and Azzalini, A., "On the Unification of Families of Skew-normal Distributions", *Scandinavian Journal of Statistics*, 33,561-574, (2006).
- [2] Banerjee, S. and Carlin, B. P. and Gelfand, A. E., "Hierarchical Modeling and Analysis for Spatial Data", Boca Raton: Chapman and Hall/CRC, (2004).
- [3] Chiles, J. P. and Delfiler, P., "Geostatistics: Modeling Spatial Uncertainty". New York: Wiley Inter-science, (1999).
- [4] Cressie, N., "Statistics for Spatial Data, Revised edition", John Wiley, New York, (1993).
- [5] Cressie, N. and Wikle, C., "statistics for Spatio-Temporal Data", New York: Wiley, (2011).
- [6] Diggle, P. J. and Ribeiro, Jr. P.J., "Model-Based Geostatistics", Springer Series in Statistics, (2007).
- [7] Genton, M. G., "Skew-Symmetric and Generalized Skew-Elliptical Distributions", In Genton, M. G., *Skew-Elliptical Distributions and Their Applications*, chapter 5, 81-100, Chapman and Hall/CRC, (2004).
- [8] Heaton, M. J. and Gelfand, E., "Spatial Regression Using Kernel Averaged Predictors", *Journal Of Agricultural, Biological, and Environmental Statistics*, 16:233-252, (2011).
- [9] Heaton, M.J. and Gelfand, A.E., "Kernel Averaged Predictors for Spatio-Temporal Regression Models," *Spatial Statistics*, 2, 15-32, (2012).
- [10] Matern, B., "Spatial Variation", (2nd,ed), Berlin: Springer, (1986).
- [11] Wakernagel, H. "Multivariate Geostatistics", Berlin: Springer, (2003).
- [12] Zhang, H. and El-Shaarawi, A., "On Spatial Skew-Gaussian Processes and Applications", *Environ-metrics*, 21, 33-47, (2010).