Journal of the Korean Statistical Society (2005), **34**: 3, pp 173–183

A STUDY ON THE EFFECT OF POWER TRANSFORMATION IN SPATIAL STATISTIC ANALYSIS

JIN-HEE LEE[1] AND KEY-IL SHIN[2]

ABSTRACT

The Box-Cox power transformation is generally used for variance stabilization. Recently, Shin and Kang (2001) showed, under the Box-Cox transformation, invariant properties to the original model under the large mean and relatively small variance assumptions in time series analysis. In this paper we obtain some invariant properties in spatial statistics.

AMS 2000 subject classifications. Primary 86A32; Secondary 91B72.
Keywords. Spatial statistics, Invariant Property, Variogram, Box-Cox power Transformation.

1. INTRODUCTION

The Box-Cox power transformation is the most frequently used in statistical analysis. Generally, most of data, in spatial statistic analysis have the long tail or unsymmetric distribution and then data transformation is usually required. Although many researches have been devoted to studying transformation, deciding whether the transformation will be better off at prediction or not is not quite simple. Even we could not get the effect of the transformation. Recently Griffith et al. (1998) discussed the importance of transformations for spatial data and examined bivariate Box-Cox/Box-Tidwell transformations of the dependent and independent variables in a spatial autoregression. And Shin and Kang (2001) showed that order and first step ahead forecast of the transformed model are approximately invariant to those of the original model under certain assumptions on the mean and variance. In this paper the power transformation proposed by

Received March 2005; accepted August 2005.
[1]Cancer Registration Branch, Research Institute for National Cancer Control and Evaluation, 809, Madu-dong, Ilsan-Gu, Goyang, Kyunggi, 411-764, Korea (e-mail : jhlee@ncc.re.kr)
[2]Department of Statistics, Hankuk University of Foreign Studies, Yongin-Si, Kyunggi-Do, 449-791, Korea (e-mail : keyshin@hufs.ac.kr, Fax: +82-31-330-4566)

Box-Cox (1964) is studied with spatial statistics. The power transformation and related theories have been applied to spatial statistic analysis. In spatial statistic analysis, many important statistical results are obtained under the stationarity assumption. See Cressie (1993) for the stationarity assumption. When a process has non-constant error variance, the Box-Cox power transformation is applied which is given by

$$Y_\lambda(X) = \begin{cases} Z^\lambda(X), & \lambda \neq 0, \\ log(Z(X)), & \lambda = 0 \end{cases} \tag{1.1}$$

where original spatial data, $Z(X)$ are positive. Here X stands for location. We assume that the transformed data $Y_\lambda(X)$ are symmetric and have the stabilized variance and λ is the power transformation parameter.

The transformed data, $Y_\lambda(X)$ are used in data analysis procedure and kriging. After kriging with transformed data, retransformation is used. Shin and Kang (2001) showed that for stationary time series, the transformation process is not sensitive with large mean relative to a small error variance. In this paper, we obtain some invariant properties of the power transformation in spatial statistics. This paper is organized as follows. Invariant properties of correlogram, variogram, confidence interval and kriging for the spatial statistic model subject to the transformation are obtained in section 2. In section 3, examples are illustrated for those invariant properties and finally, some concluding remarks are in section 4.

2. Invariant Properties

Let us consider a spatial stochastic process, $\{Z(X)|X \in D\}$ where X stands for location and $D \in R^d, d \geq 1$. Assume that this process is ergodic and satisfies the hypothesis of intrinsic stationarity and given by

$$Z(X) = \mu_Z + \delta(X)$$

where $\mu_Z = E(Z(X))$ is a unknown constant and $\delta(X)$ is an error obtained from the location X.

With this assumption and use of Taylor series expansion, we obtain invariant properties of correlogram, range of variogram, confidence interval and kriging error.

2.1. *Invariant properties on correlogram*

Let $Y_\lambda(X)$ be the transformed process defined in (1.1) and consider the Taylor series expansion up to the first-order term. Then we have

$$Y_\lambda(X) \approx \begin{cases} \mu_Z^\lambda + \lambda\mu_Z^{\lambda-1}(Z(X)-\mu_Z), & \lambda \neq 0, \\ log(\mu_Z) + \mu_Z^{-1}(Z(X)-\mu_Z), & \lambda = 0 \end{cases} \tag{2.1}$$

where $\mu_Z = E(Z(X))$.

Thus, for $\lambda \neq 0$, we have

$$Y_\lambda(X) - \mu_Y \approx Z(X)^\lambda - \mu_Z^\lambda \approx \lambda\mu_Z^{\lambda-1}(Z(X)-\mu_Z) \tag{2.2}$$

and for $\lambda = 0$,

$$Y_\lambda(X) - \mu_Y \approx log(Z(X)) - log(\mu_Z) \approx \mu_Z^{-1}(Z(X)-\mu_Z). \tag{2.3}$$

Using equation (2.2) and (2.3), we have the approximate variance of $Y_\lambda(X)$ and covariance between $Y_\lambda(X_i)$ and $Y_\lambda(X_j), i = 1,2,3\cdots,n, j = 1,2,3,\cdots,n$ for $\lambda \neq 0$:

$$\begin{aligned} Var(Y_\lambda(X)) &\approx (\lambda\mu_Z^{\lambda-1})^2 Var(Z(X)), \\ Cov(Y_\lambda(X_i), Y_\lambda(X_j)) &\approx (\lambda\mu_{Z(X_i)}^{\lambda-1})(\lambda\mu_{Z(X_j)}^{\lambda-1})Cov(Z_\lambda(X_i), Z_\lambda(X_j)). \end{aligned}$$

Then under the assumption, $\mu_Z(X_i) = E(Z(X_i)) = E(Z(X_j)) = \mu_Z(X_j)$,

$$\begin{aligned} \varrho_{Y_\lambda(X_i),Y_\lambda(X_j)} &= Corr(Y_\lambda(X_i), Y_\lambda(X_j) = \frac{Cov(Y_\lambda(X_i), Y_\lambda(X_j)}{\sqrt{Var(Y_\lambda(X_i))}\sqrt{Var(Y_\lambda(X_j))}} \\ &\approx \frac{Cov(Z(X_i), Z(X_j))}{\sqrt{Var(Z(X_i))}\sqrt{Var(Z(X_j))}} = \varrho_{Z(X_i),Z(X_j)}. \end{aligned}$$

In addition, for $\lambda = 0$, we have the same result using (2.3). Therefore the correlogram is approximately invariant under the power transformation.

2.2. *Invariant properties on variogram*

Variogram estimation is a crucial stage of spatial analysis because it determines the kriging weights. It is important to have a variogram estimator which remains close to the true underlying variogram.

$$Var(Z(X+h)-Z(X)) = 2\gamma(h), \quad \forall X, X+h \in D$$

where $2\gamma(h)$ is the variogram. We consider three variogram estimators suggested by Matheron's (1962) method, Cressie and Hawkins (1980)and Genton's (1998).

Now, let us look at the Box-Cox transformations of variogram. First, the most widely used classical variogram estimator is due to matheron (1962) which is given by

$$2\hat{\gamma}_1(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (Y_\lambda(X_i) - Y_\lambda(X_i+h))^2, \quad h \in R^d \tag{2.4}$$

where $N(h)$ is the number of pairs of observations among the available data separated by lag h.

This estimator is unbiased, but when outliers exist in the data, the results will be heavily affected. For that reason, Cressie and Hawkins (1980) proposed a robust estimator defined by

$$2\hat{\gamma}_2(h) = \frac{1}{0.457 + \frac{0.494}{N(h)}} [\frac{1}{N(h)} \sum_{i=1}^{N(h)} |Y_\lambda(X_i) - Y_\lambda(X_i+h)|^{\frac{1}{2}}]^4, \quad h \in R^d \tag{2.5}$$

where the denominator corrects for bias under Gaussianity.

And similarly, Genton's (1998) estimator of the variogram is given by

$$2\hat{\gamma}_3(h) = [2.219\{|V_i(h) - V_j(h)|; i < j\}_{(k)}]^2, \quad h \in R^d \tag{2.6}$$

where $k = \binom{[\frac{N(h)}{2}]+1}{2}$ and $V_i(h) = Y_\lambda(X_i) - Y_\lambda(X_i+h)$.

As seen in equation (2.4) through (2.5), variogram estimators are obtained only through $\{Y_\lambda(X_i) - Y_\lambda(X_i+h)\}$. Using equation (2.1), for $\lambda \neq 0$,

$$\{Y_\lambda(X_i) - Y_\lambda(X_i+h)\} \approx \lambda\mu_Z^{\lambda-1}\{Z(X_i) - Z(X_i+h)\} \tag{2.7}$$

and for $\lambda = 0$,

$$\{Y_\lambda(X_i) - Y_\lambda(X_i+h)\} \approx \frac{1}{\mu_Z}\{Z(X_i) - Z(X_i+h)\}. \tag{2.8}$$

Plugging (2.6) and (2.7) into $\hat{\gamma}_i(h), i = 1,2,3$, in (2.4) and (2.5) we have the following transformed variogram estimator.

$$\hat{\gamma}_i^Y(h) \approx \begin{cases} (\lambda \mu_Z^{\lambda-1})^2 \hat{\gamma}_i^Z(h), & \lambda \neq 0, \\ \frac{1}{\mu_Z^2} \hat{\gamma}_i^Z(h), & \lambda = 0. \end{cases}$$

If c^2 defined as the following

$$c^2 = \begin{cases} (\lambda \mu_Z^{\lambda-1})^2, & \lambda \neq 0, \\ \frac{1}{\mu_Z^2}, & \lambda = 0 \end{cases}$$

then

$$\hat{\gamma}_i^Y(h) \approx c^2 \cdot \hat{\gamma}_i^Z(h). \tag{2.9}$$

Note that c^2 is only a function of λ and μ_Z which are constant. Therefore variogram estimators of transformed data are proportional to those of the original data.

After obtaining variogram estimates, we try to fit theoretical variogram. In this section, we consider three theoretical variogram : Spherical variogram, Exponential variogram and Gaussian variogram

Spherical model :

$$\gamma_s(t, \theta) \approx \begin{cases} 0, & t = 0, \\ (\theta_0 + \theta_1 \{\frac{3}{2}\frac{t}{R} - \frac{1}{2}(\frac{t}{R})^3\}, & 0 < t \leq R, \\ \theta_0 + \theta_1, & t \geq R. \end{cases}$$

Exponential model :

$$\gamma_e(t, \theta) \approx \begin{cases} 0, & t = 0, \\ (\theta_0 + \theta_1(1 - \exp(-\frac{t}{R})), & t > 0. \end{cases}$$

Gaussian model :

$$\gamma_g(t, \theta) \approx \begin{cases} 0, & t = 0, \\ (\theta_0 + \theta_1(1 - \exp(-\frac{t^2}{R^2})), & t > 0 \end{cases}$$

where θ_0 is nugget, θ_1 is sill and R is range. It is known that actual ranges depend on the models. But these are proportional to R in models, we just use the estimate of R.

From equation (2.9), we can easily obtain the results of

$$\hat{\theta}_{0Y} \approx c^2 \cdot \hat{\theta}_{0Z}, \hat{\theta}_{1Y} \approx c^2 \cdot \hat{\theta}_{1Z}$$

and

$$\hat{R}_Y \approx \hat{R}_Z.$$

So the power transformation does not affect to the estimate of range.

2.3. Invariant properties of kriging and confidence interval

Now let us consider the ordinary kriging. Due to the unbiasness for the ordinary kriging, we have the following restriction:

$$\sum_{i=1}^{n} \pi_i = 1$$

With the above restriction, the coefficients of optimal ordinary kriging are obtained by

$$\pi = \Gamma^{-1}(\gamma + \frac{(1 - \tilde{1}'\Gamma^{-1}\gamma)}{\tilde{1}'\Gamma^{-1}\tilde{1}}\tilde{1}) \tag{2.10}$$

where matrix $\Gamma_{n\times n} = \gamma(X_i - X_j), i = 1, 2, \cdots, n, j = 1, 2, \cdots, n$ and $\gamma_{n\times 1} = \gamma(X_0 - X_i), i = 1, 2, \cdots, n$ and $\tilde{1}' = (1, 1, \cdots, 1)$. See Cressie (1993) for more details.

Denote π_Y, π_Z as kriging coefficients of the transformed data and the original data respectively. Then using (2.9) we have

$$\hat{\pi}_Y = \hat{\Gamma}_Y^{-1}(\hat{\gamma}_Y + \frac{(1 - \tilde{1}'\hat{\Gamma}_Y^{-1}\hat{\gamma}_Y)}{\tilde{1}'\hat{\Gamma}_Y^{-1}\tilde{1}}\tilde{1}) \approx \hat{\Gamma}_Z^{-1}(\hat{\gamma}_Z + \frac{(1 - \tilde{1}'\hat{\Gamma}_Z^{-1}\hat{\gamma}_Z)}{\tilde{1}'\hat{\Gamma}_Z^{-1}\tilde{1}}\tilde{1}) = \hat{\pi}_Z. \tag{2.11}$$

Therefore we have an invariant property of kriging coefficients.

Next consider the optimal kriging predictor. The optimal kriging predictor using the original data is obtained by $\hat{Z}_0 = \sum_{i=1}^{n} \hat{\pi}_{Z_i} Z(X_i)$.

Let $\hat{Z}_{Y0}^{\lambda}$ be the optimal kriging predictor of the transformed data. Then similarly we can calculate $\hat{Z}_{Y0}^{\lambda} = \sum_{i=1}^{n} \hat{\pi}_{Y_i} Y_\lambda(X_i)$. Also denote $\hat{Z}_{Y0}$ as final kriging predictor after retransformation. Then we have, $\lambda \neq 0$,

$$\hat{Z}_{Y0} = \{\sum_{i=1}^{n} \hat{\pi}_{Y_i} Y_\lambda(X_i)\}^{\frac{1}{\lambda}} \approx \{\hat{\mu}_Z^{\lambda} + \lambda\hat{\mu}_Z^{\lambda-1} \sum_{i=1}^{n} \hat{\pi}_{Yi}(Z(X) - \hat{\mu}_Z)\}^{\frac{1}{\lambda}}. \tag{2.12}$$

Using (2.11) and the result in Shin and Kang (2001), equation (2.12) becomes

$$\hat{Z}_{Y0} \approx \{\hat{\mu}_Z + \sum_{i=1}^{n} \hat{\pi}_{Yi}(Z(X_i) - \hat{\mu}_Z)\} = \sum_{i=1}^{n} \hat{\pi}_{Z_i} Z(X_i). \tag{2.13}$$

Also, one can easily obtain the same result for $\lambda = 0$. Therefore we obtain an invariant property of the optimal kriging predictor.

Next consider kriging error and confidence interval. Kriging error can be obtained (Cressie, 1993) by

$$\sigma_0^2 = \gamma'\Gamma^{-1}\gamma - \frac{(\tilde{1}\Gamma^{-1}\gamma - 1)^2}{(\tilde{1}'\Gamma^{-1}\tilde{1})}.$$

Then the kriging error for transformed data is as following:

$$\hat{\sigma}_{Y_0}^2 = \hat{\gamma}_Y'\hat{\Gamma}_Y^{-1}\hat{\gamma}_Y - \frac{(\tilde{1}\hat{\Gamma}_Y^{-1}\hat{\gamma}_Y - 1)^2}{(\tilde{1}'\hat{\Gamma}_Y^{-1}\tilde{1})}.$$

Then again using (2.9) we have

$$\hat{\sigma}_{Y_0}^2 \approx c^2 \cdot (\hat{\gamma}_Z'\hat{\Gamma}_Z^{-1}\hat{\gamma}_Z - \frac{(\tilde{1}\hat{\Gamma}_Z^{-1}\hat{\gamma}_Z - 1)^2}{(\tilde{1}'\hat{\Gamma}_Z^{-1}\tilde{1})}) = c^2 \cdot \hat{\sigma}_{Z_0}^2 \tag{2.14}$$

where ${\sigma_{Z_0}}^2$ is the kriging error for original data. Therefore $\hat{\sigma}_{Y0}^2 \approx c^2 \cdot \hat{\sigma}_{Z0}^2$. That is, kriging error obtained by using the transformed data is a constant times kriging error obtained by using the original data. We use this result for obtaining an invariant property of confidence interval. Consider a 95% confidence interval using the original data:

$$(\hat{Z}_0 - 1.96\sigma_{Z0},\ \hat{Z}_0 + 1.96\sigma_{Z0}).$$

The confidence interval for the transformed data is also defined by, for $\lambda \neq 0$, from the above equation,

$$(\hat{Z}_{Y0}^{\lambda} - 1.96\sigma_{Y0},\ \hat{Z}_{Y0}^{\lambda} + 1.96\sigma_{Y0}). \tag{2.15}$$

Then using (2.12) and (2.14) we have

$$\{\hat{\mu}_Z^{\lambda} + \lambda\hat{\mu}_Z^{\lambda-1} \sum_{i=1}^{n} \hat{\pi}_{Yi}(Z(X) - \hat{\mu}_Z) \pm 1.96\lambda\hat{\mu}_Z^{\lambda-1}\hat{\sigma}_{Z0}\}^{\frac{1}{\lambda}}.$$

Now $\sum_{i=1}^{n} \pi_{Yi} = 1$, we have

$$
\begin{aligned}
&\{\hat{\mu}_Z^{\lambda} + \lambda\hat{\mu}_Z^{\lambda-1}\sum_{i=1}^{n}\hat{\pi}_{Yi}(Z(X) - \mu_Z \pm 1.96\hat{\sigma}_{Z0})\}^{\frac{1}{\lambda}} \\
&= \{\hat{\mu}_Z + \sum_{i=1}^{n}\pi_{Yi}(Z(X) - \hat{\mu}_Z \pm 1.96\hat{\sigma}_{Z0})\} \\
&= \{\sum_{i=1}^{n}\hat{\pi}_{Yi}(Z(X) \pm 1.96\hat{\sigma}_{Z0})\} \\
&= \sum_{i=1}^{n}\hat{\pi}_{Zi}(Z(X) \pm 1.96\hat{\sigma}_{Z0})\} \\
&= \hat{Z}_0 \pm 1.96\hat{\sigma}_{Z0}. \qquad (2.16)
\end{aligned}
$$

So two confidence intervals are approximately the same. Also for $\lambda = 0$, one can get the results.

3. Analysis of Real Data

From the results of section 2, we calculate the estimates of variogram with Jura data (Goovaerts, 1997) using the following models such as spherical model, exponential models and gaussian model, and obtain the kriging values. For calculations, we use s-plus package.

3.1. Parameter estimate(range, sill and nugget)

In order to see, with large means and relatively small variance, the approximate invariant properties, range, sill and nugget of variogram are calculated with original data and the data which are added 20 and 50 each to original data respectively. Then we consider the 5 ways of transformations with $\lambda = 1, 0.5, 0, -0.5$ and -1. At this point, adding the constants(20 and 50) to original data makes the mean of data larger than original one and that can give the clue of the invariant properties on variogram.

Table 1, 2 and 3 show that the changes of range, sill and nugget with 3 different data set to 3 different models. For the values of ranges, it can be easily confirmed the invariant properties with 5 ways of transformations. Also we tabulate $\frac{1}{c^2}\hat{\theta}_{0Y}$ and $\frac{1}{c^2}\hat{\theta}_{1Y}$ to check the theoretical results on nugget and sill.

From the result of Table 1 to 3, we see that the values of ranges converge to the range of original data when the mean of the data are large. Also we confirm

TABLE 3.1 *The ranges of variogram*

λ	spherical model			exponential model			gaussian model		
	data	data +20	data +50	data	data +20	data +50	data	data +20	data +50
1.0	2.261	2.261	2.262	1.677	1.678	1.678	1.301	1.301	1.301
0.5	2.084	2.255	2.259	1.068	1.624	1.633	1.162	1.292	1.297
0	1.978	2.238	2.256	0.773	1.574	1.655	1.050	1.283	1.293
−0.5	1.842	2.219	2.252	0.590	1.528	1.612	0.950	1.274	1.290
−1	1.677	2.201	2.245	0.445	1.484	1.591	0.840	1.266	1.286

TABLE 3.2 *The sill of variogram*

λ	spherical model			exponential model			gaussian model		
	data	data +20	data +50	data	data +20	data +50	data	data +20	data +50
1.0	0.080	0.080	0.080	0.107	0.107	0.107	0.068	0.068	0.068
0.5	0.093	0.080	0.080	0.113	0.105	0.107	0.076	0.067	0.068
0	0.120	0.077	0.080	0.151	0.104	0.106	0.097	0.063	0.066
−0.5	0.171	0.080	0.080	0.237	0.104	0.106	0.138	0.067	0.066
−1	0.273	0.080	0.080	0.446	0.103	0.105	0.223	0.067	0.067

TABLE 3.3 *The nugget of variogram*

λ	spherical model			exponential model			gaussian model		
	data	data +20	data +50	data	data +20	data +50	data	data +20	data +50
1.0	0.373	0.373	0.373	0.375	0.375	0.375	0.390	0.390	0.390
0.5	0.393	0.371	0.373	0.388	0.372	0.375	0.414	0.388	0.390
0	0.445	0.370	0.372	0.423	0.372	0.373	0.470	0.388	0.389
-0.5	0.537	0.369	0.371	0.477	0.370	0.373	0.572	0.386	0.388
-1	0.691	0.367	0.371	0.520	0.368	0.372	0.743	0.385	0.388

TABLE 4.1 *The MSE on kriging*

λ	spherical model			exponential model			gaussian model		
	data	data +20	data +50	data	data +20	data +50	data	data +20	data +50
1.0	0.385	0.385	0.385	0.384	0.384	0.384	0.418	0.418	0.418
0.5	0.388	0.385	0.385	0.374	0.383	0.383	0.421	0.418	0.416
0	0.407	0.385	0.385	0.374	0.382	0.383	0.439	0.417	0.418
-0.5	0.540	0.385	0.385	0.375	0.382	0.383	0.467	0.418	0.418
-1	0.469	0.385	0.385	0.323	0.381	0.383	0.499	0.418	0.418

the results on sill and nugget.

3.2. Kriging and Kriging error

Checking invariant properties on kriging prediction, we eliminate one of the data points temperately from the data sets and then predict its value by kriging using the remaining data points. For each λ, we retransform to get proper predict values. Following Table 4 shows $MSE = \frac{1}{n}\sum_{i=1}^{n}(Z_i - Z_{(i)})^2$ for each λ. Again one can see the invariant properties.

4. CONCLUDING REMARKS

The Box-Cox transformation is generally used for variance stabilization and it makes the estimation and testing more reliable. However, sometimes this transformation causes the increasing bias (Granger and Newbold, 1976) and makes the model more complicated, so definitely interpretation of the results is difficult. This study shows the effect of the Box-Cox transformation to spatial statistic analysis. If data have the large mean and relatively small variance, specially in case of using Box-Cox transformation with first order Taylor expansion, some invariant properties can be obtained. The results obtained in this paper should be taken into account in use of the power transformation.

REFERENCES

BOX, G. E. P., COX, D. R. (1964). "An analysis of transformations", *Journal of the Royal Statistical Society Series B*, **26**, 211–252.

CRESSIE, N. A. C. (1993). "Statistics for spatial data", John Wiley and Sons, Inc.

CRESSIE, N., HAWKINS, D. M. (1980). "Robust estimation of the variogram", *Mathematical Geology*, **12**, No. 2, 115-125.

GENTON, M. C. (1998). "Highly robust variogram estimation", *Mathematical Geology,* **30**, No. 2, 213-221.

GOOVAERTS P. (1997). "Geostatistics for natural resources evaluation", *Oxford University Press,* New York.

GRANGER, C. W. J., NEWBOLD, P. (1976). "Forecasting Transformed Series", *Journal of the Royal Statistical Society, Series B(Methodological),* **38**, Issue 2, 189-203.

GRIFFITH, D., PAELINCK, J., GASTEL, R. V. (1998). "The Box-Cox Transformation: Computational and interpretation features of the parameters", *Econometric Advances in Spatial Modelling and Methodology,* 46-56.

KALUZNY, S. P., VEGA, S. C., CARDOSO, T. P., SHELLY, A. A. (1998). "S+SpatialStats User's Manual for Windows and UNIX", Springer.

MATHERON, G. (1962). "Traite de Geostatistique appliquee", *Tome I. Memoires du Bureau de Recherches Geologiques et Minieres,* No. 14, Editions Technip, Paris.

SHIN, K-I., KANG, H-J. (2001). "A study on the effect of power transformation in the ARMA(p,q) model", *Journal of Applied Statistics,* **28**, No. 8, 1019-1028.