Prediction of returns in WEEE reverse logistics based on spatial correlation

Jun LV (JLV@dbm.ecnu.edu.cn)

School of Business, East China Normal University
800, Dongchuan Road, Shanghai 200241, China.

Jia-ping XIE

School of International Business, Shanghai University of Finance & Economics,
777, Guoding Road, Shanghai 200433, China

Abstract
Under the stress of environment legislation and resource lack, logistics management turns from linear mode into closed mode. It is difficult for academic circle and industry to predict the returns of recycling WEEE products due to the uncertainty of the number of returns. From a new perspective, we consider spatial correlation of returned WEEE, introduce Kriging method of spatial statistics and build Kriging spatial model of reverse logistics return prediction. Four main properties about the return volume of recycle centers are found. Simulation experiments based on Monte Carlo are conducted to validate the effectiveness of the developed model.

Keywords: Reverse logistics, Kriging method, Spatial correlation

Introduction
Human is facing some problems about sustainability of the environment such as the increasing waste and lack of resources, etc. However, it is difficult for academic circle and industry to predict the returns of recycling WEEE products due to the uncertainty of the number of returns.
In the field of return product prediction, some articles have emerged that provided prediction modeling methods. Kelle and Silver (1989) assumed that the return product quantity is the function of past sale and return, and built the return prediction model based on sale and return history data. Yang and Williams (2009) built a logic regression model to predict the return tendency of computers. Polynomial regression analysis does well in fitting the polynomial tendency of data series, but has higher modeling error, because the highest power of polynomial is randomly chosen by researcher.

Gómez et al. (2002) developed a prediction model based on back propagation neural network to predict the return quantity of WEEE. However, the network structure and network node number are decided not by mathematics analysis, but only by user’s experience.

Toktay et al. (2000) assumed the return quantity follows binomial probability distribution, and used Bayes inferences and expectation maximization algorithm to find the total probability and delay probability parameter of reverse logistics, and then find the return quantity and time.

The above articles consider the return quantity of network nodes as time series data, and the time correlation is the key consideration, and under given conditions these methods have good results to predict return quantity. However, in the WEEE reverse logistics network, the return quantity not only has the time correlation, but also has the spatial correlation. Min et al. (2006) built nonlinear model closed-loop supply chain network based on both space and time, and researched on solution based on genetic algorithm. Graczyk and Witkowski (2011) pointed out that only if both space and time are considered, better result can be realized in reverse logistics. Three Japanese scholars from national graduate institute for policy studies of Japan found that from the space perspective, the solid waste has stronger interdependency (Daisuke et al., 2012). They used the 2006-2010 data from ministry of the environment of Japan, classified solid waste into two types: household solid waste and business solid wastes. And the spatial correlation is analyzed in 47 areas by the method of data envelopment analysis, and the result showed that the solid waste quantities in 47 areas are highly spatial correlation.

In what follows, we develop and analyze Kriging models to provide insights to inform policy managers and WEEE companies regarding spatial correlation between the two return centers. In §2, we descript the problems and second order randomness. In §3, we provide an analysis of spatial structure and variogram function of WEEE return and develop Kriging spatial model. In §4, the simulation experiment and results analysis are conducted. We end with a summary and conclusion in §5.

**Problem Descriptions and Second Order Randomness Hypothesis**

**Problem Description**

We assume that one electronic company has \( q \) return centers located within a two-dimension finite region \( D \). The return centers are \( s_1(x,y), s_2(x,y), \ldots, s_q(x,y) \). \( x \) and \( y \) represent the return center geographical location of the abscissa and ordinate in this region, the distance
between any two points is \( h \), and the vector forms of all points are \( \mathbf{S}(s_1, s_2, \ldots, s_q) \), and all return centers in \( \mathbf{D} \) are \( s_q(x, y) \subseteq \mathbf{D} \). The return quantity of each center is denoted by \( Z(s_1), Z(s_2), \ldots, Z(s_q) \). The vector form of the return quantity of all centers in this region is \( \mathbf{Z}(Z(s_1), Z(s_2), \ldots, Z(s_q)) \), which is considered as a stochastic variable.

It is assumed that in one period and under random condition, the return quantities of \( n \) centers points are \( Z(s_1), Z(s_2), \ldots, Z(s_n) \), two questions exist:

How to quantitatively describe the spatial structure correlation between the known return quantity of any \( n \) centers and the rest unknown return quantity of \( q-n \) centers? Based on the quantitative description of their spatial structure correlation, how to predict the unknown return quantity using known return quantity and to predict return quantity in the whole region?

Trying to answer the above two questions, we need to analyze the spatial structure of the return quantity and build spatial correlation stochastic model under the condition of spatial correlation and uncertainty.

**Second Order Randomness**

Kriging method is one of the most important spatial prediction methods of stochastic variable in spatial statistics. This method is first used by South African mining engineer D.R. Krige to search gold in 1951, and is named by him. And then, the Kriging method was systematically proposed by French mathematician Matheron, and became the important part of spatial statistics.

Randomness indicates that variables are arbitrary value in one period, and second order randomness indicates that variables keep on being arbitrary, and the covariance \( \text{cov}[s_i(x, y), s_j(x, y)] \) at any various locations depends only on distance between two return locations \( h = \|s_i - s_j\| \). Second order randomness is expressed as:

\[
E[Z(s)] = m, (s \in \mathbf{D})
\]

\[
\text{Cov}[Z(s + h), Z(s)] = C(h), (s, s + h \in \mathbf{D})
\]

where \( E(\cdot) \) is the mathematic expectation and \( \text{Cov}(\cdot) \) is the covariance of the increment between two return centers, \( m \) is a constant, denotes the total return quantity in one period and in one area, \( h \) is the distance between any two return centers, \( Z(s + h) \) denotes the return quantity of one center with a given distance \( h \) from the other center \( s \).
Spatial Prediction Mathematics Model of WEEE Return Quantity

Spatial Structure Analysis of WEEE Return Quantity
The variogram function of spatial statistics is used to analyze spatial variation structure of return quantity variable. For any distance $h$, the variable of $[Z(s + h) - Z(s)]$ has a definite variance, which depends on distance $h$ and for all $s$ and $h$ satisfies (Matheron, 1973):

$$\text{Var}[Z(s + h) - Z(s)] = E\{[[Z(s + h) - Z(s)] - E[Z(s + h) - Z(s)]]^2\} = 2\gamma(h)$$

(3)

where $\gamma(h)$ is called variogram in spatial statistics, denotes the degree of spatial correlation in a spatial stationary process at a given distance $h$, which will be calculated in 3.3.

**Proposition 1**: In the whole region $D$, spatial variogram function, covariance of return quantity between various centers and variance of return quantity of one center can be written as:

$$\gamma(h) = \sigma^2 - C(h)$$

(4)

$$\sigma^2 = \text{Var}[Z(s)] = E[[Z(s) - m]^2]$$

$$C(h) = \text{Cov}[Z(s + h), Z(s)] = E[Z(s + h) - m][Z(s) - m] = E[Z(s + h)Z(s)] - m^2$$

**Proposition 2**: In the whole region $D$, spatial variogram function, variance of return quantity of one center and spatial correlation coefficient can be written as:

$$\gamma(h) = \sigma^2[1 - \rho(h)]$$

(5)

where $\rho(h)$ is spatial correlation coefficient and is defined as: $\rho(h) = C(h) / \sigma^2$

WEEE Return Quantity Variogram Function
Return quantity variogram function $\gamma(h)$ explores spatial variation degree and correlation structure of the return quantity variables of various spatial positions in region $D$ for one WEEE company, and can also describes the variables’ randomness and structure.

Figure 1 shows WEEE return quantity variogram function. The three main parameters in figure 1 are explained as follows:
Figure 1 WEEE return quantity variogram function

Sill \( (C_o + C_i) \): in figure 1, as distance \( h \) gets large, the correlation between the two centers which is separated with the distance \( h \) becomes negligible and the value of variogram tends to be stationary. The value of sill is the maximum height of the variogram curve. Nugget \( (C_o) \): the variance at zero distance due to measurement or sampling error. Correlation range \( (a) \): the distance at which the variance achieves the plateau. When \( h < a \), pairs of centers have spatial correlation; when \( h \geq a \), the spatial correlation is negligible.

In spatial statistics, popular theoretical fitting variogram functions include spherical, exponential, Gauss functions, which are empirically proved that can effectively analyze spatial variation structure:

1) Spherical variogram function

\[
\gamma(h) = C_o + C_i[1.5(h/a) - 0.5(h/a)^3], \quad 0 \leq h \leq a
\]  \( \text{(6)} \)

\[
\gamma(h) = C_o + C_i, \quad h > a
\]  \( \text{(7)} \)

The covariance function of spherical variogram function is obtained:

\[
C(0) = \sigma^2 = C_o + C_i, \quad h = 0
\]

\[
C(h) = C_i[1 - 1.5(h/a) + 0.5(h/a)^3], \quad 0 < h \leq a
\]  \( \text{(8)} \)

\[
C(h) = 0, \quad h > a
\]

2) Exponential variogram function

\[
\gamma(h) = C_o + C_i(1 - e^{-3h/a})
\]  \( \text{(9)} \)

The covariance function of exponential variogram function is obtained:
$C(0) = \sigma^2 = C_0 + C_1; C(h) = C_1e^{-3(h/a)}$ \hspace{1cm} (10)

3) Gauss variogram function

$\gamma(h) = C_0 + C_1(1 - e^{-3(h/a)^2})$ \hspace{1cm} (11)

The covariance function of Gauss variogram function is obtained:

$C(0) = \sigma^2 = C_0 + C_1; C(h) = C_1e^{-3(h/a)^2}$ \hspace{1cm} (12)

**Proposition 1:** when the value of WEEE return quantity variogram function $\gamma(h)$ is increasing with the distance $h$ till a stationary constant ($\text{Sill}, C_0 + C_1$), the constant reflects the biggest variation range or total variation degree of WEEE return system.

**Proposition 2:** when distance $h$ is zero, $\gamma(0)$ equals $C_0$ (Nugget, $C_0$), this value reflects the size of randomness of return quantity variable in one center.

**Proposition 3:** Let total variation degree ($C_0 + C_1$) be divided by variation range ($C_1$) between various centers, and then we can obtain structure variance ratio $\frac{C_1}{C_0 + C_1}$. The ratio characterizes the degree of spatial heterogeneity, which means the spatial variability caused by the autocorrelation occupies the proportion of the total variation of the system. It reflects the spatial correlation. The larger the ratio is, the stronger the spatial correlation is. It is generally believed that in WEEE reverse logistics when the ratio value is less than 25%, the variables are with weaker spatial correlation; and when 25% to 75%, the variables have a moderate spatial correlation; and when greater than 75%, the variables have a stronger spatial correlation.

**Proposition 4:** In the whole region $D$, there exists a distance $a$, when pairs of centers less than this distance have spatial correlation; otherwise, pairs of centers are negligibly correlated.

**Kriging Spatial Model of the WEEE Returns**

Based on the second order randomness and spatial structure of WEEE return quantity, we can build a WEEE Kriging spatial prediction model.

The return quantity $\hat{Z}(s_0)$ at the unknown center location $s_0$ can be determined by a linear combination of $n$ known return quantity $Z(s_i)$, see equation (13):

$$\hat{Z}(s_0) = m + \sum_{i=1}^{n} \lambda_i [Z(s_i) - m]$$ \hspace{1cm} (13)

where $\hat{Z}(s_0)$ is the return quantity in unknown center $s_0$ to be predicted, $Z(s_i)$ is the known
return quantity in center \( s_i \), \( \lambda_i \) are weights characterized by the impacts of known centers on the value of the unknown center \( s_0 \), and \( n \) are the number of known centers, and the weights \( \lambda_i \) are as well determined variables, obtained from the condition that the estimation is unbiased and variance error is minimized.

1) Unbiasedness of prediction

\[
E[\hat{Z}(s_0) - Z(s_0)] = 0
\]

(14)

\[
E[\hat{Z}(s_0) - Z(s_0)] = m + \sum_{i=1}^{n} \lambda_i [E[Z(s_i)] - m] - E[Z(s_0)] = m + \sum_{i=1}^{n} \lambda_i (m - m) - m = 0
\]

(15)

Whatever the weights are, the return prediction of Kriging method is unbiased, and it is also demonstrated that Kriging method is effective.

2) Minimum variance of prediction

\[
Var[\hat{Z}(s_0) - Z(s_0)] = \sigma^2 - 2 \sum_{i=1}^{n} \lambda_i C(s_0 - s_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(s_i - s_j) = \Sigma
\]

(16)

where \( \Sigma \) denotes the variance of error between prediction value and real value.

To minimize the Kriging error variance, the weight coefficients are calculated by:

\[
\begin{bmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{bmatrix} = \begin{bmatrix}
C_{11} & \cdots & C_{1n} \\
\vdots & \ddots & \vdots \\
C_{n1} & \cdots & C_{nn}
\end{bmatrix}^{-1} \begin{bmatrix}
C_{01} \\
\vdots \\
C_{0n}
\end{bmatrix}
\]

(17)

\[ C_{ij} = Cov(s_i - s_j) \] are covariance between \( s_i \) and \( s_j \), \( C_{ii} \) are variance value.

**Proof:**

When the minimum variance of Kriging method is satisfied, the weighted Kriging linear equation is written as:

\[
\Sigma = \sigma^2 - 2 \lambda_1 C(s_0 - s_1) + \lambda_2 C(s_0 - s_2) + \ldots + \lambda_n C(s_0 - s_n) \\
+ \lambda_1 [\lambda_1 C(s_1 - s_1) + \lambda_2 C(s_1 - s_2) + \ldots + \lambda_n C(s_1 - s_n)] \\
+ \lambda_2 [\lambda_1 C(s_2 - s_1) + \lambda_2 C(s_2 - s_2) + \ldots + \lambda_n C(s_2 - s_n)] \\
+ \ldots \\
+ \lambda_n [\lambda_1 C(s_n - s_1) + \lambda_2 C(s_n - s_2) + \ldots + \lambda_n C(s_n - s_n)]
\]

(18)

Linear equation system is obtained as:

\[
\sum_{i=1}^{n} \lambda_i C(s_i - s_j) = C(s_0 - s_j), \quad j = 1, \ldots, n
\]

(19)
Kriging weights can be calculated by equation (20), and then to estimate the return quantity of unknown centers. And error variance can be calculated as well:

\[
\sigma_{\theta_k}^2 = Var(\hat{Z}(s_0) - Z(s_0)) = \sigma^2 - \sum_{i=1}^{n} \lambda_i C(s_0 - s_i)
\]  

(21)

According to equations (13) to (21), we can predict the return quantity of any unknown return center in region \(D\), and finally predict the total return quantity in the whole region.

Simulation and results analysis

Monte Carlo Simulation Data Generation
We use Monte Carlo method to generate four samples with spatial correlation (see table 1). The numbers of return centers are respectively 100, 50, 20 and 10, basically covering the possible number range of return centers in practice. For sample 2, 3 and 4, the simulation is repeated for 50 times. For sample 1, since there are 100 centers and too much computation, the simulation is repeated for 25 times. So the simulation is totally repeated for 125 times, and detailed statistics description is shown in table 1. 80% of each sample is as known return center, that is, for four samples, the number of known return centers are respectively 80, 40, 16 and 8. The rest 20% of each sample, as unknown centers, is used to be tested.

<table>
<thead>
<tr>
<th>Sample Order Number</th>
<th>Number of Centers</th>
<th>Simulation Repetition times</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
<th>Median</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>25</td>
<td>0</td>
<td>9543</td>
<td>5421</td>
<td>0.0026</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>8945</td>
<td>4580</td>
<td>0.0027</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>50</td>
<td>0</td>
<td>10000</td>
<td>5110</td>
<td>0.0025</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>50</td>
<td>0</td>
<td>7849</td>
<td>3459</td>
<td>0.0026</td>
</tr>
</tbody>
</table>

Spatial Structure Analysis and Variogram Function Calculation
In spatial statistics, popular theoretical fitting variogram functions include spherical, exponential and Gauss functions. According to section 3.2 and 3.3, three variogram functions are finally fitted after multiple repeated simulations.

Simulation result 1: Fitting result shows that for three variogram function, all the correlation length \(h\) is 90 kilometer, which means as \(h\) becomes larger than 90 kilometer, the spatial correlation disappears.

The comparative analyze results of three variogram functions are shown in table 3. The ratio
of the structural variance is the ratio of nugget to sill, \( c_i / (C_0 + c_i) \), reflects spatial variation degree. The larger the ratio is, the stronger the spatial correlation is (see section 3.2).

Simulation result 2: Prediction result using exponential function is best and the error is smallest. From table 2, we can find that the structural variance ratio in exponential mode, compared with spherical model and Gauss model, has the largest value, which means that the spatial correlation between samples fitted by exponential model is strongest.

### Table 2 Quantity analyze results of three variogram function

<table>
<thead>
<tr>
<th>Number of Centers</th>
<th>Variogram function</th>
<th>Sill ((10^5)) C0+C1</th>
<th>Nugget ((10^5)) C0</th>
<th>Ratio of structural variance (%)</th>
<th>prediction error percentage</th>
<th>standard deviation of prediction error</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Spherical function</td>
<td>0.80778</td>
<td>0.14828</td>
<td>81.64%</td>
<td>1.08%</td>
<td>0.015657</td>
</tr>
<tr>
<td></td>
<td>Exponential function</td>
<td>0.80664</td>
<td>0.10263</td>
<td>87.28%</td>
<td>0.87%</td>
<td>0.016062</td>
</tr>
<tr>
<td></td>
<td>Gauss function</td>
<td>0.81273</td>
<td>0.24242</td>
<td>70.17%</td>
<td>1.71%</td>
<td>0.012183</td>
</tr>
<tr>
<td>50</td>
<td>Spherical function</td>
<td>0.82451</td>
<td>0.14717</td>
<td>82.15%</td>
<td>1.04%</td>
<td>0.014249</td>
</tr>
<tr>
<td></td>
<td>Exponential function</td>
<td>0.84158</td>
<td>0.08912</td>
<td>89.41%</td>
<td>0.84%</td>
<td>0.012343</td>
</tr>
<tr>
<td></td>
<td>Gauss function</td>
<td>0.82170</td>
<td>0.18225</td>
<td>77.82%</td>
<td>1.76%</td>
<td>0.014894</td>
</tr>
<tr>
<td>20</td>
<td>Spherical function</td>
<td>0.90012</td>
<td>0.14104</td>
<td>84.33%</td>
<td>0.84%</td>
<td>0.014158</td>
</tr>
<tr>
<td></td>
<td>Exponential function</td>
<td>0.92157</td>
<td>0.09086</td>
<td>90.14%</td>
<td>0.48%</td>
<td>0.018421</td>
</tr>
<tr>
<td></td>
<td>Gauss function</td>
<td>0.92108</td>
<td>0.16422</td>
<td>82.17%</td>
<td>1.42%</td>
<td>0.019510</td>
</tr>
<tr>
<td>10</td>
<td>Spherical function</td>
<td>0.94212</td>
<td>0.11194</td>
<td>88.12%</td>
<td>1.07%</td>
<td>0.016518</td>
</tr>
<tr>
<td></td>
<td>Exponential function</td>
<td>0.94852</td>
<td>0.55578</td>
<td>94.15%</td>
<td>0.87%</td>
<td>0.014328</td>
</tr>
<tr>
<td></td>
<td>Gauss function</td>
<td>0.92349</td>
<td>0.21733</td>
<td>76.47%</td>
<td>2.49%</td>
<td>0.015571</td>
</tr>
</tbody>
</table>

WEEE Return Prediction Result Using Kriging Model

We use two indexes of “prediction error percentage” and “standard deviation of prediction error” to evaluate the accuracy of Kriging method. The prediction error percentage is that the prediction value is subtracted from actual value and then divided by actual value, and standard deviation of prediction error is the standard deviation of Kriging prediction error of multiple smaples. Table 2 shows the WEEE return prediction results of kriging model.

Simulation result 3: The total prediction error percentage is 1.21%, standard error is 0.001532. The result shows that the Kriging model can accurately predicate WEEE return quantity, and the method of Kriging is feasible.

Simulation result 4: As for variogram function, the error percentage of exponential function is 0.77%, lower than spherical function (1.01%) and Gauss function (1.85%). Therefore, the result of exponential function is best and the error is smallest in section 4.2 is validated.

Simulation result 5: From perspective of the number of centers affect on accuracy of prediction model, the prediction error percentages on the centers with number of 100,50,20 and 10 are 1.22%, 1.21%, 0.91% and 1.48% respectively, big prediction error does not exist. It is demonstrated that return Kriging model can be used for the WEEE return system with different numbers of centers.
Conclusions

Under the stress of environment legislation and resource lack, logistics management turns from linear mode into closed mode. It is difficult for academic circle and industry to predict the returns of recycling WEEE products due to the uncertainty of the number of returns. From a new perspective, we consider spatial correlation of returned WEEE, introduce Kriging method of spatial statistics and build Kriging spatial model of reverse logistics return prediction. The conclusions can be reached that:

1) Based on Monte Carlo simulation, the WEEE return spatial model presented in this paper can predict WEEE return accurately, and the total prediction error percentage is 1.21%.
2) Three variogram functions can be used to analyze the spatial structure of WEEE return. The prediction error percentage of exponential function, spherical function and Gauss function is 0.77%, 1.01% and 1.85% respectively, and exponential function gets a better prediction result. Therefore, when WEEE return prediction based on spatial correlation is used, exponential function is recommended.
3) WEEE return system often has different number of centers. In simulation study, the number of centers ranges from 10 to 100, which covers range of the number of centers in actual condition. The prediction error for Kriging model is comparatively stable.
4) In Monte Carlo simulation, WEEE return system exists a distance $a$ at which the variance achieves the plateau, and when $h < a$, any pairs of centers have spatial correlation, when $h \geq a$, the spatial correlation is negligible.

Reference: