APPLICATION TREND SURFACE MODELS WITH ESTIMATION

Shaymaa Riyadh Thanoon

College of Nursing, University of Mosul, Mosul, Iraq

ARTICLE INFO.

Article history:
- Received: 12 / 3 / 2018
- Accepted: 19 / 4 / 2018
- Available online: / / 2018

Keywords: parameters, residuals

Corresponding Author:
Name: Shaymaa Riyadh
E-mail: Shaymaaaluaim@yahoo.com
Tel:

Abstract

This research deal with estimation of trend surface analysis and with spatial data with three models to real spatial data represents a rising ground water. The first method in assessment is to estimation trend surface model parameters by (ml), the second method requires decision on the maximum time difference to be calculated (s), while, the third method need a resolution, and the residuals r of dots to take for the conjecture the f(Dij) parameters. The first and second methods require resolution principle of "neighbor" from determines of "W". These three approaches are applied to real data which represent the ground water levels in 47 wells in mountain region in Sin jar district in Nineveh governorate.

To a multidimensional research due to the finder [2], and circulated earth science by [3]. The basic model is often installed by the normal minimum squares (ols) assuming that residues separate, assumption residues are spatially autocorrelated, the estimates of the lower squares ineffective ,the standard errors and test morale be biased. The link between residuum can appear in the graphs, for example of groundwater location in a certain area . Statistical models are often used to represent views in terms of random variables and these models can be used to estimate the based on Probability theory. [1]. attribute actual data trend --surface using of duet dimension polynomial equation An observation at a dot i to enable to writs Yi = Y(u, v), where u and v represent a rectangular coordinate system for the area. Trend --surface analysis assume that Yi consists of two components, a trend component Ti represented by the polynomial equation and a residual component (Ri ) so that

Yi = Ti + Ri

T(u, v) = β00 + β10u + β01v + β20u² + β11uv + β21u²v + ... + βp0q0u q0v ... (1)

Then term p + q represents the order of the trend surface, first order or linear p + q = 1 second order or quadratic p + q = 2, third order or cubic p + q = 3, and so on, the technique has been documented in a number of textbooks.

Introduction

Trend --surface analysis is used for describing a certain style to represent real desktop data or highlights to represent changes that occur the spatial stochastic process. Base model is usually in the trend surface analysis in a way that ordinary least-squares and that assumes that the residuals are independent but if least-squares method is not efficient the standard errors and test morale be biased. The link between residuum can appear in the graphs, for example of groundwater location in a certain area. Statistical models are often used to represent views in terms of random variables and these models can be used to estimate the based on Probability theory. [1]. attribute actual data trend --surface using of duet dimension polynomial equation An observation at a dot i to enable to writs Yi = Y(u, v), where u and v represent a rectangular coordinate system for the area. Trend --surface analysis assume that Yi consists of two components, a trend component Ti represented by the polynomial equation and a residual component (Ri ) so that

Yi = Ti + Ri

T(u, v) = β00 + β10u + β01v + β20u² + β11uv + β21u²v + ... + βp0q0u q0v ... (1)

Then term p + q represents the order of the trend surface, first order or linear p + q = 1 second order or quadratic p + q = 2, third order or cubic p + q = 3, and so on, the technique has been documented in a number of textbooks.

Regression analysis

Defines the concept of a general regression mathematical representation for the average relationship between one variable and other variables called independent variables. Specialized Regression analysis to describe the relationship between variables in the form of have from that contains one independent variable is then called (simple linear regression model) write the following formula [4,5]

Y=β0 + β1X + e
Also, if the from contains several independent variables it is called (multiple linear regression model ). General linear model takes a written form and take more than independent variables and the formula be: 

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_m X_{im} + \epsilon_i \]

\[ i = 1,2,3, \ldots, n \]

Using arrays can formulate the following general form. \[ \text{Method 1: estimation maximum likelihood} \]

The vector of observation, denoted by \( Y \) and matrix notation the model is

\[ Y = X\beta + \epsilon \] (2)

the matrix \( X \) is

\[
X = \begin{bmatrix}
1 & u_1 & v_1 & u_1^2 & v_1^2 & u_1v_1 & \ldots & u_1^p & v_1^p \\
1 & u_2 & v_2 & u_2^2 & v_2^2 & u_2v_2 & \ldots & u_2^p & v_2^p \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & u_n & v_n & u_n^2 & v_n^2 & u_nv_n & \ldots & u_n^p & v_n^p
\end{bmatrix}
\]

and

\[ \beta = [\beta_0, \beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \cdots, \beta_{pq}] \]

\[ \beta = (X'X)^{-1}X'Y \] (3)

If there is a Spatial Autocorrelation between subjective random errors \( u_i \) because of random error in each site depends on neighbouring sites and the error in this case we

\[ E(\epsilon_i, \epsilon_j) \neq 0 \quad \forall \ i \neq j \]

From the simplest Autocorrelation is that appears in the first order autoregressive scheme \[ a_i = \rho a_{i-1} + e_i \]

That \( \rho \) autocorrelation between \( a_i \) and \( a_{i-1} \) and random error for this model the formula takes the same assumptions used in some way (OLS) and \( e_i \sim N(0, \sigma^2) \)

The relationship is found by \( (4) \)

\[ \rho(h) = \frac{C(h)}{\sigma(i)\sigma(j)} \]

That \( h = 1,2,3,4, \ldots, n \) and \( i = 1,2,3, \ldots, n-1 \)

Not that \( \sigma(i) \) and \( \sigma(j) \) represent the standard deviation of the data and \( C(h) \) represents the heterogeneity and which can be found from the relationship:

\[ C(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} (Y(s_i) - \bar{Y}(s_i))(Y(s_j) - \bar{Y}(s_j)) \]

\[ N(h) = \{(s_i, s_j) : s_i - s_j = h \} \quad i, j = 1,2, \ldots, n \]

Application

application has on real data represent groundwater levels for 47 borehole in sin jar in Nineveh. As each watch its coordinates by location \( X(S_i) \) and \( Y(S_i) \) because staying \( i = 1,2,3, \ldots, 47 \) in the exploration area and the data shown in table.

<table>
<thead>
<tr>
<th>u_i</th>
<th>v_i</th>
<th>Y(u_i, v_i)</th>
<th>u_i</th>
<th>v_i</th>
<th>Y(u_i, v_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9</td>
<td>3.7</td>
<td>6.2</td>
<td>1.5</td>
<td>3.1</td>
<td>9.6</td>
</tr>
<tr>
<td>3.8</td>
<td>4.3</td>
<td>7.4</td>
<td>1.3</td>
<td>5.1</td>
<td>7.8</td>
</tr>
<tr>
<td>1.5</td>
<td>3.7</td>
<td>6.2</td>
<td>1.1</td>
<td>5.4</td>
<td>7.6</td>
</tr>
<tr>
<td>1.4</td>
<td>3.9</td>
<td>6.2</td>
<td>1.2</td>
<td>4.4</td>
<td>5.4</td>
</tr>
<tr>
<td>4.5</td>
<td>4.4</td>
<td>7.3</td>
<td>1.4</td>
<td>4.2</td>
<td>5.3</td>
</tr>
<tr>
<td>1.8</td>
<td>3.8</td>
<td>6.1</td>
<td>3.2</td>
<td>3.9</td>
<td>7.1</td>
</tr>
<tr>
<td>1.5</td>
<td>4</td>
<td>6.1</td>
<td>1.6</td>
<td>3.5</td>
<td>6.2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3.6</td>
<td>1.8</td>
<td>3.1</td>
<td>6.1</td>
</tr>
<tr>
<td>3.3</td>
<td>4.9</td>
<td>3.4</td>
<td>1.8</td>
<td>2.9</td>
<td>6.1</td>
</tr>
<tr>
<td>4.9</td>
<td>4.2</td>
<td>7.4</td>
<td>1.1</td>
<td>5.1</td>
<td>5.8</td>
</tr>
<tr>
<td>1.7</td>
<td>3.5</td>
<td>6.2</td>
<td>1.5</td>
<td>4.9</td>
<td>6.6</td>
</tr>
<tr>
<td>4.7</td>
<td>4</td>
<td>7.3</td>
<td>1.4</td>
<td>5.4</td>
<td>5.2</td>
</tr>
<tr>
<td>1.8</td>
<td>3.4</td>
<td>6.2</td>
<td>1.4</td>
<td>4.7</td>
<td>5.2</td>
</tr>
<tr>
<td>1.7</td>
<td>2.9</td>
<td>6.2</td>
<td>1.7</td>
<td>2.8</td>
<td>6</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>3.9</td>
<td>1.5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>0.8</td>
<td>3.3</td>
<td>3.5</td>
<td>2</td>
<td>3.3</td>
<td>6.1</td>
</tr>
<tr>
<td>0.8</td>
<td>3</td>
<td>3.5</td>
<td>0.8</td>
<td>4.5</td>
<td>5.9</td>
</tr>
<tr>
<td>3.2</td>
<td>3.6</td>
<td>3.2</td>
<td>1.3</td>
<td>4.9</td>
<td>5.6</td>
</tr>
<tr>
<td>3.9</td>
<td>7.2</td>
<td>1.2</td>
<td>5.2</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>3.8</td>
<td>6.2</td>
<td>1</td>
<td>4.4</td>
<td>5.2</td>
</tr>
<tr>
<td>1.5</td>
<td>3.9</td>
<td>6.2</td>
<td>1</td>
<td>4.2</td>
<td>5.1</td>
</tr>
<tr>
<td>4.5</td>
<td>4.2</td>
<td>7.3</td>
<td>0.7</td>
<td>3.1</td>
<td>3.9</td>
</tr>
<tr>
<td>4</td>
<td>4.5</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimating the usual method model parameters

Model (1), rate first of the case where the directional orderly arrangement of the trend surface is known . [9] proposed a regression pattern for the regression models with the remaining spatial autocorrelation by following these steps: [10]

First Step. Appreciation \( \hat{\beta} \) in (1) (call this \( \hat{\beta} \) by (ols) assuming that the Residuals is in depended

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

Second Step. Gain the (ols) residuals, \( \hat{\epsilon} = Y - X\hat{\beta} \)

Third Step. Use the residuals to achieve \( \hat{\Omega} \) (or \( \hat{\Omega}^{-1} \), if this can be gained immediately.

Fourth Step. Appreciation \( \beta \) by generalized (gls).

\[ \hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y \] (6)

Fifth Step. get the (gls) residuals \( \hat{\epsilon} = Y - X\hat{\beta} \) and return to third step. Iterate until convergence.

The critical step is three, because the specification of \( \Omega \) could be approached in several different methods. I contrast a maximum likelihood (ml) approach (method 1) with two other approaches, direct estimation of the residual autocovariances (way 2) and estimation of an autocovariance function for the residue (way3).

Throughout, I shall write \( e \) for residuals, it being understood that at the first cycles the (ols) residuals \( \hat{\epsilon} \) are used, but in subsequent cycles (gls) residuals \( \hat{\epsilon} \) are used.

Method 1: estimation maximum likelihood

[11] studied the ML estimators for spatial regression model parameters where errors are automatically linked. trend-surface shape trend is a special case of regression modeling so that a natural start. point is to study this approach.
The auto covariance among Y values is stated by a parametric model function, a position two values (usually only dependent space between sites) and a set of unknown parameters ($\theta$). The autocovariances matrix must be positive, clear, and twice latent in relation to unknown parameters. Assuming a Gaussian process (ml) of $\beta \theta$ are achieved by maximize – likelihood

$$f(Y, L(\beta), L(\theta)) = \left|\Omega\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y-X\beta)'\Omega^{-1}(Y-X\beta)}$$

$$L(\beta, \theta, Y) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln |\Omega| - \frac{1}{2}(Y - X\beta)'\Omega^{-1}(Y - X\beta)$$

Where $|\Omega|$ denotes determinat of . the second model is the multivariate autoregressive scheme can be written as:

$$e = w + a \text{, where } a \sim N(0, \sigma^2 I) \text{ and } W \text{ is an } n \times n \text{ connectivity or describing linkage matrix the proximity or neighbor relationship among the sites }$$

$$i \text{ is an neighbor of }$$

$$W_{ij} = 1 \text{ if }$$

$$\text{ otherwise}$$

The next step is to create a matrix $\Omega$. $\Omega = \sigma^2 [(1 - \rho W)(1 - \rho W)]^{-1}$

$$\beta = (X\Omega^{-1}X)^{-1}X\Omega^{-1}Y$$

Then find the estimated values trend surface model $\hat{Y}$

That, $\hat{Y} = X\hat{\beta}$

Comparative selection coefficient adopted $R^2$ as well as mean square error MSE

$$\text{MSE} = \frac{\text{SSE}}{n-k}, \text{ that SSE= } Y^\beta Y - \left(\hat{\beta}\right)^\beta X^\beta Y$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

$$\text{MSE} = 0.606$$

$$R^2 = 0.7555$$

$$\text{MSE} = 1.9908$$

To improve the previous models suggested following methods.

**Method 2: lag order (autocovariance estimates).**

residuals Autocovariance the makes direct estimated . Let $C(s)$ indicate residuals Autocovariance at lag s. state

$$C(0) = \sum_{i=1}^{n} u_{ij}^2 \text{  s = 1, 2 ... }$$

$$C(s) = \sum_{i=1}^{n} u_{ij}u_{ij}^s \text{ is an s lag neighbor of i, s = 1, 2 ... , }$$

$$\hat{\omega}_{ij} = c(s) \text{ if j is an s lag neighbor of i}$$

$$\hat{\omega}_{ij} = 0 \text{ otherwise}$$

$$\Omega = \sigma^2 [(1 - \rho W)(1 - \rho W)]^{-1}$$

$$\beta = (X\Omega^{-1}X)^{-1}X\Omega^{-1}Y$$

[10] estimated several approaches of autocovariances.

Comparative selection coefficient adopted $R^2$ as well as mean square error MSE .

$$\text{MSE} = \frac{\text{SSE}}{n-k}, \text{ that SSE= } Y^\beta Y - \left(\hat{\beta}\right)^\beta X^\beta Y$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

$$R^2 = 0.7555$$

$$\text{MSE} = 1.9908$$

**Method 3: autocovariance function**

discussed [12] the Auto covariance function are parameters are estimated of a continuous in a different context He assumed that the structure of dependence among the residuals .

the function can be described by $u_1 = F(D_{ij})u_1 + e_1$, $F(D_{ij})$ refer to function of the distance among $j$ . It is more a summed , $E[e_1] = 0$

$u_{ij} = F(D_{ij})u^2_2 + e_1e_1$

providing $u_1$ and $e_1$ are independent, that

$$E[F(D_{ij})] = E[u_{ij}] = E[u_{ij}^1] = \rho_{ij}$$

Where ($\rho_{ij}$) is the autocorrelation separated among two sites by a distance $D_{ij}$. Although this seems like a approach a reasonable, $e$ and $u$ processes cannot be independent \text{, and in general the expected value will be an underestimate of the true autocorrelation function.}

Suggested approximating the distance function by the expression, $F_A(D_{ij}) = A + BD_{ij} + CD_{ij}^2$ and discuss an (ols) procedure for estimated the parameters $A$, $B$, and $C$ . $\text{f} A = 1$, then $F_A(D_{ij}) = 1 \text{ if } D_{ij} = 0$

Produce so that $F_1(D_{ij})$ can be interpreted as an autocorrelation function.

depends the way to define constant radius ($r$) around each point within which pairs of location captivated for the interpret of $A$, $B$, and $C$ past interim autocorrelation is equal $0$.

$$r < D_{ij} \text{ F}(D_{ij}) \hat{\omega}_{ij} = 0 \text{ otherwise. }$$

The way, the previous way unlike, allows the values in $\Omega$ to be adjusted for distance between location, it may be suitable for the case of irregular site distributions. For this reason. For this reason.

1 representing evaluating. When you make up $\hat{\omega}_{ij}$ in equation

$$\sigma^2 [(1 - \rho W)(1 - \rho W)]^{-1} = \Omega,$$

estimate (gls) get by

$$\hat{\beta}.\Omega^{-1}X^{-1}X\Omega^{-1}Y$$

Comparative selection coefficient adopted $R^2$ as well as mean square error MSE .

$$\text{MSE} = \frac{\text{SSE}}{n-k}, \text{ that SSE= } Y^\beta Y - \left(\hat{\beta}\right)^\beta X^\beta Y$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$

$$R^2 = 0.9222$$

$$\text{MSE} = 0.8921$$

methods comparing the used for error rate previously roads note that I said error rate significantly this proof that estimate of variance components and distance function approximation contributed significantly to reduce error and improve the model.
References

The following table shows the comparison between some values of truth and estimated values.

<table>
<thead>
<tr>
<th>Y(s)</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>3.8</td>
<td>6.1</td>
<td>6.5</td>
</tr>
<tr>
<td>7.4</td>
<td>7.9</td>
<td>5.8</td>
<td>7.9</td>
</tr>
<tr>
<td>6.2</td>
<td>4.2</td>
<td>5.6</td>
<td>6.1</td>
</tr>
<tr>
<td>6.2</td>
<td>5.2</td>
<td>5.4</td>
<td>6.4</td>
</tr>
<tr>
<td>7.3</td>
<td>7.7</td>
<td>5.3</td>
<td>6</td>
</tr>
<tr>
<td>6.1</td>
<td>6.1</td>
<td>7.1</td>
<td>6.2</td>
</tr>
<tr>
<td>6.1</td>
<td>5.5</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>3.6</td>
<td>2.6</td>
<td>6.1</td>
<td>3.1</td>
</tr>
<tr>
<td>3.4</td>
<td>1.4</td>
<td>6.2</td>
<td>3.2</td>
</tr>
<tr>
<td>3.4</td>
<td>2.4</td>
<td>6.1</td>
<td>6.1</td>
</tr>
<tr>
<td>7.4</td>
<td>6.1</td>
<td>5.8</td>
<td>7.1</td>
</tr>
<tr>
<td>6.2</td>
<td>4.3</td>
<td>6.6</td>
<td>6.9</td>
</tr>
<tr>
<td>7.3</td>
<td>5.3</td>
<td>5.2</td>
<td>7.2</td>
</tr>
<tr>
<td>6.2</td>
<td>4.2</td>
<td>5.2</td>
<td>6.2</td>
</tr>
<tr>
<td>6.2</td>
<td>6.9</td>
<td>6.6</td>
<td>6.1</td>
</tr>
<tr>
<td>3.9</td>
<td>3.1</td>
<td>6.6</td>
<td>4.2</td>
</tr>
<tr>
<td>3.5</td>
<td>2.5</td>
<td>6.1</td>
<td>3</td>
</tr>
<tr>
<td>3.5</td>
<td>3.1</td>
<td>5.9</td>
<td>3</td>
</tr>
<tr>
<td>3.2</td>
<td>3.9</td>
<td>5.6</td>
<td>3.6</td>
</tr>
<tr>
<td>7.2</td>
<td>4.2</td>
<td>5.5</td>
<td>6.7</td>
</tr>
<tr>
<td>6.2</td>
<td>6.9</td>
<td>5.2</td>
<td>6.2</td>
</tr>
<tr>
<td>6.2</td>
<td>6.1</td>
<td>5.1</td>
<td>6.1</td>
</tr>
<tr>
<td>7.3</td>
<td>5.3</td>
<td>3.9</td>
<td>7.1</td>
</tr>
<tr>
<td>7.5</td>
<td>8.5</td>
<td>8.7</td>
<td>8.9</td>
</tr>
</tbody>
</table>

The table represents a standard comparison of methods by the coefficient of determination and mean square error.

<table>
<thead>
<tr>
<th>ESTIMATON</th>
<th>METHOD 1</th>
<th>METHOD 2</th>
<th>METHOD 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.606</td>
<td>0.755</td>
<td>0.922</td>
</tr>
<tr>
<td>MSE</td>
<td>1.9908</td>
<td>1.001</td>
<td>0.8921</td>
</tr>
</tbody>
</table>

Conclusion
The present study has been considered as a repeat pattern to estimating the parameters of the trend surface models with residuals autocorrelation. The procedures discussed here would seem appropriate in those situations in which three levels of variation are suspects in the data, but they may also be useful in avoiding order misspecification trend-surface analysis. The importance of fine tests does not seem to be available, but sufficient information can be taken from the variance of the common contrast matrix to design approximate tests. Of conclusion supply, the (ml) way to be the best, Method 3 works well, and special suitable irregular Method 2 requires the evaluation of multiple autocorrelation requests and shows the progress of the method 1 in the current context, however, perhaps to a lot of autocorrelation in the data sites, from all this we can conclude that it is better on the third method way depending on the comparison criteria.

- تطبيق نماذج سطح الاتجاه مع التقدير

شيهام رياض ذنوة
كلية التمريض - جامعة الموصل, العراق

الملخص
يتناول البحث تحليل سطح الاتجاه للبيانات المكانية مع تطبيق ثلاثة نماذج على بيانات مكانية حقيقية تمثل ارتفاع مناسب المياه الجوفية. وتم تمثيل الطريقة الأولى تقييم معلمات نموذج سطح الاتجاه بواسطة (ml) (Dij). وفي نفس الوقت تحتاج الطريقة الثالثة إلى حل وحساب البواقي. وتم تصميم (Wij) مبدأ "الجوار" الذي تمثل ب Wij . تم التطبيق على بيانات حقيقية تمثل ارتفاع مناسب المياه الجوفية في 47 نبأ مع احداثيات موقعها في قضاء سنجر في محافظة بنيو.