Using Trend Surface Methodology to Compare Spatial Surfaces

Dale Zimmerman\(^1\), Zhi-Jun Liu\(^2\), and George Hallberg\(^3\)

Trend surface analysis is a very popular method for describing spatial data. When data on more than one response variable are available, it is generally desirable to compare the corresponding fitted surfaces obtained by trend surface analysis. Existing methods for such comparisons have serious deficiencies. The objective of this article is to develop statistically valid techniques for such comparisons which retain the familiarity and relative simplicity of trend surface analysis. In particular, we develop tests for surface parallelism and surface coincidence. The methods are illustrated by applying them to groundwater quality data from an aquifer in northeast Iowa.

INTRODUCTION

The multiple linear regression method known as polynomial trend surface analysis (TSA) is used widely in the physical sciences to describe geostatistical data. TSA is based on an assumption that the available data can be usefully described by a model in which each observation is equal to the sum of a trend component, generally taken to be a polynomial function of the spatial coordinates, and an error component. For example, for data taken from a two-dimensional region, a \(p\)th-order TSA model is given by

\[
Z(x, y) = \beta_{00} + \beta_{10}x + \beta_{01}y + \beta_{20}x^2 + \beta_{11}xy + \beta_{02}y^2 + \cdots + \beta_{r,s}x^ry^s + \epsilon(x, y).
\]

Here, \(Z(x, y)\) is the variate of interest's value at the site with Cartesian coordinates \((x, y)\); \(\beta_{00}, \beta_{10}, \ldots, \beta_{r,s}\) are unknown parameters; \(r + s = p\), the order of the polynomial; and \(\epsilon(x, y)\) is the error component. In the earliest applications of TSA, inference on model parameters was based on an ordinary least squares fit and thus the errors were taken, either explicitly or implicitly, to be independent normally distributed random variables with zero mean and constant variance (Grant 1957; Krumbein 1959). Standard \(F\)-tests from multiple regression analysis were used to determine the order of the trend surface. Later, however, it was argued that in many situations small-scale positive spatial dependence is likely to exist among the errors, rendering the ordinary least

---

\(^1\)Associate Professor of Statistics, University of Iowa, Iowa City, IA.
\(^2\)Research Associate, University Hygienic Laboratory, Iowa City, IA.
\(^3\)Geologist and Associate Director, University Hygienic Laboratory, Iowa City, IA.
squares estimates of model coefficients inefficient, biasing tests on the parameters, and tending to result in a fitted surface of too high an order (Cliff and Ord 1973; Ripley 1981). More recently, Agterberg (1984) and Haining (1987) have used generalized least squares to fit TSA models with spatially correlated errors.

Although some geostatistical problems in which TSA can be used involve fitting a surface to data on a single response variable, others involve fitting surfaces to data on several response variables. If these variables are commensurate or correspond to a single variable measured on several occasions, it may be of interest to compare the surfaces to each other. For example, one may want to compare the surface that describes current levels of a groundwater contaminant in an aquifer to the surface that corresponds to levels determined 10 years previously. The objective of this paper is to extend the statistical methodology of TSA to handle such comparisons.

The comparison of two or more fitted polynomial trend surfaces is not a new problem; indeed, it was an active area of research in the 1950s and 1960s as TSA was becoming popular among mathematical geologists and geographers. Three main approaches to the problem were proposed: a point correlation approach, a regression coefficient comparison, and a gradient-based approach. For a review of these, see Goodman (1983). Although these approaches are popular, they have several shortcomings, not the least of which is that none of them yield a statistically valid test for surface similarity (Zimmerman et al. 1996).

The methodology for comparing spatial surfaces that we propose here does not suffer from these shortcomings. In particular, it permits one to formally and validly test for coincidence and parallelism of the surfaces in a manner which accounts for spatial correlation.

**METHODOLOGY**

**Preliminaries**

To simplify the exposition, we consider an important particular case of the general problem. We assume that only two surfaces, with common spatial domain \( D \), are to be compared. The surfaces may correspond either to two commensurate variables or to the same variable measured on two occasions; the terminology we use, however, presumes the latter. We also assume that \( D \subset \mathbb{R}^2 \) and denote an arbitrary location in \( D \) as \( s = (x, y)' \). The cases of more than two surfaces or more than two dimensions can be handled by the same general methods to be presented here, but the notation becomes overly cumbersome.

The basis of our approach for comparing surfaces is to regard the data from which each fitted surface is constructed as a partial realization of a random process. Let \( \{Z_1(s) : s \in D\} \) and \( \{Z_2(s) : s \in D\} \) denote the random processes
corresponding to the variable of interest (or to perhaps a transformation, such as logarithms, of the variable of interest) on the two occasions. It is assumed that these processes are governed by the model

\[ Z_k(s) = \sum_{1 \leq i + j \leq p} \beta_{ijk} x^i y^j + \epsilon_k(s) \]  

where \{\beta_{ijk}: 0 \leq i + j \leq p; k = 1, 2\} are unknown parameters and \{\epsilon_1(s)\} and \{\epsilon_2(s)\} are independent, stationary, zero-mean Gaussian processes with covariance functions \( C_1(h) \equiv \text{cov}\{\epsilon_1(s), \epsilon_1(s+h)\} \) and \( C_2(h) \equiv \text{cov}\{\epsilon_2(s), \epsilon_2(s+h)\} \), respectively. We assume initially that \( p \) and the covariance functions are known; subsequently we will discuss their estimation.

Let \( s_1 = (x_1, y_1)' \), \ldots, \( s_{n_1} = (x_{n_1}, y_{n_1})' \) denote the \( n_1 \) locations where \( \{Z_1(s)\} \) is observed, and let \( s_{n_1+1} = (x_{n_1+1}, y_{n_1+1})' \), \ldots, \( s_{n_1+n_2} = (x_{n_1+n_2}, y_{n_1+n_2})' \) denote the \( n_2 \) locations where \( \{Z_2(s)\} \) is observed. Regarding the relationship of the first set of locations to the second set, two cases will be considered: (1) arbitrary sampling networks, in which no relationship between the first set and second set is specified, and (2) identical sampling networks, in which \( n_1 = n_2 = n \) and \( s_l = s_{n+l} \) for \( l = 1, \ldots, n \) (perhaps after a suitable permutation of the locations' labels). The second case permits an analysis based on paired differences among observations at the same locations and thus bears a resemblance to the case of paired-difference data in a classical two-sample problem. If, for fixed \( s \), the errors \( \epsilon_1(s) \) and \( \epsilon_2(s) \) actually exhibit positive temporal dependence rather than being independent as assumed in model (1), then the variance of the difference in these two errors will be smaller than the sum of their two variances. This reduction in variation, if large enough and consistent across locations, can result in substantially more powerful tests for surface coincidence and parallelism.

Finally, let \( Z_1 = [Z_1(s_{11}), \ldots, Z_1(s_{1n_1})]' \) and \( Z_2 = [Z_2(s_{21}), \ldots, Z_2(s_{2n_2})]' \) denote the vectors of responses observed at the two times, and let \( \epsilon_1 = [\epsilon_1(s_{11}), \ldots, \epsilon_1(s_{1n_1})]' \) and \( \epsilon_2 = [\epsilon_2(s_{21}), \ldots, \epsilon_2(s_{2n_2})]' \) denote the corresponding vectors of errors. Put \( Z = [Z_1', Z_2']' \) and \( \epsilon = [\epsilon_1', \epsilon_2']' \).

**Arbitrary Sampling Networks**

Under the stated assumptions and using the notation defined previously, the model for the observed data can be written as

\[ Z_l = \begin{cases} 
\beta_{001} + \sum_{1 \leq i + j \leq p} \beta_{1ij} x^i y^j + \epsilon_l & \text{for } l = 1, \ldots, n_1 \\
\beta_{002} + \sum_{1 \leq i + j \leq p} \beta_{2ij} x^i y^j + \epsilon_l & \text{for } l = n_1 + 1, \ldots, n_1 + n_2 
\end{cases} \]  

(2)

where \( E(\epsilon) = 0 \) and \( \text{var}(\epsilon) \equiv \text{diag}(\Sigma_{11}, \Sigma_{22}) \equiv \Sigma \), which we take to be positive definite. Here, \( \Sigma_{kk} \) is the \( n_k \times n_k \) matrix with \((r,s)\)th element given by \( C_k(s_{kr} - s_{ks}) \). Let \( q = (p+1)(p+2)/2 \) and let \( X_1 \) denote the \((n_1 + n_2) \times 2q\) model matrix corresponding to model (2), which is assumed to have full column
rank, and let $\beta$ denote the associated $2q \times 1$ vector of coefficients. The best linear unbiased estimator of $\beta$ is given by

$$\hat{\beta} = (X_1'\Sigma^{-1}X_1)^{-1}X_1'\Sigma^{-1}Z.$$

For purposes of testing for parallelism or coincidence of the two surfaces in model (2), two submodels are defined as follows.

**Parallelism Model**

$$Z_l = \begin{cases} 
\beta_{001} + \sum_{1 \leq i+j \leq p} \beta_{ij}x_i^l y_j^l + \epsilon_l & \text{for } l = 1, \ldots, n_1 \\
\beta_{002} + \sum_{1 \leq i+j \leq p} \beta_{ij}x_i^l y_j^l + \epsilon_l & \text{for } l = n_1 + 1, \ldots, n_1 + n_2
\end{cases}$$

**Coincidence Model**

$$Z_l = \beta_{00} + \sum_{1 \leq i+j \leq p} \beta_{ij}x_i^l y_j^l + \epsilon_l \text{ for } l = 1, \ldots, n_1 + n_2.$$

Now let $\text{RSS}_m = Z'\Sigma^{-1}Z - Z'\Sigma^{-1}X_m(X_m'\Sigma^{-1}X_m)^{-1}X_m'\Sigma^{-1}Z$ for $m = 1, 2, 3$, where $X_2$ and $X_3$ are the model matrices corresponding to the parallelism and coincidence models, respectively. Then, the parallelism hypothesis can be tested by comparing the value of

$$F_P = \frac{(\text{RSS}_2 - \text{RSS}_1)/(q - 1)}{\text{RSS}_1/(n_1 + n_2 - 2q)}$$

to a suitable percentage point of a (central) $F$ distribution with $q - 1$ and $n_1 + n_2 - 2q$ degrees of freedom. Similarly, we can test for coincidence by comparing the value of

$$F_C = \frac{(\text{RSS}_3 - \text{RSS}_1)/q}{\text{RSS}_1/(n_1 + n_2 - 2q)}$$

to a suitable percentage point of an $F$ distribution with $q$ and $n_1 + n_2 - 2q$ degrees of freedom.

The procedures just described are identical to the classical tests for parallelism or coincidence of regression surfaces described, for example, by Weisberg (1985) except that, to account for spatial correlation, the residual sums of squares used here are based on generalized least squares (GLS), rather than ordinary least squares (OLS), fits of the models. It should be noted that a “spatial analysis of variance,” of which the analyses of variance that correspond to these procedures are examples, has been used before for problems other than testing for parallelism or coincidence of surfaces; see Griffith (1978) and Gotway and Cressie (1990).

**Identical Sampling Networks**

Now suppose that $n_1 = n_2 = n$ and $s_l = s_{n+l}$ for $l = 1, \ldots, n$. Define the paired differences

$$D_l = Z_2(s_l) - Z_1(s_l) = Z_{n+l} - Z_l,$$

$$\delta_l = \epsilon_2(s_l) - \epsilon_1(s_l) = \epsilon_{n+l} - \epsilon_l \text{ for } l = 1, \ldots, n.$$
and let $D = (D_l)$ and $\delta = (\delta_l)$. It follows from model (2) that the model for $D_l$ is

$$D_l = \gamma_{00} + \sum_{1 \leq i+j \leq p} \gamma_{ij} x_i^j y_{i+j} + \delta_l$$

for $l = 1, \ldots, n$ (3)

where $\gamma_{ij} = \beta_{i+j} - \beta_{i-j}$, $E(\delta) = 0$, and var($\delta$) = $\Sigma_{11} + \Sigma_{22} = \Sigma_\delta$. Let $W$ denote the $n \times q$ model matrix corresponding to model (3) and let $y$ denote the associated $q \times 1$ vector of coefficients. Based on this model, the best linear unbiased estimator of $y$ is given by $\hat{y} = (W'\Sigma_\delta^{-1}W)^{-1}W'\Sigma_\delta^{-1}D$.

Testing for parallelism corresponds, in this situation, to testing that all the coefficients in model (3) except $\gamma_{00}$ equal 0 and hence can be tested by the mean-corrected overall $F$-test from a GLS analysis. That is, we compare the value of

$$F_p' = \frac{\text{RegSS}_4/(q-1)}{\text{RSS}_4/(n-q)}$$

to a suitable percentage point of an $F$ distribution with $q-1$ and $n-q$ degrees of freedom, where $\text{RegSS}_4 = D'\Sigma_\delta^{-1}W(W'\Sigma_\delta^{-1}W)^{-1}W'\Sigma_\delta^{-1}D - D'\Sigma_\delta^{-1}1_n(1_n\Sigma_\delta^{-1}1_n)^{-1}1_n\Sigma_\delta^{-1}D$ and $\text{RSS}_4 = D'\Sigma_\delta^{-1}D - D'\Sigma_\delta^{-1}W(W'\Sigma_\delta^{-1}W)^{-1}W'\Sigma_\delta^{-1}D$. Testing for coincidence corresponds in this situation to testing that all coefficients in model (3) equal 0 and hence can be tested by comparing

$$F_C' = \frac{\text{RegSS}_5/q}{\text{RSS}_4/(n-q)}$$

to a suitable percentage point of an $F$ distribution with $q$ and $n-q$ degrees of freedom, where $\text{RegSS}_5 = D'\Sigma_\delta^{-1}W(W'\Sigma_\delta^{-1}W)^{-1}W'\Sigma_\delta^{-1}D$.

Covariance Estimation and Order Determination

In practice, of course, the covariance matrices $\Sigma_{11}$ and $\Sigma_{22}$ (or $\Sigma_\delta$, if a difference-based approach is possible) and the orders $p_1$ and $p_2$ of the two surfaces are unknown. The covariance matrices can be estimated using a classical geostatistical approach based on the sample semivariogram; see Cressie (1991, ch. 2). Once $\Sigma$ (or $\Sigma_\delta$) has been estimated, the tests for parallelism and coincidence that we have described can be carried out by substituting $\hat{\Sigma}$ for $\Sigma$ in the previously given expressions for the test statistics. Of course, the resulting test statistics no longer have $F$ distributions, but provided that $\hat{\Sigma}$ is a consistent estimator of $\Sigma$, the asymptotic distributions of the test statistics are $F$ distributions and hence the tests are asymptotically valid; see Cressie (1991, pp. 242-243).

As for the determination of the order(s) of the surfaces, a stepwise procedure is described in Zimmerman et al. (1996) in which, starting with $p_0 = 0$, a $(p_0 + 1)$-order model is fitted by OLS; its residuals are tested for spatial correlation; and then an $F$-test (based on either an OLS or GLS fit) comparing the $p_0$-order and $(p_0 + 1)$-order models is carried out.
EXAMPLE: NITRATE CONCENTRATION IN BIG SPRING BASIN, IOWA

Big Spring Basin is a groundwater basin located in Clayton County in northeast Iowa. It is named after Big Spring, the largest spring in the state, which discharges from the underlying Galena aquifer. The area of the basin is 267 square km (103 square miles). Because of the proximity of the aquifer to the land surface and the existence of karst features (such as sinkholes) at the surface, the Galena aquifer in this basin is highly responsive to recharge events and susceptible to a relatively rapid influx of contaminants. The land in the basin is essentially all used for agriculture. There are no significant urban or industrial areas, landfills, or other major point sources that may affect groundwater quality. The combination of pure agricultural use and a responsive hydrogeological system in the basin offers a good opportunity for studying the impact of agricultural non-point source pollution on groundwater quality.

The Geological Survey Bureau of the Iowa Department of Natural Resources conducted a basin-wide sampling of private wells in the Galena aquifer in 1989, which was a drought year, and 1992, which was a relatively wet year. Wells were selected somewhat haphazardly, and inclusion in the sample depended on the presence and cooperation of the property owner, among other factors. Complete data were available from both years for 125 wells. All wells were sampled during a period of 7-10 days in October of the respective year so that the effects of short-term climatic and hydrologic events, as well as seasonal variability, would be minimized. Concentrations of several contaminants in the samples were analyzed by the University Hygienic Laboratory in Iowa City, a U.S. Environmental Protection Agency certified contract laboratory. Here we consider only the nitrate (NO3\(^-\)) concentrations, which were measured in mg/L and transformed to the log scale. In addition to the nitrate concentrations, the depth and \((x, y)\) coordinates of each well was recorded. All three coordinates were scaled to lie in the range \([-1, 1]\).

One of the objectives of the investigation was to determine whether the surface of nitrate concentrations over the basin was consistent with what would be expected from scientific considerations. For example, it is well known that nitrate concentrations decrease with increasing depth. Moreover, in the eastern half of the basin there were a large number of sinkholes, so higher concentrations were expected there. Another objective of the investigation was to determine how the nitrate concentration surface had changed from 1989 to 1992. In particular, since precipitation levels differed substantially in the two years, it was hoped that changes in the surface could lead to the formation of tentative hypotheses on how precipitation amounts and responsivity of the basin may interact to affect groundwater nitrate concentrations.

To address these objectives, first we performed a trend surface analysis...
separately for each year. Application of our order determination procedure resulted in a first-order model for both 1989 and 1992. Examination of the sample semivariograms of residuals, as well as Schmoyer's (1994) permutation test for spatial correlation, indicated that the spatial correlation was quite weak at the scale represented by most lags in the data set. Consequently, we fit first-order models by OLS, obtaining the fitted models

\[
\hat{Z}_1(d, x, y) = 1.66 + 2.22x - 0.79y - 1.45d
\]

for 1989 and

\[
\hat{Z}_2(d, x, y) = 2.52 + 1.90x - 0.45y - 1.72d
\]

for 1992. (Here, \(d\) represents depth and quantities in parentheses under the coefficients indicate their standard errors.) The signs and magnitudes of the estimated coefficients were consistent with what was expected from scientific considerations: (1) nitrate concentration decreased with increasing depth; (2) nitrate concentration increased as one moved from west to east; and (3) the relationship between nitrate concentration and the north-south coordinate was not as strong as the others.

Next, we compared the surfaces by testing for parallelism and coincidence using the \(F\)-tests we have previously described. The parallelism hypothesis was accepted (\(P = .95\)) but the coincidence hypothesis, assuming parallelism, was strongly rejected (\(P < .0001\)). This indicates that differences in precipitation levels (and possibly differences in other factors confounded with year) significantly affected the height, but not the shape, of the nitrate concentration surface.

The entire testing procedure was repeated using GLS fits of the two surfaces. The classical geostatistical procedure was used to estimate the covariance matrices for these fits. Results of the tests were qualitatively very similar to the results obtained under the assumption of no spatial correlation.

One interesting feature of the OLS analysis, however, was that the residuals from 1989 were significantly, positively correlated with the residuals (at the same sites) from 1992 (\(r = .70, P < .0001\)). This suggests that although it may be reasonable to assume that \(\Sigma_{11}\) and \(\Sigma_{22}\) are scalar multiples of an identity matrix for these data, it may not be reasonable to assume that \(\Sigma\) is block diagonal, as in model (2). Furthermore, there was evidence of somewhat greater variation among the measurements in 1992 than in 1989. For these reasons, and since the data lie on identical sampling networks, we carried out the difference-based approach as well. The conclusions drawn from applying this approach were not fundamentally different than those obtained using the original (i.e. undifferenced) observations.
REFERENCES


BIOGRAPHICAL SKETCH

Dale Zimmerman is an Associate Professor in the Department of Statistics and Actuarial Science at the University of Iowa. He graduated from Iowa State University in 1986 with a Ph.D. in Statistics. Dale teaches and conducts research in spatial statistics, environmental statistics, and linear models.

Zhi-Jun Liu is a Research Associate in the University Hygienic Laboratory at the University of Iowa. He received an M.S. in Statistics and a Ph.D. in Geography from the University of Iowa. His research interests include environmental modeling, spatial statistics, and geographic information systems.

George Hallberg is Associate Director of the University of Iowa Hygienic Laboratory and an Adjunct Professor in Geology and Civil and Environmental Engineering. He received his Ph.D. in Geology from the University of Iowa in 1975. Hallberg conducts research in environmental monitoring and public health surveillance.