



## USING OF SPATIAL ANALYSIS TO IMPROVE PRECISION IN AGRONOMIC TRIALS OF BREAD WHEAT (*Triticum aestivum* L.)

Mohamed, N. A.\* and A. A. Hamada\*\*

\* Cent. Lab. for Design and Stat. Anal. Res., Agric. Res. Cent. Giza, Egypt.

\*\* National Wheat Res. Prog., Field Crops Res. Inst., ARC, Giza, Egypt.

### ABSTRACT

Soil heterogeneity often decreases precision in large yield trials. Estimation of, and adjustment for fertility trends within a trial may increase precision. Two methods for estimating fertility trends (or spatial variation) are least squares smoothing and papadakis (nearest neighbor analysis) were evaluated in two experiments of bread wheat and compared them with the classical analysis such as randomized complete block (RCB) and triple lattice designs. Two experiments were carried out at Gemmeiza Research Station in 2001/2002 and 2002/2003 seasons. The experimental design was 10 X10 triple lattice. Each experiment consisted of a trial with three replications of 88 lines of bread wheat and 12 local cultivars namely; (Gemmeiza 3, Gemmeiza 5, Gemmeiza 7, Gemmeiza 9, Gemmeiza 10, Giza 164, Giza 168, Giza 170, Sakha 8, Sakha 61, Sakha 69 and Sakha 93). It could be summarized that: The results showed that the papadakis method was superior for all other used types of analyses in the two seasons. It decreased the percentage values of experimental error for RCBD from 67.604% and 57.404% to 2.85% and 2.266% in both seasons, respectively. Lattice was less than least squares smoothing and papadakis in reduction percentage in EMS. Also papadakis method was more effective for reducing CV. and raising the precision as other types of analyses. Lattice and least squares smoothing gave little reduction in CV. Correlation coefficients for adjusted means for lattice and least squares smoothing were more highly with RCBD, showing that either of these two methods gave similar results in this data set, and the general rankings of the genotypes were similar RCB means. It could be concluded that spatial analysis can improve the efficiency of field trials.

### INTRODUCTION

Many field trials in agriculture, including trials to determine genotypic yield potential, are carried out using a complete block design. The purpose of blocking is to increase precision by ensuring that treatments are evaluated with respect to similar environmental conditions within a block. Heterogeneity among plots within a block causes the estimate of a difference between two treatments to vary across blocks. Accordingly, efficiency of the randomized complete block design (RCBD) tends to be poor in trials involving a large number of treatments. To increase precision in such a trial, one approach is to reduce block size by employing an incomplete block design such as a lattice, lattice square, or one of the more flexible, but potentially less efficient designs (Patterson *et al.*, 1978). Another approach is to use a method of analysis that utilizes information on plot position to estimate and correct for spatial variation in yield potential due, for example, to differences among plots in soil fertility, moisture, or even pest populations. These analyses require contiguous or regularly spaced plots, arranged in a strip or

rectangular grid given an appropriate layout, these analyses can be applied to data from a complete or an incomplete block design.

Increasing efficiency via an analysis that includes estimation of spatial variation in yield potential has been the focus of a number of recent articles in the agronomy. (Papadakis, 1937; Williams, 1952; Wilkinson *et al.*, 1983; Green *et al.*, 1985; Gleeson and Cullis, 1987 and Lefkovitch, 1992).

Green *et al.* (1985) and Clarke *et al.* (1994) reported that Least squares smoothing effectively removed systematic variation and improved precision of treatment comparisons.

Papadakis method (Papadakis, 1937) received little attention until discussed by Bartlett (1978). Warren and Mendez, (1982); Wilkinson *et al.* (1983) and Bhatti *et al.* (1991) indicated that, the papadakis model was biased and developed a method called nearest neighbor analysis that overcomes the biased error (Nasr and Omar, 1999). Also, among the methods described are trend analyses (Kirk *et al.*, 1980; Tamura *et al.*, 1988; Bowman, 1990; Nasr, 1994; Nasr and El-Hady, 1999). Stroup and Mulitze (1991) have noted a controversy between those who favor spatial analyses and those who prefer incomplete block designs. More recently, Zimmerman and Harvill (1991) proposed a method for fitting fertility trends directly rather than by differencing or use of neighbor residuals. Cavell *et al.* (1993) reported that precision may be improved by using spatial analysis methods when spatial variation is accounted in estimation of treatment means.

Spatial analysis methods differ in their assumptions about the pattern of variability or in how the assumptions are incorporated in the analysis (Clarke and Baker, 1996). Randomized complete block and incomplete block analyses assume that the trend effects are constant within each block. Least squares smoothing (Green *et al.*, 1985) estimate trend effects from nearby plots with most weight on the closest neighbors. The papadakis method (Bartlett, 1978) places equal weight on plots that are immediate neighbors. Least squares smoothing and the papadakis method assume the trend is locally linear but differ in how this assumption is incorporated into the analysis. These concepts can be applied to other large field experiments.

Recommendations have been based on the assumption that the best method is that which produces the smallest calculated error, and there has been little attention given to validity of estimates of precision or corresponding tests of hypothesis. In making comparisons with the traditional randomized block analysis, there has also been a tendency to misrepresent the properties of this analysis by suggesting that the analysis is invalidated by the presence of systematic variation in the field.

The goal of this study was to investigate the possibility that spatial analysis by least squares smoothing and papadakis adjustment (nearest neighbor analysis) would result in greater precision compared with a randomized complete block and lattice analysis, where a large number of entries are tested in field trials.

## MATERIALS AND METHODS

Two identical experiments were conducted at Gemmeiza agricultural Research Station during the two growing seasons of 2001/2002 and 2002/2003 to evaluate some bread wheat genotypes under Egyptian conditions. Each experiment contained 88 bread wheat genotypes were introduced from Mexico and Syria and 12 local cultivars namely; (Gemmeiza 3, Gemmeiza 5, Gemmeiza 7, Gemmeiza 9, Gemmeiza 10, Giza 164, Giza 168, Giza 170, Sakha 8, Sakha 61, Sakha 69 and Sakha 93). The experimental design was 10 X 10 triple lattice. The field layout consisted of a 30 X 10 grid of plots, (Fig. 1 a and b). Each plot consisted of 6 rows with 4 m length with 20 cm between. Cultural practices were applied as usually recommended for the ordinary wheat fields. The harvested area was 3.6 m<sup>2</sup>. Grain yield was recorded in kg/plot for each experiment.

Statistical procedures:

The analysis described below assumes a rectangular (a X b) layout of plots, with row position indexed by  $R_i$ , where  $i = 1, a$ , and column position by  $C_j$ , where  $j = 1, b$ . Each of the  $t$  entries is replicated  $r$  time, so that the total number of plots in the grid is  $ab = rt$ . For the plot in row  $i$  and column  $j$ , plot (i, j) let  $Y_{ij}$  represent the observed yield,  $T_{ij}$  the unknown yield potential, and  $\tau_k(ij)$  the effect for the entry assigned to this plot. Then a model which incorporates spatial variation is:

$$Y_{ij} = \mu + \tau_k(ij) + T_{ij} + \epsilon_{ij} \quad [1]$$

Where:  $\mu$  is an overall mean,  $\tau_k(ij)$  and  $T_{ij}$  are assumed to be fixed effects and the  $\epsilon_{ij}$  are random errors.

Mapes of yield trends for field plots were calculated using the percent deviation from trial mean  $\{ [(Y_i - Y_j) / Y_j] \times 100\}$ .

At first, data from the two experiments were analyzed using randomized complete block design and 10 X10 triple lattice squares, then data were re-analyzed by using two methods of spatial analysis such as:

### A - Least squares smoothing analysis:

For field experiments, Green *et al.*(1985) considered a model that includes a smooth trend and an independent error. For any plot  $i$ , Containing genotype  $J$ , the model for yield response is:

$$Y_{ij} = \mu + \tau_k(ij) + T_{ij} + \epsilon_{ij} \quad [2]$$

The terms  $\mu + \tau_k(ij)$  represent the mean performance of genotype  $k$  occurring in the plot indexed by the  $i_{th}$  row and the  $j_{th}$  column of the experiment.  $T_{ij}$  is that part of environmental component of performance in plot  $ij$  that can be attributed to systematic spatial variation,  $\epsilon_{ij}$  constitutes the random component of the environmental effect on plot  $ij$ .

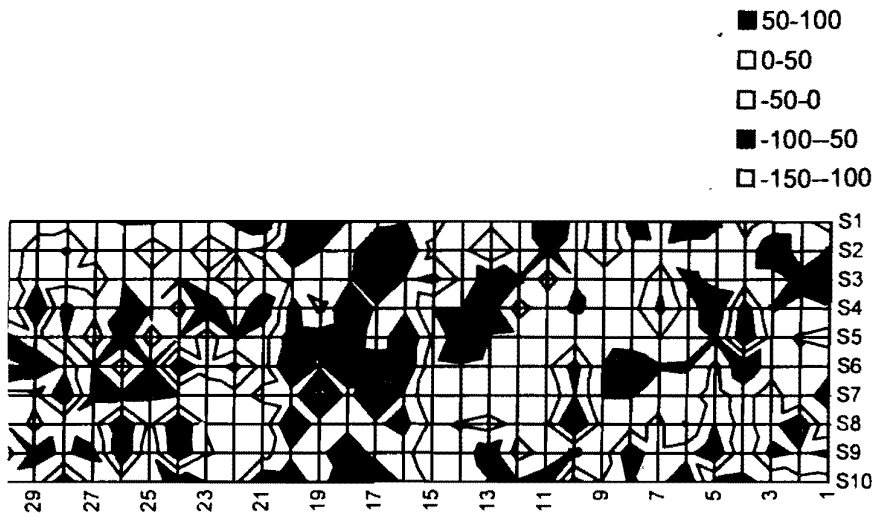


Fig. 1-a: Map of fertility gradient of field plots and variability pattern in season (2001/2002) at Gemmeiza Research Stations

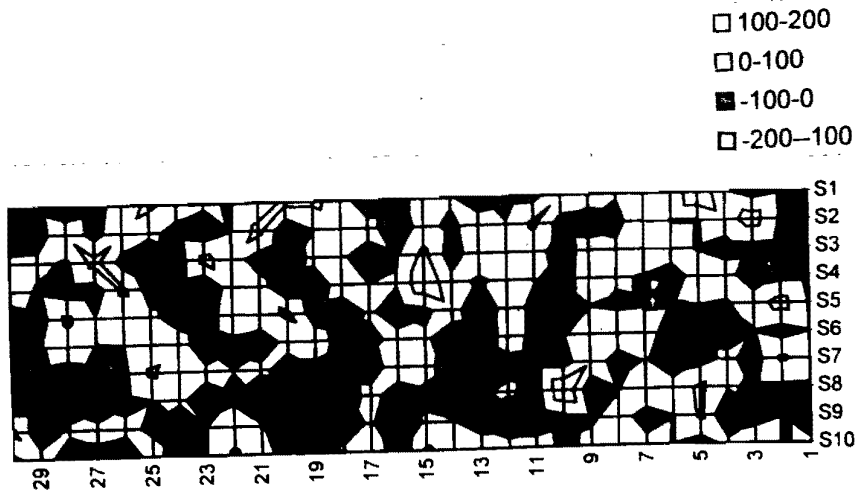


Fig. 1-b: Map of fertility gradient of field plots and variability pattern in season (2002/2003) at Gemmeiza Research Stations

For the randomized block analysis, the trend effects,  $T_{ij}$  were assumed to be constant for all plots in the same block, i.e.,  $T_{ij} = \beta_i$ . For least squares smoothing, trend effect,  $T_{ij}$  were estimated by assuming that, within rows, their second differences were approximately zero ( $T_{ij-1} - 2T_{ij} + T_{ij+1} = 0$ ) and that the independent residuals  $\varepsilon_{ij}$  were uncorrelated with each other and with the trend. Least squares smoothing requires the choice of a tuning coefficient that balances the need for a smooth trend with the requirement for uncorrelated residuals. One approach to choosing a suitable tuning coefficient is to choose one that minimizes the serial correlation between residuals (Green *et al.*, 1985). These authors pointed out that serial correlation will be negative for small tuning coefficient and positive for large tuning coefficient.

To implement that least squares smoothing analysis using SPSS program. (SPSS inc. 1995) function "Transform" can then be used to fit the model in Eq. [5] and obtain adjusted entry means.

#### **B – Papadakis analysis (neighbor analysis):**

We used an iterated one-dimensional modification of papadakis method (Stroup *et al.*, 1984) to calculate a trend index from the neighbors on the long side of each plot. A residual is calculated for each plot by subtracting the appropriate entry mean from the plot yield. A measure of yield potential,  $X_{ij}$ , is then computed for each plot as the average of the residuals for the neighboring plots. Adjusted means are obtained for entries by treating the measures  $X_{ij}$  as values of a covariate. That is, adjusted means for entries are based on fitting the model

$$Y_{ij} = \mu + \tau_K(ij) + bX_{ij} + \varepsilon_{ij} \quad [3]$$

Where. For interior plots,

$$X_{ij} = \frac{1}{4} (r_{ij-1} + r_{ij+1} + r_{i-1,j} + r_{i+1,j}), \quad [4]$$

$$r_{ij} = Y_{ij} - Y_{K(ij)}, \quad [5]$$

and  $Y_{K(ij)}$  is the mean yield for the entry assigned to plot (i, j). For border plots,  $X_{ij}$  is the mean of the  $r_{ij}$  for the two or three neighboring plots. Blocks are ignored when carrying out the papadakis analysis. The papadakis analysis is sometimes described as an analysis of covariance, but note that  $X_{ij}$  is not a true covariate because the  $r_{ij}$  values are calculated from the observed yields  $Y_{ij}$ . Version of the papadakis analysis is also referred to as nearest neighbor analysis (Pearce and Moore, 1976; Bhatti *et al.*, 1991). To implement the papadakis analysis using SAS program. (SAS institute, 1989) Proc GLM can then be used to fit the model in Eq. [3] and obtain adjusted entry means.

## **RESULTS AND DISCUSSION**

Mean performance value of 100 genotypes of wheat for grain yield / plot and coefficient of variation (CV.) are presented in Table 1. The results indicated that, there were highly significant differences among all studied genotypes mean and CV. values were high (23.21% and 26.04, respectively) during the two seasons. As soon as, Figure (1 – a & b) presents field layout and the maps of fertility gradient of field plots and variability pattern in the two experiments, showing allocation of the 100 entries arrayed in an 30 X 10 grid.

RCBD in two seasons. These results were confirmed by Cavell *et al.*, (1993); Clarke *et al.*, (1994); Stroup *et al.*, (1994); Clarke and Baker, (1996) and Nasr and Omar, (1999).

**Table 3: Percentage of error sum of square (ESS) from total SS. and reduction Percentage in error mean square of RCBD vs. other designs in two trials.**

Parameters	2001 / 2002	2002 / 2003
%of ESS from total SS:		
1- RCBD	67.604	57.404
2- Lattice	31.498	47.717
3-Least squares smoothing	34.468	23.053
4- Papadakis	2.850	2.266
Reduction% in MSS (RCBD vs. other designs):		
1- Lattice	46.173	3.767
2-Least squares smoothing	82.963	82.363
3- Papadakis	95.802	96.062

Data in Table 4 presents the coefficient of variation and percentage of gain in CV. by using RCBD, lattice, Lss. and papadakis analyses in the two seasons of experimentation, it indicated clearly that papadakis analysis was more effective for reducing CV. and raising the precision in comparison with other analyses. The percentage of gain in CV. in the two seasons were the highest for papadakis analysis recording 79.47% and 80.18%, respectively. Thus, it may be concluded that gains in efficiency due to papadakis analysis over RCBD were sufficient enough to justify the use of this type of analysis in large yield trials where spatial variation exists. This conclusion was in agreement with the results of Patterson *et al.*, (1978); Bartlett (1978); Wilkinson *et al.* (1983); Bhatti *et al.*, (1991) and Nasr and Omar (1999).

**Table 4: Coefficient of variation and percentage of gain in CV. by using RCBD, lattice, LSS. and papadakis analysis in 2001 / 2002 and 2002 / 2003 seasons.**

Parameters	2001 / 2002	2002 / 2003
Coefficient of variation (cv):		
1- RCBD	23.210	26.040
2- Lattice	17.050	25.542
3- Least squares smoothing	9.590	10.950
4- Papadakis	4.764	5.161
Gain in cv%:		
1- Lattice	26.540	1.912
2-Least squares smoothing	58.682	57.949
3- Papadakis	79.474	80.180

\* Gain in cv% = [(cv of RCBD - cv of other analysis) / cv of RCBD] X 100

The degree of similarity in genotypes mean for yield adjusted by the different methods of analysis is indicated by the correlations in Table 5.

Correlation coefficients for adjusted means of lattice and least squares smoothing were more highly with RCBD, showing that either of these two methods gave similar results in this data set, and the general rankings of the genotypes were similar to RCB means. Papadakis gave adjusted means that were high correlated with unadjusted means but less than with adjusted means from the other two analyses. Similar results were reported in other large field experiments studies such as: Stroup *et al.*, (1994)

**Table 5: Spatial correlations between genotypes unadjusted mean yields (RCBD) and adjusted means using lattice, LSS and papadakis analyses from combined data of the two seasons 2001 /2002 and 2002 / 2003.**

Parameters	Lattice	LSS	Papadakis
RCBD	0.932***	0.934***	0.821***
Lattice	-	0.871***	0.767**
Least squares smoothing(LSS)		-	0.760**

\*\* : Significant at P < 0.01 level.

\*\*\* : Significant at P < 0.001 level.

The results of this study demonstrated improved efficiency due to the use of two methods of spatial analyses in large field trials with large heterogeneous blocks. The results therefore indicated that the lattice design gave a less reduced of the values of experimental error in RCBD. Least squares smoothing and papadakis analyses confirmed the results showing that there was sufficient systematic yield variation within blocks to mask genotypes differences. The argument for using spatial analysis such as papadakis analysis (nearest neighbor analysis) is that greater gains in precision can sometimes be achieved with such an analysis than with an incomplete block design and the conventional analysis. It seems clear that precision can be improved by using a nontraditional analysis without necessarily sacrificing validity.

## REFERENCES

- Bartlett, M. S. (1978). Nearest neighbor model in the analysis of field experiments. *J. R. Stat. Soc., Ser. B*, 40: 147 – 174.
- Bhatti, A. U.; D. J. Mulla; F. E. Koehler and A. H. Gurmani. (1991). Identifying and removing spatial correlation from yield experiments. *Soil Sci. Am. J.*, 55: 1523 – 1528.
- Bowman, D. T. (1990). Trend analysis to improve efficiency of agronomic trials in flue – cured tobacco. *Agron. J.*, 82: 499 – 501.
- Cavell, B.; D. T. Bowman and J. W. Burton (1993). Estimating spatial variation in analysis of data from yield traits: A comparison of methods. *Agron. J.*, 85: 1244 – 1253.
- Clarke, F. R. and R. J. Baker (1996). Spatial analysis improves precision of seed lot comparisons. *Crop. Sci.* 36: 1180 – 1184.

- Clarke, F. R.; R. J. Baker and R. M. De Pauw (1994). Moving mean and least squares smoothing for analysis of grain yield data. *Crop Sci.* 34: 1479 – 1483.
- Gleeson, A. C. and B. R. Cullis. (1987). Residual maximum likelihood (REML) estimation of a neighborhood model for field experiments. *Biometrics*, 43: 277 – 288.
- Green, P. J.; C. Jennison and A. H. Seheult (1985). Analysis of field experiments by least squares smoothing. *J. R. Statis. Soc. B.* 47: 299 – 315.
- Kirk, H. J.; F. L. Haynes and R. J. Monroe (1980). Application of trend analysis to horticultural field trials. *J. Am. Soc. Hortic. Sci.*, 105: 189 – 193.
- Lefkovich, L. P. (1992). Fertility adjustments in rectangular field trials. *Biometrics J.*, 34: 3 – 15.
- Nasr, S. M. (1994). Use trend analysis in statistical decision making in crop research. *J. Agric. Sci. Mansoura Univ.*, 19 (7): 2127 – 2138.
- Nasr, S. M. and M. A. Omar (1999). Reduction of experimental error in large yield trials to improve precision by using some methods of spatial analysis. *J. Agric. Sci. Mansoura Univ.*, 24: 1677 – 1687.
- Nasr, S. M. and M. M. El – Hady (1999). The use of trend analysis to remove spatial variation from field plot data. *Bull. Fac. Agric., Cairo Univ.*, 50: 41 – 54.
- Papadakis, J. S. (1937). Methode statistique pour des experiences sur champ. *Bulletin de Institut d' Amelioration des plantes a Salonique* No. 23.
- Patterson, H. D.; E. D. Williams and E. A. Hunter. (1978). Block designs for variety trials. *J. Agric. Sci.*, 90: 395 – 400.
- Pearce, S. C. and C. S. Moore. (1976). Reduction of experimental error in perennial crops, using adjustment by neighbouring plots. *Exp. Agric.*, 12: 267 – 272.
- SAS institute (1989). *SAS/STAT user's guide version 6. Volume 2.* SAS Inst. Cary, NC.
- SPSS inc. (1995). *Spss for windows user's guide. Release 6.1.2 (02 May 1995)* Standard version.
- Stroup, W. W. and D. K. Mulitze. (1991). nearest neighbor adjusted best linear unbiased prediction. *Am. Stat.*, 45: 194 – 200.
- Stroup, W. W.; P. S. Baenziger and D. K. Mulitze (1984). Removing spatial variation from wheat yield trials: A comparison of methods. *Crop Sci.* 34: 62 – 66.
- Tamura, R. N.; L. A. Nelson and G. C. Naderman. (1988). An investigation of the validity and usefulness of trend analysis for field plot data. *Agron. J.*, 80: 712 – 718.
- Warren, J. A. and I. Mendez. (1982). Methods of estimating background variation in field experiments. *Agron. J.*, 74: 1004 – 1009.
- Wilkinson, G. N.; S. R. Eckert; T. W. Hancock, and O. Mayo. (1983). nearest neighbor (NN) analysis of field experiments. *J. R. Statis. Soc. B.* 45: 151 – 200.



- Williams, R. M. (1952). Experimental designs for serially correlated observations Biometrika 39: 152 – 167.
- Zimmerman, D. L. and D. A. Harville. (1991). A random field approach to the analysis of field – plot experiments and other spatial experiments. Biometrics, 47: 223 – 239.

استخدام التحليل المكاني لزيادة كفاءة التجارب الحقلية في محصول القمح  
نجدى عبد العليم محمد\* و أسعد أحمد حماد\*\*  
\*المعمل المركزى لبحوث التصميم والتحليل الاحصائى - مركز البحوث الزراعية - جيزة -  
مصر  
\*\* البرنامج القومى لبحوث القمح - معهد بحوث المحاصيل الحقلية - مركز البحوث الزراعية  
- مصر

يعتبر عدم تجانس الحقول التجريبية ذات المساحات الكبيرة من أهم العوامل التي تؤدي لانخفاض دقة نتائج التجارب الحقلية خاصة عند تقييم السلالات والأصناف ذات العدد الكبير باستخدام تصميم القطاعات الكاملة أو الناقصة فقد لا تكون التحليلات القياسية النمطية كافية لفصل أثر التباين المكاني من قيمة الخطأ التجريبي مما يقلل من كفاءة التجربة. وللتغلب على هذه المشكلة يجب على الباحث استخدام طرق التحليل المكاني كالمطلوب حديثاً وجيد بديلاً لتحليل التباين النمطي. ويهدف هذا البحث لإيضاح امكانية استخدام طريقتين من طرق التحليل المكاني وهما Least squares smoothing او Papadakis analysis ومقارنة كفاءتهما في فصل أثر التباين المكاني في التجارب ذات المساحات الكبيرة من قيمة الخطأ التجريبي بالمقارنة بتحليل القطاعات الكاملة والناقصة. لذلك أقيمت تجربتان في محطة البحوث الزراعية بالجميزة خلال موسمي ٢٠٠١/٢٠٠٢ و ٢٠٠٢/٢٠٠٣ على ٨٨ سلالة مستوردة و ١٢ صنف محلي من أقماح الخبز. وتم استخدام التصميم الشبكي الثلاثي وحللت البيانات بطرق التحليل النمطية ثم اعيد تحليلها بطريقتي التحليل المكاني ويمكن تلخيص النتائج فيما يلي:

- ١ - تفوقت طريقة P apadakis a nalysis على باقي الطرق في فصل التباين المكاني من قيمة الخطأ التجريبي مما أدى إلى زيادة دقة التحليل والنتائج. فقد إنخفضت نسبة تباين الخطأ التجريبي من التباين الكلي في الموسمين من ٢٧,٦٠٤% ، ٥٧,٤٠٤% إلى ٢,٨٥% ، ٢,٢٦٦% على التوالي.
  - ٢ - أوضحت النتائج أن تحليل القطاعات الناقصة Lattice كان الأقل كفاءة في خفض متوسط مربعات الخطأ التجريبي عن القطاعات الكاملة مقارنة بطريقتي Papadakis , Least squares smoothing.
  - ٣ - أظهرت النتائج أن Papadakis analysis أكثر كفاءة في خفض قيمة معامل الاختلاف يليها طريقة Least squares smoothing بينما كانت القطاعات الناقصة أقل كفاءة في هذا الخصوص.
  - ٤ - كما بينت دراسة معاملات الارتباط بين المتوسطات المحسوبة من القطاعات الكاملة والمتوسطات المعدلة الناتجة من طريقتي Least squares smoothing، القطاعات الناقصة ارتفاع قيمة معاملات الارتباط بينها حيث كانت المتوسطات المعدلة الناتجة عن هاتين الطريقتين أكثر تشابهاً مع المتوسطات الطبيعية عن تلك المعدلة باستخدام طريقة Papadakis analysis.
- من النتائج السابقة يمكن القول اجمالاً أن التحليل المكاني يزيد من كفاءة التحليل الاحصائي للتجارب الحقلية ذات المساحات الكبيرة.