BU-1497-M

SAS/GLM AND SAS/MIXED FOR TREND ANALYSES USING FOURIER AND POLYNOMIAL REGRESSION FOR CENTERED AND NON-CENTERED VARIATES

BY

Walter T. Federer,

Murari Singh, and

Russell D. Wolfinger

ABSTRACT

Spatial variation in field experiments often exhibits a cyclical pattern. To facilitate the use a statistical procedure to account for this type of variation, A SAS program has been written. Fourier regression is used as it involves sines and cosines. To allow exploratory selection of an appropriate model to account for the spatial variation present in the experiment, codes for a randomized complete block design, for a row-column design, and for polynomial regression with both centered and non-centered variates have been written. A code for mixing a row-column analysis with an orthogonal polynomial regression analysis has also been included.

Key words: exploratory model selection, fixed and random effects.

1498
BU-. .-M August 2000
Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, New York 14853

Title: SAS/GLM and SAS/MIXED for Trend Analyses Using Fourier and Polynomial Regression for Centered and Non-Centered Variates

Authors: Walter T. Federer, 434 Warren Hall Cornell University, Ithaca, Ny 14853, e-mail wtf1@cornell.edu, Murari Singh, ICARDA, P. O. Box 5466, Aleppo, Syria, HYPERLINK "mailto:M.Singh@CGIAR.ORG" M.Singh@CGNET.com, and Russell D. Wolfinger, SAS Institute, R52, SAS Campus Drive, Cary NC 27513, e-mail sasrdw@unx.sas.com.

Purpose: Spatial variation that is cyclic in nature should have a statistical procedure to account for this type of variation. Since Fourier polynomial regression is a procedure that fits cyclic variations, a code is given here for such analyses. The particular data set used to illustrate the application of the codes is the eight row-seven column experiment on tobacco plant heights given by Federer and Schlottfeldt (1954) {also elsewhere in this book}. The experiment was designed as a randomized complete block design but laid out as an eight-row by seven-column design. The spatial variation in the experimental area was non-cyclical and not entirely in the row-column orientation. The Fourier regression model would not be expected to perform well for this data set. If it is desired to use non-centered polynomial regression, a code for this is given. Note that regressions to retain in the model will need to be determined from a Type I rather than a Type III or IV analysis. The order of the codes for the trend analyses in the following program are Fourier regression (FRTA), non-centered variate polynomial regression (NPRTA), randomized complete block (RCBD), row-column (RCD), orthogonal polynomial (centered variates) regression (PRTA), and a mixture of row-column and orthogonal polynomial regression. The last is considered to be the appropriate model for this data set. Since only three orthogonal polynomials of degree 4 and 6 in columns and degree 4 in rows, c4, c6, and r4, were omitted in the next to last analysis, it was decided to use the last analysis listed. Then for this model, the blocking effect parameters were taken to be random for the SAS/MIXED procedure and treatment estimates and means were obtained. A code written as below is useful for exploratory model selection in patterning spatial variation.

References: Federer, W. T. (1998). Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. Biometrics 54(2):471-481.

Federer, W. T. and C. S. Schlottfeldt (1954). The use of covariance to control gradients in experiments. Biometrics 10:282-290. Errata 11:251, 1955.

SAS Codes

```
/*--The SAS codes for obtaining standard textbook RCBD and RCD analysis, FRTA,
NPRTA, and PRTA analyses  are given below:-- */
data colrow;
 infile 'colrow.dat';
 input Yield row col Trt;

/*--code for Fourier polynomials, FRTA--*/
            NTrt = 7; Nrow = 8; Ncol = 7;
Frc1 = Sin(2*3.14159*col/Ncol)     ;
Frc2 = Cos(2*3.14159*col/Ncol)     ;
Frr1 = Sin(2*3.14159*row/Nrow)     ;
Frr2 = Cos(2*3.14159*row/Nrow)     ;
/*--code for non-centered polynomials, NPRTA--*/
pc1= col; pc2=col**2;pc3=col**3;pc4=col**4;pc5=col**5;pc6=col**6;
pr1= row; pr2=row**2;pr3=row**3;pr4=row**4;pr5=row**5;
```

```
pr6=row**6; pr7 = row**7 ;
 run;
/*--code for ANOVA using Fourier series, FRTA--*/
proc glm data = colrow ;
 class Trt row col ;
 model Yield = Frc1 Frc2 Frr1 Frr2 Frc1*Frr1 Frc1*Frr2
              Frc2*Frr1 Frc2*Frr2 Trt;
run;
/*--code for ANOVA using non-centered polynomials, NPRTA--*/

proc glm data = colrow;
 class Trt row col ;
 model  Yield = pc1 pc2 pc3 pc4 pc5 pc6 pr1 pr2 pr3 pr4 pr5 pr6
pr7 pc1*pr1 pc2*pr1 pc2*pr3 pc3*pr2 pc4*pr1 pc4*pr2 Trt;  run;

/*--code to construct orthogonal polynomials--*/
Proc iml;
/*--7 columns and up to 6th degree polynomials--*/
opn4=orpol(1:7,6);
opn4[,1]=(1:7)`;
 op4=opn4;
 create opn4 from opn4[colname={'col' 'c1' 'c2' 'c3' 'c4' 'c5' 'c6'}];
 append from opn4;
 close opn4;
/*--8 rows and up to 7th degree polynomials--*/
 opn3=orpol(1:8,7);
 opn3[,1]=(1:8)`;
 op3=opn3;
 create opn3 from opn3[colname={'row' 'r1' 'r2' 'r3' 'r4' 'r5'
 'r6' 'r7'}] ;
 append from opn3;
 close opn3;  run;
/*--merge in polynomial coefficients--*/
 data rcbig;
 set colrow;
 idx = _n_;
 proc sort data = rcbig;
  by col ;
 data rcbig ;
  merge rcbig opn4;
  by col ;
 proc sort data = rcbig;
  by row ;
 data rcbig ;
  merge rcbig opn3;
  by row ;
 proc sort data = rcbig ;
  by idx ;
run;
/*--ANOVA for randomized complete blocks(rows), RCBD--*/
Proc Glm data = rcbig ;
 Class row Trt ;
 Model Yield = row  Trt ;
 run ;
/*--ANOVA for row-column design, RCD--*/
Proc Glm data = rcbig ;
 Class row  col  Trt ;
```

```
 Model Yield = row  col  Trt ;
run;

/*--ANOVA using orthogonal polynomials after omitting regressors
which had an F-value less than the 25% level, PRTA--*/
 Proc Glm data = rcbig;
  Class Trt row  col  ;
  Model Yield = c1 c2 c3 c5 r1 r2 r3 r5 r6 r7 c1*r1 c2*r1
  c2*r3 c3*r2 c4*r1 c4*r2 Trt;
 run ;

/*--ANOVA for mixture of row-column and orthogonal polynomial
regression trend analysis--this is the preferred analysis--*/
Proc Glm data = rcbig ;;
  Class row  col  Trt ;
  Model Yield = row  col  Trt c1*r1 c2*r1 c2*r3 c3*r2 c4*r1 c4*r2 ;
 Run ;
/*--random blocking effects and fixed Trt effects--*/
Proc Mixed data = rcbig ;
  Class row  col  Trt ;
  Model Yield = Trt/solution ;
  Random row  col  c1*r1 c2*r1 c2*r3 c3*r2 c4*r1 c4*r2 ;
  Lsmeans Trt ;
run ;
```

SAS Program Output (abbreviated)

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	1363645.120	97403.223	7.03	0.0001
Error	41	568131.505	13856.866		
Corrected Total	55	1931776.625			

R-Square	C.V.	Root MSE	YIELD Mean
0.705902	11.62648	117.7152	1012.475

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	6	273875.4500	45645.9083	3.29	0.0097
FRC1	1	48018.4941	48018.4941	3.47	0.0698
FRC2	1	702583.1192	702583.1192	50.70	0.0001
FRR1	1	301163.4604	301163.4604	21.73	0.0001
FRR2	1	7263.2834	7263.2834	0.52	0.4732
FRC1*FRR1	1	2486.5375	2486.5375	0.18	0.6741
FRC1*FRR2	1	26380.3457	26380.3457	1.90	0.1751
FRC2*FRR1	1	107.9593	107.9593	0.01	0.9301
FRC2*FRR2	1	1766.4703	1766.4703	0.13	0.7229

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	6	233771.3341	38961.8890	2.81	0.0220
FRC1	1	17663.3134	17663.3134	1.27	0.2655
FRC2	1	718308.7485	718308.7485	51.84	0.0001
FRR1	1	301163.4356	301163.4356	21.73	0.0001
FRR2	1	7263.2583	7263.2583	0.52	0.4732

FRC1*FRR1	1	1924.3838	1924.3838	0.14	0.7113
FRC1*FRR2	1	26805.6766	26805.6766	1.93	0.1718
FRC2*FRR1	1	62.7531	62.7531	0.00	0.9467
FRC2*FRR2	1	1766.4703	1766.4703	0.13	0.7229

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	1727731.702	69109.268	10.16	0.0001
Error	30	204044.923	6801.497		
Corrected Total	55	1931776.625			

R-Square	C.V.	Root MSE	YIELD Mean
0.894374	8.145504	82.47119	1012.475

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	6	273875.4500	45645.9083	6.71	0.0001
PC1	1	96813.2065	96813.2065	14.23	0.0007
PC2	1	535987.4627	535987.4627	78.80	0.0001
PC3	1	222783.3577	222783.3577	32.76	0.0001
PC4	1	17076.3332	17076.3332	2.51	0.1236
PC5	1	130081.1699	130081.1699	19.13	0.0001
PC6	1	2178.7015	2178.7015	0.32	0.5756
PR1	1	278087.6327	278087.6327	40.89	0.0001
PR2	1	21476.7478	21476.7478	3.16	0.0857
PR3	1	43739.6582	43739.6582	6.43	0.0167
PR4	1	1967.8963	1967.8963	0.29	0.5946
PR5	1	20330.7758	20330.7758	2.99	0.0941
PR6	1	11851.9481	11851.9481	1.74	0.1968
PR7	1	10860.2508	10860.2508	1.60	0.2161
PC1*PR1	1	9578.6523	9578.6523	1.41	0.2446
PC2*PR1	1	18255.7824	18255.7824	2.68	0.1118
PC2*PR3	1	518.4962	518.4962	0.08	0.7844
PC3*PR2	1	64.3080	64.3080	0.01	0.9232
PC4*PR1	1	27989.2845	27989.2845	4.12	0.0515
PC4*PR2	1	4214.5876	4214.5876	0.62	0.4374

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	6	99655.69428	16609.28238	2.44	0.0483
PC1	1	98.68785	98.68785	0.01	0.9049
PC2	1	23.75368	23.75368	0.00	0.9533
PC3	1	67.88482	67.88482	0.01	0.9211
PC4	1	561.62765	561.62765	0.08	0.7758
PC5	1	1513.16254	1513.16254	0.22	0.6406
PC6	1	2820.59147	2820.59147	0.41	0.5245
PR1	1	17782.73420	17782.73420	2.61	0.1164
PR2	1	16210.23452	16210.23452	2.38	0.1331
PR3	1	14741.72628	14741.72628	2.17	0.1514
PR4	1	13480.96650	13480.96650	1.98	0.1695
PR5	1	12426.54810	12426.54810	1.83	0.1866
PR6	1	11559.77509	11559.77509	1.70	0.2023
PR7	1	10860.25084	10860.25084	1.60	0.2161
PC1*PR1	1	3426.85586	3426.85586	0.50	0.4833
PC2*PR1	1	6236.00944	6236.00944	0.92	0.3460
PC2*PR3	1	1843.47705	1843.47705	0.27	0.6065
PC3*PR2	1	633.86885	633.86885	0.09	0.7623

PC4*PR1	1	20811.22981	20811.22981	3.06	0.0905
PC4*PR2	1	4214.58759	4214.58759	0.62	0.4374

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	662190.3521	50937.7194	1.69	0.1004
Error	42	1269586.2729	30228.2446		
Corrected Total	55	1931776.6250			

R-Square	C.V.	Root MSE	YIELD Mean
0.342788	17.17205	173.8627	1012.475

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW	7	388314.9021	55473.5574	1.84	0.1056
TRT	6	273875.4500	45645.9083	1.51	0.1985

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW	7	388314.9021	55473.5574	1.84	0.1056
TRT	6	273875.4500	45645.9083	1.51	0.1985

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	1667110.584	87742.662	11.93	0.0001
Error	36	264666.041	7351.834		
Corrected Total	55	1931776.625			

R-Square	C.V.	Root MSE	YIELD Mean
0.862993	8.468638	85.74284	1012.475

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW	7	388314.902	55473.557	7.55	0.0001
COL	6	1159072.132	193178.689	26.28	0.0001
TRT	6	119723.549	19953.925	2.71	0.0281

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW	7	388314.902	55473.557	7.55	0.0001
COL	6	1004920.232	167486.705	22.78	0.0001
TRT	6	119723.549	19953.925	2.71	0.0281

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	1793028.425	81501.292	19.38	0.0001
Error	33	138748.200	4204.491		
Corrected Total	55	1931776.625			

R-Square	C.V.	Root MSE	YIELD Mean
0.928176	6.404311	64.84205	1012.475

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	6	273875.4500	45645.9083	10.86	0.0001
C1	1	96813.2065	96813.2065	23.03	0.0001
C2	1	535987.4627	535987.4627	127.48	0.0001
C3	1	222783.3577	222783.3577	52.99	0.0001
C5	1	133144.7539	133144.7539	31.67	0.0001
R1	1	278087.6327	278087.6327	66.14	0.0001
R2	1	21476.7478	21476.7478	5.11	0.0305
R3	1	43739.6582	43739.6582	10.40	0.0028
R5	1	20330.7758	20330.7758	4.84	0.0350
R6	1	11851.9481	11851.9481	2.82	0.1026
R7	1	10860.2434	10860.2434	2.58	0.1175
C1*R1	1	9735.5843	9735.5843	2.32	0.1376
C2*R1	1	20003.6652	20003.6652	4.76	0.0364
C2*R3	1	11087.8978	11087.8978	2.64	0.1139
C3*R2	1	47865.7583	47865.7583	11.38	0.0019
R1*C4	1	45098.6300	45098.6300	10.73	0.0025
R2*C4	1	10285.6523	10285.6523	2.45	0.1273

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	6	160441.5837	26740.2639	6.36	0.0002
C1	1	67779.6280	67779.6280	16.12	0.0003
C2	1	443098.2755	443098.2755	105.39	0.0001
C3	1	249994.1996	249994.1996	59.46	0.0001
C5	1	132223.5073	132223.5073	31.45	0.0001
R1	1	278087.6327	278087.6327	66.14	0.0001
R2	1	21476.7478	21476.7478	5.11	0.0305
R3	1	43739.6582	43739.6582	10.40	0.0028
R5	1	20330.7758	20330.7758	4.84	0.0350
R6	1	11851.9481	11851.9481	2.82	0.1026
R7	1	10860.2434	10860.2434	2.58	0.1175
C1*R1	1	9140.3988	9140.3988	2.17	0.1498
C2*R1	1	21256.3073	21256.3073	5.06	0.0313
C2*R3	1	15800.4345	15800.4345	3.76	0.0611
C3*R2	1	48709.6471	48709.6471	11.59	0.0018
R1*	1	44314.8960	44314.8960	10.54	0.0027
R2*C4	1	10285.6523	10285.6523	2.45	0.1273

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	1799231.771	71969.271	16.29	0.0001
Error	30	132544.854	4418.162		
Corrected Total	55	1931776.625			

R-Square	C.V.	Root MSE	YIELD Mean
0.931387	6.565027	66.46925	1012.475

General Linear Models Procedure

Dependent Variable: YIELD

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW	7	388314.902	55473.557	12.56	0.0001
COL	6	1159072.132	193178.689	43.72	0.0001
TRT	6	119723.549	19953.925	4.52	0.0023
C1*R1	1	9578.652	9578.652	2.17	0.1513
R1*C2	1	18255.782	18255.782	4.13	0.0510
C2*R3	1	7858.737	7858.737	1.78	0.1923
C3*R2	1	41660.946	41660.946	9.43	0.0045
R1*C4	1	44992.650	44992.650	10.18	0.0033
R2*C4	1	9774.420	9774.420	2.21	0.1473

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW	7	388314.902	55473.557	12.56	0.0001
COL	6	1019062.385	169843.731	38.44	0.0001
TRT	6	117916.250	19652.708	4.45	0.0025
C1*R1	1	9397.488	9397.488	2.13	0.1551
R1*C2	1	20305.653	20305.653	4.60	0.0403
C2*R3	1	12900.535	12900.535	2.92	0.0978
C3*R2	1	42698.784	42698.784	9.66	0.0041
R1*C4	1	44171.272	44171.272	10.00	0.0036
R2*C4	1	9774.420	9774.420	2.21	0.1473

Covariance Parameter Estimates (REML)

Cov Parm	Estimate
ROW	7290.8987058
COL	21799.310745
C1*R1	5842.8474878
R1*C2	15988.580408
C2*R3	10848.115411
C3*R2	40466.589248
R1*C4	41571.338509
R2*C4	4747.3414225
Residual	4437.2974998

Solution for Fixed Effects

Effect	TRT	Estimate	Std Error	DF	t	Pr > |t|
INTERCEPT		903.07551781	68.35032779	6	13.21	0.0001
TRT	1	129.72384541	37.61944012	30	3.45	0.0017
TRT	2	137.78424774	36.51437318	30	3.77	0.0007
TRT	3	168.80483626	38.07134440	30	4.43	0.0001
TRT	4	148.48914899	37.87631974	30	3.92	0.0005
TRT	5	62.38611199	36.39931336	30	1.71	0.0969
TRT	6	118.60818492	35.65171322	30	3.33	0.0023
TRT	7	0.00000000	.	.	.	.

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
TRT	6	30	4.64	0.0019

Least Squares Means

Effect	TRT	LSMEAN	Std Error	DF	t	Pr > |t|
TRT	1	1032.7993632	68.49924871	30	15.08	0.0001

TRT	2	1040.8597656	68.27610042	30	15.24	0.0001
TRT	3	1071.8803541	68.98349917	30	15.54	0.0001
TRT	4	1051.5646668	69.12808466	30	15.21	0.0001
TRT	5	965.46162980	68.79788141	30	14.03	0.0001
TRT	6	1021.6837027	68.53512010	30	14.91	0.0001
TRT	7	903.07551781	68.35032779	30	13.21	0.0001