

## 밭의 비옥도를 고려한 품종실험 분석 \*

윤성철<sup>1)</sup> 강위창<sup>2)</sup> 이영조<sup>3)</sup> 임용빈<sup>4)</sup>

### 요약

농사실험에서 품종실험자료를 분석할 때, 난괴법(Completely Randomized Block Design) 모형을 많이 이용하고 있다. 이 모형에서는 각 블록내의 모든 실험단위들에서 비옥도가 같다고 가정한다. 그러나 많은 경우에 각 블록내 실험단위들의 비옥도에 규칙적인 이질성이 존재한다. 이러한 이질성을 고려하기 위하여, 본 논문에서는 다단계 일반화 선형모형(Hierarchical Generalized Linear Models)을 이용하여 품종효과와 블록내의 비옥도 효과를 함께 모형화 하고, 이 모형으로 Scottish Agricultural Colleges의 목록에 실려 있는 자료를 분석하여, 마코프체인 몬테칼로(Markov Chain Monte Carlo)방법으로 분석한 결과와 비교해 본다.

### 1. 서론

Nelder와 Wedderburn (1972)에 의해 제안되어진 일반화 선형모형(GLM : generalized linear models)은 회귀분석의 영역을 크게 확장시켰다. Lee와 Nelder (1996)는 일반화 선형모형(GLM)에서 선형 예측변수(linear predictor)에 땅의 비옥도와 같은 변량효과(random effect)를 추가시킬 수 있는 다단계 일반화 선형모형(HGLM)을 개발하였다. 그림 1.1은 Scottish Agricultural Colleges (Besag et. al., 1995)의 목록에 있는 자료를 발췌한 것으로 공간 배치실험(spatial layout experiment)자료를 보여준다. 각 블록들은 매우 긴 밭을 나타내며 서로 떨어져 있고 각 밭을 일렬로 나열된 75개의 플롯(plot)으로 나누었다. 밭내의 각 플롯에는 75종의 봄보리 중 하나를 랜덤하게 선택하여 심었고, 수확기에 얻은 산출량과 품종은 표에 순서대로 기입되어 있다. 밭과 밭사이에는 상관성이 없지만, 밭내에서 이웃하는 플롯간에는 상관성이 있을 것으로 예상된다. 밭내의 각 플롯에서의 산출량은 순수 품종효과 뿐만 아니라 다양한 요인들, 특히 날씨와 플롯의 비옥도에 의해 영향을 받을 수 있다. 품종효과를 비교하는데 있어서 날씨효과는 같다고 취급할 수 있겠다. 만약 밭내의 비옥도를 같다고 가정하고 품종효과를 비교한다면 난괴법(CRBD) 모형이 되는데, 실제 각 밭은 매우 길고 좁게 뻗어 있어서 밭내에서 상당히 떨어진 플롯들 간에는 비옥도가 같지 않을 수 있다. 이 경우에 난괴법 모형을 사용하여 자료를 분석한다면 추정되어진 품종효과는 땅의 비옥도 효과와 교락되어있어 어느 품종이 우수한 것인지 정확히 구별해내지 못할 것이다.

\* 이 연구는 1997년도 교육부 기초과학 육성 연구비(BSRI-97-1415)지원에 의한 것임.

- 1) (151-742) 서울시 관악구 신림동 산 56-1, 서울대학교 자연과학대학 통계학과, 박사과정
- 2) (151-742) 서울시 관악구 신림동 산 56-1, 서울대학교 자연과학대학 통계학과, 박사과정
- 3) (151-742) 서울시 관악구 신림동 산 56-1, 서울대학교 자연과학대학 통계학과, 부교수
- 4) (120-750) 서울시 서대문구 대현동 11-1, 이화여자대학교 자연과학대학 통계학과, 교수

block 1 (합 1)		block 2 (합 2)		block 3 (합 3)	
블록	산출량	블록	산출량	블록	산출량
57	9.29	49	7.99	63	11.77
39	8.16	16	9.56	36	12.05
3	8.97	8	9.02	14	12.05
46	8.33	69	8.91	71	10.96
75	8.66	29	9.17	22	9.84
21	9.05	59	9.49	46	9.27
66	9.01	19	9.73	6	11.05
12	9.40	39	9.38	30	11.40
30	10.16	67	8.80	16	10.75
32	10.30	57	9.72	24	10.30
59	10.73	37	10.94	40	11.27
50	9.69	26	10.55	64	11.13
5	11.49	16	9.67	8	10.55
23	10.73	6	10.17	56	12.52
14	10.71	47	11.46	32	10.95
68	10.21	36	10.05	45	10.82
41	10.52	64	11.47	54	10.77
1	11.09	63	10.63	37	11.08
64	11.39	33	11.03	21	10.22
28	11.24	74	10.55	29	10.59
46	10.65	13	11.35	62	11.35
73	10.77	43	10.25	5	11.39
37	10.92	4	10.08	70	10.59
55	12.07	53	10.25	13	11.26
19	11.03	23	9.57	11	11.79
10	11.64	62	11.34	44	12.25
35	11.37	52	10.19	36	12.23
26	10.34	12	10.80	52	10.84
17	9.52	2	10.04	60	10.82
71	8.99	32	9.69	65	10.41
8	8.34	22	9.36	3	10.96
62	9.25	42	9.43	19	9.94
44	9.88	72	11.48	67	11.27
53	9.90	73	9.29	59	11.79
74	11.04	25	10.10	2	11.51
20	10.30	45	9.53	75	11.64
56	11.56	15	10.35	27	10.84
29	9.69	35	11.34	43	9.76
2	10.68	66	11.36	51	8.86
47	10.91	5	10.88	10	10.28
11	10.05	56	11.61	35	12.15
38	10.80	46	10.33	74	10.36
65	10.06	71	10.53	66	9.59
13	10.04	51	8.67	34	10.53
31	10.50	21	9.56	18	11.26
40	9.51	1	9.95	50	10.37
4	9.20	31	11.10	42	10.10
67	9.74	11	10.11	1	9.95
22	8.84	41	9.36	58	9.60
49	9.33	61	10.23	26	10.55
58	9.51	55	11.38	41	9.31
43	9.35	14	11.30	25	9.29
7	9.01	44	10.50	33	10.03
25	10.58	34	10.97	9	9.49
61	11.03	54	12.22	17	11.52
16	9.89	24	10.10	57	12.24
52	11.39	4	11.22	65	11.64
70	11.24	65	10.01	49	10.74
34	12.18	75	10.29	73	10.29
42	10.21	38	10.95	7	10.25
24	11.08	17	9.66	23	11.39
33	11.05	68	9.31	72	13.34
51	10.29	7	8.64	55	12.73
60	10.57	27	10.64	31	12.62
69	10.42	58	9.45	39	10.19
15	10.49	48	9.66	47	11.61
6	10.60	28	9.85	15	10.52
63	9.23	60	9.24	20	9.07
54	10.57	30	10.11	61	10.76
18	10.27	70	9.63	28	9.91
45	8.86	20	9.04	53	10.17
72	9.45	9	8.43	69	8.66
9	8.03	40	10.97	45	8.74
36	9.22	50	8.98	12	9.15
27	8.70	10	9.88	4	9.39

그림 1.1: 75종의 봄보리 공간배치실험(spatial layout experiment) 자료

그림 1.1에 제시한 자료를 난괴법모형으로 분석한 후 잔차에 대해 일차 자기상관(first-order autocorrelation)의 존재를 검정하면 대단히 유의하였다. 예를 들면, 밭 1의 경우 일차 자기상관계수 추정값의 t 값(t-value)은 7.61로 p 값(p-value)이 0.000으로 매우 유의하다. 따라서 순수 품종효과를 비교하기 위해서는 난괴법에서 밭내의 비옥도가 같다고 가정하기 보다는 밭내에 자기상관 유형으로 표현되는 비옥도의 이질성이 존재한다고 가정하고 난괴법 자료를 분석할 필요가 있겠다. 변량효과(random effect)인 땅의 비옥도는 대부분의 경우 관측이 불가능하며, 공간배치실험(spatial layout experiment)자료의 분석 시에는 항상 고려되어야 한다. 본 논문에서는 그림 1.1의 자료에 Lee와 Nelder (1996)가 제안한 다단계 일반화 선형모형(HGLM)을 적용하여 비옥도의 이질성을 고려한 자료분석을 실시하고, 그 결과를 Besag et. al. (1995)이 마코프체인 몬테칼로(MCMC)방법으로 분석한 결과와 비교해본다. Lee와 Nelder (1996)의 논문에서는 변량효과들이 독립인 경우에 대한 예들을 제시하였다. 본 논문에서는 다단계 일반화 선형모형이 변량효과들 사이에 상관이 존재할 때도 유용한 방법론임을 예시하고자 한다.

## 2. 모형(MODEL)

$Y_{ij(k)}$  ( $i = 1, 2, 3, j = 1, \dots, 75, k = 1, \dots, 75$ )를  $i$ 번째 밭의  $j$ 번째 플롯에  $k$ 번의 종자를 심어 얻어지는 산출량이라 하고 그 관측값을  $y_{ij(k)}$  라 하자. 예를 들면, 그림 1.1의 밭 1에서 첫 번째 플롯의 산출량은  $Y_{1\ 1(57)}$ 로 표현된다. 또한  $Y = (Y_{1\ 1(57)}, Y_{1\ 2(39)}, \dots, Y_{3\ 75(4)})'$  로 표현하고  $Y$ 의 관측값을  $y$ 로 나타낸다. Besag (1995)등이 품종효과와 땅의 비옥도효과를 동시에 고려하기 위하여 사용한 다단계 모형을 정리하면 다음과 같다.

$$\begin{aligned} \text{단계I} \quad Y|v &\sim N(\mu', \sigma^2 I), \\ \mu' &= E(Y|v) = X\beta + Zv, \end{aligned}$$

여기서  $X, Z$ 는 각각 모델행렬(model matrices)이고,  $\beta = (\mu_1, \mu_2, \mu_3, \beta_1, \dots, \beta_{75})'$  에서  $\mu_i$ 는  $i$ 번째 밭의 총평균,  $\beta_k$ 는  $k$ 번 품종의 품종효과를 나타내며,  $v = (v_{1\ 1}, \dots, v_{1\ 75}, \dots, v_{3\ 1}, \dots, v_{3\ 75})'$  에서  $v_{ij}$ 는  $i$ 번째 밭의  $j$ 번째 플롯의 비옥도를 나타내는 변량효과(random effect)다.

$$\text{단계II} \quad \pi(v|\sigma_v^2) \propto \sigma_v^{-n} \exp \left[ -(2\sigma_v^2)^{-1} \sum_{i=1}^3 \sum_{l \sim m} (v_{il} - v_{im})^2 \right],$$

여기서  $\pi(v|\sigma_v^2)$ 는  $\sigma_v^2$  이 주어졌을 때, 비옥도의 사전분포를 나타내며,  $n$ 는 플롯의 총수이고,  $\sum_{l \sim m}$  는 이웃하는 플롯의 쌍들에 대한 합을 의미한다. 즉,  $i$ 밭에서  $j(1 < j < 75)$ 번째 플롯을 예를 들면,  $\sum_{j \sim m} (v_{ij} - v_{im})^2$ 는  $j$ 번째 플롯에 이웃하는  $j - 1$ 과  $j + 1$ 번째 플롯에서 계산된  $(v_{ij} - v_{ij-1})^2 + (v_{ij} - v_{ij+1})^2$ 이 된다. 난괴법모형은  $v_{ij} = v_i, j = 1, \dots, 75$  로 밭내의 비옥도가 같다고 가정한다. 서론에서 그림 1.1의 자료를 이와 같은 난괴법모형으로 분석하면 잔차에 일차 자기상관이 존재한다는 것을 지적하였다. 시계열자료에서는 자료의 관측이 시간축에 따라 이루어지므로 자기회귀모형(autoregressive model)등이 적절할 수 있으나, Besag (1995, pp.19-23)등이 지적한 바와 같이 그림 1.1의 공간배치실험자료에서는 이웃한 플롯의 상관을  $(v_{ij} - v_{ij-1})^2$ 과 같이 한 방향으로만 표현하기 보다는 양 방향을 모두 고

려한  $(v_{ij} - v_{ij-1})^2 + (v_{ij} - v_{ij+1})^2$ 으로 표현하는 것이 더욱 타당하다고 생각된다. 앞의 단계II에서  $\pi(v|\sigma_v^2)$ 를 확률밀도함수  $f(v; \sigma_v^2)$ 로 다음과 같이 표현할 수 있겠다.

$$f(v; \sigma_v^2) = (2\pi\sigma_v^2)^{-(n-b)/2} \exp \left[ -(2\sigma_v^2)^{-1} v' Av \right],$$

$$A = \begin{bmatrix} B & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & B \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ & & & & \ddots & & & \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \quad 75 \times 75$$

이고,  $\sigma_v^2 = \sigma_v^2/2$ ,  $b$ 는 블록의 수이며  $v_{i1} = 0 (i = 1, 2, 3)$  라고 가정한다. 행렬  $B$ 은 Whittaker (1993, pp.143-144)의 분산-공분산행렬의 역행렬에 대한 보조 정리들에 의해 발내에서 플롯의 비옥도간의 편상관(partial correlation) 관계를 나타낸다. 즉, 행렬  $B$ 에서 이웃한 플롯의 비옥도 사이에는 직접적인 상관이 존재하지만, 나머지 플롯들 사이에는 직접적인 상관은 없고 이웃한 플롯을 통해서만 상관되어 있음을 알 수 있다.

Besag (1995)등은 앞의 단계I, II에 다른 모수들에 대한 사전분포(prior)를 추가하여 마코프체인 몬테칼로(MCMC)방법을 이용한 베이지안 분석을 시행하였다. 한편 Lee와 Nelder (1996)가 제안한 다단계 일반화 선형모형에서는 모수들에 대한 사전분포를 요구하지 않고 단계I과 II만을 이용하여 분석을 실시할 수 있다.

### 3. 다단계 일반화 선형모형 적합

2절의 단계I과 II에 의해 다단계 우도(hierarchical likelihood)함수  $h$ 는 다음과 같이 정의된다.

$$\begin{aligned} h &= l(\mu', \sigma^2; y|v) + l(\sigma_v^2; v) \tag{3.1} \\ &= -\frac{1}{2\sigma^2} (y - X\beta - Zv)' (y - X\beta - Zv) - \frac{225}{2} \log(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma_v^2} [v_1^2 + v_2^2 + v_3^2 + \sum_{i=1}^3 \sum_{j=3}^{75} (v_{ij} - v_{ij-1})^2] - \frac{222}{2} \log(2\pi\sigma_v^2) \end{aligned}$$

$l(\mu', \sigma^2; y|v)$ 는  $v$ 가 주어졌을 때  $y$ 의 조건부 로그우도함수이며  $l(\sigma_v^2; v)$ 는  $v$ 의 로그우도함수이다.

#### 3.1. 모수효과(FIXED EFFECT)와 변량효과(RANDOM EFFECT)에 대한 추정

식 (3.1)에서 산포모수(dispersion parameter)  $\sigma^2$ 와  $\sigma_v^2$ 가 주어졌을 때, 모수  $\beta$ 와 변량효과(random effect)  $v$ 의 추정값은 다단계 우도함수  $h$ 를 최대로 하는  $\beta$ 와  $v$ 값으로 다음식의

해가 된다.

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + U \end{bmatrix} \begin{bmatrix} \beta \\ v \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (3.2)$$

여기서

$$U = \frac{\sigma^2}{\sigma_\epsilon^2} \begin{bmatrix} u_1 & 0 & 0 \\ 0 & u_1 & 0 \\ 0 & 0 & u_1 \end{bmatrix}, \quad u_1 = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ & & & \ddots & & & \\ & & & & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \quad 74 \times 74$$

이고, 상기식은 Henderson (1975)의 혼합선형모형(mixed linear models) 추정식과 일치한다.

### 3.2. 산포모수(DISPERSION PARAMETER)에 대한 추정

산포모수  $\sigma^2$ 와  $\sigma_\epsilon^2$ 를 추정하기 위하여, Lee와 Nelder (1996)가 제안한 수정된 단면 다단계 우도 (adjusted profile hierarchical likelihood) 함수  $h_p$ 를 사용하였으며 이는 다음과 같다.

$$h_p = h_A |_{\beta=\hat{\beta}, v=\hat{v}} \quad (3.3)$$

여기서  $h_A = h + \frac{1}{2} \log[\det(2\pi\sigma^2 H^{-1})]$  이고,

$$H = -\sigma^2 E \left[ \frac{\partial^2 h}{\partial \beta \partial v} | v \right] = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + U \end{bmatrix},$$

$\hat{\beta}$ 와  $\hat{v}$ 는 식 (3.2)를 만족하는 해이다. 산포모수  $\sigma^2$ 와  $\sigma_\epsilon^2$ 의 추정값은  $h_p$ 가 최대가 되게 하는 값들로  $\partial h_A / \partial \sigma^2 |_{\beta=\hat{\beta}, v=\hat{v}} = 0$ 와  $\partial h_A / \partial \sigma_\epsilon^2 |_{\beta=\hat{\beta}, v=\hat{v}} = 0$ 의 해를 반복하여 구함으로써 얻어진다. 이때  $\hat{\beta}$ 와  $\hat{v}$ 는 매 반복시점에서 다시 계산되어진 값이다. 식 (3.3)으로부터 산포모수에 대한 스코어함수(score function)는 다음과 같이 계산된다.

$$\frac{\partial h_A}{\partial \sigma^2} |_{\beta=\hat{\beta}, v=\hat{v}} = 0, \quad \frac{\partial h_A}{\partial \sigma_\epsilon^2} |_{\beta=\hat{\beta}, v=\hat{v}} = 0,$$

여기서

$$\begin{aligned} \frac{\partial h_A}{\partial \sigma^2} &= \frac{1}{2\sigma^4} (y - X\beta - Zv)'(y - X\beta - Zv) + \frac{37}{\sigma^2} - \frac{1}{2} \text{tr} \left( H^{-1} \frac{\partial H}{\partial \sigma^2} \right), \\ \frac{\partial h_A}{\partial \sigma_\epsilon^2} &= \frac{1}{2\sigma_\epsilon^4} [v_1^2 + v_2^2 + v_3^2 + \sum_{i=1}^3 \sum_{j=3}^{75} (v_{ij} - v_{ij-1})^2] - \frac{111}{\sigma_\epsilon^2} - \frac{1}{2} \text{tr} \left( H^{-1} \frac{\partial H}{\partial \sigma_\epsilon^2} \right) \end{aligned}$$

## Normal Probability Plot

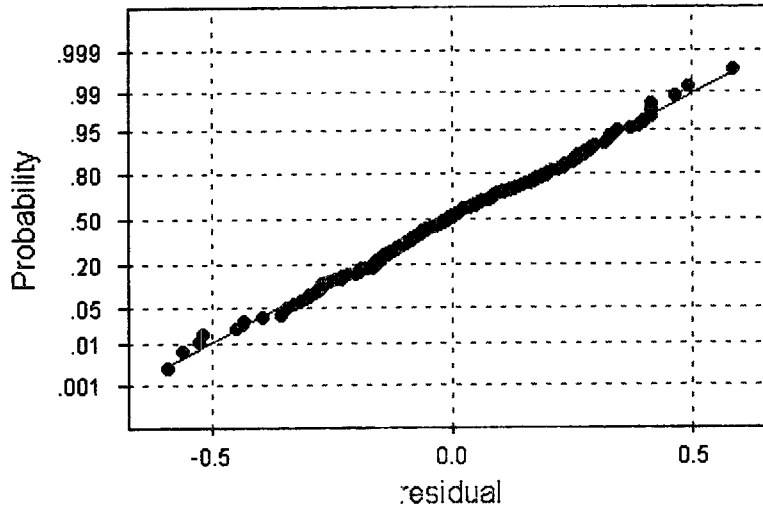


그림 4.1: 잔차에 대한 정규확률지 그림

이고, 이 스코어방정식(score equations)은 뉴턴-라프슨방법(Newton-Raphson method)으로 풀 수 있겠다.

2절의 단계I, II에서  $y|v$ 와  $v$ 의 분포가 모두 정규분포이므로, 다단계 우도함수  $h$ 는 혼합 선형모형에서 Henderson (1975)의 결합우도함수와 일치하고, 앞에서 제안한 추정법으로 구한  $\hat{\beta}$ 와  $\hat{v}$ 는 산포모수를 알 때 각각 최대우도추정량과 최량선형비편향예측량(BLUP : Best Linear Unbiased Predictor)이 된다. 또한, 수정된 단면 다단계 우도함수  $h_p$ 는 Patterson과 Thompson (1971)의 제한우도(restricted likelihood)와 상수항을 제외하면 같게 되어, 산포모수들의 추정은 제한최대우도(restricted maximum likelihood)추정이 된다. 한편, 다단계 일반화 선형모형은  $y|v$ 의 분포로 일반화 선형모형에서의 지수족(exponential family)을 가정하고,  $v$ 는 임의의 분포를 가정하므로 본 절에서 제안된 추정법은  $y|v$ 의 분포와  $v$ 의 분포가 정규분포가 아닐 때도 사용할 수 있다.

## 4. 모형 검토

Lee와 Nelder (1996)는 다단계 일반화 선형모형에서 모형의 적합도(goodness-of-fit)에 대한 판정기준으로 다음과 같은 척도편차(scaled deviance)를 제안하였고,

$$D(y, \hat{\mu}') = -2[l(\hat{\mu}', \sigma^2; y|v) - l(y, \sigma^2; y|v)],$$

상기 통계량  $D(y, \hat{\mu}')$ 의 자유도로  $n - \text{trace}(\mathbb{H}^{-1}\mathbb{H}^*)$ 를 사용하였다. 여기서  $n$ 는 자료의 총 개수

표 5.1: 우량 품종의 순위

순위	CRBD	RevG	RevU	HGLM
1	55	55	56	56
2	56	35	35	35
3	35	72	31	31
4	14	31	55	55
5	72	55	72	72
6	31	47	47	47
7	<u>64</u>	54	54	54
8	47	<u>18</u>	38	38
9	38	38	<u>18</u>	<u>18</u>
10	5	40	40	40
11	34	26	14	26
12	54	14	26	5
13	44	62	5	14
14	13	57	57	62
15	52	5	62	57

CRBD: Completely Randomized Block Design  
 RevG : Reversible Gibbs with Gaussian prior for fertility  
 RevU : Reversible Gibbs with uniform prior for fertility  
 HGLM : Hierarchical Generalized Linear Models

이고,  $H^* = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z \end{pmatrix}$  이다. 식 (3.1)의 다단계 우도함수를 이용하여 모형을 적합시켰을 때, 척도편차  $D(y, \hat{\mu}') = 75.79$ 이고 자유도  $n - \text{trace}(H^{-1}H^*)$ 는 75.78이 되어 모형에 적합성 결여(lack-of-fit)가 없음을 알 수 있었다. 앞의 그림 4.1은 잔차  $e_i = y_i - \hat{\mu}'_i$ , ( $i = 1, 2, \dots, 225$ )에 대한 정규확률지그림(normal probability plot)을 나타낸다. 그림에서 잔차들이 거의 직선 위에 있으므로  $y|v$ 는 정규분포를 따른다고 생각된다.

### 5. 분석 결과

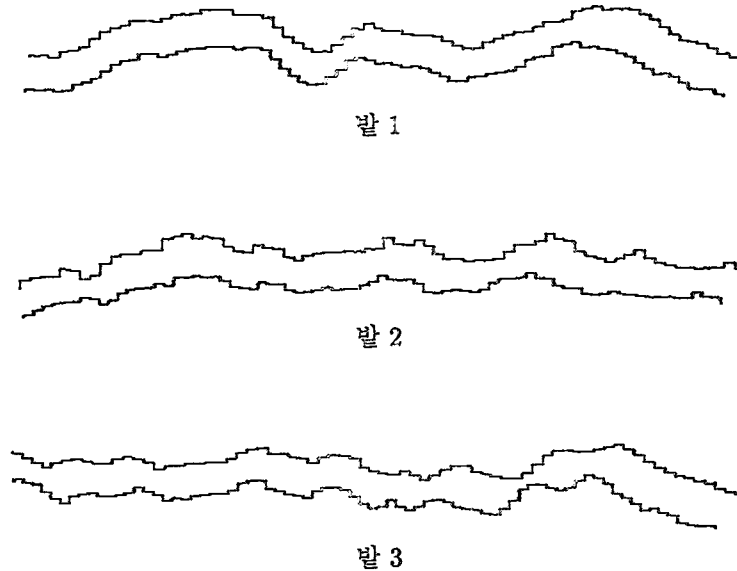
품종실험자료를 분석하는 중요한 목적은 우량품종들을 올바르게 가려내는 것이다. 이때 비옥도의 이질성을 감안하지 않으면 우량품종을 잘못 선정하는 오류를 범할 수 있다. 표 5.1은 발내 비옥도의 이질성을 고려하지 않는 난괴법(CRBD)에 의해 추정된 품종효과와, 이것을 감안한 분석법인 마코프체인 몬테칼로(MCMC)방법(RevG, RevU) 및 다단계 일반화 선형모형(HGLM)에 의해 추정된 품종효과들 중에서 1순위부터 15순위까지의 우량품종들을 기록한 것이다.

Besag (1995)등은 품종효과의 사전분포로 평균이 0이고 분산이  $1/\lambda$ 인 정규분포를 가정

하였고,  $\lambda$ 와 두 산포모수  $\sigma^2$ ,  $\sigma_\epsilon^2$ 의 사전분포로는 특정 감마분포일 때와,  $\lambda \equiv 0$  이고  $\sigma^2$ ,  $\sigma_\epsilon^2$ 은 앞에서와 동일한 사전분포를 따를 때를 고려하였다. 표 5.1에서 RevG는  $\lambda$ 의 사전분포가 감마분포일 때 마코프체인 몬테칼로(MCMC)방법의 분석결과이고, RevU는  $\lambda \equiv 0$  일 때, 즉 품종효과의 분포가 모호사전분포(vague prior)인 경우이다. 표 5.1에서 다단계 일반화 선형 모형과 마코프체인 몬테칼로(MCMC)방법이 15개의 동일한 우량품종을 찾아낸 것을 알 수 있다. 이와 같은 비옥도의 이질성을 고려한 분석방법들(RevG, RevU, HGLM)의 결과와 난괴법(CRBD)에서의 결과를 비교하면, 15순위 내에 있는 품종의 종류 뿐만 아니라, 순위에도 많은 차이가 있음을 알 수 있다. 예를 들어 64번 품종인 경우 난괴법에서는 7위이지만, 땅의 비옥도를 고려한 분석방법들에서는 15순위 내에 들지 못한다. 즉, 64번 품종은 실제 우량품종이 아닌데 실험한 땅이 비옥하여 산출량이 많았던 것으로 생각된다. 18번 품종은 이와는 반대의 경우로, 난괴법에서는 15순위 내에 들지도 못하지만, 나머지 방법들에서는 8위 혹은 9위로 실제로는 우량품종인데 실험한 땅이 비옥하지 않아 산출량이 적었다고 생각된다. 한편, 다단계 일반화 선형모형 분석결과와 품종효과에 모호사전분포(vague prior)를 가정한 마코프체인 몬테칼로(MCMC)방법의 결과(RevU)가 거의 일치하는데(10 위까지 완전히 일치한다), 마코프체인 몬테칼로방법 RevU는 두 산포모수  $\sigma^2$ ,  $\sigma_\epsilon^2$ 의 사전분포로 특정 감마분포를 가정하였고, 다단계 일반화 선형모형에서는 두 산포모수를 미지의 상수로 가정한다는 것을 감안하면, 위의 결과는 일반화 선형모형에서 최대우도추정치가 모호사전분포(vague prior)를 가정한 베이지안 추정치와 같다는 잘 알려진 사실이 다단계 일반화 선형모형에서도 성립될 수 있다는 것을 보여준다.

그림 5.1은 다단계 일반화 선형모형에서 추정된 각 발내 플롯의 비옥도를 마코프체인 몬테칼로(MCMC)방법의 결과와 같은 척도로 그린 것으로 두 그림이 매우 유사함을 알 수 있다. 여기서 x축 방향은 발내의 각 플롯을 나타내며 y축 방향은 각 플롯의 비옥도의 추정값을 나타낸다. 표 5.1과 그림 5.1에서 알 수 있는 사실은 마코프체인 몬테칼로(MCMC)방법들의 결과와 다단계 일반화 선형모형(HGLM)에 의한 결과가 품종효과의 추정과 비옥도 효과의 추정에서 거의 같은 순위와 유사한 패턴의 그림을 보여준다는 것이다. 따라서 다단계 일반화 선형모형과 Besag (1995)등이 제안한 마코프체인 몬테칼로(MCMC)방법이 모두 발내의 공간상관(spatial correlation)를 충분히 감안해 줄 수 있는 모형임을 알 수 있다. 본 논문에서 제안한 모델은 SAS PROC IML을 이용하여 프로그래밍 하였으며, 분석의 결과를 얻는데 있어서 마코프체인 몬테칼로(MCMC)방법에 비해 시간이 훨씬 짧게 걸린다(Lee와 Nelder (1996)에서 R. L. Smith교수와의 토론을 참조). 또한 다단계 일반화 선형모형(HGLM)에 의한 방법은 마코프체인 몬테칼로(MCMC)방법으로 분석할 때 발생하는 문제점 중의 하나인 수렴(convergence)에 대한 판단의 모호함이 없는 장점이 있다.





위 선 : 다단계 선형모형(HGLM).  
 아래 선 : 마코프체인 몬테칼로(MCMC)방법.  
 x축 방향 : 발내의 각 플롯.  
 y축 방향 : 각 플롯의 비육도 추정값.

그림 5.1: 각 발내에서의 비육도 추정값에 대한 그림

### 참고문헌

[1] Besag, J., Green, P. Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statistical Science*. Vol. 10. 1-66.

[2] Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*. Vol. 31. 423-447.

[3] Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society. B*. Vol. 58. 619-678.

[4] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. A*. Vol. 135. 370-384.

[5] Patterson, H. D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*. Vol. 58. 545-554.

- [6] Whittaker, J.(1993). *Graphical models in applied multivariate statistic*. John Wiley & Sons. New York.

[ 1998년 3월 접수, 1998년 7월 최종수정 ]

## Modelling Heterogeneity in Fertility for Analysis of Variety Trials \*

Seongchul Yun<sup>1)</sup> Weechang Kang<sup>2)</sup> Youngjo Lee<sup>3)</sup> Yong Bin Lim<sup>4)</sup>

### ABSTRACT

In agricultural field experiments, the completely randomized block design is often used for the analysis of variety trials. An important assumption is that every experimental unit in each block has the same fertility. But, in most agricultural field experiments there often exists a systematic heterogeneity in fertility among the experimental units. To account for the heterogeneity, we propose to use the hierarchical generalized linear models. We compare our analysis of the data from Scottish Agricultural Colleges list with that using Markov chain Monte Carlo method.

---

\* The present study was supported by the Basic Science Research Institute Program, Ministry of Education, 1997, Project No. BSRI-97-1415.

1) Department of Statistics, Seoul National University, Seoul, Korea.

2) Department of Statistics, Seoul National University, Seoul, Korea.

3) Associate Professor, Department of Statistics, Seoul National University, Seoul, Korea.

4) Professor, Department of Statistics, Ewha Womans University, Seoul, Korea.