

Bayesian structural equation modeling for analysis of climate effect on whole crop barley yield

Moonju Kim^a · Minhee Jeon^a · Kyung-Il Sung^b · Young-Ju Kim^{a,1}

^aDepartment of Statistics, Kangwon National University;

^bDepartment of Feed Science and Technology, Kangwon National University

(Received November 12, 2015; Revised January 5, 2016; Accepted January 28, 2016)

Abstract

Whole Crop Barley (WCB) is a representative self-sufficient winter annual forage crop, along with Italian Ryegrass (IRG), in Korea. In this study, we examined the path relationship between WCB yield and climate factors such as temperature, precipitation, and sunshine duration using a structural equation model. A Bayesian approach was considered to overcome the limitations of the small WCB sample size. As prior distribution of parameters in Bayesian method, standard normal distribution, the posterior result of structural equation model for WCB, and the posterior result of structural equation model for IRG (which is the most popular winter crop) were used. Also, Heywood case correction in prior distribution was considered to obtain the posterior distribution of parameters; in addition, the best prior to fit the characteristics of winter crops was identified. In our analysis, we found that the best prior was set by using the results of a structural equation model to IRG with Heywood case correction. This result can provide an alternative for research on forage crops that have hard to collect sample data.

Keywords: whole crop barley, structural equation model, Bayesian approach, Heywood cases, longitudinal data

1. 서론

청보리(Whole Crop Barley; WCB)는 이탈리아 라이그라스(Italian Ryegrass; IRG)와 함께 우리나라에서 재배되는 대표적인 동계작물로서 벼와 작부체계가 좋고, 종자가 국내에서 개발되어 자급자족되는 장점이 있다. 반면에 내한성과 내습성이 약하고 재배지역이 남부에 국한되는 단점이 있다 (Shin 등, 2012; Kim 등, 2010). 국내에서 소와 양 등의 반추기축을 위한 연구는 미흡한 실정이며, 특히 풀사료의 가격절정과 고품질을 위해 풀사료로 쓰이는 동계작물의 생산성에 영향을 미치는 요인에 대한 연구는 해결이 시급한 과제이다. 일반적으로 동계작물의 생산량에 영향을 미치는 변수로 기후, 토양, 품종, 재배기술 등이 알려져 있으며 특히 기온이 높고 건조한 환경에서 강수는 작물의 생산량에 미치는 영향

This study is supported by 2015 Research Grant from Kangwon National University (No.520150416) and by Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ009766) from Rural Development Administration, Republic of Korea.

¹Corresponding author: Department of Statistics, Kangwon National University, 1 Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Korea. E-mail: ykim7stat@kangwon.ac.kr

이 크다고 보고되었다 (Schlenker와 Roberts, 2009; Descheemaeker 등, 2014). 또한 폴사료 생산성과 기후에 대한 다양한 연구결과가 발표되었는데, 먼저 변수탐색을 위한 연구로 Peng 등 (2013)과 Lee 등 (2013)은 옥수수의 생산량과 관련된 기후 변수를 이용하여 평균온도보다 적산온도(accumulative temperature)가 생산량과 밀접하고 강수일수와 일조시간이 중요한 변수임을 확인하였다. Sung 등 (2014)은 IRG 건물수량을 예측을 위해 기온, 강수 및 일조시간을 이용한 회귀분석을 수행하였다. 그리고 Kim 등 (2014a)은 토양요인을 추가하여 이탈리아 라이그라스 건물수량 예측모형을 분석하여 기후가 토양보다 건물수량에 미치는 영향이 더 크다고 보고하였다. 그 외에도 지리적 정보를 이용하여 강원도의 지역에 따른 최적 재배지를 알아보는 연구도 진행되었다 (Kim 등, 2013).

구조방정식을 이용하여 기후요인과 생산요인간의 인과관계를 파악한 최근 연구 (Kim 등, 2014b)에서는 IRG 생산량에 대하여 가을기온의 간접적 영향과 봄기온, 봄강수 및 최한월기온의 직접적 영향이 있음을 확인하였다. 그러나 일반적으로 구조방정식모형은 표본의 크기가 충분히 큰 (600 이상) 경우 편의(bias)가 없는 해석이 가능하지만, 표본의 크기가 작은(300 미만) 경우 다변량 정규성가정을 만족시키지 못함에 대한 해석에 편의를 배제할 수 없다고 알려져 있다 (Gao 등, 2008). 이러한 표본 수의 한계를 극복하기 위해 Kim 등 (2015)은 추가조사를 통하여 갱신된 대표본의 IRG 자료에 구조방정식모형을 적용하여 편의를 배제한 해석이 가능하였다. Kim 등 (2015)은 IRG 자료와 더불어 WCB 자료분석도 수행하였는데, WCB의 경우 표본의 크기가 작아 최한월과 강수에 대한 경로가 폴사료의 생산에 부합하지 않은 결과가 발생하였다.

본 연구에서는 WCB 자료의 소표본 한계를 해결하기 위하여 베이지안 접근법(Bayesian approach)을 고려하였다. 베이지안 방법은 다양한 사전정보의 사용이 가능할 뿐만 아니라 대표본 근사이론에 크게 의존하지 않아 표본의 크기가 작은 경우에도 믿을만한 결과를 제공하고, 계산된 사후분포를 추후 갱신 자료의 사전분포로 사용 가능한 장점이 있다 (Dunson 등, 2005). 또한 베이지안 접근법은 로버스트 방법과 더불어 다변량 정규성가정과 등분산성 가정의 한계를 극복하기 위한 방법 중 하나로 알려져 있다 (Mullen 등, 1995; Gao 등, 2008; Lim과 Melville, 2009; Kazama 등, 2010). 그러므로 본 연구는 베이지안 방법을 이용하여 최적의 사전정보를 탐색하고, WCB에 대한 구조방정식모형을 통해 WCB 생산량에 기후가 미치는 인과관계를 파악하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 자료의 수집 및 가공과정을 설명하고 베이지안 구조방정식모형에 대하여 설명하였다. 3장은 실제 자료를 이용하여 표준정규분포, IRG와 WCB에 대한 구조방정식모형의 빈도론적 결과를 사전분포로 사용한 결과를 비교하였다. 마지막으로 4장은 결론 및 연구의 한계와 향후 연구 방향에 대해 논하였다.

2. 재료 및 방법

2.1. 자료의 설명

본 연구의 대상인 WCB는 농협중앙회의 목초 및 사료작물품종 수입적응성시험 심의 관련 자료, 동계/하계작물 신품종개발 공동연구보고서와 직무육성신품종선정위원회의 자료로부터 수집되었다. 중복 또는 누락된 경우를 정리하여 얻은 최종자료는 1993년부터 2012년까지(20년간) 9개 지역(대구, 수원, 아산, 예산, 익산, 제주, 진주, 천안 및 청원)에서 얻은 33개 품종으로 총 316개체이다. WCB의 생산성에 영향을 줄 수 있는 기후 변수로 기상청으로부터 얻은 일평균기온, 일최고기온, 일최저기온, 일강수량 및 일조시간을 고려하였다. 또한 WCB는 전년도 가을에 파종하여 이듬해 봄에 수확하기 때문에 기후 변수를 정의하는 기준 시기를 월동시점(평균기온이 0°C가 되는 날)으로 정하였다. 즉, 파종일부터 월동시점까지의 기간(가을)과 월동시점부터 수확 일까지의 기간(이듬해 봄)으로 나누었다. 평균온

도는 기간 내 일평균온도의 평균값이고, 누적온도는 기간 내 평균온도의 합이다. 강수는 일별 강수량의 총합이고 강수일수는 강수일의 합이다. 여기서 강수일수는 강수량이 1mm 기준으로 기록된 일수이다. 그 결과 생성된 변수는 Table 3.1과 같다. 생산량을 나타내는 변수는 수확 후 말리지 않고 측정하는 생초수량(Fresh Matter Yield; FMY)과 건조 후 측정하는 건물수량(Dry Matter Yield; DMY)이다. 가을에 측정된 변수로 파종일부터 월동시점까지 기간을 측정한 가을생육기간(Autumn Growing Day; AGD)과 동일 기간의 누적온도를 측정한 가을누적온도(Autumn Accumulate Temperature; AAT)이다. 이듬해 봄에 측정된 변수는 이듬해 1월1일 이후의 최초 월동시점부터 수확일까지의 기간을 측정한 봄생육기간(Spring Growing Day; SGD), 동일 기간에 측정된 누적온도인 봄누적온도(Spring Accumulate Temperature; SAT), 최한월(1월)의 평균온도를 측정한 최한월평균온도(Mean Temperature in January; MTJ), 0°C 이상이 되는 날부터 온도의 합이 150이 되는 날까지 걸린 기간인 150도달기간(Period to Degree 150; PD150), 1월 1일부터 수확까지의 일조시간을 측정한 봄일조시간(Spring Sunshine Time; SST), 동일기간의 강수량인 봄강수량(Spring Rain Fall; SRF), 그리고 강수량이 1mm 이상 기록된 일수의 합으로 봄강수일수(Spring Rainfall Day; SRD)이다.

2.2. 베이지안 구조방정식

구조방정식 모형은 변수들을 잠재변수(요인)로 축소하는 측정부분과 잠재변수 간의 경로를 해석하는 구조부분으로 이루어진다 (Arbuckle, 2005). 측정부분의 방정식은 다음과 같다.

$$\begin{aligned} y &= \nu_y + \Lambda_y \eta + \epsilon^y, \\ x &= \nu_x + \Lambda_x \xi + D + \delta^x. \end{aligned}$$

여기서 $y = (y_1, y_2)'$, $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9)'$, ν_y 와 ν_x 는 측정방정식의 절편벡터이다. Λ_y 와 Λ_x 는 계수행렬, η 는 반응변수로부터 축소된 요인벡터이고 ξ 는 설명변수로부터 축소된 요인벡터이다. ϵ^y 와 δ^x 은 각각 오차벡터이며 D 는 오차간 상관성을 나타내는 잠재변수 행렬이다. 본 연구에서 y_1 는 FMY, y_2 는 DMY, x_1 은 AGD, x_2 은 AAT, x_3 은 SGD, x_4 은 SAT, x_5 은 SRF, x_6 은 SRD, x_7 은 SST, x_8 은 MTJ 및 x_9 은 PD150이다. AGD와 SGD, AAT와 SAT는 가을과 봄에 측정된 같은 변수이므로 오차 간 상관성이 존재할 것으로 가정하고 계절효과를 반영하기 위해 오차 사이를 연결하는 행렬 D 를 측정방정식에 포함하였다. 측정오차벡터 ϵ^y 와 δ^x 는 독립인 정규분포를 따르며, $\epsilon^y \sim N(0, \Sigma_y)$, $\Sigma_y = \text{diag}(\sigma_{y_1}^2, \sigma_{y_2}^2)$, $\delta^x \sim N(0, \Sigma_x)$, $\Sigma_x = \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_9}^2)$ 로 가정하였다. 종단자료 특성을 나타내는 오차간 연결행렬 $D \sim N(0, \Omega_D)$ 로 가정하였다.

요인 간의 경로를 나타내는 구조부분의 방정식은 다음과 같다.

$$\eta = \alpha + B\eta + \Gamma\xi + \tau,$$

여기서 α 는 절편 벡터이고, B 는 구조부분에서 잠재변수 η 간의 연관성을 나타내는 계수행렬로써 대각원소가 모두 0이다. Γ 는 ξ 가 η 에 미치는 영향을 나타내는 계수행렬이고 τ 는 오차벡터이다. 구조방정식 모형의 공통된 가정으로 ξ 와 τ 는 서로 독립인 정규분포를 따르며 $\xi \sim N(\mu_\xi, \Omega_\xi)$, $\Omega_\xi = \text{diag}(\omega_{\xi_1}^2, \dots, \omega_{\xi_n}^2)$, $\tau \sim N(0, \Omega_\tau)$, $\Omega_\tau = \text{diag}(\omega_{\tau_1}^2, \dots, \omega_{\tau_n}^2)$ 로 가정하였다.

잠재요인을 포함한 종단자료에 대한 구조방정식모형의 우도함수는 다음과 같다.

$$\begin{aligned} L(y, x, \eta, \xi | \Theta) &= \prod_{i=1}^n N(y_i | \nu_{y,i} + \Lambda_y \eta_i, \Sigma_y) N(x_i | \nu_{x,i} + \Lambda_x \xi_i + D_i, \Sigma_x) \\ &\quad \times N(\eta_i | \alpha_i + B\eta_i + \Gamma\xi_i, \Omega_\tau) N(\xi_i | \mu_\xi, \Omega_\xi) N(D_i | 0, \Omega_D), \end{aligned}$$

여기서 절편 ν_x 와 ν_y 는 사전분포를 가정하였으나 α 는 0으로 고정하였다. 따라서 모수공간은 $\Theta = (\beta, \gamma, \nu_x, \nu_y, \lambda_x, \lambda_y, \sigma_x^2, \sigma_y^2, \omega_\tau^2, \mu_\xi, \omega_\xi^2, w_D^2)$ 이며, β 는 B 의 원소, γ 는 Γ 의 원소이며 λ_x 와 λ_y 는 각각 Λ_x 와 Λ_y 의 원소이고, w_D^2 는 Ω_D 의 대각원소를 나타낸다.

베이저안 방법은 모수공간을 구성하는 계수와 오차의 사전분포를 고려해야 한다. 본 연구에서 사전분포는 계산이 용이하도록 공액사전분포(conjugate prior distribution)를 이용하였다. 측정방정식에서 절편인 ν_x 와 ν_y , 경로계수인 λ_x 와 λ_y , 경로계수행렬의 원소인 β 와 γ , μ_ξ 의 사전분포로 정규분포를 가정하였고 오차인 σ_x^2 , σ_y^2 , ω_τ^2 , ω_ξ^2 및 w_D^2 는 독립인 역-감마분포를 사전분포로 가정하였다. 정규분포의 초모수인 평균과 분산 값으로 IRG와 WCB 자료의 빈도론적 구조방정식모형 결과나 요인분석 결과의 경로계수의 평균 및 분산 값을 사용하였다. 역-감마분포의 초모수값은 경험적으로 선택하였다.

만약 모수의 수가 많거나 모수간 연관성이 존재하여 결합사후분포로부터 주변분포를 구하는 것이 어려운 경우 주변분포의 계산없이 결합사후분포로부터 표본생성이 가능한 MCMC 알고리즘을 사용한다 (Dunson 등, 2005). MCMC 알고리즘은 마코프 연쇄(Markov chain)에 의해 목표 분포(target distribution)로부터 임의 표본을 생성한 후, 몬테카를로 적분(Monte Carlo integration)을 이용하여 모수를 추정하는 방법이다. 깁스 샘플러(Gibbs sampler)와 메트로폴리스-헤이스팅스 알고리즘(Metropolis-Hastings algorithm)은 마코프 연쇄를 이용하여 MCMC 알고리즘을 수행하는 대표적인 방법이며, 깁스 샘플러(Gibbs sampler)는 각 변수의 조건부사후분포로부터 표본을 반복적으로 생성하는 방법이다. 본 연구는 깁스 샘플러를 사용하였다.

잠재변수와 구조모수의 조건부사후분포를 얻기 위하여 Dunson 등 (2005)의 방법을 이용하였다. 먼저 잠재변수 η , ξ 와 구조모수 β , γ 의 조건부사후분포를 유도하였다. 결합사후분포가 우도함수와 사전분포의 곱에 비례한다는 성질에 근거하여 반응변수의 요인 η 의 조건부사후분포는 다음과 같이 계산하였다.

$$\pi(\eta_i | \nu_y, \Lambda_y, \Sigma_y, \tilde{\mu}_\eta, \tilde{\Omega}_\eta, y_i) \propto \pi(y_i | \nu_y + \Lambda_y \eta_i, \Sigma_y) \pi(\eta_i | \tilde{\mu}_\eta, \tilde{\Omega}_\eta),$$

여기서 $\tilde{\mu}_\eta = A[\Gamma\xi]$, $\tilde{\Omega}_\eta = A\Omega_\tau A'$ 이고, $A = [I - B]^{-1}$ 이다. 따라서 η_i 의 분포는 $N(\hat{\eta}_i, \hat{\Omega}_\eta)$ 인 정규분포를 따르며

$$\begin{aligned} \hat{\eta}_i &= \hat{\Omega}_\eta \left[\Lambda'_y \Sigma_y^{-1} (y_i - \nu_y) + \tilde{\Omega}_\eta^{-1} \tilde{\mu}_\eta \right], \\ \hat{\Omega}_\eta^{-1} &= \Lambda'_y \Sigma_y^{-1} \Lambda_y + \tilde{\Omega}_\eta^{-1} \end{aligned}$$

가 된다.

설명변수의 요인 ξ 의 조건부사후분포는 다음과 같다.

$$\begin{aligned} &\pi(\xi_i | \eta_i, \Omega_\tau, \nu_x, \lambda_x, \Sigma_x, B, \Gamma, \mu_\xi, \Omega_\xi, x_i) \\ &\propto \pi(x_i | \nu_x + \lambda_x \xi_i + D_i, \Sigma_x) \pi(\eta_i | B\eta_i - \Gamma\xi_i \Omega_\tau) \pi(\xi_i | \mu_\xi, \Omega_\xi) \pi(D_i | 0, \Omega_D) \end{aligned}$$

이므로 ξ_i 의 분포는 $N(\hat{\xi}_i, \hat{\Omega}_\xi)$ 인 정규분포를 따르며

$$\begin{aligned} \hat{\xi}_i &= \hat{\Omega}_\xi \left[\Lambda'_x \Sigma_x^{-1} (x_i - \nu_x) + \Gamma' \Omega_\tau^{-1} (\eta_i - B\eta_i) + \Omega_\xi^{-1} \mu_\xi \right], \\ \hat{\Omega}_\xi^{-1} &= \Lambda'_x \Sigma_x^{-1} \Lambda_x + \Gamma' \Omega_\tau^{-1} \Gamma + \Omega_\xi^{-1} + \Omega_D^{-1} \end{aligned}$$

로 얻어진다.

요인계수 간 영향력을 측정하는 경로계수인 β_{rj} ($r, j = 1, \dots, m$)의 조건부사후분포와 설명변수의 요인 ξ_i 가 반응변수의 요인 η_i 에 미치는 효과를 나타내는 γ_{rj} ($r = 1, \dots, m, j = 1, \dots, n$)의 조건부사후분포

는 Dunson 등 (2005)에서 유도된 결과와 동일하게 나타났다. β_{rj} ($r, j = 1, \dots, m$)의 조건부사후분포는 $N(\hat{\beta}_{rj}, \hat{\omega}_{\beta}^2)$ 인 정규분포를 따르며

$$\hat{\beta}_{rj} = \hat{\omega}_{\beta} \left[\frac{\mu_{\beta}}{\omega_{\beta}^2} + \sum_{i=1}^n \frac{\eta_i^j}{\omega_{\xi r}^2} \left(\eta_i^r - \sum_{s=1}^n \gamma_{rs} \xi_i^s - \sum_{t=1, t \neq j}^m \beta_{rt} \eta_i^t \right) \right],$$

$$\hat{\omega}_{\beta}^{-1} = \frac{\sum_{i=1}^n (\eta_i^j)^2}{\omega_{\tau r}^2} + \frac{1}{\omega_{\beta}^2}$$

이 되고, γ_{rj} ($r = 1, \dots, m, j = 1, \dots, n$)의 조건부사후분포는 $N(\hat{\gamma}_{rj}, \hat{\omega}_{\gamma}^2)$ 인 정규분포를 따르며

$$\hat{\gamma}_{rj} = \hat{\omega}_{\gamma} \left[\frac{\mu_{\gamma}}{\omega_{\gamma}^2} + \sum_{i=1}^n \frac{\xi_i^j}{\omega_{\tau r}^2} \left(\eta_i^r - \sum_{s=1}^m (\beta_{rs} \eta_i^s) - \sum_{t=1, t \neq j}^n \gamma_{rt} \xi_i^t \right) \right],$$

$$\hat{\omega}_{\gamma}^{-1} = \frac{\sum_{i=1}^n (\xi_i^j)^2}{\omega_{\tau r}^2} + \frac{1}{\omega_{\gamma}^2}.$$

위와 같이 모수들의 조건부사후분포를 계산한 후 깃스샘플러 알고리즘을 이용하여 표본을 생성하였다.

$$\beta^{(j+1)} \sim \pi \left(\beta | y, \gamma^{(j)}, \nu_x^{(j)}, \dots, \omega_{\xi}^{2(j)}, D^{(j)} \right),$$

$$\gamma^{(j+1)} \sim \pi \left(\gamma | y, \beta^{(j+1)}, \nu_x^{(j)}, \dots, \omega_{\xi}^{2(j)}, D^{(j)} \right),$$

$$\nu_x^{(j+1)} \sim \pi \left(\nu_x | y, \beta^{(j+1)}, \gamma^{(j+1)}, \dots, \omega_{\xi}^{2(j)}, D^{(j)} \right),$$

$$D^{(j+1)} \sim \pi \left(D | y, \beta^{(j+1)}, \gamma^{(j+1)}, \dots, \omega_{\xi}^{2(j+1)} \right).$$

이러한 과정을 50,000번 반복하여 얻은 50,000개의 깃스표본을 이용하여 각 경로계수 및 오차의 주변분포를 구하였다.

2.3. 분석방법

먼저 구조방정식모형의 빈도론적 접근을 위해 결측과 이상점을 처리하였다. 여기서 이상점은 Kim 등 (2014b)이 제안한 마할라노비스 거리에 대한 확률의 상자그림을 이용하였다. 그 결과 결측 16개체, 이상점 7개체를 제거하여 최종분석에 사용된 WCB 표본은 293개체이다. 베이지안 접근을 위한 사전분포로 표준정규분포, WCB와 IRG의 구조방정식모형의 빈도론적 결과를 고려하였다. 여기서 WCB를 분석하기 위한 사전분포로 IRG의 분석결과를 고려한 이유는 IRG가 우리나라에서 가장 많이 재배되고 있어 작물의 대표성이 우수한 반면, WCB는 재배지역이 주로 남부에 국한되어 표본을 추가로 조사하는 것이 시간과 비용문제로 어렵기 때문이다. 또한 선행연구에서 분석한 IRG 자료는 상대적으로 대표본($n = 658$)이고 구조방정식모형 분석결과가 생육특성과 잘 부합하였기 때문에 IRG의 정보를 WCB 베이지안 분석에 반영하고자 하였다.

Kim 등 (2015)이 제시한 IRG의 구조방정식 분석결과에서 음의 분산추정량이 계산되는 헤이우드 케이스(Heywood Cases; HC)가 발생하였다. HC가 발생하는 원인으로 소표본이거나 구조적 오지정(misspecification)이 있는 경우 (Kolenikov와 Bollen, 2012), 이상점과 영향치가 있는 경우 (Bollen, 1987), 경험적 과소식별(under-identification)인 경우 (Rindskopf, 1984), 결측이 있는 경우 (Wothke,

Table 3.1. Variable characteristics of data

Variable (unit)	Symbol	Min	Max	Mean	S.D.	Quartile		
						1st	2nd	3rd
FMY(kg/ha)	y1	11820	56710	31289.73	8446.75	24525	30925	35785
DMY(kg/ha)	y2	4490	18580	11250.64	2250.19	9820	11400	12595
AGD(day)	x1	18	71	46.45	11.75	40	47	55
AAT(°C)	x2	184.1	788	468.56	142.66	376.7	454.1	545.6
SGD(day)	x3	51	124	93.14	15.45	82.5	92	105
SAT(°C)	x4	497	1547.2	1178.16	206.92	1089.5	1183.5	1313.5
SRF(mm)	x5	86.1	736.8	324.29	139.4	224	309.9	375.5
SRD(day)	x6	34	68	49.4	8.19	44	50	53
MTJ(°C)	x7	-7.3	5.5	-1.46	3.16	-3.1	-1.1	0.1
SST(hr)	x8	518.1	1028.7	752.81	116.87	689.3	755.55	825.2
PD150(day)	x9	30	90	66.1	16.08	58	71	76

FMY = fresh matter yield of crops, DMY = dry matter yield of crops, AGD = growing days in autumn, AAT = accumulated temperature in autumn, SGD = growing days in next spring, SAT = accumulated temperature in next spring, SRD = raining days in next spring, SRF = amount of rain in next spring, MTJ = mean temperature in January, SST = time of shining in next spring, PD150 = period to reach at accumulated temperature 150.

1993) 및 표본추출의 변동 (Boomsma, 1983) 등으로 보고되고 있다. 본 연구에서 HC가 발생한 원인은 복합적으로 생각할 수 있는데 소표본 문제와 하나의 요인에 변수가 2개인 요인이 많아 발생하는 구조적 문제에 의한 것으로 추측하였다. HC 문제를 해결하는 일반적인 방법으로 HC가 발생한 오차에 대해 작은 양의 값으로 대체한다 (Arbuckle, 2005). 본 연구의 빈도론적 구조방정식모형에서는 오차에 0.005값을 직접 입력하여 HC를 수정하였다. 베이지안 구조방정식 모형에서는 HC를 수정하지 않은 빈도론적 구조방정식 결과와 HC를 수정한 빈도론적 구조방정식 결과를 이용하여 각각 사전분포의 초모수로 설정하여 HC가 개선되는지를 확인하였다. 그러므로 본 연구에 사용된 사전분포는 표준정규분포, HC를 수정하지 않은 WCB와 IRG의 빈도론적 구조방정식 결과 및 HC를 수정한 WCB와 IRG의 빈도론적 구조방정식 결과를 포함하여 모두 5가지를 사용하였다. 각각의 사전분포에 대한 베이지안 분석방법 결과를 비교하기 위해 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion) 및 DIC(Deviance Information Criterion)를 계산하였다. 알고리즘의 반복 수는 변동이 상대적으로 큰 초기 5,000번을 제외시킨 후 50,000번 반복하였다. 또한 수렴과정이 초기값의 영향을 받는지 확인하기 위해 -10.0, 0.1 및 10.0의 3가지 초기값을 사용하였다. 구조방정식의 빈도론적 적합은 AMOS 21.0을 사용하였고, 베이지안 구조방정식은 WinBUGS14를 사용하여 분석하였다.

3. 분석결과

WCB 자료의 기초분석 결과는 Table 3.1과 같다. 수확 후 말리지 않은 수량인 FMY의 평균은 31289.73(kg/ha), DMY의 평균은 11250.64로 FMY의 평균이 약 3배 정도 높게 나타났다. 즉, 건물률(percentage of dry matter)이 약 30%임을 의미한다. FMY와 DMY 모두 평균과 중위수가 비슷하게 나타났고, 중위수로부터 제 1사분위수와 제 3사분위수의 차이가 비슷하게 나타났다.

3.1. 구조방정식모형의 빈도론적 결과

Kim 등 (2015)은 IRG와 WCB의 생산량과 기후요인간 인과관계를 분석하기 위하여 가을기온과 봄기

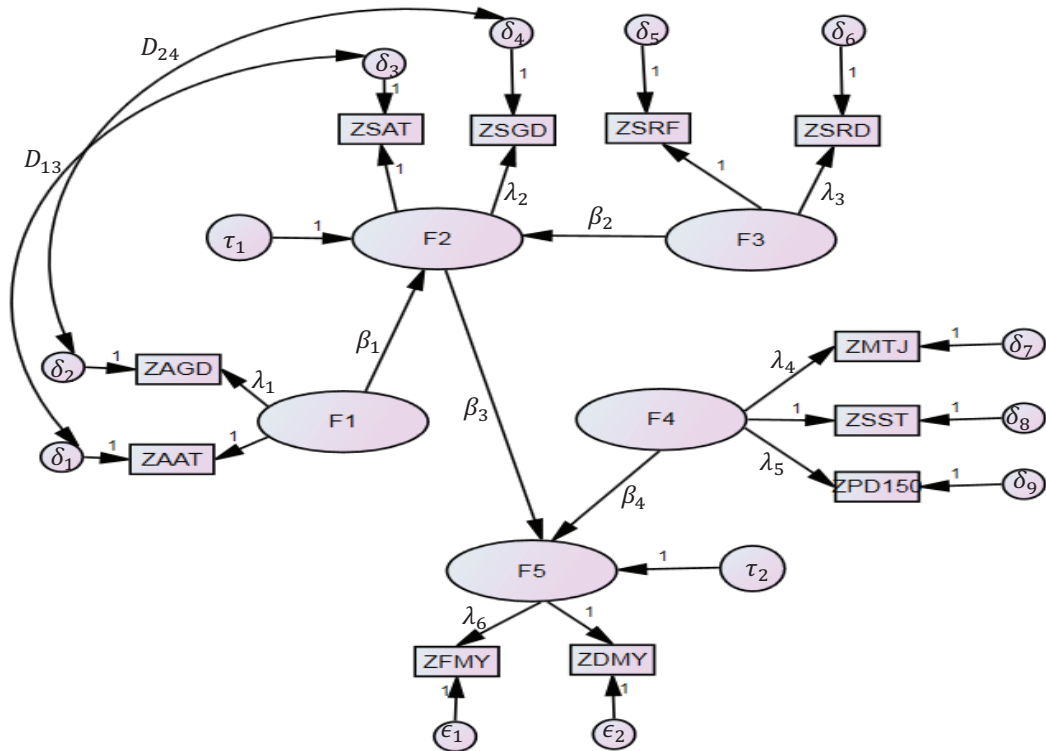


Figure 3.1. Path diagram of longitudinal structural equation model in WCB.

은 사이에 계절효과를 나타내는 종단자료 구조방정식모형을 구축하였다. 본 연구는 Kim 등 (2015)의 모형으로부터 해석이 어려운 교호효과를 제거하여 Figure 3.1과 같은 모형을 얻었다. 여기서 F1은 가을기온으로 AGD와 AAT로, F2는 봄기온으로 SGD와 SAT, F3는 봄강수로 SRF와 SRD, F4는 최한월기온으로 SST, MTJ 및 PD150로, 그리고 F5는 생산량으로 FMY와 DMY로 이루어졌다. AGD와 SGD, AAT와 SAT는 가을과 봄에 측정된 같은 변수이므로 오차가 독립이 되지 않는다. 따라서 계절 효과를 반영하기 위해 오차 사이를 연결하였다. WCB에 대한 모형의 적합도는 GFI(Goodness of Fit Index)가 0.665, CFI(Comparative Fit Index)가 0.597, NFI(Normed Fit Index)가 0.590로 나타났고, IRG에 대한 모형의 적합도는 각각 0.721, 0.702 및 0.698로 나타났다. 그러므로 종단자료 구조방정식 모형은 WCB자료보다 IRG자료에 잘 적합하였다.

WCB자료에 대한 베이지안 구조방정식모형의 사전분포를 얻기 위해 WCB와 IRG에 각각 종단자료 구조방정식모형을 적용한 빈도론적 결과는 Table 3.2와 같다. Table 3.2는 HC의 수정없이 WCB와 IRG의 빈도론적 종단자료구조방정식을 적용한 결과(각각 MLE1과 MLE2)와 HC를 수정한 WCB와 IRG의 종단자료구조방정식 결과(각각 MLE3와 MLE4)를 정리하였다. 요인간 경로계수를 나타내는 β의 부호가 모두 양수이므로 각 요인의 증가는 생산량의 증가를 의미한다. 하지만 F4(최한월기온)과 PD150(150도달기간)의 경로인 λ5는 모두 음수로 나타났다. 즉 누적온도 150에 도달하는 기간이 짧을수록, 봄이 일찍 올수록 생산량이 증가하는 것을 의미한다. 또한 WCB와 IRG는 모두 음의 분산 추정량을 갖는 HC가 발생하였다. WCB는 3개의 오차에서 HC가 발생하였다. 특히 δ5은 MLE1에서

Table 3.2. The result of maximum likelihood method in WCB and IRG(from left, MLE1: not modified Heywood case in WCB, MLE2: not modified Heywood case in IRG, MLE3: modified Heywood case in WCB, and MLE4: modified Heywood case in IRG)

Parameter	MLE1		MLE2		MLE3		MLE4	
	Mean	S.D.	Mean	S.D.	mean	S.D.	mean	S.D.
β_1	0.423	0.057	0.099	0.035	0.462	0.979	0.133	0.744
β_2	0.019	0.317	0.727	0.052	0.245	0.807	0.734	1.308
β_3	0.332	0.055	0.457	0.036	0.472	1.080	0.469	0.923
β_4	0.448	0.094	0.925	0.247	2.302	13.586	0.907	6.182
λ_1	1.103	0.135	0.665	0.141	1.128	2.318	0.921	0.436
λ_2	1.027	0.076	1.064	0.022	1.003	1.356	1.060	0.410
λ_3	0.011	0.178	0.806	0.054	0.088	0.996	0.814	1.411
λ_4	9.825	4.890	4.726	1.002	5.546	30.624	4.745	25.779
λ_5	-2.788	0.473	-4.285	0.873	-4.933	27.360	-4.253	22.111
λ_6	0.144	0.131	1.023	0.051	0.824	0.584	1.035	1.283
δ_1	0.207	0.094	-0.366	0.283	0.227	1.545	0.005	
δ_2	0.072	0.114	0.403	0.126	0.046	1.940	0.174	0.256
δ_3	0.205	0.050	0.116	0.014	0.185	0.927	0.114	0.179
δ_4	0.150	0.052	-0.002	0.015	0.173	0.944	0.005	
δ_5	-13.026	233.156	0.247	0.043	0.005		0.254	1.103
δ_6	0.988	0.086	0.510	0.039	0.982	1.391	0.505	1.000
δ_7	0.734	0.111	0.097	0.077	0.209	0.309	0.087	2.000
δ_8	0.991	0.082	0.958	0.054	0.967	1.374	0.958	1.385
δ_9	-2.161	1.171	0.258	0.065	0.005		0.266	1.642
ϵ_1	0.846	0.119	0.149	0.034	0.342	0.498	0.139	0.847
ϵ_2	-4.311	4.683	0.186	0.033	0.005		0.196	0.820
τ_1	5.192	4.675	0.526	0.043	0.715	1.048	0.509	1.077
τ_2	0.596	0.111	0.452	0.036	0.537	1.133	0.441	0.897
D_{13}	-0.028	0.036	0.002	0.013	0.024	0.635	0.003	0.256
D_{24}	0.103	0.034	-0.02	0.012	0.098	0.618	0.019	0.230

WCB = whole crop barley, IRG = Italian Ryegrass.

-13.026로 추정되었다. 이것은 다른 HC가 발생한 오차보다 매우 큰 음수이고 표준편차도 233.156으로 주의가 필요한 오차로 판단하였다. MLE2에서는 2개의 오차(δ_1 과 δ_4)에서 HC가 발생하였는데 각각 -0.366과 -0.002로 나타났다. 이러한 음의 분산추정량의 값을 0.005로 대체한 사전분포가 MLE3와 MLE4이다. F2(봄기온)에서 F5(생산량)로 이어진 경로(β_3), 가을생육기간(AGD)에서 F1으로 이어진 경로(λ_1) 및 봄생육기간(SGD)에서 F2로 이어진 경로(λ_2)가 모든 빈도론적 접근 결과에서 비슷하게 나타났다. 또한 WCB의 결과 중 HC 수정 후의 MTJ(최한월평균온도)에서 F4(최한월기온)로 이어진 경로(λ_4), PD150(150도달기간)에서 F4로 이어진 경로(λ_5) 및 FY(생초수량)에서 F5(생산량)로 이어진 경로(λ_6)가 각각 5.546, -4.933 및 0.824로 IRG의 결과인 4.745, -4.253 및 1.035와 비슷하게 나타났다. 하지만 WCB의 경우 표준크기가 작기 때문에 빈도론적 결과를 신뢰하기 어려운 한계가 있다.

3.2. 베이지안 구조방정식모형 결과

WCB자료를 분석하기 위해 베이지안 구조방정식 방법을 사용한 결과는 Table 3.3과 같다. 여기서 표준 정규분포, MLE1, MLE2, MLE3 및 MLE4를 사전분포로 사용한 결과가 각각 Posterior1, Posterior2,

Table 3.3. The result of posterior distribution by Bayesian approach in WCB and IRG with different priors(from left, Posterior1: standard normal, Posterior2: not modified Heywood case in WCB, Posterior3: not modified Heywood case in IRG, Posterior4: modified Heywood case in WCB, and Posterior5: modified Heywood case in IRG)

Parameter	Posterior1		Posterior2		Posterior3		Posterior4		Posterior5	
	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.	mean	S.D.
β_1	0.048	0.073	1.231	0.163	1.548	0.582	0.741	0.113	0.598	0.086
β_2	0.257	0.078	-0.269	0.052	0.616	0.143	0.296	0.111	0.345	0.143
β_3	0.410	0.082	0.222	0.069	0.082	0.070	2.158	0.072	0.084	0.069
β_4	-0.600	0.105	-0.869	0.348	0.417	0.144	-0.112	0.069	0.382	0.147
λ_1	0.974	0.037	2.273	0.103	1.305	0.283	1.493	0.063	1.159	0.047
λ_2	0.897	0.039	0.903	0.039	0.928	0.041	0.901	0.039	0.899	0.040
λ_3	0.099	0.071	-0.065	0.056	0.124	0.155	0.131	0.112	0.209	0.168
λ_4	1.217	0.061	9.822	0.008	0.465	0.038	5.505	0.032	4.672	0.038
λ_5	-1.170	0.063	-3.004	0.082	-0.426	0.043	-4.930	0.036	-4.234	0.043
λ_6	0.893	0.041	0.889	0.042	0.924	0.044	0.932	0.046	0.924	0.044
δ_1	-0.046	0.578	0.113	0.383	-0.365	0.138	0.068	0.589	0.005	
δ_2	-0.104	0.537	0.116	0.331	0.396	0.311	0.082	0.485	0.412	1.424
δ_3	-0.112	0.695	-0.205	0.675	-1.021	1.579	-0.078	0.895	0.330	1.460
δ_3	-0.184	0.727	-0.617	0.650	-1.000	1.575	-0.053	0.883	0.005	
δ_4	0.002	0.065	-13.03	0.017	-0.211	0.664	0.005		-0.206	0.667
δ_5	0.004	0.057	0.965	0.448	0.199	0.700	0.586	0.581	0.164	0.702
δ_6	0.009	0.075	0.783	0.338	0.053	0.460	0.025	1.098	0.044	0.455
δ_7	-0.006	0.092	0.725	0.389	0.633	0.592	0.632	0.589	0.607	0.581
δ_8	0.009	0.074	-2.156	0.033	0.217	0.524	0.005		0.223	0.554
ϵ_1	0.047	0.505	0.425	0.033	-0.119	0.885	-0.123	1.049	-0.056	0.879
ϵ_2	0.060	0.565	-4.311	0.300	-0.115	0.903	0.005		-0.044	0.918
τ_1	0.084	0.649	5.192	0.008	0.282	0.839	0.393	0.722	0.529	0.817
τ_2	-0.099	0.649	0.912	0.323	0.116	0.836	0.399	0.751	0.090	0.809
D_{13}	0.048	0.576	-1.584	0.560	-1.457	0.921	-0.808	0.765	-1.573	0.826
D_{24}	0.102	0.535	-1.052	0.647	-1.667	1.128	-0.976	0.784	-1.421	1.139

WCB = whole crop barley, IRG = Italian Ryegrass.

Posterior3, Posterior4 및 Posterior5이다. 먼저 오차를 살펴보면, Posterior1은 6개의 HC가 발생하였다. Posterior2는 사전분포(MLE1)와 사후분포에서 각각 3개와 5개의 HC가 발생하였고, 특히 δ_5 의 추정값은 -13.026에서 -13.030로 개선되지 않았다. Posterior3는 사전분포(MLE2)와 사후분포에서 각각 2개와 6개의 HC가 발생하였다. 그러므로 HC가 존재하는 사전분포를 사용하는 경우는 WCB와 IRG 모두 HC가 개선되지 않았다. 반면에 HC를 수정한 WCB와 IRG의 사전분포(MLE3과 MLE4)를 사용한 Posterior4와 Posterior5는 모두 3개의 HC가 발생하였다. 즉 사전분포에서 HC를 수정하면 사후분포에서 HC의 수가 줄어드는 것을 확인하였다. 여전히 존재하는 HC의 경우 하나의 요인에 축소되는 변수의 수가 너무 적다는 구조적 문제에 기인한 것으로 생각되며, HC가 완전하게 제거되기 위해서는 추후 연구에서 변수를 추가적으로 생성해야 할 것으로 사료된다.

각 모형의 생산량과의 경로 해석을 비교하면, F1(가을기온)에서 F2(봄기온)로 이어진 경로(β_1)는 Posterior2와 Posterior3에서 각각 1.231과 1.548로 비슷하고, Posterior4와 Posterior5에서 각각 .741과 .598로 비슷하게 나타났다. 이것은 IRG와 WCB의 초종에 의한 차이보다는 HC 수정에 의한 차이가 크게 나타난 결과이다. 또한 F3(봄강수)에서 F2(봄기온)로 이어진 경로(β_2)는 Posterior2, Posterior3,

Table 3.4. Fitness of each Bayesian approach

	Not modified Heywood case			Modified Heywood case	
	Posterior1	Posterior2	Posterior3	Posterior4	Posterior5
AIC	3283.333	5903.940	4921.170	3693.770	3691.690
BIC	3319.628	5940.238	4957.468	3730.068	3727.988
DIC	5916.830	7030.530	6813.090	6127.250	6054.870

AIC = Akaike information criterion, BIC = Bayesian information criterion, DIC = deviance information criterion.

Posterior4 및 Posterior5에서 각각 -0.269 , $.616$, $.296$ 및 $.345$ 로 나타났다. 특히 Posterior2에서 음수로 나타난 것은 봄강수의 증가가 봄기온의 감소에 영향을 미치고 이것은 생산량 감소로 이어지는 것을 나타내므로 동계작물의 특성에 부합하지 않는다. 왜냐하면 동계작물의 생장에 영향을 미치는 강수량은 대부분 강설량이며 1월부터 강설량이 많을수록 지반의 온도가 높아져 3월 이후 작물의 생장이 활발해지기 때문이다. F2(봄기온)에서 F5(생산량)로 이어진 경로(β_3)는 각각 $.222$, $.082$, 2.158 및 $.084$ 로 나타났다. 이것은 HC의 수정 전과 후의 차이가 WCB는 크지만 IRG는 작은 것을 의미한다. 마지막으로 F4(최한월기온)에서 F5(생산량)로 이어진 경로(β_4) 각각 $-.869$, $.417$, $-.112$ 및 $.382$ 로 나타났다. 이것은 WCB의 사전정보를 사용한 경우와 IRG의 사전정보를 사용한 경우에서 최한월기온이 생산량에 미치는 영향이 반대로 나타나는 결과이다. 일반적으로 생산량은 최한월 평균온도(MTJ)와 1월 1일부터 일조시간(SST)이 증가할수록, 150도달기간(PD150)은 감소할수록 증가하는 특성을 갖는다. 이러한 해석이 가능한 것은 β_4 의 부호가 양수인 Posterior3와 Posterior5이다. 대체적으로 IRG 분석에서는 HC 수정 전과 후에서 요인 사이의 경로들 차이가 작은 반면 WCB 분석에서는 차이가 크게 나타났다.

요인간의 경로 λ 를 살펴보면 대체적으로 사전분포에 관계없이 비슷하게 나타났다. 하지만 SRD(봄강수일수)에서 F3(봄강수)로 이어진 경로(λ_3)는 Posterior2에서 $-.065$ 로 다른 결과들($.124$, $.131$ 및 $.209$)과 차이를 보였다. 이것은 강수일수와 강수량이 강수요인에 미치는 영향이 다르게 해석되는 것을 의미한다. 일반적으로 강수일수가 많을수록 강수량이 많은데 이러한 해석이 가능한 것은 Posterior3, Posterior4 및 Posterior5이다. 또한 MTJ(최한월평균온도)에서 F4(최한월기온)로 이어진 경로(λ_4)는 Posterior2와 Posterior3에서 각각 9.822 와 $.465$ 로 차이가 컸지만 Posterior4와 Posterior5에서는 각각 5.505 와 4.672 로 결과가 비슷하게 나타났다. 따라서, HC를 수정한 사전분포를 사용한 경우 WCB와 IRG 결과의 해석이 유사하고, WCB 결과보다 IRG 결과를 사전분포로 사용하는 것이 사료에 대하여 알려진 이론과 부합하는 것으로 판단되었다.

WCB에 대한 빈도론적 결과인 MLE3와 IRG의 사전정보를 사용하여 WCB 자료를 분석한 Posterior5를 비교하면 베이저안 방법의 효과를 쉽게 알 수 있었다. 예를 들어 최한월기온(F4)이 생산량(F5)에 영향을 미치는 경로(β_4)는 2.302 에서 $.382$ 로 바뀌었다. 이것은 타 요인에 비해 과대 측정되었던 영향이 IRG의 사전정보가 반영된 후 풀사료 생육특성에 정상적인 범위내의 해석이 가능하도록 바뀐 것으로 판단하였다. 또한 Posterior5는 MLE3에 비해 모든 경로의 표준편차가 작게 나타났다.

베이저안 방법들을 비교하기 위하여 적합도를 계산한 결과는 Table 3.4와 같다. AIC, BIC 및 DIC 값은 Posterior1에서 가장 작고 Posterior5, Posterior4, Posterior3 및 Posterior2 순으로 나타났다. 즉 HC를 수정한 사전분포를 사용한 모형들(Posterior4와 Posterior5)이 HC를 수정하지 않은 사전분포를 사용한 모형들(Posterior2와 Posterior3)보다 적합도가 모두 좋게 나타났다. 특히 이탈도(deviance)를 이용하는 DIC보다 베이지점수(Bayes factor)로 근사하는 BIC에서 차이가 크게 나타났다. 그리고 WCB와 IRG의 빈도론적 구조방정식 결과를 사전분포로 사용한 모형을 비교하면 HC 수정여부에 관계없이 IRG의 결과를 사전분포로 사용한 모형이 더 좋은 적합을 보였다. 그러므로 기후요인과 생산량의

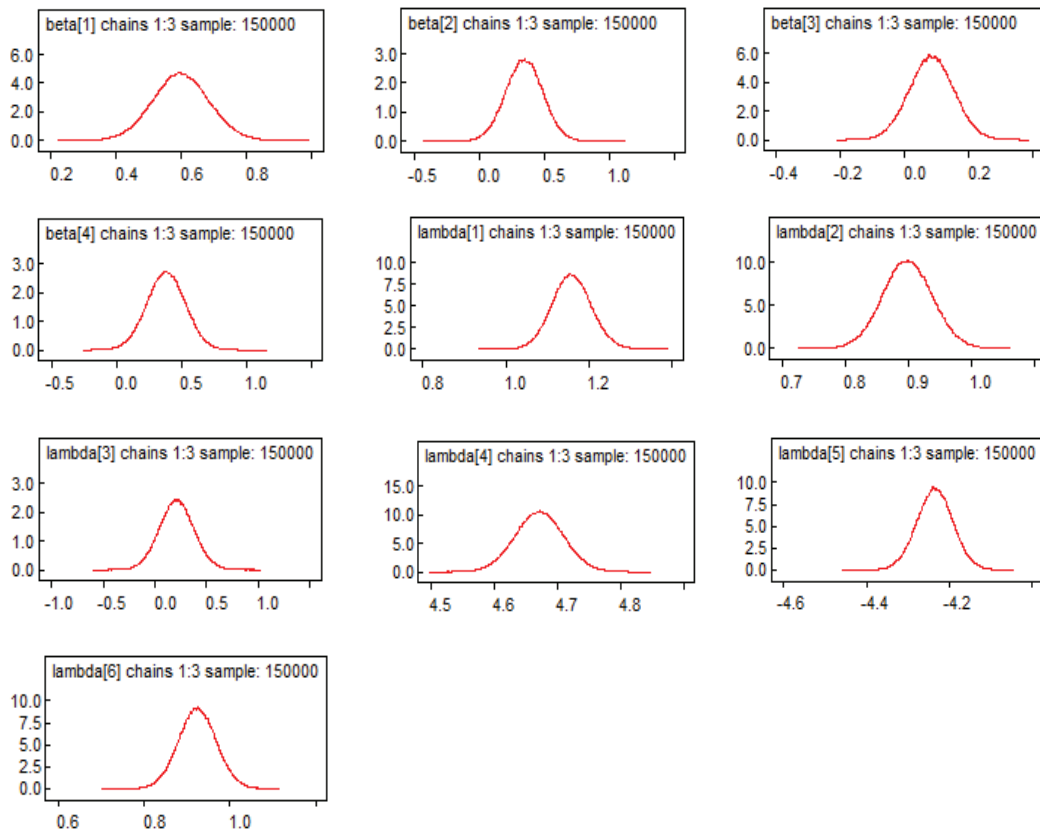


Figure 3.2. Density diagram of posterior distribution of path parameters.

인과관계에 대한 해석과 모형의 적합성을 볼 때 Posterior5가 가장 우수한 것으로 판단하였다.

Posterior5에서 경로계수의 사후분포 그림은 Figure 3.2와 같다. 모든 경로의 사후분포가 종모양으로 나타났다. 또한 베이지안 접근에서 알고리즘의 수렴 여부를 확인하기 위해서 초기 300번까지 반복 수에 따른 수렴과정을 나타낸 결과는 Figure 3.3과 같다. 세 가지 초기값(-10.0, 0.1 및 10.0)에 관계없이 모든 경로에서 빠르게 수렴하였다. 또한 변동이 상대적으로 큰 초기 5,000번은 삭제하고, 50,000번의 반복을 수행하면 잘 수렴하는 것을 확인하였다.

4. 결론

본 연구는 동계작물인 청보리 생산량에 대한 기후의 영향 및 인과관계 해석을 위하여 베이지안 방법을 이용한 종단자료 구조방정식모형으로 분석하였다. 그 결과 표본크기가 작은 청보리 자료에 작물의 대표성이 우수한 이탈리아 라이그라스에 대한 빈도론적 구조방정식 결과를 사전정보로 사용한 경우가 적절한 해석이 가능하였다. 또한 헤이우드 케이스 현상도 줄어드는 것을 확인하였다. 본 연구를 통해 청보리 자료분석에 적합한 사전정보로 이탈리아 라이그라스의 분석결과를 활용할 수 있음을 확인하였으며, 표본수집이 어려운 호밀 등의 베이지안 연구에도 사전정보로 사용될 수 있을 것으로 기대한다.

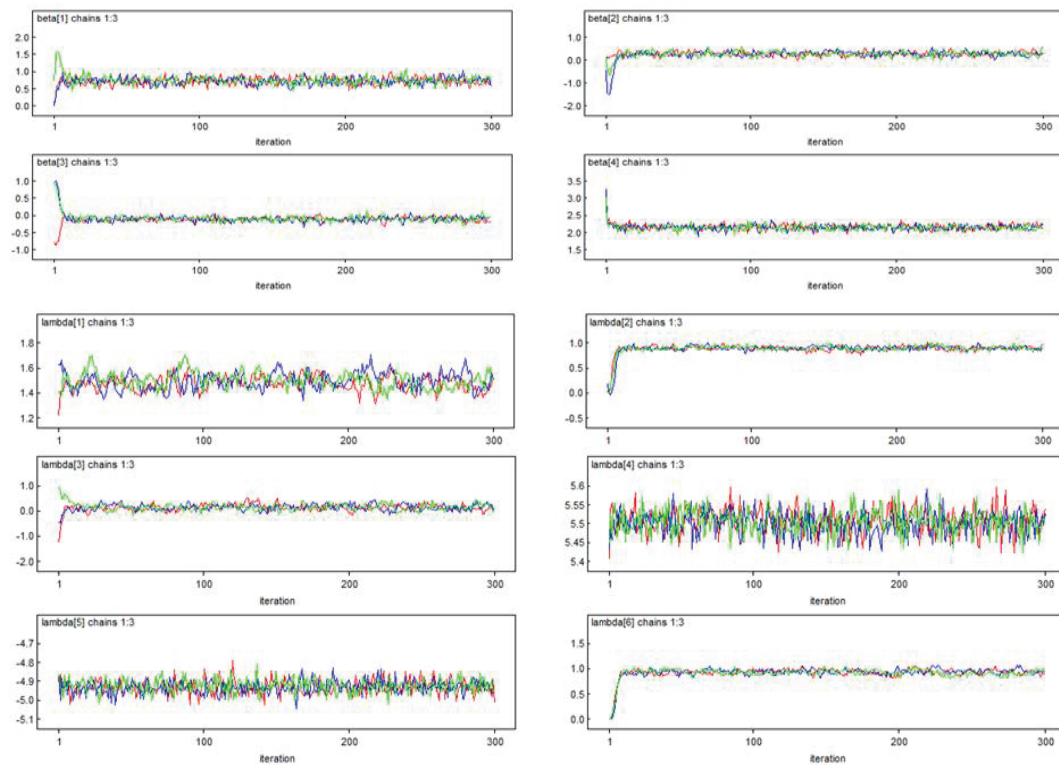


Figure 3.3. History of path parameters until 300th replications.

References

- Arbuckle, J. (2005). *Amos 6.0 user's Guide, Marketing Department*, SPSS incorporated, Chicago.
- Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models, *Sociological methodology*, **17**, 37–69.
- Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality (Doctoral dissertation, University of Groningen).
- Descheemaeker, K., Smith, A. P., Robertson, M. J., Whitbread, A. M., Huth, N. I., Davoren, W., and Llewellyn, R. (2014). Simulation of water-limited growth of the forage shrub saltbush (*Atriplex nummularia* Lindl.) in a low-rainfall environment of southern Australia, *Crop and Pasture Science*, **65**, 1068–1083.
- Dunson, D. B., Palomo, J., and Bollen, K. (2005). Bayesian structural equation modeling, (Technical report).
- Gao, S., Mokhtarian, P. L., and Johnston, R. A. (2008). Nonnormality of data in structural equation models, *Transportation Research Record, Journal of the Transportation Research Board*, **2082**, 116–124.
- Kazama, J. I., De Saeger, D., Kuroda, K., Murata, M., and Torisawa, K. (2010). A Bayesian method for robust estimation of distributional similarities, In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 247–256.
- Kim, J. D., Lee, H. J., Jeon, K. H., Yang, G. Y., Kwon, C. H., Sung, H. G., Hwangbo, S., and Jo, I. H. (2010). Effect of harvest stage, Wilting and Crushed rice on the forage production and silage quality of organic Whole crop barely, *Journal of The Korean Society of Grassland and Forage Science*, **30**, 25–34.
- Kim, K. D., Sung, K. I., Jung, Y. S., Lee, H. I., Kim, E. J., Nejad, J. G., Jo, M. H., and Lim, Y. C. (2013). Suitability classes for Italian ryegrass (*Lolium multiflorum* Lam.) using soil and climate digital database in Gangwon province, *Journal of the Korean Society of Grassland and Forage Science*, **32**, 437–446.

- Kim, M. J., Sung, K. I., Kim, B. W., Peng, J. L., Hyeon, D. H., Lee, B. H., Kim, E. J., Jo, M. H., Lim, Y. C., and Kim, K. D. (2014a). Prediction of Italian ryegrass yield affecting climate and soil, In *Proceedings of 2014 Annual Congress of Korean Society of Animal Sciences and Technology*, **204**.
- Kim, M. J., Sung, K. I., and Kim, Y. J. (2014b). Analysis of climate effects on Italian ryegrass yield via structural equation model, *The Korean Journal of Applied Statistics*, **27**, 111–120.
- Kim, M. J., Sung, K. I., and Kim, Y. J. (2015). Analysis of causal relation for climates and yield of Italian ryegrass and Whole crop barley via structural equation model(Manuscript).
- Kolenikov, S. and Bollen, K. A. (2012). Testing negative error variances is a Heywood case a symptom of misspecification, *Sociological Method and Research*, **41**, 124–167.
- Lee, B. H., Kim, Sung, K. I., Kim, E. J., Peng, J. L., Ji, D. H., Ghassemi Nejad, J., Kim, M. J., Jo, M. H., Lim, Y. C., and Kim, K. D. (2013). Comparison between groups variables influencing dry matter yield of Whole crop corn, In *Proceeding of The Korean Society of Grassland and Forage Science*.
- Lim, S. and Melville, N. P. (2009). Robustness of structural equation modeling to distributional misspecification: empirical evidence and research guidelines, Electronic copy available at <http://ssrn.com/abstract=1375251>.
- Mullen, M. R., Milne, G. R., and Doney, P. M. (1995). An international marketing application of outlier analysis for structural equations: a methodological note, *Journal of International Marketing*, **3**, 45–62.
- Peng, J. L., Kim, E. J., Lee, B. H., Ji, D. H., Ghassemi Nejad, J., Kim, M. J., Jo, M. H., Lim, Y. C., and Kim, K. D. (2013). Comparison of annual dry matter yields of whole crop corn, In *Proceeding of The Korean Society of Grassland and Forage Science*.
- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models, *Psychometrika*, **49**, 37–47.
- Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yield under climate change, In *Proceeding of the National Academy of Sciences*, **106**, 15594–15598.
- Shin, C. N., Ko, K. H., and Kim, J. D. (2012). Agronomic characteristics and forage productivity of Italian ryegrass(*Lolium multiorum* Lam.) cultivar, *Journal of The Korean Society of Grassland and Forage Science*, **32**, 229–236.
- Sung, K. I., Kim, B. W., Peng, J. L., Hyeon, D. H., Lee, B. H., Kim, E. J., Jo, M. H., Lim, Y. C., Kim, K. D., and Kim, M. J. (2014). Prediction of Italian ryegrass yield affecting climate variables, In *Proceedings of 2014 Annual Congress of Korean Society of Animal Sciences and Technology*, **203**.
- Wothke, W. (1993). Nonpositive definite matrices in structural modeling, In *Testing Structure Equation Models*, edited by K.A.Bollen and J.S. Ling. Newbury Park: Sage, 256–256.

청보리 생산량의 기후요인 분석을 위한 베이지안 구조방정식 모형

김문주^a · 전민희^a · 성경일^b · 김영주^{a,1}

^a강원대학교 정보통계학과, ^b강원대학교 사료생산과학전공

(2015년 11월 12일 접수, 2016년 1월 5일 수정, 2016년 1월 28일 채택)

요약

청보리는 국내에서 자급자족되는 중요한 동계 품사료이다. 본 연구는 구조방정식 모형을 이용하여 온도, 강수 및 일조시간과 관련 있는 기후요인이 청보리의 생산량에 미치는 경로와 영향력을 파악하였다. 청보리의 소표본 자료의 한계를 보완하기 위하여 베이지안 구조방정식 방법을 사용하였다. 베이지안 방법의 사전분포로 표준정규분포, 청보리 자료의 빈도론적 구조방정식 결과와 가장 대중적인 동계 품사료인 이탈리아인 라이그라스의 빈도론적 구조방정식 결과를 이용하였다. 또한, 사전분포의 헤이우드 케이스 수정을 하지 않은 경우와 수정한 경우에 대하여 구한 사후분포의 결과를 비교하여 동계작물의 생육특성과 잘 부합하는 사전분포를 탐색하였다. 분석 결과, 사전분포의 헤이우드 케이스를 수정하여 이탈리아인 라이그라스의 빈도론적 구조방정식 결과를 사전분포로 사용하는 것이 가장 적절하였다. 그러므로 본 연구의 베이지안 접근법을 통해 표본 수집이 어려운 품사료 연구에 좋은 제안이 될 것이다.

주요용어: 청보리, 구조방정식, 베이지안 방법, 헤이우드케이스, 종단자료

본 연구는 강원대학교 대학회계 학술연구조성비(관리번호-520150416)와 농촌진흥청 공동연구사업(과제번호: PJ009766)의 지원에 의해 이루어진 것임.

¹교신저자: (24341) 강원도 춘천시 강원대학길1, 강원대학교 정보통계학과. E-mail: ykim7stat@kangwon.ac.kr