IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 10, Issue 3 Ver. IV (May-Jun. 2014), PP 101-108
www.iosrjournals.org

A Bayesian model for a crop yield in a district of Andhra Pradesh, India

A.Nageswara Rao and M.ChattiBabu
1.Dept. of statistics, Andhra University,Waltair-530003,India
2. Dept. of statistics, Andhra University,Waltair-530003,India

Abstract: *Non-parametric and semi-parametric Bayesian regression is useful tools for practical data analysis. They provide posterior mean or median estimates, confidence bands and estimates of other functional without having to rely on approximate normality of estimators. The data that are analyzed are the rice yield in a district of Andhra Pradesh state and also the state data. We use a Bayesian threshold model with non-linear functions of four relevant continuous variables considered for the study. Inferences are based on Markov Chain Monte Carlo technique. The dependence of the results on the hyper parameters of the estimated variance components are analyzed. The effects of the covariates on rice yield in the district are compared with those of the state data.*

I. Introduction

The main scientific objective of this study is to explain the processes causing variation in the yield of rice in West Godavari district of Andhra Pradesh. The data that are analyzed are the secondary data available for the years 1971 to 2006. A Bayesian model with non-linear functions of continuous covariates such as rainfall, rice irrigated area and total fertilizer consumption is considered. In particular the relationships between rice yield and the covariates are of interest.

We cannot assume, conditional on covariate effects, yield follows a normal distribution. The explanatory variables may have non-linear effects. The dependant variable 'yield' may be considered as categorical with categories normal and not normal. Now a semi parametric binary regression is needed that allows the simultaneous non-parametric modeling and estimation of non-linear effects of the covariates.

Let the metrical or spatially correlated covariates $x_1, x_2, \ldots, x_p$ say with unknown, possibly non-linear effects and a vector of further covariates whose influence on the predictor is assumed to be linear. Therefore an additive predictor η with random effects ε to account for unobserved heterogeneity or correlation is given by

$$\eta=\sum_{i=1}^{p} f\left(x_i\right)+w^1\beta+\varepsilon \qquad \ldots\ldots\ldots \quad (1.1)$$

Together with an exponential family observation model and a suitable link function equation (1.1) defines a Generalized Additive Mixed Model (GAMM). In this paper we consider a Bayesian approach via Markov Chain Monte Carlo (MCMC) sampling for inference in GAMM. The MCMC procedure provides samples from all posteriors of interest and permits the estimation of posterior means, medians, quantiles, confidence bands and predictive distribution (Woods, S.(2006)).\

II. Bayesian semi-parametric mixed model:

Let y be a response, $X=(x_1, x_2, \ldots x_p)$ be a vector of metrical or spatial covariates and w be of further covariates.

Generalized additive and semi parametric models(Hastie and Tibshirani 1990) assume that given x_i - $=(x_{i1},x_{i2},\ldots,x_{ip})$ and w_i , the distribution of y_i belongs to an exponential family, with mean $\mu_i = E(y_i/x_i w_i)$ linked to an additive semi parametric predictor η_i by

$$\mu_i = h(\eta_i), \; \eta_i = f_1\left(x_{i1}\right)+f_2\left(x_{i2}\right)+\ldots\ldots.+f_p\left(x_{ip}\right)+w_i^1\,\eta \quad \ldots.(2.1)$$

Multi-categorical time data consist of observations (y_{it}, x_{it}, w_{it}), $i = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$ for the i^{th} individual, where the response variable y is observed in ordered or unordered categories, say r, $r \in (1,2,..,k)$. If we consider the spatial location or site s on a spatial array $\{1,2,...,s,....,S\}$ for each unit as an additional information, multi-categorical time space data on n individuals or units then consists of observations ($y_{it}, x_{it}, w_{it}, s_i$), $i=1,2,\ldots,n$; $t=1,2,\ldots,T$. where $s \in (1,2,..,S)$ is the location of individual i.

2.1 Cumulative Threshold Model:

Categorical response models may be motivated from the consideration of latent variables (Fahrmeir and Tutz, 2001). Let u be a latent variable given by

$$u = \eta + \varepsilon \qquad \ldots\ldots (2.2)$$

Where η is a predictor dependent on covariates and parameters and ε be the error variable. It is postulated that y is a categorized version of u obtained through the threshold mechanism y=r if and only if $\theta_{r-1} \le u \le \theta_r$

r=1, 2… k with thresholds $-\infty = \theta_0 < \theta_1 < \theta_2 < \ldots. < \theta_k = \infty$. If the error variable ε has distribution function F, it follows that y obeys a cumulative model

$$p(y \le r) = F(\theta_r - \eta) \qquad \ldots\ldots (2.3)$$

The most popular choices for F in (2.3) are the logistic and standard normal distribution leading to cumulative logit or probit models.

The covariates w enter the model through the predictor η. To consider possible non-linear functions of the continuous covariates, the linear predictor is given by

$$\eta = \sum_{i=1}^{p} f_i(x_i) + w^1\eta \qquad \ldots\ldots (2.4)$$

Here $f_1, f_2, \ldots, f_p$ are possibly non-linear functions of the continuous covariates $X=(x_1, x_2, \ldots x_p)^1$. The term $w^1\eta$ corresponds to effects of covariates w.

III. Prior Model

For Bayesian inference, unknown functions $f_1, f_2, \ldots, f_p$ and η are considered as random variables and have to be supplemented by appropriate prior distributions. For the fixed effect parameters θ and η. We assume diffuse priors $p(\theta)\ \alpha$ const., $p(\eta)\,\alpha$ const. and functions $f_1, f_2, \ldots, f_p$ of continuous metrical covariates are specified by p-splines, which were introduced by Eilers and Marx (1996) in a frequentist setting and by Lang and Brezger (2004) in a Bayesian version. The basic assumption is that an unknown function f_j of a covariate x_j can be approximated by a polynomial spline of degree l defined on a set of equally spaced knots $x_j^{min} = \zeta_0 < \zeta_1 < \ldots < \zeta_{d-1} < \zeta_d = x_j^{max}$ within the domain of x_j. The spline can be written in terms of a linear combination of $M_j = d + l$ B-spline functions B_m, i.e. $f_j(x_j) = \sum_{m=1}^{M_j} \beta_{jm} B_m(x_j)$

Here $\beta_j=\left(\beta_{j1},....,\beta_{jm}\right)$ corresponds to the vector of unknown regression coefficients.In a Bayesian appropriate, the first or second order random walks used as a prior for the regression coefficients. More details about the Bayesian p-splines can be found in Lang and Brezger(2004). First and second order random walks are defined by

$$\beta_{jm}=\beta_{j,m-1}+u_{jm} \quad \text{or} \quad \beta_{jm}=2\beta_{j,m-1}-\beta_{jm-2}+u_{jm}$$

With Gaussian errors $u_{jm} \sim N\left(0,\tau_j^2\right)$. For a fully Bayesian analysis hyper priors for τ_j^2 are introduced in a further stage of the hierarchy. This allows for a simultaneous estimation of the unknown function and the amount of smoothness. Common choices are highly dispersed inverse gamma priors (proper) $p\left(\tau_j^2\right) \sim IG\left(a_j,b_j\right)$. A possible choice for a_j and b_j are $a_j=1;\ b_j=0.005$. Alternatively we may set $a_j=b_j=0.001$. In some situations, the estimated non-linear functions may be sensitive to the particular choice of hyper parameters a_j and b_j. The Bayesian model specification is completed by the following conditional independence assumptions.

i) For given covariates and parameters f, β and ε, observations y_i are conditionally independent.

ii) Priors $p\left(f_j/\tau_j^2\right), j=1,2,....,p$ are conditionally independent.

iii) Priors for fixed and random effects and hyper priors τ_j^2, j=1, 2… p are mutually independent.

IV. Markov Chain Monte Carlo Inference:

Full Bayesian inference is based on the entire posterior distribution

$$p\left(\beta_1,\beta_2,......,\beta_p,\tau_1,\tau_2,....,\tau_p,\gamma,u/y\right) \alpha\ p\left(y/u\right)p\left(u/\beta_1,\beta_2,......,\beta_p,\gamma\right)\prod_{i=1}^{p}p\left(\beta_j/\tau_j\right)p\left(\tau_j\right)$$

The conditional likelihood $p\left(y_i/u_i\right)=\sum_r I\left(\theta_{r-1}<u<\theta_r\right).I\left(y_i=r\right)$ MCMC simultaneous is based on drawings from full conditionals of blocks of parameters, given the other parameters and the data. In the analysis of data the following full conditionals are used:

a) The full conditionals for the u_i are truncated normal with $u_i/. \sim TN_{t_1,t_2}\left(\eta_i,1\right)$. The truncation points depend on the observed y_i.

b) The full conditionals for the regression parameters $\beta_j, j=1,2,....,p$ are multivariate Gaussian with covariance matrix and mean given by

$$\Sigma_j=\left(X_j^1X_j+\frac{1}{\tau_j^2}K_j\right)^{-1}, \quad \mu_j=\Sigma X_j^1\left(U_j-\eta\right)$$

Where η is the part of the predictor η that is associated with the remaining effects in the model and K_j is the penalty matrix for p-splines. For example,

$$K_j = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & . & . & . & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

For p-splines with a first order random-walk penalty.

c) The full conditionals for the linear effects parameters γ are Gaussian with mean and covariance matrix given by

$$\mu_\eta = \left(w^1 w\right)^{-1} w^1 (u-\eta), \qquad \sum_\eta = \left(w^1 w\right)^{-1}.$$

d) The full conditionals for the variance parameters τ_j^2 are inverse gamma with parameters

$$a_j^1 = a_j + \mathrm{rank}\left(K_j\right)/2 \quad \text{and} \quad b_j^1 = b_j + \frac{1}{2}\beta_j^1 K_j \beta_j$$

e) The full conditionals for threshold $\theta_r, r = 1,2$ is uniform on the interval.

$$\left[\max\left\{u_i : y_i = r\right\}, \min\left\{u_i : y_i = r+1\right\}\right]$$

Since all full conditionals are known distributions we can use a Gibbs sampler, drawing successively random numbers from the conditional distributions of the parameters. Numerical efficiency is obtained by utilizing the band matrix structure of the posterior precision matrices p_j of the regression parameters. The details of the band matrix algorithms are described in George and Lim (1981). We compare different models in terms of the DIC (Spiegelhalter etal, 2002). The DIC is defined as DIC= $\overline{D}+P_D$. Where $\overline{D}$ is the posterior mean deviance and the P_D is the difference between the posterior mean deviance and the deviance evaluated at the posterior mean of y. credible intervals could be obtained by running the MCMC sampler several times. Justification for the usage of the DIC as a Bayesian analogue to the Akaike information can be found for example (Congdon (2006)).

V. Application:

This application illustrates the appropriateness of Bayesian structured additive model. The data on rice yield for the period 1971 to 2006 in west Godavari district of A.P. is considered. The most relevant covariates used here are the rice irrigated area, rainfall and total fertilizers consumption. The geo additive predictor.

$$\eta = \gamma_0 + f_1(\text{rainfall}) + f_2(\text{total fertilizer consumption}) + f_3(\text{rice irrigated area})$$

The continuous covariates assumed to have a possibly non-linear effect on the response variable 'yield' and are therefore modeled non-parametrically (as p-splines with second order random walk prior).

Data of East Godavari and Andhra Pradesh are analyzed separately using BayesX program available in the public domain.

To assess the dependence of results on the hyper parameters a_j and b_j of variance components τ_j^2 different choices are made for the hyper parameters and the corresponding models are estimated with 20 equidistant knots.

The results for the East Godavari are presented in table 5.1 to 5.6. It can be noticed from the results that the estimated non-linear functions are sensitive to the particular choice of the hyper parameters.

Table 5.1 Hyper parameters $a_i = 1, b_i = 0.001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	278.548	394.066	1.27445	29.4839	177.052	473.965	1790.47	0.247036	2191.03
TFC	153.723	120.377	2.54849	10.4962	139.82	357.369	486.407	0.581688	487.21
RICEIA	398.666	571.966	2.79646	13.3062	233.437	1050.25	2607.71	2.11046	2924.89
FIXED EFFECT (95%Posterior probability 1)	7.7518	0.040042	7.68918	7.69588	7.74859	7.80658	7.83602	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor).
Deviance information criterion (DIC) = 48.3252

Table 5.2 Hyper parameters $a_i = 0.001, b_i = 0.001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	109.601	185.8	1.07684	2.45477	17.5428	352.414	477.567	0.961894	1093.38
TFC	42.2985	63.7249	3.07838	4.578662	25.3407	74.1637	207.474	2.10046	411.275
RICEIA	148.815	162.849	2.39555	11.3903	87.0446	317.01	479.193	0.803216	921.715
FIXED EFFECT (95%Posterior probability 1)	7.75358	0.04798	7.68434	7.69477	7.7478	7.82205	7.8673	-	-

No. of iterations = 49 (at which corresponding penalized part was small relative to the linear predictor) Deviance information criterion (DIC) = 47.4345

Table 5.3 Hyper parameters $a_i = 0.0001, b_i = 0.0001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	465.435	827.012	2.06131	7.91261	152.743	1056.67	3742.43	1.12538	4407.46
TFC	30.2188	37.9546	1.62282	3.07651	17.404	55.8642	174.181	1.31665	177.169
RICEIA	164.176	233.113	2.86458	4.72198	66.0477	505.868	650.822	2.84257	1281.13
FIXED EFFECT (95%Posterior probability 1)	7.76324	0.04134	7.69316	7.69801	7.76402	7.81387	7.8311	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 45.9271

Table 5.4 Hyper parameters $a_i = 0.0005, b_i = 0.0005$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	264.489	304.137	3.46869	28.0392	136.453	634.625	1009.05	2.13984	1591.96
TFC	81.4782	159.453	0.998277	3.50616	23.199	220.276	585.784	0.786771	882.571
RICEIA	308	570.362	1.90691	5.07791	125.992	879.59	1384.95	1.01132	3527.42
FIXED EFFECT (95%Posterior probability 1)	7.75638	0.042884	7.66332	7.7074	7.76094	7.81027	7.83781	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 45.7097

Table 5.5 Hyper parameters $a_i = 1, b_i = 0.0005$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	916.488	601.307	4.20629	56.5261	846.937	1909.06	2051.93	3.33445	2515.96
TFC	676.542	871.383	7.09637	11.9	279.373	1801.99	2963.81	1.71575	4280.98
RICEIA	909.321	870.256	19.1335	158.148	669.171	2077.61	2811.28	8.93537	4756.17
FIXED EFFECT (95%Posterior probability 1)	7.7512	0.041923	7.66671	7.69207	7.76075	7.801	7.81479	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 45.2811

Table 5.6 Hyper parameters $a_j = 1$, $b_j = 0.0001$

VARIABLE	P0ST MEAN	P0ST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	P0ST MIN	P0ST MAX
RAINFALL	3523.95	4419.46	10.471	92.5797	1540.84	10319	16052.7	6.00035	19879.6
TFC	1050.34	1407.5	31.9724	47.7476	432.211	3906.68	4772.43	15.1374	4954.23
RICEIA	4116.03	4272.65	5.91861	35.5659	2551.33	10218.9	14891.5	5.22488	19706.1
FIXED EFFECT (95%Posterior probability 1)	7.74756	0.03854	7.68316	7.69876	7.7492	7.80494	7.81925	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 44.6188.

Predicted Means

LNYIELD	RAINFALL	TFC	RICEIA
7.1468	798.9	8.62	320.15
7.4128	1077.1	35.29	315.89
7.3759	887.7	41.49	289.33
7.4116	1085.3	43.61	295.51
7.5522	1201.15	19.75	316.12
7.5148	1317	51.3	328.08
7.6257	1103	62.79	329.45
7.5512	1116	54.22	330.82
7.6629	897	94.6	351.75
7.6358	1230	82.65	313.58
7.72	972	67.27	349.37
7.7711	867	72.79	360.46
7.8051	805	79.71	353.05
7.7133	743	100.52	398.82
7.7965	1045	103.24	347.03
7.7928	1062	64.93	360.11
7.4378	1166	117.33	381.12
7.7493	1577	78.87	378.19
7.8782	1988	118.4	403.08
7.8176	959	186.45	404.99
7.5838	1283	192.96	357.6
7.7174	1084	142.33	355.92
7.8517	937	152.02	336.87
8.0275	1546	148.11	380.5
7.8296	1246	156.72	379.9
7.9128	1616	156.92	367.87
7.789	1062	148.69	401.44
7.8236	1692	162.95	361.39
7.832	1012	170.66	393.81
8.0507	1021	201.19	389.98
8.0953	997	203.63	396.87
8.0953	707	199.76	346.08
8.1505	1078	192.28	295.29
8.3131	873	187.62	387.49
8.268	1389	203.55	363.09
8.3047	1167	211.27	366.64

The least Deviance Information Criterion (DIC) is obtained for the hyper parameter values a_j=1 and b_j=0.0001 as 44.6188. Therefore the corresponding linear functions as given in table 5.6 best predict the response variable. Hence the corresponding model best fits the East Godavari data.

The following tables present the model fit for the A.P data

Table 5.7 Hyper parameters $a_j = 0.0001$, $b_j = 0.0005$

VARIABLE	P0ST MEAN	P0ST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	P0ST MIN	P0ST MAX
RAINFALL	27.1342	24.1988	2.69733	3.08665	16.328	73.5589	77.2121	1.25942	77.4072
TFC	28.6037	45.2597	1.17006	2.27726	11.1909	96.3297	200.418	0.734796	204.765
RICEIA	21.4873	23.3193	0.720074	1.61207	11.3735	58.4996	81.2314	0.381212	103.201
FIXED EFFECT (95%Posterior probability 1)	7.59611	0.000217	7.59566	7.59581	7.59612	7.59635	7.59647	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 52.9127.

Table 5.8 Hyper parameters $a_i = 0.0001$, $b_i = 0.0001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	51.088	101.106	0.537833	0.7855	14.9131	117.29	374.76	0.177401	529
TFC	21.6299	28.6771	2.69547	3.64781	9.98541	57.7891	110.707	1.56569	124.687
RICEIA	29.7891	37.2306	1.15213	6.30782	15.5231	50.8447	152.701	0.880892	217.026
FIXED EFFECT (95%Posterior probability 1)	7.59609	0.000269	7.59564	7.59579	7.59606	7.59642	7.5967	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 51.2981

Table 5.9 Hyper parameters $a_i = 0.001$, $b_i = 0.001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	2.20288	2.27811	0.297555	0.473373	1.32078	6.08249	7.40076	0.296346	10.7803
TFC	12.6498	14.1193	0.596392	1.35147	8.0675	26.3543	57.4407	0.424739	73.3349
RICEIA	12.6498	14.1193	0.596932	1.35147	8.0675	26.3543	57.4407	0.424739	73.3349
FIXED EFFECT (95%Posterior probability 1)	7.59598	0.000282	7.59539	7.5956	7.59602	7.59635	7.5965	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor) Deviance information criterion (DIC) = 50.8256

Table 5.10 Hyper parameters $a_i = 1$, $b_i = 0.001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	34.8425	29	1.68948	2.89466	26.3056	86.3443	89.7078	0.92088	97.3109
TFC	26.6964	28.65	1.6617	2.05712	17.1255	60.5415	95.1239	1.13735	166.33
RICEIA	29.1694	27.6674	1.76357	2.05675	19.1987	71.2962	98.7243	0.763205	100.006
FIXED EFFECT (95%Posterior probability 1)	7.59612	0.000217	7.59576	7.59582	7.59614	7.59638	7.59655	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 49.4493

Table 5.11 Hyper parameters $a_i = 1$, $b_i = 0.0001$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	178.067	216.55	0.895918	2.04183	108.812	413.584	828.458	0.69422	1043.41
TFC	79.9969	65.5753	2.1489	15.9446	65.7255	180.988	272.567	0.67412	280.343
RICEIA	109.863	129.509	3.48709	13.8822	58.2661	325.976	430.077	0.967588	596.006
FIXED EFFECT (95%Posterior probability 1)	7.59606	0.000243	7.59563	7.59575	7.59607	7.59644	7.59649	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 48.2099

Table 5.12 Hyper parameters $a_i = 1$, $b_i = 0.0005$

VARIABLE	POST MEAN	POST S.D	POST2.5% QUANTILE	POST10% QUANTILE	POST MEDIAN	POST90% QUANTILE	POST97.5% QUANTILE	POST MIN	POST MAX
RAINFALL	83.1002	55.0759	8.67444	13.0815	81.6843	153.759	228.123	2.10443	253.993
TFC	51.1997	38.6114	7.91293	15.4581	39.3106	100.756	159.748	6.24313	190.848
RICEIA	54.6377	53.2245	2.92948	8.77207	38.1442	104.205	194.109	2.62616	311.405
FIXED EFFECT (95%Posterior probability 1)	7.59617	0.000248	7.59578	7.59581	7.5962	7.59643	7.59661	-	-

No. of iteration = 49 (at which corresponding penalized part was small relative to the linear predictor). Deviance information criterion (DIC) = 43.943.

Predicted Means

LNYIELD	RAINFALL	TFC	RICEIA
7.5863	968.88	19.956	158.153
7.5868	722.53	23.07	136.14
7.58731	725.93	21.968	130.986
7.58782	894.52	19.136	152.597
7.58832	856.96	21.092	160.935
7.58883	1088.81	27.114	175.182
7.58934	1025.15	36.587	179.264
7.58984	921.35	42.768	177.158
7.59035	1123.75	50.498	178.56
7.59085	783.15	44.575	142.752
7.59136	835.35	46.236	146.862
7.59186	953.6	52.103	156.953
7.59237	829.2	56.828	148.96
7.59287	1086.69	74.095	170.63
7.59337	793.99	80.696	143.761
7.59388	786.84	74.094	141.283
7.59438	874.29	78.206	141.69
7.59488	926.18	80.734	133.407
7.59539	1161.99	101.987	201.193
7.59589	1074.28	116.549	174.051
7.59639	1322.76	122.851	166.528
7.59689	1016.82	128.254	162.422
7.5974	860.35	123.604	148.027
7.5979	827.09	125.239	146.466
7.5984	879.41	130.365	149.798
7.5989	964.95	164.662	156.948
7.5994	1105.5	134.542	170.783
7.5999	865.5	145.92	146.657
7.6004	1083.05	150.319	180.042
7.6009	822.82	165.685	167.021
7.6014	905.55	166.662	183.667
7.6019	874.68	159.056	152.953
7.6024	621.5	153.151	122.24
7.6029	940.82	156.316	127.48
7.6034	713.95	165.541	128.25
7.6039	1146.5	198.339	129

The least DIC value of 43.943 is obtained for the hyper parameter value of a_j=1, b_j=0.0005 which are different from the model of East Godavari data. Therefore the corresponding linear functions given in table 5.12 best predict the response variable. Hence the corresponding model best fits the A.P. data.

5.5. Future Work:

Applicability of the model to the other districts can be studied. Plausible interactions between covariates can also be considered.

Reference

[1]. Augustine, N.H. et.al. (2007) A spatial model for the needle losses of pine-trees in the forests of Baden-Wurttemberg: an application of Bayesian structured additive regression. Appl. Statist.56,29-50.

[2]. Congdon, P.(2006) A model for non-parametric spatially varying regression coefficients Computnl. Statist. Data Anal. , 50,422-455.

[3]. Eilers, P. and Marx, B.(1996) Flexible smoothing using B-splines and penalizes likelihood. Statist. Sci., 11,89-121.

[4]. Fahrmeir, L. and Tutz, G. (2001) Multivariate Statistical Modeling based on Generalized Linear Models. New York: Springer.

[5]. Fahrmeir, L. and Lang, S. (2001) Bayesian semi-parametric regression analysis of multi-categorical time-space data. Ann. lnst. Statist. Math., Vol.53, No. 1, 11-30.

[6]. Hastie, T and Tibshirani, R. (1990) Generalized Additive Models. London: Chapman and Hall.

[7]. Lang, S. and Brezger, A. (2004) Bayesian P-splines J.Computnl. Graph. Statist., 13, 183-212.

[8]. Woods, S. (2006) Generalized Additive Models: an introduction with R.Boca Raton : Chapman and Hall-CRC