

# The Bayesian Polynomial Regression Model through INLA Using Gaussian Method with Empirical Bayesian, CCD and GRID Strategy

Hery Tri Sutanto<sup>1,\*</sup>, Henny Pramoedyo<sup>2</sup>, Wayan Surya Wardhani<sup>2</sup>, Suci Astutik<sup>2</sup>

<sup>1</sup> Graduate Student of Mathematics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia

<sup>2</sup> Statistics Department, Brawijaya University, Jl. Veteran, Malang, Indonesia

\*Corresponding author. Email: [herysutanto@student.ub.ac.id](mailto:herysutanto@student.ub.ac.id)

## ABSTRACT

Relationship rice between with height can be described with scatter diagram. Scatter diagram can product the quadratic polynomial regression mod els and the qubic polynomial regression mode which possible. The quadratic polynomial regression model the best variable height and height<sup>2</sup> significant. Then parameter in the quadratic polynomial regression model can be estimated with Bayesian INLA with Gaussian and Empirical Bayes Strategy compare CCD, GRID Strategy. The posterior marginal intercept  $\beta_0$ , coefficient  $\beta_1$  and  $\beta_2$  with Gaussian method and different strategy product similar. The posterior marginal of random effects  $\tau$  the grid strategy different with CCD and empirical Bayes strategy.

**Keywords:** Relationship, scatter diagram, quadratic polynomial regression models, Gaussian method, Empirical Bayes Strategy, CCD Strategy, Grid Strategy, marginal posterior and Bayesian INLA.

## 1. INTRODUCTION

Regression analysis is statistical model used to study the relationship between one dependent variable with more than one independent variable [1]. The construction of a regression model that is suitable for statistical data associated with a "scatter diagram" of the variable. One regression model that is widely used is study the relationship of independent variable with dependent variable in the form of polynomial functions is polynomial regression.

The difficulty in the polynomial regression model is determining the degree of the polynomial model [2]. Determination of the degree of polynomial with the Bayesian approach has two choices. First determine the degree of polynomial which is a Bayes Factor and second the calculation of the posterior probability model using *Bayesian Information Criterion* by [3]. The method is used to compute the posterior distribution in

the Bayesian model is the Markov Chain Monte Carlo. The Markov Chain Monte Carlo theory describes of the simulate data for several hours or days before convergence to the posterior distribution. If the model is more complex with many parameters, it will be difficult to determine the convergence and feasibility [4].

In 2009, Rue is developed Integrated Nested Laplace Approximation (INLA) as instead Bayesian MCMC Methods. INLA is approximate that has more speed and accuracy than MCMC [5]. Polynomial Regression has not been studied in more detail using the INLA approximation, so this study will discuss the use of INLA in polynomial regression in the relationship of height between with rice obtained. Based on the research background the issues that will be raised in this study are the following: 1) how to scatter diagram on data and determine polynomial regression model possible, 2) how the z parameterization of a polynomial regression model with Gaussian method which might

use the grid strategy, 3) how the z parameterization of a polynomial regression model with Gaussian methods which might use the CCD strategy, 4) how the z parameterization of a polynomial regression model with Gaussian methods which might use the empirical Bayes strategy.

## 2. STUDY LITERATURE

### 2.1 Polynomial Regression

If Y is response variable and x variable independent, then polynomial regression model with degree j ( $M_j$ ) is

$$Y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \dots + \beta_jx^j + \varepsilon, 0 < j \leq d \quad (1)$$

There are two ways in choose degree j in model

- a) Keep adding model until the added model is not significant variable
- b) Star model with the highest degree the eliminate variable not significant [6]

### 2.2 The Integrated Nested Laplace Approximation.

Compute the parameter-parameter joint posterior distribution for model with Bayesian is used Markov Chain Monte Carlo is ran too long for more the complex model and many parameters [7]. In 2009, Rue proposes a way of approximation the marginal distribution. This compute the marginal distribution faster because is used numerical integration method [5].

### 2.3 Compute Bayesian INLA

If is use the Latent Gaussian Model in compute Bayesian INLA necessary the posterior marginal for latent parameter  $\theta$  and hyperparameter  $\psi$  follows:

$$p(\theta_i|y) = \int p(\theta_i|\psi, y) p(\psi|y) d\psi \quad i = 1, 2, \dots, n_d \quad (2)$$

and

$$p(\Psi_j|y) = \int p(\Psi|y) d\Psi_j \quad j=1, 2, \dots, s \quad (3)$$

Equation (3) is not solution analytic then the posterior marginal for latent parameter  $\theta$  with approximation:

$$\tilde{p}(\theta_i|y) = \int \tilde{p}(\theta_i|\Psi, y) \tilde{p}(\Psi|y) d\Psi \quad i = 1, 2, \dots, n_d \quad (4)$$

Equation (4) is not solution analytic then the posterior marginal for hyperparameter  $\Psi$  with approximation:

$$\tilde{p}(\Psi_j|y) = \int \tilde{p}(\Psi|y) d\Psi_j \quad j = 1, 2, \dots, s \quad (5)$$

### 2.5. Approximation $P(x_i|\Psi, y)$

The approximation to the marginals of the hyperparameters can be determined by marginalizing over  $\tilde{p}(\Psi|y)$  above to obtain  $\tilde{p}(x_i|\Psi, y)$ . The approximation to the marginal of the latent effects require integrating the hyperparameters out and marginalizing over the latent effects.

### 2.6 Approximation $P(\Psi|y)$

Rue et all [5] propose two different exploration schemes which requiring a reparameterization the  $\Psi$  space of several steps.

- a) Compute mode  $p(\Psi|y)$  is obtained by maximize  $\log p(\Psi|y)$  on  $\Psi$  through Newton-Raphson Method.
- b) Compute the negative Hessian H matrix on mode.
- c) Compute the eigenvalue decomposition  $H^{-1} = V\Lambda V^T$
- d) With application the eigenvalue decomposition for each value of the hyperparameter  $\Psi$  can reparameterization z so that  $\Psi(z) = \Psi^* + V\Lambda^{1/2}z$  [8]

$\log p(\Psi|y)$  is explored using the z in three different schemes depending on the number of hyperparameters and required a reparameterization of  $\Psi$  space to make the density more regular [2]

- i) Using s regular grid on step h centered on the mode  $z=0$  and points in the grid are considered if  $|\log p(\Psi(0)|y) - \log p(\Psi(z)|y)| < \delta$  where  $\delta$  a given threshold. Working in along the axis in the z parameterization and all intermediate points that fulfill the previous condition are added. This will provide a set of configuration of the hyperparameters about the posterior mode that can be used in the numerical integration in popular the grid strategy.
- ii) With CCD (central composite design) by [9] centered at  $\Psi(0)$  can be used so that a few strategically placed points are obtained. This can be more efficient that the grid strategy as the high dimension of the hyperparameter. The CCD strategy, a number of points that fill the space are chosen using a response surface design. Rue et al. [5], this CCD strategy worked well in many cases.
- iii) Empirical Bayes strategy used the large number of hyperparameters that will plug the posterior mode of  $\Psi|y$ . The strategy will work when the variability in the hyperparameters does not impact the posterior of

the latent effects  $x$ . This strategy of plugin estimators can be used to fit high parameters model and they can provide good strategy in a number scheme.

**2.7 Prior in INLA**

Prior in Bayesian is use definition which is very carefully and describe in more detail. If not prior distribution a parameter use the default prior. Prior-prior in INLA are set in the internal representation of the parameter, which have the different scale of the parameter in the model [2].

**3. RESULT AND DISCUSSION**

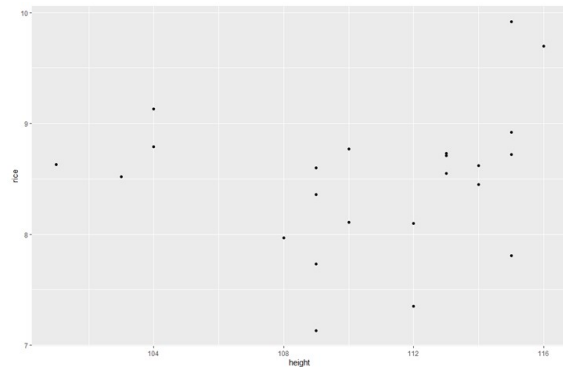
**3.1 The Quadratic Polynomial Regression Model Bayesian INLA**

**3.1.1. Data and scatter diagram**

**Table 1.** Data about height (cm) and rice (ton/ha)

Data Rice	height	city	D	
8.45	114	Luwu Timur	12996	1
8.72	115	Luwu Timur	13225	1
7.97	108	Luwu Timur	11664	1
8.62	114	Luwu Timur	12996	1
8.77	110	Luwu Timur	12100	1
8.92	115	Luwu Timur	13225	1
7.81	115	Luwu Timur	13225	1
8.55	113	Luwu Timur	12769	1
7.73	109	Luwu Timur	11881	1
8.36	109	Luwu Timur	11881	1
8.79	104	Luwu Timur	10816	1
8.63	101	Luwu Timur	10201	1
8.1	112	Pinrang	12544	0
8.71	113	Pinrang	12769	0
8.6	109	Pinrang	11881	0
8.45	114	Pinrang	12996	0
8.11	110	Pinrang	12100	0
8.73	113	Pinrang	12769	0
9.7	116	Pinrang	13456	0
9.92	115	Pinrang	13225	0
7.35	112	Pinrang	12544	0
7.13	109	Pinrang	11881	0
9.13	104	Pinrang	10816	0
8.52	103	Pinrang	10609	0

Scatter diagram between tree height (cm) with the rice weight (ton/ha) on figure 1.



**Figure 1.** Scatter diagram between tree height (cm) with the rice weight (ton/ha)

After plotting data will produce scatter diagram on figure 1. From identification on figure 1.

- a) Predict the quadratic polynomial regression model with form  $rice_i = \beta_0 + \beta_1 height_i + \beta_2 height_i^2 + \epsilon_i$
- b) Predict the cubic polynomial regression model with form

$$rice_i = \beta_0 + \beta_1 height_i + \beta_2 height_i^2 + \beta_3 height_i^3 + \epsilon_i$$

**3.1.2. The quadratic polynomial regression model**

Then quadratic polynomial regression model with equation  $rice = 235.285 - 4.190102 height + 0.019316 height^2$  with two independent variable significant.

Call:

`lm(formula = rice ~ height + height^2, data = datapol)`

Residuals:

Min	1Q	Median	3Q	Max
-1.07345	-0.20009	-0.01153	0.31008	1.03655

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	235.285499	77.478688	3.037	0.00627 **
height	-4.190102	1.422944	-2.945	0.00774 **
$height^2$	0.019316	0.006525	2.960	0.00747 **

Signif.codes: 0‘\*\*\*\*’ 0.001 ‘\*\*\*’ 0.01 ‘\*\*’ 0.05 ‘.’ 0.1 ‘.’ 1

Residual standard error: 0.5518 on 21 degrees of freedom

Multiple R-squared:0.3095, Adjusted R-squared: 0.2437

F-statistic: 4.706 on 2 and 21 DF, p-value: 0.02047

Value  $Pr(>|t|) = 0.00774 < \alpha = 0.005$  for height and  $Pr(>|t|) = 0.00774 < \alpha = 0.005$  for height<sup>2</sup> means the height and height<sup>2</sup> variable significant the impact height tree to rice.

### 3.1.3. The cubic polynomial regression model

After the height and height<sup>2</sup> variable significant is try the cubic polynomial regression model.

Then cubic polynomial regression model with equation  $rice = -3447 + 97.54 \text{ height} - 0.9166 \text{ height}^2 + 0.002867 \text{ height}^3$  with height, height<sup>2</sup>, height<sup>3</sup>, variable not significant.

Call:

lm (formula = rice ~ height + height<sup>2</sup> + height<sup>3</sup>, data = datapol)

Residuals:

Min	1Q	Median	3Q	Max
-1.15569	-0.24376	0.01995	0.30447	0.95431

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.447e+03	2.272e+03	-1.517	0.145
height	9.754e+01	6.277e+01	1.554	0.136

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	234.868	73.915	89.749	234.866	379.867	234.868	0
height	-4.182	1.357	-6.848	-4.182	-1.519	-4.182	0
height <sup>2</sup>	0.019	0.006	0.007	0.019	0.031	0.019	0

The model has no random effects

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for	3.59	1.06	1.88	3.47	6.00	3.24

Expected number of effective parameters (stdev): 3.00(0.00)

Number of equivalent replicates: 8.00

height<sup>2</sup> -9.166e-01 5.773e-01 -1.588 0.128

height<sup>3</sup> 2.867e-03 1.768e-03 1.621 0.121

Residual standard error: 0.5316 on 20 degrees of freedom

Multiple R-squared: 0.3897, Adjusted R-squared: 0.2982

F-statistic: 4.257 on 3 and 20 DF, p-value: 0.01767

Because height, height<sup>2</sup>, height<sup>3</sup>, variable not significant choose the quadratic polynomial Regression model [6].

Then compute Bayesian inla quadratic polynomial regression model with the Gaussian method and the empirical Bayesian, CCD and Grid strategy for compare.

### 3.1.4. The Bayesian INLA quadratic polynomial regression model with the Gaussian methods and the empirical Bayesian strategy

Call:

c("inla(formula = formula, family = \"gaussian\", data = datapol, \" \" control.inla =list(strategy = \"gaussian\", int.strategy = \"eb\"))")

Time used:

Pre = 1.03, Running = 0.166, Post = 0.0613, Total = 1.26

Fixed effects:

The posterior distribution contains: mean posterior and 95% interval credible which can be used as Bayesian instead. Mean posterior from parameter  $\beta_1$  is -4.182 and 95% interval credible is [-6.848, -1.519] and mean posterior from parameter  $\beta_2$  is 0.019 and 95 % interval

credible is [0.007, 0.031] with height probability. The  $kld=0$  (small value) is a diagnostic that measures the accuracy of the INLA approximation a good.

Expected number of effective parameters (standard deviation), which can be used a measure of the complexity of the model and number of equivalent replicates which is number of observations divided by the effective number of parameters. The average number of observations available to estimate each parameter in the model and higher values are better [2].

Marginal log-Likelihood -44.08

Three posterior densities of the parameters and one hyperparameter are using the integration Gaussian and the empirical Bayes strategy in Figure 2-4.

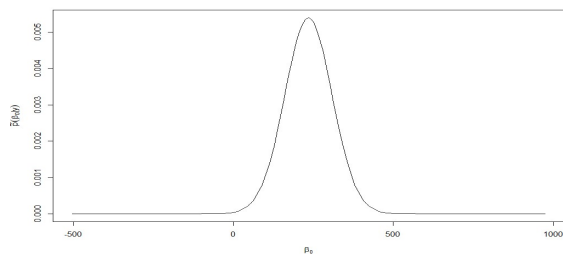


Figure 2. Posterior marginal of the intercept  $\beta_0$  with INLA

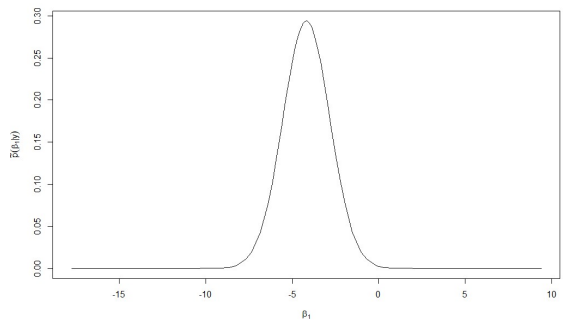


Figure 3. Posterior marginal of the coefficient  $\beta_1$  with INLA

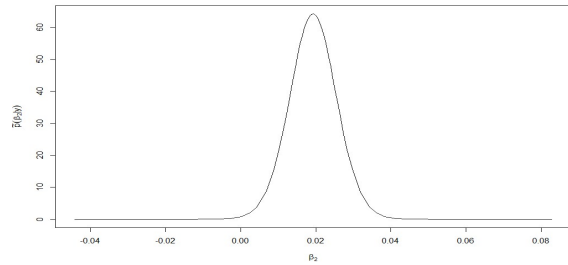


Figure 4. Posterior marginal of the coefficient  $\beta_2$  with INLA

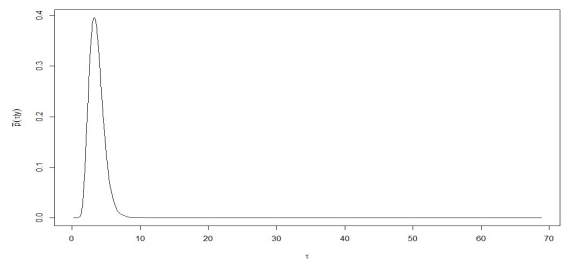


Figure 4. Posterior marginal of the precision of the random effects  $\tau$  with INLA

3.1.5. The Bayesian INLA quadratic polynomial regression model with the Gaussian methods and the central composite design(CCD) strategy

```
c("inla(formula = formula, family = \"gaussian\", data = datapol, \" \" control.inla = list(strategy = \"gaussian\", int.strategy = \"ccd\"))")
```

Time used:

Pre=1.3, Running =0.481, Post =0.0628, Total =1.84

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	234.844	75.351	85.985	234.855	383.411	234.882	0
height	-4.182	1.384	-6.914	-4.182	-1.452	-4.183	0
height <sup>2</sup>	0.019	0.006	0.007	0.019	0.032	0.019	0

The posterior distribution contains: mean posterior and 95% interval credible which can be used as Bayesian instead. Mean posterior from parameter  $\beta_1$  is -4.182 and 95 % interval credible is [-6.914, -1.452] and mean posterior from paramater  $\beta_2$  is 0.019 and 95 % interval

credible is [0.007, 0.032] with high probability. The  $kld = 0$  (small value) is a diagnostic that measures the

accuracy of the INLA approximation a good.

The model has no random effects:

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for the Gaussian observations	3.59	1.06	1.88	3.47	6.00	3.24

Expected number of effective parameters (stdev): 3.00 (0.001)

Number of equivalent replicates: 8.00

Marginal log-Likelihood: -44.08

Expected number of effective parameters (standard deviation), which can be used a measure of the complexity of the model and number of equivalent replicates which is number of observations divided by the effective number of parameters. The average number of observations available to estimate each parameter in the model and higher values are better [2].

Three posterior densities of the parameters  $\beta_0, \beta_1, \beta_2$  and hyperparameter  $\tau$  are using the integration Gaussian and the CCD strategy in Figure 5-8.

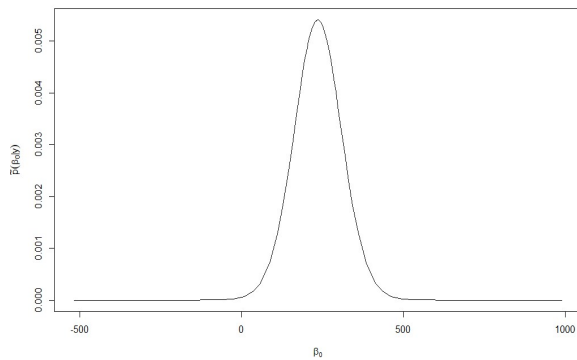


Figure 5. Posterior marginal of the intercept  $\beta_0$  with INLA

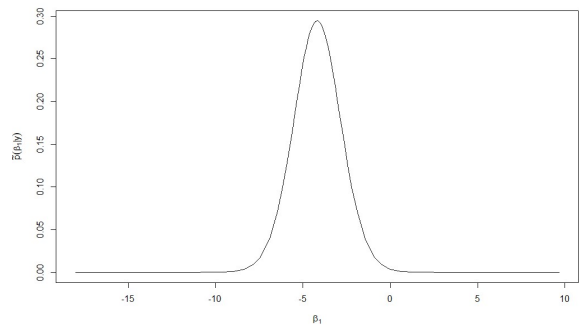


Figure 6. Posterior marginal of the coefficient  $\beta_1$  with INLA

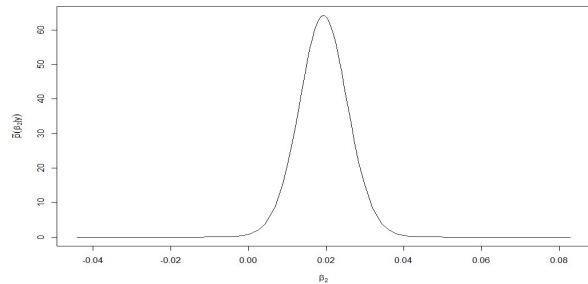


Figure 7. Posterior marginal of the coefficient  $\beta_2$  with INLA

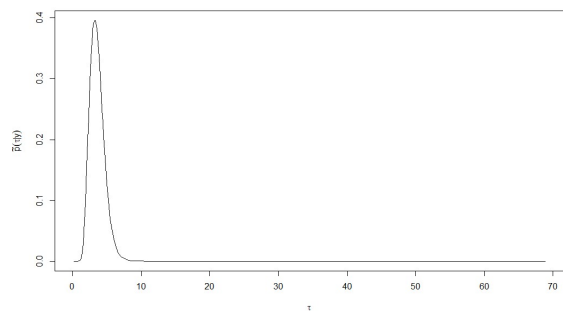


Figure 8. Posterior marginal of the precision of the random effects  $\tau$  with INLA

3.1.6. The Bayesian INLA quadratic polynomial regression model with the Gaussian methods and the grid strategy

Time used: Pre = 1.17, Running = 0.148, Post = 0.0592, Total = 1.37

```
Call:c("inla(formula = formula, family = \"gaussian\", data = datapol, \" \" control.inla = list(strategy = \"gaussian\", int.strategy = \"grid\")") )
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	234.832	77.013	82.253	234.849	387.214	234.887	0
height	4.182	1.414	-6.982	-4.182	-1.381	-4.183	0
height <sup>2</sup>	0.019	0.006	0.006	0.019	0.032	0.019	0

The model has no random effects

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for	3.60	1.06	1.83	3.49	5.95	3.29

the Gaussian observations

Expected number of effective parameters(stdev): 3.00(0.001)

Number of equivalent re plicates: 8.00

Marginal log-Likelihood: -44.08

Three densities of the parameters intercept  $\beta_0$ , coefficient  $\beta_1$  and  $\beta_2$  and hyperparameter  $\tau$  using the integration Gaussian and the grid strategy in Figure 9-12.

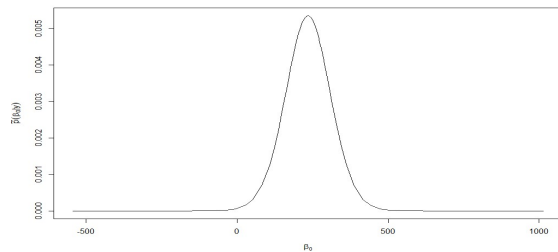


Figure 9. Posterior marginal of the intercept  $\beta_0$  with INLA

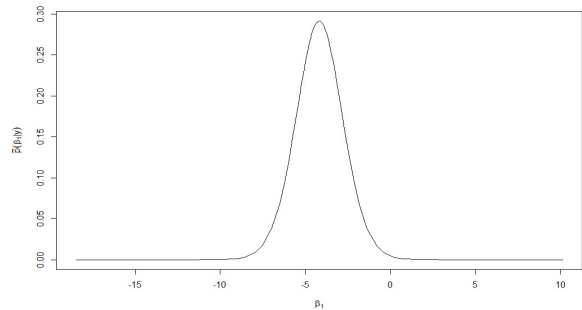


Figure 10. Posterior marginal of the coefficient  $\beta_1$  with INLA

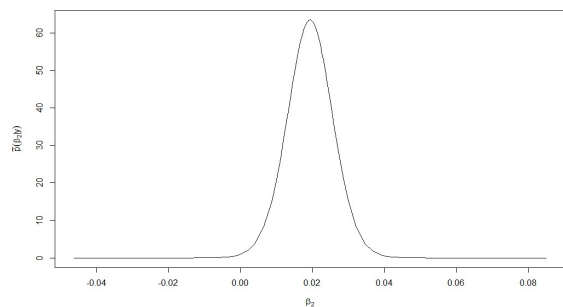


Figure 11. Posterior marginal of the coefficient  $\beta_2$  with INLA

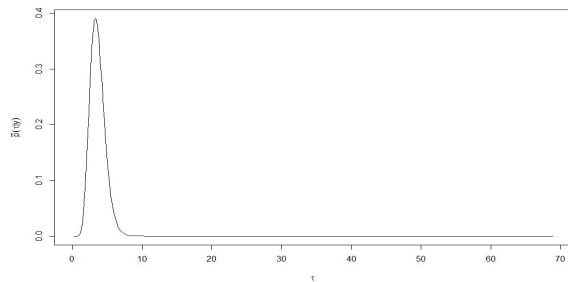


Figure 12. Posterior marginal of the precision of the random effects  $\tau$  with INLA

The estimates of the posterior marginal intercept  $\beta_0$ , coefficient  $\beta_1$  and  $\beta_2$  with Gaussian method and different strategy produce similar. The estimates of the posterior of random effects the grid strategy different with CCD and empirical Bayes strategy.

#### 4. CONCLUSION

Based problem formulation and discussion can be conclusion. Scatter diagram on data obtained two polynomial regression model possible with least square method. The quadratic polynomial regression model is model with height variable and height<sup>2</sup> significant. Posterior marginal Then quadratic Polynomial Regression Model via Bayesian INLA with high probability is height and height<sup>2</sup>. The posterior marginal intercept  $\beta_0$ , coefficient  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  with Gaussian method and different strategy produce similar. The estimates of the posterior of random effects  $\tau$  the grid strategy different with CCD and empirical Bayes strategy.

#### REFERENCES

- [1] N. Draper and H. Smith, *Applied Regression Analysis*, New York: Wiley, 1998.
- [2] V. Gomes and Rubio, *Bayesian Inference with INLA*, CRC Press, Spain, 2020.
- [3] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics*, 1978.
- [4] E. O. Okuto, *Bayesian Spatial and Spatiotemporal Modelling (Applied to precipitation dataset)*, Doctoral dissertation, University of Nairobi, Nairobi, 2013.
- [5] H. Rue and Martino, *Approximation Bayesian Inference for Latent Gaussian Models Using INLA*, Norwegian, 2009.
- [6] J. Faraway, *Linear Models with R*, USA: Chapman & Hall, 2009.
- [7] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Boca Raton, Florida: Chapman & Hall, 1996.
- [8] M. Blangiardo and M. Cameletti, *Spatial and Spatio-temporal Bayesian Models with R-INLA*, United Kingdom: John Wiley & Sons, Ltd, 2015.
- [9] G. E. P. Box and N. R. Draper, *Empirical Model Building and Response Surfaces*, New York, NY: John Wiley & Sons, 1987.