DOI: 10.15415/mjis.2015.32016

Bayesian Method in Linear Model and Constant Time Series Model Using Non-Informative Prior Under Phenology

VIJAY KUMAR PANDEY[1], RAJEEV PANDEY[2] AND MAYANK TRIVEDI[3]

[1]Research Scholar in Department of Statistics, University of Lucknow, India-226007
[2]Professor, Department of Statistics, University of Lucknow, India-226007
[3]Research Scholar in Department of Statistics, University of Lucknow, India-226007

[*]**Email: vijay.pandey550@gmail.com**

Received: 15 December 2014| Revised: 27 February 2015| Accepted: 4 March 2015

Published online: March 30, 2015

Abstract: Climate Change is very recent topic at global level for discussion for all of us. Phenology is one of the main bio- indicators to track climate change effects on ecosystem. The present study is devoted to derive results of coherent interest in the field of phenology from Bayesian point of view. In this paper we have developed the phenological probability models using linear model and constant time series model. The comparison of both the models has also been done using the concept of residual sum of square and Bayes' factor.

Keywords: Bayesian Analysis, Linear Model,Constant Time Series Model, Phenology, Non-Informative Prior.

1. INTRODUCTION

Phenology is a field of research which studies the annual rhythm of biological phenomena mainly related to climate. Centuries ago, especially for agricultural purposes, phenological knowledge improved the understanding of the variation in cycle events. Observations of phenological phases are probably the simplest way to track changes in the ecology of species in response to climate change. The use of phenological data as bio-indicator for climate variations and global change is based on the well known relationship between climate parameters and the onset of phenological phases.

According to **[5]** phenophases are regarded as an integrating climatological measurement responding to many meteorological and environmental factors

Mathematical Journal of Interdisciplinary Sciences
Vol. 3, No. 2,
March 2015
pp. 183–191

CHITKARA UNIVERSITY

such as light, photoperiod, temperature, precipitation, humidity, wind and others although their particular influence differs.

Phenological models predict time of events in an organism's development to find the information about climate change and its impacts on phenological events, it is necessary to develop the statistical tools for analysing the data. [3] used a Bayesian method to investigate phenological changes. Climate change detection employing non-parametric Bayesian function estimation is especially useful for studies of climate change impacts in natural systems where conditions are prescribed to change. [4] was among the first to explore the Bayesian approach in climate change detection. He also pointed out the need for the estimation of model uncertainities.[1] used the Bayesian analysis for climate change assesment.

The Bayesian analysis of the time series follows the methodology introduced by [2]. [2] proposed the three different models to describe the phenological time series data. The constant model, linear model and one change point model. The constant model represents the hypothesis of no change with a functional behavior constant in time and an associated zero rate of change. The linear model assumes a linear change of the observed phenomenon with associated constant rate of change. The change point model allows for a time varying rate of change and thus is an indicator for non-linear changes. The present paper deals with the linear model to analyse the phenological data.

The Bayesian approach to statistical inference offers an extremely useful tool for model comparisons called Bayes' factors which has been proposed by [3]. Bayes' factors can be used to compare models with different blocks of covariates, models with different functional forms, and can be used with any number of plausible models. As with all quantities in the Bayesian approach, Bayes' factors provide the posterior probability that model M_1 is the true data generating process compared with model M_2.

2. PROPOSED LINEAR MODEL USING NONINFORMATIVE PRIOR IN PHENOLOGY

The linear model assumes a linear change of the observed phenomenon with an associated constant rate of change. This model assumes a linear trend in the data leaving open the question whether we expect a rise or a fall as a function of observation year. The model equation for this case becomes,

$$d_i - f_i \frac{x_N - x_i}{x_N - x_1} - f_N \frac{x_i - x_1}{x_N - x_1} = \in_i \tag{1}$$

Where x_1, x_N are the first and last year of observations and f_1, f_N design functional values that specify a linear trend between x_1 and x_N but are of course unknown. Equation (1) suggests going over to matrix notation:

$$\vec{d} - A\vec{f} = \vec{\in} \tag{2}$$

Here $\in$ is the error term, Nis the number of observations and A is the N rows and two column matrix. If expected value of $\{\in_i\}$ is assumed to be zero with known variance as σ^2, then according to principle of least square theory the likelihood becomes,

$$P(\vec{d} / \vec{x}, \sigma, \vec{f}, l, I) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \cdot \exp\left\{-\frac{1}{2\sigma^2}\left(\vec{d} - A\vec{f}\right)^{'}\left(d - A\vec{f}\right)\right\} \tag{3}$$

The symbol added ^{l}in the condition reminds that we are treating a linear trend model. Let the prior distribution $P\left(\vec{f} / I\right)$ (on $\vec{f}$ is the chosen (weakly informative) to be constant over the range 2γ that is,

$$P\left(\vec{f} / I\right) = \frac{1}{2\gamma} \tag{4}$$

The range γ can be estimated from the variance of the data. A possible kdimensional generalization of (4) would be $\left(\frac{1}{2\gamma}\right)^k$. A more efficient choice of the prior volume is, however obtained if we replace the hypercube by a hyper -sphere V_s (k, γ):

$$Vs(k, \gamma) = \frac{\gamma^k \left(\sqrt{\pi}\right)^k}{\Gamma\left(\frac{k+2}{2}\right)} = P\left(\vec{f} / \gamma, k, I\right) \tag{5}$$

A similar non- informative choice is made for P(σ/I).Since σ is a scale parameter for the difference $|d_i - f_i|$, we choose a normalizedform of JEFFREYS' prior:

$$\text{P}(\sigma/\beta, \text{I}) = \frac{1}{2\ln\beta}, \frac{1}{\sigma}\frac{1}{\beta} < \sigma < \beta \tag{6}$$

From the marginalization theorem, we obtain

$$P(\vec{d} / \vec{x}, l, I) = \int d\vec{f} d\sigma . P(\vec{d}, \vec{f}, \sigma / \vec{x}, l, I) \tag{7}$$

This equation is an identity. The integral in (7) canbe expanded using the product rule:

$$P(\vec{d} / \vec{x}, l, I) = \int P(\vec{f} / I).P(\sigma / I).P(\vec{d} / \vec{x}, \sigma, \vec{f}, I) d\vec{f}\ d\sigma \tag{8}$$

$$\begin{aligned} \text{Let } \phi &= \left(\vec{d} - A\vec{f}\right)^T \left(\vec{d} - A\vec{f}\right) \\ &= \left(\vec{d} - A\left(\vec{f} - f_0 + f_0\right)\right)^T \left(\vec{d} - A\left(\vec{f} - f_0 + f_0\right)\right) \\ &= \left(\vec{f} - \vec{f_0}\right)^T Q\left(\vec{f} - \vec{f_0}\right) + R \end{aligned} \tag{9}$$

Where R is residue of the proposed model.
Hence from (1) and (7) we get,

$$\begin{aligned} P\left(d / x, l, I\right) &= \frac{1}{(2\pi)^{\frac{N}{2}}} . \frac{1}{V_s(2,\gamma)} . \frac{1}{2\ln\beta} \int_0^\infty \frac{d\sigma}{\sigma} \frac{1}{\sigma^2} \exp\left\{-\frac{R}{2\sigma^2}\right\} . \frac{2\pi\sigma^2}{\sqrt{\det(Q)}} \\ &= \frac{1}{2V_S(2,\gamma)} . \frac{1}{2\ln\beta} . \left(\frac{1}{\pi}\right)^{\frac{N-2}{2}} . \frac{1}{\sqrt{\det(Q)}} \frac{\Gamma\left(\frac{N-2}{2}\right)}{(\mathrm{R})^{\frac{N-2}{2}}} \end{aligned} \tag{10}$$

The matrix Q and residue R are unknown. These quantities follows from a comparison of coefficients in $\vec{f}$ of the two equivalent forms of ϕ in (9) and hence
$Q = A^T A$

Therefore, residual sum of square of the model (10) is given by,

$$\mathrm{R} = \vec{d}\ ^T \vec{d} - \vec{d}^T A (A^T A)^{-1} A^T \vec{d} \tag{11}$$

The expression for the residue R can be simplified considerably if we employ singular value decomposition on the matrix A.

$$\text{Let us Assume } A = \Sigma_i \lambda_i \vec{U}_i \vec{V}_i^T \tag{12}$$

yields for matrix Q:

$$Q = A^T A = \sum_i \sum_k \lambda_i \lambda_k \vec{V}_i \vec{U}_i^T \vec{U}_K \vec{V}_k^T = \sum_k \lambda_k^2 V_K V_k^T \tag{13}$$

The last equality follows from the fact that $\{\vec{U}_i\}$ and the $\{\vec{V}_i\}$ each form an orthogonal normalized vector system. Accordingly we obtain from (12) and (13) as

$$\mathrm{Det}\,(\mathrm{Q}) = \prod_k \lambda_k^2 \tag{14}$$

$$\mathrm{A}Q^{-1}A^T = \sum_k \vec{U}_K \vec{U}_k^T \tag{15}$$

$$R = \vec{d}^T \left\{1 - \sum_K \vec{U}_K \vec{U}_k^T\right\} \vec{d} \tag{16}$$

Now if we take constant time series model, we have

$$d_i - f = \in_i \ \forall i = 1,2,3\ldots.\mathrm{N} \tag{17}$$

Using the above rules we get

$$\mathrm{P}(\vec{d}\,/\,\vec{x},C,I) = \frac{1}{2}\left(\frac{1}{\pi}\right)^{\frac{N-1}{2}} \cdot \frac{1}{2\gamma}\frac{1}{2\ln\beta} \cdot \frac{\Gamma\left(\frac{N-1}{2}\right)}{\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}} \cdot \frac{1}{\sqrt{N}} \tag{18}$$

To check the reliability between the constant time series model and linear model we derive the Bayes' factor in the favour of linear model as given by,

$$B = \frac{p\left(\vec{d}\,/\,\vec{x},l,I\right)}{p\left(\vec{d}\,/\,\vec{x},C,I\right)} = \frac{\frac{1}{2V_S(2/\gamma} \cdot \frac{1}{2\ln\beta} \cdot \left(\frac{1}{\pi}\right)^{\frac{N-2}{2}}}{\frac{1}{2}\left(\frac{1}{\pi}\right)^{\frac{N-1}{2}} \cdot \frac{1}{2\gamma}\frac{1}{2\ln\beta}} \cdot \frac{\frac{1 \quad \Gamma\left(\frac{N-2}{2}\right)}{\sqrt{\det(Q)_{(R)}}\,\frac{N-2}{2}}}{\frac{\Gamma\left(\frac{N-1}{2}\right)}{\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}} \cdot \frac{1}{\sqrt{N}}} \tag{19}$$

Using equation (5) theBayes' factor becomes

$$B=\frac{2}{\sqrt{\pi}}\frac{1}{\gamma}.\frac{\frac{1}{\sqrt{\det(Q)}}.\frac{\Gamma\left(\frac{N-2}{2}\right)}{(R)^{\frac{N-2}{2}}}}{\frac{\Gamma\left(\frac{N-1}{2}\right)}{\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}}.\frac{1}{\sqrt{N}}}=\frac{2\sqrt{N}\left(N\overline{\Delta d^2}\right)^{\frac{N-1}{2}}.\Gamma\left(\frac{N-2}{2}\right)}{\sqrt{\pi}\gamma.\sqrt{\det(Q)}.(R)^{\frac{N-2}{2}}.\Gamma\left(\frac{N-1}{2}\right)} \quad (20)$$

The residual sum of square of the model(17) can be derived as,

$$E=\sum_{i=1}^{N}\epsilon_i^2=\sum_{i=1}^{N}\left(d_i-f\right)^2$$

For the minimisation of E, we shall differentiate E with respect to f. Hence, $\frac{\partial E}{\partial f}=0 \Rightarrow f=\frac{1}{N}\sum_{i=1}^{N}d_i=\bar{d}$. Therefore, the residual sum of square of the proposed constant time series model is given by

Table 1

Years	Average temperature of the years	Production of rice (in tones)
2000	26.26	21.13
2001	26.32	20.576
2002	25.47	20.022
2003	26.1	19.468
2004	24.16	18.914
2005	24.5	18.360
2006	25.112	18.935
2007	25.87	19.510
2008	25.85	20.085
2009	24.9	20.660

$$R = \sum_{i=1}^{N} \left(d_i - \bar{d}\right)^2 = N\overline{\Delta d^2}$$

where, $\overline{\Delta d^2} = \frac{1}{N}\sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2$.

3. NUMERICAL ILLUSTRATION

To get the residual sum of square and Bayes' factor of linear and constant time series model,we use the following table of 10 years average temperature data obtained from Narendra Deva University of Agriculture and Technology, Faizabad (India) as :

The temperature data is,

$$\vec{d} = \begin{pmatrix} 26.26 \\ 26.32 \\ 25.47 \\ 26.10 \\ 24.50 \\ 25.11 \\ 25.87 \\ 25.85 \\ 24.90 \end{pmatrix}$$

$$\text{and } A^T = \begin{pmatrix} 1 & 8/9 & 7/9 & 6/9 & 5/9 & 4/9 & 3/9 & 2/9 & 1/9 & 0 \\ 0 & 1/9 & 2/9 & 3/9 & 4/9 & 5/9 & 6/9 & 7/9 & 8/9 & 1 \end{pmatrix}$$

Hence residual sum of square of the linear model is

$$R = d^T\left[I - A\left(A^T A\right)^{-1} A^T\right]d = 4.364696$$

and $P(\vec{d}/\vec{x}, l, I) = 0.000052$.

Similarly, we have for constant time series model the residual sum of square R=5.16 and $P(\vec{d}/\vec{x}, C, I) = 0.0000012$.

The Bayes' factor in favour of linear model is

$$B = \frac{p(\vec{d}/\vec{x}, l, I)}{p(\vec{d}/\vec{x}, C, I)} = 43.33$$

Table 2

	Residual Sum of Squares	Efficiency	Model Probabilities	Bayes' Factor in Favourof Linear Model
Constant Time Series Model	5.16	19.38%	$P(\vec{d} / \vec{x}, C, I) = 0.0000012$	–
Linear Time SeriesModel	4.364969	22.91%	$P(\vec{d} / \vec{x}, l, I) = 0.000052$	43.33

CONCLUSION

The comparision of the of the Constant Time Series model and Linear model in a tabular form is given below in terms of Residual Sum of Squares, Efficiency and Bayes' Factors:

According to above table 2, we observe that using the Linear Model we can do better prediction of those phenological events which satisfiy the conditions of Linear Model. We can also check the accuracy of the linear model by calculating the residual sum of square. Since residual sum of square of the Linear Model is less than the residual sum of square of the Constant Time Series Model, the efficiency of Linear Model is more than the Constant Time Series Model and as the Bayes' factor in favour of Linear Model is more than unity, therefore Linear Model gives a better fit of the data than Constant Time Series Model.

ACKNOWLEDGEMENT

We are thankful to Prof. Rohit Pandey, Narendra Deva University of Agriculture and Technology, Faizabad (India) for providing phenological data as shown in table1.

REFERENCES

[1] Berliner, L.M., (2000): Bayesian Climate Change Assesment, Jour. Climate, [3805-3820, 13].

[2] Dose, V. and Menzel, A. (2004): Bayesian Analysis of Climate Change Impacts in Phenology, Global Change Biol., [259-272, 17]. http://dx.doi.org/10.1111/j.1529-8817.2003.00731.x

[3] Kass, R.E. and Raftery, A.E.(1995):Bayes Factor and Model Uncertainity, J StatistAssoc. [773-795, 90].

[4] Leroy, S.S. (1998): Detecting Climate Signals: Some Bayesian Aspects, Jour. of Climate [640-651, 11].
http://dx.doi.org/10.1175/1520-0442(1998)011<0640:DCSSBA>2.0.CO;2

[5] Menzel, A. (2002): Phenology: Its Importance to the Global Change Community, Jour. Climatic Change [379-385, 54].