Transactions of the Nigerian Association of Mathematical Physics
***Volume 3, (January, 2017*), pp 169 – 176**
© Trans. of NAMP

Hierarchically Structured Model for Predicting the Distribution of Insurance Risk

Agwuegbo S.O.N[1], Onugha E.E[2] and Ezepue P.O[3]

[1]Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria
[2]Department of Mathematics, Federal University of Technology, Owerri, Nigeria.

Abstract

Hierarchically structured models provide the basis for an entirely different approach in data analysis. It originated in an attempt at using conceptual model and has been applied to a wide variety of statistical problems. In this paper, emphasis is on exploring hierarchically structured modelling approach as an alternative model for predicting the distribution of insurance risks. The cardinal approach of the hierarchically structured model is to impose structures into state spaces and to work with subsets of states by partitioning the original state spaces into hierarchically organized subsets. Dynamics in the state space were approximated with advantage by random walk on trees. The result of the study on insurance risk shows that hierarchically structured model allows for the realization of estimated probability distribution which determines uniquely how the system evolves with time and provides a generalization of an ARIMA (1, 0, 0) models.

Keywords: State space, Dynamics, Binary Trees, Random Walk, Markov Processes

1.0 Introduction

Hierarchical structured models have gained wide spread use in statistics during the last few decades [1], and have proved to be useful tools for modeling dynamic behaviour in large dimensional state space [2,3] or exploring structures in complicated data [4].Hierarchical structured state space models are concerned with conceptual information modeling and it captures on how different real world object behaviors may be vied. Hierarchical Modelsare central to many current analysis of environmental processes [4-7],and also for predicting the spread of ecological processes [8].

The notion of hierarchically structured model is extended to insurance claim data in this paper by looking at the profile of insurance risks. The profile conventionally assumed in insurance is usually that there is a high frequency of low severity incidents and a low frequency of high severity incidents. The profile revealed basically that all estimates of the expected future claim costs involve a compounding of two functions: the number of claim in a given period and the amount of each claim. The pattern may be viewed, as the distributions of claims over possible type or categories over time. The profile of risks is a way of representing the frequency and severity of risks as leaves of a tree. This characteristics pattern has been used as the basis for algorithms for identifying the claims distributions.

More generally the tree structure (that is, the set of contexts) is known and the problem is on how to estimate the conditional probabilities at the leaves. The context (the path from the root of the tree) are possible and desirable, where there is no unit difference between successive members, but for this presentation a context is taken to be a suffix of the data string. In addition, the contexts may provide insight and define regions of state space which are interesting.

Dynamics in this study is introduced by considering a two level binary tree. Essentially the use of a binary tree structure provides a convenient way of representing the contexts as leaves of a tree. In addition, the contexts may provide insight and define regions of state space which are interesting.

2.0 Methodology: Basic Hierarchical Approach

The formal ideas of hierarchical modeling arise from simple probability rules. The idea follow those in [4], and is based on the simple fact from probability that the joint distributions of a collection of random variables can be decomposed into a

Corresponding author: Agwuegbo S.O.N, E-mail: Agwuegbo_son@yahoo.com, Tel.: +2348038004887

series of conditional models. The major advantage of modeling the conditional distribution of the data is due to the fact that substantial simplifications in model form are possible. That is, if $X,Y.Z$ are random variables, then the joint distributions can be written in terms of factorizations as

$$(X,Y,Z) = (Z/Y,X)(Y/X)(X)\ . \qquad (1)$$

This simple formula is the crux of hierarchical thinking. However, it is often much easier to specify the distribution of the relevant conditional models by conditioning the process at the present time given the past. In this case, the product of a series of relatively simple conditional models leads to a joint distribution that can be quite involving [4].

For complicated processes in the presence of data, the idea is to approach the problem as a conceptual information modeling. The hierarchical organization of subsets can be visualized as an inverted tree [2, 3, 9]. It is advantageous and useful to impose structures into state spaces and to work with subsets of states by partitioning the original state spaces into hierarchically organized subsets, where a subset on a level is further partitioned into a set of smaller subsets on the next level [2, 3]. A realization or time history on the structure of the state spaces is that not all of these numerous configurations are equally likely in equilibrium. Presumably, since not all the states are equally accessible from the initial state in a given amount of time, a structural information or constraints is build into the model state spaces in order to classify states into clusters or subsets corresponding to some measure of similarity or distances. As the measure of distance is changed, a cluster can be further subdivided into a set of sub clusters, so that clusters and sub clusters, become partially nested, or form nodes of trees. Dynamics of symbol strings is achieved by the use of binary tree as shown in Fig. 1.0 below

Claim Process

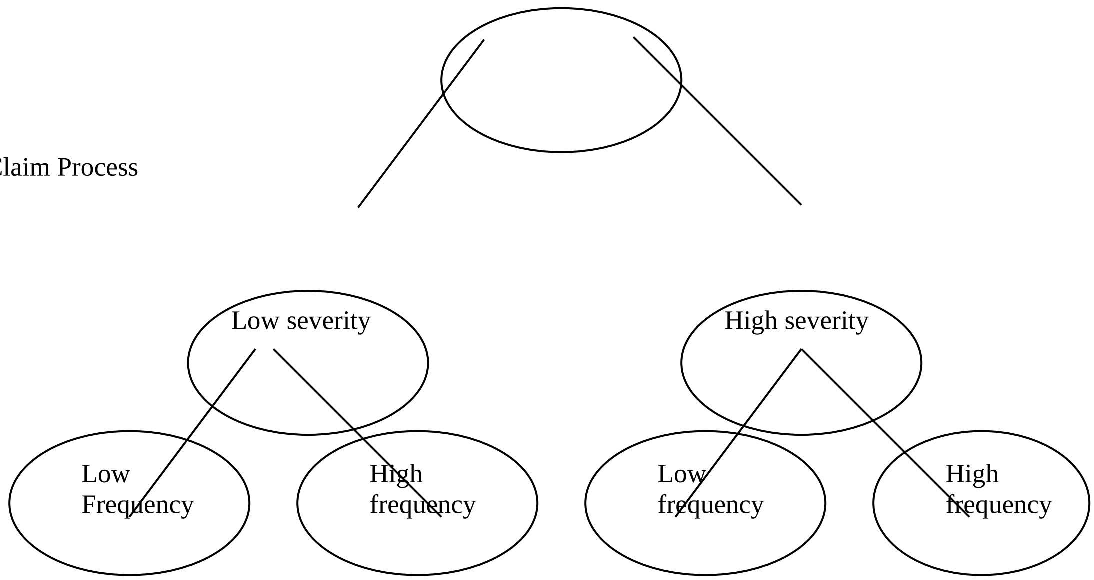

Figure 1: A simple Binary Tree for Claim Process

Essentially the use of a binary tree structure provides a convenient way of representing the claim process as leaves of a tree. To enumerate all possible system state an appropriate state description on the claim data is to classify the distribution into the following states:

State 1	Low frequency	Low severity	S_1
State 2	High frequency	Low severity	S_2
State 3	Low frequency	High severity	S_3
State 4	High frequency	High severity	S_4

The classification formed was compared by order of fineness (i.e. by inclusion). The order of fineness between the classes corresponds to the order relating to the node of a rooted tree. It is assumed that probabilities are independent when conditioned on context. The binary tree structure reproduced a system having finite memory $S_0, S_1, ...S_n$ and the system defined as a causal model. By letting X_t be the claim amount at time $t \in T,$ the collection of random variables $S = \{S_t, t = 1,2,.3,4\}$ is a stochastic process in discrete time while the claim amounts $X_t, t > 0$ have a continuous range and continuous state space.

2.1 Stochastic Processes for Insurance Risk

Instead of thinking of the process $X_t, t \in T$ as a sequence of random variable, one can visualize it as the random walks on a tree defined by the realization or sample path of the process denotedas $S_t(\omega)$. Random walks are processes with independent increments and processes with independent increments are Markov processes. The study of random walks is the study of sum of random variables and the general practice in defining such structure of state space is by partitioning the observation into organized subsets. The defining condition of the random walk $\{S_t, t = 0,1,2,.......\}$ starting at zero is obtained by cumulatively summing or integrating independent identically distributed (i.i.d) random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = X_1 + X_2 + ...+ X_t \quad \text{for } t = 1,2,........ \tag{1}$$

Where $\{X_t\}$ is i.i.d noise. If $\{X_t\}$ is the binary process, then $\{S_t, t = 0,1,2,.......\}$ is called a symmetric random walk. $\{S_t\}$ can be considered a Gaussian random walk by assuming identically, independently and normally distributed increments and is defined as

$$S_t = S_0 + \sum_{k=1}^{t} Z_k \qquad t = 1,2,....$$

Or

$$S_t = \sum_{k=1}^{t} Z_k \qquad t = 1,2,.... \tag{2}$$

using the fact that

$$S_t = Z_1 + Z_2 +,...................+ Z_t \tag{3}$$

and that the Z's are jointly independent the mean and variance of the random walk can be computed as

$$E(S_t) = E[\sum_{k=1}^{t} Z_k]$$

$$= \sum_{k=1}^{t} E(Z_k), \quad \text{since} E(Z_k) = 0$$

$$= 0 \tag{4}$$

and

$$Var(S_t) = E\left[\sum_{k=1}^{t} Z_k^2\right]$$

$$= \sum_{k=1}^{t} E(Z_k^2)$$

$$= \sum_{k=1}^{t} \sigma^2$$

$$= t\sigma^2 \tag{5}$$

In order to handle these processes within the framework of the classical time series analysis, the random walk is a martingale sequences. Since S_t is an independent increments random sequence defined for $t \geq 0$, then

$$S_t = S_0 + \sum_{k=0}^{t} Z_k \quad t \geq 0 \tag{6}$$

Then S_t in Equation (6) is a martingale, which can be show as

$$E\{(S_t / S_{t-1}, S_{t-2},........,S_0)\} = E\{(\sum_{k=1}^{t} Z_k / S_{t-1}, S_{t-2},........., S_0)\}$$

$$=\{\sum_{k=1}^{t} E(Z_k / S_{t-1}, S_{t-2}, ..., S_0)\}$$

$$= \sum_{k=1}^{t-1} Z_k + E(Z_t)$$

$$= S_{t-1} \qquad (7)$$

Viewing the conditional expectation as an estimate of the future value of the sequence based on the past, then for a martingale this estimate is just the most recent value. The random walk process in our study can be characterized as a series in which successive changes in level are determined by chance, in this case by a random drawing process from an $N(0,1)$ population. For our purpose, a random walk is any stochastic, non stationary process that can be defined by the relation

$$S_t = S_{t-1} + a_t \qquad (-\infty < t < \infty) \qquad ---- \qquad (8)$$

Where (a_t) is a white noise process with mean zero and variance σ^2.

If S_t denotes the position of a particle on the real line at time t, Equation (8) states that the position of the particle at time t is its position at time $t-1$ plus a random displacement. The particle may be said to perform a random walk. Equation (8) is identical with the equation satisfied by an $AR(1)$ process if the autoregressive parameter ϕ is set equal to 1. But the $AR(1)$ process is a well defined. Equation (8) is non stationary and can be written as

$$S_t - S_{t-1} = a_t \qquad (-\infty < t < \infty) \qquad (9)$$

If one define the variables Z_t by $Z_t = S_t - S_{t-1}$, then Equation (9) is the first differences of a martingale and is called a martingale difference (MD) written as

$$Z_t = a_t \qquad (-\infty < t < \infty) \qquad (10)$$

Note that (10) is a martingale difference. Martingale difference has known linear second order properties

2.2 The Difference Operator

Using the backward shift operator B, one may rewrite Equation (9)through recursive substitution as

$$(1-B)S_t = a_t$$

$$S_t = (1-B)^{-1} a_t$$

$$S_t = a_t + a_{t-1} + ...a_1 \qquad (11)$$

In order to handle this process within the framework of the classical time series analysis, the observed claim process must be transformed by differencing the process in order to get a stationary process. The transform process is then

$$Z_t = (1-B)^d S_t \qquad (12)$$

Such a model is called an integrated model because the stationary model that is fitted to the difference data has to be summed or integrated to provide a model for the original non stationary data. Describing the dth difference of S_t is said to be an $ARIMA(p,d,q)$ process. In practice, the first differencing is often formal to be adequate to make a series stationary. It may turn out that there is more than one plausible model and based on the use of Akaike information criterion(AIC), the goodness of fit of different models is to be compared by assuming that the data are normally distributed. The AIC is defined as

$$AIC = -2 \max imized \log - likelihood + 2n$$

$$\approx T \; In\hat{\sigma}^2 + 2n + const, \qquad (13)$$

where T is the length of the observed series after any differencing , n is the number of fitted parameters and $\hat{\sigma}^2$ is the estimated white noise variance.The model with the smallest value of the AIC is judged to be the most appropriate [10].

3.0 Results

The idea of dynamics on state spaces that are hierarchically structured is extended in our study to insurance risks as in Table 1.

Table 1: Distribution of Insurance Claims

Jan: 4469654, 991698, 1243344, 513000, 522473, 3800000, 744538, 610536, 900000, 573750, 2025000, 570978, 542400, 2592000, 574536, 705682, 719059, 933038, 696173, 665766, 581487, 700000, 750000, 661000
Feb: 684068, 750988, 3401510, 1389917, 1113524, 623396, 3036377, 2334479, 2208789, 1107032, 1050000, 691572, 612716, 537750, 1072512, 1312500, 838080, 12831900, 10159375, 1516854, 500000, 7000000, 1682558, 1012500
March: 1700000, 899965, 513000, 3000000, 1066400, 5772690, 629803, 819500, 1923000, 720000, 1624500, 2183298, 1133930, 1680120, 1938070, 1163191
April: 1091250, 1327784, 3066729, 17093431, 645188, 612000, 586528, 6277603, 864374, 3278926, 1260000, 738000
May: 678000, 577776, 869660, 501785, 777600, 965037, 1260000, 1304394, 1800000, 1083598, 1177808, 1310143, 1298097, 1968740, 760089, 2700000, 4302997, 1224172, 643204, 1344823, 2588500, 3165761.
June: 532010, 1495237, 742808, 724693, 1176641, 7000000, 2109000, 810000, 699510, 531644, 966845, 2618136, 1230618, 3162224, 1081947, 666750, 1500000, 652500, 529376, 4850000, 878500, 520600, 3298464, 1980000, 679115, 632591, 540000.
July: 627838, 723050, 504900, 2263523, 606000, 1729917, 1950000, 704660, 1000000, 2479351, 1417898, 500000, 810000, 1408603, 1256584, 1620000, 540000, 522969, 717039, 982816, 4250000, 700000.
Aug: 760000, 2353302, 546826, 531451, 823125, 515735, 1364993, 1030228, 1393273, 3244220, 1080000, 1044000, 24400000, 1075284, 1070244, 1197000, 995663, 761846, 4254817, 1162800, 24579269.
Sept: 1338750, 1338750, 6300000, 3179432, 800650, 891950, 4454095, 2307436, 559558, 559545, 2063844, 2685354, 1246300, 1245983, 4884223, 857719, 566820, 631125, 1648438, 832733, 3254900, 2061216, 1085797.
Oct: 5235988, 688500, 1411242, 2607147, 1530000, 800000, 1620000, 1067600, 826350, 1982973, 576000, 1381026, 6697192, 3265331, 3222164, 1238226, 828800, 1657500, 14552619, 1121850, 842387, 728946, 3734997, 1341743, 546950, 1134488, 544266, 1351500, 562002, 1851600, 1823018, 3054268.
Nov: 1048478, 1625570, 3886258, 3910305, 1313125, 2900000, 600750, 800000, 1026667, 14490580, 563170, 705093, 1792500, 2153730, 2920459, 643357, 7435587, 542500, 565213, 5178084, 5161160, 2207540, 513359, 746971, 1882850, 2089548, 680400, 553248, 914973, 1080000, 1346386, 27311939.
Dec: 500000, 1042321, 765000, 1344823, 746971, 1339595, 3134790, 540510, 661500, 671700, 1768000, 3587542, 1051200, 1303154, 1298996, 544000, 1744652, 3017240, 3865360, 711461, 992062, 515800, 870795, 665000, 675000, 1080000, 3451391, 524846.

By using a binomial process, the study analyzed the distribution of claims as a random walk model. Examination of the time plot of Table 1 revealed greater variability of claims as shown in Fig 2.

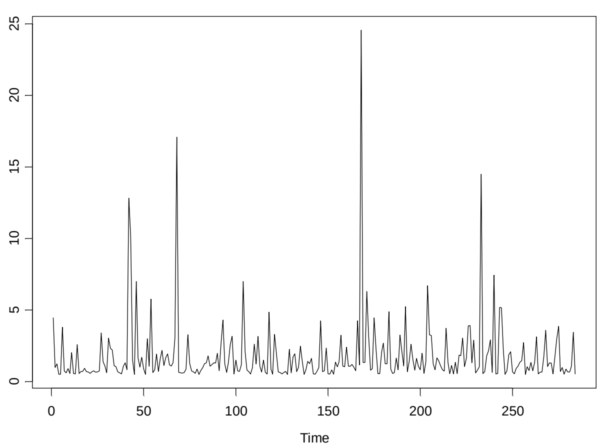

Fig. 2: Original series for the claim distribution

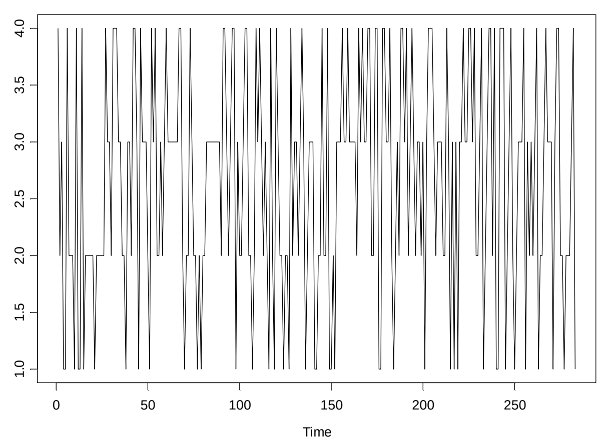

Fig 3: Realization of claims using state space hierarchically structured model

With the use of hierarchical or tree structure, claims in Table 1 were partitioned in one of four possible states as in and the distribution as shown in Table 2.

Table 2: Distribution of the state spaces

4, 3, 2, 1, 1, 4, 2, 2, 2, 1, 4, 1, 1, 4, 1, 2, 2, 2, 2, 2, 1, 2, 2,2, 2, 2, 4, 3, 3, 2, 4, 4, 4, 3, 3, 2, 2, 1, 3, 3, 2, 4, 4, 3, 1, 4, 3, 3, 3, 2, 1, 4, 3, 4, 2, 2, 3, 2, 3, 4, 3, 3, 3, 3, 3, 3, 4, 4, 2, 2, 1, 2, 2, 4, 3, 2, 2, 1, 2, 1, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 2, 4, 4, 3, 2, 3, 4, 4, 1, 3, 2, 2, 3, 4, 4, 2, 2, 1, 2, 4, 3, 4, 3, 2, 3, 2, 1, 4, 2, 1, 4, 3, 2, 2, 1, 2, 2, 1, 4, 2, 3, 3, 2, 3, 4, 3, 1, 2, 3, 3, 3, 1, 1, 2, 2, 4, 2, 2, 4, 1, 1, 2, 1, 3, 3, 3, 4, 3, 3, 4, 3, 3, 3, 3, 2, 4, 3, 4, 3, 3, 4, 4, 2, 2, 4, 4, 1, 1, 4, 4, 3, 3, 4, 2, 1, 2, 3, 2, 4, 4, 3, 4, 2, 3, 4, 3, 2, 3,3, 2, 3, 1, 3, 4, 4, 4, 3, 2, 3, 3, 3, 2, 2, 4, 3, 1, 3, 1, 3, 1, 3, 3, 4, 3, 3, 4, 4, 3, 4, 2, 2, 3, 4, 1, 2, 3, 4, 4, 2, 4, 1, 1, 4, 4, 4, 1, 2, 3, 4, 2, 1, 2, 3, 3, 3, 4, 1, 3, 2, 3, 2, 3, 4, 1, 2, 2, 3, 4, 3, 3, 3, 1, 3, 4, 4, 2, 2, 1, 2, 2, 2, 3, 4, 1

The study modeled the transition of one state to another, and arrived at the probability of each ensemble. If the actual distribution is to be assessed using binomial process over a small time interval, the Wiener process is then the ideal limiting distribution for the random walk model.

The original path of the model as depicted in Fig 1 is more of a chaotic situation. There were lots of irregularities making it difficult to define the path of the process. These sudden structural changes can be reduced by the use of hierarchical structured state space model as in Fig 2. A hierarchical structural model sets out to capture the salient features of a time series.These are often apparent from the nature of the series. Hierarchical structural model contains several disturbance terms and can be reduced as an autoregressive integrated moving average (ARIMA) model. The relationship between the structural and reduced forms gives considerable insight into the potential effectiveness of the different ARIMA models.

The plot in Fig 3 is achieved through the use of the binary tree structure and depicts more of a random walk by using S-PLUS software.

By adopting Box ~ Jenkins ARIMA (p,d,q) model approach to time series analysis, model identification, parameter estimation and diagnostic check were feasible.

The model identification according to Box and Jenkins involved using differencing, acf and pacf. The Box and Jenkins ARIMA models can be shown to be optimal and provides a systematic approach to model selection, utilizing all the information contained in the sample autocorrelaion (ACF) and partial autocorrelation (PACF) functions. The ACF and PACF are meaningful only when applied to stationary series. The study adopted this approach using the S-PLUS package and the result indicated that the given series were generated by a particular ARIMA model. The quality and amount of data available were sufficient for accurate model identification. The sample autocorrelations (ACF) function and the partial autocorrelation (PACF) function for the claim distribution using the first difference is as in Table 3 while the ACF, PACF and AIC for the structured state space models is as shown in Table 4.

Table 3: Claims Distribution Using First Difference

Lag K	ACF	PACF
1	-0.456	-0.456
2	-0.0928	-0.380
3	0.0619	-0.250
4	-0.0231	-0.151
5	-0.049	-0.154
6	-0.018	-0.177
7	0.057	-0.110
8	0.012	-0.040
9	-0.056	-0.068
10	-0.045	-0.168
11	0.099	-0.087
12	0.023	-0.016
13	-0.115	-0.079
14	0.064	-0.047
15	0.028	-0.015
16	-0.036	-0.010
17	0.018	-0.045
18	-0.054	-0.055
19	0.066	-0.011
20	-0.053	-0.056

Table 4: Claims Distribution Using Hierarchical Structured Model

Lag K	ACF	PACF	AIC
1	0.1144	0.1144	0.000
2	-0.0584	-0.0724	0.502
3	-0.0010	0.0168	2.422
4	-0.0415	-0.0488	3.743
5	0.0637	0.0775	4.026
6	0.0141	-0.0100	5.997
7	0.0173	0.0284	7.767
8	0.1354	0.1293	4.965
9	-0.0315	-0.0575	6.021
10	-0.0588	-0.0354	7.664
11	0.0560	0.0722	8.177
12	0.0069	-0.0213	10.048
13	-0.0306	-0.0420	11.544
14	0.0170	0.0255	13.360
15	0.0190	0.0204	15.241
16	-0.0365	-0.0707	15.0813
17	-0.0052	0.0229	17.663
18	-0.0185	-0.0092	19.639
19	-0.0024	-0.0237	21.479
20	0.0172	0.0193	23.372

The hierarchical tree played the role of the universal class formed from all the objects. The idea is that the tree is constructed by regarding components of the patterns as binary valued and associating martingales with nodes of the tree. Many tree based models focus on hierarchical structure. By martingale property, the tree classification collects into a node all configurations that have the same correlation coefficient between any of them. These martingale properties characterize ultrametric distance which can be thought of as Euclidean distance between two nodes. Hence any two random variables in a cluster have the same correlation. Conditional on the random variable associated with this commonly shared node, the random variable are assumed to be independent.

3.0 References

[1] Hober, J. (2000). Hierarchical Models: A current computational perspective. Journal of American Statistical Association 95, 1312 – 1316.

[2] Aoki, M. (1994): New macro economic Modelling Approach: Hierarchical Dynamics and Mean Field Approximation, Jour. Econ. Dynamics and control 18, 865 – 77.

[3] Aoki, M. (1996): New Approaches to Macro-economic Modelling: Evolutionary stochastic Dynamics, Multiple Equilibria and Externalities as Field Effects, Cambridge University Press, Cambridge.

[4] Wikle, C. K. (2003). Hierarchical Bayesian methods for predicting the spread of ecological processes. Ecology 84, 1382 – 1394.

[5] Wikle, C. K. and Hooten, M. B. (2006). Hierarchical Bayesian Spatio temporal models for population spread. In Application of computational statistics in Environmental Sciences: Hierarchical Bayes and MCMC methods, Oxford University Press, New York.

[6] Hooten, M. B.; Wikle, C. K.; Dorazio, M. R. and Royle, J. A. (2006). Hierarchical spatio temporal matrix models for characterizing invasions. Journal of International Biometric Society 63 (2) 558 – 567.

[7] Wikle, C. K.; Berliner, L. M. and Cressie, N. (1998). Hierarchical Bayesian space time models. Jour. Of Envi. Ecol. Stat., 5, 117 – 154.

[8] Haneuse, S. and Wakefield, J. (2007). Hierarchical models for combining Ecological and case – Control Data Journal of the International Biometric Society 63, No.1, 128 – 136.

[9] Wisse, P. (2001). Meta pattern context and Time in information models, Addison – Wesley Boston.

[10] Akaike, H. (1974): A New Look at statistical model identification, I.E.E.e. Transactions on Automatic control, AU – 19, 716 – 722