Research Article

Odds Prediction of Drought Category Using Loglinear Models Based on SPI in the Northeast of Thailand

Wisoot Salee and Veeranun Pongsapukdee*

Department of Statistics, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand
**Corresponding author. E-mail address: veeranun@su.ac.th; veeranun@hotmail.com.*

Received October 8, 2012; Accepted November 13, 2012

Abstract

The prediction of drought category is performed through Loglinear modeling for three dimensional contingency tables. The frequencies of drought categories are evaluated from a 12-month time scale by means of Standardized Precipitation Index (SPI), using the raw data obtained from 19 rain gauge stations in 19 provinces in the Northeast of Thailand. SPI monthly values were computed in a 12 month time scale for the period from January 1962 to December 2009, within 48 years or 576 months in each of 19 rainfall stations. The results show that the selected quasi-association loglinear model is an adequate model and appropriate tool to fit the data from 19 stations (P-value = 0.927). The values of various drought class transitions are estimated in three consecutive months. The predicted odds ratios and the corresponding confidence intervals are evaluated to predict the drought classes' transitions. Even if, most of the results display for more normal drought classes than those of moderate or severity drought classes; however, there are still some areas of investigation that drought could be developed. Therefore, future monitoring of drought watch system is needed to be set the pace in these areas to keep up with them. Furthermore, an appropriate statistical model to predict drought phenomenon in each area is necessary and is probably able to provide more effectiveness in the administrative management of hazard from drought.

Key Words: Odds ratio; Drought class transitions; Three-dimensional loglinear models

Introduction

Thailand has approximately all year warm climate and attractively is an agricultural and rice export country in the world. Drought is particularly important in affecting both agriculture and climate; especially, in the northeast of the country. Prediction or forecasting of drought initiation and ending are both essential for timely and appropriate implementation of measures to cope with drought as well as to make drought warning is possible. Frequency and severity make drought both a hazard and a disaster: a hazard because it is a natural accident of unpredictable occurrence but of recognizable recurrence; a disaster because it corresponds to the failure of the precipitation regime, causing the disruption of the water supply to natural and agriculture ecosystems and other human activities (Moreira et al., 2006; Sharma, 1997). The hazard and disaster nature of droughts makes it important to develop prediction tools, including statistical modeling using loglinear models and Markov chains models through the analysis of Standardized Precipitation Index (SPI) drought class transitions (Moreira et al., 2008; McKee, et al., 1993; Paulo, et al. 2005, 2006b; Sivakuma.and Wilhite, 2002; Nebraska-Lincoin University, 2002). The SPI was developed by McKee et al. (1993) and it is widely used as it allows a reliable comparison between different location and climate for the identification of drought events and to evaluate its severity. It is also often that drought events are becoming more frequent and/ or more severe due to the climate change.

Loglinear models (Nelder, 1974; Agresti, 1990, 2002) are of use primary when at least two variables are response variables. The models specify how the expected count in contingency table depends on levels of the categorical variables for that cell as well as associations and interactions among those variables. For the basic theory or an associated sampling distribution, consider an $(I \times J)$ contingency table that cross-classifies a multinomial sample of n subjects on two categorical responses. The cell probabilities are $\{P_{ij}\}$ and the expected frequencies are $\{nP_{ij}\}$. In addition, loglinear models formulas generally apply with Poisson sampling for independent cell counts. Conditional on the sum n of cell counts, Poisson loglinear models for expected counts become multinomial models for cell probabilities. The main purposes of loglinear modeling are not only the analysis of association and interaction patterns but also its predictions. Loglinear models for expected counts in three-way tables such that for higher dimensions are tested and analyzed in this paper. They are considered to be a more adequate tool to analyze the drought class because they have shown to be adequate to perform a monthly prediction of SPI drought class transitions and/ or contingency data (Agresti, 2002, Paulo, et al. 2005, and Fienberg, 2000). Therefore, the SPI data of drought events with the 12-month time scale are computed and analyzed through the adjusting loglinear models to the probabilities of transitions between the SPI drought classes in the form with 3-dimensional contingency table (Nelder, 1974; Paulo, et al., 2003; Pereira, et al., 2002). Then, odds prediction of drought category using loglinear models based on SPI for the whole northeast of Thailand and that for sub-area, Khonkaen province are performed and investigated to interpret through the odds ratios predictions.

Material and Methods

The general methodology of 3-dimensions loglinear models describes the association patterns among categorical variables which are performed for the cell counts in contingency tables (Agresti, 1990). A Poisson sampling model for counts is usually used for counts in contingency tables and assumes that they are independent Poisson random variable.

Let the 3-dimension criterions, A, B and C with levels i, j, and k (i=1, 2,……, 4) , (j=1,….,4) and (k=1,…, 4), respectively. The categorical variables A, B and C refer to drought classes at month's t-2, t-1 and t, respectively. The levels 1, 2, 3, 4 are associated to the drought classes: 1 to the nondrought class, 2 to the near normal drought class, 3 to the moderate drought class, and 4 to the severe/ extreme drought class. The severity drought classes which are defined by McKee et al., 1993 (McKee, et al.) and modified by Moreira et al., 2006 are shown in Table 1.

Table 1 Drought classes classified by Standardized Precipitation Index (SPI) values

Code	Drought classes	SPI values
1	Non-drought class	SPI $\geq$ 0
2	Near normal drought class	-1 < SPI < 0
3	Moderate drought class	-1.5 < SPI $\leq$ -1
4	Severe/extreme drought class	SPI $\leq$-1.5

The observations frequencies (o_{ijk}) are the response variable for the analysis using loglinear methods and it's data refer to the observed number of transitions between the drought classes i at month's t-2 (A), the drought classes j at month's t-1 (B), and the drought classes k at month's t (C). For example, the observation o_{111} is the number of times that a given site stays for three consecutive months in drought class 1 (or non-drought class). Therefore, the loglinear modeling provides the expected frequencies of drought class transitions corresponding to a 2-month step transition from drought class i to drought class j (i.e., from t-2 to t-1) and from class j to class k (i.e., from t-1 to t). The contingency table for such data is given in Table 2.

Table 2 Three-dimensional contingency table ($I\times J\times K$) or ($4\times 4\times 4$) for two consecutive transitions between drought class i at month t-2 and class j at month t-1 and class k at month t corresponding to a 2-month step transition from drought class i to class j (t-2$\rightarrow$t-1) and from class j to class k (t-1$\rightarrow$t)

Drought class month t – 2	Drought class month t (k=1, 2, 3, 4)															
	1				2				3				4			
	Drought class month t-1 (j=1, 2, 3, 4)				Drought class month t–1 (j=1, 2, 3, 4)				Drought class month t–1 (j=1, 2, 3, 4)				Drought class month t–1 (j=1, 2, 3, 4)			
(i=1,2,3,4)	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	O_{111}	O_{121}	O_{131}	O_{141}	O_{112}	O_{122}	O_{132}	O_{142}	O_{113}	O_{123}	O_{133}	O_{143}	O_{114}	O_{124}	O_{134}	O_{144}
2	O_{211}	O_{221}	O_{231}	O_{241}	O_{212}	O_{222}	O_{232}	O_{242}	O_{213}	O_{223}	O_{233}	O_{243}	O_{214}	O_{224}	O_{234}	O_{244}
3	O_{311}	O_{321}	O_{331}	O_{341}	O_{312}	O_{322}	O_{332}	O_{342}	O_{313}	O_{323}	O_{333}	O_{343}	O_{314}	O_{324}	O_{334}	O_{344}
4	O_{411}	O_{421}	O_{431}	O_{441}	O_{412}	O_{422}	O_{432}	O_{442}	O_{413}	O_{423}	O_{433}	O_{443}	O_{414}	O_{424}	O_{434}	O_{444}

1: Non-drought; 2: Near normal; 3: Moderate; 4: Severe/extreme

For selection of models' goodness-of-fit to the three dimensional contingency table, we obtained that for overall area in northeastern of Thailand, the quasi-association model is the one that has adequately best fitted (P-value = 0.927) the data of which the true model shown in (1).

$$\log(E_{ijk}) = \lambda+\lambda_i^A+\lambda_j^B+\lambda_k^C+\beta u_i v_j+\alpha u_i w_k+\eta v_j w_k+\tau u_i v_j w_k+\delta_{1i} I(i=j)+\delta_{2i} I(i=k)+\delta_{3j} I(j=k)+\delta_{4i} I(i=j=k) \quad (1)$$

where, λ is constant, λ_i^A denotes the ith level for A, λ_j^B denotes the jth level for B, and λ_k^C denotes the kth level for C, and $\beta, \alpha, \eta, \tau$ are model parameters, u_i, v_j, and w_k are, respectively, the ith, jth, kth level scores for A, B, C with i, j, k $\in \{1, 2, 3, 4\}$, and δ_{1i}, δ_{2i}, δ_{4i} are parameters associated to the ith diagonal element of A, and δ_{3j} the jth diagonal element of B. I is the indicator function defined as usual by I=0 if condition is true, and I=1 if condition is false. The expected counts are then obtained from the specified models. However, for each sub-area/province, we simplify (1) into other simpler good fit models; such as, in Khonkaen province, the true model is given by (2) (with goodness-of-fit P-value = 0.854).

$$\log (E_{ijk}) = \lambda+\lambda_i^A+\lambda_j^B+\lambda_k^C+\beta u_i v_j+\eta v_j w_k+\delta_{1i} I(i=j)+\delta_{2i} I(i=k) \qquad (2)$$

For other provinces in the northeast of Thailand we obtain similar models as given in (2) with only some different parameters (not shown here), and all of them are the sub-models of (1). All selected regions in 19 provinces (19 rainfall stations) are in northeast of Thailand, each of which the boundaries are given in Figure 1. All raw data used in this investigation were from the Institute for Water, Thailand. We collected the data and organized them using the SPI monthly values. Then data were computed in a 12 month time scale for the period from January 1962 to December 2009, all together, 576 months or 48 years in each 19 rainfall stations in the northeast of Thailand. All research works are processed and corporate using the program run with SAS® version 9.1.

Several loglinear models for three-dimensional contingency tables were fitted and tested. In loglinear models with Poisson sampling, the error is Poisson random variable. The models' parameter estimation for loglinear models is estimated with the iterative maximum likelihood

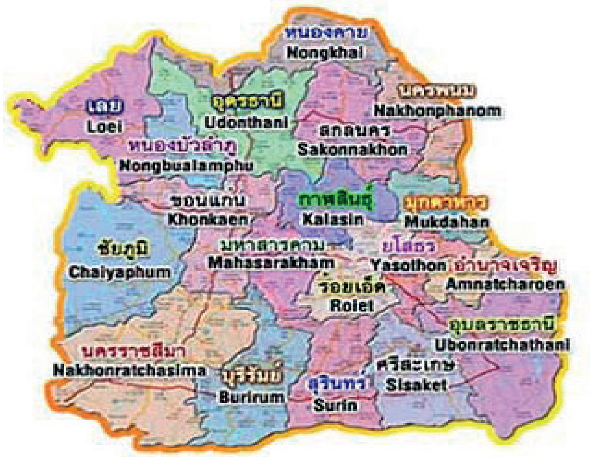

Figure 1 Selected regions in 19 provinces (19 rainfall stations) in northeast of Thailand

method using SAS. The residual deviances that have approximate chi-square distribution with degrees of freedom equal to the number cells in the contingency table minus the number of linearly independent estimated model parameters (Nelder, 1974, Agresti, 1990) were obtained for goodness-of-fit tests.

Results

The results for odds prediction of drought category using loglinear models based on SPI in the northeast of Thailand are processed for both the whole areas in northeast of Thailand and that for a sub-area such that Khonkaen province in the northeast of Thailand.

The Results for Northeast of Thailand

The null hypothesis tested is: the model fits data well. The null hypothesis is not rejected for those models having a residual deviance not exceeding the chi-square quantile for a probability 1 - a = 0.95 with the corresponding degrees of freedom. The backward elimination method (Agresti, 2002) was applied to each complete QA model adjusted to the northeast of Thailand data set to reduce the number of model parameters without significant loss of information to allow the selection of an alternative sub-model eliminating the less significant parameters of the QA model. The observed frequencies of drought class transitions from month t-2 to month t-1 to month t: for the northeast of Thailand SPI 12 month's time scale are presented in Table 3.

Table 3 Observed values of drought class transitions from month t-2 to month t-1 to month t: for the Northeast of Thailand SPI 12 month's time scale

Drought class month t – 2	Drought class month t															
	1				2				3				4			
	Drought class month t–1				Drought class month t–1				Drought class month t–1				Drought class month t-1			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Observed values																
1	216	3	0	0	21	23	0	0	0	1	0	0	0	3	3	3
2	19	21	1	1	6	127	11	0	0	18	8	0	0	1	3	5
3	1	1	0	0	0	16	6	0	0	2	9	1	0	0	0	26
4	1	0	0	0	0	3	3	5	0	2	1	3	0	4	6	34

Drought classes: 1: Non-drought; 2: Near normal; 3: Moderate; 4: Severe/extreme.

It is found that the Quasi Association loglinear model (or QA) in (1) is selected from several loglinear models performed to be the most adequate for all the whole areas in northeast of Thailand of 19 provinces at the significance level of a = 0.05 (P-value = 0.927). The odds ratios prediction and the corresponding confidence intervals, which are used to analyze and predict the drought classes' transitions under the QA model, are investigated. The predicted values of odds ratios/estimated values of various drought class transitions in three consecutive months are summarized in Table 4.

Table 4 Estimates of the Odds ($\Omega_{34|ij}$) and correspondent confidence interval referring to estimates for month t when the drought class are known for months t-1 and t-2 for the Northeast of Thailand

$\Omega_{34\|ij}$	Drought class month t-1			
Drought class month t-2	1	2	3	4
1	24.561	0.403	0.007	0.003
	(12.482 , 65.561)	(0.222 , 1.684)	(0.001 , 1.642)	(0.001 , 1.332)
2	0.608	11.043	2.475	0.038
	(0.021 , 1.568)	(2.135 , 25.211)	(0.978 , 5.644)	(0.021 , 1.245)
3	0.190	70.333	30.094	0.060
	(0.002 , 1.854)	(20.314 , 94.254)	(15.622 , 56.421)	(0.029 , 1.558)
4	0.188	0.507	0.190	**0.103**
	(0.011 , 1.255)	**(0.111 , 1.894)**	(0.023 , 2.546)	**(0.002 , 1.698)**

The upper value in each cell denotes the odds estimate and the lower one refers to the odds-confidence interval.

From Table 4, an estimated odds is a ratio (($\Omega_{kl\,|\,ij}$)) of expected frequencies (E_{ijk}) predicted from loglinear model, ranging from 0 to $+\infty$, and represented the number of times that is more or less, or equally probable the occurrence of a certain event instead of another. Where, the odds ratios for three–dimensional loglinear models are given by $\Omega_{kl\,|\,ij} = E_{ijk}/E_{ijl}$, $k \neq 1$ meaning that, 1 month from now (t), it is $\Omega_{kl\,|\,ij}$ times more, less, or equally probable that a specific site is in class k instead of class l, given that at present it is in class j, and 1 month before it was in class i, with i, j, k, and l $\in$ {1, 2, 3, 4} and $k \neq 1$.

For example, the results for the estimates of $\Omega_{34|\,ij} = E_{ij3}/E_{ij4}$ and respectively confidence intervals in Table 4. They mean that, taking the case for the drought class in January and February, which was 4 (i=j=4) or the severe/extreme drought class and the estimate of odds ratio was $\Omega_{34|44}$ = 0.103 with the confidence interval (0.002, 1.698). Since the value 1 is included in that interval, then it is concluded that it was equally probable that by March this site would be in severe/extreme drought (k=4) instead of being in moderate drought (k=3) given that in January and February it was in severe/extreme drought (i, j =4).

The Results for Khonkaen province in the northeast of Thailand

For each sub-area in northeast of Thailand, we obtain for each province quite similar models as given in (2) and that all of them are sub-models of (1) with only some different parameters. Consequently, a selected province, Khonkaen, is demonstrated. The observed frequencies of drought class transitions from month t-2 to month t-1 to month t: for Khonkaen province's SPI 12 month's time scale are presented in Table 5.

The odds ratios prediction and the corresponding confidence intervals are performed and evaluated in order to investigate and predict the drought classes' transitions under the QA loglinear model. The predicted values of odds ratios for various drought class transitions in three consecutive months are shown in Table 6.

Table 5 Observed values of drought class transitions from month t-2 to month t-1 to month t: for the Khonkaen province in Thailand

Drought class month t-2	Drought class month t-1															
	1				2				3				4			
	Drought class month t–1				Drought class month t–1				Drought class month t–1				Drought class month t-1			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Observed values																
1	239	13	0	0	27	19	2	0	2	2	1	0	1	0	0	1
2	27	20	2	0	6	88	4	0	1	7	3	1	0	3	2	2
3	2	1	1	0	1	11	4	0	0	1	11	1	0	0	4	10
4	0	0	0	0	0	1	3	2	0	1	6	10	0	0	3	27

Drought classes: 1: Non-drought; 2: Near normal; 3: Moderate; 4: Severe/extreme.

Table 6 Estimates of the odds ($\Omega_{34|ij}$) and correspondent confidence interval referring to estimates for month t when the drought class are known for months t-1 and t-2 for Khonkaen province of Thailand

$\Omega_{34\|ij}$	Drought class month t-1			
Drought class month t-2	1	2	3	4
1	1.591	127.121	102.231	0.008
	(0.987 , 3.245)	(89.754 , 154.643)	(80.647 , 130.254)	(0.002 , 0.664)
2	40.667	2.545	1.212	0.478
	(25.254 , 68.974)	(0.215 , 3.654)	(0.654 , 3.264)	(0.3254 , 1.264)
3	2.546	**13.875**	2.427	0.102
	(0.326 , 5.647)	**(5.647 , 25.699)**	(0.984 , 3.654)	(0.025 , 1.365)
4	0.254	17.564	1.549	0.425
	(0.125 , 1.658)	(8.947 , 25.644)	(0.687 , 3.2654)	(0.058 , 1.254)

The upper value in each cell denotes the odds estimate and the lower one refers to the odds confidence interval.

In Table 6, for Khonkaen province, the model given by (2) is accepted adequately of fit with the deviance = 11.23, P-value = 0.854, α = .05. The predicted values of the odds ($\Omega_{34|ij}$) for Khonkaen province referring to estimates for month t when the drought class are known for months t-1 and t-2 are evaluated. The results for odds estimates $\Omega_{34|ij} = E_{ij3}/E_{ij4}$ and respectively confidence intervals for the Khonkaen province are also reported.

Taking the case for the drought class in January and February, which was 3 (moderate drought) and 2 (near normal drought) (i = 3, j = 2), respectively and the estimate of odds $\Omega_{34|ij}$ was 13.875, with the confidence intervals (5.647, 25.699). It indicates that the value 1 is not included in that interval which means that if it was given in January and February, which was 3 (medium) and 2 (near normal) it was 13.875 times more likely that by March this province would be in moderate drought (k=3) instead of being in severe/extreme drought (l=4). For more details and information are given in Table 6.

Discussion and Conclusion

This research shows that the application of loglinear modeling to the drought data allowed the comparison of the three periods in terms of odds ratios or that in terms of probabilities of transition between drought classes. We have investigated and built the loglinear models using the backward method in SAS to compare and select the adequacy goodness-of-fit of the fitted models. Results from the odds ratios with the confidence intervals are analyzed and used to predict the drought classes together with their corresponding estimated probability values of various drought class transitions in three consecutive months. Most of the results are consistent with the existence of a long-term normal natural periodicity. Even if, most of the results display for more normal drought classes than those of moderate or severity drought classes; however, there are still some areas of investigation that drought could be developed and that could be attributed to the climate change. Therefore, future monitoring of drought watch system is needed to be set the pace in these areas in order to keep up with them as well as more analyzing data from an extended time period is also needed to detect a possible long term climate change. Especially, an appropriate statistical model used to predict drought phenomenon in each area is necessary and is probably able to provide more effectiveness in the administrative management of hazard from drought. Furthermore, future research works concerning building and extending loglinear models and/or logit models (Nelder, 1974, Agresti, 1990, Pongsapukdee, 2012) including model selection for prediction of drought category are also strongly recommended.

Acknowledgements

All raw data used in this investigation are from the Institute for Water, Thailand. We would also like to thank the editors and the referees for their very helpful comments.

References

Agresti, A. (1990) *Categorical Data Analysis.* John Wiley & Sons, New York.

Agresti, A. (2002) *Categorical Data Analysis.* 2nd ed., John Wiley & Sons, New York.

Fienberg, S. E. (2000) Contingency Tables and Log-Linear Models: Basic Results and New Developments. *Journal of the American Statistical Association* 95: 643-647.

Moreira Elsa E., Ana A. Paulo, Luis S. Pereira, and Joao T. Mexia. (2006) Analysis of SPI drought class transitions using loglinear models. *Journal of the Science Direct* 331: 349-359.

Moreira Elsa E., Carlos A. Coelho, Ana A. Paulo, Luis S. Pereira, and Joao T. Mexia. (2008) SPI-based drought category prediction using loglinear models. *Journal of the Science Direct* 354: 116-130.

Mckee, T. B., Doesken, N. J., and Kleist, J. (1993) The relationship of drought frequency and duration to time scales. In *Proceedings of the Eighth Conference on Applied Climatolgy. American Meteorological Society.* Boston: 233-236.

Nelder, J. A. (1974) *Loglinear models for contingency tables: a generalization of classical least squares.* Applied Statistics. John Wiley & Sons, New York.

Nebraska-Lincoin University. "SPI." [Online URL: www.drought.unl.edu] accessed on July 10, 2009.

Paulo, A. A., Ferreira, E., Coelho, C., and Pereira, L. S. (2005) Drought class transition analysis through Markov and Loglinear models, an approach to early warning. *Agricultural Water Management* 77:59-81.

Paulo, A. A. and Pereira, L. S. (2006b) Prediction of SPI drought class transitions using Markov chains. *Water International* 31 (in press).

Paulo, A. A., Pereira, L. S., and Matias, P. G. (2003) Analysis of local and regional droughts in southern Portugal using the theory of runs and the Standardized. *Water International* 28: 65-68.

Pereira, L. S., Cordery, I., and Iacovides, I. (2002) Coping with Water Scarcity. UNESCO IHP VI, *Technical Documents in Hydrology* No. 58, UNESCO, Paris, 267 pp. [Online URL: www.unescdoc.unesco.org/images/0012/001278/127846.pdf].

Pongsapukdee, V. (2012) *Analysis of Categorical Data: Theories and Applications with GLIM, SPSS, SAS and MTB.* 3rd ed., Silapakorn University Press, Nakhon-Pathom.

Sharma, T. C. (1997) Estimation of drought severity on independent and dependent hydrologic series. *Water Resources Management* 11: 35-50.

Sivakumar, M. V. K. and Wilhite, D. A. (2002) Drought preparedness and drought management. In: *Drought Mitigation and Prevention of Land Desertion (Proceedings of the International Conference, Bled, Slovenia)*, UNESCO and Slov. Nat. Com. ICID, Ljubljana, CD-ROM Paper 2.