

Dollapas Purparich 2009: Factor Analysis for Predictive Model of Rainfall Occurrence in the Northeastern Part of Thailand. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsrisawang Ph.D. 146 pages.

The purpose of this research was the analysis of the common factors of daily upper air data for the development of a model of rainfall occurrences in Northeastern Thailand. The daily upper air observations were provided by The Bureau of the Royal Rain Making and Agricultural Aviation and the records of rainfall in 24 hour period were obtained from the Department of Meteorology. Both data sets contained 277 inconsistent records for the period of 12 March 2004 - 14 September 2006. Each record consisted of 49 variables such as temperature, humidity, pressure, wind, stability and forecasting indices, and warm cloud seeding potential. The data on rainfall volume were categorized into rain or no-rain events. Two statistical methods of factor analysis and logistic regression were applied for developing the prediction model of rain event in Northeastern Thailand. The factor analysis was first conducted by considering three interesting issues as following: 1) the selection of 45 quantitative variables or 28 quantitative variables with $VIF \geq 10$, 2) the extraction of common factors using the Principal Component or the Maximum Likelihood method, and 3) the estimation of factor score using the method of Regression or Factor Based Score. As a result, the different eight alternatives of factor analysis were studied. Next the logistic regression was performed using the number of factors obtained from the above result of each alternative and three additional qualitative variables: convective temperature reachable, warm cloud seeding potential, and yesterday's rain event as the independent variables in the development of eight difference predictive models of rainfall occurrences.

The predictive power of the eight rainfall models were measured by the percentages of prediction accuracy or hit rate statistics. It appeared that these models could achieve the hit rates ranged from 88.8% to 91.0%. The most suitable model was the model that extracted three common factors from the combination of using 28 quantitative variables with $VIF \geq 10$, factor extraction by maximum likelihood, and estimation of factor score by the Factor Based Score method. These three factors and the three stated qualitative variables were modeled under the procedure of forward stepwise variable selection with a significance level of 0.15 for entry and 0.20 for removal. The results showed that the statistically significant factors at 0.05 level consisted of one common factor of five humidity variables and the other qualitative variable of yesterday's rain event with their odds ratios (OR) of 1.38 and 3.80, respectively. This model gave the accuracy of prediction about 90.3%

Student's signature

Thesis Advisor's signature

/ /