# Use of Regression Method in Prediction of Drought in five areas in Sudan

**Dr. Suhad Ali Osman Abdallah**

Assistant Professor at the Community College for Girls Khamis Mushait King Khalid University[1] – KSA

# استخدام طريقة الانحدار في التنبؤ بالجفاف في خمس ولايات في السودان

**د. سهاد علي عثمان عبد الله**

**أستاذ مساعد تخصص إحصاء بكلية المجتمع للبنات بخميس مشيط جامعة الملك خالد[1] – المملكة العربية السعودية**

[1] **تود الباحثة شكر جامعة الملك خالد**

**على دعمها الفني والإداري لهذا البحث**

# Abstract

*This paper aimed to explore the possibility of the use of logistic regression in the prediction of drought in Sudan. In this paper Regression models were used to predict drought in 5 states in Sudan. Results showed that the average prediction error for above-average and below-average years is reported for each station. Area-specific statistics are listed in Table1. Correlations between predictions and observations range from $r = -0.31$ to $r = -0.68$, with the model performing slightly better in predicting the number of wet days in drier years. In general, however, the simple linear model displays an average absolute error ranging between 9 and 22 days.*

**Keywords**: *Regression, Drought.*

***[1]The researcher would like to thank King Khalid University for the administrative and technical support for this research***

# Introduction

During the 20th century, Sudan experienced major drought. The most destructing ones were in 1913, 1940, and 1954 which covered many parts of the country. In 1913 and 1940, about 1.5 million people were affected. In the 1984, 4.5 million people went hungry. Some of the affected people became relief-recipient and less work-oriented. Different tribes responded differently to recurring drought(1).

Annual variability and the relative scarcity of rainfall in the north of Sudan in particular – have a dominant effect of agriculture and food security, and are strongly linked to displacement and related conflicts. Drought events also change the ecosystem, as dry spell kill long-lived trees and result in a general reduction of the vegetation cover, leaving land more vulnerable to overgrazing and erosion[1].

Insufficient and highly variable annual precipitation is a define feature of the climate of most of Sudan. A variable analysis of rainfall records from 1961 to 1990 in Northern and Southern Kordofan found that annual precipitation ranged from 350 to 850 mm, with an average annual variation of 65 percent in the northern parts of the Northern Kordofan and 15 percent in the southern parts of Southern Kordofan(2).

Degradation of grazing resources in one of the major livestock production problems as of result of drought coupled with other factors namely overgrazing and the expansion of large scale mechanized farming on marginal grazing lands. Land sat STM map of 1983/84 showed that the semi-desert (455,000 sq. km) and some parts of the northern fringes of the low rainfall woodland savannah were severely affected by drought and environmental degradation. Range and postural administration report noted that 177 million faddans of rangeland area are considered as severely degraded lands(3).

Due to its complexity with diverse origins and occurrences at different temporal and spatial scales, drought prediction has presented a major challenge to climatologists and hydrologists as well as decision and policymakers. Generally, three types of methods have been used for drought prediction: statistical, dynamical, and hybrid methods(4, 5).

The statistical prediction method uses empirical relationships of historical records, taking different influencing factors as predictors. With the increased computational capability and improved understanding of climate, drought prediction has been tackled more with state-of-the-art general circulation models (GCMs), which provide drought prediction based on the physical processes of the atmosphere, ocean, and land surface. The past decade has also witnessed the development of hybrid prediction methods that combine forecast from both statistical and dynamical methods(6).

Drought prediction generally refers to the prediction of drought severity (e.g., values of a specific drought indicator). In certain cases drought prediction also refers to other properties, such as drought duration and frequency, or phases, such as drought onset, persistence, and recovery. In this study, we mainly focus on the prediction of drought severity at the seasonal time scale, which centers around the current drought prediction efforts and is particularly related to the operational early warning to mitigate drought impacts(7). This paper based on the use of logistic regression in the prediction of drought in Sudan.

## Materials and methods

Linear regression is a traditional method for statistical prediction in hydrology and climatology. A basic formulation of the multiple linear regression model for drought prediction concerning two predictors *X* and *Z* can be expressed as (1-month lead):

$$Y_t = AX_{t-1} + BZ_{t-1} + \varepsilon_t \quad (1)$$

where $Y_t$ is the predictand (or dependent variable) in the form of drought indicator series; $X_{t-1}$ and $Z_{t-1}$ are covariates (or independent variables) that provide the predictive information of $Y_t$; A and B are regression coefficients; and $\varepsilon_t$ is the error term.

In contrast to the time series model that predicts drought severity based solely on the persistence of certain drought indicators, the regression model seeks to establish the relationship between predictand and other variables that may contribute to the predictive information of drought developments. The regression model has been used to model the relationship between drought indices to be predicted and a suite of predictors(8). Due to the high persistence of the drought indicator, the autoregressive component (e.g., $Y_{t\text{-}1}$ for 1 month lag) is commonly used in the regression framework. In this case, the drought indicator is predicted based on linear combinations of antecedent drought conditions through a lagged term as well as external factors related to drought conditions of the target period. The regressive model in equation **(1)** is generally limited by the assumption of linear relationships between the predict and predictors. Several nonlinear regression models, such as the locally weighted polynomial regression, have been employed for modeling nonlinear relationships for prediction purposes.

Define a binary response variable $Y$ ($Y = 1$ or 0). The distribution function of $Y$ is expressed as the probability $P(Y = 1) = \pi$ of success and $P(Y = 0) = 1 - \pi$ of failure. The objective is to estimate the conditional distribution of $Y$ with respect to the predictor $X$:

$$logit(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k$$

(2)

where $\alpha$ and $\beta$ are parameters; $\alpha + \beta_1 x_1 + \ldots + \beta_k x_k$ is the linear combination of explanatory variables; logit($\pi$) is the link function that links the response variable $Y$ with the predictor $X$. The commonly used link function is the logit which is defined as logit ($\pi$) = log($\pi/(1 - \pi)$). Note that a linear relationship is assumed between the transformed response logit ($\pi$) (but not the response $Y$ itself) and explanatory variables.

When there are three or more drought categories (e.g., multiple categories D0–D4 in the USDM), the ordinal regression model can be used for the prediction of multiple drought categories with respect to a suite of predictors. For $m$ drought categories, denote $F_j = P(Y \leq j)$ the cumulative non exceedance probability of drought category $j$ ($j$ = 1, 2,..., $m - 1$). By substituting $F_j$ into the logit link function, the ordinal regression model is expressed as(9, 10):

$$\log\left(\frac{P(Y \leq j)}{1-P(Y \leq j)}\right) = \alpha_j + \beta_1 x_1 + \ldots + \beta_k x_k \quad j = 1,2,\ldots,m-1$$

(3)

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_{m-1}]$ is the intercept parameter; $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k]$ is the regression coefficient.

The ordinal regression model has been employed to predict USDM drought categories based on a suite of predictors, such as precipitation, soil moisture, evaporation, runoff, or climate indices. The potential limitation in predicting multiple drought categories is that a large number of parameters may be involved. A potential solution is to combine several categories to keep the model parsimonious(11). For example, the USDM categories can be grouped into two drought categories (e.g., drought more severe than D2 or not) and the logistic regression model can be employed to predict the probability of drought occurrences with fewer parameters. Currently, several drought information systems are based on the drought categories. The logistic or ordinal regression model can be applied for drought category prediction in these systems

## Results and discussion

The first principal component of the data reported approximately 70% of the variance in the number of wet days for all 5 states with daily data. High correlations are observed between this first principal component and surface temperature within the region typically associated with rain fall. No additional regions or climate variables (e.g., sea level pressure, geopotential height) are identified using this method. Further, composite mapping and global wavelet analysis yield no additional potential predictors for incorporation into the prediction model. Thus, the wet–dry day frequency model uses only the dusty winds season ahead index as a direct predictor of wet days in any given drought season. Using the same cross-validation method already described, the number of wet days per season is predicted for each station.

In addition to correlation coefficients between the predicted and observed number of wet days, the average prediction error for above-average and below-average years is reported for each station. Area-specific statistics are listed in Table1. Correlations between predictions and observations range from r = −0.31 to r = −0.68, with the model performing slightly better in predicting the number of wet days in drier years. In general, however, the simple linear model displays an average absolute error ranging between 9 and 22 days.

**Table (1)**

| **Area** (average number of wet days) | Correlation value (r) between prediction and observation | Average absolute error in years with above-average number of wet days | Average absolute error in years with below-average number of wet days |
|---|---|---|---|
| Area 1: 120 days | - 0.67 | 18 days | 16 days |
| Area 2: 88 days | -0.54 | 11 days | 9 days |
| Area 3: 139 days | -0.31 | 22 days | 19 days |
| Area 4: 60 days | -0.48 | 14 days | 13 days |
| Area 5: 35 days | -0.33 | 9 days | 8 days |

## Conclusion

Use of applied statistics in general and logistic regression method, in particular, will help the specialists in the water resources and agriculture to predict the annual amount of water available for agriculture and grazing and this will help in turn in setting accurate plans to avoid shortages.

## References

1. National Adaptation Plan of Action,. (NAPA) (Sudan) UNDP/GEF/Sudan Inception Report 2010.

2. Babekir Hag Hassan. Food gaps in drought areas affected in Sudan, The Administration of Food Security, Ministry of Agriculture and Irrigation 2008.

3. Mohmed Elamin Abdelrahman,. The development of alternative and supplementary livelihood system to reduce pressure on land based resources in Sudan dry land 2008.

4. Mishra, A. K., and V. P. Singh, Drought modeling-A review, J. Hydrol., 2011; 403(1-2), 157–175.

5. Mariotti, A., S. Schubert, K. Mo, C. Peters-Lidard, A. Wood, R. Pulwarty, J. Huang, and D. Barrie, Advancing Drought Understanding, Monitoring, and Prediction, Bull. Amer. Meteor. Soc, 2013; 94(12), 186-188.

6. Pozzi, W., J. Sheffield, R. Stefanski, D. Cripe, R. Pulwarty, J. V. Vogt, R. R. Heim Jr, M. J. Brewer, M. Svoboda, and R. Westerhoff, Towards global drought early warning capability: expanding international cooperation for the development of a framework for global drought monitoring and forecasting, Bull. Amer. Meteor. Soc, 2013; 94(6), 776–785.

7. Wetterhall, F., H. Winsemius, E. Dutra, M. Werner, and E. Pappenberger, Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin, Hydrol. Earth Syst. Sci., 2015; 19(6), 2577-2586.

8. Sharma, T. C., and U. S. Panu, Prediction of hydrological drought durations based on Markov chains: case of the Canadian prairies, Hydrol. Sci. J., 2012; 57(4), 705-722.

9. Hao, Z., Y. Hong, Y. Xia, V. P. Singh, F. Hao, and H. Cheng, Probabilistic drought characterization in the categorical form using ordinal regression, J. Hydrol., 2016; 535, 331–339.

10. Agresti, A., An introduction to categorical data analysis, John Wiley & Sons; Hoboken, New Jersey 2007.

11. Hao, Z., Y. Xia, L. Luo, V. P. Singh, W. Ouyang, and F. Hao, Toward a categorical drought prediction system based on U.S. Drought Monitor (USDM) and climate forecast, J. Hydrol., 2017; 551, 300–305.