IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Using the Group Lasso Method to Identify the Most Important Factors Affecting Desertification in Al-Qadisiyah Governorate

Saif Hosam Raheem[1], Enas abid alhafidh mohamed[2] and Jassim Nassir Hussain[3]

[1]Department of statistics, University of Al-Qadisiyah, Collage of Administration and Economics, Al-Qadisiyah, Iraq
[2]Department of statistics, University of Karbala , Collage of Administration and Economics, Karbala, Iraq
[3]Akut university college , kut, Iraq

ARTICLE INFO

Article history:
Received 20/11/2024
Revised 20/11/2024
Accepted 14/1/2025
Available online 15/5/2025

Keywords:
Desertification
Variable Selection
Lasso
Group Lasso

ABSTRACT

This study investigates desertification in Al-Qadisiyah Governorate to better understand the environmental and economic patterns shaping the region. It also examines the impact of both natural and human factors on land degradation. Some areas suffer from more severe degradation, resulting in the loss of biodiversity and productive land.
To address this issue, the study identifies the key drivers of desertification by analyzing a range of factors—natural, human, economic, social, and technological—using the Group Lasso method, a modern variable selection technique. Annual data from 2020 to 2023 were analyzed using the R programming language.
The results show that four main factors significantly contribute to desertification: climate change, unsustainable agricultural practices, poverty, and traditional irrigation methods. These findings highlight the urgent need for targeted strategies to mitigate these challenges and ensure the sustainable use of natural resources for future generations

1. Introduction

Desertification is one of the most pressing environmental challenges worldwide, with serious implications for sustainable development and natural resource management. This study focuses on understanding land degradation and the decline in soil fertility caused by a variety of natural, human, economic, social, and technological factors. It highlights the complex interaction between people and their environment, showing how human activities, along with environmental forces, accelerate land degradation. Recognizing the significance of desertification is essential for developing strategies to mitigate its effects and promote sustainable land use.

Patterns of desertification can be understood by identifying severely affected areas and distinguishing them from regions that maintain productive capacity. The phenomenon tends to be more intense in specific areas due to a combination of contributing factors. Natural elements such as climate and topography, particularly in arid regions, play a major role in accelerating land degradation. In addition, poor land management and unsustainable agricultural practices amplify the problem. Socioeconomic factors—such as poverty and population pressure—further intensify resource depletion. Technological limitations, especially outdated irrigation methods, also contribute to inefficient resource utilization.

In Al-Qadisiyah Governorate, desertification has emerged as a serious concern, resulting in continued land degradation, loss of natural

* Corresponding author : E-**mail** address: saif.hosam@qu.edu.iq
https://doi.org/10.62933/r9tv0505

resources, and negative impacts on economic development. This multifaceted issue highlights the urgent need to investigate the underlying causes.

The objective of this study is to identify the main drivers of desertification in Al-Qadisiyah Governorate by collecting and analyzing relevant data. To achieve this, the study employs the Group Lasso method, a modern variable selection technique capable of handling large numbers of predictors and selecting the most influential ones. This approach provides deeper insights into the factors contributing to land degradation and supports the development of more effective and sustainable resource management strategies

2. Methodology

2.1 Desertification

Desertification is a serious environmental phenomenon that leads to land degradation in arid and semi-arid regions, driven by both natural and human-induced factors. Among the natural causes, climate change plays a major role, with rising temperatures and declining rainfall reducing soil fertility and increasing land vulnerability to erosion. Human activities such as overgrazing, deforestation, and unsustainable farming practices further intensify degradation, leading to reduced productivity, food insecurity, and economic instability [1].

In Iraq, desertification poses an escalating threat to both the environment and human livelihoods. The country suffers from severe water scarcity, rising temperatures, and mismanagement of natural resources—all contributing to the decline of arable land. Nearly 39% of Iraq’s territory is at risk of desertification, with significant effects on agriculture and rural stability. This has driven large numbers of people to migrate from rural areas to urban centers in search of better opportunities. A notable factor worsening this crisis is the absence of effective water governance, compounded by dam construction in neighboring countries like Turkey and Iran, which reduces water inflows into Iraq .

Al-Qadisiyah Governorate in southern Iraq is particularly vulnerable. Chronic water shortages and poor land-use practices have accelerated the degradation of farmland. The resulting loss in soil fertility threatens food security and has intensified rural migration to cities. Studies show that enhancing water resource management, modernizing irrigation infrastructure, and implementing sustainable agricultural methods are critical to combating desertification in the region. These measures would help safeguard agricultural productivity and build long-term environmental and economic resilience [3]

2.2 Affecting Factors For Desertification

Desertification is a complex environmental challenge driven by a range of factors that degrade land and ecosystems. These factors can be broadly categorized into four main groups: natural factors, human factors, economic and social factors, and technological factors. Each of these groups plays a significant role in either accelerating or mitigating desertification, particularly in regions that are vulnerable to land degradation.[1]

Natural factors are among the primary drivers of desertification. Climate changes, such as shifts in rainfall patterns and prolonged droughts, reduce water availability and contribute to soil degradation. Rising temperatures and increased evaporation rates diminish the moisture content in the soil, making it difficult to support agricultural activities. Irregular rainfall, especially when it is low or intermittent, affects groundwater recharge and makes farming less sustainable. Additionally, strong winds contribute to the erosion of fertile topsoil, reducing the quality of agricultural land. Natural disasters, such as sandstorms and floods, can further accelerate land degradation by damaging the soil and making it more susceptible to erosion.[2]

Human factors also play a significant role in exacerbating desertification. Unsustainable agricultural practices, such as over-plowing, excessive irrigation, and the use of chemical pesticides and fertilizers, deplete the soil's nutrients, leading to a decline in its fertility. Deforestation, where trees and vegetation are cleared, removes natural barriers that protect soil from erosion, worsening land degradation. Overgrazing by livestock reduces vegetation cover, leaving the soil exposed to erosion and diminishing the quality of the land. Urbanization and the expansion of infrastructure convert productive agricultural land into residential or industrial areas, decreasing the amount of arable land. Additionally, unsustainable water use, including groundwater over-extraction and inefficient irrigation systems, contributes to land degradation and soil salinity.

Economic and social factors also contribute to desertification. Poverty, especially in rural communities, forces people to rely heavily on natural resources for their livelihoods, increasing pressure on the land and environmental resources. Rapid population growth further exacerbates this issue, as more people require more land and water for food production and living spaces, leading to the overuse of these resources. Ineffective agricultural and environmental policies, or the absence of such policies, worsen the problem of desertification, as there are no measures in place to protect the environment and manage land sustainably. Rural migration, where people move from rural areas to cities, often leaves agricultural lands unattended, accelerating land degradation as these lands are no longer managed or protected from environmental factors.

Technological factors are also important in influencing desertification. The reliance on outdated agricultural technologies that inefficiently use resources leads to the depletion of soil and water. The inability to adopt modern technologies, such as drip irrigation, which can conserve water and reduce soil degradation, further exacerbates the issue. The adoption of modern technologies could serve as an effective solution to reduce desertification and promote sustainable agricultural practices.[1,2]

All these factors interact in complex ways, contributing to the ongoing challenge of desertification in affected regions. They directly influence the ability of these regions to support agriculture and maintain the sustainability of their natural resources. Understanding and addressing these factors is essential for combating desertification and ensuring the long-term health and productivity of the land.

2.3 Variable selection

Variable selection is a crucial step in developing statistical models and conducting data analysis, especially in forecasting models such as linear regression and logistic regression. The primary goal of variable selection is to identify the most effective set of independent variables that enhance model accuracy while reducing complexity, resulting in a more interpretable model that is less prone to bias or overfitting. Several methods and techniques can be used for variable selection, ranging from traditional to regularization methods.

One common approach is the traditional selection method, which involves relying on domain knowledge, experience, and intuition to choose the most relevant variables. This approach is particularly useful when the researcher or analyst has a deep understanding of the problem and the relationships between the variables. However, it can be subjective and may overlook important variables.

A regularization method that is widely used is stepwise selection [4], which combines both forward and backward selection. It starts with either no variables or all variables and adds or removes variables based on specific criteria, typically the p-value or information criteria like AIC [5] or BIC [6]. Forward selection begins with no variables, adding them one by one based on their contribution to the model, while backward elimination starts with all variables and removes the least significant ones. A bidirectional approach allows variables to be added or removed at each step.

Another widely used method is Lasso (Least Absolute Shrinkage and Selection Operator) proposed by Tibshirani (1996)[7]. Lasso is a regularization technique that applies a penalty to the regression coefficients, shrinking some of them to zero. This effectively removes variables from the model, making it a powerful tool for variable selection, especially when multicollinearity is present.

Ridge regression [8], also a regularization method, adds a penalty to the regression coefficients but does not shrink them to zero, allowing all variables to remain in the model. This method is ideal when the goal is to reduce multicollinearity without strictly performing variable selection. Elastic Net [9], a hybrid of Lasso and Ridge regression, combines the strengths of both methods, making it useful when dealing with correlated variables by selecting groups of variables together.

Other advanced methods for variable selection include group Lasso [10], which extends the Lasso method by selecting groups of correlated variables together, and Adaptive Lasso [11], which allows for a more flexible penalty, improving the selection process by adapting the penalty weights for each coefficient. Another important method is smoothly clipped absolute deviation (SCAD) proposed by Fan and Li (2001)[12], which minimizes the bias induced by traditional penalty methods like Lasso.

Additionally, Adaptive Elastic Net [13] improves on the original Elastic Net by adapting the penalty to better suit the data structure. Reciprocal Lasso [14] offers another extension of Lasso that adapts the penalty in a more flexible manner, enhancing its performance in selecting significant variables.

In addition, information criteria such as AIC and BIC [4,5] are used to evaluate models and choose the optimal set of variables by penalizing model complexity. These criteria strike a balance between model fit and simplicity, guiding the selection of variables that provide strong explanatory power without overfitting.

Principal Component Analysis (PCA) [15] is another approach, though it focuses on dimensionality reduction. PCA transforms the original variables into a smaller set of uncorrelated components, which can then be used as predictors. While it doesn't directly select variables, it reduces the number of predictors and addresses multicollinearity.

2.4 Group Lasso Method

The Group Lasso method was proposed by Yuan and Lin in 2006 as an extension of the traditional Lasso method. In the traditional Lasso approach, only individual variables are selected, which can lead to ignoring entire groups of related variables that may be important in certain applications. To address this issue, the Group Lasso method introduces an appropriate penalty that allows for selecting or discarding entire groups of variables, rather than just individual variables[16].

The Group Lasso approach organizes variables into groups and imposes a penalty based on the selection of variable groups instead of individual variables. The Group Lasso estimator can be represented by the following equation:

$$\hat{\beta} = arg\ min \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{g=1}^{G} |\beta_{I_g}|$$

where:

- G represents the number of groups (where g=1,2,…,G).
- I_g represents the set of indices belonging to the g-th group of variables.

The penalty used in Group Lasso is a combination of the l_1 and l_2 norms [7], allowing the method to balance between variable selection and regularization. This method has the advantage of selecting variables under orthogonal transformations (GroupWise), similar to Ridge regression.

The benefit of the Group Lasso method is its ability to either select or discard entire groups of variables, which simplifies the interpretation of results in applications that involve grouped data, such as genomic analyses or functional MRI studies. This makes Group Lasso a powerful tool for handling structured variable selection in a wide range of fields. at the group level, making it particularly useful for situations where variables are naturally grouped, such as in genetic studies or signal processing. Furthermore, it remains invariant

3. Results and discussion

The required data was obtained from the Central Statistical Organization of Iraq, the Iraqi Ministry of Planning, and the Iraqi Ministry of Water Resources for the period from 2020 to 2023. This data represents the factors influencing desertification in Al-Diwaniyah Governorate. These factors can be explained as follows:

Table 1: Coding of factors affecting desertification

factors	coding	factors	coding
1. Natural Factors		3. Economic and Social Factors	
- Climate changes:	x1	- Poverty:	x10
- Irregular rainfall:	x2	- Population growth:	x11
- Strong winds:	x3	- Ineffective agricultural and environmental policies:	x12
- Natural disasters:	x4	- Rural migration:	x13
2. Human Factors		4. Technological Factors	
- Unsustainable agriculture:	x5		
- Deforestation:	x6	- Outdated agricultural technologies:	x14
- Overgrazing:	x7		
- Urbanization and urban expansion:	x8		
- Unsustainable water use:	x9	- Lack of modern technologies:	x15

The real data, consisting of fifteen independent variables and one dependent variable, was analyzed using the Group Lasso method to identify the most significant factors.

The R programming language was utilized to obtain the results, which are presented in the table below.

Table 2: Estimated Model Results

Variable	β	Variable	β	Variable	β
x1	3.0154	x6	0.0012	x11	0.0005
x2	6.2145	x7	6.2147	x12	0.0054
x3	0.0004	x8	0.0003	x13	0.0047
x4	0.0023	x9	5.0124	x14	0.0004
x5	3.0142	x10	8.2145	x15	7.1021

In this table, the results of the estimated model using the Group Lasso method are presented. This method works by identifying the influential variables and excluding the non-influential ones based on the estimated coefficients (β). The interpretation of the table is as follows:

Variables with high estimated values (influential variables):

The variables climate changes, irregular rainfall, unsustainable agriculture, overgrazing, urbanization and urban expansion, unsustainable water use, and population growth have relatively high estimated values (e.g., 3.0154, 6.2145, 3.0142, 6.2147, 5.0124, 8.2145, and 7.1021, respectively). These variables are considered influential in the model, as the high β values indicate that they play a significant role in explaining the dependent variable .

Variables with estimated values close to zero (non-influential variables):

The variables strong winds, natural disasters, deforestation, ineffective agricultural and environmental policies, rural migration, poverty, lack of modern technologies, and outdated agricultural technologies have estimated values very close to zero (e.g., 0.0004, 0.0023, 0.0012, 0.0003, 0.0005, 0.0054, 0.0047, and 0.0004). These variables are considered non-influential in the model, as the Group Lasso method automatically shrinks these variables' values to nearly zero, effectively excluding them from the model.

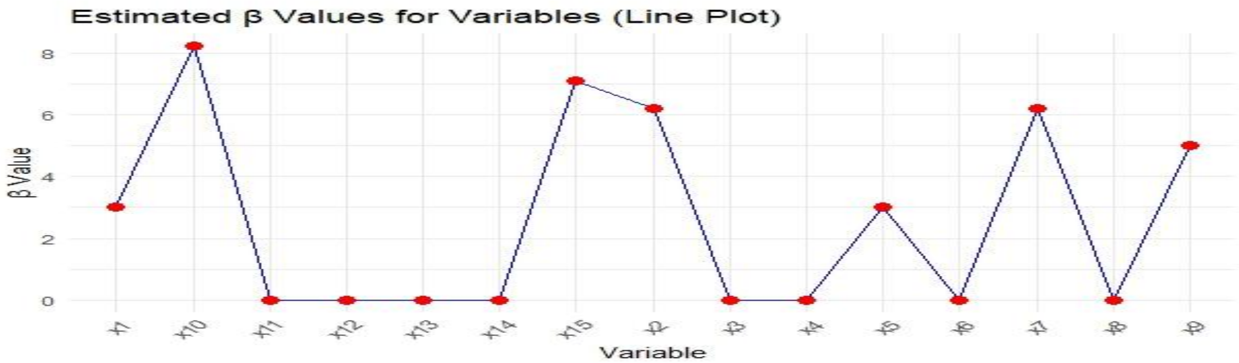

Figure 1. Estimated β Values for Variables

4. Conclusions

The findings of this study confirm the initial hypothesis presented in the introduction. Using Group Lasso analysis on data collected between 2020 and 2023, the study identified key factors contributing to desertification in Al-Qadisiyah Governorate. Climate changes, unsustainable agriculture, poverty, and traditional irrigation practices emerged as the most significant drivers of land degradation.

These outcomes align with prior expectations, emphasizing the critical role of both environmental and human-related influences. Additional factors, such as population growth and unsustainable water use, were also shown to intensify the problem, reflecting the complex interplay between ecological stress and socio-economic pressures.

The Group Lasso method proved effective in selecting the most influential variables, offering a clear interpretation of the underlying causes of desertification. This contributes a valuable framework for analyzing land degradation and supports evidence-based decision-making.

Limitations of the Group Lasso Method: Despite its strengths, the Group Lasso approach has certain limitations. It requires predefined group structures, and its performance can decline if these groups are incorrectly specified or overlapping. Additionally, it assumes uniform importance among variables within the same group, which may not always hold in practice. In high-dimensional data with limited observations, the method may also yield unstable estimates or biased selections.

These insights are important for policymakers seeking to reduce desertification impacts through informed environmental planning. Future research should focus on applying improved water management systems, advanced irrigation technologies, and sustainable agricultural methods to address the highlighted issues effectively.

References

[1] Geist, H. J., & Lambin, E. F. (2004). Dynamic causal patterns of desertification. Bioscience, 54 (9), 817-829.
[2] Al-Obaidi, J. R., Yahya Allawi, M., Salim Al-Taie, B., Alobaidi, K. H., Al-Khayri, J. M., Abdullah, S., & Ahmad-Kamil, E. I. (2022). The environmental, economic, and social development impact of desertification in Iraq: a review on desertification control measures and mitigation strategies. Environmental Monitoring and Assessment, 194(6), 440.
[3] AL-Kar, H. A. G. (2021). Environmental and Demographic Effects of Desertification in Al-Qadisiyah Province. Al-Qadisiyah Journal For Humanities Sciences, 24(3), 1-25.
[4] Efroymson, M. A. (1960). Multiple Regression Analysis. Mathematical Methods for Digital Computers/John Wiley.
[5] Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.
[6] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.
[7] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.
[8] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
[9] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.
[10] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B: Statistical Methodology, 68(1), 49-67.
[11] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.
[12] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.
[13] Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.
[14] Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. Journal of the American Statistical Association, 110(512), 1607-1620.
[15] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.
[16] Wang, H., & Leng, C. (2008). A note on adaptive group lasso. Computational statistics & data analysis, 52(12), 5277-5286.
[17] AL-Sabbah, S. A., Mohammed, L. A., & Raheem, S. H. (2021). SLICED INVERSE REGRESSION (SIR) WITH ROBUST GROUP LASSO. International Journal of Agricultural & Statistical Sciences, 17(1).