

## Representación de la retención de humedad en los suelos mediante el Biplot de regresión con rango reducido

*Willin Álvarez<sup>(1)</sup>, Teodoro García<sup>(2)</sup>*

*<sup>(1)</sup> Universidad Nacional Experimental Rómulo Gallegos, Venezuela*

*<sup>(2)</sup> Universidad de Carabobo. Grupo de Investigación en Gestión de la Calidad, Venezuela*

*Email: aprendizborroso@yahoo.com*

### Resumen

En este trabajo se analiza, mediante una aproximación Biplot de los coeficientes de regresión con rango reducido, la retención de humedad en los suelos en la depresión del Lago de Valencia, ubicada en Venezuela entre los estados Aragua y Carabobo. La técnica propuesta se llevó a un programa computacional que permite realizar representaciones Biplot de los coeficientes de regresión con rango reducido en presencia y ausencia de la multicolinealidad. Al analizar los resultados se pudo observar que las dos componentes mejor representadas en el Biplot se formaron fundamentalmente debido a las contribuciones de las variables colineales. En consecuencia, la retención de humedad puede ser explicada por estas componentes.

**Palabras clave:** Biplot de coeficientes de regresión, rango reducido, multicolinealidad, retención de humedad.

## Representation of the humidity retained by the soils through regression biplot with reduced range

### Abstract

This paper analyses through an approximation of regression coefficients Biplot with reduced rank, the humidity retention in soil obtained in the Depression of Valencia Lake, located between Carabobo and Aragua states in Venezuela. The technique proposed was programmed and this software allows obtaining the representations of regression coefficients Biplot with Reduced rank in presence or absence of multicollineality. The analysis of results revealed that the two components better represented in the Biplot were formed mainly due to contributions from collinear variables. Therefore, the humidity retention in soils can be explained by these components.

**Keywords:** Regression coefficients biplot, reduced range, multicollineality, humidity retention.

### 1. INTRODUCCIÓN

Desde el punto de vista agronómico, el suelo es el sitio donde viven y se desarrollan las plantas y animales, los cuales por razones obvias son altamente importantes en el mantenimiento de la vida humana [1].

El suelo puede definirse desde varios puntos de vista, pero en esta investigación el interés recae en la óptica de la Edafología, ésta considera al suelo en relación a su uso como un medio para el crecimiento de las plantas. Su función primordial es la conservación

y mejora de los suelos, a tales efectos, se evalúan los factores que causan variaciones en su productividad.

En todos los aspectos de la Ciencia del Suelo, tales como los procesos de meteorización, el desarrollo de horizontes, el crecimiento de las plantas, el manejo del suelo, los usos agrícolas del suelo, el agua es un elemento que debe ser considerado.

La presencia del agua en la superficie del suelo es vital para el crecimiento de las plantas, no solo porque las plantas necesitan el agua para sus procesos fisiológicos, sino también porque el agua contiene los

nutrimentos en solución esenciales para su desarrollo. Además, el agua controla otros dos componentes importantes para el crecimiento de las plantas como son el aire y la temperatura. Finalmente, el agua que escurre sobre la superficie de los suelos tiene un poder erosivo muy grande, factor que hay que tomar en consideración para la conservación y productividad de los mismos.

Las curvas de retención de humedad indican que el óptimo contenido de humedad para la mayoría de los cultivos, al cual debe mantenerse un suelo, es la capacidad de campo. La pregunta que cada agricultor o productor agrícola se hace, es cuándo y cuánta cantidad de agua es necesario aplicar para reponer el agua que se pierde del suelo por evaporación y evapotranspiración, y llevarlo a su capacidad de campo, evitando que el cultivo sufra de déficits de humedad.

La retención de agua de los suelos es una propiedad clave para el estudio de agua aprovechable por las plantas, infiltración, drenaje, conductividad hidráulica, riego, estrés de agua en las plantas y movimiento de solutos [2].

La retención de humedad en el suelo está directamente relacionada con los contenidos de arena, limo y arcilla [2, 3]. También puede ser afectada por el contenido de materia orgánica y la densidad aparente del suelo [4, 5]. Los valores de estas variables por lo general, son más fácilmente disponibles porque son determinados en análisis de rutina. Por esta razón, frecuentemente se toman en cuenta para la estimación indirecta de la humedad retenida en el suelo, a un potencial métrico determinado [6].

Distintos modelos han sido desarrollados para estimar la retención de humedad a distintas tensiones. La mayoría de ellas consisten en ecuaciones de regresión, mientras que otras consisten en modelos de simulación. [7] han propuesto el término “funciones de pedotransferencia” para referirse a estos modelos.

El éxito de un modelo de simulación como modelo de pedotransferencia depende del cumplimiento de los supuestos subyacentes al modelo [7], y la disponibilidad de los datos sobre los parámetros de entrada al modelo [5].

El modelo de [8] fue desarrollado usando modelos de análisis de regresión a partir de 43 pedones to-

mados en 10 localidades del oeste y centro de Los Estados Unidos. [9] elaboraron otro modelo a partir de 5320 datos de horizontes de 1323 sitios en 32 estados de Los Estados Unidos basado también en un análisis de regresión lineal. El modelo de [10] fue desarrollado a partir de 160 determinaciones de suelo de los altos Llanos Occidentales de Venezuela. El modelo de [11] se obtuvo a partir de diez suelos de distintos órdenes tomados como representativos de áreas agrícolas en Venezuela. Las muestras fueron tomadas mayoritariamente en: los Llanos Occidentales, la cuenca del lago de Valencia y Quibor.

Las ecuaciones de [8, 9 y 11] estiman contenidos volumétricos de humedad; mientras que, las ecuaciones de [10], y las de [12] estiman contenidos gravimétricos.

Además, [12] establecen que los modelos de regresión lineal constituyen una alternativa válida para estimar la retención de humedad del suelo en forma económica. Sin embargo, el poder de predicción de cada ecuación depende del universo de suelos representado por la muestra utilizada para su determinación y del volumen de datos incluidos en el modelo.

Los modelos utilizados hasta el momento para estimar la retención de humedad han sido utilizados con fines predictivos, descartando la bondad de la regresión para explicar la interrelación de las variables, además estas propuestas no contemplan a los modelos de Regresión Multivariante, los estudios presentados anteriormente evidencian la restricción de los mismos a la regresión múltiple, dejando a un lado el efecto global de las variables involucradas en el estudio.

Ahora bien, el Biplot de Coeficientes de Regresión permite explicar las interrelaciones de las variables dependientes y determinar la importancia de las independientes en la predicción de los modelos de regresión, esto permite sugerir lo significativo y novedoso de esta metodología en el campo de las ciencias agrícolas.

## 2. MARCO TEÓRICO

### 2.1 Representación del Biplot de regresión con rango reducido

La representación Biplot de una matriz  $K$  de orden  $(n \times p)$  y rango  $r$ , con  $n$  individuos y  $p$  variables,

consiste según [13] en su aproximación mediante la Descomposición en Valores Singulares (DVS), la cual proporciona la mejor aproximación de rango  $q$  ( $q < r$ ) en el sentido de los mínimos cuadrados [14]:

$$K \approx U_q \cdot D_q \cdot V_q' \quad (1)$$

donde  $U_q$  de orden  $(n \times q)$  contiene las primeras  $q$  columnas de  $U$ ,  $V_q$  de orden  $(p \times q)$  contiene las primeras  $q$  columnas de  $V$ , y  $D_q$  es una matriz diagonal que contiene los  $q$  mayores valores propios asociados a  $U_q$  y  $V_q$ .

Una vez realizada la DVS de la matriz se realiza la factorización en el producto de dos matrices, o sea:  $K \approx U_q D_q V_q' = AB'$ . Usualmente a las matrices  $A$  y  $B$ , se les llama matriz de marcadores filas y matriz de marcadores columnas respectivamente y representan las  $n$  filas y  $p$  columnas de  $X$  sobre el gráfico [15]

Los marcadores filas y columnas tienen propiedades que los caracterizan, y su representación permite, considerando las propiedades geométricas del producto interno entre ellos, reconstruir con alta calidad los datos originales ( $k_{ij} = a_i' b_j$ ), y aproximar gráficamente el orden de los individuos y/o la variabilidad y correlación de las variables, a través de las proyecciones de los marcadores filas  $a_i$  sobre los marcadores columnas  $b_j$ , y de la longitud y el ángulo que forman los ejes Biplot.

Se resalta el hecho de que en una representación Biplot Clásica [13] las  $p$  variables que conforman la matriz de datos constituyen un solo conjunto, por lo que la aproximación de la matriz se hace en forma simétrica, sin considerar relaciones de dependencia funcional entre variables, o sea, sin distinción entre dependientes e independientes [16].

Las representaciones Biplot desde la perspectiva del Análisis No Simétrico, pueden enriquecerse si se obtienen a través de aproximaciones Mínimo Cuadráticas como en los Modelos de Regresión Lineal Multivariante, en cuyo caso, se pueden adaptar las ideas expuestas por [17] al proponer una representación Biplot de la matriz de coeficientes de regresión con la cual es posible interpretar sobre la representación gráfica, las relaciones entre las variables, así como a sus estadísticos  $t$  asociados [18].

## 2.2 Biplot de regresión con rango reducido

Cuando la relación entre los dos conjuntos de variables es asimétrica, se parte de una regresión multivariante de  $Y$  sobre  $X$  usando el modelo

$$Y = X \Theta + E \quad (2)$$

con  $X(n \times p)$  y  $Y(n \times q)$ , matrices que contienen observaciones de  $n$  unidades para  $p$  variables independientes y  $q$  variables dependientes, respectivamente,  $\Theta$  la matriz  $(p \times q)$  de coeficientes de regresión y  $E$  la matriz  $(n \times q)$  de errores aleatorios. Los errores de las diferentes observaciones (filas de  $E$ ) son asumidos independientes e idénticamente distribuidos con esperanza cero y matriz de covarianza  $\Sigma$  ( $q \times q$ ).

A diferencia del problema en regresión lineal clásica, en este modelo no se supone que la matriz  $\Theta$  es de rango completo, pudiendo existir restricciones lineales en la estimación de los coeficientes de regresión. Al considerar a  $\Theta'$  (la transpuesta de la matriz de coeficientes de regresión) y las variables de estudio estandarizadas se tiene que:

$$\hat{\Theta}' = [(X'X)^{-1} X'Y'] = [R_{xx}^{-1} R_{yx}]' \quad (3)$$

$$\hat{\Theta}' = [(X'X)^{-1} X'Y'] = [R_{xx}^{-1} R_{yx}]' \quad (4)$$

Donde  $\hat{\Theta}'$  es la estimación  $\Theta'$ . Una forma de caracterizar a  $\hat{\Theta}'$  es mediante un plano, aprovechando así las herramientas del Análisis de Datos Multivariante a través de un Biplot, en tal sentido, la matriz  $\hat{\Theta}'$  puede factorizarse en la forma  $A_0 B_0'$ , siendo  $A_0$  y  $B_0$  matrices de orden  $q \times r$  y  $p \times r$ , respectivamente.

Para realizar la aproximación gráfica, se utiliza una aproximación por mínimos cuadrados ponderados usando como matriz de pesos  $S$  a la inversa de la matriz de covarianza estimada de  $\hat{\Theta}'$  (esto permite considerar, además de los errores estándar, a las estimaciones de su covarianza). El estimado de la  $i$ -ésima fila y la  $j$ -ésima columna de la matriz de covarianza de  $\hat{\Theta}'$  es  $(n-p-1)^{-1} s_{ij} R_{xx}^{-1}$  [19], donde  $s_{ij}$  es el  $(i,j)$  elemento de la suma de cuadrados dada a continuación:

$$S_e = R_{yy} - R_{yx} \cdot R_{xx}^{-1} \cdot R_{xy} \quad (5)$$

De allí que la función de pérdida está dada por:

$$\left\| S_e^{-1/2} \cdot (\hat{\Theta}' - A_0 B_0') \cdot R_{xx}^{1/2} \right\|^2 \quad (6)$$

$$= \left\| S_e^{-1/2} \cdot R_{yx} \cdot R_{xx}^{-1/2} - (S_e^{-1/2} \cdot A_0) \cdot (R_{xx}^{1/2} \cdot B_0) \right\|^2 \quad (7)$$

El mínimo se obtiene de la DVS [17], como:

$$S_e^{-1/2} \cdot \hat{\Theta}' \cdot R_{xx}^{1/2} = U_1 \cdot D_1 \cdot V_1' \quad (8)$$

Donde  $U_1$  y  $V_1$  son matrices ortogonales de orden  $q \times t$  y  $p \times t$  que contienen a los vectores singulares, y  $D_1$  una matriz diagonal con los valores singulares ordenados en forma decreciente sobre la diagonal ( $d_1 \geq d_2 \geq \dots \geq d_t \geq 0$ ). El mínimo de (6) es por lo tanto  $d_1 + d_2 + \dots + d_t$ , y es conseguido ajustando

$$A_0 = [S_e^{1/2} U_1 D_1]_r \quad y \quad B_0 = [R_{xx}^{-1/2} V_1]_r \quad (9)$$

Pero, la DVS en (9), según [17], se puede escribir también como

$$V_1 = V, \quad y \quad S_e^{-1/2} \cdot U_1 \cdot D_1^{-1} = R_{yy}^{-1/2} \cdot U \cdot D^{-1} \quad (10)$$

De (9) y (10) se obtienen las expresiones

$$A_0 = [R_{yy}^{1/2} U D]_r \quad y \quad B_0 = [R_{xx}^{-1/2} V]_r \quad (11)$$

Esto es,  $A_0$  contiene las correlaciones inter-set de las variables dependientes, mientras que  $B_0$  contiene los pesos canónicos de las variables independientes [20]. En conclusión, estas correlaciones inter-set y estos pesos minimizan la función de pérdida (6). Graficando cada fila de  $A_0$  y  $B_0$  como vectores en un sistema de coordenadas cartesianas  $r$ -dimensional, se obtiene un Biplot que aproxima a  $\hat{\Theta}$  en el sentido de los mínimos cuadrados ponderados o generalizados como le llaman también algunos autores (p.e. [21]).

### 2.3 Biplot de coeficiente de Regresión con rango reducido bajo la perspectiva de los biplots generalizados.

Hasta ahora en este trabajo, se han estudiado a los biplots de coeficientes de regresión con rango reducido bajo la perspectiva asimétrica, partiendo de un modelo de regresión multivariante. Si se introducen a esta teoría algunos elementos adicionales, puede enfo-

carse usando los biplots generalizados [22]. Primero, cuando se introduce a  $S_e$  como matriz de pesos en (6), debe quedar claro que es posible introducir otras matrices con tales fines, siempre y cuando esta matriz sea simétrica y definida positiva, además ésta debe ser elegida convenientemente para aportar las mejores propiedades posibles al biplot que se utilice, respondiendo a tal consideración se puede tomar una matriz genérica  $\Gamma$  con las características antes señaladas.

Segundo, el análisis efectuado se ejecuta sobre  $\hat{\Theta}$ , es decir, se busca el mínimo de  $\tilde{\Theta} = S_e^{-1/2} \cdot \hat{\Theta}' \cdot R_{xx}^{1/2}$ , donde  $\hat{\Theta}$  es la matriz de los coeficientes de regresión generalizada. Esta última expresión puede escribirse, transponiendo y sustituyendo a la matriz de pesos  $S_e^{-1/2}$  por la matriz genérica definida positiva  $\Gamma$ , además, en algunos problemas, el interés del investigador puede basarse en el análisis de datos centrados (o datos brutos), en cuyo caso, se puede sustituir  $R_{xx}^{1/2}$  por  $(X'X)^{1/2}$ ; por consiguiente, para ejecutar aproximaciones Biplots en forma más general, se busca el mínimo de la expresión:

$$\tilde{\Theta} = (X'X)^{1/2} \hat{\Theta}' \Gamma \quad (12)$$

### 2.4 Biplot generalizado de $\hat{\Theta}$

Como se señaló anteriormente, la formalización del Biplot de Coeficientes de Regresión como un Biplot generalizado, se ejecuta por medio de la descomposición en valores singulares generalizada (DVSG), por tal motivo (12) puede minimizarse así:

$$\tilde{\Theta} = (X'X)^{1/2} \hat{\Theta}' \Gamma = UDV' \quad (13)$$

Luego, la factorización Biplot de la matriz estimada de coeficientes de regresión, está dada por:

$$\begin{aligned} \hat{\Theta} &= (X'X)^{-1/2} (UDV') \Gamma^{-1} \\ &= [(X'X)^{-1/2} U D^c] \cdot [\Gamma^{-1} \cdot V D^{1-c}]' = A \cdot B' \quad (14) \end{aligned}$$

Donde:

$$A = (X'X)^{-1/2} U D^c \quad y \quad B = \Gamma^{-1} \cdot V \cdot D^{1-c} \quad (15)$$

A partir de la factorización encontrada en (15), se pueden definir los biplots clásicos de [13], para los biplot de coeficientes de regresión con rango reducido, escogiendo el valor de  $c$  en la factorización de la si-

guiente manera:

Si  $c=1$ , entonces se referirá al Biplot de Coeficientes de Regresión con Rango Reducido para el caso JK-Biplot.

$$A = (X'X)^{-1/2}UD \text{ y } B = \Gamma^{-1} \cdot V \quad (16)$$

Si  $c=0$ , entonces se referirá al biplot de coeficientes de regresión con rango reducido para el caso GH-Biplot.

$$A = (X'X)^{-1/2} \cdot U \text{ y } B = \Gamma^{-1} \cdot VD \quad (17)$$

Si  $c=1/2$ , entonces se referirá al biplot de coeficientes de regresión con rango reducido para el caso SQRT-Biplot.

$$A = (X'X)^{-1/2} \cdot UD^{1/2} \text{ y } B = \Gamma^{-1} \cdot VD^{1/2} \quad (18)$$

Por otro lado, si se factoriza adecuadamente se obtendrá el caso HJ- Biplot de coeficientes de regresión con rango reducido.

$$A = (X'X)^{-1/2} \cdot UD \text{ y } B = \Gamma^{-1} \cdot VD \quad (19)$$

En [22] se evalúan las propiedades del GH-Biplot de coeficientes de regresión con rango reducido y se comprueba la teoría dada por [23] para sus marcadores filas y columnas cuando se introduce a la matriz de peso  $\Gamma$ . Así se tiene que, si  $\Gamma = I$ , se está en presencia del análisis de redundancia [23], y el Biplot de  $\hat{\Theta}$  se puede interpretar como un GH-Biplot clásico, donde sus filas representan los pesos canónicos de las variables independientes y sus columnas reproducen a las columnas de las variables originales  $Y$ . Por otro lado, si  $\Gamma = R_{yy}^{-1/2}$  entonces:  $B = R_{yy}^{1/2} VD$ ; y  $A$  puede expresarse como  $A = R_{xx}^{-1/2}U$ . De manera que,  $A$  contiene los pesos canónicos de las variables independientes, y  $B$  contiene las correlaciones interser de las variables dependientes.

La tabla 1 resume a la regresión multivariante, en sus filas están contenidas las columnas de la matriz  $\hat{\Theta}$ , éstas muestran a los coeficientes de regresión para cada una de las variables dependientes estimadas sobre el conjunto  $X$ . La última columna contiene a los coeficientes de correlación múltiple. Un Biplot de coeficientes de regresión con rango reducido representa gráficamente la información contenida en esta tabla, el

producto interno de marcadores filas y columnas aproxima a los coeficientes de regresión estimados, esto quiere decir, que si por ejemplo se proyecta ortogonalmente el marcador correspondiente a  $x_i$  sobre el vector que representa a  $y_j$ , se obtendrá una aproximación del coeficiente  $q_{ij}$ .

Tabla 1. Resumen de la regresión multivariante de las  $Y$  sobre las  $X$

$\hat{\Theta}$	$x_1$	$x_2$	...	$x_i$	...	$x_p$	$R_{y_j, \hat{y}_j}$
$\hat{y}_1$	$\theta_{11}$	$\theta_{21}$	...	$\theta_{i1}$	...	$\theta_{p1}$	$R_{y_1, \hat{y}_1}$
$\hat{y}_2$	$\theta_{12}$	$\theta_{22}$	...	$\theta_{i2}$	...	$\theta_{p2}$	$R_{y_2, \hat{y}_2}$
:	:	:	...	:	...	:	:
$\hat{y}_j$	$\theta_{1j}$	$\theta_{2j}$	...	$\theta_{ij}$	...	$\theta_{pj}$	$R_{y_j, \hat{y}_j}$
:	:	:	...	:	...	:	:
$\hat{y}_q$	$\theta_{1q}$	$\theta_{2q}$	...	$\theta_{iq}$	...	$\theta_{pq}$	$R_{y_q, \hat{y}_q}$

Además de la interpretación anterior, con este Biplot se pueden realizar otras interpretaciones, como por ejemplo: las proyecciones de los vectores para las variables independientes sobre los vectores de las dependientes que tengan el mismo sentido tendrán correlación positiva, si ocurre al contrario, la correlación será negativa. Los vectores de las variables dependientes que tienen la misma dirección tienen correlaciones positivas, y viceversa. La longitud de un vector para una variable dependiente aproxima el valor de su correlación múltiple. Es decir, los vectores para las variables dependientes con mayores longitudes están mejor explicados por el modelo de regresión. La longitud de los vectores de las variables independientes indica qué tan bien están representados; de esta manera, longitudes de vectores cercanas al origen indican una mala representación, y en caso contrario, se dice que la variable tiene mejor representación. Puesto que los pesos canónicos son una medida de importancia de las variables independientes, la longitud de la proyección ortogonal de una variable independiente sobre la dependiente, establece la contribución de la variable dependiente al modelo de regresión múltiple para esa variable dependiente. Esta propiedad es muy importante puesto que ayuda a determinar al mejor modelo de regresión múltiple.

No obstante, esta aproximación Biplot se ve afectada por los efectos de la multicolinealidad, la cual entre otras cosas, invalida la interpretación de un coeficiente de regresión como el efecto condicional de una variable independiente, dados los valores de las otras variables independientes, y por lo tanto, los mar-

cadore de los biplots de los coeficientes de regresión son inútiles. Así que bajo la presencia de multicolinealidad no se puede realizar representaciones biplots de los coeficientes de regresión.

Una alternativa ante el problema de multicolinealidad según [23] es usar métodos de estimación sesgados, tales como mínimos cuadrados parciales (PLS) como proponen [24] y [25], la regresión ridge con restricción en el rango de la matriz de coeficientes de regresión o la regresión por componentes principales [21]. A continuación se presenta una alternativa a este problema usando la regresión por componentes principales.

### 2.5 Regresión con componentes principales y los biplots de coeficientes de regresión con rango reducido.

Si se toman en cuenta la regresión por componentes principales el modelo de regresión multivariante queda como:

$$Y = Z \cdot \Theta_c + E \tag{20}$$

Donde:  $Z = XV$  y  $\Theta_c = V' \Theta$   
y de (20) se determina la estimación del coeficiente de regresión multivariante por la expresión:

$$\hat{\Theta}_c = (ZZ')^{-1} Z'Y \tag{21}$$

Si se utiliza a (21) para la regresión multivariante, entonces la matriz de coeficientes de regresión se puede expresar como:

$$\hat{\Theta}_c = V (D^2)^{-1} V' X' Y \tag{22}$$

En adelante se denotará a la descomposición en valores singulares de X como  $X = U_1 D_1 V_1'$  y a la descomposición en valores singulares de  $\hat{\Theta}_c$  (matriz de coeficientes de regresión generalizada) como  $\hat{\Theta}_c = UDV'$ .

En tal sentido,  $Z = XV_1$  y  $\hat{\Theta}_c = V_1 (D_1^2)^{-1} Z'Y$ . La teoría que se ha utilizado para los Biplots de coeficientes de regresión bajo el enfoque generalizado, experimenta unos ligeros cambios en cuanto a las expresiones matemáticas; sin embargo, si se estudian las propiedades de estos Biplots desde el modelo (20), se

obtendrán las mismas propiedades que las dadas al tomar a (2) como modelo de referencia.

### 2.6 Procedimiento para la representación Biplot de los coeficientes de regresión fundamentada en la regresión con componentes principales

Por lo expuesto anteriormente, si existe multicolinealidad entre las variables independientes se debe transformar el modelo de (2) a (9) por medio de  $Z = XV_1$ , y luego se determina el Biplot de coeficientes de regresión tal como se hizo en (15). Es decir, se ejecuta el procedimiento para los Biplots de coeficientes de regresión con componentes principales como sigue:

a.- Realice la DVS de las Variables independientes:  $X = U_1 D_1 V_1'$ .

b.- Escoja  $Z = XV_1$

c.- Realice la DVSG de  $\tilde{\Theta}_c = (Z' Z)^{1/2} \hat{\Theta}_c \Gamma = UDV'$

d.- A partir del paso anterior para el GH-Biplot se tiene que:

$$A = (Z' Z)^{-1/2} U = V_1 (D_1)^{-1} V_1 U \text{ y } B = \Gamma^{-1} VD$$

e.- Realice la representación GH Biplot de la matriz de regresión con componentes principales.

f.- Cada eje Biplot para las variables independientes es ahora una componente principal, por lo tanto use la teoría de estas últimas para una mejor interpretación, procurando eliminar aquellas variables independientes que presenten menor contribución a las componentes.

## 3. MATERIALES

Para realizar un Biplot de coeficientes de regresión usando componentes principales que permita representar la retención de humedad en suelos, se utilizan datos generados por un estudio de suelos de la depresión del Lago de Valencia, realizado por el Ministerio del Ambiente y de los Recursos Naturales Renovables (publicado sólo parcialmente), correspondiente a suelos aluviales, y presentados en [12]. El conjunto de variables dependientes está compuesto por la cantidad de agua retenida a -33Kpa (w33) y -1500Kpa (w15); las cuales son medidas sobre 45 muestras de suelos, y las variables explicativas están compuestas por Carbono Orgánico (C.O.) en porcenta-

je, Densidad aparente (Da) en  $\text{mg.m}^{-3}$ , y tres determinaciones de texturas, a saber: Arena, Limo y Arcillas, también medidas en porcentaje. Luego, la matriz Y es de orden 45x2, mientras que la matriz X es de orden 45x5.

#### 4. RESULTADOS Y DISCUSIÓN

##### 4.1 Diagnóstico de la colinealidad.

Se utilizó el SPSS versión 9.0 para realizar la diagnosis de colinealidad entre las variables independientes, los resultados están contenidos las Tablas 2 y 3. Aquí se limitará el estudio a la Tabla 2, por considerar que su discusión es similar a la de la Tabla 3, en el primero están contenidos los resultados para la regresión de variable dependiente w33, mientras que para el otro, se visualizan los resultados de la regresión para w15. Entre los distintos valores de tolerancia se aprecia que los más pequeños son 0.372 y 0.391, estos co-

rresponden a las variables arena y arcilla, respectivamente; indicando que ambas variables pueden ser explicadas como una combinación lineal obtenida a partir de ellas. Por otro lado, los factores de inflación de la varianza (VIF), para estas variables son muy grandes (2.691 y 2.558), develando una mayor varianza para los coeficientes de regresión correspondientes.

##### 4.2 Correlaciones entre variables.

La Tabla 4, muestra las correlaciones de todas las variables involucradas en el estudio, las variables con mayores coeficientes de correlación con respecto a las variables dependientes son: Carbono Orgánico, Arena y Arcilla; mientras que, la Densidad aparente (Da), tiene una relación más baja con las dependientes, pero significativa al 5%. Por otro lado, el limo no guarda relación con ninguna de las otras variables, por lo tanto, será considerada como variable de ruido o con información irrelevante.

Tabla 4. Correlaciones simples entre las variables de estudio.

		CAPA	CO	DA	ARENA	LIMO	ARCILLA	W33	W1500
CAPA	Correlación de Pearson	1.000	-0.662	0.182	-0.061	0.051	-0.138	-0.141	-0.226
	Sig. (2-colas)	.	0.000	0.231	0.689	0.739	0.364	0.357	0.135
	N	45	45	45	45	45	45	45	45
CO	Correlación de Pearson	-0.662	1.000	-0.238	-0.143	-0.058	0.237	0.362	0.409
	Sig. (2-colas)	0.000	.	0.115	0.350	0.704	0.117	0.014	0.005
	N	45	45	45	45	45	45	45	45
DA	Correlación de Pearson	0.182	-0.238	1.000	0.342	0.036	-0.234	-0.301	-0.320
	Sig. (2-colas)	0.231	0.115	.	0.022	0.815	0.122	0.045	0.032
	N	45	45	45	45	45	45	45	45
ARENA	Correlación de Pearson	-0.061	-0.143	0.342	1.000	-0.129	-0.742	-0.865	-0.772
	Sig. (2-colas)	0.689	0.350	0.022	.	0.399	0.000	0.000	0.000
	N	45	45	45	45	45	45	45	45
LIMO	Correlación de Pearson	0.051	-0.058	0.036	-0.129	1.000	-0.105	0.146	-0.063
	Sig. (2-colas)	0.739	0.704	0.815	0.399	.	0.493	0.337	0.682
	N	45	45	45	45	45	45	45	45
ARCILLA	Correlación de Pearson	-0.138	0.237	-0.234	-0.742	-0.105	1.000	0.695	0.876
	Sig. (2-colas)	0.364	0.117	0.122	0.000	0.493	.	0.000	0.000
	N	45	45	45	45	45	45	45	45
W33	Correlación de Pearson	-0.141	0.362	-0.301	-0.865	0.146	0.695	1.000	0.857
	Sig. (2-colas)	0.357	0.014	0.045	0.000	0.337	0.000	.	0.000
	N	45	45	45	45	45	45	45	45
W1500	Correlación de Pearson	-0.226	0.409	-0.320	-0.772	-0.063	0.876	0.857	1.000
	Sig. (2-colas)	0.135	0.005	0.032	0.000	0.682	0.000	0.000	.
	N	45	45	45	45	45	45	45	45

\*\* . Correlación significativa al 1% (2-colas).

\* . Correlación significativa al 5% (2-colas).

### 4.3 Biplots de regresión por mínimos cuadrados ordinarios (OLS)

La representación Biplot de las estimaciones de los coeficientes de regresión realizada por OLS, es apreciada en la Figura 1. Este ajuste indica que las variables dependientes (w15 y w33) guardan una correlación positiva entre ellas. Por otro lado, el plano indica que los coeficientes de regresión tanto para w15 y como para w33 con las variables densidad aparente, carbono orgánico, arena y arcilla son positivos, mientras que el de limo es negativo; obsérvese que las Tablas 1 y 2 no corroboran esta información. En otros términos, puede verse que en este plano Densidad aparente está mal representada, y desde luego, no reproduce su relación con las otras variables. En general, esta representación no es capaz de aproximar las relaciones de las variables de estudio.

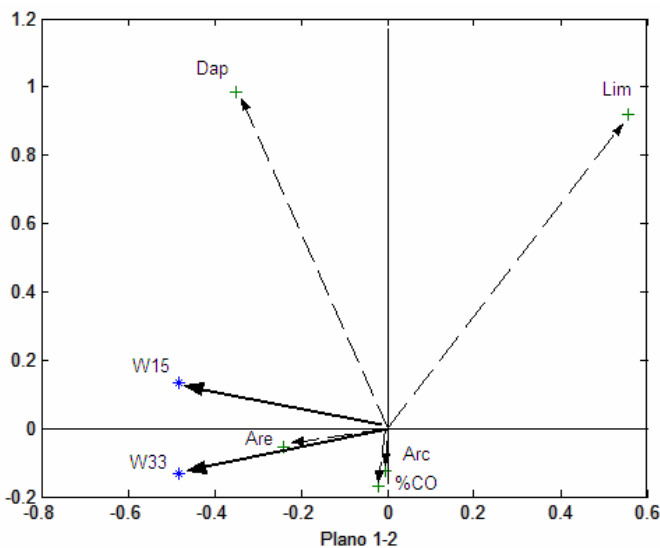


Figura 1: Biplot de regresión por OLS

### 4.4 Biplots de regresión con componentes principales

La Figura 2 muestra el Biplot de los coeficientes de regresión con componentes principales. Puede verse que las componentes mejor representadas son F1 y F4, las cuales se han formado mayormente de la contribución de las variables Arena y Arcilla (Tabla 5), obsérvese que las componentes F2 y F3 no tienen una buena representación en este plano; sin embargo, la magnitud y dirección de la representación de los coeficientes de regresión se ajustan a los coeficientes suministrados por las Tablas 6 y 7. En consecuencia, la retención de humedad a -33Kpa y -1500Kpa, puede ser explicada por estas dos componentes, y se sugiere además que la mejor explicación del modelo es suministrada por las variables Arcilla y Arena.

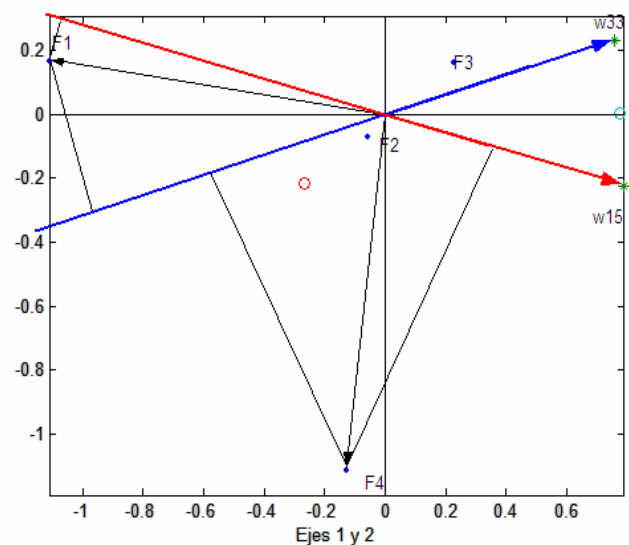


Figura 2: Biplot de regresión con componentes principales.

Tabla 5. Resumen de las componentes principales: correlaciones y contribuciones.

Variable	Correlación variable-factor				Contribución de la variable al factor en porcentaje			
	F1	F2	F3	F4	F1	F2	F3	F4
Carbono Orgánico	0.45	0.76	0.47	0.06	10.2	59.2	29	1.7
Densidad Aparente	-0.58	-0.41	0.7	0.07	16.8	17.6	64	2.3
Arcilla	-0.87	0.35	0.01	-0.34	37	13	0	50
Arena	0.86	-0.32	0.23	-0.33	36	10.2	7	46



Tabla 6. Coeficientes de regresión con componentes principales para w33.

Variable	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Std. Error	Beta		
(Constante)	25.698	0.666		38.566	0.000
Factor 1	-8.104	0.674	-0.829	-12.026	0.000
Factor 2	-1.315	0.674	-0.135	-1.952	0.058
Factor 3	1.513	0.674	0.155	2.245	0.030
Factor 4	-2.780	0.674	-0.284	-4.125	0.000

Tabla 7. Coeficientes de regresión con componentes principales para w1500.

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Std. Error	Beta		
(Constante)	14.440	0.404		35.771	0.000
Factor 1	-5.847	0.408	-0.881	-14.324	0.000
Factor 2	-0.762	0.408	-0.115	-1.867	0.069
Factor 3	1.471	0.408	0.222	3.603	0.001
Factor 4	0.642	0.408	0.097	1.572	0.124

## 5. CONCLUSIONES

Las representaciones Biplots para los coeficientes de regresión aproximan a las estimaciones de estos coeficientes por medio del producto interno de marcadores filas y columnas; y con la incorporación de los fundamentos de los Biplots generalizados, permiten, además, realizar una mejor interpretación de los Biplots de coeficientes de regresión con rango reducido, y suministran información valiosa para verificar sus propiedades. Estos principios pueden ser adaptados para realizar visualizaciones de los coeficientes de regresión con componentes principales, como lo describe el procedimiento proporcionado en este ensayo.

Cuando se está en presencia de multicolinealidad, la escogencia del Biplot de Coeficientes de Regresión con Componentes principales resulta ser una buena alternativa, debido a que la representación convencional por mínimos cuadrados ordinarios resulta ser inadecuada. Las ecuaciones de Pedotransferencia pueden tener una buena explicación con los Biplots de Coeficientes de Regresión, siempre y cuando no exista un serio problema de multicolinealidad, en caso contrario, se recomienda el uso de componentes principales para los coeficientes de regresión y la posterior aplicación de las representaciones Biplots.

Las dos componentes mejor representadas en el Biplot de Regresión con Componentes Principales se

formaron mayormente gracias a la contribución de las variables Arena y Arcilla. En consecuencia, la retención de humedad a -33Kpa y -1500Kpa puede ser explicada por estas dos componentes, y se sugiere además que la mejor explicación del modelo es suministrada por las variables que tienen un alto grado de colinealidad.

## 6. REFERENCIAS

- [1] Casanova, E. (1996). Introducción a la Ciencia del Suelo. Universidad Central de Venezuela. Consejo Científico y Humanístico.
- [2] Kern, J. (1995). Evaluation of soil water retention models based on basic soil physical properties. Soil Sci. Soc. Am. J., 59, 1134-1141.
- [3] Russel, (1973). Soil conditions and plant growth. London and New York. Editorial Longman.
- [4] Williams, R., Ahuja, L., y Naney, J. (1992). Comparison of methods to estimate soil water characteristics from soil texture, bulk density, and limited data. Soil Sci. Soc. Am. J., 153, 172-184.
- [5] Tietje, O. y Tapkenhinrichs, M. (1993). Evaluation of pedotransfer functions. Soil Sci. Soc. J., 57, 1088-1095.
- [6] Saxon, K.; Rawls, W.; Romberger, J. y Papendick (1986). Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J., 50, 1031-1036

- [7] Bouma, J y Van Lanen, J. (1987). Transfers functions and trehold values: from land characteristics to land qualities. In: K.J. Beek et. al. (ed). Quantified Land Evaluation. Proceedings of Workshop. ISSS and SSSA. ITC Publications. Enschede, The Netherlands. p.106-110.
- [8] Gupta, S. y Larson, W. (1979). Estimating soil water retention characteristics from particle size distribution, orgnic matter percent, and bula density. *Water Resour. Res.*, 15, 1633-1635.
- [9] Rawls, W. ; Brakensiek, D. y Saxton, K. (1982). Estimation of soil water properties. *Trans. ASAE*, 25, 1316-1320.
- [10] Delgado, F. y Barreto, L. (1988). Una aproximación matemática para la elaboración de curvas de retención de humedad en suelos representativos de Los Llanos Occidentales. *Revista UNELLEZ de Ciencia y Tecnología* ,6 (1-2), 45-50.
- [11] Malave, N. (1991). Estudio preliminar de las relaciones entre características de retención de humedad y distribución de tamaño de partículas, materia orgánica y densidad aparente. Trabajo de grado. Universidad Central de Venezuela, Facultad de Agronomía, Departamento de Edafología. 92p.
- [12] Pineda, C. y Viloria, J. (1997). Funciones de Pedotransferencia para estimar la retención de humedad en suelos de la cuenca del lago de Valencia. *Venesuelos*, vol. 4, N° 1 y 2, 39-44.
- [13] Gabriel, K. R (1971). The Biplot-graphic display of matrices with applications to principal component analysis. *Biometrika*, 58, 453-467.
- [14] Householder, A. y Young, G. (1938). Matrix approximation and Latent Roots. *American Mathematics Monthly* 45,165 -171.
- [15] Cárdenas, O. y Galindo, P. (2004). Biplot con Información Externa basado en Modelos Lineales Generalizados. Ediciones del Consejo de Desarrollo Científico y Humanístico de la Universidad Central de Venezuela.
- [16] Cárdenas, O., Noguera, C., Galindo, P. y Vicente-Villardón, J. (2006). Alternativa a la regresión con componentes principales basada en biplot de regresión. *Revista Interciencia* ISSN 0378-1844. Vol .31, N° 3, p.160-167.
- [17] Ter Braak, C. (1990). Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika* 55, 519 – 531.
- [18] Amaro, I. ; Vicente-Villardón, J. ; Galindo-Villardón, M.(2004). Manova biplot para arreglos de tratamientos con dos factores basado en modelos lineales generales multivariantes. *Revista Interciencia* ISSN 0378-1844. Vol .29, p.26-32.
- [19] Anderson T. (1984). An introduction to multivariate statistical analysis (2nd. ed.). New York: Wiley.
- [20] Gittins, R. (1985). Canonical Análisis a review with Applications in Ecology. Springer. Velarg. Berlin Heidelberg New York-Tokyo.
- [21] Montgomery D. y Peck E. (1992). Introduction to Linear Regression Analysis. Second Edition. John Wiley & Sons.
- [22] Álvarez W. (2007). Una contribución a los Biplots de Coeficientes de Regresión usando la Descomposición en Valores Singulares Generalizada. Trabajo de Ascenso para la categoría de Profesor Agregado. Universidad Nacional Experimental Rómulo Gallegos.Venezuela.
- [23] Ter Braak, C. and Looman, C. (1991). Biplots in reduced rank regression. *Biometrical Journal* 8, 983-1003.
- [24] Aastveit, A. and Martens, H. (1984). ANOVA interactions by partial least squares regresión. *Biometrics* 42, 829 – 844.
- [25] Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics* 2, 211- 228.