

Aplicación del Biplot en el estudio de propiedades de suelo de dos comunidades de Atlixco, Puebla.

Verónica Animas S.^a, Gladys Linares F.^b, Rossana Schiaffini A.^c, Sonia Silva G.^d

^{a,b,d} *Instituto de Ciencias de la Benemérita Universidad Autónoma de Puebla (ICUAP), Puebla, México.*

^c *Facultad de Derecho y Ciencias Sociales de la Universidad Autónoma de Puebla (BUAP).*

^a veronicad.animasserrano@viep.com.mx, ^b gladys.linaresfleites@viep.com.mx,

^c rossana.schiaffiniaponte@viep.com.mx, ^d soemsgo@gmail.com

Resumen

En este trabajo se analizaron las relaciones entre algunas propiedades fisicoquímicas del suelo, con datos recuperados de 13 muestreos de las comunidades San Juan Portezuelo y 10 de Abril, pertenecientes al municipio de Atlixco en el estado de Puebla. Se generó un análisis de correlación, donde se encontró alta correlación positiva de la M. O. con el Nitrógeno, y alta correlación negativa de la M. O. con el PH. Además, se hizo un Análisis de Componentes Principales en donde se definieron dos nuevas componentes consiguiendo acumular un 82 % de la información con una reducción de dimensionalidad de datos. Para la representación gráfica de datos multivariantes, se utilizó la técnica de Biplot, dando muestra de la variabilidad del conjunto de individuos (muestras) y variables (propiedades), y se identificaron perfiles similares de suelo.

Palabra claves: Propiedades fisicoquímicas del suelo, Reducción de Dimensionalidad, Técnica Biplot.

Introducción

Las ciencias, en general, sean éstas sociales, económicas o naturales, tienen una base estadística dada por las implicaciones y transversalidad de esta rama matemática. Dentro del proceso de Investigación Científica y la construcción del conocimiento, es necesario profundizar en la teoría estadística para esclarecer la realidad de manera confiable. Dentro de las ciencias ambientales, este análisis fehaciente de datos da pauta para el entendimiento de cada compartimento, sus interrelaciones y sus funciones como sistemas complejos.

El suelo, aunque muchas veces desvalorizado, además de albergar y conservar una gran cantidad de biota en su interior y

superficie, lleva a cabo funciones tan importantes como la producción y/o suministro de alimentos, fibras, combustibles, materiales de construcción, productos farmacéuticos (entre otros). Otras capacidades de los suelos, menos conocidas, pero de vital importancia y fundamentales para el equilibrio ecológico, son la retención de carbono, el ciclaje de nutrientes y la regulación del clima. No obstante, el suelo es susceptible a sufrir desequilibrios, situación que altera sus propiedades naturales degradándolo, y por tanto afectando sus funciones.

La Materia Orgánica del Suelo (MOS), describe a los elementos orgánicos en diversas etapas de descomposición, así como los organismos vivos y raíces del suelo (biomasa).

La MOS es dinámica y fundamental para la estabilización de la estructura del compartimento, permitiendo la retención y liberación de nutrientes requeridos por las plantas, además, facilita la disponibilidad del agua, con lo cual mejora la fertilidad del suelo y en definitiva, mejora la productividad de los alimentos. En resumen, es fundamental para mantener el equilibrio, la funcionalidad y la salud del suelo.

En este artículo se realiza un análisis exploratorio basado en datos de muestras de suelo de dos comunidades pertenecientes al municipio de Atlixco en el Estado de Puebla: San Juan Portezuelo y 10 de Abril.

En el se contemplan las siguientes propiedades: PH (en relación suelo/agua), Conductividad Eléctrica (CE), porcentaje de Materia Orgánica, porcentaje de Nitrógeno (N), Capacidad de Intercambio Catiónico (CIC) Cmol(+) kg.

Para conocer la similitud entre variables y posibilitar la construcción de un modelo predictivo de tendencias de la M. O. en dichas comunidades, se hizo un Análisis de Componentes Principales.

Para la representación gráfica de datos se eligió utilizar la técnica de Biplot, dada la utilidad en su interpretación fundamentada en conceptos geométricos sencillos.

Metodología

Aspectos generales

Los datos analizados fueron extraídos de una investigación de Fitoextracción de suelos contaminados por elementos potencialmente tóxicos en dicha región, realizado en el 2017 dentro del Instituto de Ciencias de la BUAP.

R-Commander, el software utilizado, utiliza un lenguaje de programación estadística gratuito que permite acceder a muchas capacidades del entorno estadístico R, sin necesidad de conocer el lenguaje de comandos propio de este entorno, es bastante intuitivo gráficamente y de fácil manejo a través de sus menús desplegables.

Metodología Estadística

Como análisis preliminar se buscó una homogeneidad en los datos para lo cual se seleccionaron los elementos más adecuados en cuanto a clase, resultando en la siguiente base de datos.

Tabla 1. Propiedades del suelo Comunidades de San Juan Portezuelo y 10 de Abril.

N.	Localidad	PH	CE	% M.O.	N%	CIC
1	SJP	6.64	0.22	2.30	0.15	13.0
2	SJP	6.90	0.21	1.30	0.05	23.0
3	SJP	7.04	0.18	1.20	0.06	28.0
4	SJP	6.82	0.15	1.40	0.09	20.0
5	SJP	6.74	0.19	2.10	0.11	16.0
6	SJP	7.18	0.19	1.10	0.05	22.0
7	10A	6.72	0.16	1.50	0.07	30.0
8	10A	6.22	0.29	3.50	0.19	18.0
9	10A	6.18	0.23	4.20	0.25	25.0
10	10A	6.66	0.13	2.20	0.11	28.0
11	10A	6.24	0.17	3.20	0.18	20.0
12	10A	6.46	0.09	3.00	0.15	16.0
13	10A	5.86	0.21	4.50	0.28	19.0

Análisis de correlación.

La correlación es un tipo de asociación entre dos variables numéricas, que evalúa la tendencia creciente o decreciente en los datos. Dos variables están asociadas cuando una variable nos da información acerca de la otra. De otra forma (si no nos dice nada sobre el comportamiento de la otra variable), no existe correlación.

Para encontrar correlaciones entre las propiedades (variables), se hizo un análisis de correlación, con un trazado por grupos, vinculados a sus respectivos gráficos de dispersión.

Análisis de Componentes Principales.

Para reducir la dimensionalidad y poder hacer una visualización adecuada de los datos se hizo un Análisis de Componentes Principales. Esta técnica es utilizada para describir un conjunto de datos en términos de nuevas variables (componentes) no correlacionadas. Los componentes se ordenan por la cantidad de varianza original que describen, por lo que la técnica permite reducir la dimensionalidad. A partir de la proyección de los datos de entrada sobre las direcciones de máxima varianza se obtendrá un nuevo espacio de representación de los datos en el que se puede eliminar fácilmente aquellas componentes con menor varianza, garantizando la mínima pérdida de información.

El ACP se emplea sobre todo en análisis exploratorio de datos y para la construcción de modelos predictivos.

Grafica Biplot.

Entre los métodos de análisis multivariante los investigadores han dado un gran impulso a las aplicaciones de los Biplot en diversos campos de la ciencia, dada la alta confiabilidad en sus resultados.

De la misma manera que un diagrama de dispersión muestra la distribución conjunta de dos variables, un Biplot representa tres o más variables, aproximando la distribución de una muestra multivariante en un espacio de dimensión reducida, normalmente de dimensión dos, y superpone sobre la misma, representaciones de las variables sobre las que se mide la muestra. La similitud entre individuos es una función inversa de la distancia entre los mismos, sobre la

representación biplot, las longitudes y los ángulos de los vectores que representan a las variables, se interpretan en términos de variabilidad y covariabilidad respectivamente

Resultados y Discusión

Resultados del Análisis de Correlación.

En este caso, se evidenció que la M.O. tiene alta relación directa con el Nitrogeno, es decir, que las magnitudes crecen en relación una de la otra, y alta relación inversa (negativa) con el pH.

Tabla 2. Matriz de Correlación salida R-Commander

```

Pearson correlations:
      CE      CIC      M..O.      N.      PH
CE    1.0000  -0.1799  0.3107  0.3420  -0.2746
CIC   -0.1799  1.0000  -0.3071  -0.3408  0.2769
M..O. 0.3107  -0.3071  1.0000  0.9841  -0.9658
N.     0.3420  -0.3408  0.9841  1.0000  -0.9538
PH    -0.2746  0.2769  -0.9658  -0.9538  1.0000

Number of observations: 13
  
```

Resultados del Análisis de Componentes Principales ACP.

En la siguiente tabla 3, se presenta la salida del ACP dado por el Software.

Tabla 3. Salida del ACP del software.

```

Component loadings:
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
CE    0.2465324  0.6590850  0.7087948548  0.04652984  0.01654107
CIC   -0.2459637  -0.6631474  0.7054028007  -0.04205162  -0.01938895
M..O. 0.5436330  -0.2031691  0.0016662473  -0.29740944  0.75810969
N.     0.5466237  -0.1513307  -0.0002854843  -0.52177022  -0.63722621
PH    -0.5332816  0.2483221  0.0037262252  -0.79710106  0.13624580

Component variances:
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
3.21683567  0.90325283  0.82011261  0.04595775  0.01384114

Importance of components:
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
Standard deviation  1.7935539  0.9503961  0.9056007  0.21437759  0.117648367
Proportion of Variance 0.6433671  0.1806506  0.1640225  0.00919155  0.002768228
Cumulative Proportion 0.6433671  0.8240177  0.9880402  0.99723177  1.000000000
  
```

Carga de los Componentes: Cada elemento de la columna “Component Loadings”, muestra los coeficientes de la ecuación de cada componente principal. Así,

vemos que el componente principal 1 (CP1), puede escribirse con la siguiente ecuación:

$$CP1 = 0.2465324 * Z1 - 0.2459637 * Z2 + 0.5436330 * Z3 + 0.5466237 * Z4 - 0.5332816 * Z5$$

Y el el componente principal 2 (CP2), puede escribirse con la siguiente ecuación:

$$CP2 = 0.6590850 * Z1 - 0.6631474 * Z2 - 0.2031691 * Z3 - 0.1513307 * Z4 + 0.2483221 * Z5$$

Las variables originales (CE, CIS, M.O, N, PH) han sido sustituidas por Z (Z 1 - Z 5), ya que son las variables estandarizadas y obsérvese que se destacan la M.O. y el N. En oposición al pH.

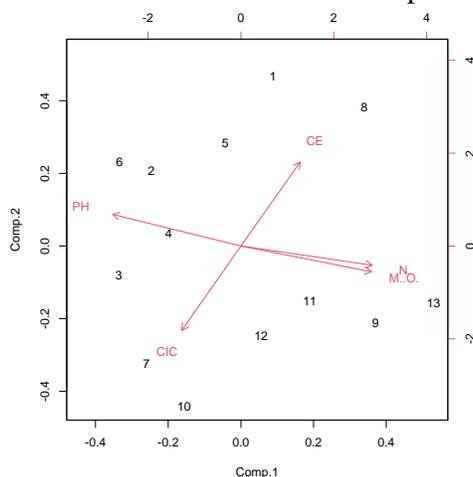
Desviación Estándar: muestra las desviaciones estándar de cada componente principal. Se calcula a partir de los datos obtenidos sustituyendo los valores de cada individuo en la ecuación de cada componente principal.

Proporción de Varianza: es la proporción de la varianza que explica cada componente principal. Su suma es igual a 1. Esta fila es realmente la importante de nuestros resultados. La componente 1 contiene la mayor proporción de varianza, 0.64, y la componente 2 apenas un 0.18 de varianza.

Porción Acumulada: es la proporción acumulada, calculada sumándolas progresivamente. Los dos primeros CP representan el 0.82 de proporción acumulada.

Resultados de la Gráfica de Biplot.

Gráfico 1. Grafica de Biplot.



En el Gráfico 1, observamos las 13 muestras de cada comunidad (SJP del 1 al 6 y 10A del 7 al 13). Podemos notar que en el eje de abscisas se representa el Componente 1 y en el de ordenadas el componente principal 2. Cuanto más largas sean las flechas rojas, más alto es el valor del coeficiente de esas variables en ese componente. Esa longitud de los marcadores columna (vectores) aproximan la desviación típica de los indicadores.

Los cosenos de los ángulos se interpretan en terminos de correlación entre las propiedades, de modo que ángulos agudos se asocian a indicadores con alta correlación positiva (variables M.O. y N), ángulos obtusos indican correlación negativa (variables M.O. y PH) y ángulos rectos señalan variables no correlacionadas (variables CE y CIC).

Otra característica notable en el gráfico, es que las distancias entre los marcadores fila se interpretan como una función inversa de sus similitudes, con ello los individuos próximos son más similares permitiendo la identificación de perfiles similares, obteniendo con ello una utilidad de agrupamiento o clusterización de grupos. Entonces, la M.O. y el N, formarían un grupo.

Discusión

En el gráfico, es posible identificar las propiedades predominantes en los suelos de cada muestra, y por tanto, de cada localidad.

Obsérvese que en la comunidad 10 de Abril se presentan mayor porcentaje de M. O. y N. que en la localidad San Juan Portezuelo. Por el contrario, el PH alcanza mayores valores en la comunidad San Juan Portezuelo que en la 10 de Abril.

Conclusiones

En la búsqueda de relaciones entre propiedades del suelo, las técnicas estadísticas

seleccionadas (Análisis de Correlación, ACP, Gráfica Biplot), fueron oportunas y eficaces. Se concluye que la Materia Orgánica del suelo, guarda alto grado de correlación con el Nitrógeno, es decir, la cantidad de M.O. que hay en el suelo nos da información de la cantidad de Nitrógeno en el mismo y viceversa. Así mismo, tienen un perfil similar, respecto a las demás propiedades muestreadas.

Además, fue posible hacer una sencilla reducción de datos para tener una mejor visualización de la dinámica de los elementos. Se resalta la división de las comunidades.

Referencias

Avila G. D. y Tamariz F. J. (2017). Fitoextracción de Suelos Contaminados por Elementos Potencialmente Tóxicos en la Región de Atlixco, Puebla. Tesis Buap. Pp. 149

FAO. (2016). Estado mundial del recurso del suelo (EMRS) - Resumen Técnico. Organización de las Naciones Unidas para la Alimentación y Agricultura Roma, Italia. Pp. 80

FAO. (2017). Carbono Orgánico del Suelo: el potencial oculto. Organización de las Naciones Unidas para la Alimentación y Agricultura. Roma, Italia. Pp. 78

Cárdenas O. Galindo P. y Vicente-Villardón J. L. (2007) Los Métodos Biplot: Evolución y Aplicaciones. Revista Venezolana de Análisis de Coyuntura, enero-junio, año/vol. XIII, número 001 Universidad Central de Venezuela. Caracas, Venezuela. Pp. 279-303

Fox J, Bouchet-Valat M (2020). Rcmdr: R Commander. R package version 2.7-1, <https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/>.

Polanco J. (2016). The role of principal component analysis in the evaluation of air quality monitoring networks. Comunicaciones en Estadística. Vol. 9, No. 2. Pp. 271–294