

# Aplicación de herramientas de software al análisis del relleno sanitario de São Giácomo de Caxias do Sul

## Application of software tools to the analysis of São Giácomo's sanitary landfill of Caxias do Sul

Muñoz H., J. A.,<sup>I</sup> Andrade, M. Z.,<sup>II</sup> Grisaa, A. M.<sup>II</sup>

**Resumen.** El presente artículo describe la aplicación de herramientas de software a los parámetros físico químicos del lixiviado del relleno sanitario de São Giácomo - Caxias do Sul, Brasil. Se presenta inicialmente una descripción del relleno sanitario, luego se describen las variables que intervienen en el proceso explicando lo relacionado con tratamiento de datos faltantes, y se organizan las matrices de datos fisicoquímicos que serán sometidas al estudio. Se aplican los algoritmos MPCA (*Multiway principal component anlysis*) para seleccionar las celdas del relleno que son representativas en el análisis final, y el algoritmo PCA para interpretar las variables correlacionadas con la degradación y la actividad biológica. En el análisis de los datos multivariados se usaron las herramientas Matlab, InfoStat y Norm, con el objetivo de estudiar la degradación de materia orgánica y actividad biológica en el relleno.

Un indicativo de la actividad biológica llevada a cabo en la celda del relleno sanitario se observa cuando la concentración de nitrógeno amoniacal es superior a la concentración de nitrógeno orgánico. Los parámetros importantes para degradación de materia orgánica son: pH, ST, Alcalinidad, DQO, DBO, Nitrógeno total y Nitrógeno amoniacal, y los parámetros que miden la biodegradabilidad son DQO y DBO.

**Palabras clave:** Relleno sanitario, análisis multivariado, degradación, componentes principales, InfoStat.

**Abstract.** This paper describes an application of advanced software tools to the physico-chemical parameters of leached in the sanitary filling of São Giácomo, located in Caxias do Sul, Brasil. At the beginning, a description of the sanitary filling is made, as long as a description of the variables that takes part in the process. An explanation of the missing data is mentioned and the matrixes containing the physico-chemical data are organized to be studied later on. Algorithms of the MPCA type are applied (Multiway Principal Component Analysis) in order to choose the representative filling cells for the final analysis; also a PCA algorithm to interpret the correlated variables with the degradation and the biological activity is used. Matlab, InfoStat and

I. Universidad del Tolima, Facultad de Ingeniería Agronómica. amunoz@ut.edu.co

II. Universidad de Caxias Do Sul, Centro de Ciências Exatas e Tecnologia.

Norm software tools were used for the analysis of the multivariate data, to cope with the objective of studying the degradation of organic matter and the biological activity in the sanitary filling of São Giacomo, at Caxias Do Sul. The important parameters to measure biodegradability of organic matter are: pH, ST, alkalinity, DQO, DBO, total nitrogen and ammonia nitrogen. DQO and DBO are parameters which measure biodegradability.

**Key words:** sanitary filling, multivariate analysis, degradation, principal components, InfoStat.

## 1. INTRODUCCIÓN

El sistema de São Giacomo, una disposición de residuos en celdas de relleno fue implantado según normas brasileras para relleno sanitario, opera con licencia del órgano ambiental estatal FEPAM (Fundação Estadual de Proteção Ambiental), y es monitoreado mensualmente.

La finalidad del monitoreo de un relleno sanitario es evaluar la evolución del proceso de descomposición de residuos sólidos en su interior y evaluar la influencia de ese proceso sobre el medio ambiente. Por medio del monitoreo se pueden evaluar los procesos que se señalan a continuación. Según Reichert (2000), los lixiviados (líquidos percolados) se definen como aquellos provenientes de tres fuentes principales: humedad natural de los residuos, agua de constitución de varios materiales que sobra durante una descomposición y líquido proveniente de la disolución de materiales orgánicos por enzimas expelidas por las bacterias. Los líquidos percolados que llegan al fondo del relleno provienen principalmente de aguas de precipitaciones, de aguas que se filtran en el lugar en que se depositan, y de otras infiltraciones.

La caracterización de los percolados líquidos permite conocer un poco mejor la evolución de los procesos biológicos que ocurren dentro del relleno sanitario, así como su edad y la composición de residuos sólidos urbanos depositados.

El trabajo que se presenta en este artículo corresponde al estudio en condiciones bióticas del relleno sanitario de São Giacomo de Caxias Do Sul-RS/ Brasil (fig. 1), utilizando el método exploratorio de análisis multivariado de 12 parámetros fisico-químicos del lixiviado: pH, alcalinidad, sólidos volátiles totales (STV), demanda química de oxígeno (DQO), demanda bioquímica de oxígeno (DBO), nitrógeno Total (N.T.), nitrógeno amoniacal (N-NH<sub>3</sub>), cadmio (Cd), cobre (Cu), hierro (Fe), manganeso (Mn) y zinc (Zn), producido en las celdas del relleno sanitario a lo largo

del tiempo, apuntando a la determinación de procesos de degradación/ biodegradación.



**Figura 1.** Celda del relleno sanitario de São Giácomo de Caxias do Sul-RS.

Según Ehrig (1983), para prever una calidad de los líquidos percolados es útil entender los procesos biológicos que ocurren en el relleno sanitario. En un relleno ocurren básicamente los procesos biológicos aeróbicos y anaeróbicos, siendo la penetración de oxígeno baja, con excepción del área de superficie. Por esta razón, el proceso anaeróbico es predominante.

El relleno sanitario está dividido en 7 celdas (cel1, cel3, cel4, cel5, cel7, cel9 y cel10). La disposición de residuos sólidos de forma controlada permite condiciones adecuadas para el desarrollo de procesos de digestión anaeróbica de materia orgánica presente en los residuos, y el desarrollo de actividad microbiana en el interior de las celdas del relleno. Éstas, en asociación con factores ambientales, promueven una percolación de cantidades de lixiviado para su posterior tratamiento. La cantidad actual diaria de residuos que entran a la unidad del relleno es de 340 toneladas. El proceso total de degradación anaeróbica ocurre básicamente en 3 fases distintas: fase aeróbica, fase acetogénica y fase metanogénica, siendo difícil que las tres fases ocurran juntas en los primeros 15 años del relleno.

Los rellenos viejos tienen una cantidad significativa de material biológicamente inerte, lo que es indicado por la baja relación DBO/DQO, o por la gran diferencia entre estos valores (Lo, 1996). La relación DBO/DQO indica el porcentaje de materia orgánica que es biodegradable y que disminuye a medida que el relleno evoluciona.

Inicialmente esta relación es de 0.5 a 0.8, pasando a 0.07 o 0.08 después de varios años. Autores como Robinson (1979) describen bandas, una relación DBO/DQO asume valores superiores a 0.4 durante la fase ácida e inferiores e iguales a 0.4 durante la fase metanogénica.

En el estudio de la aplicación de herramientas matemáticas y de software al relleno sanitario en cuestión se tiene un conjunto de variables fisicoquímicas medidas sobre diferentes muestras pertenecientes a diferentes celdas, constituyendo éste un modelo multivariado. Es posible preguntarnos si se puede predecir estadísticamente la actividad biológica de un componente y de una celda en general. El análisis que se propone está basado en el estudio de la degradación/biodegradación que se presenta en las diferentes celdas que conforman el relleno sanitario, para lo cual se ha utilizado análisis de datos multivariados con las herramientas Matlab, infoStat y Norm. PCA es uno de los diversos métodos multivariados que permiten explorar patrones en los datos, observar cómo los componentes se comportan en forma similar, encontrar las celdas que pertenecen a un grupo y en general cómo se puede predecir el comportamiento a partir de información disponible.

El principal uso del PCA está relacionado con la explicación de la estructura de varianzas y covarianzas de una serie de variables originales, mediante unas pocas combinaciones lineales de ellas, buscando generar nuevas variables que puedan expresar la información contenida en el conjunto original de datos y reducir la dimensionalidad del problema estudiado. Las nuevas variables generadas se denominan *componentes principales*. El análisis es esencialmente descriptivo y tiene una interpretación geométrica (Lema, 1996; Smith, 2002). Lo que se persigue es extraer información de una matriz multivariada conformada por variables y muestras tomadas en el tiempo.

De igual manera, en la aplicación del relleno sanitario se tienen 12 variables fisicoquímicas y muestras tomadas en el tiempo a partir de lixiviados recolectados *in situ*. La idea entonces es reducir el número de variables a uno que dé explicación a la máxima varianza de los datos, pues se tienen muchas observaciones (muestras) y un alto número de dimensiones (variables). Esta herramienta permite razonar sobre la información con más profundidad y por ello es necesario reducir los datos a un conjunto de factores más pequeños, los cuales son combinaciones de variables observadas, pueden ser bases más efectivas para profundizar, y corresponden a los *componentes principales*. Aun si el significado físico de lo que se interpreta es oscuro, los datos observados son descritos en términos de estos factores en lugar de describirlos en términos de las variables-dimensiones originales.

## 2. METODOLOGÍA

El método PCA consiste en combinar variables relacionadas, y hacer foco en las variables no correlacionadas o independientes, especialmente aquellas a lo largo de las cuales las observaciones en los datos originales tienen alta varianza. Se busca un conjunto de variables más pequeño que explique la mayor parte de la varianza en los datos originales, en una forma más compacta y profunda. Las nuevas variables-dimensiones son combinaciones lineales de las variables originales, no están correlacionadas entre sí (ortogonales en el espacio de dimensión original) y capturan la mayor parte posible de la varianza original en los datos. Dichas variables son llamadas *componentes principales*. La motivación para aplicar PCA es encontrar una nueva variable o unas pocas variables nuevas que representen tanto como sea posible la varianza en el conjunto de datos.

La descomposición en valores propios está dada por:

$$XX^T = U\Sigma U \quad (1)$$

Donde,  $U = [u_1, u_2, \dots, u_M]$  y  $\Sigma = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_M]$

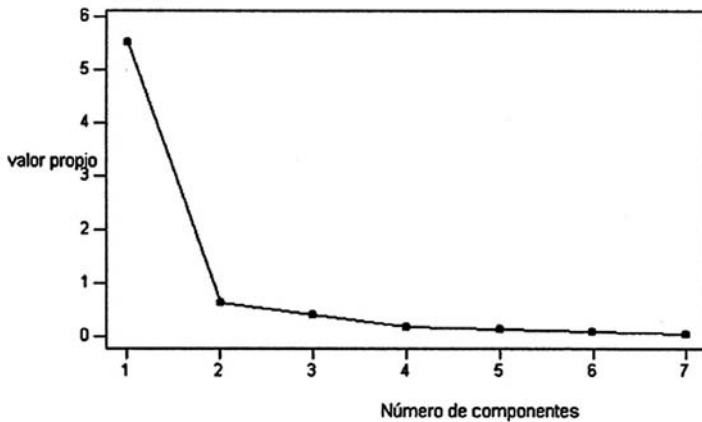
$\lambda$ : valor propio.

$x$ : vector propio asociado al valor propio.

De la misma manera se procede para los ejes siguientes, de tal forma que los nuevos ejes son los vectores propios de la matriz de correlaciones de las variables originales, los cuales capturan las similitudes de las variables originales basadas en cómo las muestras de los datos se proyectan en ellos.

La matriz  $U$  corresponde a los vectores propios y es llamada matriz *loadings*. Ésta es usada luego para calcular la matriz *scores* en el espacio de componentes principales. Esta matriz es la rotación de los datos originales, donde el nuevo conjunto de datos aparece proyectado en la dirección de los ejes de los componentes principales (Berrar, 2003).

La figura 2 muestra que los últimos valores propios están cercanos a una línea recta, así el número de PC's es el número de eigenvalores antes de esa línea (Lattin *et al.*, 2003).



**Figura 2.** Número de componentes principales, PC

Además los datos se deben escalar antes de aplicar PCA, debido a que diferentes valores promedio de las variables individuales en el conjunto de datos implican que las variables están pesadas en forma diferente, lo cual influenciará los resultados del PCA y conducirá a interpretaciones erróneas (Wise y Gallagher, 1996). Con el fin de aplicar PCA es necesario un pre-tratamiento de los datos para eliminar *outliers* y valores faltantes (*missing values*). El proceso inicia con la generación de la matriz de covarianza del conjunto de datos, una vez éstos han sido arreglados en forma apropiada. En este trabajo las columnas representan las variables fisicoquímicas y las filas las muestras en el tiempo. Se utilizan Matlab para escalar los datos (MPCA: Multi-way Principal Component Analysis) y Norm con el fin de completar los datos faltantes en cada una de las matrices de cada celda. El análisis de PCA se realizó haciendo uso del paquete de software InfoStat.

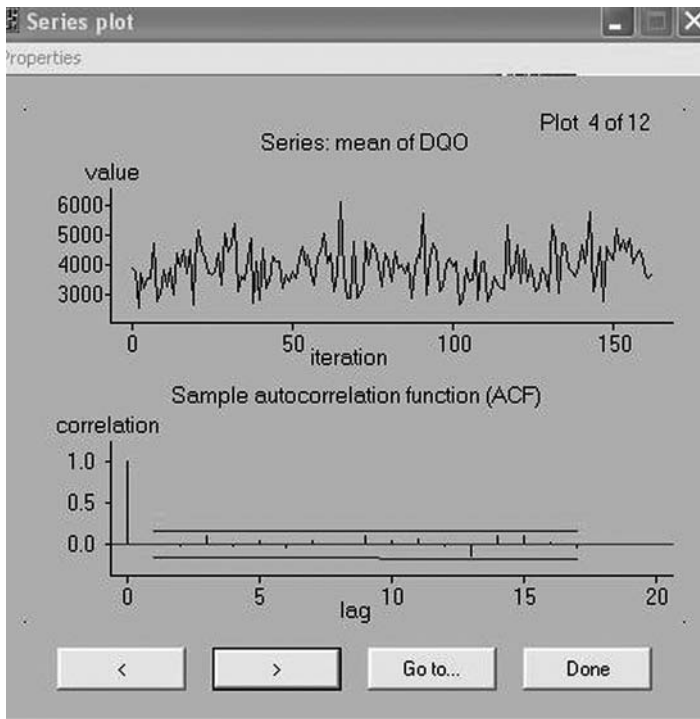
Los datos correspondientes a la información de las variables fisicoquímicas registradas para cada celda del relleno sanitario presentan la siguiente forma en la Tabla 1:

**Tabla 1.** Datos correspondientes a los valores de las variables fisicoquímicas del relleno sanitario de la celda 10

Fecha	pH	Alcal	STV	DQO	DBO	N.T.	N-NH3	Cd	Cu	Fe	Mn	Zn
05/02	6.9	4300	2308	7200	3107	367.9	292.3	0.06	0.09	107	18	0.61
06/02	6.4	2400	3392	8000	3453	292.3	206.6	0.01	0.06	115	22.6	1.97
07/02	7.3	6600	4892	11600	5012	810.9	766	0.01	0.03	183	18	0.74
08/02	7.1	4400	1728	2800	1202	509	425.6	0.02	0.04	66.7	12	0.61
09/02	6.8	1250	390	490	202	196.6	166	0.01	0.06	19	2.67	0.58
10/02	6.6	2750	3760	8700	3756	386.4	324.5	0.01	0.06	114.2	12.2	1.42
11/02	7.3	5000	2444	3200	1376	504.9	483.8	0.01	0.03	61	13	0.79
12/02	7.2	5350	1410	1800	769.8	645.1	564.5	0.01	0.06	52	6.31	0.54
01/03	7.3	5050	1176	1500	640	623.6	569.8	0.05	0.06	35	3.7	0.55

Los patrones de datos faltantes se pueden predecir a partir de la información contenida en otras variables y no de la variable que está incompleta. Dos formas de atacar el problema de la ausencia de datos son en primer lugar la de eliminar las unidades de información incompleta, que resulta muy fácil pero puede causar problemas de interpretación por la pérdida de mucha información (puede producir impacto en la precisión y potencia de las estimaciones), y en segundo lugar imputar los valores ausentes. En esta última se pueden presentar dos casos, el primero relacionado con la asignación de la media al valor ausente, y el segundo con la predicción del valor ausente mediante un modelo matemático de regresión múltiple (imputaciones múltiples), en el cual se reemplaza el valor ausente con varios valores imputados extraídos de una distribución predictiva.

El método MPCA es implementado usando Matlab de Mathworks R-2006b, mientras la técnica de PCA ha sido implementada usando InfoStat, un programa estadístico desarrollado en el ambiente Windows que ofrece una interfase avanzada, basada en el concepto de planilla electrónica, para el manejo de datos. La implementación introductoria a los datos faltantes (*missing data*) se programó usando un paquete de software llamado NORM, para Windows 95-98-NT, versión 2.03 (Schafer, 1998). La figura 3 muestra el valor promedio para el DQO y la función de auto correlación luego de que el algoritmo converge en la iteración 37.



**Figura 3.** Convergencia de la variable DQO al aplicar Norm.

Norm es un programa para imputación múltiple (MI) de datos multivariados incompletos. Su nombre se refiere a la distribución normal multivariada, modelo usado para generar las imputaciones. Utiliza el algoritmo EM (*Estimated Parameters*) para estimación eficiente de la media, las varianzas y las covarianzas (o correlaciones), y usa el procedimiento DA (*Data Augmentation*) para generar imputaciones múltiples de datos faltantes. La mayoría de los paquetes estadísticos no manipulan muy bien los datos faltantes (*missing data*), y en algunos casos son típicamente descartados, resultando en pérdida de información substancial y tal vez con bias en los resultados de manera impredecible. Usando información obtenida del conjunto de datos observados, NORM simula la parte faltante  $m > 1$  veces ( $m$  versiones igualmente plausibles de los datos completos). Cada uno de los  $m$  conjuntos de datos completos es analizado mediante técnicas de datos completos estándar. Los  $m$  conjuntos de resultados son luego combinados (Rubin, 1987), para obtener un conjunto de estimados y errores estándar (medias y desviaciones estándar) que incorporan la incertidumbre de los datos faltantes (Schafer, 1999).



### 3. RESULTADOS Y DISCUSIÓN

Se analizaron 12 variables para cada una de las celdas del relleno sanitario, inicialmente para un periodo de 5 años. Como la cantidad de datos faltantes en la mayoría de las celdas no excedía el 5%, se decidió aplicar la media a cada variable con el fin de llenar los espacios de datos faltantes (*missing data*). Diferentes tratamientos de los datos pueden ser hechos antes de aplicar diversos métodos estadísticos al conjunto de datos. Un pretratamiento que es necesario con el fin de aplicar PCA es el escalado del conjunto de datos. En el *cel-wise unfolding* (usando Matlab), las trayectorias promedio para todas las variables se remueven si se aplica el centrado promedio. Esta clase de escalado ayuda a eliminar variación en el modelo cuando se usa el MPCA.

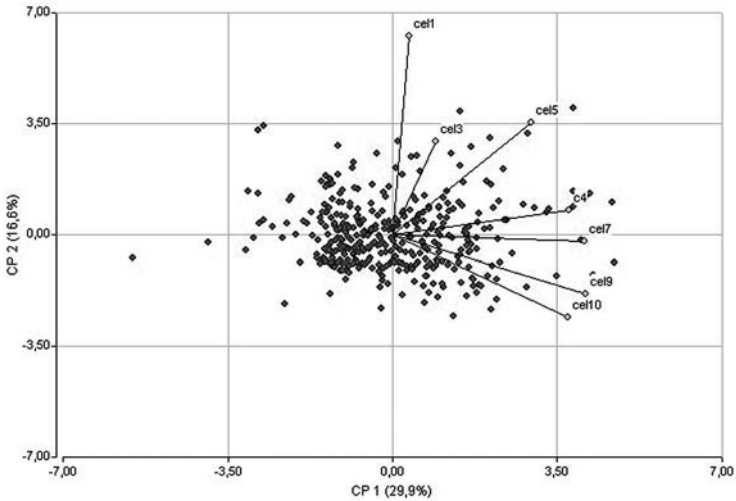
En consideración del objetivo que se perseguía, consistente en analizar las diferentes celdas para detectar celdas defectuosas, se decidió escoger el despliegue por celdas (*cel-wise unfolding*). Todas las matrices que contienen el conjunto de datos de variables fisicoquímicas y las muestras fueron arregladas de tal forma que contuvieran las mismas dimensiones (29 muestras fueron seleccionadas para cada una de las celdas del relleno), esto es, el mismo número de variables y muestras. Además, debido a limitaciones de los algoritmos aplicados, se realizó un pretratamiento de los datos con el fin de obtener matrices adaptadas para construir un arreglo tridimensional (se utilizó el programa Matlab tanto para el tratamiento como para el despliegue de las celdas).

En la etapa inicial se sesgaron las matrices para conseguir el mismo número de muestras para cada celda, utilizando el número de muestras presente en la matriz de datos de la celda más pequeña. De esta manera el tamaño o dimensión de las matrices de datos de cada una de las celdas para la aplicación del método MPCA fue de 29 muestras x 12 variables. Una vez se arreglaron en la estructura tridimensional, el conjunto de datos con las 7 celdas se representó por  $\mathbf{X}$  ( $7 \times 12 \times 29$ ).

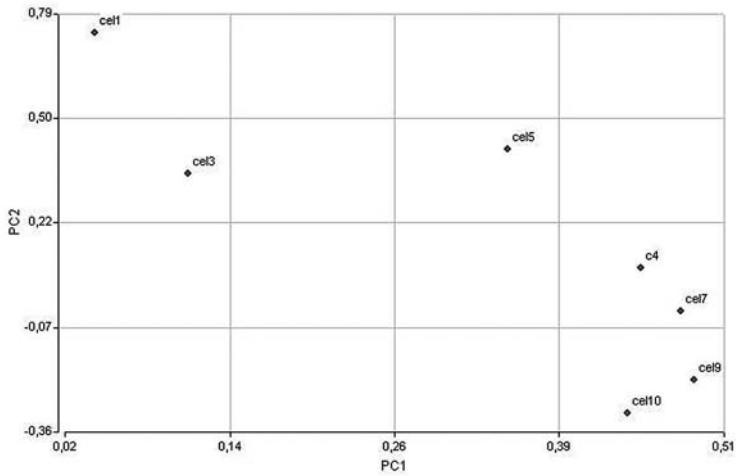
Luego a este conjunto tridimensional de datos se le aplicó el algoritmo *cel-unfolding*. Tras el despliegue de las celdas (*cel-wise unfolding*) se obtuvo el arreglo bidimensional  $\mathbf{X}$  ( $7 \times 377$ ), y de esa manera se pudo aplicar el método PCA tal como se aplica a cualquier matriz que contiene un conjunto de datos en dos dimensiones (muestras en las filas y variables en las columnas). El objetivo central con el cual se hizo este tipo de arreglo fue identificar o detectar las diferencias entre las distintas celdas.

La Figura 4 muestra los *scores* en el espacio de los componentes principales para las 7 celdas en estudio, después de haber realizado pretratamiento a los datos utilizando

auto escalado (*auto scaling*), a través del cual se subtrae la media para cada variable seguida de la división por la desviación estándar. El pretratamiento se realizó para cada una de las celdas, antes de hacer el arreglo de todos los datos para el análisis de MPCA. Posteriormente las columnas de datos resultantes de este arreglo se organizaron en otra matriz especial con el fin de aplicar el método PCA a este conjunto de datos. Nótese cómo la celda 1 tiene un comportamiento diferente a las demás, lo cual se aprecia también en la figura 5, donde se puede observar la presencia de dos grupos, uno con valores altos a lo largo del componente principal 2. La interpretación que se decidió aplicar a las celdas del relleno sanitario se fundamentó en la aplicación que realizada en el estudio de reactores tipo Batch (Haack, 2006) e implementada en una aplicación práctica en el mismo tipo de reactores (Leivas, 2007). En referencia a la posición con respecto al segundo componente principal, nótese también cómo la celda 1 aparece con el valor más alto a lo largo de dicho componente, mostrando de nuevo un comportamiento anormal posiblemente debido a la forma como se tomó la muestra en campo.



**Figura 4.** Scores en el espacio de los componentes principales para las 7 celdas. Pretratamiento usando auto scaling.



**Figura 5.** Celdas representadas en el espacio PC con pretratamiento usando *auto scaling*.

Los resultados para los *scores* de las celdas se pueden atribuir a la disposición inadecuada de los residuos sólidos.

En la celda C1 ocurrió disposición inadecuada de residuos sólidos (disposición de residuo ya degradado con residuo percolado con menor carga orgánica). Ésta no cuenta con sistema de impermeabilización de base, o permite una infiltración de aguas provenientes de nacientes existentes en el área, y una consecuente dilución del líquido percolado. La celda C3 fue la primera proyectada e implantada según los criterios del relleno sanitario, sin embargo tampoco muestra un buen comportamiento.

Cuando se usa el PCA al conjunto de datos normales se aplica el escalado a cero media y varianza unitaria, pues éste es más apropiado principalmente cuando las variables del proceso tienen diferentes unidades o diferentes rangos. La tabla 2 muestra los componentes principales para cada una de las celdas del relleno sanitario.

**Tabla 2.** Datos de componentes principales para las celdas del relleno correspondientes a 5 años.

Celdas	Componentes principales
C1	PC1 = 0.46 Alcal + 0.40 DQO + 0.36pH + 0.36STV + 0.34NT + 0.29Cd + 0.28 N-NH3 + 0.23 DBO + 0.13 Fe + 0.04 DBO/DQO PC2 = 0.50 DBO/DQO + 0.46 Cu+ 0.42 Mn +0.40 Fe+ 0.33 DBO + 0.16 STV + 0.15Zn
C3	PC1 = 0.45 DBO+ 0.42 Alcal + 0.40 STV +0.37 DBO/DQO + 0.29N-NH3 + 0.326 DQO + 0.25NT + 0.19 pH + 0.18Cu + 0.05Zn PC2 = 0.41 Zn + 0.32 Cu + 0.28 pH +0.13 STV + 0.9 DQO+ 0.05Alcal
C4	PC1=0.37 DBO + 0.35 DQO + 0.34 DBO/DQO + 0.33 Zn + 0.30 Cd+ 0.30 Cu + 0.26 STV + 0.02 Fe PC2 = 0.58 Alcal + 0.38 NT + 0.25 N-NH3 + 0.19Cu 0.18 DBO + 0.18 DQO + 0.15 pH + 0.12 STV + 0.15 Fe + 0.05 Zn+ 0.04 DBO/DQO
C5	PC1=0.43 DBO + 0.43 DQO + 0.36 Fe + 0.36 Mn + 0.31 Zn + 0.24 STV + 0.21 Cu + 0.19 DBO/DQO + 0.18Alcal + 0.12 NT + 0.06Cd PC2 = +0.52 N-NH3 + 0.51 NT + 0.39 Cu + 0.33 Alcal + 0.25 pH + 0.10Fe
C7	PC1= 0.37 DQO + 0.37 Mn + 0.36 STV + 0.36 Fe + 0.32 Zn + 0.30 DBO + 0.30 Cu + 0.04 Cd PC2 = 0.55 Alcal + 0.55 NT+ 0.44 N-NH3 + 0.25 pH +0.20 STV +0.17 DBO + 0.16 DBO/DQO + 0.12 Zn + 0.10 DQO + 0.07Mn
C9	PC1=0.35 DQO + 0.35 DBO + 0.35 Mn + 0.30 STV + 0.30 Fe + 0.25 Zn + 0.10 Cd+ 0.04 DBO/DQO PC2 = 0.47 Cu + 0.45 Alcal + 0.37 NT+ 0.35 N-NH3 + 0.31 Cd + 0.26 Zn + 0.23 DQO + 0.19 DBO + 0.18 STV + 0.12Mn

La importancia de las variables se puede inferir por el valor de sus coeficientes y la distancia desde el origen. Se observa en general, sin embargo, alto predominio de las variables DBO y DQO a lo largo del primer componente principal PC1. Por otra parte, se observa un buen predominio de las variables NT y N-NH3, al igual que de los metales a lo largo del componente PC2, con valores significativamente altos en la mayoría de las celdas.

El primer componente PC1 explica lo relacionado con la degradación anaeróbica de los microorganismos sobre la materia orgánica contenida en las celdas, por esta razón en todas ellas, excepto en la 1 y la 3 las variables DBO, DQO y STV aparecen con valores de carga altos. El segundo componente principal PC2 permite explicar lo relacionado con la actividad biológica (relacionada con el control de acidez del medio), apareciendo con mayor carga en este componente las variables NT, N-NH3 y alcalinidad en todas las celdas, excepto en la 1 y la 10. Así, se concluye que CP2 está más relacionado con la actividad biológica.

En el gráfico de *scores* (figuras 6a, 6b, 6c ) los tiempos menores para CP2 (variables más activas) están localizados en la parte correspondiente a la actividad biológica, mientras lo contrario sucede con los *scores* más jóvenes, ubicados en la zona de las variables responsables de la degradación.

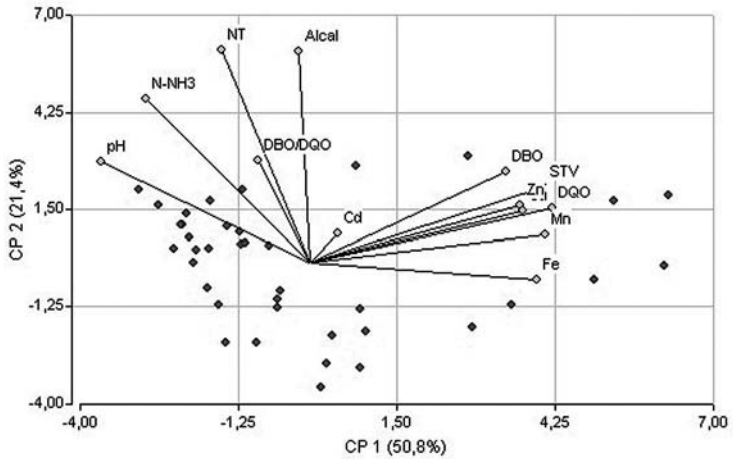


Figura 6a. *Scores* y *loadings* para la celda 7.

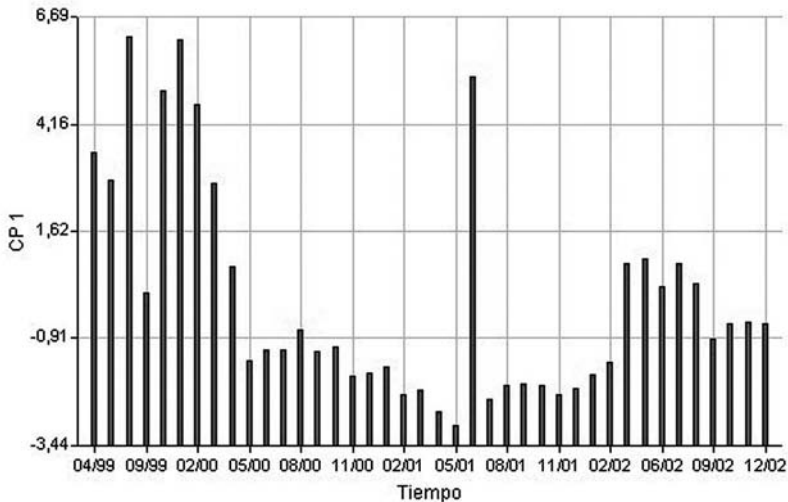
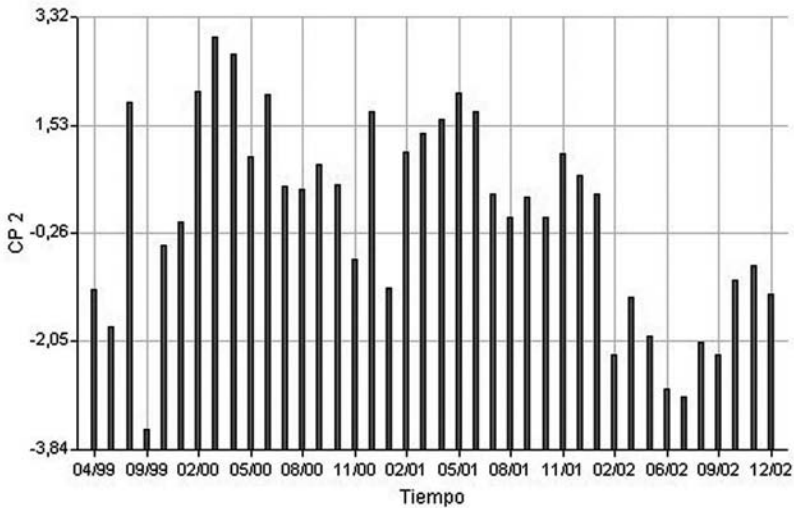


Figura 6b. *Scores* PC1 vs tiempo para la celda 7.



**Figura 6c.** Scores PC2 vs tiempo para la celda 7.

El gráfico de *scores* contiene información sobre los meses (edad de la celda) llegando a ser visibles la semejanza, las agrupaciones y las diferencias entre los mismos, con base en las variables usadas.

En la figura 6b, a medida que el tiempo pasa las variables que hacen parte del componente principal 1 (PC1), en el caso de la celda C7, disminuyen su valor, porque a medida que pasa el tiempo o cuanto mayor es el valor de la variable en meses (edad de la celda), los valores de PC1 tienden a disminuir.

A medida que pasa el tiempo de la celda los valores de PC2 tienden a aumentar, pues en el periodo de degradación de materia orgánica los microorganismos involucrados en el proceso liberan sustancias y ocurre el incremento de actividad biológica. Lo anterior está relacionado con que a medida que pasa el tiempo la actividad de degradación disminuye, y la actividad biológica se incrementa.

#### 4. CONCLUSIONES

Mediante el uso de PCA se describe la estructura e interrelación de variables que originan el fenómeno de biodegradación de la materia orgánica, obtenidas agrupando variables que tienen información semejante (altamente correlacionadas). Con las observaciones presentadas en este artículo se pretende hacer énfasis en las posi-

bilidades que ofrece el uso de estrategias adicionales al PCA y el complemento con otros paquetes de software que sirvan de soporte para el análisis del estado actual del relleno sanitario.

En casi todos los casos se utilizaron los dos primeros componentes principales estandarizados. Éstos explicaban aproximadamente el 50% de la estructura de varianzas y covarianzas de los datos, lo que equivale a pasar de 12 variables a 2 componentes principales.

La importancia de cada variable es proporcional a su longitud vectorial y, así mismo, el ángulo entre dos variables es inversamente proporcional a la correlación entre ellas. Por ejemplo el ángulo entre las variables DBO y DQO es pequeño, mostrando una alta correlación entre ellas, mientras que la variable Alcalinidad no está nada correlacionada con la variable Fe. De igual manera se observa una correlación negativa entre pH y Fe.

Los resultados del análisis MPCA-PCA describen la estructura y las interrelaciones de variables originales en el fenómeno de biodegradación de materia orgánica a partir del análisis e interpretación de los componentes principales, obtenidos agrupando variables que ofrecen informaciones semejantes que están altamente correlacionadas.

Los parámetros importantes para degradación de materia orgánica son pH, ST, Alcalinidad, DQO, DBO, Nitrógeno total y Nitrógeno amoniacal, y los parámetros que miden la biodegradabilidad son DQO y DBO.

Un indicativo de actividad biológica llevada a cabo en la celda del relleno sanitario se observa cuando la concentración de nitrógeno amoniacal es superior a la concentración de nitrógeno orgánico.

Hasta los primeros 5 años se tienen muchos microorganismos para degradación, teniéndose valores mayores para la relación DBO/DQO.

La diferencia entre el conjunto de datos de la tabla de observaciones en la cual la información faltante se completó utilizando el concepto de media, y aquel de tabla en la que el conjunto de datos se completó usando el concepto de *missing data* es mínima. Lo anterior se debe a que la cantidad de datos faltantes en la tabla no superaba el 5%, de esta forma se concluyó que reemplazar los datos faltantes simplemente por la media de la columna correspondiente a cada variable era una opción suficiente.

## BIBLIOGRAFÍA

- Berrar, D.P. *et al.* (2003). *Singular value decomposition and principal component analysis*. Kluwer: Norwell, MA,..LANL LA-UR-4001. 91-109.
- Haack, M. (2006). *Monitoring of fermentation processes using image analysis and on-line multi-wavelength fluorescence*. Doctoral Thesis, Biosentrum, Technical University of Denmark, Denmark.
- Lattin, *et al* (2003). *Analyzing multivariate data*. Duxbury Applied Series, 25, 95-108.
- Leivas, B. (2007). *Fermentation process monitoring using multi-way principal component analysis*. Master Thesis, Department of Chemical Engineering, Technical University of Denmark, Denmark.
- Lema, T. (1996). *Borrador para algunos elementos de estadística multivariada. Postgrado en bosques y Conservación Ambiental*. Facultad de Ciencias Agropecuarias, Departamento de Ciencias Forestales: Universidad Nacional de Colombia.
- Rubin, D. (1987). *Multiple imputation for no response in surveys*. New York: Wiley.
- Schafer, J. & Olsen M. (1998). *Multiple imputation for multivariate missing-data*. Doctoral thesis, Center for Social Organization of Schools Johns Hopkins University, The Pennsylvania State University.
- Schafer, J. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model*. Disponible en línea. Consultado el día 4 de julio de 2007 en: <http://www.stat.psu.edu/~jls/misoftwa.html>.
- Smith, L. (2002). *A tutorial on principal components analysis*. Disponible en línea. Consultado el día 10 de agosto de 2007 en: [http://www.csnet.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).
- Wise, B. & Gallagher, N. (1996). *The process chemometrics approach to process monitoring and fault detection*. Journal of Process Control 6, 329-348.
- Reichert, G. (2000). *Aterro Sanitário, Projeto, Implantação e Operação*. ABES Associação Brasileira de Engenharia Sanitária e Ambiental. Proceedings del XXVII congreso interamericano de ingeniería sanitaria y ambiental, diciembre 3-8, (pp. 69-74). Porto Alegre – RS, Brasil.
- Ehrig, H. (1983). *Waste Management & Research*. (Disponible: Departamento de ingeniería, Institut für Satadtbauwesen, TU Braunschweig, West Germany). 

Referencia	Recepción	Aprobación
Muñoz H., J. A.; Andrade, M. Z.; Grisaa, A. M. Aplicación de herramientas de software al análisis del relleno sanitario de São Giácomo de Caxias do Sul. Revista <i>Tumbaga</i> (2009).	Día/mes/año 14/06/2009	Día/mes/año 19/09/2009