

## CAPÍTULO 3

ISBN 978-987-543-443-1

[http://cursosihlla.bdh.org.ar/Libro\\_011/](http://cursosihlla.bdh.org.ar/Libro_011/)

### **Análisis multivariado de la relación entre variables de importancia agroecológica y el NDVI**

**Marcelo Gandini<sup>1</sup> y Eduardo Usunoff<sup>2</sup>**

<sup>1</sup> Centro de Investigaciones Integradas Sobre Agricultura Sustentable CIISAS. Facultad de Agronomía de Azul / Av. República de Italia 780, B7300, Azul, Buenos aires, Argentina.

<sup>2</sup> Instituto de Hidrología de Llanuras “Dr. Eduardo Jorge Usunoff” (IHLLA) sede Azul / Av. República de Italia 780, B7300 Azul, Buenos Aires, Argentina.

Correo electrónico: [mgandini@faa.unicen.edu.ar](mailto:mgandini@faa.unicen.edu.ar)

---

#### **1. Introducción**

El fin del presente trabajo es desarrollar una metodología objetiva para la identificación, a nivel regional, de las relaciones entre variables básicas o independientes y otras dependientes generadas por sensores remotos con el sensor SPOT Vegetation, generalizando las observaciones realizadas por [Gandini et al. \(1996, 1997\)](#).

Existe una alta correlación entre algunas variables individuales de superficie y el *NDVI*, usando datos del satélite National Oceanic Atmospheric Administration-Advanced High Resolution Radiometer (NOAA-AVHRR) ([Gandini, 2006](#)). Un conjunto de datos independientes de *NDVI* debería repetir las tendencias observadas. En este caso, variables ambientales de la cuenca del arroyo del Azul, son consideradas independientes y el *NDVI* y el Uso/cobertura de la tierra como dependientes. Mediante la utilización de datos de *NDVI* provenientes del satélite SPOT Vegetation, tomados para la misma zona, con la misma resolución pero en años diferentes, debería repetirse una asociación entre el *NDVI* y las zonas definidas, ya que es un hecho que existe estrecha relación entre la estructura y el funcionamiento de los ecosistemas a nivel regional ([Noss, 1996](#)). Una descripción de las correlaciones entre los mismos es un aporte a la definición de regiones naturales. El *NDVI* se mostró que responde diferencialmente según la región considerada ([Gandini, 2006](#)). Por otro lado, se conoce que otras variables son también importantes en la definición del proceso de precipitación-escorrentía. Tal es el caso del Uso/Cobertura de la Tierra,

concepto definido principalmente para la clasificación de la superficie terrestre mediante sensores remotos (Anderson *et al.*, 1976).

Para lograr el objetivo de relacionar variables, se propone la utilización de un Sistema de Información Geográfica (SIG) ensamblado a una base de datos y a un paquete estadístico capaz de realizar análisis multivariado, específicamente análisis de correlaciones canónicas.

La oferta actual de programas para análisis de imágenes de sensores remotos y SIG no incluye la totalidad de las opciones que existen de análisis multivariado, restringiéndose únicamente al análisis de componentes principales (e.g. Eastman, 2006), análisis discriminante (e.g. Landgrebe, 2001) y algunos tipos de clúster. Es así como, aún siendo factibles análisis como el de factores con distintos tipos de rotaciones, el análisis discriminante o las correlaciones canónicas, puede interpretarse que la ausencia de trabajos que exploren las capacidades de estos análisis usando bases de datos SIG o imágenes de satélite, esté relacionada principalmente al uso empaquetado de los programas y la carencia de estas herramientas en los paquetes. El avance en este sentido, estaría favorecido por un trabajo más interdisciplinario.

Las correlaciones canónicas constituyen una generalización de las correlaciones simples y múltiples. Las correlaciones simples estiman la relación existente entre dos variables, la variable independiente  $X$  y la dependiente  $Y$ . Las correlaciones múltiples estiman la relación entre un conjunto de variables independientes y una sola variable dependiente  $Y$ . Las correlaciones canónicas estiman la correlación existente entre un conjunto de variables independientes y otro conjunto de variables dependientes. Las correlaciones canónicas indican entonces, en base a modelos lineales, cómo es la estructura de correlación entre los grupos de variables que explican la mayor parte de la variabilidad. Desde el punto de vista metodológico el uso de las correlaciones canónicas exige varias reflexiones: la primera acerca del número de variables que componen el *grupo X* y el *grupo Y*. Si son muchas, posiblemente en cada grupo puede suceder que haya altas correlaciones entre las variables, lo cual es igual a decir que se están incluyendo variables redundantes. Si son muy pocas, se corre el riesgo de no incluir aquellas variables que realmente tienen mayor fuerza explicativa. La segunda reflexión es que el comportamiento ideal de las variables se encuentra cuando existe una muy baja correlación dentro de cada grupo, tanto en el de las  $X$  como en el de las  $Y$ , y máxima entre los dos grupos. Esto implica una elección afinada de aquellas variables, por una parte, más relevantes y significativas tanto en el grupo de las  $X$  como en el grupo de las  $Y$ , a la vez que independientes entre sí dentro de cada grupo, es decir, aquellas que midan cuestiones distintas y aparentemente inconexas aunque naturalmente referidas al tema que se esté

investigando. Por este motivo es necesario un acercamiento exploratorio durante el proceso de elección de variables para su ingreso a la matriz final de datos.

La metodología propuesta fue aplicada satisfactoriamente en un caso anterior para la misma zona con imágenes del Satélite Argentino de Observación de la Tierra (SAC-C) (Gandini y Usunoff, 2004). La relación encontrada entre el NDVI y las zonas agroecológicas homogéneas, espera ser confirmada mediante la utilización de otra fuente de datos independientes como el NDVI calculado con la misma resolución para la misma zona pero para otro año, a partir de imágenes del sensor Vegetation a bordo del satélite SPOT.

## **2. Materiales y Métodos**

Se contó con una base de datos utilizada para la regionalización de la cuenca del arroyo del Azul, consistente en el conjunto de variables ambientales que se observan en la Tabla 1, una estimación de Uso/Cobertura de la Tierra realizada con imágenes SAC-C, y datos de NDVI provenientes del sensor Vegetation a bordo del satélite SPOT 4 para el periodo junio 2002-febrero 2004 (59 escenas correspondientes a composiciones de 10 días realizadas por el centro de distribución de las imágenes SPOT (VITO) de Bélgica, <http://free.vgt.vito.be/>).

### *2.1 Estimación de Uso/Cobertura de la superficie*

Se dispuso de ocho ventanas de imágenes del sensor Multispectral Medium Resolution Scanner (MMRS) SAC-C de la cuenca del arroyo del Azul (Buenos Aires, Argentina) para el periodo entre abril de 2001 a mayo de 2002, de los path 224, 225, y 226 del *World-wide Reference System* Landsat. El tratamiento previo consistió en la calibración mediante los coeficientes provistos por la CONAE y la corrección atmosférica aplicando el modelo denominado de corrección del histograma por valores mínimos (Chávez, 1988).

Por otro lado, y debido a que se trabajó en la zona de superposición de imágenes de diferentes path, lo que implica un error geométrico considerable, se realizó una corrección geométrica, usando el método del vecino más próximo para los valores de píxeles, y registración previa al uso de las mismas, con un error de un píxel en todos los casos.

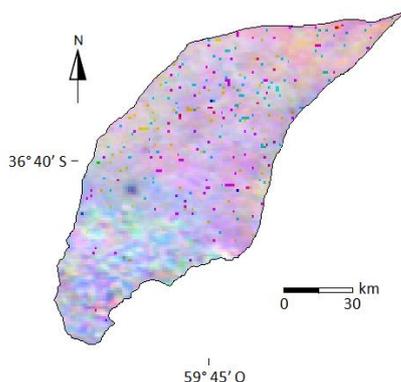
El uso de la tierra fue estimado realizando primeramente una clasificación no supervisada, y superponiendo a ésta polígonos con datos de campo, para luego realizar una clasificación supervisada, usando clases de uso similares a las definidas en las tablas de curva número del Soil Conservation Service (United States Department of Agriculture): *urbano, cuerpos de agua libres, arboledas y montes, pastizales (naturales y pasturas), cultivos de invierno y de verano.*

### *2.2 Tratamiento de imágenes Vegetation*

El índice NDVI para los distintos meses incluidos en este análisis se obtuvo de imágenes Vegetation disponibles en el sitio VITO (<http://free.vgt.vito.be/>). Las

mismas fueron georreferenciadas y recortadas para obtener el *NDVI* correspondiente al rectángulo que incluye la cuenca del arroyo del Azul (Figura 1).

Las imágenes crudas son corregidas atmosféricamente de todos los efectos relevantes para las longitudes de onda y escala en cuestión. Se corrige por vapor de agua, ozono y aerosoles entre otros, y los píxeles con nubes son marcados y enmascarados. La corrección geométrica se hace en forma sistemática con puntos de control en Latitud-Longitud. Las imágenes distribuidas finalmente son compuestas de diez días (VGT-S10; tres para cada mes): días 1-10, 11-20, y 21-30. Los datos VGT-S10 son generados seleccionando los píxeles que tienen los valores de máximo *NDVI* dentro de estos periodos. Este acercamiento ayuda a reducir al mínimo el efecto de la cubierta de nubes y la variabilidad en profundidad óptica de la atmósfera. Para más detalles sobre las imágenes utilizadas, puede consultarse a [Boles et al. \(2004\)](#).



**Figura 1.** Imagen compuesta Red, Green, Blue (RGB) de *NDVI* de la primera semana de los meses de junio (Rojo), julio (Verde) y agosto (Azul) de 2002 sobre cuenca del arroyo del Azul.

Las imágenes Vegetation en sus bandas 1, 2, 3, y 4 son equivalentes a las bandas 1, 3, 4, y 5 del satélite Landsat Thematic Mapper (TM), respectivamente, según [Boles et al. \(2004\)](#) y por lo tanto también equivalentes a las bandas correspondientes del SAC-C.

### 2.3 Tratamiento estadístico de los datos

Los datos fueron tratados con ENVI ([ITT, 2006](#)) y analizados por correlaciones canónicas previa estandarización de las variables con el programa R (2005). El uso de la tierra fue tratado como una variable categórica (no continua) en el análisis. La hipótesis estadística de las correlaciones canónicas

fue que las combinaciones lineales de las variables dependientes estaban correlacionadas con las combinaciones lineales de las variables independientes.

Los píxeles contaminados por nubes fueron descartados en el análisis de correlaciones canónicas, y luego reemplazados por el promedio e ingresados nuevamente para el cálculo de los escores factoriales, con el objetivo de contar con imágenes producto del análisis.

Se realizó primero un análisis exploratorio, con el fin de identificar aquellas variables que cumplieran con los siguientes requisitos:

- *Baja correlación en su grupo.*
- *Alta carga (loading) en los análisis preliminares de correlaciones canónicas.*

En base a los criterios definidos, se utilizaron las siguientes variables:

**Independientes:** Fisiografía, Natracualfes mólicos, Argiudoles, Uso de la Tierra y Orientación.

**Dependientes:** De las 59 imágenes disponibles inicialmente para el periodo utilizado, se decidió ingresar al análisis final una imagen de cada una de las tres disponibles para los meses de junio a diciembre de 2002, debido a las altas correlaciones y la redundancia observada. Este hecho se supone que se debe a que, al tratarse de imágenes consecutivas en el tiempo, existe una mayor probabilidad de correlación. Si bien esta correlación es interesante y será usada para explicar los resultados, las correlaciones entre fechas demasiado cercanas, pueden oscurecer las relaciones buscadas. Por construcción, estas imágenes consecutivas en el tiempo podrían contener píxeles con valores de *NDVI* muy similares. Las variables finalmente utilizadas en el análisis fueron las siguientes:

**Independientes (grupo Y):** Natracuoles, Argiudoles, Natracualfes típicos, Natracuoles mólicos, Complejos indiferenciados, Hapludoles líticos, Argialboles, Pendientes y Lluvias.

**Dependientes (grupo X):** *NDVI* (primera decena de días de cada mes) de los meses de junio 2002 a enero de 2003.

En el análisis final se decidió utilizar una rotación Varimax para los tres primeros ejes, con el objeto de potenciar la representatividad de las variables de mayor carga en los ejes canónicos.

Los escores factoriales obtenidos en el análisis final, fueron tratados como nuevas variables, y compuestos en el SIG mediante una rutina de composición de imágenes para ser analizados en conjunto. Los datos fuera de la cuenca en estudio fueron tratados como de peso cero.

### **3. Resultados y Discusión**

En la Figura 1 se muestra un ejemplo donde puede verse el recorte realizado para la cuenca del arroyo del Azul, con dos zonas de colores oscuros donde se

ubican las ciudades de Azul (36° 46' S, 59° 52' O) y Cacharí (36° 23' S, 59° 30' O).

La correlación total fue muy significativa ( $r^2 = 0,748$ ;  $p < 10^{-4}$ ), aunque menor que la obtenida utilizando todas las variables disponibles ( $r^2 = 0,981$ ;  $p < 10^{-5}$ ). Esta evidencia es suficiente para no rechazar la hipótesis de trabajo expuesta anteriormente.

La predictibilidad de cada una de las variables dependientes en base al conjunto de las independientes fue altamente significativa para todas ellas ( $p < 10^{-4}$ ). Estos resultados apoyan la idea que el conjunto de variables utilizado para la regionalización estaría definiendo el comportamiento observado del NDVI (Gandini, 2006).

Existe correlación significativa entre las variables canónicas, resultantes de las combinaciones lineales que proporcionan la máxima explicación de la variabilidad presente (Tabla 1). Los ejes canónicos por su parte están construidos con la totalidad de las variables dependientes e independientes usadas en el análisis. Esta afirmación es válida para los ejes hasta el 5 inclusive, pues las correlaciones canónicas fueron significativas hasta el factor 5 ( $p < 0,05$ ; test de Bartlett para correlaciones residuales; se muestran hasta el tercer eje en la Tabla 1).

En la Tabla 2 (a y b) se presentan las correlaciones calculadas entre las variables usadas en el análisis dentro de cada grupo, y en la Tabla 3 entre los grupos X e Y. En general se observa una baja correlación entre las variables independientes (Tabla 2-a), mientras que el NDVI presenta altos valores de correlación entre algunos meses consecutivos (Tabla 2-b). Esto último es lógico que ocurra considerando la escala de tiempo utilizada, con respecto al ciclo de los cultivos y vegetación natural. Excluyendo el efecto de los meses consecutivos, las correlaciones entre variables dependientes e independientes son más fuertes (Tabla 3). Por ejemplo, observar las correlaciones entre NDVI de junio y Natracuoles y Argiudoles, y NDVI de octubre con Argiudoles y Natracualfes típicos. Las redundancias, si bien son mayores para el grupo de variables dependientes, son del mismo orden (Tabla 1).

Es interesante destacar las altas correlaciones (todas positivas) entre los meses de julio, agosto y septiembre. En estos meses debería haber una alta influencia del Uso/Cobertura de la Tierra en toda la cuenca. La explicación (basada en el uso) de este comportamiento es que en aquellas zonas donde el terreno ha sido arado, el mismo no presentaría respuesta de la vegetación hasta septiembre, y aquellas zonas donde existía alguna cobertura vegetal en julio, la conservan al menos hasta septiembre. Otro conjunto interesante de altas correlaciones es el que ocurre entre octubre-noviembre (positiva) y el grupo anterior (negativa). Para este caso, la explicación sería que los cultivos de invierno (trigo principalmente)

alcanzan su máxima cobertura e índice de área foliar (Cañibano *et al.*, 2004) en noviembre, contrastando con el suelo desnudo del invierno.

**Tabla 1.** Carga de las variables sobre sus correspondientes ejes canónicos (VC: Eje canónico). No se consignan las cargas cuyo módulo sea menor a 0,25.

<i>Variable</i>	<i>VC1</i>	<i>VC2</i>	<i>VC3</i>
<i>Natracuoles</i>	0,859	--	--
<i>Argiudoles</i>	-0,734	--	0,429
<i>Natracualfes típicos</i>	--	--	-0,811
<i>Natracualfes mólicos</i>	--	0,415	-0,431
<i>Complejos</i>	--	0,584	--
<i>Hapludoles líticos</i>	-0,535	--	--
<i>Argialboles</i>	--	-0,429	-0,264
<i>Pendientes</i>	-0,555	--	--
<i>Lluvias</i>	--	-0,82	0,381
<i>Varianza explicada X</i>	22,674	18,453	13,883
<i>Redundancia X</i>	0,11	0,035	0,023
<i>NDVI 6/02</i>	-0,802	--	-0,402
<i>NDVI 7/02</i>	--	0,641	--
<i>NDVI 8/02</i>	--	0,536	--
<i>NDVI 9/02</i>	--	0,7	--
<i>NDVI 10/02</i>	-0,954	--	--
<i>NDVI 11/02</i>	-0,295	-0,571	--
<i>NDVI 12/02</i>	0,283	-0,603	0,613
<i>NDVI 01/03</i>	-0,365	-0,812	0,253
<i>Varianza explicada Y</i>	41,593	27,378	15,067
<i>Redundancia Y</i>	0,139	0,032	0,038
<i>Correlación Canónica</i>	0,737	0,469	0,405

Un análisis similar se puede realizar para los cultivos de verano: las parcelas de cultivos de verano aumentan su cobertura vegetal durante los meses de noviembre, diciembre y enero, habiendo tenido muy baja cobertura en los meses de julio, agosto y septiembre, momento de preparación para la siembra. Por esta razón se observan correlaciones negativas entre estos grupos de meses y positivas dentro de los mismos (Tabla 2).

Estos resultados llevaron a efectuar el análisis usando una rotación Varimax hasta los primeros 3 factores.

La diferencia entre un análisis de factores con rotación Varimax y un análisis de componentes principales, radica en que para el primero la rotación puede realizarse para un subconjunto de ejes (en este caso se eligen los primeros 3), y en el segundo se realiza para todos. Por otro lado, el término de error se construye de manera diferente (Harman, 1976).

La Figura 2 muestra el resultado de la composición en una imagen RGB de los tres primeros escores factoriales del análisis. En la Figura 2-a se observa la composición de los ejes canónicos definidos solamente por las variables independientes. Las zonas homogéneas observadas (entre 6 y 8) coinciden en general con las obtenidas por Gandini (2006). Un análisis similar se puede realizar para los cultivos de verano: las parcelas de cultivos de verano aumentan su cobertura vegetal durante los meses de noviembre, diciembre y enero, habiendo tenido muy baja cobertura en los meses de julio, agosto y septiembre, momento de preparación para la siembra. Por esta razón se observan correlaciones negativas entre estos grupos de meses y positivas dentro de los mismos (Tabla 2).

**Tabla 2-a.** Correlaciones entre las variables de un mismo grupo. Variables independientes. Se destacan los mayores valores para una mejor visualización.

	1	2	3	4	5	6	7	8	9
1. Natracuoles	1								
2. Argiudoles	<b>-0,604</b>	1							
3. Natrac_ típicos	-0,228	<b>-0,347</b>	1						
4. Natrac_mólicos	-0,227	-0,031	-0,157	1					
5. Complejos	-0,145	-0,076	-0,175	-0,101	1				
6. Haplud_líticos	-0,235	0,162	-0,155	-0,078	-0,037	1			
7. Argialboles	-0,116	-0,083	-0,083	-0,054	0,106	-0,040	1		
8. Pendientes	-0,050	0,244	-0,152	-0,059	-0,053	<b>0,478</b>	-0,050	1	
9. Lluvias	0,193	0,198	-0,171	-0,012	<b>0,325</b>	-0,127	0,193	-0,226	1

**Tabla 2-b.** Correlaciones entre las variables de un mismo grupo. Variables dependientes. Se destacan los mayores valores para una mejor visualización.

	1	2	3	4	5	6	7	8
1. NDVI 6-02	1							
2. NDVI 7-02	0,205	1						
3. NDVI 8-02	0,160	<b>0,845</b>	1					
4. NDVI 9-02	0,108	<b>0,715</b>	<b>0,785</b>	1				
5. NDVI 10-02	<b>0,728</b>	-0,009	0,017	0,150	1			
6. NDVI 11-02	0,281	<b>-0,398</b>	<b>-0,441</b>	<b>-0,347</b>	<b>0,490</b>	1		
7. NDVI 12-02	<b>-0,316</b>	-0,255	-0,257	<b>-0,37</b>	-0,186	<b>0,488</b>	1	
8. NDVI 1-03	0,261	<b>-0,305</b>	-0,213	<b>-0,355</b>	<b>0,337</b>	<b>0,483</b>	<b>0,563</b>	1

Estos resultados llevaron a efectuar el análisis usando una rotación Varimax hasta los primeros 3 factores.

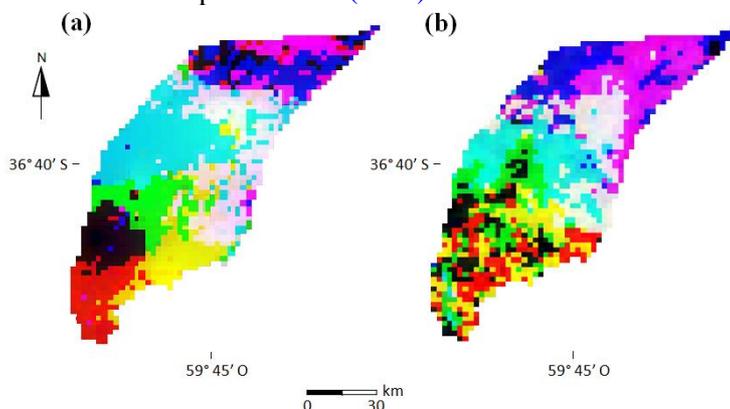
La diferencia entre un análisis de factores con rotación Varimax y un análisis de componentes principales, radica en que para el primero la rotación puede realizarse para un subconjunto de ejes (en este caso se eligen los primeros 3), y

en el segundo se realiza para todos. Por otro lado, el término de error se construye de manera diferente (Harman, 1976).

**Tabla 3.** Correlaciones entre las variables del grupo X (columna izquierda) e Y (fila superior). Se destacan los mayores valores para una mejor visualización.

	NDVI 6-02	NDVI 7-02	NDVI 8-02	NDVI 9-02	NDVI 10-02	NDVI 11-02	NDVI 12-02	NDVI 1-03
1.Natracuoles	<b>0,432</b>	0,055	0,040	0,083	0,410	0,033	<b>-0,347</b>	0,044
2.Argiudoles	<b>-0,491</b>	0,065	0,037	0,014	<b>-0,573</b>	-0,195	0,180	-0,254
3.Natrac_ típicos	0,216	-0,091	-0,069	-0,050	<b>0,375</b>	0,251	0,159	<b>0,325</b>
4.Natrac_mólicos	-0,160	0,167	0,184	0,146	-0,164	-0,192	-0,073	<b>-0,306</b>
5.Complejos	0,056	-0,064	-0,036	-0,126	-0,063	0,081	0,096	0,184
6.Haplud_líticos	<b>-0,325</b>	-0,080	-0,099	-0,012	-0,256	-0,046	0,120	-0,091
7.Argialboles	0,100	-0,224	-0,175	-0,226	0,100	0,171	0,100	0,186
8.Pendientes	-0,360	-0,047	-0,111	-0,091	-0,329	-0,149	0,118	-0,129
9.Lluvias	0,026	-0,129	-0,057	-0,209	-0,149	0,074	0,025	0,068

La Figura 2 muestra el resultado de la composición en una imagen RGB de los tres primeros escores factoriales del análisis. En la Figura 2-a se observa la composición de los ejes canónicos definidos solamente por las variables independientes. Las zonas homogéneas observadas (entre 6 y 8) coinciden en general con las obtenidas por Gandini (2006).



**Figura 2.** Imágenes compuestas RGB obtenidas con los escores factoriales; (a) Independientes y (b) Dependientes.

En la Figura 2-b se observa la composición de los tres primeros ejes de escores factoriales de las variables dependientes. Una visión general del mismo, y comparando con la Figura 2-a, apoya el valor de correlación obtenido, principalmente en la zona norte de la cuenca. En la zona sur, que como ya se dijo es principalmente agrícola, se observa un “pixelado” con un patrón típico de las

variaciones espaciales producto del uso agrícola de la misma, revelando el mismo, la naturaleza dinámica de las variables a esta escala. Esta diferencia visual lleva a proponer que, de realizarse un análisis por zonas, la mayor correlación debería observarse en las zonas con uso no agrícola, por una cuestión de escala de tiempo y espacio. De esta manera, se puede evaluar visualmente que existe una correlación significativa entre las variables independientes utilizadas, y la variabilidad espacial del *NDVI* de las zonas de pastizales naturales, bajo uso ganadero. Esta asociación permite suponer que la condición de la vegetación natural está definida por las variables utilizadas en la regionalización, que indirectamente se relacionan con factores de estrés de la vegetación, como la presencia de sales o sodio en el perfil del suelo, o la anegabilidad.

Las variaciones en el *NDVI* son muy marcadas dentro del ciclo de los cultivos. En el caso de la zona sur de la cuenca del arroyo del Azul, se practican cultivos tanto de invierno como de verano, y se siembran pasturas y verdeos en diferentes épocas del año. Esta característica de uso de la tierra provoca que existan tres momentos del año con altos valores de *NDVI*, ocurriendo estos mismos separados en el espacio. El comportamiento variable particular del sector Sur, sumado al uso diferencial de la tierra, provocaron que, por un lado se rechazara la variable Uso de la Tierra dentro de este análisis, pero por otro lado, se la considerara como una variable que, tal vez a otra escala de análisis, estaría condicionando la respuesta del *NDVI*.

En cuanto a la carga de los valores de *NDVI* en el primer eje, aparecen con mayores valores (signo negativo) los dos meses en que el suelo de la zona sur de la cuenca está desnudo, por lo que se supone un *NDVI* mínimo. Esto coincide con el mes de julio, cuando se siembra la mayor parte del trigo de ciclo intermedio y corto, y con octubre, mes de siembra de los cultivos de verano (predominantemente soja, maíz y papa). Por su parte, la carga negativa de la variable Argiudoles está indicando que estos valores se alcanzan en este subgrupo de suelos, contrariamente a lo que ocurre en los Natracuoles. La interpretación en este caso es que los cultivos de la cuenca están asentados principalmente sobre los suelos Argiudoles, hecho que concuerda con su aptitud máxima de uso agrícola y con los resultados obtenidos en [Gandini \(2006\)](#).

El segundo eje canónico estaría separando por comportamiento los usos de la tierra agrícolas y ganaderos, con altas cargas positivas en dos suelos de baja representatividad espacial en la cuenca, como los Complejos indiferenciados y los Natracuales mólicos. Estos últimos están restringidos a la zona centro-Este de la cuenca en explotaciones típicamente mixtas agrícola-ganaderas. Las cargas negativas se concentran en otro suelo de baja cobertura y las lluvias que, con alta carga, estarían definiendo los valores de *NDVI* de noviembre, diciembre y enero (también con alta carga negativa). El tercer eje parece tener un comportamiento

opuesto al segundo en valores de carga, lo que llevaría a interpretarlo de manera inversa a aquél, pero con el aporte de una alta carga de Natracualfes (negativa) y Argiudoles (positiva) afectando los NDVI de junio (negativa) diciembre y enero (positiva). La interpretación de este eje no es tan clara, pero puede arriesgarse una relación también con el uso mixto de la zona centro-Oeste, donde el uso ganadero se desarrolla principalmente en los suelos Natracualfes y el agrícola en los Argiudoles.

En la Tabla 4 se observan los valores de correlación total y los residuales según se van extrayendo los ejes canónicos. Se observa una gran absorción de la correlación por el primer eje, acompañado en parte por los dos siguientes. Aun existiendo correlaciones significativas con los ejes siguientes, los tres primeros ejes sirven para explicar la principal estructura de correlación entre las matrices utilizadas. Por otro lado, las varianzas explicadas de cada combinación lineal son mayores a 1 (explican más que cada variable individualmente), en tres ejes para el grupo Y, y en cuatro ejes para el grupo X (datos no mostrados). Por este motivo la discusión involucra los resultados basados en los tres primeros ejes canónicos.

**Tabla 4.** Autovalores de los ejes canónicos y parámetros de significación.

<i>Correlación Canónica</i>	<i>Autovalor extraído</i>	<i>Chi Cuadrado</i>	<i>Grados de Libertad</i>	<i>Probabilidad</i>
Total	--	2159,268	72	0,000
0,737	1	933,756	56	0,000
0,469	2	545,373	42	0,000
0,405	3	265,227	30	0,000
0,258	4	157,817	20	0,000
0,242	5	63,442	12	0,000
0,180	6	12,127	6	0,059
	7	4,793	2	0,091

Los escores factoriales correspondientes a cada sitio de la matriz inicial fueron combinados en una imagen compuesta RGB, donde cada color del falso color compuesto indica la predominancia de cada uno de los tres ejes canónicos, en una zona del espacio. Zonas con predominancia de píxel color azul indican un alto escore en el primer eje. La misma interpretación debe hacerse con los colores verdes/segundo eje y rojo/tercer eje (Figuras 2-a (dependientes) y 2-b (independientes)).

#### 4. Conclusiones

La utilización del análisis de correlaciones canónicas provee una aproximación interesante para la interpretación de mapas generados por medio de sensores remotos. En este caso, la variación temporal de una variable usada

ampliamente para describir la dinámica de los ecosistemas como el *NDVI*, se correlaciona con las variables ambientales.

En particular, la serie de tiempo de *NDVI* utilizada como matriz de variables dependientes, muestra una variabilidad vinculada estrechamente con las principales variables que definen las zonas homogéneas de la cuenca del arroyo del Azul (centro de la provincia de Buenos Aires, Argentina), principalmente condicionada por los suelos -entre ellos los de mayor representatividad espacial como Natracuoles y Argiudoles, las Pendientes, y las Lluvias. Las restantes variables, si bien aportan a la correlación, lo hacen en menor medida y no afectan la significación de la correlación canónica, por lo que algunas fueron excluidas.

La rotación Varimax, unida a la depuración de variables produce un realce de las variables de mayor peso en la definición de los ejes. Esto puede observarse en la Tabla 1. Las variables de mayor carga son los Natracuoles, Argiudoles, y el *NDVI* de junio y octubre para el primer eje, que separa los sitios de uso predominantemente agrícola del resto. En el segundo eje las variables Lluvias y tres de los suelos (que ocupan escaso porcentaje de la superficie de la cuenca), junto con los *NDVI* de enero y septiembre son las que tienen más carga, y este eje separaría los ambientes de la cuenca media y baja, caracterizados por un uso predominantemente ganadero, y con diferentes características edáficas. Estos dos ejes en conjunto, inducen a separar visualmente trazando límites entre zonas al norte y al sur de la cuenca. El tercer eje se define por los Natracualfes típicos y por el *NDVI* de diciembre y enero y diferencian zonas en un sentido Este-Oeste, o sea en forma perpendicular a los ejes anteriores.

La variación en el *NDVI* durante el ciclo estudiado puede considerarse afectada por la productividad de los agroecosistemas involucrados (Tucker y Sellers, 1986). Por lo tanto, puede proponerse que en este análisis se está realizando una evaluación preliminar de cuáles son las variables independientes que aportan a definir la productividad en esta cuenca.

El papel del SIG en esta metodología es la ordenación espacial de los resultados en un mapa que es combinación de los factores que explican la mayor parte de la variabilidad original (Gandini y Entraigas, 1996). La elaboración de una Imagen Compuesta con los tres factores de mayor varianza explicada para cada grupo de variables es un sustrato apropiado para el análisis regional, ya sea mediante una interpretación visual directa o por agrupamiento de los nuevos valores de los píxeles en clústeres, de acuerdo con los objetivos de la regionalización.

La utilización este conjunto de recursos estadísticos y computacionales, si bien requiere de un manejo ajustado del uso de las variables y la interpretación de resultados, presenta la ventaja de sintetizar en pocas variables un sistema complejo. La representación gráfica que se obtiene, resume en un mapa la

variabilidad espacial no perceptible a partir del análisis de las variables individuales. Este hecho, es una ventaja relevante, del método paramétrico de regionalización (Mabbut, 1968).

La metodología utilizada junto a la inclusión en el conjunto de variables de aquellas de importancia agroecológica ofrece la oportunidad de contar con una regionalización dinámica, que puede ser ajustada a posteriori de acuerdo a las tendencias en el Uso de la Tierra y/o cambios en las variables climáticas, o servir como base para la creación de planes maestros. Además, el uso de variables provenientes de sensores remotos ofrece un recurso invaluable para la detección de los cambios ecológicos como los producidos por variaciones en el uso de la tierra (Cherrill *et al.*, 1994).

Por otro lado, el uso de un SIG es una herramienta indispensable cuando además de la descripción interesa el modelado y manejo de las complejas relaciones presentes en los agroecosistemas.

En conclusión, en este estudio se presenta una combinación de herramientas útiles para la evaluación y caracterización de zonas que poseen abundante información disponible pero no sistematizada y escasamente procesada.

### **Agradecimientos**

Los Autores agradecen la puesta en línea de las imágenes Vegetation por el sitio Vito y, al Instituto de Hidrología de Llanuras de Azul y a la facultad de Agronomía de Azul por la infraestructura básica para el desarrollo del trabajo. Se agradece además a Jorge Adamoli y a Marta Collantes por la lectura crítica del manuscrito.

### **Referencias**

1. Boles, S.H.; Xiao, X.; Liub, J.; Zhang, Q.; Munkhtuya, S.; Chen, S.; Ojima, D. Land cover characterization of Temperate East Asia using multitemporal VEGETATION sensor data. *Remote Sensing of Environment* 2004, 90, 477–489.
2. Cañibano, M.A.; Gandini, M.; Sacido, M. Evaluación de la intensificación del uso de la tierra, en la cuenca del Arroyo Del Azul, Buenos Aires, Argentina. *Actas del XI Simposio Latinoamericano de percepción remota*, 2004, Santiago, Chile.
3. Chavez, P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment* 1988, 24, 459-479.
4. Cherrill, A.J; Mc Clean, C.; Watson, P.; Tucker, K.; Rushton, S.P.; Sanderson, R. Predicting the distributions of plant species at the regional scale: a hierarchical matrix model. *Landscape Ecology* 1995, 10(4), 197-207.

5. Eastman, J.R. IDRISI - A grid based geographic analysis system. Clark University, Graduate School of Geography, Worcester, Massachusetts 01610-1477, USA, 2006.
6. Gandini, M.L. Zonas homogéneas de la cuenca del arroyo del azul: algunos aspectos de su dinámica analizados usando sensores remotos y sig. Tesis doctoral UBA, 2006.
7. Gandini, M.L.; Entraigas, I.N. Regionalización agroecológica del partido de Azul. Publicación 2 Programa Institucional de Investigación y Desarrollo Tecnológico. Azul, Argentina, 1996, 17 pp.
8. Gandini, M.L.; Usunoff, E.J. SCS curve number estimation using remote sensing NDVI in a GIS environment. *Journal of environmental Hydrology* 2004, 12, paper 16, 9 pp.
9. Harman H. Modern Factor Analysis. The Univ. Chicago Press, Chicago, 1976, 487 pp.
10. ITT Visual Information Solutions. ENVI Manual. Boulder, Colorado: ITT Visual Information Solutions, 2006.
11. Landgrebe, D.A., Biehl, L., An introduction to multiSpec Version 5, 2001. School of Electrical and Computer Engineering, Purdue University. [www.ece.purdue.edu/biehl/Multispec/Intro5-01.pdf](http://www.ece.purdue.edu/biehl/Multispec/Intro5-01.pdf) (accessed 10 March 2005).
12. Mabbut, G.A. Review of concepts of land classification. In *Land Evaluation*. G. A. Stewart ed. Mc Millan. Melbourne, 1968.
13. Noss, R.F. Ecosystems as conservation targets. *Trends in Ecology and Evolution* 1996, 11, 351-351.
14. R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2005.
15. Tucker, C.J.; Sellers, C. Satellite remote sensing of primary production. *International Journal of Remote Sensing* 1986, 7, 1395-1416.