

## **Análisis de componentes principales regionalizados aplicados a la clasificación digital de una imagen de satélite LANDSAT8 sensor OLI del año 2015, estudio de caso El Espinal, Tolima**

### **Regionalized Principal Component Analysis applied to a digital OLI LANDSAT8 2015 image classification. Case of study El Espinal, Tolima**

Edgar Andrés Rico Páez<sup>1</sup>, Luis Fernando Santa Guzmán<sup>2</sup> & Héctor Javier Fuentes López<sup>3</sup>

**Para citar este artículo:** Rico, E.A., Santa, L.F. & Fuentes, H.J. (2017). Análisis de componentes principales regionalizados aplicados a la clasificación digital de una imagen de satélite LANDSAT8 sensor OLI del año 2015, estudio de caso El Espinal, Tolima. *Revista de Topografía Azimut*, (8), 7-17.

**Recibido:** 2-diciembre-2016 / **Aceptado:** 2-marzo-2017

#### **Resumen**

En la presente investigación se aplicó la técnica geoestadística multivariada de Análisis de Componentes Principales Regionalizados (ACPR), la cual es una técnica de mejoramiento espectral que a diferencia del Análisis de Componentes Principales (ACP) describe la estructura de correlación espacial de las bandas de la imagen, a partir de la cual genera un nuevo grupo de bandas artificiales de menor dimensión, menos correlacionado y más interpretable. Se obtuvieron clasificaciones supervisadas con las bandas de ACP y ACPR como insumo, donde se eligió el algoritmo de clasificación Distancia de Mahalanobis, el cual ofreció los mejores resultados en la clasificación, según lo evaluado cualitativamente con la imagen satelital y las fotografías aéreas de GoogleEarth. La calidad de estas clasificaciones se determinaron calculando las matrices de confusión e índices Kappa ( $\kappa$ ), mostrando que la clasificación obtenida a partir de las bandas de ACPR para 60m ofreció la mejor exactitud temática, con el mayor valor de  $\kappa = 0.8640$ .

**Palabras clave:** ACP, ACPR, técnica geoestadística multivariada, estructura de correlación espacial, algoritmo de clasificación, exactitud temática

#### **Abstract**

The present investigation applied the multivariate geostatistical technique Regionalized Principal Components Analysis (RPCA) that is a spectral enhancement that unlike Principal Components Analysis (PCA), describe the spatial correlation structure of the image bands, from which gets a new artificial bands group less dimensional, less correlated and more interpretable. Supervised classifications were obtained with bands PCA and RPCA as a input, where the classification algorithm Mahalanobis Distance was selected because provided the best results in the classification according with the qualitative evaluation of the image satellite and Google Earth's aerial photographs. The quality of the classifications were obtained from the confusion matrix and Kappa Coefficients ( $\kappa$ ). Showing that the

1 Universidad Distrital Francisco José de Caldas, Carrera 8 # 40-62 Bogotá DC

2 Universidad Distrital Francisco José de Caldas, Carrera 8 # 40-62 Bogotá DC

3 Universidad Distrital Francisco José de Caldas, Carrera 8 # 40-62 Bogotá DC

classification obtained from the RPCA for 60m bands group provided the best thematic accuracy with the highest Kappa Coefficient  $\kappa = 0.8640$ .

**Keywords:** PCA, RPCA, multivariate geostatistical technique, spatial correlation structure, classification algorithm, thematic accuracy.

## INTRODUCCION

Las técnicas multivariadas como el (ACP) permiten reducir la dimensionalidad de un grupo de variables, teniendo en cuenta su estructura de correlación, eliminando la información redundante, ofreciendo un nuevo conjunto de variables no correlacionadas, las cuales son una combinación lineal de las originales (Isaaks & Srivastava, 1989).

En teledetección el ACP se aplica para incrementar la interpretabilidad de las coberturas presentes en una imagen multiespectral, muy útil por su capacidad de síntesis, eliminando la información redundante que pueden presentar algunas coberturas en ciertas longitudes de onda (Chuvieco, 2010).

El análisis de componentes principales regionalizados (ACPR) es una técnica geoestadística multivariada que calcula varias matrices de correlación espacial, también llamadas de correogionalización  $\Gamma(\mathbf{h})$ , las cuales describen la estructura de dependencia espacial de las bandas a analizar (Wackernagel, 2003), a partir de las cuales se obtiene un grupo de bandas artificiales no correlacionadas, de menor dimensión, ofreciendo una alternativa en la interpretación visual de coberturas.

Se utilizó la subescena de la imagen multi-espectral Landsat 8 (Path/Row= 8-57) del sensor OLI (Operational Land Imager) correspondiente al municipio de El Espinal (Tolima), con una resolución espacial de 30 metros que cubre un área de 16.906,05 has. Se escogió esta imagen

porque el porcentaje de nubes es (0 %) para el municipio de Espinal, además de ser temporalmente la más reciente, importante al momento de verificar resultados en campo.

Se trabajó con las bandas Aerosol (b1), Espectro Visible (b2, b3, b4), IRC (b5) y SWIR (b6, b7), adicionalmente con la banda pancromática (b8) como apoyo en la interpretación visual debido a su resolución espacial de 15 m. Las bandas cirrus (b9) y las térmicas (b10, b11) no se tuvieron en cuenta en este estudio.

En el software R se implementó esta técnica geoestadística, donde se creó una función que calcula el nuevo grupo de componentes a partir de la información de las bandas de la imagen.

Utilizando como insumo las bandas artificiales obtenidas con ACP y ACPR se calculan clasificaciones supervisadas donde se evaluó su exactitud temática por medio de las matrices de confusión e índices Kappa ( $\kappa$ ), con el propósito de determinar cuál grupo de bandas ofrece la mejor clasificación.

## METODOLOGIA

A través de ATCOR 3 se corrigió el efecto atmosférico de las bandas originales, utilizando como datos de entrada los parámetros de calibración de radiancia C0 y C1 (Bias y Gain) en unidades de [mW/cm<sup>2</sup> sr micron], el Azimut Solar y el Zenith (Tabla No.1), además del DEM obtenido para la zona.

**Tabla 1.** Parámetros de calibración atmosférica

Banda	C0	C1
Banda1	-6,085921	0,00122
Banda2	-6,232057	0,00125
Banda3	-5,742788	0,00115
Banda4	-4,842642	0,00097
Banda5	-2,963455	0,00059
Banda6	-0,736984	0,00015
Banda7	-0,248403	0,00005
<b>Azimut Solar</b>	58,01	
<b>Zenit</b>	30,75	

En el software **ERDAS** se ejecutó el modelo *Principal Components Transformation* que permitió la obtención de los Componentes Principales utilizando como entrada la imagen corregida atmosféricamente. Este nuevo grupo de bandas se reescalaron a 8 bits, ya que ofrecieron un mejor contraste mayor interpretabilidad de las coberturas presentes en la zona.

El análisis de componentes principales regionalizados se realizó en el software **R** utilizando como información de entrada la imagen con correcciones atmosféricas, cuyo cálculo consistió de las siguientes etapas:

1. Se evaluaron 3 Modelos Lineales de Coregionalización (MLC), tales como: Exponencial, Esférico y Gaussiano, donde se calculó el error ajuste por medio de los índices estadísticos  $R^2$  de Pearson y Root Mean Square Deviation (**RMSD**).
2. Se creó la función **ACPR()** que permite calcular los componentes principales regionalizados para diferentes distancias de  $h$ , la cual aplica un reescalamiento a 8 bits a las bandas generadas. Obteniéndose

como resultado tres grupos de bandas para las distancias  $h$  (30m, 60m y 90m).

La leyenda definida para el presente trabajo se basó en el sistema de clasificación de coberturas **CIAF** en el nivel de RECONOCIMIENTO o NIVEL II, debido a que se trabajó con una imagen satelital de baja resolución espacial (30 metros) para la generación de cartografía a escala 1:50.000.

La clasificación se realizó a través de **ERDAS** donde se utilizó como información de entrada cada uno de los grupos de bandas (*ACP*, *ACPR<sub>30m</sub>*, *ACPR<sub>60m</sub>* y *ACPR<sub>90m</sub>*) y se seleccionó el algoritmo de clasificación Mahalanobis Distance, ya que al evaluar cualitativamente con la imagen original y las fotografías aéreas de GoogleEarth éste ofreció los mejores resultados en la clasificación.

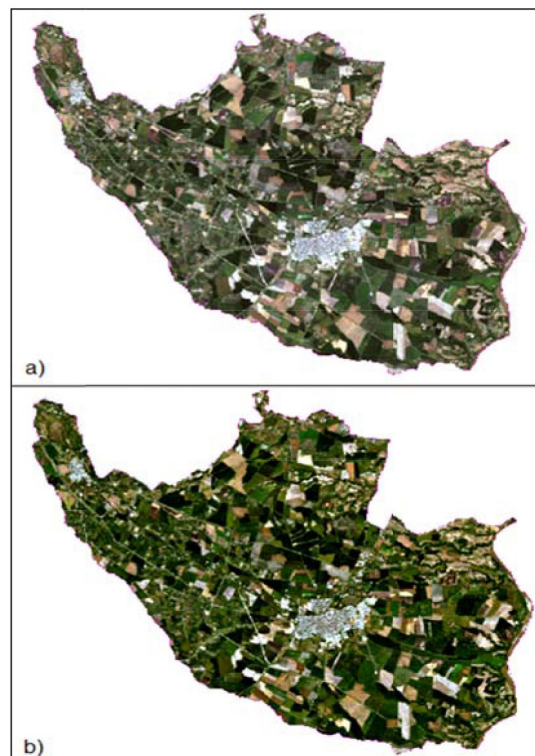
La exactitud temática se evaluó con información obtenida en campo y con 283 puntos de chequeo sobre cada imagen clasificada a partir de los cuales se calcularon las matrices de confusión y los índices Kappa ( $\kappa$ ) con sus respectivos intervalos de confianza con una significancia del 95% para una distribución normal.

## RESULTADOS

En la **Figura No.1** se presentan las imágenes de la zona de estudio antes y después de realizadas las correcciones atmosféricas en la combinación a color **RGB(432)** (*Color Natural*). En la imagen b) se observa una notable remoción del efecto de la absorción atmosférica, en la cual se percibe con mejor definición los colores, la textura y las

tonalidades de las coberturas de la zona de estudio, sin el efecto opaco causado por la dispersión o scattering.

En la Tabla No.2 se presenta la matriz de varianza-covarianza para la imagen corregida atmosféricamente, donde se observa que las mayores varianzas pertenecen a las bandas 5 y 6, aportando mayor variedad de información.



**Figura 1.** Imagen con y sin correcciones atmosféricas

**Tabla 2.** Matriz varianza-covarianza.

	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>
<b>B1</b>	60581	-	-	-	-	-	-
<b>B2</b>	74479	98644	-	-	-	-	-
<b>B3</b>	83907	117505	164178	-	-	-	-
<b>B4</b>	106121	156827	217692	314210	-	-	-
<b>B5</b>	-42112	-63107	-49535	-120323	724754	-	-
<b>B6</b>	84132	134860	213077	308486	138975	535533	-
<b>B7</b>	99465	147348	205718	297047	-43113	397976	351982

A partir de esta matriz de varianza-covarianza se calculan los autovalores y los autovectores los cuales se presentan en la Tabla No.3 y 4.

Donde los tres primeros componentes explican un 99.59% de la variabilidad total, reteniendo casi toda la información original.

**Tabla 3.** Matriz autovalores.

CP	CP1	CP2	CP3	CP4	CP5	CP6	CP7
<b>Eigenv.</b>	4513314	645205	62762	13872	4161	2762	495
<b>%Var</b>	86,09	12,31	1,2	0,26	0,08	0,05	0,01

**Tabla 4.** Matriz autovectores

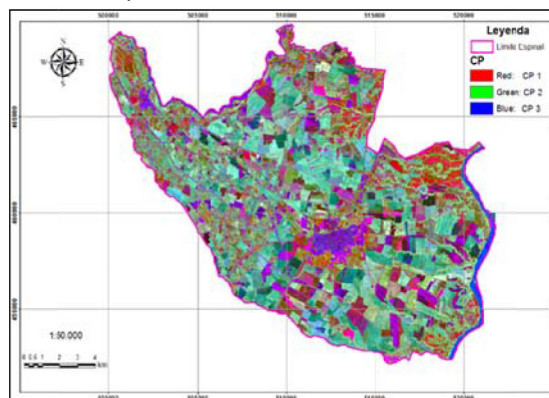
	CP1	CP2	CP3	CP4	CP5	CP6	CP7
B1	0,03	-0,16	0,40	0,53	-0,30	-0,05	0,67
B2	0,06	-0,24	0,42	0,36	-0,23	-0,26	-0,72
B3	0,15	-0,30	0,41	-0,25	0,00	0,81	-0,07
B4	0,16	-0,45	0,30	-0,61	0,09	-0,52	0,19
B5	0,75	0,59	0,27	-0,03	0,10	-0,07	0,00
B6	0,54	-0,33	-0,54	-0,01	-0,55	0,05	-0,01
B7	0,31	-0,41	-0,19	0,40	0,73	0,01	0,01

Mientras cada autovector indica la ponderación que debe aplicarse a cada banda.

En la **Figura No.2** se presenta los CP en una combinación a color RGB(123) donde se observa gran contraste, presentando la vegetación de cultivos temporales y permanentes en verde marino, mientras los bosques en verde claro, los pastos en

naranja-rojizo, los cultivos sin vegetación, el suelo desnudo y la zona urbana en morado brillante, mientras los cuerpos de agua en azul claro.

Para la obtención de los Componentes Principales Regionalizados se seleccionó el MLC Exponencial el cual presentó la mejor bondad de ajuste con los valores de R2 más cercanos a 1 y los valores más bajos de RMSD en todas las bandas como se muestra en la siguiente Tabla No.5.



**Figura 2.** Combinaciones a color de componentes principales RGB(123)

**Tabla 5.** Errores validación cruzada.

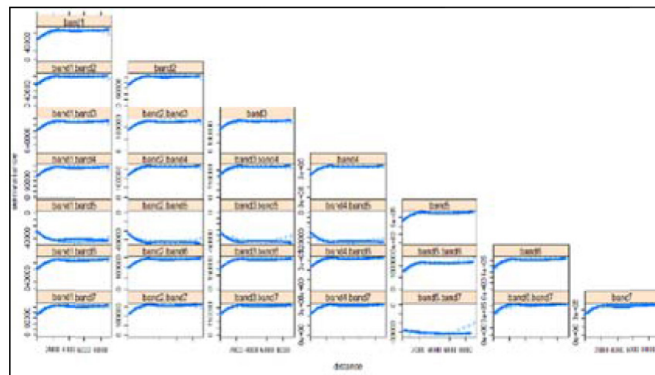
	Modelo	B1	B2	B3	B4	B5	B6	B7
Exp.	R2	0,91	0,93	0,97	0,97	0,79	0,97	0,97
	RMSD	92,3	118,0	90,3	115,8	382,6	136,3	98,0
Esf.	R2	0,89	0,92	0,94	0,94	0,78	0,92	0,94
	RMSD	93,5	118,3	90,5	117,2	387,8	138,4	99,5
Gaus.	R2	0,80	0,91	0,94	0,93	0,77	0,92	0,95
	RMSD	93,7	118,3	90,6	117,8	388,4	139,1	100,5

En la **Figura No. 4** se presenta la gráfica del MLC seleccionado, en el cual su diagonal contiene los semivariogramas de cada banda y los elementos fuera de ésta contienen los semivariogramas cruzados entre bandas, mostrando que la banda5 ofrece información única para cualquier distancia de h, ofreciendo semivariogramas cruzados negativos con las demás bandas de la imagen.

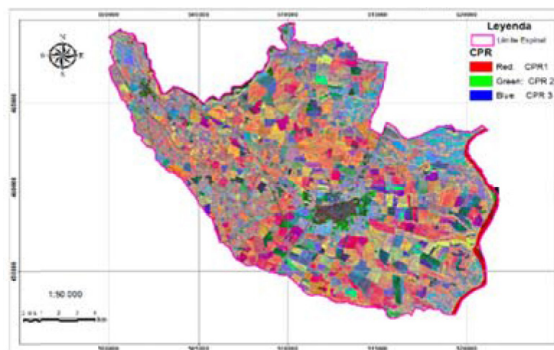
Al evaluar el MLC Exponencial para las distancias de 30m, 60m y 90m se obtienen las matrices de coregionalización, donde a partir de éstas se calculan

los autovalores y autovectores de igual forma como en ACP, necesarios para la obtención de los nuevos componentes principales regionalizados. A continuación se presentan las matrices obtenidas:

En las anteriores matrices se observa que en todas, la *banda 5* presenta los valores más bajos de correlación espacial con las demás bandas, mientras las bandas del espectro visible muestran una alta correlación espacial entre ellas. Siendo la banda5 la que muestra el valor más alto de semivarianza indicando que es la banda que más información específica aporta.

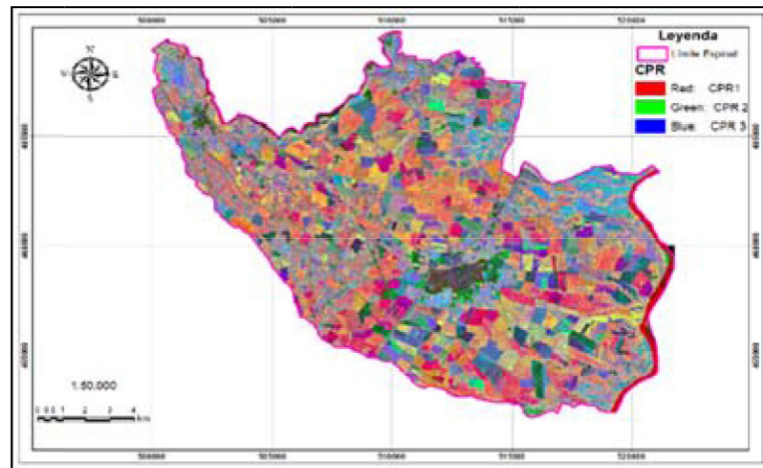


**Figura 4.** MLC Exponencial seleccionado

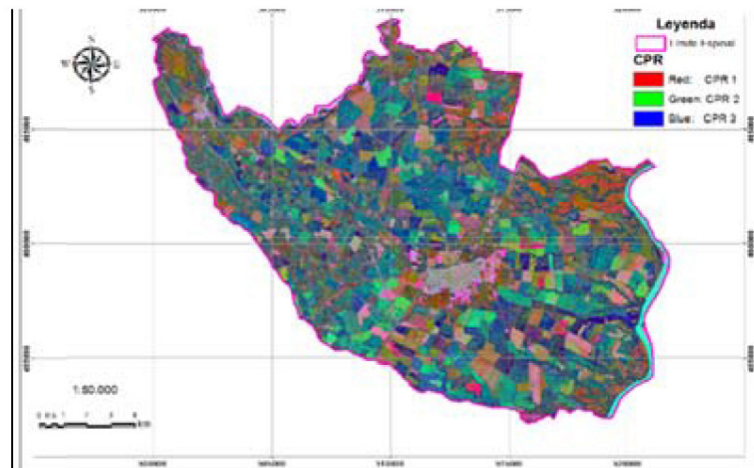


**Figura 5.** Combinación a color ACPR  $h=30$





**Figura 6.** Combinación a color ACPR  $h=60$



**Figura 7.** Combinación a color ACPR  $h=90$

A continuación en las siguientes figuras se presenta 3 combinaciones a color RGB(123) obtenidas para cada grupo de bandas o componentes.

Al comparar el grupo de bandas obtenidos con a **ACP** y **ACPR** se observa que en ACP se resalta de mejor forma los pastos en rojo cercanos al Río Magdalena, el suelo desnudo en fucsia y las zonas urbanas en morado. Sin embargo, en **ACPR**<sub>30m</sub> se presenta mayor variedad de colores en la vegetación del mosaico de cultivos y bosques, incrementando la separabilidad de estas coberturas, lograndose diferenciar de mejor forma el suelo desnudo de las zonas urbanas.

Para **ACPR**<sub>30m</sub> y **ACPR**<sub>60m</sub> no se observan diferencias significativas referentes a color, brillo, textura, tonos, etc., ofreciendo imágenes casi idénticas.

Al comparar **ACP** con **ACPR**<sub>90m</sub> se observa que en ACP el mosaico de cultivos se muestra en verde claro mientras en **ACPR**<sub>90m</sub> esta cobertura se presenta en una mezcla de colores entre azul y verde permitiendo una identificación mas clara; en ambas combinaciones a color se logra identificar con claridad las zonas urbanas, el suelo desnudo y los cuerpos de agua. Sin embargo en **ACPR**<sub>90m</sub> se observa un mejor contraste entre las coberturas de suelo desnudo (*rosado*), la vegetación de cultivos y bosques.

**Tabla 6.** Matriz de correionalización  $\Gamma(30)$ .

	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>
B1	39659	49964	56832	72437	-25433	51342	61072
B2	49964	68810	83228	112285	-42521	88341	96737
B3	56832	83228	120695	160367	-33564	143905	139663
B4	72437	112285	160367	233733	-91216	211132	205338
B5	-25433	-42521	-33564	-91216	525684	60165	-53571
B6	51342	88341	143905	211132	60165	305453	235118
B7	61072	96737	139663	205338	-53571	235118	216344

**Tabla 7.** Matriz de correionalización  $\Gamma(60)$ .

	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>
B1	40445	50880	57842	73682	-26004	52528	62470
B2	50880	69915	84493	113904	-43200	90010	98568
B3	57842	84493	122307	162442	-34047	146375	142043
B4	73682	113904	162442	236570	-92091	214547	208596
B5	-26004	-43200	-34047	-92091	531640	61894	-53846
B6	52528	90010	146375	214547	61894	312291	240154
B7	62470	98568	142043	208596	-53846	240154	220767

**Tabla 8.** Matriz de correionalización  $\Gamma(90)$ .

	<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>
B1	44044	55070	62461	79373	-28617	57951	68863
B2	55070	74969	90282	121312	-46307	97642	106945
B3	62461	90282	129677	171937	-36256	157674	152926
B4	79373	121312	171937	249545	-96092	230163	223500
B5	-28617	-46307	-36256	-96092	558881	69804	-55101
B6	57951	97642	157674	230163	69804	343566	263185
B7	68863	106945	152926	223500	-55101	263185	240996

En la Tabla No.9 se presentan la leyenda utilizada para la clasificación del grupo de bandas, mientras en la Figura No.8 se presentan las clasificaciones supervisadas obtenidas.

Se generaron 283 puntos de chequeo por el método aleatorio estratificado para cada imagen clasificada, a partir de los cuales se obtuvieron las matrices de confusión e índices Kappa.

En la Tabla No.10 se presenta los valores de Kappa obtenidos de las matrices de confusión calculadas para las clasificaciones donde se observa que el Kappa de mayor valor es el obtenido para la clasificación obtenida con las bandas de ACPR60m con un valor de 0.8640 y el rango más pequeño ofreciendo la clasificación con mayor exactitud temática del grupo de clasificaciones. De forma contraria el Kappa de menor magnitud es el obtenido para ACP y el rango más grande del grupo indicando que es la clasificación con menor exactitud temática.

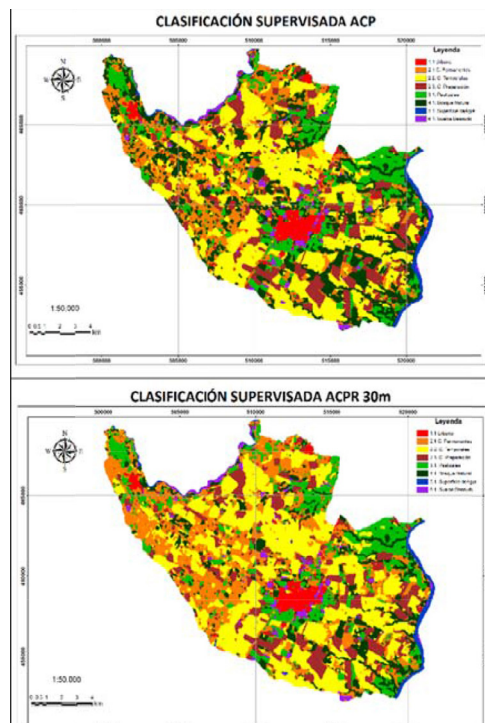


**Tabla 9.** Leyenda Nivel II

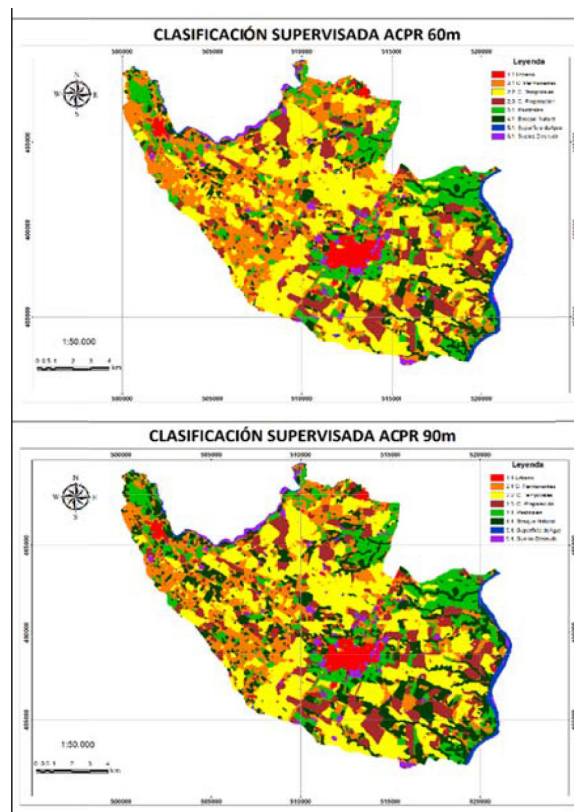
CLASE	SIMBOLO
<b>1.Construcciones</b>	
1.1Urbano	
<b>2.Cultivos y Parcela</b>	
2.1Cultivos Permanentes	
2.2Cultivos Temporales	
2.3Cultivos En Preparación	
<b>3.Vegetación Herbácea</b>	
3.1Pastizales	
<b>4.Bosques y/o Montes</b>	
4.1Bosque Natural	
<b>5.Cuerpos de Agua</b>	
5.1Superficies de Agua	
<b>6.Tierras Eriales</b>	
6.1Suelo Desnudo	

**Tabla 10.** Estadístico Kappa.

Clasificación	Kappa	$\sigma^2$	$IC_K$
ACP	0,8069	0,00066311	(0,7564;0,8574)
ACPR <sub>30m</sub>	0,8599	0,00051296	(0,8155;0,9043)
ACPR <sub>60m</sub>	0,864	0,00049902	(0,8202;0,9078)
ACPR <sub>90m</sub>	0,8274	0,00060673	(0,7791;0,8757)



**Figura 8.** Clasificaciones supervisadas.



## CONCLUSIONES

- Las matrices de correogionalización son muy útiles al momento de describir la estructura de correlación espacial de un grupo de variables especializadas como las bandas de una imagen de satélite.
- La clasificación que ofreció la mayor exactitud temática fue la obtenida a partir de las banda de ACPR para 60 metros, con el mayor valor de Kappa  $\kappa = 0.8640$  y el menor rango (0,8202; 0,9078), evidenciando que la técnica geoestadística ACPR aquí propuesta es útil para la clasificación de imágenes Landsat8.
- La exactitud temática de una clasificación supervisada depende de la información de entrada utilizada, además de la calidad del muestreo en la etapa de entrenamiento, del algoritmo de asignación, del conocimiento de la zona por parte del especialista, la habilidad y experiencia del intérprete en la clasificación de imágenes satelitales.
- Esta técnica de ACPR se puede aplicar a otras imágenes multispectrales, teniendo en cuenta que cada una de sus bandas se considera una variable regionalizada a la cual se le puede aplicar un análisis geoestadística como el realizado en la presente investigación.
- El lenguaje de programación R es una herramienta muy útil que permite al usuario crear sus propias funciones que se ajusten a sus necesidades. Además le permite realizar análisis geoestadísticos a información de teledetección que no se encuentran implementados en otros software de procesamiento de imágenes, tales como el cálculo del MLC, semivariogramas cruzados y simples, validaciones cruzadas, entre otros.

## REFERENCIAS BIBLIOGRAFICAS

- [1] Chuvieco E. Teledetección Ambiental. La observación de la tierra desde el espacio. Editorial Ariel S.A. Barcelona España. 2002.
- [2] CIAF Montoya J. Estudio Agrológico semidetallado y de recocimiento sector: Espinal-Suarez-Carmen Aplicada al Departamento del Tolima. Ibagué Tolima, 1971.
- [3] Giraldo H. R. Introducción a la Geoestadística Teoría y Aplicación. Facultad de Ciencias Departamento de Estadística Universidad Nacional de Colombia. Bogotá Colombia, 2002.
- [4] IGAC (Instituto Geográfico Agustín Codazzi) Corrección y Descripción de Productos Landsat8 LDCM. Grupo Interno de Trabajo en Percepción Remota y Aplicaciones Geográficas. Bogotá, 2013.
- [5] Isaaks E.H and Srivastava, R.M. An Introduction to Applied Geostatistics, Oxford University Press, England, 1989.
- [6] Jensen, J. R. Introductory Digital Image Processing: A Remote Sensing Perspective. New Jersey, 1996.
- [7] Melo L. Interpretación visual de imágenes de sensores remotos y su aplicación en levantamiento de cobertura y uso de tierra IGAC Centro de Investigación y Desarrollo de información Geográfica. Bogotá DC, Colombia, 2005.
- [8] Vargas, E. Análisis y Clasificación del Uso y Cobertura de la Tierra con Interpretación de Imágenes. IGAC. Bogotá DC. Colombia, 1992.
- [9] Wackernagel, H. Multivariate Geostatistics: An introduction with applications. Alemania, 2003.