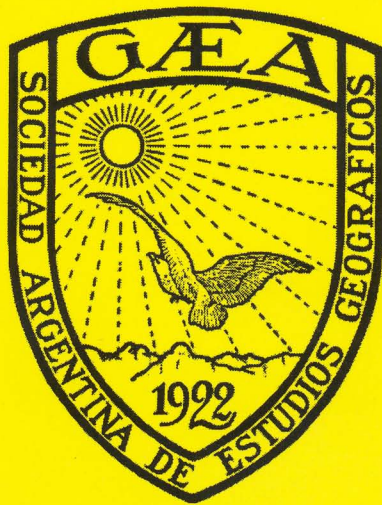


Contribuciones Científicas

GÆA - Sociedad Argentina de Estudios Geográficos

Volumen 20, 2008



AGENCIA

NACIONAL DE PROMOCION
CIENTIFICA Y TECNOLOGICA



Buenos Aires - 2008

MODELO DE PREDICCIÓN ESPACIAL PARA EL ÍNDICE DE ARIDEZ EN UNA TRANSECTA DEL CENTRO ARGENTINO

NARDECCHIA, Lucía ¹ - DIBLASI, Angela ^{1,2}
GONZÁLEZ LOYARTE, Margarita ³

1 Facultad de Ciencias Económicas, Univ. Nacional de Cuyo, Centro Universitario, Mendoza

2 Área de Ciencias Exactas/ CONICET CCT Mendoza, sede de CRICYT

3 Instituto Argentino de Investigaciones de las Zonas Áridas (IADIZA) CONICET CCT Mendoza, sede CRICYT

54-261-4135000 (interno 2436) | gnardecc@fcmail.uncu.edu.ar

Resumen

El procesamiento de imágenes satelitarias, conjuntamente con metodologías de los procesos estocásticos espaciales, ofrecen información sobre el comportamiento de variables cuyos valores son inexistentes. Particularmente, la obtención de un índice de aridez (IA) es imposible en puntos donde se carece de estaciones meteorológicas. Se utiliza la información de las imágenes NOAA-AVHRR NDVI (índice verde) para construir un modelo espacial predictivo para el NDVI en la transecta entre 69.51° W y 58.79° W y 34.40°S y 38.07°S. Mediante este modelo se obtienen predicciones de NDVI en los puntos de esta transecta donde se encuentran estaciones meteorológicas que permiten calcular el IA. Este índice es explicado mediante un modelo lineal por el NDVI. El modelo ajustado es aplicado luego a otros puntos dentro de la transecta para predecir valores del IA en puntos donde no existen estaciones meteorológicas. Así, mediante supuestos de continuidad espacial, se analizan patrones globales de comportamiento de aridez.

Palabras Clave: predicción espacial, modelos lineales, función de covarianza, imágenes NOAA-AVHRR NDVI, índice de aridez.

SPATIAL PREDICTIVE MODEL FOR THE ARIDITY INDEX IN A TRANSECT OF THE CENTER OF ARGENTINA

Abstract

Satellite image data jointly with Spatial Stochastic Process methodologies offer information about the behaviour of variables whose values do not exist. Particularly, it is impossible to calculate values of the Aridity Index (AI) in points where there are

no meteorological stations. Information based on NOAA-AVHRR NDVI images is used to build a predictive model for NDVI in the transect between 69.51° W and 58.79° W, and 34.40° S and 38.07° S. In the points where the meteorological stations are located, NDVI predictive values from this model are obtained. Then, a Linear Model is fitted to the NDVI predictive values to explain the calculated AI values in these points. This model is applied to obtain values of AI in points where there are no meteorological stations. Thus, aridity patterns are analyzed under spatial continuity assumptions.

Key words: spatial prediction, linear models, covariance function, NOAA-AVHRR NDVI images, aridity index.

Introducción

La facilidad de obtención de información relacionada con los fenómenos espaciales, tales como la proveniente de imágenes satelitarias, el desarrollo de tecnologías tales como los sistemas de información geográfica, han producido en estas últimas décadas un notable impulso a las metodologías estadísticas vinculadas a estos fenómenos. Estas metodologías permiten muchas veces establecer relaciones y construir modelos que utilizando la información disponible pueden ayudar a comprender otros cuya información directa no está suficientemente disponible. En particular, los modelos espaciales predictivos ofrecen una herramienta invaluable en este sentido. Algunos ejemplos de ellos son los trabajos de Cressie et al., 2006; Ver Hoef et al., 2004; Sanso and Guenni, (2000), etc.

En este trabajo utilizamos dos modelos estadísticos con los que, a partir de información suficientemente disponible (imagen de NDVI) analizamos patrones globales de otra variable cuyos valores son no existentes en considerables extensiones geográficas.

La carencia de estaciones meteorológicas en muchos sitios es una dificultad de la infraestructura que limita estudios sobre, por ejemplo, el comportamiento del índice de aridez en una región que los comprende. Por otra parte las imágenes satelitarias nos proveen de información que puede ser muy valiosa si podemos relacionarla adecuadamente con variables meteorológicas como la temperatura y las precipitaciones (Justice et al., 1986; Potter y Brooks, 1998). En este trabajo tomamos la transecta comprendida entre los meridianos $69^{\circ} 30' 48''$ W y $58^{\circ} 47' 27''$ W y los paralelos $34^{\circ} 24' 17''$ S y $38^{\circ} 03' 58''$ S por presentar un interesante gradiente (Fig. 1).

En este trabajamos, analizamos particularmente dos índices: NDVI o índice verde e IA o índice de aridez. Este último es de la forma P / ETP , donde P es la precipitación media anual y ETP la evapotranspiración potencial. Los valores de NDVI medios para las ubicaciones dentro de la transecta se obtienen a partir de las imágenes NOAA-AVHRR NDVI. Estos valores son transformados adecuadamente para llevarlos al intervalo $[-1, 1]$. A los efectos de construir modelos adecuados para encontrar patrones espaciales globales de variación del NDVI realizamos un análisis exploratorio de las observaciones proveídas por las imágenes mencionadas. Este modelo nos permite realizar predicciones en diferentes ubicaciones de esta región. Con las observaciones

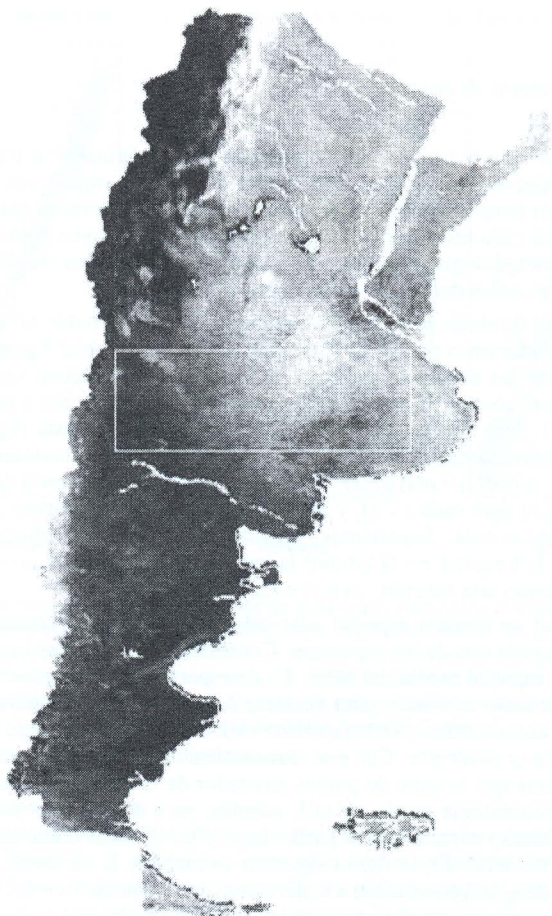


Fig. 1: El Rectángulo en una imagen de enero del NDVI señala la transecta utilizada en este trabajo.

de temperatura y precipitaciones en las estaciones meteorológicas con las contamos en el área, planteamos un modelo lineal entre los valores de NDVI predichos en cada sitio donde se encuentran las estaciones meteorológicas y del índice de aridez (IA) correspondiente. Esto nos permite ajustar un modelo entre estas dos variables (altamente correlacionadas) y aplicárselo a los valores de NDVI predichos por el otro modelo para predecir ahora valores del índice de aridez en puntos donde no se encuentran estaciones meteorológicas.

Metodología

Análisis exploratorio de datos

Para realizar una reducción del conjunto de datos originales se transforman las coordenadas espaciales a radianes, y se superpone sobre la región una grilla regular donde cada celda tiene lados de 0.004 radianes. Todos los valores originales de *NDVI* sobre puntos que caen en cada celda se promedian y se atribuyen al punto central de la celda. Esta metodología permite reducir el número de puntos sin distorsionar las características globales del índice en cuestión.

A los efectos de elaborar un modelo predictivo para la variable *NDVI* en la región en estudio, consideramos que el *NDVI* en cada ubicación espacial s puede expresarse como la suma de dos términos, uno para representar la tendencia o variación a gran escala y otro para recoger la posible dependencia local o variación a pequeña escala (Cressie, 1993). Más precisamente, consideramos a *NDVI* en esta región como un proceso estocástico continuo $\{NVDI(s): s \in D\}$ donde D es la región descripta anteriormente y donde $NVDI(s) = \mu(s) + \varepsilon(s)$ con μ una función no aleatoria de la ubicación espacial $s = (x, y)$ para cada $s \in D$ y donde $\{\varepsilon(s): s \in D\}$ es un proceso aleatorio isotrópico con media nula. Suponemos, además, que para cada ubicación espacial $s = (x, y)$ (x es la longitud e y la latitud) la variación a gran escala $\mu(s) = \mu(x, y)$ puede modelarse como una función: $\mu(x, y) = r(x) + t(y)$.

Para estimar un modelo espacial adecuado, realizamos previamente un análisis exploratorio de cada uno de los supuestos. Consideramos en primer lugar el supuesto de continuidad espacial media del *NDVI*. Es de esperar que si las observaciones provienen de un proceso continuo, para vectores h de norma suficientemente pequeña, los puntos con coordenadas $(NVDI(s), NVDI(s+h))$ deberían encontrarse próximos a la bisectriz del primer cuadrante. Con este razonamiento a medida que aumenta la norma de h es de esperar que la nube de puntos alrededor de la bisectriz debe tornarse más abierta. Esta metodología puede ser útil, además, para descubrir posibles outliers o puntos que presentan características particulares y no globales como las que estamos interesadas en modelar. En la figura siguiente se muestra la variación de *NDVI* para cada punto $s = (long, lat)$ en relación a la del punto de su entorno $s+h$. Hemos seleccionado tres direcciones: en el sentido de las latitudes (norte-sur), $h = (0, 0.004, 0)$; en el de las longitudes (este-oeste), $h = (0, 0, 0.004)$ y sudoeste-noreste $h = (0, 0.004, 0.004)$. El valor de *NDVI* para cada punto s se representa con el correspondiente al del punto $s+h$. Esperamos, entonces, que su variación en un punto del entorno sea “pequeña” ya que $\|h\|$ es “pequeña” y en consecuencia los puntos con coordenadas $(NVDI(s), NVDI(s+h))$ se ubican “aproximadamente” alrededor de la bisectriz del primer cuadrante (**Fig. 2**). En base a este análisis no descartamos la continuidad espacial del *NDVI*.

Con el propósito de analizar la estructura de la media en cada punto s como una función de la forma: $\mu(x, y) = (x) + t(y)$, estudiamos la variación media del índice *NDVI* en direcciones paralelas a los paralelos geográficos y los meridianos. La **Fig. 3**

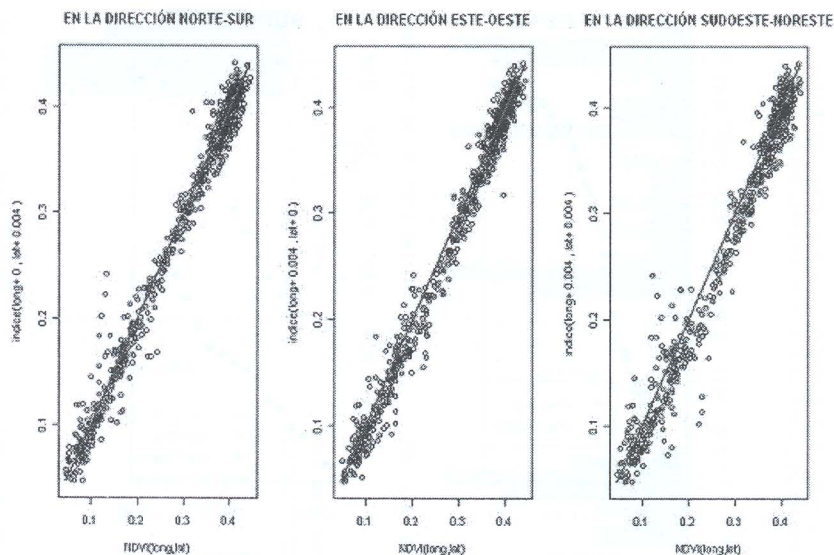


Fig. 2: En estos tres gráficos la abscisa de cada punto representa el valor de NDVI en un punto de coordenadas (longitud, latitud) de la región en estudio. En el de la izquierda, la ordenada representa el valor de NDVI en el punto de coordenadas (longitud, latitud + 0.004). En el central, la ordenada representa el valor de NDVI en el punto de coordenadas (longitud+0.004, latitud). En el de la derecha, la ordenada representa el valor de NDVI en el punto (longitud+0.004, latitud+0.004).

muestra los valores medios del *NDVI* para cada longitud (promediados sobre las latitudes) en la figura de la izquierda y para cada latitud (promediados sobre las longitudes) en la figura de la derecha. De esta figura podemos inferir que existe una tendencia creciente en los valores medios de *NDVI* en la dirección oeste-este y también sur-norte. También surge de esta figura que las distribuciones condicionales, para cada longitud fija, pueden considerarse simétricas. No sucede lo mismo con las distribuciones condicionales del *NDVI* para cada latitud fija, en el sentido de las longitudes (gráfico de la derecha) donde la mediana supera a la media indicando una fuerte influencia de algunos valores bajos.

Para analizar esta situación, analizamos la distribución probabilística del *NDVI* para algunas latitudes y longitudes fijas.

En la Fig. 4 podemos corroborar lo vislumbrado en la Fig. 3 respecto de la distribución del *NDVI* en las direcciones de los paralelos (este-oeste). En efecto, los tres histogramas de la parte superior de la Fig. 4 muestran la simetría de las distribuciones

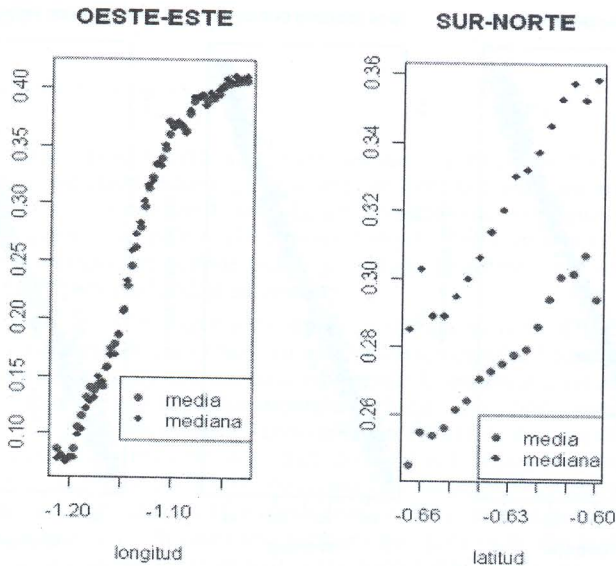


Fig. 3: En ambas figuras se muestra la media y mediana de los valores de NDVI para cada longitud fija (izquierda) y para cada latitud fija (derecha).

condicionales (para cada longitud fija). Las inferiores señalan la asimetría de las distribuciones condicionales en el sentido de las longitudes (este-oeste) para cada latitud fija. En éstas podemos observar, además, que la distribución parte de ser sesgada hacia la derecha a los $34^{\circ} 36' 23.94''$ hacia ser marcadamente más no-unimodal a medida que nos movemos sobre un paralelo más al sur, como se observa en la figura en el histograma de los $36^{\circ} 26' 24.42''$. Estas distribuciones condicionales nos limitan la posibilidad de considerar una distribución conjunta gaussiana.

Un modelo predictivo para NDVI

De acuerdo al análisis exploratorio anterior, consideramos para el NDVI un modelo de la forma

$$NDVI(x, y) = r(x) + t(y) + \varepsilon(x, y) \quad (1)$$

donde $\mu(x, y) = r(x) + t(y)$ es una función no aleatoria de la longitud (x) y la latitud (y) y donde vamos a considerar que $\{\varepsilon(x, y) : (x, y) \in D\}$ es un proceso isotrópico y con media nula. A los efectos de estimar la estructura de covarianza o variograma de $\{NDVI(x, y) : (x, y) \in D\}$ estimamos el de los residuos obtenidos a través de la metodología del pulido de medianas (Cressie, 1986) y lo ajustamos luego a un miembro de la familia exponencial por el método de mínimos cuadrados pesados (Fig. 5).

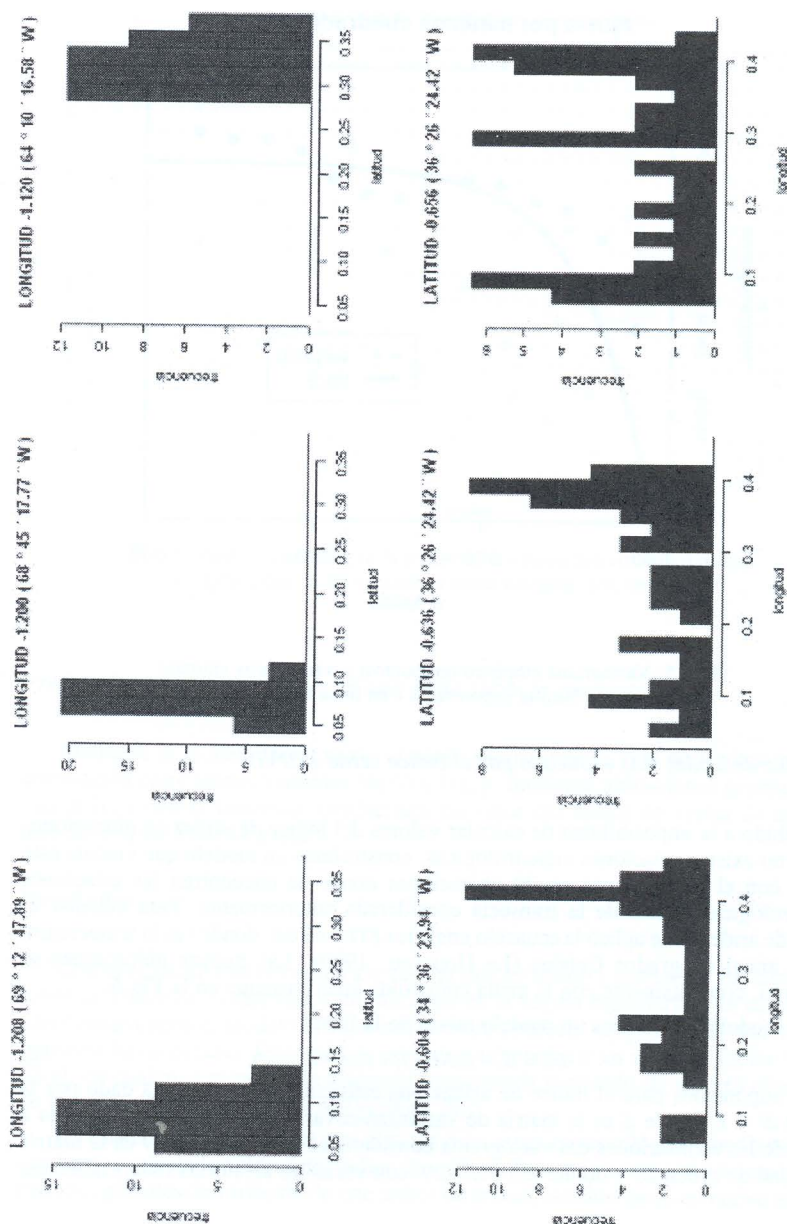


Fig. 4: En las figuras de la parte superior se muestran las distribuciones empíricas del NDVI en el sentido de las latitudes (norte-sur) para tres longitudes fijas. En las inferiores, se observan las distribuciones empíricas en el sentido de las longitudes (este-oeste) para cada latitud fija.

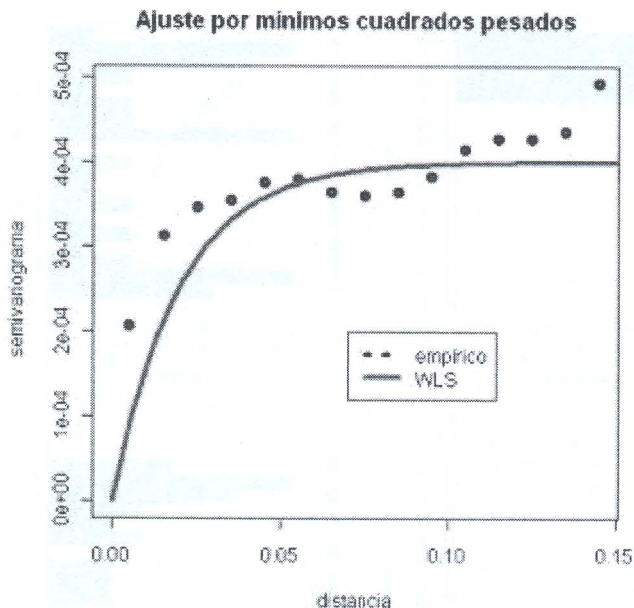


Fig. 5: Variograma empírico (en puntos) y variograma ajustado a la familia exponencial (en línea llena).

El índice de aridez (IA) explicado por el índice verde (NDVI)

Debido a la imposibilidad de calcular valores del índice de aridez en ubicaciones donde no existen estaciones meteorológicas, construimos un modelo que vincula este índice con el índice verde en 29 ubicaciones donde se encuentran las estaciones meteorológicas dentro de la transecta considerada anteriormente. Para calcular el índice de aridez IA se aplicó la ecuación empírica $ETP = 68.64t$, donde t es la temperatura media anual en grados Celsius (Le Houerou, 1989). Las nuevas ubicaciones se muestran, conjuntamente con la grilla construida anteriormente en la Fig.6.

El modelo a ajustar es un modelo mixto de la forma:

$$IA_i = \beta_0 + \beta_1 NDVI_i + \varepsilon_i, i = 1, \dots, 29 \quad (2)$$

donde suponemos para el índice de aridez una estructura de covarianza dado por la matriz $\Sigma + I$, donde Σ es la matriz de varianzas-covarianzas del $NDVI$ construida a partir de las estimaciones del variograma consideradas anteriormente, I es la matriz identidad de orden 29 y donde $\varepsilon_i, i = 1, \dots, 29$. son variables aleatorias independientes.

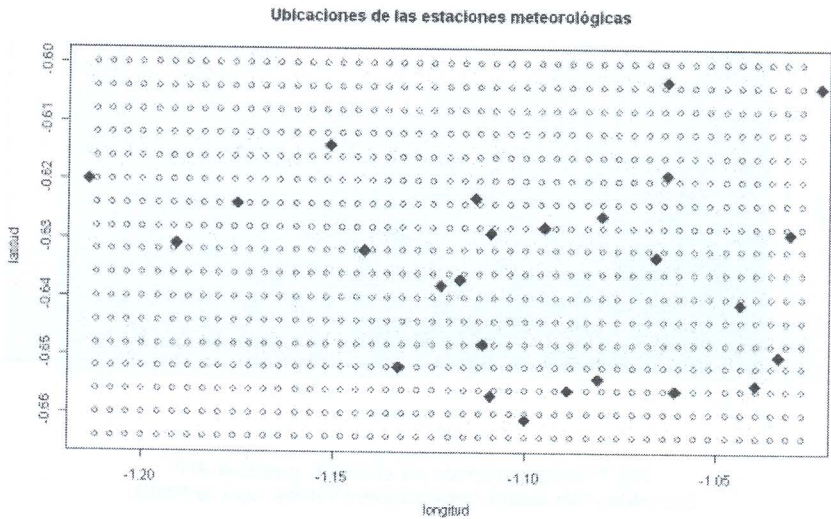


Fig. 6: Puntos generados en la transecta (en pequeñas circunferencias) y ubicaciones de las estaciones meteorológicas (en rombos).

Predicción del índice de aridez en una grilla sobre la transecta

Mediante el modelo lineal mixto ajustado anteriormente, que nos indica una alta correlación entre ambas variables ($NDVI$ y IA), y, mediante ubicaciones generadas en una grilla sobre la transecta, predecimos un valor del índice de aridez IA en cada uno de los puntos de la grilla.

Resultados

Con la estructura de covarianza o variograma estimado y ajustado y los supuestos sobre la tendencia, construimos un predictor en la región considerada de la variable $NDVI$ construimos el modelo (1) que nos permite predecir valores de $NDVI$ en toda la transecta bajo estudio. Este modelo nos permitió predecir un valor de índice verde, $NDVI$, en cualquier punto de la transecta mencionada partiendo de valores obtenidos de imágenes NOAA-AVHRR $NDVI$ (Fig. 7).

En esta transecta contamos con algunas estaciones meteorológicas mediante a las que hemos calculado valores del índice de aridez. Teniendo como objetivo encontrar patrones globales de variación de este índice de aridez, consideramos un nuevo modelo

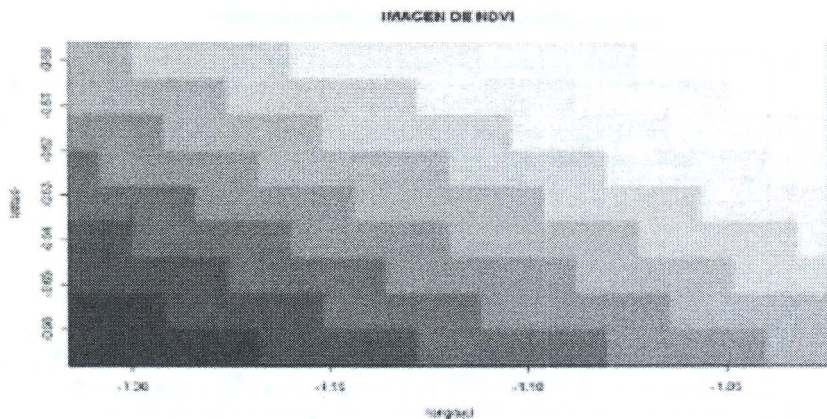


Fig. 7: Imagen generada por el modelo predictivo (1).
Los colores más oscuros corresponden a valores bajos de NDVI.

lineal (2) en el índice verde ajustado como un modelo mixto (Zhang, 2002). Este modelo nos da una relación entre el *NDVI* y el índice de aridez *IA* que nos permite predecir valores del segundo en función del primero y obtener gradientes de *IA* en la transecta. El modelo estimado resulta:

$$\hat{E}(IA) = 0.347 + 2.13NDVI$$

y la proporción de la varianza de explicada por este modelo es del 89% con un valor para el estadístico *F* del test de la bondad del ajuste de 238.4 con 1 y 27 grados de libertad al que corresponde un *p*-valor asociado de: 6.391×10^{-15} . Este modelo lo utilizamos para predecir valores del índice de aridez *IA* en otros puntos de la transecta en estudio.

La **Fig. 8** muestra la imagen construida a través de estas predicciones. Debido a la alta correlación entre ambas variables esta imagen es muy similar a la generada para *NDVI* en la misma grilla de puntos en la transecta.

Algunas reflexiones finales

La carencia de observaciones de variables muy importantes como la temperatura y las precipitaciones en alguna región limitan notablemente el análisis de fenómenos vinculados a estas variables. Sin embargo la posibilidad de contar con datos a través de imágenes de otras variables en la región en estudio nos ha permitido analizar patrones globales de comportamiento de las mencionadas variables meteorológicas o, más precisamente, de un índice de aridez, *Pet*, construido mediante ellas. Por supuesto

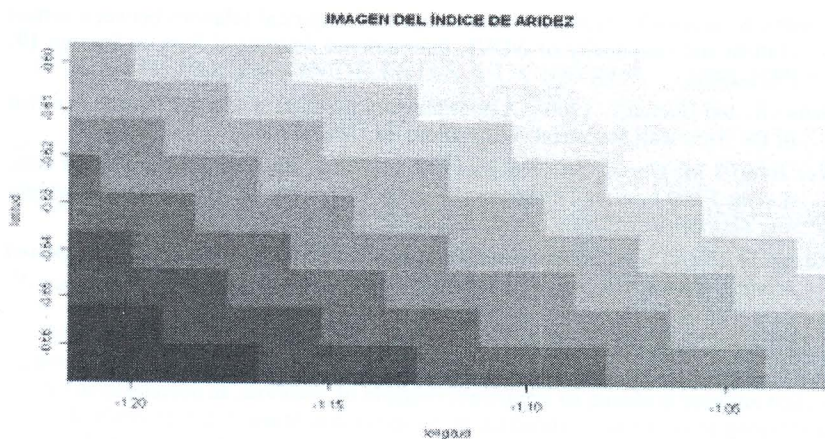


Fig. 8: Imagen generada por el modelo predictivo (1).
Los colores más oscuros corresponden a valores inferiores de IA.

estas relaciones las hemos podido realizar debido al ajuste de modelos estadísticos centrados particularmente en la predicción.

De acuerdo a la metodología utilizada, la predicción del índice de aridez (Pet) podría mejorarse en la franja oeste de la transecta donde se cuenta con pocas estaciones meteorológicas y, particularmente en la parte sur de esta franja si contáramos con información adicional.

Referencias

- Cressie, N. (1986), *Kriging Nonstationary Data*, Journal of the American Statistical Association, vol. 81, N°35.
- Cressie, N. (1993) *Statistics for Spatial Data*, Wiley series in Probability and Statistics, New York.
- Cressie, N., Frey, J., Harch, B., and Smith, M. (2006) Spatial Prediction on a River Network, *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 11, N°2.
- Justice CO, Holben BN, Gwynne MD. (1986). Monitoring East African vegetation using AVHRR data. *International Journal of Remote Sensing* 7(11): 1453-1474.
- Le Houérou HN. (1989). Classification éoclimatique des zones arides (s.l.) de l'Afrique du Nord. *Ecologia Mediterranea* 15: 95-144.

- Potter CS, Brooks V. (1998). Global analysis of empirical relations between annual climate and seasonality of NDVI. *International Journal of Remote Sensing* **19**: 2921-2948.
- Sanso, B. and Guenni, L. (2000) A nonstationary multisite model for rainfall, *Journal of the American Statistical Association*, 95, 1089-1100.
- Ver Hoeff, J.M., Cressie, N., and Barry, R.P. (2004) Flexible Spatial Models Based on the Fast Fourier Transform (FFT) for Cokriging, *Journal of Computational and Graphical Statistics*, 13, 265-282.
- Zhang, H. (2002), *On estimation and Prediction for Spatial Generalized Linear Mixed Models*, *Biometrics*, vol. 58, 129-136.