

Principales variables ambientales que explican el abandono de cultivos en el sureste de España

F. Alonso Sarría¹, C. Martínez Hernández², A. Romero Díaz², F. Cánovas García^{3,1}, F. Gomariz Castillo^{4,1}

¹ Instituto Universitario del Agua y del Medio Ambiente. Universidad de Murcia. Campus de Espinardo, 30100 Murcia

² Departamento de Geografía. Universidad de Murcia. Campus de la Merced, 30001 Murcia

³ Departamento de Geología y Minas e Ingeniería Civil. Universidad Técnica Particular de Loja. San Cayetano Alto s/n Loja, Ecuador

⁴ Fundación Instituto Euromediterráneo del Agua. Campus de la Merced, 30001 Murcia

alonsarp@um.es, carlos.martinez3@um.es, arodi@um.es, fulgencio.canovas@um.es, ffgomariz@um.es

RESUMEN: El abandono agrícola es un fenómeno global cuyas consecuencias ambientales son difíciles de evaluar. La Región de Murcia es una de las regiones más áridas del sur de Europa y también una de las más propensas al abandono. Este estudio trata de descubrir qué variables medioambientales (incluyendo variables climáticas y geomorfológicas, uso del suelo, tipo de suelo y tipo de roca), además de las distancias a áreas no cultivadas y distancias a las carreteras, son más relevantes para explicar el abandono a escala de parcela agrícola (25 metros de resolución de celda cuadrada). Las variables geomorfológicas se midieron a diferentes escalas para investigar qué escala podría ser más relevante. Para evitar el problema de colinealidad que surge cuando el número de predictores es elevado, en una primera fase se utilizó el factor de inflación de la varianza (VIF) para seleccionar un subconjunto de variables con baja correlación. Este subconjunto se ha utilizado para calibrar dos modelos diferentes: la regresión logística, modelo estadístico que permite la interpretación del efecto producido por las variables involucradas, y el *Random Forest*, modelo de aprendizaje automático con mayor capacidad predictiva pero que tiene el inconveniente de generar modelos menos interpretables. El uso combinado de estos modelos permite seleccionar un conjunto de variables predictoras que, cuando se utiliza con *Random Forest*, produce un mapa con una predicción muy exacta del abandono y, cuando se utiliza con regresión logística, producir un modelo muy interpretable. La principal conclusión es que, al menos a esta escala de trabajo, las características climáticas son más relevantes que las variables geomorfológicas para explicar el abandono de cultivos.

Palabras-clave: Abandono de cultivos, predictores ambientales, selección de variables, análisis de datos, regresión logística, *Random Forest*, aprendizaje automático, sureste de España.

1. INTRODUCCIÓN

El abandono de tierras es un fenómeno global, asociado a un descenso tanto en las prácticas agrícolas tradicionales como en la población rural, que conduce a la degradación del suelo (Moravec y Zemeckis, 2007; Lenda et al., 2012; Navarro y Pereira, 2012). En los países europeos más desarrollados el abandono alcanzó un máximo durante la industrialización del siglo XIX y tras la Segunda Guerra Mundial (Gellrich y Zimmermann, 2007). Hoy en día, en algunos casos, se relaciona con las políticas agrarias europeas (Renwick et al., 2013) y afecta a miles de kilómetros cuadrados en toda Europa.

En España, el abandono agrícola alcanzó su mayor intensidad durante los años sesenta y setenta del siglo XX (García-Ruiz et al., 2010). Más recientemente, el proceso de rápida urbanización (Jiménez-Herrero et al., 2005) y las políticas agrícolas europeas que subsidian el abandono de los cultivos menos rentables (Errea y Lasanta, 2001) han contribuido al proceso.

Desde 1980 el abandono agrícola ha afectado a un gran porcentaje de la superficie cultivada en la Región de Murcia, sobre todo en suelos frágiles en cuencas margosas neógeno-cuaternarias (Romero Díaz et al., 2007). Martínez-Hernández et al. (2013) indicaron que la tierra cultivada se redujo en un 46% en el período 1991-2001, mientras que la zona no cultivada aumentó en un 33%, especialmente las áreas de bosque y tierras abandonadas.

La evolución del uso del suelo en Murcia, como en toda la cuenca mediterránea, está relacionada con la dinámica socioeconómica, especialmente con la disminución de la población rural (Kosmas et al., 2002). Otros factores en los países del Mediterráneo occidental son la política agraria comunitaria, que trata de suprimir los cultivos menos rentables (Errea y Lasanta, 2001), la expansión urbana (Ghosh et al., 2014) o los cambios del mercado (García-Ruiz et al., 2010). Sin embargo, de acuerdo con el personal de la Consejería de Agricultura de la Comunidad Autónoma de la Región de Murcia, sólo entre el 5 y el 10% de las tierras abandonadas en Murcia se han beneficiado de subvenciones de la PAC (datos inéditos).

Dadas sus consecuencias, tanto ambientales como de gestión, es necesario evaluar el abandono e identificar tanto las áreas más propensas al mismo como los factores ambientales que pueden explicarlo. Todo ello para establecer políticas adecuadas de gestión y para atenuar los efectos negativos del abandono. En este trabajo partimos del supuesto de que, aunque los factores socioeconómicos globales pueden ser responsables de las tendencias generales que afectan a todo el territorio, las decisiones finales de los agricultores sobre el abandono de las parcelas individuales están más relacionadas con factores ambientales.

El objetivo principal de este estudio es identificar las variables ambientales más relevantes para explicar el abandono de tierras a escala local (parcela agrícola), que consideramos equivalente a una resolución de celda cuadrada de 25 metros, a partir de un mapa de áreas abandonadas en la Región de Murcia (Martínez Hernández *et al.*, 2013). Somos conscientes de que, como se ha mencionado anteriormente, las características socioeconómicas también son relevantes para explicar el abandono de tierras; sin embargo, a escala de parcela agrícola, las variables ambientales pueden ser más relevantes. Sería, además, muy difícil obtener información socioeconómica a una escala tan detallada.

2. MATERIAL Y MÉTODOS

2.1. Variables analizadas

Un gran número de variables ambientales se consideraron como posibles predictores del abandono:

- Uso del suelo, obtenido del Mapa de Cultivos y Aprovechamientos del Ministerio de Agricultura, Pesca y Alimentación con datos de campo recogidos en 2001. Este mapa fue, además, utilizado como origen del proceso de abandono, ya que sólo las áreas que aparecen cultivadas en este mapa se tuvieron en cuenta en el análisis posterior.
- Litología, a partir del mapa geológico del IGME, simplificado a 5 categorías que representan los 5 tipos de rocas principales en la región: rocas carbonatadas; detríticas; evaporíticas; metamórficas y volcánicas; y Cuaternario.
- Tipo de suelo, a partir del mapa de suelos producido en el proyecto LUCDEME (Alias y Ortiz, 1986-2004). El mapa fue reclasificado para cumplir con los criterios de 2007 (IUSS Working Group WRB, 2007). Las asociaciones menos frecuentes fueron reclasificadas en grupos mayores para evitar la presencia de clases poco frecuentes.
- Variables climáticas mensuales: precipitación (PrecTot), temperatura mínima y máxima absoluta y temperatura media (MinAbsTemp, MaxAbsTemp y AvgTemp), evapotranspiración potencial (PET) y radiación solar incidente (RAD); en total 72 capas. La precipitación y la temperatura se obtuvieron mediante modelos GLM (Gomariz-Castillo y Alonso-Sarría, 2013), la evapotranspiración potencial utilizando la modificación de Allen de la ecuación de Hargreaves (Allen et al., 1994) y la radiación solar utilizando el módulo de GRASS r.sun (Hofierka, 1997).
- Variables geomorfométricas calculadas a partir del MDE del IGN, con una resolución de 25 metros, utilizando módulos de GRASS (Neteler y Mitasova, 2008) y SAGA (Olaya y Conrad, 2009). Estas variables incluyen las más utilizadas y representativas de los diferentes grupos de variables geomorfométricas (Hengl y Reuter, 2009):
 - Elevación, el mencionado MDE.
 - Derivadas del terreno, calculadas con el módulo de GRASS r.param.scale (Wood, 1996), a partir del MDE: Pendiente (SLO), seno (SIN) y el coseno (COS) de la orientación, curvatura en el perfil (PROF) y curvatura plana (PLAN). Se calcularon con diferentes tamaños de ventana (3, 7, 11, 15, 19, 23, 27, 31, 35 y 39 píxeles) para obtener estimaciones a diferentes escalas (desde aproximadamente 0,5 ha a 95 has); en total 50 capas.

- Rugosidad usando diferentes estimadores: coeficiente de Melton (MEL) (Melton, 1965); Índice de Rugosidad del Terreno (TRI), de Riley et al. (1999); Medida Vectorial de Rugosidad (VRM) de Sappington et al. (2007); coeficiente de Iwahashi y Kamiya (IWA), citado en Iwahashi y Pike (2007). Estos cuatro coeficientes se calculan con los mismos tamaños de ventana utilizados para calcular las derivadas del terreno, lo que representa un total de 40 capas.
- Posición topográfica: con SAGA se calculó el índice de posición topográfica (TPI) de Weiss (2001), utilizando las mismas escalas que en los índices anteriores. De esta manera, se crearon 10 capas.
- Índices topográficos derivados: Índice de Humedad Topográfico (TWI) (Quinn et al., 1991; Beven et al., 1995); Índice Multiresolución de planitud del fondo del valle (MRVBF) (Gallant y Dowling, 2003) y el factor LS de la USLE usando la formulación propuesta por Moore y Burch (1986).
- Otras variables: Distancia a vías de comunicación (RoadDist), distancia a las áreas naturales (NatDist), ambos calculados con GRASS, y concentración de carbono orgánico del suelo (SOC) (Blanco-Bernardeau et al., 2014).

2.2. Selección de variables

Se trata en total de 182 capas (179 cuantitativas y 3 cualitativas). Además de la complejidad computacional, otros problemas que aparecen cuando se maneja un gran número de predictores son la colinealidad y el riesgo de sobreajuste del modelo. Uno de los métodos más habituales para hacer frente a la colinealidad es el Análisis de Componentes Principales (ACP). Sin embargo, los nuevos componentes creados con este procedimiento son difíciles de interpretar; y, debido a que el objetivo de este estudio es identificar qué variables son más relevantes para explicar el abandono, preferimos utilizar el factor de inflación de la varianza (VIF) (Zuur et al., 2007) que permite obtener un subconjunto de variables no colineales. De este modo los modelos resultantes son bastante más fáciles de interpretar.

El programa R (R Core Team, 2014) se utilizó para el análisis VIF y el resto del análisis estadístico en esta investigación. El algoritmo presentado en Zuur et al. (2009) se adaptó para calcular, de forma recursiva, el VIF de cada variable, seleccionar la variable con el VIF más alto, eliminarla del conjunto de datos, recalculando el VIF para el resto de variables y continuar con la siguiente iteración. Este procedimiento continúa hasta que no queda ninguna variable con un VIF superior a un determinado umbral. Zuur et al. (2007) recomiendan un umbral de 10 pero, siguiendo la recomendación más restrictiva de otros autores (Kutner et al., 2004; O'Brien, 2007), se utilizó un umbral de 5.

2.3. Métodos de clasificación

2.1.1. Regresión logística

La regresión logística es un caso particular de los Modelos Lineales Generalizados (McCullagh y Nelder, 1989; Zuur et al., 2009) para variables dependientes con rango entre 0 y 1. Por lo tanto, es especialmente útil para modelizar porcentajes o probabilidades. También se utiliza para clasificar respuestas binomiales y puede ser adaptado para el caso multinomial. La principal ventaja de la regresión logística sobre los métodos de aprendizaje automático más sofisticados es que el modelo resultante es fácil de interpretar, y su validez medioambiental puede evaluarse examinando los coeficientes de los predictores que entran en el modelo.

Para calibrar un primer modelo de regresión logística se utilizó un método de regresión por pasos que comienza con un modelo saturado con las variables que superaron el análisis VIF. Las variables van siendo eliminadas si su p-valor es inferior a 0,01.

2.1.2. Random Forest

Random Forest (Breiman, 2001) utiliza varios árboles de decisión (500 a 2000) entrenados a partir de un subconjunto aleatorio de casos (obtenido por *bootstrapping*) y un subconjunto aleatorio de variables. Cada nuevo caso se clasifica por un sistema de votación entre todos los árboles. La aleatoriedad añadida al proceso disminuye la correlación entre los árboles y el gran número de árboles reduce el error de generalización (Breiman, 2001; Pal, 2005; Prasad et al., 2006), proporcionando mejores resultados que otros métodos de clasificación (Breiman, 2001; Liaw y Wiener, 2002).

El principal inconveniente de *Random Forest* es que es más difícil de interpretar que un único árbol de decisión. Sin embargo, *Random Forest* ofrece varias métricas para interpretar el modelo. La importancia de una variable se evalúa en función de cómo cambiaría la predicción si los datos de ésta fueran permutados al

azar. Se pueden utilizarse varios estadísticos como estimadores de la importancia; en este estudio se utilizó la media de disminución en la exactitud para cada variable. En este trabajo se empleó la implementación del paquete de R randomForest (Liaw y Wiener, 2002).

La importancia de las variables, obtenida de *Random Forest*, se utilizó posteriormente para ordenar y seleccionar variables al calibrar nuevos modelos, tanto para *Random Forest* como para regresión logística.

2.4. Muestreo

Todo el análisis se llevó a cabo con un tamaño de pixel de 25x25 metros, lo que significa más de 18 millones de píxeles en la Región de Murcia. Para la selección de variables se obtuvo una muestra aleatoria de 2 millones de píxeles, en áreas cultivadas en 2001 según el Mapa de Cultivos y Aprovechamientos citado.

Los métodos de clasificación se calibraron con una submuestra aleatoria de 100.000 píxeles tomados de los píxeles utilizados para el análisis VIF. Las muestras se tomaron al azar, por lo que no están balanceadas; es decir, hay muchos menos casos de abandono (3,46%) que de no abandono. Otro subconjunto aleatorio de 100.000 píxeles se utilizó para construir curvas ROC (James et al., 2013) para la validación.

2.5. Validación

Tanto Random Forest como la regresión logística dan un valor de probabilidad de abandono. Sin embargo, estos valores no se deben considerar como un predictor de probabilidad directa (con 0.5 como umbral) ya que las clases no están balanceadas. En estos casos es mejor utilizar un umbral de probabilidad que maximice la exactitud de la predicción (Kuhn y Johnson, 2014). Además, en este estudio, el objetivo no es obtener una clasificación, sino un mapa de potencial de abandono cuyos usuarios puedan establecer el umbral en función de sus necesidades. En tales casos, los estadísticos de exactitud como el porcentaje de acierto o el índice kappa (Congalton y Green, 2008) no son adecuados. En su lugar, el área bajo la curva ROC (AUC) (James et al., 2013) es la opción más adecuada para estimar la exactitud del modelo. Las curvas ROC se calcularon con el paquete de R ROCR (Sing et al., 2005).

3. RESULTADOS

La Tabla 1 muestra las 42 variables cuantitativas, seleccionadas por el método basado en VIF, que resumen la variabilidad ambiental en el área de estudio, evitan la colinealidad y mantienen la interpretabilidad de variables y modelos. A estas 42 variables se añadieron las citadas 3 variables cualitativas (uso del suelo, tipo de suelo y litología), lo que hace un total de 45 variables.

Tabla 1. Variables con VIF inferior a 5 tras la selección de variables.

Tipo	Variables
Clima	Prectot8 Prectot9 Prectot10 Prectot12 MaxAbsTemp8 MinAbsTemp12 RAD12 PET4 PET9
Derivadas del terreno	SLO39 SIN3 SIN39 COS3 COS39 PROF3 PROF7 PROF11 PROF23 PROF39 PLAN3 PLAN7 PLAN11 PLAN15 PLAN19 PLAN23 PLAN27 PLAN31 PLAN35 PLAN39
Rugosidad	VRM3 VRM11 IWA3 IWA7 IWA11 IWA39
Posición topográfica	TPI3
Índices topográficos	MRVBF LS TWI
Otros	SOC NatDist RoadDist

La Figura 1, a la izquierda, muestra los resultados de ambos modelos de clasificación calibrados con las 45 variables seleccionadas y, a la derecha, las 30 variables más importantes de acuerdo con el algoritmo *Random Forest*. El modelo *Random Forest* tiene mayor capacidad de predicción que la regresión logística. Se observa que las características climáticas son más importantes que las geomorfométricas.

Un modelo con 45 variables sigue siendo demasiado grande debido al riesgo de sobreajuste y a su dificultad de interpretación. Para descubrir cuantas variables son realmente necesarias para alcanzar una exactitud suficientemente alta se calibró el modelo *Random Forest* con un número creciente de variables, comenzando por la más importante hasta llegar a un modelo con las 20 variables más importantes según

Random Forest. Los resultados aparecen en la Figura 2. Aunque sólo se analizaron las 20 más importantes, se alcanza rápidamente una meseta (AUC~0.91) con sólo 6 variables. Este resultado demuestra que una exactitud similar, o incluso mejor, a la obtenida con todas las variables puede obtenerse seleccionando las más importantes.

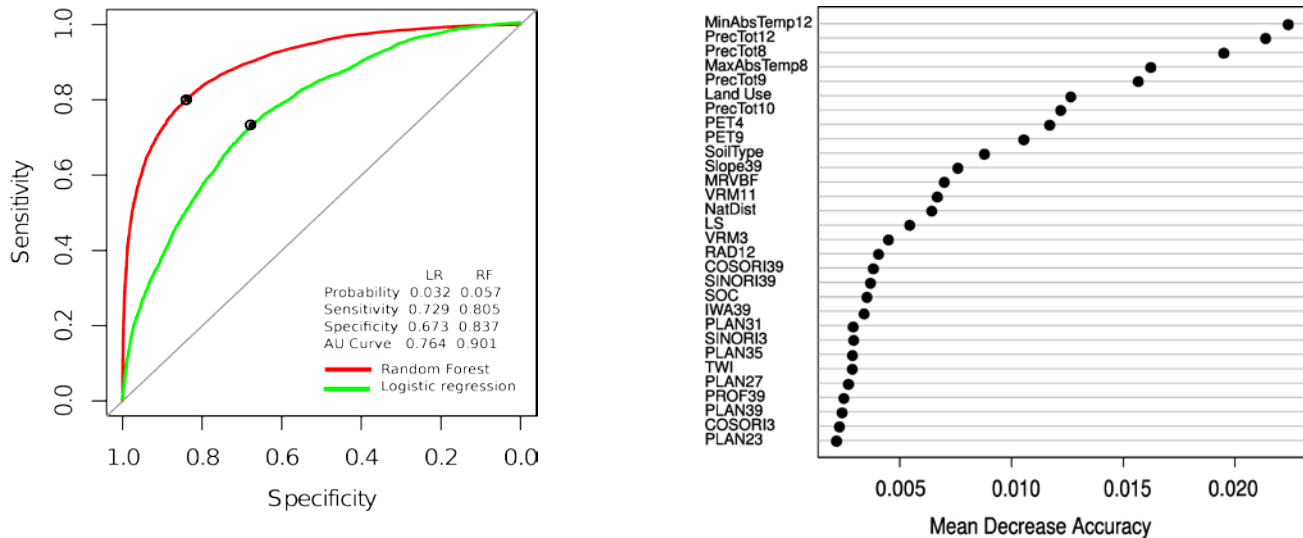


Figura 1. A la izquierda, curvas ROC para ambos modelos de clasificación con las variables seleccionadas. Se indican los puntos óptimos y los valores de umbral de probabilidad, sensibilidad, especificidad y área bajo la curva. A la derecha, importancia de las variables según el modelo *Random Forest*.

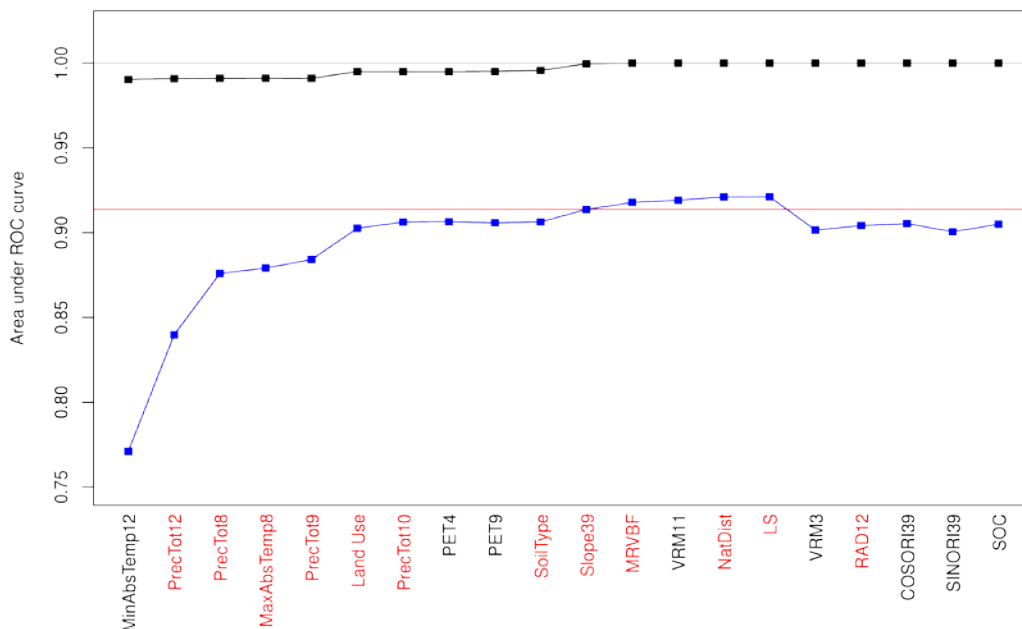


Figura 2. Aumento del área bajo la curva ROC con los datos de calibración (negro) y validación (azul), al añadir variables al modelo *Random Forest* en orden de importancia (sólo las 20 más importantes). Las variables en negro se rechazaron por no ser significativas en el modelo de regresión logística o porque el signo del coeficiente no respondía al tipo de efecto esperable. Las variables en rojo permanecieron en el modelo, la línea roja horizontal muestra la exactitud obtenida con ellas (AUC=0.917).

Random Forest no proporciona información sobre el efecto de las variables en el modelo. Para

entender qué papel juegan las diferentes variables repetimos el proceso de calibración secuencial con una regresión logística. Las variables, una vez más, se introducen en el modelo de acuerdo a su importancia según *Random Forest*; la diferencia es que las variables se mantuvieron u omitieron en el modelo en función de dos criterios: el p-valor del efecto tenía que ser menor que 0,05, y el efecto de la variable en el modelo debía ser razonable desde un punto de vista ambiental. En la figura 2 las variables que fueron rechazadas aparecen en negro, mientras que las variables en rojo se mantuvieron en el modelo. En total se mantuvieron 12 variables.

Para evaluar la capacidad predictiva del modelo basado en este subconjunto de 12 variables importantes para la predicción y cuyo efecto es además ambientalmente razonable en el modelo logístico, se repitió la calibración secuencial, pero esta vez tanto para *Random Forest* como para la regresión logística. La Figura 3 muestra, a la izquierda, el aumento en el área bajo la curva ROC al añadir variables en orden de importancia y, a la derecha, la importancia de estas en este modelo reducido. Una vez más, es posible reducir el número de variables en el modelo a sólo 8 (6 cuantitativas y 2 cualitativas) sin perder exactitud (AUC = 0,917). Durante el proceso de calibración secuencial el signo de los coeficientes y los p-valores de los predictores en el modelo de regresión logística se mantuvieron bastante estables al introducir cada nueva variable.

La Figura 4 muestra el efecto de las seis variables cuantitativas seleccionadas. Todos ellos son muy significativos (p-valores muy reducidos).

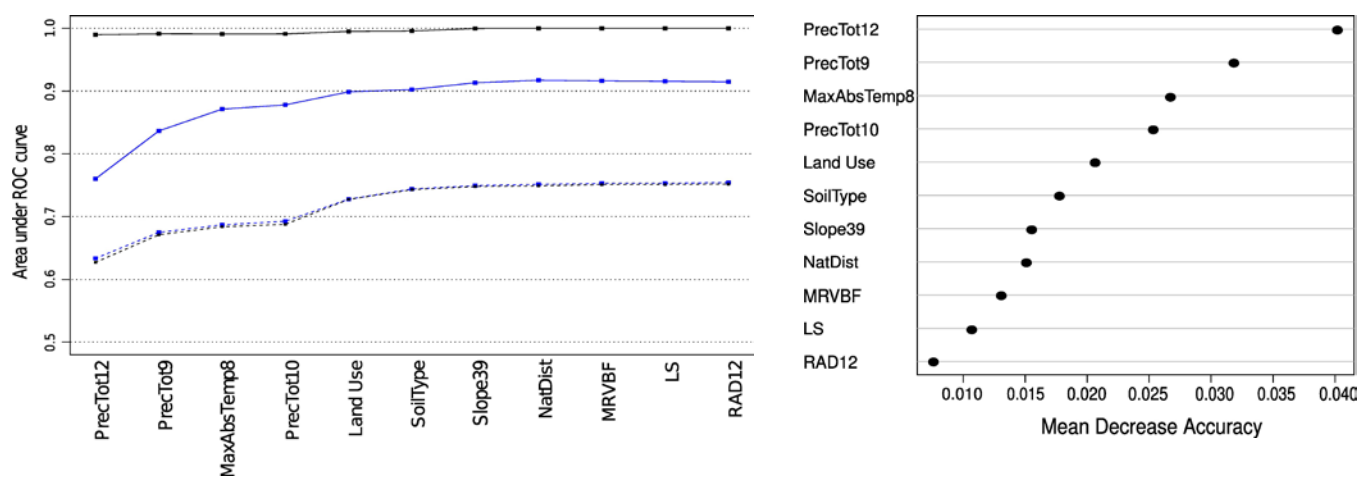


Figura 3. A la izquierda, aumento del área bajo la curva ROC al añadir a los modelos las 12 variables en orden de importancia. La línea negra muestra la exactitud calculada con los datos de calibración y la azul con los de validación. La línea sólida muestra el modelo de *Random Forest* y la punteada el de regresión logística. A la derecha, importancia de las 12 variables según el modelo *Random Forest*.

4. DISCUSIÓN

El conjunto de variables resultante del análisis VIF debe interpretarse como un subconjunto de los datos originales que resume estos reduciendo, al mismo tiempo, la colinealidad y sin comprometer la interpretabilidad. Cada variable incluida representa no solo a sí misma, sino también a todas las variables correlacionadas con ella y no incluidas.

Las variables climáticas resultaron ser más importantes que las geomorfométricas. Una de las razones que podrían explicarlo es que el clima es más relevante que la geomorfología; sin embargo, también es posible que los agricultores perciban los factores climáticos más fácilmente que otros. Entre las variables climáticas, las precipitaciones de octubre y diciembre tienen una correlación claramente negativa con el abandono. La Región de Murcia recibe precipitación tanto de sistemas frontales de invierno procedentes del océano Atlántico como de células convectivas generadas durante los meses de otoño en el mar Mediterráneo y ambos mecanismos son bastante independientes. La precipitación en octubre y diciembre representa estos mecanismos en el modelo.

El efecto en el abandono de cultivos de la precipitación en septiembre es positivo. Puede interpretarse como el efecto de las tormentas convectivas de final de verano en el interior de la región que pueden producir importantes pérdidas económicas en las zonas cultivadas. Otro efecto positivo es el de la

temperatura máxima absoluta de agosto. Dado que Murcia es una de las regiones más áridas de Europa, este efecto se relaciona con la cantidad de calor y la evapotranspiración potencial en verano, factor que conduce al abandono.

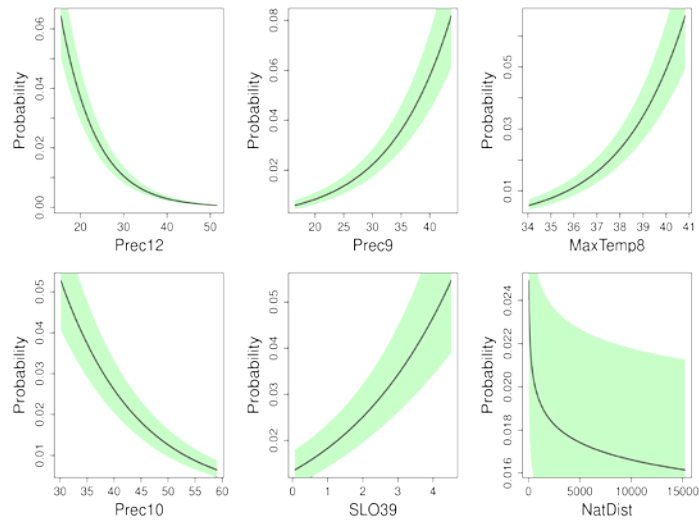


Figura 4. Efectos de las variables cuantitativas en el modelo de regresión logística, la precipitación se mide en milímetros, la temperatura en grados Celsius, la pendiente en grados y la distancia a espacios naturales en metros. Los p-valores fueron inferiores a 0.00005 en todos los casos.

Entre las características geomorfométricas, sólo la pendiente, calculada con una ventana de 39 píxeles (alrededor de 95 ha), aparece en el modelo final y con un efecto positivo. La distancia a las zonas de vegetación natural aparece también en el modelo en este caso con efecto negativo; sin embargo, su incertidumbre es mayor que en los otros casos. La interpretación es que la probabilidad de abandono de parcelas cultivadas muy cerca de cobertura de vegetación natural (menos de 500 m) es ligeramente mayor que en las parcelas más alejadas de áreas no cultivadas. Se trata posiblemente de áreas con menor vocación agrícola.

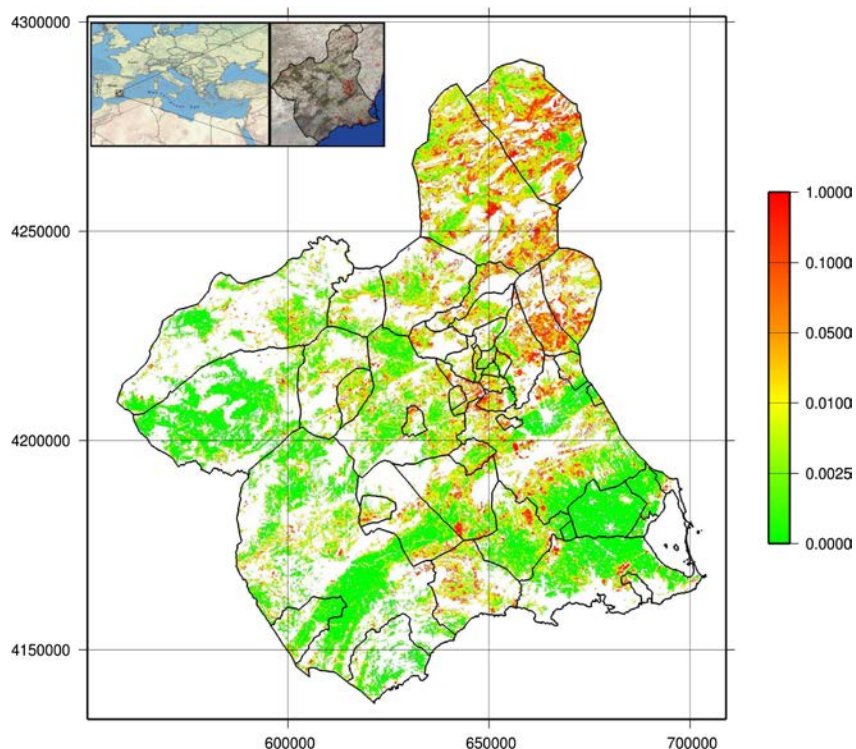


Figura 5. Mapa de abandono potencial a partir del modelo de *Random Forest*.

Los efectos de usos y tipos de suelo no se muestran en este trabajo por las limitaciones de espacio. A grandes rasgos, los cultivos de secano resultan ser más propensos al abandono que los de regadío. Los frutales de regadío y los viñedos son los cultivos más rentables y, por tanto, menos propensos al abandono. En un segundo grupo, algarrobos (un cultivo muy marginal) y cultivos herbáceos de secano son los más propensos al abandono; sin embargo, los intervalos de confianza son bastante amplios en estos casos.

La mayoría de los tipos de suelo cuyo efecto sobre el abandono es positivo presentan limitaciones para la agricultura. Los arenosoles son suelos poco desarrollados, con bajo contenido de nutrientes. Los solonchaks son suelos altamente salinos, los vertisoles son suelos con una alta proporción de arcillas expansivas, y, como resultado, forman grietas anchas y profundas cuando se secan. Los regosoles se forman a partir de materiales no consolidados, apareciendo sobre todo en las depresiones arcillosas donde se producen los procesos de erosión hídrica (Romero Díaz et al., 2007), los litosoles son suelos muy poco profundos (menos de 10 cm). Los fluvisoles cálcicos son suelos típicamente aluviales y fértiles, sin embargo, resultaron ser propensos al abandono. La causa es probablemente no tanto el suelo, sino su presencia en las zonas de inundaciones recurrentes, uno de los riesgos ambientales más importantes en la Región de Murcia.

La exactitud del modelo final que utiliza *Random Forest* con 8 variables es muy alta (AUC=0,917), ligeramente mejor que cuando se utilizan las 45 variables. La figura 5 muestra el potencial de abandono estimado con este modelo. El menor potencial de abandono aparece en zonas de regadío en el sur del área de estudio donde la disponibilidad de agua se combina con suelos fértiles; y en las zonas montañosas del noroeste, donde la precipitación es más alta y las temperaturas más suaves. El mayor potencial de abandono aparece en el noreste, donde el clima es más continental, árido y extremo.

La utilización de modelos predictivos basados en aprendizaje automático suele conducir a modelos con alta capacidad predictiva pero con un gran número de predictores difíciles de interpretar. La integración de regresión logística y *Random Forest* ha permitido obtener un modelo con un número reducido de variables cuyos efectos son fáciles de interpretar y ambientalmente razonables, sin perder capacidad predictora.

5. CONCLUSIONES

El análisis VIF permitió pasar de 182 a 45 variables que resumen la información del conjunto de datos inicial y no están correlacionadas entre sí. Combinar la importancia de las variables, calculada con *Random Forest*, con la información de regresión logística permitió reducir aún más esta cantidad a un subconjunto de 8 variables que producen un modelo interpretable y con alto poder predictivo. Las variables climáticas resultaron ser más relevantes que otras para explicar el abandono.

El mapa resultante indica qué zonas son más propensas al abandono en futuros ciclos económicos negativos, y así puede combinarse con mapas de riesgos ambientales relacionados con el abandono, como la erosión del suelo, para evaluar los niveles de riesgo en el área de estudio y destacar las áreas donde los esfuerzos para prevenir el abandono o mitigar sus consecuencias deben ser concentrados. Puesto que las variables climáticas resultaron ser las más importantes, pueden utilizarse escenarios de cambio climático para estimar los cambios resultantes en el potencial de abandono.

Una de las características más interesantes de la metodología utilizada es que se puede implementar fácilmente en cualquier otra área geográfica y también se puede adaptar a la cantidad de información disponible. Una interesante futura línea de trabajo sería utilizar MDE más detallados para repetir el análisis a mayor escala y entender cómo los resultados cambian con la escala.

AGRADECIMIENTOS

Este estudio se llevó a cabo en el marco del proyecto de investigación 15233/PI/10, financiado por la Fundación Séneca.

6. BIBLIOGRAFÍA

- Alias, L., Ortiz, R. (1986–2004): Memorias y mapas de suelos de las hojas del MTN a escala 1:50.000. Proyecto LUCDEME. Ministerio de Medio Ambiente.
- Allen, R., Smith, M., Perrie, A., Pereira L. (1994): “An update for the calculation of reference evapotranspiration”. *ICID Bulletin*, 43(2), 35-92.

- Beven, K., Lamb, R., Quinn, P., Romanowicz, R., Freer, J. (1995): "TOPMODEL". En Singh, V.P. (ed) *Computer Models of Watershed Hydrology*. Water Resour. Publ., 627–668.
- Blanco-Bernardeau A., Alonso-Sarría F., Gomariz-Castillo F. (2014): "Elaboración de un mapa de carbono orgánico del suelo en la Región de Murcia". XVI Congreso Nacional de Tecnologías de la Información Geográfica. Universidad de Alicante.
- Breiman L. (2001): "Random Forests". *Machine Learning*, 45, 5-32.
- Congalton, R.G., Green, K. (2008): *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*. CRC Press, 2 edn.
- Errea, M.P., Lasanta, T. (2001): "Les possibilités de development de l'élevage dans les champs abandonnés à la suite du gel des terres par la PAC en milieu semi-aride : L'exemple du «Campo de Saragosse» (Espagne)". *Sud-Ouest Européen*, 9, 75-84.
- Gallant, J.C., Dowling, T.I. (2003): "A multiresolution index of valley bottom flatness for mapping depositional areas". *Water Resources Research*, 39(12), 1-14.
- García-Ruiz, J.M., Lana-Renault, N., Beguería, S., Lasanta, T., Regüés, D., Nadal-Romero, E., Serrano-Muela, P., López-Moreno, J.I., Alvera, B., Martí-Bono, C., Alatorre, L.C. (2010): "From plot to regional scales: Interactions of slope and catchment hydrological and geomorphic processes in the Spanish Pyrenees". *Geomorphology*, 120, 248-257.
- Gellrich, M., Zimmermann, N. (2007): "Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: a spatial statistical modelling approach". *Landscape Urban Plan*, 79, 65-76.
- Ghosh, A., Sharma, R., Joshi, P.K. (2014): "Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion". *Applied Geography*, 48, 31-41.
- Gomariz-Castillo, F., Alonso-Sarría, F. (2013): "An R script to model monthly climatic variables with GLM to be used in hydrological modelling". 9th International R User Conference. Universidad de Albacete, Spain.
- Hengl, T., Reuter, H.I. (2009): *Geomorphometry: Concepts, Software, Applications*. Elsevier.
- Hofierka, J. (1997): "Direct solar radiation modelling within an open GIS environment". *Proceedings of JEC-GI'97 conference in Vienna, Austria*, IOS Press Amsterdam, 575-584.
- IUSS Working Group WRB (2007): *World Reference for Soil Resources 2006, first update 2007*. World Soil Resources Reports 103, FAO, Rome. http://www.fao.org/ag/agl/agll/wrb/doc/wrb2007_corr.pdf.
- Iwahashi, J, Pike, R.J. (2007): "Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature". *Geomorphology*, 86, 409–440.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013): *An introduction to Statistical Learning with Applications in R*. Springer.
- Jiménez-Herrero, L., Prieto del Campo, F., Riechmann-Fernández J., Gómez-Sal, A. (2005): *Informe de la Sostenibilidad en España 2005*. Universidad de Alcalá, Observatorio para la Sostenibilidad en España.
- Kosmas, C., Danalatos, N.G., López-Bermúdez, F., Romero-Díaz, A. (2002): "The effect of Land Use and Soil Erosion and Land Degradation under Mediterranean Conditions". En Geeson, N.A., Brandt, C.J., Thornes, J.B. (eds) *Mediterranean Desertification: a mosaic of processes and responses*, John Wiley & Sons, 57-70.
- Kuhn, M., Johnson, K. (2014): *Applied predictive modeling*. Springer.
- Kutner, M.H., Nachtsheim, C., Neter, J. (2004): *Applied linear regression models*. McGraw Hill.
- Lenda, M., Skórka, P., Knops, J.M.H., Moron, D., Tworek, S., Woyciechowski, M. (2012): "Plant establishment and invasions: an increase in a seed disperser combined with land abandonment causes an invasion of the non-native walnut in Europe". *Proc. R. Soc. B.*, 279, 1491-1497.
- Liaw, A., Wiener, M. (2002): "Classification and Regression by randomForest". *R News*, 2(3), 18–22.

- Mandal, D., Sharda, V.N. (2013): "Appraisal of soil erosion risk in the Eastern Himalayan region of India for soil conservation planning". *Land Degradation and Development*, 24, 430-437.
- Martínez-Hernández, C., Cánovas-García, F., Alonso-Sarría, F., Romero-Díaz, A., Belmonte-Serrato, F. (2013): "Cartografía de áreas agrícolas abandonadas mediante técnicas de SIG y fotointerpretación Comarcas de la Huerta y Campo de Murcia y Alto Guadalentín". *Espacios insulares y de frontera, una visión geográfica*. Universitat de les Illes Balears, 393-403.
- McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC.
- Melton, M.A. (1965): "The geomorphic and paleoclimatic significance of alluvial deposits in southern Arizona". *Journal of Geology*, 73, 1-38.
- Moore, I.D., Burch, G.J. (1986): "Modelling erosion and deposition: Topographic effects". *Transactions of the American Society of Agricultural and Biological Engineers*, 29, 1624-1630.
- Moravec, J., Zemeckis, R. (2007): "Cross Compliance and Land Abandonment, Deliverable D17 of the CC Network Project", SSPE-CT-2005-022727, 6-16.
- Navarro, L., Pereira, H. (2012): "Rewilding Abandoned Landscapes in Europe". *Ecosystems*, 15, 900-912.
- Neteler, M., Mitasova, H. (2008): *Open Source GIS: A GRASS GIS Approach*. New York, Springer.
- O'Brien, R.M. (2007): "A Caution Regarding Rules of Thumb for Variance Inflation Factors". *Quality and Quantity*, 41, 673-690.
- Olaya, V., Conrad, O. (2009): *Geomorphometry in SAGA*. en Hengl, T., Reuter, H.I. (eds) *Geomorphometry*. Elsevier, 141-169.
- Pal, M. (2005): "Random forest classifier for remote sensing classification". *International Journal of Remote Sensing*, 26, 217-222.
- Prasad, A.M., Iverson, L.R., Liaw, A. (2006): "Newer classification and regression tree techniques: bagging and random forests for ecological prediction". *Ecosystems*, 9, 181-199.
- Quinn, P., Beven, K., Chevallier, P., Planchon, O. (1991): "The prediction of hillslope paths for distributed hydrological modeling using digital terrain models". *Hydrological Processes*, 5, 59-79.
- R Core Team (2014): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Renwick, A., Jansson, T., Verburg, P.H., Revoredo-Giha, C., Britz, W., Gocht, A., McCracken, D. (2013): "Policy reform and agricultural land abandonment in the EU". *Land Use Policy*, 30, 446-457.
- Riley, S.J., De Gloria, S.D., Elliot, R. (1999): "A Terrain Ruggedness Index That Quantifies Topographic Heterogeneity". *Intermountain Journal of Sciences*, 5, 23-27.
- Romero-Díaz, A., Marín-Sanleandro, P., Sánchez-Soriano, M., Belmonte-Serrato, F., Faulkner, H. (2007): "The causes of piping in a set of abandoned agricultural terraces in Mediterranean Spain". *Catena*, 16, 282-293.
- Sappington, J.M., Longshore, K.M., Thompson, D.B. (2007): "Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert". *Journal of Wildlife Management*, 71, 1419-1426.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T. (2005): "ROCR: visualizing classifier performance in R". *Bioinformatics*, 21, 3940-3941.
- Weiss, A. (2001): "Topographic position and landforms analysis". ESRI User Conference. ESRI, San Diego, CA, USA.
- Wood, J. (1996): *The Geomorphological characterisation of Digital Elevation Models*. phThesis Diss., Department of Geography, University of Leicester, U.K.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M. (2009): *Mixed Effects Models and Extensions in Ecology with R*. Springer.
- Zuur, A.F., Ieno, E.N., Smith, G.M. (2007): *Analyzing Ecological Data*. Springer.