

FLUJO DE TRABAJO DEL RETO FINAL

Para poder predecir la idoneidad potencial de Sierra Nevada para el establecimiento de comunidades de enebrales se ha establecido un flujo de trabajo que podría dividirse en tres sub-apartados. Dos de ellos (NDVI y tendencia de nieve), están referidos a la preparación de dos variables para posteriormente incorporarlas al flujo principal, que se detalla en el tercer sub-apartado.

EXTRACCIÓN DE LOS VALORES PROMEDIO DE NDVI DEL AÑO 2014

En primer lugar, se deben extraer los valores promedio de NDVI correspondientes al año 2014, limitándolos al polígono de interés en Sierra Nevada. Mediante la utilización del lenguaje API Python se pueden extraer los datos correspondientes al NDVI en una tabla, que posteriormente será transformada en un archivo **asc*, que pueda ser importado a Rstudio.

CÁLCULO DE LA TENDENCIA DE NIEVE POR AÑO HIDROLÓGICO Y PIXEL

Tomamos como punto de partida una base de datos que contiene valores referidos a la cobertura de nieve en Andalucía (procedentes del sensor MODIS en el periodo comprendido entre 1999 y 2009). Los años 1999-2000 y 2008-2009 están incompletos por lo que se decide eliminarlos. Mediante el uso de un gestor de base de datos (Microsoft Access) se realiza una consulta que nos muestre el número de días que han tenido nieve cada uno de los píxeles por año hidrológico. Una vez obtenida la nueva tabla en formato **csv* e importada a Rstudio, se aplica un test de Mann Kendall para calcular la tendencia de duración de la nieve en un pixel y posteriormente un bucle para poder aplicar la tendencia a todos los pixeles disponibles. Ahora ya se tiene una tabla de datos con información acerca de la tendencia de cobertura de nieve por pixel durante el período establecido.

FLUJO PRINCIPAL

Tras haber extraído las capas correspondientes al NDVI promedio del año 2014 y a la tendencia de nieve por año hidrológico y píxel, se procede a juntar dichas capas con las variables disponibles inicialmente; relativas a precipitación (anual y de verano), huella humana, radiación solar, temperaturas máximas y mínimas de invierno y verano, y variables topográficas. Para ello utilizamos las funciones *brick* y *stack*, que nos permiten agrupar todas las variables en un único objeto espacial, lo cual hacemos para cada uno de los escenarios (presente y futuro) que queremos predecir.

Como hemos delimitado nuestro ámbito de estudio a Sierra Nevada, debemos ajustar la extensión de las variables, las cuales están referidas inicialmente a la totalidad de Andalucía. Para ello, contamos con un archivo **shp*, al cual hay que modificarle la extensión a **asc* (se ha utilizado Qgis en este caso) y rasterizarlo, para poder aplicar la función *crop*. Esta función nos permite ajustar los datos de nuestras variables a la extensión determinada por el archivo, ignorando todos aquellos valores que se queden fuera de los límites marcados por el mismo. Ahora ya podemos generar mapas para todas las variables en cada uno de los escenarios.

A continuación, se realiza un análisis de correlación de las variables, para observar el grado de correlación y si existe redundancia entre ellas que nos permita simplificar en la medida de lo posible nuestra base de datos. Para ello, primero debemos generar una tabla de datos a partir de los mapas creados previamente mediante la función *as.data.frame* y eliminar los valores nulos (función *na.omit*). Con la tabla generada se llevan a cabo sendas matrices de correlación (*cor*) y distancias eliminando previamente las correlaciones negativas (*abs* y *as.dist*). Una vez

aplicadas las matrices se lleva a cabo un cluster de las variables teniendo en cuenta que a mayor distancia menor grado de correlación. Observando los valores y representándolos en un gráfico cluster, se seleccionan solo aquellas variables que resulten significativas para los objetivos establecidos.

Una vez seleccionadas, se debe crear una tabla de datos que nos permita elaborar y evaluar el modelo de distribución. Para ello necesitamos importar los valores de presencia y ausencia del enebro, además de incluir las variables consideradas. Para poder juntar todas estas variables en una sola tabla debemos tener en cuenta que los registros de presencia y ausencia importados aparecen en proyección UTM y debemos transformarlos a coordenadas geográficas (y posteriormente a objeto longitud+latitud). Una vez convertidos, se extraen a una tabla y se agrupan en un único data.frame junto a las variables y las coordenadas, eliminando aquellos registros que no tengan valores para todas las variables.

Por último, computamos una regresión logística y elaboramos el modelo utilizando únicamente las variables significativas. Mediante la función *glm*, ajustamos un modelo lineal generalizado y lo aplicamos a las variables que se han establecido al principio mediante la función *predict*. Introduciendo los valores correspondientes a cada uno de los escenarios simulados, se obtienen diferentes mapas que nos permiten observar la idoneidad para la distribución potencial de enebrales en escalas temporales de 10 años desde la actualidad hasta el 2050.



