



# Aplicación de técnicas de *machine learning* al abastecimiento de agua en entornos de cooperación al desarrollo

Pedro Martínez Santos

Universidad Complutense de Madrid

Madrid  
5 Junio 2020



Las técnicas de machine learning tienen mucho que ver con el reconocimiento automático de patrones (patrones de consumo, patrones de comportamiento, patrones asociados a perfiles de riesgo, patrones de reconocimiento de imágenes, etc)



El reconocimiento de patrones también puede ser útil en la gestión de recursos hídricos

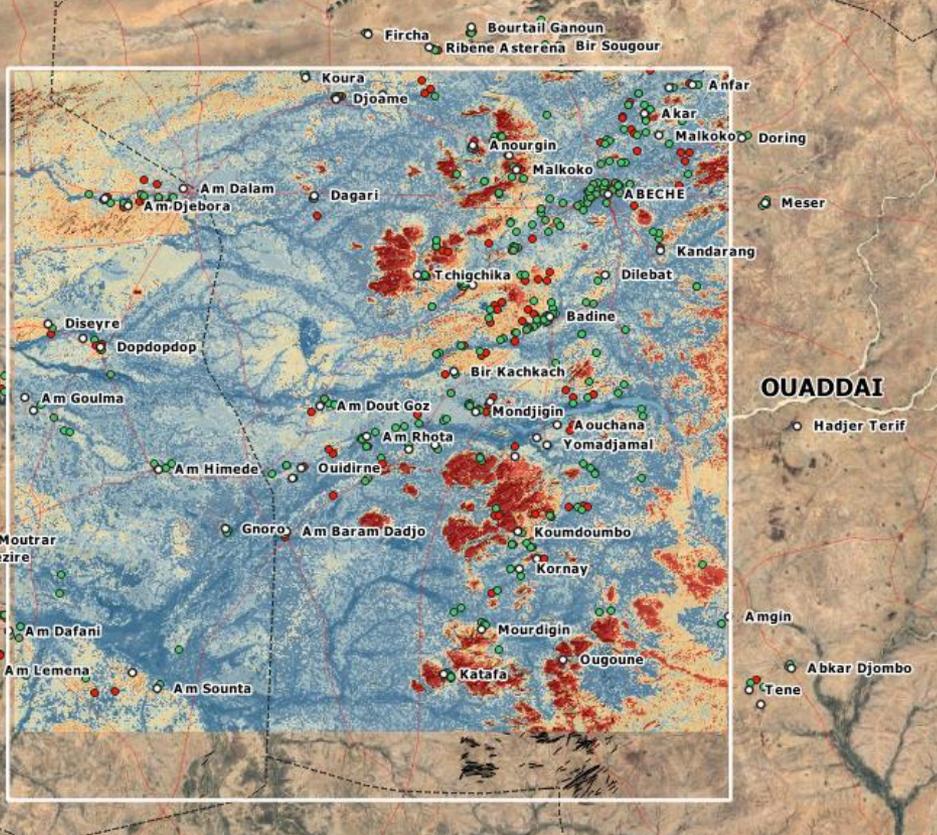
# Objetivo

- Aplicación de reconocimiento de patrones al *mapeo predictivo*, programada en QGIS y de libre acceso
- Se trata de **predecir una variable objetivo** distribuida en el espacio a partir de una serie de **variables explicativas**, también distribuidas en el espacio, y de unos pocos ejemplos resueltos
- Para ello **se utilizan trece algoritmos distintos de clasificación supervisada** (ScikitLearn), de los cuales se seleccionan solo los que mejores resultados dan
- El **resultado final es un mapa** en el cual, a partir del reconocimiento de patrones, extrapolamos desde lo que conocemos a lo que no conocemos
- Nos vamos a centrar en aplicaciones en el ámbito del agua en la cooperación al desarrollo, pero **la técnica puede extrapolarse a numerosos campos** relacionados (y no relacionados) con la gestión de recursos hídricos

### Mapeo predictivo del potencial hidrogeológico a escala país (República del Chad)

Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

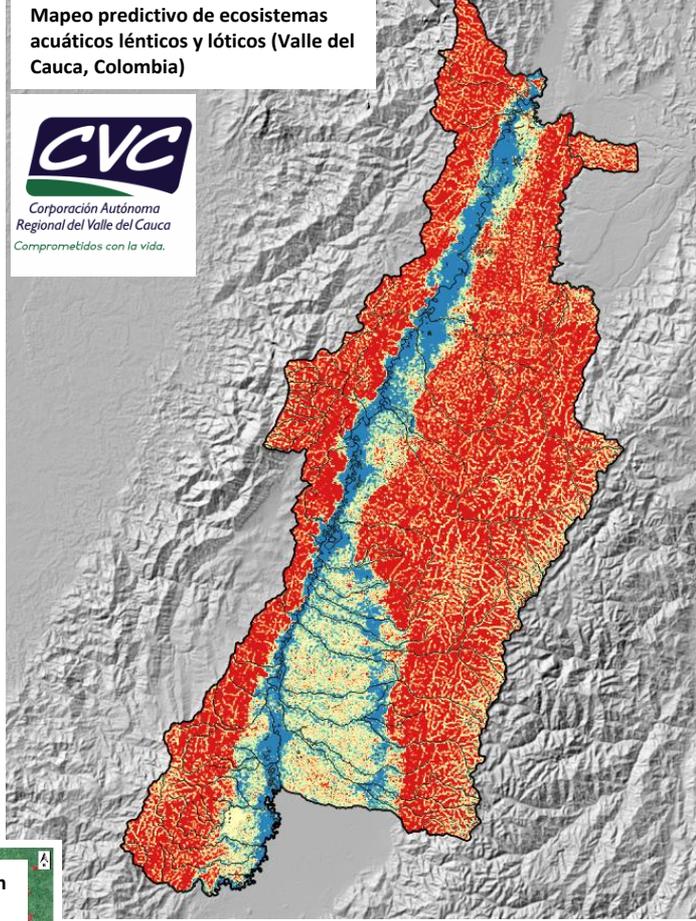
Swiss Agency for Development and Cooperation SDC



### Mapeo predictivo de ecosistemas acuáticos lénticos y lóticos (Valle del Cauca, Colombia)

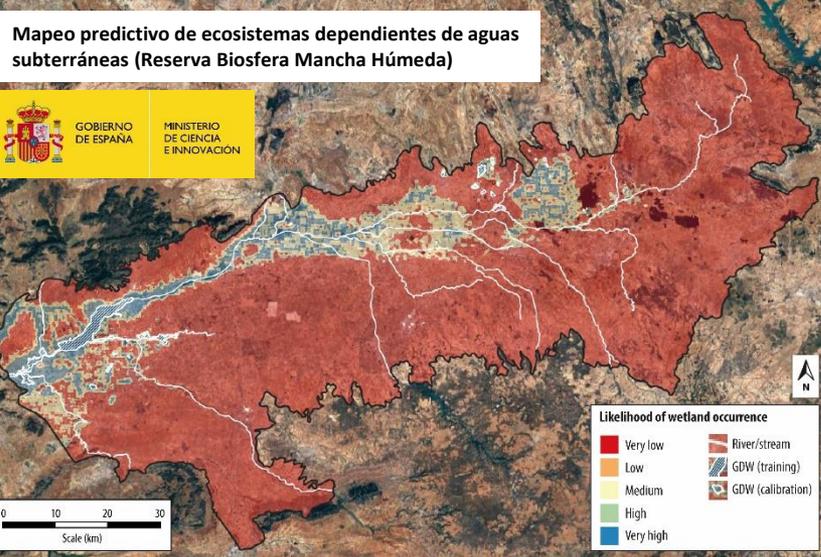


Corporación Autónoma Regional del Valle del Cauca  
Comprometidos con la vida.



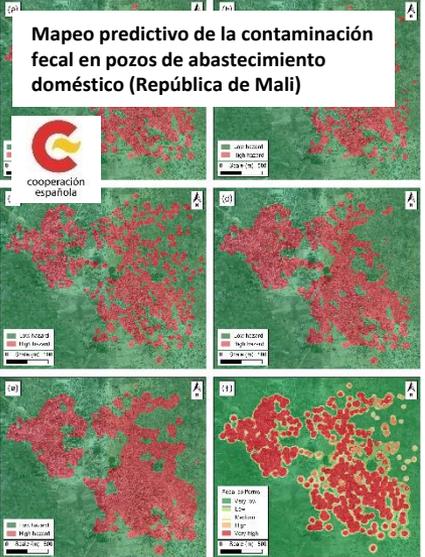
### Mapeo predictivo de ecosistemas dependientes de aguas subterráneas (Reserva Biosfera Mancha Húmeda)

GOBIERNO DE ESPAÑA  
MINISTERIO DE CIENCIA E INNOVACIÓN

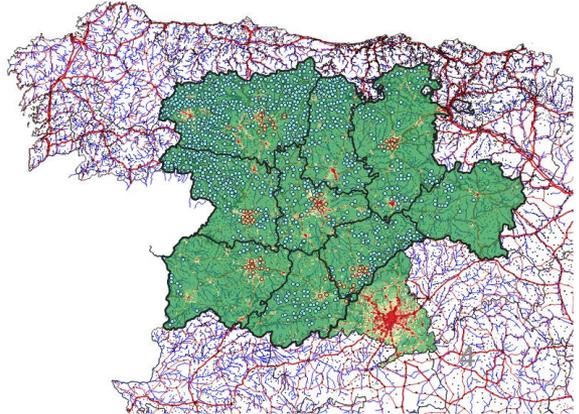


### Mapeo predictivo de la contaminación fecal en pozos de abastecimiento doméstico (República de Mali)

cooperación española



### Mapeo predictivo de los potenciales caminos de expansión del lobo ibérico en Castilla y León y Madrid

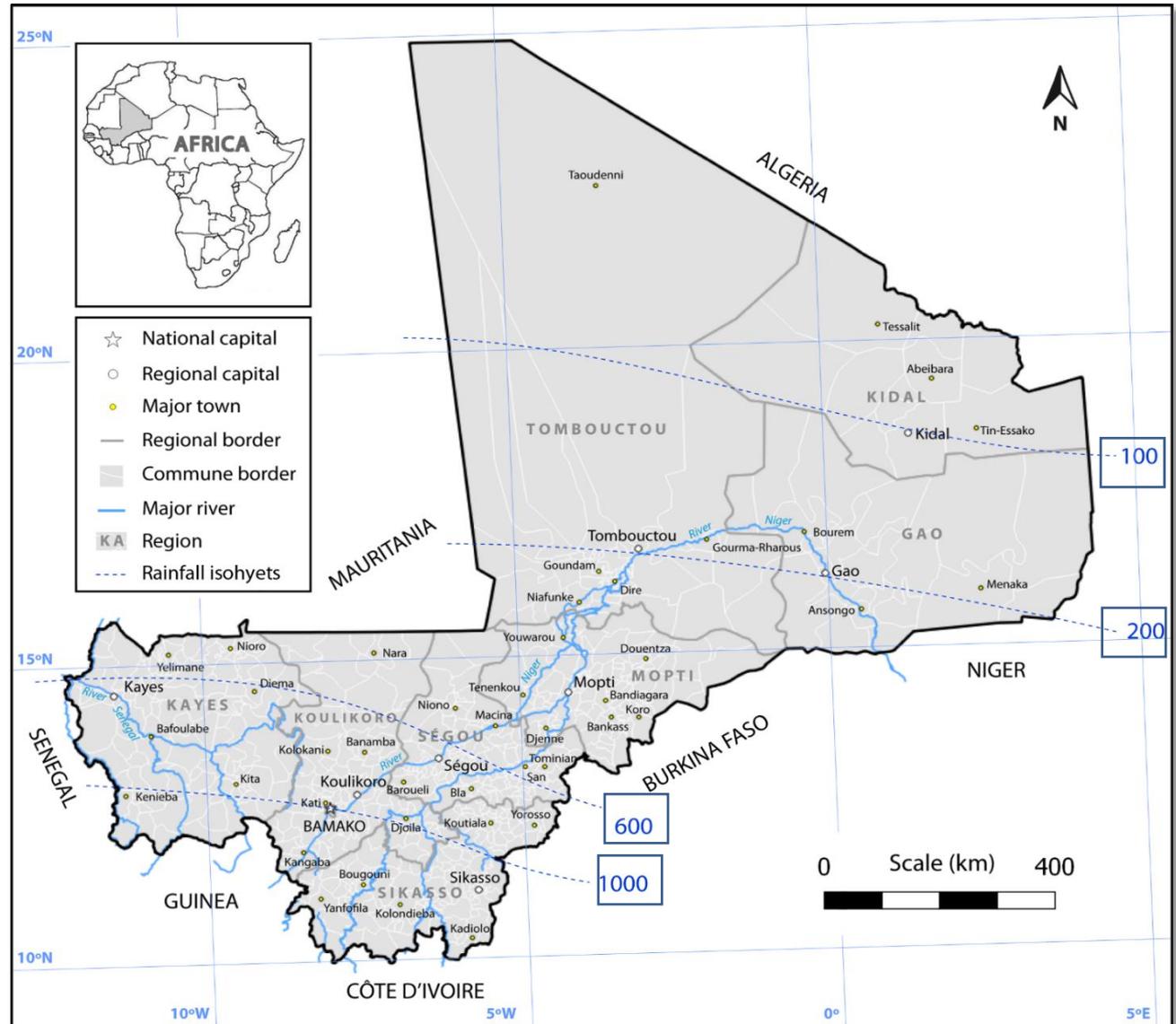


# Predicción del potencial hidrogeológico en Mali

En gran parte del país solo llueve entre junio y septiembre (monzón africano)

No hay más agua superficial permanente que los ríos Senegal y Níger (y algunos pequeños tributarios)

El agua subterránea juega un papel fundamental en el día a día de más del 80% de la población del país



# Problemática de acceso a agua potable

- Problema: perforar un pozo es muy caro.
- Está fuera del alcance de cualquier particular en zonas rurales (también fuera del alcance de muchas aldeas y municipalidades)
- Además, hay zonas donde la tasa de éxito de las perforaciones es inferior al 20%, en gran parte por falta de criterio hidrogeológico
- Esto implica que, en muchas ocasiones, la inversión resulta infructuosa
- Como consecuencia, existe una importante problemática de acceso al agua potable
- El objetivo de este trabajo es llevar a cabo una cartografía de zonas susceptibles de albergar agua subterránea en la cuenca del río Baoulé



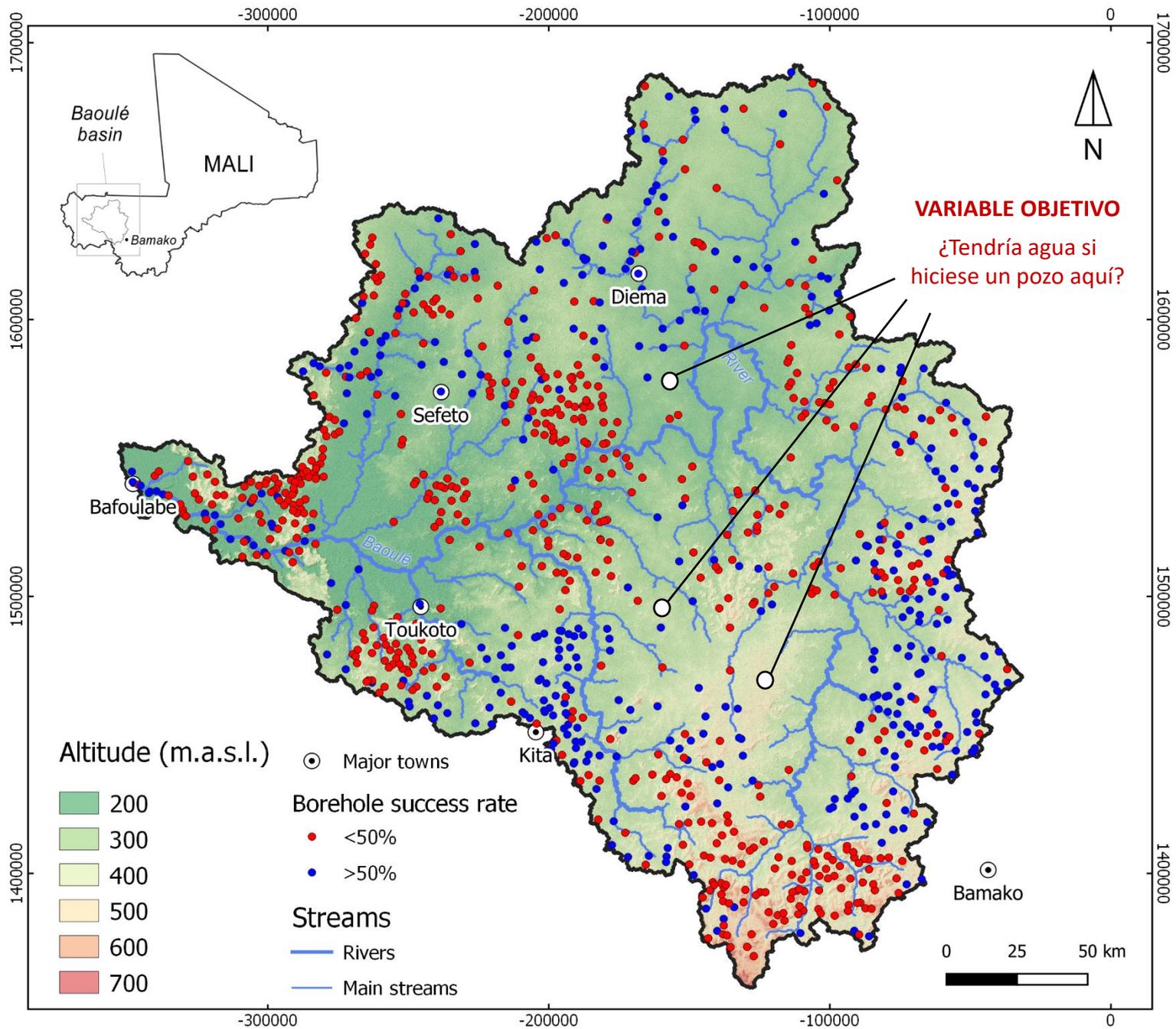
# ¿ Qué es un mapa de potencial hidrogeológico?

- Es un mapa que nos da idea de las zonas de un ámbito geográfico son “mejores” para perforar pozos de agua subterránea
- “Mejor” puede significar varias cosas: mayor caudal, mayor probabilidad de perforar con éxito, mejor calidad de las aguas, etc
- En este caso la cartografía se orienta a detectar las áreas donde existe una mayor probabilidad de éxito (el caudal que hace falta es relativamente pequeño y los posibles problemas de contaminación son fácilmente minimizables)

**IMPORTANTE:** Un mapa de potencial hidrogeológico no sustituye al trabajo de prospección hidrogeológica de terreno, pero permite optimizarlo

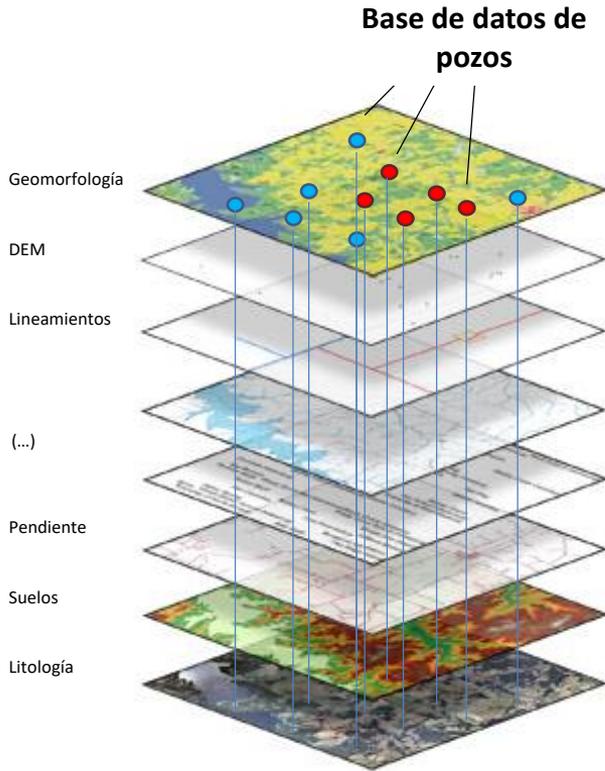
Descartar zonas de baja o nula factibilidad, optimizar recursos humanos, materiales y de tiempo, minimizar dificultades logísticas, minimizar riesgos

# Base de datos de pozos



# Modelo conceptual

**Variable objetivo:** probabilidad de éxito de una perforación



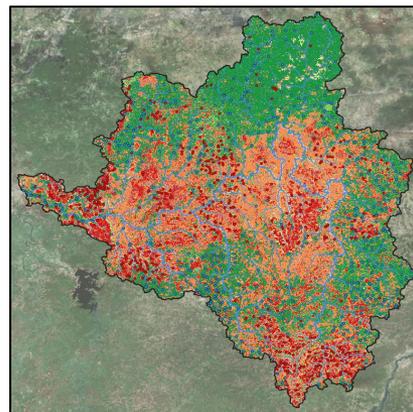
## Variables explicativas

**Presupuesto de partida:** “es posible inferir la presencia de agua subterránea a partir de variables de superficie”

Cada dato conocido (pozo +ivo o -ivo) se asocia al valor de ese pixel en cada una de las capas que representan las variables explicativas

ForagesEnhancedML :: Objetos totales: 1026, Filtrados: 1026, Seleccionados: 0

	CLULC100	CLandf100	CLDens10	CLDiet10	CINDVI100
1	1.0000000000	2.0000000000	1.0000000000	3.0000000000	1.6723310947
2	1.0510833168	2.0000000000	1.0000000000	2.0000000000	2.1402769089
3	2.0000000000	1.1608278751	1.0000000000	3.0000000000	2.4160318375
4	1.4322979450	2.0000000000	1.0000000000	3.0000000000	1.0660072565
5	1.6872192621	2.0280942917	1.0000000000	2.0000000000	4.6683654785
6	1.6666666269	2.0000000000	1.0000000000	4.0000000000	1.0921696424
7	1.0702085495	2.0000000000	2.0000000000	2.0000000000	2.5936987400
8	1.1130399704	2.0000000000	1.0000000000	1.1666666269	2.3297777176
9	1.0677404404	2.0000000000	1.0000000000	2.0000000000	2.2512276173
10	1.0000000000	2.0000000000	2.0000000000	1.3055554628	1.2050936222
11	1.0000000000	2.0000000000	1.0000000000	3.0000000000	3.6134500504
12	1.0032230616	2.0000000000	1.0000000000	4.0000000000	3.0409076214
13	1.0027145147	2.0000000000	2.0000000000	2.0000000000	2.6479818821
14	1.0000000000	2.0000000000	1.0000000000	2.0000000000	3.3593201637
15	1.0000000000	2.0000000000	1.0000000000	3.0000000000	3.2445805073
16	1.1666666269	2.0000000000	1.0000000000	2.0000000000	1.7105132341

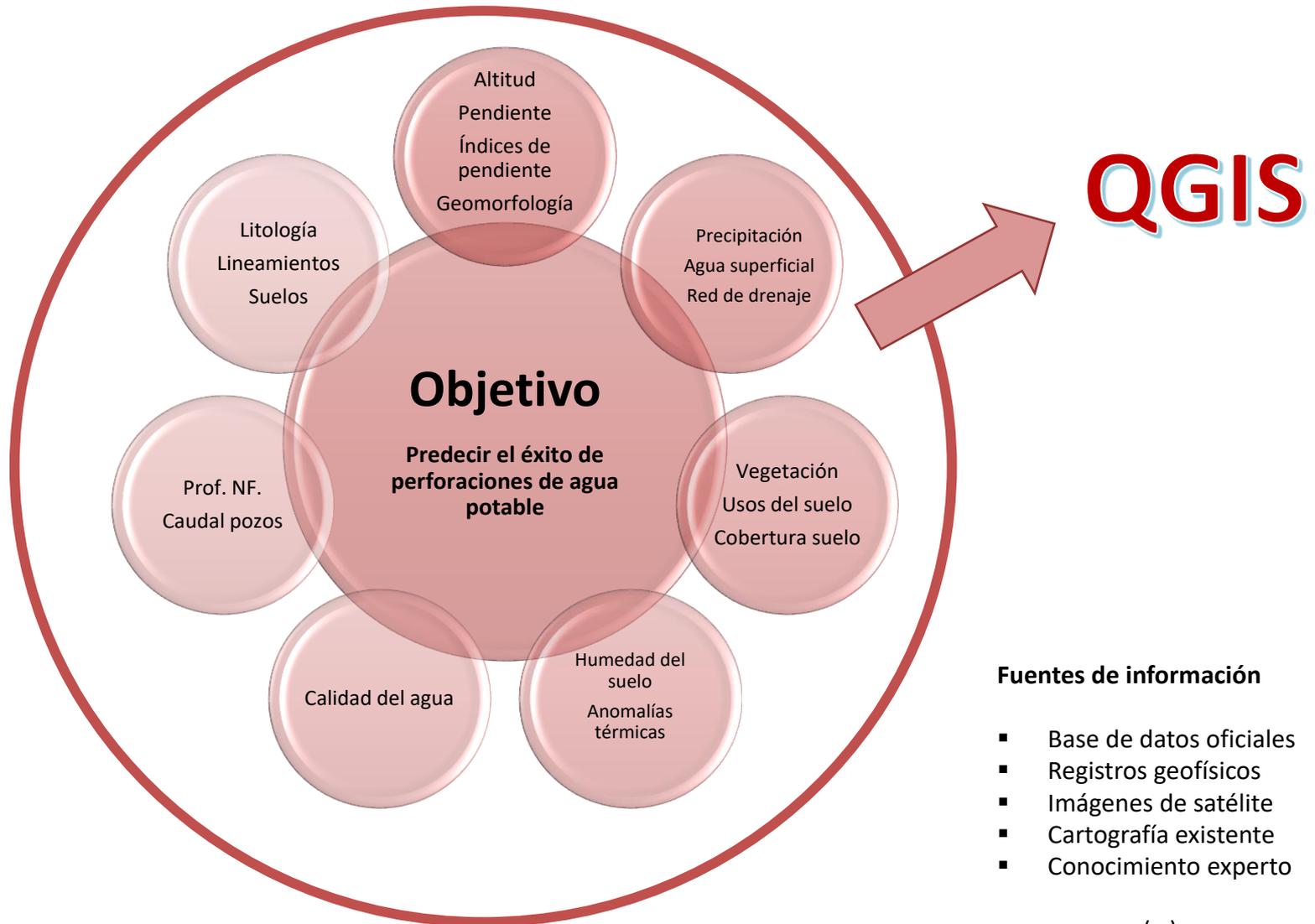


Linear vector support  
Support vector machine  
Multilayer perceptron  
K-Neighbours  
Naïve Bayes  
Logistic Regression  
Random Forest  
Decision Tree  
Linear Discriminant  
Quadratic Discriminant  
AdaBoost  
Gaussian process classifier  
Gradient boosting

Los algoritmos de *machine learning* deben encontrar los patrones que dan pozo positivo y pozo negativo. Para ello dividen la tabla en dos. Un porcentaje de los datos se usa para buscar los patrones que dan pozo positivo y pozo negativo (training set); mientras que el resto se usa para comprobar si los algoritmos “aciertan” en sus predicciones (test set)

Una vez conocemos estos patrones con una precisión satisfactoria (>90%), podemos extrapolar los resultados a cada pixel del GIS. Obtenemos así un mapa **validado** de potencial hidrogeológico

# Variables explicativas

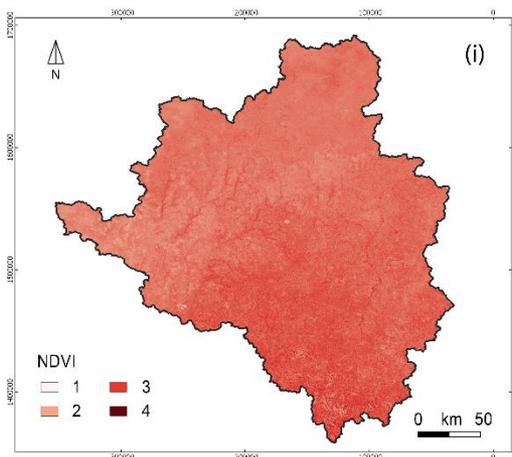
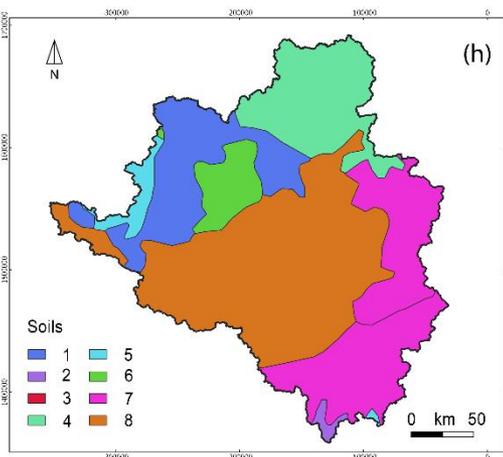
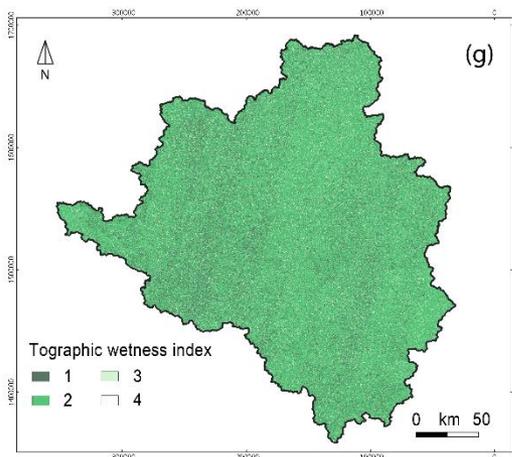
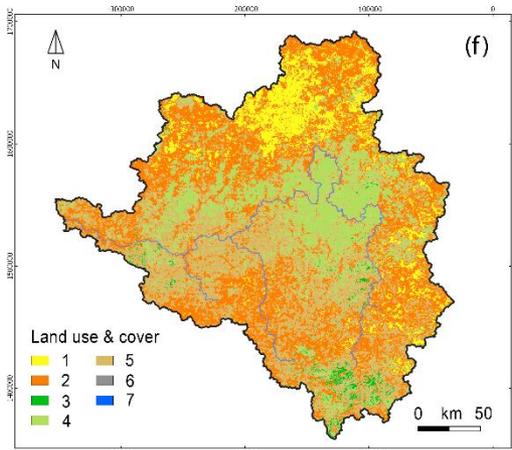
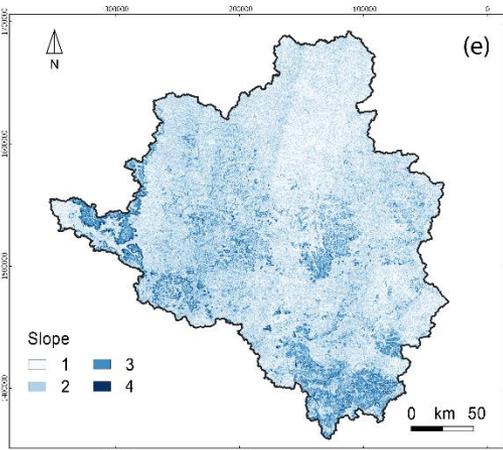
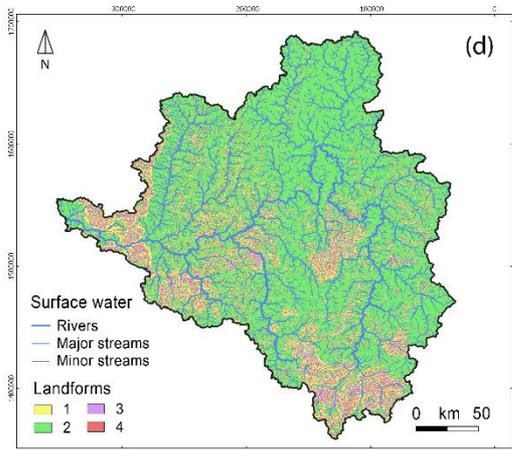
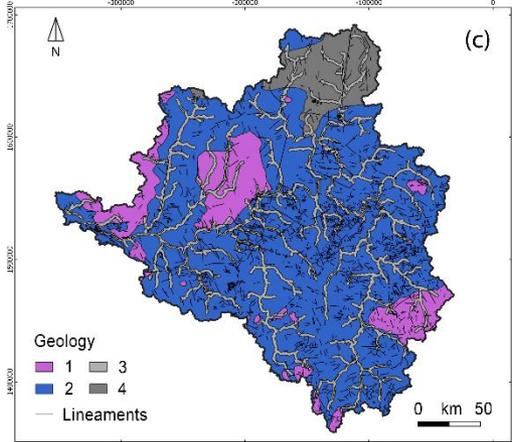
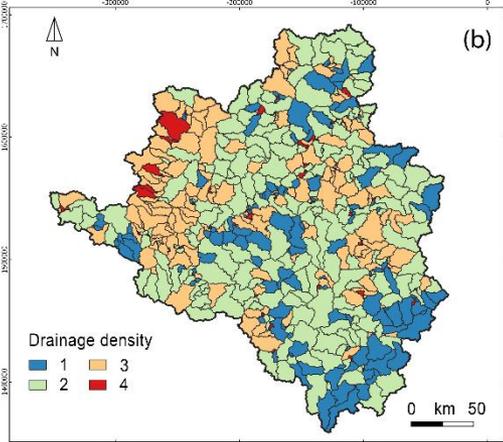
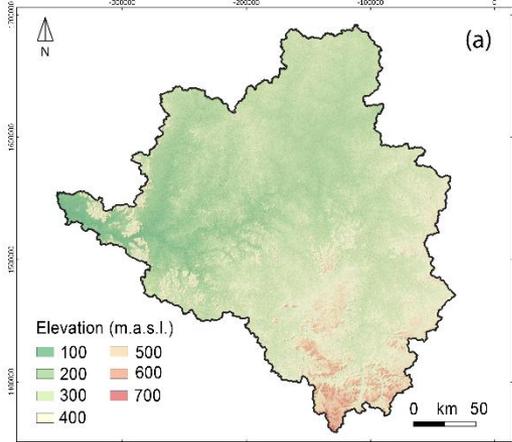


## Fuentes de información

- Base de datos oficiales
- Registros geofísicos
- Imágenes de satélite
- Cartografía existente
- Conocimiento experto

(...)

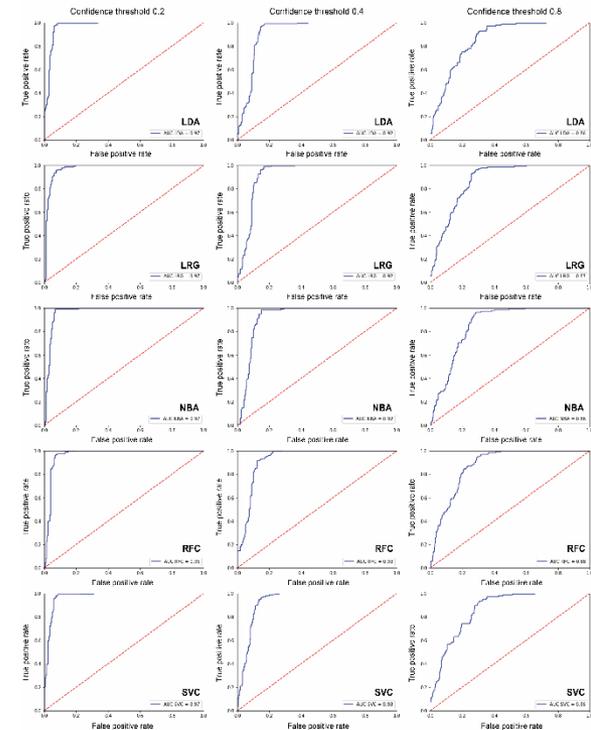
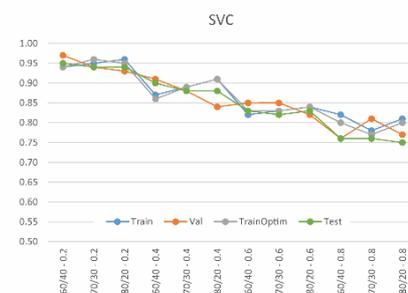
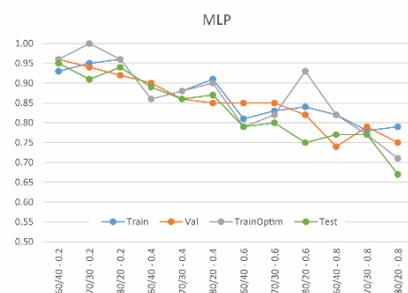
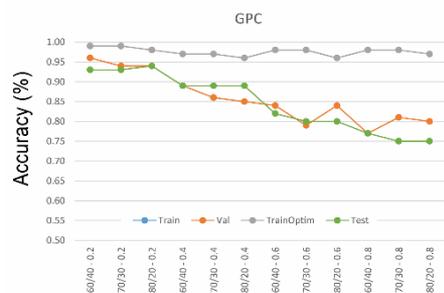
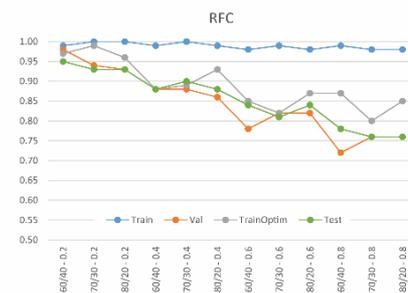
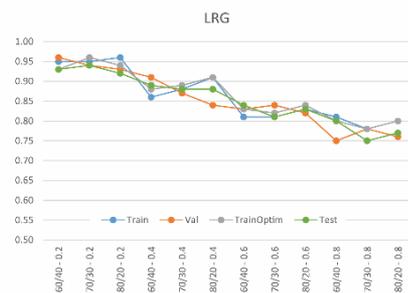
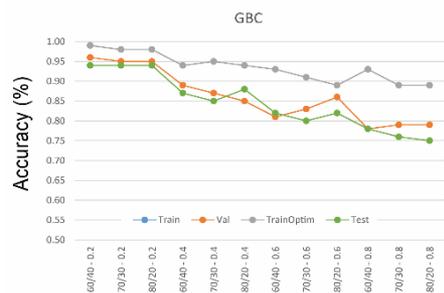
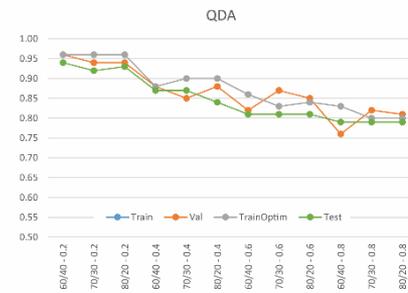
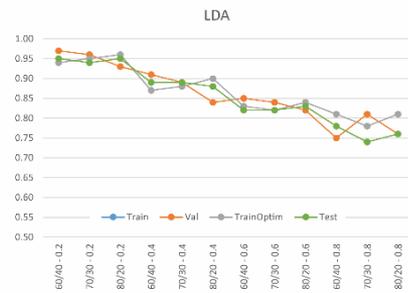
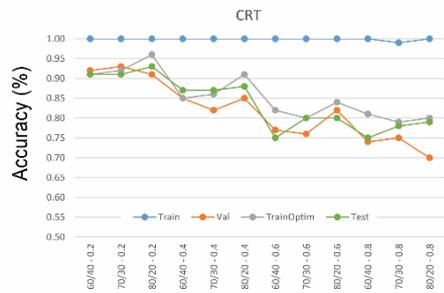
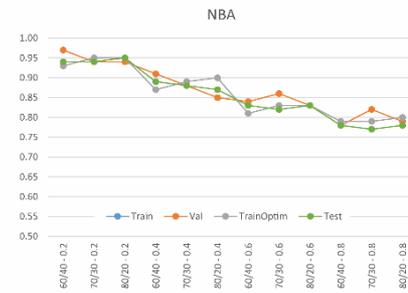
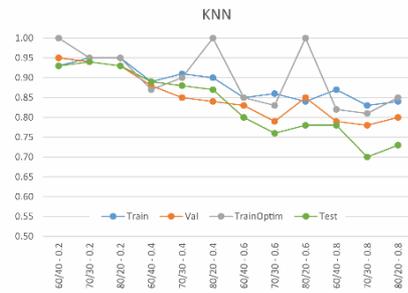
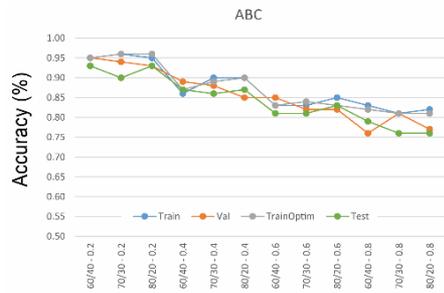
# Variables explicativas





# Métricas y resultados

Fuente: Martínez-Santos y Renard (2019)



**Eliminación recursiva de variables / optimización automática de variables / validación cruzada / "split" variable:**

Las variables más importantes son la geomorfología, la litología y los factores de pendiente

## Resumen

Test score > 0.9

AUC > 0.9

RFC, SVM y LRG son los más fiables

Split / confidence threshold

Split / confidence threshold

Split / confidence threshold



# Conclusiones

- Herramienta de **mapeo predictivo**, aplicable a cualquier variable espacialmente distribuida que pueda predecirse a partir de la combinación de otras variables espacialmente distribuidas
- Gratuita e integrada en **QGIS 3**
- El **resultado está validado** debido al procedimiento train/test implícito en las metodologías de machine learning
- No sustituye al trabajo de prospección hidrogeológica en terreno, pero **lo optimiza**
- Da **apoyo a la toma de decisiones**, permitiendo maximizar el impacto de las actuaciones cuando los recursos son limitados

## Referencias

Díaz-Alcaide S, Martínez-Santos P (2019). Review: Advances in groundwater potential mapping. Hydrogeology Journal 27(9)2307-2324

Martínez-Santos P, Renard P (2019). Mapping groundwater potential through an ensemble of big data methods. Groundwater, DOI: 10.1111/gwat.12939