

# Aranda. UN SISTEMA DE INFORMACIÓN Y DE AYUDA BASADO EN LA LÓGICA DIFUSA

---

**Resumen:** Se propone una metodología para el desarrollo de un sistema de consulta que permita a distintos tipos de usuarios obtener información sobre el cultivo de olivar y el soporte ambiental del mismo, en la provincia de Granada. El sistema de información se fundamenta en bases de datos de carácter difuso y es de naturaleza flexible, por lo que será capaz de ser consultado interactivamente. Los principales problemas que resuelve este modelo son el tratamiento de datos inciertos e imprecisos, como es el caso de la información ambiental y sobre cultivos, la representación de datos espaciales y puntuales y la fusión de datos sobre resultados de cultivo y otros de carácter científico-experimental.

## 1. INTRODUCCIÓN

La toma de decisiones en el ámbito agrícola afecta, tanto a los cultivos en todas sus fases (plantación, manejo, gestión de costes, etc.) como a la distribución y organización de los sistemas de producción. Para que las decisiones se tomen sobre una base correcta, el análisis de los sistemas de producción se debe abordar desde perspectivas comarcales y por grupos homogéneos de explotaciones, conociendo los factores de toda índole (físicos y socioeconómicos) que determinan la dinámica de esos sistemas agrarios y su comportamiento frente a las posibles modificaciones. La evaluación de Suelos basada en datos de carácter físico y socioeconómico, es por tanto una herramienta esencial para la toma de decisiones en agricultura, ya que valora en términos de aptitud a los diferentes sistemas agroecológicos presentes en un área, analizando para ello las relaciones entre las variables que condicionan dichos sistemas (Sys, 1991).

En la Evaluación de Suelos se plantean problemas como son el tratamiento de información imprecisa y el establecimiento de límites flexibles entre las clases de aptitud que pueden ser resueltos mediante la lógica difusa -fuzzy logic- (Groenemans & col., 1997; McBratney & Odeh, 1997). La lógica difusa, introducida en 1965 por L.A. Zadeh dentro de su Teoría de Subconjuntos Difusos (Zadeh, 1965), pretende manejar ese tipo de información imprecisa e incierta. Desde el momento en que fue presentada ha experimentado un fuerte auge debido a las aportaciones de muchos otros investigadores.

La metodología expuesta en este trabajo se ha desarrollado para llevar a cabo un proyecto cuyo objetivo es crear un sistema de información y de ayuda a la decisión para el cultivo de olivar en la provincia de Granada, tomando como base la utilización óptima de los recursos de acuerdo a sus potencialidades. Se pretende optimizar el binomio *Recursos* (caracterizados por información dispersa e imprecisa)/*Usuarios* (de carácter heterogéneo) mediante una planificación basada en los conceptos de *uso sostenible*. Se ha seleccionado el cultivo del olivar, porque puede decirse que es el cultivo con el que identifica a la agricultura andaluza, tiene una amplia representación en la provincia de Granada y se encuentra actualmente en expansión. Así mismo, se han empleado

Sistemas de Información Geográfica (SIG) como herramientas de modelización y análisis espacial imprescindibles en cualquier proyecto que necesite integrar información georeferenciada (Longley *et al.*, 1999).

## **2. PROPUESTA METODOLÓGICA**

### **2.1. Premisas y desarrollo del Sistema de Información y Ayuda a la Decisión**

Para desarrollar una metodología partimos de la premisa de que el óptimo de cultivo es un sistema con la máxima productividad compatible con el mantenimiento de la aptitud para el cultivo, es decir: optimización del binomio productividad/degradación. Para alcanzar esta optimización del cultivo del olivar hay que cumplir una serie de objetivos: a) máximo beneficio. b) productos de máxima calidad. c) mínimo coste, por lo que la creación de un olivar totalmente mecanizado es fundamental. d) ausencia de limitaciones importantes para el cultivo, para evitar la "marginalidad". e) emplear buenas prácticas agrícolas lo que se entiende como cuidados culturales necesarios sin un exceso en detrimento del beneficio y del aumento de los riesgos ambientales. f) usar conservando, se convierte en objetivo prioritario.

El proyecto comprende una serie de operaciones que se exponen en la [figura 1](#). La recogida de datos sobre los sistemas actuales y potenciales de cultivo del olivar puede ser bibliográfica o directa. Mediante la recopilación bibliográfica se obtienen reglas de conocimiento experto, por un proceso de extracción de conocimiento. Estas reglas nos permiten, con el establecimiento de un árbol de conocimiento, basado en la lógica difusa y aplicando un sistema experto de evaluación (ALES) elaborar una clasificación de aptitud informatizada basada en reglas difusas.

Los datos (bibliográficos y directos) sobre los sistemas de cultivo se almacenan en bases de datos de suelos y cultivo y en mapas digitalizados. De los datos directos sobre el cultivo de olivar, se deducen nuevas reglas de conocimiento que se incorporan a las anteriores y a partir de estas bases de datos se establecen, mediante la fusión del conocimiento sobre suelos y cultivos (más adelante se detalla), áreas homogéneas suelo-cultivo a las que denominaremos "unidades agroedáficas". Las unidades agroedáficas son evaluadas con la clasificación de aptitud elaborada, lo que nos permite clasificar la aptitud de los sistemas actuales y potenciales para el cultivo del olivar. Con el empleo de Sistemas de Información Geográfica (SIG) todos los resultados de evaluación y la propia información recogida en bases de datos y mapas digitalizados, puede ser expuesta espacialmente (cartografías temáticas). Finalmente, los procedimientos de consulta flexible permiten obtener información y ayuda a la decisión a partir de: a) las bases de datos, b) la aptitud de los sistemas de cultivo actuales y potenciales y c) las cartografías temáticas elaboradas. La posibilidad de introducción de nuevos datos, durante el proceso de consulta al Sistema de Información y Ayuda a la Decisión, hace que este sea interactivo con los usuarios del mismo.

La información necesaria para un Sistema de Información y Ayuda a la Decisión en el ámbito agrícola, es de suelos (en un sentido amplio) y de cultivos (formas de uso del suelo). La información de suelos comprende datos de carácter edáfico (suelo considerado en el sentido estricto de "individuo suelo"), climático, topográfico, geológico, etc., y puede ser recopilada de cartografías temáticas, bases de datos, modelos digitales del terreno, etc. Esta información se representa espacialmente como unidades cartográficas de suelos, áreas climáticas de igual temperatura o precipitación, zonas isoclinas, unidades litoestratigráficas, etc. La información sobre cultivos es de carácter agronómico (variedades de plantas cultivadas, modalidades de riego, abonos utilizados, etc.), y socioeconómico. Los datos sobre cultivo se recopilarán esencialmente con una encuesta al empresario agrícola y a través de diversos informes y anuarios. La encuesta comprende tres apartados: A) Localización y Datos Generales, B) Manejo del Cultivo y C) Datos del Suelo. La representación espacial de la información sobre cultivo tiene en el caso de la encuesta una unidad que es la parcela o propiedad del empresario agrícola, mientras que otros datos se representan a nivel de comarca, de municipio o incluso en el total de la provincia.

En el caso de la encuesta al empresario agrícola, la primera fuente de incertidumbre deriva de que sólo son encuestados un pequeño porcentaje de los olivareros granadinos; Rodríguez et al. (1998) refieren que el porcentaje mínimo de encuestas es función de la variabilidad de los sistemas agrícolas estudiados, si bien el conocimiento de esa variabilidad exige de encuestas; nos encontramos ante un círculo vicioso. La mayoría de los encuestados tienen su propiedad muy fragmentada; la encuesta se realiza sobre la parcela más emblemática por lo que puede existir una distorsión en cuanto al tamaño de la explotación. La encuesta en algunos de sus ítems puede generar incertidumbre e imprecisión por problemas de lenguaje porque algunos términos no son comunes para todos los posibles encuestados. Los costes de explotación son en muchas de las actividades agrícolas, difíciles de calcular, sobre todo cuando estas tareas son realizadas por el propio empresario y su familia. Los olivareros dudan también sobre los precios de abonos y productos fitosanitarios. Otra cuestión añadida es la tendencia a enmascarar por razones socioculturales, las prácticas agrícolas realmente desarrolladas y sus costes. En los ítems de la encuesta relacionados con el suelo, hay que partir del hecho de que el empresario agrícola tiene un concepto del suelo muy diferente al del edafólogo.

En las bases de datos de suelos se almacenan dos tipos esenciales de atributos en relación con su obtención: bibliográficos y experimentales, y dentro de estos últimos, morfológicos y analíticos. En la figura 2, se relacionan los tipos de datos con los factores y la incertidumbre e imprecisión de los mismos. Además de lo expuesto en la fig. 2, debemos resaltar que buena parte de la información sobre los suelos tiene carácter cualitativo, por lo que está sometida a la subjetividad del investigador y su tratamiento matemático es difícil (Webster, 1977; Webster & Oliver, 1990).

También se debe insistir en la dificultad que supone el análisis de un objeto, el suelo, cuya variabilidad se produce en un *continuum*, lo que plantea el uso de nuevas metodologías para su estudio y clasificación basados en la lógica difusa y los conjuntos

difusos (*McBratney & Gruijter*, 1992).

La fusión de datos a nivel espacial se ha resuelto estableciendo en el territorio granadino lo que se han denominado *Unidades Agroedáficas* (UA). La UA es un concepto abstracto definido por tipologías de suelo y otros parámetros de carácter físico (pendiente y temperatura), que influyen decisivamente en el cultivo del olivar. las UAs se establecen por superposición de tres cartografías diferentes: suelos, temperatura y pendiente. Se pretende con esta división del paisaje, establecer áreas de tamaño adecuado, para, por un lado disminuir el grado de variabilidad de las condiciones para este cultivo en la provincia, y por otro lado permitir que el número de datos por cada tipo de unidad sea suficiente. Las bases de datos de suelos y las correspondientes a las encuestas se asociarán espacialmente dentro de las UAs.

Como mapa de suelos se empleará el de la Provincia de Granada a escala 1:200.000 de *Pérez-Pujalte* (1980). Las unidades cartográficas de suelos de este mapa están establecidas a pequeña escala. Es un mapa de base geológico-fisiográfica, que supone agrupamiento de tipologías de suelo clasificatoriamente distintas pero con importantes semejanzas en morfología y propiedades.

El Mapa de temperatura se establecerá mediante correlación  $T^a$ /altitud, que en nuestra provincia muestra coeficientes muy elevados. La limitación para el olivo de  $-7^{\circ}\text{C}$  de la media de las mínimas absolutas del mes más frío, y de  $13^{\circ}\text{C}$  en la media anual, de acuerdo a las correlaciones  $T^a$ /altitud para la provincia de Granada, nos marca los umbrales siguientes: menos de 1250m, mala aptitud; de 1250m a 800m, aptitud regular; de 800m a 275m, óptimo y menos de 275m, aptitud regular. El límite de 275m es dudoso si pensamos en todos los olivares de Córdoba y de Sevilla, quizá se trate de un problema de variedad de planta.

La pendiente, o inclinación de la ladera, condiciona aspectos del cultivo como son: manejo del suelo (riego y mecanización, principalmente) y riesgos de degradación (erosión hídrica, principalmente). Empleando los límites establecidos por la FAO para la pendiente, podemos establecer cuatro clases de limitación, que serán las cuatro unidades cartográficas de pendiente establecidas en el mapa: menos de 6% de pendiente, sin limitaciones; de 6 a 13%, moderadas limitaciones; de 13 a 25%, severas limitaciones y menos de 25%, muy severas limitaciones.

La UA se puede considerar una "unidad de información", es decir, no puede ser definida por sí misma, sino por sus atributos. En principio la UA no se ha orientado como unidad "natural" de paisaje sino como una división que mezcla la unidad cartográfica de suelos basada en una clasificación natural de suelos (FAO) con unidades artificiales de paisaje establecidas con criterios utilitarios (riesgos de helada, periodo de latencia, posibilidades de mecanización, etc). No obstante, estos criterios utilitarios al estar expresados como umbrales de altitud o de pendiente, suponen también variabilidad natural y por tanto, la UA es una división del paisaje que en buena medida tiene un carácter natural; no sería exactamente una "unidad de tipos de tierras" tal y como la define la FAO (1974).

En la figura 3 se expresa gráficamente cómo se interpreta la UA. Una vez establecida la UA, se vuelca en ella la información agronómica (encuesta al empresario agrícola), de suelos y de otras bases de datos (p.e. climáticas). La UA puede ser reinterpretada desde el punto de vista del cultivo del olivar (como ya hemos indicado, encierra información utilitaria) con ayuda de la clasificación de aptitud elaborada. La aptitud de estas UAs puede ser utilizada para la ayuda a la decisión, con carácter geográfico, mediante los sistemas de consulta flexible.

Se ha empleado ARC/INFO v. 7.02 para Windows NT para la gestión de la información cartográfica, dada su capacidad de integración con sistemas de bases de datos externos al sistema, tales como Oracle, así como con software especial para la construcción de modelos de evaluación de tierras como el sistema ALES. La información cartográfica de partida ha sido incorporada al sistema mediante digitalización de los correspondientes mapas o bien ha sido adquirida directamente en formatos digitales compatibles con ARC/INFO. Esta información se ha modelizado en unos casos directamente como coberturas vectoriales, por ejemplo los mapas de suelos y de usos del suelo y en otros casos como coberturas o modelos raster en formato grid o lattice de ARC/INFO, como es el caso del modelo digital del terreno con un tamaño de pixel de 20 metros. La información climática se ha modelizado inicialmente sobre un modelo raster construido por interpolación espacial a partir del modelo digital del terreno empleando ecuaciones de correlación lineal múltiple obtenidas a partir de 40 estaciones meteorológicas de la provincia de Granada y zonas colindantes. Se construyeron diferentes modelos raster para los distintos parámetros climáticos modelizados. A partir de dichos modelos climáticos se generaron coberturas vectoriales correspondientes a los límites de parámetros climáticos empleados en la generación de las distintas cartografías.

A partir del modelo digital del terreno se generaron mapas de pendientes en forma de coberturas vectoriales empleando los márgenes de pendientes anteriormente indicados.

Las bases de datos asociadas a las coberturas vectoriales se han construido y gestionado en "Oracle" conectándose a las coberturas a través de las correspondientes tablas de atributos (Feature attribute tables).

Una vez integrada toda la información en el GIS, se emplearon herramientas de análisis espacial tales como análisis topológico (spatial join, proximities, buffering, etc.) de coberturas vectoriales junto con operaciones lógicas sobre las tablas de atributos para la obtención de resultados cartográficos como los mapas de unidades agroedáficas, mapas de evaluación de tierras, etc.

#### **2.4.- La consulta flexible como base para la información y la ayuda a la decisión**

Un sistema típico de gestión de bases de datos proporciona procedimientos para el almacenamiento, acceso y modificación de la información. En nuestro caso, nos vamos a centrar en los métodos de acceso a esa información. Mediante los mismos, un usuario va a poder obtener información contenida en la base de datos (BD), debidamente

formateada para su comprensión. Incluso le será posible especificar los criterios en base a los cuales seleccionar la información; o exigir que sólo se nos devolviera aquella información que cumpliera la condición con un cierto grado (*grado de cumplimiento*).

Cuando, además, parte de la información con la que se trabaja está dotada de un carácter difuso, es necesario proveer de métodos que permitan un eficaz acceso a esa información. Una manera de lograr esto es flexibilizar la consulta. La idea es dotar al propio lenguaje de consulta a la BD de la capacidad de manejar información imprecisa, de expresar datos difusos mediante términos propios del lenguaje natural.

Por ejemplo, una posible consulta típica que se podría realizar al sistema de BD, "*Dime qué fincas tienen una producción anual superior a los 3000 Kg.*", debidamente traducida al lenguaje de consulta a la BD, quedaría de la siguiente forma: "*SELECT nombre\_finca, produccion\_anual FROM finca WHERE produccion\_anual >= 3000*". Supongamos ahora que nos interesara conocer aquellas fincas cuya producción anual fuera "alta". La correspondiente consulta flexible (o difusa) que permitiera esto debería tener la siguiente estructura: "*SELECT nombre\_finca, produccion\_anual FROM finca WHERE produccion\_anual = ALTA*". El término "alta", tan cotidiano para nosotros, conlleva una gran imprecisión. ¿Qué cantidad de aceituna debería producir una finca exactamente para que su producción fuera considerada "alta"? Internamente, el sistema de BD podría modelar esta característica de distintas formas, pero ese tratamiento debería quedar oculto al usuario, al que sólo le interesa saber qué fincas obtienen una mayor producción. Nótese además que esta consulta flexible puede realizarse tanto sobre atributos con un marcado carácter difuso, como sobre atributos cuyo valor está perfectamente definido.

Distintos modelos sobre BD han abordado el problema de la consulta flexible. El modelo GEFRED (*Medina et al., 1994*), desarrollado por nuestro grupo de trabajo, es el más general conocido, ya que maneja imprecisión e incertidumbre conjuntamente. En posteriores trabajos, que comentamos a continuación, se han llevado a la práctica las ideas presentadas en este modelo.

Mediante este primer módulo, se le facilita al usuario la introducción de datos, imprecisos o no. Consta de dos partes bien diferenciadas, tanto en su construcción como en su funcionalidad. La parte visible es una serie de formularios accesible a través de Internet mediante cualquier navegador HTML. Por debajo de ella, tenemos una aplicación para el procesado de las entradas y la comunicación con la base de datos.

El módulo es lo suficientemente general como para permitir adaptarlo a cualquier conjunto de tablas que se deseen manejar. En nuestro caso, tenemos implementadas dos variantes del mismo, una para atender a las tablas con las encuestas y una segunda para las tablas sobre suelos.

Otra funcionalidad añadida es la de realizar consultas simples y alteraciones sobre los datos siguiendo esta misma interfaz. El objetivo de esto es el de poder modificar sobre la marcha datos ya almacenados. Para otros tipos de consulta, p.e., consultas flexibles,

tenemos la siguiente herramienta.

Este módulo se encarga de presentar al usuario la información previamente introducida en la base de datos, de acuerdo a una consulta flexible especificada por el mismo. Se divide también en dos partes bien diferenciadas.

El módulo principal es el programa FQ (Fuzzy Queries), desarrollado por nuestro grupo de trabajo. En realidad, este módulo constituye una aplicación independiente por sí mismo, aunque nosotros lo tratemos como una parte del todo. Sobre FQ hemos programado un módulo adicional, FQBuilder, para facilitar la edición y construcción de las consultas flexibles.

Mediante FQBuilder, podemos consultar el Catálogo del Sistema, y dentro del él, la Base de Metaconocimiento Difuso (FMB). En concreto, nos es posible visualizar las correspondencias etiqueta lingüística/distribución trapezoidal que pudiera haber definidas sobre atributos con carácter difuso. El usuario puede también solicitar una descripción sobre los atributos contenidos en una determinada tabla, cuando el elevado número de éstos llegue a confundir al usuario.

Una vez construida la sentencia, ésta será enviada a FQ. FQ llevará a cabo la comunicación con la base de datos para comprobar la correcta sintaxis de la sentencia antes de ejecutarla. FQ es también quien se encarga de mostrar por pantalla los resultados de la ejecución de la consulta.

Como planteamientos futuros tenemos, en primer lugar, integrar en un solo módulo FQ y FQBuilder. Otro objetivo a cumplir es el de aumentar la riqueza de la respuesta a la consulta flexible, añadiendo medidas difusas de agregación y resumen de la información contenida en la base de datos.