

# Plataforma Escalable para el Almacenamiento y Procesamiento de Imágenes Multi e Hiperespectrales con Propiedades de Acceso Espacio-Temporal

Miguel Sánchez<sup>1</sup>, Manuel Barrena<sup>1</sup>, Aurora Cuartero<sup>2</sup>

<sup>1</sup> Dpto. Ingeniería de Sistemas Informáticos y Telemáticos

<sup>2</sup> Dpto. Expresión Gráfica

Escuela Politécnica, Universidad de Extremadura, C/Avda. de la Universidad, s/n. 10003  
Cáceres, España

{ mscabrera, barrena, acuartero }@unex.es

**Resumen.** Este artículo describe el desarrollo de una plataforma escalable y eficiente para el almacenamiento y gestión masiva de imágenes multi e hiperespectrales a diferente escala; imágenes de satélite a gran altura e imágenes a media y baja altura procedentes de vuelos con vehículos aéreos no tripulados, con propiedades de acceso espacio-temporal, que sea capaz de ofrecer una mejora en los procesos de gestión y productividad de explotaciones agrarias. Esta información se combina con el uso de tecnologías actuales como bases de datos NoSQL para la gestión del almacenamiento, así como el diseño de los modelos de datos necesarios que sean capaces de soportar la variabilidad de la característica espacio-tiempo de los datos citados, y de aquellas fuentes externas que necesitan ser analizadas y procesadas antes de incorporarse al sistema. Ofreciendo mediante esta integración de datos y tecnología una infraestructura digital que logre facilitar, dentro del ámbito agrícola, los procesos de toma de decisión y la optimización de la gestión de las zonas de cultivo.

**Keywords:** Agricultura de Precisión, Datos Espaciales, Imágenes Multiespectrales, Imágenes Hiperespectrales, Georreferenciación.

## 1 Introducción

El constante proceso de investigación en la mejora de la eficiencia productiva dentro del ámbito de las explotaciones agrarias, así como de la agricultura en general, se ha visto incrementado en los últimos años debido al gran aumento y consolidación que han adquirido los paradigmas como IoT, Big Data, Cloud Computing y todas aquellas tecnologías de la información y las comunicaciones que, siendo aplicadas de forma correcta en este entorno, definen el término “Agricultura de Precisión”, cuyo propósito general es conseguir una mejora en la gestión y la productividad del entorno agrícola a partir de la implantación y el uso de las TIC [1,2,6].

En las últimas décadas se ha buscado incesantemente aumentar esta eficiencia productiva mediante el uso de muy diversas técnicas. El notable y evidente interés por

conseguir un proceso de mejora en este ámbito se ha incrementado con el rápido crecimiento de la población mundial, la demanda de alimentos y un control más exhaustivo de la economía, entre otros, intentando además combinar estos factores con el continuo avance de las tecnologías.

En el entorno de la agricultura de precisión, los medios utilizados para la gestión de la actividad agrícola incluyen métodos de sensorización local [3] (sensores físicos de medida instalados en las propias zonas de cultivo, estaciones meteorológicas, etc.), sensorización remota o semirremota (imágenes de satélite e imágenes de vuelos a gran y baja altura) y fuentes complementarias, volcado de datos a distancia, gestión de información vía web o utilización de dispositivos como instrumentos de entrada y salida de datos. El uso combinado de estos elementos proporciona al usuario del ámbito agrícola la oportunidad de mejorar y administrar de forma eficiente la gran cantidad de información generada y existente desde un punto de vista agronómico, económico y medioambiental [1,3].

En este artículo se describe la arquitectura software de un sistema de información basado en el desarrollo de una plataforma escalable y eficiente que ofrece soporte tecnológico necesario para agilizar los procesos de análisis y toma de decisiones en la gestión de explotaciones agrarias. La plataforma facilita el almacenamiento y gestión masiva de datos multi e hiperespectrales provenientes de sensores remotos de alta, media y baja altura y proporciona herramientas de acceso espacio-temporal a los datos de interés para su posterior análisis [1,6].

Las imágenes hiperespectrales se presentan como una herramienta ideal para profundizar en el estudio del patrón espectral de la vegetación, ya que permiten el estudio y análisis de parámetros tanto del entorno físico, humedad del suelo y ambiental, como del estado de la planta, desarrollo vegetal, estado de maduración del fruto, coloración foliar, presencia de hongos [4,5].

En la actualidad, la gran mayoría de las aplicaciones utilizadas en las técnicas de teledetección de datos de sensores multi e hiperespectrales sobre entornos agrícolas parten de imágenes adquiridas exclusivamente desde plataformas espaciales [1,2]. Para solucionar estos inconvenientes, el sistema complementa las fuentes citadas con información obtenida a partir de sensores remotos o semirremotos sobre vehículos aéreos no tripulados y autónomos, explotando así el auge y desarrollo de estos dispositivos, y aprovechando su capacidad de transportar diferentes tipos de sensores (humedad, temperatura, etc.).

Además, desde el punto de vista computacional, disponer de imágenes tomadas desde distintas alturas, ángulos, y en distintos instantes de tiempo, supone una formidable fuente de información para facilitar las tareas de análisis, tanto descriptivo como con capacidades predictivas [4,5]. Se trata de proporcionar soluciones de bajo coste y simples de utilizar, a través de un sistema que, mediante la adecuada combinación de tecnologías actuales y fusión de los datos de sensores a distintas alturas con distintas resoluciones, sea capaz de ofrecer la información suficiente para facilitar los procesos de monitorización exhaustiva de explotaciones agrícolas, mejorando el proceso de toma de decisiones en la optimización de prácticas agrarias.

Para describir la arquitectura del sistema y los procesos más importantes, el presente artículo se estructura en tres secciones. En la primera se lleva a cabo una descripción

de los componentes principales del sistema. Una vez descrita la arquitectura, la sección segunda aporta información sobre los procesos de extracción de información de las fuentes externas y su necesaria georreferenciación. Para terminar, la última sección presenta una visión general de la plataforma y las funcionalidades ofrecidas al usuario final en el ámbito agrícola.

## 2 Descripción del Sistema

El sistema combina la integración de varias tecnologías para crear una arquitectura software capaz de hacer frente a los objetivos y funcionalidades citados en el apartado anterior, cuya estructura general del sistema es la siguiente:

- Datos espectrales: Elementos provenientes de fuentes externas, tales como imágenes de satélite a gran altura e imágenes a media y baja altura procedentes de vuelos con vehículos aéreos no tripulados que nutren de información al sistema.
- Almacenamiento: Base de datos NoSQL encargada de almacenar los datos del sistema, metadatos extraídos de las imágenes. Sistema de archivos HDFS para el almacenamiento de las imágenes.
- Extracción de metadatos: Proceso encargado de extraer la información de cada una de las diversas fuentes externas y generar un esquema de datos homogéneo y su posterior acople eficaz al sistema.
- Georreferenciación de datos: Proceso encargado de realizar la georreferenciación de aquellos datos obtenidos que tienen ausencia de posicionamiento espacial en un sistema de coordenadas geográficas o cuyo posicionamiento no se ha realizado con la precisión exigible.
- Aplicación web del sistema: Interfaz visual de la plataforma que permite de modo sencillo e intuitivo la obtención de datos del sistema mediante la inserción de consultas atendiendo a la característica espacio-temporal de los datos, ofreciendo al usuario la visualización de los mismos.

### 2.1 Datos Espectrales

Como se ha explicado en el punto anterior, la plataforma se nutre y proporciona información sobre un conjunto de datos compuesto por imágenes multispectrales e imágenes hiperespectrales. Es importante incidir en por qué utilizar ambos tipos, ya que, a la hora de escoger las tecnologías para la captación de datos en agricultura de precisión, surge la pregunta de qué sistema resulta más adecuado para ello: teledetección mediante satélites o mediante vehículos aéreos no tripulados. Dependiendo del uso de la información que se vaya a realizar, cada uno de estos métodos tiene ventajas e inconvenientes, lo cual implica el grado de adecuación al sistema.

La teledetección es la técnica de adquisición de información de una superficie a pequeña o gran escala, de forma aérea e incluye todo el trabajo posterior de tratamiento de estas imágenes, así como su procesado e interpretación. Los vehículos aéreos no tripulados son capaces de portar diferentes tipos de sensores de medición (óptica, LIDAR, térmica, etc.), aportando múltiples opciones en cuanto al tipo y necesidades de

información se refiere. Los sistemas espaciales pueden incluir sensores ópticos y multiespectrales, existiendo también la posibilidad de sensores térmicos y de radar, pero a diferencia del caso anterior, su resolución no es la adecuada para la aplicación en la agricultura de precisión de un modo tan eficaz.

Por este motivo, existen varios puntos clave entre ambas metodologías, imágenes multiespectrales proporcionadas por satélite e imágenes hiperespectrales proporcionadas mediante vuelos con vehículos aéreos no tripulados, tales como:

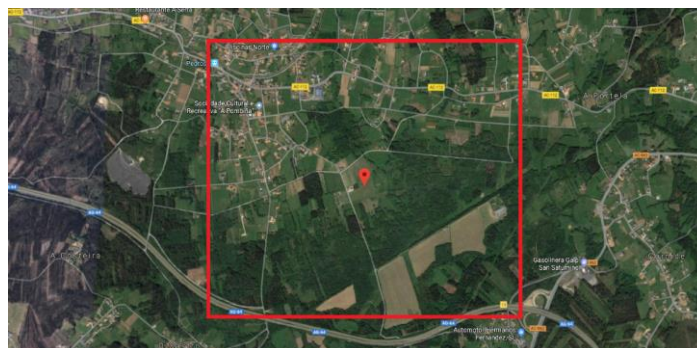
- **Calidad de la imagen:** Ambas tecnologías permiten recoger información muy valiosa, pero existen particularidades que condicionan que un sistema u otro sea más adecuado a la hora de conseguir información que sirva de utilidad aplicable en agricultura de precisión. Las imágenes de satélite son calibradas mediante correcciones geométricas y atmosféricas [10,11], pudiendo ocasionar en estas últimas, distorsiones entre imágenes adquiridas en tiempos diferentes debido a la concentración de aerosoles presentes en la atmósfera. Además, estos sensores, al tener una elevada resolución espacial pueden ocasionar interferencia en la precisión de la toma de medidas debido a la reflexión del suelo o de las malas hierbas colindantes. Podemos decir que una imagen satelital muestra de forma precisa las diferencias por comparación, pero en el aspecto radiométrico, la característica temporal de las mismas no suele ser comparable. Por el contrario, la calibración radiométrica mediante UAV (unmanned aerial vehicle) es más sencilla, ya que no necesita calibración atmosférica, permitiendo una elevada precisión y comparación temporal de imágenes, además de un análisis de la vegetación mucho más correcto gracias a su elevada resolución espacial.
- **Nivel de resolución y el tamaño del píxel:** El satélite proporciona imágenes con píxeles de gran tamaño que funcionan correctamente a nivel comparativo de zonas en una misma captura. En cambio, el UAV, con un procesamiento adecuado, permite un conjunto de píxeles de vegetación de tamaño pequeño y radiométricamente preciso que puede ser aplicado en cualquier cultivo [5,6].
- **Meteorología:** Los satélites toman sus datos por encima de la atmósfera, mientras que los vehículos aéreos no tripulados realizan sus vuelos por debajo de las nubes. Esta diferencia es clara e importante, ya que la posición repercute en la toma de datos si los factores climatológicos no son buenos en el caso del satélite, pudiendo incluso tener que posponer la fecha de adquisición de datos. Por el contrario, con el uso de UAV, al realizar sus vuelos por debajo de las nubes, este factor es mucho menos probable.

Por lo tanto, ambos sistemas pueden ser aplicados en la captación de datos sobre cultivos en agricultura de precisión. La diferencia entre utilizar la teledetección por satélite o mediante UAV, o los dos complementariamente, dependería del grado de precisión de la información que el usuario, dentro del sector agrícola, exija y la importancia o necesidad que implique aplicar las ventajas e inconvenientes citados anteriormente. Combinando ambas tecnologías de obtención de información de la superficie, garantizamos un sistema completo que, además, de forma general, contrarreste la falta actual de información de la mayoría de plataformas que únicamente se centran en un tipo concreto de datos.

Actualmente existen varias opciones en cuanto a plataformas dedicadas a ofrecer datos multispectrales obtenidos mediante teledetección satelital [7,8], pero no todos ellos proporcionan la información de forma gratuita, ni disponen de una cobertura máxima del planeta, o no mantienen una actualización continua y precisa de sus datos. Sistemas satelitales como Earth Observing-1 y OrbView-3, proporcionan datos espectrales, pero no sobre un recubrimiento total del planeta. Por el contrario, si comparamos dicha cobertura con la ofrecida por los satélites Landsat 8 y Sentinel-2, comprobamos que estos sí ofrecen un porcentaje de recubrimiento de la superficie de casi la totalidad del terreno del planeta [7]. Es por ello, y debido a la gran cantidad de información actualizada que tanto Landsat 8 como Sentinel-2 ponen a disposición de cualquier usuario, de forma libre, los datos multispectrales que va a almacenar y gestionar la plataforma de desarrollo en cuanto a la técnica de teledetección por satélite, se corresponde a estos dos sistemas citados.

En cuanto a los datos espectrales obtenidos a media y baja altura mediante teledetección con UAV, actualmente y a diferencia de lo expuesto anteriormente con los datos multispectrales, hay una diferencia muy elevada en cuanto a la existencia de plataformas que gestionen y proporcionen datos hiperespectrales [6,7] del mismo modo que sí lo hacen con datos multispectrales y que tengan una disponibilidad y un tipo de información que pueda ayudarnos en tareas relacionadas con el estudio, análisis y mejora de la agricultura de precisión, como es el caso. La mayoría de las plataformas y sistemas existentes en este ámbito se centran en proporcionar al usuario servicios de pago para el procesado sus datos espaciales, generación de mapas, ortomosaicos, modelos digitales de elevación, etc., pero no proporcionan un conjunto de datos espaciales obtenidos mediante teledetección por UAV o con el uso de drones.

La plataforma que definimos en el presente artículo, suple esa carencia de información hiperespectral mediante la obtención de imágenes a través de vuelos con vehículos aéreos no tripulados sobre zonas de cultivo y zonas de vuelo libre. Concretamente, y gracias al trabajo conjunto con el Grupo Integrado de Ingeniería de la Escuela Politécnica Superior de la Universidade da Coruña, se dispone de la tecnología necesaria para realizar, de forma progresiva y reiterada, vuelos con drones que proporcionan la obtención del material hiperespectral para el sistema.



**Fig. 1** Zona de vuelo del vehículo aéreo no tripulado.

La Figura 1 muestra una imagen de la zona de vuelo del UAV. El indicativo rojo en la imagen se corresponde con la situación habitual de los pilotos en el campo de vuelo. La zona de vuelo está situada en el Club de Aerodelismo Naron, en el municipio de Naron, La Coruña.

La superficie señalada cubre el área definida por la legislación vigente, la cual permite volar a una distancia de 500 metros para cualquier posición del piloto en el campo de vuelo. En lo referente a la obtención de los datos, el UAV, el dispositivo está preparado para la utilización en sus vuelos mediante la instalación de dos cámaras hiperespectrales, PHI V2 y UHD 185 con las siguientes características que pueden apreciarse en la tabla 1.

	PHI v2	UHD 185
Tecnología de adquisición	Pushbroom	Snapshot
Dimensiones (mm)	210x50x60	195x67x60
Peso (g)	680	470
Resolución espacial (pixel)	2048 (across-track)	50x50
Resolución espectral	En proceso de recalibración	8@532
Bits de digitalización	12	12
Bandas espectrales	360	125

**Tabla. 1** Características de las cámaras hiperespectrales del UAV.

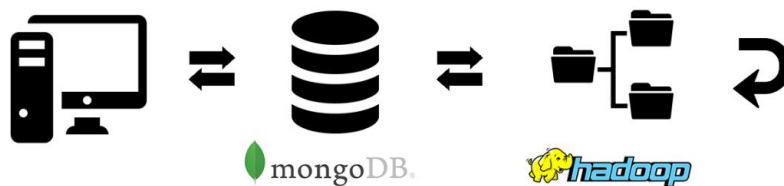
Inicialmente la disponibilidad geográfica de los datos de la plataforma estará centrada en España, tanto para los datos multispectrales que cubrirán todo el territorio nacional, como para los datos hiperespectrales obtenidos mediante los vuelos realizados que, en un principio cubrirán la zona de vuelo de Naron y se irán expandiendo a más zonas del territorio.

## 2.2 Almacenamiento de la Información

El proceso de almacenamiento de la información en el sistema está separado en dos fases atendiendo al tipo de datos.

Por un lado, una primera fase encargada de almacenar todo lo relativo a los metadatos de las imágenes espectrales, una vez extraídos y procesados. Para esta fase, la tecnología utilizada se corresponde con un sistema de base de datos NoSQL, concretamente MongoDB.

Por otro lado, una fase de almacenamiento de las imágenes espectrales sobre un sistema de archivos HDFS, debido al gran tamaño de este tipo de datos [12]. De este modo, se realiza una separación de la información para lograr que el sistema sea capaz de responder de modo adecuado en la gestión y eficiencia de los datos, como se puede apreciar en la figura 2.



**Fig. 2** Proceso del sistema encargado del almacenamiento de los metadatos e imágenes.

### Base de Datos NoSQL

El propósito del sistema desarrollado, en cuanto al almacenamiento, escalabilidad y persistencia de la información, es conseguir una mayor adaptación de los datos que con el uso de los sistemas relacionales tradicionales, por ello se ha considerado la utilización de la tecnología de los sistemas de base de datos NoSQL, por las ventajas que ofrece en estos casos, proporcionando una mayor adaptabilidad y un aumento de la eficiencia en aplicaciones intensivas de datos.

Los sistemas de bases de datos NoSQL son más versátiles y hacen frente a los problemas de alta escalabilidad de los datos cuando estos sufren un crecimiento notable dentro de un sistema. De este modo, mejoran el rendimiento al aplicar un escalado horizontal, es decir, añadiendo nuevos nodos al sistema en vez de recursos físicos como memoria, disco o CPU. Además, la gran mayoría de estos sistemas garantizan la disponibilidad de la información gracias al sharding, reduciendo el número de réplicas de los datos en cada uno de los nodos, proporcionando una ventaja al conseguir que la información se mueva por cada uno de los nodos del sistema. Muchos de estos sistemas de bases de datos NoSQL realizan operaciones directamente en memoria, volcando únicamente los datos a disco cada cierto tiempo, lo que permite que las operaciones de escritura que se realizan sean realmente rápidas.

La base de datos utilizada para el almacenamiento y acceso de los datos en nuestro sistema es MongoDB, ya que proporciona facilidad a la hora de realizar el diseño del modelo de datos. MongoDB es una base de datos NoSQL escrita en C++, libre de esquemas y orientada a documentos. El elemento base utilizado para codificar y almacenar los datos es el documento JSON, lo cual supone una ventaja al proporcionar simplicidad en el modelo de datos, ya que cualquier dato que llega al sistema se puede adecuar a este formato de una manera sencilla, característica importante a la hora de elegir el sistema de almacenamiento.

De este modo, cuando llegan al sistema datos espectrales de las fuentes externas citadas en los apartados anteriores, el propósito general es llevar al sistema toda la información contenida en los metadatos de las imágenes espectrales, cuya naturaleza es heterogénea debido a la gran variedad de información diferente que comprende. Para ello, los metadatos se extraen y se encapsulan en los esquemas diseñados para que el modelo de datos del sistema sea único, en la medida de lo posible, y poder así garantizar la eficiencia del mismo.

El almacenamiento de los datos en la base de datos se realiza para los metadatos de las imágenes espectrales, una vez realizada la extracción de los mismos y el posterior acople a las estructuras del modelo de datos en documentos JSON. Por ello, se han

definido las estructuras que establecen con precisión el significado y componentes de cada uno de los tipos de datos.

El esquema general para los principales tipos de datos del sistema realiza una división de la estructura atendiendo a:

- Tipo de dato espectral: Contiene información sobre la fuente de la imagen, es decir, si el dato ha sido obtenido mediante teledetección por satélite o utilizando UAV.
- Información general: Contiene información propia de la fuente externa, como el tipo de satélite o dispositivo UAV, identificador específico de la imagen, fecha de toma y adquisición, información acerca del software de procesado, etc.
- Información específica: Contiene información específica, según el tipo de dato en cuestión, que incluye parámetros como la fuente de elevación, el formato de imagen, tipo de sensor, coordenadas del producto (latitud, longitud), información específica de las bandas, valores para la reflectancia, brillo, valor del píxel, etc.

La estructura de estos esquemas puede estar formada por tipos simples para el almacenamiento de valores tales como cadenas de caracteres o números; o tipos estructurados más complejos, formados por elementos simples o de objetos del mismo tipo.

```

"LI_METADATA_FILE": {
  "METADATA_FILE_INFO": {
    "ORIGIN": "Image courtesy of the U.S. Geological Survey",
    "LANDSAT_SCENE_ID": "LC82030332017335LGN00",
    "FILE_DATE": "2017-12-07T08:16:01Z",
    "PROCESSING_SOFTWARE_VERSION": "LPGS_13.0.0"
  },
  "PRODUCT_METADATA": {
    "DATA_TYPE": "L1TP",
    "COLLECTION_CATEGORY": "T1",
    "DATE_ACQUIRED": "2017-12-01",
    "SCENE_CENTER_TIME": "11:08:29.0651720Z",
    "CORNER_UL_LAT_PRODUCT": "39.96111",
    "CORNER_UL_LON_PRODUCT": "-8.64757",
    "CORNER_UR_LAT_PRODUCT": "39.92193",
    "CORNER_UR_LON_PRODUCT": "-5.96597",
    "CORNER_LL_LAT_PRODUCT": "37.86057",
    "CORNER_LL_LON_PRODUCT": "-8.65781",
  }
}

```

**Fig. 3** Esquema de datos en formato JSON para el almacenamiento de los metadatos.

En la Figura 3 se observa, a modo de ejemplo, la sección del esquema de datos que alberga información de una imagen multispectral de Landsat 8, cuyos metadatos ya han sido procesados y cargados en la base de datos.

Un aspecto muy importante a tener en cuenta es el relativo a la coherencia entre las dimensiones de espacio y tiempo. La variabilidad temporal en los datos del sistema se contempla en los parámetros almacenados en los metadatos, que guardan la fecha de toma y adquisición de los datos, logrando que el sistema sea capaz de servir toda la información en diferentes momentos a lo largo del tiempo.

Por su parte, el concepto de variabilidad espacial es referido a aquellas modificaciones que pueden sufrir los cultivos o las áreas destinadas a los mismos. Como ocurre en [3], el sistema puede integrar dichos cambios añadiendo a la base de datos las correspondientes estructuras que hagan referencia a los cultivos y espacios físicos que aparecen en las imágenes espaciales y que han ido sufriendo alteraciones (diferentes distribuciones de los cultivos, aumento o decremento de las zonas destinadas a los mismos, desplazamientos o diferencias en el terreno, etc.) causados por el paso del tiempo.

De este modo, el sistema garantiza la coherencia entre el tiempo y el espacio en sus datos, siendo capaz de mantener la información actualizada y cada uno de los cambios provocados en ella.



## Sistema de Archivos HDFS

Como se ha visto anteriormente, el sistema almacena por separado los datos y sus metadatos. Utilizando un sistema de archivos distribuido para el almacenamiento de las imágenes, garantizamos una alta tolerancia a los fallos, un aumento en el rendimiento a la hora de acceder a los datos y, no menos importante, la posibilidad de ejecutar el sistema sobre un hardware económico, reduciendo los costes. Esta separación entre metadatos e imágenes en diferentes sistemas (base de datos y sistema de archivos distribuido) se realiza para lograr una mayor eficiencia a la hora de realizar consultas y carga de datos, de este modo el sistema puede liberar a la base de datos del proceso de almacenar y consultar información de gran tamaño y delegar estas tareas sobre el sistema de archivos, actuando este como un repositorio de imágenes del sistema global, conectado a la base de datos, que recupere la información necesaria de forma directa. La base de datos almacena la ruta de almacenamiento de estas imágenes dentro del sistema de archivos, por lo que una vez se realizan las consultas, la información contenida en los metadatos de las imágenes referida a la ubicación solo ha de ser leída y accedida para recuperar la imagen, de forma automática, y llevarla al sistema.

Para ello, implantamos la tecnología del sistema de archivos distribuido de Hadoop (HDFS) [12], especialmente optimizado para trabajar con datos de gran tamaño, que ofrece escalabilidad, tolerancia a fallos y alta concurrencia a través del particionado de los datos y la replicación entre nodos. De este modo, la información será dividida en bloques de gran tamaño y distribuida entre los nodos que forman el clúster de datos. Considerando el gran tamaño que tienen las imágenes espectrales suministradas por las fuentes externas, es de suma importancia lograr que nuestros datos se pueden particionar a lo largo de los nodos que componen del sistema.

Adicionalmente, HDFS realiza la replicación de los datos en los nodos, como protección a la posible fiabilidad del sistema, lo que permite la tolerancia a fallos. Si un nodo se cae o no está disponible, habrá otros nodos del sistema que tendrán una copia de los datos del nodo caído; de este modo, los datos seguirán estando disponibles. Se puede especificar el número de replications de un archivo cuando se crea, y dicho parámetro se puede cambiar posteriormente. Los bloques se replican de forma que queden próximos entre ellos para un mejor rendimiento.

Por defecto, en el particionado de los datos, el tamaño de bloque es de 64 megabytes y por el lado de la replicación, HDFS mantiene tres copias de cada bloque, aunque tanto el tamaño de bloque como la cantidad de copias pueden configurarse para adaptarse a las condiciones del sistema.

En resumen, integrando HDFS como tecnología de sistema de archivos distribuidos, conseguimos que, gracias al particionado y la replicación de los datos, sea más sencillo y seguro realizar el escalado de grandes volúmenes de datos y garantizar la fiabilidad al hacer frente a posibles fallos de hardware.

### 3 Procesos del Sistema

El sistema cuenta con varios procesos para realizar la extracción, transformación y carga de los datos desde las múltiples fuentes externas, dos de los principales se explican a continuación debido a la importancia de su funcionamiento para el sistema.

#### 3.1 Proceso de Extracción de Metadatos

Debido a la disparidad existente entre las fuentes externas que suministran de información al sistema, es necesario realizar un proceso previo de extracción y procesamiento de los metadatos incluidos en las imágenes espectrales. Dado que el almacenamiento de los mismos en la base de datos ha de seguir un esquema específico, como se ha comentado en puntos anteriores, el sistema lanza un proceso encargado de ir analizando esta información de manera automática y adaptándola al modelo de datos diseñado.

La información de las imágenes es diferente en cuanto a la técnica de teledetección [9] utilizada y, en el caso de aquellas imágenes de tipo satelital, los metadatos son diferentes en el contenido de la información y en el formato en que son proporcionados. Es decir, hay una clara diferencia entre los datos espectrales obtenidos con Landsat 8 y con Sentinel-2, ya que cada uno utiliza una metodología dispar y un formato específico para facilitar al usuario tanto las imágenes como los metadatos de las mismas [11]. Aunque existen muchos parámetros que son comunes entre uno y otro, generalmente no se utiliza la misma nomenclatura para describirlos, por lo que importante que la extracción de estos datos sea lo más exhaustiva y precisa posible.

Llegados a este punto, llevar toda esta información al sistema y acoplarla en un esquema diseñado para que sea acorde al modelo de datos utilizado, no es una tarea sencilla. Para ello se han desarrollado una serie de scripts, utilizando el lenguaje de programación Python, que se encargan de analizar la información contenida en cada uno de los datos de todas las fuentes externas e ir extrayendo los metadatos de manera precisa, lo cual garantiza que, una vez terminado dicho proceso, el sistema contará en su base de datos con la información necesaria ya adaptada, y los datos estarán disponibles para su administración y consulta.

#### 3.2 Proceso de Georreferenciación

En muchos casos, los datos obtenidos por las fuentes externas, tales como imágenes de satélite, fotografías aéreas, imágenes captadas por el vuelo de UAV o datos ráster en general, pueden no contener o poseer una referencia espacial con una exactitud insuficiente. Por este motivo, se necesita un proceso previo de georreferenciación para aquellos datos cuya información correspondiente a la ubicación y coordenadas geográficas del producto no es la adecuada.

La georreferenciación junto con las correcciones geométricas son los procesos de transformación de los datos para posicionar los píxeles en su posición correcta eliminando los posibles sesgos y distorsiones. Las imágenes obtenidas por teledetección contienen distorsiones geométricas debido a muchos factores y pueden clasificarse en sistemáticas y no sistemáticas [4]. Las distorsiones sistemáticas son: sesgo de escaneo,

velocidad del espejo de escaneo, efectos panorámicos y curvatura terrestre, velocidad de la plataforma y rotación terrestre. Las distorsiones geométricas no sistemáticas son: falta de normalidad de altitud, del terreno, de altura de vuelo del escaneo, las cuales distorsionan la escala; y la orientación del satélite en el espacio tridimensional.

La necesidad de realizar este tipo de corrección surge por las diferentes distorsiones que pueden darse por las condiciones del desplazamiento y adquisición de la información en la plataforma satelital [9].

En el contexto de teledetección, se entiende por georreferenciación y corrección geométrica a un proceso de cambio de espacio de referencia: la imagen original tomada por el sensor está definida en un sistema local donde la localización de cada píxel queda determinada por su situación en filas y columnas: es el espacio imagen. La corrección geométrica ajusta esta imagen a un nuevo espacio de referencia donde cada píxel tiene asignado el valor XY que le corresponda en un sistema de proyección geográfica determinado: es el espacio de proyección [10].

Normalmente es necesario partir de datos espaciales existentes, y alinear los datos erróneos sobre un sistema de coordenadas preciso y normalizado. Este proceso implica la identificación inicial de una serie de puntos de control del terreno para vincular las ubicaciones del dato inicial erróneo o falto de referenciación espacial con los datos correctos tomados como partida. En [10], se realiza una comparativa de los diferentes métodos de georreferenciación existentes y se discute sus posibilidades según el número y distribución de los puntos de apoyo o control empleados.

Estos puntos de control pueden identificarse con precisión con datos del mundo real. Se pueden utilizar técnicas diferentes para ello [9,10], como tomar la referencia sobre elementos o entidades identificables visualmente: intersecciones de corrientes o caminos, elementos naturales, esquinas de edificios o cualquier otra entidad. Los puntos de control son utilizados para generar una transformación polinómica que desplazará el dato, en este caso, la capa ráster de la imagen deseada, desde su ubicación existente a la ubicación espacialmente correcta. La conexión de las coordenadas entre un punto de control de partida y el punto de control correspondiente de los datos del punto de destino es un vínculo tomado por el proceso para realizar la georreferenciación.

Dependiendo del número de puntos creado, la transformación de la capa de datos hacia una correcta reubicación de las coordenadas, será más precisa cuanto mayor sea dicho número, teniendo en cuenta que la importancia reside en que dichos puntos estén correctamente expandidos sobre el terreno [10]. De lo contrario, seleccionar un número alto de puntos concentrados sobre una zona no tendrá el objetivo deseado en el proceso. Se puede tomar como referencia la selección de al menos un vínculo junto a cada esquina de la capa ráster y unos cuantos por el interior, garantizando buenos resultados.

Hay que destacar que este proceso es tan importante como costoso, pues se realiza de forma manual para aquellos datos previos a formar parte del sistema que necesiten un reajuste o un añadido en el valor de sus coordenadas. Del mismo modo, se busca que el sistema final sea capaz de realizar este proceso de forma automática y transparente, como sí ocurre con la extracción de los metadatos.

## 4 Aplicación Web del Sistema

Ya se ha contemplado en el artículo la estructura del sistema, conceptos importantes como el almacenamiento y gestión de la información, así como los procesos más relevantes y la descripción de cada uno de los elementos que alimentan al mismo. De cara a la capa de aplicación, el proyecto cuenta con una interfaz web para ofrecer y facilitar las funcionalidades del sistema al usuario. Esta aplicación web es la encargada de visualizar la información a través de las consultas realizadas por el usuario. Del mismo modo, también ejecuta funciones de gestión y almacenamiento sobre los datos.

Gracias a la amplia variedad de información contemplada en el sistema, el usuario podrá acceder a la misma mediante búsquedas de distinta tipología, ejecutando consultas sobre los elementos atendiendo a los parámetros temporales, espaciales y de clasificación de las imágenes:

- **Parámetro temporal:** Permite elegir el rango de fechas de consulta de los datos, así como una fecha en particular.
- **Parámetro espacial:** Permite elegir la ubicación exacta de los datos que se desean consultar, ya sea mediante las coordenadas geográficas de un punto concreto o mediante la creación de un polígono de coordenadas geográficas.
- **Tipo de datos:** Permite elegir el tipo de datos que se desea consultar. Imágenes satelitales (Landsat 8 y/o Sentinel-2), así como imágenes adquiridas mediante teledetección por el UAV.

La aplicación se apoya en el uso de la API de Google Maps para generar los mapas necesarios sobre los que visualizar las imágenes espectrales del sistema, como resultado de las búsquedas, como puede observarse en la Figura 4. Actualmente se está estudiando la posibilidad de combinar, como trabajo futuro, otros sistemas de visualización como OpenStreetMap en la plataforma, con el fin de aunar las ventajas de estos sistemas y así ofrecer un mejor soporte a la amplia naturaleza heterogénea de las imágenes y datos espaciales.

Gracias a la naturaleza de los datos y el amplio abanico de características que los componen, las funciones de búsqueda pueden personalizarse tanto como se necesite para lograr un resultado más preciso.

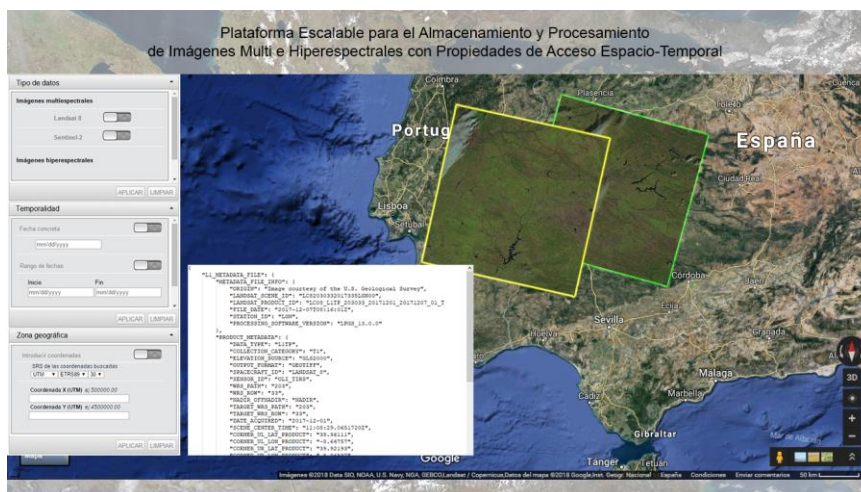


Fig. 4 Prototipo de la Plataforma de Datos.

## 5 Conclusiones

En los últimos años, el sector industrial agrónomo se ha visto favorecido por el impulso en el grado de desarrollo e innovación llevado a cabo con las nuevas tecnologías aplicadas a la agricultura de precisión, logrando imponer cada vez más este paradigma dentro del sector, y más concretamente en el entorno agrícola.

Los sistemas de información desarrollados bajo estas condiciones y enfocados al sector agrícola facilitan las tareas de gestión de explotaciones, análisis de los cultivos y toma de decisiones sobre los mismos, con el objetivo de lograr una mayor rentabilidad hacia un entorno más sostenible.

El sistema de información presentado en este artículo está desarrollado utilizando las tecnologías descritas, y tiene como objetivo principal hacer frente a las necesidades de los usuarios agrícolas, facilitando una gran cantidad de información variada. Además, se pretende seguir fielmente el concepto de agricultura de precisión, con el fin de que el sistema sea capaz de ofrecer una mejora en los procesos de gestión y productividad de explotaciones agrarias, así como una reducción en términos económicos.

El impacto que provoca conseguir un sistema eficiente que esté dotado con la amplia naturaleza de los datos espaciales a diferentes escalas que se han descrito, supone un punto de inflexión en el desarrollo de las plataformas de este tipo.

Gracias al trabajo conjunto de investigación y constante de desarrollo realizado con el Grupo Integrado de Ingeniería de la Universidade da Coruña, podemos implantar en el sistema diferentes técnicas de captación de información, consiguiendo llegar a los objetivos marcados a través de un sistema rentable y sostenible, que facilite el proceso de toma de decisiones en el ámbito agrícola.

**Agradecimientos:** Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad (ref. TIN2015-63646-C5-5-R) y la Junta de Extremadura y la Unión Europea (FEDER) a través de los fondos de ayuda a grupos de investigación GR1517.

## 6 Referencias

1. Popovic, Tomo & Latinović, Nedeljko & Pesic, Ana & Zecevic, Zarko & Krstajic, Bozo & Đukanović, Slobodan. (2017). Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study. *Computers and Electronics in Agriculture*. 2017. 255-265. 10.1016/j.compag.2017.06.008.
2. Khanal, Sami & Fulton, John & Shearer, Scott. (2017). An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture*. 139. 22-32. 10.1016/j.compag.2017.05.001.
3. Sánchez, Miguel & Barrena, Manuel & Bustos, Pablo & Campillo, Carlos & García, Pablo. (2016). Arquitectura Software Basada en Tecnologías Smart para Agricultura de Precisión. XXI Jornadas de Ingeniería del Software y Bases de Datos, Salamanca, 2016.
4. Richard, J. A. and X. Jia (2006), Remote Sensing Digital Image Analysis, Springer-Verlag Berlin Heidelberg, 4th Edition, pp. 46-50.
5. Kenji Ose, Thomas Corpetti and Laurent Demagistri, 2 - Multispectral Satellite Image Processing, In Optical Remote Sensing of Land Surface, edited by Nicolas Baghdadi and Mehrez Zribi, Elsevier, 2016, Pages 57-124, ISBN 9781785481024, <https://doi.org/10.1016/B978-1-78548-102-4.50002-8>.
6. Bhardwaj, Anshuman & Sam, Lydia, Akanksha & Martín-Torres, F. J. & Kumar, Rajesh. (2016). UAVs as remote sensing platform in glaciology: Present applications and future prospects. *Remote Sensing of Environment*. 175. 196-204. 10.1016/j.rse.2015.12.029.
7. USGS Global Visualization Viewer (GloVis). <https://glovis.usgs.gov/>. Accedido el 5 de marzo de 2018.
8. Earth Observing System (EOS). <https://eos.com/landviewer/>. Accedido el 27 de diciembre de 2017.
9. Sobrino, J. A. (2000), Teledetección, Servicio de publicaciones, Universidad de Valencia (ISBN 84-370-4220-8), Valencia (España), pp. 78-81.
10. Cuartero, A., and Felicísimo, A. M. (2003), Rectificación y ortorrectificación de imágenes de satélite: análisis comparativo y discusión, *GeoFocus (Artículos)*, N° 3, 2003, pp. 45-57. ISSN: 1578-5157.
11. Wang, Mi & Cheng, Yufeng & Chang, Xueli & Shuying, Jin & Zhu, Ying. (2017). On-orbit geometric calibration and geometric quality assessment for the high-resolution geostationary optical satellite GaoFen4. *ISPRS Journal of Photogrammetry and Remote Sensing*. 125. 63-77.
12. Ghazi, Mohd & Gangodkar, Durgaprasad. (2015). Hadoop, MapReduce and HDFS: a developers perspective. *Procedia Computer Science*. 48. 45-50. 10.1016/j.procs.2015.04.108.