

UNIVERSIDAD AUTÓNOMA CHAPINGO

**POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL
AGUA**

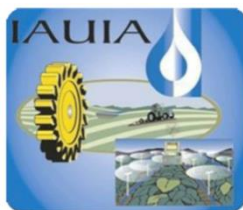
**ESTIMACIÓN DE LA EVAPOTRANSPIRACIÓN
DEL CULTIVO DE MAÍZ CON FUSIÓN DE
IMÁGENES Y REDES NEURONALES
TESIS**

Que como requisito parcial para obtener el grado de:

DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

Presenta: C. ALVARO MURGUIA COZAR

Bajo la supervisión de: GILBERTO DE JESÚS LÓPEZ CANTEÑS, DR



Chapingo, Estado de México, noviembre de 2026.



**ESTIMACIÓN DE LA EVAPOTRANSPIRACIÓN DEL CULTIVO DE MAÍZ CON
FUSIÓN DE IMÁGENES Y REDES NEURONALES**

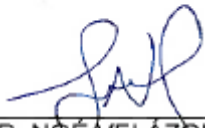
Tesis realizada por C. ALVARO MURGUIA COZAR bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DIRECTOR:


DR. GILBERTO DE JESÚS LÓPEZ CANTÉNS

ASESOR:


DR. NOÉ VELÁZQUEZ LÓPEZ

ASESOR:


DRA. ANTONIA MACEDO CRUZ

LECTOR
EXTERNO


DR. RONALD ERNESTO ONTIVEROS CAPURATA

CONTENIDO

LISTA DE CUADROS	ix
LISTA DE FIGURAS	x
DEDICATORIAS	xii
AGRADECIMIENTOS	xiii
DATOS BIOGRÁFICOS	xiv
RESUMEN GENERAL	xv
GENERAL ABSTRACT	xvi
1 Introducción General	17
1.1 Objetivo General.....	20
1.2 Objetivos Específicos	20
1.3 Hipótesis.....	20
1.4 Estructura de la tesis	20
1.5 Literatura citada	21
2 Revisión de Literatura	23
2.1 Teledetección y sensores remotos para agricultura de precisión	23
2.1.1 Plataformas de observación	24
2.1.1.1 Plataformas satelitales	24
2.1.1.2 Vehículos aéreos no tripulados (UAV)	24
2.1.1.3 Sensores proximales y plataformas terrestres	25
2.1.1.4 Integración de plataformas para agricultura de precisión	26

2.2	Fusión y superresolución de imágenes multiespectrales.....	27
2.2.1	Modelos de fusión tradicionales	28
2.2.2	Superresolución basada en aprendizaje profundo	28
2.2.2.1	Modelos SISR	31
2.2.2.2	Modelos MISR.....	32
2.2.2.3	Modelos de superresolución de imágenes Sentinel-2 y UAV.....	34
2.3	Estimación de la evapotranspiración del maíz mediante redes neuronales.....	35
2.4	Referencias	37
3	Fusión de imágenes UAV–Sentinel-2 usando redes neuronales convolucionales para análisis de teledetección*	44
1.	Introduction	47
2.	Methodology.....	50
2.1.	<i>Study area</i>	50
2.2.	<i>Multispectral image acquisition</i>	50
2.2.2.	<i>UAV imagery</i>	50
2.2.3.	<i>Data normalisation</i>	51
2.2.4.	<i>Generation of vegetation indices</i>	51
2.3.	<i>Super-resolution models</i>	52
2.3.2.	<i>VDSR Model</i>	52
2.3.3.	<i>ESRGAN model</i>	53
2.4.	<i>Model training and inference</i>	53

2.5.	<i>Evaluation metrics</i>	54
2.5.1.	<i>Root mean square error (RMSE)</i>	54
2.5.2.	<i>Peak signal-to-noise ratio (PSNR)</i>	54
2.5.3.	<i>Structural similarity index measure (SSIM)</i>	55
2.5.4.	<i>Correlation</i>	55
2.5.5.	<i>Determination of evaluation metrics</i>	55
3.	<i>Results</i>	55
3.1.	<i>Multiscale performance analysis</i>	55
3.2.	<i>Spectral index behaviour</i>	57
3.3.	<i>Training stability across dates at 2 m</i>	60
3.4.	<i>Computational efficiency</i>	62
4.	<i>Discussion</i>	62
5.	<i>Conclusions</i>	64
4	<i>Author contributions</i>	64
5	<i>Disclosure statement</i>	64
6	<i>References</i>	65
4	<i>Fusión temporal de imágenes UAV y Sentinel-2 mediante redes neuronales convolucionales para el monitoreo de cultivos</i>	68
4.1	<i>Introducción</i>	70
4.2	<i>Trabajos relacionados</i>	73
4.2.1	<i>Modelos de superresolución de imágenes múltiples</i>	73
4.2.2	<i>Modelos de superresolución espacio temporales</i>	75

4.2.3	Modelos de superresolución para la integración Sentinel-2 con UAV	77
4.3	Materiales y Métodos.....	79
4.3.1	Área de Estudio	79
4.3.2	Adquisición de datos	80
4.3.2.1	Imágenes UAV	80
4.3.2.2	Imágenes Sentinel-2	81
4.3.2.3	Índices de Vegetación	82
4.3.3	Método propuesto.....	83
4.3.3.1	Arquitectura general del modelo TVDSR	83
4.3.3.2	Módulo de extracción de características espaciales (Encoder)	86
4.3.3.3	Modelado temporal mediante ConvLSTM bidireccional	86
4.3.3.4	Modelado estacional basado en embeddings y FiLM.....	88
4.3.3.5	Fusión temporal y mecanismos de atención	89
4.3.3.5.1	Mecanismo de atención temporal	90
4.3.3.5.2	Mecanismo de atención espectral.....	91
4.3.3.5.3	Módulo de fusión paralela	92
4.3.3.6	Módulo de reconstrucción espacial (Decoder)	93
4.3.3.7	Función de pérdida y estrategia de entrenamiento	94
4.3.3.8	Implementación y detalles computacionales	95
4.3.4	Validación del modelo TVDSR	96

4.4	Resultados y discusión	99
4.4.1	Complejidad computacional de los modelos evaluados.	99
4.4.2	Desempeño de los modelos de superresolución.....	100
4.4.3	Capacidad de generalización temporal.	102
4.4.4	Desempeño de los modelos de superresolución por índice de vegetación.	105
4.4.5	Evaluación visual de las imágenes reconstruidas.	107
4.5	Conclusiones	111
4.6	Referencias	112
5	Estimación espacial de la evapotranspiración del cultivo de maíz mediante imágenes de dron, Sentinel-2 y aprendizaje profundo.....	120
5.1	Introducción	122
5.2	Metodología.....	126
5.2.1	Área de estudio	126
5.2.2	Imágenes multiespectrales UAV.....	126
5.2.3	Imágenes Sentinel-2.....	127
5.2.4	Generación de Índices de Vegetación.....	128
5.2.5	Fusión de imágenes mediante el modelo TVDSR.....	129
5.2.6	Datos meteorológicos.....	130
5.2.7	Humedad de suelo	130
5.2.8	Parámetros de suelo	130
5.2.9	Estimación de evapotranspiración.....	132

5.2.9.1	Determinación del Kcb espacial	132
5.2.9.2	Estimación del Ks y el Ke	134
5.2.9.3	Construcción del coeficiente dual.....	137
5.2.10	Estimación del coeficiente dual de cultivo mediante aprendizaje profundo	138
5.3	Resultados y discusión	141
5.3.1	Modelos de calibración de los sensores Watermark	141
5.3.2	Comparación entre sensores Watermark y balance hídrico	143
5.3.3	Desempeño del modelo de segmentación semántica	145
5.3.4	Desempeño del modelo U-Net VGG para la estimación de Kcdual	147
5.3.5	Estimación espacial de la ETc del maíz	151
5.4	Conclusiones	156
5.5	Literatura Citada	157

LISTA DE CUADROS

Cuadro 2.1. Principales métodos tradicionales de fusión aplicados a imágenes satelitales y UAV.....	29
Cuadro 4.1. Índices de Vegetación evaluados en el estudio.	83
Cuadro 4.2. Métricas de calidad empleadas para la evaluación del desempeño de los modelos de superresolución.	97
Cuadro 4.3. Desempeño y costo computacional de los modelos.....	100
Cuadro 4.4. Resultados cuantitativos consolidados para las dos fechas de evaluación.....	102
Cuadro 4.5. Variación porcentual del desempeño entre las fechas de evaluación (2024–2025).	103
Cuadro 5.1. Propiedades físicas e hidráulicas de los puntos de monitoreo de humedad del suelo.	144
Cuadro 5.2. Comparación entre sensores Watermark y balance hídrico.....	144
Cuadro 5.3. Desempeño del modelo de segmentación semántica.	146
Cuadro 5.4. Desempeño de la red U-Net VGG respecto a la fuente de datos.	148
Cuadro 5.5. Evaluación de la ETc derivada de imágenes UAV, Sentinel-2 y Sentinel-2 superresueltas mediante balance hídrico.	152

LISTA DE FIGURAS

Figura 2.1. Niveles de procesamiento en una fusión de imágenes (Pohl & Van Genderen, 1998)	27
Figura 4.1. Localización geográfica y delimitación del área de estudio en la Universidad Autónoma Chapingo.	80
Figura 4.2. Flujo metodológico para la fusión de imágenes Sentinel-2 y UAV mediante el modelo TVDSR.	81
Figura 4.3. Estructura general del modelo TVDSR.	85
Figura 4.4. Estructura del encoder del modelo TVDSR.	86
Figura 4.5. Estructura del decoder del modelo TVDSR.	94
Figura 4.6. Mapa de calor del desempeño de los modelos de superresolución por índice de vegetación.	105
Figura 4.7. Comparación visual de las imágenes reconstruidas por los modelos de superresolución para la fecha del 3 de agosto de 2025.	108
Figura 5.1. Ubicación del área de estudio y puntos de monitoreo. (a) Localización en el Estado de México. (b) Distribución de parcelas y puntos de monitoreo.	127
Figura 5.2. Modelo espaciotemporal TVDSR.	129
Figura 5.3. Puntos de monitoreo de humedad del suelo (a) diagrama de instalación; (b) Punto de monitoreo.	131
Figura 5.4. Arquitectura U-Net-VGG utilizada para la segmentación semántica del cultivo de maíz.	136
Figura 5.5. Arquitectura U-Net-VGG utilizada para la predicción del coeficiente del cultivo dual.	139
Figura 5.6. Modelos ajustados para los puntos de monitoreo (a) punto 1, (b) punto 2 y (c) punto 3.	142
Figura 5.7. Segmentación de parcelas del 3 de agosto de 2025.	146

Figura 5.8. Mapa de Kcdual predicho para el 3 de agosto de 2025 generado con la imagen superresuelta. 151

Figura 5.9. Mapas de evapotranspiración del maíz obtenidos a partir de imágenes a) UAV, y b) Sentinel-2 superresuelta y c) Sentinel-2. 154

DEDICATORIAS

A ti, Laura, y a ti, mi pequeña Miranda, quiero dedicar este gran logro. Sin ustedes, nada de esto habría sido posible. Han sido mi inspiración, mi fortaleza y el motor que me impulsa a seguir adelante en los momentos más difíciles. ¡Las amo!

AGRADECIMIENTOS

Secretaría de Ciencia, Humanidades, Tecnología e Innovación, SECIHTI de México por la beca prestada para el desarrollo de esta investigación.

A la Universidad Autónoma Chapingo y al Posgrado en Ingeniería Agrícola y Uso Integral del Agua por permitirme continuar con mi preparación profesional.

Al Dr. Gilberto de Jesús López Canteñs, por brindarme la oportunidad de desarrollar la presente investigación bajo su valiosa guía y orientación.

la Dra. Antonia Macedo Cruz, por su invaluable orientación y por continuar guiándome en mi proceso de formación y preparación profesional.

Al Dr. Noé Velázquez López por todo su apoyo y guía.

A mi madre Marina, mi padre Jesús y mis hermanos Néstor, Jesús y Rolando por todo su amor y cariño que me demuestran día con día.

Un agradecimiento especial a Mayra, quien siempre estuvo atenta a las necesidades del posgrado y de los estudiantes. Con tu apoyo, amabilidad y disposición para ayudar, los procesos administrativos resultaron siempre más sencillos. Gracias por tu invaluable labor y por hacer más fácil este camino.

A mis profesores y compañeros, quienes con su amistad, apoyo y compañía hicieron que los retos de este proceso fueran más fáciles de afrontar. Gracias por formar parte de esta importante etapa de mi vida.

DATOS BIOGRÁFICOS



Datos Personales:

Nombre: Alvaro Murguía Cozar
Fecha de Nacimiento: 31 de mayo de 1989
Lugar de Nacimiento: Progreso, Atoyac, Veracruz
No. Cartilla Militar:
CURP: MUCA890531HVZRZL09
Profesión: Maestro en Ciencias en Hidrociencias
Cédula Profesional: 12964980

Desarrollo Académico:

Bachillerato: Preparatoria Agrícola (Universidad Autónoma Chapingo)
Licenciatura: Ingeniería Mecánica Agrícola (Universidad Autónoma Chapingo)
Maestría: Maestría en Ciencias en Hidrociencias (Colegio de Postgraduados)

RESUMEN GENERAL

ESTIMACIÓN DE LA EVAPOTRANSPIRACIÓN DEL CULTIVO DE MAÍZ CON FUSIÓN DE IMÁGENES Y REDES NEURONALES

La creciente necesidad de optimizar el uso del agua en la agricultura requiere herramientas capaces de estimar con precisión variables biofísicas relacionadas con la evapotranspiración a escalas espaciales detalladas. Recientemente, la teledetección agrícola ha tomado gran auge en este tipo de aplicaciones mediante la integración de datos satelitales y de drones utilizando algoritmos de superresolución. En esta investigación se propuso una metodología para determinar la distribución espacial de la evapotranspiración del maíz basada en la integración de información multispectral de Sentinel-2 y UAV mediante técnicas de fusión de imágenes y modelos de aprendizaje profundo. Se realizó el monitoreo de 50 hectáreas cultivadas con maíz en los ciclos agrícolas 2024 y 2025 y se integraron imágenes UAV-Sentinel-2 a nivel índices de vegetación. Se evaluaron modelos de superresolución espacial y se propuso un modelo espaciotemporal TVDSR para fusionar las imágenes UAV-Sentinel-2 bajo el enfoque multiimagen, además se entrenó un modelo U-Net VGG para predecir el coeficiente dual del cultivo (K_{cdual}) y estimar la evapotranspiración del maíz. Los resultados demostraron que la superresolución espacial permitió incrementar la resolución de Sentinel-2 a 2m con métricas de reconstrucción adecuadas ($RMSE = 0.041$; $PSNR = 33.83$ dB; $SSIM = 0.983$; $Corr = 0.981$). El modelo TVDSR, que integra la resolución temporal, permitió reconstruir las imágenes satelitales a la misma resolución con métricas altas ($RMSE=0.149$, $MAE=0.115$, $PSNR=22.60$ dB y $SSIM=0.85$) y con la menor degradación entre fechas, lo que evidenció una mayor capacidad de generalización temporal. Aunque las imágenes UAV obtuvieron las mejores métricas en la predicción del K_{cdual} y la evapotranspiración del cultivo, las imágenes superresueltas redujeron el RMSE en 16 % y el MAE en 17 % respecto a Sentinel-2 original. En conjunto, los resultados obtenidos demuestran que la combinación de información Sentinel-2 y UAV mediante métodos de superresolución basados en aprendizaje profundo constituye una alternativa eficaz para garantizar un balance entre el nivel de detalle espacial y la extensión de superficie monitoreada fortaleciendo la estimación de la distribución espacial del cultivo de maíz.

Palabras Clave: Superresolución; TVDSR; Índice de Vegetación; Aprendizaje profundo.

Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Programa de Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: M.C. Alvaro Murguía Cozar.

Director de tesis: Dr. Gilberto de Jesús López Canteñs.

GENERAL ABSTRACT

ESTIMATION OF MAIZE CROP EVAPOTRANSPIRATION WITH IMAGE FUSION AND NEURAL NETWORKS

The growing need to optimize water use in agriculture requires tools capable of accurately estimating biophysical variables related to evapotranspiration at fine spatial scales. Recently, agricultural remote sensing has gained considerable relevance for this type of application through the integration of satellite and UAV data using super-resolution algorithms. In this study, a methodology was proposed to determine the spatial distribution of maize evapotranspiration based on the integration of Sentinel-2 and UAV multispectral information through image fusion techniques and deep learning models. A total of 50 hectares cultivated with maize were monitored during the 2024 and 2025 growing seasons, and UAV-Sentinel-2 imagery was integrated at the vegetation-index level. Spatial super-resolution models were evaluated, and a spatiotemporal TVDSR model was proposed to fuse UAV and Sentinel-2 imagery under a multi-image approach. In addition, a U-Net VGG model was trained to predict the dual crop coefficient (K_{cdual}) and estimate maize evapotranspiration. The results showed that spatial super-resolution increased the spatial resolution of Sentinel-2 imagery to 2 m, with adequate reconstruction metrics (RMSE = 0.041; PSNR = 33.83 dB; SSIM = 0.983; Corr = 0.981). The TVDSR model, which incorporates temporal information, reconstructed the satellite imagery at the same spatial resolution with high performance metrics (RMSE = 0.149, MAE = 0.115, PSNR = 22.60 dB, and SSIM = 0.85) and the lowest degradation across dates, indicating greater temporal generalization capability. Although UAV imagery achieved the best metrics for K_{cdual} prediction and crop evapotranspiration estimation, the super-resolved imagery reduced RMSE by 16% and MAE by 17% compared with the original Sentinel-2 imagery. Overall, the results demonstrate that integrating Sentinel-2 and UAV information through deep learning-based super-resolution methods provides an effective alternative for balancing spatial detail and the extent of the monitored area, thereby improving the estimation of the spatial distribution of maize evapotranspiration.

Keywords: Super-resolution; TVDSR; Vegetation Index; Deep Learning.

Doctoral Thesis. Graduate Program in Agricultural Engineering and Integrated Water Resources Management, Universidad Autónoma Chapingo.

Author: Msc. Alvaro Murguía Cozar.

Thesis director: Phd. Gilberto de Jesús López Canteñs.

1 Introducción General

El crecimiento sostenido de la población mundial es un fenómeno innegable que plantea serios retos para la seguridad alimentaria global. De acuerdo con las Naciones Unidas (2025), la población mundial pasó de 2,500 millones en 1950 a más de 8,000 millones en 2022, y se estima que alcanzará los 9,700 millones para 2050 y 10,400 millones en 2100. Este incremento implica una demanda creciente de alimentos y, en consecuencia, una presión cada vez mayor sobre los sistemas agrícolas y los recursos hídricos necesarios para sostener la producción.

En este contexto, uno de los principales retos consiste en incrementar la productividad agrícola sin aumentar proporcionalmente el consumo de agua. Actualmente, la agricultura utiliza alrededor del 70 % del agua dulce extraída a nivel mundial (Banco Mundial, 2025) , por lo que la optimización del riego y la cuantificación precisa de las necesidades hídricas se han convertido en elementos fundamentales para garantizar la sostenibilidad de los sistemas productivos. Desde esta perspectiva, la Agricultura 5.0 propone la integración de tecnologías digitales, inteligencia artificial y sistemas avanzados de observación para mejorar la eficiencia en el uso de los recursos, aumentar la productividad y fortalecer la resiliencia frente al cambio climático (Méndez-Jocik et al., 2021)

La estimación de las necesidades hídricas de los cultivos depende principalmente de la evapotranspiración del cultivo (ETc), variable que representa la cantidad de agua requerida para mantener un desarrollo adecuado de las plantas. Sin embargo, la obtención de ETc a escala regional continúa siendo compleja debido a la limitada disponibilidad de estaciones meteorológicas, los costos asociados al monitoreo de campo y la dificultad de implementar balances hídricos detallados en grandes superficies agrícolas (De Caro et al., 2023).

Aunque la teledetección ha permitido superar parcialmente estas limitaciones, persiste un problema importante relacionado con la resolución espacial de la información disponible. Las imágenes satelitales de libre acceso, como Sentinel-2, ofrecen una elevada cobertura espacial y temporal, pero su resolución suele resultar insuficiente para capturar la variabilidad, dentro de la parcela, requerida en aplicaciones de agricultura de precisión. Por el contrario, las imágenes obtenidas mediante vehículos aéreos no tripulados proporcionan un nivel de detalle mayor, aunque presentan limitaciones asociadas a costos operativos, cobertura espacial reducida, imágenes sin calibración radiométrica absoluta y campañas de adquisición esporádicas. Esta diferencia ha impulsado el desarrollo de estrategias de fusión de datos y técnicas de superresolución basadas en inteligencia artificial, cuyo objetivo es combinar la cobertura temporal de los satélites con el detalle espacial proporcionado por los drones, lo que permite generar información de alta resolución con potencial para mejorar la estimación de variables relacionadas con la evapotranspiración de cultivo.

En este contexto, la teledetección y los sensores remotos se han consolidado como herramientas valiosas para estimar variables agronómicas claves como el coeficiente de cultivo y la evapotranspiración. Diversos estudios han demostrado su aplicación para evaluar el desarrollo vegetal, el estrés hídrico y la gestión del riego, mediante la generación de mapas espaciales y temporales de ETc (Blanta et al., 2011; Kadam et al., 2017). Asimismo, el modelo FAO Penman-Monteith se ha consolidado como el estándar internacional para la estimación de la evapotranspiración mediante la combinación de variables climáticas y coeficientes de cultivo (Pereira et al., 2025)

Dado el alto costo y la limitada disponibilidad de datos climáticos en muchas regiones agrícolas, una solución viable ha consistido en utilizar sensores remotos, tanto satelitales como montados en drones, para extraer información relacionada con el estado fenológico y fisiológico de los cultivos. Esta información

puede emplearse para estimar la ETc de forma más precisa y a escalas espaciales extensas (Maimaitijiang et al., 2020).

Las imágenes satelitales de libre acceso, como Landsat 8 y Sentinel-2, han sido fundamentales para el monitoreo agrícola a gran escala; sin embargo, su baja resolución espacial limita la identificación de patrones detallados dentro de los cultivos. Por otra parte, los satélites comerciales de alta resolución y los drones proporcionan imágenes con un elevado nivel de detalle, aunque presentan restricciones asociadas a costos de adquisición, logística operativa y frecuencia de observación (Alvarez-Vanhard et al., 2021). Con el fin de aprovechar las ventajas de ambas fuentes de información, diversos investigadores han explorado estrategias de fusión entre imágenes de alta y baja resolución para mejorar el detalle espacial sin sacrificar cobertura temporal y regional (Gao et al., 2006; Jiang et al., 2023; Song et al., 2020).

En los últimos años, los avances en aprendizaje profundo han impulsado el desarrollo de técnicas de superresolución capaces de reconstruir imágenes de alta resolución a partir de observaciones de menor resolución espacial. Estas técnicas representan una alternativa prometedora para generar productos con mayor nivel de detalle sin incrementar significativamente los costos de adquisición de datos. Desde la perspectiva de la gestión agrícola, la superresolución puede contribuir a mejorar la representación espacial de variables biofísicas relacionadas con la evapotranspiración, el estado hídrico y el desarrollo de los cultivos. No obstante, aún persisten interrogantes respecto a la capacidad de estos modelos para preservar la información espectral necesaria para aplicaciones agronómicas y mantener un comportamiento estable frente a diferentes condiciones temporales de observación.

1.1 Objetivo General

Estimar la distribución espacial de la evapotranspiración del cultivo de maíz mediante imágenes Sentinel-2, drones y técnicas de superresolución basadas en aprendizaje profundo.

1.2 Objetivos Específicos

- Fusionar composiciones multibanda derivadas de índices de vegetación obtenidos a partir de información Sentinel-2 y UAV mediante redes neuronales convolucionales, para generar productos de alta resolución espacial.
- Desarrollar y evaluar un modelo de superresolución espaciotemporal basado en aprendizaje profundo para generar composiciones multibanda de alta resolución espacial derivadas de índices de vegetación obtenidos a partir de información Sentinel-2 y UAV.
- Entrenar, validar y probar una red neuronal tipo U-Net VGG mediante el procesamiento de cinco índices de vegetación para estimar la distribución espacial del coeficiente dual del cultivo y la evapotranspiración del cultivo de maíz.

1.3 Hipótesis

Las redes neuronales basadas en aprendizaje profundo permiten mejorar la resolución espacial de información derivada de imágenes Sentinel-2 mediante el uso de datos UAV, lo que genera productos capaces de estimar con precisión la distribución espacial de la evapotranspiración del cultivo de maíz.

1.4 Estructura de la tesis

El presente documento de tesis está estructurado en cinco capítulos, organizados para describir la metodología propuesta para la estimación espacial de la evapotranspiración del cultivo de maíz.

En el Capítulo 1 se presenta la introducción general, donde se expone el contexto, la problemática y la relevancia de la estimación espacial de la evapotranspiración en sistemas agrícolas. En el Capítulo 2 se desarrolla el marco teórico y se revisan los antecedentes relacionados con la aplicación de la teledetección, el aprendizaje profundo y los modelos basados en inteligencia artificial para la estimación de variables agronómicas vinculadas a la evapotranspiración.

El Capítulo 3 presenta una metodología de fusión basada en tres algoritmos clásicos de superresolución aplicados a composiciones multibanda derivadas de índices de vegetación obtenidos a partir de información Sentinel-2 y UAV, con el propósito de determinar la escala espacial más adecuada para el proceso de reconstrucción.

El Capítulo 4 describe el modelo de superresolución espaciotemporal TVDSR propuesto en esta investigación, el cual integra información Sentinel-2 y UAV de diferentes fechas para generar productos de alta resolución espacial con una mayor estabilidad temporal.

Finalmente, el Capítulo 5 presenta el desarrollo, entrenamiento, validación y evaluación de una red neuronal tipo U-Net VGG para estimar espacialmente el coeficiente dual del cultivo y, a partir de éste, determinar la distribución espacial de la evapotranspiración del cultivo de maíz.

1.5 Literatura citada

- Alvarez-Vanhard, E., Corpetti, T., & Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. In *Science of Remote Sensing* (Vol. 3). Elsevier B.V. <https://doi.org/10.1016/j.srs.2021.100019>
- Banco Mundial. (2025, April 2). *Agua*. El Agua Es La Base de Un Planeta Habitable. <https://www.bancomundial.org/es/topic/water/overview>

- Blanta, A., Nicolas, D. R., Maliara, A., & Spyropoulos, N. (2011). Monitoring cotton crop evapotranspiration based on satellite data. *Sustainable Agriculture Production and Environment*, 1–12.
- De Caro, D., Ippolito, M., Cannarozzo, M., Provenzano, G., & Ciraolo, G. (2023). Assessing the performance of the Gaussian Process Regression algorithm to fill gaps in the time-series of daily actual evapotranspiration of different crops in temperate and continental zones using ground and remotely sensed data. *Agricultural Water Management*, 290, 108596. <https://doi.org/10.1016/j.agwat.2023.108596>
- Gao, F., Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207–2218. <https://doi.org/10.1109/TGRS.2006.872081>
- Jiang, J., Atkinson, P. M., Chen, C., Cao, Q., Tian, Y., Zhu, Y., Liu, X., & Cao, W. (2023). Combining UAV and Sentinel-2 satellite multi-spectral images to diagnose crop growth and N status in winter wheat at the county scale. *Field Crops Research*, 294. <https://doi.org/10.1016/j.fcr.2023.108860>
- Kadam, S. A., Gorantiwar, S. D., Das, S. N., & Joshi, A. K. (2017). Crop Evapotranspiration Estimation for Wheat (*Triticum aestivum* L.) Using Remote Sensing Data in Semi-Arid Region of Maharashtra. *Journal of the Indian Society of Remote Sensing*, 45, 297–305. <https://doi.org/10.1007/s12524-016-0594-1>
- Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., & Fritschi, F. B. (2020). Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sensing*, 12(9). <https://doi.org/10.3390/RS12091357>
- Méndez-Jocik, C. Alberto., Núñez-Pérez, A. de J., & González-Robaina, F. (2021). Vista de Aplicación de la teledetección espacial al manejo de los recursos hídricos. *Revista Ciencias Técnicas Agropecuarias*, 30(1), 17–25.
- Naciones Unidas. (2025). *Paz, dignidad e igualdad en un planeta sano*. Desafíos Globales Población. <https://www.un.org/es/global-issues/population#:~:text=D%C3%ADA%20de%20los%208000%20millones,hito%20en%20el%20desarrollo%20humano>
- Pereira, L. S., Allen, R. G., Paredes, P., López-Urrea, R., Raes, D., Smith, M., Kilic, A., & Salman, M. (2025). Crop evapotranspiration-Guidelines for computing crop water requirements. In *Crop evapotranspiration* (Second). FAO. <https://doi.org/10.4060/cd6621en>
- Song, Z., Zhang, Z., Yang, S., Ding, D., & Ning, J. (2020). Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Computers and Electronics in Agriculture*, 179. <https://doi.org/10.1016/j.compag.2020.105812>

2 Revisión de Literatura

2.1 Teledetección y sensores remotos para agricultura de precisión

La agricultura de precisión (AP) se define como una estrategia de gestión basada en datos que recoge, procesa y analiza información temporal, espacial e individual para apoyar decisiones de manejo según la variabilidad estimada, con el fin de mejorar la eficiencia en el uso de recursos, la productividad, la calidad, la rentabilidad y la sostenibilidad de la producción agrícola (Ammoniaci et al., 2021). En la práctica, la AP se sintetiza en la idea de "hacer lo correcto, en el lugar correcto y en el momento correcto", y su punto de partida es la recolección de datos mediante sensores proximales o remotos, cuya interpretación agronómica se traduce en recomendaciones de manejo o en insumos para máquinas de aplicación de tasa variable (Ammoniaci et al., 2021) .

La teledetección aporta la capacidad de caracterizar de forma no destructiva, repetitiva y geoespacialmente el estado de los cultivos a lo largo del ciclo agrícola, desde escalas centimétricas (sensores proximales y drones) hasta escalas regionales y globales (satélites) (Weiss et al., 2020). Esta información es la base para realizar el seguimiento de cultivos, gestión del riego, fertilización nitrogenada, control de malezas y enfermedades, y predicción de rendimiento (Sishodia et al., 2020). En términos generales, la teledetección se concibe como la adquisición de información sobre un objeto a distancia mediante un sensor montado en una plataforma (satélite, aeronave, UAV o sonda), que mide la radiación electromagnética reflejada por la superficie (Weiss et al., 2020) .

Los sistemas de teledetección empleados en AP se clasifican de acuerdo con la plataforma del sensor y el tipo de sensor. La plataforma condiciona la resolución espacial, temporal y radiométrica; el tipo de sensor determina la resolución espectral y la naturaleza de la variable recuperable (Sishodia et al., 2020) .

2.1.1 Plataformas de observación

2.1.1.1 Plataformas satelitales

La era de la teledetección satelital aplicada a la agricultura comienza con el lanzamiento del Landsat 1 en 1972, al que siguieron misiones como SPOT, IRS, IKONOS, RapidEye, WorldView, Landsat 8/9, Sentinel, entre otras. Los satélites Sentinel, y en particular Sentinel-2, se han convertido en el estándar para el monitoreo agrícola por su carácter gratuito, su resolución espacial adecuada a escala de parcela y su alta frecuencia temporal (Ammoniacci et al., 2021). Su utilidad se ha demostrado para estimar nitrógeno y variables biofísicas con modelos paramétricos y de aprendizaje automático (Perich et al., 2021), caracterizar biomasa en viñedos (Sassu et al., 2021), correlacionar la humedad del suelo a través de índices como NDVI y NDMI para la gestión precisa del agua (Ibrahim et al., 2023) .

Sin embargo, la resolución espacial de Sentinel-2 puede resultar insuficiente para representar adecuadamente la variabilidad dentro de la parcela requerida por aplicaciones avanzadas de agricultura de precisión. Un análisis comparativo entre información satelital y UAV mostró que Sentinel-2 representa adecuadamente tendencias generales del cultivo, pero presenta limitaciones para identificar variaciones locales a escalas finas, especialmente en fechas individuales de adquisición (Bukowiecki et al., 2021) .

2.1.1.2 Vehículos aéreos no tripulados (UAV)

Los vehículos aéreos no tripulados han transformado la adquisición de información agrícola debido a su capacidad para generar imágenes con resolución centimétrica y gran flexibilidad operativa. La incorporación de cámaras multiespectrales, hiperespectrales y térmicas ha permitido el monitoreo detallado del estado de los cultivos, la detección temprana de enfermedades, la evaluación nutricional y el análisis del estrés hídrico. Guebsi et al. (2024) destacaron que los UAV constituyen herramientas fundamentales para el monitoreo de cultivos, la gestión del riego, la estimación de rendimiento y la detección temprana de

problemas agronómicos mediante la integración de sensores avanzados e inteligencia artificial.

Asimismo, los UAV permiten obtener observaciones bajo demanda, superando parcialmente las limitaciones asociadas a la nubosidad y frecuencia de revisita de los sensores satelitales. La combinación de UAV y algoritmos de aprendizaje profundo ha mostrado altos niveles de precisión para el monitoreo de cultivos y detección de enfermedades, con niveles de exactitud superiores al 95 % en algunos escenarios agrícolas (Zhu et al., 2024).

A pesar de su elevado detalle espacial, los UAV presentan limitaciones relacionadas con cobertura espacial, duración de vuelo, costos operativos y volumen de datos generado. Por ello, su uso suele complementarse con sistemas satelitales, para aprovechar la naturaleza complementaria de ambas plataformas (Guebsi et al., 2024; Phang et al., 2023) .

2.1.1.3 Sensores proximales y plataformas terrestres

Los sensores proximales corresponden a dispositivos instalados sobre maquinaria agrícola, plataformas móviles o sistemas portátiles que operan muy cerca de la superficie cultivada. Estos sensores permiten obtener mediciones directas y continuas con elevadas resoluciones espaciales, lo que minimiza los efectos atmosféricos que afectan a los sistemas remotos convencionales.

Desarrollos recientes incluyen sensores espectrales activos, cámaras térmicas, sistemas de fluorescencia, LiDAR y plataformas integradas con Internet de las Cosas. Estas tecnologías permiten estimar variables como índice de área foliar, biomasa, contenido de nitrógeno, estado hídrico, vigor vegetativo y rendimiento potencial de los cultivos. Laveglia et al. (2024) destacaron que los sensores proximales permiten identificar estrés, deficiencias nutricionales, enfermedades y plagas con altos niveles de precisión, lo que promueve las prácticas agrícolas más sostenibles.

Las investigaciones más recientes muestran que la combinación de sensores proximales y plataformas UAV genera resultados especialmente favorables para el monitoreo de cereales y otros cultivos extensivos (Grbović et al., 2025).

2.1.1.4 Integración de plataformas para agricultura de precisión

La tendencia actual en agricultura de precisión se orienta hacia la integración de sensores satelitales, UAV y sensores proximales dentro de marcos de análisis unificados apoyados por inteligencia artificial. Los sensores satelitales proporcionan cobertura continua a gran escala, los UAV ofrecen información de alta resolución espacial para caracterizar la variabilidad dentro de la parcela y los sensores proximales suministran mediciones de referencia de elevada precisión.

Diversos trabajos recientes coinciden en que la integración de estas plataformas mejora significativamente la calidad de la información utilizada para la gestión del riego, la estimación del rendimiento y la evaluación del estado fisiológico de los cultivos. Además, la combinación de teledetección e inteligencia artificial ha abierto nuevas oportunidades para generar productos agrícolas de mayor resolución espacial y temporal, lo que favorece aplicaciones avanzadas asociadas al monitoreo agrícola y a la estimación de la evapotranspiración (Laveglia et al., 2024; Phang et al., 2023; Xing et al., 2026) .

En este contexto, las técnicas de fusión de datos y superresolución basadas en aprendizaje profundo representan una línea de investigación emergente orientada a aprovechar simultáneamente la cobertura de los sensores satelitales y el elevado detalle espacial proporcionado por los UAV, aspecto que constituye la base metodológica de la presente investigación.

2.2 Fusión y superresolución de imágenes multiespectrales

La combinación de imágenes multiespectrales es un tema amplio que data de la década de los 90's que refiere al proceso de combinar dos imágenes de diferentes fuentes para mejorar la información de una de las dos. De acuerdo con Pohl & Van Genderen (1998) la fusión de imágenes se puede ejecutar en tres niveles de procesamiento: nivel pixel, nivel característica y nivel decisión (Figura 2.1)

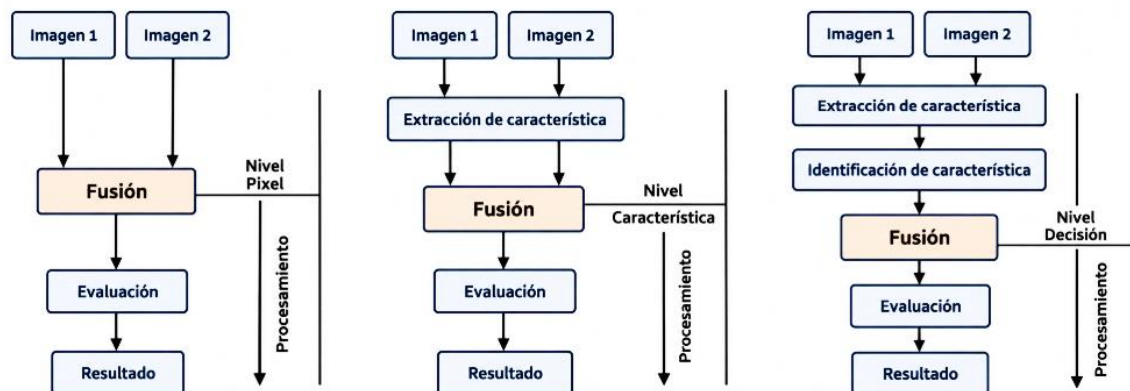


Figura 2.1. Niveles de procesamiento en una fusión de imágenes (Pohl & Van Genderen, 1998).

La fusión a nivel píxel manipula directamente los valores radiométricos de las imágenes originales. Este enfoque requiere que las imágenes se encuentren correctamente corregidas geométricamente y compartan la misma referencia espacial. Por su parte, la fusión a nivel de características utiliza objetos o patrones previamente extraídos de las imágenes, tales como regiones segmentadas, índices de vegetación o descriptores texturales. Finalmente, la fusión a nivel de decisión combina resultados obtenidos a partir de diferentes clasificadores o algoritmos de análisis para producir una única salida final (Alvarez-Vanhard et al., 2021; Pohl & Van Genderen, 1998).

2.2.1 Modelos de fusión tradicionales

El Cuadro 2.1 resume los principales métodos de fusión y mejoramiento espacial reportados en la literatura para la integración de información proveniente de Satélite y UAV. Los enfoques tradicionales se han basado principalmente en transformaciones espectrales, modelos geoestadísticos y algoritmos de desagregación espacial. Aunque estos métodos han demostrado ser efectivos para incrementar la resolución espacial o temporal de los productos generados, presentan limitaciones asociadas a la introducción de artefactos, dependencia de parámetros específicos o pérdida de información espectral. Estas limitaciones han impulsado el desarrollo de métodos basados en aprendizaje profundo, particularmente redes neuronales convolucionales, que constituyen la línea de investigación predominante en los estudios recientes de superresolución de imágenes multiespectrales.

2.2.2 Superresolución basada en aprendizaje profundo

El surgimiento de las redes neuronales convolucionales y el aprendizaje profundo marcó un cambio importante en la evolución de los métodos de fusión de imágenes. La aparición del aprendizaje profundo permitió reemplazar estos procedimientos por modelos capaces de aprender automáticamente relaciones no lineales entre imágenes de baja y alta resolución. En consecuencia, la fusión clásica de imágenes fue sustituida por la superresolución, entendida como el proceso de reconstruir imágenes de mayor resolución espacial con la información espectral y estructural de los datos originales. Este cambio impulsó el desarrollo de nuevas arquitecturas especializadas basadas en redes convolucionales que actualmente constituyen una de las principales líneas de investigación en teledetección y agricultura de precisión.

Cuadro 2.1. Principales métodos tradicionales de fusión aplicados a imágenes satelitales y UAV.

Método	Datos utilizados	Principio general	Aplicación principal	Limitaciones
Pansharpening (Scarpa & Ciotola, 2022)	Imagen multiespectral + banda pancromática	Integra resolución espacial y multirresolución, optimización	Generación de imágenes multiespectrales de alta resolución	Requiere banda pancromática y puede introducir distorsión espectral
IHS (Intensity-Hue-Saturation) (Scarpa & Ciotola, 2022)	Imagen multiespectral + banda pancromática	Sustituye la componente de intensidad por la banda pancromática	Fusión espacial de imágenes multiespectrales	Distorsión espectral significativa cuando aumenta la diferencia radiométrica
PCA (Principal Component Analysis) (Shahdoosti & Ghassemian, 2011)	Imagen multiespectral + banda pancromática	Sustituye el primer componente principal por la banda pancromática	Mejoramiento espacial y reducción de dimensionalidad	Modifica las relaciones espectrales originales
Gram-Schmidt (Laben & Brower, 2000)	Sentinel-2 y UAV	Transformación ortogonal para eliminar correlaciones entre bandas e incorporar detalle espacial	Clasificación de coberturas del suelo	Sensible a diferencias radiométricas entre sensores
Wavelet Transform (Shahdoosti, 2015)	Imágenes multiespectrales + banda pancromáticas	Integra componentes espaciales y mediante descomposición en frecuencia	Conservación espectral en imágenes fusionadas	Mayor complejidad computacional
ATPRK (Wang et al., 2015)	MODIS, Sentinel-2 y Landsat	Regresión kriging área-punto para desagregar píxeles de baja resolución	Incremento de resolución espacial multiespectral	Alta complejidad computacional y dependencia de supuestos geoestadísticos
TMM desagregado (Zurita-Milla et al., 2008)	Landsat, MERIS y otros sensores multirresolución	Estima la contribución de píxeles de alta resolución mediante clasificación no supervisada y desagregación espacial	Monitoreo de cultivos y variables biofísicas	Puede generar artefactos y depende del número de clases y tamaño de ventana
STARFM (Gao et al., 2006)	Landsat + MODIS	Combina imágenes de alta resolución espacial y alta frecuencia temporal mediante ponderaciones espaciales y temporales	Monitoreo fenológico y generación de imágenes sintéticas multitemporales	Menor desempeño ante cambios abruptos en la cobertura
ESTARFM (Li et al., 2023)	Landsat + MODIS; Sentinel-2 + UAV	Extensión de STARFM para regiones heterogéneas	Monitoreo agrícola y cambios temporales	Mayor complejidad y dependencia de imágenes parejas

Método	Datos utilizados	Principio general	Aplicación principal	Limitaciones
FIT-FC (Li et al., 2023)	Landsat y MODIS	Ajuste de curvas temporales y espacialización de reflectancias	Generación de series temporales de alta resolución	Sensible a errores radiométricos
FSDAF (Li et al., 2023)	Landsat y MODIS	Integración de cambios espaciales y temporales mediante mezcla adaptativa	Monitoreo de cultivos y uso de suelo	Requiere múltiples imágenes de entrada
SIFT (Song et al., 2020)	UAV multitemporal	Extracción y correspondencia de características invariantes a escala	Identificación de acame y segmentación agrícola	No incrementa directamente la resolución espacial
MVIF (Fortuna et al., 2020)	RGB de alta resolución + sensor multispectral de baja resolución	Modelo multivariado basado en mínimos cuadrados para recuperar información espectral	Reconstrucción espectral de alta resolución	Requiere alta correlación entre sensores

En términos generales, los modelos de superresolución basados en aprendizaje profundo pueden clasificarse en dos grandes categorías: superresolución de imagen única (SISR) y superresolución multiimagen (MISR). Los enfoques SISR utilizan una única imagen de entrada para reconstruir una versión de mayor resolución espacial, mientras que los modelos MISR explotan múltiples observaciones de una misma escena adquiridas en diferentes instantes o desde sensores distintos para extraer información complementaria durante la reconstrucción.

2.2.2.1 Modelos SISR

El primer modelo de superresolución basado en CNN fue Super-Resolution Convolutional Neural Network, propuesto por Dong et al. (2014) , que estructuró el problema en tres etapas: extracción de características, mapeo no lineal y reconstrucción. Posteriormente, los mismos autores desarrollaron Faster Super-Resolution Convolutional Neural Network, sin la interpolación bicúbica inicial e incorporando una capa de deconvolución para realizar el aumento de resolución dentro de la propia red (Dong, & Tang, 2016) .

La evolución posterior se centró en arquitecturas más profundas apoyadas en aprendizaje residual. En esta línea, Very Deep Super-Resolution introdujo redes profundas con aprendizaje residual global, que mejora la convergencia y la reconstrucción de detalles de alta frecuencia (Kim et al., 2016) . Sobre esta base surgieron modelos que incorporaron mecanismos de atención para priorizar información relevante. Entre ellos destacan Residual Channel Attention Network, que incorpora módulos de atención por canales para enfatizar información de alta frecuencia (Zhang et al., 2018), y Dense Channel Attention Network, que complementa la atención espectral con mecanismos de atención espacial para mejorar la localización de características relevantes (Ma et al., 2021) . Más recientemente, High-Frequency Mamba Network (HFMN) integró mecanismos de autoatención para combinar información local y global durante la reconstrucción (Wu et al., 2025).

Modelos basados en redes generativas adversariales, orientados a mejorar la calidad perceptual de las reconstrucciones. SRGAN fue el primer modelo en incorporar un esquema generador-discriminador junto con pérdida perceptual basada en VGG para producir imágenes visualmente más realistas (Ledig et al., 2017). Posteriormente, Enhanced Super-Resolution Generative Adversarial Network mejoró esta arquitectura mediante bloques residuales densos, un discriminador relativista y una pérdida perceptual optimizada, convirtiéndose en una de las referencias más importantes dentro del área (Wang et al., 2018). A partir de ESRGAN surgieron variantes como A-ESRGAN, que incorpora mecanismos de atención multiescala (Wei et al., 2023), y arquitecturas orientadas a fusión multispectral como MSCGAN, diseñada para aplicaciones de pansharpening mediante la integración conjunta de información espacial y espectral (Zhang et al., 2022).

En conjunto, la evolución de los modelos SISR ha transitado desde arquitecturas CNN relativamente simples hacia modelos profundamente integrados que combinan aprendizaje residual, mecanismos de atención y redes generativas adversariales. Estos desarrollos han permitido mejorar significativamente la calidad de reconstrucción espacial y la preservación de la información espectral, lo que consolidó al aprendizaje profundo como una de las principales herramientas para la superresolución de imágenes multispectrales.

2.2.2.2 Modelos MISR

El aprendizaje profundo permitió generar arquitecturas convolucionales capaces de aprender directamente la relación entre múltiples observaciones LR y la reconstrucción HR. En una primera etapa, los modelos dependían de algoritmos externos para estimar el movimiento. Ejemplos representativos son los trabajos de (Liao et al., 2015) y VSRNet desarrollado por Kappeler et al. (2016), donde la alineación mediante flujo óptico y la interpolación bicúbica precedían al proceso de reconstrucción mediante redes convolucionales.

Una evolución importante ocurrió con la incorporación de mecanismos de alineación directamente dentro de las redes neuronales. En este contexto, Tao et al. (2017) introdujeron la Sub-Pixel Motion Compensation (SPMC), una capa capaz de realizar simultáneamente la compensación de movimiento y el proceso de aumento de resolución espacial. Esta contribución eliminó la dependencia de módulos externos y permitió desarrollar arquitecturas completamente integradas.

A partir de entonces surgieron modelos más sofisticados que incorporaron alineación, fusión y reconstrucción dentro de una misma estructura. Entre ellos destacan RBPN propuesto por Haris et al. (2019) , EDVR desarrollado por (X. Wang, Chan, et al., 2019) y HighRes-Net presentado por Deudon et al. (2020) . Estas arquitecturas integran mecanismos de atención espacial y temporal, convoluciones deformables y esquemas de retroproyección recurrente que permiten explotar simultáneamente la información temporal y espacial presente en múltiples observaciones.

La creciente disponibilidad de secuencias temporales satelitales impulsó el desarrollo de modelos explícitamente espaciotemporales. En este contexto surgieron arquitecturas recurrentes basadas en ConvLSTM capaces de modelar dependencias temporales de largo alcance. Entre ellas destaca BRCN propuesta por Huang et al. (2018) , que incorpora convoluciones tridimensionales y estructuras recurrentes bidireccionales para procesar simultáneamente información pasada y futura. En años posteriores, diferentes variantes de ConvLSTM fueron utilizadas para integrar información multitemporal en tareas de superresolución, restauración e integración multimodal (Heng et al., 2020; Li et al., 2024; Xu et al., 2019; Zhang et al., 2020).

Más recientemente, los modelos basados en mecanismos de atención y Transformers han adquirido gran relevancia. Arquitecturas como BSRT (Luo et al., 2022) , MAST (Li et al., 2023) , TR-MISR (An et al., 2022), EpiMISR (Aira et al., 2024) e HiRT (Choi & Kim, 2026) explotan mecanismos de atención global

capaces de establecer relaciones entre regiones distantes dentro de la imagen y entre observaciones adquiridas en distintos instantes. Estas arquitecturas han demostrado una capacidad superior para preservar estructuras complejas y capturar dependencias espaciotemporales a largo plazo.

2.2.2.3 Modelos de superresolución de imágenes Sentinel-2 y UAV

La aplicación de técnicas de superresolución a imágenes Sentinel-2 ha evolucionado desde esquemas centrados únicamente en la mejora de las bandas nativas del sensor hacia enfoques que integran múltiples fuentes de información. Los primeros trabajos utilizaron las bandas de alta resolución de Sentinel-2 para mejorar las bandas de menor resolución. Ejemplos representativos incluyen DSen2 (Lanaras et al., 2018), el modelo residual propuesto por Palsson et al. (2018) y FUSE desarrollado por (Gargiulo et al., 2019).

Posteriormente, comenzaron a utilizarse fuentes externas de información espacial, como PlanetScope, WorldView y UAV. En esta línea, (Galar et al., 2020b) emplearon imágenes PlanetScope para generar pares de entrenamiento de un modelo EDSR adaptado a Sentinel-2. Latte & Lejeune (2020) desarrollaron una arquitectura residual para fusionar Sentinel-2 y PlanetScope a 2.5 m, mientras que (Jia et al., 2020) introdujeron DL-SDFM, un modelo que integra simultáneamente relaciones espaciales y temporales mediante una red convolucional de dos flujos.

La incorporación de modelos generativos adversariales también ha mostrado resultados prometedores. Entre ellos destacan RS-ESRGAN (Salgueiro Romero et al., 2020a), MAF-GAN (Wang et al., 2025) y GS-SRGAN (Farah et al., 2026), los cuales buscan preservar simultáneamente la calidad espacial y la consistencia espectral de las imágenes reconstruidas. En paralelo, modelos más recientes han explorado enfoques basados en difusión, como DiffFuSR (Sarmad et al., 2025), así como mecanismos de atención espectral y espacial adaptativa para mejorar la estabilidad de la reconstrucción (Nguyen-Vi et al., 2025).

En el contexto de la integración Sentinel-2 y UAV, RarefyNet (Mazzia et al., 2020), UAV-Net (Xiao et al., 2023) y S³-ESRGAN (Toosi et al., 2026) representan algunos de los desarrollos más relevantes. Estos modelos aprovechan la complementariedad entre la elevada frecuencia temporal de Sentinel-2 y la alta resolución espacial de los UAV para generar productos capaces de preservar simultáneamente información espacial, espectral y temporal. Como señalan (Ren et al., 2025), la integración de ambas plataformas constituye actualmente una de las líneas más prometedoras para la generación de productos de alta resolución aplicados al monitoreo agrícola y ambiental.

A pesar de los avances recientes, persisten desafíos relacionados con la alineación geométrica y temporal entre sensores, la disponibilidad de datos de referencia de alta calidad, el costo computacional asociado a arquitecturas profundas y la necesidad de preservar la consistencia espectral en escenarios multitemporales. Estas limitaciones continúan impulsando el desarrollo de modelos más robustos, eficientes y adaptativos capaces de integrar de forma coherente información multiescala, multisensor y multitemporal.

2.3 Estimación de la evapotranspiración del maíz mediante redes neuronales.

La evapotranspiración del cultivo (ET_c) constituye una de las variables fundamentales para la programación del riego y la gestión eficiente del agua en agricultura. En los últimos años, la disponibilidad de imágenes satelitales, sensores remotos e inteligencia artificial ha impulsado el desarrollo de metodologías capaces de estimar la ET_c de forma espacialmente distribuida y con niveles crecientes de precisión. Diversas revisiones recientes coinciden en que los modelos de aprendizaje automático y aprendizaje profundo superan frecuentemente a los enfoques empíricos tradicionales para la estimación de la evapotranspiración, particularmente cuando se integran variables biofísicas derivadas de sensores remotos (Amani & Shafizadeh-Moghadam, 2023).

En el cultivo de maíz, los modelos de inteligencia artificial han mostrado una elevada capacidad para estimar tanto la evapotranspiración como los coeficientes de cultivo. Por ejemplo, Chen et al. (2020) desarrollaron modelos basados en redes de convolución temporal (TCN) que alcanzaron $R^2 = 0.95$, MAE = 0.309 mm d^{-1} y MSE = 0.144 mm d^{-1} , superando arquitecturas LSTM y redes neuronales profundas convencionales. De manera similar, Zhao et al. (2024) obtuvieron $R^2 = 0.842$, RMSE = 0.433 mm d^{-1} y MAE = 0.316 mm d^{-1} mediante modelos híbridos optimizados con algoritmos bioinspirados.

La incorporación de información proveniente de sensores remotos ha permitido extender estos enfoques hacia estimaciones espacialmente distribuidas. Diversos estudios han demostrado que índices de vegetación como NDVI, NDWI, NDRE, EVI y GNDVI mantienen relaciones sólidas con el coeficiente de cultivo y la evapotranspiración. Utilizando imágenes Sentinel-2, se han reportado correlaciones de hasta 0.95 entre índices espectrales y coeficientes de cultivo, así como modelos de ETc con errores inferiores a 0.44 mm d^{-1} (Nagy et al., 2024). De igual forma, series temporales de NDVI derivadas de Sentinel-2 permitieron estimar evapotranspiración con coeficientes de determinación entre 0.95 y 0.96 respecto a mediciones obtenidas mediante sistemas Eddy Covariance (French et al., 2023).

Las imágenes obtenidas mediante UAV han mostrado niveles de precisión aún mayores debido a su alta resolución espacial. En maíz, se han obtenido estimaciones de coeficiente de cultivo con $R^2 = 0.94$, RMSE = 0.055 y MAE = 0.043 utilizando índices multiespectrales derivados de imágenes UAV (Marcial-Pablo et al., 2021). Asimismo, la combinación de índices de vegetación, coeficientes de estrés hídrico y el enfoque FAO-56 permitió estimar la evapotranspiración con $R^2 = 0.81$, RMSE = 0.95 mm d^{-1} y un índice de concordancia de 0.94 (Tang et al., 2019). Más recientemente, modelos multimodales basados en CNN, Transformers y mecanismos de atención alcanzaron $R^2 = 0.889$, RMSE = 0.939 mm d^{-1} y MAE = 0.713 mm d^{-1} ,

evidenciando el potencial del aprendizaje profundo para integrar variables meteorológicas, espectrales y fisiológicas en la estimación de ET_c (Zhangzhong et al., 2026).

La evidencia disponible demuestra que la resolución espacial constituye un factor determinante para mejorar la representación de variables biofísicas relacionadas con la evapotranspiración. Sin embargo, la mayoría de los estudios se han desarrollado utilizando imágenes Sentinel-2 nativas o información UAV de muy alta resolución. Existe todavía una cantidad limitada de investigaciones que evalúan el efecto de imágenes Sentinel-2 superresueltas sobre la estimación espacial de la evapotranspiración. Esta situación representa una oportunidad de investigación relevante, ya que la combinación de superresolución, teledetección multiespectral e inteligencia artificial podría generar productos con una resolución intermedia entre Sentinel-2 y UAV, manteniendo simultáneamente cobertura regional, frecuencia temporal y coherencia espectral. Precisamente en este contexto se enmarca la presente investigación.

2.4 Referencias

- Aira, L. S., Valsesia, D., Molini, A. B., Fracastoro, G., Magli, E., & Mirabile, A. (2024). Deep 3D World Models for Multi-Image Super-Resolution Beyond Optical Flow. *IEEE Access*, 12, 188902–188913. <https://doi.org/10.1109/ACCESS.2024.3514188>
- Alvarez-Vanhard, E., Corpetti, T., & Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. In *Science of Remote Sensing* (Vol. 3). Elsevier B.V. <https://doi.org/10.1016/j.srs.2021.100019>
- Amani, S., & Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. In *Agricultural Water Management* (Vol. 284). Elsevier B.V. <https://doi.org/10.1016/j.agwat.2023.108324>
- Ammoniaci, M., Kartsiotis, S.-P., Perria, R., & Storchi, P. (2021). State of the Art of Monitoring Technologies and Data Processing for Precision Viticulture. *Agriculture*, 11(3), 201. <https://doi.org/10.3390/agriculture11030201>
- An, T., Zhang, X., Huo, C., Xue, B., Wang, L., & Pan, C. (2022). TR-MISR: Multiimage Super-Resolution Based on Feature Fusion With Transformers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 1373–1388. <https://doi.org/10.1109/JSTARS.2022.3143532>

- Bukowiecki, J., Rose, T., & Kage, H. (2021). Sentinel-2 data for precision agriculture? —a uav-based assessment. *Sensors*, 21(8). <https://doi.org/10.3390/s21082861>
- Chen, Z., Sun, S., Wang, Y., Wang, Q., & Zhang, X. (2020). Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. *Computers and Electronics in Agriculture*, 169. <https://doi.org/10.1016/j.compag.2019.105206>
- Choi, Y. J., & Kim, B. G. (2026). Hierarchical recurrent transformer network for video super-resolution. *Engineering Applications of Artificial Intelligence*, 166. <https://doi.org/10.1016/j.engappai.2025.113714>
- Deudon, M., Kalaitzis, A., Goytom, I., Arefin, M. R., Lin, Z., Sankaran, K., Michalski, V., Kahou, S. E., Cornebise, J., & Bengio, Y. (2020). *HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery*. <https://arxiv.org/abs/2002.06460>
- Dong, C., Loy, C. C., He, K., & Tang, X. (2014). *Learning a Deep Convolutional Network for Image Super-Resolution* (pp. 184–199). https://doi.org/10.1007/978-3-319-10593-2_13
- Dong, C., Loy, C. C., & Tang, X. (2016). *Accelerating the Super-Resolution Convolutional Neural Network* (pp. 391–407). https://doi.org/10.1007/978-3-319-46475-6_25
- Farah, S., Saadaoui, H., & Rhinane, H. (2026). Enhancing Spatial Resolution of Sentinel-2 Imagery through Deep Learning and Generative Adversarial Networks: GS-SRGAN. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W17-2025, 123–134. <https://doi.org/10.5194/isprs-archives-xxviii-4-w17-2025-123-2026>
- Feng Gao, Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207–2218. <https://doi.org/10.1109/TGRS.2006.872081>
- Fortuna, J., Martens, H., & Johansen, T. A. (2020). Multivariate image fusion: A pipeline for hyperspectral data enhancement. *Chemometrics and Intelligent Laboratory Systems*, 205. <https://doi.org/10.1016/j.chemolab.2020.104097>
- French, A. N., Sanchez, C. A., Wirth, T., Scott, A., Shields, J. W., Bautista, E., Saber, M. N., Wisniewski, E., & Gohardoust, M. R. (2023). Remote sensing of evapotranspiration for irrigated crops at Yuma, Arizona, USA. *Agricultural Water Management*, 290. <https://doi.org/10.1016/j.agwat.2023.108582>
- Galar, M., Sesma, R., Ayala, C., Albizua, L., & Aranda, C. (2020). Super-resolution of Sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18). <https://doi.org/10.3390/RS12182941>
- Gargiulo, M., Mazza, A., Gaetano, R., Ruello, G., & Scarpa, G. (2019). Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks. *Remote Sensing*, 11(22). <https://doi.org/10.3390/rs11222635>
- Grbović, Ž., Ivošević, B., Budjen, M., Waqar, R., Pajević, N., Ljubičić, N., Kandić, V., Pajić, M., & Panić, M. (2025). Integrating UAV multispectral imaging and proximal

- sensing for high-precision cereal crop monitoring. *PLOS ONE*, 20(5 May). <https://doi.org/10.1371/journal.pone.0322712>
- Guebsi, R., Mami, S., & Chokmani, K. (2024). Drones in Precision Agriculture: A Comprehensive Review of Applications, Technologies, and Challenges. In *Drones* (Vol. 8, Number 11). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/drones8110686>
- Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent Back-Projection Network for Video Super-Resolution. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3892–3901. <https://doi.org/10.1109/CVPR.2019.00402>
- Heng, H., Ye, H., & Huang, R. (2020). Defocus Blur Detection by Fusing Multiscale Deep Features with Conv-LSTM. *IEEE Access*, 8, 97279–97288. <https://doi.org/10.1109/ACCESS.2020.2996200>
- Huang, Y., Wang, W., & Wang, L. (2018). Video Super-Resolution via Bidirectional Recurrent Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 1015–1028. <https://doi.org/10.1109/TPAMI.2017.2701380>
- Ibrahim, G. R. F., Rasul, A., & Abdullah, H. (2023). Assessing how irrigation practices and soil moisture affect crop growth through monitoring Sentinel-1 and Sentinel-2 data. *Environmental Monitoring and Assessment*, 195(11). <https://doi.org/10.1007/s10661-023-11871-w>
- Jia, D., Song, C., Cheng, C., Shen, S., Ning, L., & Hui, C. (2020). A novel deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions using a two-stream convolutional neural network. *Remote Sensing*, 12(4). <https://doi.org/10.3390/rs12040698>
- Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video Super-Resolution With Convolutional Neural Networks. *IEEE Transactions on Computational Imaging*, 2(2), 109–122. <https://doi.org/10.1109/TCI.2016.2532323>
- Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1646–1654. <https://doi.org/10.1109/CVPR.2016.182>
- Laben, C. A., & Brower, B. V. (2000). *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*.
- Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., & Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 305–319. <https://doi.org/10.1016/j.isprsjprs.2018.09.018>
- Latte, N., & Lejeune, P. (2020). PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. *Remote Sensing*, 12(15). <https://doi.org/10.3390/RS12152366>
- Laveglia, S., Altieri, G., Genovese, F., Matera, A., & Di Renzo, G. C. (2024). Advances in Sustainable Crop Management: Integrating Precision Agriculture and Proximal

- Sensing. In *AgriEngineering* (Vol. 6, Number 3, pp. 3084–3120). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/agriengineering6030177>
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105–114. <https://doi.org/10.1109/CVPR.2017.19>
- Li, C., Song, S., Wang, X., Liu, Y., & Bi, L. (2024). RCFNC: a resolution and contrast fusion network with ConvLSTM for low-light image enhancement. *Visual Computer*, 40(4), 2793–2806. <https://doi.org/10.1007/s00371-023-02986-9>
- Li, J., Lv, Q., Zhang, W., Zhu, B., Zhang, G., & Tan, Z. (2023). Multi-Attention Multi-Image Super-Resolution Transformer (MAST) for Remote Sensing. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174183>
- Li, Y., Yan, W., An, S., Gao, W., Jia, J., Tao, S., & Wang, W. (2023). A Spatio-Temporal Fusion Framework of UAV and Satellite Imagery for Winter Wheat Growth Monitoring. *Drones*, 7(1). <https://doi.org/10.3390/drones7010023>
- Liao, R., Tao, X., Li, R., Ma, Z., & Jia, J. (2015). Video Super-Resolution via Deep Draft-Ensemble Learning. *2015 IEEE International Conference on Computer Vision (ICCV)*, 531–539. <https://doi.org/10.1109/ICCV.2015.68>
- Luo, Z., Li, Y., Cheng, S., Yu, L., Wu, Q., Wen, Z., Fan, H., Sun, J., & Liu, S. (2022). BSRT: Improving Burst Super-Resolution with Swin Transformer and Flow-Guided Deformable Alignment. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 997–1007. <https://doi.org/10.1109/CVPRW56347.2022.00113>
- Ma, Y., Lv, P., Liu, H., Sun, X., & Zhong, Y. (2021). Remote sensing image super-resolution based on dense channel attention network. *Remote Sensing*, 13(15). <https://doi.org/10.3390/rs13152966>
- Marcial-Pablo, M. de J., Ontiveros-Capurata, R. E., Jiménez-Jiménez, S. I., & Ojeda-Bustamante, W. (2021). Maize crop coefficient estimation based on spectral vegetation indices and vegetation cover fraction derived from uav-based multispectral images. *Agronomy*, 11(4). <https://doi.org/10.3390/AGRONOMY11040668>
- Mazzia, V., Comba, L., Khaliq, A., Chiaberge, M., & Gay, P. (2020). UAV and machine learning based refinement of a satellite-driven vegetation index for precision agriculture. *Sensors (Switzerland)*, 20(9). <https://doi.org/10.3390/s20092530>
- Nagy, A., Kiss, N. É., Buday-Bódi, E., Magyar, T., Cavazza, F., Gentile, S. L., Abdullah, H., Tamás, J., & Fehér, Z. Z. (2024). Precision Estimation of Crop Coefficient for Maize Cultivation Using High-Resolution Satellite Imagery to Enhance Evapotranspiration Assessment in Agriculture. *Plants*, 13(9). <https://doi.org/10.3390/plants13091212>
- Nguyen-Vi, K., Bui-Quoc, B., & Kamel, N. (2025). Deep learning-based Sentinel-2 super-resolution via channel attention and high-frequency feature enhancement. *Frontiers in Remote Sensing*, 6. <https://doi.org/10.3389/frsen.2025.1644460>

- Palsson, F., Sveinsson, J. R., & Ulfarsson, M. O. (2018). Sentinel-2 image fusion using a deep residual network. *Remote Sensing*, 10(8). <https://doi.org/10.3390/rs10081290>
- Perich, G., Aasen, H., Verrelst, J., Argento, F., Walter, A., & Liebisch, F. (2021). Crop nitrogen retrieval methods for simulated sentinel-2 data using in-field spectrometer data. *Remote Sensing*, 13(12). <https://doi.org/10.3390/rs13122404>
- Phang, S. K., Chiang, T. H. A., Happonen, A., & Chang, M. M. L. (2023). From Satellite to UAV-Based Remote Sensing: A Review on Precision Agriculture. *IEEE Access*, 11, 127057–127076. <https://doi.org/10.1109/ACCESS.2023.3330886>
- Pohl, C., & Van Genderen, J. L. (1998). Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. In *International Journal of Remote Sensing* (Vol. 19, Number 5, pp. 823–854). <https://doi.org/10.1080/014311698215748>
- Ren, A., Zhou, X., Li, G., Li, J., Ma, Y., Qiao, L., Yan, J., & Shen, J. (2025). Integrated UAV and Satellite Remote Sensing: A Survey on Integration Strategies and Data Fusion Methods. *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, 8835–8840. <https://doi.org/10.1109/IGARSS55030.2025.11242522>
- Salgueiro Romero, L., Marcello, J., & Vilaplana, V. (2020). Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks. *Remote Sensing*, 12(15), 2424. <https://doi.org/10.3390/rs12152424>
- Sarmad, M., Kampffmeyer, M., & Salberg, A.-B. (2025). DiffFuSR: Super-Resolution of All Sentinel-2 Multispectral Bands Using Diffusion Models. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–13. <https://doi.org/10.1109/TGRS.2025.3638896>
- Sassu, A., Gambella, F., Ghiani, L., Mercenaro, L., Caria, M., & Pazzona, A. L. (2021). Advances in unmanned aerial system remote sensing for precision viticulture. In *Sensors (Switzerland)* (Vol. 21, Number 3, pp. 1–21). MDPI AG. <https://doi.org/10.3390/s21030956>
- Scarpa, G., & Ciotola, M. (2022). Full-Resolution Quality Assessment for Pansharpening. *Remote Sensing*, 14(8). <https://doi.org/10.3390/rs14081808>
- Shahdoosti, H. R. (2015, November 9). MS and PAN image fusion by combining Brovey and wavelet methods. *International Conference on Information Technology, Computer & Communication*.
- Shahdoosti, H. R., & Ghassemian, H. (2011). Multispectral and Panchromatic Image Fusion by Combining Spectral PCA and Spatial PCA Methods. *MODARES JOURNAL OF ELECTRICAL ENGINEERING*, 11(3).
- Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sensing*, 12(19), 1–31. <https://doi.org/10.3390/rs12193136>
- Song, Z., Zhang, Z., Yang, S., Ding, D., & Ning, J. (2020). Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Computers and Electronics in Agriculture*, 179. <https://doi.org/10.1016/j.compag.2020.105812>

- Tang, J., Han, W., & Zhang, L. (2019). UAV multispectral imagery combined with the FAO-56 dual approach for maize evapotranspiration mapping in the North China Plain. *Remote Sensing*, 11(21). <https://doi.org/10.3390/rs11212519>
- Tao, X., Gao, H., Liao, R., Wang, J., & Jia, J. (2017). Detail-Revealing Deep Video Super-Resolution. *2017 IEEE International Conference on Computer Vision (ICCV)*, 4482–4490. <https://doi.org/10.1109/ICCV.2017.479>
- Toosi, A., Samadzadegan, F., & Javan, F. D. (2026). S³-ESRGAN: Enhanced Super-Resolution Generative Adversarial Network for Remote Sensing Imagery Spatial Resolution Improvement—An Application Using Sentinel-2 and UAV Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19, 2149–2172. <https://doi.org/10.1109/JSTARS.2025.3640940>
- Wang, Q., Shi, W., Atkinson, P. M., & Zhao, Y. (2015). Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, 166, 191–204. <https://doi.org/10.1016/j.rse.2015.06.003>
- Wang, X., Chan, K. C. K., Yu, K., Dong, C., & Loy, C. C. (2019). EDVR: Video Restoration With Enhanced Deformable Convolutional Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1954–1963. <https://doi.org/10.1109/CVPRW.2019.00247>
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*. <http://arxiv.org/abs/1809.00219>
- Wang, Z., Tan, H., Wang, Z., Ci, J., & Zhai, H. (2025). MAF-GAN: A Multi-Attention Fusion Generative Adversarial Network for Remote Sensing Image Super-Resolution. *Remote Sensing*, 17(24), 3959. <https://doi.org/10.3390/rs17243959>
- Wei, Z., Huang, Y., Chen, Y., Zheng, C., & Gao, J. (2023). A-ESRGAN: Training Real-World Blind Super-Resolution with Attention U-Net Discriminators. In J. Siekmann (Ed.), *PRICAI 2023: Trends in Artificial Intelligence* (pp. 16–27). Springer. https://doi.org/https://doi.org/10.1007/978-981-99-7025-4_2
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236. <https://doi.org/10.1016/j.rse.2019.111402>
- Wu, T., Xu, W., & Wu, Y. (2025). A lightweight high-frequency mamba network for image super-resolution. *Scientific Reports*, 15(1), 25973. <https://doi.org/10.1038/s41598-025-11663-x>
- Xiao, J., Aggarwal, A. K., Rage, U. K., Katiyar, V., & Avtar, R. (2023). Deep Learning-Based Spatiotemporal Fusion of Unmanned Aerial Vehicle and Satellite Reflectance Images for Crop Monitoring. *IEEE Access*, 11, 85600–85614. <https://doi.org/10.1109/ACCESS.2023.3297513>
- Xing, Y., Liu, X., & Wang, X. (2026). Integrating UAVs, satellite remote sensing, and machine learning in precision agriculture: pathways to sustainable food production, resource efficiency, and scalable innovation. In *Frontiers in Agronomy* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fagro.2025.1670380>
- Xu, F., Ma, H., Sun, J., Wu, R., Liu, X., & Kong, Y. (2019). LSTM Multi-modal UNet for Brain Tumor Segmentation. *2019 IEEE 4th International Conference on Image*,

- Vision and Computing (ICIVC)*, 236–240. <https://doi.org/10.1109/ICIVC47709.2019.8981027>
- Zhang, J., Xu, T., Li, J., Jiang, S., & Zhang, Y. (2022). Single-Image Super Resolution of Remote Sensing Images with Real-World Degradation Modeling. *Remote Sensing*, 14(12). <https://doi.org/10.3390/rs14122895>
- Zhang, Y., Chen, Z., & Liu, S. (2020). Video Super Resolution Using Temporal Encoding ConvLSTM and Multi-Stage Fusion. *2020 IEEE International Conference on Visual Communications and Image Processing, VCIP 2020*, 298–301. <https://doi.org/10.1109/VCIP49819.2020.9301823>
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). *Image Super-Resolution Using Very Deep Residual Channel Attention Networks*. <http://arxiv.org/abs/1807.02758>
- Zhangzhong, L., Shen, Y., Li, J., Li, T., Wang, J., Zhu, Q., & Zheng, W. (2026). M2E2Net model: Multimodal fusion for real-time estimation of daily maize evapotranspiration. *Information Processing in Agriculture*. <https://doi.org/10.1016/j.inpa.2026.06.003>
- Zhao, L., Qing, S., Li, H., Qiu, Z., Niu, X., Shi, Y., Chen, S., & Xing, X. (2024). Estimating maize evapotranspiration based on hybrid back-propagation neural network models and meteorological, soil, and crop data. *International Journal of Biometeorology*, 68(3), 511–525. <https://doi.org/10.1007/s00484-023-02608-y>
- Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., Xu, J. L., & He, Y. (2024). Intelligent agriculture: deep learning in UAV-based remote sensing imagery for crop diseases and pests detection. In *Frontiers in Plant Science* (Vol. 15). Frontiers Media SA. <https://doi.org/10.3389/fpls.2024.1435016>
- Zurita-Milla, R., Clevers, J. G. P. W., & Schaepman, M. E. (2008). Unmixing-based landsat TM and MERIS FR data fusion. *IEEE Geoscience and Remote Sensing Letters*, 5(3), 453–457. <https://doi.org/10.1109/LGRS.2008.919685>

3 Fusión de imágenes UAV–Sentinel-2 usando redes neuronales convolucionales para análisis de teledetección*

Resumen

La teledetección ha consolidado el uso de imágenes satelitales para aplicaciones agrícolas como el monitoreo de cultivos, esto impulsa el desarrollo de enfoques de integración multisensor orientados hacia la mejora de la resolución espacial y la calidad radiométrica. En este estudio se implementó una metodología de fusión UAV–Sentinel-2 mediante composiciones espectrales basadas en siete índices de vegetación. Esta metodología permitió la evaluación de cuatro resoluciones espaciales objetivo (0.5, 1, 1.5 y 2 m) a través de tres modelos de superresolución: FSRCNN, VDSR y ESRGAN, entrenados de manera independiente para cuatro fechas de adquisición. El análisis multiescala permitió identificar 2 m como la resolución óptima de fusión, al maximizar simultáneamente métricas de fidelidad radiométrica y preservación estructural. A esta escala, VDSR presentó el mejor desempeño (RMSE = 0.041; PSNR = 33.83 dB; SSIM = 0.983; Corr = 0.981). El análisis de dispersión de las métricas entre bandas evidenció mayor uniformidad de reconstrucción a resoluciones gruesas, mientras que la evaluación entre fechas mostró alta consistencia de los tres modelos a lo largo de cuatro fechas de adquisición a 2m (CV<12.9% para el RMSE, < 4.33% para el PNSR, < 1.24% para el SSIM y < 1.25% para la correlación). En conjunto, los resultados validan la propuesta metodológica y subrayan la importancia de equilibrar mejoramiento espacial y coherencia espectral en procesos de fusión UAV–Sentinel-2 para aplicaciones agrícolas cuantitativas y monitoreo biofísico de precisión.

Palabras clave: CNN, Superresolución, Aprendizaje Profundo, Índices de Vegetación, Consistencia espectral.

Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Programa de Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: M.C. Alvaro Murguía Cozar.

Director de tesis: Dr. Gilberto de Jesús López Canteñs.

*Artículo publicado en la revista **INTERNATIONAL JOURNAL OF IMAGE AND DATA FUSION** en 10.1080/19479832.2026.2701715

Fusion of UAV.Sentinel.2 images using convolutional neural networks for remote-sensing analysis

Abstract

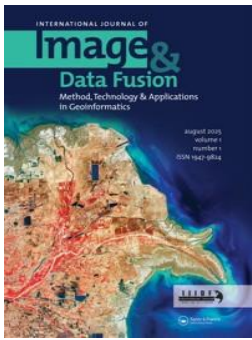
Remote sensing has consolidated the use of satellite imagery for agricultural applications such as crop monitoring, which drives the development of multisensor integration approaches aimed at improving spatial resolution and radiometric quality. In this study, a UAV–Sentinel-2 fusion methodology was implemented through spectral compositions based on seven vegetation indices. This methodology enabled the evaluation of four target spatial resolutions (0.5, 1, 1.5 and 2 m) through three super-resolution models: FSRCNN, VDSR, and ESRGAN, independently trained for four acquisition dates. Multiscale analysis allowed the identification of 2 m as the optimal fusion resolution, by simultaneously maximising radiometric fidelity and structural preservation metrics. At this scale, VDSR demonstrated the best performance (RMSE = 0.041; PSNR = 33.83 dB; SSIM = 0.983; Correlation = 0.981). The dispersion analysis of metrics across bands showed greater reconstruction uniformity at coarser resolutions, while the inter-dates evaluation revealed high consistency of the three models across four acquisition dates at 2 m (CV < 12.9% for RMSE, <4.33% for PSNR, <1.24%, for SSIM and <1.25% for correlation). Overall, the results validate the proposed methodological approach and highlight the importance of balancing spatial enhancement and spectral coherence in UAV–Sentinel-2 fusion processes for quantitative agricultural applications and precision biophysical monitoring.

Keywords: CNN; super-resolution; deep learning; vegetation indices; spectral consistency.

Doctoral Thesis. Graduate Program in Agricultural Engineering and Integrated Water Resources Management, Universidad Autónoma Chapingo.

Author: Msc. Alvaro Murguía Cozar.

Thesis director: Phd. Gilberto de Jesús López Canteñs.



Fusion of UAV–Sentinel-2 images using convolutional neural networks for remote-sensing analysis

Alvaro Murguia-Cozar , Gilberto de Jesús López-Canteñs , Antonia Macedo-Cruz , Noé Velazquez-López & Ronald Ernesto Ontiveros-Capurata

To cite this article: Alvaro Murguia-Cozar , Gilberto de Jesús López-Canteñs , Antonia Macedo-Cruz , Noé Velazquez-López & Ronald Ernesto Ontiveros-Capurata (2026) Fusion of UAV–Sentinel-2 images using convolutional neural networks for remote-sensing analysis, International Journal of Image and Data Fusion, 17:1, 2701715, DOI: [10.1080/19479832.2026.2701715](https://doi.org/10.1080/19479832.2026.2701715)

To link to this article: <https://doi.org/10.1080/19479832.2026.2701715>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 11 Jul 2026.



Submit your article to this journal [↗](#)



Article views: 509



View related articles [↗](#)



View Crossmark data



Fusion of UAV–Sentinel-2 images using convolutional neural networks for remote-sensing analysis

Alvaro Murguía-Cozar^a, Gilberto de Jesús López-Canteñs^a, Antonia Macedo-Cruz^b,
Noé Velazquez-López^a and Ronald Ernesto Ontiveros-Capurata^c

^aPosgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo, Chapingo, México; ^bPosgrado de Hidrociencias, Colegio de Postgraduados, Texcoco, México; ^cCoordinación de Seguridad Hídrica, Paseo Cuauhnáhuac, Instituto Mexicano de Tecnología del Agua, Jiutepec, México

ABSTRACT

Remote sensing has consolidated the use of satellite imagery for agricultural applications such as crop monitoring, which drives the development of multisensor integration approaches aimed at improving spatial resolution and radiometric quality. In this study, a UAV–Sentinel-2 fusion methodology was implemented through spectral compositions based on seven vegetation indices. This methodology enabled the evaluation of four target spatial resolutions (0.5, 1, 1.5 and 2 m) through three super-resolution models: FSRCNN, VDSR, and ESRGAN, independently trained for four acquisition dates. Multiscale analysis allowed the identification of 2 m as the optimal fusion resolution, by simultaneously maximising radiometric fidelity and structural preservation metrics. At this scale, VDSR demonstrated the best performance (RMSE = 0.041; PSNR = 33.83 dB; SSIM = 0.983; Correlation = 0.981). The dispersion analysis of metrics across bands showed greater reconstruction uniformity at coarser resolutions, while the inter-dates evaluation revealed high consistency of the three models across four acquisition dates at 2 m (CV < 12.9% for RMSE, <4.33% for PSNR, <1.24% for SSIM and <1.25% for correlation). Overall, the results validate the proposed methodological approach and highlight the importance of balancing spatial enhancement and spectral coherence in UAV–Sentinel-2 fusion processes for quantitative agricultural applications and precision biophysical monitoring.

ARTICLE HISTORY

Received 16 April 2026

Accepted 5 July 2026

KEYWORDS

CNN; super-resolution; deep learning; vegetation indices; spectral consistency

1. Introduction

Remote sensing is a fundamental tool for obtaining spatial information about the Earth's surface through the analysis of multispectral imagery, which enables continuous monitoring of changes in land cover. Its utility is reflected in recent applications such as the detection of bacterial blight in rice (Yuan *et al.* 2025), the assessment of regional suitability for cotton cultivation (He *et al.* 2025), the estimation of biophysical variables such as leaf area, aboveground biomass, and nitrogen concentration in soybean (Maimaitijiang *et al.* 2020), as well as soil salinity monitoring (Ma *et al.* 2020, 2022). Likewise, it has been used to calculate the leaf area index in maize (Yao *et al.* 2025) and to estimate biomass and canopy moisture content in mango plantations (Liu *et al.* 2025). These applications confirm that data obtained from remote sensors (particularly multispectral imagery) represent a reliable tool for crop monitoring and the optimisation of agricultural processes (Murugan *et al.* 2017).

The growing use of satellite imagery and unmanned aerial vehicles (UAVs) has further expanded these capabilities, revolutionising agricultural monitoring through the generation of more precise, continuous, and accessible data that enable a more detailed understanding of land-use changes (Lottering *et al.* 2025). As remote-sensing platforms and sensors increase, the amount, variety, and complexity of available data also expand, creating new challenges for data processing and analysis. Information derived from a single sensor is often insufficient to meet spatial, spectral, temporal, and radiometric resolution requirements in remote-sensing applications; therefore, multisensor data integration has been proposed to improve data quality and

CONTACT Gilberto de Jesús López-Canteñs alelopez10@hotmail.com Universidad Autónoma Chapingo, Posgrado en Ingeniería Agrícola y Uso Integral del Agua. Carretera México-Texcoco Km 38.5, Chapingo, C. P. 56230, México

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

continuity (Swain *et al.* 2024). To achieve this, it is essential that the different data sources be comparable and coherent with one another. However, satellite sensors and UAVs differ in their inherent acquisition system characteristics, which necessitates the application of calibration, normalisation, and harmonisation processes to ensure compatibility in joint analyses (Aasen *et al.* 2018).

Image fusion is an area of research that has been widely studied since the 1990s and refers to the process of integrating two images from different sources to enhance the information available in one of them. Fusion can be carried out at three processing levels: pixel level, feature level, and decision level. In pixel-level fusion, reflectance values of all bands are directly manipulated using different mathematical methods, which requires the images to be at least coregistered and, preferably, geocoded. Feature-level fusion involves the extraction of objects or patterns from the different data sources, for example through segmentation, vegetation indices, or textural measures. Finally, decision-level fusion is based on the integration of information previously extracted at the earlier levels; this information is combined by applying decision rules that reinforce the common interpretation and resolve discrepancies between sources, in order to obtain a more robust representation of the observed objects or phenomena (Pohl and Van Genderen 1998, Alvarez-Vanhard *et al.* 2021).

Traditional image fusion methods, widely used since early pansharpener studies, include techniques such as Intensity-Hue-Saturation (IHS), Principal Component Analysis (PCA), Brovey, and wavelet transforms. These approaches, such as IHS- and Brovey-based methods, tend to distort multispectral image information, adversely affecting the radiometric fidelity of the fused product. The PCA method can introduce spectral or spatial distortions depending on the statistical structure of the data (Shahdoosti and Ghassemian 2011). The Brovey method exhibits dependence on linear combinations, which can result in excessive injection of spatial details and unrealistic information into the low-resolution image (Shahdoosti 2015). Wavelet-based methods are not translation invariant due to subsampling operators, which can generate artefacts and inconsistencies in the fusion process (Shahdoosti 2015). Reported correlation metrics for the quality evaluation of enhanced images obtained with these types of methods range between 0.74 and 0.85, which represent low levels of data fusion performance (Shahdoosti and Ghassemian 2011, Shahdoosti 2015).

More flexible and robust methods have been presented, such as the Gram-Schmidt (GS) method based on Gram-Schmidt transformations to combine UAV and Sentinel-2 images and discriminate 10 land-cover classes, increasing discrimination accuracy by 11% to 16% compared with Sentinel-2 alone (Zhao *et al.* 2019); feature-based fusion techniques (SIFT) employed by Song *et al.* (2020) for the identification of lodging in sunflower, wheat, and rice crops; and advanced geostatistical approaches, such as Area-to-Point Regression Kriging (ATPRK) proposed by Wang *et al.* (2015) as an alternative for enhancing the spatial resolution of MODIS, Sentinel-2, and Landsat-8 imagery.

In the last decade, the development of deep learning has transformed the field of image fusion. Convolutional neural networks (CNNs) have demonstrated superior capability to capture complex spatial and spectral relationships, as well as to model nonlinear interactions among bands. CNN-based techniques have contributed to image super-resolution by enabling the generation of more accurate and efficient reconstructions compared to traditional methods. Dong *et al.* (2016) implemented the first Super-Resolution Convolutional Neural Network (SRCNN), which enhances a low-resolution image through three stages: feature extraction, nonlinear mapping, and image reconstruction. In the same year, Dong *et al.* (2016) developed the Faster Super-Resolution Convolutional Neural Network (FSRCNN) model as a redesign of SRCNN by removing the initial bicubic interpolation and dividing the nonlinear mapping into three stages: shrinking, mapping, and expanding; furthermore, in the reconstruction stage, a deconvolution layer was implemented to increase the spatial resolution of the processed image.

The Very Deep Super-Resolution (VDSR) approach introduced the use of very deep networks together with global residual learning, which facilitates directly modelling the missing details between low- and high-resolution images, accelerating convergence, and improving reconstruction quality compared to previous models such as SRCNN (Kim *et al.* 2016). Zhang *et al.* (2018) proposed the Residual Channel Attention Network (RCAN) as an adjustment to the VDSR model through a Residual-in-Residual (RIR) structure to prioritise high-frequency information by means of channel attention modules. The Dense Channel Attention Network (DCAN), unlike RCAN, implements a dense learning scheme to ensure the persistence of original information and adds a spatial attention component, which allows filtering which information is relevant and localising where such relevance resides in the image space (Ma *et al.* 2021).

A more recent model, the High-Frequency Mamba Network (HFMN), implemented self-attention mechanisms to integrate local and global information in order to balance the importance that the model assigns to each type of information; HFMN includes a local feature extraction module, a global feature extraction module, and a dual-attention fusion module (Wu *et al.* 2025).

The first Generative Adversarial Network (GAN) model formally applied to super-resolution was SRGAN (Ledig *et al.* 2017), which introduced a generator-discriminator scheme combined with a perceptual loss based on the Visual Geometry Group (VGG) to produce realistic reconstructions from low-resolution images. Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) consolidated the use of adversarial learning to achieve realistic super-resolution through a generator based on Residual-in-Residual Dense Blocks, a relativistic discriminator, and an improved perceptual loss using features extracted from VGG before activation (Wang *et al.* 2018).

These architectures demonstrate how deep networks, residual learning, and adversarial strategies drive super-resolution performance towards more accurate, efficient, and visually realistic results. A-ESRGAN is a generative adversarial network with a multiscale discriminator based on a U-Net structure composed of an encoding module with downsampling, a decoding module with upsampling, and several attention blocks (Wei *et al.* 2023). Zhang *et al.* (2022) designed MSCGAN, a GAN network with joint multistream architecture and spectral compensation for pansharpening. The methodology extracts structural information from the panchromatic (PAN) image, which has high spatial resolution, and introduces it into the generator together with the previously rescaled low-resolution multispectral (LRMS) image. Through multistream convolutional architectures, structural and spectral features are separately extracted and integrated at the feature level. Likewise, spectral compensation blocks allow the injection of LRMS spectral information into the generation process, reducing spectral distortion and improving the fidelity of the generated high-resolution multispectral image.

Modern multisource fusion approaches explicitly integrate spatial-spectral interactions with the aim of injecting high-resolution spatial details without compromising the spectral consistency of the original image. HyperNet, proposed by Li *et al.* (2022), employs multiscale attention enhancement blocks that analyse the image at multiple scales and adaptively select the most relevant features through spatial and spectral attention mechanisms. Similarly, P2Sharpen, introduced by Zhang *et al.* (2023), incorporates a Spectral Transformation Network (STNet), which learns a local nonlinear transformation between multispectral and panchromatic images, thereby improving the coherence between both data sources.

The Super-Resolution-Guided Progressive Pansharpening Neural Network (SRPPNN) proposes a progressive framework in which super-resolution is first applied to low-resolution multispectral images to extract intrinsic spatial details. These features are combined with high-frequency information from the panchromatic image through a progressive pansharpening process that incrementally increases the resolution across multiple stages. Finally, a high-frequency residual module is incorporated to inject fine spatial details derived from the PAN image (Cai and Huang 2021).

In addition, the work by Peng *et al.* (2021) addresses the problem from a more comprehensive perspective by jointly modelling spatial, spectral, and temporal domains. This approach represents the data as a four-dimensional tensor (width, height, spectral, and temporal), which is decomposed into dictionaries and a sparse core tensor. The interaction between spatial and spectral dimensions is captured by the coefficients of the core tensor, allowing for a more flexible representation of the relationships among different information sources.

Despite considerable progress in image enhancement and fusion techniques, existing UAV-Satellite fusion strategies still present notable limitations. Traditional interpolation-based methods, such as bicubic or bilinear interpolation, are computationally simple but often produce overly smooth images and fail to recover high-frequency spatial details. Deep learning-based super-resolution approaches, including CNN- and GAN-based models, improve visual quality but may introduce hallucinated textures and do not guarantee spectral fidelity, which is critical in remote-sensing applications. Spatial-spectral fusion frameworks have been developed to exploit complementary information from multisource data (PAN, multispectral, and hyperspectral images). While these methods better preserve spectral characteristics, they often require complex architectures and are not specifically tailored to UAV-satellite integration scenarios.

Therefore, there remains a need for fusion strategies that can effectively enhance spatial resolution while preserving meaningful spectral information in a physically consistent manner. This motivates the proposed.

vegetation index-based fusion framework, which aims to leverage domain-specific spectral characteristics to improve fusion quality and ensure better consistency for vegetation analysis.

In this context, the present study proposes a new multispectral image fusion methodology that combines information from Sentinel-2 satellite sensors and UAV platforms, using exclusively vegetation indices organised into band compositions. This approach makes it possible to preserve the radiometric resolution of Sentinel-2 images while incorporating high-fidelity spatial and spectral patterns captured by UAV sensors, thus avoiding the traditional absolute radiometric calibration processes required for direct reflectance images.

This research demonstrates that feature-level fusion between Sentinel-2 images and UAV images, based on vegetation indices, significantly enhances spatial detail while maintaining the quantitative consistency of the data. The proposed methodology opens new opportunities for the development of high-precision agricultural monitoring products through the efficient combination of heterogeneous data sources.

2. Methodology

2.1. Study area

The present research was conducted at the ‘San Juan’ experimental field of the Universidad Autónoma Chapingo (UACH), located at 19°29′32″ N latitude and 98°51′32″ W longitude, at an elevation of 2,211 metres above mean sea level. The experimental field comprises 143 ha of agricultural land use, where different research projects related to maize, oats, and horticultural crops are established. The soil is characterised as sandy loam, with organic matter contents below 1% and depths ranging from 0.3 to 0.4 metres. The plots are distributed into sections known as ‘tabla’ each ‘tabla’ is composed of between 17 and 25 agricultural plots. In the present research, 42 agricultural plots (49 ha) were selected, including 12 uncultivated plots (11 ha), 5 plots with hawthorn (9 ha), and 25 plots with maize (29 ha). [Figure 1](#) shows the geographical location and spatial delineation of the study area corresponding to the ‘San Juan’ experimental field of the UACH.

2.2. Multispectral image acquisition

2.2.1. Sentinel-2 satellite imagery

Sentinel-2 satellite images corresponding to four specific dates were downloaded: 9 July, 3 August, 13 August, and 2 September 2025. The images were obtained from the official Copernicus program portal of the European Space Agency (ESA): <https://browser.dataspace.copernicus.eu>. The red (B04, 665 nm), green (B03, 560 nm), red-edge (B06, 740 nm), and near-infrared (B08A, 864 nm) bands were used, all at processing level L2A; therefore, no atmospheric correction was required for the images.

2.2.2. UAV imagery

2.2.2.1. Ground control point for georeferencing. To improve the spatial accuracy of the generated products and ensure proper georeferencing of the flights, a ground control point was established within the study area using a Trimble R8 GNSS RTK receiver. The equipment was operated in static post-processing mode with an observation time of 5 hours. This point was linked to the ICMX station of the Red Nacional Activa (RNA), managed by the Instituto Nacional de Estadística y Geografía (INEGI), which constitutes the official reference framework in Mexico, based on ITRF 2008 (epoch 2010). Through a differential post-processing procedure, the precise position of the control point was obtained by applying trilateration techniques using Emlid Studio software.

2.2.2.2. UAV flights. Flights were conducted using the DJI Mavic 3 M multispectral UAV on the same dates as the Sentinel-2 imagery acquisitions. During the flights, a D-RTK 2 antenna was used to receive real-time corrections and ensure high-precision georeferencing. The D-RTK 2 antenna was installed directly over the established ground control point and configured with the calculated coordinates, thereby ensuring accurate spatial georeferencing of the flights. The flights were executed at an altitude of 120 m, with 70% forward and side overlap, and altitude optimisation based on the digital elevation model of the area, which was preloaded into the aircraft control system. The captured data were processed using Pix4D Mapper software,

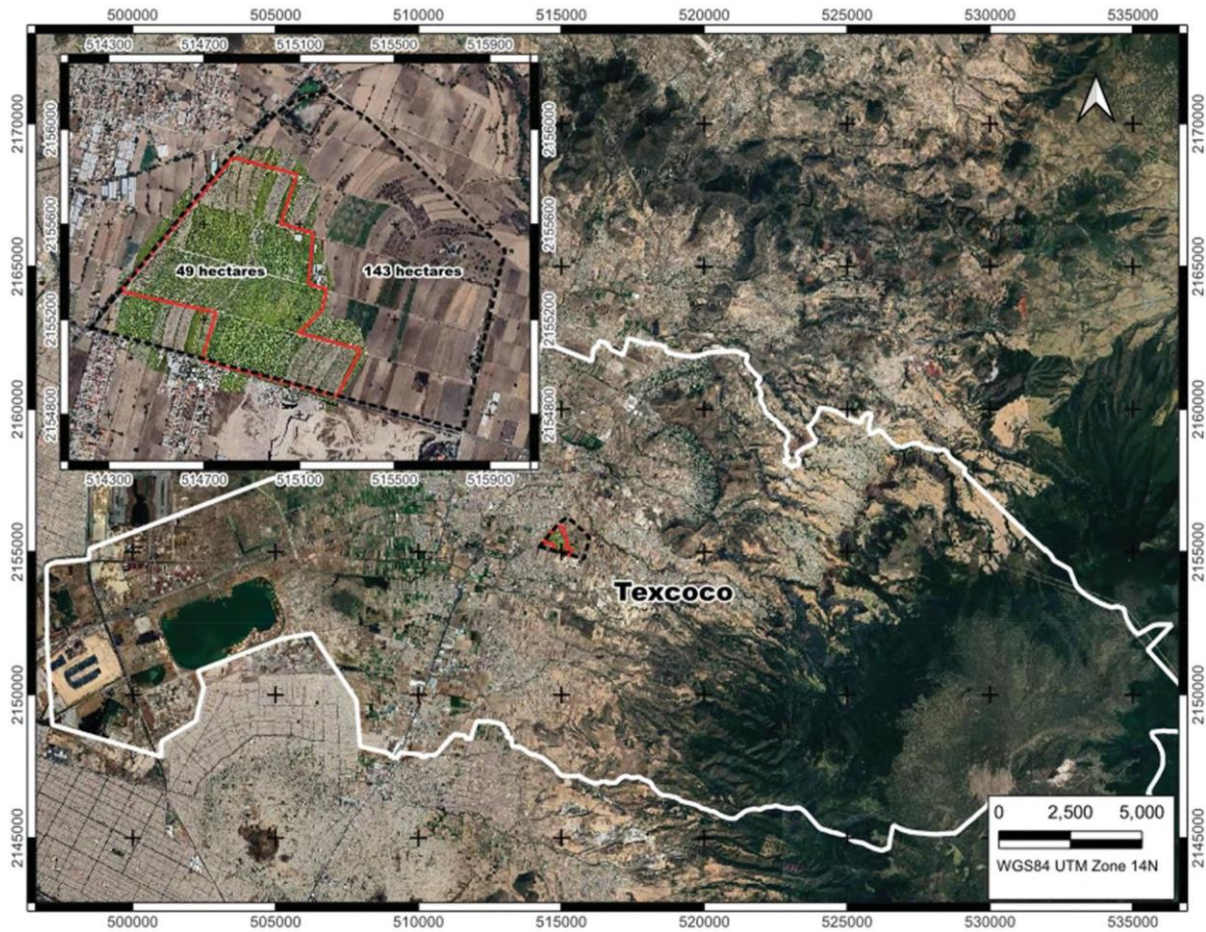


Figure 1. San Juan experimental field of the Universidad Autónoma Chapingo.

through which reflectance maps for the red (650 nm), green (560 nm), red-edge (740 nm), and near-infrared (860 nm) bands were generated with a spatial resolution of 6 cm.

2.2.3. Data normalisation

The obtained data, consisting of spectral bands from Sentinel-2 images and UAV reflectance maps, were normalised using the min-max method. Transformation (1) was applied to ensure the homogenisation of scales among variables.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x is the original value, $x_{\min}$ is the minimum value of the analysed band, $x_{\max}$ is the maximum value of the analysed band, and x_{norm} is the normalised value.

2.2.4. Generation of vegetation indices

From these normalised bands, multispectral compositions were generated consisting of seven vegetation indices: Normalised Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Green Normalised Difference Vegetation Index (GNDVI), Normalised Difference Red Edge (NDRE), Normalised Difference Water Index (NDWI), Red-Edge Normalised Difference Vegetation Index (NDVI_{re}), and Red-Edge Chlorophyll Index (CI_{re}), estimated using Equations 2-8.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

$$SAVI = \frac{NIR - RED}{NIR + RED + L} (1 + L) \quad (3)$$

$$GNDVI = \frac{NIR - GREEN}{NIR + GREEN} \quad (4)$$

$$NDRE = \frac{NIR - RE}{NIR + RE} \quad (5)$$

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (6)$$

$$NDVIRE = \frac{RE - RED}{RE + RED} \quad (7)$$

$$Clre = \frac{NIR}{RE} - 1 \quad (8)$$

Finally, the compositions generated from both satellite data and UAV-acquired data were resampled to 0.5, 1.0, 1.5, and 2 metres to form band compositions by date and data source. Upsampling and downsampling processes were applied to homogenise spatial resolution. For the UAV images, a downsampling process was applied using the mean of the pixels contained within the analysis filter. For the Sentinel-2 compositions, an upsampling process was applied using bicubic interpolation. Through this process, the spectral bands with original resolutions of 20 m and 10 m were resampled to resolutions of 0.5, 1, 1.5, and 2 m. Finally, the compositions were clipped to the size of the study area, using the spatial extent shown in [Figure 1](#) as a reference.

2.3. Super-resolution models

2.3.1. FSRCNN model

The FSRCNN model is a convolutional neural network specifically designed to perform image super-resolution efficiently, rapidly, and with high reconstruction quality. It was proposed as an accelerated and improved version of the SRCNN model, maintaining or exceeding its performance while drastically reducing its computational cost (Dong *et al.* 2016).

FSRCNN presents an ‘hourglass-type’ architecture, characterised by a larger number of features at the beginning and end and a narrower central region. The network is composed of five main modules: feature extraction, shrinking, deep nonlinear mapping, expanding, and deconvolution or upsampling.

In the feature extraction layer, the network begins by applying a convolution $\text{Conv}(f, d, c)$ directly to the low-resolution (LR) image, where f is the filter size, d is the number of filters, and c is the number of channels. This layer transforms each input patch into a feature vector of dimension d , using a smaller filter size (5×5). For the present case study, the structure proposed by the author was preserved, with only the number of channels adapted to $\text{Conv}(5, 56, 7)$.

During dimensionality reduction, FSRCNN introduces a $\text{Conv}(1, s, d)$ layer capable of reducing the feature dimensionality from $d = 56$ to $s = 12$, thereby decreasing the computational load without significant information loss. The nonlinear mapping stage consists of $m = 4$ convolutional layers of size 3×3 , all with s filters, which enables the construction of a deep and efficient mapping in the reduced feature space. The expanding stage is the inverse process of shrinking, using a $\text{Conv}(1, d, s)$ layer symmetric to the reduction stage. Deconvolution for upsampling applies to $\text{DeConv}(9, 1, d)$ layer with a stride equal to the desired scaling factor. For the present study, the scaling factor employed was $\text{stride} = 1$, as upsampling was applied to the LR images prior to their input into the model.

2.3.2. VDSR Model

The VDSR model, proposed by Kim *et al.* (2016), is based on a very deep convolutional neural network specifically designed to address the image super-resolution problem with high accuracy. The model

introduces substantial improvements over methods such as SRCNN, both in performance and convergence speed, due to its depth, residual-learning strategy, and multiscale treatment.

The VDSR architecture consists of 20 convolutional layers with 3×3 filters, inspired by VGG networks used for image recognition. All intermediate layers use 64 filters of size 3×3 and ReLU activation functions, except for the final layer, which reconstructs the residual image using a single 3×3 filter. The architecture operates on a low-resolution image that has been previously interpolated to the target resolution. The model does not generate the high-resolution image directly; instead, it predicts the residual between the high-resolution (HR) image and the enhanced version. The residual is added to the interpolated image to produce the final super-resolved image. This allows the network to focus on reconstructing high-frequency details and textures, while interpolation accounts for the low-frequency components.

2.3.3. ESRGAN model

The ESRGAN model constitutes an improved version of SRGAN (Ledig *et al.* 2017) and is designed to produce super-resolved images with significantly more realistic and detailed textures. Its advances are based on three fundamental improvements. First, it introduces the Residual-in-Residual Dense Block (RRDB), a deep block that combines multilevel residual learning and dense connections without using Batch Normalisation, which increases the representational capacity of the generator and avoids the artefacts observed in SRGAN, especially in deep networks or those trained under adversarial settings. Second, it replaces the traditional discriminator with a relativistic discriminator (RaGAN) that does not evaluate whether an image is real or fake in absolute terms, but rather whether it is 'more realistic' than another, providing more informative gradients and promoting the recovery of natural edges and textures. Third, it redefines the perceptual loss by using VGG features before activation, which corrects brightness issues, avoids the extreme dispersion present in activated features, and provides stronger supervision for recovering fine textures and realistic details. In addition, ESRGAN incorporates a network-interpolation strategy that allows combining a PSNR-oriented model with a perception-oriented model to control the balance between sharpness, naturalness, and artefact suppression without the need to retrain the model (Wang *et al.* 2018).

2.4. Model training and inference

All experiments were conducted on a high-performance workstation running Microsoft Windows 11 Pro for Workstations (64-bit). The system was equipped with an Intel Core i9-12900F processor featuring 16 physical cores and 24 logical threads, and 128 GB of DDR5 RAM, enabling efficient handling of large-scale remote-sensing datasets and deep-learning training processes. GPU acceleration was provided by an NVIDIA GeForce RTX 4090 with 24 GB of dedicated memory, which was used for both model training and inference. The GPU operated with NVIDIA driver version 591.86. This configuration allowed for efficient parallel processing and significantly reduced training times across the evaluated super-resolution models.

The FSRCNN, VDSR, and ESRGAN models were implemented in PyTorch using Python 3.10.19, following the original architectures provided by the authors of each model (Dong *et al.* 2016, Kim *et al.* 2016, Wang *et al.* 2018). Hyperparameters were manually tuned to achieve the best performance for each architecture. In the case of VDSR and FSRCNN, both models shared the same hyperparameters, which are presented below:

- Patch size = 96
- Number of patches per image = 500
- Batch size = 16
- Epochs = 1000
- Learning rate = 0.001

The ESRGAN model used different hyperparameters, which are presented below:

- Patch size = 96
- Number of patches per image = 300
- Batch size = 8
- Epochs = 200
- Learning rate = 0.0001

To ensure reproducibility and methodological clarity, the experimental setup is described in detail. The dataset was divided into training and validation subsets following an 80-20% split. To avoid spatial leakage, training and testing samples were extracted from non-overlapping regions. All models were trained using the Adam optimiser with a learning rate of 0.0001 and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The VDSR and FSRCNN models were optimised using MSE loss function, while the ESRGAN model employed a combination of content loss L1 and adversarial loss.

The models were trained for the four available spatial resolutions, using exclusively the acquisition date of 13 August 2025, to identify the optimal scale for performing data fusion. The acquisition was selected due to its representativeness in terms of vegetation conditions and data consistency. On this date, more than 80% of the study area was covered by well-developed vegetation, providing suitable spatial patterns for evaluating super-resolution performance under realistic agricultural conditions. Additionally, the UAV data acquisition was conducted under conditions that ensured temporal proximity with the Sentinel-2 overpass (~17:00 local time), minimising radiometric inconsistencies between datasets. Although weather conditions during the 2025 rainy season limited the availability of synchronised UAV flights on other dates, this acquisition provided the most suitable balance between vegetation coverage, illumination conditions, and temporal alignment. The spatial resolution that yielded the best model performance, according to the quality metrics obtained from the enhanced images, was selected.

Subsequently, the three models were retrained using the remaining available dates, but exclusively with images resampled to the selected resolution. This second stage aimed to validate the temporal consistency of the models, that is, to verify that the observed performance was not the result of conditions specific to August 13 (illumination, phenological stage, or atmospheric conditions) but rather reflected a robust and reproducible behaviour under varying local conditions. This additional verification ensures that the models maintain their generalisation capability and that the selected resolution is suitable for multitemporal UAV-Sentinel-2 fusion throughout the entire study period.

2.5. Evaluation metrics

Radiometric metrics are those that evaluate radiometric similarity between a reference image and an evaluated image by directly comparing intensity values (DN or reflectance) pixel by pixel and band by band (Wald 2002, Zhou *et al.* 2004). Global spectral metrics are those that assess the spectral fidelity of an image by simultaneously considering all spectral bands; that is, they analyse the complete spectral vector of each pixel and generate a global quality indicator (Vivone *et al.* 2021). Four radiometric evaluation metrics were applied per spectral band: Root Mean Square Error, Peak Signal-to-Noise Ratio, Structural Similarity Index Measure, and Correlation.

2.5.1. Root mean square error (RMSE)

RMSE is a statistical measure used, in the context of image fusion, to evaluate how faithfully a method reproduces the original spectral information in each band. RMSE is easy to interpret, has a clear scale (smaller values indicate greater similarity between images), and effectively summarises the average error incurred during the synthesis process (Wald 2002). Equation (9) was used to estimate this metric:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (9)$$

Where x_i is the reference image (UAV), y_i is the low-resolution image (enhanced image), i is the analysed pixel, and n is the number of pixels per spectral band.

2.5.2. Peak signal-to-noise ratio (PSNR)

Although it originated as an indicator in analogue systems, PSNR has become a standard tool for comparing the quality of image restoration, compression, and fusion algorithms (Lin and Huang 2008, Liu *et al.* 2008). PSNR quantifies the ratio between the maximum possible signal and the level of noise introduced by a transformation process. Because it is based on the MSE, it allows a direct evaluation of the error introduced

at each pixel. Equation (10) was used to determine the PSNR values of the images generated by the employed super-resolution models.

$$PSNR = 10 \log \left(\frac{\max_i I_i^2}{MSE} \right) \quad (10)$$

Where $\max_i I_i$ is the maximum possible pixel value of the enhanced image and $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ is the mean squared error between the reference image (UAV) and the enhanced image

2.5.3. Structural similarity index measure (SSIM)

The Structural Similarity Index Measure evaluates the degree of similarity between two images based on how the human visual system (HVS) perceives structural changes. SSIM assesses image degradation by considering luminance, chromaticity, and structure. The metric compares the average brightness between the two images, the variation of intensities, and the local correlation between them. SSIM was determined using Equation (11).

$$SSIM(x, y) = \frac{\left((2 \cdot \mu_x \cdot \mu_y + C_1) \cdot (2 \cdot \sigma_{xy} + C_2) \right)}{\left((\mu_x + \mu_y + C_1) \cdot (\sigma_x + \sigma_y + C_2) \right)} \quad (11)$$

Where μ_x is the mean reflectance value of the reference image (UAV) x , μ_y is the mean reflectance value of the enhanced image y , σ_x is the variance of the pixels of the reference image (UAV) x , σ_y is the variance of the pixels of the enhanced image y , σ_{xy} is the covariance between pixels of the reference and enhanced images. The constants are defined as $C_1 = (K_1 \cdot L)$, $C_2 = (K_2 \cdot L)$ con $K_1 = 0.01$ y $K_2 = 0.03$ y $L = 2$ as the dynamic range of the image, since the vegetation indices used are normalised from -1 to 1.

2.5.4. Correlation

In image fusion, correlation reveals whether spatial variations and patterns from the reference image are preserved in the enhanced image. In multispectral image fusion, correlation can be used to determine whether linear relationships among bands are maintained after fusion, which is critical for remote sensing applications. Equation (12) was applied to assess the quality of the enhanced images.

$$corr = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (12)$$

Where x is the reference image, y is the enhanced image, $\bar{x}$ is the mean of the pixels in the reference image, $\bar{y}$ is the mean of the pixels in the enhanced image, i is the analysed pixel, and n is the total number of pixels contained in the images.

2.5.5. Determination of evaluation metrics

The enhanced images generated by the models were compared with their respective reference images. This comparison was performed using image quality metrics (RMSE, SSIM, PSNR, and Correlation) to quantify the degree of similarity between the super-resolved image and the original reference, and thus objectively evaluate the performance of each super-resolution model. The employed metrics allowed the analysis of several aspects of the enhancement, such as the preservation of fine details, structural similarity, radiometric fidelity, and the reduction of pixel-by-pixel errors.

3. Results

3.1. Multiscale performance analysis

The results obtained from the training of the super-resolution models (Table 1) reveal differential behaviour across architectures and spatial resolutions. Overall, VDSR exhibited the best performance across all metrics, systematically improving as the target resolution approached 2 metres, where it achieved its highest

Table 1. Evaluation metrics obtained from model inference by spatial resolution.

Resolution	Model	RMSE ↓	PSNR (dB) ↑	SSIM ↑	Corr ↑
0.5 m	FSRCNN	0.121	24.74	0.818	0.820
	VDSR	0.114	24.57	0.842	0.869
	ESRGAN	0.123	18.17	0.814	0.816
1 m	FSRCNN	0.106	25.28	0.864	0.862
	VDSR	0.068	29.05	0.946	0.943
	ESRGAN	0.125	17.69	0.743	0.817
1.5 m	FSRCNN	0.098	26.00	0.900	0.898
	VDSR	0.052	31.66	0.974	0.972
	ESRGAN	0.076	22.85	0.944	0.943
2 m	FSRCNN	0.088	27.24	0.923	0.920
	VDSR	0.041	33.83	0.983	0.981
	ESRGAN	0.061	24.93	0.963	0.963

The models were trained using images corresponding to 13 August 2025.

performance with an RMSE of 0.041, a PSNR of 33.83 dB, an SSIM of 0.983, and a correlation of 0.981, positioning it as the model with the greatest capacity to preserve structural and radiometric details.

The FSRCNN model, although consistently inferior to VDSR, also presented progressive improvements as the resolution increased, achieving its best performance at 2 m, with an SSIM of 0.923 and a PSNR of 27.24 dB, which evidences a positive response of the model at intermediate resolutions. In contrast, ESRGAN exhibited a different trend: its performance was more stable at medium resolutions, achieving higher metric values from 1.5 m onward and presenting its best behaviour at 2 m, with a PSNR of 24.93 dB and an SSIM of 0.963, reflecting the ability of its GAN-based architecture to reconstruct textures in a more realistic manner at less demanding scales.

Figure 2 presents a visual comparison of the reference images (UAV), the interpolated Sentinel-2 images, and the super-resolved versions generated using FSRCNN, VDSR, and ESRGAN at four scale levels. In general, it can be observed that all three models improve the blurred appearance exhibited by interpolated Sentinel-2 imagery, achieving better recovery of edges and transitions between plots.

When compared with the UAV reference, VDSR is the method that most clearly reproduces object shapes and crop boundaries (Figure 2(d,g,j,m)). Its results show sharper edges and less distortion at all scales, especially from 1.0 m onward, where image sharpness improves consistently.

FSRCNN also improves upon interpolation, although its enhancement is more moderate (Figure 2(c,f,i,l)). The image appears more stable and freer of exaggerated effects; however, contrast and texture definition are lower than those produced by VDSR. Nevertheless, at 2 m it provides sufficiently good visual quality and is close to the reference (Figure 2(a)).

In the case of ESRGAN, a clear increase in texture and local contrast is observed, making the image appear more detailed (Figure 2(e,h,k,n)). However, in some areas, it may generate patterns or textures that are not necessarily present in the reference, particularly when the target scale is very fine (0.5 m). While this may be useful for visual analysis, it should be interpreted with caution for quantitative applications.

In terms of scale, the four levels show that: 0.5 m is the most demanding resolution, as no model fully reaches the level of detail of the UAV imagery (Figure 2(a)); 1.0 m and 1.5 m exhibit the greatest visual gain, with improved shape recovery and reduced noise; and 2.0 m is the scale at which all three methods behave most consistently, also showing the best preservation in terms of inter-date consistency.

Overall, the comparison shows that all models provide visible improvements over interpolated Sentinel-2 imagery; however, VDSR achieves the greatest recovery of detail and spatial continuity (Figure 2(m,g,j) and Figure 3(d)), followed by FSRCNN (Figures 2(l,f,i) and 3(c)), while ESRGAN (Figures 2(n,h,k) and 3(e)) offers a perceptual enhancement that may introduce additional textures.

This superiority of the 2 m scale is explained by its ability to reduce spatial noise in the images without losing spectral representativeness, while simultaneously facilitating the learning of relevant patterns in heterogeneous agricultural regions. In addition, this resolution represents an appropriate compromise between spatial detail and radiometric consistency, which is essential to ensure that super-resolution models generalise correctly and produce robust results across different points in the temporal series.

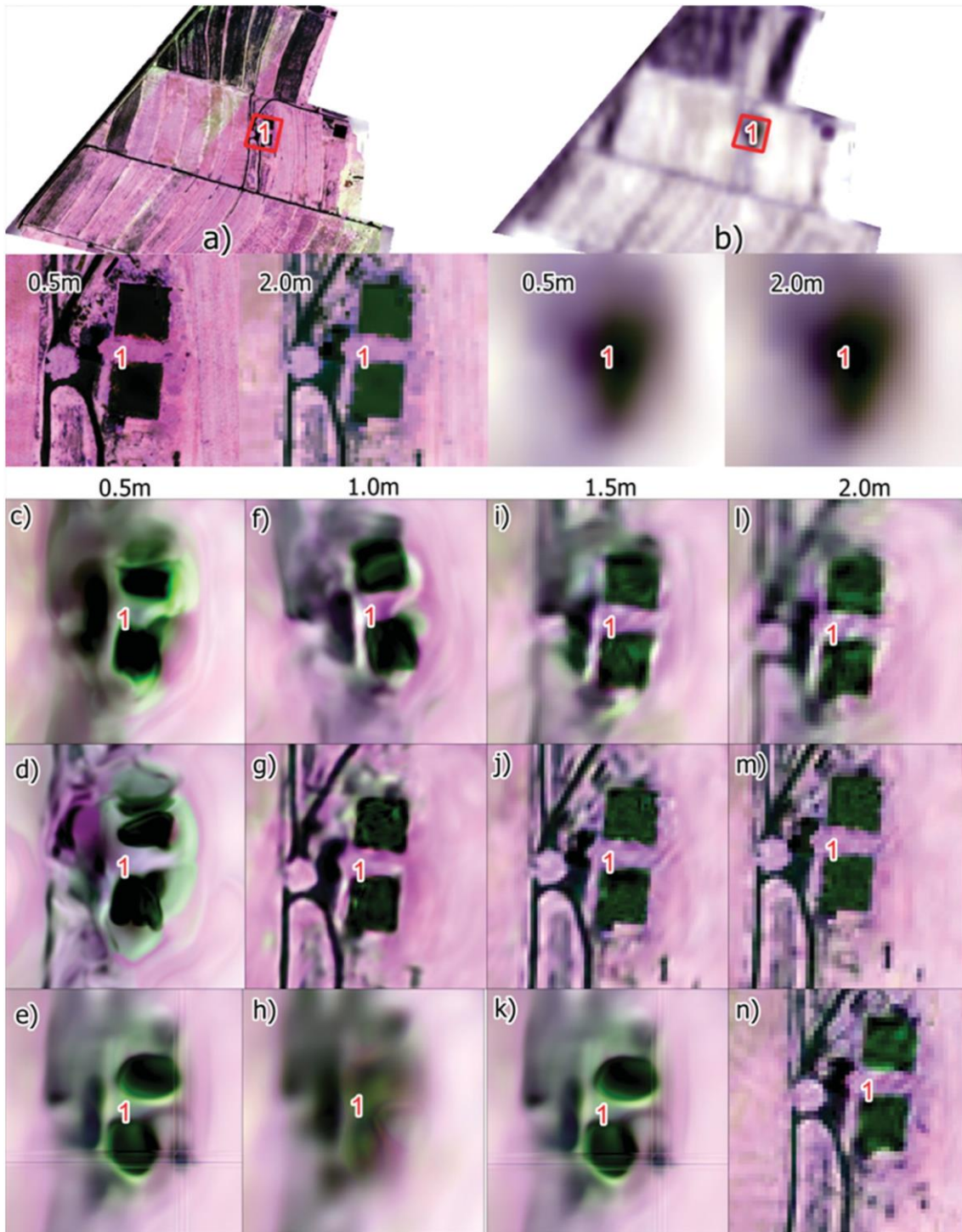


Figure 2. Multiscale visual comparison of Sentinel-2 super-resolution results using FSRCNN, VDSR, and ESRGAN with UAV reference imagery. a) UAV image with two views resampled to 0.5 m and 2.0 m; b) Sentinel-2 image with two views resampled to 0.5 m and 2.0 m; c), f), i) and l) FSRCNN images; d), g), j) and m) VDSR images; e), h), k) and n) ESRGAN images.

3.2. Spectral index behaviour

Figure 4 shows the evolution of the SSIM metric for the three super-resolution models applied to different vegetation indices derived from UAV and Sentinel-2 bands on 13 August 2025. The trend observed in the three graphs indicates that SSIM increases as the target spatial resolution becomes coarser; that is, the models exhibit greater stability and spectral coherence when generating products at 1.5 m and 2.0 m.

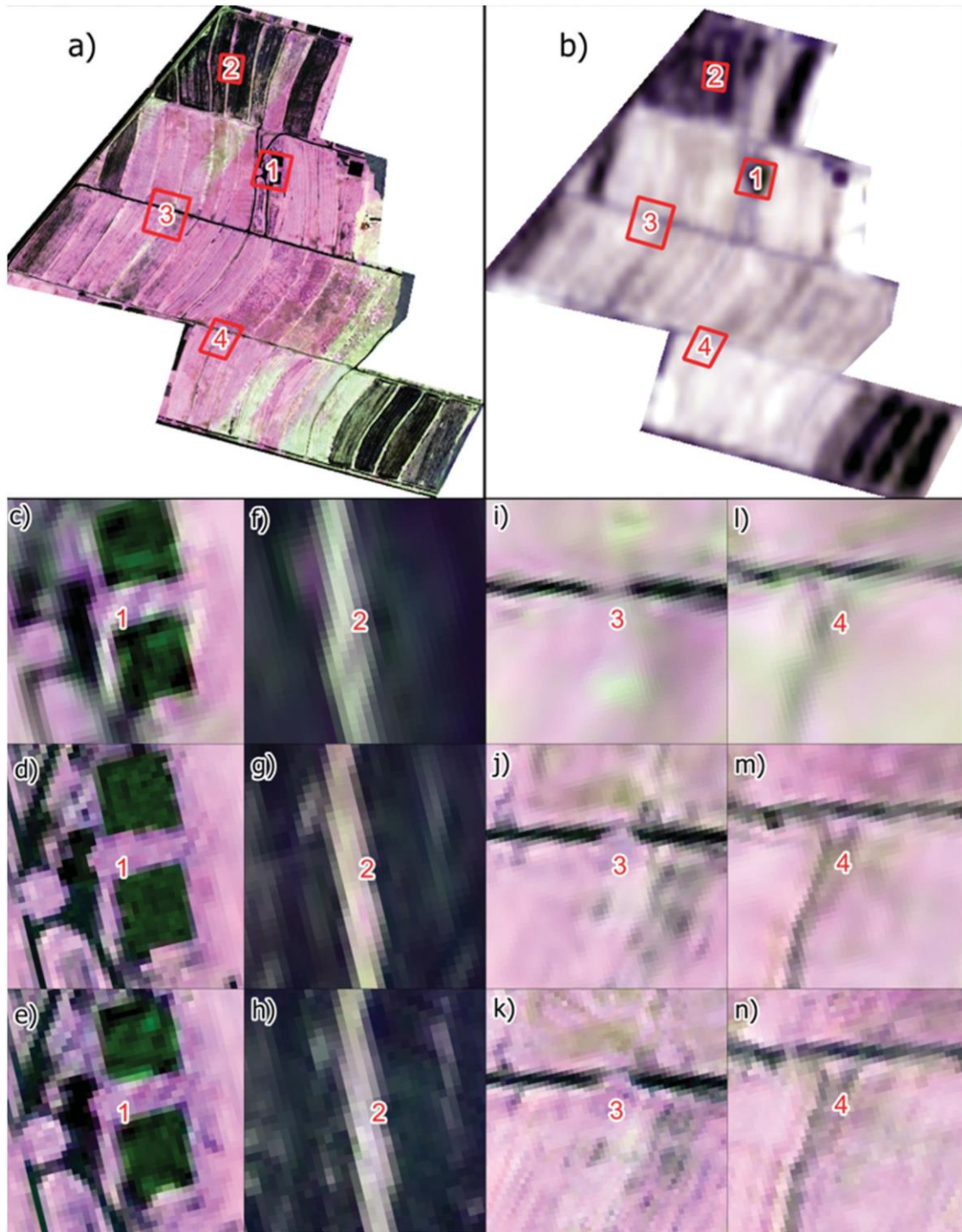


Figure 3. Visual comparison of super-resolution reconstruction and local detail enhancement in UAV–Sentinel-2 fusion. c), f), i), and l) correspond to FSRCNN results; d), g), j), and m) correspond to VDSR results; and e), h), k), and n) correspond to ESRGAN results.

The FSRCNN model (Figure 4(a)) exhibited a progressive improvement in SSIM from 0.5 m to 2.0 m for all evaluated indices. The NDVI, NDVI_{re}, NDWI, and GNDVI indices show the highest and most consistent values, while SAVI, NDRE, and Clre present lower and less consistent values. This suggests that FSRCNN preserves vegetation-related structural information well, although its spectral stability is more pronounced at intermediate and coarse resolutions.

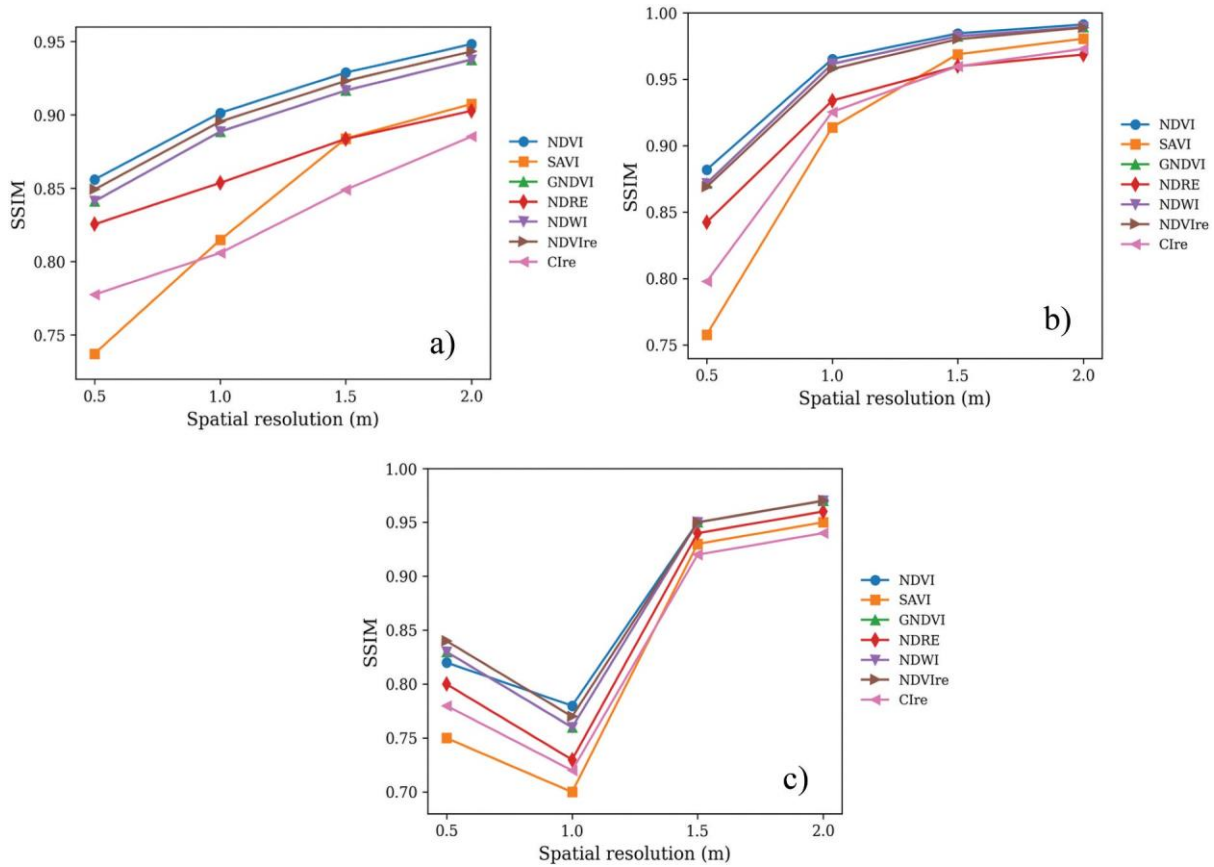


Figure 4. Evolution of the SSIM metric obtained from model inference for (a) FSRCNN, (b) VDSR, and (c) ESRGAN for each spatial resolution and spectral band on 13 August 2025.

The graph corresponding to VDSR (Figure 4(b)) showed the most robust behaviour among the three models. SSIM reaches values close to 1.0 at resolutions of 1.5 m and 2.0 m for all indices, reflecting very high structural similarity with respect to the UAV reference. Even at 0.5 m and 1.0 m, VDSR maintains values higher than those of FSRCNN and ESRGAN, confirming its ability to reconstruct details without compromising spectral coherence. This behaviour is consistent with the global metrics, where VDSR obtained the best PSNR and SSIM values at 2 m across the four evaluated dates.

In contrast, ESRGAN (Figure 4(c)) exhibits a different pattern, showing a notable decrease in SSIM at the 1.0 m resolution, with values below those obtained by FSRCNN and VDSR. However, from 1.5 m and 2.0 m onward, its performance stabilises and the values converge towards levels comparable to those of FSRCNN and VDSR. This behaviour suggests that ESRGAN, being GAN-based, tends to generate additional textures that affect structural similarity when the target resolution is too fine for the available spectral information, although this effect diminishes when working at coarser resolutions.

Table 2 summarises the metrics dispersion among seven vegetation indices for each model and spatial resolution using the SSIM, RMSE, PSNR, and correlation metrics for August 13. Overall, it is observed that dispersion across vegetation indices decreases as the spatial resolution becomes coarser, indicating that the models produce more consistent responses among indices when operating at 1.5 m and 2 m.

In the case of FSRCNN, both the standard deviation and the coefficient of variation show a decreasing trend from 0.5 m to 2 m, suggesting a progressive improvement in model stability across indices. For VDSR, dispersion is lower than observed for FSRCNN and ESRGAN in nearly all metrics, particularly for SSIM and correlation, where the coefficient of variation falls below 1-2% at 2 m, making it the most consistent model across resolutions. In contrast, ESRGAN exhibits marked instability at 1 m, especially in SSIM (CV = 20%) and PSNR (CV = 17.9%); however, it stabilises at 1.5 m and 2 m, with dispersion values comparable to those of the other models.

Table 2. Dispersion metrics across vegetation indices for each model and spatial resolution (13 August).

Model	Resolution	SSIM		RMSE		PSNR		COR R	
		SD (%)	CV	SD (%)	CV	SD (%)	CV	SD	CV (%)
FSRCNN	0.5 m	0.044	5.40	0.031	24.4	2.87	12.1	0.040	4.9
FSRCNN	1 m	0.039	4.61	0.025	21.5	2.75	10.9	0.034	4.0
FSRCNN	1.5 m	0.029	3.22	0.022	18.9	2.68	10.3	0.028	3.1
FSRCNN	2 m	0.024	2.64	0.021	18.3	2.63	9.7	0.026	2.8
VDSR	0.5 m	0.047	5.56	0.030	22.1	2.91	11.9	0.041	4.7
VDSR	1 m	0.021	2.20	0.019	23.5	2.59	8.9	0.019	2.0
VDSR	1.5 m	0.011	1.12	0.014	22.3	2.41	7.6	0.012	1.2
VDSR	2 m	0.009	0.92	0.007	17.1	2.18	6.4	0.011	1.1
ESRGAN	0.5 m	0.046	5.46	0.035	24.9	2.78	13.9	0.040	4.9
ESRGAN	1 m	0.163	20.0	0.034	23.8	3.18	17.9	0.037	4.6
ESRGAN	1.5 m	0.015	1.60	0.018	22.6	2.64	12.3	0.014	1.5
ESRGAN	2 m	0.013	1.35	0.016	22.9	2.62	10.8	0.013	1.4

SD Standard Deviation, CV Coefficient of Variation.

3.3. Training stability across dates at 2 m

Table 3 summarises the mean, standard deviation, and coefficient of variation of the RMSE, PSNR, SSIM, and correlation metrics for the three models evaluated at a spatial resolution of 2 m. The results show that VDSR is the model with the best performance and greatest stability, while FSRCNN presents intermediate performance and ESRGAN provides a notable perceptual enhancement but with lower metrics relative to VDSR.

In terms of error, VDSR achieved the lowest RMSE (0.0387 ± 0.00495), indicating greater radiometric proximity to the reference. FSRCNN exhibits a higher RMSE (0.0891 ± 0.0113), and ESRGAN falls between the two models (0.0601 ± 0.0066). The coefficients of variation (CV $\approx$ 11–13%) reveal that the relative stability of RMSE is similar across models, although VDSR maintains more favourable absolute values.

With respect to PSNR, VDSR reaches the highest value (34.35 ± 1.16 dB), confirming its ability to preserve the signal with lower distortion. FSRCNN achieves moderate levels (27.49 ± 1.19 dB), whereas ESRGAN presents the lowest value (24.19 ± 0.90 dB), suggesting a greater presence of noise or radiometric alterations in its reconstruction. Nevertheless, PSNR CVs (3–4%) indicate that all three models are stable for this metric.

For SSIM, which reflects structural similarity, VDSR again stands out with the highest value (0.986 ± 0.0057 , CV = 0.58%). ESRGAN also exhibits high SSIM (0.964 ± 0.0017 , CV = 0.18%), although lower than that of VDSR; FSRCNN attains a lower value (0.920 ± 0.0113) but remains stable (CV = 1.23%). This trend indicates that VDSR better preserves spatial structure, whereas ESRGAN is highly consistent, albeit not as structurally accurate as VDSR.

Finally, for correlation, the same patterns are maintained: VDSR shows the highest value (0.985 ± 0.0056), followed by ESRGAN (0.963 ± 0.0017) and FSRCNN (0.919 ± 0.0115). The CVs indicate that VDSR and ESRGAN are the most stable (<1.3%), while FSRCNN exhibits greater relative variability (1.25%). Likewise, although the metrics of VDSR are better than those of ESRGAN, it should be noted that the variability of ESRGAN results is lower than that of VDSR, indicating that the ESRGAN model exhibited greater consistency in the enhanced images than VDSR.

Overall, the results demonstrate that VDSR is the most reliable and stable model at 2 m, with the best balance between low error, high structural similarity, and high correlation. FSRCNN maintains solid but more limited performance, whereas ESRGAN, although very stable in terms of SSIM and correlation, exhibits a relatively lower PSNR, suggesting that its perceptual enhancement does not always correspond to a real radiometric improvement.

Table 3. Comparative performance of super-resolution models across different dates.

Model	RMSE (Mean $\pm$ SD)	CV RMSE (%)	PSNR (Mean $\pm$ SD)		SSIM (Mean $\pm$ SD)		Corr (Mean $\pm$ SD)	
			CV PSNR (%)		CV SSIM (%)		CV Corr (%)	
FSRCNN	0.0891 ± 0.0113	12.7	27.49 ± 1.19	4.32	0.920 ± 0.0113	1.23	0.919 ± 0.0115	1.25
VDSR	0.0387 ± 0.00495	12.8	34.35 ± 1.16	3.37	0.986 ± 0.0057	0.58	0.985 ± 0.0056	0.57
ESRGAN	0.0601 ± 0.0066	10.9	24.19 ± 0.90	3.72	0.964 ± 0.0017	0.18	0.963 ± 0.0017	0.18

The three super-resolution models were evaluated using four acquisition dates: July 9, August 3, August 13, and September 2 2025.

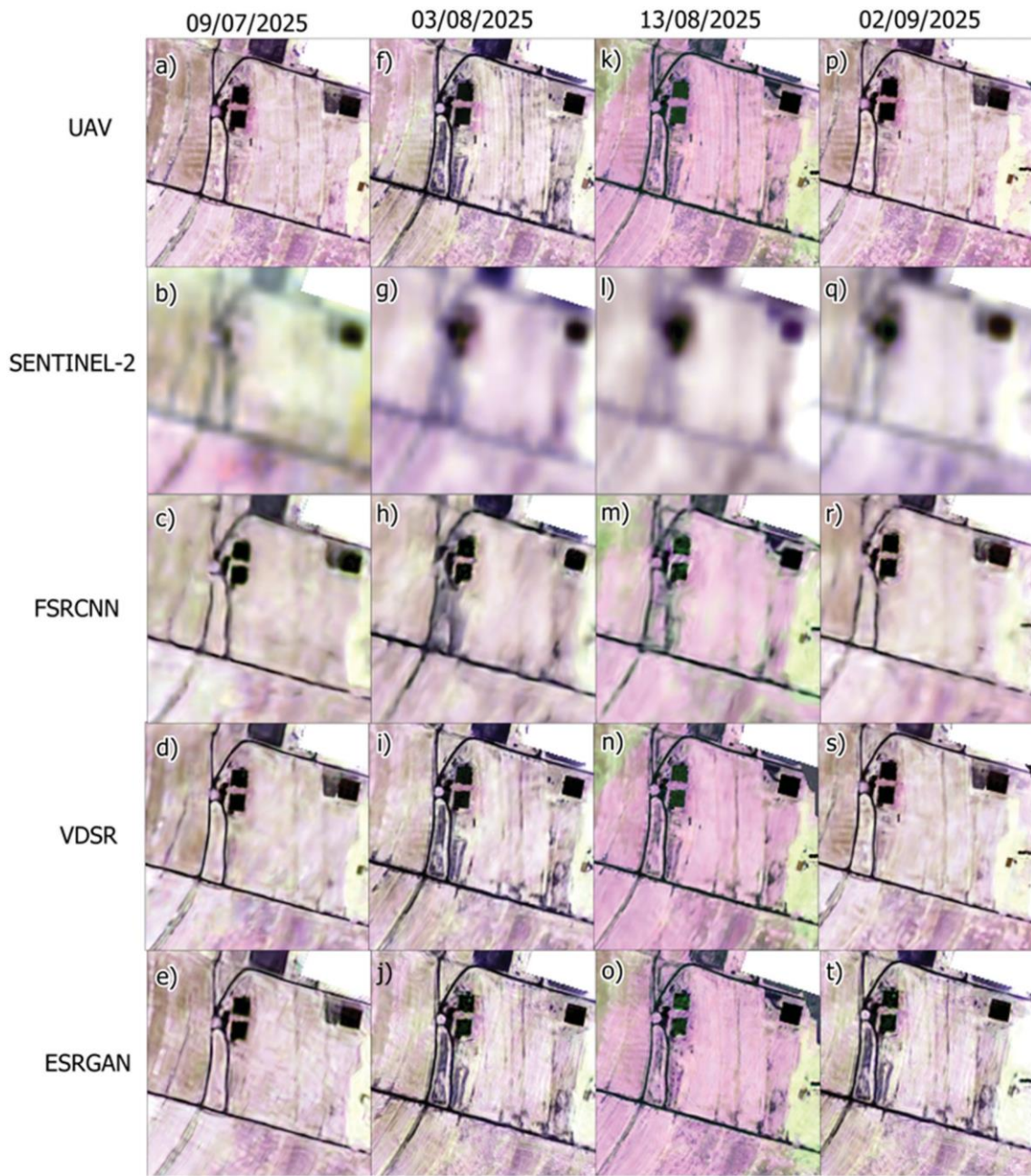


Figure 5. Visual evaluation of super-resolved satellite images at 2 m.

Figure 5 evaluates the consistency of the three models across four acquisition dates at 2 m with respect to the UAV reference image. It can be observed that VDSR maintains the highest visual coherence across dates, consistently reproducing plot boundaries, shadows, and land-cover transitions without introducing strong temporal inconsistencies.

However, it is important to note that the models were trained independently for each acquisition date using synchronised UAV imagery. Therefore, the observed consistency should not be interpreted as temporal generalisation, but rather as stable performance under similar conditions across dates. This distinction is particularly relevant when assessing robustness, as the proposed approach does not explicitly model temporal dependencies. Instead, it demonstrates that each model configuration can produce consistent outputs when trained and evaluated under comparable acquisition settings.

The FSRCNN model provided a stable improvement over interpolation, with reasonably well-defined contours and a low risk of artefacts. Although its level of detail is lower than that of VDSR, its response does

not fluctuate across dates, making it predictable for time series and therefore useful when consistency is prioritised over maximum enhancement.

In the ESRGAN model, a perceptual enhancement that increases micro-contrast is observed; however, this enhancement is not uniform across dates, and in certain scenes the magnitude of textures varies (for example, the excerpts illustrate slight differences in apparent grain/noise between July 9 and August 3).

3.4. Computational efficiency

Table 4 summarises the model parameters, training and inference times, and architectural characteristics of the evaluated approaches. The results reveal differences in computational cost among the models. While FSRCNN represents the most lightweight approach, with only 20,035 parameters, the GAN-based model (ESRGAN) increases this complexity, exceeding 13 million parameters. This increase is reflected in its higher computational requirements, resulting in the longest training time (6596.11 s) and inference time (0.98 s). Notably, the VDSR model achieves an efficient training process (403.08 s) despite having a denser residual architecture than FSRCNN. In addition, VDSR exhibits the fastest inference performance, with a processing time of 0.45 s per image, making it a favourable trade-off between computational efficiency and model complexity.

4. Discussion

The fusion of Sentinel-2 and UAV imagery is a complex process that depends on multiple interacting factors. The obtained results highlight three particularly determining aspects: (i) the adopted fusion level, (ii) the spatial resolution at which data integration is performed, and (iii) the spectral and structural consistency of the fused information.

The fusion level of Sentinel-2 and UAV images is closely related to the type of sensor acquiring the data; therefore, when integration is performed at the reflectance value level, the application of absolute radiometric calibration processes using calibration panels is recommended to ensure data coherence (Padró *et al.* 2018, Naethe *et al.* 2023, Swaminathan *et al.* 2024). In contrast, the approach proposed in this study, based on feature-level fusion using vegetation indices, although it does not incorporate absolute radiometric calibration with panels, achieved high structural similarity values ($SSIM > 0.85$), indicating robust preservation of spatial patterns and relative spectral relationships.

A conceptually similar approach was adopted by Zhang *et al.* (2025), who fused Sentinel-2 and UAV imagery at the feature level (NDVI) without applying absolute radiometric calibration to the UAV images, achieving a reduction of approximately 40% in the Mean Absolute Percentage Error (MAPE). This suggests that, for applications oriented towards super-resolution and spectral index analysis, the absence of panel-based calibration does not prevent the generation of products with high structural and spectral quality.

Determining the optimal spatial resolution for data integration is a key factor within the workflow, as it defines the boundary between detail enhancement and signal degradation. This limitation is inherently related to the native spatial resolution of Sentinel-2 imagery, which constrains the amount of spatial information available for reconstruction. The results suggest that fusion should not be forced towards excessively fine scales, since aggressive increases in spatial resolution tend to induce structural instability and radiometric artefacts. Conventional interpolation processes, which only resample the original signal without adding new spatial information, further illustrate this limitation by producing smoothed representations and blurred edges. This phenomenon occurs when the super-resolution model attempts to synthesise information at a scale where the coherence between the coarse Sentinel-2 pixel and the fine UAV detail breaks down, resulting in greater dispersion of spectral indices and a loss of crop biophysical fidelity.

Table 4. Computational complexity and efficiency of the evaluated super-resolution models.

Model	Parameters	Training Time (s)	Inference Time (s)	Architecture Type
FSRCNN	20035	4557.65	0.49	Lightweight CNN
VDSR	672839	403.08	0.45	Deep Residual CNN
ESRGAN	13348720	6596.11	0.98	Generative Adversarial

This finding becomes particularly relevant when contrasted with recent studies such as Toosi *et al.* (2025), who, using a metaheuristic approach, determined optimal integration resolutions ranging between 2.03 m and 2.35 m. Other authors have reported improvements in Sentinel-2 imagery down to resolutions of 2 m (Salgueiro Romero *et al.* 2020), 2.5 m (Galar *et al.* 2020, Major *et al.* 2025) and 5 m (Kramer *et al.* 2024). The convergence among these values and the 2 m resolution identified in this study suggests that this scale is not arbitrary but rather represents a level of compatibility at which high-resolution spatial information and Sentinel-2 spectral content reach an optimal balance.

In contrast, the present study demonstrates that the 2 m resolution is not only consistent with the geometric relationship between sensors, but also jointly optimises PSNR and SSIM while reducing inter-index variability. This result is consistent with the findings of Kramer *et al.* (2024) and Shao *et al.* (2019), who emphasise that increases in spatial resolution must preserve spectral coherence for quantitative applications. The correlation greater than 0.98 achieved by VDSR at 2 m confirms that the global spectral structure is adequately preserved, which is a fundamental aspect for vegetation index derivation and subsequent biophysical analyses.

In contrast to conventional single image super-resolution models such as FSRCNN, VDSR, and ESRGAN, which primarily focus on recovering spatial details from a single low-resolution input, recent multisource fusion frameworks explicitly model the interaction between spatial and spectral features. These methods leverage complementary information from multiple sensors and incorporate specialised mechanisms such as channel spatial attention (Li *et al.* 2022), spectral transformation networks (Zhang *et al.* 2023), progressive fusion strategies (Cai and Huang 2021) and semi-coupled tensor factorisation (Peng *et al.* 2021) to ensure both spatial enhancement and spectral fidelity. The spatial spectral cross-fusion framework proposed by Fu *et al.* (2025) further advances this paradigm by introducing dedicated cross-attention mechanisms, including spectral cross attention and spatial cross attention blocks, to explicitly fuse multiresolution hyperspectral, multispectral, and panchromatic data. These designs enable more effective integration of complementary spatial and spectral information, leading to improved reconstruction accuracy and stronger spectral consistency.

Obtaining an enhanced image with high sharpness and an optimal visual appearance is not sufficient. In remote sensing applications, the utility of the final product lies in the spectral and structural consistency of the fused information, which ensures that the data faithfully represent the biophysical reality in the field. The inter-date analysis exhibited high stability in performance metrics for the three evaluated models. In terms of SSIM, the coefficients of variation were below 1.3% in all cases; ESRGAN stood out as the model with the lowest relative variability (CV = 0.18%), followed by VDSR (CV = 0.58%) and FSRCNN (CV = 1.23%).

This robustness in the metrics demonstrates that the models maintain consistent performance over time when trained independently for each date. However, such consistency does not imply intrinsic temporal generalisation capability, as the proposed approach requires a contemporaneous reference pair (UAV and Sentinel-2 acquired on the same date). Therefore, the value of these results lies in the fidelity with which the model transfers UAV spatial structure without altering the spectral identity of the satellite, a condition that is indispensable for multitemporal crop monitoring.

Multisource remote sensing data fusion has yielded substantial improvements in vegetation analysis, with quantitative metrics such as PSNR and SSIM increasingly serving as bridges between methodological rigour and real-world application. For instance, spatiotemporal fusion frameworks that integrate MODIS and Landsat imagery achieve PSNR values above 35 dB and SSIM exceeding 0.90, directly translating into reliable leaf area index time series that capture rapid phenological transitions in croplands and grasslands (Yang *et al.* 2020, Schreier *et al.* 2021). Similarly, pansharpening algorithms optimised for vegetation indices produce fused images with PSNR > 38 dB and spectral fidelity SSIM > 0.95 sufficient to estimate canopy chlorophyll content with R^2 above 0.85 across multiple crop types (Qian *et al.* 2022). The synergistic fusion of SAR and optical data, yielding correlation coefficients (R) above 0.80 between predicted and measured aboveground biomass, demonstrates that high fusion quality measured through reduced RMSE and improved spatial detail preservation enables operational monitoring of carbon stocks in heterogeneous forest landscapes (Zeng *et al.* 2020, Padalia *et al.* 2023). The use of multispectral and hyperspectral imagery combined with super-resolution models incorporating attention mechanisms enabled an improvement of approximately 5% in marsh vegetation classification accuracy, yielding super-resolved images with quality metrics of PSNR = 28.87 and SSIM = 0.76 (Yao *et al.* 2025). Furthermore, the integration of hyperspectral imagery with LiDAR data increased classification accuracy in karst wetlands by up to 9.5% compared to single-source datasets (Fu *et al.* 2023). These quantitative benchmarks confirm that robust fusion quality (PSNR > 33 dB, SSIM > 0.88) is

a necessary precondition for downstream vegetation products to meet the accuracy thresholds required by precision agriculture and ecological management. Although the present study focuses on reconstruction metrics, these findings suggest that the proposed framework is expected to provide similar improvements in downstream vegetation applications due to its ability to preserve spatial detail and spectral consistency.

A key limitation of the evaluated models is that they are based on a single-image super-resolution paradigm. As a result, they rely on the availability of high-resolution reference data, such as UAV imagery, to effectively enhance the spatial resolution of Sentinel-2 images. This dependency limits their applicability in scenarios where high-resolution auxiliary data are not available or cannot be acquired synchronously. In contrast, multi-image or multi-temporal super-resolution approaches offer a more flexible framework by exploiting temporal redundancy and complementary information across multiple acquisitions. These methods can enhance robustness and reduce reliance on external high-resolution sources by leveraging inherent temporal dynamics within the dataset (Tarasiewicz *et al.* 2023, Toosi *et al.* 2026). Therefore, while single-image super-resolution models provide strong reconstruction capabilities under controlled conditions, their dependence on high-resolution reference imagery represents a significant limitation compared to multi-image approaches that enable more scalable and temporally robust solutions.

To extend the framework towards a truly temporal super-resolution scheme, it is necessary to incorporate architectures capable of modelling spatiotemporal dependencies, such as recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), or temporal attention mechanisms. Such structures would enable the capture of persistent patterns in spatial and spectral dynamics over time, facilitating the reconstruction of new acquisitions without necessarily requiring a synchronised UAV image. Consequently, future research should focus on integrating spatiotemporal components into super-resolution architectures, with the aim of advancing towards a robust multitemporal enhancement scheme for specific regions, while maintaining the balance between structural accuracy and spectral consistency demonstrated in the present study.

5. Conclusions

The use of deep learning-based super-resolution models enables the effective integration of UAV and Sentinel-2 data for remote sensing applications. The feature-level fusion approach, using vegetation indices, ensures the generation of products with a spatial resolution of 2 m that preserve the radiometric and structural integrity of the original scene. This level of enhancement represents a scaling of satellite information by a factor of 5× without compromising the fidelity of spectral signatures or data coherence.

The effectiveness of feature-level fusion lies in its ability to stabilise vegetation index responses under variations in illumination and in the absence of absolute calibration. The low dispersion observed in index values derived from images generated with the VDSR model at a 2 m resolution ensures that the produced information is technically valid for the estimation of biophysical variables.

The stability of the metrics obtained from the inter-date analysis indicates that the proposed methodology provides consistent performance when models are trained under comparable acquisition conditions. Consequently, the proposed workflow offers a reliable tool for phenological interpretation and supports agronomic decision-making based on high spatial fidelity data, although its application to fully continuous high-resolution time series remains subject to the availability of reference data for each acquisition.

4 Author contributions

CRediT: **Alvaro Murguía-Cozar**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft; **Gilberto de Jesús López-Canteñs**: Conceptualization, Supervision, Writing – review & editing; **Antonia Macedo-Cruz**: Conceptualization, Supervision, Writing – review & editing; **Noé Velázquez-López**: Writing – review & editing; **Ronald Ernesto Ontiveros-Capurata**: Writing – review & editing.

5 Disclosure statement

No potential conflict of interest was reported by the author(s).

6 References

- Aasen, H., *et al.*, 2018. Quantitative Remote Sensing at ultra-high resolution with UAV spectroscopy: a review of sensor technology, measurement procedures, and data correction workflows. *Remote Sensing*, 10 (7), 1091. doi:[10.3390/rs10071091](https://doi.org/10.3390/rs10071091)
- Alvarez-Vanhard, E., Corpetti, T., and Houet, T., 2021. UAV & satellite synergies for optical remote sensing applications: a literature review. *Science of Remote Sensing*, 3, 100019. doi:[10.1016/j.srs.2021.100019](https://doi.org/10.1016/j.srs.2021.100019)
- Cai, J. and Huang, B., 2021. Super-resolution-guided progressive pansharpening based on a deep convolutional neural network. *IEEE Transactions on Geoscience & Remote Sensing*, 59 (6), 5206–5220. doi:[10.1109/TGRS.2020.3015878](https://doi.org/10.1109/TGRS.2020.3015878)
- Dong, C., *et al.*, 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 38 (2), 295–307. doi:[10.1109/TPAMI.2015.2439281](https://doi.org/10.1109/TPAMI.2015.2439281)
- Dong, C., Loy, C.C., and Tang, X., 2016. Accelerating the super-resolution convolutional neural network Computer Vision – ECCV 2016 9906 Cham Leibe B Matas J Sebe N Welling M (Springer)391–407 doi:[10.1007/978-3-319-46475-6_25](https://doi.org/10.1007/978-3-319-46475-6_25).
- Fu, B., *et al.*, 2023. Combination of super-resolution reconstruction and SGA-Net for marsh vegetation mapping using multi-resolution multispectral and hyperspectral images. *International Journal of Digital Earth*, 16 (1), 2724–2761. doi:[10.1080/17538947.2023.2234340](https://doi.org/10.1080/17538947.2023.2234340)
- Fu, B., *et al.*, 2025. GFHMP: gradual fusion framework of hyperspectral, multispectral, and panchromatic images using a novel spatial-spectral cross-fusion network. *IEEE Transactions on Geoscience and Remote Sensing* 63. doi:[10.1109/TGRS.2025.3599189](https://doi.org/10.1109/TGRS.2025.3599189).
- Galar, M., *et al.*, 2020. LEARNING SUPER-RESOLUTION FOR SENTINEL-2 IMAGES WITH REAL GROUND TRUTH DATA FROM a REFERENCE SATELLITE. *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, V-1–2020 (1), 9–16. doi:[10.5194/isprs-annals-V-1-2020-9-2020](https://doi.org/10.5194/isprs-annals-V-1-2020-9-2020)
- He, J., *et al.*, 2025. Regional-scale precision mapping of cotton suitability using UAV and satellite data in arid environments. *Agricultural Water Management*, 307, 109215. doi:[10.1016/j.agwat.2024.109215](https://doi.org/10.1016/j.agwat.2024.109215)
- Kim, J., Lee, J.K., and Lee, K.M., 2016. Accurate image super-resolution using very deep convolutional networks IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 1646–1654. doi:[10.1109/CVPR.2016.182](https://doi.org/10.1109/CVPR.2016.182) Las Vegas.
- Kramer, P., Steinhardt, A., and Pedretsch, B., 2024. Enhancing Sentinel-2 image resolution: evaluating advanced techniques based on convolutional and Generative neural networks arXiv [Preprint]. <http://arxiv.org/abs/2410.00516> [Accessed 8 April 2026].
- Ledig, C. *et al.*, 2017. Photo-realistic single image super-resolution using a Generative adversarial network IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017 105–114 doi:[10.1109/CVPR.2017.19](https://doi.org/10.1109/CVPR.2017.19) Honolulu.
- Li, K., *et al.*, 2022. HyperNet: a deep network for hyperspectral, multispectral, and panchromatic image fusion. *ISPRS Journal of Photogrammetry & Remote Sensing*, 188, 30–44. doi:[10.1016/j.isprsjprs.2022.04.001](https://doi.org/10.1016/j.isprsjprs.2022.04.001)
- Lin, P.-L. and Huang, P.-Y., 2008. Fusion methods based on dynamic-segmented morphological wavelet or cut and paste for multifocus images. *Signal Processing*, 88 (6), 1511–1527. doi:[10.1016/j.sigpro.2007.12.018](https://doi.org/10.1016/j.sigpro.2007.12.018)
- Liu, J., *et al.*, 2025. Assessment of mango canopy Water content through the fusion of multispectral unmanned aerial vehicle (UAV) and Sentinel-2 Remote Sensing data. *Forests*, 16 (1), 167. doi:[10.3390/f16010167](https://doi.org/10.3390/f16010167)
- Liu, Z., Forsyth, D.S., and Laganère, R., 2008. A feature-based metric for the quantitative evaluation of pixel-level image fusion. *Computer Vision and Image Understanding*, 109 (1), 56–68. doi:[10.1016/j.cviu.2007.04.003](https://doi.org/10.1016/j.cviu.2007.04.003)
- Lottering, R., Peerbhay, K., and Adelabu, S., 2025. Remote Sensing applications in agricultural, Earth and environmental Sciences. *Applied Sciences*, 15 (8), 4537. doi:[10.3390/app15084537](https://doi.org/10.3390/app15084537)
- Ma, Y., *et al.*, 2020. Spectral index fusion for salinized soil salinity inversion using Sentinel-2A and UAV images in a coastal area. *IEEE Access*, 8, 159595–159608. doi:[10.1109/ACCESS.2020.3020325](https://doi.org/10.1109/ACCESS.2020.3020325)
- Ma, Y., *et al.*, 2021. Remote sensing image super-resolution based on dense channel attention network. *Remote Sensing*, 13 (15), 2966. doi:[10.3390/rs13152966](https://doi.org/10.3390/rs13152966)
- Ma, Y., *et al.*, 2022. Fusion level of satellite and UAV image data for soil salinity inversion in the coastal area of the Yellow River Delta. *International Journal of Remote Sensing*, 43 (19–24), 7039–7063. doi:[10.1080/01431161.2022.2155080](https://doi.org/10.1080/01431161.2022.2155080)
- Maimaitijiang, M., *et al.*, 2020. Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sensing*, 12 (9), 1357. doi:[10.3390/rs12091357](https://doi.org/10.3390/rs12091357)
- Major, D., *et al.*, 2025. A holistic approach for multi-spectral Sentinel-2 super-resolution and spectral evaluation. *International Journal of Remote Sensing*, 46 (20), 7437–7464. doi:[10.1080/01431161.2025.2549132](https://doi.org/10.1080/01431161.2025.2549132)
- Murugan, D., Garg, A., and Singh, D., 2017. Development of an adaptive approach for precision Agriculture monitoring with drone and satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 10 (12), 5322–5328. doi:[10.1109/JSTARS.2017.2746185](https://doi.org/10.1109/JSTARS.2017.2746185)
- Naethe, P., *et al.*, 2023. Calibration and Validation from ground to airborne and satellite level: joint application of time-synchronous field spectroscopy, drone, aircraft and Sentinel-2 imaging. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 91 (1), 43–58. doi:[10.1007/s41064-022-00231-x](https://doi.org/10.1007/s41064-022-00231-x)
- Padalia, H., Prakash, A., and Watham, T., 2023. Modelling aboveground biomass of a multistage managed forest through synergistic use of Landsat-OLI, ALOS-2 L-band SAR and GEDI metrics. *Ecological Informatics*, 77, 102234. doi:[10.1016/j.ecoinf.2023.102234](https://doi.org/10.1016/j.ecoinf.2023.102234)

- Padró, J.C., et al., 2018. Radiometric correction of Landsat-8 and Sentinel-2A scenes using drone imagery in synergy with field spectroradiometry. *Remote Sensing*, 10 (11), 1687. doi:10.3390/rs10111687
- Peng, Y., et al., 2021. Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal-spectral fusion of remote sensing images. *Information Fusion*, 65, 21–36. doi:10.1016/j.inffus.2020.08.013
- Pohl, C. and Van Genderen, J.L., 1998. Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, 19 (5), 823–854. doi:10.1080/014311698215748
- Qian, B., et al., 2022. A sentinel-2-based triangular vegetation index for chlorophyll content estimation. *Agricultural and Forest Meteorology*, 322, 109000. doi:10.1016/j.agrformet.2022.109000
- Salgueiro Romero, L., Marcello, J., and Vilaplana, V., 2020. Super-resolution of Sentinel-2 imagery using Generative adversarial networks. *Remote Sensing*, 12 (15), 2424. doi:10.3390/rs12152424
- Schreier, J., Ghazaryan, G., and Dubovyk, O., 2021. Crop-specific phenomapping by fusing Landsat and Sentinel data with MODIS time series. *European Journal of Remote Sensing*, 54 (sup1), 47–58. doi:10.1080/22797254.2020.1831969
- Shahdoosti, H.R., 2015. MS and PAN image fusion by combining Brovey and wavelet methods. In: *International Conference on Information Technology, Computer & Communication*. India.
- Shahdoosti, H.R. and Ghassemian, H., 2011. Multispectral and panchromatic image fusion by combining spectral PCA and spatial PCA methods. *Modares Journal of Electrical Engineering*, 11 (3), 19–27. https://mjee.modares.ac.ir/article_12895.html.
- Shao, Z., et al., 2019. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment*, 235, 111425. doi:10.1016/j.rse.2019.111425
- Song, Z., et al., 2020. Identifying sunflower lodging based on image fusion and deep semantic segmentation with UAV remote sensing imaging. *Computers and Electronics in Agriculture*, 179, 179. doi:10.1016/j.compag.2020.105812
- Swain, R., Paul, A., and Behera, M.D., 2024. Spatio-temporal fusion methods for spectral remote sensing: a comprehensive technical review and comparative analysis. *Tropical Ecology*, 65 (3), 356–375. doi:10.1007/s42965-023-00318-5
- Swaminathan, V., et al., 2024. Radiometric calibration of UAV multispectral images under changing illumination conditions with a downwelling light sensor. *The Plant Phenome Journal*, 7 (1). doi:10.1002/ppj2.70005
- Tarasiewicz, T., et al., 2023. Multitemporal and multispectral data fusion for super-resolution of Sentinel-2 images. *IEEE Transactions on Geoscience & Remote Sensing*, 61, 1–19. doi:10.1109/TGRS.2023.3311622
- Toosi, A., Samadzadegan, F., and Javan, F.D., 2025. Toward the optimal spatial resolution ratio for fusion of UAV and Sentinel-2 satellite imagery using metaheuristic optimization. *Advances in Space Research*, 75 (7), 5254–5282. doi:10.1016/j.asr.2025.02.019
- Toosi, A., Samadzadegan, F., and Javan, F.D., 2026. S³-ESRGAN: enhanced super-resolution Generative adversarial network for Remote Sensing imagery spatial resolution improvement—an application using Sentinel-2 and UAV images. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 19, 2149–2172. doi:10.1109/JSTARS.2025.3640940
- Vivone, G., et al., 2021. A New benchmark based on recent advances in multispectral pansharpening: revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9 (1), 53–81. doi:10.1109/MGRS.2020.3019315
- Wald, L., 2002. *Data fusion. Definitions and architectures-fusion of images of different spatial resolutions*. Paris, France: Presses de l'Ecole, Ecole des Mines de Paris.
- Wang, Q., et al., 2015. Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, 166, 191–204. doi:10.1016/j.rse.2015.06.003
- Wang, X. et al., 2018. ESRGAN: enhanced super-resolution Generative adversarial networks Computer Vision – ECCV 2018 Workshops. ECCV 2018 Cham eds. Leal-Taixé L Roth S Springer. doi:10.1007/978-3-030-11021-5_5.
- Wei, Z., et al., 2023. A-ESRGAN: training real-world blind super-resolution with attention U-Net discriminators. In: J. Siekmann, ed. *PRICAI 2023: trends in artificial intelligence*. Berlin: Springer, 16–27.
- Wu, T., Xu, W., and Wu, Y., 2025. A lightweight high-frequency mamba network for image super-resolution. *Scientific Reports*, 15 (1), 25973. doi:10.1038/s41598-025-11663-x
- Yang, J., et al., 2020. A robust method for generating high-spatiotemporal-resolution surface reflectance by fusing MODIS and Landsat data. *Remote Sensing*, 12 (14), 2312. doi:10.3390/rs12142312
- Yao, H., et al., 2025. Combination of UAV hyperspectral and LiDAR data for classifying karst wetland vegetation species using the AEL-Stacking and LIME. *Journal of Remote Sensing*, 5. doi:10.34133/remotesensing.0452
- Yao, Y., et al., 2025. Interpretable LAI fine inversion of maize by fusing satellite, UAV multispectral, and thermal infrared images. *Agriculture*, 15 (3), 243. doi:10.3390/agriculture15030243
- Yuan, L., et al., 2025. Integrating UAV and high-resolution satellite remote sensing for multi-scale rice disease monitoring. *Computers and Electronics in Agriculture*, 234, 110287. doi:10.1016/j.compag.2025.110287
- Zeng, N., et al., 2020. The utility of fusing multi-sensor data spatio-temporally in estimating grassland aboveground biomass in the three-river headwaters region of China. *International Journal of Remote Sensing*, 41 (18), 7068–7089. doi:10.1080/01431161.2020.1752411
- Zhang, F., et al., 2025. A novel image fusion method based on UAV and Sentinel-2 for environmental monitoring. *Scientific Reports*, 15 (1). doi:10.1038/s41598-025-13049-5
- Zhang, H., et al., 2023. P2Sharpen: a progressive pansharpening network with deep spectral transformation. *Information Fusion*, 91, 103–122. doi:10.1016/j.inffus.2022.10.010

- Zhang, L., *et al.*, 2022. A generative adversarial network with joint multistream architecture and spectral compensation for pansharpening. *Applied Soft Computing*, 117, 108451. doi:[10.1016/j.asoc.2022.108451](https://doi.org/10.1016/j.asoc.2022.108451)
- Zhang, Y. *et al.*, 2018. Image super-resolution using very deep residual channel attention networks Proceedings of the European Conference on Computer Vision (ECCV) 2018 286–301. https://openaccess.thecvf.com/content_ECCV_2018/html/Yulun_Zhang_Image_Super-Resolution_Using_ECCV_2018_paper.html.
- Zhao, L., *et al.*, 2019. Finer classification of crops by fusing UAV images and Sentinel-2A data. *Remote Sensing*, 11 (24), 3012. doi:[10.3390/rs11243012](https://doi.org/10.3390/rs11243012)
- Zhou, W., *et al.*, 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4), 600–612.

4 Fusión temporal de imágenes UAV y Sentinel-2 mediante redes neuronales convolucionales para el monitoreo de cultivos

Resumen

El uso de imágenes satelitales Sentinel-2 para el monitoreo de cultivos se ve limitado por la resolución espacial nativa. Ante esta restricción, las técnicas de superresolución basadas en fusión de datos han emergido como una alternativa para generar imágenes de mayor detalle espacial. En este estudio se presenta el modelo TVDSR, un modelo de fusión de imágenes Sentinel-2 y UAV que integra información temporal y espectral mediante una arquitectura basada en ConvLSTM bidireccional. Las imágenes empleadas para la superresolución fueron construidas a partir composiciones espectrales de cinco índices de vegetación. Para evaluar su desempeño se utilizaron 15 fechas de adquisición, correspondientes a 11 imágenes satelitales Sentinel-2 de 2024 y 4 de 2025. Dos fechas independientes se reservaron para la validación mediante métricas de calidad de reconstrucción. El desempeño del modelo propuesto se comparó con tres modelos de referencia del estado del arte: S³-ESRGAN, DeepSent y UAV-Net. UAV-Net obtuvo el mejor desempeño para la imagen del 2024 (RMSE=0.128, MAE=0.10, PSNR=24.06dB y SSIM=0.86); sin embargo, TVDSR presentó las mejores métricas para la evaluación del 2025 (RMSE=0.149, MAE=0.115, PSNR=22.60 dB y SSIM=0.85) y mostró la menor degradación entre fechas, lo que evidencia una mayor capacidad de generalización temporal. Asimismo, TVDSR preservó de manera más consistente los índices de vegetación derivados, utilizados para evaluar la conservación de información espectral relevante para el monitoreo del estado del cultivo, y generó reconstrucciones visualmente más próximas a las imágenes UAV de referencia. La integración conjunta de información temporal y espectral constituye una estrategia efectiva para mejorar la robustez y la capacidad de generalización de la superresolución de imágenes Sentinel-2 en aplicaciones de monitoreo de cultivos.

Palabras Clave: Superresolución; Índices de Vegetación; Espaciotemporal; ConvLSTM; Aprendizaje profundo.

Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Programa de Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: M.C. Alvaro Murguía Cozar.

Director de tesis: Dr. Gilberto de Jesús López Canteñas.

Fusión temporal de imágenes UAV y Sentinel-2 mediante redes neuronales convolucionales para el monitoreo de cultivos

Abstract

The use of Sentinel-2 satellite imagery for crop monitoring is constrained by its native spatial resolution. To address this limitation, data fusion-based super-resolution techniques have emerged as a promising approach for generating imagery with enhanced spatial detail. This study presents TVDSR, a Sentinel-2 and UAV image fusion model that integrates temporal and spectral information through a bidirectional ConvLSTM-based architecture. The images used for super-resolution were constructed from spectral compositions of five vegetation indices. To evaluate its performance, 15 acquisition dates were used, including 11 Sentinel-2 images acquired in 2024 and 4 acquired in 2025. Two independent dates were reserved for validation using image reconstruction quality metrics. The performance of the proposed model was compared against three state-of-the-art benchmark models: S³-ESRGAN, DeepSent, and UAV-Net. UAV-Net achieved the best performance for the 2024 evaluation image (RMSE = 0.128, MAE = 0.10, PSNR = 24.06 dB, and SSIM = 0.86). However, TVDSR achieved the best results for the 2025 evaluation image (RMSE = 0.149, MAE = 0.115, PSNR = 22.60 dB, and SSIM = 0.85) and exhibited the lowest performance degradation across acquisition dates, indicating superior temporal generalization capability. Furthermore, TVDSR more consistently preserved the derived vegetation indices, which were used to assess the retention of spectral information relevant to crop condition monitoring, and produced reconstructions that were visually closer to the reference UAV imagery. The joint integration of temporal and spectral information constitutes an effective strategy for improving the robustness and generalization capability of Sentinel-2 image super-resolution for crop monitoring applications.

Keywords: Super-resolution; Vegetation Indices; Spatiotemporal Analysis; ConvLSTM; Sentinel-2; Deep learning.

Doctoral Thesis. Graduate Program in Agricultural Engineering and Integrated Water Resources Management, Universidad Autónoma Chapingo.

Author: Msc. Alvaro Murguía Cozar.

Thesis director: Phd. Gilberto de Jesús López Canteñs.

4.1 Introducción

El monitoreo de la superficie Terrestre, mediante el procesamiento de imágenes, se ha convertido en una tarea fundamental en diversos campos, particularmente en el ámbito agrícola, donde proporciona una base sólida para el seguimiento de cultivos, caracterización de parcelas y la predicción de variables biofísicas orientadas a la optimización en la gestión de recursos. En los últimos años ha aumentado la disponibilidad de fuentes para la adquisición de imágenes espectrales en distintos rangos del espectro electromagnético, lo que ha permitido obtener datos multiespectrales e hiperespectrales con mayor resolución espacial, temporal y espectral. Esto ha hecho posible complementar las limitaciones inherentes al uso de una sola fuente mediante la fusión de datos provenientes de múltiples sensores (Swain et al., 2024)

En este contexto, han surgido diversas técnicas orientadas a mejorar la calidad de las imágenes. Entre ellas, destaca la superresolución, una metodología que surge en la década de 1980 en el campo del procesamiento de imágenes, con el enfoque de la restauración de imágenes de alta resolución a partir de un conjunto de imágenes de baja resolución de una misma superficie (Elad & Feuer, 1997; Irani & Peleg, 1990). Con el desarrollo de la teledetección satelital, este concepto evolucionó hacia métodos de fusión espacial como el pansharpening, en el cual una imagen pancromática de alta resolución es utilizada para mejorar la calidad espacial de imágenes multiespectrales (Pohl & Van Genderen, 1998). Este proceso de mejoramiento de imágenes multiespectrales, en el cual una imagen de alta resolución se utiliza como guía, se ha popularizado y potenciado con el surgimiento del aprendizaje profundo. Estas técnicas han permitido integrar información multiescala proveniente de vehículos aéreos no tripulados (UAV's) y plataformas satélites (Bazrafkan et al., 2025; Maimaitijiang et al., 2020b; J. Xiao et al., 2023b).

Los métodos de superresolución de datos multifuente pueden dividirse en dos enfoques principales: superresolución de imagen única (SISR) y superresolución de imágenes múltiples (MISR). Mientras que MISR usa múltiples imágenes de la misma escena para mejorar una imagen de baja resolución (LR), SISR emplea una sola imagen LR (Fernandez-Beltran et al., 2016; Xie et al., 2026). Entre los modelos SISR se encuentran los enfoques basados en redes convolucionales como Superresolution Convolutional Neural Network (SRCNN) (Dong et al., 2014), Very Deep Super-Resolution VDSR (Kim et al., 2016) y Enhanced Deep Super-Resolution (EDSR)(Lim et al., 2017); los métodos generativos adversarios como Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)(Wang et al., 2019) y Real-ESRGAN (Wang et al., 2021); así como arquitecturas más recientes basadas en Transformers como SwinIR (Liang et al., 2021) y modelos de difusión como SinSR (Wang et al., 2024), los cuales han demostrado buen desempeño en la reconstrucción de detalles espaciales y calidad perceptual.

A pesar de los avances logrados por métodos SISR, estos presentan limitaciones de generalización hacia nuevas imágenes LR. Estas limitaciones se deben en parte, a que dichos métodos parten de un problema mal planteado, en el cual múltiples soluciones de alta resolución pueden corresponder a una misma imagen de baja resolución (J. Zhang et al., 2022b). Es así como, durante el entrenamiento los modelos deben inferir información espacial a partir de datos incompletos, ya que no consideran información adicional más allá de una sola observación. Como consecuencia, el modelo no cuenta con la capacidad de capturar adecuadamente la variabilidad espacial, temporal y espectral, lo cual es importante en aplicaciones de teledetección donde la dinámica del suelo y la reflectancia varían en el tiempo (Cao et al., 2026).

En este contexto, los enfoques basados en MISR son una alternativa más adecuada para la integración de imágenes UAV y satelitales, ya que permiten incorporar información adicional proveniente de diferentes observaciones de una

misma región, adquiridas en distintas fechas (Ding et al., 2025). Los enfoques MISR aprovechan los datos temporales para enriquecer la reconstrucción espacial y capturar información no observable en una sola imagen debido a las limitaciones del sensor, ruido o condiciones atmosféricas (W. Liu et al., 2025a). En este sentido la integración de secuencias temporales permite modelar la evolución temporal de la superficie, lo que conduce a representaciones más consistentes frente a cambios interanuales, iluminación variable y diferencias en condiciones de adquisición.

En este trabajo se propone un modelo de superresolución multitemporal denominado Temporal Very Deep Super Resolution (TVDSR), diseñado para la fusión de datos Sentinel-2 y UAV a nivel de índices de vegetación, con la finalidad mejorar la resolución espacial de imágenes Sentinel-2 para el monitoreo de cultivos mediante la integración conjunta de información temporal y espectral, con el fin de obtener reconstrucciones más precisas y robustas a lo largo del tiempo. El modelo aprende un mapeo no lineal espacio temporal que utiliza como entrada una secuencia de composiciones satelitales para predecir una imagen de alta resolución correspondiente a la fecha objetivo. La arquitectura combina un encoder convolucional profundo para la extracción de características espaciales, un módulo ConvLSTM para capturar dependencias temporales, un mecanismo de atención temporal y un mecanismo de atención espectral que ponderan de forma dinámica las contribuciones en función de su relevancia para la reconstrucción.

Adicionalmente, el modelo incorpora un embedding del día juliano que permite actualizar las activaciones de las compuertas del módulo ConvLSTM y representar explícitamente la naturaleza cíclica de los procesos estacionales, además facilita la generación entre diferentes años y condiciones de adquisición. Asimismo, el modelo aplica una formulación de aprendizaje residual, en la cual se predice la diferencia entre en la imagen de entrada y salida de alta resolución,

lo que contribuye a preservar la estructura espectral y mejorar la estabilidad del entrenamiento.

Los resultados obtenidos demuestran que el enfoque propuesto es capaz de capturar relaciones espacio temporales complejas y generalizar adecuadamente con baja variabilidad interanual. El modelo TDVSR superó a modelos recientes basados en información espacial y con enfoques MISR. Los resultados resaltan la importancia de incorporar series temporales en la integración de imágenes Sentinel-2 y UAV. En la siguiente sección se presenta una revisión de los trabajos relacionados en superresolución multitemporal y fusión de datos satelitales y UAV.

4.2 Trabajos relacionados

4.2.1 Modelos de superresolución de imágenes múltiples

Los modelos MISR surgieron a partir de modelos de formación de imagen, donde cada observación de baja resolución se representa como una versión degradada de una imagen de alta resolución desconocida. En este contexto, la reconstrucción se formuló como un problema inverso regularizado, en el cual se emplearon funciones de error basadas en normas L2 y L1 para estimar la imagen que mejor explicara el conjunto de observaciones disponibles (Elad & Feuer, 1999; Farsiu et al., 2004). Aunque estos enfoques demostraron la viabilidad de la reconstrucción multimagen, su desempeño dependía fuertemente de la precisión del registro geométrico entre imágenes.

Para reducir la sensibilidad a errores de alineación, se desarrollaron técnicas capaces de explotar redundancias espaciales y temporales sin depender exclusivamente de la estimación explícita del movimiento. Entre ellas destacan los enfoques basados en Non-Local Means y regresión kernel multidimensional, que incorporaron mecanismos más robustos para modelar la similitud entre observaciones y mejorar la reconstrucción en presencia de ruido o movimientos

complejos (Protter et al., 2009; Takeda et al., 2009). No obstante, estos métodos seguían limitados por la necesidad de definir explícitamente modelos de degradación y reglas de reconstrucción.

La introducción del aprendizaje profundo transformó el problema de la superresolución al permitir aprender directamente la relación entre secuencias de imágenes de baja resolución y sus correspondientes reconstrucciones de alta resolución. Los primeros enfoques combinaron algoritmos convencionales de alineación con redes neuronales convolucionales encargadas de la fusión y reconstrucción, como fue el caso de los modelos propuestos por Liao et al. (2015) y el modelo VSRNet propuesto por Kappeler et al. (2016). Posteriormente, arquitecturas como SPMC integraron la compensación de movimiento y el proceso de upsampling dentro de la propia red que eliminó la necesidad de módulos externos y permitió una optimización conjunta del proceso de reconstrucción (Tao et al., 2017).

Los avances recientes se han centrado en integrar de manera conjunta la alineación, la fusión temporal y la reconstrucción de imágenes. Modelos como RBPN incorporaron mecanismos recurrentes para refinar iterativamente la información temporal proveniente de múltiples imágenes (Haris et al., 2019), mientras que EDVR combinó convoluciones deformables y mecanismos de atención temporal y espacial para mejorar la selección dinámica de características relevantes durante la reconstrucción (Chan, et al., 2019). De forma similar, HighRes-net demostró que la fusión jerárquica de representaciones profundas permite integrar eficazmente información procedente de múltiples observaciones (Deudon et al., 2020).

Más recientemente, la investigación se ha orientado al modelado de dependencias temporales de largo alcance. BasicVSR y BasicVSR++ establecieron nuevos referentes al utilizar propagación bidireccional de características y esquemas avanzados de alineación basados en flujo óptico y

convoluciones deformables, logrando mejoras significativas en tareas de superresolución de video (Chan et al., 2021, 2022). Estos avances evidencian la importancia de explotar explícitamente la dimensión temporal para mejorar la calidad de reconstrucción.

4.2.2 Modelos de superresolución espacio temporales

Los modelos de superresolución espaciotemporal surgieron como una extensión de los enfoques tradicionales de superresolución multimagén desarrollados inicialmente para restauración de video. A diferencia de los métodos basados únicamente en la fusión espacial de múltiples observaciones, estas arquitecturas buscan modelar explícitamente las dependencias temporales presentes entre imágenes consecutivas para mejorar la reconstrucción de detalles de alta resolución (Jo et al., 2018; Tao et al., 2017). Esta evolución permitió aprovechar de forma más eficiente la información redundante contenida en secuencias temporales y reducir la dependencia de procedimientos explícitos de estimación de movimiento.

Posteriormente, la investigación se orientó hacia arquitecturas capaces de integrar conjuntamente las etapas de alineación, fusión y reconstrucción. En este contexto, DeepSUM y DeepSUM++ explotan correlaciones entre múltiples adquisiciones satelitales mediante redes convolucionales profundas e incorporan mecanismos de información no local para capturar dependencias de largo alcance entre regiones similares de la imagen (Bordone Molini et al., 2020; Molini et al., 2020). De manera paralela, los modelos de superresolución por ráfagas incorporaron mecanismos de alineación adaptativa y atención espacial para mejorar el aprovechamiento de microdesplazamientos entre imágenes, destacando los trabajos de Bhat et al. (2021) y Luo et al. (2021).

Los avances más recientes han incorporado mecanismos de atención, modelos probabilísticos y arquitecturas híbridas para mejorar la representación temporal. Entre ellos destacan la reformulación probabilística propuesta por Bhat et al. (2021),

las estrategias de pseudo-burst (Dudhane et al., 2025), así como BSRT, que integra Transformers, flujo óptico y convoluciones deformables para combinar información local y global durante la reconstrucción (Luo et al., 2021). De forma similar, QMambaBSR introduce modelos de espacio de estados y mecanismos de fusión multiescala para capturar interacciones entre y dentro del frame de manera más eficiente (Di et al., 2025).

Cuando la consistencia temporal constituye un factor crítico, las arquitecturas basadas en ConvLSTM han demostrado ser especialmente efectivas debido a su capacidad para modelar simultáneamente dependencias espaciales y temporales. Zhang et al. (2020) emplearon ConvLSTM dentro de una arquitectura encoder-decoder para generar representaciones globales de secuencias temporales, mientras que Li et al. (2024) incorporaron esta estructura en un esquema de fusión multiescala para mejorar la reconstrucción mediante la eliminación de información redundante. Trabajos posteriores han extendido este tipo de arquitecturas a tareas de restauración, fusión multimodal y reconstrucción multitemporal mediante mecanismos recurrentes capaces de ponderar dinámicamente la contribución de cada observación (Fu et al., 2025; Heng et al., 2020; Liu et al., 2025).

Paralelamente, los Transformers se han consolidado como una de las principales líneas de investigación en superresolución multiimagen. Modelos como TR-MISR, MAST y EpiMISR utilizan mecanismos de atención para establecer correspondencias entre múltiples observaciones y capturar dependencias globales difíciles de representar mediante convoluciones convencionales (Aira et al., 2024; An et al., 2022; Li et al., 2023). Más recientemente, HiRT incorporó una arquitectura recurrente jerárquica en el dominio de la frecuencia para explotar simultáneamente dependencias espaciales, temporales y multiescala (Choi & Kim, 2026). Asimismo, arquitecturas híbridas como TRCFNet combinan atención, mecanismos de compuertas y fusión entre capas para mejorar la recuperación

de detalles finos y reducir redundancias en la representación de características (Wang et al., 2026).

4.2.3 Modelos de superresolución para la integración Sentinel-2 con UAV

La superresolución de imágenes Sentinel-2 ha evolucionado desde enfoques basados exclusivamente en la información proporcionada por el propio sensor hacia estrategias que integran múltiples fuentes de datos para superar las limitaciones de resolución espacial. Los primeros modelos basados en aprendizaje profundo utilizaron las bandas de mayor resolución espacial de Sentinel-2 para mejorar las bandas de menor resolución. Entre ellos destacan DSen2 (Lanaras et al., 2018), que explota correlaciones espectrales mediante redes convolucionales profundas, así como los enfoques propuestos por Palsson et al. (2018) y Gargiulo et al. (2019), orientados a la recuperación de detalles de alta frecuencia y a la preservación de estructuras espaciales mediante aprendizaje residual y técnicas de fusión de información.

Posteriormente, modelos incorporaron sensores externos para proporcionar información espacial adicional. En esta línea, Galar et al. (2020b) utilizaron imágenes PlanetScope para generar conjuntos de entrenamiento supervisados mediante una arquitectura EDSR, mientras que Latte & Lejeune (2020) propusieron una estrategia de fusión Sentinel-2/PlanetScope basada en redes convolucionales residuales. Estos trabajos demostraron que la integración multisensor puede superar las limitaciones inherentes a la resolución nativa de Sentinel-2 y mejorar significativamente la capacidad de reconstrucción espacial.

Paralelamente, surgieron enfoques orientados a aprovechar la dimensión temporal de las observaciones. Jia et al. (2020) introdujeron DL-SDFM, un modelo de doble flujo capaz de modelar simultáneamente relaciones espaciales y temporales mediante una combinación adaptativa de ambas fuentes de información. Más recientemente, DeepSent incorporó información temporal y espectral dentro de una arquitectura unificada para mejorar la reconstrucción de

imágenes Sentinel-2 a escalas métricas más finas (Tarasiewicz et al., 2023). Asimismo, OBSTFM propuso una representación basada en objetos para mejorar la consistencia temporal y reducir los efectos del ruido en escenarios con cambios espaciales complejos (Zhang et al., 2021).

Otro enfoque importante corresponde a los modelos generativos. RS-ESRGAN (Salgueiro Romero et al., 2020a) demostró el potencial de las redes generativas adversariales para incrementar la resolución espacial de Sentinel-2 mediante información de sensores de muy alta resolución. Posteriormente, arquitecturas más avanzadas como MAF-GAN (Z. Wang et al., 2025) y GS-SRGAN (Farah et al., 2026) incorporaron mecanismos de atención espacial y espectral para mejorar simultáneamente la calidad visual y la consistencia radiométrica de las reconstrucciones. De forma complementaria, los modelos basados en difusión, como DiffFuSR (Sarmad et al., 2025), han surgido como una alternativa prometedora al integrar explícitamente modelos de degradación y procesos probabilísticos de reconstrucción

Dentro del contexto específico de la integración satélite-UAV, los trabajos recientes han buscado explotar la complementariedad entre la alta resolución espacial de los vehículos aéreos no tripulados y la cobertura temporal de los sensores satelitales. RarefyNet utiliza información UAV para refinar índices de vegetación derivados de imágenes satelitales (Mazzia et al., 2020), mientras que UAV-Net combina información multiescala procedente de imágenes satelitales y UAV mediante mecanismos de codificación profunda y fusión jerárquica (J. Xiao et al., 2023a). Más recientemente, se han desarrollado enfoques híbridos que integran superresolución y armonización espectral. Entre ellos destacan la metodología propuesta por Masrur et al. (2026), basada en la generación de referencias espectralmente consistentes mediante datos hiperespectrales UAV, y S³-ESRGAN (Toosi et al., 2026a), que incorpora mecanismos de atención espacial, atención de canal y preservación espectral para mantener la coherencia radiométrica durante la reconstrucción.

En conjunto, estos avances confirman que la integración de datos UAV y Sentinel-2 constituye una estrategia efectiva para combinar resolución espacial, temporal y espectral (Ren et al., 2025). No obstante, a pesar de los avances recientes en técnicas de fusión y superresolución basadas en aprendizaje profundo, persisten varios desafíos que limitan la adopción generalizada de estos enfoques. Entre los principales problemas se encuentran las dificultades en la alineación geométrica, espectral y temporal entre sensores, debido a diferencias en resolución, ángulos de adquisición, tiempos de captura y respuestas espectrales; la escasez de datos de referencia de alta calidad, corregistrados y multitemporales, necesarios para entrenar modelos supervisados de manera efectiva; y el alto costo computacional asociado con arquitecturas profundas que integran múltiples fuentes y secuencias temporales. Asimismo, la integración temporal introduce retos adicionales relacionados con la variabilidad dinámica de la escena, ya que los cambios reales en el terreno pueden interferir con el proceso de aprendizaje, por lo que se vuelve compleja la separación entre diferencias temporales legítimas y artefactos de reconstrucción. Estas limitaciones resaltan la necesidad de desarrollar modelos más robustos, eficientes y adaptativos, capaces de integrar de manera coherente información multiescala, multisensor y multitemporal, mantener la consistencia espectral a lo largo del tiempo, con el fin de facilitar su aplicación en sistemas operativos de monitoreo agrícola y ambiental a gran escala

4.3 Materiales y Métodos

4.3.1 Área de Estudio

La presente investigación se realizó en el campo experimental “San Juan” de la Universidad Autónoma Chapingo, el cual se localiza a 19°29'35.48" latitud Norte, -98°51'35.11" longitud Oeste y 2211 metros sobre el nivel de mar. El clima en la zona de estudio es, de acuerdo con la clasificación de Köppen (1936) modificado por Enriqueta García (1964), seco semiárido semicálido con temperatura media

anual menor a 18° con lluvias en verano y sequía marcada en invierno. En la Figura 4.1 se presenta la ubicación de la zona estudio con una superficie de 37.9 hectáreas, dividida en 28 parcelas cultivadas con maíz, en el 2024 y 2025.

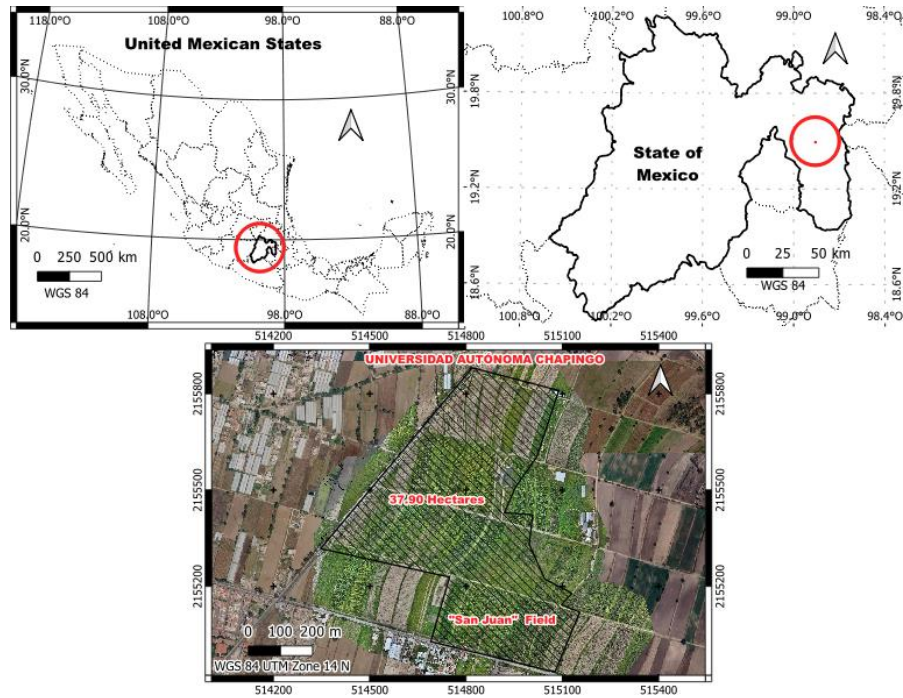


Figura 4.1. Localización geográfica y delimitación del área de estudio en la Universidad Autónoma Chapingo.

4.3.2 Adquisición de datos

4.3.2.1 Imágenes UAV

Se empleó la aeronave DJI Mavic 3M para ejecutar vuelos durante los ciclos agrícolas 2024 y 2025, lo que permite la adquisición de imágenes aéreas de alta resolución para el análisis de la vegetación. Integra un sistema de cámaras compuesto por una RGB (CMOS 4/3 y 20MP) y cuatro sensores multispectrales (CMOS de 1/2.8 pulgadas y 5MP), verde (560nm $\pm$ 16 nm), rojo (650nm $\pm$ 16 nm), borde rojo (730nm $\pm$ 16 nm) e infrarrojo cercano (860nm $\pm$ 26 nm). El equipo cuenta con un módulo GNSS con capacidad RTK (Real Time Kinematic), que proporciona una precisión centimétrica durante la georreferenciación de los vuelos.

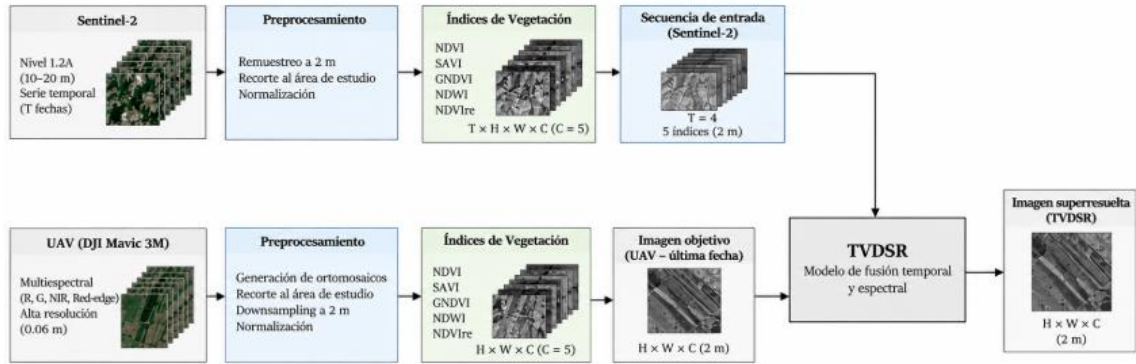


Figura 4.2. Flujo metodológico para la fusión de imágenes Sentinel-2 y UAV mediante el modelo TVDSR.

Se realizaron 11 vuelos sobre la superficie de estudio en los meses de julio a noviembre del 2024 y cuatro vuelos de julio a septiembre del 2025. Los vuelos UAV se realizaron a 120 m de altura, con un traslape frontal y lateral del 70 %. Las imágenes fueron georreferenciadas mediante una estación D-RTK 2, instalada sobre un punto de control previamente establecido en el marco de referencia ITRF2008 (época 2010). Las coordenadas precisas del punto se determinaron mediante observaciones estáticas con un Trimble R8 y posprocesamiento diferencial en Emlid Studio ligado la estación ICMX de la Red Nacional Activa de México. Las imágenes fueron procesadas en Pix4Dmapper para generar mapas de reflectancia de las bandas Red, Green, NIR y Red Edge, con una resolución espacial de 0.06 m. Finalmente, los mapas fueron remuestreados mediante downsampling a 2 m para su integración con los datos Sentinel-2 (Figura 4.2).

4.3.2.2 Imágenes Sentinel-2

Para cada fecha de adquisición UAV se descargaron imágenes Sentinel-2 nivel L2A correspondientes a las bandas verde (560 nm), roja (665 nm), borde rojo (740 nm) e infrarrojo cercano (865 nm). Estas bandas fueron seleccionadas por su relevancia en el cálculo de índices de vegetación utilizados en el estudio. Posteriormente, todas las imágenes fueron remuestreadas mediante

interpolación bicúbica a una resolución espacial de 2 m para garantizar la consistencia durante el proceso de fusión y superresolución.

4.3.2.3 Índices de Vegetación

Las imágenes obtenidas a partir de los datos Sentinel-2 y UAV fueron normalizadas mediante la transformación mínimo-máximo (Ecuación 1). Este procedimiento permitió escalar los valores de reflectancia al intervalo [0,1], lo que favorece la consistencia, comparabilidad y estabilidad de los índices de vegetación en el análisis multitemporal y multisensor.

$$I_{norm} = \frac{I - R_{MIN}}{R_{MAX} - R_{MIN}} \quad (1)$$

donde I_{norm} es la banda normalizada, I corresponde al valor de reflectancia original del píxel, R_{MIN} es el valor mínimo de reflectancia presente en la imagen y R_{MAX} es el valor máximo de reflectancia presente en la imagen.

A partir de las imágenes normalizadas se calcularon cinco índices de vegetación para cada fecha de adquisición, tanto para Sentinel-2 como para UAV. Los índices seleccionados fueron NDVI, SAVI, GNDVI, NDWI y NDRE, debido a que proporcionan información complementaria sobre vigor vegetal, contenido de clorofila, estrés hídrico, condición de la vegetación y respuesta espectral del borde rojo, variables ampliamente utilizadas en el monitoreo de cultivos. Posteriormente, los cinco índices se integraron en una composición multibanda de cinco canales y se recortaron al área de estudio mostrada en la Figura 4.1. Las expresiones empleadas para el cálculo de cada índice se presentan en la Cuadro 4.1.

Cuadro 4.1. Índices de Vegetación evaluados en el estudio.

Index	Ecuación
Normalized Difference Vegetation Index	$NDVI = \frac{NIR - RED}{NIR + RED}$
Soil Adjusted Vegetation Index	$SAVI = \frac{NIR - RED}{NIR + RED + L} (1 + L)$
Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}$
Normalized Difference Water Index	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
Red-Edge Normalized Difference Vegetation Index	$NDVI_{re} = \frac{RE - RED}{RE + RED}$

RED banda roja, GREEN banda verde, NIR banda infrarrojo cercano, RE banda borde rojo, L=0.5 factor de corrección del brillo del suelo

4.3.3 Método propuesto

4.3.3.1 Arquitectura general del modelo TVDSR

El modelo TVDSR propuesto es un modelo híbrido que integra diferentes estructuras provenientes de trabajos previos en el área de superresolución de imágenes y modelado espaciotemporal (Figura 4.3). En particular, el modelo toma como referencia la arquitectura VDSR (Kim et al., 2016a), la cual se caracteriza por el uso de redes convolucionales profundas junto con una estrategia de aprendizaje residual global, lo que permite la extracción de características con un costo computacional bajo. Dado que VDSR se concibió como un modelo de superresolución SISR, TVDSR extiende su formulación para incorporar la dimensión temporal bajo el enfoque MISR. Para ello se introduce un esquema de modelado secuencial basado en ventanas, donde cada muestra de entrada está compuesta por múltiples observaciones.

El problema que se aborda en este trabajo se formula como una tarea de superresolución espaciotemporal supervisada, en la cual se busca aprender una función de mapeo no lineal entre secuencias de imágenes Sentinel-2 y una imagen UAV de alta resolución correspondiente a la última fecha de observación. Dado un conjunto de observaciones multitemporales de Sentinel-2, representadas como una secuencia ordenada cronológicamente:

$$X = \{X_{t_1}, X_{t_2}, \dots, X_{t_T}\} \quad (2)$$

donde $X_t \in R^{C \times H \times W}$ corresponde a una imagen multicanal conformada por cinco índices de vegetación, se busca estimar una imagen de alta resolución espacial adquirida por UAV en el tiempo final t_T :

$$Y_{t_T} = \mathcal{F}(X) \quad (3)$$

donde $Y_{t_T} \in R^{C \times H \times W}$ representa la imagen UAV de referencia y $\mathcal{F}(\cdot)$ una función compleja que modela relaciones espaciales, espectrales y temporales.

A diferencia de enfoques de superresolución SISR, el presente enfoque incorpora explícitamente la dimensión temporal, lo que permite aprovechar la redundancia multitemporal presente en las observaciones satelitales. Esta redundancia es relevante en contextos agrícolas, donde las dinámicas fenológicas generan patrones temporales consistentes que pueden utilizarse para mejorar la reconstrucción espacial. Cada muestra de entrenamiento se construye mediante una estrategia de ventanas deslizantes sobre las secuencias disponibles, de modo que el modelo recibe como entrada un conjunto de T imágenes Sentinel-2 consecutivas:

$$X = [X_{t_i}, X_{t_{i+1}}, \dots, X_{t_{i+T-1}}] \quad (4)$$

y como objetivo la imagen UAV correspondiente al último instante de la secuencia $Y_{t_{i+T-1}}$.

Con la finalidad de capturar las dependencias espaciotemporales dentro de las ventanas, el modelo incorpora un módulo ConvLSTM, el cual permite integrar la información temporal con la preservación simultánea de la estructura espacial de los datos (Shi et al., 2015). Además, se adopta un enfoque bidireccional donde la secuencia es procesada hacia adelante y hacia atrás, lo que permite

aprovechar la información del pasado y del futuro (Chan et al., 2022; Choi & Kim, 2026).

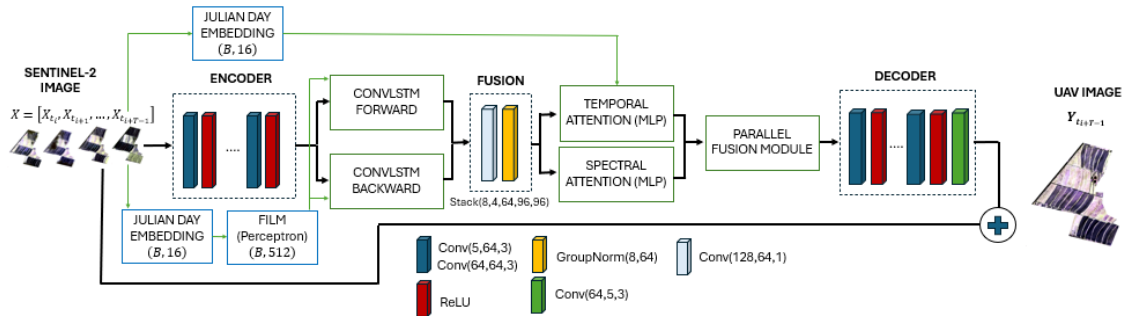


Figura 4.3. Estructura general del modelo TVDSR.

Asimismo, el modelo incorpora información temporal explícita mediante el uso de embeddings sinusoidales (Vaswani et al., 2023). Basados en el día juliano, lo que facilita el modelado de patrones periódicos asociados a la dinámica temporal de las observaciones. Estos embeddings son empleados para modular el comportamiento del ConvLSTM a través de un mecanismo Feature-wise Linear Modulation (FiLM) (Perez et al., 2017), el cual ajusta las compuertas internas del modelado mediante parámetros de escala y desplazamiento condicionados por la información temporal.

Posteriormente, el modelo incluye un módulo de atención temporal que permite asignar pesos adaptativos de los estados ocultos, basado en los embeddings sinusoidales, a lo largo de la secuencia, lo que permite darles más peso a los detalles temporales más relevantes para la tarea de reconstrucción; asimismo de forma paralela se agrega un mecanismo de atención espectral que al igual que el módulo de atención temporal determina pesos adaptativos por canal. Los tensores de salida de los módulos de atención se fusionan mediante un módulo de fusión paralela que establece dos parámetros que funcionan como pesos para una correcta fusión. Finalmente, la representación resultante es procesada por un decoder convolucional, el cual, en conjunto con la conexión residual, permite reconstruir la imagen de salida.

4.3.3.2 Módulo de extracción de características espaciales (Encoder)

El módulo de extracción de características espaciales transforma las imágenes de entrada en tensores de características que preservan la estructura espacial de los datos. El bloque está compuesto por diez capas convolucionales 2D con activaciones no lineales ReLU, como se muestra en la Figura 4.4, donde se parte de una convolución inicial que proyecta los canales de entrada hacia un espacio de características de mayor dimensión para pasar de 5 a 64 canales, seguido de nueve capas convolucionales intermedias que mantienen la dimensionalidad y permiten una extracción progresiva de características conforme se aplican las capas convolucionales. En particular, el encoder está formado por un conjunto de filtros espaciales de 3×3 , con un padding para mantener la resolución espacial, lo cual favorece la captura de patrones locales. Este procesamiento se aplica de manera independiente a cada elemento de la ventana temporal, lo que genera una secuencia de mapas de características de dimensión (B, C, H, W) por cada instante temporal.

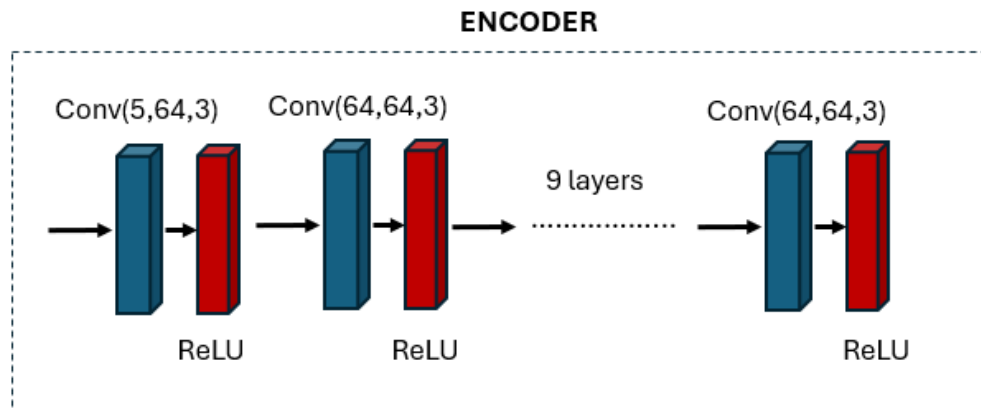


Figura 4.4. Estructura del encoder del modelo TVDSR.

4.3.3.3 Modelado temporal mediante ConvLSTM bidireccional

Después de extraer las características espaciales de las imágenes Sentinel-2, el modelo incorpora una capa Convolutional Long Short-Term Memory (ConvLSTM). A diferencia LSTM convencional, en la que los valores de las

compuertas se calculan mediante operaciones matriciales, la ConvLSTM sustituye dichas operaciones por convoluciones bidimensionales. Esta mejora permite preservar la estructura espacial de los tensores de características extraídos en el encoder.

Sea F_t el tensor de características extraído para el instante temporal t . La ConvLSTM actualiza sus estados ocultos y de memoria mediante las compuertas de entrada i_t , olvido f_t y salida o_t , además de un estado candidato de memoria $\tilde{C}_t$. Las ecuaciones de actualización de parámetros son las siguientes:

$$i_t = \sigma(W_{xi} * F_t + W_{hi} * H_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{xf} * F_t + W_{hf} * H_{t-1} + b_f) \quad (6)$$

$$\tilde{C}_t = \tanh(W_{xc} * F_t + W_{hc} * H_{t-1} + b_c) \quad (7)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (8)$$

$$o_t = \sigma(W_{xo} * F_t + W_{ho} * H_{t-1} + b_o) \quad (9)$$

$$H_t = o_t \odot \tanh(C_t) \quad (10)$$

donde H_t representa el estado oculto, C_t el estado de memoria, $\sigma(\cdot)$ la función sigmoide, $\tanh(\cdot)$ la función tangente hiperbólica, W_{xi} , W_{xf} , W_{xo} y W_{xc} representan los filtros aplicados a la entrada actual; W_{hi} , W_{hf} , W_{ho} y W_{hc} corresponden a los filtros asociados al estado oculto previo H_{t-1} ; b_i , b_f , b_o y b_c corresponden a los términos de sesgo y $\odot$ es el producto elemento por elemento.

En la implementación de la capa ConvLSTM propuesta, las activaciones de las tres compuertas y el estado candidato de memoria se calculan, simultáneamente, con una convolución 2D conformada por 256 filtros de 3×3 . Para ello, el tensor de características de 64 canales se concatena con estado oculto previo de 64 canales, lo que genera un tensor de 128 canales. Después de aplicar el proceso

de convolución, genera G_t que contiene 256 mapas de características, los cuales se dividen en cuatro grupos de 64 canales para formar i_t, f_t, o_t y $\tilde{C}_t$ respectivamente.

Con el fin de aprovechar la información temporal del futuro y del pasado, se implementó una arquitectura bidireccional compuesta por dos ramas ConvLSTM independiente (Figura 4.3). La primera procesa las ventanas de forma cronológica hacia adelante y la segunda recorre en sentido inverso. Como resultado se generan dos estados ocultos H_t^{fwd} y $H_t^{bwd} \in \mathbb{R}^{B \times 64 \times H \times W}$ para cada instante temporal.

4.3.3.4 Modelado estacional basado en embeddings y FiLM

La dinámica espectral observada en sistemas agrícolas presenta una dependencia estacional asociada al desarrollo de los cultivos. Como consecuencia, imágenes con fechas de captura cercanas en un mismo ciclo presentan características espectrales similares, a diferencia de las imágenes separadas por largas distancias temporales. Para incorporar este contexto, se empleó una representación temporal basada en el día juliano. Para cada imagen Sentinel-2, la fecha de captura se convirtió, con la ecuación 11, a un día del año normalizado d_t , definido en el intervalo $[0,1)$:

$$d_t = \frac{DOY_t - 1}{365} \quad (11)$$

Donde DOY_t corresponde al día juliano.

Para preservar la naturaleza cíclica de los procesos estacionales, los valores d_t se transformaron en vectores sinusoidales con la ecuación 12. Esta representación genera un embedding temporal E_t compuesto por funciones seno y coseno de distintas frecuencias.

$$E_t = [\sin(k2\pi d_t), \cos(k2\pi d_t)]_{k=1}^8 \quad (12)$$

donde k representa los diferentes armónicos utilizados para describir la periodicidad anual.

Para incorporar la información estacional al proceso de aprendizaje, se implementó un mecanismo de modulación de características basado en FiLM. El vector embedding temporal E_t se proyectó mediante una capa lineal desde un espacio de 16 dimensiones hacia un vector de 512 parámetros como sigue:

$$P_t = W_f E_t + b_f \quad (13)$$

donde W_f y b_f representan los pesos y el sesgo de la capa lineal.

Posteriormente, P_t se divide en dos subconjuntos de igual tamaño correspondientes a los factores de escala γ_t con $\gamma_t \in \mathbb{R}^{256}$ y desplazamiento β_t con $\beta_t \in \mathbb{R}^{256}$. Estos parámetros son utilizados para modular las activaciones generadas por la convolución interna de la ConvLSTM mediante la operación:

$$G'_t = \gamma_t \odot G_t + \beta_t \quad (14)$$

donde G_t representa las activaciones originales de la celda recurrente.

4.3.3.5 Fusión temporal y mecanismos de atención

Con el propósito de integrar la información temporal, extraída por la capa ConvLSTM, en ambas direcciones, se concatenaron los estados ocultos a lo largo de la dimensión de canales como se muestra en la ecuación 15.

$$H_t^{cat} = [H_t^{fwd}, H_t^{bwd}] \quad \text{con } H_t^{cat} \in \mathbb{R}^{B \times 128 \times H \times W} \quad (15)$$

Posteriormente se aplicó un proceso de fusión de estados conformado por una convolución, que aplica 64 filtros de 1×1 , para proyectar H_t^{cat} a un espacio de menor dimensionalidad, lo que permitió generar $H_t^{fusion} \in \mathbb{R}^{B \times 64 \times H \times W}$. Posteriormente, se aplicó una capa de normalización GroupNorm con ocho

grupos y 64 canales de salida. Este proceso normaliza los valores de los mapas de características dentro de grupos de canales, de esta forma se reduce la sensibilidad al tamaño del lote y mejora la estabilidad del entrenamiento.

Aunque la ConvLSTM bidireccional proporciona una representación temporal enriquecida, no todas las observaciones de la secuencia contribuyen de igual manera a la reconstrucción de la fecha objetivo. Para identificar automáticamente las observaciones más relevantes, se incorporó un mecanismo de atención temporal guiado por la información estacional contenida en los embeddings del día juliano, un mecanismo de atención espectral basado en canales que procesan H_t^{fusion} paralelamente y un módulo de fusión paralela que asigna pesos de importancia de las características temporales y espectrales.

4.3.3.5.1 *Mecanismo de atención temporal*

Sea E_t el embedding asociado al instante t y E_T el embedding de la fecha objetivo. Ambos vectores son concatenados para formar una representación temporal de 32 dimensiones S_t . Esta representación es procesada mediante un perceptrón multicapa compuesto por una capa oculta de 16 neuronas con activación ReLU y una neurona de salida. Para una secuencia de longitud T , el perceptrón genera un conjunto de puntajes s :

$$s = [s_1, s_2, \dots, s_T] \quad (16)$$

Posteriormente los puntajes son normalizados mediante una función Softmax para obtener los coeficientes de atención con la expresión 17.

$$\alpha_t = \frac{e^{s_t}}{\sum_{j=1}^T e^{s_j}} \quad \text{con} \quad \sum_{t=1}^T \alpha_t = 1 \quad (17)$$

Cada coeficiente α_t representa la importancia relativa de una observación dentro de la secuencia temporal. Finalmente, la representación utilizada para la

reconstrucción espacial se calcula como una combinación ponderada de todos los estados fusionado (ecuación 18).

$$H_{final} = \sum_{t=1}^T \alpha_t H_t^{fusion} \quad (18)$$

4.3.3.5.2 Mecanismo de atención espectral

Para reforzar la representación espectral generada por el bloque ConvLSTM, se incorporó un mecanismo de atención espectral basado en canales. Este módulo permite asignar pesos adaptativos a cada característica espectral aprendida por la red. El mecanismo enfatiza aquellas más relevantes para la reconstrucción de la imagen objetivo y reduce la influencia de características menos informativas.

Sea $H_t^{fusion} \in \mathbb{R}^{B \times T \times C \times H \times W}$ la salida del bloque ConvLSTM reducida, donde B representa el tamaño del lote, T el número de observaciones temporales, C el número de canales de características y $H \times W$ las dimensiones espaciales. En una primera etapa, se obtiene una representación temporal agregada mediante el promedio de las características a lo largo de la dimensión temporal con la ecuación 19.

$$H_{mean} = \frac{1}{T} \sum_{t=1}^T H_t^{fusion} \quad (19)$$

Ahora $H_{mean} \in \mathbb{R}^{B \times C \times H \times W}$.

Posteriormente, se calculan dos descriptores globales para cada canal: el promedio global (Global Average Pooling, GAP) y el máximo global (Global Max Pooling, GMP). Estos descriptores permiten resumir la información espacial contenida en cada canal mediante las ecuaciones 20 y 21.

$$z_{avg} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W H_{mean}(c, i, j) \quad (20)$$

$$z_{max}(c) = \max_{i,j} H_{mean}(c, i, j) \quad (21)$$

Ambos vectores son procesados mediante una red perceptrón multicapa (MLP) compartida. La estructura de la MLP sigue una estructura $64 \times 8 \times 64$. Esta estrategia reduce el número de parámetros y permite modelar dependencias globales entre canales espectrales:

$$f(z) = W_2 \delta(W_1 z) \quad (22)$$

Donde W_1 y W_2 representan las matrices de pesos de la MLP y $\delta(\cdot)$ denota la función ReLU. Las respuestas obtenidas a partir de GAP y GMP son combinadas y normalizadas mediante una función sigmoide para generar los pesos de atención espectral con (23).

$$A = \sigma \left(f(z_{avg}) + f(z_{max}) \right) \quad (23)$$

donde $A \in \mathbb{R}^C$ contiene la importancia relativa estimada para cada canal. Finalmente, los pesos calculados son aplicados sobre la representación agregada temporalmente mediante una multiplicación canal por canal con la expresión 24.

$$H_{spec} = A \odot H_{mean} \quad (24)$$

donde $\odot$ representa la multiplicación elemento a elemento.

4.3.3.5.3 Módulo de fusión paralela

Con el propósito de aprovechar simultáneamente la información generada por los módulos de atención temporal y atención espectral, se incorporó un módulo de fusión paralela adaptativa. A diferencia de una combinación fija de características, este mecanismo permite que la red aprenda automáticamente la contribución relativa de cada rama durante el proceso de entrenamiento.

Sea H_{final} la representación obtenida por el módulo de atención temporal y H_{spec} la representación generada por el módulo de atención espectral. Ambas ramas producen mapas de características con dimensiones compatibles, los cuales son integrados mediante una combinación ponderada. Para ello, se define un vector de parámetros entrenables 25.

$$\theta = [\theta_t, \theta_s] \quad (25)$$

donde θ_t y θ_s representan los parámetros asociados a las ramas temporal y espectral, respectivamente. Durante el entrenamiento, estos parámetros se optimizan mediante retropropagación y descenso por gradiente. Con el fin de obtener coeficientes positivos y normalizados, se aplica una función Softmax como se muestra en (26).

$$\omega_i = \frac{\exp(\theta_i)}{\sum_{j=1}^2 \exp(\theta_j)} \quad (26)$$

obteniéndose los pesos

$$\omega_t, \omega_s \quad \text{con } \omega_t + \omega_s = 1 \quad (27)$$

La representación fusionada se calcula mediante una combinación lineal ponderada de ambas ramas de atención con 28.

$$F = \omega_t H_{final} + \omega_s H_{spec} \quad (28)$$

4.3.3.6 Módulo de reconstrucción espacial (Decoder)

La representación espaciotemporal F es procesada mediante un decoder convolucional profundo que se encarga de reconstruir los detalles espaciales ausentes en las composiciones Sentinel-2. Como se muestra en la Figura 4.5, el decoder está conformado por nueve capas convolucionales con 64 filtros de 3×3 con su correspondiente función de activación ReLU. Este tipo de convoluciones permiten transformar progresivamente esta representación espaciotemporal en

un residuo espacial. La última capa convolucional proyecta la representación latente de 64 canales hacia un espacio de salida de cinco canales, mediante la aplicación de 5 filtros de tamaño 3×3.

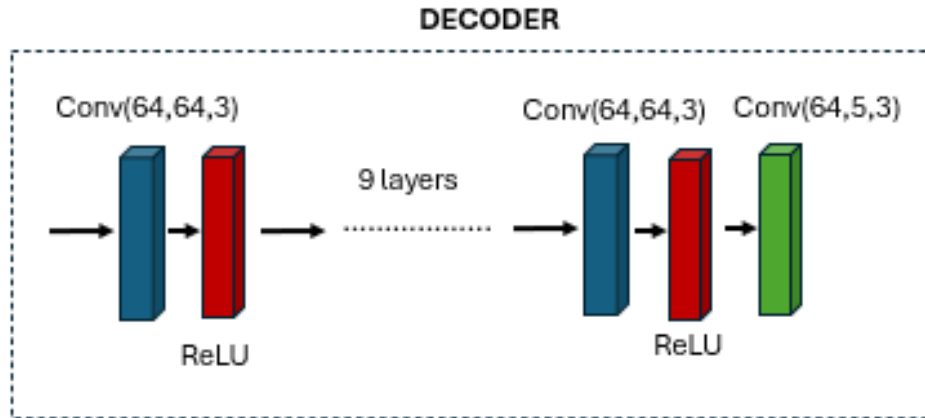


Figura 4.5. Estructura del decoder del modelo TVDSR.

Finalmente se aplica la estrategia residual propuesta en el modelo VDSR, la cual determina la salida final del modelo TVDSR mediante una conexión de salto que suma el residuo reconstruido a la última composición Sentinel-2 de la secuencia temporal, como se muestra en la expresión (29).

$$Y = X_T + R(F) \quad (29)$$

donde X_T corresponde a la imagen Sentinel-2 asociada a la fecha objetivo y Y representa la predicción final generada por TVDSR.

4.3.3.7 Función de pérdida y estrategia de entrenamiento

El modelo TVDSR fue formulado como un problema de regresión supervisada cuyo objetivo consiste en aprender una transformación espaciotemporal no lineal entre una secuencia de composiciones Sentinel-2 y la composición UAV correspondiente a la última fecha de la ventana temporal. Durante el entrenamiento, la predicción generada por la red se comparó con la imagen UAV

de referencia utilizando la función de pérdida de error cuadrático medio (Mean Square Error, MSE), definida como:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2 \quad (30)$$

donde $\hat{Y}_i$ representa el valor predicho por el modelo, Y_i corresponde al valor observado en la imagen UAV de referencia y N es el número total de píxeles evaluados.

La minimización de esta función permitió optimizar de manera conjunta todos los parámetros entrenables de la arquitectura mediante retropropagación. En consecuencia, la pérdida no solo ajustó los pesos del encoder convolucional, la ConvLSTM y el decoder, sino también los parámetros de la proyección lineal utilizada por el módulo FiLM y los pesos del perceptrón multicapa empleado en el mecanismo de atención temporal. De esta forma, el modelo completo fue entrenado de extremo a extremo.

La optimización se realizó utilizando el algoritmo Adam con una tasa de aprendizaje inicial $\eta_0 = 3 \times 10^{-4}$. Para favorecer la convergencia del modelo durante las etapas finales del entrenamiento, se empleó un programador de tasa de aprendizaje tipo Step Learning Rate Scheduler, que reduce el valor de la tasa de aprendizaje en un factor de 0.5 cada 50 épocas. Adicionalmente, se aplicó una estrategia de recorte de gradiente (gradient clipping) con una norma máxima de 0.5 para evitar inestabilidades numéricas asociadas a gradientes excesivamente grandes.

4.3.3.8 Implementación y detalles computacionales

El modelo TVDSR fue implementado en Python utilizando la biblioteca PyTorch para el desarrollo y entrenamiento de la arquitectura de aprendizaje profundo. Las operaciones de lectura y escritura de datos geoespaciales se realizaron mediante Rasterio, mientras que NumPy fue utilizado para el procesamiento

numérico de los datos de entrada. La arquitectura fue desarrollada bajo un enfoque modular que permitió integrar los componentes de extracción espacial, modelado temporal, modulación estacional y reconstrucción residual dentro de una única red entrenable de extremo a extremo.

Las muestras de entrenamiento fueron generadas mediante ventanas temporales deslizantes de longitud $T=4$ de 13 fechas. El modelo fue entrenado durante 500 épocas utilizando lotes de ocho muestras. Cada muestra estuvo compuesta por una secuencia de cuatro fechas consecutivas de Sentinel-2 y una imagen UAV correspondiente a la última fecha de la secuencia. Con la finalidad de incrementar el número de ejemplos de entrenamiento, las imágenes fueron divididas en 200 parches de 96×96 píxeles extraídos aleatoriamente. Todos los parches pertenecientes a una misma muestra temporal fueron obtenidos de la misma ubicación espacial, garantizando la coherencia espaciotemporal necesaria para el aprendizaje de la dinámica temporal.

Con la finalidad de evaluar la capacidad de generalización del modelo, las secuencias temporales disponibles fueron divididas en conjuntos de entrenamiento y validación con una proporción 80:20. La partición se realizó a nivel de secuencia, esto evita que observaciones pertenecientes a una misma serie temporal estuvieran simultáneamente presentes en ambos conjuntos. Posteriormente, los parches utilizados para el entrenamiento y la validación fueron extraídos exclusivamente de las secuencias correspondientes a cada subconjunto.

4.3.4 Validación del modelo TVDSR

Para evaluar la capacidad de generalización del modelo TVDSR, con nuevas observaciones no utilizadas durante el entrenamiento, se realizó una evaluación cuantitativa utilizando imágenes Sentinel-2 y UAV correspondientes al 22 de octubre de 2024 y el 3 de agosto de 2025 independientes del conjunto de entrenamiento. Esta estrategia permitió analizar la capacidad de reconstrucción

del modelo en escenarios representativos de aplicación real, donde las fechas de inferencia no han sido previamente observadas por la red.

Adicionalmente, las métricas obtenidas se utilizaron para comparar el desempeño de TVDSR con tres modelos representativos del estado del arte en superresolución de imágenes satelitales: DeepSent, UAV-Net y S³-ESRGAN, los cuales fueron adaptados a las características de las imágenes empleadas en este estudio. La selección de estos modelos respondió a que representan diferentes estrategias de reconstrucción. DeepSent incorpora información temporal y espectral para mejorar imágenes Sentinel-2 mediante aprendizaje profundo; UAV-Net está orientado a la integración de imágenes satelitales y UAV a través de mecanismos de fusión multiescala; mientras que S³-ESRGAN utiliza una arquitectura generativa adversarial con mecanismos de atención espacial y preservación espectral para la fusión Sentinel-2/UAV. De esta forma, la comparación permitió evaluar el desempeño de TVDSR frente a enfoques basados en integración temporal, fusión multisensor y modelos generativos, respectivamente.

Para caracterizar la calidad de las reconstrucciones se emplearon métricas espaciales y espectrales complementarias, permitiendo evaluar simultáneamente la preservación de detalles espaciales y la consistencia de la información espectral asociada a los índices de vegetación, las cuales se presentan en la Cuadro 4.2.

Cuadro 4.2. Métricas de calidad empleadas para la evaluación del desempeño de los modelos de superresolución.

Metric	Equation	Description
Root Mean Square Error	$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{N}}$	Es una métrica radiométrica que determina la raíz del error promedio cometido durante la reconstrucción que permite evaluar la fidelidad de reconstrucción de un

		método de fusión (Wald, 2002).
Mean Absolute Error	$MAE = \frac{1}{N} \sum_{i=1}^N x_i - y_i $	Es una métrica radiométrica que cuantifica la diferencia promedio absoluta entre los valores de la imagen reconstruida y los valores de la imagen de referencia.
Peak Signal-to-Noise Ratio	$PSNR = 10 \log \left(\frac{\max_I^2}{MSE} \right)$	Es una métrica radiométrica que cuantifica la relación entre la intensidad máxima posible de la señal y el nivel de ruido introducido en cada píxel, lo que permite conocer la distorsión radiométrica (Lin & Huang, 2008; Z. Liu et al., 2008).
Estructural Similarity Index Measure	$SSIM(x, y) = \frac{((2 \cdot \mu_x \cdot \mu_y + C_1) \cdot (2 \cdot \sigma_{xy} + C_2))}{((\mu_x^2 + \mu_y^2 + C_1) \cdot (\sigma_x^2 + \sigma_y^2 + C_2))}$	Es una métrica estructural que analiza la degradación de una imagen respecto a la luminancia, el contraste y la estructura.
Correlation	$corr = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$	Es una métrica radiométrica que determina si las variaciones espaciales y los patrones presentes en la imagen de referencia se mantienen en la imagen superresuelta.
Spectral Angle Mapper	$SAM = \arccos \left(\frac{\sum_{j=1}^B y_j x_j}{\sqrt{\sum_{j=1}^B y_j^2} \sqrt{\sum_{j=1}^B x_j^2}} \right)$	Es una métrica espectral global que calcula el ángulo formado por el vector espectral de referencia y el vector espectral reconstruido, lo que permite saber si se preservan las relaciones espectrales entre las bandas.

Relative Dimensionless Global Error of Synthesis	$ERGAS = 100 \cdot \frac{h}{l} \sqrt{\frac{1}{N} \sum_{j=1}^B \left(\frac{RMSE_j}{\mu_j} \right)^2}$	Es una métrica espectral global que combina los errores de todas las bandas espectrales y sus valores medios para generar un indicador que describe la calidad de reconstrucción de la imagen superresuelta (Wald, 2002).
--	---	---

x_i es la imagen superresuelta, y_i imagen de referencia, i píxel analizado, N número de píxeles por banda, max_i valor máximo posible que pueden tener los píxeles de la imagen mejorada, MSE error cuadrático medio, μ_x y μ_y son las medias de las imágenes superresuelta y de referencia; σ_x^2 y σ_y^2 son las varianzas de las imágenes superresuelta y de referencia, σ_{xy} covarianza entre los valores de los píxeles de las dos imágenes, $C_1 = (K_1 \cdot L)^2$, $C_2 = (K_2 \cdot L)^2$ con $K_1 = 0.01$, $K_2 = 0.03$ y $L=2$ son constantes; $\bar{x}$ y $\bar{y}$ son las medias de los píxeles contenidos en las imágenes superresuelta y de referencia respectivamente; $B=5$ número de bandas, j número de banda analizada, $\frac{h}{l}$ relación entre resoluciones espaciales $h=high$ y $l=low$ y μ_j es la media de los píxeles de la banda j .

4.4 Resultados y discusión

4.4.1 Complejidad computacional de los modelos evaluados.

Se analizó la complejidad computacional de las arquitecturas DeepSent, TVDSR Temporal, UAV-Net y S³-ESRGAN mediante el número de parámetros entrenables, tiempo de entrenamiento, tiempo de inferencia, velocidad de procesamiento (throughput), uso máximo de memoria y costo computacional expresado en GFLOPs.

Los resultados mostraron diferencias importantes entre los modelos (Cuadro 4.3). DeepSent presentó la arquitectura más compacta, el menor consumo de memoria y la mayor velocidad de procesamiento. Por su parte, UAV-Net presentó el mayor número de parámetros entrenables, aproximadamente 21 veces más que el modelo TVDSR. Sin embargo, el costo computacional de UAV-Net fue el más bajo, además registró el tiempo de entrenamiento más bajo. Este comportamiento se atribuye a la estrategia de downsampling implementada en su módulo encoder, la cual reduce progresivamente el tamaño espacial de los mapas de características y disminuye el número de operaciones requeridas

durante la propagación hacia adelante. El modelo TVDSR Temporal presentó una complejidad intermedia con valores de tiempo de entrenamiento, velocidad costo computacional mayores a los modelos DeepSent y UAV-Net.

Cuadro 4.3. Desempeño y costo computacional de los modelos.

Modelo	Parámetros entrenables	Tiempo de entrenamiento (min)	Tiempo de inferencia (seg)	Velocidad (iter/s)	Memoria Pico (MB)	Costo computacional (GFLOPs)
DeepSent	882,437	32.20	4.4	136.46	87.10	7.71
TVDSR	1,288,304	71.40	4.2	40.32	180.50	33.31
UAV-Net	27,001,285	30.63	4.1	57.72	516.40	4.56
S³-ESRGAN	9,845,801	406.50	4.3	13.17	552.90	71.22

GFLOPs: Giga Floating-Point Operations Per Second, iter: iteración, MB: Megabytes, min: minutos, seg: segundos.

4.4.2 Desempeño de los modelos de superresolución.

El Cuadro 4.4 presenta el desempeño cuantitativo de los modelos S³-ESRGAN, DeepSent, UAV-Net y TVDSR al inferir las fechas de evaluación. Los resultados mostraron que UAV-Net obtuvo el mejor desempeño global para la imagen de 2024, al alcanzar los menores errores radiométricos y las mejores métricas de reconstrucción espacial. No obstante, esta superioridad no fue uniforme para todas las propiedades evaluadas. DeepSent presentó el menor SAM, lo que evidencia una mayor preservación de las relaciones espectrales entre las bandas, mientras que DeepSent y TVDSR alcanzaron ligeramente mayores coeficientes de correlación que UAV-Net. Estos resultados ponen de manifiesto que un mejor desempeño en métricas orientadas a la reconstrucción espacial no implica necesariamente una conservación equivalente de la información espectral. Esta observación coincide con lo señalado por Aybar et al. (2026) quienes advierten que los métodos de superresolución pueden incrementar el nivel de detalle espacial a costa de introducir distorsiones radiométricas o inconsistencias

espectrales cuando dichas propiedades no son consideradas explícitamente durante el entrenamiento.

El comportamiento observado para el 3 de agosto de 2025 fue diferente al de la primera fecha de evaluación. TVDSR alcanzó el mejor desempeño global, al obtener simultáneamente los menores errores radiométricos, espectrales y los mayores niveles de similitud estructural. Estos resultados sugieren que la incorporación explícita de dependencias temporales como ConvLSTM bidireccional favorece la reconstrucción bajo condiciones distintas a las observadas durante el entrenamiento. En este sentido, Chang & Luo (2019) reportaron mejoras en la calidad de reconstrucción al incorporar estructuras BiConvLSTM capaces de integrar información complementaria procedente de múltiples observaciones. De manera similar, Michel & Inglada (2026) demostraron que la integración conjunta de información multitemporal y multirresolución puede superar enfoques que explotan de forma independiente las dimensiones espacial, espectral o temporal.

El cambio de comportamiento entre ambas fechas constituye uno de los resultados más relevantes de este estudio. En escenarios de fusión de imágenes Sentinel-2 y UAV, donde la dinámica fenológica, los cambios en la cobertura vegetal y las diferencias radiométricas entre adquisiciones pueden modificar significativamente la respuesta espectral de la superficie, la capacidad de explotar simultáneamente información previa y posterior representa una ventaja importante para la reconstrucción de imágenes de alta resolución. En conjunto, los resultados evidencian un compromiso entre reconstrucción espacial, preservación espectral y capacidad de generalización. UAV-Net obtuvo los mejores resultados espaciales en la primera fecha, DeepSent preservó mejor la información espectral y TVDSR mostró el desempeño más estable entre fechas y el mejor rendimiento global en la segunda evaluación.

Cuadro 4.4. Resultados cuantitativos consolidados para las dos fechas de evaluación.

Fecha	Modelo	RMSE	MAE	PSNR (dB)	SSIM	Corr	ERGAS	SAM (°)
22/10/2024	S ³ -ESRGAN	0.154	0.117	22.5	0.79	0.8	32.42	14.96
22/10/2024	DeepSent	0.152	0.128	22.45	0.85	0.88	32.05	9.86
22/10/2024	UAV-Net	0.128	0.100	24.06	0.86	0.87	27.06	12.89
22/10/2024	TVDSR	0.152	0.126	22.8	0.84	0.88	32.88	17.71
03/08/2025	S ³ -ESRGAN	0.203	0.147	19.94	0.69	0.73	47.79	26.23
03/08/2025	DeepSent	0.182	0.129	20.83	0.78	0.8	43.13	19.58
03/08/2025	UAV-Net	0.181	0.139	20.92	0.75	0.78	41.76	23.29
03/08/2025	TVDSR	0.149	0.115	22.6	0.85	0.87	34.94	18.19

Corr: Correlación, dB: decibeles, °: grados

4.4.3 Capacidad de generalización temporal.

El Cuadro 4.5 presenta como variaron las métricas analizadas entre las dos fechas de validación. Mientras S³-ESRGAN, DeepSent y UAV-Net experimentaron degradaciones significativas en varias métricas, TVDSR mantuvo un desempeño prácticamente constante entre ambas fechas. En términos generales, los incrementos observados en RMSE, MAE, ERGAS y SAM, junto con las reducciones en PSNR, SSIM y correlación, indican que la fecha de agosto de 2025 representó un escenario más complejo para el proceso de superresolución. Dado que ambas evaluaciones corresponden a diferentes momentos dentro de la misma serie temporal, los resultados sugieren que las variaciones ocurridas en la superficie tuvieron un impacto directo sobre la capacidad de reconstrucción de los modelos.

Los resultados evidenciaron diferencias importantes en la capacidad de generalización temporal de los modelos. S³-ESRGAN presentó la mayor degradación entre fechas, con incrementos sustanciales en los errores radiométricos y espectrales, lo que sugiere que la incorporación de información espacial y espectral no fue suficiente para capturar adecuadamente la

variabilidad temporal de la escena. DeepSent mostró un comportamiento más estable, especialmente en términos de MAE, aunque registró incrementos importantes en RMSE, ERGAS y SAM, lo que indica una pérdida apreciable de fidelidad espectral al extrapolar hacia condiciones temporales diferentes. UAV-Net, pese a haber obtenido el mejor desempeño global en la evaluación de 2024, experimentó una degradación considerable en todas las métricas durante la evaluación de 2025, lo que demuestra que un excelente desempeño en una fecha específica no garantiza una elevada capacidad de generalización temporal. En contraste, TVDSR presentó las menores variaciones entre fechas, mantuvo constantes las métricas de error, similitud estructural y preservación espectral. Particularmente, los cambios observados en RMSE, PSNR, SSIM, correlación, ERGAS y SAM fueron significativamente inferiores a los registrados por los demás modelos, lo que evidencia una mayor robustez frente a variaciones temporales.

Cuadro 4.5. Variación porcentual del desempeño entre las fechas de evaluación (2024–2025).

Modelo	Δ RMSE (%)	Δ MAE (%)	Δ PSNR (%)	Δ SSIM (%)	Δ Corr (%)	Δ ERGAS (%)	Δ SAM (%)
S ³ -ESRGAN	31.8	25.6	-11.4	-12.7	-8.8	47.4	75.3
DeepSent	19.7	0.8	-7.2	-8.2	-9.1	34.6	98.6
UAV-Net	41.4	39	-13.1	-12.8	-10.3	54.3	80.7
TVDSR	-2	-8.7	-0.9	1.2	-1.1	6.3	2.7

Δ : variación, Corr: Correlación

La capacidad de generalización constituye uno de los principales desafíos de los modelos de superresolución aplicados a teledetección. Xiong et al. (2020) demostraron que el desempeño de los modelos puede variar cuando estos son transferidos fuera de las condiciones empleadas durante el entrenamiento, particularmente ante cambios de localización geográfica y sensor. Los autores indican que los métodos basados en aprendizaje profundo dependen en gran medida de la representatividad de los datos de entrenamiento y encuentran

dificultades cuando enfrentan condiciones de adquisición diferentes a las observadas durante la fase de aprendizaje.

La dimensión temporal adquiere especial importancia en entornos agrícolas debido a que las características espectrales y estructurales de la superficie evolucionan continuamente como resultado de procesos fenológicos, ambientales y de manejo. Xiao et al. (2022) plantean que para conservar el rendimiento de los modelos bajo condiciones variables deben capturarse adecuadamente las relaciones espaciotemporales. Asimismo, señalan que una descripción insuficiente de estas relaciones limita la capacidad de generalización del modelo. Wang et al. (2022) identifican la generalización y la adaptación a escenarios reales como desafíos prioritarios dentro de la superresolución de imágenes satelitales. De forma similar, H. Chen et al. (2024) sostienen que persiste una brecha importante entre los entornos de entrenamiento y las condiciones reales debido a la complejidad y diversidad de las degradaciones presentes en aplicaciones prácticas. En consecuencia, un desempeño elevado en una fecha específica no implica necesariamente una capacidad equivalente de reconstrucción cuando las condiciones de la escena presentan cambios temporales.

Bajo esta perspectiva, la degradación observada en S³-ESRGAN, DeepSent y UAV-Net entre las fechas de evaluación puede interpretarse como una manifestación de una capacidad limitada para enfrentar distribuciones no observadas durante el entrenamiento. Los incrementos registrados en RMSE, MAE, ERGAS y SAM, junto con las disminuciones en PSNR, SSIM y correlación, sugieren que la fecha de 2025 representó un escenario más alejado del dominio aprendido por dichos modelos, lo que dificultó la preservación simultánea de la fidelidad radiométrica, estructural y espectral. Por el contrario, TVDSR mostró un comportamiento estable, con variaciones mínimas en todas las métricas analizadas. Esta estabilidad sugiere la presencia de representaciones menos

dependientes de las condiciones particulares observadas durante el entrenamiento y más resistentes frente a cambios temporales de la escena.

4.4.4 Desempeño de los modelos de superresolución por índice de vegetación.

Se evaluó el efecto de la superresolución sobre variables derivadas de la reflectancia debido a que numerosos productos utilizados en monitoreo agrícola, estimación de biomasa, estado hídrico y contenido de clorofila se calculan a partir de índices espectrales. Por esta razón, la preservación de estos índices constituye un indicador complementario de la capacidad de los modelos para conservar información biofísica relevante después del proceso de reconstrucción. Para ello se utilizaron cinco índices ampliamente empleados en agricultura y teledetección: NDVI, SAVI, GNDVI, NDWI y NDVire. Con el propósito de sintetizar el comportamiento de los modelos a través de múltiples métricas, se calculó un puntaje compuesto definido como $Composite\ score = (1 - RMSE) + SSIM + Corr$. La Figura 4.6 resume los resultados obtenidos para las dos fechas de evaluación mediante mapas de calor.

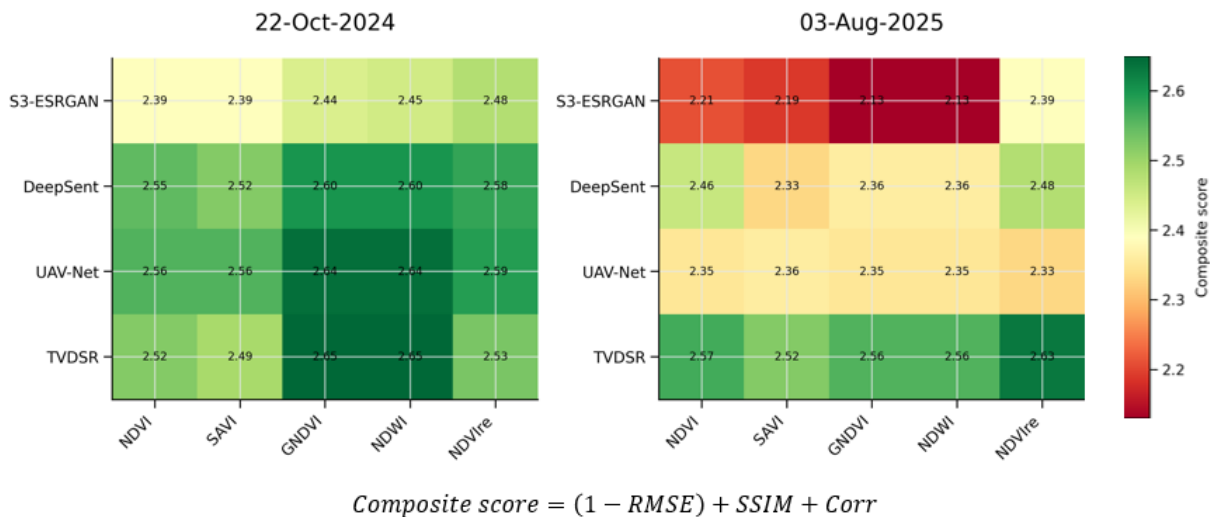


Figura 4.6. Mapa de calor del desempeño de los modelos de superresolución por índice de vegetación.

La evaluación correspondiente al 2024 muestra que UAV-Net obtuvo el mejor desempeño global en la mayoría de los índices analizados. Los mayores valores del puntaje compuesto se observaron para NDVI, SAVI y NDVI_{re}, mientras que para GNDVI y NDWI los resultados de UAV-Net fueron similares a los obtenidos por TVDSR. Estos resultados indican una elevada capacidad para preservar las relaciones espectrales necesarias para la estimación precisa de índices biofísicos bajo condiciones temporales similares a las observadas durante el entrenamiento. DeepSent mostró resultados intermedios, mantuvo valores homogéneos en todos los índices evaluados, con puntajes comprendidos entre 2.52 y 2.60. En contraste, S³-ESRGAN registró los puntajes más bajos en todos los índices, particularmente para NDVI y SAVI, lo que refleja una menor capacidad para conservar las relaciones espectrales requeridas para la estimación de variables biofísicas.

La situación cambió de manera importante durante la evaluación del año 2025. En esta fecha, TVDSR presentó el mejor desempeño para los cinco índices analizados. El modelo alcanzó los mayores puntajes compuestos. TVDSR presentó los mejores puntajes en NDVI y NDVI_{re}, donde se observan las mayores diferencias respecto a los demás modelos. Estos resultados muestran una elevada capacidad para preservar la información biofísica incluso cuando las condiciones temporales de la escena difieren significativamente de las observadas inicialmente.

DeepSent se posicionó como el segundo mejor modelo durante esta segunda evaluación. Aunque sus puntajes fueron inferiores a los de TVDSR, mantuvo valores relativamente estables para todos los índices, especialmente en NDVI y NDVI_{re}. En contraste, UAV-Net experimentó una degradación notable respecto a la primera fecha de evaluación. Sus puntajes disminuyeron en todos los índices, especialmente en NDVI_{re}, donde pasó de 2.59 a 2.33. Un comportamiento similar se observó en GNDVI y NDWI, cuyos puntajes se redujeron aproximadamente el 11%. La degradación más severa se observó en S³-ESRGAN, cuyos puntajes

disminuyeron de manera consistente en todos los índices. Los menores valores se registraron para GNDVI y NDWI. El modelo presenta una pérdida considerable de capacidad para preservar las relaciones espectrales asociadas al contenido de clorofila y al estado hídrico de la vegetación.

4.4.5 Evaluación visual de las imágenes reconstruidas.

La Figura 4.7 presenta una comparación visual entre la imagen UAV de referencia, la imagen Sentinel-2 original y las reconstrucciones generadas por S³-ESRGAN, DeepSent, UAV-Net y TVDSR para la fecha del 3 de agosto de 2025. Las regiones señaladas en las imágenes permiten analizar diferencias en la recuperación de bordes parcelarios, patrones internos de cultivo y áreas afectadas por degradaciones presentes en la imagen de entrada. De manera general, todos los modelos mejoraron significativamente la calidad visual respecto a la imagen Sentinel-2 original. Los modelos recuperaron parte de la estructura espacial observada en la imagen UAV. No obstante, la calidad de la reconstrucción presentó diferencias notables entre arquitecturas, particularmente en la definición de límites parcelarios y en la preservación de la variabilidad espectral dentro de las parcelas agrícolas.

S³-ESRGAN mostró el desempeño visual más limitado. Aunque el modelo incrementó la resolución aparente de la escena, las reconstrucciones presentan suavizado excesivo, pérdida de detalle fino y menor definición de los patrones internos de los cultivos. Esta situación es especialmente evidente en las regiones 2 y 3, donde los límites entre parcelas aparecen menos definidos que en la referencia UAV. Además, se observan artefactos cromáticos y anomalías espectrales en la región 4, lo que sugiere una mayor sensibilidad a las degradaciones presentes en la imagen original.

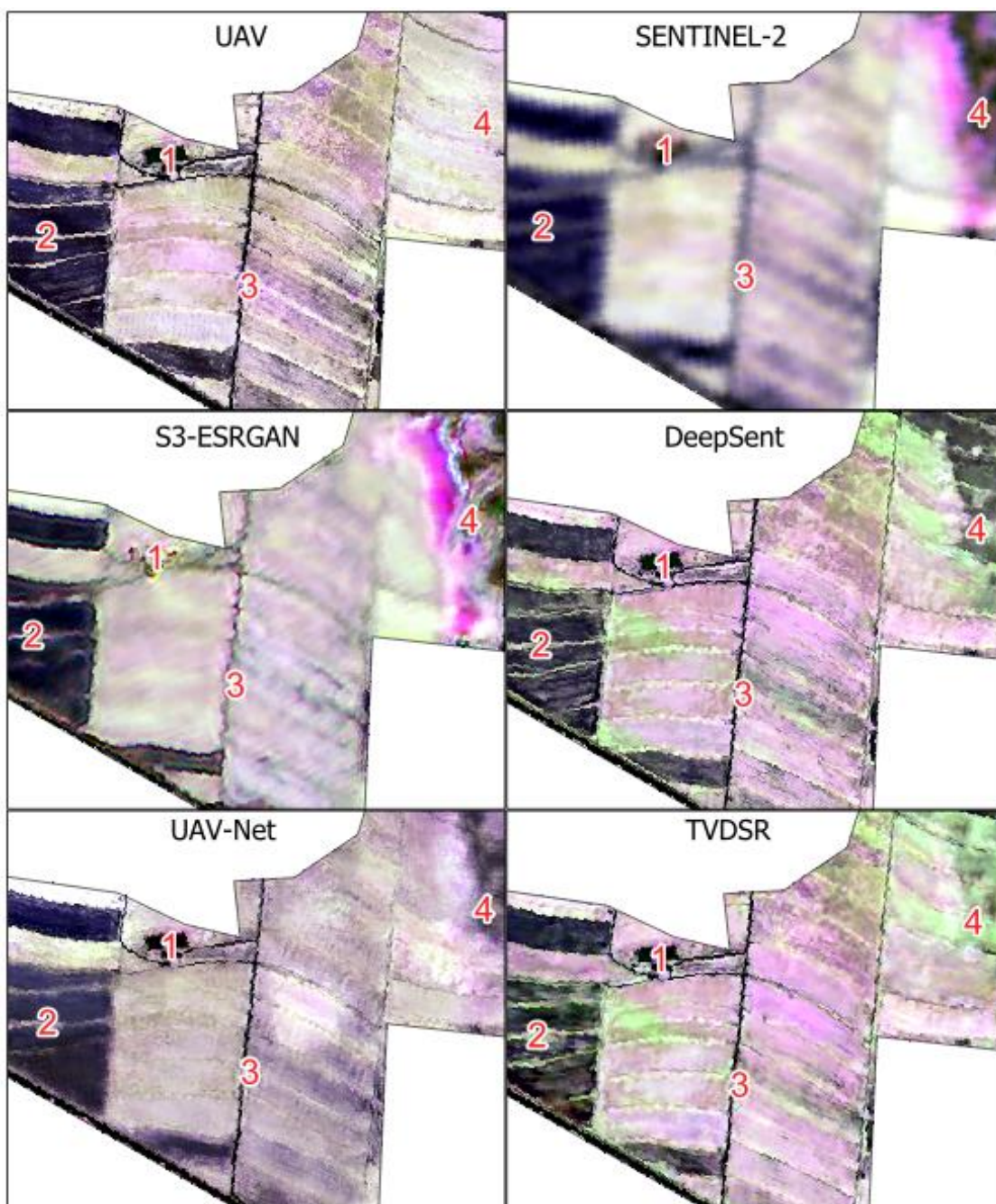


Figura 4.7. Comparación visual de las imágenes reconstruidas por los modelos de superresolución para la fecha del 3 de agosto de 2025.

DeepSent logró una recuperación más consistente de la estructura agrícola. El modelo reprodujo con mayor fidelidad los límites parcelarios y parte de la heterogeneidad interna de los cultivos. Las regiones 1 y 2 muestran una definición

espacial notablemente superior a la obtenida por S³-ESRGAN, mientras que los gradientes espectrales dentro de las parcelas presentan una distribución más próxima a la observada en la imagen UAV. En la región 3 se observa una adecuada definición de caminos y los límites parcelarios. Por su parte, UAV-Net produjo reconstrucciones visualmente detalladas y con una adecuada representación de la estructura general de la escena. Los límites entre parcelas son claramente identificables, aunque menos definidas que en la imagen generada por DeepSent. La respuesta radiométrica mantiene una buena correspondencia con la referencia. Sin embargo, algunas texturas agrícolas aparecen más homogéneas que en la imagen UAV y parte de la degradación observada en la región 4 permanece visible en la reconstrucción.

Los resultados más favorables se obtuvieron con TVDSR, cuya reconstrucción presenta la mayor similitud visual con la imagen UAV. En las regiones 1, 2 y 3 se aprecia una mejor recuperación de los bordes parcelarios, una mayor continuidad espacial de las estructuras agrícolas y una representación más detallada de la variabilidad interna de los cultivos. Asimismo, la distribución espacial de los patrones espectrales es más consistente con la observada en la referencia de alta resolución. Cabe resaltar que en la región 1 la definición de parcelas y de caminos fue similar a la región generada por DeepSent

Un aspecto particularmente relevante se observa en la región 4, donde la imagen Sentinel-2 presenta una zona afectada por nubosidad o por los efectos radiométricos asociados a la presencia de nubes. Esta anomalía permanece claramente visible en las reconstrucciones generadas por S³-ESRGAN y UAV-Net, ambos modelos conservan parte de la información degradada contenida en la imagen de entrada. En contraste, DeepSent y TVDSR muestran una reducción parcial de este efecto, los modelos recuperaron patrones espaciales y tonalidades más coherentes con los observados en la imagen UAV de referencia. Aunque la eliminación de la degradación no es completa, ambas arquitecturas reconstruyen parcialmente las zonas afectadas.

En conjunto, la evaluación visual confirma las tendencias observadas previamente en las métricas de calidad de imagen y en los índices de vegetación. Mientras que todos los modelos mejoraron la resolución espacial de la imagen Sentinel-2, TVDSR y DeepSent mostraron una mayor capacidad para preservar simultáneamente la estructura espacial y la coherencia espectral bajo condiciones temporalmente más complejas. La capacidad de TVDSR para recuperar detalles finos y reducir parcialmente los efectos de degradaciones locales refuerza los resultados cuantitativos obtenidos en las Secciones anteriores, donde este modelo presentó el comportamiento más robusto frente a variaciones temporales entre fechas de observación.

Aunque los resultados obtenidos demuestran el potencial de los modelos evaluados para la superresolución de imágenes Sentinel-2 mediante información UAV, existen aspectos que pueden ser mejorados en investigaciones futuras. Una de las principales limitaciones de este estudio corresponde a la disponibilidad de datos para el entrenamiento, ya que únicamente se contó con 13 imágenes UAV, de las cuales diez fueron adquiridas durante 2024 y tres durante 2025. Si bien este conjunto de datos permitió entrenar satisfactoriamente los modelos y obtener reconstrucciones de alta calidad, la limitada representación temporal disponible pudo haber reducido su capacidad para capturar completamente la variabilidad fenológica y radiométrica presente en la zona de estudio. Esta situación pudo haber contribuido, al menos parcialmente, a la degradación observada en algunos modelos durante la evaluación realizada en agosto de 2025.

No obstante, aunque TVDSR mostró la mayor estabilidad entre las fechas evaluadas y presentó la menor degradación de desempeño respecto a los demás modelos, los resultados disponibles aún no permiten afirmar de manera concluyente que se trata de una arquitectura inherentemente robusta frente a la variabilidad temporal. La evaluación se realizó únicamente sobre dos fechas de observación, lo que limita la capacidad para caracterizar adecuadamente su

comportamiento ante cambios fenológicos, climáticos y de manejo a lo largo de ciclos agrícolas completos. En consecuencia, estudios futuros deberán incorporar un mayor número de fechas distribuidas a través de diferentes etapas fenológicas y temporadas de cultivo, con el fin de determinar si la estabilidad observada corresponde a una propiedad consistente del modelo o a un comportamiento específico de las condiciones evaluadas en este trabajo.

Asimismo, el desarrollo del presente trabajo se realizó con información procedente de una única zona agrícola, por lo que la capacidad de generalización espacial de los modelos aún requiere ser evaluada en escenarios más diversos. Diferencias en tipos de cultivo, prácticas de manejo, condiciones climáticas, propiedades del suelo y patrones fenológicos podrían influir en el comportamiento de las arquitecturas de superresolución cuando son aplicadas en otras regiones.

4.5 Conclusiones

Este estudio demostró la viabilidad de emplear modelos de superresolución basados en aprendizaje profundo para incrementar la resolución espacial de imágenes Sentinel-2 mediante información UAV. Los resultados confirman que este tipo de enfoques permite generar productos con mayor utilidad para aplicaciones agrícolas, ya que conservan información espectral relevante para el cálculo de variables biofísicas derivadas de la reflectancia. Además, evidenció que la evaluación de modelos de superresolución no debe limitarse a métricas de calidad de imagen. La preservación de índices biofísicos constituye un criterio complementario cuando los productos generados se enfocan al monitoreo de cultivos, ya que permite determinar si la información espectral mantiene utilidad para aplicaciones agronómicas posteriores.

Los resultados reflejan la importancia de considerar la dimensión temporal en la evaluación de modelos de superresolución para teledetección agrícola. Las diferencias observadas entre fechas indican que el comportamiento de los

modelos puede variar cuando cambian las condiciones de la superficie, por lo que los procedimientos de validación convencional podrían no reflejar completamente su desempeño en escenarios operativos. En este sentido, la incorporación de protocolos de validación temporal aporta una perspectiva más realista sobre la aplicabilidad práctica de estas metodologías.

TVDSR mostró un comportamiento consistente durante las evaluaciones realizadas, lo que sugiere que la incorporación explícita de mecanismos temporales y espectrales es una estrategia potencial para el desarrollo de modelos de superresolución aplicados a agricultura de precisión. Sin embargo, la evidencia obtenida debe considerarse preliminar, ya que la evaluación temporal se realizó únicamente sobre dos fechas. Por tanto, se requiere un conjunto más amplio de observaciones distribuidas a lo largo de diferentes etapas fenológicas, campañas agrícolas y condiciones ambientales para determinar con mayor certeza su capacidad de generalización temporal.

Este trabajo aporta evidencia acerca de la necesidad de integrar criterios radiométricos, estructurales, espectrales y biofísicos dentro de esquemas de evaluación unificados. Este enfoque puede contribuir al desarrollo de modelos de superresolución más robustos y facilitar su adopción en aplicaciones de monitoreo multitemporal de cultivos basadas en imágenes Sentinel-2 y UAV.

4.6 Referencias

- Aira, L. S., Valsesia, D., Molini, A. B., Fracastoro, G., Magli, E., & Mirabile, A. (2024). Deep 3D World Models for Multi-Image Super-Resolution Beyond Optical Flow. *IEEE Access*, 12, 188902–188913. <https://doi.org/10.1109/ACCESS.2024.3514188>
- An, T., Zhang, X., Huo, C., Xue, B., Wang, L., & Pan, C. (2022). TR-MISR: Multiimage Super-Resolution Based on Feature Fusion With Transformers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 1373–1388. <https://doi.org/10.1109/JSTARS.2022.3143532>
- Aybar, C., Contreras, J., Donike, S., Portalés-Julià, E., Mateo-García, G., & Gómez-Chova, L. (2026). A radiometrically and spatially consistent super-resolution

- framework for Sentinel-2. *Remote Sensing of Environment*, 334. <https://doi.org/10.1016/j.rse.2025.115222>
- Bazrafkan, A., Igathinathane, C., Bandillo, N., & Flores, P. (2025). Optimizing integration techniques for UAS and satellite image data in precision agriculture — a review. *Frontiers in Remote Sensing*, 6. <https://doi.org/10.3389/frsen.2025.1622884>
- Bhat, G., Danelljan, M., Van Gool, L., & Timofte, R. (2021). Deep Burst Super-Resolution. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9205–9214. <https://doi.org/10.1109/CVPR46437.2021.00909>
- Bhat, G., Danelljan, M., Yu, F., Van Gool, L., & Timofte, R. (2021). Deep Reparametrization of Multi-Frame Super-Resolution and Denoising. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2440–2450. <https://doi.org/10.1109/ICCV48922.2021.00246>
- Bordone Molini, A., Valsesia, D., Fracastoro, G., & Magli, E. (2020). DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3644–3656. <https://doi.org/10.1109/TGRS.2019.2959248>
- Cao, H., Tan, Z., Zhu, B., Xiong, H., & Lv, Q. (2026). A Review on Super-Resolution Reconstruction of Single-Frame Remote Sensing Images via Diffusion Models. *Remote Sensing*, 18(11), 1702. <https://doi.org/10.3390/rs18111702>
- Chan, K. C. K., Wang, X., Yu, K., Dong, C., & Loy, C. C. (2021). BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4945–4954. <https://doi.org/10.1109/CVPR46437.2021.00491>
- Chan, K. C. K., Zhou, S., Xu, X., & Loy, C. C. (2022). BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5962–5971. <https://doi.org/10.1109/CVPR52688.2022.00588>
- Chang, Y., & Luo, B. (2019). Bidirectional convolutional LSTM neural network for remote sensing image super-resolution. *Remote Sensing*, 11(20). <https://doi.org/10.3390/rs11202333>
- Chen, H., Li, W., Gu, J., Ren, J., Sun, H., Zou, X., Zhang, Z., Yan, Y., & Zhu, L. (2024). Low-Res Leads the Way: Improving Generalization for Super-Resolution by Self-Supervised Learning. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25857–25867. <https://doi.org/10.1109/CVPR52733.2024.02443>
- Choi, Y. J., & Kim, B. G. (2026). Hierarchical recurrent transformer network for video super-resolution. *Engineering Applications of Artificial Intelligence*, 166. <https://doi.org/10.1016/j.engappai.2025.113714>
- Deudon, M., Kalaitzis, A., Goytom, I., Arefin, M. R., Lin, Z., Sankaran, K., Michalski, V., Kahou, S. E., Cornebise, J., & Bengio, Y. (2020). *HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery*. <http://arxiv.org/abs/2002.06460>

- Di, X., Peng, L., Xia, P., Li, W., Pei, R., Cao, Y., Wang, Y., & Zha, Z.-J. (2025). QMambaBSR: Burst Image Super-Resolution with Query State Space Model. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23080–23090. <https://doi.org/10.1109/CVPR52734.2025.02149>
- Ding, X., Wang, W., Zhang, B., & Zhang, J. (2025). Attention-based multi-image super-resolution reconstruction for remote sensing. *Signal, Image and Video Processing*, 19(6). <https://doi.org/10.1007/s11760-025-04080-0>
- Dong, C., Loy, C. C., He, K., & Tang, X. (2014). *Learning a Deep Convolutional Network for Image Super-Resolution* (pp. 184–199). https://doi.org/10.1007/978-3-319-10593-2_13
- Dudhane, A., Zamir, S. W., Khan, S., Khan, F. S., & Yang, M.-H. (2025). Burst Image Restoration and Enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(11), 9454–9467. <https://doi.org/10.1109/TPAMI.2024.3356188>
- Elad, M., & Feuer, A. (1997). Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12), 1646–1658. <https://doi.org/10.1109/83.650118>
- Elad, M., & Feuer, A. (1999). Super-Resolution Reconstruction of Image Sequences. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 21(21), 817–834.
- Farah, S., Saadaoui, H., & Rhinane, H. (2026). Enhancing Spatial Resolution of Sentinel-2 Imagery through Deep Learning and Generative Adversarial Networks: GS-SRGAN. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W17-2025, 123–134. <https://doi.org/10.5194/isprs-archives-xxviii-4-w17-2025-123-2026>
- Farsiu, S., Robinson, M. D., Elad, M., & Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10), 1327–1344. <https://doi.org/10.1109/TIP.2004.834669>
- Fernandez-Beltran, R., Latorre-Carmona, P., & Pla, F. (2016). REVIEW Single-Frame Super-Resolution in Remote Sensing: a Practical Overview. In *International Journal of Remote Sensing IJRS-SR-FernandezBeltranLatorreCarmonaPla-Final-NoBlack* (Vol. 00, Number 00). <http://database.eohandbook.com/database/missiontable.aspx>
- Fu, C., Yuan, H., Jiang, S., Zhang, G., Shen, L., & Hamzaoui, R. (2025). Global Spatial-Temporal Information-Based Residual ConvLSTM for Video Space-Time Super-Resolution. *IEEE Transactions on Multimedia*, 27, 5212–5224. <https://doi.org/10.1109/TMM.2025.3542970>
- Galar, M., Sesma, R., Ayala, C., Albizua, L., & Aranda, C. (2020). Super-resolution of Sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18). <https://doi.org/10.3390/RS12182941>
- Gargiulo, M., Mazza, A., Gaetano, R., Ruello, G., & Scarpa, G. (2019). Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks. *Remote Sensing*, 11(22). <https://doi.org/10.3390/rs11222635>

- Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent Back-Projection Network for Video Super-Resolution. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3892–3901. <https://doi.org/10.1109/CVPR.2019.00402>
- Heng, H., Ye, H., & Huang, R. (2020). Defocus Blur Detection by Fusing Multiscale Deep Features with Conv-LSTM. *IEEE Access*, 8, 97279–97288. <https://doi.org/10.1109/ACCESS.2020.2996200>
- Irani, M., & Peleg, S. (1990). Super resolution from image sequences. *[1990] Proceedings. 10th International Conference on Pattern Recognition, ii*, 115–120. <https://doi.org/10.1109/ICPR.1990.119340>
- Jia, D., Song, C., Cheng, C., Shen, S., Ning, L., & Hui, C. (2020). A novel deep learning-based spatiotemporal fusion method for combining satellite images with different resolutions using a two-stream convolutional neural network. *Remote Sensing*, 12(4). <https://doi.org/10.3390/rs12040698>
- Jo, Y., Oh, S. W., Kang, J., & Kim, S. J. (2018). Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3224–3232. <https://doi.org/10.1109/CVPR.2018.00340>
- Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video Super-Resolution With Convolutional Neural Networks. *IEEE Transactions on Computational Imaging*, 2(2), 109–122. <https://doi.org/10.1109/TCI.2016.2532323>
- Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1646–1654. <https://doi.org/10.1109/CVPR.2016.182>
- Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., & Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 305–319. <https://doi.org/10.1016/j.isprsjprs.2018.09.018>
- Latte, N., & Lejeune, P. (2020). PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. *Remote Sensing*, 12(15). <https://doi.org/10.3390/RS12152366>
- Li, C., Song, S., Wang, X., Liu, Y., & Bi, L. (2024). RCFNC: a resolution and contrast fusion network with ConvLSTM for low-light image enhancement. *Visual Computer*, 40(4), 2793–2806. <https://doi.org/10.1007/s00371-023-02986-9>
- Li, J., Lv, Q., Zhang, W., Zhu, B., Zhang, G., & Tan, Z. (2023). Multi-Attention Multi-Image Super-Resolution Transformer (MAST) for Remote Sensing. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174183>
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image Restoration Using Swin Transformer. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1833–1844. <https://doi.org/10.1109/ICCVW54120.2021.00210>

- Liao, R., Tao, X., Li, R., Ma, Z., & Jia, J. (2015). Video Super-Resolution via Deep Draft-Ensemble Learning. *2015 IEEE International Conference on Computer Vision (ICCV)*, 531–539. <https://doi.org/10.1109/ICCV.2015.68>
- Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1132–1140. <https://doi.org/10.1109/CVPRW.2017.151>
- Lin, P.-L., & Huang, P.-Y. (2008). Fusion methods based on dynamic-segmented morphological wavelet or cut and paste for multifocus images. *Signal Processing*, 88(6), 1511–1527. <https://doi.org/10.1016/j.sigpro.2007.12.018>
- Liu, W., Che, S., Wang, W., Du, Y., Tuo, Y., & Zhang, Z. (2025). Research on remote sensing multi-image super-resolution based on N. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-93049-7>
- Liu, Z., Forsyth, D. S., & Laganière, R. (2008). A feature-based metric for the quantitative evaluation of pixel-level image fusion. *Computer Vision and Image Understanding*, 109(1), 56–68. <https://doi.org/10.1016/j.cviu.2007.04.003>
- Luo, Z., Yu, L., Mo, X., Li, Y., Jia, L., Fan, H., Sun, J., & Liu, S. (2021). EBSR: Feature enhanced burst super-resolution with deformable alignment. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 471–478. <https://doi.org/10.1109/CVPRW53098.2021.00058>
- Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., & Fritsch, F. B. (2020). Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sensing*, 12(9). <https://doi.org/10.3390/RS12091357>
- Masrur, A., Olsen, P., Jackson, C., & Adler, P. (2026). Learning to See More: A Spectral Extension Super-Resolution Framework for Harmonized Satellite-UAS Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2026.3667863>
- Mazzia, V., Comba, L., Khaliq, A., Chiaberge, M., & Gay, P. (2020). UAV and machine learning based refinement of a satellite-driven vegetation index for precision agriculture. *Sensors (Switzerland)*, 20(9). <https://doi.org/10.3390/s20092530>
- Michel, J., & Inglada, J. (2026). Temporal attention multi-resolution fusion of satellite image time-series, applied to Landsat-8/9 and Sentinel-2: all bands, any time, at best spatial resolution. *Remote Sensing of Environment*, 334. <https://doi.org/10.1016/j.rse.2025.115159>
- Molini, A. B., Valsesia, D., Fracastoro, G., & Magli, E. (2020). Deepsum++: Non-Local Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 609–612. <https://doi.org/10.1109/IGARSS39084.2020.9324418>
- Palsson, F., Sveinsson, J. R., & Ulfarsson, M. O. (2018). Sentinel-2 image fusion using a deep residual network. *Remote Sensing*, 10(8). <https://doi.org/10.3390/rs10081290>
- Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, A. (2017). *FiLM: Visual Reasoning with a General Conditioning Layer*. <http://arxiv.org/abs/1709.07871>

- Pohl, C., & Van Genderen, J. L. (1998). Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. In *International Journal of Remote Sensing* (Vol. 19, Number 5, pp. 823–854). <https://doi.org/10.1080/014311698215748>
- Protter, M., Elad, M., Takeda, H., & Milanfar, P. (2009). Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 18(1), 36–51. <https://doi.org/10.1109/TIP.2008.2008067>
- Ren, A., Zhou, X., Li, G., Li, J., Ma, Y., Qiao, L., Yan, J., & Shen, J. (2025). Integrated UAV and Satellite Remote Sensing: A Survey on Integration Strategies and Data Fusion Methods. *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, 8835–8840. <https://doi.org/10.1109/IGARSS55030.2025.11242522>
- Salgueiro Romero, L., Marcello, J., & Vilaplana, V. (2020). Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks. *Remote Sensing*, 12(15), 2424. <https://doi.org/10.3390/rs12152424>
- Sarmad, M., Kampffmeyer, M., & Salberg, A.-B. (2025). DiffFuSR: Super-Resolution of All Sentinel-2 Multispectral Bands Using Diffusion Models. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–13. <https://doi.org/10.1109/TGRS.2025.3638896>
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). *Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting*. <http://arxiv.org/abs/1506.04214>
- Swain, R., Paul, A., & Behera, M. D. (2024). Spatio-temporal fusion methods for spectral remote sensing: a comprehensive technical review and comparative analysis. *Tropical Ecology*, 65(3), 356–375. <https://doi.org/10.1007/s42965-023-00318-5>
- Takeda, H., Milanfar, P., Protter, M., & Elad, M. (2009). Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing*, 18(9), 1958–1975. <https://doi.org/10.1109/TIP.2009.2023703>
- Tao, X., Gao, H., Liao, R., Wang, J., & Jia, J. (2017). Detail-Revealing Deep Video Super-Resolution. *2017 IEEE International Conference on Computer Vision (ICCV)*, 4482–4490. <https://doi.org/10.1109/ICCV.2017.479>
- Tarasiewicz, T., Nalepa, J., Farrugia, R. A., Valentino, G., Chen, M., Briffa, J. A., & Kawulok, M. (2023). *Multitemporal and multispectral data fusion for super-resolution of Sentinel-2 images*. <https://doi.org/10.1109/TGRS.2023.3311622>
- Toosi, A., Samadzadegan, F., & Javan, F. D. (2026). S³-ESRGAN: Enhanced Super-Resolution Generative Adversarial Network for Remote Sensing Imagery Spatial Resolution Improvement—An Application Using Sentinel-2 and UAV Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 19, 2149–2172. <https://doi.org/10.1109/JSTARS.2025.3640940>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>

- Wald, L. (2002). *Data Fusion. Definitions and Architectures-Fusion of Images of Different Spatial Resolutions* (Number 2). Presses de l'Ecole, Ecole des Mines de Paris. <https://minesparis-psl.hal.science/hal-00464703v1>
- Wang, X., Chan, K. C. K., Yu, K., Dong, C., & Loy, C. C. (2019). EDVR: Video Restoration With Enhanced Deformable Convolutional Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1954–1963. <https://doi.org/10.1109/CVPRW.2019.00247>
- Wang, X., Wang, D., Song, C., Zhang, X., Liu, L., Sun, M., Cai, S., Han, S., Liu, L., & Zhang, Y. (2026). FRCFD-Net: Multi-frame super-resolution through two-branch feature refinement, cross-layer fusion and difference enhancement. *Neurocomputing*, 681. <https://doi.org/10.1016/j.neucom.2026.133412>
- Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1905–1914. <https://doi.org/10.1109/ICCVW54120.2021.00217>
- Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., & Min, H. (2022). A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing. *Remote Sensing*, 14(21). <https://doi.org/10.3390/rs14215423>
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Loy, C. C. (2019). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks* (pp. 63–79). https://doi.org/10.1007/978-3-030-11021-5_5
- Wang, Y., Yang, W., Chen, X., Wang, Y., Guo, L., Chau, L.-P., Liu, Z., Qiao, Y., Kot, A. C., & Wen, B. (2024). SinSR: Diffusion-Based Image Super-Resolution in a Single Step. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25796–25805. <https://doi.org/10.1109/CVPR52733.2024.02437>
- Wang, Z., Tan, H., Wang, Z., Ci, J., & Zhai, H. (2025). MAF-GAN: A Multi-Attention Fusion Generative Adversarial Network for Remote Sensing Image Super-Resolution. *Remote Sensing*, 17(24), 3959. <https://doi.org/10.3390/rs17243959>
- Xiao, J., Aggarwal, A. K., Rage, U. K., Katiyar, V., & Avtar, R. (2023a). Deep Learning-Based Spatiotemporal Fusion of Unmanned Aerial Vehicle and Satellite Reflectance Images for Crop Monitoring. *IEEE Access*, 11, 85600–85614. <https://doi.org/10.1109/ACCESS.2023.3297513>
- Xiao, J., Aggarwal, A. K., Rage, U. K., Katiyar, V., & Avtar, R. (2023b). Deep Learning-Based Spatiotemporal Fusion of Unmanned Aerial Vehicle and Satellite Reflectance Images for Crop Monitoring. *IEEE Access*, 11, 85600–85614. <https://doi.org/10.1109/ACCESS.2023.3297513>
- Xiao, Y., Su, X., Yuan, Q., Liu, D., Shen, H., & Zhang, L. (2022). Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection. *IEEE Transactions on Geoscience and Remote Sensing*, 60. <https://doi.org/10.1109/TGRS.2021.3107352>

- Xie, C., Xing, F., Zhu, Y., & Ye, S. (2026). Multi-scale partial convolution residual distillation network for efficient image super-resolution. *Signal, Image and Video Processing*, 20(6), 360. <https://doi.org/10.1007/s11760-026-05417-z>
- Xiong, Y., Guo, S., Chen, J., Deng, X., Sun, L., Zheng, X., & Xu, W. (2020). Improved SRGAN for remote sensing image super-resolution across locations and sensors. *Remote Sensing*, 12(8). <https://doi.org/10.3390/RS12081263>
- Zhang, H., Sun, Y., Shi, W., Guo, D., & Zheng, N. (2021). An object-based spatiotemporal fusion model for remote sensing images. *European Journal of Remote Sensing*, 54(1), 86–101. <https://doi.org/10.1080/22797254.2021.1879683>
- Zhang, J., Xu, T., Li, J., Jiang, S., & Zhang, Y. (2022). Single-Image Super Resolution of Remote Sensing Images with Real-World Degradation Modeling. *Remote Sensing*, 14(12). <https://doi.org/10.3390/rs14122895>
- Zhang, Y., Chen, Z., & Liu, S. (2020). Video Super Resolution Using Temporal Encoding ConvLSTM and Multi-Stage Fusion. *2020 IEEE International Conference on Visual Communications and Image Processing, VCIP 2020*, 298–301. <https://doi.org/10.1109/VCIP49819.2020.9301823>

5 Estimación espacial de la evapotranspiración del cultivo de maíz mediante imágenes de dron, Sentinel-2 y aprendizaje profundo

Resumen

La estimación espacial de la evapotranspiración de los cultivos constituye un reto fundamental para mejorar la eficiencia en el uso del agua. La teledetección ha facilitado el proceso mediante imágenes satelitales libres Sentinel-2, sin embargo, su resolución espacial limita la representación de la variabilidad del cultivo. En este estudio se propone una metodología para determinar la distribución espacial de la evapotranspiración del maíz mediante la integración de imágenes UAV, Sentinel-2, el modelo de superresolución TVDSR y redes neuronales convolucionales. Se aplicó la metodología FAO-Penman-Monteith para determinar la distribución espacial del coeficiente dual del cultivo (K_{cdual}) a partir de mapas de altura y cobertura del suelo. Se integraron imágenes Sentinel-2 y UAV mediante el modelo TVDSR. Se entrenó una red U-Net VGG para determinar la cobertura de suelo y tres redes U-Net VGG para predecir K_{cdual} a partir de imágenes UAV, Sentinel-2 y Sentinel-2 superresueltas. Finalmente, se determinó la evapotranspiración del maíz a partir de los mapas K_{cdual} y la evapotranspiración de referencia. La clasificación de cobertura alcanzó valores de precisión, recall y F1-score superiores a 0.99. Las imágenes UAV presentaron el mejor desempeño en la predicción de K_{cdual} ($R^2 = 0.883$; RMSE = 0.093; MAE = 0.043), seguidas por las imágenes superresueltas ($R^2 = 0.827$) y Sentinel-2 ($R^2 = 0.774$). Para la evapotranspiración estimada con los mapas K_{cdual} predichos, las imágenes UAV mostraron la mejor concordancia ($R^2 = 0.875$; RMSE = 0.40 mm d⁻¹; MAE = 0.18 mm d⁻¹). La superresolución redujo el RMSE en 16 % y el MAE en 17 % respecto a Sentinel-2 original. Los resultados demuestran que la combinación de superresolución espacial, teledetección multiespectral y aprendizaje profundo constituye una alternativa eficaz para mejorar la estimación de coeficientes de cultivo y evapotranspiración del maíz a escala parcelaria.

Palabras Clave: U-Net VGG, Segmentación semántica, TVDSR e Índice de Vegetación.

Tesis de Doctorado en Ingeniería Agrícola y Uso Integral del Agua, Programa de Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo.

Autor: M.C. Alvaro Murguía Cozar.

Director de tesis: Dr. Gilberto de Jesús López Canteñs.

Spatial Estimation of Maize Crop Evapotranspiration Using UAV Imagery, Sentinel-2, and Deep Learning

Abstract

The spatial estimation of crop evapotranspiration constitutes a fundamental challenge for improving water use efficiency. Remote sensing has facilitated this process through freely available Sentinel-2 satellite imagery; however, its spatial resolution limits the representation of crop variability. This study proposes a methodology to determine the spatial distribution of maize evapotranspiration through the integration of UAV imagery, Sentinel-2 imagery, the TVDSR super-resolution model, and convolutional neural networks. The FAO-Penman-Monteith methodology was applied to determine the spatial distribution of the dual crop coefficient (K_{cdual}) from crop height and ground cover maps. Sentinel-2 and UAV images were integrated using the TVDSR model. A U-Net VGG network was trained to determine ground cover, and three U-Net VGG networks were trained to predict K_{cdual} from UAV, Sentinel-2, and super-resolved Sentinel-2 imagery. Finally, maize evapotranspiration was determined from the K_{cdual} maps and reference evapotranspiration. Ground-cover classification achieved precision, recall, and F1-score values above 0.99. UAV imagery showed the best performance in K_{cdual} prediction ($R^2 = 0.883$; RMSE = 0.093; MAE = 0.043), followed by super-resolved imagery ($R^2 = 0.827$) and Sentinel-2 imagery ($R^2 = 0.774$). For evapotranspiration estimated from the predicted K_{cdual} maps, UAV imagery showed the best agreement ($R^2 = 0.875$; RMSE = 0.40 mm d⁻¹; MAE = 0.18 mm d⁻¹). Super-resolution reduced RMSE by 16% and MAE by 17% compared with the original Sentinel-2 imagery. The results demonstrate that the combination of spatial super-resolution, multispectral remote sensing, and deep learning constitutes an effective alternative for improving the estimation of crop coefficients and maize evapotranspiration at the field scale.

Keywords: U-Net VGG; Semantic segmentation; TVDSR; Vegetation index.

Doctoral Thesis. Graduate Program in Agricultural Engineering and Integrated Water Resources Management, Universidad Autónoma Chapingo.

Author: Msc. Alvaro Murguía Cozar.

Thesis director: Phd. Gilberto de Jesús López Canteñs.

5.1 Introducción

La evapotranspiración constituye uno de los procesos fundamentales del ciclo hidrológico y representa el principal componente del consumo de agua en los sistemas agrícolas. Su cuantificación precisa es esencial para la programación del riego, la estimación de requerimientos hídricos y la gestión sostenible de los recursos hídricos, especialmente en regiones donde la disponibilidad de agua se encuentra limitada por el crecimiento de la demanda agrícola y los efectos del cambio climático (Feng et al., 2023). En cultivos de alta importancia agronómica como el maíz (*Zea mays L.*), una estimación inadecuada de la evapotranspiración puede provocar pérdidas de rendimiento, disminución de la eficiencia en el uso del agua y sobreexplotación de fuentes hídricas superficiales y subterráneas (Feng et al., 2023).

La estimación de la evapotranspiración se ha realizado mediante métodos físicos basados en el balance de energía superficial, los cuales integran información sobre radiación neta, flujo de calor sensible, flujo de calor latente y temperatura superficial. Entre los modelos más utilizados destacan SEBAL, METRIC, SEBS, TSEB y DisALEXI, los cuales emplean principalmente información térmica proveniente de sensores satelitales o plataformas aéreas para determinar la transferencia de energía entre la superficie y la atmósfera (Ankela et al., 2026) . Diversos estudios han demostrado que estos enfoques permiten obtener estimaciones precisas de evapotranspiración real a escala de parcela y regional, especialmente cuando se dispone de observaciones de temperatura superficial de alta calidad (Ankela et al., 2026; N. Liu et al., 2026; Sabharwal et al., 2026; Tunca, 2024). Sin embargo, estos modelos suelen requerir información térmica, múltiples variables biofísicas y procesos de calibración relativamente complejos, lo que puede limitar su aplicación operacional en grandes superficies agrícolas (Tunca, 2024) .

Otro enfoque ampliamente adoptado para la estimación de la evapotranspiración consiste en la utilización de la metodología FAO-Penman-Monteith (Pereira et al., 2025), donde la evapotranspiración del cultivo (ET_c) se obtiene mediante la combinación de la evapotranspiración de referencia (ET_o) y un coeficiente de cultivo (K_c). En este contexto, la teledetección ha permitido sustituir gradualmente los coeficientes de cultivo tabulados por estimaciones dinámicas derivadas de índices de vegetación obtenidos mediante sensores remotos. Diversos trabajos han demostrado relaciones robustas entre el NDVI y variables biofísicas como la cobertura vegetal, el índice de área foliar (LAI) y los coeficientes basales de cultivo, lo que permite estimar espacial y temporalmente los requerimientos hídricos de diferentes cultivos (Pereira et al., 2020; Pôças et al., 2020). Asimismo, estudios realizados con Sentinel-2 han mostrado que la utilización de series temporales de NDVI permite identificar etapas fenológicas y estimar coeficientes de cultivo con elevada precisión, mejorando la representación de la variabilidad espacial y temporal de la evapotranspiración respecto a los enfoques convencionales basados en calendarios fijos (French et al., 2023; Hassan et al., 2022; Nagy et al., 2024).

La disponibilidad de imágenes de alta resolución espacial provenientes de vehículos aéreos no tripulados (UAV) ha impulsado aún más este tipo de metodologías (Shao et al., 2021). Los drones permiten obtener información espectral con resolución centimétrica, facilitando la estimación detallada de variables estructurales del cultivo, tales como altura de planta, cobertura vegetal, biomasa y coeficientes de cultivo. Investigaciones han demostrado que índices espectrales derivados de imágenes multiespectrales UAV pueden estimar coeficientes de cultivo con coeficientes de determinación superiores a 0.90, así como generar mapas detallados de evapotranspiración a escala de parcela para apoyar estrategias de riego de precisión (Marcial-Pablo et al., 2021; G. Shao et al., 2021; Tang et al., 2019).

Investigaciones recientes han evolucionado hacia enfoques que integran múltiples fuentes de información y técnicas avanzadas de inteligencia artificial. El proceso de integrar varias fuentes de información permite explotar las bondades de cada fuente. La integración de diferentes fuentes permite combinar información complementaria obtenida a distintas escalas espaciales, temporales y espectrales, lo que proporciona una representación más completa de los fenómenos agrícolas. Además, la complementariedad entre fuentes contribuye a reducir incertidumbres, mejorar la robustez de los modelos y aumentar la precisión de las estimaciones obtenidas (Phang et al., 2023). El aprendizaje automático y el aprendizaje profundo han demostrado una elevada capacidad para representar las complejas relaciones no lineales existentes entre variables meteorológicas, espectrales, edáficas y fisiológicas asociadas al consumo de agua de los cultivos (Xing et al., 2026). Revisiones recientes identifican a algoritmos como Random Forest, XGBoost y redes neuronales profundas como algunas de las herramientas más prometedoras para mejorar la estimación de la evapotranspiración mediante la integración de datos de teledetección y observaciones de campo (Amani & Shafizadeh-Moghadam, 2023). En particular, modelos basados en redes convolucionales temporales y arquitecturas multimodales han alcanzado coeficientes de determinación superiores a 0.84 para la estimación de la evapotranspiración del maíz a partir de variables meteorológicas, condiciones del suelo y características del cultivo (Z. Chen et al., 2020; Zhangzhong et al., 2026; Zhao et al., 2024).

La fusión de datos provenientes de plataformas satelitales, sensores UAV, información meteorológica y variables biofísicas mejora significativamente la capacidad para generar estimaciones continuas de evapotranspiración con alta resolución espacial y temporal (Fehér et al., 2026). Las imágenes satelitales como Sentinel-2 presentan limitaciones para representar la variabilidad espacial dentro de la parcela debido a su resolución espacial nativa, mientras que las imágenes UAV aportan un nivel de detalle mayor. Por lo que la fusión de estas

dos fuentes permite mejorar el detalle espacial de las imágenes Sentinel-2 y fortalecer los modelos de aprendizaje profundo utilizados para la estimación de variables biofísicas (Murguía-Cozar et al., 2026). La integración de imágenes Sentinel-2, Landsat, MODIS, sensores térmicos, modelos de balance de energía y algoritmos de aprendizaje automático ha permitido desarrollar productos operativos para el monitoreo del consumo hídrico agrícola tanto a escala parcelaria como regional (Fehér et al., 2026; N. Liu et al., 2026; Zhu et al., 2025). Dichas investigaciones evidencian una tendencia clara hacia esquemas multifuente que aprovechan simultáneamente la alta resolución espacial de los UAV, la cobertura temporal de los satélites y la capacidad predictiva de los modelos de inteligencia artificial.

Bajo este contexto, la presente investigación propone estimar la distribución espacial de la ET_c del cultivo de maíz a partir de imágenes Sentinel-2 y UAV, mediante redes neuronales convolucionales capaces de predecir el coeficiente dual de cultivo obtenido con la metodología FAO-Penman-Monteith. Este enfoque busca aprovechar las ventajas complementarias de cada fuente de información: la cobertura temporal, extensión y disponibilidad operativa de Sentinel-2, el nivel de detalle espacial proporcionado por los UAV y el potencial de las técnicas de superresolución para integrar el detalle espacial y la extensión en un producto mejorado. De esta manera, se pretende generar estimaciones espacialmente explícitas de los coeficientes de cultivo y, a partir de éstos y de la evapotranspiración de referencia obtenida mediante información meteorológica, estimar la evapotranspiración del maíz con un mayor nivel de detalle espacial que el proporcionado por las imágenes Sentinel-2 originales, para aplicaciones de agricultura de precisión y gestión eficiente del agua.

5.2 Metodología

5.2.1 Área de estudio

La investigación se realizó en el campo experimental ‘San Juan’ de la Universidad Autónoma Chapingo, ubicada en 19°20’34.06” latitud Norte, 98°51’18.45” longitud Oeste y una elevación de 2211 metros sobre el nivel medio del mar. Se realizó el seguimiento de 24 parcelas cultivadas con maíz forrajero, las cuales suman una superficie total de 19.21 hectáreas sembradas en el periodo comprendido de mayo-junio de 2025. Los suelos predominantes en la zona de estudio presentan texturas: franco, franco-arenosa y franco-arcillo-arenosa; además, contenidos de materia orgánica cercanos al 1 %, características que influyen en la capacidad de retención de agua y en la dinámica de humedad del suelo. En la Figura 5.1 se presenta la localización de la zona de trabajo y la distribución de parcelas y puntos de monitoreo de humedad de suelo.

5.2.2 Imágenes multiespectrales UAV

Se utilizó una aeronave DJI Mavic 3 equipada con cámara multiespectral y sistema de posicionamiento RTK para la georreferenciación precisa de la aeronave y de las imágenes adquiridas durante los vuelos. Se realizaron 13 vuelos fotogramétricos sobre el área de estudio, los días 14 y 29 de julio; 3, 8, 23 y 28 de agosto; y 12 y 22 de octubre de 2024; así como el 9 de julio, 3, 8 y 13 de agosto, y 2 de septiembre de 2025, coincidiendo con los tiempos de revisita del Satélite Sentinel-2. El conjunto completo de imágenes adquiridas durante ambos años fue empleado para el entrenamiento, ajuste y validación del modelo de superresolución espacial. Adicionalmente, las imágenes correspondientes al ciclo agrícola 2025 fueron utilizadas para el monitoreo del cultivo y los análisis posteriores relacionados con la estimación de la evapotranspiración. En todas las campañas se mantuvo una altura de vuelo de 120 m sobre el nivel del terreno, con un traslape frontal y lateral del 70 %, y un horario de adquisición comprendido entre las 10:00 y 10:33 h, con el fin de garantizar condiciones de iluminación

relativamente homogéneas y una adecuada reconstrucción fotogramétrica de la superficie del cultivo con GSD de 0.06m.

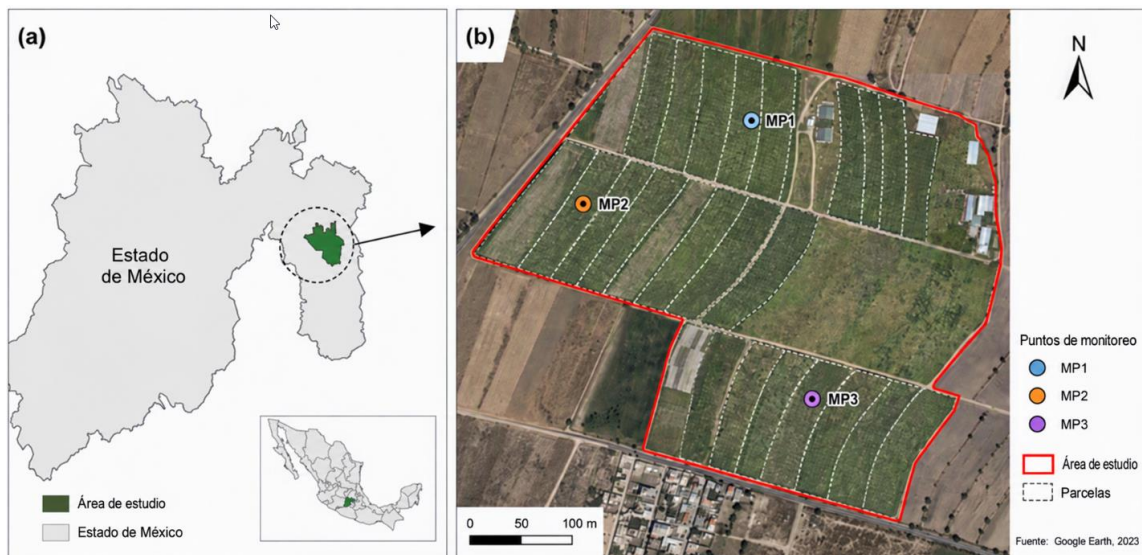


Figura 5.1. Ubicación del área de estudio y puntos de monitoreo. (a) Localización en el Estado de México. (b) Distribución de parcelas y puntos de monitoreo.

Se empleó el programa Pix4Dmapper en el procesamiento fotogramétrico de los vuelos y se generaron el ortomosaico RGB y los mapas de reflectancia, Infrarrojo Cercano (860 nm), Borde Rojo (740 nm), Rojo (650 nm) y Verde (560 nm), además se generaron los modelos digitales de elevación de superficie y de terreno.

5.2.3 Imágenes Sentinel-2

Se descargaron imágenes de la misión Sentinel-2 correspondientes al área de estudio desde la plataforma Copernicus Data Space Ecosystem de la Agencia Espacial Europea (<https://browser.dataspace.copernicus.eu/>), seleccionando aquellas con una cobertura nubosa inferior al 50 % para las 13 fechas de análisis. Se utilizaron productos con nivel de procesamiento L2A, los cuales incluyen corrección atmosférica y reflectancias de superficie. Para el presente estudio se emplearon las bandas verde (B03, 560 nm), roja (B04, 665 nm), borde rojo (B06, 740 nm) e infrarrojo cercano de banda angosta (B08A, 865 nm), debido a su

utilidad en la caracterización del estado fisiológico, vigor vegetativo y desarrollo fenológico del cultivo de maíz.

5.2.4 Generación de Índices de Vegetación

Los datos satelitales y UAV fueron normalizados mediante el método min-max (1), con el fin de homogeneizar los rangos de valores entre sensores y fechas de adquisición. Con la finalidad de facilitar la fusión espacial y temporal de la información, las imágenes fueron ajustadas a una resolución espacial intermedia de 2m, propuesta por Murguía-Cozar et al. (2026) como escala óptima para la superresolución de imágenes Sentinel-2 y UAV. Para las composiciones derivadas de UAV, el remuestreo se realizó mediante la agregación por promedio de los píxeles contenidos dentro de cada celda de 2 m, mientras que las composiciones Sentinel-2 fueron remuestreadas utilizando interpolación bicúbica. Finalmente, todas las capas generadas fueron recortadas al área de estudio utilizando la extensión espacial mostrada en la Figura 5.1.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

donde x es el valor de reflectancia original, x_{min} es el valor mínimo de la banda analizada, x_{max} es el valor máximo de la banda analizada y x_{norm} es el valor normalizado.

A partir de las imágenes multiespectrales adquiridas mediante UAV y Sentinel-2 normalizadas se generaron composiciones espectrales correspondientes a cinco índices de vegetación: NDVI (Normalized Difference Vegetation Index), SAVI (Soil Adjusted Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index) y NDVI_{re} (Red-Edge Normalized Difference Vegetation Index).

5.2.5 Fusión de imágenes mediante el modelo TVDSR

Para incrementar la resolución espacial de las imágenes Sentinel-2 se empleó el modelo Temporal Very Deep Super-Resolution (TVDSR), una arquitectura de aprendizaje profundo diseñada para realizar superresolución espaciotemporal en entornos agrícolas. El modelo aprende una función de mapeo entre secuencias temporales de imágenes Sentinel-2 y observaciones de alta resolución adquiridas mediante UAV. El modelo utiliza como variables de entrada composiciones multitemporales de índices de vegetación Sentinel-2 y como referencia las composiciones derivadas de las imágenes multiespectrales obtenidas con dron.

La arquitectura TVDSR combina un encoder convolucional profundo para la extracción de características espaciales, una capa Bidireccional ConvLSTM para modelar las dependencias temporales entre fechas de observación; mecanismos de atención temporal y espectral que identifican automáticamente las observaciones y características más relevantes para la reconstrucción. Adicionalmente, incorpora información fenológica mediante embebidos temporales basados en el día juliano y un mecanismo de modulación FiLM, lo que permite representar la dinámica estacional de los cultivos. Finalmente, un decoder convolucional residual reconstruye la imagen de salida de alta resolución a partir de la representación espaciotemporal aprendida, para generar productos con mayor detalle espacial que las imágenes Sentinel-2 originales (Figura 5.2).

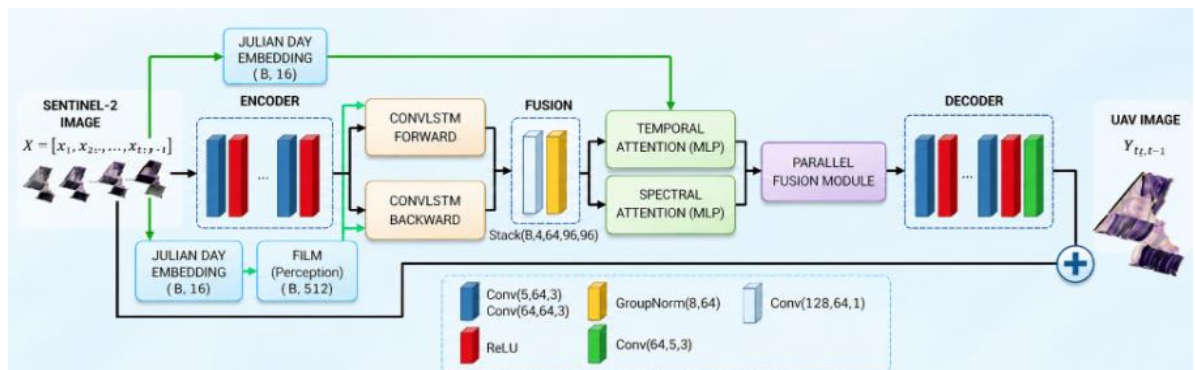


Figura 5.2. Modelo espaciotemporal TVDSR.

5.2.6 Datos meteorológicos

Se instaló una estación meteorológica en el área experimental basada en un microcontrolador ESP32 para el monitoreo de temperatura del aire, humedad relativa, radiación solar global, velocidad y dirección del viento, así como precipitación. Las seis variables meteorológicas se registraron a intervalos de 15 minutos. De acuerdo con Servín-Palestina et al. (2025), el sistema proporciona mediciones con un nivel de precisión y confiabilidad suficiente para su uso en la estimación de la evapotranspiración, y constituye una alternativa accesible para apoyar la toma de decisiones en el manejo agrícola.

5.2.7 Humedad de suelo

Se instalaron tres puntos de monitoreo de humedad residual del suelo (Figura 5.1), cada punto estuvo conformado por dos sensores de humedad de suelo de matriz granular de la marca Watermark 200SS5, un sensor se colocó a una profundidad de 0.15m y el otro a 0.30m. En la Figura 5.3a se presenta el diagrama de instalación. Los sensores se instalaron en tubos de PVC de 1/2" y conectados a un módulo de guardado construido a base de un ESP32 y una terminal microSD para el guardado de la información. Se programó el dispositivo para realizar toma de lecturas a intervalos de una hora.

Se realizó la calibración de los sensores mediante la aplicación del método gravimétrico a muestras inalteradas de suelo. Se extrajeron tres muestras inalteradas, con un muestreador de núcleos y cilindros de volumen conocido, a 0.15m y 0.30 m de profundidad e intervalos de dos días, después de un fenómeno de lluvia o la aplicación de un riego, por cada punto de monitoreo (FAO, 2023).

5.2.8 Parámetros de suelo

Para la caracterización de las condiciones físicas e hidráulicas de los suelos en la zona de estudio, se obtuvieron ocho muestras alteradas distribuidas en las

áreas representativas del experimento (tres muestras coincidieron con los tres puntos de monitoreo de humedad).

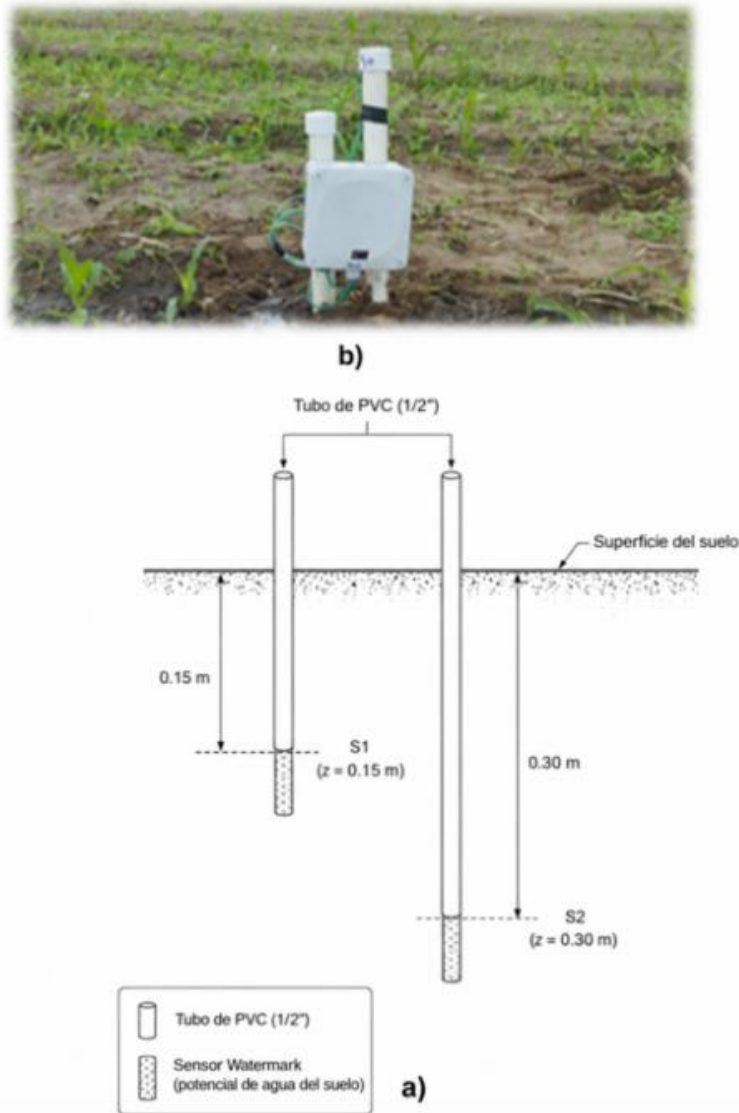


Figura 5.3. Puntos de monitoreo de humedad del suelo (a) diagrama de instalación; (b) Punto de monitoreo.

La textura del suelo se determinó mediante el método de Bouyoucos modificado (Beretta et al., 2014), mientras que la densidad aparente (DAP) se estimó a partir de la relación masa-volumen del suelo seco. Adicionalmente, se determinaron los contenidos de humedad correspondientes a capacidad de campo (CC), punto de

marchitez permanente (PMP), saturación (SAT) y las curvas de retención de humedad del suelo conforme a la norma oficial mexicana vigente (SEMARNAT, 2002).

Los parámetros obtenidos permitieron caracterizar la capacidad de almacenamiento de agua del perfil edáfico y estimar el agua disponible para el cultivo. Esta información sirvió como base para la ejecución del balance hídrico del suelo, la interpretación de la dinámica de humedad observada mediante los sensores Watermark y la estimación de los coeficientes de estrés hídrico conforme a la metodología propuesta por Pereira et al. (2025).

5.2.9 Estimación de evapotranspiración

El enfoque propuesto se basa en la metodología FAO para la estimación de la ET_c del maíz y la amplía para su determinación espacial mediante la rasterización del coeficiente dual de cultivo. La metodología comprende tres etapas principales: (i) la determinación del coeficiente basal del cultivo (K_{cb}) ajustado a las condiciones locales, (ii) la estimación del coeficiente de estrés hídrico (K_s) y del coeficiente de evaporación del suelo (K_e), y (iii) la construcción espacial del coeficiente dual de cultivo.

5.2.9.1 Determinación del K_{cb} espacial

Para determinar el K_{cb} espacial se tomaron como referencia los valores de K_{cb}, propuestos por Pereira et al. (2025), del cultivo de maíz forrajero 0.15, 1.10 y 0.80 para las etapas inicial, media y final, respectivamente. La delimitación de las etapas fenológicas se realizó mediante grados día de desarrollo (GDD), se emplearon los umbrales de 250 °C·día para el término de la etapa inicial, 750 °C·día para el inicio de la etapa media, 1350 °C·día para el final de la etapa media y 1550 °C·día para la madurez fisiológica del cultivo. Se ajustaron los valores de K_{cb} correspondientes a las etapas media y final mediante la ecuación (3).

$$K_{cbadjusted} = K_{cb} + [0.04(u_2 - 2) - 0.004(HR_{min} - 45)] \left(\frac{h}{3}\right)^{0.3} \quad (3)$$

Con el propósito de incorporar la variabilidad espacial del desarrollo del cultivo en la estimación de K_{cb} , se emplearon modelos digitales de superficie (DSM) generados a partir de imágenes UAV. La altura del cultivo se obtuvo como la diferencia entre el DSM correspondiente al terreno desnudo y los DSM generados durante el ciclo de crecimiento del maíz. Como superficie de referencia, se realizó un vuelo cuando las parcelas se encontraban preparadas para la siembra y libres de vegetación.

Para cada fecha de análisis se utilizó el mapa de altura del cultivo para ajustar espacialmente los valores de K_{cb} de las etapas media y final mediante la ecuación (3). A diferencia del procedimiento convencional, que emplea una única altura representativa del cultivo, en este estudio el ajuste se realizó a nivel de píxel con la altura estimada a partir de los DSM. De esta manera, para cada píxel se calcularon los valores ajustados de $K_{cb, mid}$ y $K_{cb, end}$, con los cuales se reconstruyó una curva específica GDD- K_{cb} .

Los grados día de desarrollo acumulados se calcularon mediante la ecuación (4) a partir de la fecha de siembra registrada para cada parcela.

$$GDD = \sum_{i=1}^n \frac{T_{max,i} + T_{min,i}}{2} - T_b \quad (4)$$

donde GDD representa los grados día desarrollo; $T_{max,i}$ corresponde a la temperatura máxima registrada en el día i después de la siembra; $T_{min,i}$ representa la temperatura mínima registrada en el mismo día; y $T_b = 10^\circ C$ es la temperatura base para el cultivo de maíz.

Finalmente, para cada píxel se determinó el valor acumulado de GDD, se identificó su posición dentro de la curva K_{cb} -GDD previamente ajustada y se

obtuvo mediante interpolación el valor correspondiente de K_{cb} . Este procedimiento permitió generar mapas espacialmente distribuidos de K_{cb} , lo que permitió incorporar simultáneamente la variabilidad temporal asociada al desarrollo fenológico y la variabilidad espacial derivada de las diferencias de altura observadas dentro de las parcelas experimentales.

5.2.9.2 Estimación del K_s y el K_e

Para determinar los parámetros K_s y K_e se ejecutó un balance hídrico del suelo en tres ubicaciones representativas del área de estudio. Debido a que cada punto se localiza dentro de grupos de parcelas con fechas de siembra similares, se asumió que las condiciones de desarrollo fenológico del cultivo eran homogéneas dentro de cada zona. Con este propósito se delimitó un polígono de 3 m × 3 m alrededor de cada punto de monitoreo y se aplicó en análisis del 1 de agosto al 2 de septiembre de 2025. El agotamiento de agua en la capa superficial del suelo y el agotamiento de humedad en la zona radicular al final de cada día se calcularon mediante las ecuaciones (5) y (6), respectivamente.

$$D_{e,i} = D_{e,i-1} - (P_i - RO_i) - \frac{I_i}{f_w} + \frac{E_i}{f_{ew}} + T_{ew,i} + DP_{e,i} \quad (5)$$

donde $D_{e,i}$ representa la lámina acumulada de evaporación (agotamiento) al final del día i [mm], $D_{e,i-1}$ corresponde a la lámina acumulada de evaporación al final del día anterior [mm], P_i es la precipitación registrada en el día i [mm], RO_i representa el escurrimiento superficial i [mm], I_i es la lámina de riego infiltrada en el suelo [mm], E_i corresponde a la evaporación del suelo ($E_i = K_e ET_o$) [mm], $T_{ew,i}$ representa la transpiración asociada a la fracción expuesta y humedecida de la superficie del suelo [mm], $DP_{e,i}$ corresponde a las pérdidas por percolación profunda en la capa superficial [mm], f_w es la fracción de suelo humedecida por riego, f_{ew} representa la fracción de suelo simultáneamente expuesta y humedecida.

$$D_{r,i} = D_{r,i-1} - (P_i - RO_i) - I_i - CR_i + ET_{c,i} + DP_i \quad (6)$$

Donde $D_{r,i}$ representa el agotamiento de humedad en la zona radicular al final del día i [mm], $D_{r,i-1}$ corresponde al agotamiento registrado al final del día anterior [mm], CR_i es el ascenso capilar proveniente de la capa freática [mm], $ET_{c,i}$ representa la evapotranspiración del cultivo [mm], DP_i corresponde a las pérdidas por percolación profunda desde la zona radicular [mm].

A partir del balance hídrico se obtuvieron los parámetros K_s , K_r y K_{cmax} correspondientes a las fechas de vuelo del 3, 8 y 13 de agosto, así como del 2 de septiembre de 2025. Los valores de K_r y K_{cmax} se utilizaron posteriormente en la ecuación (7) para calcular espacialmente el parámetro K_e .

$$K_e = \min(K_r(K_{cmax} - K_{cb}), f_{ew}K_{cmax}) \quad (7)$$

donde K_e representa el coeficiente de evaporación del suelo; K_r corresponde al coeficiente de reducción de evaporación; K_{cmax} es el valor máximo que puede alcanzar el coeficiente de cultivo; K_{cb} representa el coeficiente basal del cultivo; f_{ew} corresponde a la fracción de suelo expuesta y humedecida durante un evento de riego o precipitación.

La estimación de la fracción de suelo expuesta se realizó mediante un modelo supervisado de segmentación semántica en PyTorch. Se empleó una red neuronal tipo U-Net complementada con bloques de extracción de características inspirados en VGG (Figura 5.4), para segmentar las clases maíz, suelo y maleza. El encoder está conformado por cinco niveles de extracción de características y operaciones de MaxPooling, las cuales reducen progresivamente la resolución espacial hasta alcanzar un tamaño de 3×3 píxeles en la capa de cuello de botella. Durante este proceso, la profundidad de los tensores aumenta de 64 a 512 canales, lo que permite una representación más detallada de los patrones espaciales y espectrales presentes en las imágenes.

La reconstrucción de la segmentación se realizó mediante cinco bloques decoder que ejecutan la operación inversa al encoder. En cada nivel se emplearon capas

de convolución transpuesta (UpConvolution), las cuales duplican las dimensiones espaciales de la representación interna y recuperan progresivamente el detalle de la imagen original. La arquitectura incorpora conexiones de salto (skip connections) entre encoder y decoder para preservar la información espacial de alta resolución y mejorar la delimitación de las clases.

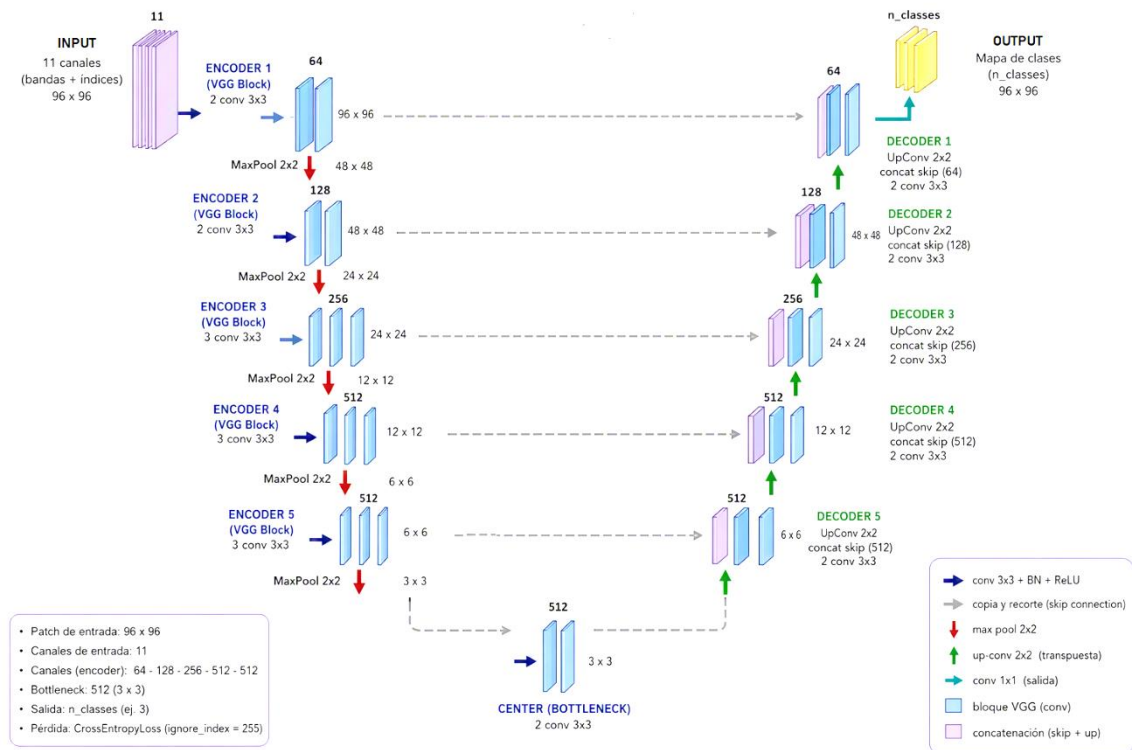


Figura 5.4. Arquitectura U-Net-VGG utilizada para la segmentación semántica del cultivo de maíz.

El entrenamiento del modelo utilizó composiciones multiespectrales integradas por once capas: bandas roja, verde, infrarrojo cercano y borde rojo, así como los índices de vegetación NDVI, SAVI, GNDVI, NDRE, NDWI, NDVI_{re} y Cl_{re}. A partir de estas composiciones y de las máscaras de referencia se extrajeron parches de 96 × 96 píxeles. Para cada clase se seleccionaron aleatoriamente hasta 1500 muestras, obteniéndose un conjunto de entrenamiento balanceado compuesto

por un máximo de 4500 parches. El conjunto de datos se dividió aleatoriamente en 80 % para entrenamiento y 20 % para validación, empleando una semilla aleatoria de 42 para garantizar la reproducibilidad del experimento. Se utilizó una tasa de aprendizaje de 0.001, un tamaño de lote de 12 muestras y un total de 100 épocas de entrenamiento. La optimización de los pesos de la red se realizó mediante el algoritmo Adam, mientras que la función de pérdida empleada fue la entropía cruzada categórica, la cual minimiza las discrepancias entre las clases observadas y las probabilidades estimadas por el modelo para cada píxel. La capa de salida utilizó una función de activación Softmax para obtener las probabilidades de pertenencia a las clases maíz, suelo y maleza. Finalmente, para evaluar el desempeño del modelo se determinó con las muestras de validación la precisión general, Recall y F1-score.

Una vez obtenidos los mapas de segmentación, la fracción de suelo expuesto f_{ew} se calculó como la proporción de píxeles clasificados como suelo dentro de celdas ráster de 2 m × 2 m. Posteriormente, se estimó de forma espacial el K_e mediante la ecuación (7) a partir de los valores de K_r obtenidos del balance hídrico, los valores de K_{cmax} correspondientes a cada fecha de análisis y los K_{cb} previamente ajustados a las condiciones climáticas locales. Este procedimiento permitió generar mapas espacialmente distribuidos del componente evaporativo de la evapotranspiración e incorporar la variabilidad espacial de la cobertura vegetal y de las condiciones hídricas presentes en el área experimental.

5.2.9.3 Construcción del coeficiente dual

El coeficiente dual de cultivo (K_{cdual}) se calculó mediante la expresión $K_c = K_s \cdot K_{cb} + K_e$. Los valores del coeficiente K_s se obtuvieron de los balances de agua determinados en los tres puntos de monitoreo de humedad del suelo. Posteriormente, cada parcela se asoció con el punto de monitoreo correspondiente según su ubicación y fecha de siembra, por lo que los valores de K_s derivados de cada balance hídrico se asignaron a las parcelas

representadas por dicho punto de control. Finalmente, los valores de K_{cdual} se multiplicaron por la evapotranspiración de referencia para determinar la ET_c del maíz.

5.2.10 Estimación del coeficiente dual de cultivo mediante aprendizaje profundo

La estimación espacial de K_{cdual} se realizó mediante una red neuronal convolucional tipo U-Net con encoder inspirado en la arquitectura VGG. La Figura 5.5 muestra la estructura del modelo empleado. La red recibe como entrada composiciones de cinco bandas correspondientes a los cinco índices de vegetación calculados en las secciones 5.2.4 y 5.2.5; y genera como salida mapas continuos de K_{cdual} con la misma resolución espacial de la imagen de entrada. El encoder está compuesto por cuatro bloques de extracción de características formados por dos capas convolucionales de tamaño 3×3 , seguidas de normalización por lotes (*Batch Normalization*) y funciones de activación ReLU. Entre cada bloque se aplicó una operación MaxPooling 2×2 que redujo progresivamente la resolución espacial de las características extraídas. El cuello de botella (*bottleneck*) se construyó con un bloque convolucional de mayor profundidad que permitió representar patrones espaciales asociados a la variabilidad del coeficiente dual de cultivo.

La reconstrucción espacial se realizó mediante cuatro bloques decoder. Cada bloque utilizó capas de convolución transpuesta (*Conv2DTranspose*) para incrementar la resolución espacial de las características extraídas y recuperar progresivamente el detalle de la imagen original. La arquitectura incorporó conexiones de salto (*skip connections*) entre los niveles equivalentes del encoder y decoder con la finalidad de preservar la información espacial de alta resolución y mejorar la representación de los patrones espaciales del cultivo. Finalmente, la capa de salida utilizó una convolución de tamaño 1×1 seguida de una función

de activación sigmoide. Posteriormente, una capa de escalamiento ajustó los valores predichos al rango observado en los mapas de referencia de Kcdual.

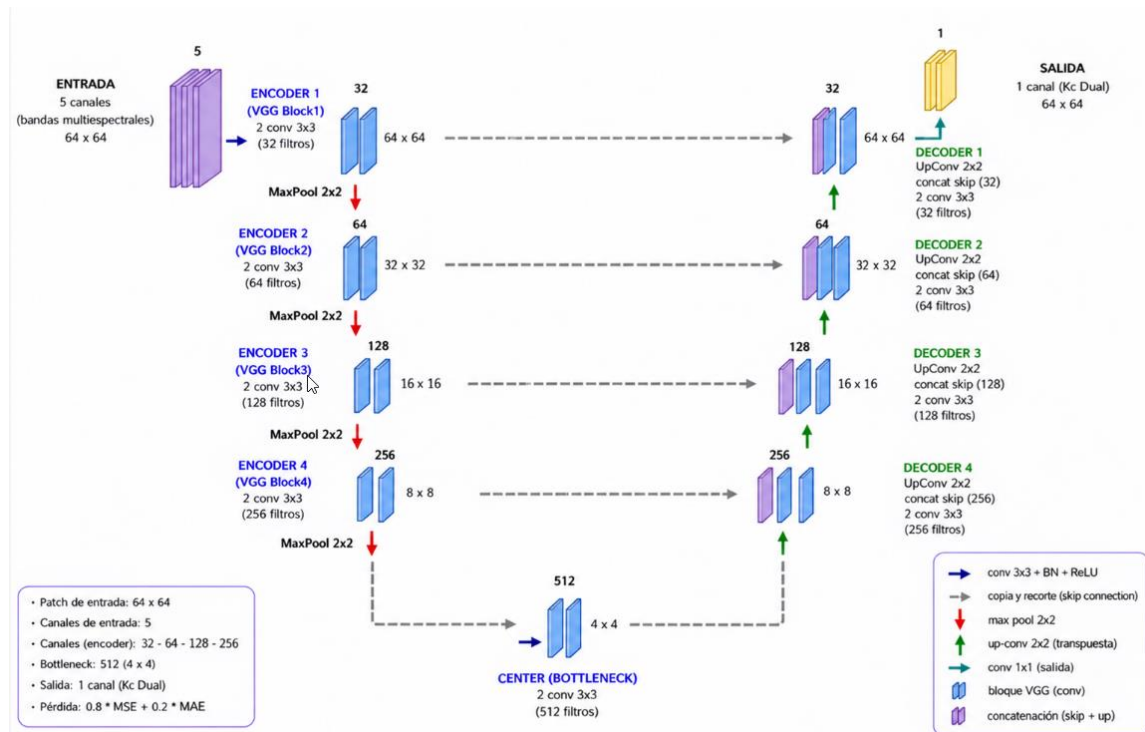


Figura 5.5. Arquitectura U-Net-VGG utilizada para la predicción del coeficiente del cultivo dual.

El entrenamiento del modelo empleó composiciones multispectrales derivadas de imágenes UAV, imágenes Sentinel-2 e imágenes superresueltas generadas mediante el modelo TVDSR. Para cada fuente de datos fue entrenado un modelo U-Net-VGG con la finalidad de evaluar el desempeño del modelo con cada tipo de datos. Los mapas de Kcdual obtenidos mediante la metodología FAO-56 dual fueron utilizados como variable objetivo para el entrenamiento del modelo. Con el propósito de incrementar el número de muestras disponibles, se extrajeron parches de 64×64 píxeles mediante una ventana móvil con un desplazamiento de 48 píxeles, lo que permitió generar muestras parcialmente solapadas. Únicamente se consideraron los píxeles localizados dentro de las parcelas experimentales delimitadas mediante un archivo vectorial, mientras que los

píxeles externos fueron excluidos del entrenamiento mediante máscaras espaciales.

El conjunto de datos se dividió aleatoriamente en 70 % para entrenamiento y 30 % para validación, se empleó una semilla aleatoria de 42 para garantizar la reproducibilidad de los resultados. Adicionalmente, se aplicaron técnicas de aumento de datos (*data augmentation*) mediante reflexión horizontal, reflexión vertical y rotación de 90°, con el propósito de incrementar la diversidad espacial de las muestras y mejorar la capacidad de generalización del modelo. El entrenamiento se realizó durante 500 épocas, se empleó un tamaño de lote (batch size) de 8 muestras y una tasa de aprendizaje de 0.0001.

La optimización de los pesos de la red se realizó mediante el algoritmo Adam. Como función de pérdida se empleó una combinación ponderada del error cuadrático medio (MSE) y del error absoluto medio (MAE), con ponderaciones de 0.8 y 0.2, respectivamente. Esta estrategia permitió penalizar tanto errores de gran magnitud como desviaciones absolutas en la estimación de K_{dual} , favoreciendo una representación más robusta de la variabilidad espacial observada en los mapas de referencia.

Finalmente, el desempeño del modelo se evaluó mediante las métricas coeficiente de determinación (R^2), MSE, raíz del error cuadrático medio (RMSE) y error absoluto medio (MAE), calculadas a partir de las predicciones obtenidas de las imágenes completas. Adicionalmente, se compararon los valores promedio, la desviación estándar y el sesgo medio entre los mapas observados y estimados para verificar la capacidad del modelo para reproducir la distribución espacial del coeficiente dual de cultivo dentro de las parcelas de estudio.

5.3 Resultados y discusión

5.3.1 Modelos de calibración de los sensores Watermark

Se realizaron siete muestreos gravimétricos en el punto de monitoreo 1 y ocho muestreos gravimétricos en los puntos de monitoreo 2 y 3, a partir del evento de precipitación ocurrido el 18 de agosto de 2025, que registró una lámina acumulada de 57 mm. Se ajustaron tres modelos no lineales con el propósito de estimar la humedad gravimétrica del suelo a partir de las lecturas de resistencia eléctrica registradas por los sensores Watermark. Los suelos monitoreados presentaron valores de capacidad de campo entre 17.3 % y 21.3 % de humedad gravimétrica y valores de punto de marchitez permanente entre 7 % y 9 %, características representativas de suelos de textura franco-arcillo-arenosa. Se evaluaron modelos de tipo potencial y exponencial, así como una adaptación del modelo biexponencial propuesto por Omuto (2009).

La Figura 5.6 presenta los modelos seleccionados para cada punto de monitoreo a partir de los mejores indicadores estadísticos de ajuste. Para el punto de monitoreo 1, el modelo potencial mostró un desempeño satisfactorio con un R^2 de 0.732 y un RMSE de 0.966 puntos porcentuales. En los puntos de monitoreo 2 y 3, la adaptación del modelo de Omuto presentó los mejores resultados, el cual alcanzó coeficientes de determinación de 0.784 y 0.860, respectivamente, así como errores RMSE de 0.842 y 0.682 puntos porcentuales. Estos resultados evidencian una relación consistente entre la resistencia medida por los sensores Watermark y la humedad gravimétrica del suelo, lo que permitió utilizar los modelos ajustados para la estimación continua del contenido de humedad durante el periodo de estudio.

Aunque los tres puntos de monitoreo se localizaron sobre suelos franco-arcillo-arenosos, los ajustes obtenidos presentaron diferencias entre sitios, lo que se reflejó en coeficientes de determinación entre 0.73 y 0.86. Estas diferencias sugieren que factores locales, como la estructura del suelo, la densidad aparente

y la distribución espacial de la humedad, influyen en la respuesta del sensor. Este comportamiento coincide con los resultados presentados por Servin-Palestina et al. (2025), quienes demostraron que las ecuaciones de conversión deben desarrollarse para cada tipo de suelo y condición local, ya que no existe una relación universal entre la señal del sensor y el contenido de agua.

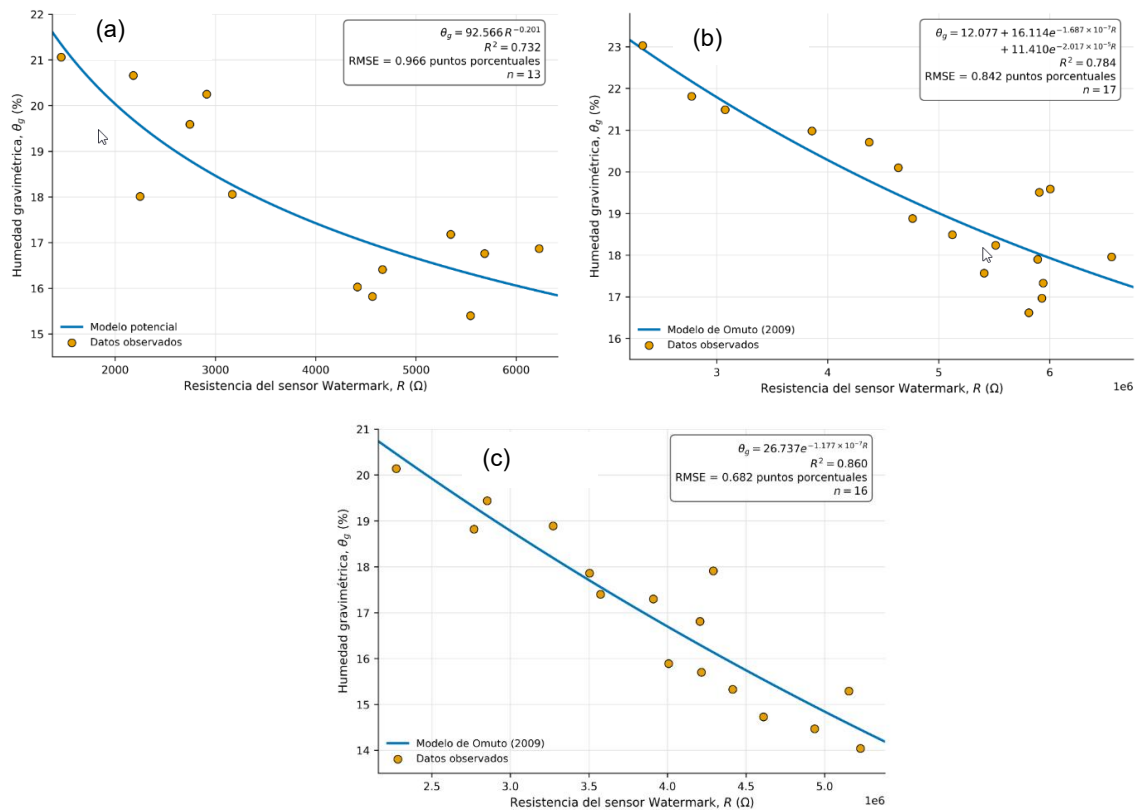


Figura 5.6. Modelos ajustados para los puntos de monitoreo (a) punto 1, (b) punto 2 y (c) punto 3.

La adaptación del modelo de omuto presentó el mejor desempeño en los puntos 2 y 3, con valores de R^2 de 0.784 y 0.860, respectivamente. En el presente estudio, el modelo se aplicó directamente a la resistencia eléctrica medida por los sensores Watermark, una variable para la cual no fue concebido originalmente. Sin embargo, tanto el potencial matricial como la resistencia eléctrica responden a cambios en la distribución y continuidad del agua dentro de la red de poros del

suelo, lo que explica que una formulación desarrollada para potencial matricial también pueda representar adecuadamente la respuesta eléctrica del sensor.

El mejor ajuste se obtuvo en el punto 3 ($R^2 = 0.860$; $RMSE = 0.682 \%$), precisamente donde se registró la mayor densidad aparente (1.42 g cm^{-3}), mientras que los puntos 1 y 2 presentaron densidades aparentes de 1.25 g cm^{-3} . Un incremento de la densidad aparente suele asociarse con una reducción de la macroporosidad y una menor heterogeneidad estructural del suelo. Bajo estas condiciones, la redistribución del agua alrededor del sensor puede ocurrir de forma más uniforme, lo que favorece una relación más estable entre humedad y resistencia eléctrica. Esta situación podría haber contribuido a la reducción de la dispersión observada en los datos del punto 3 y, por consiguiente, a una mayor precisión de la calibración.

En conjunto, los resultados demuestran que los sensores Watermark constituyen una herramienta adecuada para el monitoreo continuo de la humedad del suelo, siempre que se disponga de calibraciones locales adaptadas a las condiciones edáficas del sitio. Además, el buen desempeño del modelo biexponencial de Omuto sugiere que su uso podría extenderse más allá de la representación tradicional de curvas de retención de agua, por lo que se podría modelar variables indirectamente relacionadas con el estado hídrico del suelo, como la resistencia eléctrica registrada por sensores de matriz granular.

5.3.2 Comparación entre sensores Watermark y balance hídrico

La Cuadro 5.1 presenta los resultados de los análisis de las propiedades físicas e hidráulicas de los suelos presentes en los puntos de monitoreo empleados para los balances de agua en el suelo. Al respecto, los puntos de monitoreo asociados a suelos franco arcillo-arenosos mostraron una mayor capacidad para almacenar agua en comparación con el suelo franco arenoso, el cual presentó además una mayor densidad aparente.

Cuadro 5.1. Propiedades físicas e hidráulicas de los puntos de monitoreo de humedad del suelo.

Punto de monitoreo	CC (%)	PMP (%)	DAP (g cm ⁻³)	Textura
1	19.00	9.0	1.25	Franco arcillo-arenoso
2	21.26	7.0	1.25	Franco arcillo-arenoso
3	17.30	9.0	1.42	Franco arenoso

CC=Capacidad de Campo, PMP=Punto de Marchitez Permanente, DAP=Densidad Aparente

La comparación entre el agotamiento de agua estimado mediante el balance hídrico y el determinado a partir de los sensores Watermark durante el periodo del 4 de agosto al 4 de septiembre mostró RMSE comprendidos entre 2.94 y 3.5 mm, y valores de MAE entre 2.49 y 2.73 mm. El mejor desempeño se obtuvo en el punto 3 con el coeficiente de determinación mayor y el menor MAE (Cuadro 5.2).

Cuadro 5.2. Comparación entre sensores Watermark y balance hídrico.

Punto	n	R ²	RMSE (mm)	MAE (mm)
1	17	0.68	2.94	2.52
2	32	0.51	3.85	2.73
3	32	0.83	3.49	2.49

Diversos estudios han señalado que los sensores Watermark requieren calibraciones específicas para las condiciones locales del suelo. Varble & Chávez (2011) reportaron que las ecuaciones de fábrica pueden sobreestimar el contenido volumétrico de agua entre 0.08 y 0.20 m³ m⁻³, mientras que la calibración de campo redujo significativamente los errores de estimación. De forma similar, Marković et al. (2021) encontraron que el desempeño del sensor depende de la profundidad de instalación y de las características hidráulicas del suelo. Los autores obtuvieron mejores resultados después de realizar ajustes específicos para cada condición experimental. Estos antecedentes respaldan la necesidad de las calibraciones desarrolladas en el presente estudio y justifican las diferencias observadas entre los puntos de monitoreo.

Los resultados obtenidos también pueden estar influenciados por la variabilidad espacial inherente al muestreo. Las muestras gravimétricas utilizadas para la calibración se obtuvieron dentro de un radio aproximado de 1.5 m respecto a cada sensor, por lo que la humedad determinada en laboratorio no necesariamente correspondió al mismo volumen de suelo monitoreado por el dispositivo. Nolz & Loiskandl (2017) demostraron que pequeñas diferencias en la localización de los puntos de monitoreo pueden generar variaciones significativas en la humedad observada dentro de una misma parcela agrícola, mientras que Janssens et al. (2015) reportaron que incluso la ubicación del sensor respecto a la zona húmeda del emisor puede afectar considerablemente sus lecturas. En consecuencia, parte de la dispersión observada entre el balance hídrico y los sensores Watermark podría atribuirse a la heterogeneidad espacial del suelo.

5.3.3 Desempeño del modelo de segmentación semántica

En el Cuadro 5.3 se presentan las métricas de evaluación obtenidas por el modelo con el conjunto de datos de validación, asimismo en la Figura 5.7 se presentan los resultados visuales de la segmentación de algunas parcelas para el 3 de agosto. Los resultados obtenidos evidencian una alta capacidad del modelo para discriminar las clases maíz, suelo y malezas presentes en las imágenes UAV. La combinación de valores altos de precisión y recall indica que el modelo mantuvo simultáneamente una baja proporción de falsas asignaciones y una alta capacidad para recuperar los píxeles pertenecientes a cada clase. La elevada concordancia entre las máscaras de referencia y las predicciones generadas indica que la arquitectura utilizada fue capaz de capturar adecuadamente tanto la información espectral como los patrones espaciales asociados a cada cobertura. Este comportamiento es consistente con la literatura reciente, donde las redes convolucionales profundas han demostrado una capacidad superior para clasificar y segmentar cultivos agrícolas respecto a los enfoques tradicionales basados en características espectrales individuales (Xing et al., 2026; H. Zhu et al., 2024) .

Cuadro 5.3. Desempeño del modelo de segmentación semántica.

Clase	Precisión	Recall	F1-score	IoU	Píxeles evaluados
Maíz	0.998	0.997	0.998	0.995	4,757,227
Suelo	0.999	0.999	0.999	0.998	7,032,496
Maleza	0.999	0.999	0.999	0.998	6,847,216

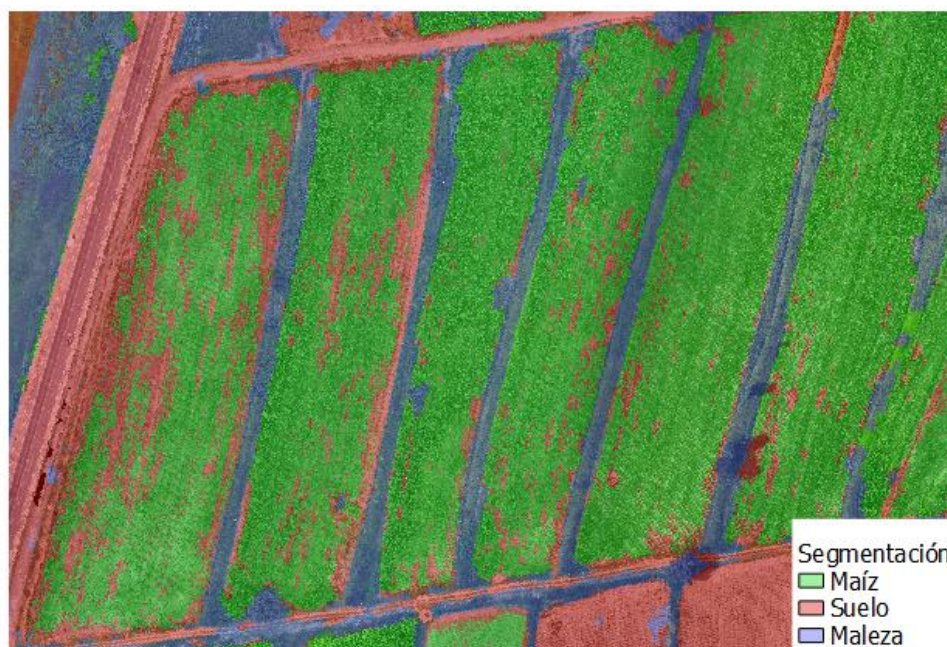


Figura 5.7. Segmentación de parcelas del 3 de agosto de 2025.

La precisión alcanzada se encuentra dentro del rango superior reportado para aplicaciones de segmentación de maíz mediante aprendizaje profundo. Por ejemplo, Kitano et al. (2024) reportaron una precisión de validación cercana a 0.999 en tareas de conteo de plantas mediante U-Net, mientras que Vong et al. (2021) obtuvieron valores de F1 entre 0.85 y 0.96 en diferentes sistemas de manejo agrícola. Liu et al. (2021) implementaron mejoras al U-Net original mediante el uso de SE-ResNeXt-50 como backbone y módulos ASPP (Atrous Spatial Pyramid Pooling), lo que les permitió obtener un F1-score de 0.9034 y un índice de Jaccard de 0.8287 para la segmentación y gradación de sequía en

maíz. De forma similar, Li et al. (2024) reportaron valores de F1 cercanos a 0.91 e índices de intersección sobre unión cercanos a 0.85 mediante la arquitectura EENet. Los resultados obtenidos en el presente estudio muestran un desempeño comparable o superior a estos trabajos, lo que confirma el potencial de las arquitecturas de segmentación profunda para la caracterización espacial de cultivos de maíz.

La elevada capacidad de discriminación observada puede atribuirse a dos factores principales. En primer lugar, la arquitectura U-Net permite conservar información espacial detallada mediante conexiones de salto entre el encoder y el decoder, mejorando la delimitación de los objetos presentes en la imagen. En segundo lugar, el uso simultáneo de bandas multiespectrales e índices de vegetación proporciona información complementaria sobre vigor, cobertura y estructura de la vegetación, incrementando la separabilidad espectral entre cultivo, suelo y maleza. Estudios recientes han demostrado que la combinación de información espectral e índices derivados mejora significativamente el desempeño de los modelos de clasificación agrícola frente al uso exclusivo de bandas multiespectrales (Zhang et al., 2025).

5.3.4 Desempeño del modelo U-Net VGG para la estimación de Kcdual

El Cuadro 5.4 presenta las métricas de evaluación promedio de las imágenes inferidas por el modelo entrenado de las cuatro fechas de analizadas. Los resultados obtenidos muestran que la arquitectura U-Net VGG fue capaz de reproducir adecuadamente la distribución espacial del coeficiente dual de cultivo en las cuatro fechas evaluadas. El mejor desempeño se obtuvo con las imágenes UAV, seguidas por las imágenes superresueltas y finalmente por las imágenes Sentinel-2 originales. Las diferencias observadas sugieren que la capacidad predictiva del modelo depende principalmente de la cantidad y calidad de la información espacial y espectral disponible en los datos originales. En este sentido, las imágenes UAV conservaron una representación más detallada de la

estructura del cultivo, mientras que las imágenes superresueltas lograron recuperar parcialmente patrones espaciales que no pueden identificarse en las imágenes Sentinel-2 originales.

La superioridad observada en las imágenes UAV coincide con diversos trabajos que han demostrado que la información adquirida mediante drones permite estimar variables directamente relacionadas con los coeficientes de cultivo, como la cobertura vegetal, la fracción de cobertura verde, la altura del cultivo y el índice de área foliar. Marcial-Pablo et al. (2021) obtuvieron un R^2 de 0.94 en la estimación de coeficientes de cultivo mediante índices espectrales derivados de UAV, mientras que Shao et al. (2021) alcanzaron un R^2 de 0.65 mediante Random Forest e índices espectrales y LAI. Estos resultados sugieren que la alta resolución espacial proporcionada por los UAV permite representar con mayor precisión las variaciones espaciales de Kc dentro de una parcela agrícola.

Cuadro 5.4. Desempeño de la red U-Net VGG respecto a la fuente de datos.

Fuente de datos	R^2	RMSE	MAE	Bias
Dron	0.883	0.093	0.043	-0.018
Sentinel-2	0.774	0.125	0.061	-0.023
Superresueltas	0.827	0.115	0.054	-0.018

La reducción del desempeño observada en Sentinel-2 puede atribuirse principalmente al efecto de mezcla espectral generado por la resolución espacial de 10 m. Bukowiecki et al. (2021) señalaron que las imágenes Sentinel-2 representan adecuadamente las tendencias generales del desarrollo del cultivo, pero presentan limitaciones para capturar la variabilidad espacial dentro de la parcela. Sin embargo, aún bajo estas condiciones, Nagy et al. (2024) reportaron correlaciones superiores a 0.83 entre índices espectrales derivados de Sentinel-2 y coeficientes de cultivo en maíz, mientras que French et al. (2023) obtuvieron coeficientes de determinación entre 0.95 y 0.96 en estimaciones de

evapotranspiración basadas en series temporales derivadas de Sentinel-2. De manera similar, Salvador-Castillo et al. (2021) demostraron que los índices espectrales obtenidos mediante sensores remotos constituyen predictores confiables de variables relacionadas con la evapotranspiración de los cultivos. Lo anterior demuestra la utilidad de esta fuente de información para aplicaciones de predicción de ETc.

La mejora observada al utilizar imágenes superresueltas indica que la información espacial recuperada mediante el modelo TVDSR permitió representar con mayor detalle los patrones asociados al desarrollo del cultivo. El incremento del coeficiente de determinación y la reducción de los errores respecto a Sentinel-2 sugieren que la resolución espacial constituye uno de los factores más importantes para la estimación de Kcdual. Este comportamiento coincide con la tendencia reciente hacia la integración de técnicas de superresolución y aprendizaje profundo para mejorar la representación espacial de variables biofísicas agrícolas derivadas de sensores remotos.

Los resultados obtenidos confirman que la información espectral contenida en las imágenes multiespectrales posee una relación consistente con la distribución espacial del coeficiente dual de cultivo. Diversos autores han demostrado que variables derivadas de sensores remotos, como NDVI, SAVI, NDWI, fracción de cobertura vegetal, índice de área foliar y altura del cultivo, presentan relaciones significativas con los coeficientes de cultivo utilizados en FAO-56. Pereira et al. (2021) y Pôças et al. (2020) destacan que estas variables constituyen indicadores robustos para estimar Kc y Kcb a distintas escalas espaciales, mientras que Kamble et al. (2013) demostraron que los índices de vegetación derivados de sensores remotos permiten estimar coeficientes de cultivo con alta precisión. En consecuencia, el desempeño alcanzado por la arquitectura U-Net VGG sugiere que la red fue capaz de extraer patrones espaciales relevantes asociados con las

variables biofísicas que controlan la variabilidad de K_{cdual} dentro de las parcelas experimentales.

La capacidad de las redes profundas para modelar este tipo de relaciones ha sido ampliamente documentada recientemente. Chen et al. (2020) alcanzaron un R^2 de 0.95 en la estimación de evapotranspiración y coeficientes de cultivo del maíz mediante Redes Convolucionales Temporales (TCN), mientras que Zhangzhong et al. (2026) reportaron un R^2 de 0.89 mediante una arquitectura multimodal basada en EfficientNet y Transformers. Asimismo, Zhao et al. (2024) demostraron que la incorporación simultánea de variables meteorológicas, humedad del suelo y características de crecimiento mejora significativamente la representación de los procesos que controlan el consumo hídrico del cultivo. Estos antecedentes respaldan la utilización de redes neuronales convolucionales como alternativa para la espacialización de K_{cdual} a partir de información espectral.

La distribución espacial observada en la Figura 5.8 muestra que la red U-Net VGG generó mapas continuos de K_{cdual} con una marcada variabilidad espacial entre parcelas y dentro de las mismas. Los patrones observados indican que el modelo fue capaz de diferenciar áreas con distintos valores de K_{cdual} y representar la heterogeneidad espacial presente en el cultivo. La ausencia de transiciones abruptas y la coherencia espacial de las zonas reconstruidas sugieren que la red logró preservar relaciones espaciales relevantes asociadas a las condiciones biofísicas de la vegetación.

Finalmente, los resultados obtenidos sugieren que las redes convolucionales constituyen una herramienta eficaz para estimar coeficientes de cultivo espacialmente distribuidos a partir de imágenes multiespectrales. La capacidad del modelo para generar mapas continuos de K_{cdual} permite superar el uso de coeficientes promedio asociados a parcelas completas y avanzar hacia representaciones espacialmente explícitas del desarrollo del cultivo. Esta característica es relevante en agricultura de precisión, donde pequeñas diferencias en vigor vegetal, cobertura del suelo o disponibilidad hídrica pueden

traducirse en variaciones significativas de los requerimientos de agua dentro de una misma parcela.

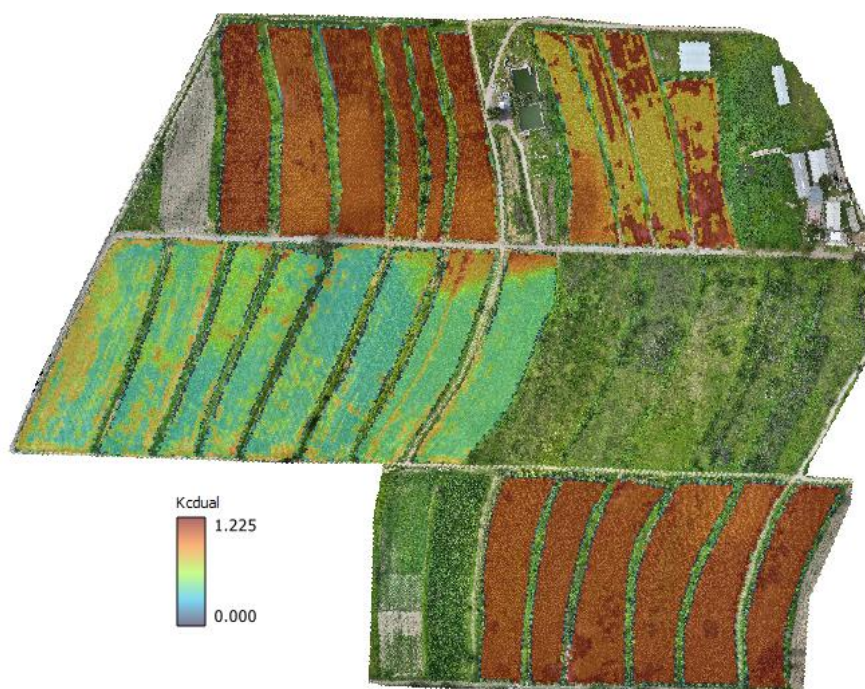


Figura 5.8. Mapa de Kcdual predicho para el 3 de agosto de 2025 generado con la imagen superresuelta.

5.3.5 Estimación espacial de la ETc del maíz

El Cuadro 5.5 presenta las métricas de comparación entre la ETc estimada a partir de los mapas de Kcdual generados mediante las diferentes fuentes de imágenes y la evapotranspiración calculada mediante los mapas Kcdual objetivo. Los resultados muestran que la evapotranspiración derivada de las imágenes UAV presentó la mejor concordancia con los valores objetivo, los cuales alcanzaron un coeficiente de determinación de 0.88, un RMSE de 0.40 mm d⁻¹ y un MAE de 0.18 mm d⁻¹.

Cuadro 5.5. Evaluación de la ETc derivada de imágenes UAV, Sentinel-2 y Sentinel-2 superresueltas mediante balance hídrico.

Fuente	No. fechas	R ²	RMSE (mmd ⁻¹)	MAE (mmd ⁻¹)	Bias (mmd ⁻¹)
Dron	4	0.875	0.398	0.180	-0.068
Sentinel-2	4	0.789	0.525	0.257	-0.045
Sentinel-2 superresuelto	4	0.857	0.442	0.213	-0.054

La superresolución produjo una mejora evidente respecto a Sentinel-2 nativo. El coeficiente de determinación aumentó 8.6 %, mientras que el RMSE y el MAE disminuyeron 15.8 % y 17.1 %, respectivamente (Cuadro 5.5). Sin embargo, el aspecto más relevante es la proximidad alcanzada respecto a las imágenes UAV. Aunque las imágenes UAV presentaron el mejor desempeño global, las diferencias observadas respecto a las imágenes superresueltas fueron reducidas. El coeficiente de determinación difirió únicamente en 0.026 unidades, mientras que las diferencias en RMSE y MAE correspondieron a 0.044 mm d⁻¹ y 0.033 mm d⁻¹, respectivamente. En consecuencia, los resultados sugieren que las imágenes Sentinel-2 superresueltas pueden constituir una alternativa viable para la estimación espacial de la evapotranspiración en superficies extensas, donde la adquisición periódica de imágenes UAV implica costos operativos elevados, restricciones logísticas o limitaciones de cobertura espacial.

Este comportamiento es consistente con los resultados obtenidos en modelos de balance de energía como SEBAL, METRIC, SEBS y TSEB, donde la exactitud de la evapotranspiración depende de la capacidad para representar adecuadamente la variabilidad espacial de las variables biofísicas utilizadas como entrada. Guzinski & Nieto (2019) demostraron que el incremento de la resolución espacial mejora la representación de parcelas individuales y reduce los efectos de mezcla espectral, mientras que Guzinski et al. (2020) observaron que la combinación de Sentinel-2 y Sentinel-3 mejora significativamente la estimación de

evapotranspiración respecto a productos de resolución más gruesa. De forma similar, Barrios et al. (2018) reportaron que una mayor resolución espacial reduce el sesgo y los errores de los mapas de evapotranspiración al representar con mayor precisión la heterogeneidad agrícola. Estas observaciones sugieren que la información espacial recuperada mediante superresolución permitió representar de manera más consistente los patrones biofísicos que controlan el consumo hídrico del cultivo.

Los errores obtenidos en el presente estudio son comparables con los reportados por investigaciones que estiman evapotranspiración mediante coeficientes de cultivo derivados de sensores remotos. Por ejemplo, Tang et al. (2019) reportaron un RMSE de 0.95 mm d^{-1} con imágenes UAV y el enfoque FAO-56 dual, mientras que Méndez-Jocik et al. (2021) obtuvieron errores de 1.16 mm d^{-1} mediante Sentinel-2 y Landsat. De manera similar, Hassan et al. (2022) reportaron valores de RMSE cercanos a 0.98 mm d^{-1} en la estimación de evapotranspiración con relaciones entre NDVI y coeficientes de cultivo. Los errores obtenidos en este trabajo, inferiores a 0.53 mm d^{-1} para las tres fuentes evaluadas, se encuentran dentro del rango reportado por la literatura especializada.

Los errores obtenidos también son comparables a los reportados por metodologías basadas en balance de energía. Tunca (2024) obtuvo un RMSE cercano a 1.1 mm d^{-1} utilizando TSEB e imágenes UAV en sorgo, mientras que Meza et al. (2025) reportaron errores cercanos a 0.6 mm d^{-1} mediante TSEB aplicado a imágenes térmicas y multiespectrales adquiridas con UAV. Del mismo modo, Sabharwal et al. (2026) alcanzaron un RMSE de 0.65 mm d^{-1} mediante el modelo SEBAL y datos Landsat 8 para estimar la evapotranspiración del maíz. En comparación, los errores obtenidos en este estudio se encuentran dentro del rango reportado para modelos propuestos anteriormente, lo que respalda la viabilidad del enfoque basado en K_{cdual} derivado mediante aprendizaje profundo.

Las tres fuentes presentaron valores negativos de Bias, lo que indica una ligera tendencia a subestimar la evapotranspiración. Este comportamiento coincide con observaciones reportadas por Guzinski et al. (2023), quienes obtuvieron sesgos negativos al estimar evapotranspiración mediante la integración de Sentinel-2, Sentinel-3 y Landsat.

La Figura 5.6 muestra la distribución espacial de la evapotranspiración del cultivo de maíz estimada para el 3 de agosto de 2025, en cuatro parcelas de maíz, mediante tres fuentes de información: a) imágenes UAV, b) imágenes Sentinel-2 y c) imágenes Sentinel-2 superresueltas. Se observa que las tres fuentes reproducen patrones espaciales similares de consumo hídrico dentro de las parcelas, identifican las mismas zonas de mayor y menor evapotranspiración. Esto sugiere que, independientemente de la fuente utilizada, los modelos fueron capaces de capturar la estructura espacial dominante del cultivo y su variabilidad a escala parcelaria.

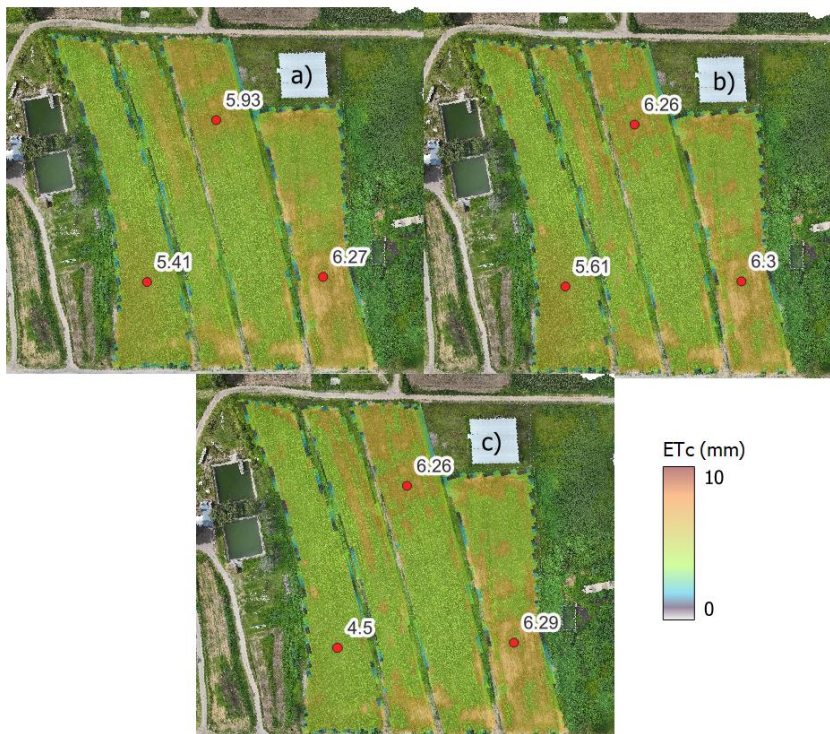


Figura 5.9. Mapas de evapotranspiración del maíz obtenidos a partir de imágenes a) UAV, y b) Sentinel-2 superresuelta y c) Sentinel-2.

Sin embargo, aunque los patrones espaciales son similares, se observan diferencias en la magnitud de los valores puntuales de ET_c , lo cual indica que las diferencias observadas entre fuentes no están asociadas a la identificación de las zonas de mayor o menor consumo hídrico, sino a la magnitud de la evapotranspiración asignada a determinados sectores de la parcela. French et al. (2023) señalaron que las imágenes Sentinel-2 permiten representar adecuadamente la variabilidad espacial y temporal de la evapotranspiración mediante coeficientes de cultivo dinámicos, mientras que Nagy et al. (2024) reportaron una elevada capacidad de los índices derivados de Sentinel-2 para estimar coeficientes de cultivo del maíz y generar mapas espaciales de ET_c . Asimismo, Jia et al. (2024) demostraron que la superresolución mejora la delimitación de parcelas y aumenta el número de píxeles puros y favorece una representación más precisa de las variables biofísicas utilizadas para la estimación de ET.

Como línea futura de investigación, resulta necesario validar las estimaciones de evapotranspiración obtenidas mediante la metodología propuesta utilizando mediciones directas realizadas con instrumentos de alta precisión, tales como lisímetros de pesada o sistemas Eddy Covariance. Esto permitiría cuantificar con mayor exactitud el error asociado a las estimaciones generadas mediante imágenes Sentinel-2, UAV y productos superresueltos. Asimismo, es conveniente extender la metodología a otros cultivos agrícolas con características morfológicas y fenológicas diferentes, con el propósito de evaluar la capacidad de generalización de los modelos desarrollados y determinar su potencial aplicación en diversos escenarios de agricultura de precisión. De manera complementaria, la incorporación de información térmica obtenida mediante cámaras termográficas podría fortalecer la capacidad predictiva de los modelos, al aportar información directa sobre el estado hídrico del cultivo y la temperatura del dosel vegetal. Finalmente, sería recomendable comparar el desempeño de la metodología propuesta frente a modelos basados en balance de energía, con el

fin de evaluar sus ventajas y limitaciones bajo diferentes condiciones de operación y determinar su viabilidad para aplicaciones operativas de monitoreo agrícola y gestión del riego.

5.4 Conclusiones

Se desarrolló una metodología para estimar la distribución espacial de la evapotranspiración del cultivo de maíz mediante redes neuronales convolucionales e índices de vegetación obtenidos a partir de información Sentinel-2 y UAV. Los resultados demostraron que la superresolución espacial permite recuperar información relevante para la estimación de coeficientes de cultivo y evapotranspiración, lo cual reduce la brecha de desempeño existente entre las imágenes satelitales de resolución moderada y las observaciones UAV de alta resolución espacial.

El modelo de segmentación semántica basado en U-Net presentó una buena capacidad para discriminar las clases de maíz, suelo y maleza con métricas de precisión, recall y F1-score superiores 0.99 para todas las categorías evaluadas. Asimismo, la arquitectura U-Net VGG demostró un desempeño alto, con valores de $R^2 > 0.79$, para la estimación del K_{cdual} a partir de imágenes multiespectrales.

La estimación de la ET_c mostró que las imágenes UAV alcanzaron la mejor correspondencia con la ET_c obtenida mediante los mapas de K_{cdual} objetivo ($R^2 = 0.875$; $RMSE = 0.398 \text{ mm d}^{-1}$; $MAE = 0.18 \text{ mm d}^{-1}$). La superresolución incrementó el coeficiente de determinación 8.6 %, redujo el RMSE en 16 % y el MAE el 17% respecto a Sentinel-2 nativo, lo que evidencia que la recuperación de detalle espacial mejora la representación de la variabilidad biofísica relacionada con el consumo hídrico del cultivo. Los resultados obtenidos confirman que la integración de imágenes UAV, imágenes Sentinel-2, técnicas de superresolución espacial y redes neuronales convolucionales permite generar

mapas espaciales de coeficientes de cultivo y evapotranspiración con precisión suficiente para aplicaciones de agricultura de precisión.

5.5 Literatura Citada

- Amani, S., & Shafizadeh-Moghadam, H. (2023). A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. In *Agricultural Water Management* (Vol. 284). Elsevier B.V. <https://doi.org/10.1016/j.agwat.2023.108324>
- Ankela, C. B., Neelima, T. L., Chandrasekar, K., Misra, N., Avil Kumar, K., & Noorzai, A. (2026). Spatiotemporal assessment of maize evapotranspiration and surface energy fluxes under varying irrigation regimes using UAV based METRIC. *Scientific Reports*, 16(1). <https://doi.org/10.1038/s41598-025-33916-5>
- Barrios, J. M., Ghilain, N., Arboleda, A., Sachs, T., & Gellens-Meulenberghs, F. (2018). Daily evapotranspiration at sub-kilometre spatial resolution by combining observations from geostationary and polar-orbit satellites. *International Journal of Remote Sensing*, 39(23), 8984–9003. <https://doi.org/10.1080/01431161.2018.1504340>
- Beretta, A. N., Silbermann, A. V., Paladino, L., Torres, D., Bassahun, D., Musselli, R., & García-Lamohte, A. (2014). Análisis de textura del suelo con hidrómetro: Modificaciones al método de Bouyoucus. *Ciencia e Investigacion Agraria*, 41(2), 263–271. <https://doi.org/10.4067/S0718-16202014000200013>
- Bukowiecki, J., Rose, T., & Kage, H. (2021). Sentinel-2 data for precision agriculture?—a uav-based assessment. *Sensors*, 21(8). <https://doi.org/10.3390/s21082861>
- Chen, Z., Sun, S., Wang, Y., Wang, Q., & Zhang, X. (2020). Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. *Computers and Electronics in Agriculture*, 169. <https://doi.org/10.1016/j.compag.2019.105206>
- FAO. (2023). *Standard operating procedure for soil moisture content by gravimetric method* (FAO, Ed.). FAO. <https://openknowledge.fao.org/handle/20.500.14283/cc4831en>
- Fehér, Z. Z., Nxumalo, G. S., & Nagy, A. (2026). Multi-Sensor NDVI Fusion for Daily Crop Evapotranspiration Mapping: A Six-Year Irrigated Maize Assessment Using MODIS–Sentinel-2–Landsat (2020–2025). *Sensors*, 26(14). <https://doi.org/10.3390/s26144470>
- Feng, X., Liu, H., Feng, D., Tang, X., Li, L., Chang, J., Tanny, J., & Liu, R. (2023). Quantifying winter wheat evapotranspiration and crop coefficients under sprinkler irrigation using eddy covariance technology in the North China Plain. *Agricultural Water Management*, 277. <https://doi.org/10.1016/j.agwat.2022.108131>
- French, A. N., Sanchez, C. A., Wirth, T., Scott, A., Shields, J. W., Bautista, E., Saber, M. N., Wisniewski, E., & Gohardoust, M. R. (2023). Remote sensing of

- evapotranspiration for irrigated crops at Yuma, Arizona, USA. *Agricultural Water Management*, 290. <https://doi.org/10.1016/j.agwat.2023.108582>
- Guzinski, R., & Nieto, H. (2019). Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations. *Remote Sensing of Environment*, 221, 157–172. <https://doi.org/10.1016/j.rse.2018.11.019>
- Guzinski, R., Nieto, H., Ramo Sánchez, R., Sánchez, J. M., Jomaa, I., Zitouna-Chebbi, R., Roupsard, O., & López-Urrea, R. (2023). Improving field-scale crop actual evapotranspiration monitoring with Sentinel-3, Sentinel-2, and Landsat data fusion. *International Journal of Applied Earth Observation and Geoinformation*, 125. <https://doi.org/10.1016/j.jag.2023.103587>
- Guzinski, R., Nieto, H., Sandholt, I., & Karamitilios, G. (2020). Modelling high-resolution actual evapotranspiration through Sentinel-2 and Sentinel-3 data fusion. *Remote Sensing*, 12(9). <https://doi.org/10.3390/RS12091433>
- Hassan, D. F., Abdalkadhum, A. J., Mohammed, R. J., & Shaban, A. (2022). Integration Remote Sensing and Meteorological Data to Monitoring Plant Phenology and Estimation Crop Coefficient and Evapotranspiration. *Journal of Ecological Engineering*, 23(4), 325–335. <https://doi.org/10.12911/22998993/146267>
- Janssens, P., Diels, J., Vanderborght, J., Elsen, F., Elsen, A., Deckers, T., & Vandendriessche, H. (2015). Numerical calculation of soil water potential in an irrigated “conference” pear orchard. *Agricultural Water Management*, 148, 113–122. <https://doi.org/10.1016/j.agwat.2014.09.023>
- Kamble, B., Kilic, A., & Hubbard, K. (2013). Estimating crop coefficients using remote sensing-based vegetation index. *Remote Sensing*, 5(4), 1588–1602. <https://doi.org/10.3390/rs5041588>
- Kitano, B. T., Mendes, C. C. T., Geus, A. R., Oliveira, H. C., & Souza, J. R. (2024). Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/LGRS.2019.2930549>
- Li, J., Pu, F., Chen, H., Xu, X., & Yu, Y. (2024). Crop Segmentation of Unmanned Aerial Vehicle Imagery Using Edge Enhancement Network. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. <https://doi.org/10.1109/LGRS.2024.3358983>
- Liu, C., Li, H., Su, A., Chen, S., & Li, W. (2021). Identification and Grading of Maize Drought on RGB Images of UAV Based on Improved U-Net. *IEEE Geoscience and Remote Sensing Letters*, 18(2), 198–202. <https://doi.org/10.1109/LGRS.2020.2972313>
- Liu, N., Liu, Q., Liu, X., Zhang, X., & Shao, L. (2026). Upscaling instantaneous ET obtained using UAV multispectral and thermal data into daily ET with and without UAV flights. *European Journal of Agronomy*, 172. <https://doi.org/10.1016/j.eja.2025.127865>
- Marcial-Pablo, M. de J., Ontiveros-Capurata, R. E., Jiménez-Jiménez, S. I., & Ojeda-Bustamante, W. (2021). Maize crop coefficient estimation based on spectral vegetation indices and vegetation cover fraction derived from uav-based multispectral images. *Agronomy*, 11(4). <https://doi.org/10.3390/AGRONOMY11040668>

- Marković, M., Krizmanić, G., Brkić, A., Atilgan, A., Japundžić-Palenkić, B., Petrović, D., & Barač, Ž. (2021). Sustainable management of water resources in supplementary irrigation management. *Applied Sciences (Switzerland)*, 11(6). <https://doi.org/10.3390/app11062451>
- Méndez-Jocik, C. Alberto., Núñez-Pérez, A. de J., & González-Robaina, F. (2021). Vista de Aplicación de la teledetección espacial al manejo de los recursos hídricos. *Revista Ciencias Técnicas Agropecuarias*, 30(1), 17–25.
- Meza, K., Torres-Rua, A. F., Hipps, L., Kustas, W. P., Gao, R., Christiansen, L., Kopp, K., Nieto, H., Burchard-Levine, V., Martín, M. P., Coopmans, C., & Gowing, I. (2025). Spatial estimation of actual evapotranspiration over irrigated turfgrass using sUAS thermal and multispectral imagery and TSEB model. *Irrigation Science*, 43(1), 5–28. <https://doi.org/10.1007/s00271-023-00899-y>
- Murguía-Cozar, A., López-Canteñs, G. de J., Macedo-Cruz, A., Velazquez-López, N., & Ontiveros-Capurata, R. E. (2026). Fusion of UAV–Sentinel-2 images using convolutional neural networks for remote-sensing analysis. *International Journal of Image and Data Fusion*, 17(1). <https://doi.org/10.1080/19479832.2026.2701715>
- Nagy, A., Kiss, N. É., Buday-Bódi, E., Magyar, T., Cavazza, F., Gentile, S. L., Abdullah, H., Tamás, J., & Fehér, Z. Z. (2024). Precision Estimation of Crop Coefficient for Maize Cultivation Using High-Resolution Satellite Imagery to Enhance Evapotranspiration Assessment in Agriculture. *Plants*, 13(9). <https://doi.org/10.3390/plants13091212>
- Nolz, R., & Loiskandl, W. (2017). Evaluating soil water content data monitored at different locations in a vineyard with regard to irrigation control. *Soil and Water Research*, 12(3), 152–160. <https://doi.org/10.17221/9/2016-SWR>
- Omuto, C. T. (2009). Biexponential model for water retention characteristics. *Geoderma*, 149(3–4), 235–242. <https://doi.org/10.1016/j.geoderma.2008.12.001>
- Pereira, L. S., Allen, R. G., Paredes, P., López-Urrea, R., Raes, D., Smith, M., Kilic, A., & Salman, M. (2025). Crop evapotranspiration-Guidelines for computing crop water requirements. In *Crop evapotranspiration* (Second). FAO. <https://doi.org/10.4060/cd6621en>
- Pereira, L. S., Paredes, P., Hunsaker, D. J., López-Urrea, R., & Mohammadi Shad, Z. (2021). Standard single and basal crop coefficients for field crops. Updates and advances to the FAO56 crop water requirements method. In *Agricultural Water Management* (Vol. 243). Elsevier B.V. <https://doi.org/10.1016/j.agwat.2020.106466>
- Pereira, L. S., Paredes, P., Melton, F., Johnson, L., Wang, T., López-Urrea, R., Cancela, J. J., & Allen, R. G. (2020). Prediction of crop coefficients from fraction of ground cover and height. Background and validation using ground and remote sensing data. *Agricultural Water Management*, 241, 106197. <https://doi.org/10.1016/j.agwat.2020.106197>
- Phang, S. K., Chiang, T. H. A., Happonen, A., & Chang, M. M. L. (2023). From Satellite to UAV-Based Remote Sensing: A Review on Precision Agriculture. *IEEE Access*, 11, 127057–127076. <https://doi.org/10.1109/ACCESS.2023.3330886>

- Pôças, I., Calera, A., Campos, I., & Cunha, M. (2020). Remote sensing for estimating and mapping single and basal crop coefficients: A review on spectral vegetation indices approaches. In *Agricultural Water Management* (Vol. 233). Elsevier B.V. <https://doi.org/10.1016/j.agwat.2020.106081>
- Sabharwal, G., Vaiyapuri, K., Selvakumar, S., Jagadeeswaran, R., Raju, M., Pazhanivelan, S., Periyasamy, P., & Sivakumar, S. (2026). Quantification and mapping of maize crop evapotranspiration using the surface energy balance algorithm for land model and medium-resolution satellite data. *Frontiers in Agronomy*, 8. <https://doi.org/10.3389/fagro.2026.1795832>
- Salvador-Castillo, J. M., Bolaños-González, M. A., Rodríguez, J. C., Palacios-Vélez, E., AlbertoPalacios-Sánchez, L., Watts, C., Lizárraga-Celaya, C., Ortega-Farías, S., & Er-Raki, S. (2021). Estimation Of Evapotranspiration Of A Vineyard Of Table Grapes (*Vitis Vinifera*) Using Sentinel-2 Satellite Imagery. *Agrociencia*, 55(5), 369–387. <https://doi.org/10.47163/agrociencia.v55i5.2514>
- SEMARNAT. (2002, December 31). *NOM-021-RECNAT-2000*. Diario Oficial. <https://www.ordenjuridico.gob.mx/Documentos/Federal/wo69255.pdf>
- Servín-Palestina, M., Vargas-Cano, D., Casa-Flores, J. I., & Garcia-Martinez, J. A. (2025). Validación de una estación meteorológica de bajo costo basada en ESP32 para el monitoreo agroclimático. *14° Congreso Internacional de Investigación en Ciencias Básicas y Agronómicas*, 63–76. https://dicea.chapingo.mx/wp-content/uploads/publicaciones/congreso-ciema/memoria-mesa-1_14CIIBA_2025.pdf
- Servin-Palestina, M., Vargas-Cano, D., García-Martínez, J. A., Ramírez-Valle, O., Jiménez-Jiménez, S. I., & Reyes-Gonzales, A. (2025). Calibración del sensor de humedad Watermark 200SS para programación de riego en cinco tipos de suelos de Zacatecas, México. *Agro Productividad*. <https://doi.org/10.32854/tpnv7y27>
- Shao, G., Han, W., Zhang, H., Liu, S., Wang, Y., Zhang, L., & Cui, X. (2021). Mapping maize crop coefficient Kc using random forest algorithm based on leaf area index and UAV-based multispectral vegetation indices. *Agricultural Water Management*, 252. <https://doi.org/10.1016/j.agwat.2021.106906>
- Tang, J., Han, W., & Zhang, L. (2019). UAV multispectral imagery combined with the FAO-56 dual approach for maize evapotranspiration mapping in the North China Plain. *Remote Sensing*, 11(21). <https://doi.org/10.3390/rs11212519>
- Tunca, E. (2024a). Evaluating the performance of the TSEB model for sorghum evapotranspiration estimation using time series UAV imagery. *Irrigation Science*, 42(5), 977–994. <https://doi.org/10.1007/s00271-023-00887-2>
- Tunca, E. (2024b). Evaluating the performance of the TSEB model for sorghum evapotranspiration estimation using time series UAV imagery. *Irrigation Science*, 42(5), 977–994. <https://doi.org/10.1007/s00271-023-00887-2>
- Varble, J. L., & Chávez, J. L. (2011). Performance evaluation and calibration of soil water content and potential sensors for agricultural soils in eastern Colorado. *Agricultural Water Management*, 101(1), 93–106. <https://doi.org/10.1016/j.agwat.2011.09.007>

- Vong, C. N., Conway, L. S., Zhou, J., Kitchen, N. R., & Sudduth, K. A. (2021). Early corn stand count of different cropping systems using UAV-imagery and deep learning. *Computers and Electronics in Agriculture*, 186. <https://doi.org/10.1016/j.compag.2021.106214>
- Xing, Y., Liu, X., & Wang, X. (2026). Integrating UAVs, satellite remote sensing, and machine learning in precision agriculture: pathways to sustainable food production, resource efficiency, and scalable innovation. In *Frontiers in Agronomy* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fagro.2025.1670380>
- Zhang, S., Wang, X., Lin, H., Dong, Y., & Qiang, Z. (2025). A review of the application of UAV multispectral remote sensing technology in precision agriculture. *Smart Agricultural Technology*, 12, 101406. <https://doi.org/10.1016/j.atech.2025.101406>
- Zhangzhong, L., Shen, Y., Li, J., Li, T., Wang, J., Zhu, Q., & Zheng, W. (2026). M2E2Net model: Multimodal fusion for real-time estimation of daily maize evapotranspiration. *Information Processing in Agriculture*. <https://doi.org/10.1016/j.inpa.2026.06.003>
- Zhao, L., Qing, S., Li, H., Qiu, Z., Niu, X., Shi, Y., Chen, S., & Xing, X. (2024). Estimating maize evapotranspiration based on hybrid back-propagation neural network models and meteorological, soil, and crop data. *International Journal of Biometeorology*, 68(3), 511–525. <https://doi.org/10.1007/s00484-023-02608-y>
- Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., Xu, J.-L., & He, Y. (2024). Intelligent agriculture: deep learning in UAV-based remote sensing imagery for crop diseases and pests detection. *Frontiers in Plant Science*, 15. <https://doi.org/10.3389/fpls.2024.1435016>
- Zhu, P., Han, Q., Li, S., Liu, H., Li, C., Ma, Y., & Wang, J. (2025). A Novel Framework Based on Data Fusion and Machine Learning for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. *Remote Sensing*, 17(23). <https://doi.org/10.3390/rs17233813>