



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Geodésica,
Cartográfica y Topográfica

Modelización del contenido de Humedad de Combustible
Vivo en la Comunitat Valenciana usando índices
espectrales extraídos de Sentinel-2 durante los meses de
verano y otoño desde 2019 hasta 2021.

Trabajo Fin de Máster

Máster Universitario en Ingeniería Geomática y Geoinformación

AUTOR/A: Pachacama , Kenneth

Tutor/a: Balaguer Beser, Ángel Antonio

CURSO ACADÉMICO: 2021/2022



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIERÍA
GEODÉSICA, CARTOGÁFICA
TOPOGRÁFICA

MÁSTER UNIVERSITARIO EN INGENIERÍA GEOMÁTICA Y GEOINFORMACIÓN

TRABAJO DE FIN DE MÁSTER

**Modelización del contenido de Humedad de Combustible
Vivo en la Comunitat Valenciana usando índices
espectrales extraídos de Sentinel-2 durante los meses de
verano y otoño desde 2019 hasta 2021.**

Autor: Kenneth Ricardo Pachacama Vallejo

Tutor: Ángel Antonio Balaguer Beser

Centro: E.T.S.I. Geodésica, Cartográfica y Topográfica
Universitat Politècnica de Valencia

Septiembre, 2022

AGRADECIMIENTOS

A mi tutor Ángel Antonio Balaguer Beser, por estar en cada paso demostrando voluntad e interés, el haberme brindado sus conocimientos, consejos y ayuda constante permitiéndome salir adelante con la realización de mis estudios, de igual modo ha demostrado paciencia y disposición a solucionar problemas académicos y también personales creando un buen ambiente de trabajo con confianza y honestidad hasta poder obtener mi meta final, un gracias infinito en esta etapa de mi vida.

A María Alicia Arcos y Roberto Edo Botella miembros del equipo de trabajo, quienes estuvieron siempre dispuestos a compartir conocimiento e información, y con su calidad humana me hicieron sentir a gusto en cada reunión de trabajo.

Expreso mi profunda gratitud a la U.P.V., E.T.S.I.G.C.T. y sus profesores por brindarme conocimientos de actualidad científica y tecnológica, siempre demostrando preocupación por el aprendizaje con carácter ético y moral, que me servirán para el desarrollo tanto profesional como humano.

Un agradecimiento especial a mis dos amigos incondicionales Juan Casado y Davide Blanchetti, que fueron un gran apoyo emocional en los momentos más duros de esta etapa y quienes me hicieron sentir como en casa desde el primer día y nunca me dejaron solo. Siempre nos apoyamos y nunca faltaron las risas, gracias por hacer de esta experiencia una de las mejores de mi vida.

Al equipo técnico de la Empresa VAERSA por facilitar la información requerida para este trabajo, proporcionada al tutor de este proyecto a través de la Direcció General de Preveñció d'Incendis Forestals de la Generalitat Valenciana.

DEDICATORIA

Este logro se lo dedico a mi padre Richard Pachacama y mi madre Rosa Vallejo excelentes profesionales, el mejor regalo que Dios me puedo regalar. En todo momento estuvieron presentes todas sus enseñanzas y valores, cada mensaje de apoyo me ayudo a seguir adelante sabiendo que tengo una hermosa familia que es mi más grande soporte. A mis hermanos Vlady y Kevin a quienes siempre pude recurrir a contarles mis problemas y siempre encontraba un refugio.

A toda mi familia, por estar apoyándome siempre e incondicionalmente.

COMPROMISO

El presente documento ha sido realizado completamente por el firmante, no ha sido entregado como otro trabajo académico previo y todo el material tomado de otras fuentes ha sido convenientemente entrecomillado y citado su origen en el texto, así como referenciado en la bibliografía.

Los datos de humedad de los combustibles vivos forestales utilizados en este trabajo han sido proporcionados por la empresa VAERSA con la autorización de la Generalitat Valenciana, gracias a la colaboración mantenida para este fin entre los tutores de este trabajo y la Dirección General de Prevención de Incendios Forestales de la Generalitat Valenciana.

RESUMEN

Se entiende por Humedad de combustible vivo (HCV), al contenido de agua que existe en cualquier tipo de vegetación viva y aunque puede ser muy variable dependiendo del tipo de vegetación y factores ambientales, es uno de los parámetros más importantes al momento de determinar la cantidad de combustible que se encuentra disponible a quemar en un área determinada, convirtiéndose así en un factor decisivo para entender, cuanto combustible puede ser consumido, que tan rápido puede consumirse y las etapas que existen durante un incendio forestal. Se puede monitorizar este parámetro con trabajo de campo, donde por métodos de laboratorio se obtiene la HCV a partir de muestras tomadas en algunas parcelas, aunque esta metodología requiere mucho tiempo y recursos tanto humanos como económicos. En el presente trabajo se obtienen modelos de estimación del contenido de HCV en la Comunitat Valenciana usando datos obtenidos en campo durante los meses de verano y otoño, desde 2019 hasta 2021, los cuales han sido obtenidos por la empresa pública VAERSA y proporcionados al tutor de este trabajo por la Dirección General de Prevención de Incendios Forestales de la Generalitat Valenciana. Como variables predictoras se consideran distintos índices espectrales obtenidos de imágenes satelitales Sentinel-2, que ayudan a describir el comportamiento de la vegetación en una fecha dada. A este conjunto de datos se pueden sumar otros parámetros meteorológicos provenientes de estaciones cercanas. Todos estos datos han sido analizados y depurados para obtener una serie temporal que sea coherente con la realidad del terreno y así poder obtener la variable dependiente HCV mediante modelos estadísticos de análisis multivariante.

Los modelos de regresión lineal multivariante fueron evaluados a partir de varios estadísticos como: R-cuadrado ajustado, Cp de Mallows y varios criterios de información. También se ha analizado la multicolinealidad de las variables usadas a través del factor de inflación de la varianza (V.I.F por sus siglas en inglés). En la provincia de Valencia gracias a la disponibilidad de datos, se realizó la validación de los modelos con parcelas que no fueron usadas para el entrenamiento, mientras que para las provincias de Alicante y Castellón se usó el método de validación cruzada para medir la capacidad predictiva del modelo.

Palabras claves: Humedad de combustible vivo; modelos estadísticos de análisis multivariante; índices espectrales; parámetros meteorológicos; multicolinealidad.

ABSTRACT

Live fuel moisture content (LFMC) is understood as the water content that exists in any type of live vegetation and although this parameter can be highly variable depending on the type of vegetation and environmental factors, it is one of the most important when determining the amount of fuel that is available to burn in a given area, thus becoming a decisive factor in understanding how much fuel can be consumed, how quickly it can be consumed and the stages that exist during a forest fire. This parameter can be monitored with field work, where LFMC is obtained by laboratory methods from samples taken some plots, although this methodology requires a lot of time and both human and financial resources. In the present work, LFMC estimation models are obtained in the Valencian Community using data obtained in the field during the summer and autumn months, from 2019 to 2021, which have been obtained by the public company VAERSA and provided to the tutor of this work by the Dirección General de Prevención de Incendios Forestales of the Generalitat Valenciana. Different spectral indices obtained from Sentinel-2 satellite images are considered as predictor variables, which help to describe the behavior of the vegetation on a given date. Other meteorological parameters from nearby stations have also been added to this dataset. All these data have been analyzed and refined to obtain a time series that is consistent with the reality of the terrain and thus be able to obtain the dependent variable HCV through statistical models of multivariate analysis.

The multivariate linear regression models were evaluated based on various statistics such as: adjusted R-squared, Mallows' Cp and various information criteria. The multicollinearity of the variables used was also analyzed through the variance inflation factor (V.I.F). In the province of Valencia, thanks to the availability of data, the validation of the models was carried out with plots that were not used for training, while for the provinces of Alicante and Castellón the cross-validation method was used to measure the predictive capacity of the model.

Keywords: Live fuel moisture content; statistical models of multivariate analysis; spectral indices; meteorological parameters; multicollinearity.

ÍNDICE

RESUMEN	4
ABSTRACT	5
1. INTRODUCCIÓN Y OBJETIVOS	12
1.1 Introducción.....	12
1.2 Objetivos.....	15
1.2.1 <i>Objetivo General</i>	15
1.2.2 <i>Objetivos Específicos</i>	15
2. DATOS	16
2.1 Ubicación de la zona de estudio y puntos de muestreo	16
2.2 Humedad de combustible vivo	23
2.3 Datos meteorológicos	24
2.4 Datos de índices espectrales	27
3. METODOLOGÍA	31
3.1 Estructura de procesos	31
3.2 Obtención de índices espectrales a partir de imágenes satelitales Sentinel-2 32	
3.2.1 <i>Tratamiento de imágenes satelitales para la descarga de índices espectrales usando Google Earth Engine (GEE)</i>	32
3.2.1.1 Insumos de entrada	32
3.2.1.2 Máscara de nubes.....	33
3.2.1.3 Generación de índices espectrales	34
3.2.1.4 Aplicación de ReduceRegions	35
3.2.2 <i>Tratamiento de base de datos usando Python</i>	36
3.2.2.1 Criterio NDVI y coeficiente de variación (CV).....	36
3.2.2.2 Criterio de la mediana.....	38
3.2.2.3 Aplicación de filtro Savitsky-Golay	39
3.2.3 <i>Obtención de estadísticos de índices espectrales</i>	41
3.3 Métodos de regresión	42
3.3.1 <i>Análisis de regresión lineal múltiple</i>	42
3.3.2 <i>Variables utilizadas en el análisis de regresión lineal</i>	44

3.4	Desarrollo del análisis de regresión lineal múltiple.	47
3.4.1	<i>Regresión lineal paso a paso hacia delante</i>	48
3.4.2	<i>Selección de modelos de regresión (R-cuadrado ajustado, Cp de Mallows y criterios de información)</i>	51
3.4.2.1	R-cuadrado ajustado	52
3.4.2.2	Cp de Mallows.....	53
3.4.2.3	Mejor criterio de información.....	55
3.4.3	<i>Modelos lineales generales</i>	57
3.4.3.1	Uso de variable categórica en los modelos lineales generales.....	59
4.	RESULTADOS	61
4.1	Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la provincia de Valencia	61
4.2	Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la Provincia de Alicante.	72
4.3	Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la Provincia de Castellón.....	78
4.4	Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Valencia.....	83
4.5	Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Alicante.	89
4.6	Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Castellón.....	92
5.	PRESUPUESTO	97
6.	CONCLUSIONES	99
7.	BIBLIOGRAFÍA	102
8.	CARTOGRAFÍA	107
9.	ANEXOS	111
9.1	Anexo 1: Código para la creación de índices espectrales en GEE	111
9.2	Anexo 2: Código criterio de la mediana en Python.....	114
9.3	Anexo 3: código del filtro <i>Savitzky-Golay</i> en Python	116
9.4	Anexo4: código para unir datos de campo con índices espectrales por fechas más cercanas	117

ÍNDICE DE FIGURAS

Figura 1. Ecosistemas de la Comunitat Valenciana. Fuente: Cartografía del ICV	17
Figura 2. Distribución de los puntos de muestreo de HCV en la Comunitat Valenciana.	19
Figura 3. Modelos de combustible de la Comunitat Valenciana. Feunte: (Conselleria de Agricultura, 2019)	20
Figura 4. Series temporales de HCV en la Parcela 1 para las especies presentes.....	24
Figura 5. Comportamiento de la variable sin_DOY a lo largo del año.	26
Figura 6. Estructuras de procesos de la metodología.....	31
Figura 7. Buffers creados alrededor de los puntos de muestreo.	32
Figura 8. Aplicación de la máscara de nubes en las imágenes satelitales.....	33
Figura 9. Generación de índices espectrales.	34
Figura 10. Aplicación de “ReduceRegions” en una imagen. Fuente: (GEE, Statistics of an Image Region, 2022).....	35
Figura 11. Exportación de índices espectrales a formato CVS.....	35
Figura 12. Diagrama del flujo seguido para el tratamiento de índices espectrales.	36
Figura 13. Anomalías presentes en el índice espectral NDVI de la parcela 62.	37
Figura 14. Índice espectral después de aplicar el criterio de NDVI y CV	38
Figura 15. Índice espectral después de aplicar el criterio de la mediana	39
Figura 16. Índice espectral después de aplicar el filtro Savitsky-Golay	40
Figura 17. Serie temporal original o descargada (superior) y la serie temporal corregida (inferior).....	41
Figura 18. Variabilidad de fechas entre el dato de campo y el índice espectral.	41
Figura 19. Estadísticos del índice espectral NDVI.	42
Figura 20. Configuración de umbral del valor-P para la selección paso a paso hacia delante.....	43
Figura 21. Mapa de la disponibilidad de los datos de campo.	47
Figura 22. Gráfico número de filas versus residuos.....	50
Figura 23. Estimación de la HCV en la Parcela 2 Año 2020.....	50
Figura 24. Gráfica de R-cuadrado ajustado para la humedad de combustible vivo versus el número de variables aplicadas	53
Figura 25. Gráfica de distribución de Cp frente al número de coeficientes.....	55
Figura 26. Gráfico de HCV predicho versus el residuo estudentizado del modelo	59
Figura 27. Gráfico de observado vs predicho, Valencia 2019. <i>Media</i> ponderada (izquierda), media matorral (derecha)	62
Figura 28. Gráfico de observado vs predicho, Valencia 2020. <i>Media</i> ponderada (izquierda), media matorral (derecha)	63
Figura 29. Gráfico de observado vs predicho, Valencia 2021. <i>Media</i> ponderada	

(izquierda), media matorral (derecha)	64
Figura 30. Gráfico de observado vs predicho, Valencia 2019-2020-2021. Media ponderada (izquierda), media matorral (derecha).....	67
Figura 31. HCV de <i>Rosmarinus officinalis</i> (rojo) y la humedad de combustible vivo de la especie <i>Pinus Pinaster</i> (azul).....	69
Figura 32. Modelos de evaluación para cada municipio de la provincia de Valencia. HCV muestreada en campo (azul), Modelo media ponderada (rojo), Modelo media ponderada categórica (rosa).....	70
Figura 33. Modelos de evaluación para <i>Rosmarinus officinalis</i> en la provincia de Valencia.	72
Figura 34. Gráfico de observado vs predicho, Alicante 2020-2021. Media ponderada (izquierda), media matorral (derecha), <i>Rosmarinus officinalis</i> (abajo).....	74
Figura 35. Modelos de validación para la provincia de Alicante.....	77
Figura 36. Gráfico de observado vs predicho, Castellón 2020-2021. Media ponderada (izquierda), media matorral (derecha), <i>Rosmarinus officinalis</i> (abajo).....	80
Figura 37. Modelos de validación para la provincia de Castellón.	83
Figura 38. Modelos para la validación de la parcela 11. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2 TU3 (rosa).....	87
Figura 39. Modelos para la validación de la parcela 14. HCV de campo (azul). Modelos de validación media ponderada TU3 (rojo). Modelo de validación media ponderada TU2 TU3 (rosa).....	88
Figura 40. Modelos de validación de la parcela 82. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2 TU3 (rosa).....	92
Figura 41. Modelo de validación para la parcela 51. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2 TU3 (rosa).....	95
Figura 42. Modelo de validación para la parcela 55. HCV de campo (azul). Modelos de validación media ponderada TU3 (rojo). Modelo de validación media ponderada TU2 TU3 (rosa).....	96

ÍNDICE DE TABLAS

Tabla 1. Tipologías del suelo en la Comunitat Valenciana. Fuente: (PATFOR, s.f). ...	16
Tabla 2. Puntos de muestreo: Coordenadas, Municipio y modelo de combustible.....	21
Tabla 3. Variables meteorológicas.	25
Tabla 4. Características de las imágenes Sentinel-2. Fuente: (GEE, 2022).	27
Tabla 5. Índices espectrales usados para el presente estudio.	28
Tabla 6. Clasificación de parcelas de estudios divididas por provincias y modelo de combustible.....	44
Tabla 7. Disponibilidad de datos en cada municipio de la Comunitat Valenciana	45
Tabla 8. Regresión múltiple paso a paso hacia adelante en la provincia de Valencia en el año 2020	49
Tabla 9. Coeficiente de correlación entre las variables independientes.	51
Tabla 10. Selección de modelos usando 8 variables.	52
Tabla 11. Modelos con su valor de Cp de Mallows	54
Tabla 12. Tabla de mejor criterio de información.	56
Tabla 13. Significancia estadística de cada factor conforme fue introducido al modelo	57
Tabla 14. Variables y sus factores de inflación de varianza (V.I.F).....	58
Tabla 15. Residuos atípicos de todas las observaciones que tienen un residuo estudentizado mayor que 2	58
Tabla 16. Aplicación de la variable categórica en el resultado del modelo lineal general	59
Tabla 17. Parcelas de matorral (SH4) de la provincia de Valencia.....	61
Tabla 18. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2019.	62
Tabla 19. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2020.	63
Tabla 20. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2021.	64
Tabla 21. Modelos resultantes para las fechas 2019, 2020 y 2021 en conjunto.....	66
Tabla 22. Parcelas de validación para la provincia de Valencia.	67
Tabla 23. Modelo de Rosmarinus officinalis para la provincia de Valencia.....	71
Tabla 24. Parcelas de evaluación para Rosmarinus officinalis en la provincia de Valencia.....	71
Tabla 25. Parcelas de matorral (SH4) de la Provincia de Alicante.	72
Tabla 26. Modelos de matorral (SH4) de la Provincia de Alicante.....	73
Tabla 27. modelo de regresión lineal usando solamente los datos del verano de 2021.	78
Tabla 28. Parcelas usadas para los modelos en la provincia de Castellón.	78
Tabla 29. Modelos de regresión para la media ponderada y media matorral en la	

provincia de Castellón	79
Tabla 30. Modelos de regresión lineal tomando únicamente los datos de junio a octubre del 2021.	83
Tabla 31. Parcelas de arbolado (TU2 y TU3) de la Provincia de Valencia.....	84
Tabla 32. Modelos de HCV para las categorías de TU2, TU3 y TU2 TU3	85
Tabla 33. Parcelas de arbolado (TU2 y TU3) de la Provincia de Alicante.	89
Tabla 34. Modelos de arbolado (TU2 y TU3) de la Provincia de Alicante.....	89
Tabla 35. Parcelas de arbolado (TU2 y TU3) de la Provincia de Castellón.....	92
Tabla 36. Modelos creados de la provincia de Castellón	93
Tabla 37. Presupuesto TFM.....	98

1. INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

Décadas de análisis de datos e informes han dejado claro que hay menos incendios, pero son más destructivos y algunos son muy difíciles de extinguir por los equipos de emergencia. Estos fenómenos pueden arrasarse miles de hectáreas de terreno forestal suponiendo una amenaza para el medio ambiente y para las personas (WWF, 2020). Debido a los perseverantes y bruscos cambios en el equilibrio de los ecosistemas, producto del cambio climático, el fenómeno ha llegado incluso hasta el Ártico, el cual se calienta dos veces más rápido que el resto del planeta y como consecuencia, se están generando incendios de altísima intensidad que liberan enormes cantidades de dióxido de carbono a la atmósfera (Borunda, 2021).

Existe una relación directa entre incendios, deforestación y pandemias. La destrucción de ecosistemas forestales vírgenes especialmente los tropicales, hace que el ser humano entre en contacto con fauna silvestre portadora de patógenos. En este sentido la Organización de las Naciones Unidas, Fondo mundial para la naturaleza (WWF por sus siglas en inglés) y *One Health*, coinciden en que “Unos bosques bien conservados se convierten en la mejor vacuna contra la zoonosis como el coronavirus” (WWF, 2020).

España no ha estado exenta y como les sucede a todos los países de la región mediterránea, es especialmente vulnerable, debido a los escenarios de sequía y desertificación. Los grandes incendios forestales en España apenas suponen el 0.18% del total, pero en estos arde el 40% de la superficie afectada. Aunque el número total de incendios forestales y superficie quemada ha disminuido, se producen de media 11700 siniestros al año, una cifra insostenible e inaceptable (MITECO, 2020). Durante la elaboración de este estudio, el país ha vivido varios escenarios de incendios forestales calificados por las autoridades como “descomunales”, en los municipios de Bejís en la provincia de Castellón y Vall d’Ebo en la provincia de Alicante, en los cuales las llamas han calcinado más de 30000 hectáreas de terreno forestal de la Comunitat Valenciana. Aunque, gracias al esfuerzo realizado por todos los efectivos que han intervenido en la extinción (bomberos, forestales, UME, Guardia Civil, policía de la Generalitat Valenciana y alcaldes), no se han tenido que lamentar pérdidas humanas, la afectación social es innegable, pues las autoridades han recurrido a confinamientos parciales o desalojos de poblaciones con el fin de garantizar su seguridad (SER, 2022).

Según la asociación WWF España y (Ferreiro Gómez, 2016), en España se destinan 1.000 millones de euros al año para aplicar medidas de extinción y, sin embargo, tan sólo 300 millones de euros a la prevención. Esta dinámica económica llega a ser muy descompensada, ya que, el actual escenario de bosques altamente inflamables, pocas lluvias, cambio climático y desertificación augura incendios cada vez más peligrosos. Las acciones de extinción son necesarias y beneficiosas, sin embargo, no es suficiente sin una

política de reordenación y planificación del territorio, el dinero invertido en extinción nunca será suficiente (Hernández & et al, 2020).

Solo reduciendo la vulnerabilidad del paisaje a la propagación de las llamas se evitarán los grandes incendios forestales. Se necesita un plan estructural de planificación del territorio. El modelo de los años 90, basado en medidas de extinción, ya no sirve. Su estrategia basada en aumentar los medios de extinción no solucionará el problema. Los grandes incendios no se apagan con agua, sino con gestión forestal y planificación territorial (EPRIF, s.f).

A este respecto, la Generalitat Valenciana cuenta con un Sistema Integrado de Gestión de Incendios Forestales (SIGIF, <https://agroambient.gva.es/es/web/prevencion-de-incendios/sistema-integrado-de-gestion-de-incendios-forestales-sigif>), cuyo origen comenzó en el año 2002 y reúne en un solo sitio la información disponible relacionada con la prevención de incendios forestales. El SIGIF forma parte del Plan de Prevención de Incendios Forestales de la Comunitat Valenciana como un elemento más de unión y recogida de información, permitiendo conocer las previsiones de riesgo para las distintas zonas del territorio de la Comunitat Valenciana.

Un incendio forestal necesita de tres componentes básicos: oxígeno, combustible y calor que produzca la ignición, para que se inicie y esparza a sus alrededores. Entender estas componentes es crucial para la planificación, extinción y prevención de incendios. Las características únicas de la biomasa de un ecosistema juegan el papel de combustible y determinarán que la ignición sea más o menos probable, y que el fuego tenga continuidad, o no (Mataix-Solera & Cerdà, 2009).

El contenido de humedad de combustible vivo (HCV), se calcula como el porcentaje de agua que contiene una especie vegetal en relación con su masa seca total y juega un papel fundamental en la dinámica que puede tener un incendio forestal tanto en sus inicios como en su desarrollo. Está directamente relacionada con la cantidad de energía necesaria para evaporar el agua antes de la ignición. Por tanto, un alto porcentaje de humedad contenida en la vegetación reduce o impide totalmente la inflamabilidad y consecuente propagación del fuego (Cunill Camprubi, González-Moreno, & Resco de Dios, 2022).

El cálculo de la HCV de las especies vegetales en un determinado ecosistema se suele realizar mediante el trabajo de campo, donde técnicos calificados están encargados de planificar, recoger y analizar muestras de campo. Este método directo conlleva un alto coste económico y de esfuerzo humano que puede tardar tiempo en rendir resultados (Yebra, y otros, 2013). En la actualidad, como una potente alternativa, son varios los autores que han analizado el uso de sensores remotos para la estimación indirecta de la HCV a través de la creación de modelos estadísticos o físicos, los cuales se basan en analizar secciones del espectro electromagnético, normalmente espectro visible e infrarrojos cercano y lejano, capturados por sensores multi-espectrales o hiper-espectrales

montados sobre satélites, que logran capturar amplias imágenes de la superficie terrestre en un corto tiempo. A través de modelos estadísticos, varios estudios han estimado la HCV principalmente en Australia (Yebra, y otros, 2018), ecosistemas de vegetación mediterránea mixta (Costa-Saura, Balaguer-Beser, Ruiz, Pardo-Pascual, & Soriano-Sancho, 2021) o en el suroeste de Estados Unidos (Myoung, y otros, 2018) (Peterson, Roberts, & Dennison, 2008).

Los índices espectrales y de vegetación son el producto del análisis de imágenes satelitales y han servido como variables predictoras para la estimación de la HCV en varios ecosistemas de España (Chuvieco, y otros, 2004). Estos índices caracterizan a la vegetación usando principalmente la sección del espectro infrarrojo entre otros, que tienen una alta respuesta ante la dinámica de la clorofila en hojas, ramas y tallos de una planta. También han sido usados para otros estudios como; la agricultura de precisión (Haboudane, Miller, Tremblay, Zarco-Tejada, & Dextraze, 2002) o análisis hidrológicos (Moreno Alcaraz, Llacer, & Heras Ibáñez, 2006). Con la ayuda de estos índices se han logrado crear modelos basados en datos de campo para predecir la HCV en lugares como al oeste de Italia (Pellizzaro, Cesaraccio, Duce, Ventura, & Zara, 2007), o al sur de Francia (Martin-StPaul, y otros, 2018). El NDVI es uno de los índices más usados para estimar la HCV, otros usados son: el índice de agua de diferencia normalizada (NDWI), el índice de vegetación verde (Vgreen), el índice de resistencia atmosférica visible (VARI) y el índice de vegetación mejorado (EVI) (Peterson, Roberts, & Dennison, 2008).

El presente estudio consiste en modelizar el contenido de Humedad de Combustible Vivo en la Comunitat Valenciana usando como variables independientes índices espectrales extraídos de Sentinel-2 y datos meteorológicos recopilados por la Agencia estatal de meteorología de España (AEMET), durante los meses de verano y otoño desde 2019 hasta 2021, mientras que para la variable dependiente se usan los datos de HCV obtenidos por la empresa Pública VAERSA, para la Dirección General de Prevención de Incendios Forestales de la Generalitat Valenciana, en diferentes especies muestreadas en 70 parcelas distribuidas por las tres provincias de la Comunitat Valenciana. También se han tomado en cuenta los modelos de combustible forestal de la Comunitat Valenciana para categorizar los datos y por consiguiente obtener modelos más restringidos.

1.2 Objetivos

1.2.1 Objetivo General

Generar modelos de Humedad de Combustible Vivo (HCV) de la Comunitat Valenciana para los meses comprendidos desde junio hasta octubre de los años 2019, 2020 y 2021, a través de índices espectrales extraídos de imágenes satelitales Sentinel-2 y otras variables meteorológicas.

1.2.2 Objetivos Específicos

- Procesar las imágenes satelitales Sentinel-2 a través de *Google Earth Engine (GEE)* para la creación de una base de datos de índices espectrales para el periodo 2019-2021.
- Depurar la base de datos usando programación *Python* para la reducción y eliminación de ruido proveniente de la presencia de nubes en las imágenes Sentinel-2 y posterior unificación con los datos de HCV provenientes de campo.
- Construir modelos de regresión lineal multivariante de HCV en la Comunitat Valenciana para los meses desde junio hasta octubre de los años 2019, 2020 y 2021, usando como variables explicativas los índices espectrales extraídos de Sentinel-2 junto con datos meteorológicos proporcionados por la Agencia Estatal de Meteorología (AEMET). Para la muestra de entrenamiento se dispone de la información de HCV obtenida en campo por la empresa Pública VAERSA con periodicidad quincenal, para dicho periodo de tiempo y en diferentes puntos de muestreo, la cual ha sido proporcionada al tutor de este proyecto a través de la Direcció General de Prevenció d'Incendis Forestals de la Generalitat Valenciana.
- Examinar la capacidad predictiva de los modelos de regresión, mediante el uso de parcelas de validación y validación cruzada, para una evaluación de los modelos.

2. DATOS

2.1 Ubicación de la zona de estudio y puntos de muestreo

La zona de estudio se encuentra ubicada en la Comunitat Valenciana situada al sur-este de la Península Ibérica, en el litoral del mar Mediterráneo, según el Instituto Cartográfico Valenciano cubre un área de 23266 km² siendo la octava región de España y representa el 4.60% de la extensión nacional. Su territorio corresponde a una variedad de usos de suelo, en varias ocasiones diferenciar de forma estricta entre un uso agrícola y un uso forestal puede ser complicado. Según el Plan Nacional Territorial Forestal (PATFOR), para determinar que una extensión de suelo es forestal, se propone como criterio inequívoco que la fracción de cabida cubierta de especies forestales arbóreas sea superior al 30% medido como máximo a nivel de subparcela catastral. El suelo forestal de la Comunitat Valenciana está establecido en tres tipologías: Suelo Forestal, Suelo Mosaico y suelo No Forestal (PATFOR, s.f), distribuidos como se ve en la Tabla 1. Se determina como suelo forestal aquellas en las que más del 90 % de la superficie tiene uso forestal. Se ha considerado mosaico aquellos terrenos en los que se dan varios usos y el forestal ocupa entre el 10% al 90% de la parcela, y se ha identificado como no forestal aquellas parcelas en las que el uso forestal representa menos del 10% de la superficie (PATFOR, s.f).

Tabla 1. *Tipologías del suelo en la Comunitat Valenciana. Fuente: (PATFOR, s.f).*

Provincia	Forestal		Mosaico		No forestal	
	Ha	%	Ha	%	Ha	%
Castellón	435,690	65.65	14,663	2.21	213,272	32.12
Valencia	604,544	55.94	19,183	1.78	456,989	42.29
Alicante	256,339	44.06	13,545	2.33	311,965	53.62
Comunitat Valenciana	1,296,573	55.74	47,397	2.04	982,226	42.22

Como se muestra en la Figura 1, la Comunitat Valenciana cuenta con 23 ecosistemas forestales, 10 catalogados como arbolados, 12 como no arbolados y uno determinado como mixto. Se puede apreciar que los ecosistemas que más destacan son: *Pinus halepensis* proveniente del ecosistema arbolado y Garriga que proviene de la especie *Quercus Coccifera* del ecosistema arbustivo. Además de estas especies, para el estrato arbolado las especies más representativas son *Quercus ilex*, *Pinus nigra* y *Pinus pinaster*, mientras que en el estrato de arbustivo las especies más representativas son: *Rosmarinus officinalis*, *Ulex parviflorus*, *Pistacea lentiscus* y *Juniperus oxycedrus*.

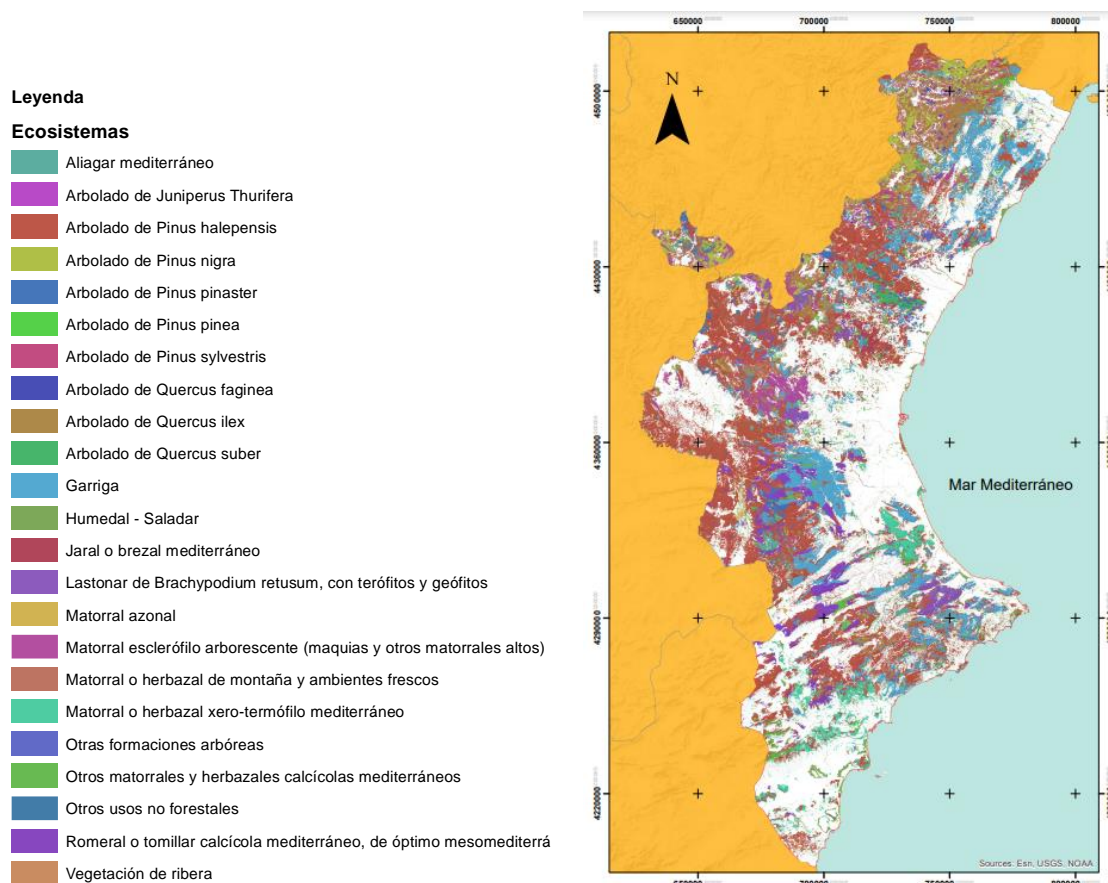


Figura 1. Ecosistemas de la Comunitat Valenciana. Fuente: Cartografía del ICV

Para este estudio los puntos de muestreo fueron seleccionados por la empresa Valenciana de Aprovechamiento Energético de Residuos, SA (VAERSA) perteneciente a la Conselleria de Agricultura, Medio Ambiente, Cambio Climático y Desarrollo Rural de la Generalitat Valenciana. Los datos campo de HCV fueron cedidos al tutor de este trabajo fin de máster por la Direcció General de Prevenció d'incendis Forestals de la Generalitat Valenciana para la creación de modelos de estimación de HCV en la Comunitat Valenciana. En el año 2019 se tomaron datos en 42 parcelas de la provincia de Valencia desde el mes de junio hasta octubre, con una periodicidad quincenal. Se elaboraron algunos modelos para ese periodo de tiempo gracias al contrato CNME19/0304/42, suscrito entre la UPV y la Direcció General de Prevenció d'incendis Forestals de la Generalitat Valenciana, los cuales se describen en (Costa-Saura, Balaguer-Beser, Ruiz, Pardo-Pascual, & Soriano-Sancho, 2021). A partir de julio de 2020 se continuó con la toma de datos en dichas parcelas, ampliando la toma de datos a otros puntos de muestreo en las provincias de Castellón y Alicante, contando con un convenio de colaboración entre la UPV y la empresa Red Eléctrica de España S.A.U para el periodo 2020-2023, para desarrollar el proyecto que lleva por título: Modelización y cartografiado de la humedad del combustible vivo en la Comunitat Valenciana, bajo la dirección del doctor Ángel Antonio Balaguer Beser (tutor de este trabajo fin de máster). En total se han tomado datos en 88 puntos de muestreo de la Comunitat Valenciana para el periodo junio de 2020-

noviembre de 2021, de los cuales se han seleccionado los datos de 70 puntos de muestreo para la elaboración de modelos de HCV en este trabajo fin de máster. Dicha selección se corresponde con las parcelas pertenecientes a los modelos de combustible más representativos en la Comunitat Valenciana. Adicionalmente, a la toma de datos en dichas parcelas también se cuenta con datos obtenidos con periodicidad semanal entre enero y octubre del año 2021 en tres parcelas de la provincia de Valencia, los cuales serán usados para validar los modelos.

Este trabajo de campo es una extensión de estudios previos realizados por la empresa VAERSA, algunos de los cuales se describen en el trabajo final de máster denominado “Análisis de la Humedad del Combustible Vivo (HCV) en la Comunitat Valenciana y la influencia de las variables meteorológicas” (Soriano Sancho, 2016). El muestreo cumple con condiciones muy específicas que son:

- Áreas de muestreo que contengan variedad de especies vegetales endémicas de la Comunitat Valenciana en lo que respecta a matorrales, herbazales y arbolado.
- El grupo de puntos de muestreo deben representar a toda la zona de análisis, es decir a la cobertura vegetal existente en la Comunitat Valenciana.
- Se consideraron aspectos climatológicos, topográficos, accesibilidad, entre otros.

Con estos criterios se han tomado 70 puntos de muestreo distribuidos en varios municipios de toda la Comunitat Valenciana como se puede ver en la Figura 2.

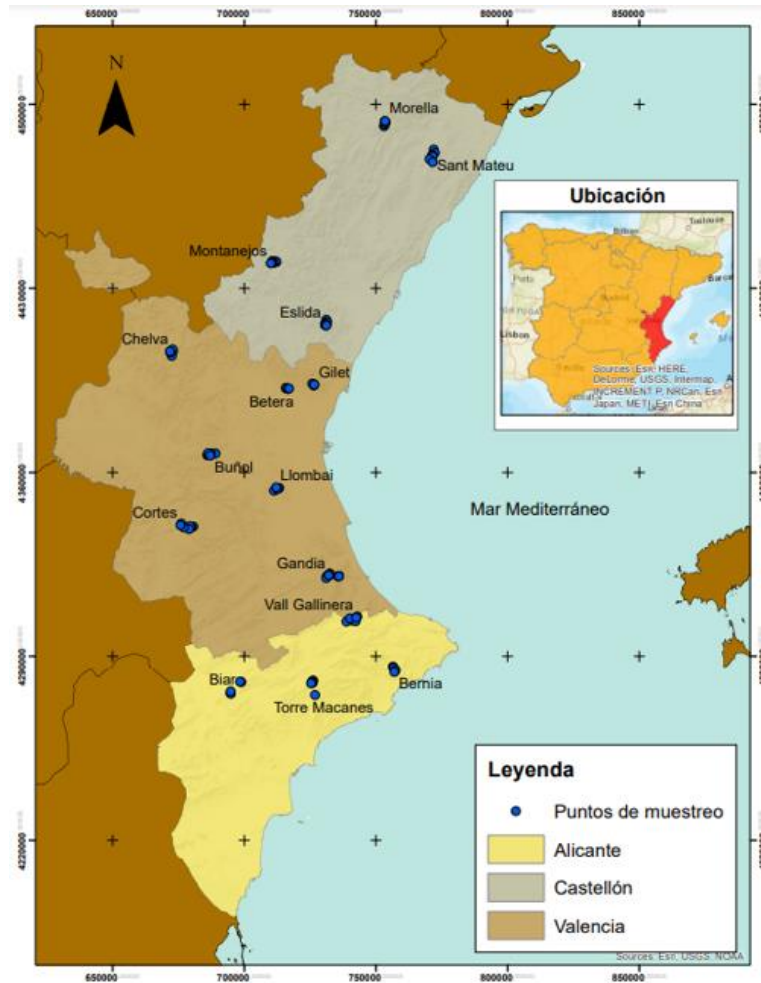


Figura 2. Distribución de los puntos de muestreo de HCV en la Comunitat Valenciana.

La Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica de la Generalitat Valenciana tiene entre sus competencias la gestión del territorio y prevención de amenazas entre los cuales se encuentran los incendios forestales. Para esto, ha desarrollado los modelos de combustible forestal que son estructuras de la vegetación tipo a las que les presuponen un comportamiento frente al fuego común. Estas estructuras permiten predecir el comportamiento del fuego, definiendo parámetros como la velocidad de propagación o la longitud de la llama (Conselleria de Agricultura, 2019). En la Figura 3 se puede apreciar la categorización de los modelos de combustible en la Comunitat de Valencia.

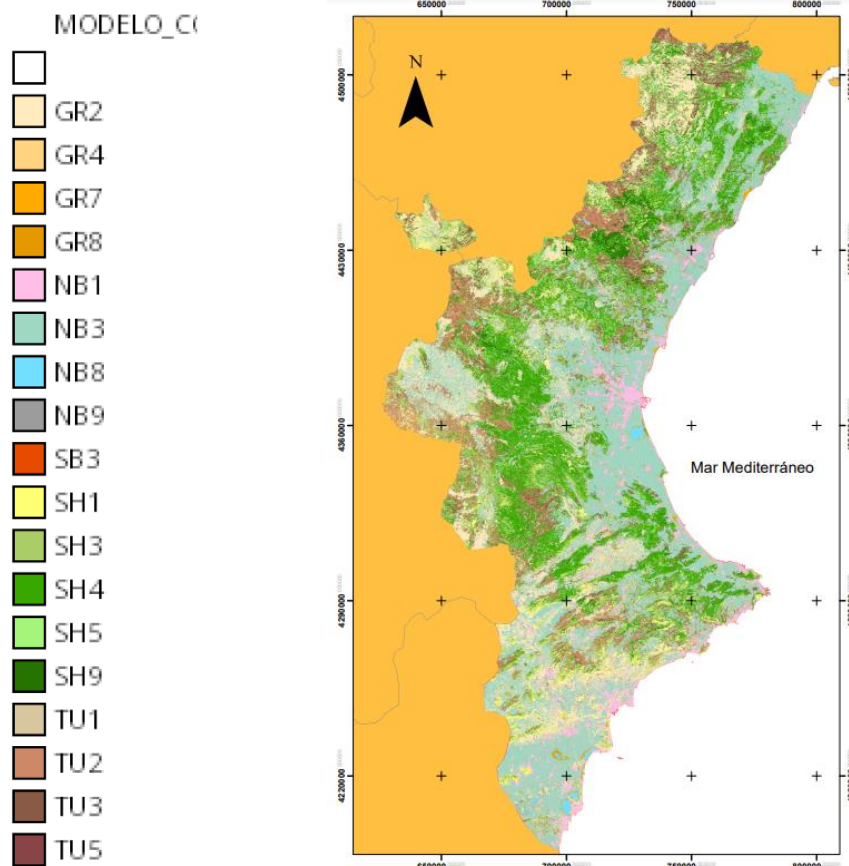


Figura 3. Modelos de combustible de la Comunitat Valenciana. Fuente: (Conselleria de Agricultura, 2019)

Según esta categorización, los modelos de combustible que más presencia tienen sobre los puntos de muestreo y sus zonas aledañas son:

- SH4: caracterizados por arbustos leñosos y hojarasca de arbusto y de elevada continuidad. Las especies características de este modelo serían el romero, la aliaga, la coscoja con pequeño porte.
- TU2: Matorral de menos de 1 m de altura, bajo dosel de arbolado adulto tanto de pináceas como de quercíneas.
- TU3: Matorral de más de 1 m de altura, bajo dosel de arbolado adulto tanto de pináceas como de quercíneas.

Ya que en este estudio se analiza la humedad de combustible vivo, factor directamente relacionado con fenómenos naturales como los incendios forestales, se han categorizado los puntos de muestreo y su modelación de HCV de acuerdo a los modelos de combustible forestal. La tabla 2 muestra las coordenadas UTM, el municipio y el modelo de combustible de todos los puntos de muestreo usados en este trabajo. La numeración dada en la primera columna corresponde a la de la muestra de 88 parcelas facilitada por el tutor de este proyecto, de las cual se han seleccionado los 70 puntos de muestreo pertenecientes

a las categorías SH4, TU2 y TU3, de modelos de combustible, usados para calcular los modelos en este trabajo. En adición a estas parcelas se contó con las parcelas de evaluación 91, 92 y 93 pertenecientes a la provincia de Valencia y catalogadas según el modelo de combustible como SH4.

Tabla 2. *Puntos de muestreo: Coordenadas, Municipio y modelo de combustible.*

Nº	Coordenadas UTM			Municipio	Mod_comb
	X	Y	Z		
1	726493	4392961	267.6	Gilet	TU3
2	726474	4393223	257.0	Gilet	SH4
3	726614	4393412	309.9	Gilet	SH4
4	725927	4393646	299.6	Gilet	SH4
5	726679	4393308	297.5	Gilet	SH4
6	715917	4392096	213.0	Bétera	SH4
7	715970	4391931	202.1	Bétera	SH4
8	716676	4391632	174.8	Bétera	TU2
11	672897	4406725	975.9	Chelva	TU2
12	672552	4404256	751.3	Chelva	SH4
13	672054	4405490	950.8	Chelva	SH4
14	671858	4406029	929.5	Chelva	TU3
15	713355	4354041	319.5	Llombai	SH4
16	712535	4353729	292.9	Llombai	SH4
17	711390	4353003	231.9	Llombai	SH4
18	712616	4353497	291.0	Llombai	SH4
19	712168	4354242	266.0	Llombai	SH4
20	689150	4367100	550.3	Buñol	SH4
21	686081	4366485	678.7	Buñol	SH4
23	687395	4367165	642.1	Buñol	TU2
24	686399	4367241	677.9	Buñol	SH4
25	687167	4366381	651.9	Buñol	TU3
27	680022	4339150	888.5	Cortes	SH4
28	679612	4339118	890.3	Cortes	SH4
29	679512	4339436	888.5	Cortes	TU2
30	679136	4338383	914.3	Cortes	TU2
31	677125	4339062	923.1	Cortes	TU2
32	676154	4340493	871.1	Cortes	SH4
33	675694	4339951	863.8	Cortes	SH4
34	731367	4320387	566.4	Gandía	SH4
35	731174	4319768	573.3	Gandía	SH4
36	735641	4320359	387.2	Gandía	TU3
38	732701	4321149	522.8	Gandía	TU3
41	735989	4320416	396.4	Gandía	SH4

42	732146	4320783	538.6	Gandía	TU3
43	712253	4439959	634.0	Montanejos	TU2
44	712164	4440107	659.3	Montanejos	TU2
46	710705	4440212	779.0	Montanejos	TU2
51	731194	4417982	411.2	Eslida	TU2
52	731550	4417173	556.0	Eslida	TU3
53	730816	4416509	677.3	Eslida	SH4
55	731164	4416130	683.4	Eslida	TU3
56	731257	4415808	783.8	Eslida	SH4
57	753122	4491662	853.2	Morella	TU2
58	753472	4492428	848.9	Morella	TU3
59	753681	4493018	822.0	Morella	SH4
60	753389	4493440	818.8	Morella	TU3
61	753578	4493584	833.0	Morella	SH4
63	772442	4481471	417.0	Sant Mateu	SH4
64	771487	4480751	551.0	Sant Mateu	SH4
66	771628	4480160	529.0	Sant Mateu	TU2
67	770419	4479059	396.0	Sant Mateu	TU2
68	771600	4477908	498.0	Sant Mateu	SH4
69	726224	4280718	1049.4	Torre Macanes	TU2
71	726058	4279822	1005.2	Torre Macanes	SH4
72	725439	4279722	1048.1	Torre Macanes	SH4
73	726936	4275243	980.7	Torre Macanes	TU2
74	738849	4303169	698.2	Vall Gallinera	SH4
75	740994	4303634	650.0	Vall Gallinera	SH4
76	742392	4303306	560.0	Vall Gallinera	SH4
77	740233	4304438	571.1	Vall Gallinera	SH4
78	742995	4305117	512.5	Vall Gallinera	TU3
79	742597	4304663	496.4	Vall Gallinera	TU3
82	698934	4280159	873.2	Biar	TU2
83	698467	4280352	847.4	Biar	TU2
84	694871	4276628	829.4	Biar	TU2
85	756529	4286014	615.3	Bernia	TU2
86	756766	4285528	618.7	Bernia	SH4
87	757572	4284628	532.2	Bernia	SH4
88	757055	4284052	620.0	Bernia	SH4
91	669272	4368453	692.0	Requena	SH4
92	685018	4370184	550.3	Buñol	SH4
93	712809	4353794	319.5	Llombai	SH4

2.2 Humedad de combustible vivo

La extracción de muestras de las especies presentes en cada punto de muestreo, y el cálculo de la HCV ha sido realizado por la empresa VAERSA siguiendo las recomendaciones del laboratorio del fuego del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), considerando las siguientes formulas:

$$\text{Peso verde} = \text{Peso total} - \text{Peso bote} \quad (1)$$

En la fórmula (1) Peso total es el peso del bote con la muestra recogida en campo y Peso bote es el peso sin la muestra recogida en campo. Restando estos valores se obtiene el valor de peso verde. Con este dato se procede a calcular la HCV de la siguiente forma:

$$\text{HCV} = \frac{\text{Peso verde} - \text{Peso seco}}{\text{Peso seco}} * 100 \quad (2)$$

El Peso seco usando en la fórmula (2) se obtiene al eliminar la humedad de las muestras- Este proceso se lo realiza secando la muestra a 100°C durante 24 horas y el resultado es la obtención de la HCV en cada parcela muestreada. También se han calculado los estadísticos de Media ponderada, Media matorral y Media arbolado, esto realizando una ponderación de la HCV con la ayuda de Fracción de cabida cubierta (FCC) observada por los técnicos de VAERSA en campo como se muestra a continuación:

$$\text{Media ponderada} = \frac{\sum_{\text{todas especies}} (\text{HCV}_{\text{especie}} \text{ FCC}_{\text{especie}})}{\sum_{\text{todas especies}} \text{ FCC}_{\text{especie}}} \quad (3)$$

$$\text{Media matorral} = \frac{\sum_{\text{especies matorral}} (\text{HCV}_{\text{especie}} \text{ FCC}_{\text{especie}})}{\sum_{\text{especies matorral}} \text{ FCC}_{\text{especie}}} \quad (4)$$

$$\text{Media arbolado} = \frac{\sum_{\text{especies arbolado}} (\text{HCV}_{\text{especie}} \text{ FCC}_{\text{especie}})}{\sum_{\text{especies arbolado}} \text{ FCC}_{\text{especie}}} \quad (5)$$

En las fórmulas anteriores $\text{HCV}_{\text{especie}}$ indica el valor de HCV obtenido en cada una de las especies consideradas en el trabajo de campo en el punto de muestreo considerado. Como resultado se obtiene una base de datos de la HCV entre las fechas de junio del 2019 a octubre del 2021. En la Figura 4 se muestra como ejemplo la parcela 1 con la humedad de combustible vivo para sus diferentes especies y sus estadísticos. En esta parcela la especie de arbolado *Pinus halepensis* tiene la mayor FCC con un 99%, siendo la más predominante de la parcela, mientras que las demás especies están por debajo del 10%. Se puede apreciar que en esta parcela la HCV de media arbolado (rosa, que coincide con la curva de *Pinus halepensis* (roja)) es más alta que la media matorral (verde) y que media ponderada (verde oscuro) se encuentra en la mitad de las dos líneas de tendencia.

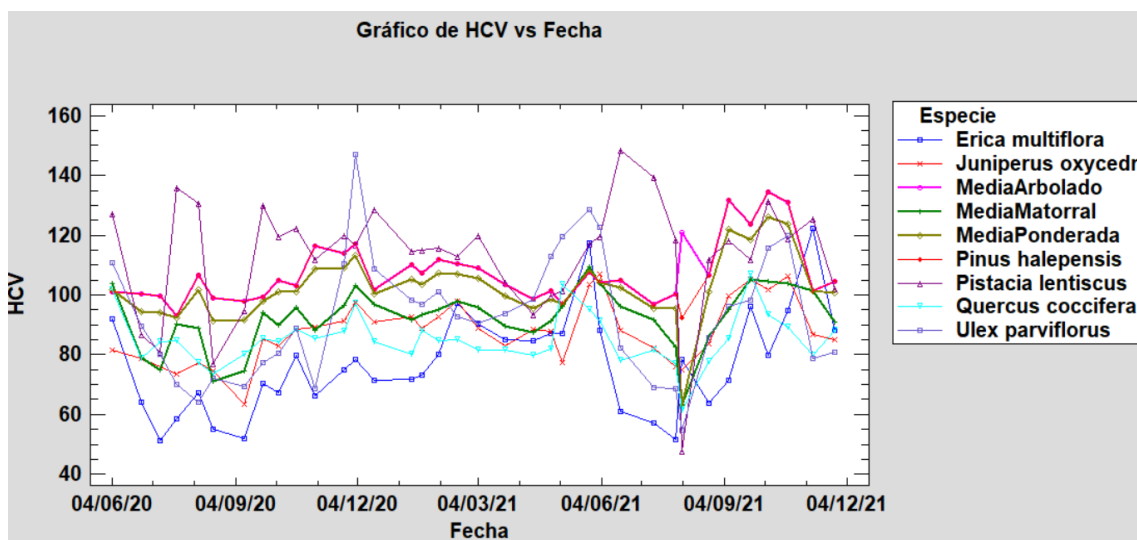


Figura 4. Series temporales de HCV en la Parcela 1 para las especies presentes.

2.3 Datos meteorológicos

Los datos meteorológicos usados para este estudio fueron facilitados por la Agencia Estatal de Meteorología (AEMET) al tutor de este trabajo fin de máster a través de la Direcció General de prevenció d'incendis forestals de la Generalitat Valenciana. Los datos fueron tratados por los miembros del grupo de investigación del tutor de este trabajo (Grupo CGAT-UPV), siguiendo el procedimiento descrito en el artículo (Costa-Saura, Balaguer-Beser, Ruiz, Pardo-Pascual, & Soriano-Sancho, 2021) denominado “Empirical Models for Spatio-Temporal Live Fuel Moisture Content Estimation in Mixed Mediterranean Vegetation Areas Using Sentinel-2 Indices and Meteorological Data”. A continuación, se describe el procedimiento seguido por el mencionado trabajo; consiste en recoger los datos diarios de precipitación, máximo, mínimo, y media diaria de temperatura, media de velocidad en km/h de las rachas máximas de viento diarias y mínimo de humedad relativa diaria de estaciones meteorológicas cercanas a los puntos de muestreo y pertenecientes a la Agencia Estatal de Meteorología de España (AEMET). Con los valores observados por las estaciones meteorológicas se realizó una interpolación hacia los puntos de muestreo de HCV usando la herramienta “*Meteoland R package*”. Este es un enfoque similar a la ponderación inversa de la distancia, el cual interpola las variables climáticas usando filtros gaussianos truncados (Cáceres & Granda, 2022).

De todas las variables meteorológicas también se obtuvieron los valores acumulados con el uso de ventanas de días previos respecto de la fecha de toma de datos en campo de HCV. Para la temperatura se obtuvieron valores promedio de 7, 15, 30 y 60 días previos a la fecha de muestreo de HCV. Los datos acumulados para la precipitación fueron calculados para 3, 7, 15, 30 y 60 días previos. Para la media de humedad relativa mínima y media de velocidad del viento, se obtuvieron los datos acumulados de 3, 7 y 15 días

previos. En la Tabla 3 se muestran todas las variables meteorológicas acumuladas calculadas para este estudio.

Tabla 3. *Variables meteorológicas.*

Tipo	Variable	Unidades	Descripción
Precipitación	P3	mm	Precipitación acumulada en los últimos 3 días.
	P7	mm	Precipitación acumulada en los últimos 7 días.
	P15	mm	Precipitación acumulada en los últimos 15 días.
	P30	mm	Precipitación acumulada en los últimos 30 días.
	P60	mm	Precipitación acumulada en los últimos 60 días.
Temperatura	MeanTemperature	°C	Media de la temperatura diaria
	MaxTemperature	°C	Máxima de la temperatura diaria
	MinTemperature	°C	Mínima de la temperatura diaria
	T7	°C	Media de las temperaturas promedio diaria de los últimos 7 días.
	T15	°C	Media de las temperaturas promedio diaria de los últimos 15 días.
	T30	°C	Media de las temperaturas promedio diaria de los últimos 30 días.
	T60	°C	Media de las temperaturas promedio diaria de los últimos 60 días.
Humedad Relativa	MeanRelativeHumidity	%	Media de la humedad relativa diaria
	MaxRelativeHumidity	%	Máxima de la humedad relativa diaria
	MinRelativeHumidity	%	Mínima de la humedad relativa diaria
	HUmin3	%	Media de la humedad relativa mínima de los últimos 3 días

	Humin7	%	Media de la humedad relativa mínima de los últimos 7 días
	Humin15	%	Media de la humedad relativa mínima de los últimos 15 días
Viento	W3	Km/h	Media de la velocidad máxima del viento de los últimos 3 días
	W7	Km/h	Media de la velocidad máxima del viento de los últimos 7 días
	W15	Km/h	Media de la velocidad máxima del viento de los últimos 15 días
Radiación	Radiation	MJ/m2	Radiación solar diaria incidente

También se ha añadido como variable el día del año DOY para intentar describir como las estaciones del año pueden afectar a la HCV. Para usarlo como variables predictoras se ha normalizado el rango de $[0, 366]$, correspondiente a días del año, a un rango entre $[-\pi, \pi]$. Obtenido el rango normalizado se ha calculado el seno (\sin_DOY) y coseno (\cos_DOY). De este modo se logra obtener una curva que describe entre -1 y 1 las épocas del año más húmedas o secas respectivamente (Zhu, Webb, Yebra, Scortchini, & Miller, 2021). Como se puede ver en la Figura 5, los valores de la variable \sin_DOY son positivos para los días de verano y otoño sobre los que centraremos la atención en este trabajo.

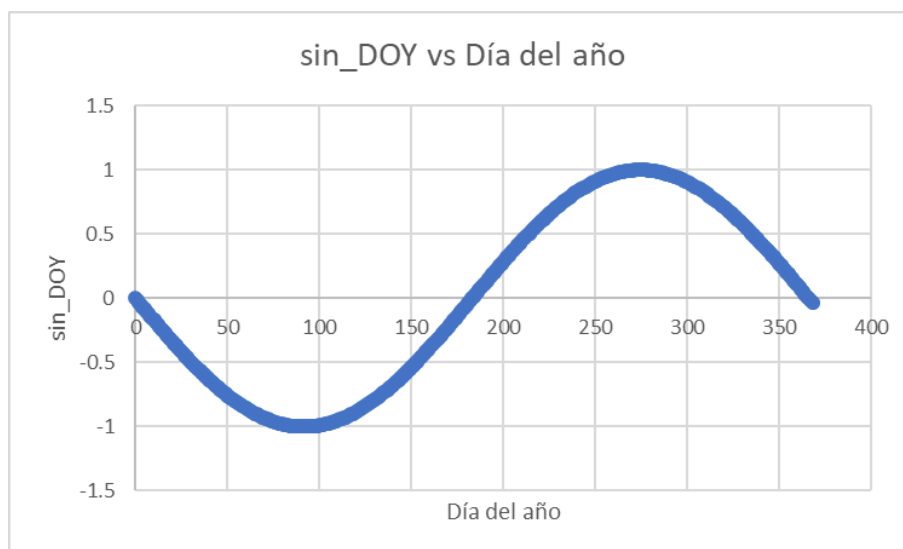


Figura 5. Comportamiento de la variable \sin_DOY a lo largo del año.

2.4 Datos de índices espectrales

La vegetación del territorio comprende un factor del medio físico determinante para el equilibrio de un ecosistema, por esto son varios los estudios previos que demuestran la aportación de los índices de vegetación para la modelación de la HCV (Yebra, y otros, 2013). A través de imágenes satelitales es posible el monitoreo espacio temporal de la cobertura vegetal de la superficie terrestre, y extraer sus características espectrales en forma de bandas separadas por rango de longitud de onda del espectro electromagnético, que dependiendo del tipo de sensor que se use ofrecerá una serie de posibilidades para describir la vegetación. Por lo tanto, un índice espectral es el resultado de la operación algebraica entre las bandas de una imagen satelital, que nos servirán para analizar las formaciones vegetales características de un lugar y tiempo.

Sentinel es una flota de satélites diseñada específicamente para brindar abundantes datos e imágenes sobre la superficie terrestre y vegetación en el programa Copernicus de la Comisión Europea. Específicamente Sentinel-2A lanzada en 2015, posee un sensor multiespectral de 13 bandas que proporciona imágenes de alta resolución, que varía de 10 a 60 m y cubre toda la superficie terrestre cada 5 días (ESA, 2015). Los detalles de las bandas de las imágenes Sentinel-2 Nivel-2A se muestran en la Tabla 4.

Tabla 4. *Características de las imágenes Sentinel-2. Fuente: (GEE, 2022).*

Nombre	Resolución	Longitud de onda	Descripción
B1	60 m	443.9nm	<i>Aerosol</i>
B2	10 m	496.6nm	<i>Blue</i>
B3	10 m	560nm	<i>Green</i>
B4	10 m	664.5nm	<i>Red</i>
B5	20 m	703.9nm	<i>Red Edge 1</i>
B6	20 m	740.2nm	<i>Red Edge 2</i>
B7	20 m	782.5nm	<i>Red Edge 3</i>
B8	10 m	835.1nm	<i>NIR</i>
B8A	20 m	864.8nm	<i>Red Edge 4</i>
B9	60 m	945nm	<i>Water vapor</i>
B11	20 m	1613.7nm	<i>SWIR 1</i>
B12	20 m	2202.4nm (<i>SWIR 2</i>
AOT	10 m		<i>Aerosol Optical Thickness</i>
WVP	10 m		<i>Water Vapor Pressure.</i>
SCL	20 m		<i>Scene Classification Map</i>
TCL_R	10 m		<i>True Color Image, Red channel</i>

TCI_G	10 m	<i>True Color Image, Green channel</i>
TCI_B	10 m	<i>True Color Image, Blue channel</i>
MSK_CLDPRB	20 m	<i>Cloud Probability Map</i>
MSK_SNWPRB	10 m	<i>Snow Probability</i>
QA10	10 m	<i>Always empty</i>
QA20	20 m	<i>Always empty</i>
QA60	60 m	<i>Cloud mask</i>

Gracias a la versatilidad que presenta *Google Earth Engine (GEE)*, se generaron varios índices de vegetación que en trabajos previos han servido para la estimación del HCV, y otros índices espectrales que también caracterizan el comportamiento de una planta y describen el verdor de una hoja o la cantidad de agua presente que puede afectar a la producción de clorofila o la estructura interna de la hoja y por lo tanto la cantidad de humedad que una planta puede almacenar (Gilbert, González-Piqueras, & García-Haro, 1997). En la Tabla 5 se muestran los 22 índices espectrales usados en este estudio, con su respectiva fórmula.

Tabla 5. *Índices espectrales usados para el presente estudio.*

Índice	Fórmula	Referencia
Índice de vegetación normalizada	$NDVI = \frac{NIR - ROJO}{NIR + ROJO}$	(Huang, Tang, Hupy, Wang, & Shao, 2021)
Índice de vegetación mejorado	$EVI = 2.5 * \frac{NIR - ROJO}{(NIR + 6 * ROJO - 7.5 * AZUL) + 1}$	(Huete, y otros, 2002)
Índice de vegetación ajustado al suelo	$SAVI = 1.5 * \frac{NIR - ROJO}{NIR + ROJO + 0.5}$	(Ahamed, Tian, Zhang, & Ting, 2011)
Índice de humedad de diferencia normalizada	$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1}$	(Wilson & Sader, 2002)

Índice de resistencia atmosférica visible	$VARI = \frac{VERDE - ROJO}{VERDE + ROJO - AZUL}$	(Hunt Jr., Daughtry, Eitel, & Long, 2011)
Índice de agua de diferencia normalizada	$NDWI = \frac{VERDE - NIR}{VERDE + NIR}$	(Gao, 1996)
Índice de diferencia verde rojo normalizado	$V_{green} = \frac{VERDE - RED\ EDGE1}{VERDE + RED\ EDGE1}$	(Tucker, 1979)
Índice de estrés por humedad	$MSI = \frac{SWIR1}{NIR}$	(Heiskanen, 2006)
Simple Ratio	$RVI = \frac{NIR}{ROJO}$	(Pearson & Miller, 1972)
Índice de Vegetación Ajustado al Suelo Optimizado	$OSAVI = (1 + 0.16) \frac{NIR - ROJO}{NIR + ROJO + 0.16}$	(Wu, Niu, Tang, & Huang, 2008)
Índice de reflectancia de absorción de clorofila transformada	$TCARI = 3 * ((RED\ EDGE1 - ROJO) - 0.2 * (RED\ EDGE1 - VERDE) * \frac{RED\ EDGE1}{ROJO})$	(Hunt Jr., Daughtry, Eitel, & Long, 2011)
Índice de reflectancia de absorción de clorofila modificado	$MCARI = \frac{TCARI}{3}$	(Hunt Jr., Daughtry, Eitel, & Long, 2011)
Índice de sequía multibanda normalizado	$NMDI = \frac{RED\ EDGE4 - (SWIR1 - SWIR2)}{RED\ EDGE4 + (SWIR1 - SWIR2)}$	(Wang & Qu, 2007)
TCARI/OSAVI	$TCARI_OSAVI = \frac{TCARI}{OSAVI}$	(Haboudane, Miller, Pattey, P.J., &

		Strachan, 2004)
MCARI/OSA VI	$MCARI_OSAVI = \frac{MCARI}{OSAVI}$	(Main, y otros, 2011)
Índice de vegetación de área foliar específica	$SLA = \frac{NIR}{RED\ EDGE1 + SWIR2}$	(Lymburner, Beggs, & Jacobson, 2000)
Índice de vegetación resistente a la atmósfera	$ARVI = \frac{NIR - 2 * ROJO + AZUL}{NIR + 2 * ROJO + AZUL}$	(Ehammer, Fritsch, Conrad, Lamers, & Dech, 2010)
Índice de vegetación avanzado	$AVI = (NIR * (1 - ROJO) - (NIR - ROJO))^{1/3}$	(Anand, Singh, & Kanga, 2018)
Índice de clorofila verde	$GCI = \frac{NIR}{VERDE} - 1$	(Ali, Martelli, Lupia, & Barbanti, 2019)
Índice de vegetación de diferencia normalizada verde	$GNDVI = \frac{NIR - VERDE}{NIR + VERDE}$	(Gitelson, Kaufman, & Merzlyak, 1996)e
Índice de Vegetación Ajustado al Suelo Modificado	$MSAVI2 = \frac{2 * NIR + 1 - ((2 * NIR + 1)^2 - 8 * (NIR - ROJO))^{1/2}}{2}$	(Wu, Wang, & Bauer, 2007)
Índice de clorofila de pigmento normalizado	$NPCI = \frac{ROJO - AZUL}{ROJO + AZUL}$	(Main, y otros, 2011)

3. METODOLOGÍA

3.1 Estructura de procesos

A continuación, se muestra en un esquema, los procesos realizados desde la obtención de datos, los análisis estadísticos y la obtención de los resultados del presente trabajo.

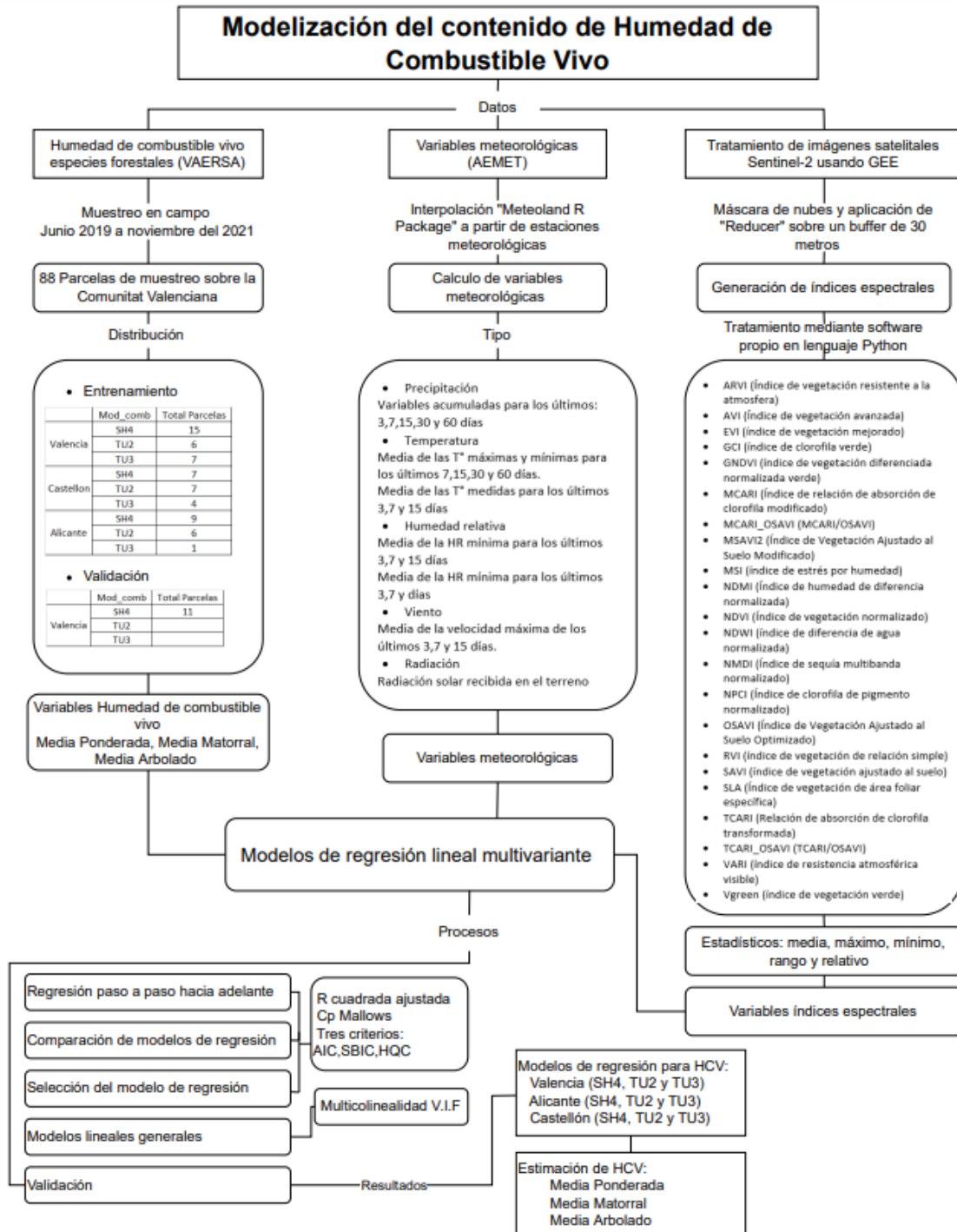


Figura 6. Estructuras de procesos de la metodología.

3.2 Obtención de índices espectrales a partir de imágenes satelitales Sentinel-2

El tratamiento de imágenes Sentinel-2 para la creación de índices espectrales se realizó a través de la plataforma *Google Earth Engine (GEE)* que combina un catálogo de varios *petabytes* de imágenes satelitales y conjuntos de datos geoespaciales con capacidades de análisis a escala planetaria (Earthengine, 2022). *GEE* facilita el tratamiento de las imágenes en línea gracias a *Earth Engine JavaScript API*, donde a través de programación se accede a las imágenes Sentinel-2 Nivel-2A para las fechas comprendidas entre junio del 2019 a octubre del 2021 y se extraen las bandas necesarias para la creación de los índices espectrales. Los códigos de *JavaScript* completos se los encuentra en el apartado de ANEXOS.

3.2.1 Tratamiento de imágenes satelitales para la descarga de índices espectrales usando *Google Earth Engine (GEE)*

3.2.1.1 Insumos de entrada

Obtenida la posición del centro de cada parcela correspondiente a los puntos de muestreo proporcionados por la empresa VAERSA, se realizó un *buffer* de 30 metros alrededor de cada centro y, como se muestra en la Figura 7, se ha precautelado que cada polígono formado por el *buffer*, cubra áreas homogéneas de la vegetación de la zona y, sobre todo, no entre en conflicto con estructuras artificiales como podrían ser caminos o edificios. El resultado es una capa geográfica -conocida como *shape*- de polígonos que contiene todos los *buffers* con la información de su parcela respectiva. Una vez obtenido el *shape* de polígonos a través de la herramienta *Assets* se cargaron los ficheros de *shp*, *dbf*, *prj*, *shx* y *cpg*. De esta forma ya se tiene la capa de *buffers* en la *API*.

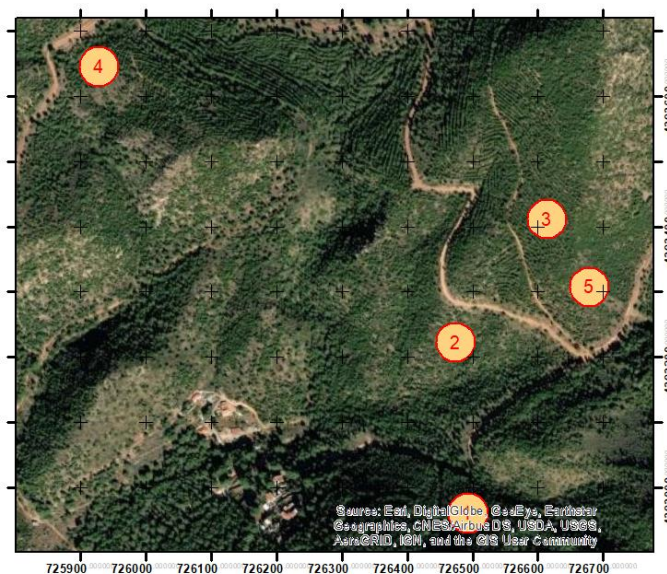


Figura 7. Buffers creados alrededor de los puntos de muestreo.

La API de *Google Earth Engine* permite hacer la búsqueda de las imágenes satelitales con las que se quiere trabajar. En este caso se usó la colección de “*COPERNICUS/S2_SR*” que son imágenes corregidas atmosféricamente a través de *Sen2cor* (GEE, 2022). Mediante el uso de los métodos “*ee.FeatureCollection*” y “*ee.ImageCollection*” se puede cargar en variables el archivo *shape* e imágenes satelitales respectivamente. Para seleccionar las imágenes satelitales que se desean en el tiempo, se aplicó un filtro de la forma “*filterDate('2019-06-01', '2021-10-31')*” y se seleccionaron las fechas que se desean estudiar, en este caso el periodo de junio del 2019 a octubre del 2021. Para seleccionar las imágenes satelitales que se desean en el espacio, se aplicó el filtro de la forma “*filterBounds(table2)*” y se seleccionó la variable que corresponde al *shape* de polígonos.

3.2.1.2 Máscara de nubes

Obtenidos los insumos de entrada se realizó una máscara de nubes para eliminar en las imágenes satelitales los píxeles con presencia de nubes, puesto que no representan información para este caso de estudio. Para ello, se ha creado la función *maskS2clouds*, la cual accede a la banda “QA60” que contiene información sobre las nubes marcando entre 0 y 1 la presencia o ausencia en cada píxel de la escena; luego, se realiza un bandeo a través del método “*bitwiseAnd*” y la actualización de la máscara a través del método “*updateMask*”. Creada la función, se aplicó a todas las imágenes de la colección. El resultado es la imagen satelital sin píxel de nubes, como se ve en la Figura 8. Aunque el uso de esta función de enmascaramiento ayuda a eliminar una gran porción de la presencia de nubes, no lo hace de forma total, pues se encontraron valores de píxel que aun presentan nubosidades y que sobre todo siguen afectando a los *buffers* de los puntos de muestreo de humedad de combustible vivo en ciertos sectores y épocas del año. Por esto será necesario a posteriori, realizar un procesamiento de datos para eliminar el posible efecto negativo de las nubes sobre estos datos.



Figura 8. Aplicación de la máscara de nubes en las imágenes satelitales.

3.2.1.3 Generación de índices espectrales

Una vez que se obtuvo la colección de imágenes filtradas en tiempo y espacio, y que se realizó un proceso para eliminar lo mejor posible la presencia de nubes, se analizaron los índices de espectrales sobre toda la escena. Un índice espectral es la relación numérica que existe en valores de reflectancia de distintas longitudes de onda, que tienen una alta respuesta a la cobertura vegetal (Muñoz, 2013). Para su cálculo se accede a las bandas de cada imagen y se opera algebraicamente; esto se hizo a través del método “*expression*” el cual permite aplicar la fórmula correspondiente al índice y acceder a las bandas necesarias de cada imagen.

Para cada uno de los índices se creó una función a través del método “*addBands*” que devuelve la misma imagen, pero añadida una banda extra correspondiente al cálculo del índice espectral. Como se puede ver en la Figura 9, el resultado es el cálculo del índice sobre toda la escena.

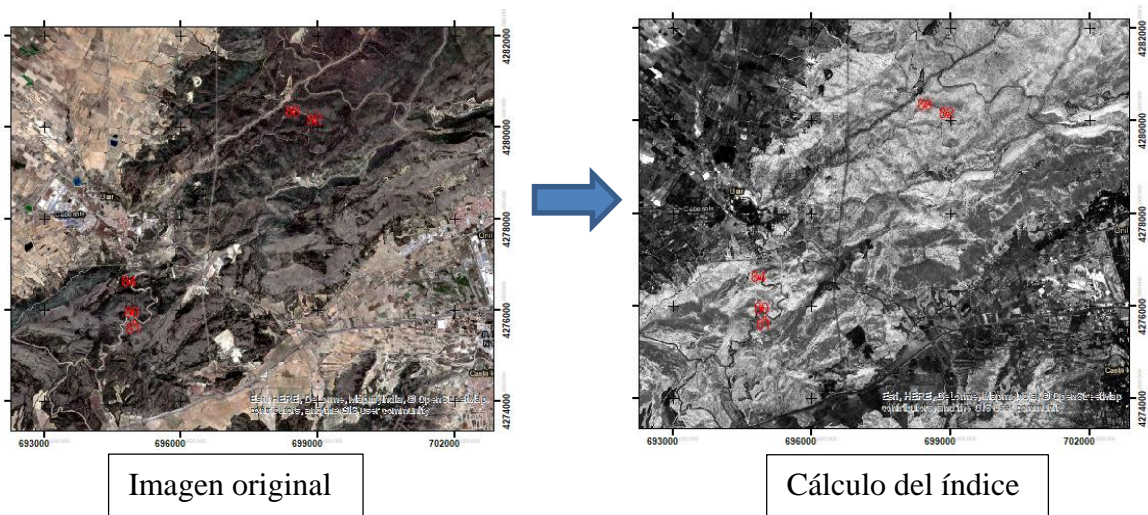


Figura 9. Generación de índices espectrales.

En la operación de bandas de cada imagen puede suceder que intervengan dos bandas con diferentes escalas; en este caso, la imagen resultante tendrá la escala del tamaño de píxel más grande entre estas dos imágenes, sin embargo, más adelante se aplica un reductor en el cual se estandariza la escala resultante de las imágenes.

3.2.1.4 Aplicación de ReduceRegions

Obtenidas las imágenes de índices espectrales para cada escena del periodo de estudio, se calculó la media de los píxeles que caen dentro de cada *buffer* de 30 metros correspondiente a las parcelas de muestreo de Humedad de Combustible Vivo (HCV). Para esto, se aplicó el método “*ReduceRegions*”, que reduce todos los píxeles de la(s) región(es) a una estadística u otra representación compacta de los datos de píxeles en la región (por ejemplo, histograma). La región se representa como una geometría, que puede ser un polígono que contiene muchos píxeles, o puede ser un solo punto, en cuyo caso solo habrá un píxel en la región. En cualquier caso, como se ilustra en la Figura 10, el resultado es una estadística: en nuestro caso “*mean*” derivada de los píxeles de la región (GEE, 2022). Para próximos análisis, también se ha descargado los datos correspondientes a la desviación estar “*stdDEV*”.

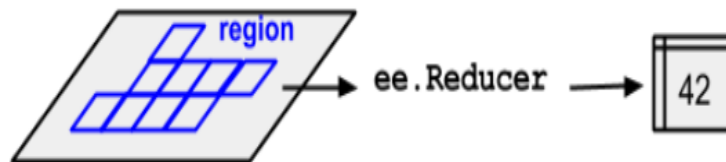


Figura 10. Aplicación de “ReduceRegions” en una imagen. Fuente: (GEE, Statistics of an Image Region, 2022)

Para obtener todos los estadísticos resultantes se aplicó el método “*flatten*”, el cual permite formar una tabla con base en datos resultantes y exportarla hacia un formato CSV como se ve en la Figura 11.

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Name	Date	ARVI	EVI	MCARI	MCARI_OSAV	MSI	NDMI	NDVI	NDWI	NMDI	OSAVI	RVI	
2		1	05/06/2019	0.35938117	0.36405824	0.03062164	0.06130085	0.61250321	0.24097191	0.68687352	0.47594188	0.59264767	0.49985007	5.40350605
3		1	10/06/2019	0.32906603	0.36135704	0.03246781	0.06879731	0.65821401	0.20727631	0.65026859	0.43365938	0.59803903	0.47236472	4.73876066
4		1	15/06/2019	0.33233378	0.3486706	0.03249042	0.06777634	0.64139006	0.21893366	0.6616685	0.4562972	0.59038407	0.47940418	4.92200881
5		1	20/06/2019	0.31016135	0.34916453	0.03291796	0.07116511	0.66891584	0.19892797	0.63555165	0.43116414	0.59148021	0.46266831	4.49779174
6		1	25/06/2019	0.12568897	0.34943832	0.0366429	0.13163791	0.77845259	0.12485451	0.32720911	0.33093863	0.58012021	0.28107241	1.98774291
7		1	30/06/2019	0.31087745	0.33520497	0.03230345	0.06949564	0.69881955	0.17780976	0.63914679	0.41008055	0.57643161	0.46498619	4.55168668
8		1	05/07/2019	0.19387998	0.28628467	0.03484423	0.0975627	0.75344628	0.14207762	0.46748938	0.36856093	0.58912804	0.35841474	2.76385529
9		1	10/07/2019	0.2973567	0.34164323	0.0347777	0.07459112	0.67975135	0.19123791	0.65088594	0.43089838	0.58224341	0.46636512	4.73845253
10		1	15/07/2019	0.26630429	0.3297946	0.03780055	0.08187622	0.67392998	0.19546298	0.63873111	0.44656074	0.57924618	0.46168718	4.54946549
11		1	20/07/2019	0.21645101	0.32617674	0.0369323	0.09213433	0.70193362	0.17569902	0.53217492	0.40132429	0.59028853	0.40106002	3.27655075
12		1	25/07/2019	0.25861514	0.32028286	0.0380734	0.08345287	0.67618267	0.19372476	0.62176575	0.44147074	0.57825395	0.45634625	4.29706954
13		1	30/07/2019	0.25495602	0.34882602	0.04006467	0.0884048	0.65274836	0.2107861	0.62094688	0.45084868	0.59326885	0.4531969	4.28508084
14		1	04/08/2019	0.25175495	0.32942652	0.04060498	0.08678517	0.65616448	0.20816223	0.64234604	0.45704105	0.58969914	0.46803582	4.60697102
15		1	09/08/2019	0.28231417	0.35122921	0.03873823	0.0799728	0.62983295	0.22791624	0.67678594	0.47436021	0.58723806	0.48448316	5.19882647
16		1	14/08/2019	0.23791602	0.33267307	0.04223595	0.09010052	0.63420834	0.22452857	0.64893275	0.47634165	0.59070607	0.46900228	4.7153894
17		1	19/08/2019	0.20983303	0.40404871	0.05080454	0.1130214	0.63273721	0.22579507	0.57548226	0.46780805	0.60358278	0.44957926	3.71437957
18		1	24/08/2019	0.24685652	0.34631838	0.04249268	0.08611199	0.57780645	0.26835194	0.68928435	0.52053573	0.60455358	0.49368742	5.45797337
19		1	03/09/2019	0.18334824	0.45788285	0.05967547	0.13033613	0.56626895	0.27727136	0.56010668	0.51965736	0.62579544	0.45789615	3.54990595
20		1	18/09/2019	0.22304262	0.38217418	0.04485444	0.09314255	0.58425713	0.26410038	0.62339657	0.53571852	0.57611993	0.48174522	4.34530486

Figura 11. Exportación de índices espectrales a formato CVS.

Uno de los parámetros de “*ReduceRegion*” es “*scale*”, el cual permite decidir cuál será la escala o tamaño de pixel de la imagen resultado. Las unidades de este valor dependerán del sistema de referencia que usen las imágenes que en este caso es metros; para este estudio se determinó una escala resultante de 20 metros.

3.2.2 Tratamiento de base de datos usando Python

3.2.2.1 Criterio NDVI y coeficiente de variación (CV)

Obtenida la base de datos que contiene todos los índices espectrales, se realizó un análisis del comportamiento de estos índices en las fechas establecidas y su tratamiento, como se muestra en la Figura 12.

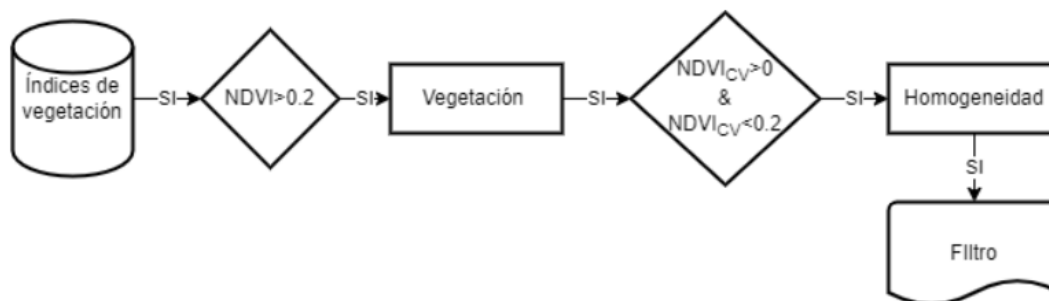


Figura 12. Diagrama del flujo seguido para el tratamiento de índices espectrales.

En la Figura 13 se puede ver el cálculo del índice espectral NDVI para la parcela 62 en la serie temporal desde junio del 2019 hasta octubre del 2021. La serie temporal presenta varios picos que se deben a la presencia de valores de NDVI muy bajos teniendo una mínima de -0.029 en la fecha julio del 2021, puesto que, este índice determina la respuesta del espectro en la longitud onda de las bandas del rojo y del infrarrojo, que son mayormente sensitivas a la cobertura vegetal. Existen varios autores que determinan que valores de $NDVI < 0.2$ no corresponden a la detección de un cuerpo vegetal; por tanto, estos valores pueden hacer referencia a infraestructura o suelo desnudo (Yang et al., 2019) (Yang, Luo, Huang, Wu, & Sun, 2019). En este caso, estos valores negativos y por debajo del umbral 0.2 podrían corresponder a la presencia de nubes y se considera al dato como una anomalía, haciendo que todos los índices correspondientes a esa fecha sean descartados para el análisis.

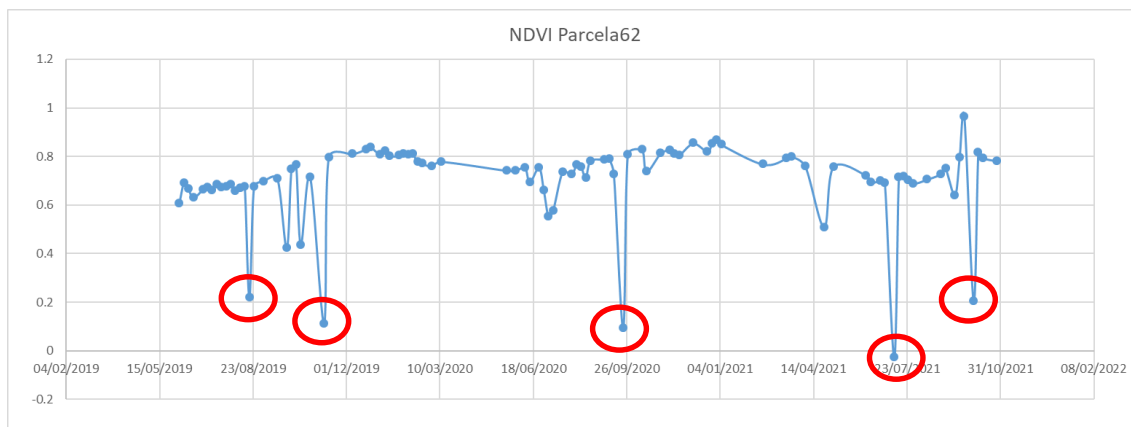


Figura 13. Anomalías presentes en el índice espectral NDVI de la parcela 62.

También, en este mismo sentido, se analiza el coeficiente de variación del índice espectral NDVI, resultante de dividir la desviación estándar para la media. Estos dos estadísticos fueron previamente descargados en el apartado anterior. El coeficiente de variación toma valores entre 0 y 1. Si el coeficiente es próximo al 0, significa que existe poca variabilidad en los datos y es una muestra muy homogénea. En cambio, si tienden a 1, es una muestra muy dispersa y la media pierde confiabilidad.

$$CV = \frac{S_x}{\bar{X}} \quad (6)$$

Para el coeficiente de variación se seleccionó un umbral que va desde 0 a 0.2; de este modo, si el CV está fuera de este rango, se considera poca homogeneidad en la parcela, por lo cual se considera que el dato es una anomalía y todos los índices espectrales correspondientes a esa fecha, serán descartados (Yebra, y otros, 2013) (Cunill Camprubi, González-Moreno, & Resco de Dios, 2022). En la Figura 14, se ven los resultados de aplicar este primer criterio NDVI y coeficiente de variación (CV). Así, la mayoría de picos inferiores que representaban a valores bajos de NDVI han sido filtrados; sin embargo, visualmente, se puede apreciar que la serie temporal todavía presenta ciertos picos alrededor de las fechas de octubre del 2019 y de octubre del 2021, que a priori se podría decir que no responden a la realidad del NDVI sobre la Parcela 62, pues es muy difícil que en lapsos tan cortos de tiempo se produzca tal variación del índice.

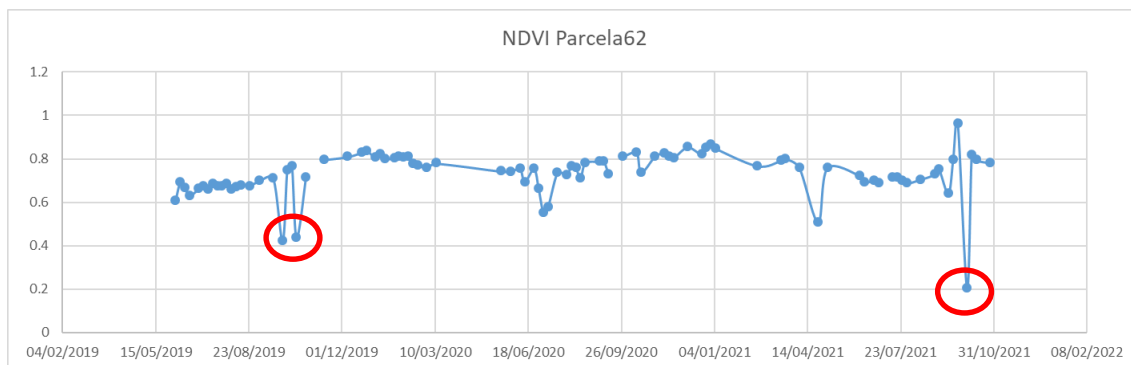


Figura 14. Índice espectral después de aplicar el criterio de NDVI y CV

3.2.2.2 Criterio de la mediana

Ya que la serie temporal de los datos de índices espectrales todavía presenta picos que no podrían representar un cambio biofísico en la realidad y que podrían afectar a la estimación de modelo, se ha aplicado a la serie temporal un nuevo criterio, en el cual, se usa una ventana de tres días para comparar la mediana con la desviación estándar de la siguiente manera. Esta metodología fue propuesta por (Lymburner, y otros, 2011).

$$\frac{\text{Índice}_i - \mu}{\sigma} > x \quad (7)$$

Donde Índice_i representa al valor del índice espectral en determinada fecha y parcela, μ representa la mediana calculada en los datos usando una ventana de 3 fechas, una fecha adelante y una atrás, σ corresponde a la desviación estándar a lo largo de toda la serie temporal de la parcela que se está estudiando y x es un valor umbral. (Yebra, y otros, 2018) proponen $x = 1$ ya que evita cambios abruptos poco realistas en el índice espectral dentro de fechas de muestreo consecutivas.

Como se ve en la Figura 15, para la parcela 62, los picos, que anteriormente se presentaban en octubre del 2019 y octubre del 2021, han desaparecido, lo cual ha hecho que la serie temporal final se vea mucho más homogénea y se eliminen aquellos casos, donde se presentaban valores que no podrían corresponder con un cambio biofísico real de la vegetación.

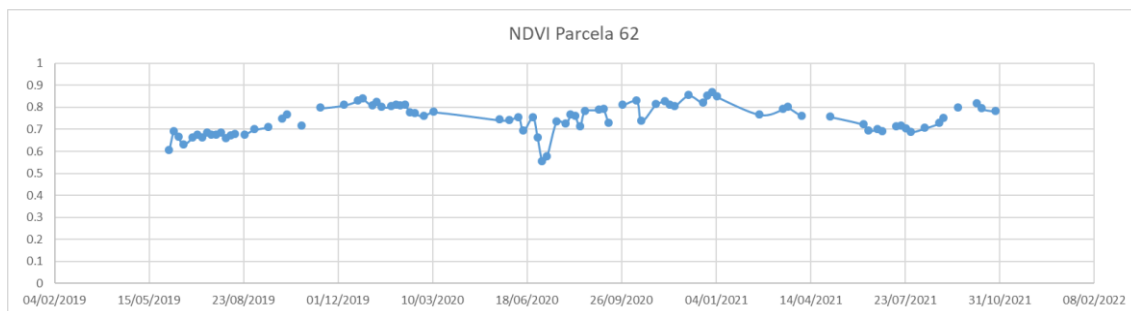


Figura 15. Índice espectral después de aplicar el criterio de la mediana

3.2.2.3 Aplicación de filtro Savitzky-Golay

Una vez eliminados aquellos valores que muestran un comportamiento anómalo en la serie temporal, se han generado espacios en la serie temporal que corresponde a fechas con datos vacíos, por lo cual es necesario realizar una interpolación para completar la serie temporal. Para esto se aplicó el filtro de *Savitzky-Golay*. Este proceso se realizó en *Python* usando la librería *scipy.signal.savgol_filter*, como se muestra en el código adjunto, en el apartado ANEXOS.

El método *Savitzky-Golay* utiliza funciones polinómicas locales, para suavizar los datos de la serie temporal y suprimir las perturbaciones, sustituyendo cada uno de los datos por una combinación lineal de valores vecinos, dentro de una ventana definida por el usuario. Este método permite ajustar bien pequeñas fluctuaciones en los datos, siendo también el más sensible al ruido (De Castro, Plant, Six, & Peña, 2014). La aplicación de este método se realizó usando una ventana de 5 fechas para todos los índices y usando un polinomio de orden 2. Una vez que se obtiene la interpolación *Savitzky-Golay* sus valores servirán para reemplazar los espacios vacíos que han generado el Criterio NDVI, Criterio de Coeficiente de Variación (CV) y Criterio de la Mediana.

Como se ve en la Figura 16, en la serie temporal para el índice espectral NDVI en la parcela 62, se eliminaron los valores que corresponderían al comportamiento anómalo del índice espectral y, como resultado, se obtuvo una serie temporal mucho más homogénea.

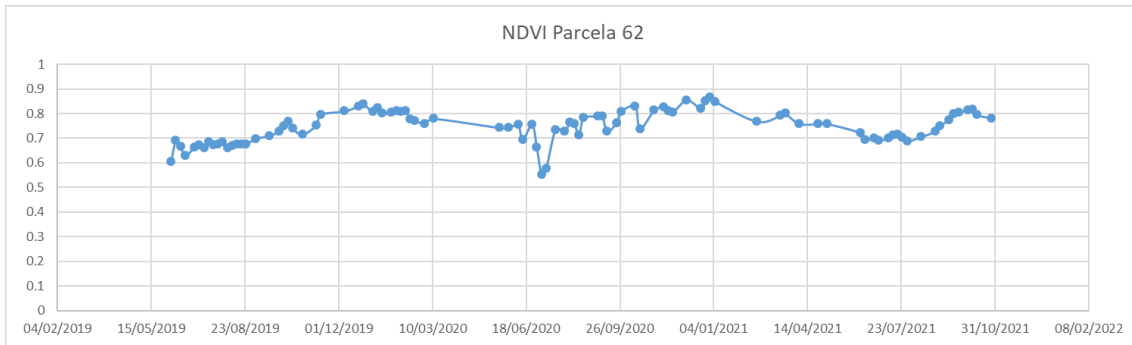
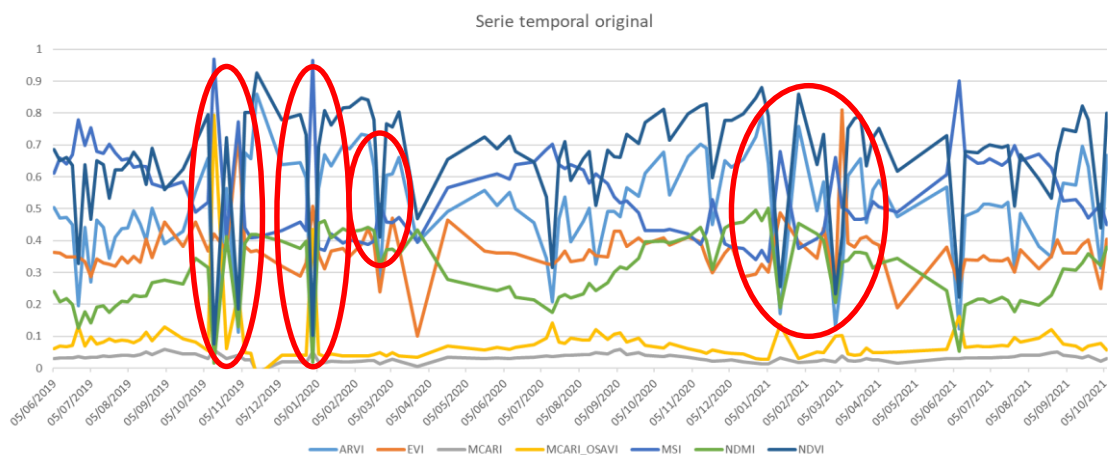


Figura 16. Índice espectral después de aplicar el filtro Savitsky-Golay

En la Figura 17 se aprecia la serie temporal original de varios índices espectrales descargados y sin tratar. También se observa la misma serie temporal corregida, donde los datos ya han sido tratados; en la serie temporal original, existe mucha presencia de ruido, haciendo que las curvas muestren altos picos, que no explican un comportamiento biofísico real de la vegetación en la parcela. Por ejemplo, para las fechas de octubre del 2019 y enero de 2020, las curvas se entrelazan de tal forma, que hace imposible distinguir la tendencia del índice espectral; en este caso, incluso la variable MCARI_OSAVI, que presenta valores más bajos alejados de las demás curvas alrededor de 0.1, llega a tener valores de 0.8 más próximos a lo que se esperaría de índices como el NDVI. Este mismo comportamiento se repite para enero y marzo del 2021, en varios índices.

Por otro lado, se encuentra la serie temporal corregida, donde se aprecia que el ruido ha desaparecido y que las líneas de tendencia son mucho más homogéneas a lo largo del tiempo, haciendo que los índices representen de manera más homogénea la realidad del terreno.



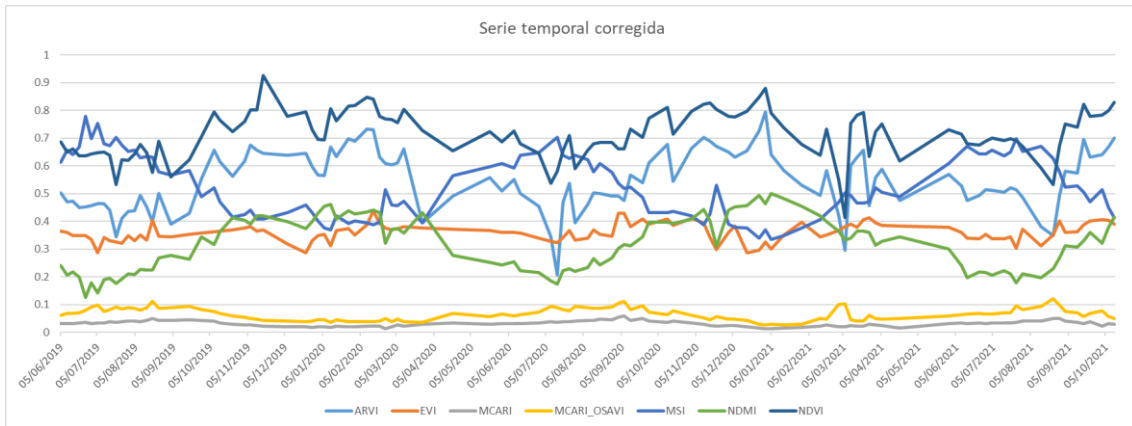


Figura 17. Serie temporal original o descargada (superior) y la serie temporal corregida (inferior).

Los datos de índices espectrales fueron unidos a los datos de humedad de combustible vivo, a través de la fecha más próxima. Este proceso fue automatizado en *Python*, el principio fue tomar como referencia el número de parcela y la fecha de toma de dato de HCV en campo; luego, buscar en la base de datos de espectrales depurados de la parcela correspondiente y la fecha más aproximada posible. En el 99% de los casos, la variabilidad de fechas entre el dato de campo y el índice espectral es menor de los 12 días; sin embargo, existen dos fechas con una variabilidad mayor correspondiente a las parcelas 34 y 36 con fecha 28/10/2021 que tiene una variabilidad de 16 días y la segunda de 21, como se ve en la Figura 18.

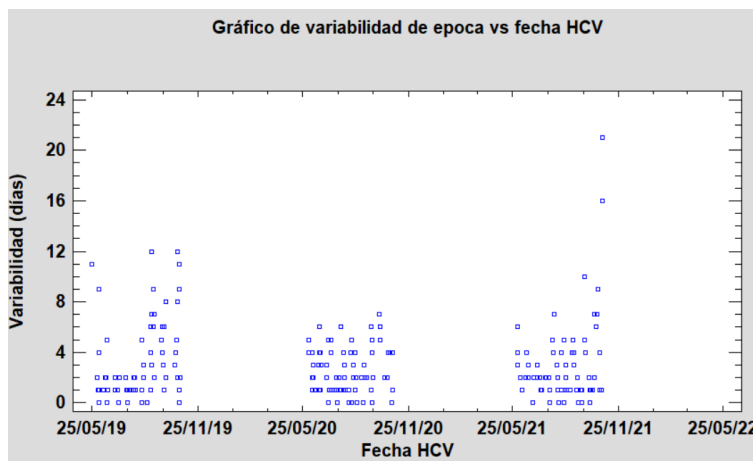


Figura 18. Variabilidad de fechas entre el dato de campo y el índice espectral.

3.2.3 Obtención de estadísticos de índices espectrales.

Una vez que se tiene una serie temporal depurada y que los datos de humedad de combustible vivo tomados en campo, han sido unidos a los índices espectrales extraídos de imágenes satelitales, se pueden calcular estadísticos de los índices espectrales, que

servirán como una variable más al momento de estimar la humedad de combustible vivo. Su objetivo es caracterizar el máximo, mínimo, media, rango de los índices espectrales, en cada parcela, a lo largo de la serie temporal. También, se estudió aplicar un estadístico que midiera, de forma relativa, la variación del índice respecto de su máximo y mínimo, de la siguiente forma.

$$\text{Índice relativo} = \frac{\text{índice}_i - \text{min}}{\text{max} - \text{min}} \quad (8)$$

Donde max es el valor máximo del índice sobre una parcela determinada en su línea temporal y min es el valor mínimo del índice sobre una parcela determinada en su línea temporal. En la figura 19 se observa como ejemplo al índice NDVI, frente a sus respectivos estadísticos.

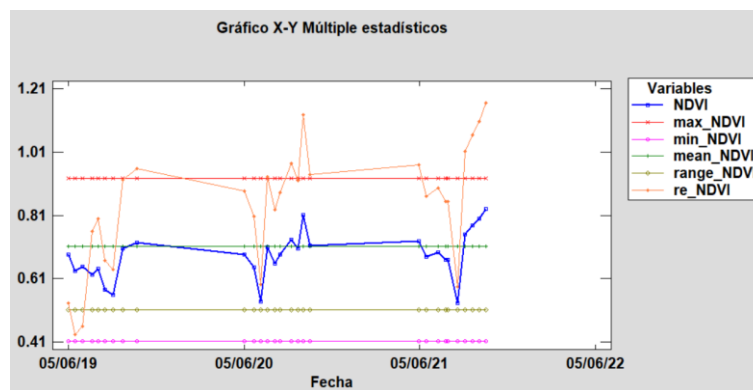


Figura 19. Estadísticos del índice espectral NDVI.

3.3 Métodos de regresión

3.3.1 Análisis de regresión lineal múltiple

En este apartado, se describe la aplicación del análisis de regresión múltiple, cuyo objetivo es encontrar la relación existente entre la variable dependiente, representada por la humedad de combustible vivo de la vegetación (HCV), con las variables independientes, las cuales están compuestas de índices espectrales y sus respectivos estadísticos, las variables meteorológicas y otras variables descritas en el apartado de Datos. Se busca que, a través de la selección de variables independientes, se encuentre un modelo capaz de describir la relación existente. En este tipo de regresiones, la descripción de la relación entre variables dependientes e independientes viene dada por medio de una ecuación.

Para el entrenamiento de un modelo de regresión, las variables independientes o predictoras, en este caso representadas por índices espectrales y parámetros meteorológicos, se pretende que expliquen el comportamiento de la variable dependiente, representada por la HCV.

En los modelos de regresión lineal múltiple, la ecuación final puede incluir todas las variables que se han de introducir; sin embargo, esto puede dar paso a la existencia de redundancias, produciendo que las variables introducidas tengan multicolinealidad, es decir que dos o más variables sean muy parecidas, aportando poco o reproduciendo errores en la estimación final. Además, el modelo resultaría muy extenso y de difícil interpretación. El mejor criterio será considerar como más robusto, un modelo de menor cantidad de variables con representatividad de diferentes fuentes y un ajuste aceptable (Pons, y otros, 2014).

En este estudio, el método que se usó para seleccionar las variables fue *Forward* o paso a paso hacia delante, método automatizado por pasos, que utiliza como criterios el estadístico de valor-P para incluir o excluir una variable. El valor-P ayuda a diferenciar resultados que han derivado del azar del muestreo, de resultados que estadísticamente son significativos. Los métodos automáticos de paso a paso ayudan a calcular varios modelos, con fines predictivos, que favorece el criterio del investigador a determinar la aplicación de las variables predictoras (Berlanga-Silvente & Vilá-Baños, 2014). Para la aplicación de esta metodología se usó el *software Statgraphics Centurion*.

También, en varios modelos, se aplicó el método *Backward* o paso a paso hacia atrás. En ambos métodos, las variables eliminadas pueden ser introducidas en un paso posterior, si se requieren como predictores y, las variables introducidas pueden ser eliminadas posteriormente, si no aportan suficiente significatividad al modelo. Es erróneo asumir, en este punto, que las variables predictoras con P-valor por encima de 0.05, deben eliminarse. Debido a un alta multicolinealidad entre algunas de las variables predictoras, los valores-P pueden cambiar drásticamente, si alguna de las variables se elimina del modelo (STATGRAPHICS, 2020). Como se puede ver en la Figura 20, para este trabajo se usó un umbral de 0.05 para el valor P, tanto para agregar variables, como para quitar variables.

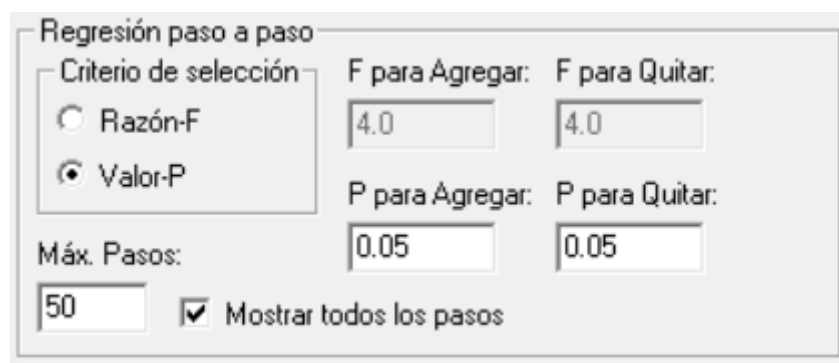


Figura 20. Configuración de umbral del valor-P para la selección paso a paso hacia delante.

3.3.2 Variables utilizadas en el análisis de regresión lineal

Como se ha mencionado en el apartado de Datos se calcularon las variables independientes que constan de 22 índices espectrales y a cada uno se le añadieron los estadísticos de media, máximo, mínimo, rango e índices espectrales relativos. También se consideraron las variables meteorológicas de temperatura, precipitación, humedad relativa, radiación solar y velocidad del viento. Por último, se han considerado también como variables predictoras el seno y coseno de *DOY*, donde *DOY* es el día del año estandarizado entre π y $-\pi$. Sin embargo, al momento de realizar la regresión en algunos casos no se pudo hacer uso de todas estas variables, ya que el número de observaciones de humedad de combustible vivo se ve reducido porque los datos de campo de HCV no están disponibles en todas las parcelas con periodicidad quincenal durante el periodo desde junio a octubre de los años 2019, 2020 y 2021.

En la Tabla 6 se muestran las parcelas usadas para el entrenamiento, divididas por provincias y por modelo de combustible. Se puede apreciar que Valencia cuenta con la mayor cantidad de parcelas teniendo un total de 28 parcelas, le sigue Castellón con un total de 18 parcelas y por último está Alicante con un total de 16 parcelas. Aunque la numeración más alta de las parcelas más es 88, varias de estas fueron descartadas ya que la categoría de modelo de combustible no tenía suficiente representatividad de observaciones por lo cual no se podría trabajar modelos de regresión lineal (ver apartados de Datos).

Tabla 6. *Clasificación de parcelas de estudios divididas por provincias y modelo de combustible.*

Entrenamiento	Mod_com	Total Parcelas	Parcelas
Valencia	SH4	15	2-5-6-7-13-17-19-20-21-24-27-28-32-35-41
	TU2	6	8-11-23-29-30-31
	TU3	7	1-14-25-36-38-42-78
Castellón	SH4	7	53-56-59-61-63-64-68
	TU2	7	43-44-46-51-57-66-67
	TU3	4	52-55-58-60
Alicante	SH4	9	71-72-74-75-76-77-86-87-88
	TU2	6	69-73-82-83-84-85
	TU3	1	79

Cada una de estas parcelas tiene una cantidad de observaciones diferente dependiendo del trabajo realizado en campo entre el 2019 a 2021. En la Tabla 7 se puede ver la disponibilidad, por ejemplo, el municipio de Gilet tiene datos de junio del 2019 a octubre del 2021. Por otro lado, el municipio de Bernia cuyo muestreo de humedad de

combustible vivo se realizó más tarde, solo se tiene disponibilidad de datos para el periodo de octubre del 2020 a octubre del 2021.

Tabla 7. *Disponibilidad de datos en cada municipio de la Comunitat Valenciana*

Municipio	Disponibilidad
Gilet	Desde junio a octubre, años 2019, 2020 y 2021
Bétera	Desde junio a octubre, años 2019, 2020 y 2021
Chelva	Desde junio a octubre, años 2019, 2020 y 2021
Llombai	Desde junio a octubre, años 2019, 2020 y 2021
Buñol	Desde junio a octubre, años 2019, 2020 y 2021
Cortes	Desde junio a octubre, años 2019, 2020 y 2021
Gandía	Desde junio a octubre, años 2019, 2020 y 2021
Montanejos	Desde julio a octubre, años 2020 y 2021
Eslida	Desde agosto a octubre, años 2020 y 2021
Morella	Desde agosto a octubre, años 2020 y 2021
Sant_Mateu	Desde agosto a octubre, años 2020 y 2021
Torre_Macanes	Desde agosto a octubre, años 2020 y 2021
Vall_Gallinera	Desde septiembre a octubre, años 2020 y 2021
Biar	Desde septiembre a octubre, años 2020 y 2021
Bernia	Desde octubre del 2020 a octubre del 2021

De este modo, la disponibilidad varía dependiendo la fecha de toma de datos de HCV en cada provincia, este factor afecta directamente al entrenamiento de modelos de regresión y la aplicabilidad del número de variables predictoras que se puede usar por cada modelo, ya que, para realizar una regresión múltiple el número de variables más la constante debe superar al número de observaciones, de tal modo que al hacer una selección que no tenga

el número suficiente de observaciones respecto de las variables que se prueban, el modelo no podrá ser calculado.

Cabe mencionar que el presente trabajo se ha centrado en obtener estimaciones de HCV para los meses desde junio hasta octubre de los años 2019, 2020 y 2021, con el objetivo de obtener predicciones del contenido de HCV en el periodo de mayor riesgo de incendios. Sin embargo, la base de datos de campo cuenta también con datos desde noviembre de 2020 hasta mayo de 2021 y también en noviembre de 2021. En la Figura 21 se presenta un mapa con el número de observaciones disponibles en cada parcela para los años del 2020 al 2021. En este mapa se puede apreciar el comportamiento de la toma de datos en campo. Los puntos de color rojo representan mayor cantidad de observaciones y la mayoría se concentran en los municipios de la provincia de Valencia. Por otro lado, los puntos de color verde indican menos cantidad de observaciones y, estos se concentran en su mayoría sobre la provincia de Alicante. De esta forma, en algunos modelos con bajo número de datos de campo en las provincias de Castellón y Alicante se han usado también datos en el periodo desde junio de 2020 hasta noviembre de 2021 para construir los modelos, centrando la validación de los mismos en los meses desde junio hasta octubre objeto de este trabajo.

En este mapa no se muestran las observaciones del año 2019 ya que en el año 2019 solo se tomaron muestras de la provincia de Valencia y solo durante el periodo desde junio hasta octubre.

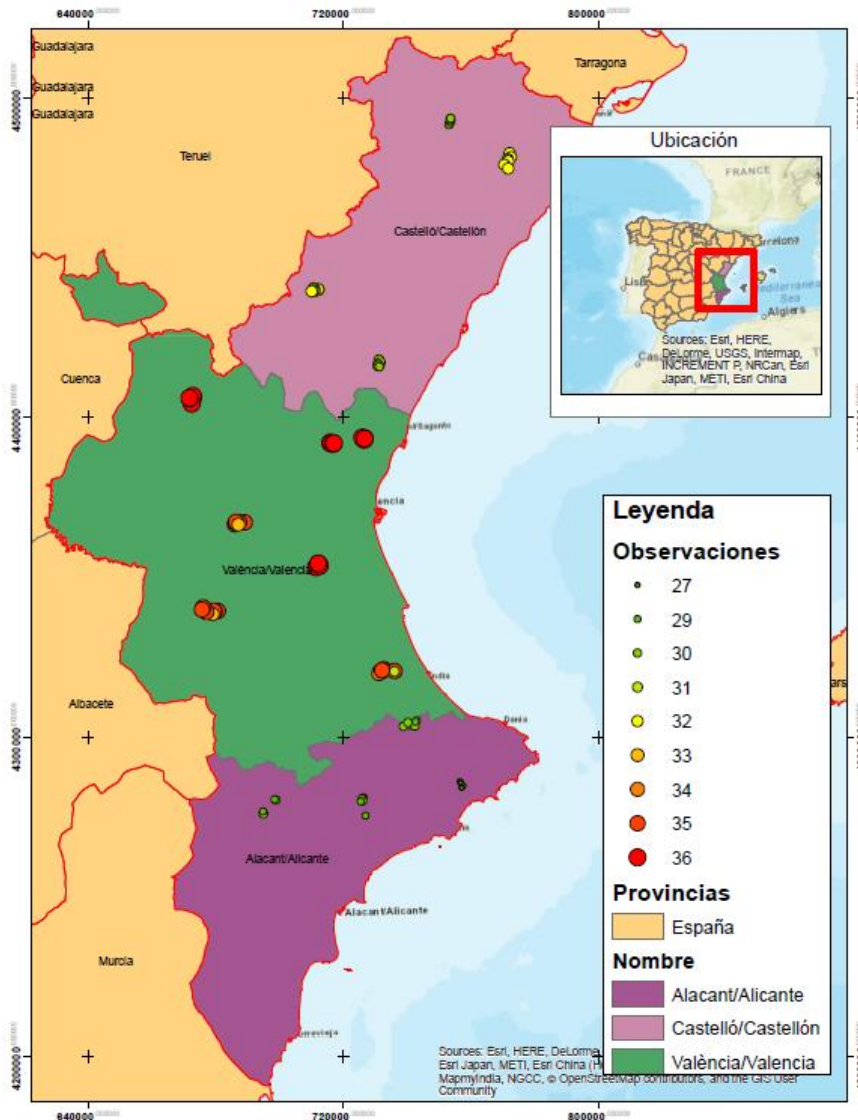


Figura 21. Mapa de la disponibilidad de los datos de campo.

3.4 Desarrollo del análisis de regresión lineal múltiple.

El análisis de diferentes modelos se realizó considerando la división por provincias y las categorías de modelo de combustible. En cada modelo se intentó determinar la menor cantidad de variables que tengan representatividad de diferentes fuentes y que alcancen un R-cuadrado ajustado bueno. Hay que mencionar que estos son modelos posibles y puede que algunas variables no sean significativamente representativas en el modelo.

La evaluación del modelo se realizó con el estadístico R-cuadrado ajustado, el cual mide la bondad del ajuste o el porcentaje de variación de la variable dependiente, en nuestro caso la humedad de combustible vivo, que es explicado por las variables predictoras o independientes. También se calculó la raíz cuadrada del error cuadrático medio (RMSE, por sus siglas en inglés). Si bien es cierto que un modelo puede mejorar con el uso de

muchas variables independientes reflejando un R-cuadrado ajustado más alto, puede ser que dos o más variables independientes muestren una multicolinealidad lo que hace que no sean representativas para el modelo. Por este motivo, fue necesario analizar la calidad de aporte que las variables independientes presentan al modelo.

3.4.1 Regresión lineal paso a paso hacia adelante

El método de selección paso a paso hacia adelante es un método iterativo que comienza con la constante del modelo y va introduciendo en cada paso la variable estadísticamente más significativa. También puede eliminar una variable del modelo si ella ya no es estadísticamente significativa después de introducir nuevas variables. En este trabajo se ha usado el criterio del Valor-P para introducir o eliminar variables, usando un umbral de 0.05, estableciendo como criterio de parada cuando no exista ninguna variable con Valor-P < 0.05 para ser introducida en el modelo. En la Tabla 8 se observan los resultados de aplicar la regresión múltiple paso a paso hacia adelante: como variable dependiente se usó la humedad de combustible vivo y como variables independientes los índices espectrales y estadísticos de las mismas más variables meteorológicas, dando un total de 116 variables predictoras. Este ejemplo muestra la aplicación de la metodología para la provincia de Valencia en el año 2020 considerando como variable dependiente la media ponderada de la humedad de combustible de todas las especies presentes en cada parcela, usando como pesos para esa media ponderada la fracción de cabida cubierta de cada especie en dicha parcela.

Se puede ver que el método de paso a paso hacia adelante ha seleccionado 8 variables predictoras representativas para este modelo que son: sin_DOY, MinTemperature, Precipitación, p7, p3, p60, NMDI y Vgreen.

Para determinar si el modelo puede simplificarse, se analizó el valor-P más alto de las variables independientes que es 0.0470 y corresponde a MinTemperature, puesto que el valor-P es menor que 0.05, ese término es estadísticamente significativo con un nivel de confianza del 95.0%. De este modo se podría decir que, en el modelo creado con 8 variables, todas estas son significativas y no habría necesidad de quitar ninguna.

El estadístico R-Cuadrado indica que el modelo así ajustado explica 65.2326% de la variabilidad de la media ponderada de HCV. El estadístico R-Cuadrado ajustado, que es más apropiada para comparar modelos con diferente número de variables independientes, es 63.26%. El error estándar del estimado muestra que la desviación estándar de los residuos es 6.45251, el cual coincide con la raíz cuadrada del error cuadrático medio (RMSE por sus siglas en inglés). Además, puesto que el valor-P en la tabla ANOVA es menor que 0.05, existe una relación estadísticamente significativa entre las variables con un nivel de confianza del 95.0%.

Tabla 8. Regresión múltiple paso a paso hacia adelante en la provincia de Valencia en el año 2020

<i>Parámetro</i>	<i>Estimación</i>	<i>Error Estándar</i>	<i>Estadístico T</i>	<i>Valor-P</i>
CONSTANTE	26.2954	10.6377	2.47191	0.0146
sin_DOY	-10.5027	1.84463	-5.69367	0
MinTemperature	0.316421	0.157905	2.00388	0.047
Precipitation	1.6732	0.535654	3.12367	0.0022
p7	0.600821	0.0873314	6.87979	0
p3	-0.569073	0.134309	-4.23704	0
p60	0.0636204	0.0183436	3.46827	0.0007
NMDI	73.9442	16.9127	4.37212	0
Vgreen	-49.7724	21.6071	-2.30352	0.0227

Análisis de la Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	11014.6	8	1376.83	33.07	0
Residuo	5870.52	141	41.6349		
Total (Corr.)	16885.1	149			

R-cuadrado = **65.2326** %

R-cuadrado (ajustado para g.l.) = **63.26** %

Error estándar del est. = **6.45251**

En la Figura 22 se presenta la gráfica de número de filas versus residuos, puesto que gráficamente no se encuentra un patrón que nos indique que la varianza de residuo se ve condicionada por el número de fila en el que las observaciones fueron hechas, se puede decir que el modelo tiene una independencia serial.



Figura 22. Gráfico número de filas versus residuos

La salida de la regresión muestra la ecuación resultante de aplicar la selección paso a paso hacia adelante para predecir la HCV y 8 variables independientes. La ecuación del modelo ajustado es:

$$\text{HCV} = 26.2954 - 10.5027 \cdot \sin_DOY + 0.316421 \cdot \text{MinTemperature} + 1.6732 \cdot \text{Precipitation} + 0.600821 \cdot p7 - 0.569073 \cdot p3 + 0.0636204 \cdot p60 + 73.9442 \cdot \text{NMDI} - 49.7724 \cdot V_{\text{green}}$$

En la Figura 23, se comparan los valores observados en campo con la predicción efectuada usando el modelo calculado a partir de la regresión múltiple para la parcela 2 en el año 2020, el modelo sigue la tendencia de los valores de HCV, sobre todo en los datos pertenecientes a junio a septiembre, pero para el mes de octubre se ve un desfase entre el valor observado y el valor pronosticado, a pesar de que los valores predichos marcan la tendencia de los observados en campo.

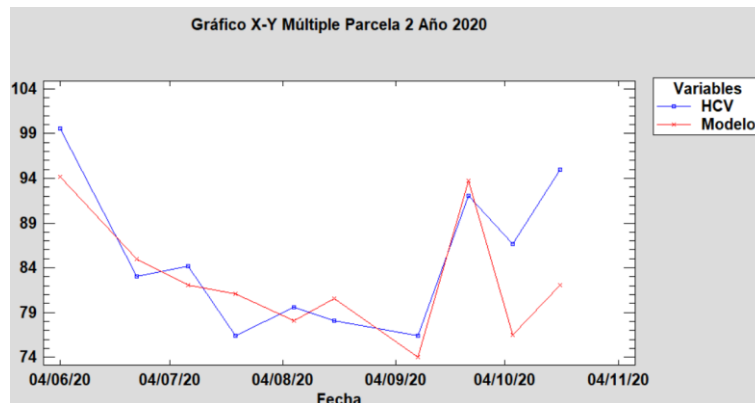


Figura 23. Estimación de la HCV en la Parcela 2 Año 2020

En la Tabla 9 se muestra el coeficiente de correlación entre las variables independientes del modelo de regresión, el cual puede ser usado para efectuar un análisis de multicolinealidad para las variables usadas en el modelo donde: 1 = sin_DOY, 2 = MinTemperature, 3 = Precipitación, 4 = p7, 5 = p3, 6 = p60, 7 = NMDI y 8 = Vgreen. Se encuentra una correlación estimada de -0.76 entre las variables p7 y p3. Estas dos variables al referirse a la precipitación acumulada 7 días y 3 días previos, es probable que sus datos sean muy parecidos aportando la misma cantidad de información a este modelo en específico. También se encuentra una correlación estimada de 0.58 entre las variables sin_DOY y p60. Aunque este valor es más bajo, analizando la naturaleza de las variables se podría decir que no son parecidas y que es muy poco probable que estas variables juntas presenten multicolinealidad.

Tabla 9. *Coefficiente de correlación entre las variables independientes.*

	1	2	3	4	5	6	7	8
1	1	0.368	0.1033	-0.287	0.3695	0.5873	0.0623	0.4311
2	0.368	1	0.0601	-0.0451	0.1748	0.3087	0.2074	-0.0102
3	0.1033	0.0601	1	0.0703	-0.1589	0.1508	0.026	-0.0793
4	-0.287	-0.0451	0.0703	1	-0.7682	-0.2577	-0.0509	-0.068
5	0.3695	0.1748	-0.1589	-0.7682	1	0.1677	0.0124	0.1289
6	0.5873	0.3087	0.1508	-0.2577	0.1677	1	-0.1788	-0.006
7	0.0623	0.2074	0.026	-0.0509	0.0124	-0.1788	1	0.083
8	0.4311	-0.0102	-0.0793	-0.068	0.1289	-0.006	0.083	1

Con el método de regresión lineal paso a paso hacia adelante se ha creado un primer modelo que nos ayuda a analizar la correlación de las variables independiente con la variable de humedad de combustible vivo. A partir de este análisis se puede determinar las variables independientes que van a influir en el modelo teniendo cuidado de que se mantenga una representatividad, tanto de variables con origen meteorológico, como variables que tengan origen en los índices espectrales.

3.4.2 Selección de modelos de regresión (R-cuadrado ajustado, Cp de Mallows y criterios de información)

Para la selección de modelos de regresión se utiliza la humedad de combustible vivo como variable a predecir, y como variables predictoras se usa el conjunto de variables resultantes de la regresión múltiple, que fue realizada previamente. Es importante que la selección de observaciones sean las mismas en este caso, la provincia de Valencia en el año 2020 y la media ponderada de las especies existentes.

Este método consiste en el análisis de los estadísticos R-cuadrado ajustado, Cp Mallows y criterios de información resultantes, resultantes de realizar diferentes regresiones con combinaciones de las variables independientes ingresadas, de tal forma que, dependiendo de la combinación realizada, el modelo resultante cambia y sus estadísticos de ajuste indicaran cuál de estas combinaciones representa una mejora para el modelo. También con este método se puede analizar la posibilidad de reducir el número de variables independientes que están presentes en la ecuación final, dependiendo de cuanto aportan al ajuste del modelo.

Ya que en el modelo de regresión con el uso de las variables: sin_DOY, MinTemperature, Precipitación, p7, p3, p60, NMDI y Vgreen, en total 8 variables independientes, se ha alcanzado un R-cuadrado ajustado de 63.26%, para la selección de modelos de regresión se intenta disminuir el número de variables y con ello la complejidad de la ecuación final, intentando siempre tener un buen ajuste lo más cercano posible al R-cuadrado ajustado de las 8 variables.

3.4.2.1 R-cuadrado ajustado

Este estadístico representa el coeficiente de determinación ajustado y mide la variabilidad de HCV que es explicada por el modelo. En la Tabla 10 se aprecian los resultados de aplicar la selección de modelos usando las 8 variables. Al analizar la columna de R-cuadrado ajustado, el estadístico disminuye poco a poco según el número de variables ingresadas y su combinación, al hacer uso de 5 variables independientes el R-cuadrado ajustado disminuye a 59.61%, siendo este valor muy cercano a 63.26% que usa la totalidad de variables ingresadas ABCDEFGH; siguiendo el principio de robustez de un modelo, el uso de 5 variables con la combinación ADEFG sería apropiado lo que se debe contrastar con otros criterios de selección de variables.

A=sin_DOY B=MinTemperature C=Precipitación D=p7 E=p3
 F=p60 G=NMDI H=Vgreen

Tabla 10. Selección de modelos usando 8 variables.

<i>CME</i>	<i>R-cuadrado</i>	<i>R-cuadrado Ajustado</i>	<i>Cp</i>	<i>VARIABLES Incluidas</i>
41.6349	65.2326	63.26	9.0	ABCDEFGH
42.519	64.2425	62.4798	11.0155	ACDEFGH
42.8975	63.9243	62.1459	12.3062	ABCDEF
43.7396	62.957	61.4028	14.2289	ACDEF
44.2026	62.8267	60.9942	16.7573	ABDEFGH
44.8352	62.0292	60.436	17.9917	ACDEGH
44.8573	62.0104	60.4164	18.0679	ADEFGH
44.8686	62.2666	60.4065	19.0289	ABCDEGH
45.1319	61.7779	60.1742	19.0109	ABDEF
45.764	60.9716	59.6164	20.281	ADEFG
46.0193	60.7538	59.3911	21.164	ACDEG
46.072	60.9817	59.3446	22.2399	ABCDEG
46.5037	60.3407	58.9636	22.8394	ADEGH
46.6054	60.806	58.8739	24.9525	ABCDFGH
47.4177	59.2804	58.1571	25.1395	ADEG
47.4981	59.4926	58.0861	26.2789	ABDEG
48.5037	58.6351	57.1988	29.7567	ABDFG
50.0052	57.0584	55.8738	34.1509	ADFG
50.5154	56.6203	55.4236	35.9278	ADEF
52.3144	55.0753	53.836	42.1933	ACDG
52.3655	55.0315	53.791	42.3712	ABDG
52.6749	54.4539	53.518	42.7137	ADG
54.3829	52.977	52.0108	48.7031	ADF
54.4696	52.902	51.9342	49.0073	ADE
58.7511	49.1999	48.1561	64.0211	AFG
59.0082	48.9777	47.9293	64.9224	ADH

59.7252	48.0039	47.2965	66.8714	AD
63.7541	44.4965	43.7413	81.0961	DF
64.5201	43.8296	43.0654	83.8005	AF
64.8532	43.5396	42.7715	84.9766	AG
67.5487	41.1929	40.3928	94.4939	FG
72.2764	36.649	36.2209	110.922	F
75.9684	33.4129	32.963	124.046	A

En la Figura 24, se aprecia la gráfica de R-cuadrado ajustado para la humedad de combustible vivo versus el número de variables aplicadas. Se puede ver que el aumento del estadístico se va reduciendo a medida que ingresan variables al modelo. A partir de la aplicación de las 5 variables ADEFG la curva tiende a estabilizarse alrededor del 60%.

Cabe mencionar que las 8 variables ingresadas en este método de selección del modelo de regresión han pasado previamente por el método de selección paso a paso hacia delante de la regresión múltiple, por tal motivo la R cuadrada ajustada llega a ser alta con el uso de las 8 variables independiente.

Se usa este método para considerar las posibilidades de reducir variables en la ecuación final haciendo que el modelo cumpla con un criterio de robustez, también analizar las posibles combinaciones de variables que se pueden dar para alcanzar un buen ajuste, teniendo representatividad de variables meteorológicas y variables de índices espectrales.

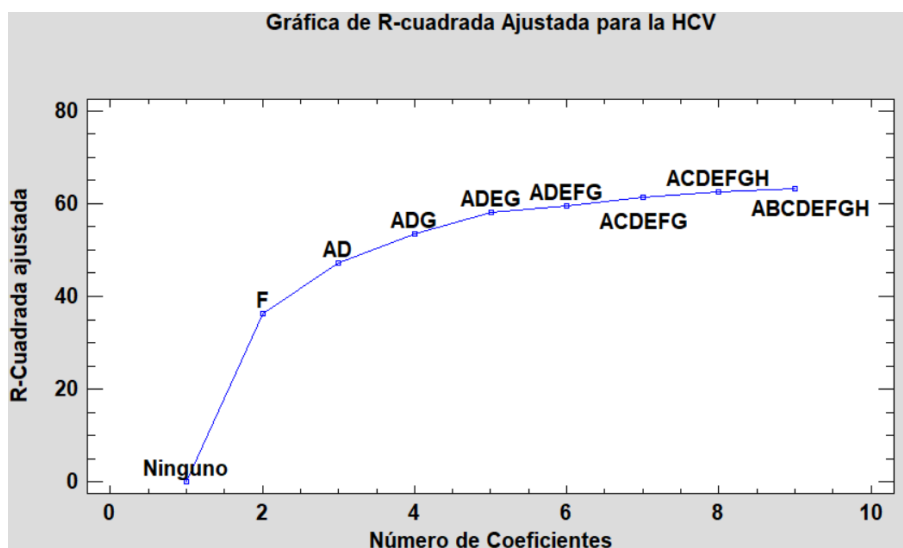


Figura 24. Gráfica de R-cuadrado ajustado para la humedad de combustible vivo versus el número de variables aplicadas

3.4.2.2 Cp de Mallows

Cp es una medida del sesgo en el modelo, basada en la comparación entre el cuadrado medio del error total y la varianza del error verdadero. Modelos sin sesgo tienen un valor

esperado de aproximadamente $p+1$, en donde p es el número de coeficientes en el modelo ajustado, en este caso se tiene 8 variables independientes. Se debe tomar en cuenta modelos con Valores de C_p cercanos a $p+1$. La Tabla 11 muestra los modelos con su valor de C_p , como se puede ver el modelo con 8 variables tiene un valor de C_p de 9 muy cercano a p , según el criterio C_p de Mallows este sería el mejor modelo disponible. El modelo creado con 5 variables independientes tiene un C_p de 20.28 un valor muy alejado de p y una R -cuadrado ajustado de 59.61%.

A=sin_DOY B=MinTemperature C=Precipitación D=p7 E=p3
 F=p60 G=NMDI H=Vgreen

Tabla 11. Modelos con su valor de C_p de Mallows

<i>CME</i>	<i>R-cuadrado</i>	<i>R-cuadrado Ajustado</i>	<i>Cp</i>	<i>Variables Incluidas</i>
41.6349	65.2326	63.26	9.0	ABCDEFGH
42.519	64.2425	62.4798	11.0155	ACDEFGH
42.8975	63.9243	62.1459	12.3062	ABCDEF
43.7396	62.957	61.4028	14.2289	ACDEF
44.2026	62.8267	60.9942	16.7573	ABDEFGH
44.8352	62.0292	60.436	17.9917	ACDEGH
44.8573	62.0104	60.4164	18.0679	ADEFGH
45.1319	61.7779	60.1742	19.0109	ABDEF
44.8686	62.2666	60.4065	19.0289	ABCDEGH
45.764	60.9716	59.6164	20.281	ADEFG
46.0193	60.7538	59.3911	21.164	ACDEG
46.072	60.9817	59.3446	22.2399	ABCDEG
46.5037	60.3407	58.9636	22.8394	ADEGH

En la Figura 25 se muestra la gráfica de distribución de C_p frente al número de coeficientes. La línea en color azul representa a p y se puede ver que modelos con el uso de 5 o más variables está muy cercano a la línea p y modelos con menos de 5 variables se alejan considerablemente de la línea p . Al analizar esta grafica se puede decir que, aunque el modelo ADEFG no cumple estrictamente con el criterio de C_p de Mallows, donde C_p debe ser muy próximo a p , si es un modelo que no se aleja demasiado de este criterio comparándolo con otros modelos de menos variables.

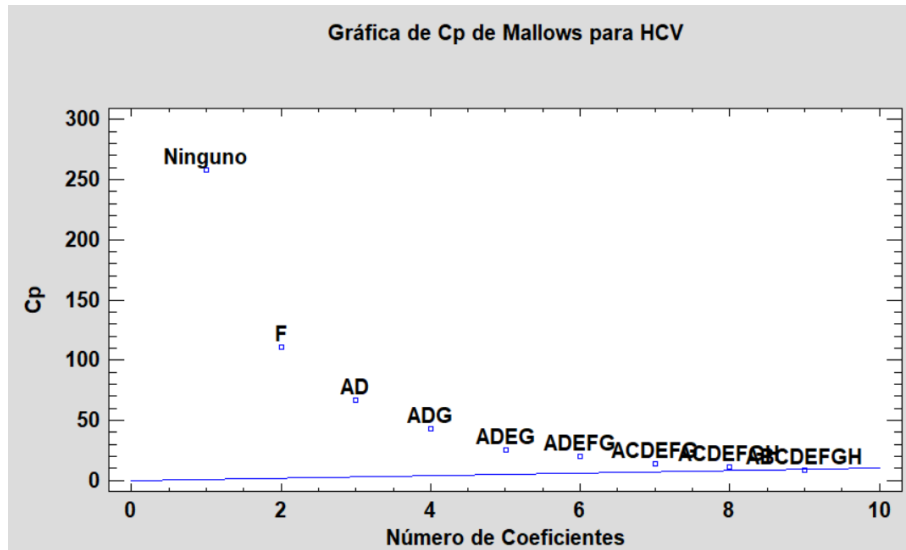


Figura 25. Gráfica de distribución de Cp frente al número de coeficientes

3.4.2.3 Mejor criterio de información

Ya que con los estadísticos de R-cuadrado ajusta y de Cp Mallows se ha escogido un modelo que tenga 5 variables ahora se debe estudiar la combinación de variables que sea mejor, para esto se usa los criterios de información siguientes: Criterio de información de Akaike (AIC), Criterio de Hannan-Quinn (HQC) y Criterio de Información Schwarz-Bayesian (SBIC). La diferencia de estos criterios radica en el uso de la penalización diferente para el número de parámetros estimados ($k+1$) teniendo en cuenta el número de datos (n) (Balaguer Beser & Ruiz Fernández, 2021). Las fórmulas usadas por Statgraphics Centurion se explican a continuación:

- Criterio de información de Akaike (AIC): es una medida de la calidad relativa de un modelo estadístico descrito por:

$$AIC = 2Ln(RMSE) + \frac{2(k + 1)}{n} \quad (9)$$

- Criterio de Hannan-Quinn (HQC):

$$HQC = 2Ln(RMSE) + \frac{2(k + 1)Ln(n)}{n} \quad (10)$$

- Criterio de Información Schwarz-Bayesian (SBIC):

$$SBIC = 2Ln(RMSE) + \frac{(k + 1)Ln(n)}{n} \quad (11)$$

(Balaguer Beser & Ruiz Fernández, 2021)

En la Tabla 12 se puede ver la combinación de 5 variables con mejores valores para los

criterios AIC, HQC y SBIC, la mejor combinación de variables es ADEFG, que está compuesto por las variables independientes: sin_DOY, p7, p3, p60 y NMDI la segunda mejor combinación está dada por ACDEG compuesto por las variables sin_DOY, Precipitación, p7, p3 y NMDI, en las dos combinaciones presentadas existe la presencia de variables meteorológicas y variables de índices espectrales. Estas combinaciones u otras seleccionadas serán validadas más adelante, analizando el ajuste y la multicolinealidad existente entre las diferentes combinaciones.

A=sin_DOY B=MinTemperature C=Precipitación D=p7 E=p3
 F=p60 G=NMDI H=Vgreen

Tabla 12. *Tabla de mejor criterio de información.*

<i>MSE</i>	<i>Coefficientes</i>	<i>AIC</i>	<i>HQC</i>	<i>SBIC</i>	<i>Variables incluidas</i>
45.764	6	3.9035	3.95242	4.02392	ADEFG
46.0193	6	3.90906	3.95799	4.02949	ACDEG
46.5037	6	3.91953	3.96846	4.03996	ADEGH
47.4177	5	3.92566	3.96643	4.02602	ADEG
47.4981	6	3.94069	3.98962	4.06112	ABDEG
48.5037	6	3.96164	4.01056	4.08207	ABDFG
50.0052	5	3.97879	4.01956	4.07915	ADFG
50.5154	5	3.98894	4.02972	4.0893	ADEF
52.6749	4	4.01747	4.05009	4.09776	ADG
52.3144	5	4.02394	4.06471	4.12429	ACDG
52.3655	5	4.02492	4.06569	4.12527	ABDG
54.3829	4	4.04938	4.082	4.12967	ADF
54.4696	4	4.05098	4.08359	4.13126	ADE
58.7511	4	4.12664	4.15926	4.20693	AFG
59.7252	3	4.12975	4.15422	4.18997	AD
59.0082	4	4.13101	4.16363	4.21129	ADH
63.7541	3	4.19503	4.2195	4.25525	DF
64.5201	3	4.20698	4.23144	4.26719	AF
64.8532	3	4.21213	4.23659	4.27234	AG
67.5487	3	4.25285	4.27731	4.31306	FG
72.2764	2	4.30717	4.32347	4.34731	F
75.9684	2	4.35698	4.37329	4.39713	A
91.2409	2	4.54017	4.55648	4.58031	D
95.9761	2	4.59077	4.60707	4.63091	G
102.506	2	4.65659	4.6729	4.69673	E
113.323	1	4.74358	4.75173	4.76365	

3.4.3 Modelos lineales generales

En este apartado se genera la ecuación que describa la relación existente entre las variables independiente con la HCV. Se usan las combinaciones de variables resultantes del proceso de selección de modelos de regresión realizado en el apartado anterior. La combinación a usar es: sin_DOY, p7, p3, p60 y NMDI. La selección de observaciones es provincia de Valencia en el año 2020 y la media ponderada de las especies; misma que fue utilizada en los apartados anteriores.

La Tabla 13 indica la significancia estadística de cada factor conforme fue introducido al modelo. El valor-P más alto corresponde a la variable meteorológica p60; dado que el valor-P es menor que 0.05, ese término es estadísticamente significativo con un nivel de confianza del 95.0%. El estadístico R-cuadrado indica que el modelo ajustado, explica 60.9716% de la variabilidad de la HCV. El estadístico R-Cuadrado ajustado, el cual es más adecuado para comparar modelos con diferente número de variables independientes, es 59.6164%.

Tabla 13. *Significancia estadística de cada factor conforme fue introducido al modelo*

<i>Variable</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
sin_DOY	2091.98	1	2091.98	45.71	0.0000
p7	1846.19	1	1846.19	40.34	0.0000
p3	660.744	1	660.744	14.44	0.0002
p60	285.553	1	285.553	6.24	0.0136
NMDI	734.723	1	734.723	16.05	0.0001
Residuo	6590.01	144	45.764		
Total (corregido)	16885.1	149			

R-cuadrado = 60.9716 %

R-cuadrado (ajustado por g.l.) = 59.6164 %

Error estándar del est. = 6.76491

Error medio absoluto = 5.24484

En la Tabla 14 se incluyen los factores de inflación de varianza (V.I.F por sus siglas en inglés), los cuales pueden usarse para medir la extensión con la que están correlacionadas las variables predictoras entre ellas mismas. V.I.F's por encima de 10, que en este caso no existen, generalmente se considera que indican una seria multicolinealidad. Una seria multicolinealidad aumenta considerablemente la estimación del error de los coeficientes del modelo, en comparación con una muestra ortogonal. En el desarrollo de los diferentes modelos puede existir combinaciones de variables que presenten estadístico V.I.F mayor a 10, en estos casos se debe realizar un análisis multivariado de correlaciones de la combinación de variables, para analizar que variable puede retirarse del modelo, y como ello afecta al ajuste de R-cuadrado.

Tabla 14. *Variabes y sus factores de inflación de varianza (V.I.F)*

<i>Parámetro</i>	<i>Estimado</i>	<i>Error</i>			<i>V.I.F.</i>
		<i>Estándar</i>	<i>Límite Inferior</i>	<i>Límite Superior</i>	
CONSTANTE	47.0399	8.88004	29.4878	64.592	
sin_DOY	-10.6384	1.57347	-13.7485	-7.5283	1.78893
p7	0.578236	0.0910393	0.39829	0.758183	2.77869
p3	-0.515096	0.135561	-0.783042	-0.24715	2.80161
p60	0.0452466	0.0181136	0.00944368	0.0810494	1.87548
NMDI	69.2234	17.2764	35.0752	103.371	1.1364

La Tabla 15 presenta los residuos atípicos de todas las observaciones que tienen un residuo estudentizado mayor que 2 en valor absoluto. Los residuos estudentizados miden cuántas desviaciones estándar se desvía cada valor observado de HCV de un modelo ajustado utilizando todas las observaciones menos esa. En este caso, hay 5 residuos estudentizados mayores que 2, pero ninguno mayor que 3.

Tabla 15. *Residuos atípicos de todas las observaciones que tienen un residuo estudentizado mayor que 2*

<i>Fila</i>	<i>Y</i>	<i>Y Predicha</i>	<i>Residuo</i>	<i>Residuo Estudentizado</i>
1131	98.93	85.5111	13.4189	2.05
1549	66.59	82.0563	-15.4663	-2.34
5565	59.12	75.1218	-16.0018	-2.43
6317	100.43	81.16	19.27	2.97
8194	70.92	85.4722	-14.5522	-2.20

Para analizar la linealidad y homocedasticidad se analiza la Figura 26, donde se presenta un gráfico de HCV predicho versus el residuo estudentizado del modelo. Se observa que acepta la hipótesis de linealidad ya que los datos se encuentran por encima y por debajo de la recta $y=0$, aunque la homocedasticidad es discutible ya que los valores predichos no poseen la misma varianza.

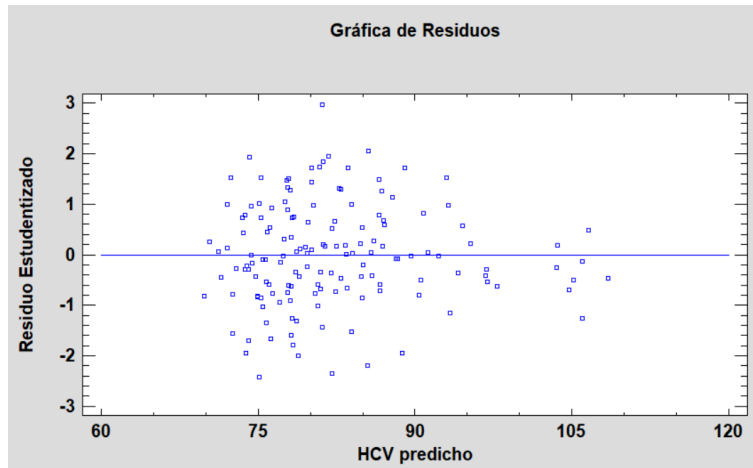


Figura 26. Gráfico de HCV predicho versus el residuo estudentizado del modelo

El resultado final de realizar este proceso es la ecuación que relaciona la HCV con las variables sin_DOY, p7, p3, p60 y NMDI:

$$\text{HCV} = 47.0399 - 10.6384 * \text{sin_DOY} + 0.578236 * \text{p7} - 0.515096 * \text{p3} + 0.0452466 * \text{p60} + 69.2234 * \text{NMDI}.$$

3.4.3.1 Uso de variable categórica en los modelos lineales generales

La aplicación de una variable categórica en los modelos lineales generales ayuda a que la ecuación final del modelo se ajuste a cada categoría contenida en la variable categórica a través del condicional “SI”, aunque al introducir esta nueva variable la R-cuadrado ajustada del modelo y la multicolinealidad de las variables independientes pueden verse afectadas. Por tal motivo es necesario una nueva revisión de estos estadísticos. Para este estudio las variables categóricas usadas son:

- Year: define los años de toma de datos de la HCV, puede ser 2019, 2020 o 2021
- Mod_comb: define la categoría del modelo de combustible de las especies de arbolado, puede ser por ejemplo TU2 o TU3

Como ejemplo, en la Tabla 16 se muestra la aplicación de la variable categórica en el resultado del modelo lineal general para la provincia de Alicante en los periodos de Junio – Octubre del 2021 usando la variable categórica Mod_comb. Se puede apreciar que el ajuste de R-cuadrado y demás estadísticos varía muy poco.

Tabla 16. *Aplicación de la variable categórica en el resultado del modelo lineal general*

Con variable categórica	Sin variable categórica
R-cuadrado = 65.139 %	R-cuadrado = 64.7975 %

R-cuadrado (ajustado por g.l.) =	R-cuadrado (ajustado por g.l.) =
64.4347 %	64.2668 %
Error estándar del est. = 9.58835	Error estándar del est. =
	9.61097
Error medio absoluto = 6.51624	Error estándar del est. =
	9.61097

La ecuación final contiene la nueva variable I1(1) la cual dependerá de la categoría de la variable Mod_comb para tomar un valor, la ecuación se presenta de la siguiente forma:

$$\text{HCV} = 21.719 + 2.81638 * \text{I1}(1) + 25.4285 * \text{ARVI} + 57.3299 * \text{min_RVI} - 187.41 * \text{min_SAVI}$$

donde I1(1) = 1 si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario.

4. RESULTADOS

4.1 Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la provincia de Valencia

En esta sección se analizan los resultados de los modelos de regresión lineal multivariante, usando como variables independientes, tanto los índices espectrales como las variables meteorológicas y, como variable dependiente, el contenido de HCV usando los datos de campo de las parcelas de la provincia de Valencia, que han sido catalogadas como SH4 según los modelos de combustible de la Comunitat Valenciana. Los modelos presentados fueron construidos con 15 parcelas obtenidas en los periodos junio-octubre del 2019, 2020 y 2021, en varios municipios de la provincia como se muestra en la Tabla 17.

Tabla 17. Parcelas de matorral (SH4) de la provincia de Valencia

Municipio	Parcelas
Gilet	2, 5
Bétera	6, 7
Chelva	13
Llombai	17, 19
Buñol	20, 21, 24
Cortes de Pallás	27, 28, 32
Gandía	35, 41

Los modelos de estimación de HCV fueron realizados a partir de la regresión múltiple donde se seleccionan variables más influyentes, para luego realizar una selección de variables, usando el mejor criterio de información y el tipo de variable introducida. Seleccionadas las variables, se usa una regresión lineal multivariante para analizar el R-cuadrado ajustado resultante, así como el factor de inflación de la varianza (V.I.F.). Esta metodología fue descrita en el apartado anterior. Ya que los datos de campo permiten discriminar entre la media ponderada y media matorral de HCV, se usaron estas dos categorías para separar el análisis.

En la Tabla 18 se aprecian los resultados del estudio del año 2019. El R-cuadrado ajustado junto a las variables que intervienen y el modelo resultante, los modelos de media ponderada y media matorral, presentan un ajuste de R-cuadrado ajustado de 63.62% y 65.06% respectivamente, es decir, que en los dos casos los modelos resultantes pueden explicar más del 60% de la humedad de combustible vivo para las parcelas de la provincia de Valencia. El valor de la raíz del error cuadrático medio (RMSE) no supera un 9% de la HCV.

En estos modelos se ha procurado obtener la menor cantidad de variables posibles, que representen a los índices espectrales. Los dos modelos resultantes, presentan solamente 3 variables independientes, haciendo que los modelos sean más robustos. Así, se encontró que, la variable meteorológica T30 es un buen predictor de la HCV, junto a los índices espectrales de AVI, NMDI y MSAVI2. En la Figura 27 se puede apreciar que los valores observados versus predichos, tanto la media de matorral como la media ponderada se acercan mucho a la diagonal y solo hay un sesgo en algunas observaciones de valores más altos.

Tabla 18. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2019.

Año	Sp	R2 ajust/ RMSE	Variables	Modelo
2019	Media Ponderada	63.61/ 7.83	T30 AVI NMDI	$HCV = 74.5516 - 2.49306 * t30 + 104.075 * AVI + 75.364 * NMDI$
	Media Matorral	65.06/ 8.93	T30 Range_NMDI MSAVI2	$HCV = 78.3198 - 3.01569 * t30 - 105.698 * range_NMDI + 124.641 * MSAVI2$

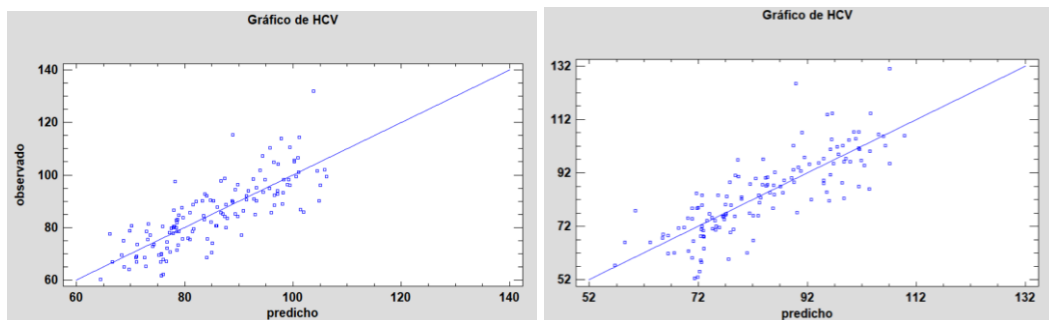


Figura 27. Gráfico de observado vs predicho, Valencia 2019. Media ponderada (izquierda), media matorral (derecha)

En el siguiente estudio (Tabla 19), se analiza el año 2020, en el cual se encuentran valores de R-cuadrado ajustado, más bajos en relación con el año anterior. Para el modelo de la media ponderada, el R-cuadrado ajustado es de 59.62% y en este caso fue necesario incluir 5 variables independientes, entre las cuales, en mayor cantidad corresponden a variables meteorológicas, como los son p3, p7, p60; variables que miden la precipitación en la parcela. También, se encontró que, solamente el índice espectral NMDI, fue incluido en el modelo y se incluye la variable sin_DOY, que representa al día normalizado del

año. Para el modelo proveniente de la media matorral, el R-cuadrado ajustado es algo más bajo, siendo 54.20% y, aunque en este modelo existen 4 variables independientes, se puede apreciar que las variables meteorológicas p7, p60 siguen siendo de gran importancia para predecir la humedad de combustible vivo. A esto se le suma el índice espectral AVI y el estadístico de índice espectral range_NDVI. En los dos modelos el valor de RMSE no supera el 9% de HCV. En la Figura 28 la gráfica de observados versus predichos de la media matorral y ponderada se acercan a la diagonal. En este caso, nuestros modelos no son capaces de predecir los valores más bajos de HCV.

Tabla 19. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2020.

Año	Sp	R2 ajust/ RMSE	Variables	Modelo
2020	Media Ponderada	59.61/ 6.76	sin_DOY p7 p3 p60 NMDI	$HCV = 47.0399 - 10.6384 * \sin_DOY + 0.578236 * p7 - 0.515096 * p3 + 0.0452466 * p60 + 69.2234 * NMDI$
	Media Matorral	54.20/ 8.71	P7 P60 AVI Range_NDVI	$HCV = 18.676 + 0.284751 * p7 + 0.107351 * p60 + 206.274 * AVI - 36.4453 * range_NDVI$

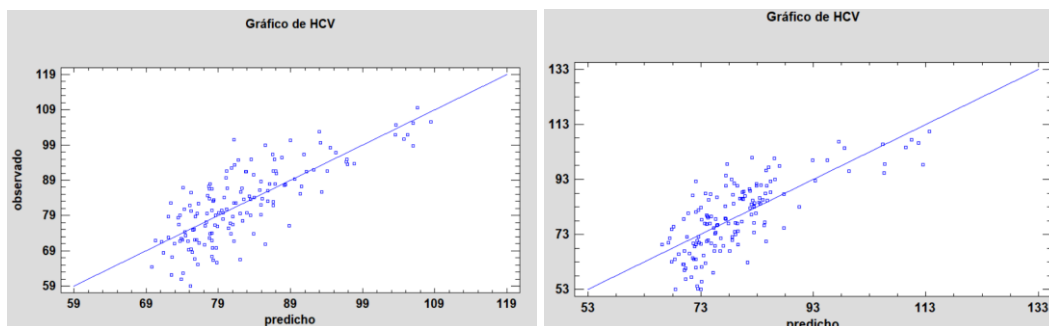


Figura 28. Gráfico de observado vs predicho, Valencia 2020. Media ponderada (izquierda), media matorral (derecha)

Los modelos de regresión lineal multivariante generados para los periodos del 2021 (Tabla 20), muestran ajustes de R-cuadrado más bajos respecto al año anterior. Para estos modelos se alcanzó un valor de 50.06% y 51.59% para modelos de media ponderada y media matorral respectivamente; sin embargo, para este año se puede apreciar que las variables espectrales tienen mayor importancia en los modelos. Para la media ponderada

se incluyen cinco variables, de las cuales dos son los índices espectrales SAVI y VARI acompañadas de una variable meteorológica, que es p60 y dos estadísticos de índices espectrales correspondientes a max_EVI y max_NDMI. El modelo de media matorral, para la cual, el R-cuadrado ajustado es muy parecido, cambia de variables y deja al modelo con 4 variables en total; de estas, los índices espectrales siguen teniendo importancia con NDVI y Vgreen. También, se suma el estadístico de índice espectral max_MSI. En el modelo del año 2021 se puede apreciar que los valores de RMSE son más altos respecto a los años anteriores obteniendo 9.71% y 11.64 respectivamente. En la Figura 29 las dos graficas de observado versus predicho se puede apreciar que existen más valores que se alejan de la diagonal.

Tabla 20. Modelos de regresión en las parcelas de matorral (SH4) de la provincia de Valencia año 2021.

Año	Sp	R2 ajust/ RMSE	Variables	Modelo
2021	Media Ponderada	50.05/ 9.71	p60 SAVI VARI max_EVI max_NDMI	$HCV = 112.041 + 0.0606447 * p60 + 139.043 * SAVI + 78.4984 * VARI - 156.084 * max_EVI - 50.702 * max_NDMI$
	Media Matorral	51.58/ 11.64	P60 NDVI Vgreen Max_MSI	$HCV = -28.9857 + 0.0747027 * p60 + 125.774 * NDVI + 135.028 * Vgreen + 69.1103 * max_MSI$

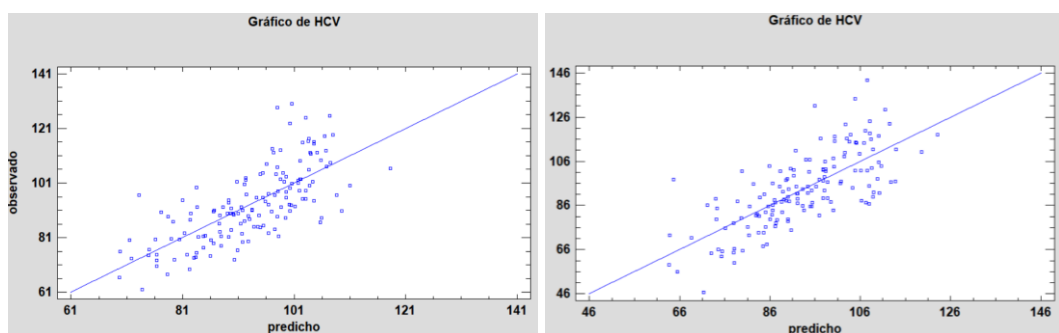


Figura 29. Gráfico de observado vs predicho, Valencia 2021. Media ponderada (izquierda), media matorral (derecha)

Después de analizar los periodos junio-octubre de los tres años por separado, se puede notar que los índices espectrales ayudan a la estimación de la HCV sobre la provincia de

Valencia, pues en todos los casos existe la inclusión de una o más de estas variables. También, se observa que los modelos resultantes no contemplan muchas variables, pues existen modelos que van desde el uso de 3 variables hasta un total de 5 variables, dependiendo del caso. También, se puede apreciar que el estadístico de R-cuadrado ajustado, se reduce en el transcurso de los años estudiados.

Luego de haber estudiado por separado los modelos de HCV, se realizó un análisis para la creación de una predicción, uniendo los años 2019, 2020 y 2021, estudiados anteriormente. Igual a lo realizado anteriormente, este análisis está categorizado por la media ponderada y media matorral, pero también se suma el aumento de una variable categórica llamada *Year*, la cual describe el año en el que se ha realizado la toma del dato de campo. Esta variable categórica, permitirá que el modelo final sea ajustado por año, mediante la aplicación de nuevas variables que dependerán de un condicional “SI” que estará en función del año de la toma del dato de campo de HCV.

Para la aplicación de la variable categórica, es necesario partir del modelo de regresión lineal multivariada, como se ha seguido en pasos anteriores. Una vez obtenida la ecuación y variables resultantes, se añade la variable categórica *Year* y se tiene cuidado de verificar que el R-cuadrado ajustado del modelo, sea parecido o mejor al modelo sin variable categórica y que las variables usadas sigan siendo significativas y no muestren multicolinealidad. Esto último revisando el estadístico V.I.F.

De esta forma, en la Tabla 21 se muestran los modelos resultantes para las fechas 2019, 2020 y 2021 en conjunto. Se obtuvo un R-cuadrado ajustado de 53.03% y 51.91% para la media ponderada y media matorral respectivamente. En dichos modelos intervienen 6 variables, entre las cuales se repite el *sin_DOY*, *p60*, *mean_VARI* y *VARI*, para la media ponderada. Las variables de índices espectrales son *ARVI* y *range_GNDVI* mientras que para la media de matorral intervienen *MCARI_OSAVI* y *range_NDVI*.

De la aplicación de la variable categórica *Year* se obtienen R-cuadrado ajustados de 56.15% y 54.75% respectivamente, siendo estos muy similares a los modelos de media ponderada y media matorral anterior. En estos modelos, se presenta un incremento de ajuste 3%, lo cual podría no ser significativo; sin embargo, como se lo mencionó anteriormente, las ecuaciones de los modelos presentan una mayor versatilidad al crear dos nuevas variables, llamadas, en este caso I1(1) y I1(2), que hacen que el modelo se adapte según el año de toma de datos de HCV 2019, 2020 o 2021 a través del condicional SI como se puede ver en la tabla. La efectividad de la aplicación de la variable categórica, será analizada más adelante, cuando se evalúen los modelos resultantes. Los valores de RMSE para los modelos de media ponderada no superan el 9%, mientras que para los modelos media matorral alcanzan el 11.12% de HCV.

Se puede evidenciar que, para la construcción de los modelos de media ponderada y media matorral en los años establecidos, la variable *VARI* y la variable meteorológica *p60*

ayudan a describir el comportamiento de la HCV sobre la provincia de Valencia. La variable VARI, al ser un índice que resalta la vegetación bajo un alto impacto atmosférico, se vuelve propicia para mitigar los posibles efectos atmosféricos que sufren las imágenes satelitales en diferentes fechas, haciendo que los resultados de los valores RGB se vean mucho más armonizados en el tiempo. Este puede ser uno de los motivos por los cuales, la variable VARI sirve para predecir en un modelo que junta diferentes años. En la Figura 30 muchos valores de observados versus predichos se alejan de la diagonal, especialmente en valores de HCV los observados más elevados.

Tabla 21. Modelos resultantes para las fechas 2019, 2020 y 2021 en conjunto

Años	Sp	R2 ajust/ RMSE	Variables	Modelo
2019- 2020- 2021	Media Ponderada	53.02/ 9.13	sin_DOY p60 ARVI VARI range_GNDVI mean_VARI	$HCV = 84.3116 - 10.973 * \sin_DOY + 0.0534445 * p60 + 49.2073 * ARVI + 73.8789 * VARI - 31.1276 * range_GNDVI - 120.325 * mean_VARI$
	Media Ponderada Categorica YEAR	56.15/ 8.82	sin_DOY p60 ARVI VARI range_GNDVI mean_VARI	$HCV = 83.6047 + 1.98065 * I1(1) - 3.60212 * I1(2) - 11.1681 * \sin_DOY + 0.0418284 * p60 + 53.8145 * ARVI + 81.4385 * VARI - 29.6629 * range_GNDVI - 133.308 * mean_VARI$ $I1(1) = 1$ si Year=2019, -1 si Year=2021, 0 de lo contrario $I1(2) = 1$ si Year=2020, -1 si Year=2021, 0 de lo contrario
	Media Matorral	51.91/ 11.12	MCARI_OSAVI VARI mean_VARI range_NDVI	$HCV = 124.222 - 12.3636 * \sin_DOY + 0.0661979 * p60 - 280.684 * MCARI_OSAVI + 110.723 * VARI - 120.596 * mean_VARI - 44.4529 * range_NDVI$

Media
Matorral
Categoría
YEAR

54.75/
10.79

sin_DOY
p60
MCARI_OSAVI
VARI
mean_VARI
range_NDVI

$$\begin{aligned}
 & \text{HCV} = 123.169 + \\
 & 1.35877 * \text{I1}(1) - \\
 & 4.00831 * \text{I1}(2) - \\
 & 12.3076 * \text{sin_DOY} + \\
 & 0.0551313 * \text{p60} - \\
 & 269.219 * \text{MCARI_OSAVI} \\
 & + 116.651 * \text{VARI} - \\
 & 124.003 * \text{mean_VARI} - \\
 & 41.4052 * \text{range_NDVI} \\
 & \text{I1}(1) = 1 \text{ si Year}=2019, -1 \\
 & \text{si Year}=2021, 0 \text{ de lo} \\
 & \text{contrario} \\
 & \text{I1}(2) = 1 \text{ si Year}=2020, -1 \\
 & \text{si Year}=2021, 0 \text{ de lo} \\
 & \text{contrario}
 \end{aligned}$$

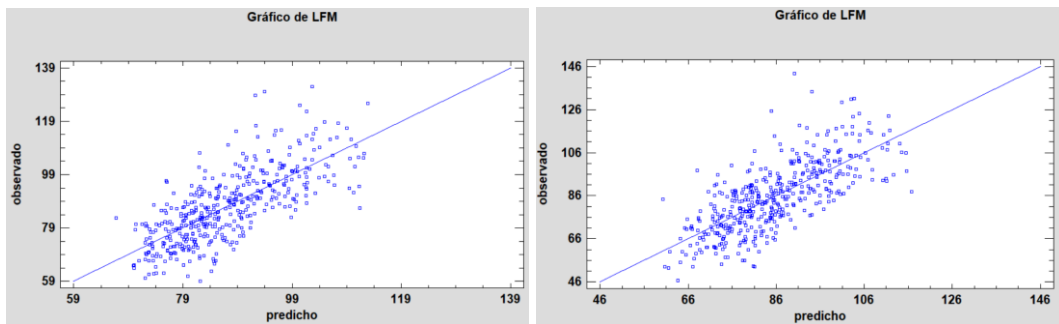


Figura 30. Gráfico de observado vs predicho, Valencia 2019-2020-2021. Media ponderada (izquierda), media matorral (derecha)

Una vez que se ha calculado los modelos de regresión lineal multivariante para los años 2019, 2020 y 2021 en conjunto para la provincia de Valencia, es necesario evaluar su ajuste. En esta ocasión, gracias a la disponibilidad de datos de campo, se usaron parcelas de evaluación, que no intervinieron en el proceso de entrenamiento. Dichas parcelas, también están catalogadas del tipo SH4, según el modelo de combustible y se encuentran en la provincia de Valencia, como se muestra a continuación en la Tabla 22.

Tabla 22. Parcelas de validación para la provincia de Valencia.

Municipio	Parcela
Gilet	4
Chelva	12
Llombai	15
Cortes de Pallás	33
Gandía	34

Buñol	92
Requena	91

En la Figura 32, se muestra los gráficos X-Y múltiple, resultado de evaluar los modelos descritos en la Tabla 21 en parcelas de evaluación en cada municipio, tanto para la media ponderada (rojo) como para la media ponderada categórica (rosa). También, se visibiliza la HCV muestreada en campo (azul). Como se puede ver en el municipio de Gilet, los modelos se ajustan muy bien para el verano del 2019; sin embargo, a medida que avanzan los años, el ajuste del modelo a la realidad se va perdiendo; esto se relaciona mucho con el análisis por años separados hecho anteriormente, donde el R-cuadrado ajustado, iba disminuyendo al pasar de los años. En la parcela de Gilet existe un porcentaje muy alto de la especie *Stipa tenacissima* (FCC=40), la cual tiene valores muy bajos de HCV. Solo existe otra parcela, la número 7, con presencia de esta especie vegetal, pero en menor proporción. Ello influye en que los valores observados de la media ponderada de HCV sean más bajos en la parcela 4, respecto a las parcelas usadas como entrenamiento del modelo, especialmente en los veranos de los años 2020 y 2021. Por este motivo la predicción en estos dos años funciona peor. En otras parcelas, como las correspondientes a Chelva y Gandía, se ve un ajuste aceptable de los modelos, donde las curvas de comportamiento se aproximan mucho a la realidad, así como sus máximos y mínimos. En el municipio de Llombai la tendencia general observada en las predicciones se acercan mucho a la HCV tomada en campo, pero las curvas generadas por el modelo pueden distanciarse de la realidad. Para el municipio de Cortes de Pallás, las curvas de los modelos muestran un comportamiento parecido al de los valores observados. Sin embargo, presentan un desplazamiento, haciendo que los valores calculados sean menores a los datos de campo. Este desplazamiento puede ser debido a que en la parcela 33 de Cortes de Pallás existe la especie *Pinus pinaster* y en las parcelas usadas para construir el modelo no existe esa especie. En la figura 31 se puede apreciar la diferencia existente entre la HCV de *Rosmarinus officinalis* -especie estudiada en este apartado– y la humedad de combustible vivo de la especie *Pinus Pinaster*. Se puede apreciar que en la HCV de *Pinus pinaster* llega a ser más alta en los años 2019 y 2020, y muy parecida para el año 2021.

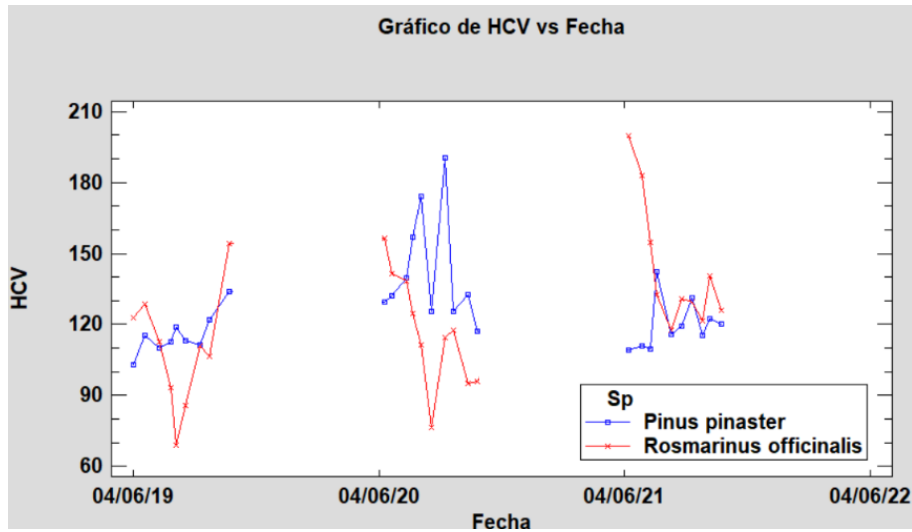
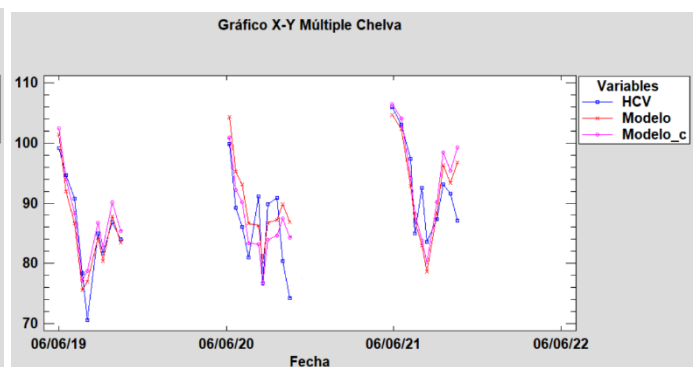
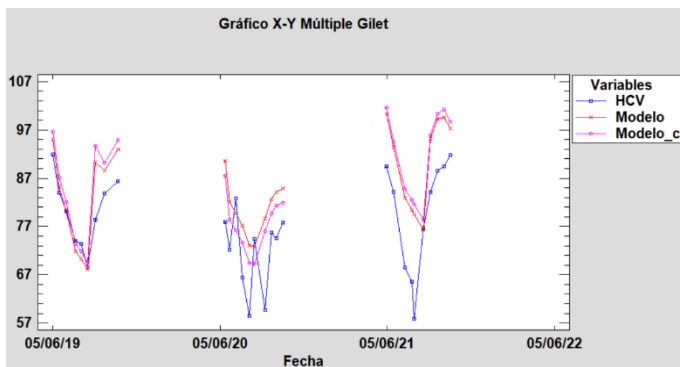


Figura 31. HCV de *Rosmarinus officinalis* (rojo) y la humedad de combustible vivo de la especie *Pinus Pinaster* (azul).

Para los municipios de Buñol y Requena, los datos de campo permiten realizar la evaluación únicamente en el verano del 2021; en la primera se puede apreciar que el ajuste de los modelos está por debajo a la HCV de campo, y para el segundo municipio el modelo no estima ciertos picos a inicios de julio y finales de agosto del 2021, que muestran los datos de campo.

En todos los modelos la introducción de la variable categórica *Year*, se asimila mucho a un modelo sin la aplicación de la misma y en ninguno de estos casos estudiados, el modelo categórico presentó una mejora significativa, tanto en R-cuadrado ajustado como en el análisis de gráficos X-Y múltiple.



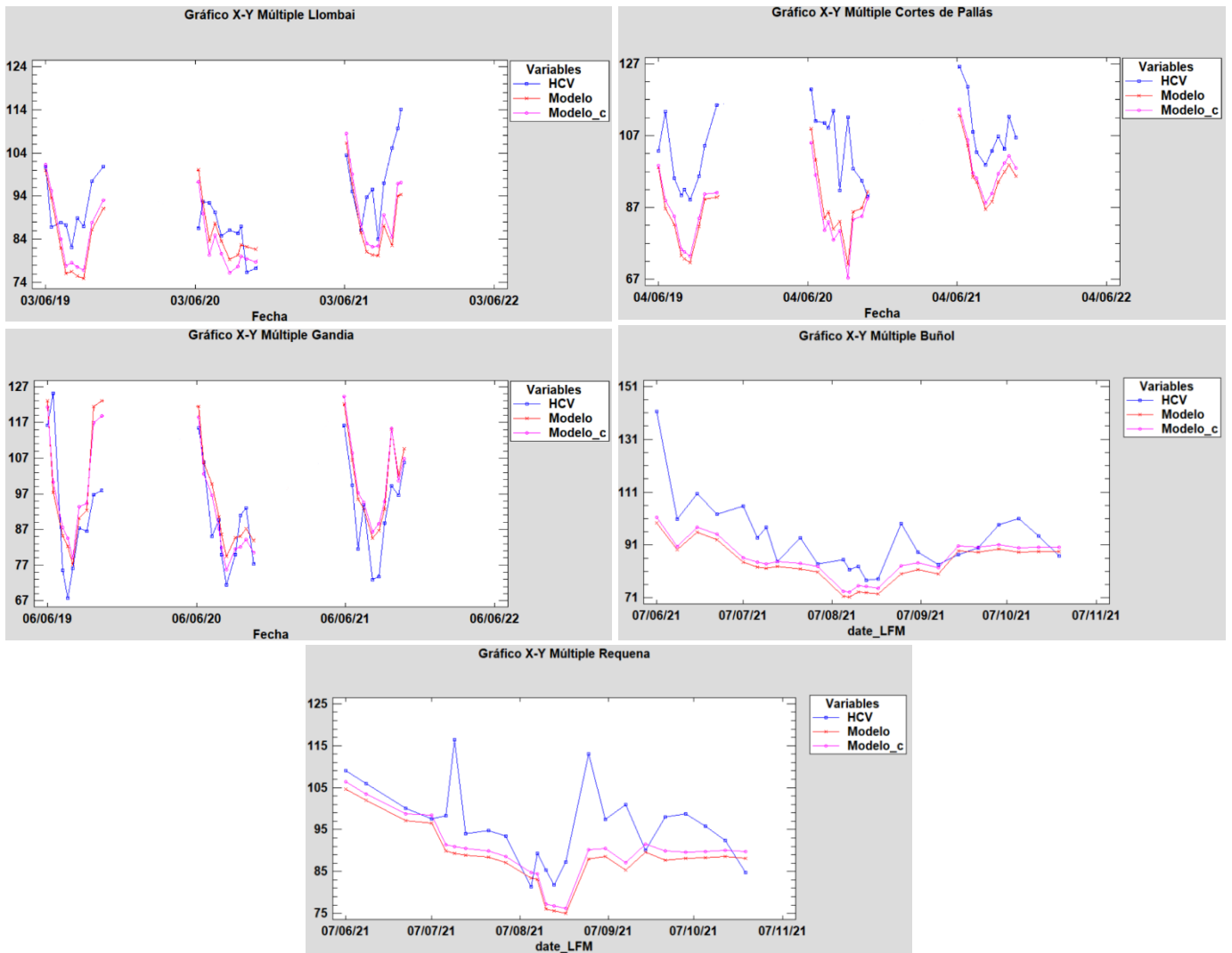


Figura 32. Modelos de evaluación para cada municipio de la provincia de Valencia. HCV muestreada en campo (azul), Modelo media ponderada (rojo), Modelo media ponderada categórica (rosa).

Ya que las parcelas catalogadas como SH4 caracterizan a un paisaje de matorral a continuación, se analiza un modelo de una de las especies más representativas de este tipo de parcelas en la zona de estudio, la especie *Rosmarinus officinalis*. Como se ve en la Tabla 23 el modelo fue creado para todos los años en conjunto alcanzando un R-cuadrado ajustado de 55.79% con el uso de seis variables entre las cuales se tiene a NDVI y Vgreen como índices espectrales, p60 como variable meteorológica, mean_SLA y range_NDVI como variables de estadísticos de índices espectrales.

Tabla 23. *Modelo de Rosmarinus officinalis para la provincia de Valencia.*

Años	Sp	R2 ajustado	VARIABLES	Modelo
2019 – 2020 - 2021	Rosmarinus officinalis	55.7953	sin_DOY2	HCV = 172.321 - 6.88127*sin_DOY2 +
			p60	0.152083*p60 +
			NDVI	238.275*NDVI +
			Vgreen	347.902*Vgreen -
			mean_SLA	104.699*mean_SLA -
			range_NDVI	65.9159*range_NDVI

La evaluación de este modelo fue realizada con las parcelas usadas en anteriores evaluaciones que no han sido parte del entrenamiento del modelo, pero en este caso es necesario que la parcela contenga HCV de la especie *Rosmarinus officinalis*. Las parcelas seleccionadas se muestran en la Tabla 24.

Tabla 24. *Parcelas de evaluación para Rosmarinus officinalis en la provincia de Valencia.*

Municipio	Parcela
Gilet	3
Chelva	12
Lombai	18
Cortes de Pallás	33

Las parcelas 3 y 18 no fueron usadas en la validación de la media ponderada pues, a pesar de pertenecer a la categoría SH4 el valor de FCC de la especie *Pinus halepensis* observada en campo es superior a 70, y ello no ocurre en el resto de parcelas de entrenamiento del modelo. A la hora de evaluar la predicción de la especie *Rosmarinus officinalis* se tiene en cuenta que en estas parcelas convive con otras especies de arbolado que están en una proporción importante.

Analizando los gráficos X-Y múltiple de la Figura 33 se puede notar que para los municipios de Gilet y Chelva los ajustes son muy buenos en los 3 años, los modelos siguen la curva de tendencia de la HCV calculada en campo, sus mínimos también se acercan mucho a la realidad, sin embargo, se pueden apreciar que los máximos pueden estar desplazados. Para los municipios de Llombai y Cortes de Pallás los ajustes presentan un desfase principalmente en el verano del 2020, haciendo que el modelo calculado sea superior en el primer caso e inferior para el segundo. Ello es debido a que en estas parcelas la especie *Rosmarinus officinalis* no es una especie predominante alcanzando una FCC

igual a 10 en la parcela de Llombai y con FCC=5 en la de Cortes de Pallás. La señal espectral recoge la presencia de otras especies además de la presencia de *Rosmarinus officinalis* y no puede captar su comportamiento con precisión.

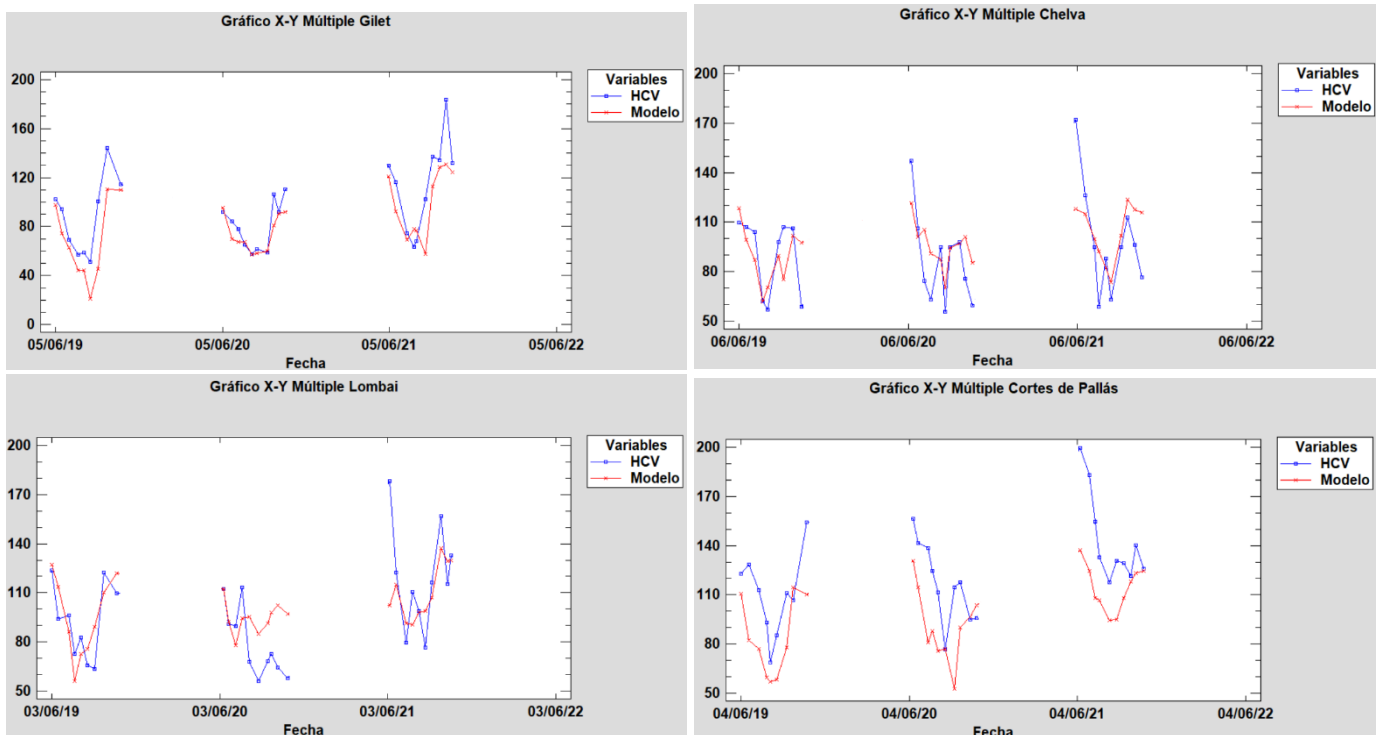


Figura 33. Modelos de evaluación para *Rosmarinus officinalis* en la provincia de Valencia.

4.2 Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la Provincia de Alicante.

Para la creación de modelos en la provincia de Alicante se tomaron las parcelas en tres municipios como se muestran en la Tabla 25. Debido a que en estas parcelas la toma de datos de campo de HCV se inició más tarde, la disponibilidad de datos se encuentra únicamente en las fechas a partir de septiembre del 2020 hasta noviembre del 2021, por lo cual al tener menos cantidad de datos que representen veranos. Por este motivo, se ha optado por la creación de modelos con todas las fechas disponibles, es decir los siguientes modelos están entrenados con datos desde septiembre del 2020 hasta noviembre del 2021.

Tabla 25. Parcelas de matorral (SH4) de la Provincia de Alicante.

Municipio	Parcela
La Torre de les Maçanes	71,72
La Vall de Gallinera	74,75,76,77
Bernia	86,87,88

Como se puede ver en la Tabla 26 se han calculado modelos para la media ponderada y media matorral, donde se ha alcanzado un R-cuadrado ajustado de 67.93% y 67.46% respectivamente, con la aplicación de seis variables en las cuales solamente el NPCI representa a los índices espectrales y las variables meteorológicas MeanRelativeHumidity y p60 se repiten en todos los modelos. También se ha construido un modelo para la especie *Rosmarinus officinalis* alcanzando un R-cuadrado ajustado de 68.13%, en este modelo influyen 5 variables en donde MCARI_OSAVI representa a los índices espectrales. Los modelos creados para la provincia de Alicante en parcelas de matorral (SH4) describen la HCV en más de un 60% llegando a ser un buen estimador, pero hay que tener en cuenta que los datos disponibles para realizar este análisis son menores y se encuentran únicamente en tres municipios de la provincia de Alicante por lo que, la focalización de la información ayuda a que se pueda llegar a un R-cuadrado ajustado más alto. Además, el periodo de tiempo usado para construir los modelos en la provincia de Alicante es distinto al utilizado en los modelos de la sección anterior, con la provincia de Valencia. En esta sección se ha escogido el periodo completo desde septiembre del 2020 hasta noviembre del 2021, mientras que en el apartado anterior solo se han escogido datos en los meses desde junio hasta octubre de cada año. Por este mismo motivo, el valor de RMSE sube significativamente de 14% para los modelos de media matorral y media ponderada a 23% para el modelo de *Rosmarinus officinalis*.

Para la realización del modelo correspondiente a la especie *Rosmarinus officinalis* se debe tener en cuenta que el número de observaciones baja puesto que las parcelas 71 y 86 no intervienen en el entrenamiento debido a que estas parcelas no contienen datos sobre esa especie de estudio. En la Figura 34 se puede apreciar que muchos valores de observados versus predichos se acercan a la diagonal.

Tabla 26. Modelos de matorral (SH4) de la Provincia de Alicante

Años	Sp	R2 ajust/ RMSE	VARIABLES	Modelo
2020 - 2021	Media Ponderada	67.92/ 14.25	sin_DOY MeanRelativeHumidity p60 NPCI mean_Vgreen range_NDWI	HCV = -120.095 - 14.4645*sin_DOY2 + 0.374933*MeanRelativeHumidity + 0.128989*p60 - 31.8003*NPCI - 745.592*mean_Vgreen + 54.218*range_NDWI
	Media Matorral	67.45/ 14.75	sin_DOY MeanRelativeHumidity p60	HCV = -102.533 - 15.9179*sin_DOY2 + 0.411774*MeanRelativeHumidity

			NPCI	ity + 0.143928*p60 -
			mean_Vgreen	28.4467*NPCI -
			range_NDWI	660.319*mean_Vgreen +
				52.0231*range_NDWI
Rosmarinu			sin_DOY	HCV = 98.1224 -
s	68.12/		MeanRelativeHumi	34.2275*sin_DOY2 +
officinalis	23.64		dity	0.691056*MeanRelativeHumid
71,86			p60	ity + 0.217949*p60 -
			MCARI_OSAVI	106.663*MCARI_OSAVI +
			min_Vgreen	51.8545*min_Vgreen

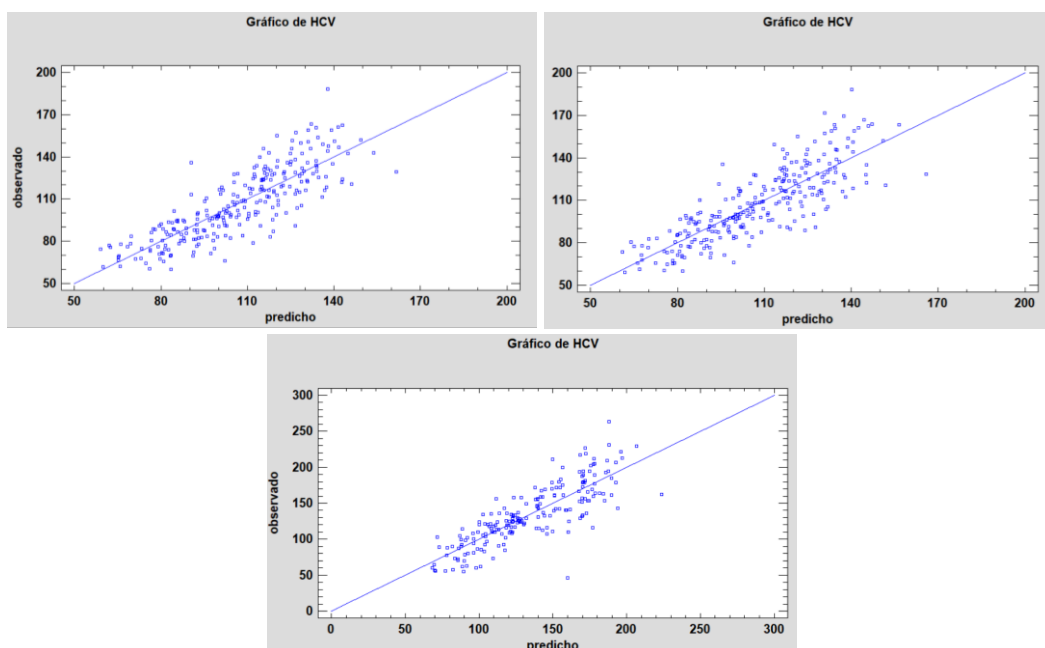


Figura 34. Gráfico de observado vs predicho, Alicante 2020-2021. Media ponderada (izquierda), media matorral (derecha), Rosmarinus officinalis (abajo)

Para la provincia de Alicante existen menos cantidad de puntos de muestreo, por lo cual no se cuenta con parcelas de evaluación disponibles y se ha procedido a realizar una validación cruzada de los datos, donde, con el uso de las mismas variables se crea un modelo descartando una de las parcelas de forma consecutiva. Así se puede analizar el ajuste del modelo tanto gráficamente como revisando los estadísticos resultantes. A continuación, se muestran los resultados de la validación cruzada del modelo de media ponderada. Se aprecia que en todos los casos el R-cuadrado ajustado está entre 66.45% a 69.27% haciendo que en todos los casos se pueda decir que el ajuste es similar. También el error estándar y el error medio absoluto son similares en cada validación cruzada. Sumado a esto, analizando las ecuaciones resultantes para cada modelo de validación, se aprecia que los coeficientes que acompañan a las variables independientes tienen una

similitud haciendo que no varíen considerablemente. Se puede decir que el modelo creado para la provincia de Alicante presenta una ecuación representativa de todas las parcelas usadas para el entrenamiento y que ninguna de las parcelas por si sola marca una gran diferencia en los resultados. La nomenclatura “Pre_alipon71” significa; la predicción realizada para la media ponderada -pon- en la provincia de Alicante -ali- sin tomar en cuenta la parcela número 71.

Pre_alipon71

$$\text{HCV} = -128.091 - 15.7647 * \sin_DOY2 + 0.461057 * \text{MeanRelativeHumidity} + 0.134902 * p60 - 27.5468 * \text{NPCI} - 743.571 * \text{mean_Vgreen} + 56.7078 * \text{range_NDWI}$$

R-cuadrado = 68.6651 %

R-cuadrado (ajustado por g.l.) = 67.822 %

Error estándar del est. = 13.9024

Pre_alipon72

$$\text{HCV} = -119.21 - 15.6788 * \sin_DOY2 + 0.4109 * \text{MeanRelativeHumidity} + 0.128925 * p60 - 31.4165 * \text{NPCI} - 742.929 * \text{mean_Vgreen} + 45.4994 * \text{range_NDWI}$$

R-cuadrado = 68.0143 %

R-cuadrado (ajustado por g.l.) = 67.1575 %

Error estándar del est. = 14.3727

Pre_alipon74

$$\text{HCV} = -106.339 - 13.5218 * \sin_DOY2 + 0.309878 * \text{MeanRelativeHumidity} + 0.126227 * p60 - 31.9664 * \text{NPCI} - 726.651 * \text{mean_Vgreen} + 37.4523 * \text{range_NDWI}$$

R-cuadrado = 69.445 %

R-cuadrado (ajustado por g.l.) = 68.6229 %

Error estándar del est. = 13.6315

Pre_alipon75

$$\text{HCV} = -118.86 - 15.1717 * \sin_DOY2 + 0.38007 * \text{MeanRelativeHumidity} + 0.121358 * p60 - 29.6938 * \text{NPCI} - 738.115 * \text{mean_Vgreen} + 55.624 * \text{range_NDWI}$$

R-cuadrado = 69.3117 %

R-cuadrado (ajustado por g.l.) = 68.4861 %

Error estándar del est. = 14.3682

Pre_alipon76

$$\text{HCV} = -119.304 - 13.3984 \cdot \sin_DOY2 + 0.353318 \cdot \text{MeanRelativeHumidity} + 0.13131 \cdot p60 - 26.7349 \cdot \text{NPCCI} - 744.594 \cdot \text{mean_Vgreen} + 53.8262 \cdot \text{range_NDWI}$$

R-cuadrado = 67.333 %

R-cuadrado (ajustado por g.l.) = 66.454 %

Error estándar del est. = 14.7618

Pre_alipon77

$$\text{HCV} = -116.858 - 13.8065 \cdot \sin_DOY2 + 0.336305 \cdot \text{MeanRelativeHumidity} + 0.120526 \cdot p60 - 28.4156 \cdot \text{NPCCI} - 745.133 \cdot \text{mean_Vgreen} + 54.1771 \cdot \text{range_NDWI}$$

R-cuadrado = 67.8489 %

R-cuadrado (ajustado por g.l.) = 66.9839 %

Error estándar del est. = 14.0071

Pre_alipon86

$$\text{HCV} = -117.664 - 14.6802 \cdot \sin_DOY2 + 0.355524 \cdot \text{MeanRelativeHumidity} + 0.127755 \cdot p60 - 32.288 \cdot \text{NPCCI} - 743.004 \cdot \text{mean_Vgreen} + 54.0947 \cdot \text{range_NDWI}$$

R-cuadrado = 69.5869 %

R-cuadrado (ajustado por g.l.) = 68.7795 %

Error estándar del est. = 14.1519

Pre_alipon87

$$\text{HCV} = -119.659 - 13.7811 \cdot \sin_DOY2 + 0.367447 \cdot \text{MeanRelativeHumidity} + 0.135003 \cdot p60 - 35.5018 \cdot \text{NPCCI} - 747.86 \cdot \text{mean_Vgreen} + 52.9235 \cdot \text{range_NDWI}$$

R-cuadrado = 68.6354 %

R-cuadrado (ajustado por g.l.) = 67.8027 %

Error estándar del est. = 14.5318

Pre_alipon88

$$\text{HCV} = -134.495 - 14.158 \cdot \sin_DOY2 + 0.404398 \cdot \text{MeanRelativeHumidity} + 0.138722 \cdot p60 - 48.5146 \cdot \text{NPCCI} - 780.583 \cdot \text{mean_Vgreen} + 77.8396 \cdot \text{range_NDWI}$$

R-cuadrado = 70.0709 %

R-cuadrado (ajustado por g.l.) = 69.2763 %
Error estándar del est. = 14.0799

En la Figura 35 se muestran los gráficos X-Y múltiple de la validación del modelo de la media ponderada en parcelas de diferentes municipios y que más diferencia de R-cuadrado ajustado presentan, en este caso las parcelas 72, 76 y 88 y se analiza el periodo de junio-octubre del 2021 ya que es el verano que más datos tiene a disposición por lo tanto mejor ajuste se podría esperar. Como se ve el ajuste para la parcela 76 sigue la curva de tendencia para todo el periodo analizado y sus máximos y mínimos se ajustan muy bien en la mayoría del recorrido, en la parcela 72 y 76 se observa que la curva de tendencia se sigue muy bien, pero que los máximos y mínimos pueden estar un poco desfasados sin que este desfase sea significativo pues los puntos medidos en campo de HCV y los puntos de predicción del modelo están muy cercanos. En estas dos últimas parcelas se nota que hacia el final del mes de octubre la predicción parece empeorar siendo el punto central agosto, donde mejor funciona la predicción.

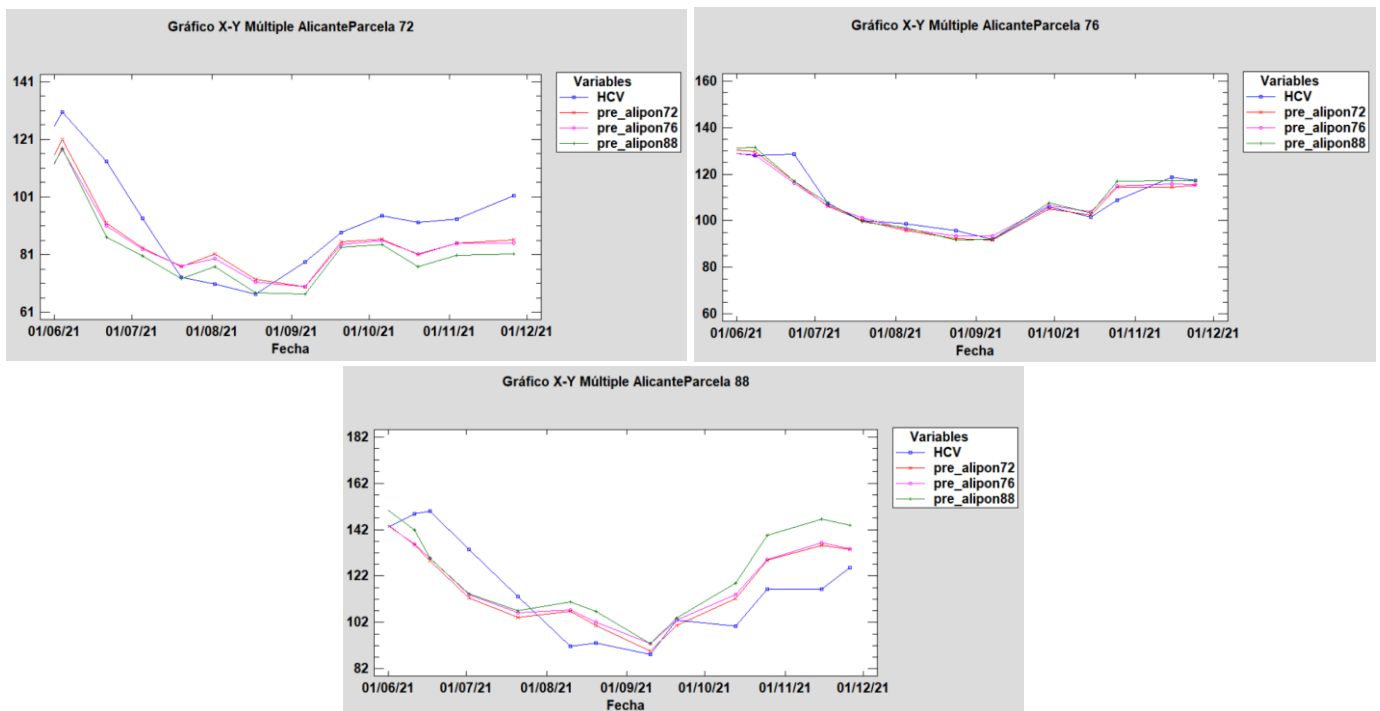


Figura 35. Modelos de validación para la provincia de Alicante.

En adición a los modelos ya presentados se ha realizado un nuevo análisis para determinar un modelo de regresión lineal usando solamente los datos del verano de 2021, como se ve en la Tabla 27. El R-cuadrado ajustado para media ponderada, media matorral y *Rosmarinus officinalis* es de 82.25, 80.87 y 79.77% respectivamente. Estos valores altos de ajuste se deben a la menor cantidad de datos de entrada que tiene los modelos de regresión. De igual forma al analizar solo con datos del año del 2021 las variables de índices espectrales tienen menos importancia frente a las variables meteorológicas puesto

que en cada caso existen dos variables del tipo meteorológico frente a una variable del tipo índice espectral.

Tabla 27. *modelo de regresión lineal usando solamente los datos del verano de 2021.*

Años	Sp	R2 ajust	Variables	Modelo	
	Media Ponderada	82.2495	p60 t60 AVI mean_Vgreen	HCV = 16.9019 + 0.14857*p60 - 3.34781*t60 + 39.8543*AVI - 580.441*mean_Vgreen	
	2021 Jun- oct		80.8747	MeanRelativeHumidity p60 t60 AVI mean_Vgreen	HCV = 17.141 + 0.245711*MeanRelativeHumidity + 0.149188*p60 - 4.10308*t60 + 63.0089*AVI - 549.669*mean_Vgreen
	Rosmarinus officinalis		79.7724 71,86	sin_DOY2 p7 t30 NPCI mean_Vgreen	HCV = 124.566 - 47.5109*sin_DOY2 + 0.700727*p7 - 6.44995*t30 + 71.4638*NPCI - 627.174*mean_Vgreen

4.3 Análisis de modelos de regresión lineal multivariante en las parcelas de matorral (SH4) de la Provincia de Castellón.

Para la creación de modelos de regresión en la provincia de Castellón se tomaron datos de campo de HCV en tres municipios como se muestra en la Tabla 28. Debido a que la recolección de datos en dichas parcelas comenzó más tarde, solamente se tiene datos de HCV en los periodos desde julio del 2020 hasta noviembre del 2021. Al tener menos datos de entrenamiento en los veranos se ha optado por realizar el entrenamiento del modelo con todas las fechas disponibles, esto sería el periodo de julio 2020 a noviembre 2021.

Tabla 28. *Parcelas usadas para los modelos en la provincia de Castellón.*

Municipio	Parcela
Eslida	53, 56
Morella	59, 61
Sant Mateu	63, 64, 68

La Tabla 29 muestra los modelos de regresión para la media ponderada y media matorral. Se obtienen un R-cuadrado ajustado de 57.73% y 54.03% respectivamente. En dichos modelos fue necesario la inclusión de 6 variables independientes entre los cuales están SLA y Vgreen como índices espectrales y como variable meteorológica a p60. También existe la presencia de varios estadísticos de índices espectrales. En el caso de la media matorral son tres de este tipo de variables que intervienen, lo cual nos hace suponer que para este modelo es mucho más difícil predecir la HCV a través de los índices espectrales.

Para el modelamiento de *Rosmarinus officinalis* se debe mencionar que la cantidad de parcelas de entrenamiento se reduce considerablemente puesto que las parcelas 59, 61, 63 y 64 no tienen presencia de esta especie, haciendo que el modelo que en los anteriores casos fue entrenado con 7 parcelas y ahora se han usado únicamente 3 parcelas. Esto influye directamente al ajuste de un R-cuadrado, pues en este caso se obtuvo un 71.97% con el uso de 3 variables. Este sería un valor muy aceptable si no fuera porque los datos de entrada para la realización del modelo se vieron muy reducidos, por tal caso no se puede decir que el modelo para *Rosmarinus officinalis* describa más del 70% de las parcelas de entrenamiento totales. Los valores de RMSE son altos respecto a modelos anteriores y se incrementan alcanzando el máximo valor de 24.48% debido a que en las estaciones de primavera e invierno los valores de HCV son más elevados y el error aumenta. En la Figura 36 de valores observados versus predichos, se puede apreciar que únicamente los valores de *Rosmarinus officinalis* siguen la diagonal de forma aceptable.

Tabla 29. Modelos de regresión para la media ponderada y media matorral en la provincia de Castellón

Años	Sp	R2 ajust/ RMSE	Variables	Modelos
2020- 2021	Media Ponderada	57.72/ 19.4985	sin_DOY2 p60 SLA Vgreen max_NMDI mean_NPCI	HCV = -147.623 - 14.2799*sin_DOY2 + 0.116851*p60 + 27.1328*SLA + 300.641*Vgreen + 297.25*max_NMDI + 461.639*mean_NPCI
	Media Matorral	54.02/ 22.149	sin_DOY2 p60 Vgreen mean_NPCI mean_SAVI min_NDWI	HCV = 209.885 - 15.2249*sin_DOY2 + 0.169853*p60 + 259.491*Vgreen + 235.291*mean_NPCI - 556.311*mean_SAVI + 297.711*min_NDWI

Rosmarinus
 officinalis
 59-61-63-
 64

71.9707/
 24.4803

sin_DOY2
 p60
 MCARI_OSAVI

HCV = 143.762 -
 41.6708*sin_DOY2 +
 0.373629*p60 -
 836.802*MCARI_OSAVI

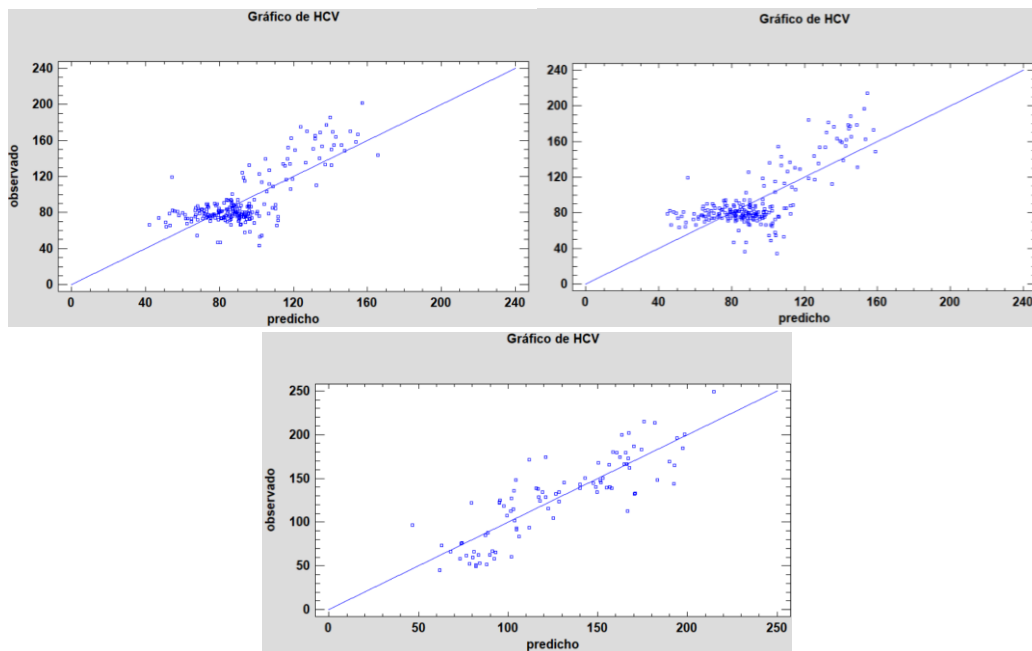


Figura 36. Gráfico de observado vs predicho, Castellón 2020-2021. Media ponderada (izquierda), media matorral (derecha), Rosmarinus officinalis (abajo)

La validación del modelo de regresión de la media ponderada de HCV para la provincia de Castellón, se realizó a través de la validación cruzada, de la misma forma que fue realizada en la provincia de Alicante. Se tiene que el R-cuadrado ajustado para cada caso de la validación cruzada varía de forma considerable, para el caso de la validación en la parcela 56 tan solo se obtiene un R-cuadrado ajustado de 48.42% un error estándar de 16.56 y un error medio absoluto de 11.52, esto nos hace entender que esta parcela aporta mucha información al proceso de modelación por tal, al momento de retirarla se obtiene un ajuste tan bajo. Por otro lado, está la parcela 68 la cual en el proceso de validación tiene un R-cuadrado ajustado de 61.81%, un error estándar de 19.77 y un error medio absoluto de 15.54, ya que al momento de realizar la validación cruzada retirando esta parcela se obtiene un ajuste alto y se puede decir que dicha parcela aporta muy poco al modelo de regresión lineal multivariado planteado.

Pre_caspon53

HCV = 49.5652 - 10.1914*sin_DOY2 + 0.0633007*p60 + 37.0762*SLA +
 255.945*Vgreen +
 56.7196*max_NMDI + 95.0633*mean_NPCI

R-cuadrado = 56.6461 %
R-cuadrado (ajustado por g.l.) = 55.1929 %
Error estándar del est. = 17.1067

Pre_caspon56

HCV = -168.27 - 8.83796*sin_DOY2 + 0.0933273*p60 - 0.76417*SLA +
189.82*Vgreen + 316.666*max_NMDI
+ 519.053*mean_NPCI
R-cuadrado = 50.0828 %
R-cuadrado (ajustado por g.l.) = 48.4189 %
Error estándar del est. = 16.5637

Pre_caspon59

HCV = -150.525 - 15.7852*sin_DOY2 + 0.12773*p60 + 26.9783*SLA +
298.169*Vgreen +
299.973*max_NMDI + 461.824*mean_NPCI
R-cuadrado = 59.3091 %
R-cuadrado (ajustado por g.l.) = 57.9527 %
Error estándar del est. = 20.5207

Pre_caspon61

HCV = -153.724 - 15.7305*sin_DOY2 + 0.13453*p60 + 24.1945*SLA +
318.943*Vgreen +
319.246*max_NMDI + 455.864*mean_NPCI
R-cuadrado = 61.0193 %
R-cuadrado (ajustado por g.l.) = 59.7199 %
Error estándar del est. = 20.1566

Pre_caspon63

HCV = -152.24 - 16.1811*sin_DOY2 + 0.136942*p60 + 26.1382*SLA +
272.635*Vgreen +
300.642*max_NMDI + 443.453*mean_NPCI

R-cuadrado = 61.9411 %
R-cuadrado (ajustado por g.l.) = 60.6583 %
Error estándar del est. = 20.0991

Pre_caspon64

HCV = -168.515 - 15.7685*sin_DOY2 + 0.135671*p60 + 30.3007*SLA +
289.693*Vgreen +

$$307.031 * \text{max_NMDI} + 492.254 * \text{mean_NPCI}$$

R-cuadrado = 59.7565 %

R-cuadrado (ajustado por g.l.) = 58.4 %

Error estándar del est. = 20.3853

Pre_caspon68

$$\text{HCV} = -143.624 - 16.475 * \text{sin_DOY2} + 0.132339 * \text{p60} + 27.3879 * \text{SLA} + 323.75 * \text{Vgreen} + 297.421 * \text{max_NMDI} + 465.559 * \text{mean_NPCI}$$

R-cuadrado = 63.0581 %

R-cuadrado (ajustado por g.l.) = 61.8129 %

Error estándar del est. = 19.7705

En la siguiente Figura 37 se muestra los gráficos X-Y múltiples resultantes de la validación cruzada en las parcelas 53, 56 y 68. En el primer caso, la parcela 53 no se puede decir que ninguna curva de validación es satisfactoria pues las curvas no corresponden ni siguen la tendencia a los valores de HCV observados en campo. En el segundo caso de la parcela 56 se puede ver que realizar un modelo sin tomar en cuenta los datos de esta parcela resta mucho al resultado final, pues los modelos de predicción realizados con los datos de esta parcela siguen ligeramente la tendencia del valor de HCV observado. La predicción sin los datos de la parcela 56, curva color rosa, se aleja y no sigue la curva del valor observado. En el último caso de la parcela 68 se puede apreciar que todas las curvas de validación siguen al valor observado por lo cual esta parcela aporta muy poco al modelo de regresión lineal.

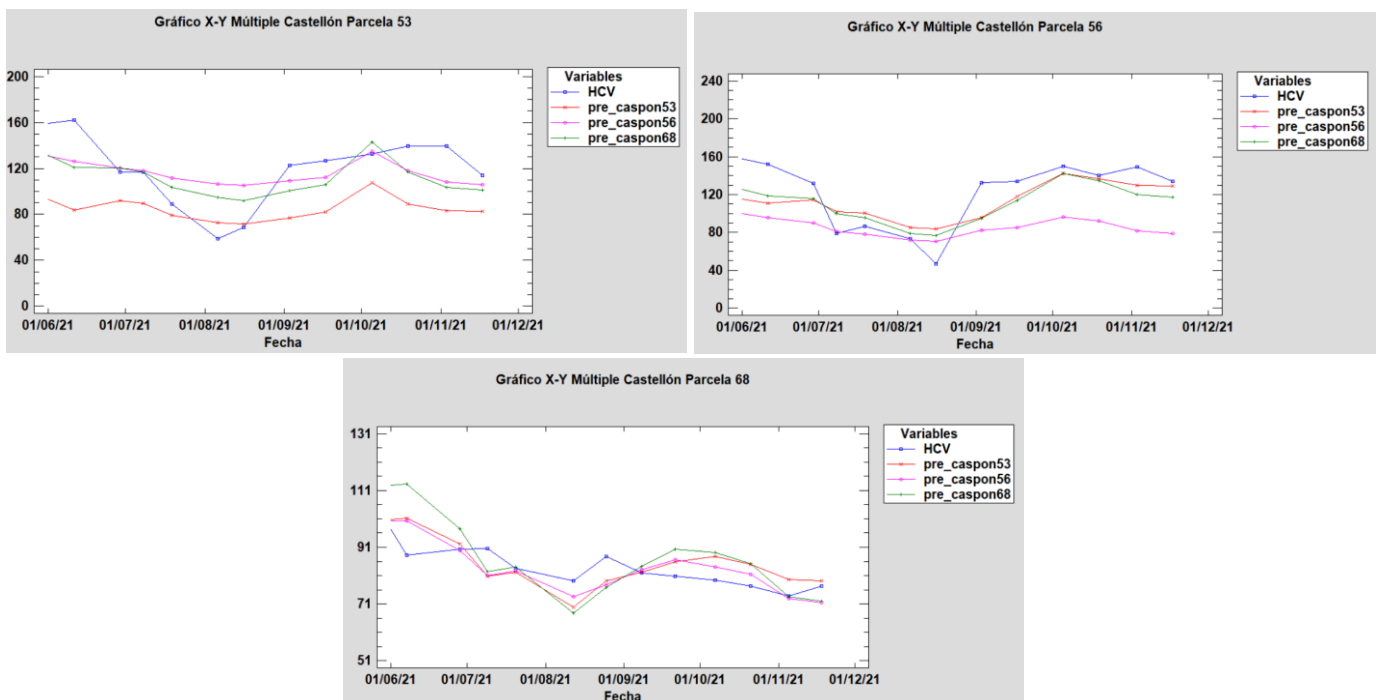


Figura 37. Modelos de validación para la provincia de Castellón.

Al igual que en el caso de Alicante, para Castellón también se ha realizado modelos de regresión lineal tomando únicamente los datos de junio a octubre del 2021 (ver Tabla 30.). En este caso el R-cuadrado ajustado para media ponderada, media matorral y *Rosmarinus officinalis* es de 61.58, 62.29 y 83.88% respectivamente, representando estos valores una ligera mejora del caso anterior donde se usa todos los datos. Sin embargo, se debe mencionar que esta mejora se debe a la reducción de datos del periodo junio-octubre del 2021, disponibles para el entrenamiento del modelo. En los modelos de este periodo de tiempo, también toman una mayor importancia las variables meteorológicas en comparación con los modelos realizados con todos los datos disponibles.

Tabla 30. Modelos de regresión lineal tomando únicamente los datos de junio a octubre del 2021.

Años	Sp	R2 ajustado	VARIABLES	Modelos
2021 Jun-oct	Media Ponderada	61.57	p3 t60 NMDI max_SLA	HCV = -63.9602 + 0.700662*p3 - 2.45935*t60 + 465.892*NMDI - 24.2164*max_SLA
	Media Matorral	62.29	p3 t60 NMDI VARI mean_SLA	HCV = -49.2416 + 0.61423*p3 - 2.09031*t60 + 496.1*NMDI + 86.0165*VARI - 72.2068*mean_SLA
	Rosmarinus officinalis 59-61-63-64	83.88	sin_DOY2 p15 p60 MCARI	HCV = 126.388 - 44.3377*sin_DOY2 + 0.25158*p15 + 0.52837*p60 - 1858.08*MCARI

4.4 Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Valencia.

En este apartado se analiza los resultados procedentes de la regresión lineal en las parcelas catalogadas como arbolado de la provincia de Valencia. En este análisis se suma una categoría dividida entre TU2 y TU3 (ver Tabla 31) la cual está determinada por el modelo de combustible de la provincia de Valencia, explicado en el capítulo de DATOS, también

se analiza los resultados de generar modelos con las dos categorías en conjunto TU2|TU3.

Tabla 31. *Parcelas de arbolado (TU2 y TU3) de la Provincia de Valencia.*

Mod_comb	Municipio	Parcela
TU2	Bétera	8
	Chelva	11
	Buñol	23
	Cortes	29, 30, 31
TU3	Gilet	1
	Chelva	14
	Buñol	25
	Gandía	36, 38, 42
	Vall de Gallinera	78

Para la generación de los modelos de regresión para la provincia de Valencia se han tomado los periodos que van desde junio del 2020 hasta noviembre del 2021. Como se ve en la Tabla 32, los resultados en los modelos de la media ponderada para las categorías de TU2, TU3 y TU2|TU3 obtienen valores de R-cuadrado ajustado muy diferentes llegando a un 66.1, 74.09 y 60.33% respectivamente, aunque en el análisis por separado se obtienen ajustes altos. Al momento de intentar generar un modelo de estas categorías en conjunto, el ajuste reduce su efectividad de forma significativa. También se nota que en el análisis por separado para las dos categorías intervienen las variables sin_DOY, p60 y Vgreen, mientras que para el modelo en conjunto se puede observar un comportamiento muy diferente de las variables que intervienen en el modelamiento, tomando mucha más importancia las variables relacionadas con los estadísticos de índices espectrales.

Después de analizar los modelos de media ponderada se analizaron los modelos de media de arbolado. En este caso las categorías de TU2 y TU3 obtienen valores de R-cuadrado ajustado muy parecidas 67.4 y 66.5% respectivamente. Se puede entender que se obtengan ajustes muy parecidos ya que los datos de entrenamiento describen precisamente la vegetación de tipo arbolado. Sin embargo, las variables que influyen difieren mucho en cada modelo mientras que para TU2 se usa seis variables entre las cuales la presencia de variables meteorológicas Radiation y t30 más el índice espectral GNDVI, el modelo de TU3 solamente usa cuatro variables entre las cuales está la variable MSI como índice espectral.

El modelo creado para la media de arbolado usando la categoría TU2|TU3 es de 55.04% lo cual muestra una reducción de R-cuadrado ajustado, el modelo usa 5 variables independientes de las cuales únicamente el índice espectral SLA participa. Tanto en los modelos de regresión de media ponderada como de media arbolado el ajuste disminuye considerablemente al momento de analizar la categoría TU2|TU3, del mismo modo en

estos modelos las variables meteorológicas pierden total peso y existe una mayor influencia de los estadísticos de índices espectrales.

Tabla 32. Modelos de HCV para las categorías de TU2, TU3 y TU2/TU3

Años	Sp	MC	R2 ajust	Variables	Modelos
2020-2021	Media Ponderada	TU2	66.10	sin_DOY p60 GCI Vgreen min_NDWI min_NPCI	HCV = 0.237871 - 7.315*sin_DOY2 + 0.0689337*p60 - 90.0138*GCI + 84.9747*Vgreen + 221.443*min_NDWI - 40.7955*min_NPCI
		TU3	74.08	sin_DOY p60 Vgreen min_NDMI min_RVI range_GNDVI	HCV = 0.586816 - 3.33906*sin_DOY2 + 0.0281706*p60 + 31.3638*Vgreen + 194.019*min_NDMI + 13.9809*min_RVI + 137.921*range_GNDVI
	TU2 TU3	60.33	sin_DOY Vgreen max_EVI max_MCARI_OSAVI mean_GNDVI mean_TCARI_OSAVI	HCV = 571.828 + 0.124379*I1(1) - 5.72145*sin_DOY + 49.8526*Vgreen - 86.6504*max_EVI + 805.472*max_MCARI_OSAVI - 403.016*mean_GNDVI - 1401.3*mean_TCARI_OSAVI	
				en donde	I1(1) = 1 si Mod_comb=TU2, - 1 si Mod_comb=TU3, 0 de lo contrario
	Media Arbolado	TU2	67.40	Radiation sin_DOY t30 GNDVI max_TCARI range_Vgreen	HCV = 331.486 - 0.970514*Radiation - 7.76005*sin_DOY2 + 1.33787*t30 + 30.8511*GNDVI - 1953.13*max_TCARI - 209.619*range_Vgreen

TU3	66.50	MSI min_SAVI range_ARVI range_MSAVI2	HCV = 150.219 - 25.9322*MSI - 208.613*min_SAVI + 83.5812*range_ARVI - 123.756*range_MSAVI2
TU2 TU3	55.03	SLA max_EVI max_NDMI mean_TCARI_OSAVI min_TCARI	HCV = 179.225 + 1.13598*I1(1) + 7.18255*SLA - 146.77*max_EVI + 63.6152*max_NDMI - 165.832*mean_TCARI_OSAVI - 421.317*min_TCARI
			en donde I1(1) = 1 si Mod_comb=TU2, - 1 si Mod_comb=TU3, 0 de lo contrario

Para realizar la validación cruzada de las parcelas de arbolado TU2 y TU3 en la provincia de Valencia se probaron los modelos de la media ponderada. A continuación, se puede ver el modelo “Pred_ValponTU2_11”, que indica la regresión de Valencia para la media ponderada en la categoría TU2 excluyendo la parcela 11 proveniente de Chelva, también se muestra el modelo “Pred_ValponTU2TU3_11” que indica la misma regresión, pero agregando las dos categorías TU2 y TU3. Se puede apreciar que en los dos casos el R-cuadrado ajustado es muy similar a los modelos iniciales obteniendo 68.226 y 61.34% respectivamente. Ya que los ajustes son muy similares en la validación cruzada para la parcela 11 se puede decir que estos dos modelos se ajustan bien para este caso en específico. En la Figura 38 se puede ver los modelos creados para la validación de la parcela 11, se aprecia que el modelo de categoría TU2 se aproxima mejor a los picos mínimos y máximos de la HCV original, mientras que el modelo de categorías TU2 y TU3, aunque sigue bien la línea de tendencia de la HCV los hace algo alejado de los valores máximos y mínimos, acercándose más a una media. Se podría decir que el modelo usando las categorías TU2 y TU3 es más conservador.

Pred_ValponTU2_11

$$HCV = -6.4834 - 8.39709*\sin_DOY + 0.0640292*p60 - 96.5481*GCI + 77.0747*Vgreen + 222.163*\min_NDWI - 38.0855*\min_NPCI$$

R-cuadrado = 69.3279 %

R-cuadrado (ajustado por g.l.) = 68.226 %

Error estándar del est. = 9.12095

Pred_ValponTU2TU3_11

$$\text{HCV} = 572.164 + 0.115437 * \text{I1}(1) - 5.92958 * \sin_DOY + 47.4149 * \text{Vgreen} - 86.775 * \max_EVI + 804.656 * \max_MCARI_OSAVI - 404.015 * \text{mean_GNDVI} - 1401.97 * \text{mean_TCARI_OSAVI}$$

en donde

$\text{I1}(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario

R-cuadrado = 62.0105 %

R-cuadrado (ajustado por g.l.) = 61.3441 %

Error estándar del est. = 10.1284

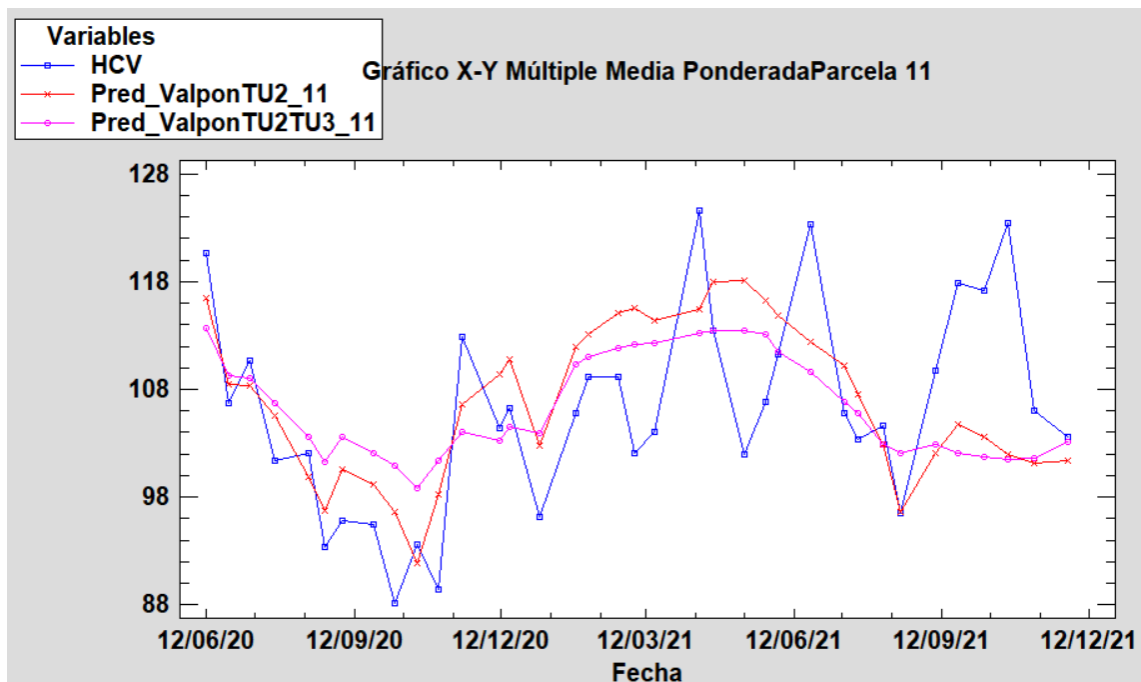


Figura 38. Modelos para la validación de la parcela 11. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2|TU3 (rosa).

Se realizó el mismo análisis de validación cruzada con la parcela 14, también perteneciente al municipio de Chelva pero catalogada en el modelo de combustible TU3. A continuación se puede apreciar que el modelo generado para la categoría TU3 sin la parcela 14 alcanza un ajuste de R-cuadrado de 73.35%, siendo muy similar al ajuste del modelo de HCV inicial. Para la validación del modelo con categorías TU2 y TU3 el comportamiento es el mismo pues se alcanza un ajuste de 59.32%, siendo muy similar al modelo correspondiente original. En la Figura 39 se presentan las curvas resultantes de la validación cruzada para la parcela 14 en las categorías TU3 y TU2 con TU3. Se puede apreciar que, en este caso, ninguno de los modelos describe un pico de la fecha de junio del 2021. Sin embargo, se puede decir que el modelo “Pred_ValponTU2TU3_14” se acerca más a la curva de HCV muestreada en campo, pues se aproxima mejor a los mínimos y a los máximos.

El modelo “Pred_ValponTU3_14” se muestra con valores más cercanos a los máximos y se podría decir que tienen un sesgo hacia los valores altos.

Pred_ValponTU3_14

$$\text{HCV} = -2.65386 - 3.37206 \cdot \sin_DOY + 0.0257704 \cdot p60 + 31.1857 \cdot V_{\text{green}} + 197.31 \cdot \min_NDMI + 15.9309 \cdot \min_RVI + 137.333 \cdot \text{range_GNDVI}$$

R-cuadrado = 74.1733 %

R-cuadrado (ajustado por g.l.) = 73.3577 %

Error estándar del est. = 8.68197

Pred_ValponTU2TU3_14

$$\text{HCV} = 572.957 + 0.295744 \cdot I1(1) - 5.94022 \cdot \sin_DOY + 50.4795 \cdot V_{\text{green}} - 85.5199 \cdot \max_EVI + 816.71 \cdot \max_MCARI_OSAVI - 404.816 \cdot \text{mean_GNDVI} - 1409.46 \cdot \text{mean_TCARI_OSAVI}$$

en donde

$I1(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario

R-cuadrado = 60.0265 %

R-cuadrado (ajustado por g.l.) = 59.3252 %

Error estándar del est. = 10.2219

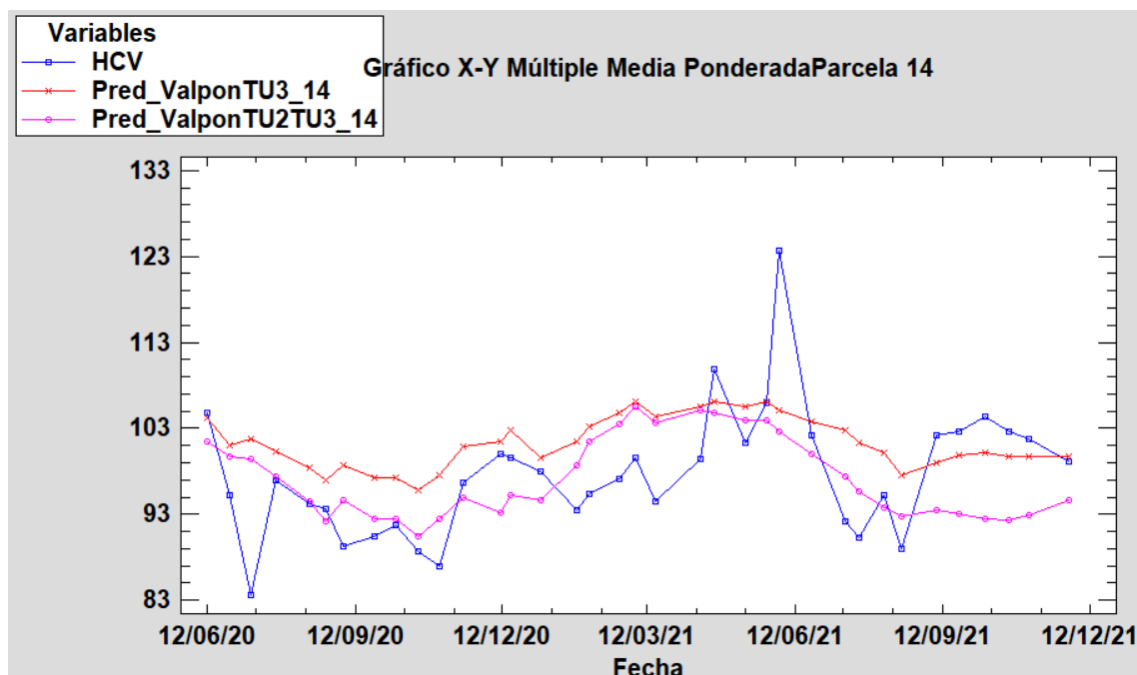


Figura 39. Modelos para la validación de la parcela 14. HCV de campo (azul). Modelos de validación media ponderada TU3 (rojo). Modelo de validación media ponderada TU2|TU3 (rosa).

4.5 Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Alicante.

Ya que la disponibilidad de datos en las parcelas procedentes de Alicante se ve reducido, para el entrenamiento de modelo de regresión lineal multivariante se ha usado únicamente datos del periodo junio-octubre del 2021, por el mismo motivo la creación del modelo en la categoría TU3 no puede ser realizado ya que los datos son insuficientes, en esta categoría solo se tiene la parcela 79. En la Tabla 33 se presenta las parcelas usadas para el entrenamiento.

Tabla 33. Parcelas de arbolado (TU2 y TU3) de la Provincia de Alicante.

Mod_comb	Municipio	Parcela
TU2	Torre de les Macanes	69, 73
	Biar	82, 83, 84
	Bernia	85
TU3	Vall de Gallinera	79

Se ve en la Tabla 34 que todos los modelos describen en más 60% la HCV de las parcelas de Alicante, para el modelo de media ponderada tanto la categoría TU2 como TU2|TU3 obtienen ajustes similares de 62.64%, sin embargo las variables que influyen en el modelo cambian, en el primer caso se tiene 5 variables entre las cuales VARI es el único índice espectral, también se tiene a la variable meteorológica MeanRelativeHumidity, en el segundo caso los índices espectrales toman más influencia interviniendo NDVI y VARI.

Para el análisis de media arbolado El ajuste de R-cuadrado también es muy similar alcanzado para TU2 y TU2|TU3 los valores de 60.28 y 64.43% respectivamente. Resalta el hecho de que, para el modelo de las categorías combinadas, aunque el R-cuadrado ajustado es muy similar existe una reducción en el número de variables que intervienen en el modelo. En este caso solo intervienen 3 variables, siendo ARVI el único índice espectral.

Tabla 34. Modelos de arbolado (TU2 y TU3) de la Provincia de Alicante.

Años	Sp	MC	R2 ajustado	Variables	Modelos
	Media Ponderada	TU2	62.63	sin_DOY MeanRelativeHumidity VARI mean_EVI mean_RVI	HCV = 77.3791 - 18.5059*sin_DOY + 0.200123*MeanRelativeHumidity + 54.2267*VARI +

				68.0951*mean_E VI - 1.49403*mean_R VI
		TU3	Sin resultados	Solo la parcela 79 cumple TU3
		TU2 TU3	62.63	$\begin{aligned} \text{HCV} = & 46.1365 + \\ & 1.02362 * I1(1) - \\ & 8.13937 * \text{sin_DO} \\ & Y + 0.585598 * t30 \\ & + 20.9855 * \text{NDVI} \\ & + 32.0003 * \text{VARI} \\ & + \\ & 17.3553 * \text{min_RVI} \end{aligned}$ <p>en donde $I1(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario</p>
		TU2	60.28	$\begin{aligned} \text{HCV} = & -323.694 + \\ & 0.128322 * \text{Humin3} \\ & + 10.4999 * \text{VARI} \\ & + \\ & 497.819 * \text{mean_N} \\ & \text{MDI} - \\ & 18.2507 * \text{mean_R} \\ & \text{VI} - \\ & 1019.04 * \text{mean_V} \\ & \text{green} \end{aligned}$
	Media Arbolado	TU3		Sin resultados
		TU2 TU3	64.43	$\begin{aligned} \text{HCV} = & 21.719 + \\ & 2.81638 * I1(1) + \\ & 25.4285 * \text{ARVI} + \\ & 57.3299 * \text{min_RVI} \end{aligned}$

$$187.41 * \min_SAV$$
$$I$$

en donde
 $I(1) = 1$ si
Mod_comb=TU2,
-1 si
Mod_comb=TU3,
0 de lo contrario

Se realizó la validación cruzada para la media ponderada de los modelos de la provincia de Alicante usando la parcela 82 perteneciente al municipio de Biar. A continuación, se muestran los resultados de generar la regresión descartando esta parcela, tanto para la categoría TU2 como TU2 y TU3. Para el primer modelo “Pred_AliponTU2_82” el R-cuadrado ajustado baja considerablemente respecto al modelo inicial, alcanzado un ajuste de 55.77%, haciendo que la parcela 82 aporte mucha información al modelo. El modelo “Pred_AliponTU2TU3_82” tiene un ajuste muy similar al original siendo de 58.75%, por lo cual para esta regresión la parcela 82 aporta menos. No se realizó la validación cruzada de la categoría TU3 ya que no se obtuvieron modelos por la falta de observaciones.

En la Figura 40 se puede ver los resultados de los modelos de validación de la parcela 82, los mínimos y máximos del modelo “Pred_AliponTU2_82” tienen acercarse más a la HCV vista en campo. Sin embargo, hacia el final del periodo del 2021 los dos modelos se alejan.

Pred_AliponTU2_82

$$HCV = 71.2138 - 18.1868 * \sin_DOY + 0.246016 * \text{MeanRelativeHumidity} + 43.0519 * \text{VARI} + 84.7469 * \text{mean_EVI} - 1.63379 * \text{mean_RVI}$$

R-cuadrado = 60.2905 %

R-cuadrado (ajustado por g.l.) = 55.7781 %

Error estándar del est. = 8.60987

Pred_AliponTU2TU3_82

$$HCV = 44.6425 + 1.69801 * I(1) - 7.42848 * \sin_DOY + 0.531558 * t30 + 20.5267 * \text{NDVI} + 23.9888 * \text{VARI} + 19.5726 * \text{min_RVI}$$

en donde

$I(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario

R-cuadrado = 60.1919 %

R-cuadrado (ajustado por g.l.) = 58.7531 %

Error estándar del est. = 8.56375

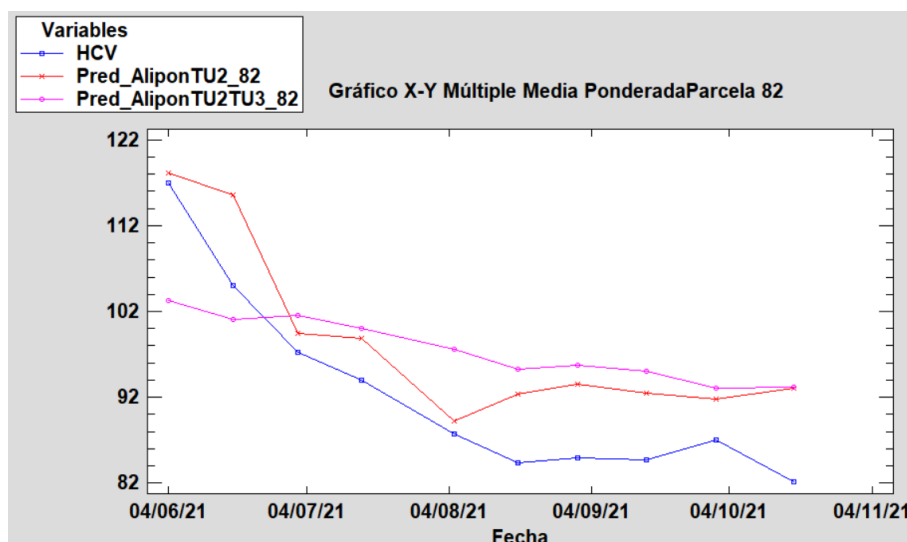


Figura 40. Modelos de validación de la parcela 82. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2|TU3 (rosa).

4.6 Análisis de modelos de regresión lineal multivariante en las parcelas de arbolado (TU2 y TU3) de la Provincia de Castellón.

Los modelos de regresión lineal generados en la provincia de Castellón fueron realizados con datos de los periodos junio-octubre del año 2021, esto debido a la disponibilidad de datos de HCV tomados en campo. En la siguiente tabla se muestra las parcelas usadas para este estudio, así como su categoría de modelo de combustible respectiva.

Tabla 35. Parcelas de arbolado (TU2 y TU3) de la Provincia de Castellón.

Mod_comb	Municipio	Parcela
TU2	Montanejos	43, 44, 46
	Eslida	51
	Morella	57
	Sant Mateu	66, 67
TU3	Eslida	52, 55
	Morella	58, 60

Como se puede ver en la Tabla 36, los modelos creados para la media ponderada muestran un ajuste similar. Para las categorías TU2 y TU2|TU3 los valores de R-cuadrado ajustado son de 61.55 y 56.85% respectivamente. En estos dos modelos las variables se asemejan pues, tanto las variables meteorológicas p30 y t60 y el índice espectral EVI son comunes en dichos modelos, Para el modelo de TU3 el ajuste del modelo llega a ser también similar a los demás, obteniendo un R-cuadrado ajustado de 57.99%. Sin embargo, las variables usadas se ven reducidas a tres, siendo p60, MCARI y NMDI suficientes para describir esta categoría de modelo de combustible.

Los modelos de regresión lineal realizados para la media arbolado presentan una notable diferencia de ajuste en las categorías TU2 y TU3 alcanzado R-cuadrado ajustado de 74.01% y 53.23% respectivamente. También se la puede ver la diferencia en la cantidad de variables que se incluyen en el modelo. En el primer caso se han usado en total cinco variables, entre las cuales se tiene a p60 y MCARI, mientras que en el segundo caso para el modelo han intervenido solamente dos variables, entre las cuales MCARI es el único índice espectral. El modelo generado para la categoría en conjunto TU2|TU3 tiene un R-cuadrado ajustado de 63.25% con cuatro variables de las cuales MCARI, NPCI y VARI son índices espectrales.

La reducción de variables para la media ponderada y para la media de arbolado en la categoría TU3 puede deberse a la reducción de parcelas que existe en esta categoría, pues mientras que en la categoría TU2 se analizan cuatro municipios, en la categoría TU3 se analizan dos municipios. Esto hace que el análisis de esta última categoría se vuelva más focalizado en el espacio.

Tabla 36. Modelos creados de la provincia de Castellón

Años	Sp	MC	R2 ajustado	Variables	Modelos
2021 Jun- oct	Media Ponderada	TU2	61.5509	p30 t60 EVI mean_GNDVI	$HCV = 328.321 +$ $0.0813901 * p30 -$ $1.86602 * t60 +$ $120.972 * EVI -$ $353.063 * mean_GNDVI$
		TU3	57.9967	p60 MCARI NMDI	$HCV = -101.334 +$ $0.138817 * p60 +$ $1878.18 * MCARI +$ $228.232 * NMDI$
		TU2 TU3	56.8452	p30 t60 EVI mean_MCARI mean_Vgreen	$HCV = 400.64 +$ $1.2993 * I1(1) +$ $0.092238 * p30 -$ $2.38725 * t60 +$ $103.958 * EVI -$ $3377.9 * mean_MCARI$ $+ 810.522 * mean_Vgreen$
					en donde $I1(1) = 1$ si $Mod_comb = TU2, -1$ si

			Mod_comb=TU3, 0 de lo contrario
Media Arbolado	TU2	74.0124	$\begin{aligned} \text{HCV} = & 259.222 + \\ & 0.0494877 * p60 + \\ & 768.525 * \text{MCARI} + \\ & 405.323 * \text{mean_Vgreen} \\ & - 40.161 * \text{min_SLA} + \\ & 108.312 * \text{min_Vgreen} \end{aligned}$
	TU3	53.2342	$\begin{aligned} \text{HCV} = & -100.163 + \\ & 1405.42 * \text{MCARI} + \\ & 279.218 * \text{mean_NMDI} \end{aligned}$
	TU2 TU3	63.2451	$\begin{aligned} \text{HCV} = & 205.028 + \\ & 1.73634 * I1(1) + \\ & 950.793 * \text{MCARI} + \\ & 21.0875 * \text{NPCI} + \\ & 42.1813 * \text{VARI} + \\ & 540.053 * \text{mean_Vgreen} \end{aligned}$

La validación cruzada para la media ponderada de la provincia de Castellón se realizó usando la parcela 51 de categoría TU2 y la parcela 55 de categoría TU3, ambas parcelas pertenecen al municipio de Eslida. A continuación, se muestran los resultados de realizar la validación cruzada en la parcela 51. En el modelo “Pred_CasponTU2_51” el ajuste sube considerablemente llegando a ser de 77%. Se puede inferir de esto que, la parcela 51 aporta de forma negativa al ajuste del modelo, pues al retirarla el ajuste ha mejorado. Sin embargo, como se puede ver en el apartado de Datos el número de observaciones para esta provincia es menor, por lo que, la causa de obtener un mejor R-cuadrado ajustado puede deberse a la reducción del número de observaciones haciendo que el modelo sea más focalizado. Este mismo comportamiento se lo ve en el modelo de validación “Pred_CasponTU2TU3_51” que mejora su ajuste respecto del modelo inicial, alcanzando un R-cuadrado ajustado de 61.87%

Pred_CasponTU2_51

$$\text{HCV} = 308.288 + 0.0836218 * p30 - 1.11849 * t60 + 126.273 * \text{EVI} - 348.787 * \text{mean_GNDVI}$$

R-cuadrado = 78.717 %

R-cuadrado (ajustado por g.l.) = 77.1692 %
 Error estándar del est. = 6.12222

Pred_CasponTU2TU3_51

$$HCV = 424.373 - 0.116695 * I1(1) + 0.103172 * p30 - 2.0547 * t60 + 95.8056 * EVI - 4490.17 * \text{mean_MCARI} + 815.822 * \text{mean_Vgreen}$$

en donde

$I1(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario

R-cuadrado = 64.1873 %

R-cuadrado (ajustado por g.l.) = 61.8769 %

Error estándar del est. = 9.85582

En la Figura 41 se aprecia la gráfica de los modelos de validación. Se puede ver que, aunque los dos modelos se alejan de la HCV observada en campo en el inicio del periodo 2021, el modelo “Pred_CasponTU2TU3_51” se acerca mucho más a la HCV hacia el final del periodo 2021 estudiado. El modelo “Pred_CasponTU2_51” tiene una tendencia de valores más altos que los observados en campo.

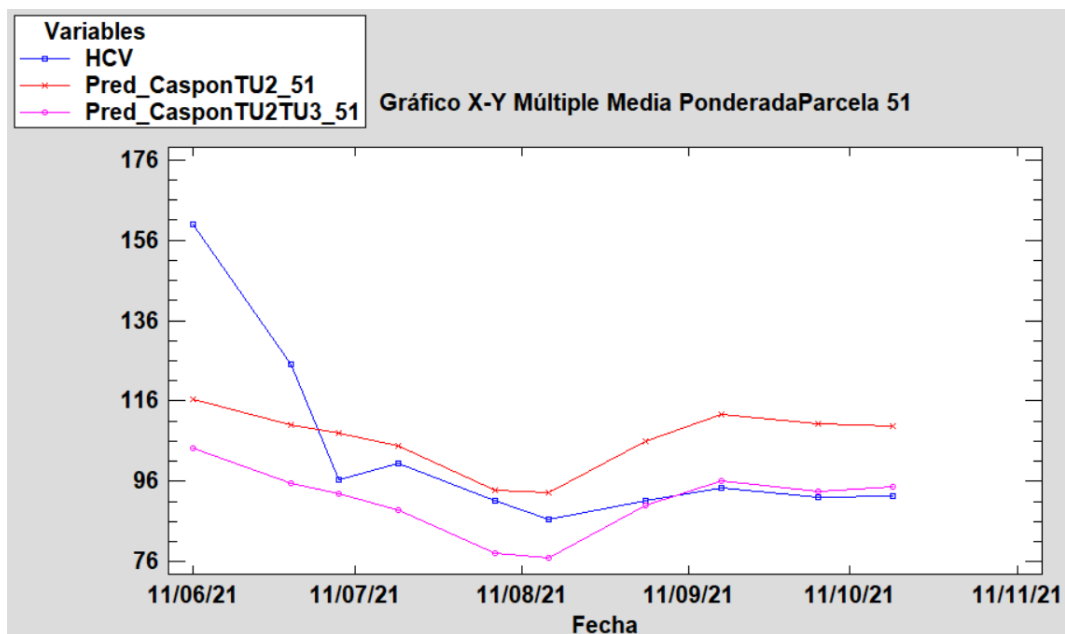


Figura 41. Modelo de validación para la parcela 51. HCV de campo (azul). Modelos de validación media ponderada TU2 (rojo). Modelo de validación media ponderada TU2|TU3 (rosa).

En la validación cruzada para la categoría TU3 usando la parcela 55, el modelo “Pred_CasponTU3_55” alcanza un ajuste muy similar a la regresión inicial con un 55.46%. Se tiene un comportamiento parecido en el análisis del modelo de validación “Pred_CasponTU2TU3_55”, que tiene un R-cuadrado ajustado de 56.73%, siendo un valor muy similar al modelo inicial. Ya que los valores de ajuste de validación cruzada y los modelos iniciales son muy parecidos se puede decir que la parcela 55, en este caso

tiene un bajo aporte en la modelación de la HCV

Pred_CasponTU3_55

$$\text{HCV} = -78.7755 + 0.114794 * p60 + 1792.54 * \text{MCARI} + 197.065 * \text{NMDI}$$

R-cuadrado = 60.0754 %

R-cuadrado (ajustado por g.l.) = 55.4688 %

Error estándar del est. = 11.0403

Pred_CasponTU2TU3_55

$$\text{HCV} = 390.587 + 2.29565 * I1(1) + 0.0659339 * p30 - 2.02081 * t60 + 109.752 * \text{EVI} - 3651.96 * \text{mean_MCARI} + 779.978 * \text{mean_Vgreen}$$

en donde

$I1(1) = 1$ si Mod_comb=TU2, -1 si Mod_comb=TU3, 0 de lo contrario

R-cuadrado = 59.357 %

R-cuadrado (ajustado por g.l.) = 56.7348 %

Error estándar del est. = 10.6357

En la Figura 42 se representan los modelos de validación para la provincia de Castellón. En los dos casos los modelos se aproximan de forma similar a los máximos y mínimos de la HCV observada en campo, pero ninguno de estos logra predecir el “pico” inferior que se genera en el mes de agosto del 2021. El modelo “Pred_CasponTU3_55” predice de mejor manera el inicio del periodo estudiado.

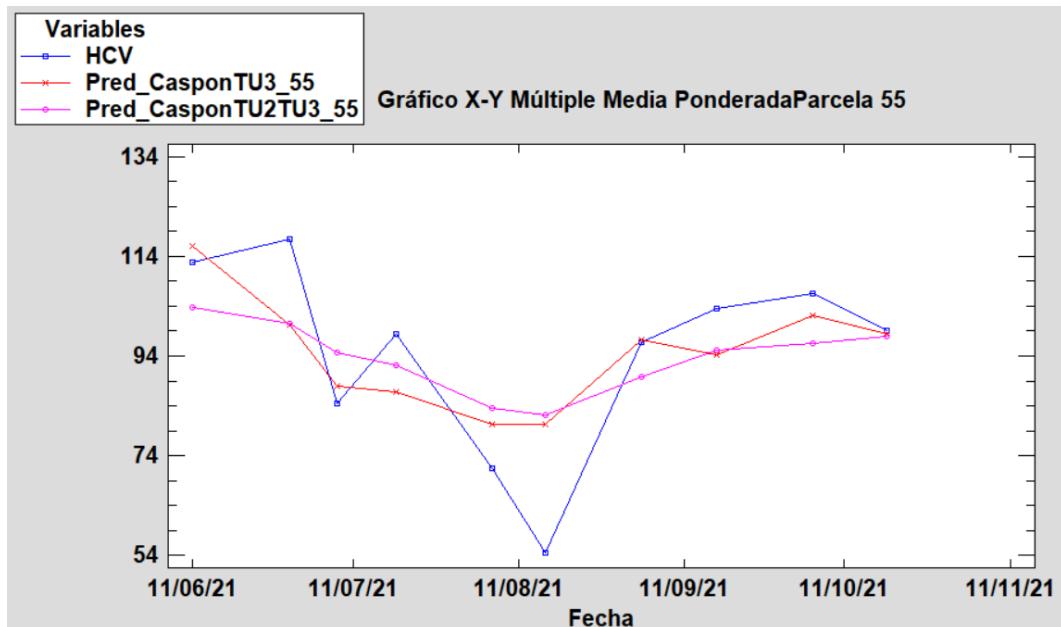


Figura 42. Modelo de validación para la parcela 55. HCV de campo (azul). Modelos de validación media ponderada TU3 (rojo). Modelo de validación media ponderada TU2|TU3 (rosa).

5. PRESUPUESTO

En la Tabla 37, se muestra el presupuesto aproximado para este estudio. En este se toma en cuenta tres grupos: Personal técnico, Datos y software. Aunque los datos de campo fueron provistos de forma gratuita por la Empresa VAERSA al tutor del proyecto a través de la direcció General de prevenció d'incendis forestals de la Generalitat Valenciana, los datos meteorológicos fueron calculados por el software libre Meteoland R, en la tabla se calcula el costo de contratar personal técnico para el muestreo de campo. La contratación de un técnico calificado es de aproximadamente 1200 euros por mes.

En este estudio se trabaja con software libre como Google Earth Engine o Python. El único software de pago es *StatGraphics XVIII-X64*, aunque gracias a la UPV se cuenta con licencia de estudiante y se calcula el precio de una licencia estudiantil para un año.

Tabla 37. Presupuesto TFM

	Mes insumos	Descripción	Costo (Observatorio/variables)	Junio	Julio	Agosto	Septiembre	Octubre	Total €
Personal técnico	Personal Técnico nivel medio	Contratación		1200	1200	1200	1200	1200	6000
	Seguridad social			703,44	703,44	703,44	703,44	703,44	3517,2
Datos		Número de estaciones	18						
	Datos Meteorológicos	Numero de variables	4						7283,52
		Tiempo (meses)	36						
	Datos HCV	Provisos por VAERSA de forma gratuita		0	0	0	0	0	0
	Imágenes satelitales	Provistas por Copernicus de manera gratuita		0	0	0	0	0	0
Software	Software Estadístico R	Software Libre		0	0	0	0	0	0
	Python	Software Libre		0	0	0	0	0	0
	QGIS	Software Libre		0	0	0	0	0	0
	StatGraphics XVII - X64*	Licencia estudiante un año	105,66						105,66
Total									16.906,38

6. CONCLUSIONES

- En este trabajo se describe la modelización del contenido de Humedad de Combustible Vivo (HCV) en la Comunitat Valenciana usando índices espectrales extraídos de Sentinel-2 durante los meses de verano y otoño desde 2019 hasta 2021.
- El procesamiento de imágenes satelitales Sentinel-2 a través de la plataforma de Google Earth Engine (GEE) que fue realizado en este estudio, presentó una gran ventaja para el rendimiento y optimización de recursos y tiempo empleado. Se procesaron imágenes satelitales desde el mes de junio del 2019 hasta el mes de octubre del 2021 con gran versatilidad en: selección de metadatos, aplicación de máscaras de nubes, manejo de bandas para la creación de índices espectrales, aplicación de un método reductor sobre cada una de las 70 parcelas estudiadas y finalmente exportación de una base de datos de índices espectrales hacia un formato plano.
- La presencia de nubes generó anomalías en el cálculo de índices espectrales. Aunque se aplicaron mascararas que usan la banda “Q60” de las imágenes Sentinel-2, se verificó visualmente que varias de las parcelas estudiadas aún se venían afectadas por formaciones nubosas en determinadas fechas, lo cual hace que el resultado del índice espectral calculado no corresponda a la realidad biofísica de la vegetación en la parcela.
- La base de datos de índices espectrales fue depurada a través de la aplicación de 3 criterios: criterio de NDVI, criterio de coeficiente de variación (CV) y criterio de la mediana. También se aplicó el filtro de *Savitzky-Golay* con una ventana de 5 fechas y polinomio de orden 2, para interpolar aquellos datos que fueron considerados como anomalías. Este procedimiento mostró eficacia para la eliminación de anomalías o picos tanto máximos como mínimos, que no responden a la realidad biofísica de la vegetación del terreno.
- Los datos de índices espectrales fueron unidos a los datos de HCV muestreados en campo a partir de su fecha más cercana. En el 99% de los casos, la variabilidad de fechas entre el dato de campo y el índice espectral es menor de los 12 días, esto representa un desfase de tiempo entre el valor de HCV y su correspondiente índice espectral, por tal motivo, se debe estudiar a futuro cual es la ventana de tiempo máxima para que un índice espectral caracterice a la vegetación de las parcelas de forma adecuada.
- De las especies muestreadas en campo sobresalen: *Pinus halepensis* y *Rosmarinus officinalis*, siendo estas las principales especies que dieron ponderación al momento de calcular la media ponderada, media matorral y media arbolado de la HCV. También se clasificó a las parcelas siguiendo los modelos de combustible de la Comunitat Valenciana, obteniendo el modelo SH4 característico de matorrales y los

modelos TU2 y TU3 característico de matorral bajo arbolado.

- El método de selección paso a paso hacia adelante con un valor-P de 0.05 fue usado para seleccionar las variables independientes que más significancia tenían para un modelo de regresión lineal. A partir de estas variables seleccionadas se realizó una selección de modelos de regresión para encontrar la combinación de variables que tuviera un valor de R-cuadrado ajustado óptimo, un valor bajo Cp de Mallows y un valor pequeño en los estadísticos que usan estos tres criterios de información: Akaike, Hannan-Quinn y Schwarz-Bayesian. Finalmente se analizó la multicolinealidad de las variables independientes, mediante el factor de inflación de la varianza V.I.F, donde valores mayores a 10 hacen referencia a que existe la probabilidad de que se tenga variables que aporten el mismo tipo de información. Esto ocurrió en varias ocasiones en variables meteorológicas como valores acumulados de precipitación y temperatura o índices espectrales que eran combinación lineal entre sí como MACARI y TCARI. En estos casos, algunas variables fueron eliminadas del modelo para conseguir que el V.I.F. de todas las variables del modelo fuese inferior a 10.
- Los modelos de regresión lineal multivariante para parcelas de matorral SH4 en la provincia de Valencia, tuvieron el mejor ajuste para el año 2019 donde el R-cuadrado ajustado para la media ponderada y media matorral fue de 63.62% y 65.06% respectivamente y en ambos casos se usaron tres variables independientes teniendo mayor importancia la variable T30. Al momento de realizar el modelo para los años 2019, 2020 y 2021 en conjunto, tanto para la media ponderada como media matorral, el R-cuadrado ajustado bajo a 56.15% y 54.75 respectivamente, y se pudo apreciar la importancia de las variables sin_DOY, p60 y VARI.
- En la validación de los modelos para la provincia de Valencia se verificó la importancia de analizar las especies vegetales presentes en cada parcela, como es el caso de la parcela 4 correspondiente al municipio de Gilet, donde el modelo creado se ajusta bien para el año 2019 pero se aleja en los años 2020 y 2021, lo cual puede deberse a la presencia de la especie *Stipa tenacissima* que solamente se encuentra en la muestras de Gilet y tienen valores muy bajos de HCV, lo cual hace que el modelo creado muestre desfases con las parcelas de validación.
- También se construyeron modelos para la especie *Rosmarinus officinalis*, una de las más representativas en las parcelas catalogadas como SH4. El R-cuadrado ajustado obtenido fue de 55.79% ajustándose muy bien en las parcelas de evaluación de Gilet y Chelva, pero no así para Llombai y Cortes de Pallás. Esto es debido a que, en estas parcelas la especie estudiada no es predominante y por tanto el modelo creado de HCV muestra un desajuste.
- Los modelos creados en las parcelas de matorral de las provincias de Alicante y Castellón obtuvieron ajustes de R-cuadrado que rodean el 60%. Aunque este ajuste es muy bueno se tiene que mencionar que las parcelas disponibles para el

entrenamiento son mucho menores, haciendo que el modelo sea mucho más restringido y por lo tanto pueda tener mejores resultados. En estas parcelas se realizó validación cruzada obteniendo R-cuadrado ajustados superiores al 60%.

- Se construyeron modelos de regresión lineal para las parcelas de arbolado considerando las categorías de modelos de combustible TU2, TU3 y en conjunto TU2|TU3, tanto en la media ponderada como en la media arbolado. Para la provincia de Valencia los modelos en su mayoría superan el 60% de R-cuadrado ajustado necesitando aproximadamente de 6 variables independientes. Se puede evidenciar que la aplicación de las categorías TU2 y TU3 por separado mejoran el ajuste, pero al momento de aplicarlas en conjunto el ajuste tiende a empeorar. Esto puede deberse a que el modelo de combustible TU2 caracteriza matorral de menos de 1m de altura bajo el dosel arbolado, mientras que el modelo TU3 caracteriza matorral de más de 1m de altura bajo el dosel arbolado.
- Para la provincia de Alicante no se pudo calcular el modelo TU3 debido a la falta de datos de campo, los resultados de TU2 y TU2|TU3, son muy similares superando el 60%. En la provincia de Castellón los modelos de media ponderada y media arbolado generados para TU2 fueron los que mejores resultados ofrecieron, alcanzaron obteniendo un R-cuadrado ajustado igual a 61.55% y 74.01% respectivamente. Se pudo ver que en la provincia de Alicante el modelo de TU3 tiene un ajuste bajo llegando a 57.99% para la media ponderada y 53.23% para la media de arbolado. Esto puede deberse a que esta provincia tiene la menor cantidad de puntos de muestreo por lo cual las observaciones de HCV se reducen considerablemente.

7. BIBLIOGRAFÍA

- Ahamed, T., Tian, L., Zhang, Y., & Ting, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass & Bioenergy*, 35(7), 2455-2469.
- Ali, A., Martelli, R., Lupia, F., & Barbanti, L. (2019). Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. *Remote sensing*, 11(20), 2384.
- Anand, A., Singh, S. K., & Kanga, S. (2018). Estimating the change in forest cover density and predicting NDVI for west Singhbhum using linear regression. *Int. J. Environ. Rehabil. Conserv.*, (9), 193-203.
- Balaguer Beser, A., & Ruiz Fernández, L. Á. (09 de 06 de 2021). *Selección de un modelo de regresión lineal múltiple para el cálculo de la precipitación media en verano*. Obtenido de RiuNet: <https://riunet.upv.es/handle/10251/167659>
- Berlanga-Silvente, V., & Vilá-Baños, R. (2014). Cómo obtener un Modelo de Regresión Logística Binaria. *REIRE, Revista d'Innovació i Recerca en Educació*, 7-8.
- Borunda, A. (20 de 05 de 2021). *Los incendios «zombis» del Ártico están vinculados al cambio climático*. Obtenido de <https://www.nationalgeographic.es/medio-ambiente/2021/05/los-incendios-zombis-del-artico-estan-vinculados-al-cambio-climatico>
- Cáceres, M., & Granda, V. (05 de 01 de 2022). *User guide to meteoland*. Obtenido de <https://cran.r-project.org/web/packages/meteoland/vignettes/UserGuide.html#interpolation-basics-and-parameters>
- Chuvieco, E., Cocero, D., Riaño, D., Martín, P., Martínez-Vega, J., de la Riva, J., & Pérez, F. (2004). Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote sensing environment*, 92(3), 322-331.
- Conselleria de Agricultura, D. R. (2019). *Modelos de combustible forestal*. Obtenido de <https://agroambient.gva.es/es/web/prevencion-de-incendios/models-de-combustible>
- Costa-Saura, J., Balaguer-Beser, Á., Ruiz, L., Pardo-Pascual, J., & Soriano-Sancho, J. (2021). Empirical Models for Spatio-Temporal Live Fuel Moisture Content Estimation in Mixed Mediterranean Vegetation Areas Using Sentinel-2 Indices and Meteorological Data. *Remote sensing*, 13(18),3726.
- Cunill Camprubi, Á., González-Moreno, P., & Resco de Dios, V. (2022). Live Fuel Moisture Content Mapping in the Mediterranean Basin Using Random Forests

- and Combining MODIS Spectral and Thermal Data. *Remote sensing*, 14(13),3162.
- De Castro, A. I., Plant, R. E., Six, J., & Peña, J. M. (2014). Mapas de calendario de cultivo y variables fenológicas mediante el análisis de imágenes MODIS y ASTER basado en objetos. *Revista de Teledetección*, (41), 29-36.
- Earthengine. (01 de 07 de 2022). *Google Earth Engine*. Obtenido de <https://earthengine.google.com/>
- Ehammer, A., Fritsch, S., Conrad, C., Lamers, J., & Dech, S. (2010). Statistical derivation of fPAR and LAI for irrigated cotton and rice in arid Uzbekistan by combining multi-temporal RapidEye data and ground measurements. *SPIE Remote Sensing*, (7824), 66-75.
- EPRIF, E. d. (s.f). *Los EPRIF Planificación y desarrollo de actuaciones*. MITECO. Obtenido de https://www.miteco.gob.es/es/biodiversidad/temas/incendios-forestales/loseprifplanificacionydesarrollodeactuaciones_compressed_tcm30-541706.pdf
- ESA. (01 de 07 de 2015). *SENTINEL 2*. Obtenido de https://www.esa.int/Space_in_Member_States/Spain/SENTINEL_2
- Ferreiro Gómez, A. (2016). *Análisis socioeconómico de los incendios forestales españoles y propuesta de rediseño de la estrategia de prevención-concienciación*. Valencia, España: Universitat Politècnica de València.
- Gao, B.-C. (1996). NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space . *Elsevier Science Inc*, 258.
- GEE. (01 de 06 de 2022). *Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. Obtenido de https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR#description
- GEE. (01 de 06 de 2022). *Statistics of an Image Region*. Obtenido de https://developers.google.com/earth-engine/guides/reducers_reduce_region
- Gilabert, M. A., González-Piqueras, J., & García-Haro, J. (1997). Acerca de los índices de vegetación. *Revista de Teledetección*, 8(1), 1-10.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289-298.
- Haboudane, D., Miller, J., Pattey, E., P.J., Z.-T., & Strachan, I. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sens Environ*, 90(3), 337-352.

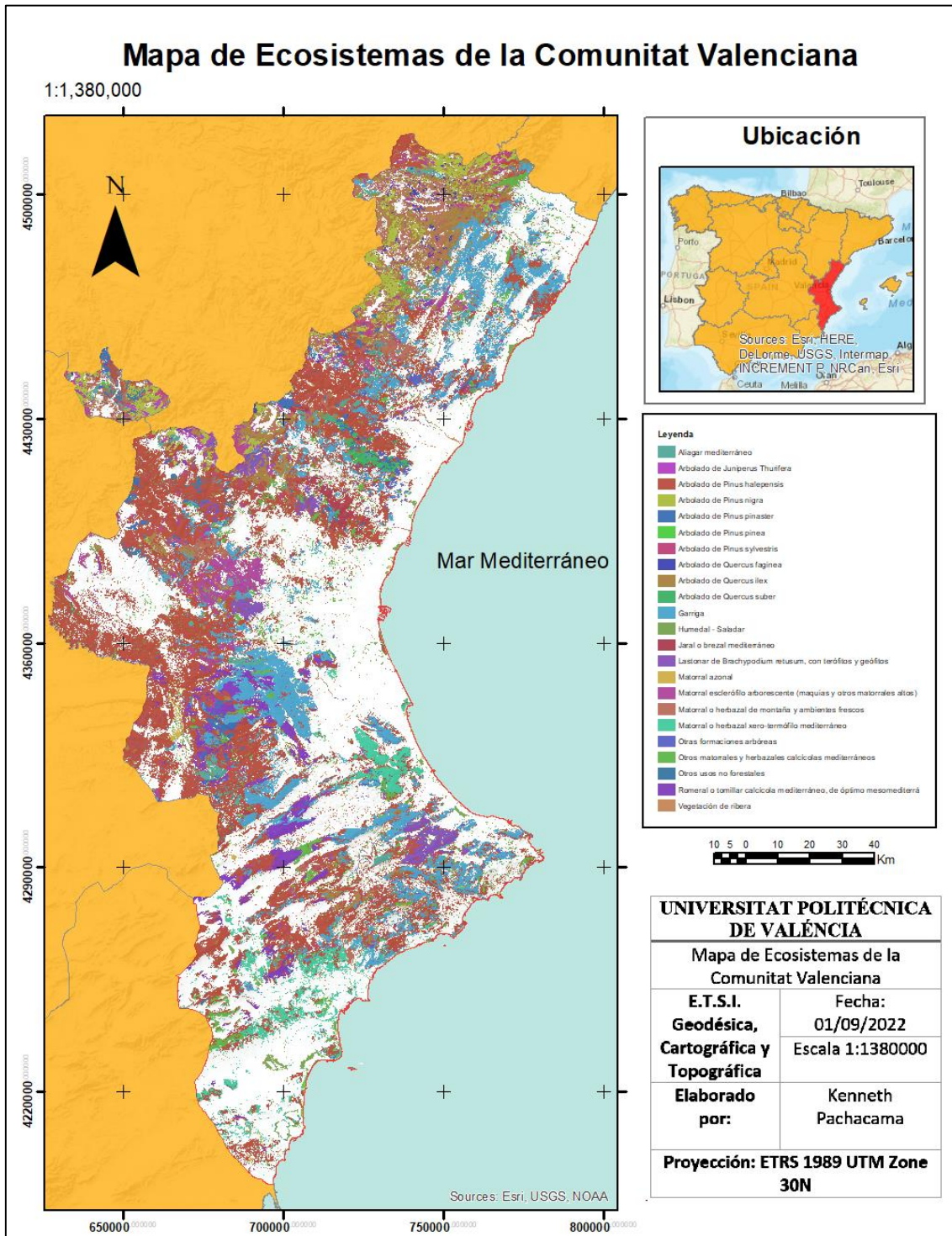
- Haboudane, D., Miller, J., Tremblay, N., Zarco-Tejada, P., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote sensing environment*, 81(2-3), 416-426.
- Heiskanen, J. (2006). Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*, 27(6), 1135-1158.
- Hernández, L., & et al. (2020). *EL PLANETA EN LLAMAS PROPUESTA IBÉRICA DE WWF PARA LA PREVENCIÓN DE INCENDIOS*. WWF España.
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1), 1-6.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.*, 83,195-213.
- Hunt Jr., E. R., Daughtry, C. S., Eitel, J. U., & Long, D. S. (2011). Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agronomy Journal*, 1090-1099.
- Lymburner, L., Beggs, P., & Jacobson, C. (2000). Estimation of canopy-average surface-specific leaf area using Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66(2), 183-192.
- Lymburner, L., Tan, P., Mueller, N., Thackway, R., Thankappan, M., Islam, A., . . . Senarath, U. (2011). The National Dynamic Land Cover Dataset. *Geoscience Australia*.
- Main, R., Cho, M. A., Mathieu, R., Oâ€™Kennedy, M. M., Ramoelo, A., & Koch, S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 751-761.
- Martin-StPaul, N., Pimont, F., Dupuy, J., Rigolot, E., Ruffault, J., Fargeon, H., . . . Toutchkov, M. (2018). Live fuel moisture content (LFMC) time series for multiple sites and species in the French Mediterranean area since 1996. *Annals of Forest Science*.
- Mataix-Solera, J., & Cerdà, A. (2009). *Incendios forestales en España. Ecosistemas terrestres y suelos. Efectos de los incendios forestales sobre los suelos en España*. 25-53: Universidad Miguel Hernández.
- MITECO. (2020). *Los incendios forestales en España 1 enero – 31 diciembre 2020 Avance Informativo* . Obtenido de Ministerio para la Trascición ecologica y el reto Demográfico:

https://www.miteco.gob.es/es/biodiversidad/estadisticas/avance_1_enero_31_diciembre_2020_tcm30-530816.pdf

- Moreno Alcaraz, J. L., Llacer, N., & Heras Ibáñez, J. D. (2006). Propuesta de un índice de vegetación acuática (IVAM) para la evaluación del estado trófico de los ríos de Castilla-La Mancha: A Comparación con otros índices bióticos. *Limnetica*, 821-838.
- Muñoz, P. (2013). *Apuntes de Teledetección: Índices de vegetación*. Chile: Alfabetización Satelital de Profesionales del Ministerio de Agricultura.
- Myoung, B., Kim, S. H., Nghiem, S. V., Jia, S., Whitney, K., & Kafatos, M. C. (2018). Estimating Live Fuel Moisture from MODIS Satellite Data for Wildfire Danger Assessment in Southern California USA. *Remote sensing*, 10(1), 87.
- PATFOR. (s.f). *PLAN DE ACCIÓN TERRITORIAL FORESTAL DE LA COMUNITAT VALENCIANA – PATFOR*. Obtenido de https://agroambient.gva.es/auto/montes-bosques/PATFOR/01_MEMORIA/PATFOR_Memoria_version_final.pdf
- Pearson, R. L., & Miller, L. D. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the short-grass Prairie, Pawnee National Grasslands, Colorado. *Proceedings of the Eighth International Symposium on Remote Sensing of Environment*, 1357-1381.
- Pellizzaro, G., Cesaraccio, C., Duce, P., Ventura, A., & Zara, P. (2007). Relationships between seasonal patterns of live fuel moisture and meteorological drought indices for Mediterranean shrubland species. *Int. J. Wildland Fire*, 232–241.
- Peterson, S. H., Roberts, D. A., & Dennison, P. E. (2008). Mapping live fuel moisture with MODIS data: A multiple regression approach. *Remote Sensing of Environment*, 4272–4284.
- Pons, X., Sevillano, E., Moré, G., Serra, P., Cornford, D., & Ninyerola, M. (2014). Distribución espacial de la incertidumbre en mapas de cubiertas obtenidos mediante teledetección. *Revista de la Asociación Española de Teledetección*, (42), 1-10.
- SER. (19 de 08 de 2022). *El incendio de Bejís supera las 20.000 hectáreas y la previsión meteorológica augura rebrotes y tormentas secas*. Obtenido de <https://cadenaser.com/comunitat-valenciana/2022/08/19/mas-de-25000-hectareas-arrasadas-por-los-incendios-de-vall-debo-y-bejis-que-amenaza-el-parque-natural-de-sierra-calderona-radio-valencia/>
- Soriano Sancho, J. L. (2016). *Análisis de la Humedad del Combustible Vivo (HCV) en la Comunitat Valenciana y la influencia de las variables meteorológicas*. Valencia, España: Universidad Politécnica de Valencia .

- STATGRAPHICS, U. M. (01 de 01 de 2020). *STATGRAPHICS Centurion XVI*. Obtenido de <https://www.statgraphics.net/wp-content/uploads/2015/03/Centurion-XVI-Manual-Principal.pdf>
- Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.*, 127–150.
- Wang, L., & Qu, J. J. (2007). NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical research letters*.
- Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 385-396.
- Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 1230-1241.
- Wu, J., Wang, D., & Bauer, M. E. (2007). Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field Crops Research*, 102(1), 33-42.
- WWF. (2020). *BOSQUES: INCENDIOS FORESTALES*. Obtenido de https://www.wwf.es/nuestro_trabajo/bosques/incendios_forestales/
- Yang, Y., Luo, J., Huang, Q., Wu, W., & Sun, Y. (2019). Weighted Double-Logistic Function Fitting Method for Reconstructing the High-Quality Sentinel-2 NDVI Time Series Data Set. *Remote sensing*, 11(20), 2342.
- Yebra, M., Dennison, P., Chuvieco, E., Riaño, D., Zylstra, P., Hunt Jr, R., . . . Jurdao, S. (2013). A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. *Remote Sensing Environment*, 455-468.
- Yebra, M., Quan, X., Riaño, D., Larraondo, P. R., Dijk, A. I., & Carya, G. J. (2018). A fuel moisture content and flammability monitoring methodology for continental Australia based on optical remote sensing. *Remote Sensing of Environment*, 261-262.
- Zhu, L., Webb, G., Yebra, M., Scortchini, G., & Miller, L. (2021). Live fuel moisture content estimation from MODIS: A deep learning approach. *Photogrammetry and remote sensing*, 83-84.

8. CARTOGRAFÍA



10 5 0 10 20 30 40 Km

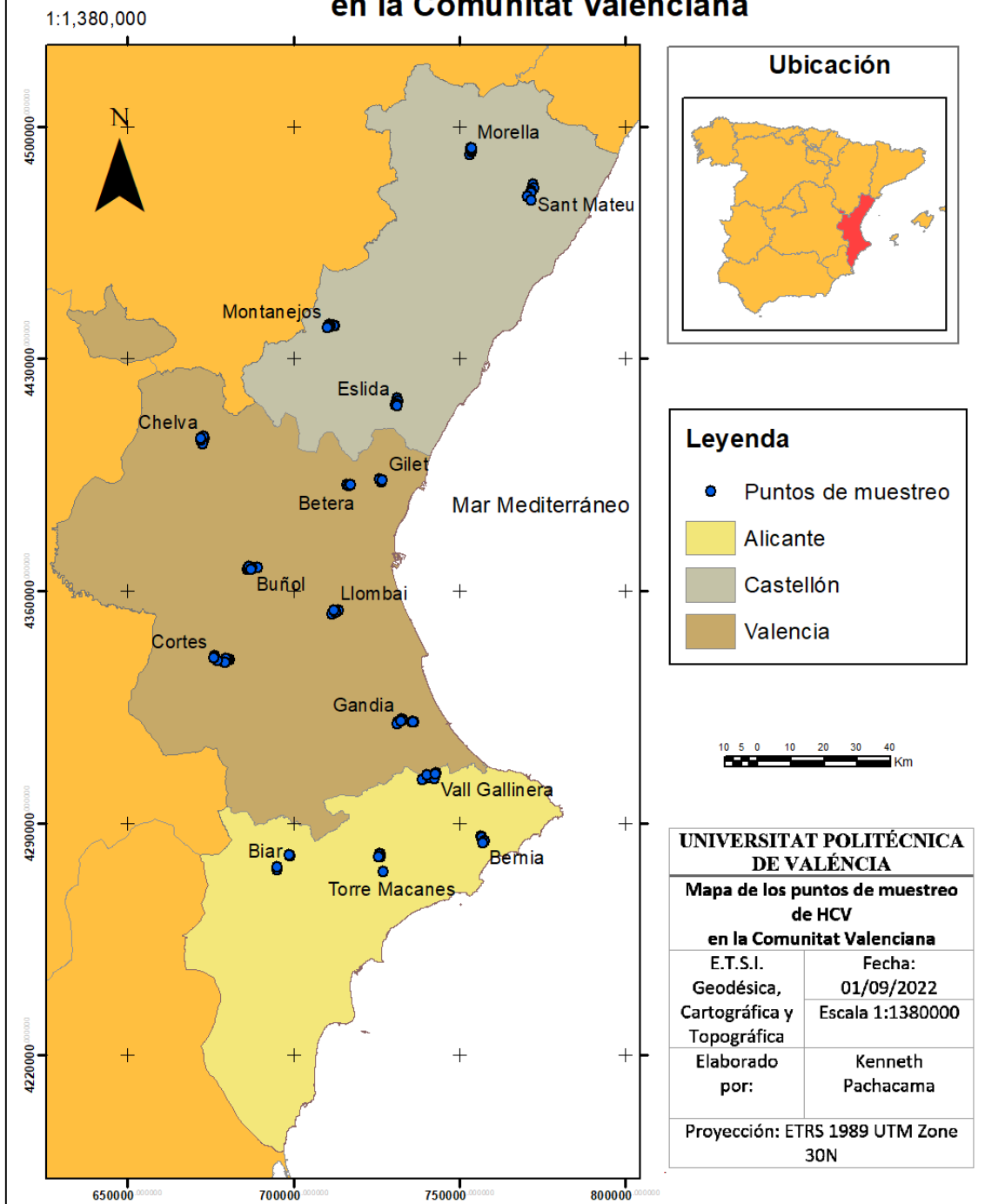
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Mapa de Ecosistemas de la Comunitat Valenciana

E.T.S.I. Geodésica, Cartográfica y Topográfica	Fecha: 01/09/2022
Elaborado por: Kenneth Pachacama	Escala 1:1380000

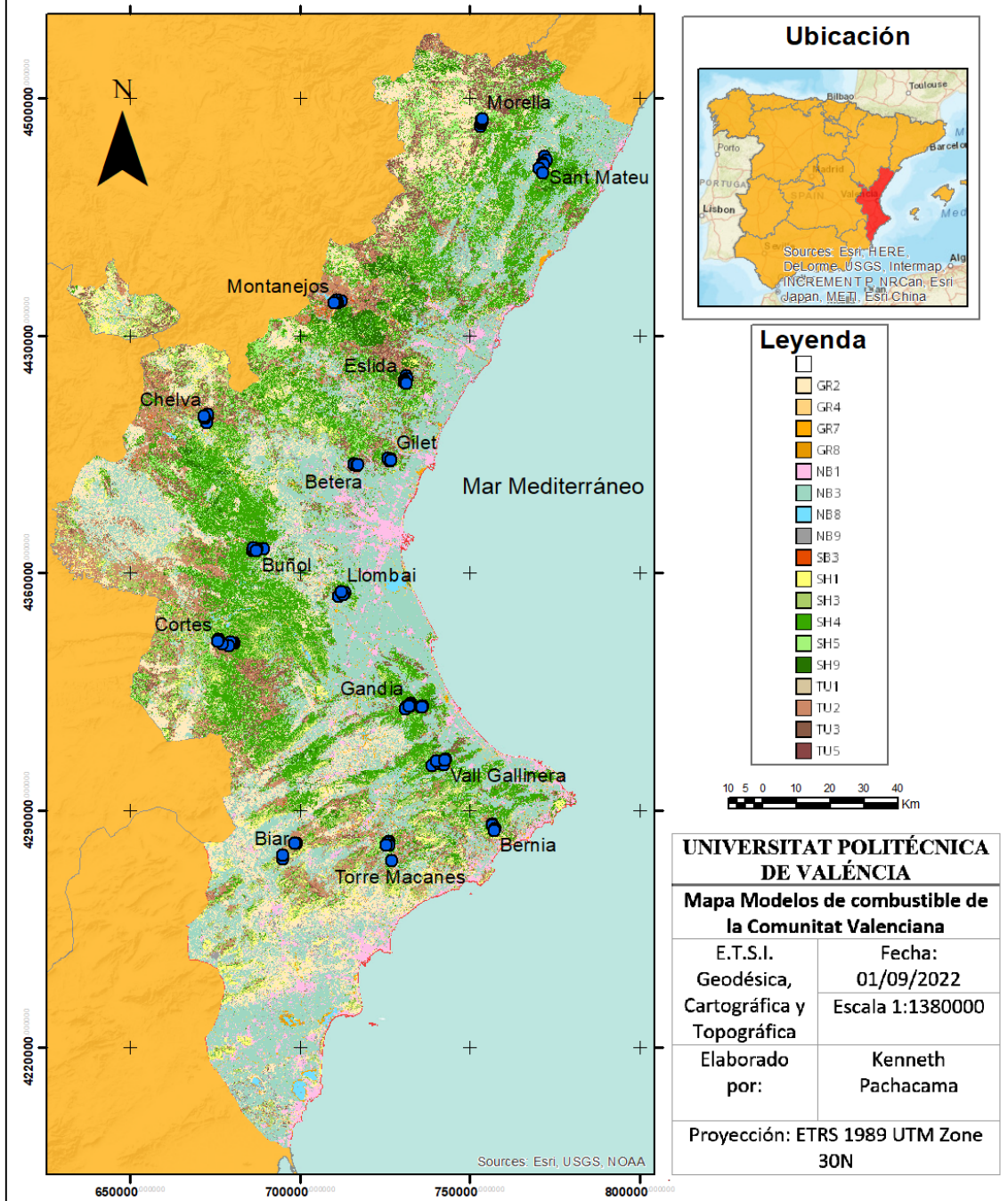
Proyección: ETRS 1989 UTM Zone 30N

Mapa de los puntos de muestreo de HCV en la Comunitat Valenciana

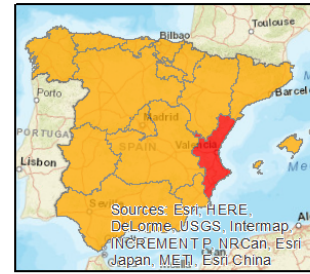


Mapa Modelos de combustible de la Comunitat Valenciana

1:1,380,000



Ubicación



Legenda

-
- GR2
- GR4
- GR7
- GR8
- NB1
- NB3
- NB8
- NB9
- SB3
- SH1
- SH3
- SH4
- SH5
- SH9
- TU1
- TU2
- TU3
- TU5

10 5 0 10 20 30 40 Km

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

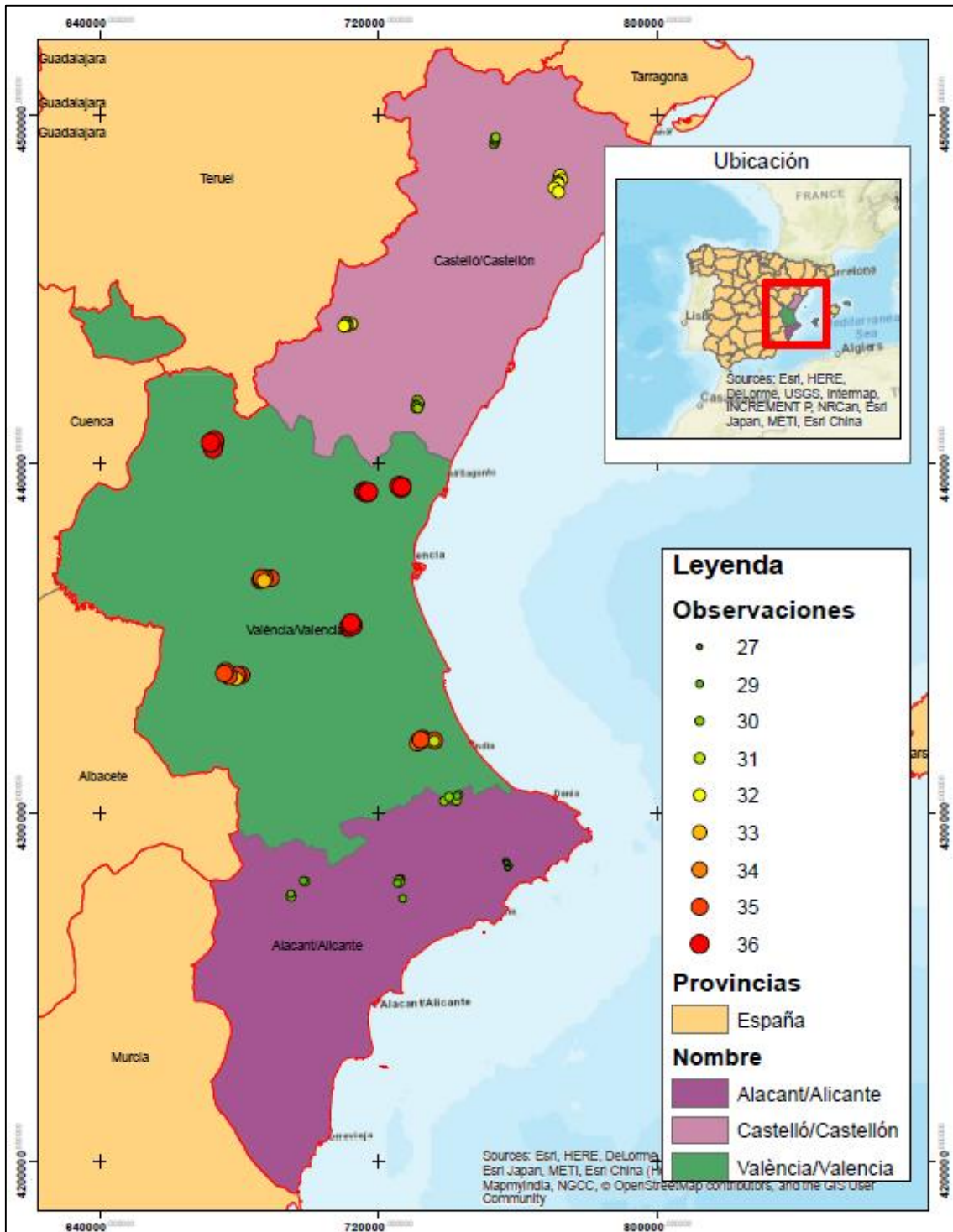
Mapa Modelos de combustible de la Comunitat Valenciana

E.T.S.I. Geodésica, Cartográfica y Topográfica	Fecha: 01/09/2022
	Escala 1:1380000

Elaborado por:	Kenneth Pachacama
----------------	-------------------

Proyección: ETRS 1989 UTM Zone 30N

Mapa Densidad de observaciones de los puntos de muestreo en la Comunitat Valenciana



9. ANEXOS

9.1 Anexo 1: Código para la creación de índices espectrales en GEE

```
//colección de imágenes

var plots = ee.FeatureCollection("users/kenneth/3_parcelas_B30");
var sent2a=ee.ImageCollection("COPERNICUS/S2_SR")
    .filterDate('2020-01-01', '2021-11-30')
    .filterMetadata('MGRS_TILE', "contains","30")
    .filterBounds(table2);

//función de máscara de nubes
function maskS2clouds(image) {
  var qa = image.select('QA60')

  // Bits 10 and 11 are clouds and cirrus, respectively.
  var cloudBitMask = 1 << 10;
  var cirrusBitMask = 1 << 11;

  // Both flags should be set to zero, indicating clear conditions.
  var mask = qa.bitwiseAnd(cloudBitMask).eq(0).and(
    qa.bitwiseAnd(cirrusBitMask).eq(0))

  // Return the masked and scaled data, without the QA bands.
  return image.updateMask(mask).divide(1)
    .select("B.*")
    .copyProperties(image, ["system:time_start"])
}

print(sent2a.first())
Map.addLayer(sent2a.first().select('QA60'),undefined,'nubes')
Map.addLayer(sent2a.first(),imageVisParam,'con nubes')

//aplicación función de máscara de nubes
var sent3a=sent2a.map(maskS2clouds)
//Map.addLayer(sent3a.first(),undefined,'sin nubes')

/////índices espectrales
/////función para calcular NDVI
function CalcularNDVI(image) {
  var ndvi = image.normalizedDifference(['B8', 'B4']).rename('NDVI');
  return image.addBands(ndvi);
}
/////función para calcular EVI
function CalcularEVI(image) {
  var evi=image.expression('2.5*((B8-B4)/(B8 + 6 * B4 - 7.5 * B2 + 10000))',{ 'B8':image.select('B8'),'B4':image.select('B4'),'B2':image.select('B2')}).rename('EVI');
  return image.addBands(evi);
}
/////función para calcular SAVI
function CalcularSAVI(image) {
  var savi=image.expression('((B8-B4)/(B8+B4+5000))*1.5',{ 'B8':image.select('B8'),'B4':image.select('B4')}).rename('SAVI');
  return image.addBands(savi);
}
```



```

////función para calcular NDMI
function CalcularNDMI(image) {
    var ndmi=image.expression('float (B8-
B11)/float (B8+B11)', {'B8':image.select('B8'),'B11':image.select('B11')
}).rename('NDMI');
    return image.addBands(ndmi);
}
////función para calcular VARI
function CalcularVARI(image) {
    var vari=image.expression('float (B3-B4)/float (B3+B4-
B2)', {'B3':image.select('B3'),'B4':image.select('B4'),'B2':image.selec
t('B2')}).rename('VARI');
    return image.addBands(vari);
}
////función para calcular NDWI
function CalcularNDWI(image) {
    var ndwi=image.expression('float (B8-
B12)/float (B8+B12)', {'B8':image.select('B8'),'B12':image.select('B12')
}).rename('NDWI');
    return image.addBands(ndwi);
}
////función para calcular Vgreen
function CalcularVgreen(image) {
    var vgreen=image.expression('float (B3-
B5)/float (B3+B5)', {'B3':image.select('B3'),'B5':image.select('B5')}).r
ename('Vgreen');
    return image.addBands(vgreen);
}
////función para calcular MSI
function CalcularMSI(image) {
    var
msi=image.expression('float (B11)/float (B8)', {'B11':image.select('B11')
,'B8':image.select('B8')}).rename('MSI');
    return image.addBands(msi);
}
////función para calcular RVI
function CalcularRVI(image) {
    var
rvi=image.expression('float (B8)/float (B4)', {'B4':image.select('B4'),'B
8':image.select('B8')}).rename('RVI');
    return image.addBands(rvi);
}
////función para calcular OSAVI
function CalcularOSAVI(image) {
    var osavi=image.expression('(1+0.16)*(B8-
B4)/(B8+B4+1600)', {'B4':image.select('B4'),'B8':image.select('B8')}).r
ename('OSAVI');
    return image.addBands(osavi);
}
////función para calcular TCARI
function CalcularTCARI(image) {
    var tcari=image.expression('3*((B5-B4)/10000-0.2*((B5-
B3)/10000)*(B5/B4)', {'B4':image.select('B4'),'B5':image.select('B5'),
'B3':image.select('B3')}).rename('TCARI');
    return image.addBands(tcari);
}
////función para calcular MCARI
function CalcularMCARI(image) {
    var

```

```

mcari=image.expression('TCARI/3',{ 'TCARI':image.select('TCARI')}).rename('MCARI');
return image.addBands(mcari);
}
////función para calcular NMDI
function CalcularNMDI(image) {
var nmdi=image.expression('float(B8A-(B11-B12))/float(B8A+(B11-B12))',{ 'B8A':image.select('B8A'),'B11':image.select('B11'),'B12':image.select('B12')}).rename('NMDI');
return image.addBands(nmdi);
}
////función para calcular TCARI_OSAVI
function CalcularTCARI_OSAVI(image) {
var
tcariosavi=image.expression('TCARI/OSAVI',{ 'TCARI':image.select('TCARI'),'OSAVI':image.select('OSAVI')}).rename('TCARI_OSAVI');
return image.addBands(tcariosavi);
}
////función para calcular MCARI_OSAVI
function CalcularMCARI_OSAVI(image) {
var
mcariosavi=image.expression('MCARI/OSAVI',{ 'MCARI':image.select('MCARI'),'OSAVI':image.select('OSAVI')}).rename('MCARI_OSAVI');
return image.addBands(mcariosavi);
}
////función para calcular SLA
function CalcularSLA(image) {
var
sla=image.expression('float(B8)/float(B5+B12)',{ 'B8':image.select('B8'),'B5':image.select('B5'),'B12':image.select('B12')}).rename('SLA');
return image.addBands(sla);
}
////función para calcular ARVI
function CalcularARVI(image) {
var arvi=image.expression('(B8-(2*B4)+B2)/(B8+(2*B4)+B2)',{ 'B8':image.select('B8'),'B4':image.select('B4'),'B2':image.select('B2')}).rename('ARVI');
return image.addBands(arvi);
}

////aplicación de funciones para calcular índices
var NDVICol=sent3a.map(CalcularNDVI);
var NDVICol=NDVICol.map(CalcularEVI);
var NDVICol=NDVICol.map(CalcularSAVI);
var NDVICol=NDVICol.map(CalcularNMDI);
var NDVICol=NDVICol.map(CalcularVARI);
var NDVICol=NDVICol.map(CalcularNDWI);
var NDVICol=NDVICol.map(CalcularVgreen);
var NDVICol=NDVICol.map(CalcularMSI);
var NDVICol=NDVICol.map(CalcularRVI);
var NDVICol=NDVICol.map(CalcularOSAVI);
var NDVICol=NDVICol.map(CalcularTCARI);
var NDVICol=NDVICol.map(CalcularMCARI);
var NDVICol=NDVICol.map(CalcularNMDI);
var NDVICol=NDVICol.map(CalcularTCARI_OSAVI);
var NDVICol=NDVICol.map(CalcularMCARI_OSAVI);
var NDVICol=NDVICol.map(CalcularSLA);
var NDVICol=NDVICol.map(CalcularARVI);

```

```

print(NDVICol)
Map.addLayer(NDVICol.select('NDVI').first(),undefined,'NDVI')

///// aplicar el reductor

var results = NDVICol.map(function(image) {
  return image.reduceRegions({
    collection: plots,
    reducer: ee.Reducer.stdDev(),
    scale: 20
  })
}).filter(ee.Filter.neq('Bl'," ")).flatten()
print(results.first())

// Exportar resultados
Export.table.toDrive(results,'17Indices_st')

Map.centerObject(sent2a.first(),10)
Map.addLayer(table,{color:'red'},'parcelas')

```

9.2 Anexo 2: Código criterio de la mediana en Python

```

# librerías
import pandas as pd
import statistics as st
import numpy as np
import math
#Lectura de archivo
archivo='D:/Master/TFM/analisis_data/datos/ajustado/Validacion/6indices
  3parcelas.xlsx'
sheet_name = 4
header = 0
df = pd.read_excel(archivo, sheet_name, header)
#print(df.head())
#Lista de parcelas
lista_plot=list(df['Name'].unique())
lista_plot.sort()
lista_indices=[]
for i in df:
  lista_indices.append(i)
lista_indices=lista_indices[2:]
# función para detectar anomalías
def anomalos(df,indice):
  lista=[]
  for i in df.index:
    data=df[indice][i]
    lista.append(data)
  #eliminar valores nan
  lista = [x for x in lista if np.isnan(x) == False]
  #calcular desviación estándar
  st_dev = np.std(lista,ddof=1)
  print(st_dev,indice)
  #print(st_dev)
  lista_secuencia=[]
  resultado=[lista[0]]
  for i in df.index:
    try:

```

```

        lista_secuencia.append(float(df[indice][i-1]))
        lista_secuencia.append(float(df[indice][i]))
        lista_secuencia.append(float(df[indice][i+1]))
        mediana=st.median(lista_secuencia)
        #print (mediana,lista_secuencia)
        lista_secuencia=[]
        name=df['Name'][i]
        date=df['Date'][i]
        valor=df[indice][i]
        #aplicacion del criterio de la mediana
        if abs(float(df[indice][i]-mediana))>(st_dev*1):
            resultado.append("")
        else:
            resultado.append(valor)
    except:
        pass
if len(df)>1:
    resultado.append(lista[len(lista)-1])

    return resultado
##formación y exportación del archivo resultante
f =
open("D:/Master/TFM/analisis_data/datos/ajustado/Validacion/6indices_c
orregido.csv", "w")
f.write("Name, Date, ARVI, EVI, MCARI, MCARI_OSAVI, MSI, NDMI, NDVI, NDWI, NMDI,
OSAVI, RVI, SAVI, SLA, TCARI, TCARI_OSAVI, VARI, Vgreen\n")
lista_name=list(df['Name'])
lista_date=[]
for i in df.index:
    dia=str(df['Date'][i].day).rjust(2, '0')
    mes=str(df['Date'][i].month).rjust(2, '0')
    año=str(df['Date'][i].year).rjust(2, '0')
    fecha=dia+"/"+mes+"/"+año
    lista_date.append(fecha)
lista_entera=[lista_name,lista_date]
for j in lista_indices:
    lista=[]
    for i in lista_plot:
        df_filter=df.loc[df['Name'] == i].reset_index()
        #print (len(df_filter))
        a=anomalos(df_filter,j)
        #print (len(a))
        lista=lista+a
    lista_entera.append(lista)
for j in range(len(lista_entera[0])):
    txt=[]
    for i in lista_entera:
        txt.append(i[j])
    txt=str(txt).replace(",","").replace(" ","").replace("'",'"')+"\n"
    f.write(txt)
f.close()

```

9.3 Anexo 3: código del filtro *Savitzky-Golay* en Python

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.signal import savgol_filter
import statistics as st
#y_filtered=savgol_filter(y,99,3)
#lectura de datos
archivo='D:/Master/TFM/ analisis_data/datos/ajustado/Validacion/6indices_3parcelas_corregido.xlsx'
sheet_name = 0
header = 0
df = pd.read_excel(archivo, sheet_name, header)
lista_plot=list(df['Name'].unique())
lista_plot.sort()
lista_indices=[]
for i in df:
    lista_indices.append(i)
lista_indices=lista_indices[2:]
##funcion para filtro savgol
def filtro_savgol(df,indice):
    print(indice)
    lista=[]
    for i in df.index:
        data=df[indice][i]
        lista.append(data)
    #print(lista)
    lista = pd.Series(lista)
    lista_interp = lista.interpolate(method="linear")
    #print(lista_interp)
    #aplicacion del filtro savgol ventana 5 polinomio 2
    lista_filtered=savgol_filter(lista_interp,5,2)
    #print(lista_filtered)
    return lista_filtered.tolist()
    # return lista_filtered.tolist(), lista
# df_filter=df.loc[df['Name'] == 88].reset_index()
# y_filtered,y=filtro_savgol(df_filter,'NDVI')
# x=list(range(0,len(y_filtered)))

# fig = plt.figure()
# ax=fig.subplots()
# p=ax.plot(x,y,'-*')
# p=ax.plot(x,y_filtered,'g-*')
# plt.show()
## formacion y exportacion de resultados
f =
open("D:/Master/TFM/ analisis_data/datos/ajustado/Validacion/6indices_3
parcelas_savgol.csv", "w")
f.write("Name,Date,ARVI,EVI,MCARI,MCARI_OSAVI,MSI,NDMI,NDVI,NDWI,NMDI,
OSAVI,RVI,SAVI,SLA,TCARI,TCARI_OSAVI,VARI,Vgreen\n")
lista_name=list(df['Name'])
lista_date=[]
for i in df.index:
    dia=str(df['Date'][i].day).rjust(2, '0')
    mes=str(df['Date'][i].month).rjust(2, '0')
    año=str(df['Date'][i].year).rjust(2, '0')
    fecha=dia+"/"+mes+"/"+año
```

```

    lista_date.append(fecha)
lista_entera=[lista_name,lista_date]
for j in lista_indices:
    lista=[]
    for i in lista_plot:
        df_filter=df.loc[df['Name'] == i].reset_index()
        #print(len(df_filter))
        a=filtro_savgol(df_filter,j)
        #print(len(a))
        lista=lista+a
    lista_entera.append(lista)

for j in range(len(lista_entera[0])):
    txt=[]
    for i in lista_entera:
        txt.append(i[j])
    txt=str(txt).replace("[", "").replace("]", "").replace("'", "")+"\n"
    f.write(txt)

f.close()

```

9.4 Anexo4: código para unir datos de campo con índices espectrales por fechas más cercanas

```

import pandas as pd
import numpy as np
#lectura de archivos
archivo='D:/Master/TFM/analisis_data/datos/ajustado/88_Parcelas_Base_d
atos_junio2020_noviembre2021_copia.xlsx'
sheet_name = 0
header = 0
df_campo = pd.read_excel(archivo, sheet_name, header)
lista_parcelas=list(df_campo['Parcela'].unique())
archivo='D:/Master/TFM/analisis_data/datos/ajustado/6Indices_19_21_cor
regido.xlsx'
sheet_name = 2
header = 0
df_indices = pd.read_excel(archivo, sheet_name, header)
lista_indices=[]
for i in df_indices:
    lista_indices.append(i)
lista_indices=lista_indices[2:]
#función para ordenar los datos de campo y lo índices espectrales
def order(parcela,df_campo,df_indices,indice):
    df_campo=df_campo.loc[df_campo['Parcela'] ==
parcela].reset_index()
    lista_fechas=list(df_campo['date_LFM'])
    #lista_parcelas=list(df_campo['Parcela'].unique())
    #####lectura archivo #####
    df_indices=df_indices.loc[df_indices['Name'] ==
parcela].reset_index()
    nan_value=float("Nan")
    df_indices.replace(" ",nan_value,inplace=True)
    df_indices.dropna(subset=[indice],inplace=True)
    #print(df_indices)
    lista_fechas2=[]
    for i in df_indices['Date'].index:

```

```

        fecha=df_indices['Date'][i]
        #fecha=numpy.datetime64(fecha)
        lista_fechas2.append(fecha)
# print(lista_fechas)
# print(lista_fechas2)
##Función buscar fechas más cercanas#####
##i=fecha de campo, a es igual fecha más cercana del índice
def nearest(items, pivot):
    return min(items, key=lambda x: abs(x - pivot))
for i in lista_fechas:
    #print('resultado:', i, "->", a)
    #print(parcela)
    try:
        a=nearest(lista_fechas2, i)
        filterbase=df_indices.loc[df_indices['Date'] ==a
].reset_index()

txt=str(i)+", "+str(parcela)+", "+str(parcela)+", "+str(a)+", "+str(filter
base[indice][0])+"\n"
        f.write(txt)
    except:
        a='no hay fecha'

txt=str(i)+", "+str(parcela)+", "+str(parcela)+", "+str(a)+", "+Error+"\n"
        f.write(txt)
##formación y exportación de resultados
for i in lista_indices:
    print(i)
    ruta="./resultados/"+i+".csv"
    print(ruta)
    f = open(ruta, "w")
    f.write("date_LFM,Parcela,Name_10mS,Date_50buf,indice_50buf+"\n")
    for j in lista_parcelas:
        order(j, df_campo, df_indices, i)
    f.close()

```