



Escuela
Politécnica
Superior

Técnicas de aprendizaje automático aplicadas a la estimación del estado de cultivos mediante series temporales



Grado en Ingeniería en Sonido e
Imagen en Telecomunicación

Trabajo Fin de Grado

Autor:

Anaida Fernández García

Tutor/es:

Juan Manuel López Sánchez

Tomás Martínez Marín

Julio 2020

Técnicas de aprendizaje automático aplicadas a la estimación del estado de cultivos mediante series temporales

Autor

Anaida Fernández García

Tutor/es

Juan Manuel López Sánchez

Dpto. de Física, Ing. Sistemas y Teoría de la Señal

Tomás Martínez Marín

Dpto. de Física, Ing. Sistemas y Teoría de la Señal



Grado en Ingeniería en Sonido e Imagen en Telecomunicación



Escuela
Politécnica
Superior



Universitat d'Alacant
Universidad de Alicante

ALICANTE, Julio 2020

Justificación y Objetivos

“Las razones que me han llevado a realizar este Trabajo Fin de Grado (TFG) son colaborar en una investigación en la que el Departamento de Física, Ingeniería de Sistemas y Teoría de la Señal de la Escuela Politécnica Superior lleva trabajando más de 6 años con amplias expectativas de futuro, profundizar mi conocimiento sobre área de las Telecomunicaciones que engloba los sistemas radar y las comunicaciones vía satélite y, por último, adentrarme en las tecnologías emergentes que tanta repercusión van a tener en nuestra vida como son las técnicas de aprendizaje automático o Machine Learning (ML).”

Agradecimientos

Es a ellos a quien dedico este trabajo.

La distancia, que es el impedimento principal del progreso de la humanidad, será completamente superada, en palabra y acción.

*La humanidad estará unida, las guerras serán imposibles,
y la paz reinará en todo el planeta.*

Nikola Tesla.

1 Resultados

1.1 Método por parcelas

1.1.1 Optimización

La optimización de los datos de entrada tiene su base en la elección de las variables de entrada y de los sets de parcelas de entrenamiento y test. Los mejores resultados para todos los casos el uso de un conjunto de 6 parcelas para entrenamiento y 1 para la evaluación, y como datos de entrada al sistema todos los mencionados anteriormente: media por día y parcela de VV, VH y ratio VH/VV y la desviación estándar de cada uno de ellos. En cuanto a la optimización del regresor, se basa en la determinación del número de árboles que lo componen. La relación entre el número de árboles para un modelo y el coeficiente de determinación es un buen descriptor para la elección de este parámetro, como se presenta en la imagen de ejemplo 1.1 con el caso de salida Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH). En ella, se puede ver que una vez alcanzado cierto nivel de coeficiente, la mejora de este en relación al aumento del número de árboles no es significativa con respecto al costo computacional y a la complejidad del sistema que se crea.

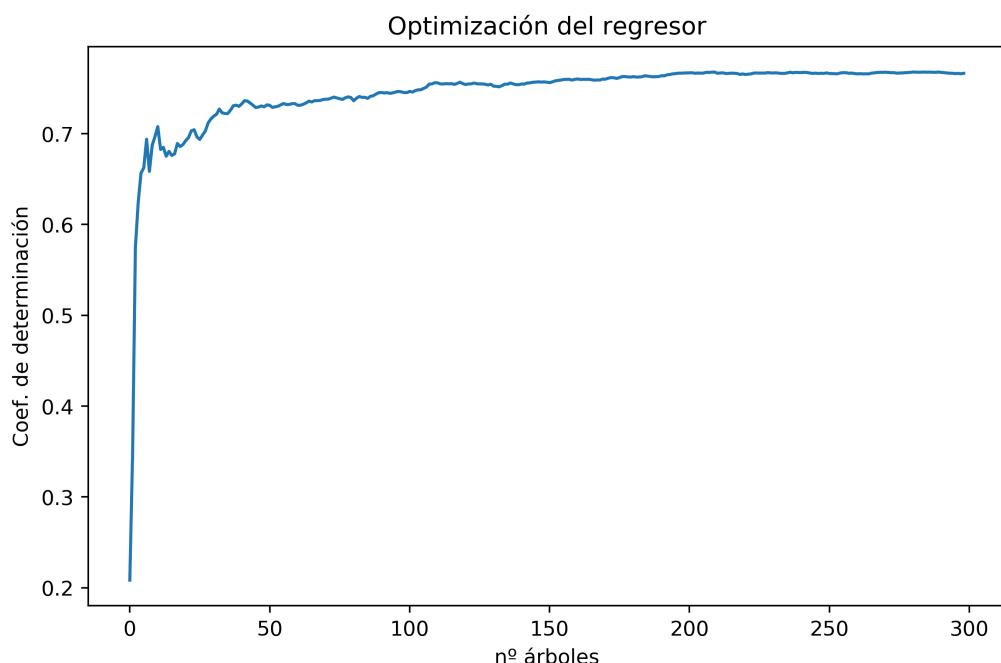


Figura 1.1: Optimización del número de árboles para Random Forest Regressor (RFR) en el modelo de salida BBCH

Finalmente, los parámetros utilizados en este método para cada caso se presentan en la tabla 1.1, donde se pueden ver: la parcela utilizada para el periodo de test del modelo; siendo el resto de parcelas utilizadas en el entrenamiento, y el número de árboles óptimo para RFR, de acuerdo con la evolución del coeficiente de determinación para cada caso.

	BBCH	Altura	BBCH&Altura
Parcela de test	'Mínima'	'Puntal'	'Mínima'
Número de árboles	42	77	54

Tabla 1.1: Parámetros de optimización de entrada y modelo

Como se puede observar, los 3 casos de este método constan de un número óptimo de árboles de, al menos, el mismo orden. Cabe destacar que el aumento de complejidad en el modelo, sobre todo en el caso de la altura como salida del sistema.

1.1.2 Salidas del modelo

1.1.2.1 Salidas de función de densidad de probabilidad

Como ya se ha comentado, las salidas del modelo son por defecto predicciones únicas, pero el marco de trabajo anterior demanda salidas en formato Probability Density Function (PDF) para ser integradas con el resto del modelo. En las figuras 1.2 (modelo de salida BBCH), 1.3 (modelo de salida altura) y 1.4 (modelo de ambas salidas) se puede apreciar cómo son algunas de estas salidas, siendo ejemplos extraídos a partir de los mismos datos para los 3 casos estudiados.

Las PDFs se presentan normalizadas con respecto a la salida con mayor probabilidad. En estos ejemplos se puede ver como, aunque hay una salida que predomina con respecto al resto, no se excluyen las demás soluciones posibles. Esto facilita la integración con el modelo de predicción temporal ya que se pueden combinar ambas salidas PDF para ver dónde coinciden y con qué probabilidad. En general, los resultados de la predicción temporal son bastante certeros, con oportunidad de fallo para ajustes finos en cultivos que hayan podido sufrir retrasos o adelantos en su desarrollo típico. Es ahí donde la contribución de este modelo es importante, aportando información extra en tiempo real y creando una predicción de cuáles serían los posibles estados de desarrollo en los que se encuentra el cultivo según la información actual de satélite.

Las 3 figuras 1.2, 1.3 y 1.4 han sido generadas con los mismos datos de entrada, es decir, los mismos datos de satélite en la misma parcela y fecha, por lo que las salidas para los modelos generados de estimación de BBCH (Figura 1.2) y altura (Figura 1.3) como modelos independientes se pueden comparar con las salidas del modelo de estimación de ambas (Figura 1.4). La principal diferencia que encontramos entre los mismos tipos de datos de salida para cada modelo es que, en general, las salidas individuales presentan una estimación principal con una diferencia de probabilidad mucho mayor con respecto al resto que las estimaciones del modelo de doble salida. El rango para cada salida se mantiene bastante similar en ambos modelos, además del valor con la probabilidad más alta, aunque la disminución de diferencias

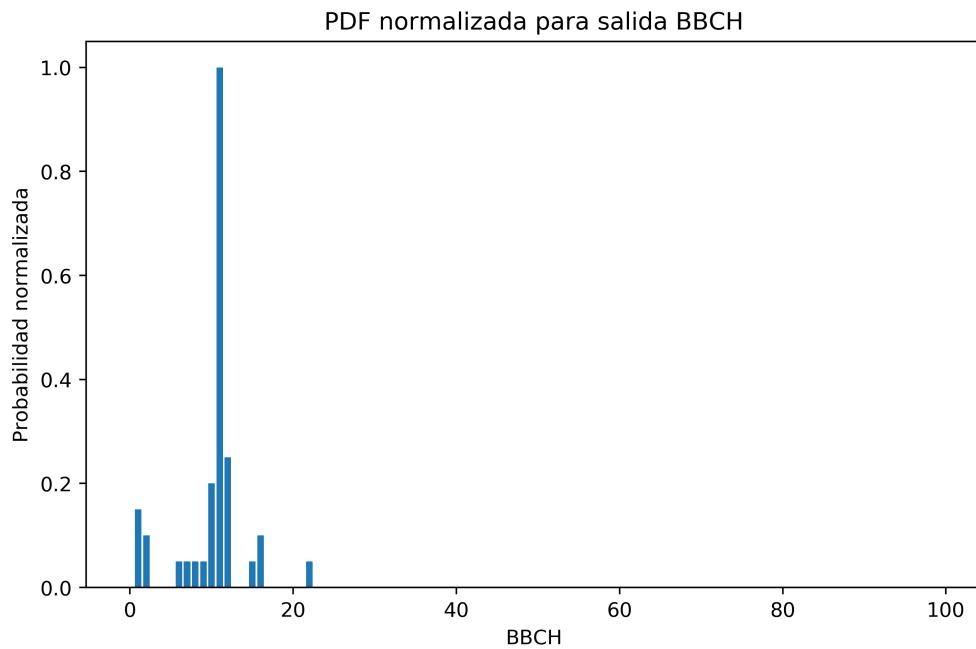


Figura 1.2: Ejemplo de salida PDF normalizada del modelo para estimación de BBCH.

con las demás estimaciones, sobre todo en el caso de la altura (1.4b), indica que ese sistema es menos preciso y estable. El hecho de que sea la salida de la altura la que se vea más perjudicada en este modelo de dos salidas puede deberse a que la optimización de este, en cuanto a número de árboles del regresor, es más similar, y por lo tanto más beneficiosa, con respecto al caso individual de estimación de BBCH.

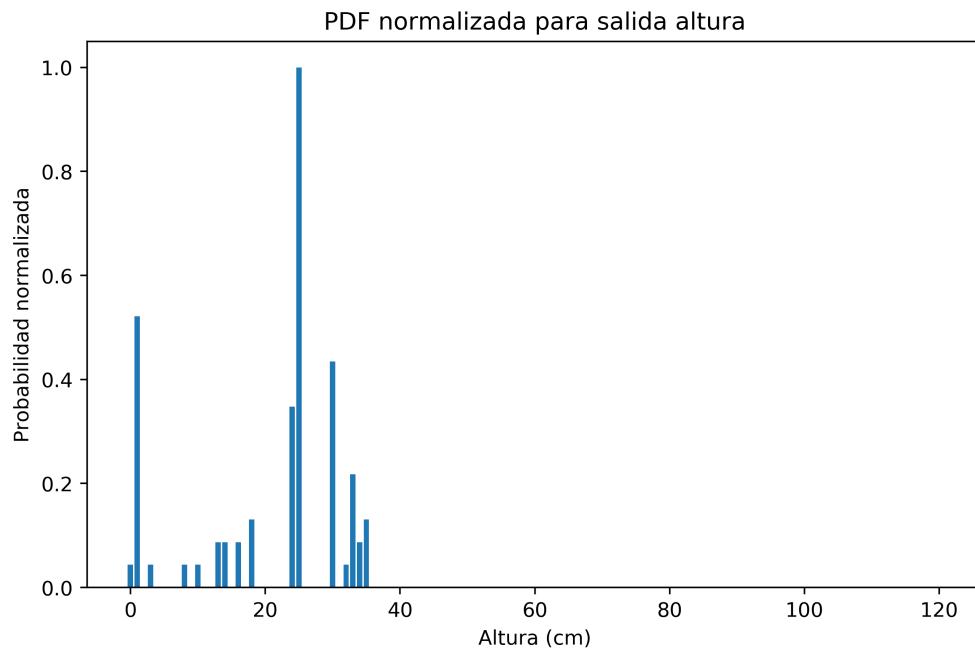
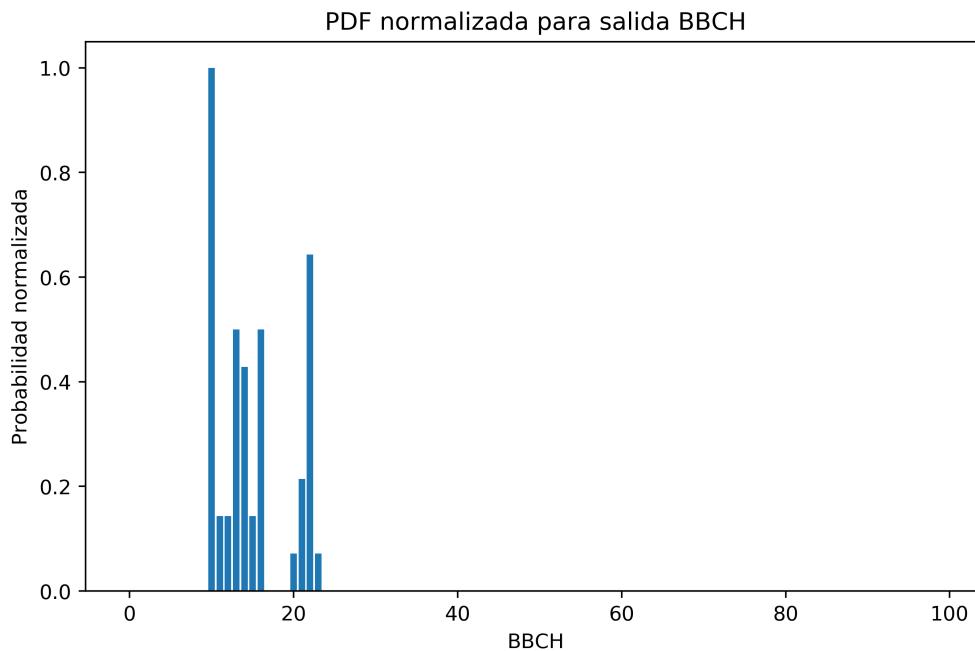
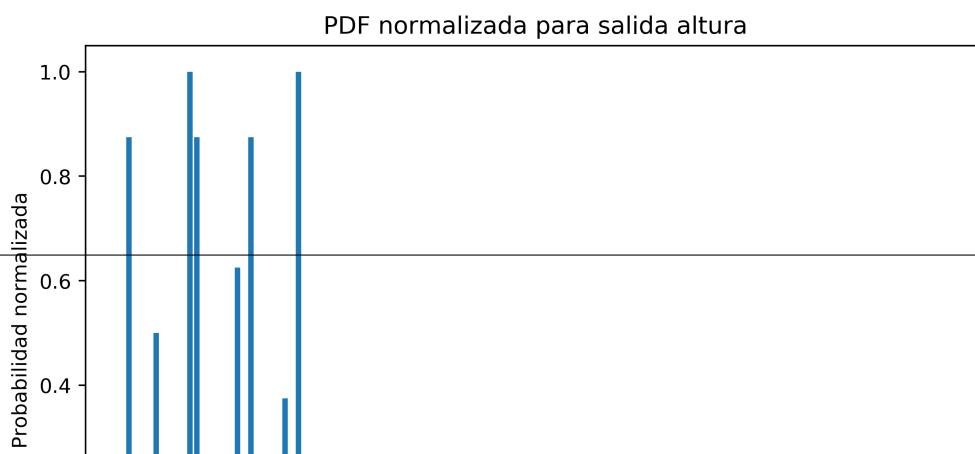


Figura 1.3: Ejemplo de salida PDF normalizada del modelo para estimación de la altura.



(a) Ejemplo del modelo de doble salida: PDF estimación de BBCH



1.1.2.2 Salidas de valor único

Para continuar con la presentación de los resultados obtenidos, se ilustran en las figuras 1.5, 1.6 y 1.7 las comparaciones de las soluciones únicas obtenidas (salida con máxima probabilidad) y los datos de tierra tomados correspondientes para la parcela de test en 2018: BBCH general, máxima y mínima en la parcela, casos individual (1.5) y de doble salida (1.7a), o la altura general del cultivo, también para ambos casos (figuras 1.6 y 1.7b, respectivamente). Estas salidas, aunque carecen de la información completa, son las más sencillas de comparar y evaluar con los resultados reales, ya que estos también son un único valor.

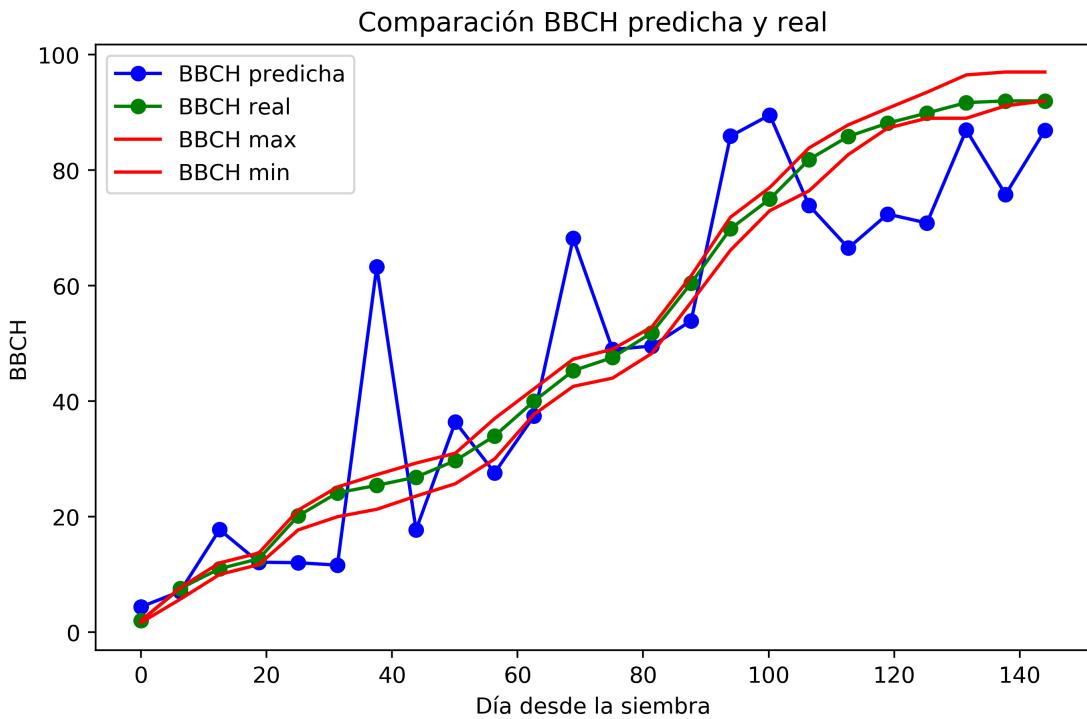


Figura 1.5: Comparación de la salida predicha y la verdad de tierra del modelo para estimación de BBCH.

Comenzando por los modelos de predicción de BBCH representados en las figuras 1.5 y 1.7a, se pueden observar estimaciones y errores muy parecidos en ambos modelos. Estos son fácilmente comparables ya que la parcela de test coincide para los dos casos. Ambos modelos, presentan estimaciones que, en su mayoría, están fuera de los límites marcados por las líneas de BBCH máximas y mínimas medidas en campo. En general, se ve una tendencia creciente a lo largo de la variable estimada entorno a los valores reales, aunque con 2 o 3 zonas con picos de error mayor en las que el sistema predice un estado fenológico mayor al real. De hecho, el primer pico en torno a los días 30-40 después de la siembra está presente tanto en la estimación de BBCH como de la altura en las figuras 1.6 y 1.7b. Teniendo en cuenta que los valores de entrada son iguales para todos los casos, se puede atribuir este fenómeno a anomalías presentes en los datos de satélite, los cuales, en una etapa concreta

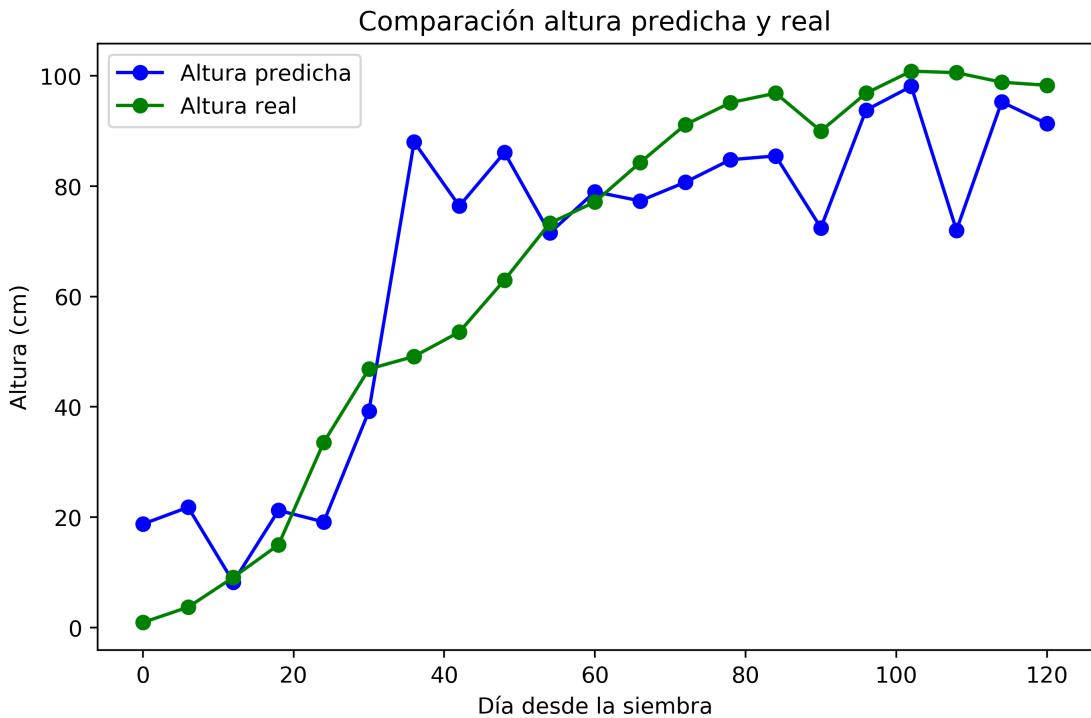
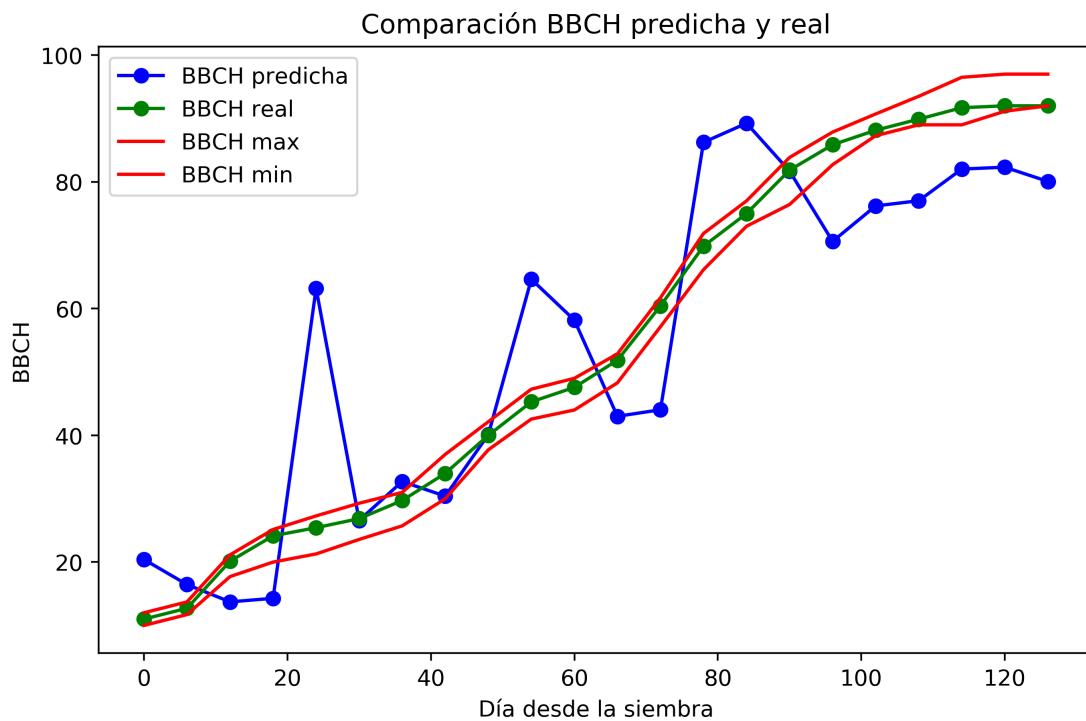


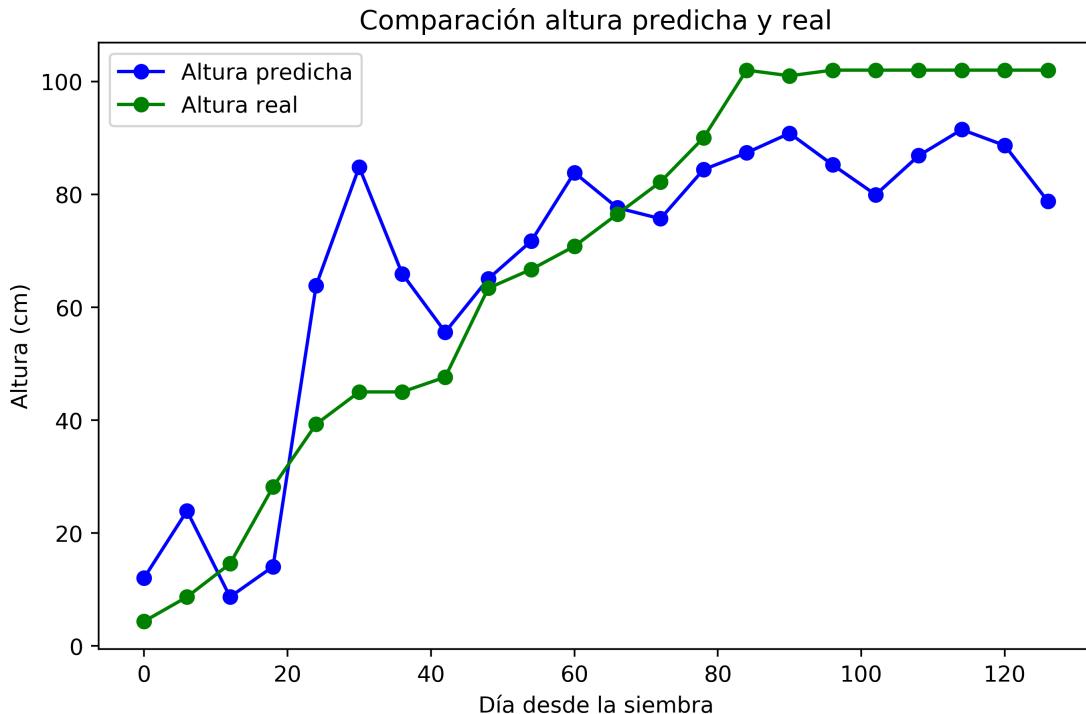
Figura 1.6: Comparación de la salida predicha y la verdad de tierra del modelo para estimación de la altura.

del desarrollo del cultivo, pueden presentar valores muy similares a los obtenidos para etapas finales, y por ello cometer un error de predicción bastante grande. Este error sería corregible bien detectando la anomalía y eliminando o modificando esta parte de los datos de entrada, tanto en entrenamiento como en test o futuros usos, o bien comprobando que el modelo está generando predicciones correctas con menor densidad de probabilidad, estando representadas en las salidas de PDF, y pudiendo ser consideradas sin un procesamiento extra de los datos, simplemente al contrastarlos con los resultados del modelo temporal, ya que es una diferencia de etapas muy grande y en este modelo no se daría un error de etapa de tal dimensión.

Visualizando las salidas para la altura, se puede también comparar, aunque no directamente, el funcionamiento de ambos modelos. En este caso las parcelas de test no son las mismas debido a la optimización para cada modelo, por lo que la comparación simplemente visual entre las dos representaciones es menos intuitiva. Aún así, se aprecian coincidencias como el error en la etapa cercana a 40 días, como se ha mencionado antes, o la estimación de mayor altura para los días de 0 a 10 después de la siembra. En general, ambas representaciones tienen también una tendencia creciente bastante similar a la real, con aparentemente mejor estimación para las etapas finales en el modelo de salida única de altura, ya que en la figura 1.6 se observa una estimación menor constante para todas las etapas desde el día 80 tras la siembra hasta la cosecha (día 120 aproximadamente).



(a) Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH.



(b) Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de la altura.

Figura 1.7: Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH y la altura.

1.1.3 Evaluación de resultados

Para la evaluación de los resultados obtenidos para el método por parcelas se van a utilizar, como se ha mencionado anteriormente, las salidas de valor único por su facilidad para ser representadas y comparadas con los valores únicos medidos.

La primera evaluación realizada se basa en la relación directa entre las salidas estimadas y las medidas reales, lo cuál muestra la fiabilidad del modelo de una forma más clara y así puede verse en las figuras 1.8, para el modelo de única salida de BBCH, 1.9, para el modelo de única salida de altura y la figura 1.10 para el modelo de doble salida. En estas representaciones se añade, además, una recta de pendiente (m) unidad, la cual sirve de referencia ya que esa relación representaría un modelo ideal. Se han utilizado para todas las representaciones los valores contrastados de la parcela de test correspondiente a cada caso, teniendo en cuenta ambos años de datos, 2017 y 2018.

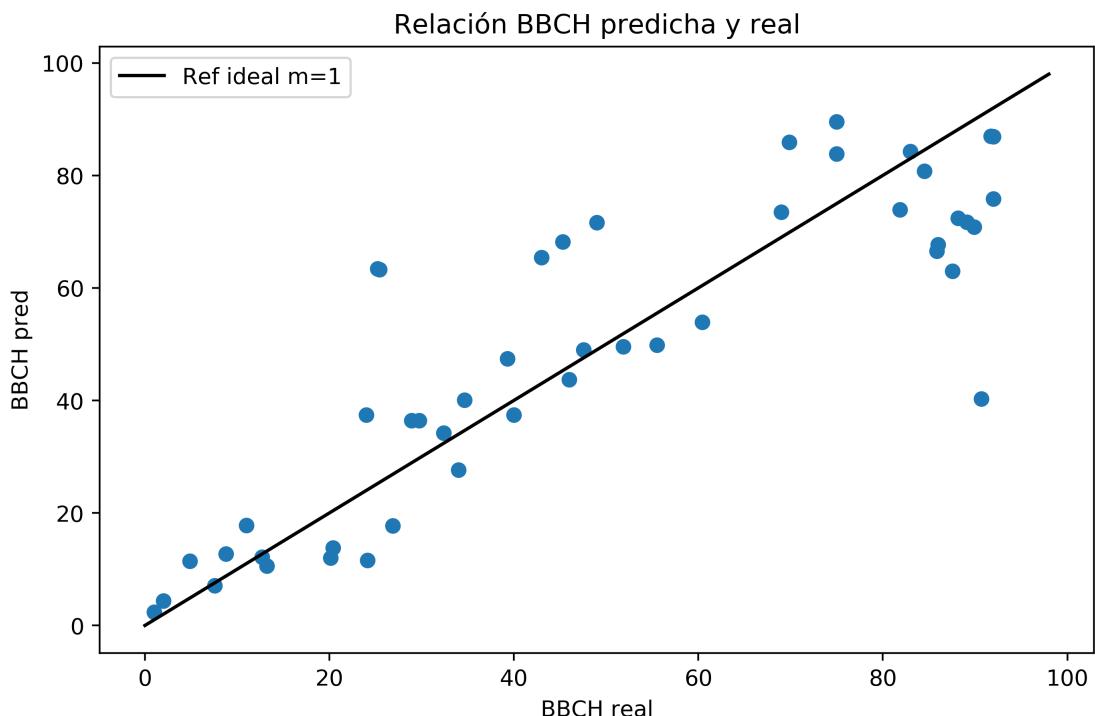


Figura 1.8: Relación de la salida predicha y la verdad de tierra del modelo para estimación de BBCH.

En la figura 1.8 se aprecia una distribución uniforme, lo cuál indica que se disponen de datos suficientes para cubrir las distintas etapas del desarrollo del cultivo. Además, las nubes de puntos se mantienen, en general, cercanos a la recta de referencia, por lo que, excepto para puntos concretos como el pico de las etapas tempranas y las últimas etapas del cultivo, se obtienen resultados bastante buenos. Comparando esta información con la obtenida para el modelo de dos salidas, representado en la figura 1.10a, se puede asegurar a simple vista qué modelo da mejores resultados. En esta última figura las nubes de puntos se presentan

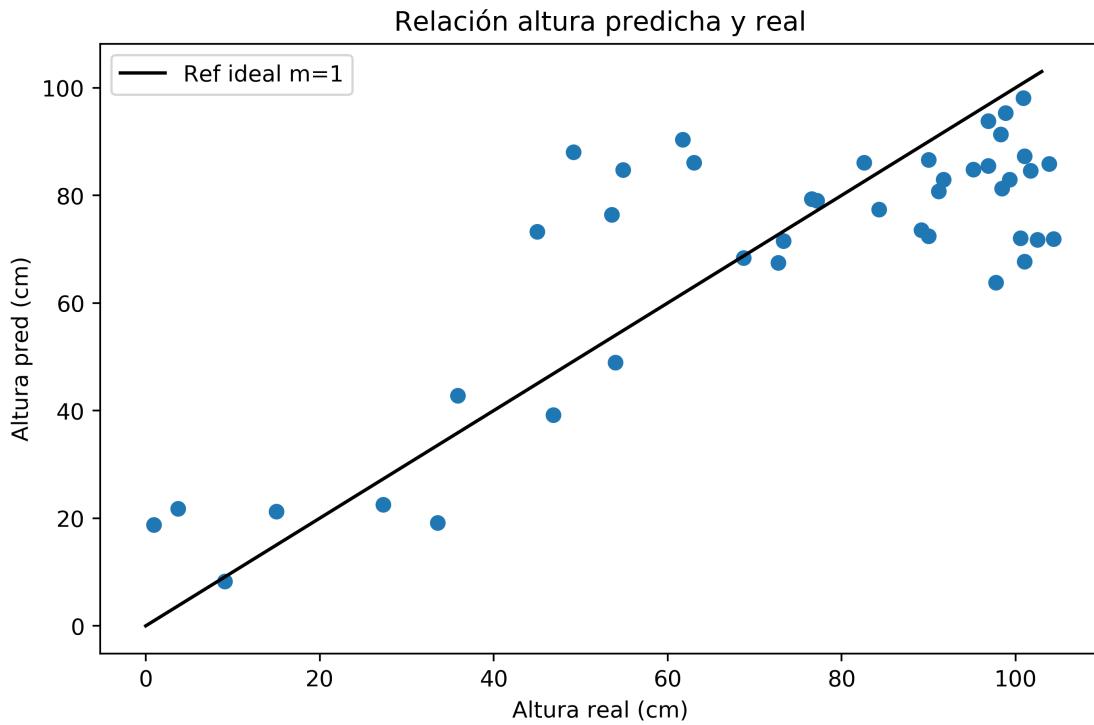
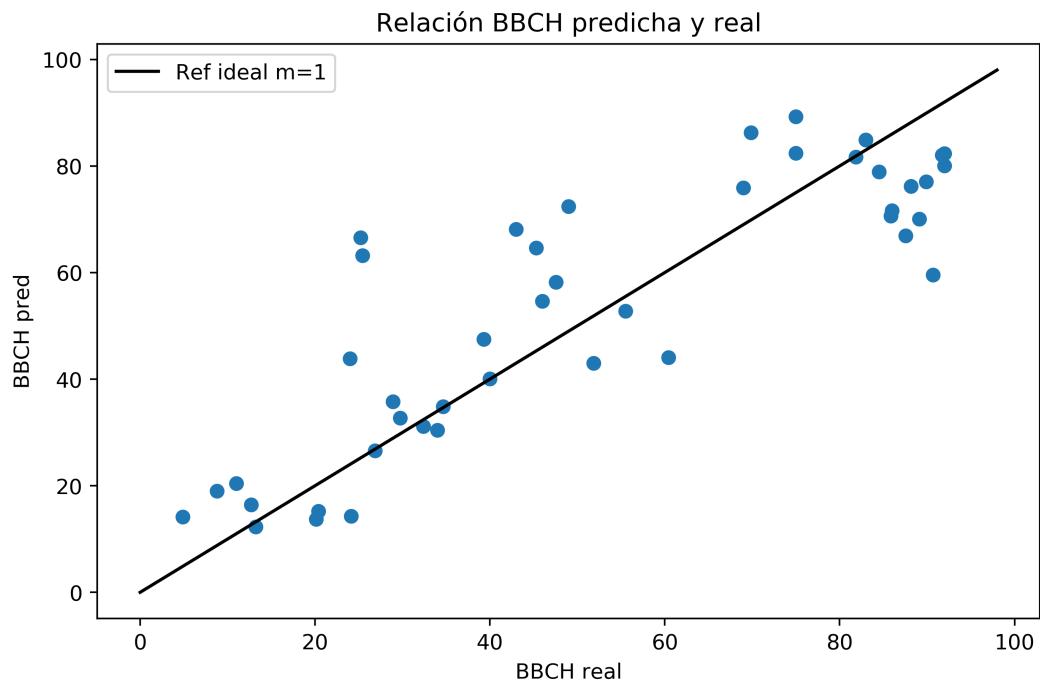


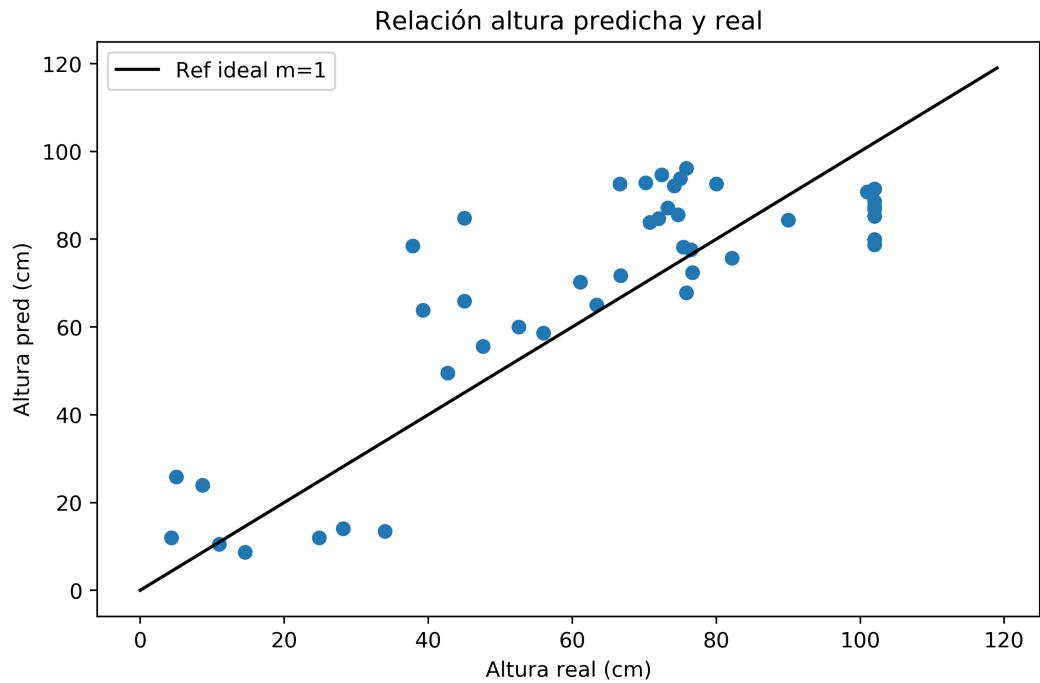
Figura 1.9: Relación de la salida predicha y la verdad de tierra del modelo para estimación de la altura.

más dispersas que en el anterior, aunque siguiendo la tendencia de la recta de referencia. Vemos, contrariamente, agrupaciones de puntos y huecos, por lo que algunas etapas no están siendo estimadas correctamente. Esto se puede deber al hecho de que el procesamiento de este modelo incluye una etapa de limpieza de datos corruptos que se encuentran en los datos de altura, por lo que en algunas etapas, tanto para la BBCH como para la altura, faltan datos que aporten la información completa al sistema.

Comparando a continuación las figuras de los modelos con salida de altura del cultivo, 1.10 para salida única y 1.10b para salida doble, se observa en ambos una mayor inestabilidad en general debida a la concentración de puntos en algunas etapas. La altura de los cultivos no tiene un crecimiento tan lineal progresivo como la fenología de un cultivo, por lo que, a parte de por la corrección de datos, se encuentran concentraciones de datos debido a la estabilidad de la altura durante distintas etapas. Es por ello que, sobre todo en las etapas finales, hay una mayor concentración de puntos, cuando el arroz se mantiene en altura. Aún así, en la figura 1.10 se aprecia que todas las estimaciones de altura tienden a la baja para casi todos los datos de altura mayor a 80cm. Una posible explicación sería una tendencia generalizada más baja en las parcelas utilizadas para entrenar el modelo o la poca variación en los parámetros de entrada para alturas desde 50cm hasta 110cm, intervalo en el que la altura estimada se mantiene en torno a la misma franja de valores.



(a) Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH.



(b) Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de la altura

Figura 1.10: Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH y la altura.

A parte de la evaluación con descriptores estadísticos que se realizan a la hora de comparar los dos métodos, se realiza también una evaluación de los datos de entrada utilizados. La tabla 1.2 representa el peso que tiene cada una de las variables de entrada en el modelo generado para cada uno de los 3 casos tratados.

	BBCH	Altura	BBCH&Altura
VV	0.12	0.12	0.12
VH	0.52	0.52	0.53
Ratio VH/VV	0.11	0.11	0.10
Dev est. VV	0.09	0.09	0.10
Dev est. VH	0.07	0.09	0.09
Dev est. ratio	0.07	0.07	0.06

Tabla 1.2: Influencia de los parámetros de entrada en el modelo de estimación.

En la tabla 1.2 se presentan proporciones altamente similares para los 3 casos contemplados, siendo incluso idénticos para los modelos de salidas independientes. El parámetro de entrada que destaca principalmente sobre el resto es el coeficiente de backscattering para la polarización VH, con más del 50% de la predicción elaborada en base a él. Los siguientes parámetros más influyentes son el coeficiente de backscattering para la polarización VV y el ratio de ambos, y, por último, las desviaciones estándar, que suponen menos del 10% de la predicción final. Para comprender mejor porqué existen estas diferencias, se representan en la figura 1.11 la relación entre cada una de las variables con BBCH.

Lo primero que se observa en la figura 1.11 es que todas las subfiguras, en general, presentan una tendencia creciente conforme BBCH va creciendo, que a los primeros estados corresponden valores más bajos y que en todas aparece el pico de detección entorno a un valor de BBCH de 20-30, que se asemeja a los valores obtenidos por los mismos parámetros para BBCH cercanas a la cosecha. Esto confirma que las predicciones erróneas cercanas a estas etapas vienen dadas por esta anomalía a la que se le podría buscar corrección.

Comparando los valores que se obtienen para los distintos parámetros a partir de una BBCH de 40, se aprecia en la subfigura 1.11b que su evolución es la más estable y además creciente. Esto quiere decir que, al contrario que en las subfiguras 1.11a y 1.11c, no hay oscilaciones en los valores de entrada y, por tanto, la predicción a partir de estos valores es más sencilla que si para un mismo valor de entrada pueden corresponder distintas BBCH o alturas. Además, estas etapas finales en la subfigura 1.11b se presentan estables pero no constantes, que sería el peor caso que se puede hallar, ya que no aportaría ninguna distinción para intervalos de BBCH muy amplios.

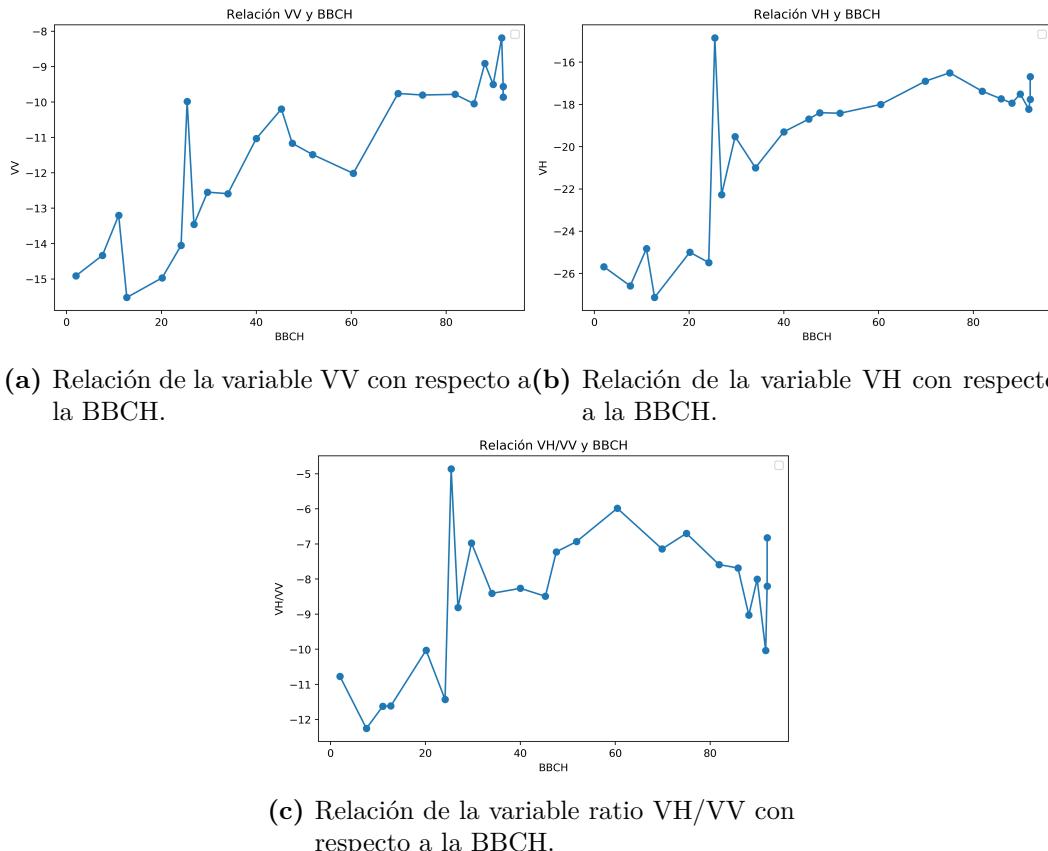


Figura 1.11: Relación de las variables de datos de entrada al modelo con respecto a la BBCH

1.2 Método por píxeles

1.2.1 Optimización

La optimización de los datos de entrada para este método se realiza de igual manera que el anterior, coincidiendo los mejores resultados para todos los casos en el uso de un conjunto de 6 parcelas para entrenamiento y 1 para la evaluación y los datos de entrada, se reducen a los tres principales (VV, VH y ratio), ya que la desviación estándar no aporta una gran mejora para este método. En cuanto a la optimización del regresor, se utiliza también la relación entre el número de árboles para un modelo y el coeficiente de determinación, presentado en la imagen de ejemplo 1.12 con el caso de salida BBCH. En ella, se puede ver que una vez alcanzado cierto nivel de coeficiente, la mejora de este en relación al aumento del número de árboles no es significativa con respecto al costo computacional y a la complejidad del sistema que se crea.

Finalmente, los parámetros utilizados en este método para cada caso se presentan en la tabla 1.3, donde se pueden ver: la parcela utilizada para el periodo de test del modelo; siendo el resto de parcelas utilizadas en el entrenamiento, y el número de árboles óptimo para RFR, de acuerdo con la evolución del coeficiente de determinación para cada caso.

Como se puede observar, los 3 casos coinciden en el set de parcelas de entrenamiento y

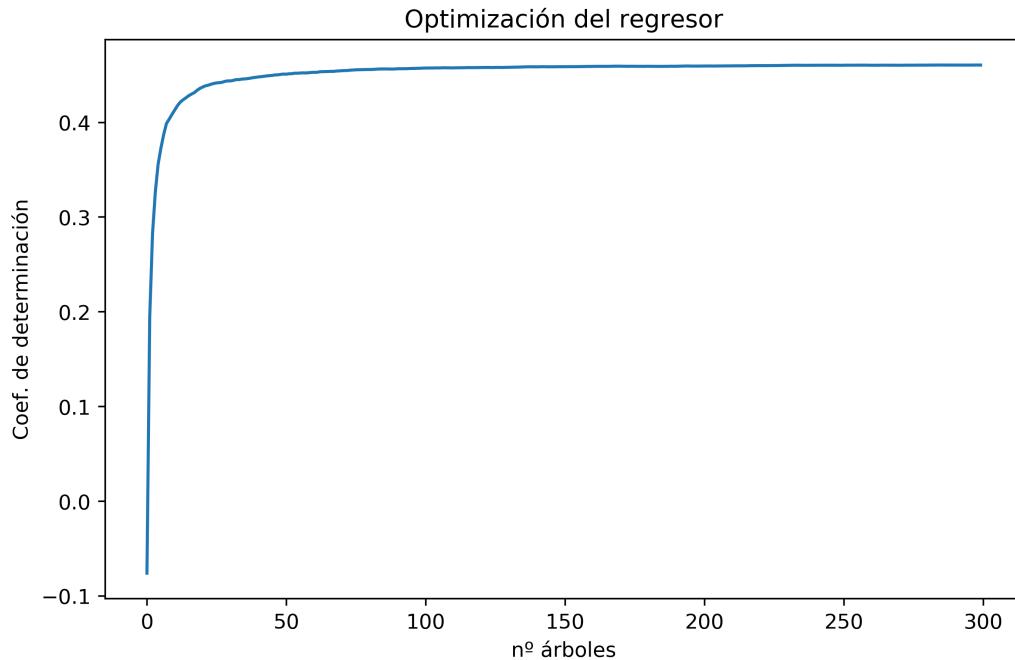


Figura 1.12: Optimización del número de árboles para RFR en el modelo de salida BBCH

	BBCH	Altura	BBCH&Altura
Parcela de test	‘Mínima’	‘Mínima’	‘Mínima’
Número de árboles	46	48	47

Tabla 1.3: Parámetros de optimización de entrada y modelo

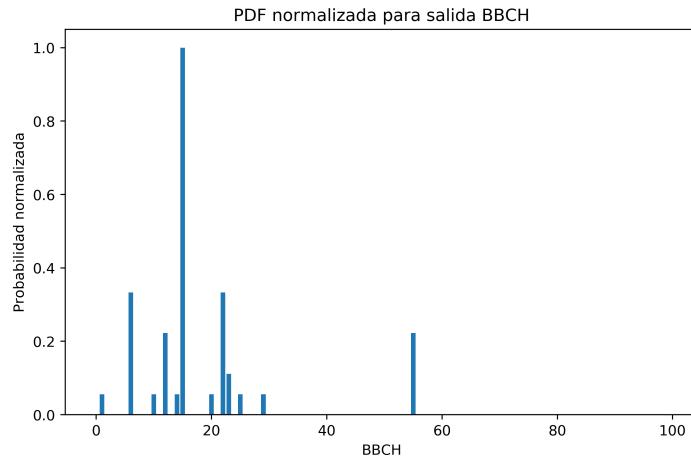
test con el que se obtienen mejores resultados. Probablemente esto se debe a anomalías en el desarrollo de algunas parcelas, las cuales, si se toman como set de test, el modelo no las habría podido tener en cuenta en el aprendizaje y el error en la predicción sería mayor. Los 3 casos de este método constan de un número óptimo de árboles muy similares, con una diferencia de 1 y 2 con respecto al caso menor. Cabe destacar que el aumento de complejidad, aunque escaso, en el modelo de salida de la altura, como ocurre para el método anterior. Además, tanto para la metodología por píxeles como por parcelas, se obtiene un número óptimo de árboles para el caso de 2 salidas intermedio a los óptimos para las mismas salidas en modelos independientes. Esto se debe a una compensación en la estimación de ambas variables, para que una variable sea óptima sin que su mejora sea a costa de la otra se llega a un valor intermedio en el que ninguna de las salidas está totalmente optimizada pero tienen el mejor resultado del sistema compartido completo.

1.2.2 Salidas del modelo

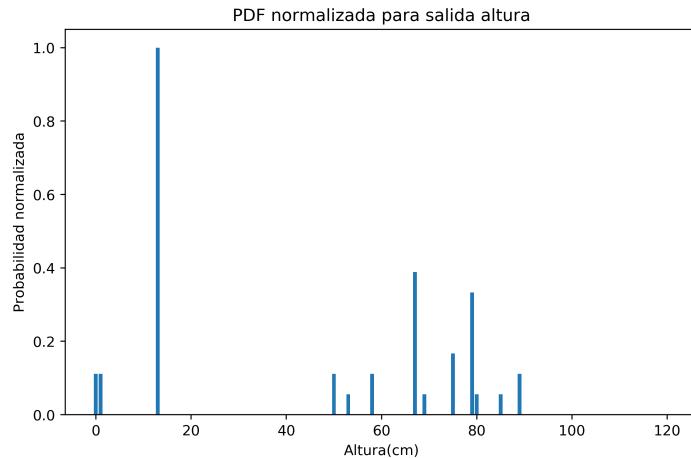
1.2.2.1 Salidas de función de densidad de probabilidad

Como se ha mencionado en la metodología anterior, las salidas útiles de este trabajo son las PDF. En las figuras 1.13a (modelo de salida BBCH), 1.13b (modelo de salida altura)

y 1.14 (modelo de ambas salidas) se puede apreciar cómo son algunas de estas salidas en esta metodología, siendo ejemplos extraídos a partir de los mismos datos para los 3 casos estudiados.



(a) Ejemplo de salida PDF normalizada del modelo para estimación de BBCH.



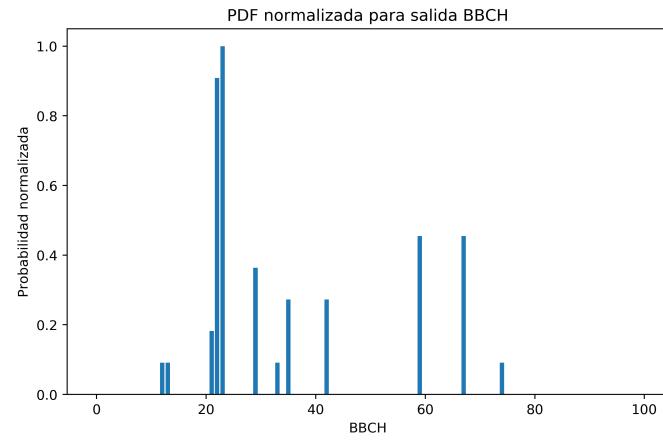
(b) Ejemplo de salida PDF normalizada del modelo para estimación de la altura.

Figura 1.13: Ejemplo de salida PDF normalizada los modelos para estimación de las variables independientes.

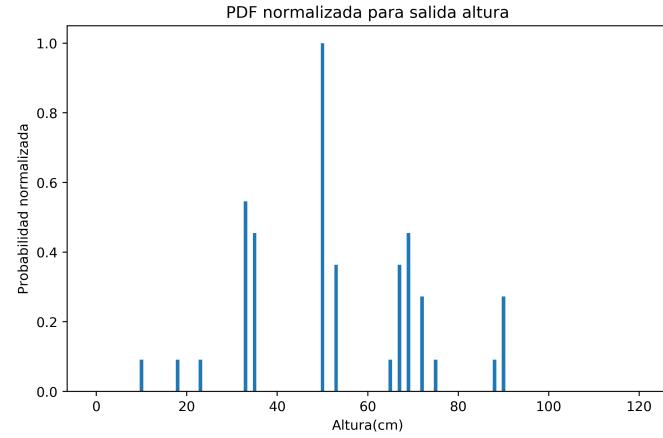
Estas representaciones tienen las mismas características mencionadas anteriormente: están normalizadas con respecto a la salida con mayor probabilidad, destaca una salida predominante pero no se excluyen el resto a la hora de integrarse con el modelo de predicción temporal.

Las figuras 1.13 y 1.14 han sido generadas con los mismos datos de entrada, es decir, los mismos datos de satélite en la misma parcela y fecha, por lo que las salidas para los

modelos generados de estimación de BBCH (Figura 1.13a) y altura (Figura 1.13b) como modelos independientes se pueden comparar con las salidas del modelo de estimación de ambas (Figura 1.14). La principal diferencia que se vuelve a encontrar entre los mismos tipos de datos de salida para cada modelo es que, las salidas individuales presentan una estimación principal con una diferencia de probabilidad mayor con respecto al resto que las estimaciones del modelo de doble salida. El rango para cada salida es bastante más amplio y disperso para el modelo de doble salida, esto junto a la disminución de diferencias con las demás estimaciones, indica que ese sistema es menos preciso y estable. En este modelo de dos salidas no se ve una clara optimización para una de ellas con respecto a la otra, ya que además ambas tiene valores muy cercanos al sistema individual.



(a) Ejemplo del modelo de doble salida: PDF estimación de BBCH

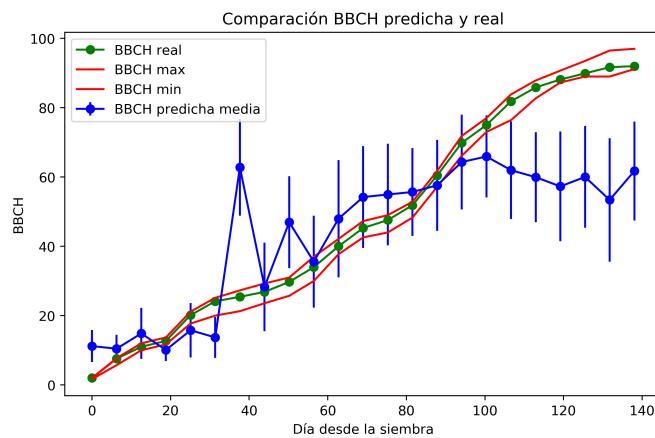


(b) Ejemplo del modelo de doble salida: PDF estimación de altura

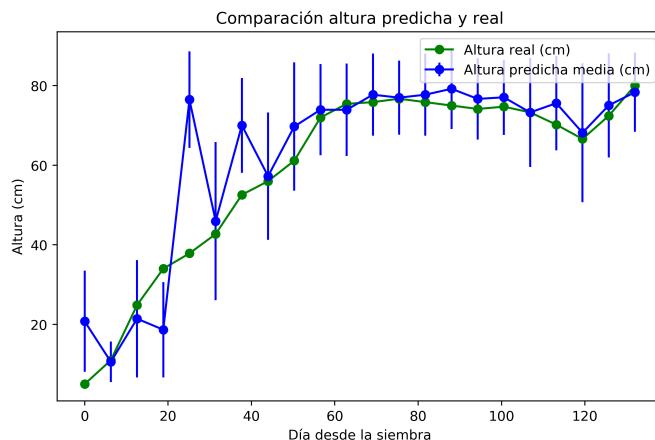
Figura 1.14: Ejemplo de salidas PDF normalizadas del modelo para estimación de BBCH y la altura.

1.2.2.2 Salidas de valor único

Para continuar con la presentación de los resultados obtenidos, se ilustran en las figuras ??, 1.15a y 1.16 las comparaciones de las soluciones únicas obtenidas (salida con máxima probabilidad) y los datos de tierra tomados correspondientes para la parcela de test en 2018: BBCH general, máxima y mínima en la parcela, casos individual (1.5) y de doble salida (1.16a), o la altura general del cultivo, también para ambos casos (figuras 1.15a y 1.16b, respectivamente). Se ha realizado una adaptación en los datos para poder representarlos de esta manera, ya que aquí la estimación de valor único se realiza a nivel de pixel y no de parcela, como los datos para contrastar. Es por ello que, para una representación y comprensión más sencillas, se representa la media y la desviación estándar para los píxeles a nivel de parcela y día.



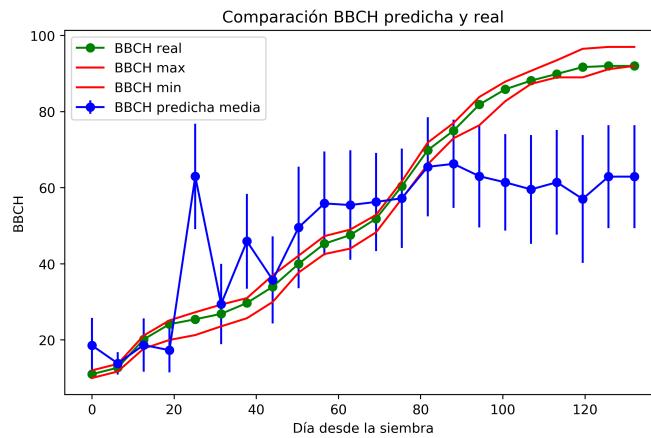
(a) Comparación de la salida predicha y la verdad de tierra del modelo para estimación de la altura.



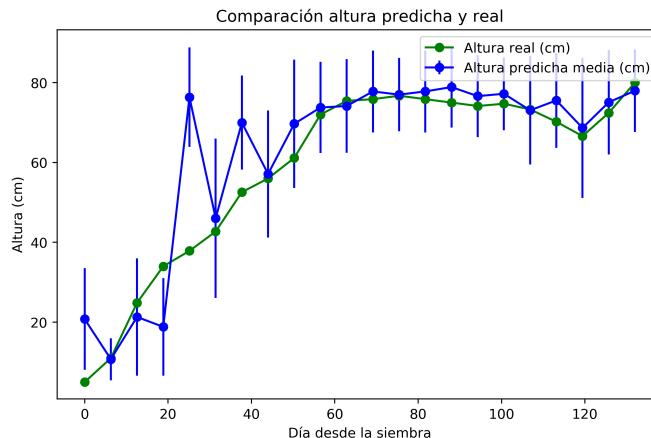
(b) Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de la altura

En este método ambos modelos son comparables para las dos salidas posibles ya que todos ellos comparten la misma parcela de test, por lo que se compara la eficiencia para los mismos

datos. Comenzando por los modelos de predicción de BBCH representados en las figuras ?? y 1.16a, se pueden observar estimaciones y errores de nuevo muy parecidos. Ambos modelos presentan una tendencia creciente, que estima a la baja la fenología para las últimas etapas del desarrollo y los mismos picos de predicción errónea sobre los días desde la siembra de 30-40 ya comentados anteriormente. Exceptuando las últimas etapas, se aprecian medias bastante cercanas a los valores reales, aunque con unas desviaciones estándar muy grandes a partir de los 40 días desde la siembra. Estas desviaciones se deben tanto a diferencias reales en la parcela, áreas con diferente nivel de desarrollo, como a un mal ajuste del modelo ya que en el entrenamiento existían estas diferencias en la misma parcela y se ha considerado el mismo valor de BBCH por la disponibilidad de datos de verdad de tierra. En esta metodología no aparecen diferencias tan claras entre los dos modelos implementados, ya que eran bastante similares.



(a) Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH.



(b) Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de la altura

Figura 1.16: Comparación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH y la altura.

Considerando las salidas para la altura, también se puede apreciar la gran similitud que presentan las representaciones comparadas: figuras 1.15a y 1.16b, difícilmente diferenciables a simple vista. Las coincidencias más obvias son las mencionadas con la estimación de la BBCH: la tendencia creciente, los picos de predicciones erróneas para las mismas etapas y grandes desviaciones estándar, en este caso, durante todo el proceso. Hay que puntualizar, que las medias para cada parcela tienen valores muy próximos a los reales, sobretodo en las etapas intermedias y finales, cuando la mayoría de los modelos presentan mayor inexactitud. Aunque los resultados globales evaluados para la metodología a nivel de pixel puedan ser numéricamente peores debido a la evaluación independiente de cada pixel, este es el método que mejores resultados ha obtenido para la estimación de la altura.

1.2.3 Evaluación de resultados

Para la evaluación de los resultados obtenidos para el método por parcelas se van a utilizar, como se ha mencionado anteriormente, las salidas de valor único por su facilidad para ser representadas y comparadas con los valores únicos medidos.

La primera evaluación realizada se basa en la relación directa entre las salidas estimadas y las medidas reales, lo cuál muestra la fiabilidad del modelo de una forma más clara y así puede verse en las figuras 1.17, para el modelo de única salida de BBCH, 1.18, para el modelo de única salida de altura y la figura 1.19 para el modelo de doble salida. En estas representaciones se añade, además, una recta de pendiente (m) unidad, la cual sirve de referencia ya que esa relación representaría un modelo ideal. Se han utilizado para todas las representaciones los valores contrastados de la parcela de test correspondiente a cada caso, teniendo en cuenta ambos años de datos, 2017 y 2018.

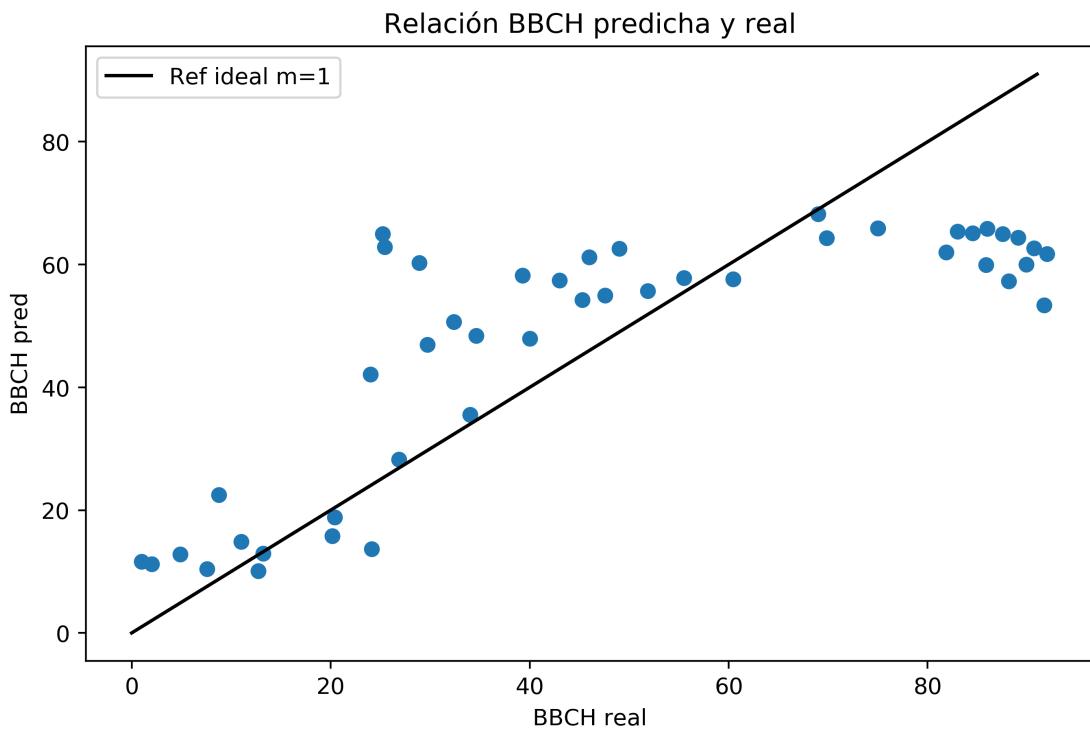


Figura 1.17: Relación de la salida predicha y la verdad de tierra del modelo para estimación de BBCH.

En la figura 1.17 se aprecia una distribución uniforme, lo cuál indica que se disponen de datos suficientes para cubrir las distintas etapas del desarrollo del cultivo. Además, las nubes de puntos se mantienen, en general, cercanos a la recta de referencia, por lo que, excepto para puntos concretos como el pico de las etapas tempranas y las últimas etapas del cultivo, se obtienen resultados bastante buenos. Comparando esta información con la obtenida para el modelo de dos salidas, representado en la figura 1.19a, se puede asegurar a simple vista

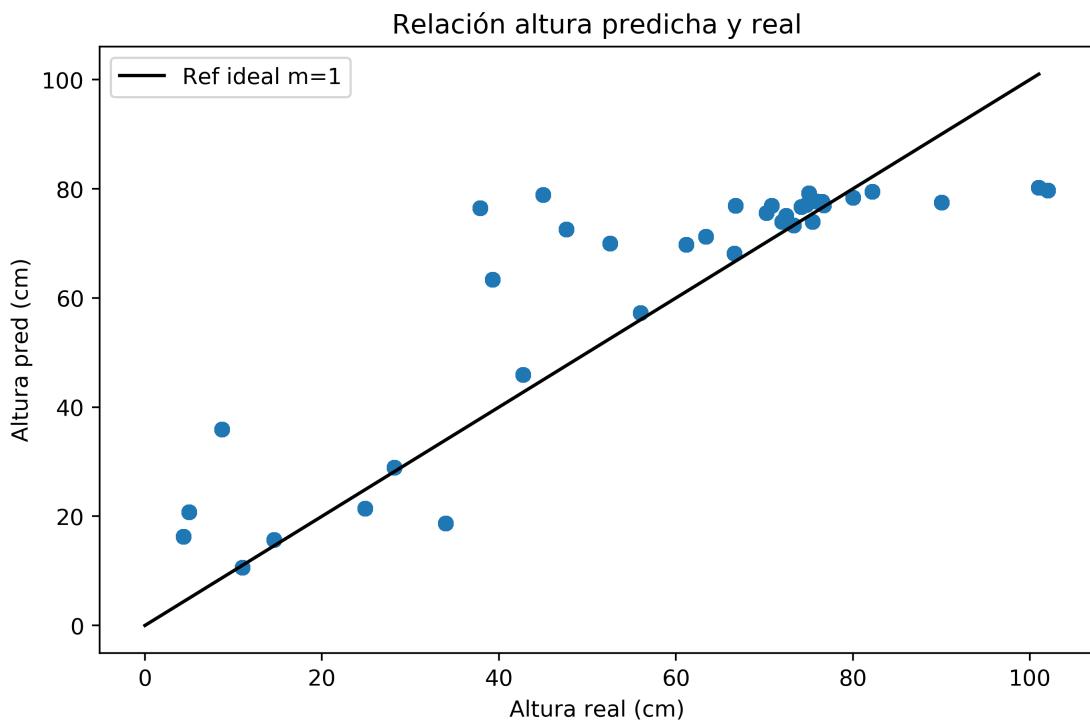
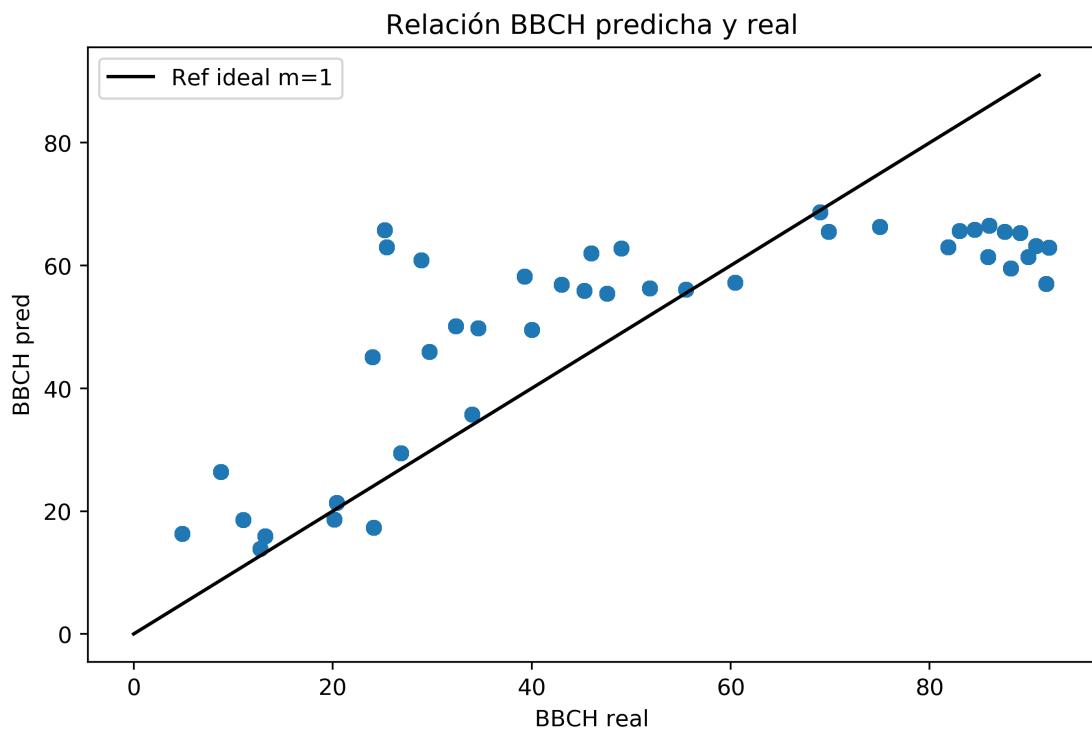


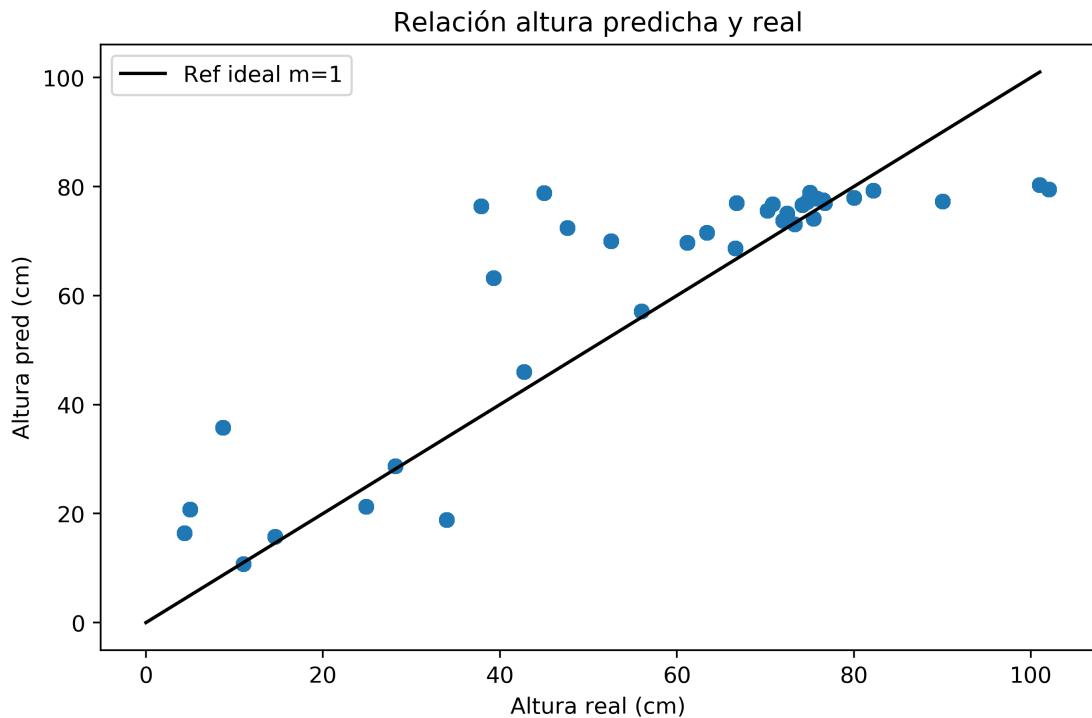
Figura 1.18: Relación de la salida predicha y la verdad de tierra del modelo para estimación de la altura.

qué modelo da mejores resultados. En esta última figura las nubes de puntos se presentan más dispersas que en el anterior, aunque siguiendo la tendencia de la recta de referencia. Vemos, contrariamente, agrupaciones de puntos y huecos, por lo que algunas etapas no están siendo estimadas correctamente. Esto se puede deber al hecho de que el procesamiento de este modelo incluye una etapa de limpieza de datos corruptos que se encuentran en los datos de altura, por lo que en algunas etapas, tanto para la BBCH como para la altura, faltan datos que aporten la información completa al sistema.

Comparando a continuación las figuras de los modelos con salida de altura del cultivo, 1.19 para salida única y 1.19b para salida doble, se observa en ambos una mayor inestabilidad en general debida a la concentración de puntos en algunas etapas. La altura de los cultivos no tiene un crecimiento tan lineal progresivo como la fenología de un cultivo, por lo que, a parte de por la corrección de datos, se encuentran concentraciones de datos debido a la estabilidad de la altura durante distintas etapas. Es por ello que, sobre todo en las etapas finales, hay una mayor concentración de puntos, cuando el arroz se mantiene en altura. Aún así, en la figura 1.19 se aprecia que todas las estimaciones de altura tienden a la baja para casi todos los datos de altura mayor a 80cm. Una posible explicación sería una tendencia generalizada más baja en las parcelas utilizadas para entrenar el modelo o la poca variación en los parámetros de entrada para alturas desde 50cm hasta 110cm, intervalo en el que la altura estimada se mantiene en torno a la misma franja de valores.



(a) Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH.



(b) Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de la altura

Figura 1.19: Relación de la salida predicha y la verdad de tierra del modelo de doble salida para la estimación de BBCH y la altura.

A parte de la evaluación con descriptores estadísticos que se realizan a la hora de comparar los dos métodos, se realiza una evaluación de los datos de entrada utilizados. La tabla 1.4 representa el peso que tiene cada una de las variables de entrada en el modelo generado para cada uno de los 3 casos tratados.

	BBCH	Altura	BBCH&Altura
VV	0.19	0.16	0.19
VH	0.61	0.61	0.59
Ratio VH/VV	0.19	0.23	0.22

Tabla 1.4: Influencia de los parámetros de entrada en el modelo de estimación.

1.3 Comparativa de métodos

La evaluación general de resultados realizada es la comparación entre los dos métodos de procesamiento de datos de entrada mencionados: a nivel de parcela o de pixel. En las siguientes tablas se pueden ver la evaluación de los resultados, divididos en entradas a nivel de parcela 1.5 y a nivel de pixel 1.6, según los índices estadísticos de Mean Absolute Error (MAE), Root-Mean-Square Error (RMSE) y el coeficiente de determinación (R^2). Ambos conjuntos corresponden al mejor caso de cada método, esto es, habiendo seleccionado sus parámetros de entrada óptimos en cuanto a número de variables y set de parcelas de entrenamiento, y habiendo optimizado también los parámetros del regresor.

Datos de entrada a nivel de parcela			
	BBCH	Altura	BBCH&Altura
R^2	0.74	0.65	0.70
RMSE	15.39	17.87	15.69
MAE	11.13	14.21	12.63

Tabla 1.5: Índices estadísticos de las predicciones con datos de entrada a nivel de parcela.

Datos de entrada a nivel de pixel			
	BBCH	Altura	BBCH&Altura
R^2	0.45	0.52	0.45
RMSE	22.23	19.89	21.17
MAE	17.26	15.64	16.58

Tabla 1.6: Índices estadísticos de las predicciones con datos de entrada a nivel de pixel.

Como se puede observar, para las 3 posibles salidas para las que se han diseñado los modelos, el método que emplea datos de entrada a nivel de parcela consigue notablemente mejores resultados. Se obtienen valores de coeficiente de determinación más cercanos a 1, valor ideal, apreciándose en el modelo de predicción de BBCH una mejora para el caso de parcelas de un

64.4% con respecto al de pixel. Lo mismo ocurre en los índices de error, cuyo valor óptimo es 0: se obtienen valores inferiores para el método a nivel de parcela en todos los casos estudiados.

Los mejores resultados obtenidos para el procesamiento a nivel de parcelas pueden recaer en que todos los datos recogidos de verdad de tierra se presentan por parcelas, por lo que no existen unos datos para contrastar cada pixel extraído de la información de satélite con el terreno. Dentro de cada parcela el cultivo no tiene porqué desarrollarse de manera homogénea, pero la información recibida de este consta del valor de BBCH (con el requisito de ser el correcto para al menos el 50% del cultivo) y la altura generalizados por parcela, además de los máximos y mínimos de cada uno, por lo que las variaciones que se puede hallar en distintas zonas del cultivo no se tienen en cuenta. Por ello, el método de pixel implementado asigna un solo valor de salida para conjuntos de píxeles de la misma parcela, los cuales pueden estar representando unos niveles de BBCH o altura distintos, y esto conlleva un ajuste del modelo erróneo y, por tanto, peores índices estadísticos de evaluación. Aún así, este método no ha sido descartado, ya que el análisis a nivel de pixel es de gran interés para detectar esas variaciones de desarrollo de un cultivo dentro de una misma parcela.

Bibliografía

- [1] BERNARDIS, C.D., VICENTE-GUILBALA, F., MARTINEZ-MARIN, T., and LOPEZ-SANCHEZ, J.M. *Contribution to real-time estimation of crop phenological states in a dynamical framework based on ndvi time series: Data fusion with sar and temperature*, 2016.
- [2] DE INGENIERÍA, R.A. *Diccionario español de ingeniería*, 2014. URL <http://diccionario.raing.es/es>.
- [3] VICENTE-GUILBALA, F., MARTINEZ-MARIN, T., and LOPEZ-SANCHEZ, J.M. *Dynamical approach for real-time monitoring of agricultural crops*, 2014.
- [4] WANGA, H., MAGAGIA, R., GOÏTAA, K., TRUDELA, M., McNAIRNB, H., and POWERS, J. *Crop phenology retrieval via polarimetric sar decomposition and random forest algorithm*. Elsevier, 2019.
- [5] FLACH, P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 2012.
- [6] RUSSELL, S.J. and NORVIG, P. *Artificial Intelligence A Modern Approach*, 2010.
- [7] FÜRNKRANZ, J. and HÜLLERMEIER, E. *Preference Learning: An Introduction*, 2011.
- [8] DECOSTE, D. and SCHOLKOPF, B. *Training invariant support vector machines*. Kluwer Academic Publishers, 2002.
- [9] JOACHIMS, T. *Text categorization with support vector machines: Learning with many relevant features*. Machine Learning: ECML-98. Lecture Notes in Computer Science, 1998.
- [10] BARGHOUT, L. *Spatial-taxon information granules as used in iterative fuzzy-decision-making for image segmentation*. Granular Computing and Decision-Making, 2015.
- [11] MAITY, A. *Supervised classification of radarsat-2 polarimetric data for different land features*. CoRR, 2016. URL <https://dblp.org/rec/journals/corr/Mait16.bib>.
- [12] EUROPEAN SPACE AGENCY (ESA). *Sentinel-1: Radar mission*, 2014. URL <https://www.youtube.com/watch?v=FJWzLxdSMyA>.
- [13] ESA. *Copernicus overview*, 2016. URL https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Overview4.
- [14] ESA. *Sentinel-1 constellation*, 2014. URL https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Satellite_constellation.

- [15] EARTH OBSERVING SYSTEM (EOS). *Sentinel 1:*, 2014. URL <https://eos.com/sentinel-1/>.
- [16] CASTILLO, F.E. and SENTÍS, F.C. *Agrometeorología*, 2001.
- [17] MEIER, U. *Growth stages of mono-and dicotyledonous plants*. Federal Biological Research Centre for Agriculture and Forestry, 2001. URL <https://web.archive.org/web/20180427154542/https://ojs.openagrar.de/index.php/BBCH/article/download/515/464>.
- [18] LANCASHIRE, P.D., BLEIHOLDER, H., BOOM, T.V.D., LANGELUDDEKE, P., STAUSS, R., WEBER, E., and WITZENBERGER, A. *A uniform decimal code for growth stages of crops and weeds*. *Annals of Applied Biology*, 1991.
- [19] VERDIN, J., PEDREROS, D., and EILERTS, G. *Índice diferencial de vegetación normalizado (ndvi)*. FEWS - Red de Alerta Temprana Contra la Inseguridad Alimentaria, Centroamérica, USGS/EROS Data Center, 2003.
- [20] ESA. *Missions: Sentinel -1. data products*, 2020. URL <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/data-products>.
- [21] SINERGISE LABORATORY FOR GEOGRAPHICAL INFORMATION SYSTEMS, L. *Sentinel hub playground*. URL <https://apps.sentinel-hub.com/sentinel-playground/?source=S2&lat=40.4&lng=-3.730000000000018&zoom=12&preset=1-NATURAL-COLOR&layers=B01,B02,B03&maxcc=20&gain=1.0&gamma=1.0&time=2019-12-01%7C2020-06-29&atmFilter=&showDates=false>.

Lista de Acrónimos y Abreviaturas

AEMA	Agencia Europea de Medio Ambiente.
BBCH	Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie.
EKF	Extended Kalman Filter.
EOS	Earth Observing System.
ESA	European Space Agency.
GRD	Ground Range Detected.
MAE	Mean Absolute Error.
ML	Machine Learning.
MMC	Método de Mínimos Cuadrados.
NDVI	Índice de Vegetación de Diferencia Normalizada.
PDF	Probability Density Function.
RAI	Real Academia de Ingeniería.
RF	Random Forest.
RFR	Random Forest Regressor.
RMSE	Root-Mean-Square Error.
SAR	Synthetic Aperture Radar.
SST	Señales, Sistemas y Telecomunicación.
TFG	Trabajo Fin de Grado.
UA	Universidad de Alicante.