

# ESTIMACION DEL NIVEL DE ESTRES HIDRICO EN FRUTALES MEDIANTE TÉCNICAS MACHINE LEARNING PARA APLICACIÓN EN SISTEMAS DE RIEGO INTELIGENTES

Juan D. González-Teruel<sup>1</sup>, Victor Blanco<sup>3</sup>, Pedro José Blaya-Ros<sup>2</sup>, Rafael Domingo<sup>2</sup>, Fulgencio Soto-Valles<sup>1</sup>, Roque Torres-Sánchez<sup>1</sup>

<sup>1</sup>Universidad Politécnica de Cartagena, Dpto. Automática, Ingeniería Eléctrica y Tecnología Electrónica, [juando.gonzalez@upct.es](mailto:juando.gonzalez@upct.es); [pencho.soto@upct.es](mailto:pencho.soto@upct.es); [roque.torres@upct.es](mailto:roque.torres@upct.es)

<sup>2</sup>Universidad Politécnica de Cartagena, Dpto. Ingeniería Agronómica, [pedro.blaya@upct.es](mailto:pedro.blaya@upct.es); [rafael.domingo@upct.es](mailto:rafael.domingo@upct.es)

<sup>3</sup>Department of Horticulture, Tree Fruit Research and Extension Center, Washington State University, Wenatchee, WA 98801, USA, [victor.blanco@wsu.edu](mailto:victor.blanco@wsu.edu)

## Resumen

*El agua es un bien escaso, especialmente en las regiones áridas y semiáridas. Este es el caso de la Cuenca Mediterránea, donde sus condiciones demográficas y climáticas la hacen idónea para el cultivo de frutas y hortalizas, demandando un volumen mayor de recursos hídricos. Las estrategias de riego deficitario se han mostrado exitosas, pero resulta primordial el control del estrés hídrico de los cultivos. La medida directa del mismo se encuentra actualmente asociada al potencial hídrico de tallo a mediodía, cuya medida es costosa en tiempo y labores asociadas. A nivel agrario sería interesante definir unos niveles cualitativos del estrés hídrico del cultivo y poder estimarlos a partir de variables cuya medida sea automatizable, de manera que se puedan implementar sistemas de riego inteligente basados en las necesidades hídricas del cultivo. En este trabajo se presenta un estudio preliminar para la obtención de un modelo capaz de predecir cinco niveles de estrés del cultivo a partir de los datos temporales de potencial matricial y contenido volumétrico de agua en el suelo y de diferentes variables agro-climáticas. Se han evaluado múltiples algoritmos de Machine Learning, obteniéndose una precisión máxima en la estimación del 72,4 %.*

**Palabras clave:** Machine Learning, riego deficitario, estrés hídrico, riego inteligente.

## 1 INTRODUCCIÓN

La escasez de agua es un problema generalizado que se agudiza bajo condiciones de clima árido y semiárido. El informe de la FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura) “Climate Smart Agriculture

Sourcebook” [1] estima para el año 2050 un incremento de la población mundial del 30 % (incremento de dos mil millones de personas) que requerirá de un aumento de la producción agrícola del 60 % para asegurar una demanda creciente de alimentos y establecer una cierta seguridad alimentaria. Este incremento en la producción agrícola de alimentos estará afectado significativamente por los efectos adversos del cambio climático que pueden empeorar la situación, como son el aumento de la temperatura y reducción de la precipitación y de los recursos hídricos disponibles [2]. En gran parte de la Cuenca Mediterránea, región caracterizada por un clima semiárido, el sector agrícola es el principal demandante de los recursos hídricos disponibles y debe convivir con la escasez de agua [3], afrontando en muchas ocasiones importantes recortes en las asignaciones de agua disponible para riego. Concretamente, la Cuenca del Segura afronta anualmente un déficit hídrico medio de 460 hm<sup>3</sup> que afecta a 2.700.000 ha de regadío [4].

Estos desequilibrios hídricos justifican la búsqueda de nuevas soluciones que mantengan e incluso incrementen las altas eficiencias del uso del agua y las producciones obtenidas con la modernización de sus regadíos. Consecuentemente, se asume globalmente que las soluciones deben fomentar un uso más eficiente del agua y la energía, para lo que las estrategias de riego deficitario han resultado ser una herramienta muy útil [5], junto a los riegos de precisión basados en la monitorización del continuo suelo-planta-atmósfera con sensores [6].

En relación a la reducción de los consumos de agua y el uso más eficiente, los esfuerzos deben centrarse en maximizar la productividad del agua más que en incrementar la producción [7], ya que no es posible,

en la mayoría de casos, poder satisfacer las necesidades hídricas máximas de los cultivos. De hecho, muchas comunidades de regantes españolas disponen de una dotación de riego para toda la campaña muy por debajo de los requerimientos teóricos. En estas condiciones, la distribución del riego a lo largo del ciclo de cultivo debe realizarse de manera que resulte efectiva en aliviar el estrés en las fases fenológicas más sensibles [8], [9]. Este es el objetivo de las estrategias de Riego Deficitario Controlado (RDC), que consisten en aportar un volumen de agua inferior a las necesidades teóricas, reduciendo el riego sólo en períodos del ciclo de cultivo donde el efecto sobre el rendimiento y calidad de la cosecha sea mínimo (períodos no críticos). En este sentido, resulta vital conocer el nivel de estrés hídrico al que está sometido el cultivo y el que puede soportar en cada fase fenológica.

El método más aceptado para la determinación del estado hídrico de los cultivos es la medida del potencial hídrico de tallo a mediodía,  $\psi_{stem}$ , con cámara de presión [10]. Sin embargo, se trata de un método destructivo y costoso en tiempo y labores asociadas. De forma alternativa, y con objeto de facilitar y automatizar la obtención del mismo, se ha estudiado la relación que pudieran tener ciertas variables agro-climáticas, como la temperatura del aire, la radiación solar, el Déficit de Presión de Vapor (DPV) o la Evapotranspiración de Referencia (ET<sub>o</sub>) con el  $\psi_{stem}$ , obteniéndose una correlación limitada [11], [12]. Intrigliolo y Castel [12] también buscaron una relación entre  $\psi_{stem}$  y el potencial matricial del suelo,  $\psi_{soil}$ , medido con sensores Watermark, encontrando cierta correlación entre ambas variables, pero con una alta dispersión, especialmente para  $\psi_{soil} > -45$  kPa. El potencial matricial o mátrico del agua en el suelo representa la fuerza con la que el agua es atraída por la superficie de las partículas sólidas del suelo, así como la fuerza de atracción entre las propias moléculas de agua. Martí et al. [13] utilizaron redes neuronales artificiales (ANN) para estimar el valor de  $\psi_{stem}$  a partir de un conjunto limitado de datos de variables climáticas y contenido de agua en el suelo, obteniendo un coeficiente de determinación de 0,926, mientras que Valdés-Vela et al. [14], a partir de las mismas variables, establecieron un enfoque diferente aplicando reglas fuzzy, lo que permitió discretizar las variables de entrada del sistema en clases cualitativas, haciendo su interpretación más accesible.

En este trabajo se plantea un enfoque inverso al de Valdés-Vela et al. [14], utilizando el valor continuo de variables climáticas y de contenido y potencial matricial del suelo para estimar no el valor exacto de  $\psi_{stem}$ , sino un nivel de estrés cualitativo del cultivo en base a un informe emitido por un experto a partir de  $\psi_{stem}$  medido y de la evolución de las distintas

variables climáticas y de suelo. El objetivo es, por tanto, obtener un modelo capaz de sustituir al experto, que estime un nivel de estrés cualitativo del cultivo a partir de variables cuya medida se pueda automatizar. En este estudio se presenta una primera aproximación al enfoque planteado, donde se evalúa la capacidad de una colección de algoritmos de Machine Learning para estimar el nivel de estrés de un cultivo frutal a partir de los datos temporales de variables agro-climáticas y de contenido y potencial matricial del suelo.

## 2 MATERIALES Y MÉTODOS

### 2.1 INSTALACIÓN EXPERIMENTAL

#### 2.1.1 Descripción del cultivo y estrategias de riego

El experimento se llevó a cabo en una finca comercial de 0,5 ha de extensión localizada en Jumilla, Murcia, España (38° 8'N; 1°22'W, altitud 670 m) desde mayo de 2015 hasta agosto de 2018 en cerezos (*P. avium* L. "Prime Giant") adultos con 15 años, injertados en rizoma SL64 y con las variedades "Early Lory" y "Brooks" como polinizadores.

El método de riego empleado fue por goteo, con una línea de goteros por cada línea de árboles y tres emisores autocompensantes de 4 L h<sup>-1</sup> por árbol. El riego fue aplicado desde marzo hasta noviembre. Se aplicaron cinco tratamientos de riego diferentes, con dos repeticiones cada uno: (i) tratamiento de control (CTL), regado para satisfacer la evapotranspiración del cultivo (ET<sub>c</sub>) máxima a lo largo del periodo de crecimiento (110 % ET<sub>c</sub>); (ii) tratamiento de déficit sostenido (DS), regado al 85 % de la ET<sub>c</sub> durante pre-cosecha y post-cosecha, a excepción de los 15-20 días después de la primera cosecha (diferenciación floral), donde el riego correspondió al 100 % de la ET<sub>c</sub>; (iii) dos tratamientos de riego deficitario controlado: RDC-1, regado al 90 % durante pre-cosecha, 100 % en diferenciación floral y 65 % de la ET<sub>c</sub> durante post-cosecha; y RDC-2, al 100 % durante pre-cosecha y diferenciación floral y al 55 % de la ET<sub>c</sub> durante post-cosecha y (iv) tratamiento agricultor (AGR), regado de acuerdo a la práctica normal del agricultor local, que consiste en regar por encima de los requerimientos hídricos del cultivo durante pre-cosecha y aplicar un déficit hídrico según su experiencia durante la post-cosecha.

#### 2.1.2 Medida de variables agro-climáticas

Para el estudio se utilizaron variables climáticas como temperatura y humedad relativa del aire, radiación solar y velocidad del viento, cuyos valores medios horarios fueron obtenidos a partir de una estación meteorológica cercana a la instalación

experimental, perteneciente al Sistema de Información Agroclimática para el Regadío (SiAR; <http://crea.uclm.es/siar/datmeteo/>). A partir de la temperatura y de la humedad relativa del aire se calculó el DPV, de acuerdo a [15].

### 2.1.3 Estado hídrico del suelo

El suelo del sitio en estudio es moderadamente pedregoso y presenta una textura franco-arenosa, con una distribución del tamaño de partículas de 67,5 % arena, 17,5 % limo y 15 % arcilla. El agua de riego, procedente de pozo, presentó una conductividad eléctrica (EC) media de 0,8 dS m<sup>-1</sup> a 25 °C. El contenido volumétrico de agua en el suelo,  $\theta_v$ , se obtuvo para cada repetición a una profundidad de 20, 40 y 70 cm, utilizando sensores dieléctricos Enviroscan (Sentek Pty. Ltd., Adelaide, Australia). Se instaló un tubo de acceso por cada repetición, localizado a 0,23 m del emisor de riego y a 1,5 m del tronco del árbol. Asimismo, se obtuvo el valor del potencial matricial del suelo,  $\psi_{soil}$ , en cada repetición a 25 y 50 cm de profundidad utilizando sensores Decagon MPS6 (Decagon Devices Inc., Pullman, WA, USA), localizados igualmente a 0,23 m del emisor de riego. Tanto  $\theta_v$  como  $\psi_{soil}$  fueron registrados por un datalogger Campbell Scientific CR1000 (CampbellScientific, Logan, UT, USA), programado para medir cada 30 s y el proporcionar el valor medio cada 10 min.

### 2.1.4 Medida del potencial de tallo a mediodía

$\psi_{stem}$

El potencial hídrico de tallo a mediodía,  $\psi_{stem}$ , se midió aproximadamente cada 10 días con una cámara de presión Scholander (Model 3000, Soil Moisture Equipment, Santa Barbara, CA, USA), de acuerdo a la metodología propuesta en [16]. Se seleccionaron hojas maduras y sanas del cuadrante norte cerca del tronco para evitar la exposición solar. Las hojas se cubrieron con bolsas de polietileno y papel de aluminio con al menos dos horas de antelación a la medida de  $\psi_{stem}$ . La medida de  $\psi_{stem}$  se llevó a cabo en los períodos fenológicos relevantes, es decir, desde abril hasta noviembre, ambos inclusive, realizándose dos medidas por repetición.

## 2.2 MODELO PARA LA ESTIMACIÓN DEL NIVEL DE ESTRÉS

### 2.2.1 Dataset

Para estimar el nivel de estrés del cultivo en base a variables agro-climáticas y de suelo se plantea un problema de clasificación basado en entrenamiento supervisado. Se consideran 10 entradas en el modelo, como son temperatura y humedad relativa del aire, radiación solar, velocidad del viento, DPV,  $\psi_{soil}$  a 25

y 50 cm de profundidad y  $\theta_v$  a 20, 40 y 70 cm de profundidad. La salida del modelo es el nivel de estrés del cultivo, definido en cuatro posibles clases: 0, ausencia de estrés; 1, estrés hídrico incipiente; 2, estrés hídrico ligero; 3, estrés hídrico moderado y 4, estrés hídrico severo.

La frecuencia con la que se muestrearon las variables de suelo y clima y  $\psi_{stem}$  es dispar, debido a las propias limitaciones asociadas a la medida de este último y ya descritas con anterioridad. Dado que los niveles de estrés del cultivo fueron determinados por el experto a partir de cada medida de  $\psi_{stem}$ , su registro también presenta una periodicidad diferente a la de las variables climáticas y de suelo, que son las variables a partir de las cuales se pretende realizar la estimación del nivel de estrés. De esta forma, mientras que las variables de suelo y agro-climáticas fueron registradas cada 10 minutos, la estimación del nivel de estrés se obtuvo cada 15 días, aproximadamente. Así, entre cada registro del nivel de estrés se dispone de una sucesión de datos temporales de las variables de suelo y agro-climáticas. Como consecuencia, y ante la previsión de utilizar modelos de clasificación de Machine Learning para aprendizaje supervisado, ha sido necesario adaptar la dimensión entre los datos disponibles de las variables de entrada (suelo y agro-climáticas) y de salida del modelo (nivel de estrés). En este sentido, teniendo en cuenta que la dinámica del continuo suelo-planta-atmósfera presenta periodos transitorios dilatados, se considera que la evolución temporal de las variables físicas de entrada del sistema, y más concretamente la energía almacenada por las mismas a lo largo de un periodo de tiempo, puede ser un indicador relevante en la determinación del nivel de estrés del cultivo. Para representar dicha energía almacenada por las variables físicas, se ha calculado el área bajo la curva descrita por las mismas a lo largo de un periodo de tiempo  $T$ , utilizando un método de integración discreta basada en el cálculo de áreas trapezoidales, implementado en Matlab bajo la función *trapz*. Así, la sucesión de datos temporales de cada una de las variables de entrada queda reducida a un único valor representativo del periodo  $T$ .

Intuitivamente, el periodo de tiempo considerado en la integración de las variables de entrada es un factor a considerar, ya que a priori se desconoce la influencia que a corto y largo plazo puedan tener las mismas sobre la variable de salida. De esta forma, considerando  $T$  como el periodo de tiempo inmediatamente anterior al registro del nivel de estrés, se han definido de manera arbitraria tres valores para el mismo:  $T_1 = 6$  días,  $T_2 = 10$  días y  $T_3 = 30$  días. Dado que la fecha en la que se evaluó por primera vez el nivel de estrés del cultivo es solo 6 días posterior al inicio del registro de las variables

agro-climáticas y de suelo, al considerar  $T_2$  y  $T_3$  el tamaño del *dataset* se ve disminuido, ya que se pierden los ejemplos de las primeras semanas de ensayo. Teniendo en cuenta la pérdida de medidas por fallo de sensores y averías en el sistema de adquisición de datos, el número de ejemplos en el *dataset* es de 373, 366 y 351 para  $T_1$ ,  $T_2$  y  $T_3$ , respectivamente. Los valores de las variables de entrada, una vez calculada su integral definida en  $T$ , fueron normalizados a valores en el intervalo [0,1], correspondiendo 0 al valor original mínimo y 1 al máximo de cada variable en el *dataset*.

**2.2.2 Modelos Machine Learning para estimar el nivel de estrés hídrico**

Haciendo uso de la aplicación “Classification Learner” de Matlab se evaluaron todos los modelos disponibles en la misma para realizar la clasificación de los niveles de estrés, los cuales vienen enumerados en la Tabla 1. Los parámetros propios de cada algoritmo se mantuvieron por defecto.

Tabla 1: Modelos Machine Learning utilizados en el estudio.

Modelo	Variante
Árbol de decisión	Fino
	Medio
	Grueso
Análisis discriminante	Lineal
	Cuadrático
Support Vector Machine (SVM)	Lineal
	Cuadrático
	Cúbico
	Gaussiano fino
	Gaussiano medio
K-Nearest Neighbours (KNN)	Gaussiano grueso
	Fino
	Medio
	Grueso
	Coseno
	Cúbico
Clasificadores agrupados	Ponderado
	Boosted trees
	Bagged trees
	Subspace discriminant
	Subspace KNN
	RUSBoosted trees

Para evitar problemas de overfitting, se ha aplicado la técnica k-fold para realizar el cross-validation, que divide el *dataset* en *training set* (conjunto de entrenamiento) y *validation set* (conjunto de validación). Esta técnica divide el *dataset* en  $k$  *subsets* y realiza  $k$  iteraciones, de manera que en cada una de ellas asigna un *subset* diferente al *validation set* y realiza el entrenamiento con el resto de *subsets*.

Una vez finalizadas todas las iteraciones, la precisión del modelo se calcula como la media de las precisiones conseguidas en el *validation set* de todas las iteraciones. En este caso se han evaluado tres valores de  $k$  diferentes: 5, 15 y 50.

**2.2.3 Evaluación de los modelos**

Para evaluar la idoneidad de los modelos ensayados se ha recurrido a un indicador básico en problemas de clasificación, como es la precisión, definiéndose ésta como la relación entre el número de ejemplos que se han clasificado correctamente y el número total de los mismos:

$$\text{Precisión (\%)} = \frac{\text{Nº ejemplos clasificados correctamente}}{\text{Nº de ejemplos totales}} \cdot 100 \quad (1)$$

**3 ANÁLISIS Y DISCUSIÓN DE RESULTADOS**

**3.1 RELACIÓN TEMPORAL ENTRE LAS VARIABLES DE SUELO Y EL ESTRÉS HÍDRICO DEL CULTIVO**

En la Figura 1 se presenta, a modo de ejemplo, la evolución temporal del potencial matricial del suelo medido a las profundidades de 25 y 50 cm en una de las repeticiones del tratamiento de riego CTL, desde finales de abril hasta finales de julio (período fenológico relevante), junto con los eventos de medida del nivel de estrés hídrico del cultivo. El potencial matricial del suelo se mide en unidades de presión y se suele expresar en valores negativos que representan la oposición del suelo al movimiento del agua. Se observa que, a 25 cm,  $\psi_{soil}$  generalmente presenta valores superiores que a 50 cm, lo que indica una mayor facilidad para extraer agua del suelo a 25 cm de profundidad. Esta diferencia puede deberse a una mayor compactación del suelo a medida que aumenta la profundidad, dado que el aumento de la densidad aparente del suelo aumenta su capacidad de retención de agua. No obstante, se registran valores de  $\theta_v$  superiores a 20 cm que a 40 cm de profundidad, lo que podría indicar que la diferencia de  $\psi_{soil}$  entre profundidades se debiera a una diferencia en la cantidad de agua disponible en el suelo como consecuencia del tiempo de infiltración.

Atendiendo a la relación entre los valores registrados de  $\psi_{soil}$  y los niveles de estrés hídrico del cultivo determinados por el experto, se observa que las dos primeras muestras de nivel de estrés coinciden con dos valles de  $\psi_{soil}$ , siendo el segundo (inmediatamente posterior al 13 de mayo) un tanto más acusado y generando un incremento del nivel de estrés, que asciende del nivel 2 al 3. Los dos siguientes eventos (previo y posterior al 27 de mayo)

vienen precedidos de un ligero restablecimiento del agua disponible en el suelo, volviendo el nivel de estrés al valor 2. Previo al 10 de junio se observa una caída de  $\psi_{soil}$  hasta aproximadamente -35 kPa en ambas profundidades. Sin embargo, la siguiente medida del nivel de estrés dio lugar a una reducción de éste con respecto a las medidas previas. A priori, el valor de  $\psi_{soil}$  en el instante en que se determina el nivel de estrés inmediatamente posterior al 10 de junio, no difiere mucho de otros casos anteriores en los que se habían obtenido un nivel de estrés 2 y la evolución temporal previa al 10 de junio de  $\psi_{soil}$  no sugiere que se vaya a producir una reducción del nivel de estrés, en base a la tendencia atisbada en los ejemplos anteriores. Por su parte, las últimas tres muestras de nivel de estrés mostradas en la Figura 1 vuelven a situarse en el nivel 2 y vienen precedidas de tendencias decrecientes de  $\psi_{soil}$ .

### 3.2. RELACIÓN ENTRE LA ENERGÍA ALMACENADA POR LAS VARIABLES FÍSICAS Y EL ESTRÉS HÍDRICO DEL CULTIVO

En la Figura 2 se muestran los diagramas de cajas y bigotes para diferentes variables de entrada del modelo en relación con los niveles de estrés hídrico del cultivo. Los valores de las variables de entrada se han calculado de acuerdo con el método descrito en la sección 2.2.1, teniendo en cuenta el período de integración  $T_3$ . Atendiendo a la mediana, la mayor diferenciación entre niveles de estrés está provista por la temperatura del aire, DPV y  $\psi_{soil}$ , aunque en el caso de este último, apenas existe diferenciación entre los niveles de estrés 0 y 1, y para ambos niveles se presentan numerosos datos atípicos. En general, se observa para todas las variables la existencia de una gran varianza en todos los niveles de estrés, algo más reducida en los niveles 3 y 4, aunque el nivel 4 no es representativo, ya que sólo existen 2 ejemplos en los que se haya llegado a dicho nivel de estrés. La alta varianza hace que se produzcan solapamientos entre niveles de estrés, por lo que, a pesar de apreciarse diferencias notables en la mediana entre niveles de estrés para ciertas variables, existen fronteras demasiado difusas entre los mismos.

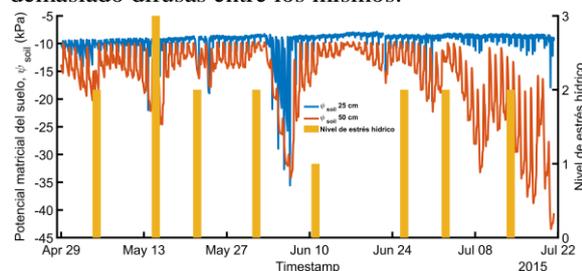


Figura 1: Evolución temporal de  $\psi_{soil}$  a 25 y 50 cm de profundidad y niveles de estrés obtenidos entre mayo y agosto de 2015.

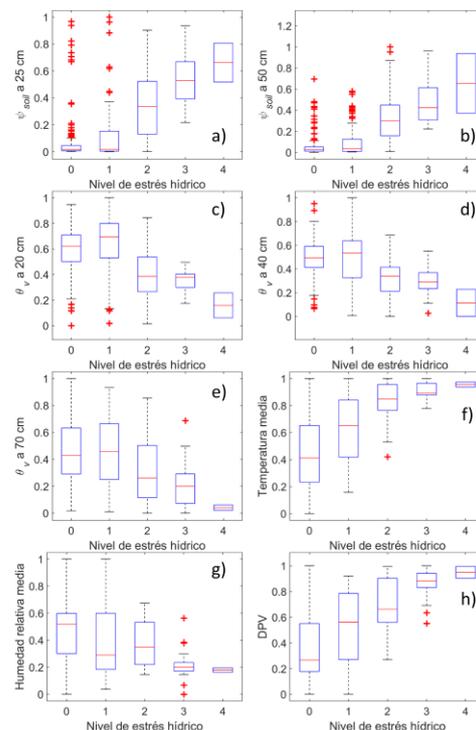


Figura 2: Diagrama de cajas y bigotes de los valores de energía normalizada almacenada por (a)  $\psi_{soil}$  a 25 cm, (b)  $\psi_{soil}$  a 50 cm, (c)  $\theta_v$  a 20 cm, (d)  $\theta_v$  a 40 cm, (e)  $\theta_v$  a 70 cm, (f) temperatura media del aire, (g) humedad relativa media del aire y (h) DPV para cada nivel de estrés hídrico del cultivo.

### 3.3. ESTIMACIÓN DEL NIVEL DE ESTRÉS HÍDRICO DEL CULTIVO

Tras aplicar los modelos descritos en la sección 2.2.2 para  $k = 5, 15$  y  $50$ , y para  $T_1, T_2$  y  $T_3$ , se muestra en la Tabla 2 el modelo con el que se ha conseguido la mayor precisión y el valor de la misma, de acuerdo a (1) para cada combinación de  $k$  y  $T$ . A excepción de la combinación  $k = 15, T_3$ , donde la mayor precisión se ha conseguido con un árbol de decisión medio, en el resto de los casos el modelo óptimo siempre es *Bagged trees*, también conocido como *Random Forest*. Los parámetros con los que se ha utilizado dicho modelo son: número máximo de bifurcaciones, 350; número máximo de predictores, 30; tasa de aprendizaje, 0,1 y dimensión del sub-espacio, 1.

No se observan diferencias significativas entre las precisiones logradas con las diferentes combinaciones de  $k$  y  $T$ , siendo la precisión máxima lograda de 72,4 % para  $k = 5$  y  $T_3$ . Para dicho caso, se presenta en la Figura 3 la tabla de verdad obtenida, donde aparecen la tasa de verdaderos positivos y falsos negativos. Se aprecia que el nivel de estrés 0 es

el que mayor tasa de acierto presenta, con un 89 % de los casos clasificados correctamente. El 11 % restante se distribuye en 10 % clasificado como nivel 1 y un 1 % como nivel 2. En cierto modo, este efecto viene favorecido por el hecho de que el nivel 0 es el que mayor representación tiene en el *dataset*, con casi un 50 %, mientras que las clases 1, 2, 3 y 4 poseen una representación de 28, 15, 6,5 y 0,6 %, respectivamente, lo que podría haber dado lugar a una “especialización” del modelo en la estimación del nivel 0. El nivel 2 presenta una tasa de verdaderos positivos del 54 %, habiéndose predicho la mayoría de los casos restantes como nivel 0. El nivel 2, en cambio, mejora la tasa de acierto del nivel 1, obteniéndose un 65 %, situándose la mayoría de falsos negativos en el nivel 1. En el caso del nivel 3, la tasa de acierto y de fallo se distribuye prácticamente por partes iguales, albergando el nivel 2 de manera exclusiva todos los falsos positivos. El nivel 4, como ya se citó anteriormente, no es representativo y, como consecuencia, el modelo no tiene capacidad de distinguirlo, asociándolo directamente al nivel inmediatamente inferior.

Tabla 2: Modelos óptimos y precisión lograda con los mismos, de acuerdo a (1), en la estimación del nivel de estrés del cultivo para diferentes combinaciones de  $k$  y  $T$ .

$k$	$T$		
	$T_1$	$T_2$	$T_3$
5	Bagged trees (67,8 %)	Bagged trees (68,9 %)	Bagged trees (72,4 %)
15	Bagged trees (71,8 %)	Bagged trees (68,6 %)	Árbol de decisión Medio (69,5 %)
50	Bagged trees (68,4 %)	Bagged trees (67,5 %)	Bagged trees (71,5 %)

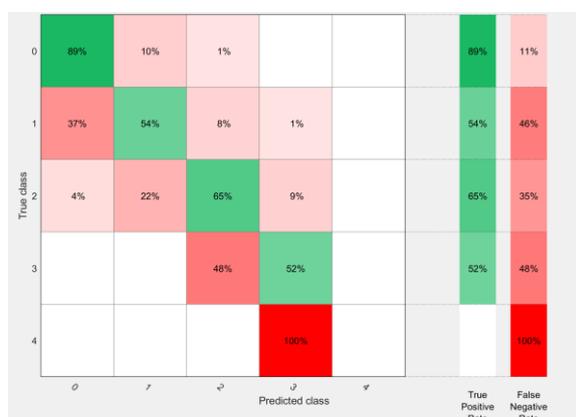


Figura 3: Tabla de verdad de la predicción del nivel de estrés hídrico con el modelo óptimo obtenido.

## 4 CONCLUSIONES

Se ha llevado a cabo un estudio preliminar para la estimación del nivel de estrés hídrico de cerezos adultos, sometidos a diferentes tratamientos de riego deficitario, a partir de la información temporal proporcionada por sensores de estado hídrico del suelo y de variables agro-climáticas. El nivel de estrés hídrico del cultivo ha venido determinado por un experto en base a las medidas de potencial hídrico de tallo a mediodía y de la evolución temporal del resto de variables físicas estudiadas. Dado que la medida del potencial hídrico de tallo es costosa en tiempo y labores, la periodicidad de su obtención es mucho más dilatada que la del resto de variables de suelo y agro-climáticas, por lo que para adaptar la información disponible a la hora de introducirla en un modelo de estimación, para cada muestra del nivel de estrés se ha definido un período de integración,  $T$ , inmediatamente previo a la obtención de la muestra, para el cual se ha calculado el área bajo la curva descrita por las diferentes variables de suelo y agro-climáticas. Se han evaluado múltiples modelos de Machine Learning para realizar la estimación del nivel de estrés hídrico del cultivo a partir de las modificadas variables de suelo y agro-climáticas. Finalmente, el modelo que mejor precisión ha obtenido es el *Bagged Trees* o *Random Forest* bajo el modo de *ensemble Bagging*, para  $k = 5$  y  $T_3$ , logrando una precisión de 72,4 %. El valor de  $k$  y del período de integración de las variables de entrada del modelo ( $T$ ) no se han mostrado excesivamente significativos en la precisión lograda por el modelo.

Se han observado dos posibles causas en la limitación de la precisión del modelo. Por un lado, la tabla de verdad presentada en la Figura 3 muestra una clara tendencia a la distribución de los falsos negativos entre los niveles contiguos al evaluado. Esto indica la existencia de una frontera difusa entre niveles de estrés que lleva al modelo a errar cuando se encuentra un ejemplo situado en la misma. Y, por otro lado, el método de adaptación del formato temporal de las variables de entrada mediante la integral definida en  $T$  de las mismas, tiene en cuenta la información temporal acumulada previa a cada muestra del nivel de estrés, pero no tiene en cuenta el orden en que se han sucedido los valores de las variables y, por tanto, la tendencia. Por lo tanto, en trabajos futuros se prevé trabajar con modelos adaptados a variables de entrada en formato de serie temporal o sucesiones y se revisará el tratamiento previo de los datos de las variables de entrada con el objetivo de reducir la difusividad de las fronteras entre niveles de estrés.

## Agradecimientos

Financiado por el Ministerio de Ciencia e Innovación, el Ministerio de Educación y Formación Profesional y los Fondos Europeos Agrícolas de Desarrollo Rural, números de subvención: AGL2016-77282-C33-R, FPU17/05155, PID2019-106226-C22/AEI/10.13039/501100011033.

## English summary

### ESTIMATION OF THE WATER STRESS LEVEL IN FRUIT TREES USING MACHINE LEARNING FOR APPLICATION IN INTELLIGENT IRRIGATION SYSTEMS

*Water is a limited resource in arid and semi-arid regions. This is the case of the Mediterranean area, where its demographic and climatic conditions make it particularly prone to farming, demanding a major percentage of water resources. Deficit irrigation strategies have proved to be successful, but it is essential to control crop water stress. The measurement of crop water stress is currently associated with midday stem water potential, which is very time-consuming. At an agricultural perspective, it would be interesting to define qualitative levels of crop water stress and to be able to estimate them from variables whose measurement can be automated, so that intelligent irrigation systems can be implemented based on the water needs of the crop. In this work we present a preliminary study to obtain a model capable of predicting five levels of crop water stress from time data of water potential and volumetric water content in the soil and different agro-climatic variables. Multiple Machine Learning algorithms have been evaluated, obtaining a maximum estimation accuracy of 72.4%.*

**Keywords:** Machine Learning, deficit irrigation, crop water stress, smart irrigation.

## Referencias

- [1] J. M. Faurès *et al.*, *Climate-Smart Agriculture Sourcebook*. Rome: FAO, 2013.
- [2] IPCC, "Climate change and water technical paper of the intergovernmental panel on climate change," Geneva, 2008.
- [3] L. Santos Pereira, I. Cordery, and I. Iacovides, *Coping with water scarcity: Addressing the challenges*. Springer Netherlands, 2009.

- [4] EGDH, "Estudio General Demarcación Hidrográfica del Segura." Confederación Hidrográfica del Segura, Murcia, 2007.
- [5] M. C. Ruiz-Sanchez, R. Domingo, and J. R. Castel, "Review. Deficit irrigation in fruit trees and vines in Spain," *Spanish J. Agric. Res.*, vol. 8, no. S2, pp. 5–20, 2010.
- [6] J. Vera, I. Abrisqueta, W. Conejero, and M. C. Ruiz-Sánchez, "Precise sustainable irrigation: a review of soil-plant-atmosphere monitoring," *Acta Hortic. (in Press.)*, 2016.
- [7] E. Fereres and M. A. Soriano, "Deficit irrigation for reducing agricultural water use," *J. Exp. Bot.*, vol. 58, no. 2, pp. 147–159, 2007.
- [8] N. Katerji, M. Mastroianni, and G. Rana, "Water use efficiency of crops cultivated in the Mediterranean region: Review and analysis," *Eur. J. Agron.*, vol. 28, no. 4, pp. 493–507, May 2008.
- [9] R. G. Evans and E. J. Sadler, "Methods and technologies to improve efficiency of water use," *Water Resour. Res.*, vol. 44, 2008.
- [10] K. A. Shackel *et al.*, "Plant water status as an index of irrigation need in deciduous fruit trees," *Horttechnology*, vol. 7, no. 1, pp. 23–29, 1997.
- [11] M. F. Ortuño *et al.*, "Relationships between climatic variables and sap flow, stem water potential and maximum daily trunk shrinkage in lemon trees," *Plant Soil*, vol. 279, no. 1–2, pp. 229–242, Feb. 2006.
- [12] D. S. Intrigliolo and J. R. Castel, "Continuous measurement of plant and soil water status for irrigation scheduling in plum," *Irrig. Sci.*, vol. 23, no. 2, pp. 93–102, May 2004.
- [13] P. Martí, M. Gasque, and P. González-Altozano, "An artificial neural network approach to the estimation of stem water potential from frequency domain reflectometry soil moisture measurements and meteorological data," *Comput. Electron. Agric.*, vol. 91, pp. 75–86, Feb. 2013.
- [14] M. Valdés-Vela, I. Abrisqueta, W. Conejero, J. Vera, and M. C. Ruiz-Sánchez, "Soft computing applied to stem water potential estimation: A fuzzy rule based approach," *Comput. Electron. Agric.*, vol. 115, pp. 150–160, Jul. 2015.
- [15] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration — guidelines for computing crop water requirements.," in *FAO Irrigation and drainage paper 56. Food and Agriculture Organization, Rome.*, Food and Agriculture Organization, Ed. Rome, 1998.
- [16] H. McCutchan and K. A. Shackel, "Stem-water Potential as a Sensitive Indicator of

Water Stress in Prune Trees (*Prunus domestica* L. cv. French),” *Journal of the American Society for Horticultural Science*, vol. 117, no. 4, pp. 607–611, 1992.



© 2021 by the authors.  
Submitted for possible  
open access publication  
under the terms and conditions of the Creative  
Commons Attribution CC BY-NC-SA 4.0 license  
(<https://creativecommons.org/licenses/by-ncsa/4.0/deed.es>).