



Predicción de la producción de raigrases mediante técnicas de aprendizaje automático (*machine learning*)

En este estudio presentamos los resultados de nuestro trabajo de investigación, cuyo objetivo fue obtener un modelo predictivo que permita estimar la producción de materia seca anual de cultivares de diferentes tipos de raigrases a partir de datos meteorológicos probables en las localidades de evaluación.

José Alberto Oliveira Prendes¹, Silverio García Cortés², Nieves Díaz Díaz², Juan Piñeiro Andión³

¹Área de Producción Vegetal, Departamento de Biología de Organismos y Sistemas. Universidad de Oviedo, Asturias, España

²Área de Ingeniería Cartográfica, Geodésica y Fotogrametría, Departamento de Explotación y Prospección de Minas. Universidad de Oviedo, Asturias, España

³Investigadores jubilados del Departamento de Pastos y Cultivos, Centro de Investigaciones Agrarias de Mabegondo, Agencia Gallega de Calidad Alimentaria (Agacal), Xunta de Galicia, Galicia, España

INTRODUCCIÓN

Entre 1978 y 2014 en el Centro de Investigaciones Agrarias de Mabegondo (CIAM), dentro de las actividades del Departamento de Pastos y Cultivos se llevó a cabo la evaluación de variedades comerciales de especies pratenses con el fin de conocer su valor agronómico (Díaz *et al.*, 2015). La experimentación sobre el terreno requiere mucho tiempo y recursos, por lo que es importante la integración de la

información obtenida en dichos ensayos de campo en una herramienta que facilite la toma de decisiones (Sau *et al.*, 2010).

El rendimiento de los cultivos es el resultado de interacciones complejas entre genotipos, ambientes y prácticas de manejo (McCauley, 2020), por lo que se intenta predecir con modelos de simulación basados en procesos físicos, químicos y biológicos. Dichos modelos de simulación utilizan datos obtenidos en

ensayos de campo con el fin de simular el crecimiento y el desarrollo de los cultivos teniendo en cuenta el ambiente y las prácticas de manejo (Muchow *et al.*, 1990).

Con el fin de simular el crecimiento y el desarrollo de cuatro cultivares representativos de raigrases en tres localidades gallegas (Mabegondo, Marco da Curra y A Pobra do Brollón), Oliveira *et al.* (2020) iniciaron la adaptación del modelo CROPGRO-PFM incluido en



► ESTOS MODELOS DE SIMULACIÓN UTILIZAN DATOS OBTENIDOS EN ENSAYOS DE CAMPO CON EL FIN DE SIMULAR EL CRECIMIENTO Y EL DESARROLLO DE LOS CULTIVOS TENIENDO EN CUENTA EL AMBIENTE Y LAS PRÁCTICAS DE MANEJO

el paquete informático del Sistema de Apoyo a las Decisiones para la Transferencia de Agrotecnología (DSSAT) (Hoogenboom *et al.*, 2019).

Por otra parte, la disponibilidad de grandes cantidades de datos de orígenes diversos, como las imágenes satélites, datos meteorológicos, evaluaciones de campo y sensores ha permitido el desarrollo de modelos de aprendizaje automático (*machine learning*) dentro del campo de la Inteligencia Artificial, para predecir el rendimiento de los cultivos (Dhaliwal y Williams, 2024). Estos modelos de aprendizaje automático son más adecuados que las herramientas estadísticas clásicas (análisis de varianza, regresión, etc.) para modelizar relaciones complejas y no lineales en grandes conjuntos de datos sin hacer suposiciones fuertes, mientras que las técnicas estadísticas clásicas ofrecen mayor interpretabilidad y se centran más en la inferencia y la comprensión de las relaciones entre variables (Roberts *et al.*, 2017).

MATERIALES Y MÉTODOS

Para la creación del conjunto de datos se recopilaron datos de los ensayos de evaluación de variedades comerciales de especies pratenses realizados por el CIAM en tres localidades de evaluación: Mabegondo (43° 14' 5" N, 8° 15' 45" O, altitud = 100 m) en la provincia de A Coruña, A Pobra do Brollón (42° 35' 40" N, 7° 24' 31" O, altitud = 400 m) en la provincia de Lugo y Grado (43° 22' 35" N, 6° 3' 45" O, altitud = 50 m), en Asturias. Se consideraron dos cultivares de raigrás italiano no alternativos ("Combata", 2x, 11 años de evaluación y "Barmultra", 4x, siete años de evaluación), un cultivar de raigrás híbrido ("Sabel", 4x, 9-10 años de evaluación según la localidad) y dos cultivares de raigrás inglés ("Brigantía", 2x, 16-17 años de evaluación según la localidad y Reveille", 4x, 12 años de evaluación).

Como variables climáticas se consideraron las siguientes: temperaturas medias de mínimas (Tmin, °C) y de máximas anuales (Tmax, °C), precipitaciones anuales (mm) y radiación solar media diaria (MJ m⁻² día⁻¹). Los datos meteorológicos de las localidades de evaluación se obtuvieron de la base de datos climáticos del sistema NASA-POWER (<https://power.larc.nasa.gov/>).

Como variable obtenida en los ensayos de los raigrases se consideró el rendimiento en kg de materia seca (MS) por ha y año del primer año de ensayo, tras la siembra en el otoño previo.

El fichero de datos comprendía 168 registros y 7 variables (localidad, cultivar, Tmax, Tmin, precipitación, radiación solar y kg MS ha⁻¹).

Se realizó un análisis de varianza considerando los factores cultivar, localidad considerados como factores fijos y su interacción.

Se probaron diferentes técnicas de aprendizaje automático (Random Forest Regressor, Support Vector Machine, Light Gradient Boosting Machine y Gradient Boosting Regressor) para ajustar un modelo al fichero de datos anterior con el fin de predecir la variable kg MS ha⁻¹ a partir del resto de las variables.

Se usó el lenguaje Python como lenguaje de programación dentro del entorno interactivo de Jupyter Notebook (Granger y Pérez, 2021).

Los análisis de varianza se realizaron con el programa estadístico IBM SPSS Statistics 24 (Field, 2013).

RESULTADOS Y DISCUSIÓN

El análisis de varianza, considerando los factores localidad, cultivar como factores fijos y su interacción (tabla 1), mostró diferencias significativas para la localidad, $F(2,153) = 19,4$, $p < 0,001$ y el cultivar $F(4, 153) = 7,4$, $p < 0,001$, pero no para la interacción localidad x cultivar, $F(8,153) = 1,3$, NS.

En la tabla 2 se puede observar que las localidades más productivas fueron Mabegondo y Grado. ►►

Tabla 1. Análisis de varianza de los factores localidad y cultivar sobre la producción de los raigrases en kg MS ha⁻¹

Fuentes de variación	Grados de libertad	Cuadrados medios	F
Localidad	2	81.167.231,7	19,4***
Cultivar	4	30.883.381,6	7,4***
Localidad x cultivar	8	5.521.304,2	1,3NS
Error	153	4.177.360,5	
Total	168		

NS = no significativo ($p \geq 0,05$), *** = significativo a $p < 0,001$, F = F-test

Tabla 2. Medias de la variable kg MS ha⁻¹ en función de la localidad y el cultivar. Medias de las localidades o de los cultivares compartiendo una misma letra no son significativamente diferentes según el test de Tukey ($p < 0,05$), siendo a > b

Factores	kg MS ha ⁻¹	
	Medias	Desviación estándar
Localidad		
A Pobra do Brollón	11149b	2249,2
Grado	12881a	2419,1
Mabegondo	13602a	1946,7
Cultivar		
Reveille	11761b	1956,2
Brigantía	11801b	2386,2
Sabel	12699ab	2353,4
Barmultra	13379a	2749,2
Combata	13881a	2150,0

Los cultivares más productivos en el primer año tras la siembra en el otoño anterior fueron los dos cultivares de raigrás italiano no alternativo (Combita y Barmultra) y el cultivar de raigrás híbrido Sabel.

La figura 1 muestra los valores medios de las producciones de MS por ha por localidades y para los cinco cultivares de raigrases.

De las técnicas *machine learning* probadas (Random Forest Regressor, Support Vector Machine, Light Gradient Boosting Machine y Gradient Boosting Regressor), Gradient Boosting Regressor ha sido la que ha dado mejores resultados empleando validación cruzada con un K-fold ($n.^{\circ}$ de subgrupos) = 5 y una optimización de los hiperparámetros con el método Optuna (Akiba *et al.*, 2019).

En el caso de juegos de datos con pocas muestras (168 muestras x 7 variables) es común emplear la técnica de “validación cruzada”, más conocida por su término en inglés K-fold *cross-validation*. Este método permite disponer de mayor número de muestras para entrenar (*train*) y también para comprobar (test o validación) el modelo, capturando la mayor parte de la información de juego total de datos y permitiendo esperar mejores resultados que en lugar de usar diferentes conjuntos fijos de datos para entrenamiento y validación como ocurre en otras técnicas.

En la figura 2 se puede observar que la predicción de los kg MS ha⁻¹ por el modelo Gradient Boosting Regressor presenta bastante dispersión con respecto a la línea diagonal. Esto indica que el modelo no captura toda la variabilidad presente en los datos. Sin embargo, existe una dispersión equilibrada por encima y por debajo de la línea de igualdad mostrando que probablemente la mayor parte de estas diferencias se deben a un ruido con distribución aleatoria. Quizá sería necesario encontrar nuevas variables descriptoras que pudieran aportar más información para disminuir la dispersión aleatoria de estas diferencias. En general, puede verse también que la dispersión podría aumentar con el valor predicho (heterocedasticidad) y, además, existe una mayor concentración de buenos resultados en rangos intermedios.

En la figura 3 se observa que las variables que más influyeron en

Figura 1. Valores medios y sus desviaciones estándar de las producciones (kg MS ha⁻¹) de cada uno de los cultivares en las tres localidades de evaluación

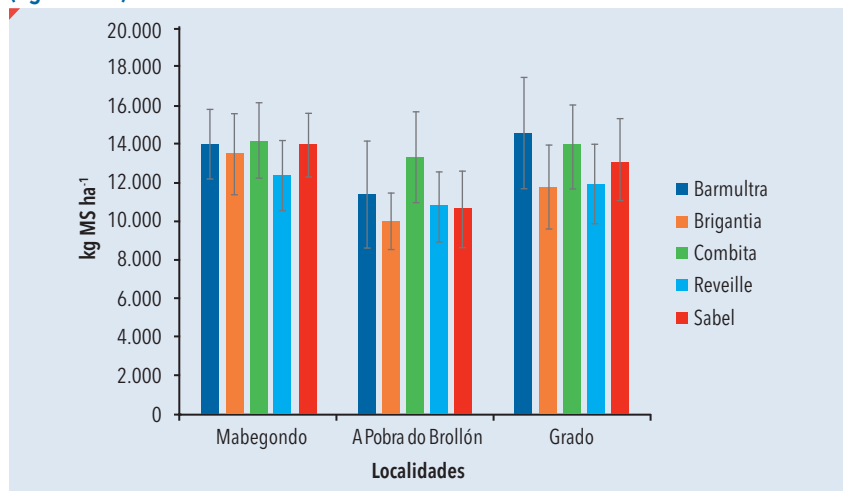


Figura 2. Gráfico de dispersión de valores predichos, mediante el modelo Gradient Boosting Regressor optimizado con Optuna, frente a valores observados de la variable kg MS ha⁻¹. La línea diagonal representa la relación de igualdad entre los valores observados y predichos

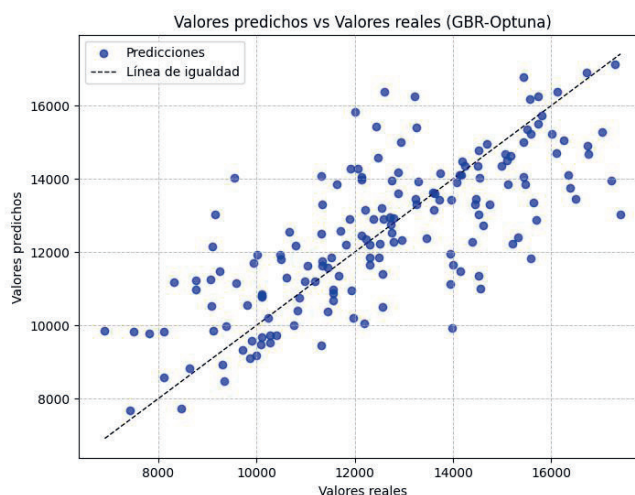
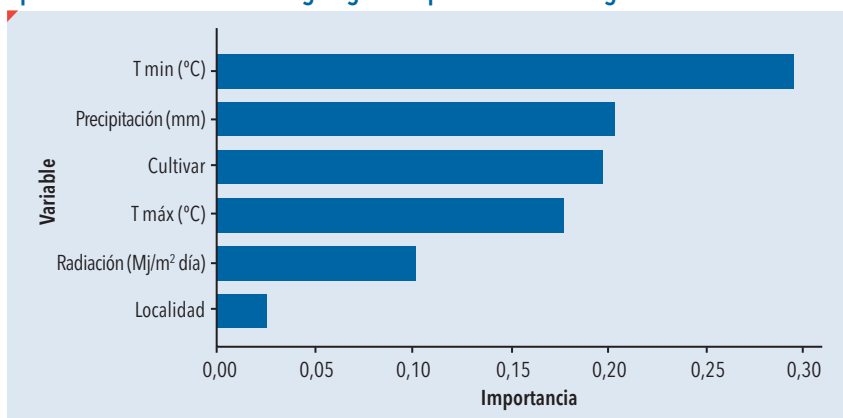


Figura 3. Importancia relativa de las variables calculadas mediante el modelo optimizado Gradient Boosting Regressor para la variable kg MS ha⁻¹



la predicción de kg MS ha⁻¹ fueron Tmin, Precipitación y Cultivar.

Las métricas (que informan sobre la calidad del ajuste) obte-

nidas por el modelo tras la optimización de los hiperparámetros mediante el método Optuna, fueron las siguientes: ►►

Mezclas Forrajeras

NUEVO CATÁLOGO

6 fórmulas anuales, adaptadas para cada tipo de aprovechamiento

9 fórmulas plurianuales, para cada condición, ambiente, terreno y aprovechamiento



Mezclas con leguminosas y gramíneas



Incluyen especies mejorantes

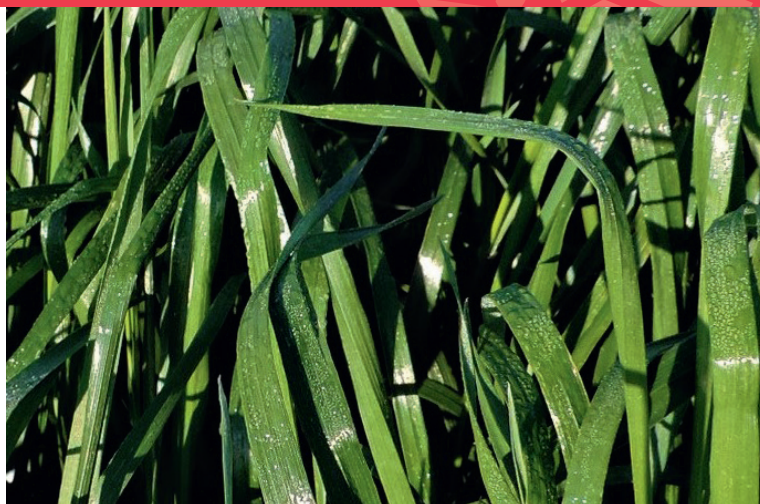
Adaptadas para dar **cumplimiento a los requisitos PAC 2025**

Fijan nitrógeno en el suelo que será aprovechado por el cultivo siguiente en la rotación

Raigrás italiano: garantía de producción y calidad

ELIRIX: Rápido en establecimiento y rebrote tras corte. Muy productivo y apetente al ganado.

ADOR: Rústico y vigoroso. Seguridad de producción en todas condiciones.



▶ AUNQUE EL MODELO LOGRÓ EXPLICAR EL 54 % DE LA VARIABILIDAD TOTAL Y EL RMSE REPRESENTA UN 13 % DEL VALOR MEDIO OBSERVADO, ESTOS RESULTADOS INDICAN UN RENDIMIENTO PREDICTIVO (DESEMPEÑO) MODERADO

- 1) error absoluto medio (MAE): 1.276,8 kg MS ha⁻¹ (cuanto más bajo mejor),
- 2) raíz del error cuadrático medio (cuanto más bajo mejor) o error de predicción (RMSE): 1.649,2 kg MS ha⁻¹ (13 % del rendimiento medio, 125.48 kg MS ha⁻¹),
- 3) coeficiente de determinación o capacidad del modelo para predecir futuros resultados (R²): 0,54 (cuanto más cerca de 1, mejor).

A la vista de la figura 3, se confirma que las variables climáticas y genéticas (Tmin, Precipitación y Cultivar) son las que más influyen en las predicciones del modelo.

La elevada importancia de Tmin sugiere que el rendimiento es particularmente sensible a las temperaturas medias de mínimas anuales. Las temperaturas mínimas demasiado bajas pueden causar daños, retrasar el crecimiento y reducir la producción.

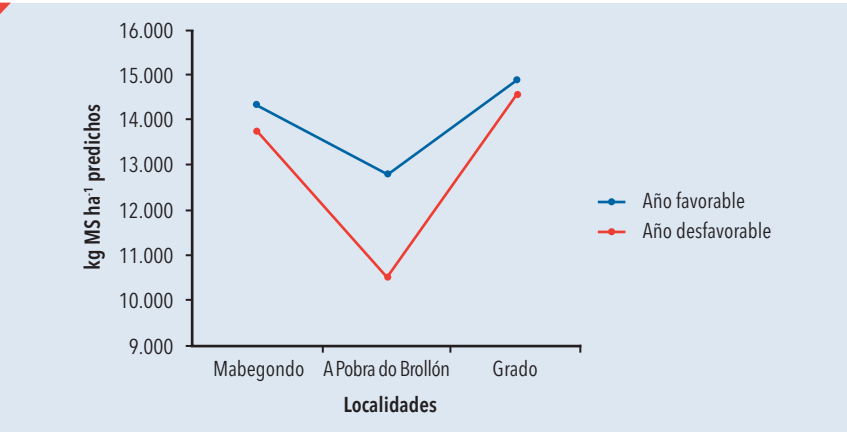
Aunque el modelo logró explicar el 54 % de la variabilidad total y el RMSE representa un 13 % del valor medio observado, estos resultados indican un rendimiento predictivo (desempeño) moderado. Para mejorar la capacidad predictiva, sería conveniente ampliar el conjunto de datos y considerar la inclusión de nuevas variables adicionales como índices fenológicos (NDVI), manejo agronómico y características edafi-

Tabla 3. Análisis de varianza de los factores año, cultivar y localidad sobre la producción de kg MS ha⁻¹ predicha

Fuentes de variación	Grados de libertad	Cuadrados medios	F
Año	1	252.206.272,6	230,2***
Cultivar	4	2.289.726,9	2,1NS
Localidad	2	774.782.161,4	707,1***
Año x cultivar	4	1.163.423,0	1,1NS
Año x localidad	2	88.756.329,3	81,0***
Cultivar x localidad	8	2.519.549,2	2,3*
Año x cultivar x localidad	8	2.549.141,6	2,3*
Error	870	1.095.754,8	
Total	900		

NS = no significativo ($p \geq 0,05$), * = significativo a $p < 0,05$, *** = significativo a $p < 0,001$, F = F-test
La interacción triple entre año, cultivar y localidad fue estadísticamente significativa, $F(8, 870) = 2,3$, $p < 0,05$

Figura 4. Medias de las producciones predichas de los cultivares de los raigrases en cada una de las localidades de evaluación para los años favorables y desfavorables



cas que podrían contribuir a mejorar la capacidad predictiva del modelo.

Con el fin de poder hacer predicciones de la producción de kg MS ha⁻¹ con el modelo obtenido, se generaron un conjunto de 30 datos para las variables Tmin y precipitación para cada localidad y cultivar mediante una distribución uniforme que genera datos de manera aleatoria dentro de un intervalo definido. Se crearon aleatoriamente 30 datos para un año favorable considerando temperaturas mínimas altas y precipitaciones altas (entre el valor medio y el valor máximo en los años de evaluación).

Para un año desfavorable se repite el proceso con temperaturas mínimas bajas y precipitaciones bajas (entre el valor medio y el valor mínimo de los años de evaluación). En el caso de las variables temperatura máxima y radiación diaria, se tuvieron los valores medios obtenidos en los años de evaluación.

Posteriormente, se realizó un análisis de varianza con los factores año, cultivar y localidad, considerados como factores fijos y sus interacciones para ver si había diferencias significativas en los valores predichos de producción en kg MS ha⁻¹ (tabla 3).

Hubo una interacción doble estadísticamente significativa entre año y localidad, $F(2,870) = 81,0$, $p < 0,001$ y entre cultivar y localidad, $F(8, 870) = 2,3$, $p < 0,05$, pero no entre año y cultivar, $F(4,870) = 1,1$, NS. Estas interacciones indican que las localidades influyeron de manera diferente según los años y los cultivares.

Cuando se observan interacciones dobles significativas (por ej., año x localidad), es el efecto de esta interacción lo que es de interés estudiar, debido a que los efectos de los factores principales (año, localidad y cultivar) tienen menos interés, pues en este caso el efecto de la localidad varía según el año. ▶▶

CALCIMAG

GRANULADO



CALCIMAG

ES MÁS QUE UNA
ENMIENDA CÁLCICA,
ES UN MEJORANTE
DE SUELOS



La única enmienda líquida a base de
carbonato de calcio para mezclar con el purín

Encale y fertilice en la misma aplicación

- ✓ Estimula la nitrificación
- ✓ Reduce los malos olores
- ✓ Reduce la emisión de gases de efecto invernadero

Descargue la **presentación completa** del
producto con todas sus características:



► AUNQUE EL ANÁLISIS DE VARIANZA NO ENCONTRÓ DIFERENCIAS SIGNIFICATIVAS ENTRE CULTIVARES EN CUANTO A LA PRODUCCIÓN PREDICHA, EL MODELO GRADIENT BOOSTING REGRESSOR MOSTRÓ QUE LA VARIABLE "CULTIVAR" FUE LA TERCERA MÁS INFLUYENTE PARA LAS PREDICCIONES



Hubo un efecto estadísticamente significativo para los factores año, $F(1,870) = 230,2$, $p < 0,001$, y localidad, $F(2,870) = 707,1$, $p < 0,001$ pero no para el factor cultivar, $F(4,870) = 2,1$, NS.

Los valores de producción de kg MS ha⁻¹ predichos considerando los valores de todos los cultivares en conjunto y la interacción año x localidad se representan en la figura 4.

Esta interacción de tipo cuantitativo o de escala (Allard y Bradshaw, 1964) indica que el valor de la producción varía entre localidades, siendo consistentemente mayor en el año favorable que en el desfavorable para todas ellas. En particular, la localidad de A Pobra presenta la mayor diferencia de producción entre ambos años, mientras que Grado destaca por alcanzar los valores más altos en ambas condiciones.

Por otro lado, aunque el análisis de varianza no encontró diferencias significativas entre cultivares en cuanto a la producción predicha, el modelo Gradient Boosting Regressor mostró que la variable "Cultivar" fue la tercera más influyente para las predicciones. Esta discrepancia puede deberse a que el análisis de varianza evalúa principalmente efectos lineales y aditivos, mientras que el modelo de aprendizaje automático puede capturar relaciones no lineales y complejas entre las variables de entrada y las variables de salida.

CONCLUSIONES

El modelo de aprendizaje automático obtenido tiene el uso potencial de ser una herramienta para predecir la producción de los raigrases teniendo en cuenta la localidad de cultivo y valores probables de la temperatura media de mínimas y la precipitación anual.

Aunque los resultados del modelo son prometedores, sería necesario mejorarlo con un mayor número de datos y nuevas variables descriptoras. ■

BIBLIOGRAFÍA

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. En Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2623–2631). ACM. <https://doi.org/10.1145/3292500.3330701>

Allard, R.W., Bradshaw, A.D., 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Science*, 4: 503-508.

Dhaliwal, D.S., Willimas, M.M., 2024. Sweet corn yield prediction using machine learning models and field-level data. *Precision Agriculture*, 25: 51-64.

Díaz, N., Díaz, D., Crecente, S., Flores, G., 2015. Valor agronómico de las variedades comerciales de gramíneas y leguminosas pratenses. *Evaluación de 2014. Afriga*, año XXI, 118: 95-106.

Field, A., 2013. *Discovering Statistics Using IBM SPSS Statistics*; Sage: London, UK.

Granger, B., Pérez, F., 2021. Jupyter: Thinking and Storytelling with Code and Data. *Computing in Science & Engineering*, 23(2): 7-14.

Hoogenboom, G., Porter, C.H., Shelia, V., Boote, K.J., Singh, U., White, J.W., Hunt, L.A., Ogoshi, R., Lizaso, J.I., Koo, J., Asseng, S., Singels, A., Moreno, L.P., Jones, J.W., 2019. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.7.5 (<https://DSSAT.net>). DSSAT Foundation, Gainesville, Florida, USA.

McCauley, D.J., 2020. Predictive agriculture. *Crop modeling for the future*. *CSA News Magazine*, 65 (5): 3-9.

Muchow, R.C., Sinclair, T.R., Bennet, J.M., 1990. Temperature and solar radiation effects on potential maize yield across locations. *Agronomy Journal*, 82(2): 338-343.

Oliveira, J.A., Boote, K.J., López, J.E., Díaz, N., Piñeiro, J., Flores, G., 2020. Adaptación del modelo CROPGRO-Perennial-Forage para simular la producción de los raigrases. *Revista Vaca Pinta* 19: 140-155.

Roberts, M.J., Braun, N.O., Sinclair, T.R., Lobell, D.B., Schlenker, W., 2017. Comparing and combining process-based models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9), 095010: 1-15.

Sau, F., López-Cedrón, F.X., Confalone, A.E., Lizaso, J.I., 2010. Modelos de simulación del cultivo del maíz: Fundamentos y aplicación en España. *Pastos*, 40(2): 117-138.

AMPLIA GAMA DE PRODUCTOS PARA UN **MEJOR RENDIMIENTO DE TU PRADERA**

GALICAL

CALES Y CARBONATOS AGRÍCOLAS

Para reducir el efecto limitante del pH y controlar el aluminio en las tierras de cultivo **es recomendable aplicar enmiendas calizas o magnésicas.**

EXTENDIDO REGULADO POR GPS



- **ENMIENDA CALIZA, CAL VIVA GRANULADA** (90 % CaO)
Alto porcentaje en calcio. Valor neutralizante: 90 %

- **ENMIENDA CALIZA, CAL VIVA GRANULADA DOLOMÍTICA**
(35 % MgO / 60 % CaO)
Alto porcentaje de magnesio. Valor neutralizante: 95 %

- **ENMIENDA CALIZA, CAL VIVA** (80 % CaO)
Gran poder de neutralización. Valor neutralizante: 80 %

- **ENMIENDA CALIZA, CAL APAGADA** (65 % CaO)
Potencia el rendimiento agrícola. De fácil asimilación.
Valor neutralizante: 65 %

- **ENMIENDA CALIZA, CAL APAGADA + DOLOMÍA**
(53 % CaO / 23 % MgO)
Aporta magnesio. Favorece la actividad clorofílica de la planta.
Valor neutralizante: 83 %

- **ENMIENDA CALIZA, CARBONATO CÁLCICO** (56 % CaO)
Para tierra y camas higiénicas. Efi caz en la reducción de
mamitis ambientales y dermatitis. Apropiado para la producción
de todo tipo de piensos. Valor neutralizante: 56 %

- **ENMIENDA CALIZA DE CARBONATO CÁLCICO MAGNÉSICO,
GALIMAG** (33 % CaO / 17 % MgO)
Aporta magnesio. Valor neutralizante: 60 %

- **ENMIENDA CALIZA, GRANICAL G60 SUPRA** (54 % CaO / 1 % MgO)
Fácil aplicación y asimilación. Valor neutralizante: 56 %

- **ENMIENDA CALIZA, GRANICAL G60 PLUS** (47,35 % CaO / 6 % MgO)
Fácil aplicación y asimilación con aporte de magnesio.
Valor neutralizante: 59 %

Transporte a cualquier punto de España y Portugal



EN VÍDEO

ENCALADO AGRÍCOLA · ENVASADO DE CALES Y DOLOMÍAS · ALIMENTACIÓN ANIMAL · CAMAS DE VACUNO



GALICAL CALES Y CARBONATOS AGRÍCOLAS SL

📍 C/ Gallastegui Unamuno. Vial G - N.º 7 Polígono Industrial As Gándaras 27003 Lugo

☎ 982 221 484

✉ info@galical.es



www.galical.es