



Predicción da produción de raigrases mediante técnicas de aprendizaxe automática (*machine learning*)

Neste estudo presentamos os resultados do noso traballo de investigación, cuxo obxectivo foi obter un modelo predictivo que permita estimar a produción de materia seca anual de cultivares de diferentes tipos de raigrases a partir de datos meteorolóxicos probables nas localidades de avaliación.

José Alberto Oliveira Prendes¹, Silverio García Cortés², Nieves Díaz Díaz³, Juan Piñeiro Andión³

¹Área de Producción Vexetal, Departamento de Bioloxía de Organismos e Sistemas. Universidade de Oviedo, Asturias, España

²Área de Enxeñaría Cartográfica, Xeodésica e Fotogrametría, Departamento de Explotación e Prospección de Minas. Universidade de Oviedo, Asturias, España

³Investigadores xubilados do Departamento de Pastos e Cultivos, Centro de Investigacións Agrarias de Mabegondo, Axencia Galega de Calidade Alimentaria (Agacal), Xunta de Galicia, Galicia, España

INTRODUCCIÓN

Entre 1978 e 2014 no Centro de Investigacións Agrarias de Mabegondo (CIAM), dentro das actividades do Departamento de Pastos e Cultivos levouse a cabo a avaliación de variedades comerciais de especies pratenses co fin de coñecer o seu valor agronómico (Díaz *et al.*, 2015).). A experimentación sobre o terreo require moito tempo e recursos, polo que é importante a integración da información obtida nos

devanditos ensaios de campo nunha ferramenta que facilite a toma de decisións (Sau *et al.*, 2010).

O rendemento dos cultivos é o resultado de interaccións complexas entre xenotipos, ambientes e prácticas de manexo (McCauley, 2020), polo que se tenta predicir con modelos de simulación baseados en procesos físicos, químicos e biolóxicos. Os devanditos modelos de simulación utilizan datos obtidos en ensaios de campo co fin de simular

o crecemento e o desenvolvemento dos cultivos tendo en conta o ambiente e as prácticas de manexo (Muchow *et al.*, 1990).

Co fin de simular o crecemento e o desenvolvemento de catro cultivares representativos de raigrases en tres localidades galegas (Mabegondo, Marco da Curra e A Pobra do Brollón), Oliveira *et al.* (2020) iniciaron a adaptación do modelo CROPGRO-PFM incluído no paquete informático do Sistema de Apoio ás



► ESTES MODELOS DE SIMULACIÓN UTILIZAN DATOS OBTIDOS EN ENSAIOS DE CAMPO CO FIN DE SIMULAR O CRECEMENTO E O DESENVOLVEMENTO DOS CULTIVOS TENDO EN CONTA O AMBIENTE E AS PRÁCTICAS DE MANEXO

Decisións para a Transferencia de Agrotecnoloxía (DSSAT) (Hoogenboom *et al.*, 2019).

Así mesmo, a dispoñibilidade de grandes cantidades de datos de orixes diversas, coma as imaxes satélites, datos meteorolóxicos, avaliacións de campo e sensores permitiu o desenvolvemento de modelos de aprendizaxe automática (*machine learning*) dentro do campo da Intelixencia Artificial, para predicir o rendemento dos cultivos (Dhaliwal e Williams, 2024). Estes modelos de aprendizaxe automática son máis adecuados que as ferramentas estatísticas clásicas (análise de varianza, regresión etc.) para modelizar relacións complexas e non lineais en grandes conxuntos de datos sen facer suposicións fortes, mentres que as técnicas estatísticas clásicas ofrecen maior interpretabilidade e céntranse máis na inferencia e a comprensión das relacións entre variables (Roberts *et al.*, 2017).

MATERIAIS E MÉTODOS

Para a creación do conxunto de datos recompiláronse datos dos ensaios de avaliación de variedades comerciais de especies pratenses realizados polo CIAM en tres localidades de avaliación: Mabegondo (43° 14' 5" N, 8° 15' 45" O, altitude = 100 m) na provincia da Coruña, A Pobra do Brollón (42° 35' 40" N, 7° 24' 31" O, altitude = 400 m) na provincia de Lugo e Grado (43° 22' 35" N, 6° 3' 45" O, altitude = 50 m), en Asturias. Consideráronse dous cultivares de raigrás italiano non alternativos ("Combita", 2x, 11 anos de avaliación e "Barmultra", 4x, sete anos de avaliación), un cultivar de raigrás híbrido ("Sabel", 4x, 9-10 anos de avaliación segundo a localidade) e dous cultivares de raigrás inglés ("Brigantia", 2x, 16-17 anos de avaliación segundo a localidade e Reveille", 4x, 12 anos de avaliación).

Como variables climáticas consideráronse as seguintes: temperaturas medias de mínimas (Tmin, °C) e de máximas anuais (Tmax, °C), precipitacións anuais (mm) e radiación solar media diaria (MJ m⁻² día⁻¹). Os datos meteorolóxicos das localidades de avaliación obtivéronse da base de datos climáticos do sistema NASA-POWER (<https://power.larc.nasa.gov/>).

Como variable obtida nos ensaios dos raigrases considerouse o rendemento en kg de materia seca (MS) por ha e ano do primeiro ano de ensaio, tras a sementeira no outono previo.

O ficheiro de datos comprendía 168 rexistros e 7 variables (localidade, cultivar, Tmax, Tmin, precipitación, radiación solar e kg MS ha⁻¹).

Realizouse unha análise de varianza, tendo en conta os factores cultivar e localidade, considerados como factores fixos e a súa interacción.

Probáronse diferentes técnicas de aprendizaxe automática (Random Forest Regressor, Support Vector Machine, Lixeiro Gradient Boosting Machine e Gradient Boosting Regressor) para axustar un modelo ao ficheiro de datos anterior co fin de predicir o variable kg MS ha⁻¹ a partir do resto das variables.

Usouse a linguaxe Python como linguaxe de programación dentro do marco interactivo de Jupyter Notebook (Granger e Pérez, 2021).

As análises de varianza realizáronse co programa estatístico IBM SPSS Statistics 24 (Field, 2013).

RESULTADOS E DISCUSIÓN

A análise de varianza, considerando os factores localidade e cultivar como factores fixos e a súa interacción (táboa 1), mostrou diferenzas significativas para a localidade, $F(2,153) = 19,4$, $p < 0,001$ e o cultivar $F(4, 153) = 7,4$, $p < 0,001$, pero non para a interacción localidade x cultivar, $F(8,153) = 1,3$, NS.

Na táboa 2 pódese observar que as localidades máis produtivas ►

Táboa 1. Análise de varianza dos factores localidade e cultivar sobre a produción dos raigrases en kg MS ha⁻¹

Fontes de variación	Graos de liberdade	Cadrados medios	F
Localidade	2	81.167.231,7	19,4***
Cultivar	4	30.883.381,6	7,4***
Localidade x cultivar	8	5.521.304,2	1,3NS
Erro	153	4.177.360,5	
Total	168		

NS = non significativo ($p \geq 0,05$), *** = significativo a $p < 0,001$, F = F-test

Táboa 2. Medias da variable kg MS ha⁻¹ en función da localidade e do cultivar. Medias das localidades ou de cultivos compartindo unha mesma letra non son significativamente diferentes segundo o test de Tukey ($p < 0,05$), sendo a>b

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

foron Mabegondo e Grado. Cultivos máis produtivos no primeiro ano tras a sementeira no outono anterior foron os dous cultivares de raigrás italiano non alternativo (Combita e Barmultra) e o cultivar de raigrás híbrido Sabel.

A figura 1 mostra os valores medios das producións de MS por ha por localidades e para os cinco cultivares de raigrases.

Das técnicas *machine learning* probadas (Random Forest Regressor, Support Vector Machine, Lixeiro Gradient Boosting Machine e Gradient Boosting Regressor), Gradient Boosting Regressor foi a que deu mellores resultados empregando validación cruzada cun K-fold (n.º de subgrupos) = 5 e unha optimización dos hiperparámetros co método Optuna (Akiba *et al.*, 2019).

No caso de xogos de datos con poucas mostras (168 mostras x 7 variables) é común empregar a técnica de “validación cruzada”, máis coñecida polo seu termo en inglés K-fold *cross-validation*. Este método permite dispor dun maior número de mostras para adestrar (*train*) e tamén para comprobar (test ou validación) o modelo, capturando a maior parte da información de xogo total de datos e permitindo esperar mellores resultados que en lugar de usar diferentes conxuntos fixos de datos para adestramento e validación como ocorre noutras técnicas.

Na figura 2 pódese observar que a predición do kg MS ha⁻¹ polo modelo Gradient Boosting Regressor presenta bastante dispersión con respecto da liña diagonal. Isto indica que o modelo non captura toda a variabilidade presente nos datos. Non obstante, existe unha dispersión equilibrada por riba e por baixo da liña de igualdade, o que mostra que probablemente a maior parte destas diferenzas se deben a un ruído con distribución aleatoria. Quizá sería necesario atopar novas variables descritoras que puidesen achegar máis información para diminuír a dispersión aleatoria destas diferenzas. En xeral, pode verse tamén que a dispersión podería aumentar co valor predito (heterocedasticidade) e, ademais, existe unha maior concentración de bos resultados en rangos intermedios.

Na figura 3 obsérvase que as variables que máis influíron na predición de kg MS ha⁻¹ foron Tmin, Precipitación e Cultivar.

As métricas (que informan sobre a calidade do axuste) obtidas polo

Figura 1. Valores medios e as súas desviacións estándar das producións (kg MS ha⁻¹) de cada un dos cultivares nas tres localidades de avaliación

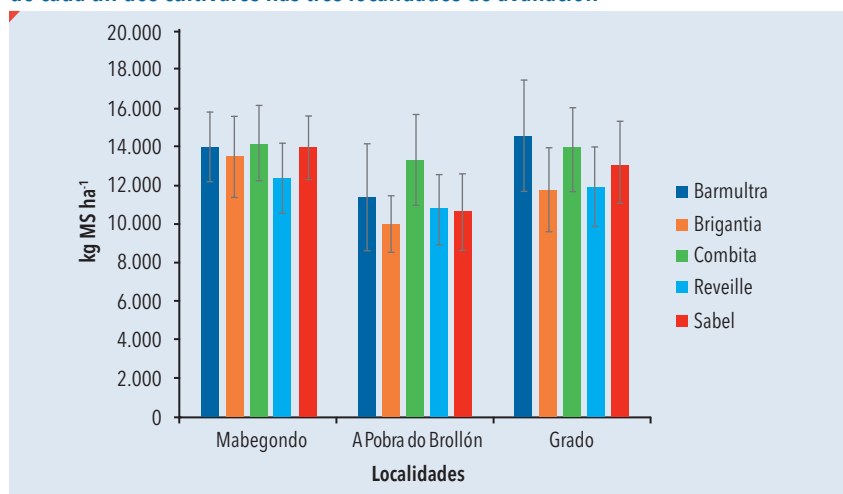


Figura 2. Gráfico de dispersión de valores preditos, mediante o modelo Gradient Boosting Regressor optimizado con Optuna, fronte a valores observados do variable kg MS ha⁻¹. A liña diagonal representa a relación de igualdade entre os valores observados e preditos

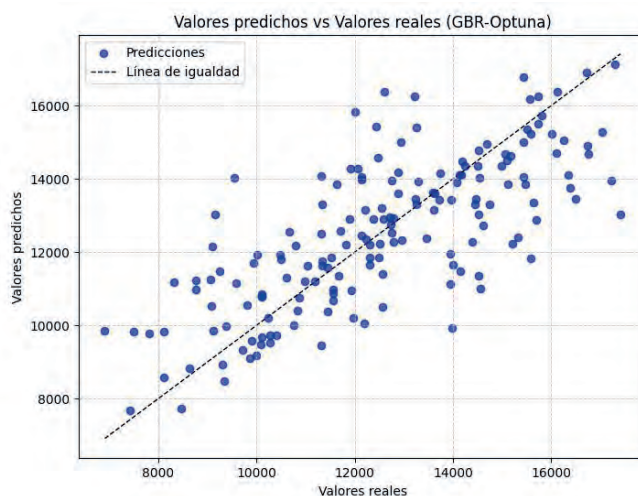
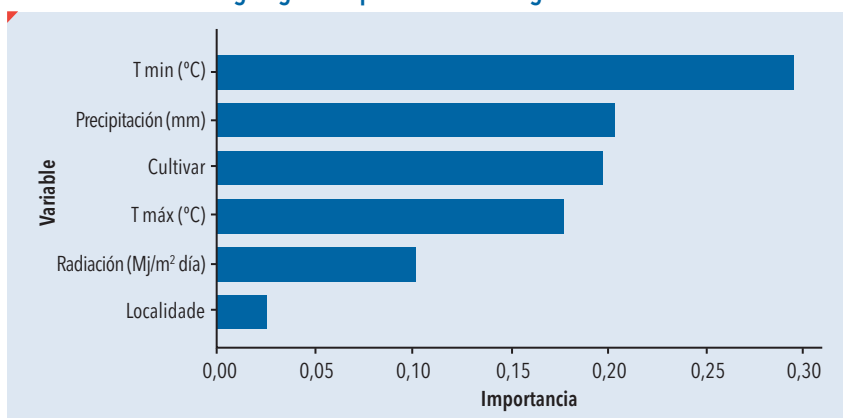


Figura 3. Importancia relativa das variables calculadas mediante o modelo optimizado Gradient Boosting Regressor para o variable kg MS ha⁻¹



modelo tras a optimización dos hiperparámetros mediante o método Optuna, foron as seguintes:

1) erro absoluto medio (MAE): 1.276,8 kg MS ha⁻¹ (canto máis baixo mellor), ►►

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.



▶ AÍNDA QUE O MODELO LOGROU EXPLICAR O 54 % DA VARIABILIDADE TOTAL E O RMSE REPRESENTA UN 13 % DO VALOR MEDIO OBSERVADO, ESTES RESULTADOS INDICAN UN RENDEMENTO PREDITIVO (DESEMPEÑO) MODERADO

2) raíz do erro cuadrático medio (canto máis baixo mellor) ou erro de predición (RMSE): 1.649,2 kg MS ha⁻¹ (13 % do rendemento medio, 125.48 kg MS ha⁻¹),

3) coeficiente de determinación ou capacidade do modelo para predecir futuros resultados (R²): 0,54 (canto máis preto de 1, mellor).

Á vista da figura 3, confirmase que as variables climáticas e xenéticas (Tmin, Precipitación e Cultivar) son as que máis influen nas predicións do modelo. A elevada importancia de Tmin suxire que o rendemento é particularmente sensible ás temperaturas medias de mínimas anuais. As temperaturas mínimas demasiado baixas poden causar danos, atrasar o crecemento e reducir a produción.

Aínda que o modelo logrou explicar o 54 % da variabilidade total e o RMSE representa un 13 % do valor medio observado, estes resultados indican un rendemento predictivo (desempeño) moderado. Para mellorar a capacidade predictiva, sería conveniente ampliar o conxunto de datos e considerar a inclusión de novas variables adicionais como índices fenolóxicos (NDVI), manexo agronómico e características edáficas que poderían contribuír a mellorar a capacidade predictiva do modelo.

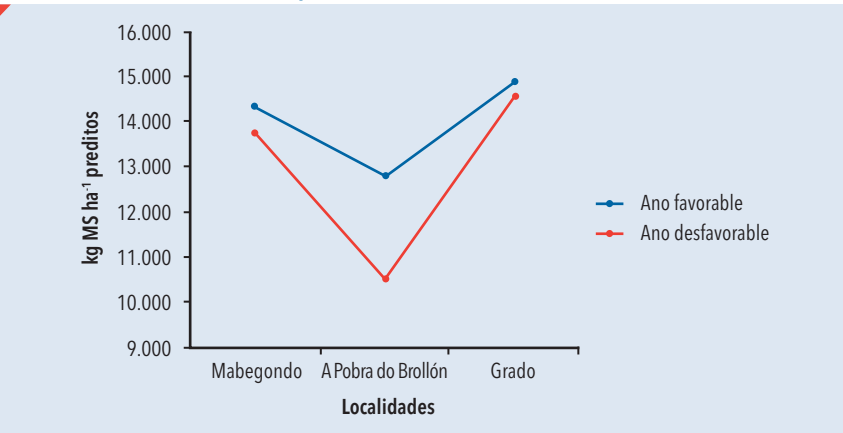
Co fin de poder facer predicións da produción de kg MS ha⁻¹ co mo-

Táboa 3. Análise de varianza dos factores ano, cultivar e localidade sobre a produción de MS ha⁻¹ predita

Fontes de variación	Graos de liberdade	Cadrados medios	F
Ano	1	252.206.272,6	230,2***
Cultivar	4	2.289.726,9	2,1NS
Localidade	2	774.782.161,4	707,1***
Ano x cultivar	4	1.163.423,0	1,1NS
Ano x localidade	2	88.756.329,3	81,0***
Cultivar x localidade	8	2.519.549,2	2,3*
Ano x cultivar x localidade	8	2.549.141,6	2,3*
Erro	870	1.095.754,8	
Total	900		

NS = non significativo (p ≥ 0,05), * = significativo a p < 0,05, *** = significativo a p < 0,001, F = F-test
A interacción tripla entre ano, cultivar e localidade foi estatisticamente significativa, F (8, 870) = 2,3, p < 0,05

Figura 4. Medias das producións preditas dos cultivares dos raigrases en cada unha das localidades de avaliación para os anos favorables e desfavorables



delo obtido, xeráronse un conxunto de 30 datos para as variables Tmin e precipitación para cada localidade e cultivar mediante unha distribución uniforme que xera datos de maneira aleatoria dentro dun intervalo definido. Creáronse aleatoriamente 30 datos para un ano favorable considerando temperaturas mínimas altas e precipitacións altas (entre o valor medio e o valor máximo nos anos de avaliación). Para un ano desfavorable repítese o proceso con temperaturas mínimas baixas e precipitacións baixas (entre o valor medio e o valor mínimo dos anos de avaliación). No caso das variables temperatura máxima e radiación diaria, tivéronse os valores medios obtidos nos anos de avaliación.

Posteriormente, realizouse unha análise de varianza cos factores ano, cultivar e localidade considerados como factores fixos e as súas interaccións para ver se había diferenzas

significativas nos valores preditos de produción en kg MS ha⁻¹ (táboa 3).

Houbo unha interacción dobre estatisticamente significativa entre ano e localidade, F(2,870) = 81,0, p<0,001) e entre cultivar e localidade, F(8, 870) = 2,3, p <0,05, pero non entre ano e cultivar, F(4,870) = 1,1, NS. Estas interaccións indican que as localidades influíron de maneira diferente segundo os anos e os cultivares.

Cando se observan interaccións dobres significativas (por ex., ano x localidade), é o efecto desta interacción o que é de interese estudar, debido a que os efectos dos factores principais (ano, localidade e cultivar) teñen menos interese, pois neste caso o efecto da localidade varía segundo o ano.

Houbo un efecto estatisticamente significativo para os factores ano, F(1,870) = 230,2, p < 0,001, e localidade, F(2,870) = 707,1, p < 0,001, pero non para o factor cultivar, F(4,870) = 2,1, NS. ▶▶

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:



► AÍNDA QUE A ANÁLISE DE VARIANZA NON ATOPOU DIFERENZAS SIGNIFICATIVAS ENTRE CULTIVARES EN CANTO Á PRODUCCIÓN PREDITA, O MODELO GRADIENT BOOSTING REGRESSOR MOSTROU QUE A VARIABLE "CULTIVAR" FOI A TERCEIRA MÁIS INFLUENTE PARA AS PREDICIÓNS



Os valores de produción de kg MS ha⁻¹ preditos considerando os valores de todos os cultivares en conxunto e a interacción ano x localidade represéntanse na figura 4.

Esta interacción de tipo cuantitativo ou de escala (Allard e Bradshaw, 1964) indica que o valor da produción varía entre localidades, sendo consistentemente maior no ano favorable que no desfavorable para todas elas. En particular, a localidade da Pobra presenta a maior diferenza de produción entre ambos os anos, mentres que Grado destaca por alcanzar os valores máis altos en ambas as condicións.

Doutra banda, aínda que a análise de varianza non atopou diferenzas significativas entre cultivares en canto á produción predita, o modelo Gradient Boosting Regressor mostrou que a variable "Cultivar" foi a terceira máis influente para as predicións. Esta discrepancia pode deberse a que a análise de varianza avalía principalmente efectos lineais e aditivos, mentres que o modelo de aprendizaxe automática pode capturar relacións non lineais e complexas entre as variables de entrada e as variables de saída.

CONCLUSIÓN

O modelo de aprendizaxe automática obtido ten o uso potencial de ser unha ferramenta para predicir a produción dos raigrases tendo en conta a localidade de cultivo e valores probables da temperatura media de mínimas e a precipitación anual.

Aínda que os resultados do modelo son prometedores, sería necesario melloralo cun maior número de datos e novas variables descritoras. ■

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.

AMPLA GAMA DE PRODUTOS PARA UN **MELLOR RENDEMENTO DA TÚA PRADEIRA**

GALICAL

CALES E CARBONATOS AGRÍCOLAS

Para reducir o efecto limitante do pH e controlar o aluminio nas terras de cultivo **é recomendable aplicar emendas calcarias ou magnésicas.**

ESTENDIDO REGULADO POR GPS



• **EMENDA CALCARIA, CAL VIVO GRANULADO** (90 % CaO)
Alta porcentaxe en calcio. Valor neutralizante: 90 %

• **EMENDA CALCARIA, CAL VIVO GRANULADO DOLOMÍTICO**
(35 % MgO / 60 % CaO)
Alta porcentaxe de magnesio. Valor neutralizante: 95 %

• **EMENDA CALCARIA, CAL VIVO** (80 % CaO)
Gran poder de neutralización. Valor neutralizante: 80 %

• **EMENDA CALCARIA, CAL APAGADO** (65 % CaO)
Potencia o rendemento agrícola. De fácil asimilación.
Valor neutralizante: 65 %

• **EMENDA CALCARIA, CAL APAGADO + DOLOMÍA**
(53 % CaO / 23 % MgO)
Achega magnesio. Favorece a actividade clorofílica da planta.
Valor neutralizante: 83 %

• **EMENDA CALCARIA, CARBONATO CÁLCICO** (56 % CaO)
Para terra e camas hixiénicas. Eficaz na redución de mamites ambientais e dermatites. Apropiado para a produción de todo tipo de pensos. Valor neutralizante: 56 %

• **EMENDA CALCARIA DE CARBONATO CÁLCICO MAGNÉSICO, GALIMAG** (33 % CaO / 17 % MgO)
Achega magnesio. Valor neutralizante: 60 %

• **EMENDA CALCARIA, GRANICAL G60 SUPRA** (54 % CaO / 1 % MgO)
Fácil aplicación e asimilación. Valor neutralizante: 56 %

• **EMENDA CALCARIA, GRANICAL G60 PLUS** (47,35 % CaO / 6 % MgO)
Fácil aplicación e asimilación con achega de magnesio.
Valor neutralizante: 59 %



EN VÍDEO

Transporte a calquera punto de España e Portugal

ENCALADO AGRÍCOLA · ENVASADO DE CALES E DOLOMÍAS · ALIMENTACIÓN ANIMAL · CAMAS DE VACÚN



GALICAL CALES E CARBONATOS AGRÍCOLAS SL

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

☎ 982 221 484

✉ info@galical.es



www.galical.es