


Estimación de masa de forraje en una pradera mixta por aprendizaje automatizado, datos del manejo de la pradera y meteorológicos satelitales



Aurelio Guevara-Escobar ^a

Mónica Cervantes-Jiménez ^{a*}

Vicente Lemus-Ramírez ^b

Adolfo Kunio Yabuta-Osorio ^b

José Guadalupe García-Muñiz ^c

^a Universidad Autónoma de Querétaro. Facultad de Ciencias Naturales. 76230 Juriquilla, Santiago de Querétaro, Querétaro, México.

^b Universidad Nacional Autónoma de México. Facultad de Medicina Veterinaria y Zootecnia, Centro de Enseñanza, Investigación y Extensión en Producción Animal en Altiplano CEIEPAA. Querétaro, México.

^c Universidad Autónoma Chapingo. Departamento de Zootecnia, Posgrado en Producción Animal. Estado de México, México.

* Autor de correspondencia: monica.cervantes@uaq.mx

Resumen:

Medir la masa de forraje (MF) en la pradera, antes del pastoreo, es fundamental para determinar la asignación diaria de forraje en sistemas pastoriles de producción animal. La MF se estima por corte de forraje en áreas conocidas, utilizando ecuaciones alométricas, o con el uso de sensores de percepción remota (PR); sin embargo, la exactitud y practicidad de los distintos métodos para estimar la MF, es variable. El objetivo fue obtener modelos predictivos usando variables ambientales y del manejo de la pradera para predecir la MF. Se ajustaron modelos de regresión para estimar la MF con base en variables del manejo de la pradera (MP) o mediciones obtenidas por PR, como reflectancia, temperatura del aire y lluvia. Por tres años se estudió una pradera mixta pastoreada con bovinos productores de carne. Con 80 % de datos se modeló por mínimos

cuadrados ordinarios (OLS) o por algoritmos de aprendizaje automatizado (ML). El 20 % restante de los datos se utilizó para validar los modelos usando el coeficiente de determinación y el sesgo promedio entre valores estimados y observados. El modelo base de estudio fue la relación entre la altura de la pradera antes del pastoreo y la MF de este modelo se ajustó usando OLS; la r^2 fue 0.43. Cuando se ajustaron modelos que incluyeron variables del MP, la r^2 fue 0.45 para OLS y 0.63 para ML. Al ajustar modelos con variables de MP y PR, la r^2 fue 0.71 para OLS y 0.96 para ML. Los ensambles de modelos ajustados con ML redujeron el sesgo de estimados de MF de la pradera examinada. En general, los modelos de ML representaron mejor la relación entre altura de la pradera antes del pastoreo y MF que los de modelos de OLS, al ajustarlos con variables de manejo de la pradera y con información de PR. Los modelos de ML pueden usarse como herramienta para la toma de decisiones diaria en sistemas productivos pastoriles.

Palabras clave: Alfalfa, Forraje, Lluvia, Temperatura, Sensores remotos.

Recibido: 08/03/2022

Aceptado: 18/07/2022

Introducción

La producción animal en pastoreo depende de la tasa de acumulación de la masa de forraje (MF), así como de la asignación oportuna de una carga animal adecuada para aprovechar la MF; otros aspectos importantes son la calidad nutricia y la estacionalidad en la tasa de acumulación de la MF. El manejo rentable de una pradera a través del pastoreo directo implica, entre otras cosas, implementar un manejo del pastoreo sin comprometer el rebrote de la cubierta vegetal, así como conocer de forma precisa la MF en la pradera antes y después del pastoreo⁽¹⁾. Tradicionalmente, la MF se mide directamente con cortes de forraje en cuadrantes de área conocida, distribuidos de manera espacialmente representativa y en un número suficiente que represente la variabilidad de la cubierta vegetal en la pradera^(2,3). El corte de cuadrantes es laborioso y por eso se han desarrollado métodos y dispositivos para la estimación indirecta de la MF⁽⁴⁻⁶⁾. La altura del dosel de la pradera, medida con una regla graduada (sward stick) es útil para representar la MF, aunque la relación puede ser diferente en función de la composición botánica, densidad del dosel de la pradera y estación del año⁽⁷⁻⁹⁾. La altura del forraje comprimido medida con un plato de aluminio (rising plate meter), estima la MF considerando la densidad del dosel y es una práctica muy común a nivel de granja en países como Nueva Zelanda⁽²⁾. La relación entre altura del dosel y MF en praderas de pasto ballico y trébol blanco es bien conocida y de aplicación rutinaria en Nueva Zelanda⁽¹⁰⁾; para praderas con otras especies forrajeras como la alfalfa, se necesita más investigación para determinar la relación entre altura del dosel y MF⁽⁸⁾.

La percepción remota por satélites orbitales (PR) mide la reflectancia espectral, la proporción de la energía incidente reflejada por la superficie terrestre en distintas longitudes de onda; estas mediciones se han asociado a los procesos de actividad de la vegetación⁽¹¹⁾. Con la información de PR también es posible estimar variables ambientales como la temperatura, la precipitación pluvial, y otras⁽¹²⁾. La amplia disponibilidad y acceso libre de productos de PR es una oportunidad para explorar la dinámica de los cultivos y establecer relaciones con parámetros productivos, como la MF. Las series de tiempo disponibles para distintos productos de PR permiten hacer estudios retrospectivos, lo cual es valioso para evaluar prácticas de manejo de praderas y estudios regionales de pastizales. Sin embargo, la escala espacial de medición es gruesa en algunos sensores PR y es una desventaja importante en estos estudios.

Recientemente se han incorporado al análisis de regresión una variedad de algoritmos de aprendizaje automatizado o machine learning (ML) y son una alternativa a la regresión por mínimos cuadrados ordinarios (OLS). La fotosíntesis en ecosistemas, nombrada productividad primaria bruta y la productividad primaria neta (descontando las pérdidas por respiración) se han modelado con enfoques empíricos o mecanísticos, desde modelos OLS hasta aquellos que simulan los procesos ecofisiológicos a nivel global a partir de PR⁽¹³⁾. La productividad primaria neta incluye la partición fotosintética hacia la biomasa aérea y de raíces y por eso no refleja la MF disponible para el pastoreo. Lang *et al*⁽¹⁴⁾ estimaron la producción del pastizal árido usando mediciones de sensores PR de lluvia, reflectancia espectral obtenida del satélite Landsat 7 y un algoritmo ML con random forest. Utilizando Redes Neuronales, otro algoritmo de tipo ML, Chen *et al*⁽¹⁵⁾ relacionaron la reflectancia espectral medida por el satélite Sentinel-2 y la MF en granjas lecheras de Tasmania en Australia. En estos estudios el coeficiente de determinación (r^2) en distintos modelos fue entre 0.6 y 0.7. Conceptualmente, es importante incorporar las condiciones de humedad, a corto o mediano plazo para explicar la capacidad de carga del pastizal⁽¹⁶⁾, ya que el agua es el principal recurso limitante de las plantas en los ambientes áridos y semiáridos. Las condiciones de disponibilidad de agua para las plantas se pueden representar por la precipitación pluvial ocurrida (P), el agua disponible en el suelo o el déficit de vapor en la atmósfera. Sin embargo, para explicar la MF no sólo es importante la P ocurrida en el periodo de acumulación de la MF (mes en que se midió la MF), sino las condiciones de humedad ocurridas en meses precedentes.

En el presente trabajo se examinó la relación entre la MF y la altura del forraje como un punto de partida para comparar otros modelos que usen variables meteorológicas obtenidas por PR o en conjunto con variables representativas de las condiciones de manejo de la pradera (MP); como son los periodos de descanso y ocupación del área de pastoreo o la misma altura del forraje. En particular, se exploró la utilidad de los modelos para predecir la MF en función de condiciones precedentes de lluvia y temperatura en distintas ventanas temporales; por ejemplo, la P acumulada en el mes anterior, en dos meses o tres meses antes de la medición de la MF. El objetivo fue obtener un modelo predictivo de la MF que pueda incorporarse en la planeación del pastoreo. Con este

propósito se usaron tres años de mediciones de una pradera mixta de alfalfa y pasto, pastoreada con ganado productor de carne.

Material y métodos

Sitio

El trabajo se efectuó en el Centro de Enseñanza, Investigación y Extensión en Producción Animal en el Altiplano, a cargo de la Facultad de Medicina Veterinaria y Zootecnia de la Universidad Nacional Autónoma de México. El sitio se ubica a 20° 36' 13.88" N, 99° 55' 02.91" O y altitud de 1,913 msnm. El clima es seco extremoso tipo Ganges sin canícula, BS₁ 0w(e)g, de acuerdo con los registros climatológicos históricos (1951 a 2006) de la estación meteorológica 22025; la más cercana al sitio donde los promedios anuales de precipitación y temperatura son 458 mm y 23.5 °C⁽¹⁷⁾.

La pradera se estableció en 2004 con una mezcla de 50 % alfalfa (*Medicago sativa*) y gramíneas como pasto ovinillo (*Dactylis glomerata*), festuca alta (*Festuca arundinaceae*) y ballico perenne (*Lolium perenne*). La superficie de pastoreo fue de 19 ha dividida en 16 secciones de igual dimensión y delimitadas a través de cerco eléctrico móvil. Con riego por aspersión bajo la modalidad de pivote central se regó la pradera; no se contó con registros de la lámina o de calendario de riego. El grupo de pastoreo se conformó por 88 vientres de la raza Limousin y sus crías. El tiempo de ocupación en cada división se estableció con base en la estimación de la MF, el análisis químico proximal de muestras de MF y el requerimiento de materia seca (MS) del grupo de pastoreo en turno. El manejo reproductivo fue principalmente con inseminación artificial y pariciones distribuidas a lo largo del año.

Datos

De 2008 a 2010 se obtuvieron 399 observaciones de MF antes del pastoreo del área asignada. Cada observación de MF correspondió al inicio de un ciclo de pastoreo del hato. Las observaciones se consideraron unidades experimentales y cada una consistió de ocho mediciones aleatorias obtenidas con la técnica del marco metálico modificada, para proteger el rebrote de la alfalfa se cortó el forraje a 10 cm de altura en un área de 0.25 m²⁽¹⁸⁾. Las muestras de forraje se deshidrataron en una estufa de aire forzado durante 48 h para determinar el contenido de MS y el dato se expresó en kg MS ha⁻¹. En cada ciclo de pastoreo, se registró la altura de la pradera (A_pradera), la fecha de pastoreo (Dia_pastoreo y Mes_pastoreo), el tiempo de ocupación (T_ocupación), días de descanso del área de pastoreo desde el pastoreo anterior (D_descanso), mes del inicio del crecimiento en el ciclo anterior (Mes_ini_crec) y tasa de acumulación de MS promedio mensual (TAF, kg MS ha⁻¹ d⁻¹). Estas variables se denominaron en conjunto como variables de manejo de la pradera (MP).

Usando la Application for Extracting and Exploring Analysis Ready Samples del Land Processes Distributed Active Archive Center de la National Aeronautics and Space Administration (NASA), se solicitó el producto MCD43A4 versión 6⁽¹⁹⁾. El producto MCD43A4 se genera a partir de las mediciones que efectúan los sensores del Moderate-Resolution Imaging Spectroradiometer (MODIS) a una resolución espacial de 500 m². Este producto consiste en siete bandas de reflectancia ajustada por la Bidirectional Reflectance Distribution Function y producido diariamente, los cuales son un promedio móvil de los 16 días contiguos. Se descargaron los datos de ocho pixeles contiguos correspondientes al polígono de coordenadas: 99.93 O, 20.60 N a 99.92 O, 20.61 N. El espectro de radiación (nm) cubierto por las bandas uno a la siete es (b1-b7): 620-670, 841-876, 459-479, 545-565, 1230-1250, 1628-1652 y 2105-2155. Los datos de P fueron del producto 3IMERG versión 6 de la misión Global Precipitation Measurement de la NASA obtenidos a través del portal Giovanni (<https://giovanni.gsfc.nasa.gov/giovanni>). El dato P (mm) fue el acumulado mensual para la coordenada 99.92 O, 20.60 N; la resolución espacial del 3IMERG es de 10 km². A través del portal Giovanni también se obtuvo el producto MODIS MOD11A2 versión 6 de temperatura de superficie diaria durante el día (LST_d) y la noche (LST_n).

Para MODIS se determinó la buena calidad de acuerdo con los datos de calidad adjuntos a los productos respectivos. En el lenguaje R⁽²⁰⁾, se generó un código para encontrar las fechas de medición del MCD43A4 más cercanas a la fecha de medición de la MF. Usando Qgis v3.16.4⁽²¹⁾ y una imagen satelital de Google Maps⁽²²⁾ como plantilla guía, se determinó una capa vectorial correspondiente área de riego por pivote central; el círculo comprendió diferente área de los pixeles muestreados del MCD43A4. Para cada banda de reflectancia se obtuvo el promedio correspondiente al vector usando la función extract del paquete raster.

Generación de variables

La reflectancia en las bandas b2 y b1 se asocian a la capacidad de la vegetación para absorber luz fotosintéticamente activa y existen distintos índices para representar dicha actividad de la vegetación. Se calculó el índice normalizado de la vegetación (NDVI) y el índice mejorado de la vegetación (EVI) usando las bandas espectrales del producto MCD43A4:

$$NDVI = \frac{b2-b1}{b2+b1} \quad 1)$$

$$EVI = 2.5 \frac{(b2-b1)}{(b2+2.4b1+1)} \quad 2)$$

Con la serie de tiempo de P se calcularon las siguientes variables: la P acumulada en el mes anterior (P_lag_1), la P acumulada en los dos meses anteriores (P_lag_2) y así sucesivamente hasta la P acumulada en seis meses anteriores: (P_lag_3, P_lag_4, P_lag_5 y P_lag_6). Para LST_d y LST_n se calculó el promedio del mes anterior (LST_x_avg_1), de los dos meses anteriores (LST_x_avg_2) o de los tres meses

anteriores (LST_x_avg_3), donde x representa el indicativo d o n, para día o noche. Estas variables representaron el ambiente prevalente antes de medir la MF.

Modelación

El modelo base para comparación fue la regresión lineal simple entre MF y A_pradera. Se exploraron cuatro escenarios de modelación según el tipo de algoritmo: ML u OLS y el tipo de variables disponibles para la modelación: sólo usando variables explicativas de origen PR (ML_PR y OLS_PR) o variables PR y además las de MP (ML_PR_MP y OLS_PR_MP). Los modelos se entrenaron con el 80 % de las observaciones elegidas al azar y se reservó el 20 % para su evaluación. La evaluación del modelo es un concepto de caja negra sobre la relevancia del resultado del modelo⁽²³⁾. Los procedimientos estadísticos se efectuaron en el lenguaje R, el nombre de los paquetes se indica donde es pertinente. Se ajustó un modelo de regresión ortogonal (major axis regression) entre valores observados y valores predichos usando el paquete smatr 3, dado que los valores observados de MF se miden con error⁽²⁴⁾. Se calculó el coeficiente de determinación (r^2), la raíz del cuadrado medio del error (RMSE), los criterios de información de Akaike (AIC) y bayesiano (BIC), la devianza y el sesgo. Estos indicadores cuantitativos, así como la evaluación gráfica son técnicas comúnmente usadas para evaluar modelos matemáticos con fines predictivos⁽²⁵⁾.

En el caso de OLS, se usó el valor de inflación de varianza (VIF) para identificar multicolinealidad utilizando las funciones stepAIC y vif⁽²⁶⁾; 10.0 fue el valor máximo permitido de VIF para retener variables en el modelo de regresión múltiple OLS. El nivel de significancia se fijó en 0.05 para los análisis paramétricos y del cumplimiento de sus supuestos estadísticos de la regresión OLS.

Se generó el modelo ML con la función h2o.automl del paquete H2O⁽²⁷⁾, éste produce un conjunto de modelos con diferentes realizaciones de algoritmos: deep learning (DL), feedforward artificial neural network (NN), general linear models (GLMs), gradient-boosting machine (GBM), extreme gradient-boosting (XGBoost), default distributed random forest (DRF) y extremely randomized trees (XRT). Cada modelo individual se puede usar para predecir la respuesta, pero también para generar dos tipos de ensamble de modelos: uno es a partir de todos los algoritmos usados en los modelos generados, y el segundo tipo de ensamble sólo considera los mejores modelos de cada clase o familia de algoritmos; ambos tipos de ensamble generalmente producen mejores predicciones que los modelos individuales⁽²³⁾.

La función h2o.automl se ejecutó veinte veces con los siguientes parámetros: a) max_runtime_secs = 500 el tiempo máximo de ejecución antes de entrenar un ensamble final de modelos, b) nfolds = 15, número de pliegues para evaluación cruzada (k-folds), c) seed = un valor entero aleatorio con valor entre 1 y 50; cada una de las ejecuciones usó un valor semilla elegido al azar, d) nthreads = 50, el número de hilos de procesamiento disponibles, e) max_mem_size = 100GB, la memoria RAM disponible en Gigabytes. El

tiempo aproximado de ejecución fue de 50 min en un equipo con doble procesador Xeon 2680 v4 de 14 núcleos y doble hilo cada uno y 128 GB de memoria RAM.

Con la función `h2o.explain`, se obtuvo la importancia de las variables en los modelos ML individuales y figuras de dependencia⁽²⁷⁾. Se utilizó la devianza como estadístico de bondad de ajuste para ordenar los modelos generados. El aprendizaje automatizado tiene dos elementos para el aprendizaje supervisado: la pérdida de entrenamiento y la regularización. La tarea de entrenamiento intenta encontrar los mejores parámetros para el modelo mientras minimiza la función de pérdida de entrenamiento; esta función es el cuadrado medio del error u otras. El término de regularización controla la complejidad del modelo, ayudando a reducir el sobreajuste. El sobreajuste se hace evidente cuando el modelo funciona con precisión durante el entrenamiento, pero la precisión disminuye durante la evaluación del modelo. Un buen modelo necesita un ajuste extenso de parámetros ejecutando el algoritmo varias veces para explorar el efecto en la regularización y la precisión de la evaluación cruzada⁽²⁸⁾. En esta investigación, la función de la pérdida de entrenamiento fue la devianza, que es una generalización de verosimilitud de la suma de cuadrados del error; los valores más bajos o negativos indican un mejor desempeño del modelo⁽²⁹⁾.

Resultados y discusión

El promedio de MF de la pradera fue de 2,134 kg MS ha⁻¹ con un patrón estacional de menor producción en invierno y mayor en verano (Figura 1a). La MF fue diferente entre los tres años 2,121, 1,770 y 2,392 kg ha⁻¹ para 2008 a 2010 ($P < 0.05$). La lluvia fue de 636, 382 y 552 mm, respectivamente. La mayor cantidad de lluvia fue de julio a septiembre; para el año 2010 fue atípico el mes de febrero con 151 mm (Figura 1b) y posiblemente impactando positivamente la MF a partir de marzo en ese año. La lluvia registrada por el producto IMERG en 2008 y 2010 fue superior a la registrada por la estación climatológica más cercana al sitio del estudio; este estimado de lluvia se consideró acertado porque este producto ha mostrado buena concordancia con registros terrestres de precipitación⁽³⁰⁾. El comportamiento estacional de la MF sugirió un efecto importante de la lluvia, aun tratándose de una pradera irrigada. Abril y mayo fueron los meses con mayor LST_d promedio (Figura 1c). La diferencia entre la LST_d y LST_n fue mayor en los meses de abril a mayo (28.5 y 27.3 °C) y menor en julio a septiembre (17.3, 16.4 y 15.6 °C); lo cual indica la característica extrema del clima durante la primavera en el sitio de estudio. Estas condiciones ambientales también se reflejaron en cambios estacionales en el manejo de la pradera en los días de descanso, la altura del forraje y la TAF (Figura 2).

Figura 1: Variables ambientales y producción de masa de forraje de pradera mixta de alfalfa y pastos templados pastoreada con ganado de carne: a) masa de forraje (MF), b) lluvia (P) y c) temperatura de superficie (LST) diurna (●) y nocturna (○)

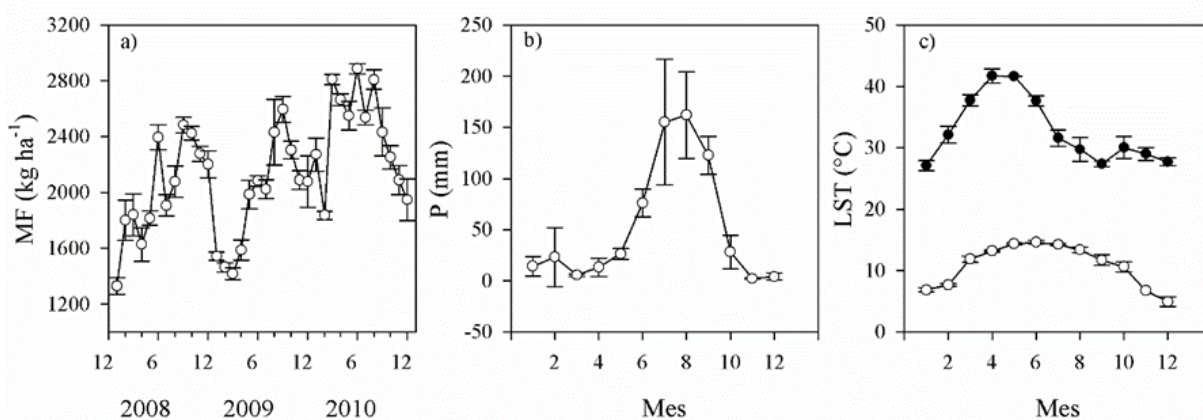
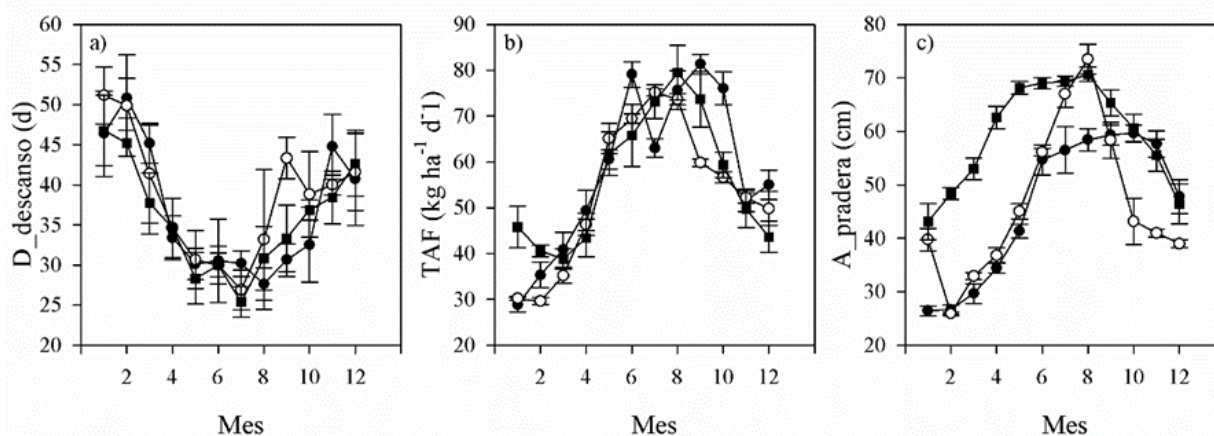


Figura 2: Manejo de pradera mixta de alfalfa y pastos templados pastoreada con ganado de carne durante 2008 (●), 2009 (○) y 2010 (■): a) Días de descanso antes del pastoreo, b) tasa de acumulación de forraje (TAF) del periodo, c) altura del forraje



En la regresión MA el intercepto fue numéricamente cercano a 0 en el escenario ML_PR_MP y su pendiente fue igual a 1, un modelo con pendiente igual a 1 e intercepto igual a 0 indica buen ajuste. El valor menor del RMSE, AIC, BIC y devianza sugirieron una mejor representación de la MF con el escenario ML_PR_MP (Cuadro 1). En el análisis de devianza la comparación entre dos o más modelos será válida si se ajustan al mismo conjunto de datos, este requisito no se cumplió porque los valores predichos de MF fueron inherentemente distintos para cada modelo generado. La diferencia de devianzas se distribuye aproximadamente como X^2 con grados de libertad iguales a la diferencia en el número de parámetros entre los modelos⁽¹⁴⁾, siendo esta diferencia 0 para el caso de los modelos de regresión lineal simple usados para representar la relación entre valores estimados y predichos en cada escenario de modelación. Por estas dos razones la elección del mejor modelo se basó únicamente en el valor numérico de las medidas de

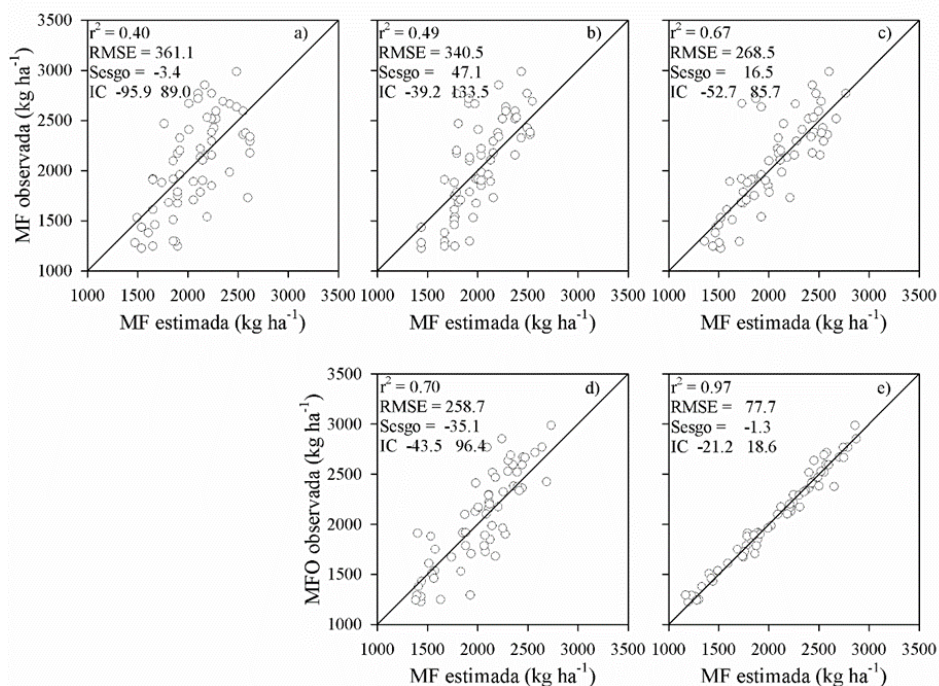
bondad de ajuste. El peor modelo fue la regresión lineal entre MF y A_pradera, no solo en función de las medias de bondad de ajuste sino también en la representación gráfica de los valores estimados vs. observados (Figura 3).

Cuadro 1: Medidas de bondad de ajuste entre valores observados y estimados de MF resultantes de escenarios de modelación usando algoritmos de mínimos cuadrados ordinarios (OLS) o aprendizaje automatizado (ML) en combinación con variables explicativas referentes al manejo de la pradera (MP) únicamente o en conjunto con variables de sensores remotos (MP_PR)

	OLS_A_pradera	OLS_PR	OLS_PR_PM	ML_PR	ML_PR_PM
r ²	0.40	0.49	0.67	0.70	0.97
RMSE	361.0	341.0	269.0	259.0	78.0
AIC	734.0	724.0	686.0	691.0	542.0
BIC	738.0	728.0	690.0	695.0	546.0
Devianza	8079684.0	6874194.0	4377078.0	4003784.0	363954.0
Sesgo	-3.4	47.1	16.5	-35.1	-1.3
IC 2.5 %	-95.9	-39.2	-52.7	-43.5	-21.2
IC 97.5 %	89.0	133.5	85.7	96.4	18.6
MA intercepto	-1799.0	-2044.0	-594.0	-735.0	27.0
IC 2.5 %	-3386.0	-3395.0	-1137.0	-1257.0	62.0
IC 97.5 %	-831.0	-1162.0	-162.0	-316.0	112.0
MA pendiente	1.9	2.0	1.3	1.4	1.0
IC 2.5 %	1.4	1.6	1.1	1.2	0.9
IC 97.5 %	2.6	2.7	1.6	1.6	1.0

r²= coeficiente de determinación; RMSE= raíz del cuadrado medio del error; AIC= criterio de información de Akaike; BIC= criterio de información de Bayesiano; MA= regresión ortogonal (major axis regression); IC= intervalo de confianza.

Figura 3: Evaluación entre valores observados y estimados de MF usando algoritmos de mínimos cuadrados ordinarios (OLS) o aprendizaje automatizado (ML)



a) OLS, variable predictora altura del forraje; b) escenario OLS_PR; c) escenario OLS_PR_MP; d) escenario ML_PR; e) escenario ML_PR_MP. Coeficiente de determinación (r^2), raíz del cuadrado medio del error (RMSE), sesgo y su intervalo de confianza (IC) al 95%.

Las variables TAF y A_pradera de MP fueron las más importantes (Cuadro 2), tanto en los modelos ML como OLS; la variable D_descanso tuvo mucho menor importancia (Cuadro 2). Las variables PR más importantes fueron: LST_n, P, P_lag_3 o P_lag_5, LST_d_avg_3 o LST_n_avg_3; indicando la relevancia de las condiciones ambientales de precipitación y temperatura no sólo del mes en curso, sino de las condiciones precedentes a la medición de la MF. La reflectancia (b1 – b7) y los índices de la vegetación se incorporaron en modelos ML, pero el procedimiento stepwise no los eligió para el OLS. Comparadas con TAF y A_pradera, las variables de reflectancia fueron de baja importancia en los escenarios PR_MP del ML. Las bandas espectrales de reflectancia fueron más importantes que el EVI y NDVI; este hallazgo coincide con el estudio de MF para praderas mixtas de clima templado⁽¹⁵⁾. Aunque se consideró adecuada la predicción de la biomasa fresca en praderas de *Brachiaria* con base en el NDVI con $r^2 = 0.73$ ⁽³¹⁾.

Cuadro 2: Variables importantes incluidas en los escenarios usando dos posibles algoritmos: mínimos cuadrados ordinarios (OLS) o aprendizaje automatizado (ML) y dos tipos de variable explicativa: sólo sensores remotos (PR) o variables del manejo de la pradera y las PR (PR_MP)

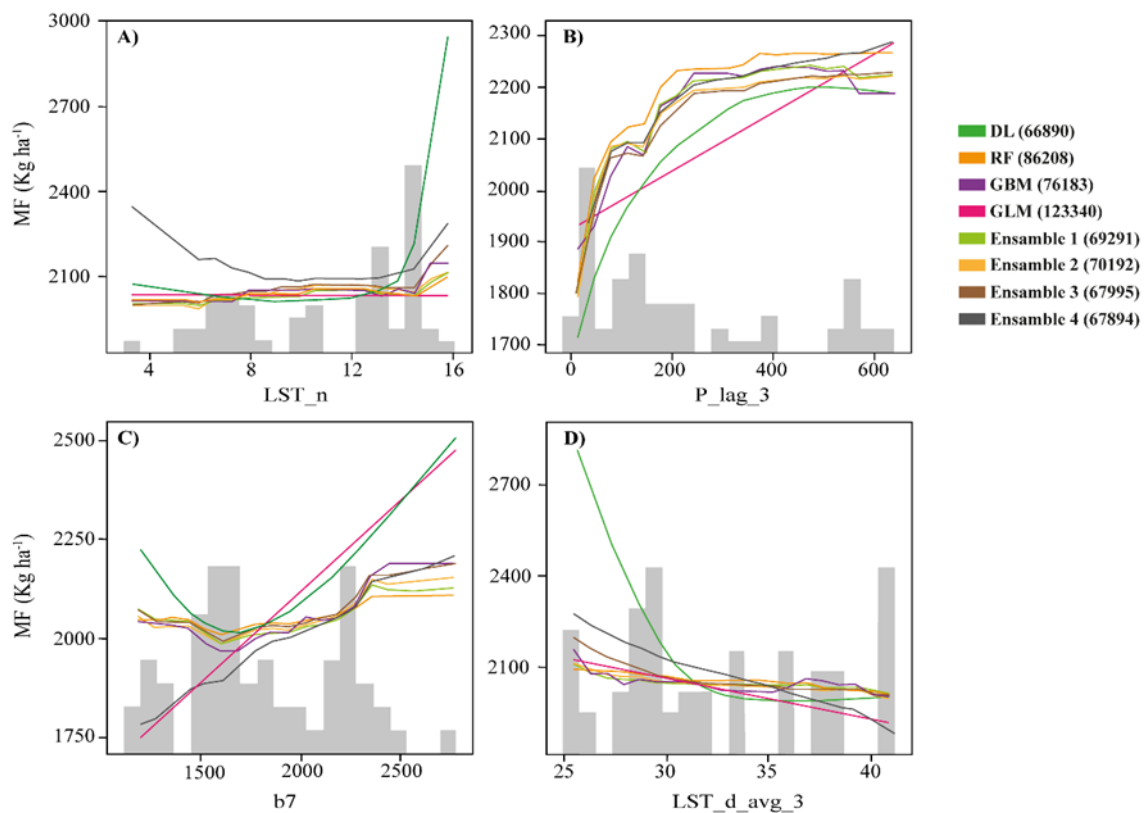
Variable	Aprendizaje automatizado (ML)		Mínimos cuadrados ordinarios (OLS)	
	Sensores remotos (PR)	Sensores remotos (PR)_Manejo de la pradera (MP)	PR	PR_MP
LST_d_avg_3	0.081	0.023	0.036	0.027
LST_n_avg_3	0.064	0.017		
LST_d	0.036	0.036		
LST_n	0.161	0.007	0.060	
b1	0.027	0.008		
b2	0.034	0.014		
b3	0.028	0.003		
b4	0.033	0.004		
b5	0.044	0.008		
b6	0.048	0.010		
b7	0.096	0.008		
P	0.058	0.008	0.048	
P_lag_3	0.099	0.023		
P_lag_5			0.270	
NDVI		0.001		
EVI	0.018	0.001		
A_pradera		0.303		0.231
Mes_ini_crec		0.006		0.020
D_descanso		0.101		0.072
TAF		0.417		0.368

Para los modelos ML la suma de la importancia es 1, para los modelos OLS la suma de la importancia es igual a la r^2 .

La dependencia parcial que existió entre la predicción de MF y el valor de algunas de las variables de mayor importancia en algunos modelos ML se muestra en la Figura 4, en el escenario ML_PR y en la Figura 5 para el ML_PR_MP. Los ensambles de modelos ML tuvieron menor devianza en comparación con algún algoritmo ML en los dos escenarios y por tanto se consideraron mejores representaciones de la MF. Las figuras de dependencia parcial indican cómo la variable explicativa influye en las predicciones de uno de los modelos o ensambles, después de estandarizar el efecto de otras variables. Para modelos de regresión lineal (como el modelo GLM obtenido por ML), la figura es una línea recta con pendiente igual al parámetro del modelo⁽³²⁾. La MF dependió directa y

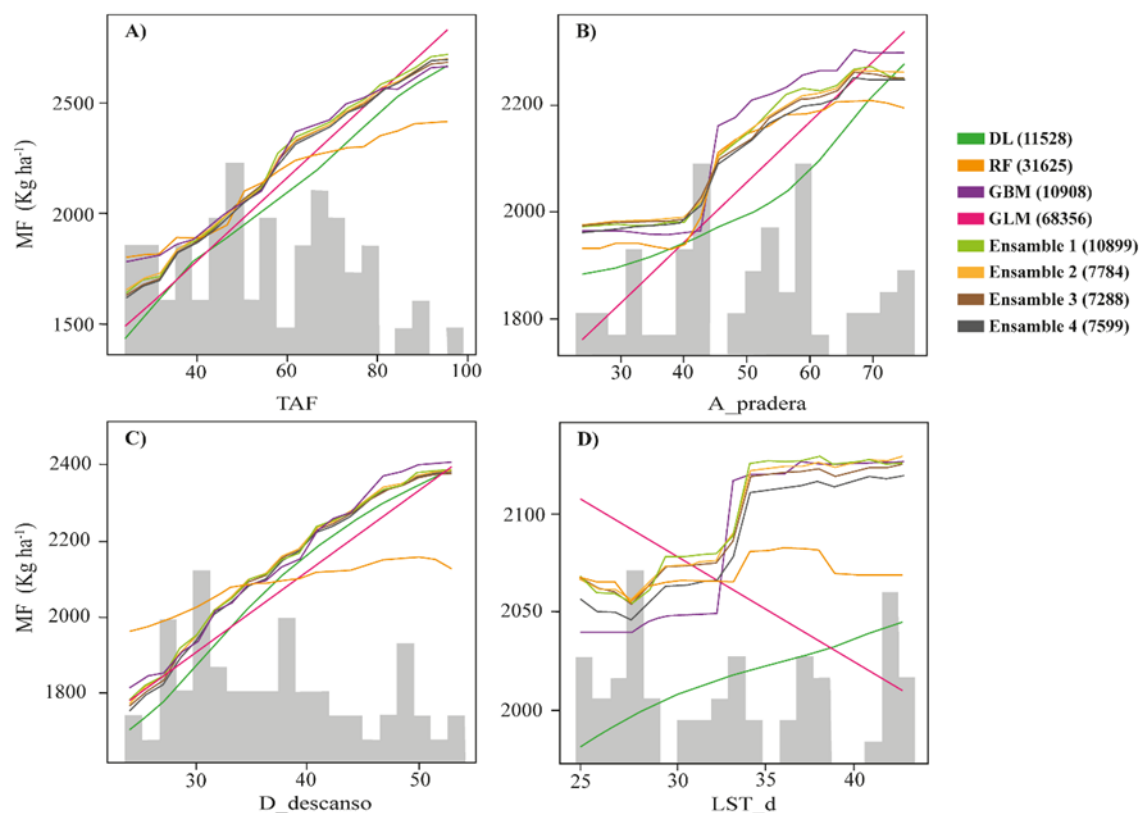
proporcionalmente de las variables TAF, A_pradera y D_descanso en diferentes modelos incluso para un modelo GLM (línea rosa), pero para las variables P_lag_3 y LST_d la dependencia difirió entre el modelo GLM y los modelos ML, particularmente el modelo de tipo DL (línea verde oscura) que fue el mejor modelo ML individual (Figura 5). La interpretación de las figuras se mejora con el histograma de frecuencia de las observaciones, en función del valor de la variable. Donde hubo menor frecuencia de datos se interpretó que la dependencia no fue soportada por suficiente evidencia. Un ejemplo de esta situación fue la dependencia de LST_n en la Figura 4 donde el modelo de tipo DL tiene un ascenso abrupto, pero los dos últimos intervalos de clase del histograma tienen pocas observaciones.

Figura 4: Dependencia parcial de la MF y: A) promedio mensual de la temperatura de superficie nocturna (LST_n), B) precipitación acumulada en los tres meses anteriores (P_lag_3), C) banda de reflectancia b7 del producto MCD43A4 de MODIS, D) promedio mensual de la temperatura de superficie diurna en los tres meses anteriores (LST_d_avg_3)



Las barras grises son la frecuencia de datos según intervalos de clase de la variable. Se presentan solo modelos de menor devianza (valor entre paréntesis) obtenidos por aprendizaje automatizado en el escenario usando solo variables medidas con sensores remotos (ML_PR).

Figura 5: Dependencia parcial de la MF y: A) tasa de acumulación de forraje (TAF), B) altura del forraje (A_pradera), C) días de descanso de la pradera (D_descanso), D) promedio mensual de la temperatura de superficie diurna (LST_d)



Las barras grises son la frecuencia de datos según intervalos de clase de la variable. Se presentan solo modelos de menor devianza (valor entre paréntesis) obtenidos por aprendizaje automatizado en el escenario usando variables medidas con sensores remotos y del manejo de la pradera (ML_PR_MP).

El escenario ML_PR_MP incluye la variable TAF y esto podría ser una limitante para la aplicación práctica del modelo. Para aclarar este aspecto se construyó un modelo ML sin esta variable y usando los mismos datos de entrenamiento, resultando en una r^2 de 0.76, RMSE de 232.2 y sesgo de -35.6 (IC -94.4 a 23.1), siendo mejor que el obtenido en el escenario ML_PR (datos no mostrados). Este resultado tiene dos aspectos de importancia: otras variables disponibles para la modelación pueden sustituir a una variable identificada como la más importante y segundo, es posible incurrir en una solución de óptimo local, aun cuando el algoritmo ML explora un espacio de solución con diferentes parámetros de optimización. Una alternativa posible sería aumentar el número de veces de ejecución de la función `h2o.automl` e incrementar el valor de la constante `max_runtime_secs`.

A pesar de la resolución espacial gruesa de los sensores remotos MODIS y GPM (250 m^2 y 10 km^2) se estimó la MF adecuadamente en el escenario ML_PR (Figura 3d), siendo que la $r^2 = 0.70$ de este modelo se encontró dentro del rango reportado recientemente en la literatura para modelos ML que estiman la biomasa con datos PR^(14,15) o la productividad primaria bruta⁽³³⁾. Un modelo con base en datos PR únicamente resulta

atractivo para el manejo de pastizales de gran área. Cuando las variables PR se usaron en combinación con variables del manejo de la pradera sencillas de medir (A_pradera) o registrar (D_descanso y Mes_ini_crec) la estimación fue muy buena (Figura 3e); la $r^2=0.97$ fue parecida a la estimación de biomasa con datos PR de resolución espacial de 30 m fue de $r^2=0.96$ ⁽⁵⁾. La predicción de la masa de forraje obtenida con mediciones de altura de forraje, mejoró cuando se incorporaron variables del manejo de la pradera y datos meteorológicos locales en un algoritmo ML de random forest ($r^2=0.82$), este enfoque se juzgó práctico para los productores, con excepción del costo de los instrumentos meteorológicos⁽²⁾.

A pesar de ser una pradera irrigada, la lluvia antecedente de corto plazo fue información importante para los modelos OLS y ML. En un estudio reciente identificaron que la variación espacio-temporal de productividad primaria bruta no sólo se explicó por bandas de reflectancia del MCD43A4 de MODIS, sino que también se relacionó con el déficit de presión de vapor⁽³³⁾. Similar al resultado de esos autores, aquí resultó útil incluir otras bandas de reflectancia además de la b1, b2 e índices de la vegetación como el NDVI y el EVI. Desde el punto de vista práctico, el modelo del escenario ML_PR_MP se consideró muy factible de implementar ya que usó mediciones rutinarias del manejo de la pradera y los datos de sensores remotos de la NASA que son de acceso público.

La producción animal en pastoreo es sostenible cuando se garantiza un consumo de alimento que satisfaga las necesidades de nutrientes. En el manejo del pastoreo esto depende de ajustar la carga animal en función de la etapa fenológica de la planta, de la MF antes y después del pastoreo y del forraje que se decida dejar como remanente en la pradera. Para el ganado bovino productor de carne, la MF antes del pastoreo adecuada, se puede establecer en $2,500 \text{ kg ha}^{-1}$ y la MF después del pastoreo alrededor de los $1,200 \text{ kg ha}^{-1}$ ⁽¹⁰⁾; aunque estos umbrales dependerán de la etapa reproductiva y fisiológica del animal, la época del año y diferentes estrategias de manejo de la pradera para el racionamiento del alimento, el control fenológico o el balance en la composición botánica⁽³⁴⁾. Por estas razones es importante que el modelo predictivo de MF ajuste bien en los extremos de su rango y con excepción del escenario ML_PR_MP, existió una sobreestimación de la MF cuando fue inferior a $1,500 \text{ kg ha}^{-1}$ aproximadamente (Figura 3).

La masa de forraje en la pradera tiene una variabilidad espacial dada por diferencias en humedad, fertilidad, deposición de excretas, alteraciones en la comunidad vegetal por el pastoreo selectivo y otros factores. El método de cortes de forraje en cuadrantes es limitado para representar y capturar esta variabilidad espacial en la pradera y por eso el método estadístico de muestreo es importante. Los sensores a bordo de vehículos aéreos no tripulados o drones son una alternativa para capturar la variabilidad en la reflectancia de la vegetación en la escala espacial de centímetros, pero debe considerarse el costo del equipo multispectral, el procesamiento de datos y la limitante operativa para cubrir el territorio⁽³⁵⁾, además de la necesidad de una función de calibración para la masa de forraje.

Conclusiones e implicaciones

La predicción de la MF tuvo menor sesgo con modelos de ML que con modelos OLS, sobre todo cuando se incorporaron en los modelos variables de sensores remotos y de manejo de la pradera. Los ensambles de ML tuvieron menor devianza que algunos de los modelos individuales de ML. El uso de variables PR predice la MF de manera semejante que la relación A_pradera vs. MF, aunque el modelo ML tuvo menor sesgo. Los modelos explorados tendrán que probarse en otras condiciones de pradera para tener una aplicación espacial, poder representar ecosistemas y valorar el servicio ambiental de captura de carbono. En la escala local de granja, estos modelos tendrán aplicación para el uso cotidiano en la planeación de la presupuestación forrajera o evaluación retrospectiva del manejo del pastoreo de la pradera. En estos casos el resultado presentado aquí tiene una aplicación promisoriosa.

Agradecimientos

El estudio fue producto del apoyo para estancia sabática del primer autor por parte de la Universidad Autónoma de Querétaro.

Literatura citada:

1. Sheath GW, Hay RJM, Giles KH. Managing pastures for grazing animals. In: Nicol, AM, editor. Livestock feeding on pasture. NZ Soc Anim Prod occasional publication. 1987;65–74.
2. Murphy D, O'Brien B, Hennessy D, Hurley M, Murphy M. Evaluation of the precision of the rising plate meter for measuring compressed sward height on heterogeneous grassland swards. *Precis Agric* 2021;22(3):922–946.
3. Radcliffe J. Cutting techniques for pasture yields on hill country. *Proc NZ Grassland Association*. 1971;33:91–104.
4. Jáuregui JM, Delbino FG, Bonvini MIB, Berhongaray G. Determining yield of forage crops using the Canopeo mobile phone app. *J NZ Grassl* 2019;41–46.
5. Marsett RC, Qi J, Heilman P, Biedenbender SH, Watson MC, Amer S, *et al*. Remote sensing for grassland management in the arid southwest. *Rangel Ecol Manag* 2006;59(5):530–540.
6. O'Donovan M, Dillon P, Rath M, Stakelum G. A comparison of four methods of herbage mass estimation. *Ir J Agric Food Res* 2002;17–27.
7. Mills A, Smith M, Moot DJ. Relationships between dry matter yield and height of rotationally grazed dryland lucerne. *J NZ Grassl* 2016;(78):185–196.

8. Moot DJ, Yang X, Ta HT, Brown HE, Teixeira EI, Sim RE, *et al.* Simplified methods for on-farm prediction of yield potential of grazed lucerne crops in New Zealand. *NZ J Agric Res* 2021;65(4-5):1–19.
9. Robertson S. Mass to height relationships in annual pastures and prediction of sheep growth rates. *Anim Prod Sci* 2014;54(9):1305–1310.
10. Nicol AM, Nicoll GB. Pastures for beef cattle. In: Nicol, AM. editor. *Feeding livestock on pasture*. Society of Animal Production. Lincoln, New Zealand. 1987;119–131.
11. Zhang Y, Ye A. Would the obtainable gross primary productivity (GPP) products stand up? A critical assessment of 45 global GPP products. *Sci Total Environ* 2021;783:146965.
12. Jiao W, Wang L, McCabe MF. Multi-sensor remote sensing for drought characterization: current status, opportunities and a roadmap for the future. *Remote Sens Environ* 2021;256:112313.
13. Anav A, Friedlingstein P, Beer C, Ciais P, Harper A, Jones C, *et al.* Spatiotemporal patterns of terrestrial gross primary production: A review. *Rev Geophys* 2015;53(3):785–818.
14. Lang M, Mahyou H, Tychon B. Estimation of rangeland production in the arid oriental region (Morocco) combining remote sensing vegetation and rainfall indices: challenges and lessons learned. *Remote Sens* 2021;13(11):2093.
15. Chen Y, Guerschman J, Shendryk Y, Henry D, Harrison MT. Estimating pasture biomass using sentinel-2 imagery and machine learning. *Remote Sens* 2021;13(4):603.
16. Hacker R, Smith W. An evaluation of the DDH/100 mm stocking rate index and an alternative approach to stocking rate estimation. *Rangel J* 2007;29(2):139–148.
17. CICESE C de IC y de ES de E. Base de datos climatológica nacional (CLICOM). [Internet]. Tequisquiapan, Querétaro; 2021. Estación 22025: Consultada 6 Mar, 2021. <http://clicom-mex.cicese.mx/>.
18. Hodgson J. *Grazing management. Science into practice*. Longman Group UK Ltd. 1990.
19. Schaaf C, Wang Z. MCD43A4 MODIS/Terra+ Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global 500 m V006. NASA EOSDIS Land Processes DAAC. 2015.
20. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 2009. <https://www.r-project.org>

21. QGIS.org. QGIS Geographic Information System. QGIS Association. 2021. <http://www.qgis.org>
22. Google maps. Mapa satelital. México; 2021.
23. Pasquel D, Roux S, Richetti J, Cammarano D, Tisseyre B, Taylor JA. A review of methods to evaluate crop model performance at multiple and changing spatial scales. *Precis Agric* 2022;23:1489–1513.
24. Warton DI, Duursma RA, Falster DS, Taskinen S. smatr 3-an R package for estimation and inference about allometric lines. *Methods Ecol Evol* 2012;3(2):257–259.
25. Tedeschi LO. Assessment of the adequacy of mathematical models. *Agric Syst* 2006;89(2):225–247.
26. Hall P, Gill N, Kurka M, Phan W, Bartz A. Machine learning interpretability with H2O driverless AI. Bartz A. Editor. California, U.S.: H2O.ai Inc.; 2019.
27. LeDell E, Poirier S. H2o automl: Scalable automatic machine learning. In 2020. https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf.
28. Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput Sci* 2017;3:e127.
29. McElreath R. Statistical rethinking: A bayesian course with examples in R and STAN. Boca Raton, FL. U.S.: CRC Press; 2020.
30. Wang J, Petersen WA, Wolff DB. Validation of satellite-based precipitation products from TRMM to GPM. *Remote Sens* 2021;13(9):1745.
31. Bretas IL, Valente DS, Silva FF, Chizzotti ML, Paulino MF, D’Áurea AP, *et al.* Prediction of aboveground biomass and dry-matter content in *Brachiaria* pastures by combining meteorological data and satellite imagery. *Grass Forage Sci* 2021;76(3):340–352.
32. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;1189–1232.
33. Joiner J, Yoshida Y. Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales. *Agric For Meteorol* 2020;291:108092.
34. Griffiths W, Dodd M, Kuhn-Sherlock B, Chapman D. Management options to recover perennial ryegrass populations and productivity in run-out pastures. NZGA: Research and Practice Series. 2021;17.
35. Ahmad A, Ordonez J, Cartujo P, Martos V. Remotely piloted aircPART (RPA) in agriculture: A pursuit of sustainability. *Agronomy* 2021;11(1):7.