

UNIVERSIDAD NACIONAL DEL LITORAL



MAESTRÍA EN COMPUTACIÓN APLICADA A LA CIENCIA Y  
LA INGENIERÍA

**ESTIMACIÓN DEL CONSUMO DE PASTURAS  
A PARTIR DE REGISTROS SONOROS CON  
MODELOS NO LINEALES**

Mariela Noelia Uhrig

FICH  
FACULTAD DE INGENIERÍA  
Y CIENCIAS HÍDRICAS

**sinc**(*i*)  
INSTITUTO DE INVESTIGACIÓN EN SEÑALES,  
SISTEMAS E INTELIGENCIA COMPUTACIONAL

CICYTTP (CONICET-Prov. ER-UADER)  
CENTRO DE INVESTIGACIÓN CIENTÍFICA  
Y DE TRANSFERENCIA TECNOLÓGICA A LA PRODUCCIÓN

Tesis de Maestría 2019

**Maestría en  
Computación Aplicada a la Ciencia y la Ingeniería**

Título de la obra:

**Estimación del consumo de pasturas a partir de  
registros sonoros con modelos no lineales**

Autor: Mariela Noelia Uhrig

Lugar: Santa Fe, Argentina

Palabras Claves:

Redes neuronales artificiales

Comportamiento ingestivo en rumiantes

Regresión no lineal

Máquinas de aprendizaje extremo



UNIVERSIDAD NACIONAL DEL LITORAL  
Facultad de Ingeniería y Ciencias Hídricas  
Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

# ESTIMACIÓN DEL CONSUMO DE PASTURAS A PARTIR DE REGISTROS SONOROS CON MODELOS NO LINEALES

**Mariela Noelia Uhrig**

Tesis remitida al Comité Académico de la Maestría  
como parte de los requisitos para la obtención  
del grado de  
MAGÍSTER EN COMPUTACIÓN  
APLICADA A LA CIENCIA Y LA INGENIERÍA  
de la  
UNIVERSIDAD NACIONAL DEL LITORAL

**2019**

Comisión de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria, Paraje  
“El Pozo”, S3000, Santa Fe, Argentina.





UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas

Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

## ESTIMACIÓN DEL CONSUMO DE PASTURAS A PARTIR DE REGISTROS SONOROS CON MODELOS NO LINEALES

Mariela Noelia Uhrig

### Lugar de Trabajo:

**sinc**(*i*)

Instituto de Señales, Sistemas e Inteligencia Computacional  
Facultad de Ingeniería y Ciencias Hídricas  
Universidad Nacional del Litoral

### Director:

Dr. Diego Humberto Milone

**sinc**(*i*)-CONICET-UNL

### Co-director:

Dr. Leonardo Rufiner

**sinc**(*i*)-CONICET-UNL

### Jurado Evaluador:

Dra. Elizabeth Tapia

CIFASIS-UNR

Dr. Omar Chiotti

INGAR-CONICET-UTN

Dr. Leonardo Giovanini

**sinc**(*i*)-CONICET-UNL

Dr. César Martínez

UNL-UNER





**UNIVERSIDAD NACIONAL DEL LITORAL**  
**Facultad de Ingeniería y Ciencias Hídricas**

Santa Fe, 14 de junio de 2019

Como miembros del Jurado Evaluador de la Tesis de Maestría titulada *“Estimación del consumo de pasturas a partir de registros sonoros con modelos no lineales”*, desarrollada por la Lic. Mariela Noelia UHRIG, certificamos que hemos evaluado la Tesis y recomendamos que sea aceptada como parte de los requisitos para la obtención del título de Magíster en Computación Aplicada a la Ciencia y la Ingeniería. La aprobación final de esta disertación está condicionada a la presentación de dos copias encuadernadas de la versión final de la Tesis ante el Comité Académico de la Maestría en Computación Aplicada a la Ciencia y la Ingeniería.

  
 .....  
 Dr. Elizabeth Tapia (A)

  
 .....  
 Dr. César Martínez

  
 .....  
 Dr. Omar Chiotti

  
 .....  
 Dr. Leonardo Giovanini

Santa Fe, 14 de junio de 2019

Certifico haber leído esta Tesis preparada bajo mi dirección y recomiendo que sea aceptada como parte de los requisitos para la obtención del título de Magíster en Computación Aplicada a la Ciencia y la Ingeniería.

  
 .....  
 Dr. Leonardo Rufiner  
 Codirector de Tesis

  
 .....  
 Dr. Diego Milone  
 Director de Tesis

(\*) La Dra Elizabeth Tapia participo por video conferencia

  
 DR. JOSÉ LUIS MACOR  
 SECRETARIO DE POSGRADO  
 Facultad de Ingeniería y Cs. Hídricas

Universidad Nacional del Litoral  
 Facultad de Ingeniería y  
 Ciencias Hídricas  
 Secretaría de Posgrado

Ciudad Universitaria  
 C.C. 217  
 Ruta Nacional Nº 168 - Km. 472,4  
 (3000) Santa Fe  
 Tel: (54) (0342) 4575 229  
 Fax: (54) (0342) 4575 224  
 E-mail: posgrado@fich.unl.edu.ar





## **DECLARACIÓN LEGAL DEL AUTOR**

Esta tesis ha sido remitida como parte de los requisitos para la obtención del grado académico de Magíster en Computación Aplicada a la Ciencia y la Ingeniería ante la Universidad Nacional del Litoral y ha sido depositada en la Biblioteca de la Facultad de Ingeniería y Ciencias Hídricas para que esté a disposición de sus lectores bajo las condiciones estipuladas por el Reglamento de la mencionada Biblioteca. Citaciones breves de esta tesis son permitidas sin la necesidad de un permiso especial, en la suposición de que la fuente sea correctamente citada. Solicitudes de permiso para la citación extendida o para la reproducción parcial o total de este manuscrito serán concebidos por el portador legal del derecho de propiedad intelectual de la obra.



---

# Agradecimientos

Mientras comienzo a escribir estas líneas, se me viene a la mente toda la gente que he conocido durante esta etapa. Quiero expresar mi agradecimiento a mi director, Dr. Diego Milone por abrirme las puertas al mundo de la investigación y guiarme en todo el camino de la formación de posgrado. De igual manera quiero agradecer a mi codirector Dr. Leonardo Rufiner, por su pleno acompañamiento, dedicación y constante ayuda en la realización de esta tesis. A ambos por la abrirme las puertas del sinc(i) y darme la posibilidad de participar de él, poder llevar adelante este trabajo de manera agradable y amena.

Quiero agradecer a todos mis compañeros del sinc(i), con quienes he vivido lindos momentos dentro y fuera del trabajo. En especial a Sebastián Vanrell por su compañerismo y buena predisposición para atender cualquier inquietud que tuviera. A Marcelo Albornoz, Victoria Peterson, José Chelotti y Nicolás Nieto por estar dispuestos siempre a brindarme su ayuda. También les quiero agradecer a David, Matias, Marina, Leandro, Tomas por su amistad y buenos momentos compartidos.

Quiero agradecer al Dr. Jorge Noriega por darme la oportunidad y el espacio para comenzar esta carrera de posgrado. De igual manera, al actual director del CICYTTP, Dr. Carlos Piña, por confiar y darme el tiempo necesario en este camino de perfeccionamiento y aprendizaje.

A mi familia y amigos por estar siempre.

*A todos y cada uno de ustedes, mi mas sincero agradecimiento.*

Mariela Noelia Uhrig  
Sante Fe, 2019



---

# Índice general

<b>Agradecimientos</b>	<b>XI</b>
<b>Resumen</b>	<b>XIX</b>
<b>Abstract</b>	<b>XXI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos de la tesis . . . . .	1
1.2. Motivación . . . . .	1
1.3. Estado del arte . . . . .	3
1.4. Organización de la tesis . . . . .	5
<b>2. Redes neuronales artificiales</b>	<b>6</b>
2.1. Generalidades . . . . .	6
2.2. Perceptrón multicapa . . . . .	8
2.3. Máquinas de aprendizaje extremo . . . . .	11
<b>3. Modelos de estimación no lineal de consumo</b>	<b>13</b>
3.1. Enfoques tradicionales . . . . .	20
3.2. Selección de características . . . . .	22
<b>4. Experimentos y resultados</b>	<b>25</b>
4.1. Datos . . . . .	25
4.2. Metodologías de evaluación y validación . . . . .	27
4.3. Experimentos . . . . .	29
4.3.1. Redes neuronales . . . . .	29
4.3.2. Selección de características . . . . .	31
4.4. Resultados y discusión . . . . .	33
4.4.1. Redes neuronales . . . . .	33
4.4.2. Selección de características . . . . .	39
<b>5. Conclusiones y trabajo futuro</b>	<b>43</b>

**Bibliografía**

**45**

---

# Índice de figuras

1.1. Esquema del proceso de obtención de los sonidos y las variables acústicas para la estimación de consumo. . . . .	3
2.1. Arquitectura de un MLP completamente conectado. . . . .	8
3.1. Una oveja pastando en macetas de plástico (izquierda). Un cráneo de oveja y los lugares donde se producen los sonidos de alimentación en un rumiante (derecha). . . . .	14
3.2. Señal acústica típica de los eventos masticatorios: arranque (bite), masticación (chew) y masticación-arranque (chewbite). . . . .	17
3.3. Correlación entre variables estimadas sobre los datos experimentales utilizados en esta tesis. . . . .	19
3.4. Relación entre el <i>DMI</i> estimado y observado basado en variables predictoras acústicas del modelo lineal de referencia. . . . .	21
4.1. Ilustración esquemática de cómo se registran los sonidos (Galli et al 2011). . . . .	26
4.2. Esquema de validación cruzada LOO. . . . .	28
4.3. Esquema de particiones para la fase de entrenamiento y la de prueba. . . . .	30
4.4. Esquema de validación cruzada para los experimentos de la Sección 4.3.2. . . . .	33
4.5. <i>DMI</i> estimado y observado con MLP (izquierda) y los resultados con ELM (derecha). . . . .	35
4.6. Relación entre el <i>DMI</i> estimado y observado para el MLP (izquierda), y para el ELM (derecha). Se muestran las particiones 20 y 10 de la Tabla 4.1, para MLP y ELM respectivamente. . . . .	36
4.7. Relación entre <i>DMI</i> observado y estimado, a la izquierda para el MLP y a la derecha para el ELM con todos los datos originales. . . . .	36
4.8. Relación entre <i>DMI</i> observado y estimado con MLP. . . . .	37
4.9. Relación entre <i>DMI</i> observado y estimado con ELM-Sigmoide. . . . .	38
4.10. Relación entre <i>DMI</i> observado y estimado con ELM-Lineal. . . . .	38
4.11. <i>DMI</i> observado y estimado MLP con CFS. . . . .	40

4.12. <i>DMI</i> observado y estimado ELM con CFS. . . . .	40
4.13. <i>DMI</i> observado y estimado MLP con WRAPPER. . . . .	41
4.14. <i>DMI</i> observado y estimado ELM con WRAPPER. . . . .	42



---

# Índice de tablas

4.1. Valores de desempeño de los tres modelos del experimento utilizando LOO-CV. . . . .	37
4.2. Valores de correlación y <i>RMSE</i> discriminados por red y método de evaluación. . . . .	39
4.3. Atributos seleccionados a lo largo de todos los experimentos. . . . .	39
4.4. Resumen de resultados de todos los experimentos. . . . .	42



---

# Resumen

La producción ganadera de rumiantes en sistemas de pastoreo es una actividad importante en la región pampeana. Los niveles de productividad dependen de las dietas adecuadas de los animales, en calidad y cantidad, para satisfacer sus necesidades diarias. En este contexto, el estudio del comportamiento ingestivo y la estimación del consumo de rumiantes de pastoreo es extremadamente útil ya que permite obtener indicadores del estado nutricional, la salud y el bienestar animal. Las prácticas de manejo de pastoreo afectan fuertemente el consumo de materia seca, el rendimiento de los pastos y el desempeño de los animales. En consecuencia, la medición precisa y rápida del comportamiento de pastoreo y el consumo de forraje permite mejorar la eficiencia de la gestión de los recursos hídricos y alimentarios. Sin embargo, las mediciones directas durante el tiempo total de pastoreo no son prácticas y son extremadamente exigentes en tiempo. En este marco, el uso de algoritmos de procesamiento inteligente de señales que permitan extraer la información relevante de registros sonoros de rumiantes se presenta como una opción válida para predecir el consumo de forraje de rumiantes en condiciones de pastoreo. Más aún, teniendo en cuenta también que debe ser un método sencillo desde el punto de vista tecnológico y que además no perturbe el comportamiento natural del animal. En esta tesis se propone utilizar dos tipos de redes neuronales artificiales, ambas actuando como modelos de regresión no lineal multivariada, para estimar el consumo de materia seca en ovejas. Dado que la cantidad de datos disponibles es relativamente pequeña en relación a la cantidad de parámetros ajustables de las redes neuronales, se propone generar mayor cantidad de datos utilizando un método de remuestreo. Otra estrategia que se explora es la selección de características para trabajar con un subconjunto más pequeño de atributos con el objetivo de mejorar el desempeño predictivo para ambas redes. Con el fin de comparar y validar los resultados, se implementó un sistema de referencia basado en un modelo lineal propuesto recientemente. En todos los casos se utilizó la metodología de validación de "dejar uno afuera". Los resultados obtenidos con las redes neuronales muestran una mejora importante respecto de las técnicas lineales que fueron utilizadas hasta el momento. Las máquinas de aprendizaje extremo presentan una significativa mejora respecto del modelo lineal y al perceptrón multicapa para iguales condiciones de validación.



---

# Abstract

Livestock production by ruminants in grazing systems is an important activity in the Pampas region. The productivity levels depends on adequate animals diets, in quality and quantity to meet their daily nutritional requirements. In this context, the study of the ingestive behavior and the estimation of the intake of grazing ruminant is extremely useful since it allows to obtain indicators of nutritional status, health and animal welfare. Grazing management practices strongly affects dry matter intakes, pasture yields and animal performances. Consequently, the accurate and rapid measurement of grazing behavior and forage intake allows to improve the efficiency of herd and food resources managements. However, direct measurements during the total grazing time are impractical and extremely demanding in time. In this framework, the use of intelligent signal processing algorithms that allow the extraction of relevant information from sound records of ruminants is presented as a valid option to predict the forage intake of ruminants in grazing conditions. Bearing in mind also, that it must be a simple method from the technological point of view and that in addition it does not disturb the natural behavior of the animal. In this thesis it is proposed to use two types of artificial neural networks, both acting as multivariate non-linear regression models, to estimate dry matter intake. Given that the amount of data available is relatively small in relation to the number of adjustable parameters of neural networks, it is proposed to generate more data using a resampling method. Another strategy that is explored is the feature selection to work with a smaller subset of attributes in order to improve the predictive performance for both prediction models. In order to compare and validate the results, a reference system based on a recently proposed linear model was implemented. In all cases, the validation methodology of leaving one out was used. The results obtained with the neural networks show an important improvement with respect to the linear techniques that were used up to now. The extreme learning machines present a significant improvement with respect to the linear model and the multilayer perceptron for the same validation conditions.



---

# Capítulo 1

## Introducción

Este capítulo se organiza en cinco subsecciones. En la Sección 1.1 se introduce la necesidad de tener un modelo que estime la materia seca consumida. En la Sección 1.2 se detalla el estado del arte de la problemática en cuestión. Por último se expresan los objetivos de la presente tesis y la organización de los capítulos de la misma en las Secciones 1.3 y 1.4 respectivamente.

### 1.1. Objetivos de la tesis

El objetivo general de la presente tesis es desarrollar métodos que permitan estimar en forma precisa el consumo de los rumiantes aportando información importante para el estudio del comportamiento animal a partir de las señales registradas por un sistema lo más elemental y menos invasivo posible.

#### Objetivos específicos

- Desarrollar un modelo de regresión no lineal para estimar el consumo con un perceptrón multicapa.
- Desarrollar un modelo de regresión no lineal para estimar el consumo con máquinas de aprendizaje extremo.
- Realizar la validación y análisis comparativo de los resultados obtenidos.

### 1.2. Motivación

La mayoría de los sistemas de pastoreo buscan la utilización eficiente del forraje y la producción animal mediante prácticas que sean económicas y ecológicamente adecuadas. De allí surge la necesidad de monitorear, evaluar y gestionar rutinariamente las relaciones entre los recursos de pastoreo, la ingesta de forraje y la producción

animal. Por lo que, la medición precisa del consumo es fundamental para avanzar en el conocimiento de la ecología de los rumiantes de pastoreo.

En virtud de encontrar un modelo que prediga con facilidad y precisión, se han realizado investigaciones aplicando diferentes técnicas en la estimación del consumo. En el avance de las investigaciones se descubrió que los sonidos de morder y masticar pueden revelar características importantes del comportamiento de forrajeo de los animales. En ellos se develó que, la energía contenida en los sonidos de masticación está relacionada con la ingesta y las características de los forrajes consumidos, lo que sugiere que la ingesta de materia seca (DMI, del inglés *Dry Matter Intake*) podría predecirse bastante bien en función de los parámetros medibles del sonido de masticación. De hecho, en [1], encontraron un modelo de regresión lineal, donde los mejores predictores de DMI en el pastoreo de ovejas experimentales fueron la energía de sonido de masticación por mordida y el flujo de energía total de los sonidos de masticación, obteniendo un modelo con cinco variables predictoras con una correlación  $R^2=0,92$  y un error  $RMSE=4,1$ . En virtud de obtener una mejora en el modelo de estimación, con mejor desempeño y con menor complejidad, se propone utilizar un modelo de regresión no lineal con selección de características, a partir del modelado con dos tipos de redes neuronales. Los ovinos son rumiantes, y como tales, su alimentación requiere de una serie de eventos específicos en la alimentación. Las actividades de alimentación incluyen: pastoreo, rumia (regurgitación del alimento ingerido para continuar con la masticación) y silencio. El sonido durante el pastoreo contiene una secuencia de eventos intercalados por silencios, donde cada evento se produce como manifestación de dos etapas bien diferenciadas durante la ingestión del forraje en pastoreo: el arranque del forraje (bite), donde el alimento es cortado e ingresa a la boca del animal, y la masticación (chew), cuando el forraje es triturado por los dientes y reducido de tamaño para favorecer su digestión. Existe un tercer tipo de evento que corresponde a un movimiento compuesto que realiza el animal, cuando produce una masticación y un arranque en el mismo movimiento mandibular (chew-bite). Clasificarlos es una condición necesaria para medidas confiables de comportamiento de pastoreo e incluso para la estimación de la ingesta de forraje mediante métodos acústicos. En la Figura 1.1 se ilustra de manera general el proceso de la toma de la señal sonora, la etapa de preprocesamiento de la señal importante para la identificación de las actividades de alimentación (pastoreo y rumia), utilizando el CBTK [2]; los sonidos de masticar y morder se separaron y analizaron como en [3] para obtener el número de bites, el número de chews, el número de mordidas de chewbites, entre otras características importantes para el cálculo de las variables acústicas que son utilizadas en el modelo predictor de la materia seca consumida por el ovino. (Para mayor detalle de cada proceso de análisis de la señal y obtención de las variables ver [1]).



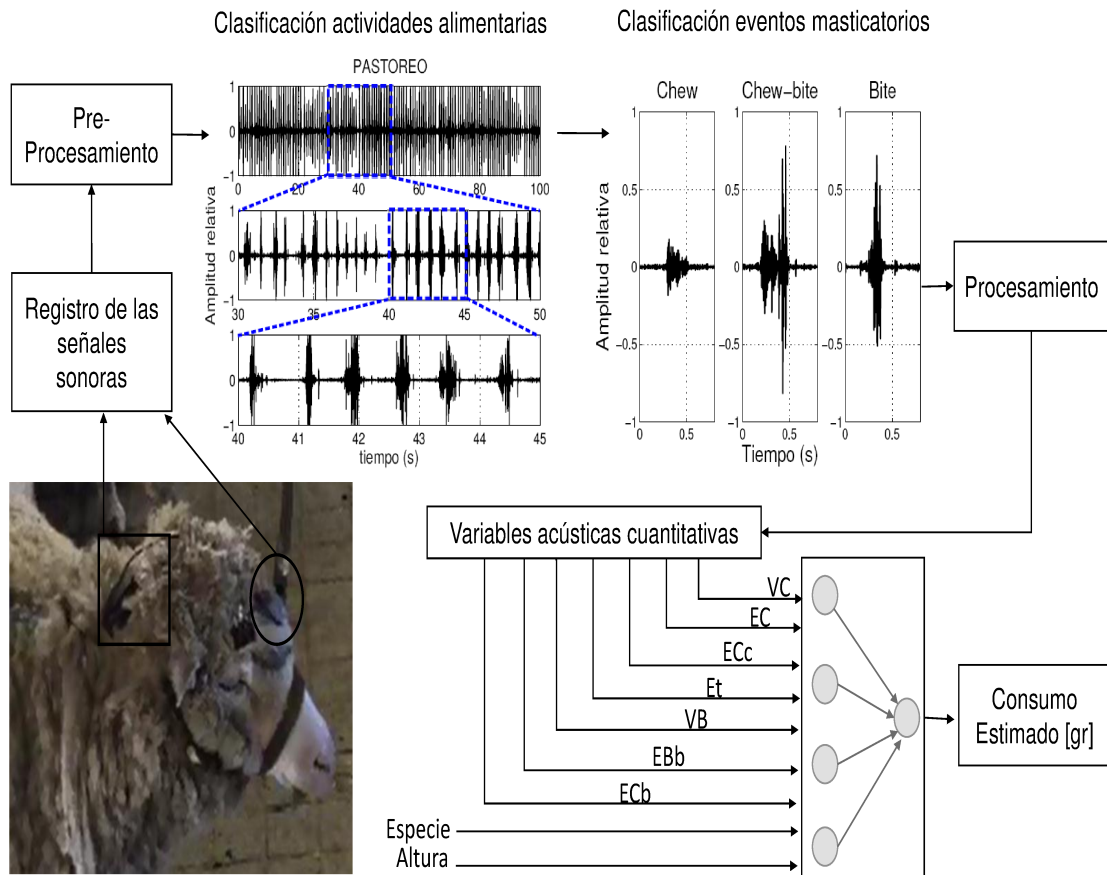


Figura 1.1: Esquema del proceso de obtención de los sonidos y las variables acústicas para la estimación de consumo.

### 1.3. Estado del arte

Disponer de una metodología precisa y económica para la medición del consumo de pasturas de rumiantes es importante para mejorar el desempeño de los sistemas de producción basados en el aprovechamiento directo de las pasturas. Este aspecto es esencial para lograr una gestión eficiente del rodeo y los recursos alimenticios, así como para la salud y bienestar animal. El proceso de pastoreo comprende la ingestión, rumia y descanso (actividades relacionadas con el comportamiento del animal durante el día). Dichas actividades están afectadas por diversos factores como el tipo de pastura [4], el estado de salud del animal, el clima y el manejo del pastoreo, que está fuertemente relacionado con la cantidad y calidad de la producción obtenida. Disponer de información clara y precisa sobre este proceso permite mejorar el manejo del pastoreo y así la productividad. Antes de que se comenzaran a investigar métodos automáticos y semiautomáticos para el estudio del comportamiento ingestivo de

los rumiantes, se utilizaba la observación directa del animal. Pero esta es una técnica que requiere mucho esfuerzo y es difícil de implementar simultáneamente en un grupo numeroso de animales y durante largos periodos de tiempo. Posteriormente, se desarrollaron sistemas basados en distintos tipos de dispositivos que, colocados en la cabeza del animal, miden el movimiento mandibular. Por ejemplo, colocando una banda que cruza por debajo de la mandíbula y mediante circuitos eléctricos se detectaba cuando esta se abría. Luego el sistema convertía estos datos en información binaria [5]. A fines de la década del '80 se comenzó a utilizar el sonido en estudios de comportamiento animal. Por ejemplo, para el estudio del comportamiento en puercoespines se diseñó un prototipo de sistema transmisor-receptor capaz de clasificar correctamente las acciones de los animales (comer, beber, caminar) solo mediante el registro remoto del sonido [6]. También se demostró que el sonido que producen los animales al comer es una fuente de información que permite caracterizar y cuantificar las actividades de alimentación en vacas y ovejas (arranques, masticaciones y movimientos combinados), el tipo de pastura que ingieren y la cantidad total ingerida [2, 1], que la energía del sonido está altamente relacionada con la cantidad de materia seca consumida por el animal [7, 8] y que se obtiene información valiosa para cuantificar el comportamiento de ingestión de los animales criados en libertad [9] y las vacas lecheras en pastoreo [10]. Con ello, se comprueba que las características acústicas poseen gran cantidad de información relacionada con el consumo y se han demostrado las ventajas de utilizar métodos acústicos en estudios comparativos de estos con métodos mecánicos [11]. En un trabajo reciente Galli y colaboradores [12], informaron que la densidad de flujo de energía de masticación total, el número de eventos masticación-arranque y la altura de la planta (alto/corto) fueron los predictores más importantes de la materia seca consumida. Sin embargo, utilizar algoritmos de procesamiento inteligente de señales que permitan extraer la información relevante de registros sonoros de rumiantes y predecir el consumo de materia seca es todo un desafío. Actualmente, una de las mejores formas de realizar este tipo de tareas de predicción es mediante el aprendizaje maquina y dentro de éste las redes neuronales. Éstas basan su funcionamiento en interconexiones (sinapsis) entre nodos simples (neuronas), y dichas interconexiones ajustan sus pesos sinápticos en base a los datos de entrenamiento. Las redes neuronales artificiales han sido creadas

con base en una analogía con los sistemas nerviosos naturales y permiten lograr modelos de regresión no lineales adecuados en muchos casos de aplicación. Por ejemplo, las redes de perceptrones multicapa (MLP, del inglés *Multilayer Perceptron*), han sido usadas en diversos problemas tales como en la predicción de inundaciones [13], en la predicción de enfermedades cardíacas [14], o para estimar productividad en plantaciones forestales [15]. Otro método reciente son las máquinas de aprendizaje extremo (ELM, del inglés *Extreme Learning Machines*), que han mostrado gran capacidad de predicción en diferentes problemas [16]. Por ejemplo, para la aplicación de diagnóstico médico en diabetes [17], o en la predicción de energía eléctrica en edificios [18]. La principal ventaja de las ELM es la ausencia del ajuste en la capa oculta del modelo neuronal, por lo que su tiempo de aprendizaje es mucho menor que un MLP clásico y su capacidad de generalización es mayor [17].

## 1.4. Organización de la tesis

El resto de la tesis está organizada de acuerdo a los siguientes capítulos:

- El Capítulo 2 presenta generalidades y conceptos de redes neuronales artificiales, explicando el funcionamiento del perceptrón multicapa y de las máquinas de aprendizaje extremo que se utilizan en este trabajo.
- El Capítulo 3 se presentan los modelos lineales y no lineales desarrollados para la estimación del consumo. Además se introducen los conceptos y criterios utilizados para la selección de características.
- El Capítulo 4 se describen los datos utilizados y métodos que se proponen, se describen los experimentos llevados a cabo y los resultados obtenidos.
- El Capítulo 5 resume las conclusiones y trabajos a realizar a futuro.

---

# Capítulo 2

## Redes neuronales artificiales

Este capítulo se divide en tres secciones. La Sección 2.1 introduce conceptos generales de la inteligencia artificial. En la Sección 2.2 se abordan los conceptos del perceptrón multicapa, estructura y algoritmo de entrenamiento, y por último, en la Sección 2.3 se explican las máquinas de aprendizaje extremo, sus definiciones y algoritmo de aprendizaje.

### 2.1. Generalidades

La inteligencia artificial (IA) intenta descubrir y describir aspectos de la inteligencia humana que pueden ser simulados mediante máquinas, especialmente por sistemas informáticos. En algunas ramas de la IA se conceptualiza al hardware de una computadora como un cerebro, y al software como el procesamiento de la información llevada a cabo por la mente. Este procesamiento incluye el aprendizaje (la adquisición de información y reglas para su uso), el razonamiento (usando las reglas para llegar a conclusiones aproximadas o definitivas) y la autocorrección.

De la analogía con las redes neuronales biológicas surgen las redes neuronales artificiales, que son estructuras que se modelan mediante un conjunto de unidades de procesamiento conectadas por enlaces, los cuales distribuyen los patrones de activación generados entre las unidades. De esta forma, se puede definir a una red neuronal artificial (RNA) como un sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona. Como ya se sabe, el pensamiento tiene lugar

en el cerebro, que consta de billones de neuronas interconectadas. Así, el secreto de la “inteligencia” se sitúa dentro de estas neuronas y de su interacción. Lo que básicamente ocurre en una neurona biológica es lo siguiente: la neurona es estimulada o excitada a través de sus entradas y cuando se alcanza un cierto umbral, la neurona se dispara o activa, pasando una señal hacia el axón. También, es bien conocido que los humanos son capaces de aprender. Aprendizaje significa que aquellos problemas que inicialmente no pueden resolverse, pueden ser resueltos después de obtener más información acerca del mismo. Por lo tanto, las RNA:

- Se componen de unidades de procesamiento que intercambian datos o información.
- Se utilizan para clasificación o predicción.
- Tienen capacidad de aprender y mejorar su funcionamiento.

La RNA más simple fue introducida por Rosenblatt en 1958 [19], el Perceptrón "simple", constituida solamente por una neurona. Esta neurona recibe estímulos de su entorno por medio de un conjunto de entradas, realiza sobre ellas un procesamiento elemental y obtiene un valor de salida que representa la actividad de la neurona. El perceptrón es capaz de lograr un ajuste de los pesos con error cero solamente cuando las entradas pertenezcan a dos clases diferentes, y además las clases puedan ser separadas por un hiperplano en el espacio N-dimensional de los datos. Estas condiciones llevan a que el perceptrón solamente pueda resolver problemas que sean linealmente separables. Para superar estas limitaciones se han desarrollado numerosas arquitecturas neuronales que son capaces de actuar como aproximadores universales de funciones: una red de retropropagación conteniendo al menos una capa oculta con suficientes unidades no lineales puede aproximar cualquier tipo de función o relación continua entre un grupo de variables de entrada y salida. Esta propiedad convierte a las redes en herramientas de propósito general, flexibles y no lineales [20, 21]. A continuación se exponen las características y algoritmos de entrenamiento de los dos tipos de modelos de redes neuronales utilizados en esta tesis.

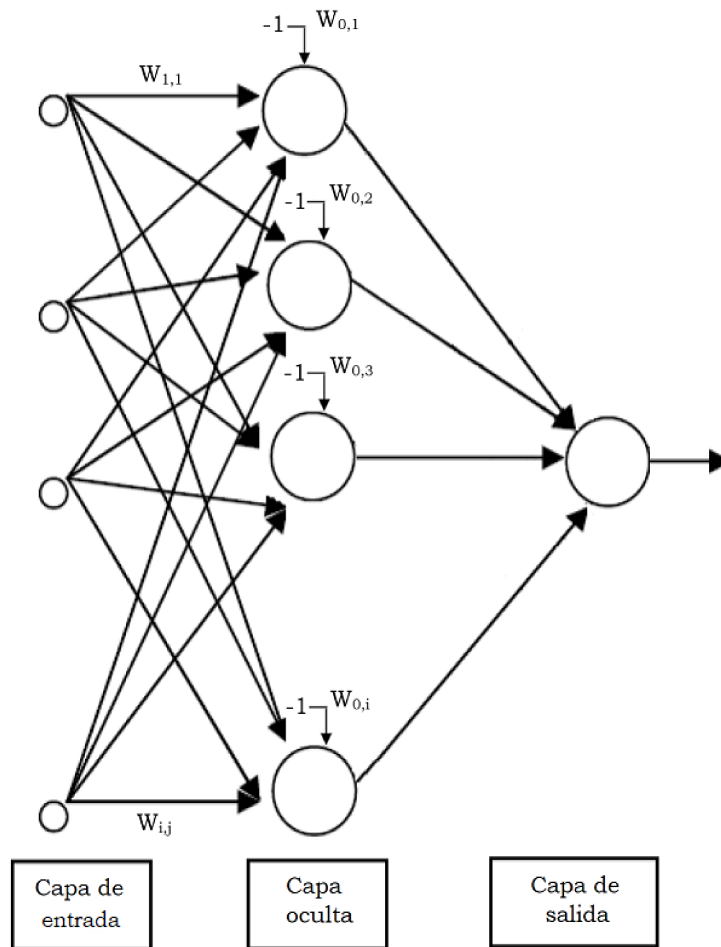


Figura 2.1: Arquitectura de un MLP completamente conectado.

## 2.2. Perceptrón multicapa

El Perceptrón Multicapa está basado en múltiples perceptrones como nodos de procesamiento y su topología está definida por una capa de entrada, capas ocultas y una capa de salida (ver Fig. 2.1). La capa de entrada recibe la información que entra del exterior a la red. Las capas ocultas están formadas por aquellas neuronas cuyas entradas provienen de capas anteriores y cuyas salidas pasan a neuronas de capas posteriores. Todas las capas están interconectadas nodo a nodo, con un peso asociado, mientras que los nodos de una misma capa no están conectados entre sí. Estos pesos son los que luego se ajustan durante el entrenamiento.

Un algoritmo muy popular que permite entrenar al perceptrón multicapa de manera supervisada es el conocido como algoritmo de retropropagación del error. Este algoritmo se basa en la regla de aprendizaje por corrección de error. Básicamente, el algoritmo consta de dos pasadas a través de las diferentes capas de la red: una pasada hacia adelante y otra hacia atrás. En la propagación hacia adelante, la información es propagada por la red desde la capa de entrada a cada neurona de la capa oculta, y desde cada neurona de la capa oculta hasta las neuronas de la capa de salida. Las entradas se multiplican por los pesos sinápticos y la suma total de estos productos llega a las neuronas de la capa oculta. La salida de cada neurona se calcula como la combinación lineal de sus entradas y pesos, resultando en

$$H_j = \sum_{i=0}^N x_i w_{ij}, 1 \leq j \leq M \quad (2.1)$$

$$Y_k = h_j w_{jk} \quad (2.2)$$

siendo  $H_j$  e  $Y_k$  las salidas lineales de las capas oculta y de salida, respectivamente. A la suma resultante en cada neurona se le aplica una función no lineal, denominada función de activación, y el resultado de esta operación es considerado como entrada de la siguiente capa

$$h_j = f(H_j) \quad (2.3)$$

$$y_k = f(Y_k) \quad (2.4)$$

siendo  $f(\cdot)$  la función sigmoidea, función de activación más utilizada, que se define como

$$f(x) = \frac{1}{(1 + e^{-\beta x})}. \quad (2.5)$$

Por otro lado, durante la pasada hacia atrás (entrenamiento), se realiza un proceso de ajuste de los pesos sinápticos, para acercar las salidas del perceptrón multicapa a los valores objetivo o salidas deseadas. Esto es equivalente al proceso de minimización del error cuadrático medio,

$$E(w) = \frac{1}{2} \sum_k (y'_k - y_k)^2 \quad (2.6)$$

donde  $w$  representa de forma genérica a todos los pesos de la red,  $y'_k$  es la  $k$ -ésima salida deseada y  $y_k$  la  $k$ -ésima salida que da la red neuronal. La salida del perceptrón multicapa se compara con el objetivo y si el error no es nulo se ajustan los pesos de la red para que la diferencia entre la salida de la red y la salida deseada sea menor en la siguiente iteración. Los pesos se ajustan por [22],

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j x_i \quad (2.7)$$

donde  $w_{ij}(t)$  es el peso del nodo oculto  $i$  (o de una entrada) al nodo  $j$  en el tiempo  $t$ ,  $x_i$  es la salida del nodo  $i$  (o es una entrada),  $\eta$  es un factor de ganancia o velocidad de aprendizaje, y  $\delta_j$  es el error para el nodo  $j$ . Si el nodo  $j$  es un nodo de salida, entonces

$$\delta_j = y_j(1 - y_j)(y'_j - y_j) \quad (2.8)$$

donde  $y'_j$  es la salida deseada del nodo  $j$  y  $y_j$  la salida real. Si el nodo  $j$  es un nodo oculto interno, entonces

$$\delta_j = x_i(1 - x_i) \sum_k \delta_k w_{jk} \quad (2.9)$$

donde  $k$  está sobre todos los nodos en la capa siguiente a la del nodo  $j$ . Los umbrales de los nodos internos se adaptan de manera similar al suponer que son los pesos de conexión en los enlaces de las entradas de valores constantes auxiliares.

La convergencia es a veces más rápida si se agrega un término de momento y los cambios de pesos se suavizan con

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j x_i + \alpha(w_{ij}(t) - w_{ij}(t-1)), \quad \text{con } 0 < \alpha < 1. \quad (2.10)$$

Lo que se hace es tener en cuenta el ajuste realizado en la iteración  $t$ , para hacer el ajuste en la iteración  $t+1$ , con un coeficiente de momento  $\alpha$ . De esta forma se pretende obviar saltos bruscos en las direcciones de optimización que marca el gradiente descendente.

El aprendizaje finaliza tras un número finito de iteraciones, una vez alcanzado un mínimo de la función error [23]. En una aplicación práctica del algoritmo de retropropagación, el aprendizaje surge como resultado de muchas presentaciones de



un conjunto de ejemplos de entrenamiento al perceptrón multicapa. A una completa presentación de un conjunto de entrenamiento determinado durante el proceso de aprendizaje se le llama *época*. El proceso de aprendizaje es realizado época a época hasta que los pesos sinápticos y los niveles de umbral de la red se estabilizan y el error cuadrático medio del conjunto de entrenamiento global converge a un mínimo.

### 2.3. Máquinas de aprendizaje extremo

Las máquinas de aprendizaje extremo (ELM, del inglés *Extreme Learning Machine*) fueron desarrolladas para redes hacia delante de una sola capa oculta o de tipo *feedforward* (SLFNs, del inglés *Single Layer Feedforward Networks*). En general, una SLFN estándar con  $M$  neuronas ocultas y función de activación  $f(\cdot)$  puede aprender de forma exacta  $n$  patrones distintos (con  $n \geq M$ ), inicializando aleatoriamente los pesos de entrada de las neuronas ocultas [16]. El algoritmo de aprendizaje extremo se basa en este hecho, tomando como base la estructura del perceptrón de una capa con  $M$  neuronas. Los pesos de salida se pueden obtener analíticamente mediante la matriz pseudoinversa [24, 25]. Así, dado un conjunto de  $n$  vectores de entrada  $\mathbf{x}_j$ , existen los parámetros  $\beta_i$ ,  $\mathbf{w}_i$  y  $b_i$  con

$$\sum_{i=1}^M \beta_i f(\mathbf{w}_i \mathbf{x}_j^T + b_i) = y'_j, \quad j = 1, \dots, n^1 \quad (2.11)$$

donde  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{iN}]$  es el vector de pesos aleatorios que conecta el nodo oculto  $i$  con los  $N$  nodos de entrada,  $\beta_i$  es el peso ajustable que conecta el nodo oculto  $i$  con la salida,  $b_i$  es el umbral o sesgo del nodo oculto  $i$  y  $y'_j$  es la  $j$ -ésima salida objetivo. Esta ecuación se puede expresar de forma matricial como

$$H\boldsymbol{\beta} = Y. \quad (2.12)$$

Como en [16, 17], a  $H$  se la llama matriz de salida de la capa oculta de la red neuronal. La columna  $i$  de  $H$  es la salida del nodo oculto  $i$  con respecto a las

---

<sup>1</sup>Nótese que no hay función de activación en la capa de salida

entradas  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ . Esto es,

$$H(\mathbf{w}_1, \dots, \mathbf{w}_M; b_1, \dots, b_M; \mathbf{x}_1, \dots, \mathbf{x}_n) = \begin{bmatrix} f(\mathbf{w}_1 \mathbf{x}_1^T + b_1) & \dots & f(\mathbf{w}_M \mathbf{x}_1^T + b_M) \\ \cdot & & \cdot \\ \cdot & \dots & \cdot \\ \cdot & & \cdot \\ f(\mathbf{w}_1 \mathbf{x}_n^T + b_1) & \dots & f(\mathbf{w}_M \mathbf{x}_n^T + b_M) \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix}, \quad Y = \begin{bmatrix} y'_1 \\ \cdot \\ \cdot \\ \cdot \\ y'_n \end{bmatrix}.$$

El entrenamiento consiste en un problema de mínimos cuadrados con pesos óptimos,

$$\boldsymbol{\beta}' = H^\dagger Y$$

donde

$$H^\dagger = (H^T)^{-1} H^T$$

es la pseudoinversa de Moore-Penrose [22,23]. La solución a este problema se calcula resolviendo la pseudoinversa de la matriz de salida de la capa oculta. Una de las características más destacadas de las ELM es que la capa oculta no necesita ser entrenada.

Este algoritmo funciona para cualquier función de activación infinitamente diferenciable  $f(x)$ . Tales funciones de activación incluyen las funciones sigmoideas, así como la base radial, seno, coseno, exponencial y muchas funciones no regulares, como se muestra en [26]. En nuestro caso, realizamos experimentos con dos tipos de funciones: lineal y sigmoide.

Las ELM son fáciles de utilizar y proporcionan un entrenamiento rápido y eficiente. Sin embargo, como en cualquier red neuronal, es necesario fijar la arquitectura de la red, es decir, el número de neuronas ocultas. En general, se desconoce el valor óptimo de  $M$  y es necesario realizar una búsqueda del mismo, lo que puede resultar costoso computacionalmente.

---

## Capítulo 3

# Modelos de estimación no lineal de consumo

Este capítulo consta de tres secciones. En la primera parte se explica a modo introductorio, el fenómeno acústico, las variables utilizadas y el modelo lineal de referencia. En la Sección 3.1 se introducen los conceptos tradicionales de los modelos lineales y en la Sección 3.2 se detalla la etapa que puede preceder al aprendizaje, que es la selección de características, con sus ventajas y los métodos utilizados en la presente tesis.

### Fenómeno acústico

Los rumiantes producen sonidos durante la ingesta de los alimentos. La adquisición de esos sonidos es importante para extraer información relevante para la estimación del consumo. Los sonidos producidos en la alimentación se deben a los movimientos de la mandíbula del animal que se propagan a través del cráneo por conducción ósea en forma de vibraciones (Figura 3.1). Los sonidos de arranque se producen en la parte frontal de la boca, donde los incisivos están presentes solo en la mandíbula inferior (Figura 3.1) y una almohadilla de piel dura (almohadilla córnea) reemplaza los incisivos superiores. Además, se producen sonidos de masticación en la parte posterior de la boca, donde las ovejas mastican los alimentos con movimientos de lado a lado utilizando los premolares y los molares (Figura 3.1) para triturar las plantas en pequeñas partículas que son más fáciles de digerir. Este fenómeno ha sido ampliamente estudiado en humanos para aplicaciones en audición [27, 28]. En

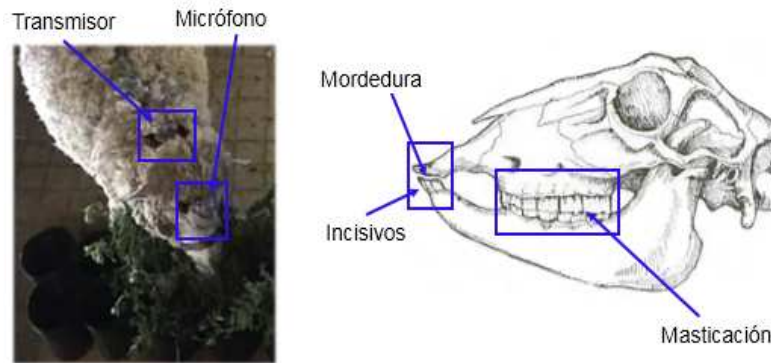


Figura 3.1: Una oveja pastando en macetas de plástico (izquierda). Un cráneo de oveja y los lugares donde se producen los sonidos de alimentación en un rumiante (derecha).

los animales, estas vibraciones se pueden sentir usando un micrófono colocado en la frente apuntando hacia "adentro", de una manera no invasiva, para no interferir con el comportamiento natural del animal.

En [9] los autores informaron que usando la señal de sonido era posible discriminar adecuadamente entre arranques y masticaciones por diferencias en el espectro (0 - 8 kHz) y otras variables acústicas (duración del sonido, densidad de flujo de energía, entre otras). Esta información de los sonidos de pastoreo tiene el potencial de superar muchos de los problemas asociados con la medición de la ingesta de pastoreo. En la Figura 3.2 se puede observar un segmento de señal acústica típica obtenida de esta forma, conteniendo los tres eventos mandibulares<sup>1</sup> de los rumiantes (arranque, masticación y movimiento compuesto de masticación-arranque). Este es el tipo de señales registradas y procesadas en los experimentos para obtener las 10 características originales que fueron utilizadas en [1]. En dicho trabajo concluyeron que el DMI se puede estimar con precisión usando medidas acústicas de comportamiento ingestivo. Luego éstas medidas se utilizaron como punto de partida en este trabajo, así como también para poder compararnos con su modelo en igualdad de condiciones.

### Mediciones experimentales y cálculo de variables

El DMI se estimó como la diferencia entre la biomasa forrajera pre y post sesión de pastoreo. Las macetas se pesaron individualmente con una precisión de 0,1 g

<sup>1</sup>Mandibulares o masticatorios son usados como sinónimos.

utilizando una balanza digital Setra 140 CP. Para estimar las pérdidas por evaporación, se pesaron dos macetas por prueba antes y después de cada sesión de pastoreo. Todos los días se seleccionó al azar un subconjunto de macetas de cada especie y altura para medir la altura del forraje en cinco hojas extendidas (en *orchardgrass*) o tallos (en *alfalfa*). Las muestras representativas del pastoreo obtenidas de macetas se secaron a 65°C, se pesaron y se analizaron para determinar el contenido de fibra detergente neutra (NDF, [29]). Las pistas de sonido de las cintas de video se analizaron utilizando CBTK, un software propio para el reconocimiento de eventos [2], y el software Cool Edit Pro versión 2 [30]. La frecuencia de muestreo fue de 44,100 kHz y el tamaño de la muestra (resolución) fue de 16 bits. Para calcular la tasa de consumo (consumo materia seca/tiempo que estuvo comiendo), la tasa de arranque (número de bocados/tiempo que estuvo comiendo) y la masa de arranque (consumo materia seca/número de bocados), se determinó el número de bocados y el tiempo de ingesta a partir de las pistas de sonido de las cintas de video. Para delimitar el tiempo de ingesta, se consideró como inicio al primer arranque de pastura y como final al tragado del último bolo alimenticio. Los arranques se identificaron por el sonido de molienda producido cuando las ovejas cortan el forraje; las masticaciones se identificaron por el sonido de cada masticación y los arranques se evidenciaron por una masticación anterior y se superponía parcialmente a un arranque dentro de un solo movimiento de la mandíbula. Los sonidos de masticación-arranque se producen cuando el forraje que ya está en la boca se mastica a medida que las mandíbulas se cierran para cortar más forraje. Los sonidos de masticar y arrancar se separaron y analizaron como en [3] para obtener el número de arranques (B), el número de masticaciones (C), el número de masticación-arranque (ChB), el tiempo de arranque (TB), el tiempo de masticación (TC), la intensidad promedio (en decibelios) de los arranques ( $\log VB$ ) e intensidad de masticación ( $\log VC$ ). Luego, los movimientos totales de la mandíbula (TJM) fueron  $(B+C-ChB)$ , la tasa total de movimientos de la mandíbula fue  $TJM/T$ , la tasa de masticación ( $C_T$ ) fue  $C/T$ , la masticación por arranque fue  $C/B$  y la masticación exclusiva por arranque fue  $((C-ChB)/B)$ . Los movimientos de la mandíbula que no produjeron sonido fueron ignorados. El número de masticaciones por g de DMI fue  $C/DMI$ , y el número de masticaciones por g de NDFI fue  $C/NDFI$ . La densidad de flujo de energía acústica (EFD) es el producto

de la intensidad acústica y la duración del sonido. En arranques y masticaciones, EFD se relaciona mecánicamente con la cantidad de forraje cortado y triturado. Las variables  $\log VB$  y  $\log VC$  se midieron mediante la opción de estadísticas de Cool Edit Pro, y otras variables se calcularon como:

- Intensidad de arranque:  $VB = 10^{(\log VB/10)} I_{ref}$
- Intensidad de masticación:  $VC = 10^{(\log VC/10)} I_{ref}$
- Energía total de densidad de flujo de masticación:  $EC = VC TC$
- Energía de densidad de flujo de arranque por arranque:  $EBb = EB/B$
- Energía total de densidad de flujo de masticación por masticación:  $ECc = EC/C$
- Energía de densidad de flujo de masticación por arranque:  $ECb = EC/B$
- Energía de densidad de flujo masticación por unidad de tiempo de comida:  $Et = EC/T$

donde  $VB$  y  $VC$  son las intensidades promedio en  $W/m^2$  de arranques y masticaciones,  $\log VB$  y  $\log VC$  son las intensidades promedio en dB de arranques y masticaciones,  $I_{ref}$  es la intensidad de referencia en el aire (se asumió arbitrariamente que era 1 pW para tener sentido). Las dimensiones, el tiempo de masticación y el tiempo de arranque son la duración de la señal, excluyendo todos los "silencios" entre arranques o masticaciones. La duración de la masticación y la duración del arranque son medidas del tiempo durante el cual los dientes aplastan y cortan el forraje, y no es necesariamente una medida del tiempo total que se tarda en completar todo el movimiento de la mandíbula.

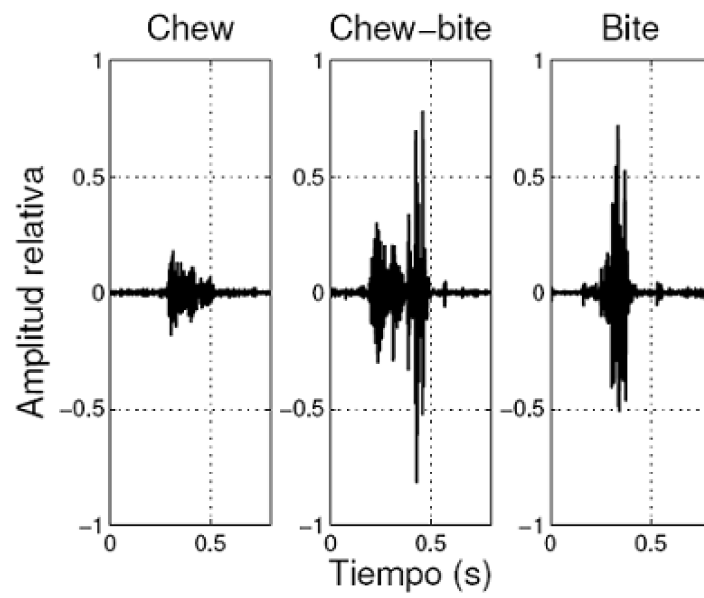


Figura 3.2: Señal acústica típica de los eventos masticatorios: arranque (bite), masticación (chew) y masticación-arranque (chewbite).

En la Figura 3.3 se muestra la correlación existente entre todas las variables explicadas en el párrafo anterior, que fueron utilizadas como entradas a los modelos en los distintos experimentos. Podemos observar que las relaciones de varias de ellas no son lineales, como vemos por ejemplo en el caso de  $Et$  con el  $DMI$  y  $EC$ ; o bien entre  $EBb$  con  $DMI$ , con  $ECc$  y con  $VC$ . Estos ejemplos permiten tomar dimensión de la complejidad de las relaciones entre estas variables.



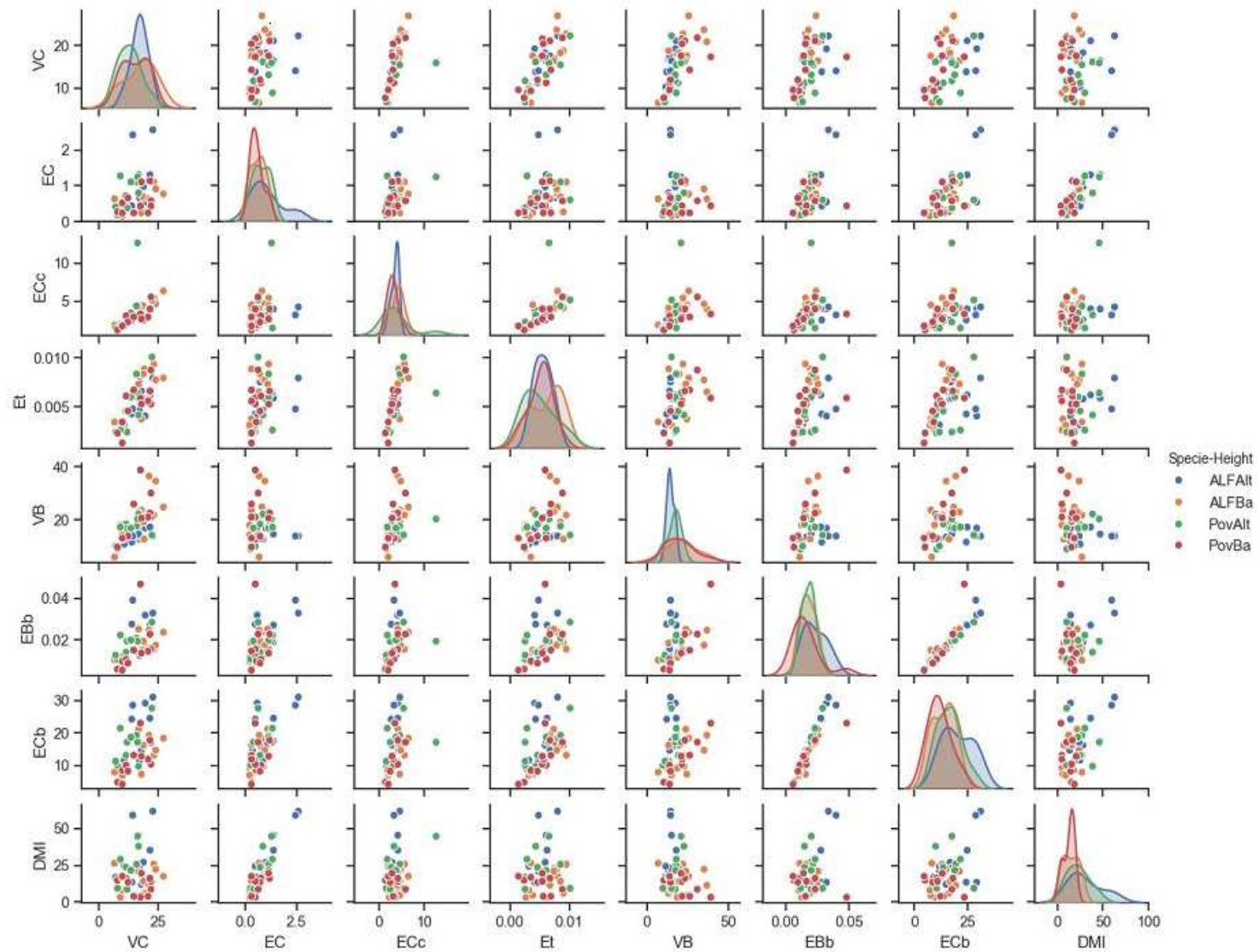


Figura 3.3: Correlación entre variables estimadas sobre los datos experimentales utilizados en esta tesis.

### 3.1. Enfoques tradicionales

Desde un enfoque matemático-estadístico, el problema de estimación de consumo puede ser abordado mediante técnicas de regresión. En este sentido, un proceso que tiene variables independientes  $X = [x_1, x_2, \dots, x_k]$  que afecten o puedan afectar a una variable dependiente  $Y$ , puede ser representado por modelos estadísticos de regresión que permiten explicar y predecir su comportamiento. Suponiendo que se cumple la hipótesis de causalidad en las variables analizadas, y con el modelo adecuado, es posible realizar estimaciones sobre la variable  $Y$  a partir de las variables  $X$ . El modelo clásico lineal sería

$$Y = \beta_0 + \beta X + \epsilon. \quad (3.1)$$

donde  $\beta = [\beta_1, \beta_2, \dots, \beta_k]$ ;  $X = [x_1, x_2, \dots, x_k]$

De acuerdo con la ecuación 3.1,  $Y$  es la variable dependiente,  $x_1, x_2, \dots, x_k$  son las variables independientes,  $\beta_0$  es la ordenada al origen del plano de regresión,  $\beta_1$  es el parámetro que indica el cambio esperado de la respuesta  $Y$  por cambio unitario en  $x_1$ , cuando las otras  $x$ 's se mantienen constantes.  $\beta_k$  es el parámetro que indica el cambio esperado de la respuesta  $Y$  por cambio unitario en  $x_k$ , cuando las otras  $x$ 's se mantienen constantes, y  $\epsilon$  es el error experimental asociado.

En el sistema de referencia se modeló el consumo de materia seca mediante un modelo de regresión lineal múltiple basado en las variables acústicas y utilizando la metodología de validación cruzada dejar-uno-afuera. Se implementó en Matlab y Weka, inspirado en el trabajo de investigación de Galli et al. [1]. El modelo lineal implementado obtuvo un desempeño equivalente al de [1], con  $R^2 = 0,928$  y  $RSME = 4,98$ .

En la Figura 3.4 se muestran los valores de  $DMI$  observados y los estimados por el modelo de regresión lineal. Los resultados indican que es posible estimar con un error bastante reducido los valores del  $DMI$  de ovejas en pastoreo, a partir de las variables acústicas consideradas mediante el modelo mencionado.

Sin embargo, hay fenómenos observables que no pueden ser explicados por modelos lineales, y es necesario un modelo no lineal en los parámetros para poder modelar

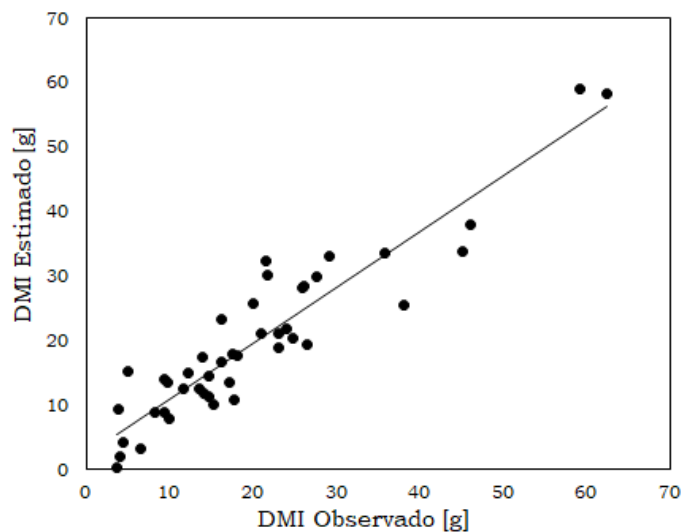


Figura 3.4: Relación entre el *DMI* estimado y observado basado en variables predictoras acústicas del modelo lineal de referencia.

el fenómeno. Un modelo de regresión no lineal puede expresarse como:

$$Y = f(X, \boldsymbol{\theta}) + \epsilon \quad (3.2)$$

donde  $Y$  es el valor medido de una o más respuestas (las variables dependientes),  $\epsilon$  es el error experimental asociado con esta medida, y  $f(X, \boldsymbol{\theta})$  es una función no lineal que contiene  $p$  parámetros desconocidos  $\theta_1, \dots, \theta_p$  y  $k$  variables  $x_1, \dots, x_k$ . Como mínimo, se pretende obtener los valores de los parámetros asociados con la mejor curva de ajuste (con el método de los mínimos cuadrados).

Como se mencionó en el Capítulo 1, las redes neuronales permiten lograr modelos no lineales adecuados en muchos casos de aplicación, por ejemplo: las MLP han sido usadas en la estimación de diversos problemas para predicción [31, 32, 13, 14, 15], las ELM han demostrado buenos resultados en problemas de regresión [16]. El desempeño predictivo se puede mejorar para ambas redes utilizando métodos de selección de características que permiten remover características poco relevantes y reducir el tiempo de entrenamiento.

## 3.2. Selección de características

A nivel conceptual, uno puede dividir la tarea del aprendizaje en dos subtareas: decidir cómo combinar esas características y decidir qué características (atributos) usar en el modelo. En la práctica, agregar atributos irrelevantes a un conjunto de datos a menudo confunde a los sistemas de aprendizaje automático. Debido a este efecto de los atributos irrelevantes, es común preceder al aprendizaje con una etapa de selección de atributos para eliminar todos los atributos excepto los más relevantes [33]. Según [34] la utilización de métodos de selección de características en un proceso de aprendizaje aporta algunas ventajas importantes entre las que destacan:

- Una mejora considerable en la eficacia del algoritmo de aprendizaje. En muchos algoritmos el costo en tiempo y en memoria cuando mantenemos constantes la capacidad computacional crece de forma superlineal al aumentar el número de características.
- Se disminuyen los costos asociados a la recopilación de datos. Muchas de las características con un bajo potencial descriptivo de la clase pueden ser ignoradas.
- Se mejora la precisión de los resultados obtenidos. Si una característica tiene irrelevancia conocida entonces podemos aumentar la calidad del modelo derivado si nuestro algoritmo parte de un conjunto relevante de atributos.

En la práctica, necesitamos definir:

- Cómo se construye el espacio de todos los posibles subconjuntos de atributos.
- Cómo se mide el rendimiento de aprendizaje para guiar la búsqueda y detenerla alcanzados ciertos criterios.
- Qué método de búsqueda utilizar, teniendo en cuenta que esta elección condiciona la complejidad computacional de la selección que podría volverse intratable al aumentar el número de atributos considerados.

Particularmente, para este trabajo, se cuenta con una base de datos de eventos masticatorios de ovejas conformada por 44 muestras representadas por 10 atributos cada

una. A priori, es posible decir que la cantidad de datos es reducida en relación a la cantidad de atributos. Es así que, luego de los primeros experimentos, fue posible comprobar esta premisa y por lo tanto se decidió explorar métodos de selección de características para mejorar el sistema. Existen dos enfoques principales que constituyen las dos familias de técnicas más utilizadas en la selección de características: los métodos de filtrado y los métodos de wrapper. Teniendo en cuenta la complejidad se seleccionó un método de cada familia. El primero es un evaluador basado en correlaciones: maximiza la correlación con la clase y minimiza la intercorrelación del subconjunto de características seleccionado [35, 36]. El segundo método, de tipo wrapper, usa un algoritmo de aprendizaje automático para evaluar los diferentes conjuntos que se obtienen con el método de búsqueda y con validación cruzada de 5 particiones.

### Método de filtro

El método utilizado fue el de selección de subconjuntos de características basadas en la correlación (CFS, del inglés *correlation-based feature selection*) [35]. CFS evalúa un subconjunto de atributos considerando la capacidad predictiva individual de cada característica junto con el grado de redundancia entre ellas. Se prefieren subconjuntos de características que están altamente correlacionadas con la clase mientras tienen baja intercorrelación. Se utiliza, de hecho, una correlación de Pearson donde todas las variables se han estandarizado. Los subconjuntos de características “buenas” contienen características altamente correlacionadas con (predictivas de) la clase, pero no correlacionadas (no predictivas) entre sí. Para CFS se utilizó el algoritmo de búsqueda “el mejor primero”, estrategia que permite realizar un seguimiento en la ruta de búsqueda. Se mueve a través del espacio de búsqueda al hacer cambios locales en el subconjunto de características actual. Si el camino que se está explorando comienza a parecer menos prometedor, este algoritmo de búsqueda puede retroceder a un subconjunto anterior más prometedor y continuar la búsqueda desde allí.

### Método de wrapper

Este método evalúa conjuntos de atributos mediante el uso de un esquema de aprendizaje. Utiliza la validación cruzada para estimar la precisión del esquema de

aprendizaje para un conjunto de atributos [36]. Los wrappers utilizan la máquina de aprendizaje de interés como una caja negra para anotar subconjuntos de variables de acuerdo con su poder predictivo. La metodología consiste en utilizar el rendimiento de predicción de una máquina de aprendizaje dada, para evaluar la utilidad relativa de los subconjuntos de variables. En este trabajo la máquina de aprendizaje utilizada fue el perceptrón multicapa. Para el wrapper se utilizó un algoritmo genético de búsqueda. Los algoritmos genéticos son técnicas de búsqueda adaptativa basadas en los principios de la selección natural en biología [37]. Emplean una población de soluciones competidoras, evolucionadas a lo largo del tiempo, para converger en una solución óptima. El espacio de la solución se busca en paralelo, lo que ayuda a evitar los óptimos locales. El algoritmo es un proceso iterativo en el que cada generación sucesiva se produce mediante la aplicación de operadores genéticos, como la cruce y la mutación, entre los miembros de la generación actual. La mutación cambia algunos de los valores al azar (agregando o eliminando características). La cruce combina diferentes características de un par de subconjuntos en un nuevo subconjunto. La aplicación de los operadores genéticos a los miembros de la población está determinada por su aptitud (qué tan bueno es un subconjunto de características). Los mejores subconjuntos de funciones tienen una mayor probabilidad de ser seleccionados para formar un nuevo subconjunto a través de cruces y mutaciones. De esta manera, los buenos subconjuntos “evolucionan” con el tiempo [38].

---

# Capítulo 4

## Experimentos y resultados

Este capítulo está dividido en cinco secciones. La Sección 4.1 explica la base de datos utilizada para los experimentos y los aspectos de implementación de los mismos. La Sección 4.2 detalla la metodología de evaluación de los modelos. La Sección 4.3 describe todos los experimentos realizados en la presente tesis. La Sección 4.4 presenta los resultados de los experimentos y por último se presentan las discusiones en la Sección 4.5.

### 4.1. Datos

Se utilizaron las variables acústicas de los datos obtenidos en [1]. Éstos corresponden a registros sonoros de tres ovejas mestizas de 2 a 4 años de edad, con un peso promedio de  $85 \pm 6,0$  kg, realizados en el Departamento de Ciencia Animal de la Universidad de California en Davis (UCDavis). Las ovejas realizaron sesiones cortas de pastoreo en cuatro tipos de pasturas puras (alfalfa alta, alfalfa corta, pasto ovillo alto y pasto ovillo corto) ofrecidas en microparcels. Para asegurar el muestreo, los animales fueron ayunados durante una hora previa al inicio de las sesiones de pastoreo. Cada sesión de pastoreo fue grabada usando una cámara digital estándar (Sony DCR-PC100 digital camcorder) y un sistema de micrófono inalámbrico (Nady Systems 151 VR) colocado contra la frente del animal (Figura 4.1). La ingesta de materia seca se estimó como la diferencia entre la biomasa forrajera de la sesión previa y posterior al pastoreo. Los potes se pesaron individualmente con precisión de 0,1 g usando una balanza digital Setra 140 CP. Se pesaron dos potes por ensayo

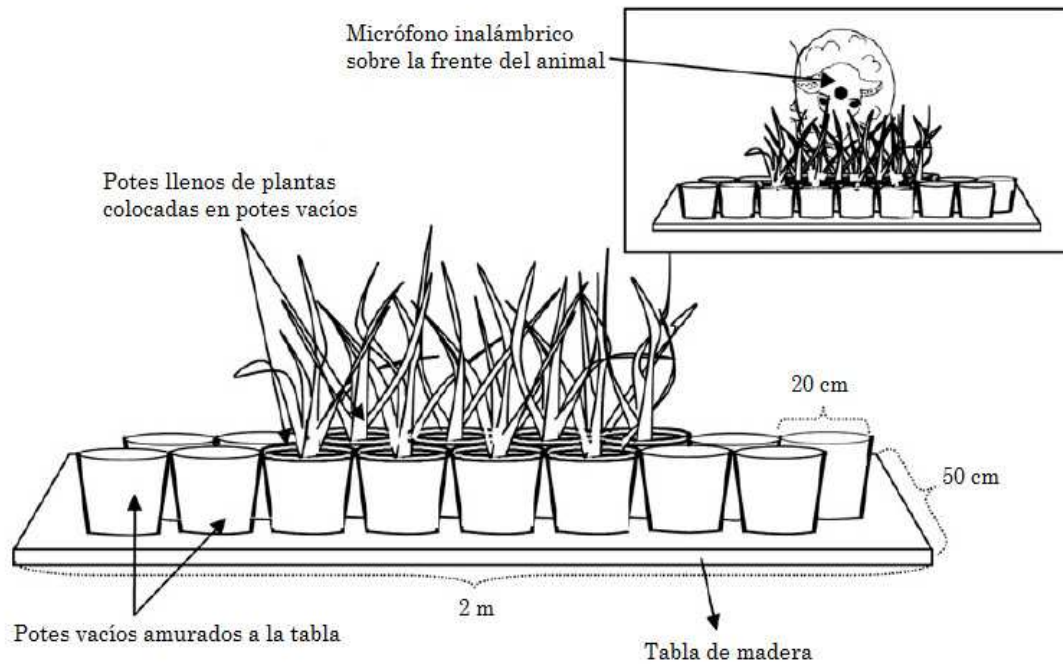


Figura 4.1: Ilustración esquemática de cómo se registran los sonidos (Galli et al 2011).

antes y después de cada sesión de pastoreo para estimar las pérdidas por evapotranspiración. Cada día se seleccionó al azar un subconjunto de potes de cada especie y altura para medir la altura del forraje en cinco hojas extendidas (en pasto ovillo) o tallos (en alfalfa).



### Aspectos de implementación

Las pruebas fueron realizadas en una notebook DELL Latitude E6230, Procesador Intel Core i5-3340M 2,70GHz, Memoria 8GB DDR III. Sistema Operativo Windows 10 Pro. La implementación de los algoritmos de todos los experimentos se realizó mediante la herramienta Matlab R2013b y el software Weka 3.6.12 (<http://www.cs.waikato.ac.nz>), que provee una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Este software fue desarrollado por la Universidad de Waikato, Nueva Zelanda. Las herramientas de Weka utilizadas fueron: para el modelo lineal de referencia: LinearRegression, para el MLP: MultilayerPerceptron. En la selección de características los métodos de evaluación utilizados fueron: CfsSubsetEval y WrapperSubsetEval, y para la búsqueda: BestFirst y GeneticSearch. Para el ELM se utilizó código implementado en Matlab para un regresor basado en ELM [17].

## 4.2. Metodologías de evaluación y validación

### Medidas para la evaluación del desempeño

Para medir el desempeño de cada prueba se utilizó el estadístico R cuadrado, que se basa en la correlación

$$R^2 = \frac{\sigma_{y_i y'_i}}{\sigma_{y_i} \sigma_{y'_i}},$$

y la raíz cuadrada del error cuadrático medio,

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(y_i - y'_i)^2}{N}}$$

donde  $y_i$  son las salidas deseadas,  $y'_i$  son las estimaciones,  $\sigma_{y_i y'_i}$  es la covarianza de las salidas deseadas y las estimadas,  $\sigma_{y_i}$  es la desviación estándar de la salida deseada,  $\sigma_{y'_i}$  es la desviación estándar de la salida estimada y  $N$  es el número de instancias o casos de prueba.

## Metodología de validación

### Cruzada

La metodología de validación cruzada divide al azar el conjunto de datos en  $k$  subconjuntos (las particiones)[39]. Luego se unen los  $k - 1$  subconjuntos, en un nuevo conjunto de entrenamiento que se utiliza para entrenar el modelo y así determinar los parámetros del mismo, y la parte que queda la utiliza para probar el rendimiento del modelo. Este conjunto es también llamado subconjunto de validación, y que se utiliza para estimar el error de generalización, es decir, una estimación del desempeño que tendrá el predictor con datos diferentes a los utilizados en el proceso de entrenamiento. El proceso de validación cruzada es repetido durante  $k$  iteraciones para probar con los  $k$  subconjuntos.

### Validación dejando uno afuera

La validación cruzada dejando uno afuera (LOO-CV, del inglés *Leave-one-out cross-validation*) es un caso especial de validación cruzada donde el número de particiones es igual al número de instancias en el conjunto de datos. Por lo tanto, el algoritmo de aprendizaje se aplica una vez para cada instancia, utilizando todas las demás instancias como conjunto de entrenamiento y utilizando la instancia seleccionada como conjunto de prueba de un solo elemento [40]. En la Figura 4.2 se ilustra el esquema descrito anteriormente.

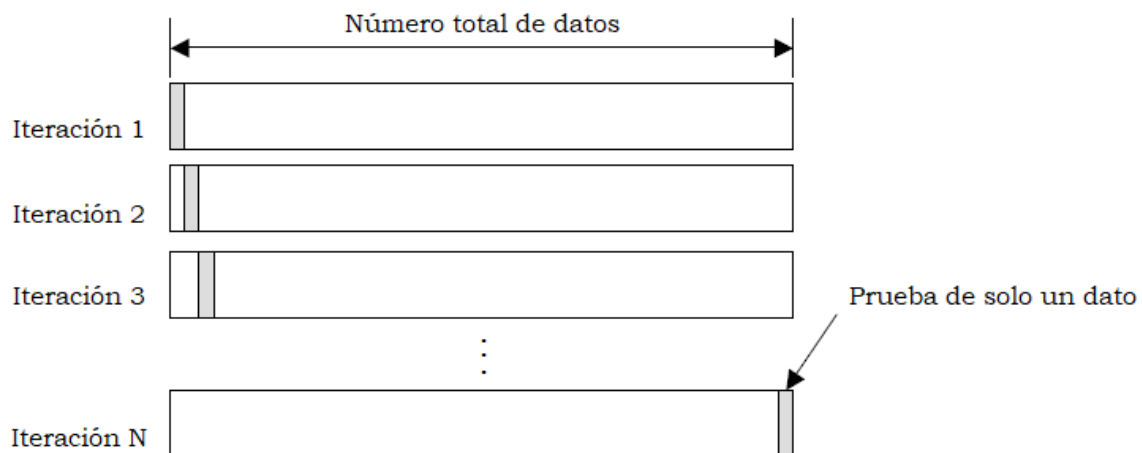


Figura 4.2: Esquema de validación cruzada LOO.

## 4.3. Experimentos

### 4.3.1. Redes neuronales

#### Validación cruzada y determinación de hiperparámetros

Para este experimento se utilizaron las redes MLP y ELM, donde se generaron particiones para el entrenamiento y prueba con los 44 registros utilizados en [1]. Debido a que la cantidad de datos disponibles es relativamente pequeña en relación a la cantidad de parámetros ajustables de las redes neuronales, se generó mayor cantidad de datos, tanto de entrenamiento como de prueba, utilizando un método de remuestreo. En una primera etapa se generaron 20 particiones madre, sólo con datos reales (esto es, sin datos generados por remuestreo). Se definió un tamaño de partición 70/30, lo que resulta en 31 registros para entrenamiento y 13 para prueba. En una segunda etapa se utilizó un método de remuestreo [38] que agregó a los datos reales, nuevos datos artificiales a partir del promedio de algunos patrones reales tomados al azar. Con este método se generaron, en cada partición de entrenamiento madre, 40 particiones de entrenamiento y prueba con relación 70/30, pero con 140 registros para entrenamiento y 60 registros para prueba. En la Figura 4.3 se ilustra el esquema de particiones descrito anteriormente y utilizado particularmente en este experimento.

Para realizar el ajuste de los hiperparámetros, tanto del MLP como del ELM, se utilizaron las 40 particiones de remuestreo. Luego, con los hiperparámetros determinados, para el ajuste de los pesos se utilizó la partición de entrenamiento madre sin remuestreo. Para el modelo MLP se optimizaron los siguientes parámetros de entrenamiento:

- Neuronas en la capa oculta: 1, 3, 5, 10 y 11.
- Tasa de aprendizaje: 0,15, 0,30 y 0,60.
- El término de momento: 0,15, 0,30 y 0,60.
- Épocas: 200 y 500.

Para el modelo ELM sólo se optimizó la cantidad de nodos en la capa oculta considerando: 100, 500, 1000, 1500 y 2000 neuronas.

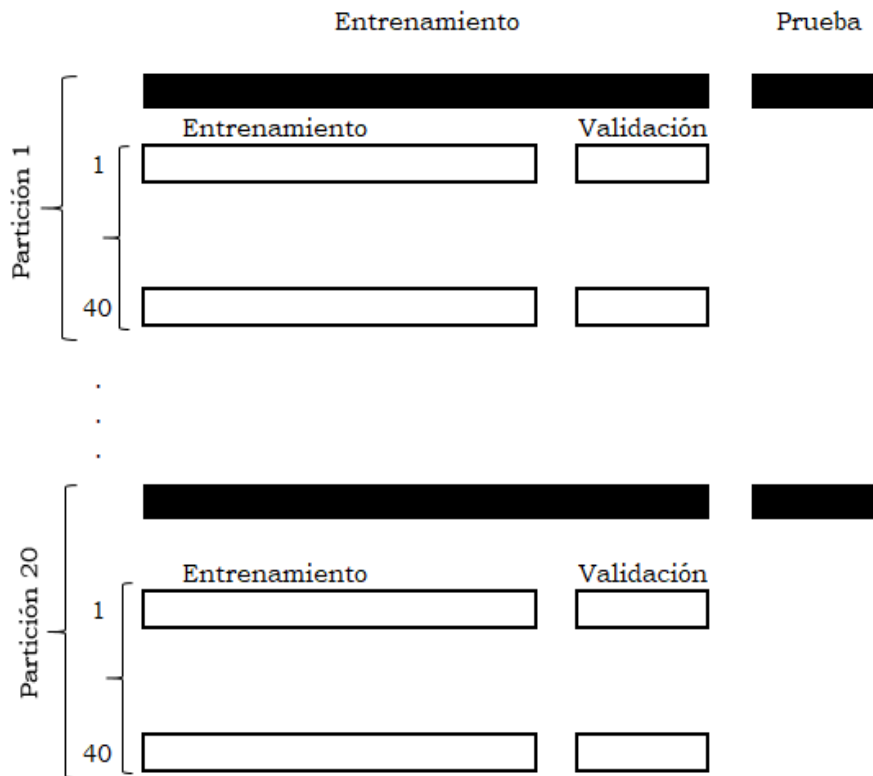


Figura 4.3: Esquema de particiones para la fase de entrenamiento y la de prueba.

### Validación cruzada con LOO

En este experimento se probaron los modelos MLP y ELM con la metodología de LOO-CV y la mejor configuración encontrada para cada red en los experimentos detallados en la sección anterior [41]. Para el modelo MLP se utilizaron los siguientes hiperparámetros:

- Neuronas en la capa oculta: 5.
- Tasa de aprendizaje: 0,15.
- El término de momento: 0,15.
- Épocas: 200.

Para el modelo ELM se utilizaron 2000 neuronas en la capa oculta.

Luego de observar los resultados de la ELM se realizó otra prueba, incorporando como hiperparámetro la no linealidad de la función de activación. Se utilizaron las funciones: lineal y sigmoide.

El esquema del experimento para ambas redes consta de un script principal con LOO-CV para el ajuste de los pesos, donde se invoca a otro script que realiza la

optimización de los parámetros, con la partición de entrenamiento correspondiente y también en un esquema interno de LOO-CV.

### 4.3.2. Selección de características

Se realizaron cuatro experimentos con selección de características, para poder evaluar el desempeño de los modelos con validación cruzada con LOO, con menos características predictoras de la materia seca consumida. En las siguientes subsecciones se describen los mismos.

#### Método de selección basado en la correlación

Se utilizó la red MLP con la metodología de LOO-CV y el algoritmo de selección de atributos basado en la correlación con búsqueda de “*el mejor primero*”. El esquema del experimento consta de un script principal del MLP con LOO-CV, donde se invoca a otro script que realiza, con los datos de entrenamiento correspondientes, la optimización de los siguientes parámetros por LOO-CV:

- Neuronas en la capa oculta: 1, 3, 5, 10 y 11.
- Tasa de aprendizaje: 0,15, 0,30 y 0,60.
- Término de momento: 0,15, 0,30 y 0,60.
- Épocas: 200 y 500.

A su vez, en cada caso se llama a un tercer script donde se lleva a cabo la selección de características con la metodología de validación cruzada de 5 particiones.

Para el caso de la red ELM también se utilizó LOO-CV, con el algoritmo de selección de atributos basado en la correlación y búsqueda según “*el mejor primero*”. En el proceso de selección de características se utilizó la metodología de validación cruzada de 5 particiones. El esquema del experimento consta de un script principal de la ELM con LOO-CV, donde se invoca a otro script que realiza la optimización de los siguientes parámetros con LOO-CV:

- Neuronas en la capa oculta: 100, 500, 1000, 1500 y 2000.
- Funciones de activación: lineal y sigmoide.

Desde aquí se llama a un tercer script donde se lleva a cabo la selección de características con la metodología de validación cruzada de 5 particiones.

### Método de selección basado en algoritmos genéticos

En este caso se utilizó la red MLP con LOO-CV y el algoritmo de selección de atributos wrapper con búsqueda genética. En el proceso de selección de características se utilizó la metodología de validación cruzada de 5 particiones para medir la aptitud de cada conjunto de atributos evaluado en la evolución. De forma similar al caso anterior, el esquema del experimento consta de un script principal del MLP con LOO-CV, donde se invoca a un script que realiza la optimización de los siguientes parámetros con LOO-CV:

- Neuronas en la capa oculta: 1, 3, 5, 10 y 11.
- Tasa de aprendizaje: 0,15, 0,30 y 0,60.
- Término de momento: 0,15, 0,30 y 0,60.
- Épocas: 200 y 500.

Mediante la llamada a un tercer script se lleva a cabo la selección de características con la metodología de validación cruzada de 5 particiones.

Para el caso de la red ELM, también con LOO-CV, se utilizó en este experimento la misma selección de atributos wrapper con búsqueda genética, con validación cruzada interna de 5 particiones. Como en los casos anteriores aquí también se desarrolló un script principal de la ELM con LOOCV, desde donde se invoca a un script que realiza la optimización de los siguientes parámetros con LOO-CV:

- Neuronas en la capa oculta: 100, 500, 1000, 1500 y 2000.
- Funciones de activación: lineal y sigmoide.

Un tercer script interno lleva a cabo la selección de características con la metodología de validación cruzada de 5 particiones.

En la Figura 4.4 se ilustra el esquema utilizado en los experimentos explicados en esta sección. El script principal hace LOO-CV dejando 1 caso para prueba y usando 43 para entrenamiento. De estos 43 de entrenamiento, se realiza la optimización

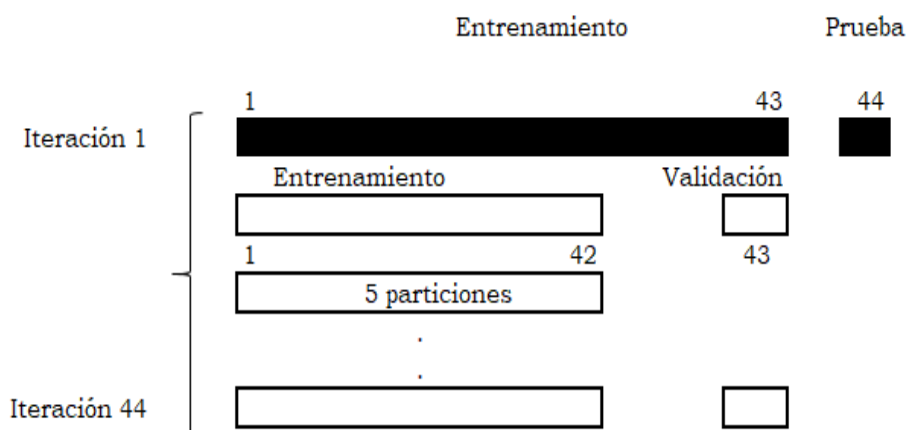


Figura 4.4: Esquema de validación cruzada para los experimentos de la Sección 4.3.2.

de parámetros con LOO-CV dejando 1 caso para prueba y 42 para entrenamiento. El tercer script, para cada partición de optimización, hace validación cruzada de 5 particiones con los 42 casos de entrenamiento.

## 4.4. Resultados y discusión

### 4.4.1. Redes neuronales

#### Validación cruzada

En la Tabla 4.4.1 se muestra por columna el desempeño de cada método en la fase de prueba, medido por correlación y error cuadrático medio. En cada fila se observan las 20 particiones de prueba para ambos métodos. En la anteúltima fila se muestra la media de los valores por columna y en la última fila el desvío por cada columna. Considerando el valor promedio de ambas medidas podemos decir que el modelo MLP fue mejor que ELM para predecir la *DMI*, aunque requirió un mayor tiempo de aprendizaje, del orden de 60 veces más que las ELM. Los resultados a los que se abordó con este experimento fueron publicados en [41].

Partición	$R^2$		$RMSE$	
	MLP	ELM	MLP	ELM
1	0,87	0,67	9,52	9,98
2	0,93	0,76	6,32	12,19
3	0,91	0,74	4,22	9,06
4	0,89	0,66	6,84	10,96
5	0,83	0,62	9,58	13,03
6	0,90	0,64	6,44	10,55
7	0,90	0,87	7,03	5,62
8	0,74	0,65	12,25	9,05
9	0,81	0,65	8,65	11,77
10	0,78	0,92	5,68	3,65
11	0,95	0,56	7,37	11,85
12	0,83	0,79	7,39	8,33
13	0,86	0,83	6,75	5,43
14	0,95	0,67	7,85	8,79
15	0,88	0,52	5,65	13,04
16	0,87	0,28	8,09	11,12
17	0,75	0,68	6,78	9,67
18	0,95	0,42	6,40	15,35
19	0,67	0,63	6,96	6,93
20	0,98	0,47	6,05	16,13
$\bar{x}$	0,86	0,65	1,74	10,13
$\sigma$	0,08	0,15	7,29	3,20

Para el MLP los hiperparámetros que mejor desempeño tuvieron en la fase de entrenamiento fueron: 5 neuronas en la capa oculta, 0,15 para la velocidad de aprendizaje, término de momento en 0,15 y 200 épocas de entrenamiento. Para el ELM el mejor desempeño se obtuvo con 2000 neuronas. A continuación se analizan estos resultados con mayor detalle. En la Figura 4.5 se muestran los valores de  $DMI$  observados y los estimados por el MLP (a la izquierda) y la ELM (a la derecha), para cada una de las 13 muestras del conjunto de prueba. Podemos ver que ambos modelos logran una buena aproximación de los valores  $DMI$ .

La Figura 4.6 compara los valores de  $DMI$  observados y estimados por el MLP y el ELM para los casos de más alta correlación según la Tabla 4.1. La nube de



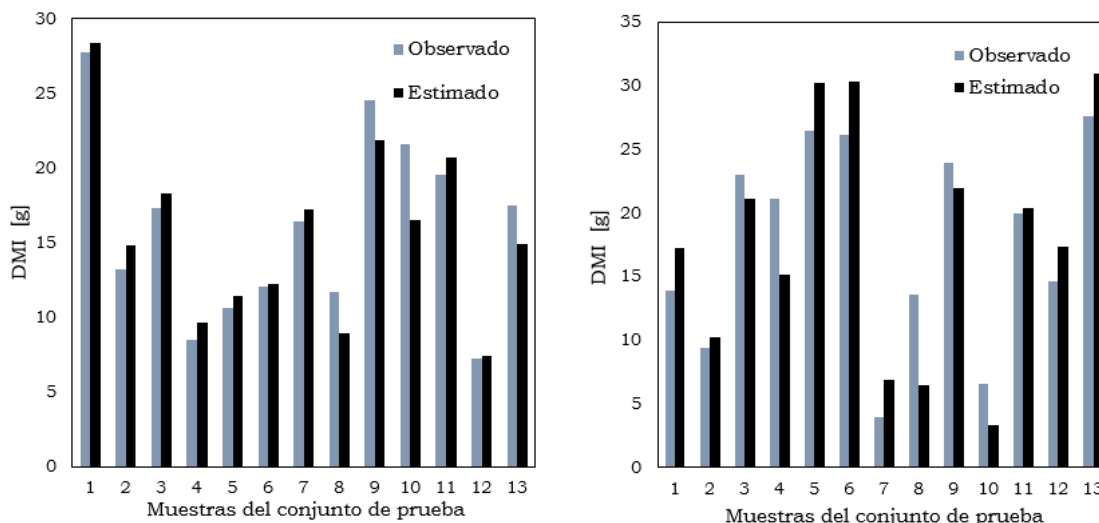


Figura 4.5: *DMI* estimado y observado con MLP (izquierda) y los resultados con ELM (derecha).

puntos se observa cercana a la recta indicando una correlación positiva y alta, lo que nos muestra la bondad de estimación obtenida con ambos modelos. En el caso particular de las ELM, coincidió que el valor más alto de correlación tiene asociado el menor valor de error, situación que no siempre sucede.

Para una mejor comparación con el modelo lineal del trabajo de Galli et al. (2011), se presentan en la Figura 4.7 ambos modelos de red neuronal, con los hiperparámetros finalmente seleccionados, graficando los 44 registros originales. Como se puede observar, se obtuvieron resultados muy buenos a partir del grupo de datos originales. Con MLP se obtuvo  $R^2 = 0,96$  y un  $RMSE = 3,76$  y con ELM un  $R^2 = 0,99$  y un  $RMSE = 1,33$ .

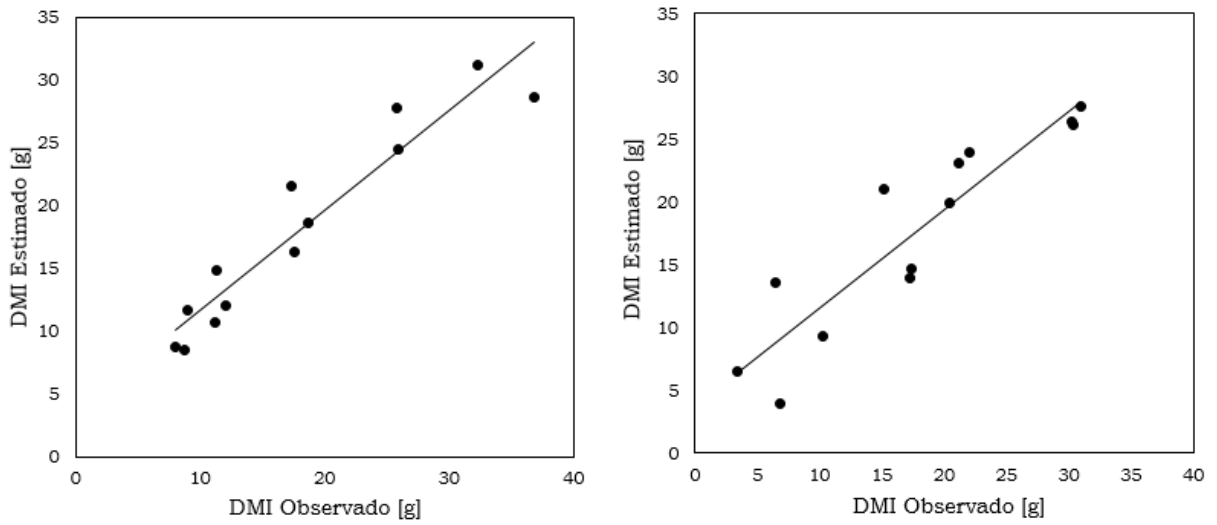


Figura 4.6: Relación entre el *DMI* estimado y observado para el MLP (izquierda), y para el ELM (derecha). Se muestran las particiones 20 y 10 de la Tabla 4.1, para MLP y ELM respectivamente.

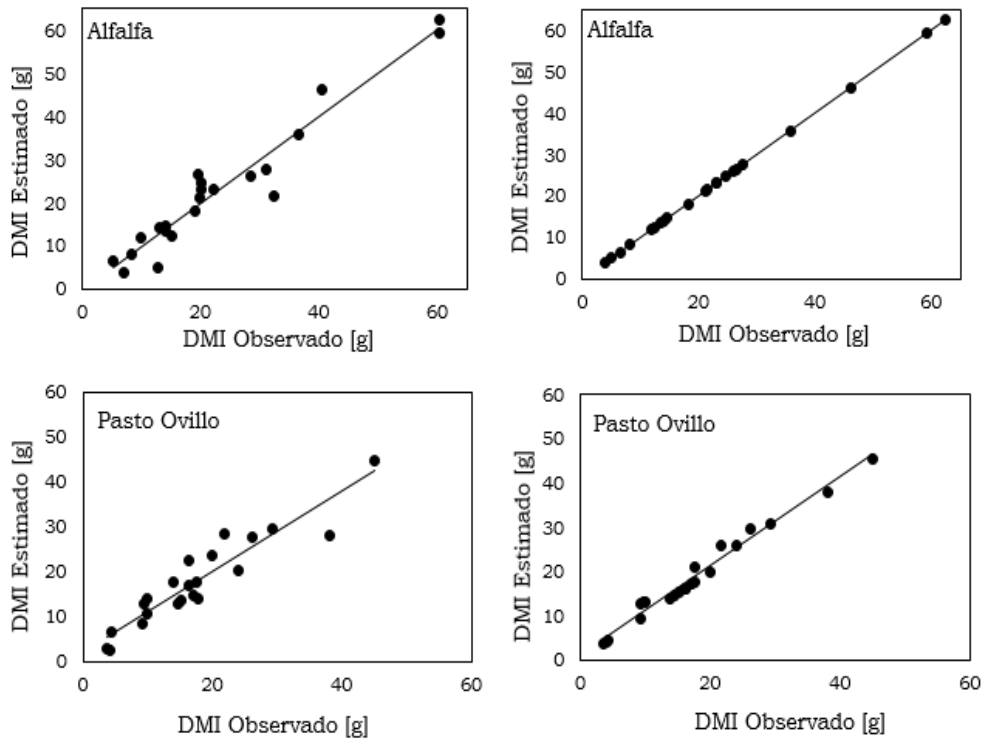


Figura 4.7: Relación entre *DMI* observado y estimado, a la izquierda para el MLP y a la derecha para el ELM con todos los datos originales.

Tabla 4.1: Valores de desempeño de los tres modelos del experimento utilizando LOO-CV.

	LOO-CV	
	$R^2$	$RMSE$
Regresión Lineal	0,92	4,9
MLP	0,90	6,12
ELM-Sigmoidea	0,59	11,01
ELM-Lineal	0,92	5,27

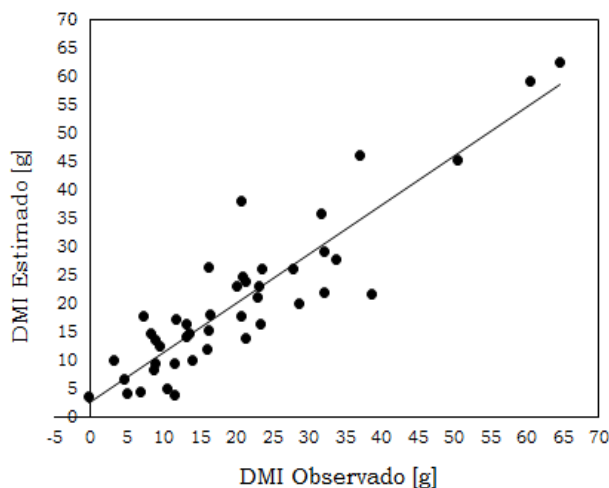


Figura 4.8: Relación entre  $DMI$  observado y estimado con MLP.

### Validación por LOO

En la Tabla 4.1 se muestra por columna el desempeño de cada método, medido por la correlación y el error cuadrático medio. En cada fila se observan los cuatro modelos que utilizaron el mismo esquema de validación. Como se puede observar, el desempeño del modelo ELM-Lineal es prácticamente equivalente al del modelo de referencia con regresión lineal en cuanto al  $R^2$  y presenta algo más de error en términos de  $RMSE$ .

La Figura 4.8 muestra los valores de  $DMI$  observados y estimados por el MLP con LOO-CV. Como es visible en esta figura los valores predichos por el modelo están muy cerca a los datos observados.

La Figura 4.9 compara los valores de  $DMI$  observados y estimados por la ELM-Sigmoidea con LOO-CV. La nube de puntos se observa bastante dispersa lo que se condice con la correlación baja de la Tabla 4.1. En la Figura 4.10 se observa una mayor correlación, lo que también se condice con los mejores resultados de ELM-Sigmoidea en la Tabla 4.1.

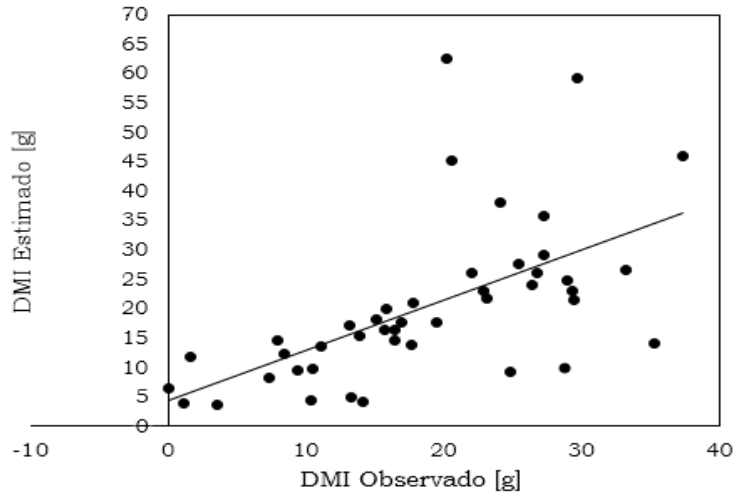


Figura 4.9: Relación entre *DMI* observado y estimado con ELM-Sigmoide.

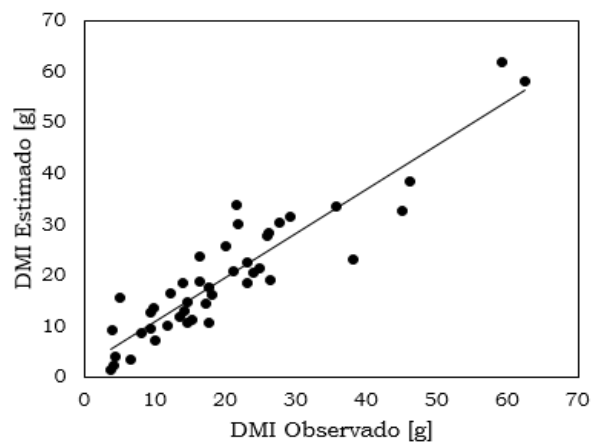


Figura 4.10: Relación entre *DMI* observado y estimado con ELM-Lineal.

#### 4.4.2. Selección de características

En la Tabla 4.2 se muestra por columna el desempeño de CFS y WRAPPER, medido por la correlación y el error cuadrático medio. Por fila se muestra el desempeño de cada red utilizada. Los resultados muestran una mejora significativa en el desempeño para las ELM en combinación con el método evolutivo para la selección de características.

Tabla 4.2: Valores de correlación y  $RMSE$  discriminados por red y método de evaluación.

	CFS		WRAPPER	
	$R^2$	$RMSE$	$R^2$	$RMSE$
MLP	0,58	10,86	0,88	6,28
ELM	0,88	6,32	0,94	3,18

En la Tabla 4.3 se muestran por columna las 9 características que fueron sometidas a los dos métodos de selección. En las filas, para cada método, se muestra el porcentaje de veces que fue seleccionada cada característica a lo largo de todos los experimentos.

Tabla 4.3: Atributos seleccionados a lo largo de todos los experimentos.

	Specie	Heigth	$VC$	$EC$	$ECc$	$Et$	$VB$	$EBb$	$ECb$
ELM-WRAPPER	34 %	89 %	2 %	100 %	16 %	11 %	52 %	9 %	75 %
ELM-CFS	55 %	55 %	100 %	100 %	-	100 %	-	-	-
MLP-WRAPPER	100 %	100 %	-	100 %	-	-	100 %	100 %	100 %
MLP-CFS	100 %	100 %	100 %	-	-	-	-	-	-

Los experimentos con selección de características muestran que el método de selección por algoritmos genéticos con ELM, que provee los mejores resultados en la Tabla 4.2, selecciona la mayor parte de las veces a la altura (Height, con 89 %), la energía total de densidad de flujo ( $EC$ , con 100 %, es decir, en todos los casos), y la energía de densidad de flujo de masticación por arranque ( $ECb$ ) en un 75 % de las veces.

### Método basado en la correlación

La Figura 4.11 compara los valores de *DMI* observados y estimados por el MLP con LOO-CV. La nube de puntos se observa cercana a la recta pero más dispersa que en los otros experimentos.

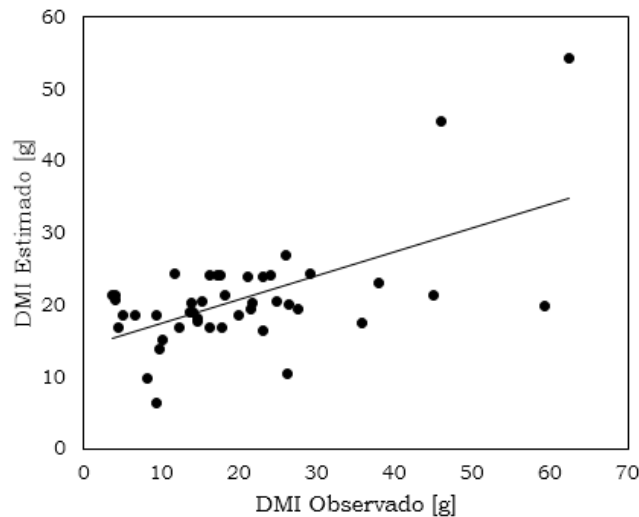


Figura 4.11: *DMI* observado y estimado MLP con CFS.

En este experimento los hiperparámetros que mejor desempeño tuvieron en la fase de entrenamiento fueron: 1 neurona en la capa oculta, 0,15 para la velocidad de aprendizaje, término de momento en 0,15 y 500 épocas de entrenamiento.

Para el caso de ELM, la Figura 4.12 compara los valores de *DMI* observados y estimados en la validación con LOO-CV. Se aprecia una correlación positiva, con una nube de puntos cercanos a la recta de máxima correlación.

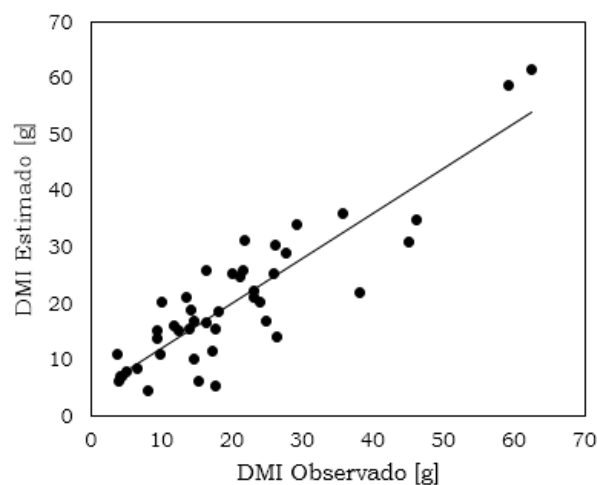


Figura 4.12: *DMI* observado y estimado ELM con CFS.

Para este experimento los hiperparámetros que mejor desempeño tuvieron en la fase de entrenamiento fueron las funciones de activación lineales y 100 neuronas en

la capa oculta.

### Método de selección con algoritmo genético

La Figura 4.13 compara los valores de *DMI* observados y estimados por el MLP con LOO-CV. La nube de puntos se observa más cercana a la recta que en el experimento anterior (Figura 4.12).

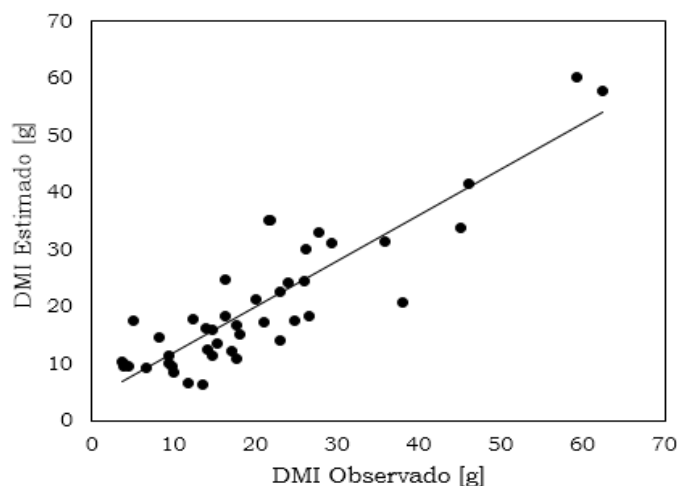


Figura 4.13: *DMI* observado y estimado MLP con WRAPPER.

Para el MLP los hiperparámetros que mejor desempeño tuvieron en la fase de entrenamiento fueron: 1 neurona en la capa oculta, 0,15 para la velocidad de aprendizaje, término de momento en 0,15 y 500 épocas de entrenamiento.

En el caso de las ELM, la Figura 4.14 compara los valores de *DMI* observados y estimados en el esquema de LOO-CV. La nube de puntos se observa cercana a la recta, lo que también indica una alta correlación positiva.

Para este experimento los hiperparámetros que mejor desempeño tuvieron en la fase de entrenamiento coinciden con el experimento que utiliza la red ELM con el selector por correlación: funciones de activación lineales y 100 neuronas en la capa oculta.

En la Tabla 4.4 se puede observar un resumen de los resultados más relevantes de este capítulo. Se puede observar que el método que mejor desempeño tuvo fue el de ELM con selección de características por algoritmos genéticos, con una correlación de 0,94. En este caso el conjunto de características más elegido fue de cuatro: altura, energía total de densidad de flujo de masticación (*EC*), intensidad de arranque (*VB*) y energía de densidad de flujo de masticación por arranque (*ECb*).

Mediante el experimento detallado la Sección 4.4.1 se logró reproducir un modelo de regresión lineal inspirado en el trabajo de Galli et al [1]. De esta manera se obtuvo un modelo de referencia para la comparación con los modelos de los siguien-

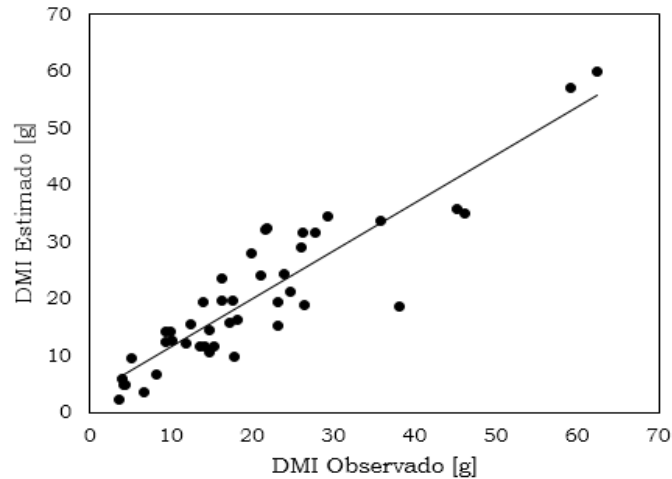


Figura 4.14: *DMI* observado y estimado ELM con WRAPPER.

Tabla 4.4: Resumen de resultados de todos los experimentos.

	$R^2$	$RMSE$
Modelo Lineal de Referencia	0,92	4,9
MLP	0,90	6,12
ELM-Lineal	0,92	5,27
MLP-CFS	0,58	10,86
MLP-WRAPPER	0,88	6,28
ELM-Lineal-CFS	0,88	6,32
ELM-Lineal-WRAPPER	0,94	3,18

tes experimentos, respetando el método de validación empleado en la publicación original.

En los experimentos de validación por LOO, cuyos resultados se presentaron en la Subsección 4.4, no se logra una gran mejora de los resultados utilizando los modelos MLP y ELM. Se aprecia una mejora en el modelo ELM al utilizar una función lineal en la capa oculta, en lugar de una función sigmoide, pero los resultados no superan al modelo lineal de referencia. Si bien se esperaba una mejora con la selección de características, al utilizar el método basado en la correlación los resultados con la red neuronal MLP no fueron buenos. Sin embargo, cuando se utilizaron las ELM se pudo apreciar una mejora importante en el desempeño. El experimento con el método evolutivo para la selección de características mejoró los resultados anteriores. Con MLP los resultados ya fueron más competitivos, aunque no mejores que la referencia del modelo lineal. Por último, el experimento con el método evolutivo para la selección de características y las ELM fue el que logró el mejor desempeño de todos los modelos estudiados en este trabajo de tesis: con el 50 % menos de las variables acústicas se obtuvo un  $R^2 = 0,94$  y  $RMSE = 3,18$ .



---

## Capítulo 5

# Conclusiones y trabajo futuro

En esta tesis se abordó el problema de estimación del consumo de rumiantes en pastoreo, mediante la utilización de dos redes neuronales actuando como modelos de regresión no lineal. Las principales conclusiones de esta tesis se resumen en los siguientes puntos:

- Se modeló el consumo de materia seca mediante un modelo de regresión lineal múltiple basado en las variables acústicas y utilizando la metodología de validación cruzada leave-one-out. Se implementó el modelo de referencia en Matlab y Weka, inspirado en el trabajo de investigación de Galli et al. [1]. Se logró obtener un modelo similar, en cuanto a las variables predictoras y los coeficientes, con el mismo valor de correlación y un error cuadrático medio equivalente. Ambos modelos basados en variables acústicas incluyeron la energía total de densidad de flujo de masticación y la energía de densidad de flujo de masticación por arranque, con iguales valores en sus coeficientes. Este resultado permite concluir que el método base puede ser tomado como referencia de modelo lineal para futuras comparaciones.
- Se presentaron dos modelos con dos tipos de redes neuronales artificiales que abordan el problema de la automatización de la estimación del consumo de rumiantes. Los resultados que se presentan indican que es posible estimar el consumo de materia seca en el pastoreo de ovejas mediante regresión no lineal multivariada utilizando redes neuronales artificiales, y además apoyan la investigación de trabajos previos sobre la riqueza de la información contenida en el sonido de las masticaciones de las ovejas al ingerir forraje.
- Con la inclusión de los algoritmos de selección de características, habiendo probado con distintas combinaciones de redes, métodos de evaluación y búsqueda, se logró obtener un modelo que no sólo tuvo el mejor desempeño por sobre todos los experimentos realizados anteriormente sino que además la complejidad se redujo a cuatro variables acústicas de las diez que fueron analizadas

en los distintas pruebas. Dentro de las variables acústicas, en este modelo se seleccionaron: la intensidad de arranque, la energía total de densidad de flujo de masticación, la energía de densidad de flujo de masticación por arranque y la altura. Es importante destacar que las tres primeras variables mencionadas fueron también las incluidas en los modelos lineales de referencia.

Como trabajo futuro se planea probar los modelos de mejor desempeño con variables acústicas obtenidas de otros rumiantes. En particular, ya se está confeccionando una base de datos de señales de vacas. Además, sería de interés extender los resultados obtenidos a señales de mayor duración en el tiempo, para realizar estimaciones de consumos que puedan abarcar períodos de días hasta semanas. Dado que todos los experimentos fueron realizados con pocos datos se planea explorar también la regularización en las ELM y alternativas para la normalización de los datos de entrada.

---

# Bibliografía

- [1] J.R. Galli, C.A. Cangiano, D.H. Milone, and E. Laca. Computational method for segmentation and classification of ingestive sounds in sheep. *Livestock Science*, 140,issues 1–3:32—412, 2011.
- [2] D. H. Milone, H. L. Rufiner, J. R. Galli, Laca E. A., and Cangiano C. A. Computational method for segmentation and classification of ingestive sounds in sheep. *Computers and Electronics in Agriculture, Elsevier*, 65:228–237, 2009.
- [3] J. Galli, H. Laforcada, D. Milone, and C.A. Cangiano. Evaluación de un programa de computación para la clasificación automática de sonidos masticatorios. *Rev. Arg. Prod. Anim.*, 25:190–191, 01 2005.
- [4] M. Gibbs and R. Orr. Grazing behaviour of ruminants. *IGER Innovations*, pages 53–57, 1997.
- [5] J. M. Luginbuhl, K. Pond, and J. Burns. A simple electronic device and computer interface system for monitoring chewing behaviour of stall-fed ruminant animals. *In Journal Dairy Science*, 70:1307–1312, 1987.
- [6] Y. Alkon, Cohen and P. Jordan. A simple electronic device and computer interface system for monitoring chewing behaviour of stall-fed ruminant animals. *In Journal Wildl. Manage*, 53:658–662, 1989.
- [7] E. A. Laca and M. F. Wallis De Vries. Acoustic measurement of intake and grazing behaviour of cattle. *In Grass Forage Sci*, 55:97–104, 2000.
- [8] J.R. Galli, C.A. Cangiano, M. Demment, and E.A. Laca. Grazing behaviour of ruminants. *Animal Feed Science and Technology, Elsevier*, 128:14—30, 2006.
- [9] M.F. WallisDeVries, E.A. Laca, and M.W. Demment. From feeding station to patch: scaling up food intake measurements in grazing cattle. *Applied Animal Behaviour Science, Elsevier*, 60:301—315, 1998.
- [10] J.R. Galli, Pece M.A. Cangiano, C.A., M.J. Larripa, and E.A. Laca. Uso del sonido en el análisis de la tasa de consumo de bovinos. *Revista Argentina de Producción Animal*, 26:165—167, 2006.

- [11] E. D. Ungar and S. Rutter. Classifying cattle jaw movements: Comparing iger behaviour recorder and acoustic techniques. *Applied Animal Behaviour Science, Elsevier*, 98:11–27, 2005.
- [12] J. R. Galli, C. A. Cangiano, M. A. Pece, M. J. Larripa, D. H. Milone, S. A. Utsumi, and E. A. Laca. Monitoring and assessment of ingestive chewing sounds for prediction of herbage intake rate in grazing cattle. *animal*, 12(5):973–982, 2018.
- [13] F. A. Ruslan, A. M. Samad, M. Tajuddin, and R. Adnan. Flood prediction modeling using improved *MLPNN* structure: Case study kuala lumpur. *IEEE Conference on Systems, Process and Control*, 2016.
- [14] C. Mena Frau and R. Montecinos Guajardo. Comparación de redes neuronales y regresión lineal para estimar productividad de sitio en plantaciones forestales, utilizando geomática. *Bosque*, 27:35–43, 2006.
- [15] Durairaj M. and Revathi V. Prediction of heart disease using back propagation *MLP* algorithm. *International Journal of Scientific & Technology Research*, 4, issue 08:ISSN 2277–8616, 2015.
- [16] G-B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems Man, and Cybernetics-Part B: Cybernetics*, 42, 2012.
- [17] G. Huang, Q. Zhu, Siew, and C. Extreme learning machine: Theory and applications. *Neurocomputing*, 70:489—501, 2006.
- [18] G. Vergara, J. Cózar, C. Romero-González, J.A. Gámez, Soria-Olivas, and E. Soria-Olivas. Comparing *ELM* against *MLP* for electrical power prediction in buildings. In: *Ferrández Vicente, J.M., Álvarez Sánchez, J.R., de la Paz López, F., Toledo-Moreo, F.J., Adeli, H. (eds.) Bioinspired Computation in Artificial Systems, Lecture Notes in Computer Science, Springer International Publishing*, 9108:409—418, 2015.
- [19] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Cornell Aeronautical Laboratory, Psychological Review*, 65:386–408, 1958.
- [20] G. Cybenko. Approximations by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 1989.
- [21] Hornik. Kurt. Approximation capabilities of multilayer feedforward networks. *neural Networks*, 4(2):251—257, 1991.

- 
- [22] R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 1987.
- [23] I. Chauvin and D. E. Rumelhart. Backpropagation: Theory, architectures, and applications. *Lawrence Erlbaum Associates, Inc., Publishers*, 1995.
- [24] C.R. Rao and S.K. Mitra. Generalized inverse of matrices and its applications. *Wiley, New York*, 1971.
- [25] D. Serre. Matrices: Theory and applications. *Springer, New York*, 2002.
- [26] G.-B. Huang and H.A. Babri. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Trans. Neural Networks*, 9:224—229, 1998.
- [27] J. D. Hood. Bone conduction: A review of the present position with especial reference to the contributions of dr. georg von békésy. *The Journal of the Acoustical Society of America*, 34(9B):1325–1332, 1962.
- [28] B. Hakansson, P. Carlsson, A. Brandt, and S. Stenfelt. Linearity of sound transmission through the human skull in vivo. *The Journal of the Acoustical Society of America*, 99:2239—2243, 1996.
- [29] J.B. Robertson and P.J. Van Soest. The detergent system of analysis and its application to human foods. in: James, w.p.t., theander, o. (eds.). *The Analysis of Dietary Fiber in Foods. Marcel Dekker Inc, New York, USA*, pages 123—158, 1980.
- [30] Syntrillium Software Corporation. Cool edit pro version 2. user’s manual. *Syntrillium Software Corporation, Phoenix, AZ, USA*, 2002.
- [31] J. Parinet, M. Julien, P. Nun, R. J. Robins, G. Remaud, and P. Höhener. Predicting equilibrium vapour pressure isotope effects by using artificial neural networks or multi-linear regression – a quantitative structure property relationship approach. *Chemosphere, Elsevier*, 134:521—527, 2015.
- [32] S. Azadi and A. Karimi-Jashni. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of fars province, iran. *Waste Management, Elsevier*, 48:14—23, 2016.
- [33] I. H. Witten, E. Frank, and M. A. Hall. Data mining: Practical machine learning tools and techniques morgan kaufmann. *Elsevier*, 2005.

- 
- [34] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
- [35] M. A. Hall. Correlation-based feature subset selection for machine learning. *Hamilton, New Zealand*, 1998.
- [36] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.
- [37] J. H. Holland. Adaptation in natural and artificial systems. *University of Michigan Press, Ann Arbor, MI*, 1975.
- [38] D. E. Goldberg. Genetic algorithms in search, optimization, and machine learning reading. *MA: Addison-Wesley*, 1989.
- [39] T. M. Cover. Learning in pattern recognition. in satoshi watanabe, editor, methodologies of pattern recognition. *Academic press, New York*, pages 111–132, 1969.
- [40] Sammut C. and Webb G.I. Leave-one-out cross-validation. *Encyclopedia of Machine Learning. Springer, Boston, MA*, 2011.
- [41] M. N. Uhrig, J. R. Galli, Rufiner H. L., and D. H. Milone. Estimación del consumo en rumiantes en pastoreo utilizando redes neuronales artificiales. *Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação, [S.l.]*, 1:ISSN 2446–7634, 2018.