

“Modelamiento dinámico de la producción ganadera en el Servicio Nacional
de Aprendizaje SENA - Seccional Popayán, Cauca”

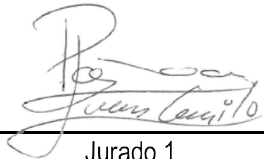
Luis Felipe Guevara Gómez

Nota de Aceptación

Certificamos que el presente Trabajo de Grado Satisface,
en alcances y calidad, todos los requisitos que demanda
un Trabajo de Grado de Maestría.



Director



Jurado 1



Jurado 2

Aprobado en cumplimiento de los requisitos exigidos por la
Pontificia Universidad Javeriana Cali, para optar el título de
Magister en Ingeniería con énfasis en electrónica.



HERNÁN CAMILO ROCHA NIÑO Ph. D.
Decano Facultad de Ingeniería y Ciencias



JUAN CARLOS MARTÍNEZ ARIAS
Director Posgrados de Ingeniería y Ciencias

Santiago de Cali, 24 de noviembre de 2023

Santiago de Cali, 3 de octubre de 2023.

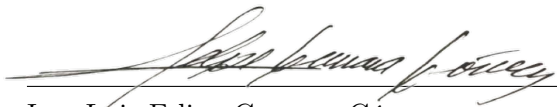
Ingeniero:

Dr. Juan Carlos Martinez Arias
Director Posgrados de Ingeniería
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana – Cali

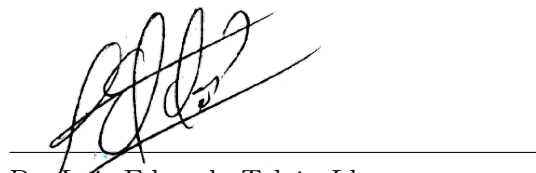
Con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el Trabajo de Grado y posteriormente optar por el título de Magíster en Ingeniería – Énfasis en Electrónica, nos permitimos presentar a su consideración el proyecto de Trabajo de Grado denominado “Modelamiento dinámico de la producción ganadera en el Servicio Nacional de Aprendizaje SENA - Seccional Popayán, Cauca”, el cual será realizado por el estudiante Ing. Luis Felipe Guevara Gómez, con código 2117293, perteneciente al énfasis en Electrónica, bajo la dirección del profesor Dr. Luis Eduardo Tobón Llano.

El suscrito director del Trabajo de Grado autoriza para que se proceda a hacer la evaluación de este Proyecto ante el Tribunal que para el efecto se designe, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado oficialmente.

Atentamente,



Ing. Luis Felipe Guevara Gómez
C.C: 1.144.095.556 de Cali



Dr. Luis Eduardo Tobón Llano
C.C: 9.726.670 de Armenia.



Pontificia Universidad
JAVERIANA
Cali

SONA - CFP 19-9113
Radicación recibida
Nº 19-1-2022-001489
28/06/2022 15:59:44

CARTA DE PRESENTACIÓN PARA POSIBILIDAD DE ESTUDIOS DE INVESTIGACIÓN

Coord. de formación profesional Centro Agropecuario SENA
María del Carmen Agredo

24 de junio de 2022

Referencia: Manifestación de apoyo para posibilidad de estudio de investigación para el estudiante de Maestría en Ingeniería con énfasis en Ingeniería Electrónica, Luis Felipe Guevara Gómez

Cordial Saludo,

De manera atenta nos permitimos manifestar nuestro apoyo hacia el estudiante Luis Felipe Guevara Gómez, quien actualmente está finalizando sus estudios de Maestría en Ingeniería en Electrónica, en la Pontificia Universidad Javeriana Cali (PUJC); y quien en pro de aplicar sus conocimientos adquiridos durante el posgrado, busca realizar un proyecto de investigación en el sector agropecuario mediante un trabajo de grado orientado al Modelamiento dinámico de producciones ganaderas en el departamento del Cauca.

El estudiante nos ha expresado que ustedes pueden contar con un conjunto de datos e información potencialmente útil que podrían resultar provechoso, tanto para él como estudiante en pro de realizar su trabajo de grado; así como también para ustedes, en pro de beneficiarse de los aportes que Luis Felipe Guevara Gómez pueda proporcionarles debido a sus respectivas investigaciones

Así pues, nosotros PhD. Luis Eduardo Tobón Llano, Director del programa de Ingeniería Electrónica; y MSc. Juan Carlos Martínez Arias, Director de posgrados, Fac. de Ingeniería y Ciencias; reiteramos nuestro apoyo hacia Luis Felipe Guevara Gómez mediante esta carta de presentación.

Atentamente,

Pontificia Universidad Javeriana Cali
MSc. Juan Carlos Martínez Arias
Director de posgrados, Fac. de Ingeniería y Ciencias

NIT 860.013.720-1

Calle 18 #118-250 Avenida Cañasgordas, Cali-Colombia, A.A. 26239, Código Postal: 760031,
PBX (+57-2) 321 8200 / 485 6400 - Línea gratuita nacional 01-8000-180556 - www.javerianacali.edu.co
Vigilada Mineducación Res. 12220 de 2016

CARTA DE SOLICITUD DE PRÉSTAMO DE DOCUMENTACIÓN FÍSICA & DIGITAL CENTRO AGROPECUARIO SENA

R/ 197
28-06-2022

Coord. de formación profesional Centro Agropecuario SENA,
Unidad de Ganadería Centro Agropecuario SENA
Maria del Carmen Agredo, Jose Alirio Cobo

28 de junio de 2022

Referencia: Solicitud de Documentos, registros físicos y registros digitales sobre datos de producción lechera SENA - Tesis de Maestría en Ingeniería - Luis Felipe Guevara Gómez

Cordial Saludo,

De manera atenta solicito se me permita fotocopiar y/o tomar prestado de manera temporal (1 día o 2) una serie archivos físicos (enumerados mas adelante), que contienen información críticamente útil para transformarlos a archivos digitales que puedan ser usados tanto para los aprendices del SENA en sus prácticas de seguimiento de producción lechera y estimación de curvas de lactancia; así como para mi, en pro de los estudios estadísticos pertinentes para el desarrollo de mi trabajo de grado orientado al Modelamiento dinámico de producciones ganaderas en el departamento del Cauca.

El inventario de documentos a solicitar es el siguiente:

1. Carpeta blanca nombrada como "Potreros": que contiene la distribución e información de espacio y hectáreas de los potreros destinados a ganadería.
2. Carpeta celuguía cartón café, nombrada como "Pesaje de Leche", que contiene los registros diarios de producciones de leche de las distintas reses entre el 31 de Diciembre de 2018 y el 12 de Abril de 2020 a ganadería.
3. Carpeta celuguía cartón café, nombrada como "Pesaje de Leche #2", que contiene los registros diarios de producciones de leche de las distintas reses entre el 23 de Marzo de 2020 y el 29 de Noviembre de 2020. a ganadería.
4. Carpeta celuguía color verde, nombrada como "Pesaje de Leche 2021", que contiene los registros diarios de producciones de leche de las distintas reses entre el 25 de Enero de 2021 y el 26 de Diciembre de 2021.
5. Carpeta celuguía color verde, nombrada como "Pesaje/Producción de Leche 2022", que contiene los registros diarios de producciones de leche de las distintas reses entre el 27 de Diciembre de 2021 hasta la fecha.
6. Carpeta celuguía cartón café, nombrada como "Nacimientos", que contiene los registros de reses que parieron crías, información que esta directamente relacionada con la producción neta de leche y los periodos en los que las reses interrumpieron o no su participación en la producción neta del hato ganadero.

7. Carpeta celuguía cartón café, nombrada como "R. Monta", que contiene los registros correspondientes a las montas de ganado.
8. Carpetas celuguía cartón café, de producciones individuales de reses que poseen este tipo de seguimiento y que puede que ya no se encuentran en el centro agropecuario, ya sea por venta o por fallecimiento:
 - 8.1. Carpeta nombrada como "Danna", Con periodo de producción iniciando el 28 de Agosto de 2021 y finalizando el 07 de Junio de 2022
 - 8.2. Carpeta nombrada como "Gina", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 2 de Septiembre de 2019
 - 8.3. Carpeta nombrada como "Lina", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 31 de Julio de 2019
 - 8.4. Carpeta nombrada como "Lunita", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 13 de Octubre de 2019
 - 8.5. Carpeta nombrada como "Nancy", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 12 de Mayo de 2019
 - 8.6. Carpeta nombrada como "Lucia", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 13 de Octubre de 2019
 - 8.7. "Fernanda", Con periodo de producción iniciando el 31 de Diciembre de 2018 y finalizando el 13 de Octubre de 2019
 - 8.8. "Juanita", Con periodo de producción iniciando el 28 de Enero de 2019 y finalizando el 16 de Septiembre de 2019
9. Carpeta celuguía cartón café, nombrada "Registro CMT", que contiene los registros correspondientes a las reses que han presentado (o no) mastitis. Con periodo iniciando el 1 de Febrero de 2021 hasta la fecha.
10. Carpeta celuguía cartón café, nombrada "Aplicación de Medicamentos", que contiene los registros correspondientes a los medicamentos suministrados a las reses en caso de tratamiento o enfermedad. Con periodo iniciando el 27 de Octubre de 2021 hasta la fecha
11. Carpeta celuguía cartón café, nombrada "Potrero 45", a cargo de Alexa Fernandez y Reinel Chamorro, que contiene información correspondiente a las acciones de mantenimiento del potrero #45 destinado para el proyecto denominado "Botón de oro".

Adicionalmente, de ser posible, me gustaría solicitar un acceso a las bases de datos digitales, que tal y como me lo comentaron los encargados de la unidad de ganadería del SENA; se encuentran registrados en el software TaurusWeb cuyo encargado es el señor Jose Alirio Cobo.

Así pues, la solicitud de prestamos de esta información se realiza con el objetivo de transformar a digital la información física existente mediante de escáner o foto y transcripción, aunque dado la cantidad de información, este proceso podría tomar entre medio día y 1 día dependiendo de cual sea el caso por lo que se considera más apropiado y económico hacer esta labor en un centro de fotocopiado especializado.

Aprovecho la oportunidad para agradecer nuevamente por la oportunidad la atención tan atenta y servicial con la que fui atendido por diferentes encargados e instructores del Centro Agropecuario quienes me facilitaron el acceso a este tipo de información.

Quedo atento a sus comentarios.

Cordialmente,



Pontificia Universidad Javeriana Cali
Luis Felipe Guevara Gómez
Maestrante en Ingeniería en Electrónica



Respuesta Electrónica

Número de Radicación

19-9-2022-004661

Fecha

08/07/2022 11:35:00 a. m.

NIS

2022-05-022358

Antecedente

1-2022-001489

Asunto *

PETICION

Descripción del Asunto *

RESPUESTA SOLICITUD

Destinatario Externo

Nombre *

JUAN CARLOS MARTINEZ/LUIS FELIPE GUEVARA GOMEZ

Email *

JUANCMARTINEZ@JAVERIANACALI.EDU.CO

Cuerpo

*

19-9113-101

Popayán, 08 de julio del 2022

MSc:

Juan Carlos Martínez Arias

Director de posgrados, Fac. de Ingeniería y Ciencias

lfguevarag97@gmail.com

juancmartinez@javerianacali.edu.co

Ciudad

Asunto: Respuesta Petición Radicado No. 19-1-2022-001489 de Fecha (28/06/2022)

NIS 2022-05-022358

Cordial saludo:

El Servicio Nacional de Aprendizaje SENA, trabaja día a día para ser una Entidad ágil, dinámica y en donde sus servicios estén orientados para el beneficio de los Ciudadanos.

Para dar respuesta a su petición relacionada con apoyo para la posibilidad de estudio de investigación para el estudiante de Maestría en Ingeniería con énfasis en Ingeniería Electrónica, Luis Felipe Guevara Gómez, cordialmente le informamos que al SENA Centro Agropecuario le complace aprobar la solicitud de estudio de investigación en nuestro Centro de Formación, y conforme ARL adjunta. Cualquier inquietud adicional con gusto la atenderé en mi correo institucional: magredo@sena.edu.co

Manifestamos que nuestro compromiso, es atender con oportunidad y eficacia, el (los) requerimiento (S) que nos presenta.

Cordialmente,

María del Carmen Ágredo

Coordinadora FPI

Centro Agropecuario

Regional Cauca

Proyectó: Daniela Mosquera Erazo - danmosquera@sena.edu.co

Cargo: Técnico G02

Anexos (0)

Document Name	Attachment Type
---------------	-----------------

Copias Internas

Copias Externas

Nombre Destinatario *
ONBASE AGROPECUARIO

Email Destinatario *
ONBASEAGROPECUARIO@GMAIL.COM

Nombre Destinatario *
LUIS FELIPE GUEVARA

Email Destinatario *
LFGUEVARAG97@GMAIL.COM

Asociar a otras comunicaciones

N.I.S

No. Radicado Relacionado



Pontificia Universidad
JAVERIANA
Cali

Facultad de Ingeniería
y Ciencias
Ingeniería Electrónica

**FICHA RESUMEN
TRABAJO DE GRADO DE MAESTRÍA**

TITULO: Modelamiento dinámico de la producción ganadera en el Servicio Nacional de Aprendizaje SENA - Seccional Popayán, Cauca

1. **ÉNFASIS:** Ingeniería - Énfasis en electrónica.
2. **TIPO DE PROYECTO:** Proyecto de investigación
3. **ÁREA DE TRABAJO:** Agropecuaria
4. **ESTUDIANTE:** Ing. Luis Felipe Guevara Gómez
5. **CORREO ELECTRÓNICO:** lfguevarag97@gmail.com
6. **DIRECCIÓN Y TELÉFONO:** Carrera 87 #10 - 55, Multicentro 10, Apartamento 408.
(+57) 305 - 379 80 30
7. **DIRECTOR:** Dr. Luis Eduardo Tobón Llano
8. **VINCULACIÓN DEL DIRECTOR:** Profesor de planta, Director de carrera de Ingeniería Electrónica
9. **CORREO ELECTRÓNICO DEL DIRECTOR:** letobon@javerianacali.edu.co
10. **CO-DIRECTOR(ES):** N/A
11. **GRUPO QUE LO AVALA:** Servicio Nacional de Aprendizaje SENA, Seccional Popayán
12. **OTROS GRUPOS:** N/A
13. **PALABRAS CLAVE:** Modelación Dinámica, Ajuste de curvas, Producción ganadera, Regresión no lineal. Splines, Ecuaciones Diferenciales Ordinarias (EDO), Levenberg-Marquardt (LMA), Efectos no lineales mixtos (NLMEA).

14. **ODS QUE APLICA EL PROYECTO (AGENDA 2030):**

- ODS #9 - Industria, innovación e infraestructura: Apoyar el desarrollo de tecnologías, la investigación y la innovación nacionales en los países en desarrollo, incluso garantizando un entorno normativo propicio a la diversificación industrial y la adición de valor a los productos básicos, entre otras cosas (9.b).
- ODS #12 - Producción y consumo responsables: Ayudar a los países en desarrollo a fortalecer su capacidad científica y tecnológica para avanzar hacia modalidades de consumo y producción más sostenibles (12.a).

15. **FECHA DE INICIO:** 1 de Febrero de 2023.

16. **RESUMEN:** Ver siguiente página.

Resumen

Modelar dinámicamente un sistema es importante para representar de manera simplificada su comportamiento sin recurrir a la experimentación, que para sistemas complejos, puede representar un alto consumo de recursos. También es útil para predecir comportamientos a largo plazo; observar, analizar, e interpretar resultados, frente a diferentes entradas que afectan directa o indirectamente a el o los subsistemas presentes; establecer conclusiones y apoyar la toma de decisiones futuras.

Este trabajo estudia el comportamiento productivo de ganado bovino y bufalino del Centro Agropecuario del Servicio Nacional de Aprendizaje (SENA), ubicado en la ciudad de Popayán (CA-SENA-POP), teniendo en consideración diferentes variables, entre ellas, la producción de leche, el consumo de agua, consumo de forrajes, consumo de concentrado, generación de desperdicios y finalmente utilidades por venta de producciones de leche y carne. El objetivo es estructurar y evaluar un modelo basado en ecuaciones diferenciales que facilite observar respuestas del sistema productor ganadero ante diferentes entradas, en pro de predecir comportamientos a largo plazo.

Este proyecto adapta de manera flexible la metodología CRISP-DM, con la que se busca ajustar conjuntos de datos a curvas lineales y no lineales, de tal forma que se proceda a modelar dinámicamente las respuestas del sistema productor ganadero frente a diferentes entradas. Este proyecto también utiliza conjuntos de datos debidamente registrados y proveídos por el SENA tanto de manera análoga como digital. Los resultados obtenidos pueden suponer un punto de referencia para prever situaciones futuras y evaluar el rendimiento del modelo frente a nuevos datos.

Palabras y Conceptos Clave: Ajuste de curvas, Modelamiento Dinámico, Producción ganadera.

Abstract

Modeling a system dynamically is important to represent its behaviour in a simplified way without resorting to experimentation, which for complex systems, it could represent a high resource consumption. It is also useful to predict long term responses; observe, analyze and interpret results against different inputs that directly or indirectly affect subsystem(s) present; draw conclusions and support future decision making.

This work studies the functioning of livestock production in the Agricultural Center of the National Apprenticeship Service, Popayan sectional (CA-SENA-POP); taking into account a variety of variables such as milk yield, water, fodder and concentrated food consumption; waste generation and last but not least, profit from milk and meat production sales. The objective is to structure and assess a model based on differential equations that facilitates the visualization of the CA-SENA-POP livestock system against different situations for long term behaviour prediction.

This project adapts, in a flexible way, the CRISP-DM methodology, with which it is sought to adjust data sets to linear and nonlinear curves in such a way that the responses of the livestock production system to different inputs are dynamically modeled. This project also uses properly registered and organized databases provided by the SENA institution, both analog and digital. The results obtained may be used as reference to foresee future situations and evaluate the performance of the model against new data.

Keywords: Curve fitting, Dynamic modeling, Livestock production.



Pontificia Universidad
JAVERIANA
Cali

**Facultad de Ingeniería
y Ciencias**
Ingeniería Electrónica

MONOGRAFÍA DE TRABAJO DE GRADO

Modelamiento dinámico de la producción ganadera en
el Servicio Nacional de Aprendizaje SENA - Seccional
Popayán, Cauca

Ing. Luis Felipe Guevara Gómez

Director
Dr. Luis Eduardo Tobón Llano

3 de octubre de 2023

Agradecimientos

Quiero agradecer a mi director de proyecto, el Dr. Luis Eduardo Tobón Llano, por su guía, conocimiento, paciencia y apoyo en la finalización de este proyecto. También agradecer al equipo del Centro Agropecuario del Servicio Nacional de Aprendizaje, seccional Popayán (CA-SENA-POP) por brindarme su apoyo, experticia y confianza al suministrarme los datos y la información necesaria para realizar mi proyecto.

A mis familiares y amigos por su presencia y apoyo moral.

Y sobre todo, especial dedicatoria a mis padres, hermanos y abuelos, por inspirarme, alentarme, apoyarme y siempre darme ánimos y fuerza para seguir adelante en la búsqueda de mis sueños y mi futuro.

Nomenclatura

El siguiente apartado describe los símbolos y acrónimos que serán usados más adelante en el contenido de este documento:

Símbolos

α_s	Ingesta total de alimento. [%]
α_x	Ingesta total de: forrajes (x=f), concentrado (x=c) o agua (x=a). [%]
χ^2 , χ_{red}^2	Factor “chi” cuadrado, Factor “chi” cuadrado reducido

Acrónimos

<i>AIC</i>	Criterio de Información de Akaike
<i>BIC</i>	Criterio de Información Bayesiano
<i>CA</i>	Centro Agropecuario.
<i>EDO</i>	Ecuaciones Diferenciales Ordinarias
<i>FM, FU</i>	Fit Múltiple, Fit Único.
<i>IGAC</i>	Instituto Geográfico Agustín Codazzi.
<i>LMA</i>	Levenberg-Marquardt.
<i>ME</i>	Efectos Mixtos (en inglés: Mixed Effects)
<i>NL</i>	No Lineal (en inglés: Non-Linear)
<i>OE</i>	Objetivo específico
<i>PUJC</i>	Pontificia Universidad Javeriana, Cali.
<i>SENA</i>	Servicio Nacional de Aprendizaje.
<i>WP</i>	Weight-Produced

Índice

1. Introducción	1
1.1. Planteamiento del Problema	2
2. Objetivos y restricciones	5
2.1. Objetivo General	5
2.2. Objetivos Específicos	5
2.3. Delimitaciones y Alcances	5
2.4. Entregables	6
3. Metodología	7
3.1. Metodología de trabajo propuesta	7
3.2. Metodología CRISP-DM	7
3.2.1. Entendimiento del negocio	8
3.2.2. Entendimiento de los datos	8
3.2.3. Preparación de los datos	9
3.2.4. Modelación	9
3.2.5. Evaluación	10
3.2.6. Despliegue	10
4. Marco teórico	11
4.1. Contexto ganadero multipropósito	11
4.1.1. Generalidades agropecuarias	11
4.1.2. Caracteres zootécnicos del ganado vacuno	12
4.1.3. Sistemas de explotación del ganado vacuno	14
4.1.4. Ciclo productivo de la leche	15
4.1.5. Ciclo productivo de la carne	17
4.1.6. Alimentación y Componentes básicos de la dieta	19
4.1.7. Balance de raciones y dietas especializadas	21
4.1.8. Leyes y Normatividad	21
4.2. Modelación matemática	22
4.2.1. Regresión lineal	22
4.2.2. Regresión no lineal	22
4.2.3. Interpolación con Splines	23
4.2.4. Ajuste de curvas	24
4.2.5. Métodos de ajuste no lineal	24
4.2.6. Modelación por “Ecuaciones Diferenciales Ordinarias” (EDOs)	26
4.2.7. Criterios de evaluación	27

4.3. Descripción general del proyecto realizado bajo la metodología CRISP-DM	30
5. Entendimiento del negocio	31
5.1. Servicio Nacional de Aprendizaje (SENA)	31
5.1.1. Centro Agropecuario	31
5.2. Empresa donde se realiza el estudio: CA-SENA-POP	31
5.3. Negocio y comercialización de productos	32
6. Entendimiento de los datos	33
6.1. Adquisición de datos	33
6.1.1. Registros de “Leche”	33
6.1.2. Registros de “Peso”	34
7. Preparación de los datos	37
7.1. Recolección de datos	37
7.2. Digitalización de los datos de producción lechera	38
7.3. Reestructuración de los datos de leche	39
7.3.1. Reestructuración inicial	39
7.3.2. “Debugging” y verificación de datos	40
7.3.3. Reestructuración numérica	41
7.3.4. Clasificación, Selección y Agrupación de datos de leche	42
7.4. Reestructuración de los datos de peso	49
7.4.1. Reorganización de datos	49
7.4.2. Agrupación de datos de peso	49
8. Modelado	51
8.1. Ajuste de Curvas mediante funciones No lineales	51
8.1.1. Aplicación del método de ajuste no lineal: Algoritmo de Levenberg-Marquardt (LMA)	52
8.1.2. Aplicación del método de ajuste no lineal: Algoritmo de Efectos Mixtos No Lineales (NLMEA)	53
8.1.3. Aplicación del método de ajuste no lineal: Algoritmo de Efectos Mixtos No Lineales con Efectos Aleatorios (NLMEA-RE)	55
8.2. Modelación dinámica	57
8.2.1. Planteamiento del modelo	57
8.2.2. Modelo EDO-WP	58
8.2.3. Ajuste del modelo EDO-WP	65
9. Evaluación	73
9.1. Objetivos específicos (OE)	73
9.1.1. OE#1: Identificar entradas, salidas, variables y parámetros asociados al sistema de producción ganadero objetivo.	73

9.1.2. OE#2: Identificar modelos dinámicos que permitan pronosticar la producción de un sistema ganadero objetivo.	74
9.1.3. OE#3: Implementar métodos de solución de los modelos dinámicos de producción ganadera seleccionados.	74
9.1.4. OE#4: Evaluar la estimación de los modelos seleccionados comparando la solución con datos empíricos recogidos.	74
9.2. Metodología CRISP-DM	75
9.3. Cumplimiento del OE#4 en la etapa de Modelación: Ajuste de curvas por métodos no lineales	75
9.3.1. Producción neta real Vs Producción teórica por ajustes no lineales	77
9.3.2. Modelo EDO: Ajuste por mínimos cuadrados	77
9.4. Retroalimentación del proyecto con el cliente final	78
9.4.1. Socialización del proyecto con el CA-SENA-POP	78
10. Conclusiones y trabajos futuros	81
10.1. Conclusiones	81
10.2. Posibles mejoras y trabajos a futuro	82
10.2.1. Modelo EDO en 3D: MEW - Leche, Estiércol, Peso	82
10.2.2. Modelado por Autómata híbrido	82
10.2.3. Modelado por programación lineal entera	83
Bibliografía	85
10.3. Modelo de optimización por programación lineal entera	89
10.3.1. Conjuntos	89
10.3.2. Parámetros	89
10.3.3. Variables	90
10.3.4. Funciones objetivo	90
10.3.5. Restricciones	91
10.4. Tabla 10.1 - Parámetros de ajuste para cada método usado.	91
10.5. Figura 10.1 - Matrices de covarianza	93
10.6. Relación matemática entre los parámetros k_1 , k_2 , k_3 y k_4	94
10.6.1. Observaciones	94
10.7. Representación gráfica del modelo EDO-WP	95

Índice de figuras

3.1. Ciclo de Vida de minería de datos. Modificada de [1].	7
4.1. Partes físicas de un bovino. Tomada de [2]	12
4.2. Aptitud lechera. Tomada de [3].	12
4.3. Aptitud de tipo carne. Tomada de [3].	13
4.4. Aptitud de doble propósito. Tomada de [3].	13
4.5. Sistema de explotación estabulada. Tomada de [3].	14
4.6. Sistema de explotación silvopastoril. Tomada de [4].	14
4.7. En la primer fila de izquierda a derecha se observan ejemplares de las razas Normando, Simmental; seguido de las razas Holstein y Jersey en la fila inferior. Tomada de [4]	16
4.8. Ciclo productivo en la ganadería de la carne. Modificada de [4].	17
4.9. En la primer fila de izquierda a derecha se observan ejemplares de las razas Charolais, Hereford; seguido de las razas Angus y Criolla en la fila inferior. Tomada de [4].	19
4.10. Hidratación del ganado. Tomada de [4].	19
4.11. Suplementos energéticos.	20
4.12. Suplementos proteínicos. Tomada de [2].	20
4.13. Forraje y leguminosas. Tomada de [2].	21
4.14. Comparación entre una interpolación de datos por Splines Cúbicas vs Una función polinómica ideal	23
4.15. Planteamiento general de las etapas 2, 3, 4 y 5 de la metodología usada en la ejecución del proyecto.	30
5.1. Granja del CA-SENA-POP.	32
6.1. Registros manuales de producción lechera en el CA-SENA-POP.	33
6.2. Registro de datos digital con software TaurusWebs.	34
6.3. Ejemplo de seguimiento digital de peso en la res “Dianita”.	35
7.1. Ejemplares de datos físicos con malas prácticas de escritura.	38
7.2. Reestructuración inicial de los datos de producción de leche transcritos.	39
7.3. Clasificación de datos por #Parto para cada vaca.	43
7.4. Desbalance de datos en agrupación inicial por partos.	44
7.5. Gráfico de dispersión para representar la producción lechera de las vacas primíparas.	45
7.6. Gráfico de dispersión para Muestras matutinas y vespertinas.	45
7.7. Gráfico de dispersión para Muestras matutinas y vespertinas.	46
7.8. Gráfico de dispersión para cada grupo de partos.	47
7.9. Reorganización inicial de los datos de peso.	49

7.10. Formato de nombre de columna para conjunto de datos. Varía en cuanto al nombre de la vaca, cantidad de muestras existentes y cantidad de parto. El código de colores se mantiene según la Tabla 7.1	50
8.1. Ajustes de curvas de lactancia para Acerada y Lunita, Parto#1.	52
8.2. Convergencia de parámetros estimados del modelo de Wood para Acerada y Lunita.	52
8.3. Estimación de parámetros generales mediante LMA, para el grupo de parto #1	53
8.4. Diagrama de cajas y bigotes para las reses del parto 1.	53
8.5. Estimación de parámetros individuales en el parto #1, Usando NLMEA-FE.	54
8.6. Diagrama de cajas y bigotes individualizado para las reses del parto 1.	54
8.7. Comparación final entre: a) Ajuste por NLMEA-RE y b) Ajuste por LMA	56
8.8. Gráfica de probabilidad normal en los residuos. Parto #1.	56
8.9. Diagrama de bloques del modelo EDO-WP en Matlab-Simulink.	59
8.10. Comportamiento del modelo con $(k_i = 1)$, para todo i	60
8.11. Comportamiento del modelo con $(k_1 > k_2)$ y $(k_3 > k_4)$	60
8.12. Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 < k_4)$	61
8.13. Comportamiento del modelo con $(k_1 < k_3)$ y $(k_2 < k_4)$	61
8.14. Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 \gtrsim k_4)$	62
8.15. Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 \gtrsim k_4)$	62
8.16. Comportamiento del modelo con $(k_1 > k_3)$ y $(k_2 > k_4)$	63
8.17. Comportamiento del modelo con $(k_1 \gtrsim k_3)$ y $(k_2 \gtrsim k_4)$	64
8.18. Comportamiento del modelo con $(k_1 \gtrsim k_3)$ y $(k_2 \gtrsim k_4)$	64
8.19. Primeros 4 registros del conjunto de datos del grupo de parto #3.	66
8.20. Conjunto de datos del parto #3 con solo 2 variables W y P para cada res.	66
8.21. Variables de estado W y P y sus respectivas condiciones iniciales, para cada res.	66
8.22. Modelo EDO-WP, implementación en Python.	67
8.23. Restricciones del modelo EDO-WP implementado en Python.	67
8.24. Soluciones de los modelos con ajuste de parámetros por mínimos cuadrados.	67
8.25. Ajuste del modelo EDO-WP, a los datos de la res “Fernanda”.	68
8.26. Variables y datos estadísticos del ajuste.	68
8.27. Ajuste del modelo EDO-WP, a los datos de la res “Guapa”.	69
8.28. Variables y datos estadísticos del ajuste.	69
8.29. Ajuste del modelo EDO-WP, a los datos de la res “Juanita”.	70
8.30. Variables y datos estadísticos del ajuste.	70
8.31. Ajuste del modelo EDO-WP, a los datos de la res “Sol”.	71
8.32. Variables y datos estadísticos del ajuste.	71
10.1. Matrices de covarianza del método “nlmefit”.	93
10.2. Representación gráfica de las ecuaciones 8.2 y 8.3.	95

Índice de tablas

7.1. Código de colores por número de parto	43
7.2. Agrupación final de vacas por # de partos	47
7.3. Agrupación general de vacas por # de partos	48
7.4. Cantidad de muestras para cada vaca de cada parto	50
9.1. Comparación de error cuadrático medio (MSE) para cada método de ajuste no lineal.	75
9.2. Comparación de la raíz del error cuadrático medio (RMSE) para cada método de ajuste no lineal.	76
9.3. Comparación del Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) para los métodos de ajuste no lineales por efectos mixtos fijos y aleatorios	76
9.4. Comparación de producción real y teórica para cada método de modelación utilizada.	78
10.1. Resumen de los parámetros de ajuste para cada método usado y su respectivo MSE.	92

Introducción

En los últimos años, el sector agropecuario ha ido retomando gran importancia en Latinoamérica y el mundo, debido a múltiples factores entre los cuales se puede resaltar el crecimiento poblacional, que lleva consigo el incremento del consumo de alimentos. Esto representa una necesidad que puede y debe ser resuelta con el crecimiento de la producción agropecuaria de hasta un 70 % para satisfacer la demanda alimenticia a nivel mundial. Para ello es de vital importancia que los países en desarrollo como Colombia, entren a formar parte del mercado altamente productivo y competitivo haciendo uso de las fuentes que le aportan una ventaja comparativa en la región que radican, siendo estas la abundancia de recursos hídricos y terrestres [5].

Actualmente, Brasil es el segundo mayor proveedor de alimentos y productos agropecuarios a nivel mundial y cuenta con mejoras continuas en los procesos de producción, que respaldan el rápido crecimiento de las exportaciones. La investigación agropecuaria ha sido uno de los factores clave para el incremento de la productividad en la región durante las últimas décadas, con el fin de hacer seguimiento al registro, tratamiento, manipulación y análisis los datos y el desempeño de los sistemas de investigación y desarrollo agropecuario mediante la evaluación de resultados. Estos datos son una herramienta indispensable cuando se trata de evaluar el aporte de la I+D agropecuaria y de manera más general, el crecimiento económico de la región y el país.

En América Latina y más precisamente en Colombia, sobresale la formación de 55 Comités de Investigación Agrícola elegidos Localmente (CIAL). Estos están constituidos por agricultores experimentados que administran y conducen investigación y desarrollo en beneficio de toda la comunidad [6]; así pues se representa la alta participación colombiana en este ámbito de desarrollo. En Colombia también se han incorporado nuevas entidades que aportan al desarrollo investigativo, productivo y competitivo como la Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA), que de mano con la expedición de la Ley 1876 de 2017 donde se crea el Sistema Nacional de Innovación Agropecuaria (SNIA), buscan integrar la ciencia, y el desarrollo tecnológico del sector junto con la formación y prestación del servicio de extensión agropecuaria. En este sentido, la ciencia permite lograr avances en términos productivos, de innovación y de competitividad [7].

En el Cauca, de acuerdo con el Instituto Geográfico Agustín Codazzi, los suelos del departamento presentan casi todos los pisos térmicos de variadas fertilidades y profundidades, y con diversas vocaciones para su uso [8]. Por otra parte, aunque comúnmente pueda considerarse un sector primitivo o de poca inversión tecnológica, el sector agropecuario cuenta con el mayor potencial tecnológico para el desarrollo sostenible y es uno de los más tecnificados en la región. Esto debido a la inclusión

de nuevas tecnologías, no solo para la ejecución de procesos, sino también para la gestión y el manejo de las grandes cantidades de datos, impulsando así, el desarrollo competitivo del campesino e incrementando su participación en el mercado a nivel regional, nacional y mundial.

No obstante, la inclusión de las actuales tecnologías puede significar una gran inversión económica para pequeños y medianos productores y cultivadores. Los altos costos significan un impedimento a la participación competitiva en el mercado ganadero y de cultivos varios, con lo que el mejoramiento de pequeñas partes del sistema total puede manifestarse en mejoras significativas en el día a día de los nuevos emprendedores ganaderos. Estos costos se pueden taxonomizar en costos de producción, mantenimiento, transporte, control, toma, registro, seguimiento, manipulación y representación de datos; que diariamente aportan al consumo de recursos limitados como los económicos, personales y tiempo. Para ello se requiere de nuevos enfoques que permitan mejorar algunas de las condiciones de trabajo de los cultivadores agrícolas promoviendo la integración científica, innovadora y competitiva.

Con base en todo lo anterior, se considera indispensable la inclusión de tecnologías y metodologías de tratamiento, simulación y modelamiento computacional de los datos presentes en la producción ganadera.

1.1. Planteamiento del Problema

El sector ganadero abarca diferentes especies de ganado tales como los bovinos, los ovinos, bufalinos, entre otros. En Colombia se tiene una alta presencia de producción de estos tipos especialmente en el ganado bovino y ovino. De igual forma, estos tipos de ganado tienen diferentes modalidades, entre las que se pueden mencionar el ganado para crianza, para producción de leche y sus derivados, y ganadería de la carne o doble propósito. En Colombia y el mundo se presentan constantes inversiones en materia tecnológica y metodológica para mejorar la forma en cómo el ganado es productivo y puede participar de forma competitiva en un mercado de consumo de calidad [5].

Aunque en el Cauca se cuente con vastas hectáreas de diferentes fertilidades y condiciones térmicas apropiadas para el cultivo y desarrollo de la ganadería [8], aún se presentan falencias en materia de inversión tecnológica, económica e investigativa, en donde los movimientos migratorios de campesinos desplazados y los efectos colaterales del conflicto armado que se ha presentado en el país, son las principales causas de estas falencias [5]. El emprendimiento en el ámbito de la producción ganadera, especialmente para pequeños y medianos productores, presenta oportunidades significativas para la implementación de enfoques avanzados, como la ingeniería, el modelamiento matemático y la incorporación de metodologías de análisis, clasificación y manejo cuantitativo de datos, incluyendo la Inteligencia Artificial.

Sin embargo, la adopción de software sofisticado y modelos matemáticos puede plantear desafíos para los productores rurales, ya sea debido a la falta de capacitación técnica o a la complejidad de las herramientas de software utilizadas para el manejo, análisis y procesamiento de datos. Esto, a menudo, conduce a que las decisiones se tomen utilizando enfoques tradicionales en lugar de

enfoques sistémicos y basados en datos.

A lo largo del último siglo, la investigación a nivel global ha permitido la estimación y predicción del comportamiento de variables asociadas a la producción agropecuaria, como la producción de leche, carne y el consumo de insumos, a través de modelos matemáticos de diferentes complejidades, tanto lineales como no lineales, como se documenta en [9], [10], [11]. Adicionalmente a esto, en las últimas décadas se han desarrollado múltiples investigaciones en las que, computacionalmente se modelan sistemas dinámicos de producción lechera tales como [12], [13], [14], [15]; Estos avances proporcionan herramientas valiosas para respaldar la toma de decisiones en los sistemas de producción ganadera [16].

Como se menciona en [7], [6] y [5]; la inclusión de la ciencia y el apoyo de instituciones en el sector agrario pueden tener un impacto significativo en el rendimiento de las operaciones ganaderas. En Colombia, se han llevado a cabo estudios de investigación con ganado bovino y caprino en diversas regiones del país, como se ejemplifica en [9], [17], [18]; Por lo tanto, se busca fomentar la investigación similar en la producción ganadera del departamento del Cauca, tomando como referencia la producción ganadera tecnificada del Centro Agropecuario (CA) del Servicio Nacional de Aprendizaje (SENA) en el municipio de Popayán.

Basándonos en lo anterior, se propone el desarrollo de un modelo dinámico de producción ganadera en el Centro Agropecuario (CA) del Servicio Nacional de Aprendizaje (SENA) en Popayán, con el propósito de estimar los consumos, las producciones y los rendimientos ganaderos. Este modelo se plantea como una herramienta esencial para respaldar la toma de decisiones futuras de los productores agropecuarios, permitiéndoles mejorar la gestión y la planificación de sus actividades ganaderas, lo que, a su vez, puede contribuir al crecimiento sostenible del sector.

Objetivos y restricciones

2.1. Objetivo General

Estructurar y evaluar un modelo dinámico que simule el comportamiento de sistemas de producción ganadera en el Centro Agropecuario del Servicio Nacional de Aprendizaje, ubicado en el municipio de Popayán, en departamento del Cauca (CA-SENA-POP).

2.2. Objetivos Específicos

Los siguientes objetivos propuestos se ejecutarán a lo largo del desarrollo del este proyecto:

- Identificar entradas, salidas, variables y parámetros asociados al sistema de producción ganadero objetivo.
- Identificar modelos dinámicos que permitan pronosticar la producción de un sistema ganadero objetivo.
- Implementar métodos de solución de los modelos dinámicos de producción ganadera seleccionados.
- Evaluar la estimación de los modelos seleccionados comparando la solución con datos empíricos recogidos.

2.3. Delimitaciones y Alcances

- No se tendrán en consideración a animales que no participen en la producción del hato ganadero bovino y bufalino del CA-SENA-POP
- El modelo se plantea con bovinos y bufalinos de producción lechera pertenecientes a distintos ciclos de lactancia, indiferentemente de su procedencia genética.
- Las codificaciones, algoritmos o “scripts” requeridos para analizar datos de la producción ganadera se realizan en software especializado como R, Matlab o Python.
- El ajuste de datos de producción lechera se realiza mediante métodos numéricos de interpolación lineales o no lineales según corresponda.

- El modelo se basa en datos existentes, verificables, análogos o digitales, y aprobados para su uso en este estudio, por parte de la producción ganadera del CA-SENA-POP.
- El modelo resultante podrá estimar comportamientos de variables de la producción ganadera y será evaluada usando datos existentes como referencia.
- Los datos suministrados por parte de CA-SENA-POP no serán considerados como entregables por motivos de acuerdos de confidencialidad.

2.4. Entregables

- Documentación descriptiva, explicativa y argumentativa que evidencie el proceso de entendimiento, preparación, modelado, análisis, diseño, evaluación y desarrollo del proyecto.
- Modelo dinámico de referencia para estimación de producciones ganaderas en el CA-SENA-POP.

3.1. Metodología de trabajo propuesta

Este trabajo consiste principalmente en la recolección de datos que posteriormente deben pasar por distintas etapas tales como análisis, organización, clasificación y, en algunos casos, normalización. Una vez se haya logrado el entendimiento de los datos se procede a establecer una serie de funciones matemáticas de estimación y/o regresión, y posteriormente un modelo dinámico mediante ecuaciones diferenciales. Finalmente, este modelo deberá ser evaluado para verificar si éste puede estimar variables de manera apropiada.

Teniendo en cuenta las etapas principales anteriormente mencionadas, y teniendo en consideración la literatura disponible para proyectos de este tipo; se considera plausible proponer una adaptación de la metodología CRISP-DM, a los intereses del proyecto.

3.2. Metodología CRISP-DM

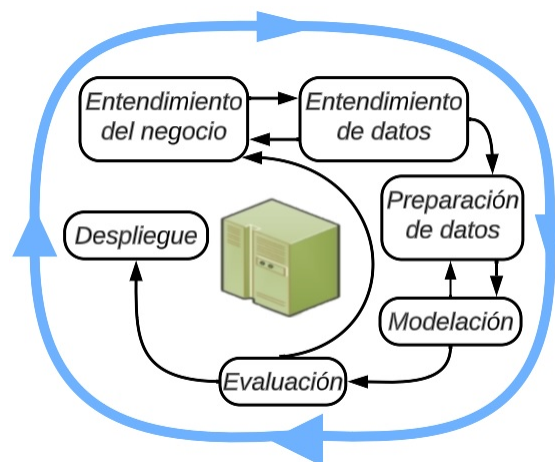


Figura 3.1: Ciclo de Vida de minería de datos. Modificada de [1].

La metodología CRISP-DM, por sus siglas en Inglés, Cross Industry Standar Process for Data Mining, propone un método probado para orientar trabajos de minería de datos. Puede ser abarcada

de 2 maneras; como proceso, ofreciendo un resumen del ciclo vital de la minería de datos; y como una metodología, en la que se incluye descripciones de las fases normales de un proyecto, incluyendo las tareas necesarias de cada etapa y una explicación inter-relacional entre las tareas [1].

Una de las características de esta metodología es su flexibilidad y personalización de un modelo de minería de datos que se adapte a las necesidades del proyecto. Esta metodología cuenta con un ciclo de vida compuesto por 6 fases o etapas no necesariamente secuenciales; y es en estas etapas donde se indican las dependencias más importantes y frecuentes entre las fases [1].

Tal y como se observa en la Figura 3.1, las etapas de la metodología son:

3.2.1. Entendimiento del negocio

Esta primera etapa consiste principalmente en obtener la máxima cantidad de información posible de los objetivos comerciales. De esta forma se puede identificar de manera clara los objetivos, metas, problemas, y recursos con los que se cuentan para realizar el estudio. La metodología sugiere la recopilación de información sobre la situación comercial actual, los recursos personales, materiales e intangibles. Por otra parte se debe tener claro conocimiento de la estructura de la empresa, patrocinadores internos o externos y unidades comerciales que se verán afectadas positiva o negativamente tras la realización del proyecto de minería de datos [1].

3.2.2. Entendimiento de los datos

Esta etapa implica estudiar más de cerca los datos disponibles de minería. Esto es un paso imprescindible para evitar problemas inesperados durante la fase de preparación de datos pues suele ser la más larga en los proyectos [1].

En esta etapa se busca acceder a los datos, explorarlos mediante tablas, gráficos o análisis estadísticos. De esta forma se podrá determinar la calidad de los datos y describir resultados en la documentación del proyecto.

En este punto de la metodología se puede catalogar los datos de 3 maneras distintas

- Datos existentes: registros, transacciones, encuesta, registro web, entre otros
- Datos adquiridos: Datos adicionales que pueden ser, o no, necesario incluirlos.
- Datos adicionales: Datos complementarios si los requiere

Con base en lo anterior y con las delimitaciones de la Sección 2.3, se puede establecer que para este proyecto se trabajará principalmente con los datos existentes. También es importante realizar una descripción detallada de los datos, ya sea en cantidad o calidad; tipos de valores, ya sean numéricos, categóricos o booleanos; esquemas de codificación

3.2.3. Preparación de los datos

La preparación de datos es uno de los aspectos más importantes pues se busca que los datos cuenten con una estructura indicada para que los modelos que se propongan pueden satisfacer las necesidades para los que son propuestos. Es posible que para llevar a cabo esta etapa se deba realizar una combinación o agrupación de datos o registros, realizar una selección de muestra de un subconjunto de datos, agregación de registros, clasificación de datos para el modelado, eliminación y/o sustitución de valores en blanco o perdidos, ó la división en conjuntos de datos de prueba, entrenamiento o evaluación [1].

En cuanto a la selección de datos, es necesario que los datos sean relevantes por lo que se pueden seleccionar datos basados en la selección de elementos (filas de registros) o en la selección de atributos (columnas o características). Posteriormente a esto se debe realizar una limpieza de datos en donde se debe tratar con datos perdidos o en blanco, metadatos erróneos o perdidos, incoherencias y errores de datos, en cuyo caso deben ser corregidos [1].

3.2.4. Modelación

Este es el punto donde los datos preparados pueden incorporarse a las herramientas analíticas y cuyos resultados podrán arrojar primeros resultados deseados respecto al problema planteado en la comprensión del negocio. El modelado se suele ejecutar en múltiples iteraciones. Normalmente, los analistas de datos ejecutan varios modelos utilizando los parámetros predeterminados y ajustan los parámetros o vuelven a la fase de preparación de datos para las manipulaciones necesarias por su modelo. Es necesario que el modelo sea acorde a los tipos de datos disponibles, a los objetivos y patrones, y a los requisitos específicos de modelado como por ejemplo si se busca que los resultados sean fácilmente presentables.

En cuanto a la generación de los modelos es necesario dedicar el tiempo suficiente para experimentar con distintos modelos antes de llegar a conclusiones definitivas. Estos modelos pueden ofrecer resultados interesantes en cuanto a interpretaciones y conclusiones. También puede realizarse una comparativa de modelos antes de integrarlos o desplegarlos. Para registrar el proceso con una amplia variedad de modelos es necesario asegurarse de registrar los cambios que se realicen en cada modelo en pro de corroborar y/o analizar los resultados cuando sea necesario.

Al final de esta etapa se dispondrá de 3 tipos de información que pueden ser usados para la toma de decisiones posteriores; estos datos son:

1. Configuración de parámetros donde se incluyen las notas que se han tomado sobre los parámetros que ofrecen mejores resultados.
2. Los modelos producidos.
3. Las descripciones de resultados de los modelos, incluyendo resultados asociados a problemas con los datos, rendimiento y exploración de resultados

3.2.5. Evaluación

En este punto, se habrá completado la mayor parte del proyecto. También se habrá determinado, en la fase de modelado, que los modelos son técnicamente correctos y efectivos en función de los objetivos que se han definido previamente. Esta etapa de la metodología produce 2 tipos de resultados:

- Los modelos finales seleccionados en la fase anterior de CRISP-DM.
- Las conclusiones o inferencias obtenidas de los modelos y del proceso de minería de datos; recibiendo el nombre de *descubrimientos*.

Es en esta etapa donde se debe verificar si los resultados se expresan con claridad y de forma que se pueden representar con facilidad; si se han realizado descubrimientos especiales o particularidades que se deban resaltar; y si los resultados se adaptan a los objetivos del proyecto. Una vez haya evaluado los resultados, es recomendable realizar una lista de el (los) modelo(s) aprobado(s) para incluir en el informe final.

3.2.6. Despliegue

En general, la fase de despliegue de CRISP-DM incluye dos tipos de actividades:

- Planificación y control del despliegue de los resultados
- Finalización de tareas de presentación como la producción de un informe final y la revisión del proyecto

Marco teórico

En este capítulo se describen las bases teóricas concernientes al contexto ganadero en Colombia y el Cauca, así como también las bases teóricas de los métodos matemáticos y algoritmos computacionales aplicados a lo largo del desarrollo de este proyecto.

4.1. Contexto ganadero multipropósito

4.1.1. Generalidades agropecuarias

La fortaleza principal del sector primario colombiano radica en la producción de recursos de la naturaleza. Esto se refiere a las actividades económicas en relación a los procesos de producción en el sector agropecuario. Este sector abarca el conjunto de técnicas y las acciones que permiten trabajar y cultivar la tierra para producir materias primas así como también la producción de recursos de la ganadería. Por su parte, el sector ganadero alude a los diferentes tipos de ganado de una zona y a las actividades que se emplean para criar y comercializar a estos animales y a los productos derivados de estos. En Colombia se practican diferentes modalidades de ganadería con base en el tipo de ganado de enfoque tales como:

- Bovinos ó Vacunos
- Equinos
- Porcinos
- Mulares y asnales
- Ovinos

De igual forma, estos tipos de ganado son trabajados con diferentes designaciones, entre las cuales se pueden mencionar:

- Crianza y material reproductivo: Tiene como objetivo la reproducción de ganado mediante el cruce de individuos con las mejores características biológicas para mejorar así la genética de los individuos. Los mejores ejemplares son comercializados como material reproductivo y producción de leche dependiendo de su sexo (macho y hembra respectivamente), y los demás ejemplares pasan a ser parte de la ganadería de la carne.
- Producción de leche y derivados lácteos: Como su nombre lo indica, es el ganado designado para la producción cuantitativa y cualitativa de leche mediante ordeño manual, mecánico o eléctrico.
- Producción de carne: Encaminado a la producción cuantitativa y cualitativa de carne.

4.1.2. Caracteres zootécnicos del ganado vacuno

La destinación del ganado vacuno para producción de leche, material reproductivo o engorde depende de algunos factores fisonómicos y biológicos como los mencionados a continuación:

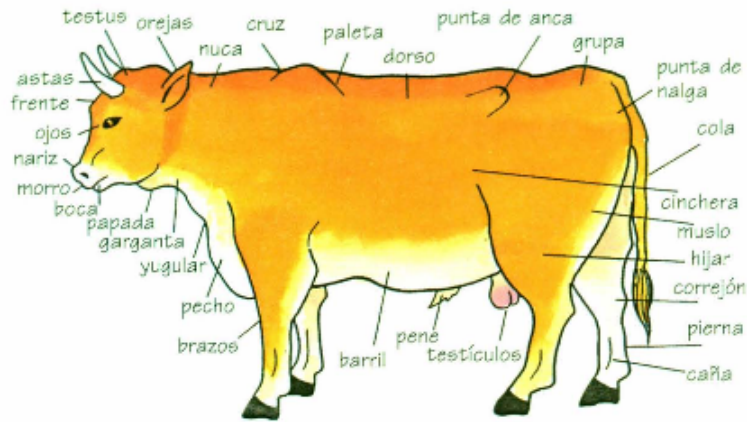


Figura 4.1: Partes físicas de un bovino. Tomada de [2]

4.1.2.1. Clasificación vacuna por aptitud productiva

- **Lechero:** (Cabe resaltar que un requerimiento o propiedad fisonómica básica es que posean ubre, esto quiere decir solo las hembras pueden hacer parte de esta aptitud productiva). Poseen formación triangular, cuentan con cuerpo y extremidades largas y delgadas con poca voluminosidad cárnica



Figura 4.2: Aptitud lechera. Tomada de [3].

- **Carne:** Poseen una contextura rectangular. Además de poseer un cuerpo ancho con sus extremidades bien dotadas de carne.



Figura 4.3: Aptitud de tipo carne. Tomada de [3].

- **Doble propósito:** Cuentan con propiedades mixtas de los 2 tipos ya mencionados, por lo que tanto su contextura, como volumen cárnico, ubre y extremidades son medianas. Este tipo de animales poseen una mayor facilidad en la adaptación del clima, alimentación y manejo.



Figura 4.4: Aptitud de doble propósito. Tomada de [3].

En el Cauca, aunque se cuente con abundantes hectáreas de diferentes fertilidades y condiciones térmicas apropiadas para el cultivo y desarrollo de la ganadería, se presentan muchas falencias en materia tecnológica y económica, en donde los movimientos migratorios de campesinos desplazados y los efectos colaterales del conflicto armado que se ha presentado en el país, son las principales causas de estas falencias. Sin embargo, el emprendimiento de los pequeños y medianos productores da paso a nuevas oportunidades de intervención por parte de la ingeniería y la implementación de nuevas metodologías que permitan mejorar la competitividad a la población campesina.

Lo anterior se diseña y se plantea acorde con el manual de Buenas Prácticas de Ganadería (BPG) y con las legislaciones pertinentes al manejo de alimentos para el consumo humano establecidas por la ley colombiana mencionados en la Sección 4.1.8.

4.1.3. Sistemas de explotación del ganado vacuno

Indiferentemente de su aptitud productiva, el ganado vacuno debe situarse en una infraestructura o ambiente apropiado para su desarrollo. Sin embargo, por cuestiones geográficas, geológicas y/o climatológicas, los animales son criados mediante métodos de explotación como el estabulado, semiestabulado y el silvopastoril:



Figura 4.5: Sistema de explotación estabulada. Tomada de [3].

El sistema “estabulado” (Ver Figura 4.5) es una forma de crianza de ganado en la cual los animales pasan la mayor parte del tiempo en establos donde realizan sus actividades diarias. En estos establos se recrea la vida del ganado con la diferencia que el mismo se encuentra protegido bajo techo sin tener exposición directa al sol y a las condiciones medio ambientales. En este sistema se pretende una mayor producción y mejor calidad de la carne en el menor tiempo posible [19]. Por su parte el sistema “silvopastoril” (ver Figura 4.6)’ es una forma de cultivo agropecuario que involucra la presencia de árboles interactuando con gramíneas y los animales sometidos a un manejo determinado para incrementar la productividad y el beneficio neto de la explotación a mediano y corto plazo [20].



Figura 4.6: Sistema de explotación silvopastoril. Tomada de [4].

Finalmente, en el sistema semiestabulado se combinan los 2 sistemas mencionados. En este sistema mixto, los animales ingieren sus raciones de alimento principal en áreas extensas de pastos y forrajes vegetales, mientras que el cuidado y el suministro especializado de vitaminas y proteínas se realiza bajo techo.

4.1.4. Ciclo productivo de la leche

El ciclo productivo de la leche de un ganado destinado para tal fin consta de 3 etapas principales que abarcan desde la gestación de la cría y máxima producción de leche, seguido de un periodo de declive y persistencia de producción y finalizando con el periodo de secado en donde se da descanso a la vaca para que se regenere el tejido glandular productor de leche para la siguiente lactancia [21].

Estas etapas son brevemente descritas a continuación:

- **0**Etapa de crianza (Entre 0 y 2 años): Es la etapa de producción temprana en donde el animal (generalmente crías hembras denominadas terneras) son alimentadas y criadas hasta alcanzar 2 años de edad.
 - No requiere de estaciones climatológicas específicas y se puede dar en todo el año (Esta característica varía dependiendo de la ubicación geográfica y las exposiciones climatológicas).
 - Los elementos principales en la dieta de estas criaturas son los nutrientes provenientes del Calostro, la leche materna, forraje y vitaminas.

- **Etapa de gestación y ordeño (Desde los 2 años en adelante y tras el primer parto respectivamente):** Es la etapa inmediatamente siguiente a la etapa de crianza, en donde la novilla esta desarrollada y en la capacidad de ser inseminada para iniciar su producción de leche.
 - El objetivo de esta etapa es que la novilla produzca leche mientras que se da la gestación del novillo dentro de ella.
 - El alimento se basa en pasturas de calidad, provenientes de las extensiones de tierra donde comen las reses y se crían.
 - El animal gana mayor peso debido a su etapa de crecimiento, por lo que entre mejor sea su alimentación se obtendrán mejores resultados.
 - Es de vital importancia que el animal se encuentre en la capacidad de alimentarse Ad-Libitum, es decir a placer y a voluntad.
 - En caso de estar gestando su primera cría, la novilla es denominada “primípara”, pues se encuentra abarcando el tiempo de gestación de su primera cría.

- **Etapa seca o de secado:** Es la etapa final que abarca los últimos 2 meses de gestación previos al parto en donde se da paso a la regeneración de tejidos glandulares que producen leche y a la curación de las glándulas mamarias para reducir el riesgo de mastitis. En sus características principales se pueden mencionar que aún cuando la novilla primípara esta en capacidad de producir leche, no se debe recurrir al ordeño hasta que no se realice el parto.

4.1.4.1. Razas

Debido a la gran variedad de especies, condiciones ambientales, y al proceso evolutivo del ganado, históricamente se ha podido clasificar y seleccionar las razas más representativas para este tipo de explotación ganadera. Más precisamente se puede hacer mención de las siguientes:

- Simmental
- Jersey
- Holstein
- Normando

Éstas son seleccionadas principalmente por la contextura voluminosa pero factores como la calidad de la leche pueden sobrepasar los criterios cuantitativos. A continuación se muestran algunos ejemplares de estas razas:



Figura 4.7: En la primer fila de izquierda a derecha se observan ejemplares de las razas Normando, Simmental; seguido de las razas Holstein y Jersey en la fila inferior. Tomada de [4]

4.1.5. Ciclo productivo de la carne

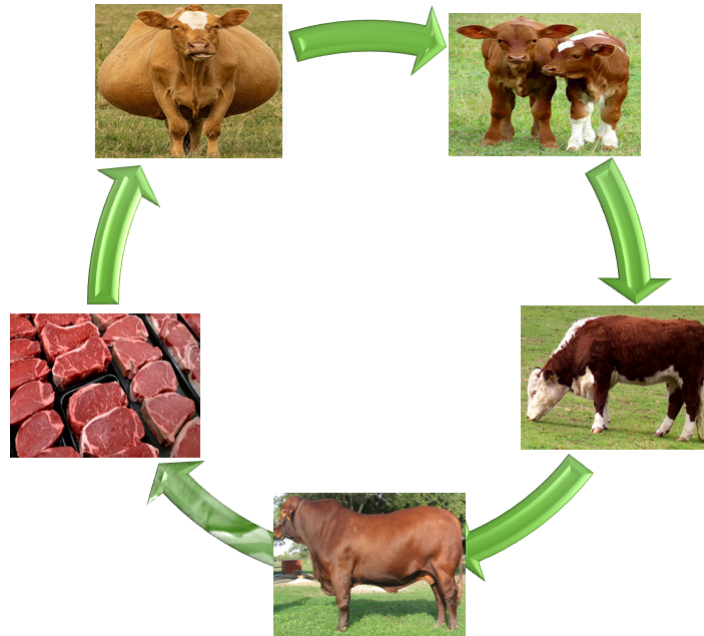


Figura 4.8: Ciclo productivo en la ganadería de la carne. Modificada de [4].

El ciclo productivo de la carne (Ver Figura 4.8) de un ganado destinado para tal fin consta de 3 etapas principales que abarcan desde el nacimiento de la cría, su desarrollo y crecimiento y finalizando con su comercialización en el momento en que alcanza las condiciones apropiadas para ser transformado en carne para consumo humano [21]. Estas etapas son brevemente descritas a continuación:

- **Ganadería de Cría:** Es la etapa de producción temprana en donde el animal (generalmente crías macho llamadas bovinos) es alimentado y criado desde su nacimiento hasta los primeros seis (6) meses de edad.
 - En esta etapa se requiere de grandes extensiones de tierra para producir crías
 - No requiere de estaciones climatológicas específicas y se puede dar en todo el año.
 - Los elementos principales en la dieta de estas criaturas son los nutrientes provenientes del Calostro y la leche materna.

- **Ganadería de Levante:** Es la etapa inmediatamente siguiente a la etapa de crianza, en donde el bovino se desarrollará entre los siete (7) y diez y ocho (18) meses de edad.

- En esta etapa se estima que el peso del animal puede alcanzar un mínimo de 230[kg] de peso en adelante.
 - Es considerada como la etapa más rentable debido a las pocas exigencias en materia de calidad alimenticia.
 - El objetivo de esta etapa es que el sujeto en cuestión alcance el peso deseado en el menor tiempo posible con el menor esfuerzo posible
 - El alimento se basa en pasturas de calidad, provenientes de las extensiones de tierra donde comen las reses y se crían.
 - El animal gana mayor peso debido a su etapa de crecimiento, por lo que entre mejor sea su alimentación se obtendrán mejores resultados.
 - Es de vital importancia que el animal se encuentre en la capacidad de alimentarse *Ad Libitum*, es decir a placer y a voluntad.
- **Ganadería de engorde o Ceba:** Es la etapa final que abarca desde los 19 meses de edad hasta los 24 o 36 meses, dependiendo de su crecimiento y otros factores como el interés del productor, la demanda del mercado, entre otras. En sus características principales se pueden mencionar:
- El límite se define por los intereses del productor, la demanda del mercado y en general por el peso ideal del animal que es aproximadamente mayor o igual a los 450[kg].
 - Para la alimentación se requieren de buenas, grandes y controladas cantidades con el fin de mejorar la carne. Para esto se utilizan dietas que requieren pasturas y concentrados dietarios.
 - Se debe monitorear la alimentación tomando registro en la ganancia de gramaje diaria de cada uno de los animales.

4.1.5.1. Razas

Debido a la gran variedad de especies, condiciones ambientales, y al proceso evolutivo del ganado, históricamente se ha podido clasificar y seleccionar las razas más representativas para este tipo de explotación ganadera. Más precisamente se puede hacer mención de las siguientes:

- | | | | |
|--------------|-------------|------------|------------|
| ■ Beefmaster | ■ Simmental | ■ Brangus | ■ Limousin |
| ■ Charolais | ■ Angus | ■ Hereford | ■ Cebú |

Éstas son seleccionadas principalmente por la contextura voluminosa pero factores como la calidad de la carne pueden sobrepasar los criterios cuantitativos. A continuación se muestran algunas de las razas más usadas para este propósito:



Figura 4.9: En la primer fila de izquierda a derecha se observan ejemplares de las razas Charolais, Hereford; seguido de las razas Angus y Criolla en la fila inferior. Tomada de [4].

4.1.6. Alimentación y Componentes básicos de la dieta

La alimentación de un cultivo de ganado requiere de una dieta o ración con diferentes componentes básicos o nutrientes que deben ser suministrados día a día de forma balanceada para lograr un crecimiento óptimo y que los animales puedan expresar su potencial genético [22].

Los componentes principales que conforman la dieta alimenticia del ganado son:

4.1.6.1. Agua

Componente principal de la alimentación. Esta debe ser suministrada en cantidad y calidad para ser aprovechada por cada animal llegando a ocupar más del 50 % de la masa corporal de un ejemplar adulto y hasta un 90 % de un recién nacido.



Figura 4.10: Hidratación del ganado. Tomada de [4].

4.1.6.2. Energía

Este componente se suministra mediante azúcares, almidones, celulosa, entre otros, los cuales aportan grandes cantidades de energía mas no de proteína, razón por la cual se deben suministrar de forma complementaria.



Figura 4.11: Suplementos energéticos.

4.1.6.3. Proteínas

Estos nutrientes son fundamentales especialmente durante los periodos de sequía, por consiguiente, se optan por fuentes altas en proteína como leguminosas forrajeras, el Maní, Leucaena y el más común, los pastos de forraje verde.



Figura 4.12: Suplementos proteínicos. Tomada de [2].

4.1.6.4. Minerales

Son indispensables en la ganancia de peso de los novillos durante la etapa de Cría y Levante. Este complemento alimenticio debe estar siempre a la disposición para que el ganado pueda abastecer sus necesidades. Estos minerales se suelen proporcionar mediante mezclas de macrominerales y microminerales que se ofrecen de libre consumo al ganado.

4.1.6.5. Vitaminas

Suministradas en cantidades pequeñas aplicadas comúnmente en animales cuya alimentación se basa en forrajes secos, o en animales enfermos convalecientes, desnutridos o durante épocas de sequía prolongada.



Figura 4.13: Forraje y leguminosas. Tomada de [2].

4.1.7. Balance de raciones y dietas especializadas

Estas dietas son suministradas por personal técnico calificado que prepara una dieta acorde a la cantidad de nutrientes de cada animal de forma particular e independiente, considerando su peso actual, su velocidad de crecimiento y estado fisiológico.

4.1.8. Leyes y Normatividad

Requerida para que el avance ganadero se realice de manera coordinada o estandarizada basada en las buenas prácticas ganaderas (Decreto 072 de 2007) en pro de mejorar la productividad y ayudarla a alcanzar niveles de ganadería bovina del mundo [23], [24], [25].

- **Resolución 0012 de 2007:** Por la cual se establece el sistema de pago de la leche cruda al productor, diseñado por la Unidad de Seguimiento de Precios.

4.2. Modelación matemática

4.2.1. Regresión lineal

La regresión lineal es un método estadístico usado para modelar la relación entre una variable dependiente denotada como y y una o más variables dependientes denotadas como x_i . El objetivo de este método es encontrar una ecuación lineal que mejor represente cómo las variables independientes influyen a la variable dependiente. La ecuación lineal estimada se puede definir como en la Ecuación #4.1:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \varepsilon$$

$$y = \beta_0 + \sum_{i=1}^n (\beta_i x_i) + \varepsilon \quad (4.1)$$

Donde y es la variable dependiente, x_i son las variables independientes, β_i son los coeficientes a estimar; y ε representa el término de error por posibles variaciones en la variable dependiente y . El objetivo de este método es encontrar los valores de los coeficientes β_i que minimizan la suma de las diferencias cuadradas entre los valores observados de la variable dependiente y y los valores predichos por la ecuación. Típicamente se logra utilizando el método de regresión por mínimos cuadrados.

4.2.2. Regresión no lineal

En comparación al método de regresión lineal, cuando la variable dependiente posee una naturaleza no lineal, el método estadístico usado para encontrar las relaciones entre las variables independientes x_i y la variable dependiente y es denominado como regresión no lineal. En este tipo de regresión se utiliza una función $f(x_i; \theta)$ con variables independientes x_i y parámetros θ , que puede tener distintas formas dependiendo del problema. Una forma general de describir un modelo de regresión no lineal es:

$$y = f(x_i; \theta) + \varepsilon \quad (4.2)$$

Donde y es la variable dependiente, x_i son las variables independientes, θ_i son los parámetros a estimar; y ε representa el término de error por posibles variaciones en la variable dependiente y . En este caso el objetivo de este método es encontrar los valores de los parámetros θ y del error ε de tal forma que los valores estimados de y se ajusten a los datos existentes. Aún cuando la regresión lineal es directa e intuitiva, este tipo de aproximación solo es válida cuando la incertidumbre no está relacionada ni con la variable independiente ni con la variable dependiente por lo que no es muy útil para situaciones en las que los datos han sufrido transformaciones [26]. Para ello pueden utilizarse otros tipos de regresión tales como la polinómica, Splines cúbica, entre otras (ver ejemplo comparativo en la Figura 4.14).

4.2.3. Interpolación con Splines

Los splines son un enfoque matemático y computacional usado en el ajuste e interpolación de curvas tanto lineales como no lineales. Dado su naturaleza, las splines son funciones matemáticas que describen el dominio de una curva dividiéndola en segmentos representados por funciones polinómicas por partes de bajo orden. Cuando éstas funciones son de bajo orden, son usadas para representar de manera sencilla a funciones lineales, y cuando son de alto orden (por lo general cúbicas), son usadas para representar de manera precisa a las curvas de funciones no lineales.

4.2.3.1. Splines cúbicas

Los splines cúbicos son splines en los que cada segmento está representado por una función polinómica cúbica, siendo este el grado máximo del polinomio en cada tramo. Suponiéndose un conjunto de puntos de datos (x_i, y_i) , se utilizan splines cúbicos para construir un polinomio cúbico por partes que pasa por cada dupla de valores, además de existir primera y segunda derivada continua en cada pareja de datos, de esta forma se asegura una curvatura suave y visualmente atractiva (ver Figura 4.14). La definición matemática de cada segmentos entre cada par de valores (x_i, x_{i+1}) se representa de la siguiente manera:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (4.3)$$

Donde $S_i(x)$ es la función de spline cúbica para el segmento i -ésimo, a_i , b_i , c_i y d_i son los coeficientes específicos para el segmento i -ésimo.

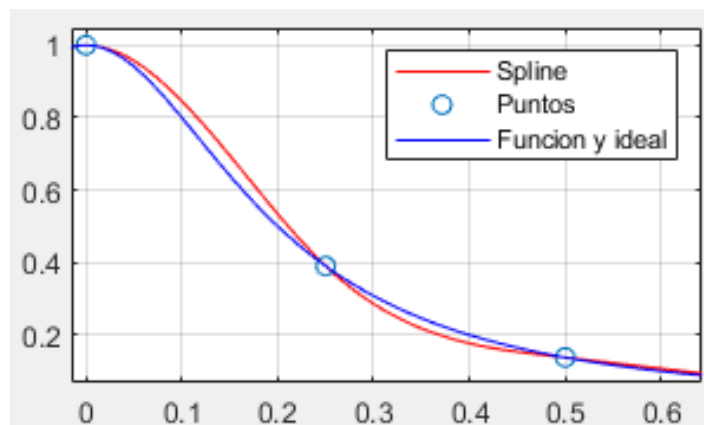


Figura 4.14: Comparación entre una interpolación de datos por Splines Cúbicas vs Una función polinómica ideal

4.2.4. Ajuste de curvas

El ajuste de curvas es un proceso matemático y estadístico que determina la curva o función de mejor ajuste que describe a un conjunto de puntos. El objetivo principal de esta técnica consiste en encontrar el modelo matemático que se aproxime a la relación existente en los datos entre la variable independiente y la variable dependiente. De esta forma puede que se establezcan valores de la variable dependiente que pueden no estar presentes en el conjunto de datos original. El ajuste de curvas también involucra la estimación de valores inexistentes o ausentes entre los puntos de datos haciendo uso de técnicas de interpolación. El modelo matemático utilizado está determinado por la naturaleza de los datos y el problema que se desea satisfacer, y este ajuste puede ser logrado tanto por modelos lineales así como también modelos no lineales que dependerán de la naturaleza de la conexión entre la variable independiente, comúnmente x y la variable dependiente, comúnmente y .

4.2.5. Métodos de ajuste no lineal

En los casos en los que un modelo lineal no es capaz de capturar efectivamente la conexión entre las variables, se aplican métodos de ajuste de curvas no lineales. Estos métodos implican ajustar funciones no lineales a los datos para capturar relaciones más complejas [27]. Entre los métodos más comunes de ajuste de curvas no lineales podemos encontrar al ajuste por mínimos cuadrados no lineales (en inglés, Nonlinear Least Squares ó NLS), en donde éste método minimiza la suma de las diferencias cuadradas entre los valores observados y los valores estimados, el ajuste de suavizado de curvas exponenciales, usualmente utilizado cuando los datos evidencian un crecimiento o decrecimiento exponencial, el ajuste de curvas logísticas, frecuentemente usados en casos donde se presentan modelos de crecimiento y/o fenómenos de saturación con curvas de comportamiento sigmoideo, el método de Gauss-Newton, comúnmente usado en ingeniería para lograr la regresión no lineal mediante una optimización iterativa; entre otros métodos explicados a continuación.

4.2.5.1. Algoritmo de “Levenberg-Marquardt” (LMA)

El algoritmo de Levenberg-Marquardt (LMA), comúnmente conocido como el enfoque de mínimos cuadrados amortiguados, en inglés: “Damped Least Squares” (DLS, por sus siglas en inglés), es una metodología matemática y computacional utilizada para manejar problemas de mínimos cuadrados no lineales, especialmente el ajuste de curvas. Cuando intentamos identificar los parámetros de mejor ajuste para una curva modelo dada que minimiza la suma de los cuadrados de las desviaciones entre los puntos de datos observados y la curva, nos encontramos con varias dificultades de minimización. Si un modelo es lineal en sus parámetros, el objetivo de mínimos cuadrados es cuadrático en los parámetros.

Este objetivo puede minimizarse con respecto a los parámetros en un solo paso mediante la solución de una ecuación matricial lineal. Este tipo de solución reduce la suma de los cuadrados de los errores entre la función del modelo y los puntos de datos a través de una secuencia de actualizaciones bien elegidas de los valores de los parámetros del modelo [28]. Siguiendo el método

de Gauss-Newton (GNA), la suma de los errores al cuadrado se reduce asumiendo que la función de mínimos cuadrados es localmente cuadrática en los parámetros y encontrando el mínimo de esta cuadrática. Esta medida de bondad de ajuste con valores escalares se denomina criterio de error de chi-cuadrado (χ^2), porque la suma de los cuadrados de las variables aleatorias normalmente distribuidas se distribuye como la distribución de (χ^2). Si la función y es no lineal en los parámetros del modelo, entonces la minimización con respecto a los parámetros debe realizarse de forma iterativa [28].

El ajuste de curvas es un proceso que se utiliza para encontrar una función matemática que mejor se ajuste a un conjunto de datos experimentales. Una de las técnicas más populares para el ajuste de curvas es el algoritmo de Levenberg-Marquardt (LMA), que es una combinación de los métodos de GNA y el método del gradiente conjugado [28]. El método de LMA es muy eficiente en la resolución de problemas de ajuste de curvas no lineales, y es ampliamente utilizado en diversas áreas de la ciencia incluyendo la física, la química, la biología, la economía y la ingeniería.

En términos generales, el método de LMA se basa en un enfoque iterativo que comienza con una estimación inicial de los parámetros de la función de ajuste y se ajusta sucesivamente hasta que se alcanza una solución que minimiza la función de error cuadrático, incluso si comienza lejos del mínimo final. Este método utiliza una matriz jacobiana para calcular la dirección del gradiente y la tasa de aprendizaje de cada paso iterativo. Si la tasa de aprendizaje es grande, el método funciona como el método del gradiente conjugado. Si la tasa de aprendizaje es pequeña, el método funciona como el método de Gauss-Newton. Por lo tanto, el método de Levenberg-Marquardt es una combinación de ambos métodos [28]. Este problema puede formularse matemáticamente como:

$$\text{Min} \quad \sum_{i=1}^N [y_i - f(x_i; \theta)]^2, \quad (4.4)$$

Donde y_i es la variable dependiente observada, y $f(x_i; \theta)$ representa la función del modelo con los parámetros θ . El objetivo es minimizar la suma de los cuadrados de las desviaciones entre la variable dependiente observada y_i y la función modelo $f(x_i; \theta)$ con parámetros θ , tomados en todos los puntos de datos ($i = 1$ a N).

4.2.5.2. Algoritmo de “Efectos Mixtos No Lineales” (NLMEA)

Otro de los métodos de ajuste más utilizados para el análisis individualizado de múltiples objetivos dependientes del tiempo es el ajuste por efectos mixtos lineales y no lineales. Estos modelos de efectos mixtos se basan en un modelo generalizado para describir poblaciones aún cuando los parámetros de crecimiento del modelo pueden ser únicos para los sujetos de estudio [29]. El modelado con el Algoritmo de Efectos Mixtos No Lineales (Nonlinear Mixed Effects Algorithm ó NLMEA, por sus siglas en inglés) es una técnica estadística para analizar datos que presentan mediciones repetidas o patrones jerárquicos. Es especialmente útil cuando se trata con datos que contienen varianzas

individuales y heterogeneidad. El método NLMEA amplía los modelos clásicos de regresión no lineal al incluir efectos fijos (Nonlinear Mixed Effects Algorithm - Fixed Effects ó NLMEA-FE, por sus siglas en inglés) donde se identifican parámetros a nivel de población; así como efectos aleatorios (Nonlinear Mixed Effects Algorithm - Random Effects ó NLMEA-RE, por sus siglas en inglés) donde se identifican parámetros específicos asociados a cada sujeto estudiado, lo que permite tener en cuenta tanto la variabilidad entre individuos como la variabilidad dentro de cada individuo en el tiempo. Para este método, agregar efectos aleatorios a un modelo extiende la confiabilidad de las inferencias más allá de la muestra específica de individuos.

Considerando un modelo no lineal con parámetros específicos para cada individuo, el modelo NLMEA se representa matemáticamente de la siguiente manera:

$$y_{ij} = f(x_{ij}; \theta_i) + \varepsilon_{ij} \quad (4.5)$$

Donde y_{ij} es la respuesta observada para la medición j -ésima en el individuo i -ésimo. x_{ij} Es la correspondiente variable predictora para la medición j -ésima en el individuo i -ésimo. $f(x_{ij}, \theta_i)$ representa el modelo no lineal con parámetros específicos para cada individuo (θ_i). ε_{ij} es el termino de error aleatorio que se asume con distribución normal con media cero (0) y varianza constante.

El modelado por NLMEA intenta estimar tanto los parámetros a nivel de población como los parámetros específicos de cada individuo al mismo tiempo. El modelo incorpora la variabilidad poblacional al tener en cuenta influencias aleatorias, lo que permite calcular la variación individual aparte de reconocer las correlaciones dentro de los subgrupos de muestra. El enfoque NLMEA se utiliza con frecuencia en una variedad de campos, incluyendo la farmacocinética, la investigación longitudinal y el modelado de crecimiento, donde la variabilidad individual y las mediciones repetidas son propiedades comunes de los datos.

4.2.6. Modelación por “Ecuaciones Diferenciales Ordinarias” (EDOs)

Este es un enfoque matemático utilizado para representar, analizar y evaluar sistemas dinámicos en los que el comportamiento de un sistema está influenciado por su estado actual, su evolución en el tiempo, los factores de entrada involucrados y las respuestas a estos estímulos. Las EDOs son ampliamente usadas en múltiples disciplinas científicas tales como la ingeniería, la física, la biología, entre otras. Las EDOs permiten modelar los cambios presentes en un sistema en donde se presentan tasas de cambio de una o más variables dependientes con respecto a sus valores existentes.

Dependiendo del orden del modelo, las EDOs pueden clasificarse en primer orden al considerar la primer derivada de la variable, ó de orden superior cuando se involucran derivadas de orden 2 o superior [30].

4.2.6.1. Planteamiento de un modelo por EDOs

Para proponer un modelo por ecuaciones diferenciales ordinarias es necesario tener en cuenta las siguientes consideraciones:

1. **Definir el sistema:** Se deben definir el sistema que se desea modelar, las variables clave y las relaciones existentes entre éstas.
2. **Formular las EDOs:** Se debe escribir las ecuaciones que describen los cambios o evoluciones que presentan las variables con respecto a otra variable, comúnmente dependientes del tiempo.
3. **Especificar parámetros:** Se deben asignar valores o rangos para los parámetros incluidos en las EDOs, estos parámetros influenciarán significativamente al comportamiento del sistema.
4. **Condiciones iniciales:** Estos son los valores iniciales de las variables dependientes en la primer muestra, comúnmente en tiempo $t = 0$.
5. **Viabilidad:** Dependiendo de la complejidad del modelo, es posible que pueda ser solucionado de manera determinista, o mediante el uso de métodos numéricos y simulaciones computacionales.
6. **Validación y análisis:** Verificar si es posible validar el modelo frente a datos o resultados experimentales, analizando el comportamiento del modelo frente a diferentes condiciones.

4.2.7. Criterios de evaluación

4.2.7.1. Error Cuadrático Medio (MSE) y Raíz de Error Cuadrático Medio (RMSE)

El MSE y el RMSE son métricas típicamente utilizadas en estadística y en modelos de aprendizaje automático (“Machine learning” en inglés) que sirven para evaluar la precisión del modelo estimador. Por su parte, el MSE mide la diferencia cuadrática promedio entre los valores observados (reales) y los valores esperados (teóricos). El MSE se calcula como el promedio del cuadrado de los residuos tal y como se observa a continuación:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.6)$$

Este criterio es comúnmente usado en problemas de regresión para evaluar la bondad de ajuste del modelo. En pocas palabras, la interpretación de esta métrica indica que entre menor sea el valor, mejor es el ajuste. No obstante, Es importante considerar que al elevar los errores al cuadrado, se penalizan los errores más grandes a comparación de los errores pequeños ocasionando que este criterio sea sensible a datos atípicos. Por su parte el RMSE se obtiene aplicando la raíz cuadrada al MSE, este proceso se realiza teniendo en cuenta que mejora su interpretación puesto que las unidades de esta métrica son las mismas que las de los datos usados [31].

4.2.7.2. Máxima verosimilitud ($\log \mathcal{L}$)

La estimación por máxima verosimilitud es una medida que permite saber qué tan bien un modelo estadístico explica los datos observados. Este criterio cuantifica la probabilidad de observar los datos bajo los supuestos del modelo. Se cree que los modelos que poseen mayores probabilidades logarítmicas se ajustan mejor a los datos. En muchos escenarios, la estimación de parámetros estadísticos implica maximizar la probabilidad logarítmica, como resultado se obtienen estimaciones de máxima verosimilitud de los parámetros del modelo. Para calcular la máxima verosimilitud de un modelo se debe:

1. Calcular la verosimilitud para un única observación x_i obteniéndose:

$$\mathcal{L}(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

2. Suponiéndose que hay independencia entre los datos, se calcula la verosimilitud de todo el conjunto de datos de la siguiente manera

$$\mathcal{L}(x_1, x_2, \dots, x_n|\mu, \sigma) = \prod_i^n \mathcal{L}(x_i|\mu, \sigma)$$

3. Finalmente, se procede a determinar el logaritmo natural para obtener la verosimilitud logarítmica, obteniéndose que:

$$\log \mathcal{L}(x_1, x_2, \dots, x_n|\mu, \sigma) = \sum_i^n \left[-\left(\frac{1}{2}\right) \log(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] \quad (4.7)$$

4.2.7.3. Criterio de Información de Akaike (AIC)

El criterio de información de Akaike en un modelo estadístico es la medida para la calidad de ajuste del modelo. Este logra un balance entre ajuste y complejidad del modelo, esto quiere decir, un balance entre una buena captura de datos y el número de parámetros usados. El AIC mide qué tan bien se ajusta el modelo a los datos y al mismo tiempo penaliza la complejidad para evitar el sobreajuste (“overfitting” en inglés). En esencia, el criterio AIC es utilizado para comparar y elegir modelos candidatos y dependiendo de su valor, se puede sugerir un mejor equilibrio entre el ajuste y la simplicidad del modelo [32]. Este criterio se puede calcular mediante la siguiente ecuación:

$$AIC = -2\ln(\mathcal{L}) + 2k \quad (4.8)$$

Donde \mathcal{L} es la función de máxima verosimilitud, k es la cantidad de parámetros, para un tamaño de muestra n .

4.2.7.4. Criterio de Información Bayesiano (BIC)

También conocido como el criterio de información de Schwarz, es un criterio similar al AIC pero éste criterio aplica una mayor penalización en la complejidad del modelo. El uso de este criterio está enfocado en la selección de modelos, especialmente cuando se necesita una penalización más estricta por la complejidad del modelo. A diferencia del AIC, el BIC favorece a los modelos más simples, lo que lo hace más eficaz al momento de seleccionar un modelo candidato [32]. Para calcular este criterio, se debe:

$$BIC = -2 \ln(\mathcal{L}) + 2k \ln(n) \quad (4.9)$$

Donde \mathcal{L} es la función de máxima verosimilitud, k es la cantidad de parámetros, para un tamaño de muestra n .

4.2.7.5. “Chi-cuadrado” (χ^2) y “Chi-cuadrado” reducido (χ_{red}^2)

- **Factor χ^2 :** Este es un número real que **no puede ser negativo** y se determina sumando las diferencias al cuadrado entre los valores observados y predichos, dividiendo el resultado por los valores esperados (ver Ecuación 4.10). El objetivo principal de la prueba de χ^2 se utiliza para determinar si existe o no una diferencia estadísticamente significativa entre los valores observados y estimados. Un valor de χ^2 bajo sugiere una estrecha coincidencia entre los valores observados y esperados, lo que indica un buen ajuste, en caso contrario el ajuste sería deficiente.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (4.10)$$

Donde O_i representa el valor observado para un valor específico y E_i representa el valor esperado para un valor en específico. Es importante considerar los grados de libertad (degrees of freedom en inglés ó DOF) que se definen como el número de categorías o valores menos 1, y esto es lo que determina la forma de la distribución χ^2 . Por En cuanto a la bondad de ajuste, χ^2 es usado para evaluar qué tan bien los datos observados se ajustan a los valores estimados por el modelo teórico.

- **Factor χ^2 reducido:** Este es una versión modificada del factor χ^2 que hace referencia al número de grados de libertad en un test estadístico. Este valor es comúnmente usado para comparar modelos con distinta cantidad de parámetros con el objetivo de evitar el sobreajuste de los datos. Para calcular el factor χ^2 reducido se debe dividir el factor χ^2 , por el número de grados de libertad:

$$\chi_{red}^2 = \frac{\chi^2}{DOF} \quad (4.11)$$

El factor decisivo radica en escoger el modelo que cuente con el valor más pequeño de χ_{red}^2 implicando que ese modelo brinda un mejor balance entre la bondad de ajuste y la simplicidad del modelo.

4.3. Descripción general del proyecto realizado bajo la metodología CRISP-DM

Tal y como se mencionó en la Sección 3.2 La metodología CRISP-DM se compone de distintas etapas, comenzando desde el entendimiento del negocio hasta completar el despliegue e implementación de los modelos obtenidos. Para tener una perspectiva general del desarrollo general de este proyecto se plantea un diagrama de bloques general evidenciado en la Figura 4.15. En éste diagrama se quiere mostrar cómo se ha planteado el desarrollo de las etapas de Entendimiento de los datos, Preparación de datos, Modelación y Evaluación, las cuales componen la estructura principal del desarrollo de este proyecto. La primer etapa no son consideradas dentro de este diagrama en tanto que la etapa inicial de Entendimiento del negocio, la cual se logra bajo el entendimiento del contexto al rededor del negocio, no están bajo el control del autor; esto mismo sucede para la etapa de despliegue, pues esta decisión depende únicamente del SENA quien tomará la decisión de implementar (o no) las ventajas que ofrece este proyecto de investigación.

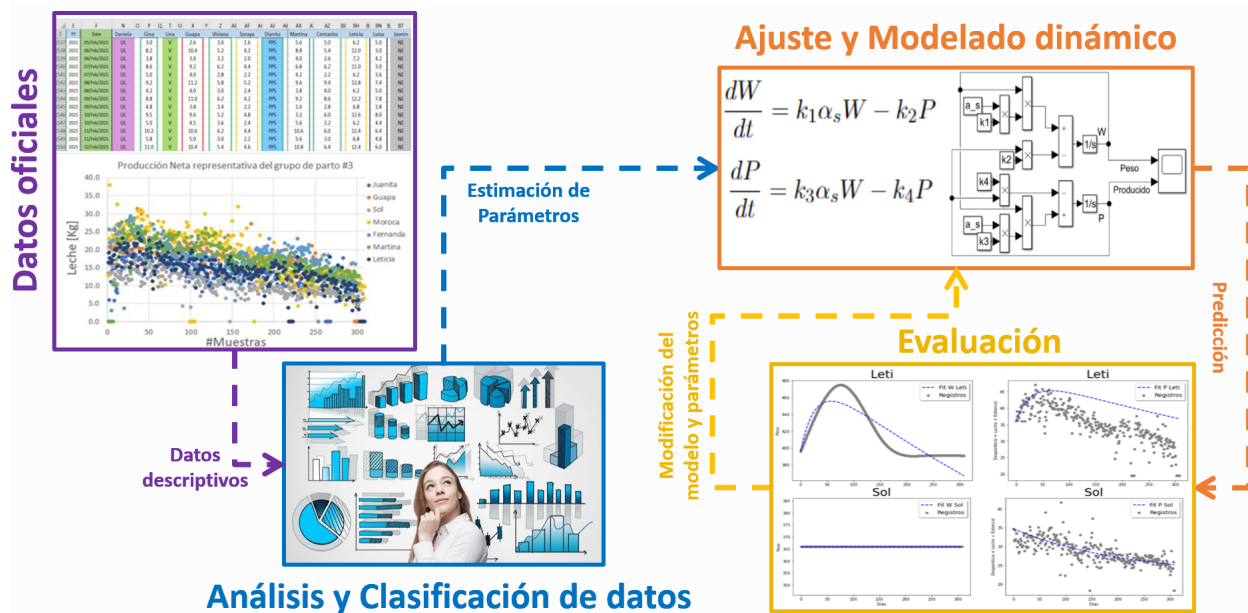


Figura 4.15: Planteamiento general de las etapas 2, 3, 4 y 5 de la metodología usada en la ejecución del proyecto.

Es importante considerar este diagrama de bloques como una guía para comprender el plan de acción llevado a lo largo de éste proyecto de investigación. No obstante, cada una de las etapas previamente mencionadas estarán explicadas en mayor detalle a lo largo del desglose de las etapas de la metodología CRISP-DM de Entendimiento de los datos (ver Capítulo 6), Preparación de los datos (ver Capítulo 7), Modelación (ver Capítulo 8) y Evaluación (ver Capítulo 9).

Entendimiento del negocio

5.1. Servicio Nacional de Aprendizaje (SENA)

El SENA es un establecimiento público de educación en Colombia que ofrece formación gratuita con programas técnicos, tecnológicos y complementarios; que busca la incorporación y el desarrollo de las personas en actividades productivas que contribuyan al desarrollo social, económico y tecnológico del país. Entre las diferentes áreas de formación ofrecidas por el SENA, se encuentra el área de las ciencias agropecuarias, siendo esta la que tiene mayor influencia en la institución.[33]

5.1.1. Centro Agropecuario

El centro agropecuario (CA) tiene como objetivo liderar la formación de los sub-sectores agrícola, pecuario, pesquero y forestal de la región; mediante alianzas estratégicas con agremiaciones, administraciones municipales, campesinas y centros de educación técnica, media y superior. El CA incorpora tecnologías alternativas en las explotaciones agropecuarias para mejorar la calidad de los productos, reduciendo a su vez costos de producción. Dentro de los servicios ofertados por el CA-SENA se encuentran la producción y comercialización bovina y bufalina, procesamiento de alimentos y control de calidad (Cárnicos y Lácteos), entre otros[34].

5.2. Empresa donde se realiza el estudio: CA-SENA-POP

El CA-SENA seccional Cauca, se encuentra ubicado en el norte de la ciudad de Popayán. Cuenta con una granja de aproximadamente 6 hectáreas en las que realizan sus actividades productoras con unidades de producción bovina, porcina, avícola, bufalina, caprina, entre otras. Su formación y capacitación consta en gran medida de las prácticas realizadas por los aprendices en cada una de las unidades productoras. La granja se encuentra dividida en 2 zonas, el área usada para las instalaciones de la seccional Popayán y los lotes donde se realizan las diferentes prácticas y servicios mencionados con anterioridad.

Se cuenta con un total de 50 lotes de aproximadamente $500m^2$ cada uno (ver figura 5.1). 30 de ellos son usados para la producción silvopastoril bovina, 10 de ellos son usados para la producción de frutas, hierbas y hortalizas, entre ellas el botón de oro que es usado como forraje complementario en la dieta del ganado vacuno; y los lotes restantes son usados por las otras unidades productoras. Enfocándonos hacia la producción bovina, el CA-SENA-POP, se dedica a explotación del ganado vacuno con fines de producción lechera. Tal y como se ha mencionado antes, manejan un sistema

de explotación silvopastoril semiestabulado, dado que distribuyen a sus animales en los diferentes lotes destinados para tal fin y llevan el registro de las cantidades de alimento concentrado y dietas secas, en un establo donde dosifican las cantidades correspondientes de sales minerales y proteínas concentradas.



Figura 5.1: Granja del CA-SENA-POP.

5.3. Negocio y comercialización de productos

El CA-SENA-POP, realiza la venta de productos ganaderos entre los que se puede mencionar, la producción de leche cruda, reses productoras de leche y terneros para engorde. El CA-SENA-POP ha manifestado que los precios de compra y venta son los precios manejados en el mercado tanto por kilogramo de peso del animal como por kilogramo de leche producida.

En el capítulo siguiente, se realiza una descripción de los datos presentes en el CA-SENA-POP usados para este proyecto de investigación.

Entendimiento de los datos

6.1. Adquisición de datos

6.1.1. Registros de “Leche”

En el establo de alimentación y dosificación de concentrado, es donde se realizan las tareas de ordeño para las vacas que aportan a la producción neta de leche. El ordeño y la extracción de la leche se ha realizado de manera manual desde sus inicios, pero cuentan con sistemas de ordeño mecánico desde los últimos años. El pesaje de la leche se realiza en kilogramos y el registro de los datos de producción lechera se ha realizado de manera manual en diarios históricos de producción neta. Un ejemplar de estos diarios puede observarse a continuación en la Figura 6.1:

REGISTRO PRODUCCIÓN DE LECHE																	
UNIDAD GANADERÍA																	
F008-BOVI-VERSION 2																	
NÚMERO DE ANIMALES	FECHA INICIO DE SEMANA: 20-06-22								FECHA TERMINACIÓN DE SEMANA: 26-06-22								OBSERVACIONES
	NOMBRE/ NÚMERO	DIAS EN LACTANCIA	LUNES		MARTES		MIÉRCOLES		JUEVES		VIERNES		SABADO		DOMINGO		
		AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM	AM	PM		
1	Soraya	6.0	3.0	5.0	3.0	5.2	3.2	5.2	2.8	5.8	3.4	5.4	3.2	5.4	2.0		
2	Antuñito	11.0	5.2	9.4	5.6	6.8	5.6	9.0	5.4	7.2	7.0	7.8	2.4	11.0	4.2		
3	Andreo	10.8	6.8	9.8	6.2	9.6	7.0	9.8	2.0	13.8	5.6	10.6	4.4	10.4	4.6		
4	Guapa	11.6	7.4	12.6	5.6	11.2	6.0	13.0	6.0	13.2	6.2	12.6	5.4	12.4	5.0		
5	Muchiro	4.4	3.4	4.8	3.2	4.4	3.2	4.2	4.0	6.6	3.2	5.0	1.2	6.6	2.2		
6	Pacha	8.8	6.8	9.0	5.6	8.4	4.4	10.0	4.0	11.2	6.0	10.0	3.6	11.0	5.0		
7	Yezmin	9.0	4.0	8.2	4.0	8.8	4.2	7.2	5.4	8.4	5.0	3.8	6.2	8.0	3.0		
8	Dianita	11.4	4.6	10.4	5.4	8.6	6.4	9.0	2.8	11.0	4.0	9.0	4.4	11.4	3.0		
9	Sol	6.2	4.8	7.2	3.0	7.0	2.8	6.2	2.0	7.8	3.8	6.4	3.8	7.8	1.8		
10	Leticia	8.0	4.8	7.2	4.4	8.2	3.0	7.6	2.0	7.6	3.0	7.0	4.0	7.6	2.8		
11	Enfermiza?	4.2	2.0	4.2	4.0	4.0	2.6	3.8	2.4	3.6	2.4	3.8	1.6	6.0	1.4		
12	Yolima	9.0	3.0	6.0	3.0	7.2	4.4	6.8	4.0	11.0	3.8	9.2	2.4	8.0	3.8		
13	Martina	8.8	5.6	9.0	4.0	7.6	5.0	6.0	1.0	10.0	3.8	6.0	4.8	8.6	3.0		
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
PRODUCCIÓN REAL		91	51	93		87	51										
TERNEROS		4	4	4	4	4	4	4	4	4	4	4	4	4	4		
DESCARTE		9.0	3.0	6.0	3.0	7.2	4.4	6.8	4.0	11.0	3.8	9.2	2.4	8.0	3.8		
TOTAL TANQUE																	
TOTAL DÍA																	

Figura 6.1: Registros manuales de producción lechera en el CA-SENA-POP.

Tal y como se puede observar, se lleva un registro diario de cada una de las vacas que se encuentran en etapa de ordeño. Los ordeños se realizan en 2 franjas horarias, en las mañanas, aproximada-

mente a las 6:00 am, que se realiza el ordeño matutino; y en las tardes, aproximadamente a las 4:00 pm, que se realiza el ordeño vespertino. El registro de los datos se realiza en hojas cuadriculadas donde se puede evidenciar, la producción diaria de cada animal, así como también la producción diaria neta de la unidad de ganadería. Los datos aquí registrados también permiten observar si hay leche que ha sido suministrada a los terneros o si ha sido descartada por sospecha de mastitis, entre otros motivos.

Aún cuando se lleva un registro manual de los datos, en los últimos 5 años se dio inicio a una migración de datos de manera tecnológica, mediante licencias de software de pago para manejo de producciones ganaderas, entre ellas la licencia del software “TaurusWebs” (ver Figura 6.2). Sin embargo, esta transición tecnológica no ha sido implementada en su totalidad, por lo que el registro de datos, se realiza de ambas formas en pro de que los datos no se pierdan. Por otra parte, el manejo de datos manual es utilizado para que sirva de material de estudio y aprendizaje en las diferentes actividades y servicios ofrecidos por el SENA, tales como los cursos de formación y capacitación de ganadería.

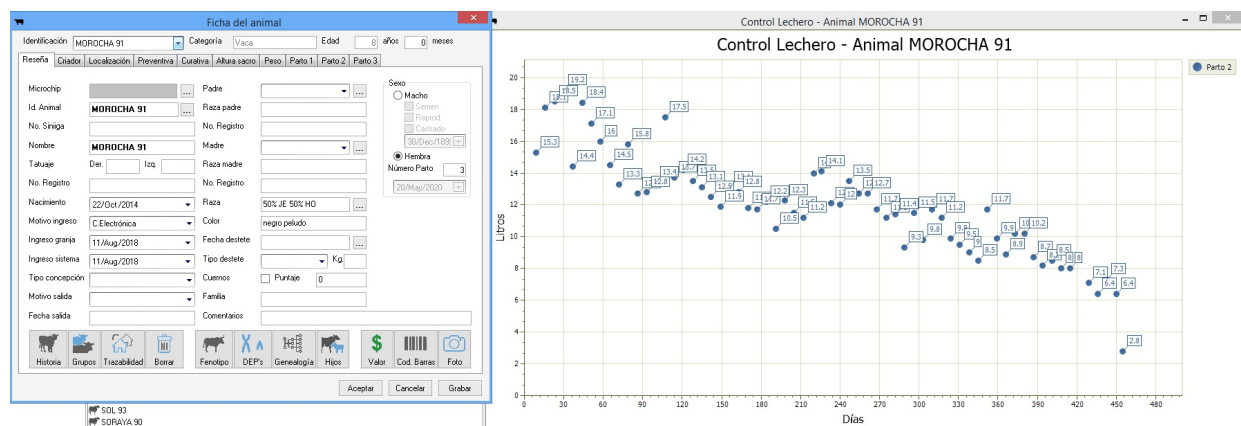


Figura 6.2: Registro de datos digital con software TaurusWebs.

Una de las ventajas de este software es que alberga una gran variedad de aplicaciones de seguimiento entre las que se pueden mencionar el seguimiento a los registros de leche, registros de peso, registros de antecedentes familiares, trazabilidad, historia, genealogía, entre otros; información que será útil para etapas posteriores en este trabajo.

6.1.2. Registros de “Peso”

En cuanto a los registros de peso de los animales, El software “TaurusWebs” mencionado anteriormente, facilita una hoja de datos donde se pueden almacenar los valores asociados al peso, así como también las fechas en las que fueron realizadas; permitiendo representar dichos datos en gráfica de seguimiento como las que se pueden evidenciar en la Figura 6.3. Estos datos guardados

con su fecha de pesaje permiten hacer análisis y algunas observaciones de variabilidad de peso para los distintos partos en los que se agrupan las reses cuyos datos modelan al sistema.

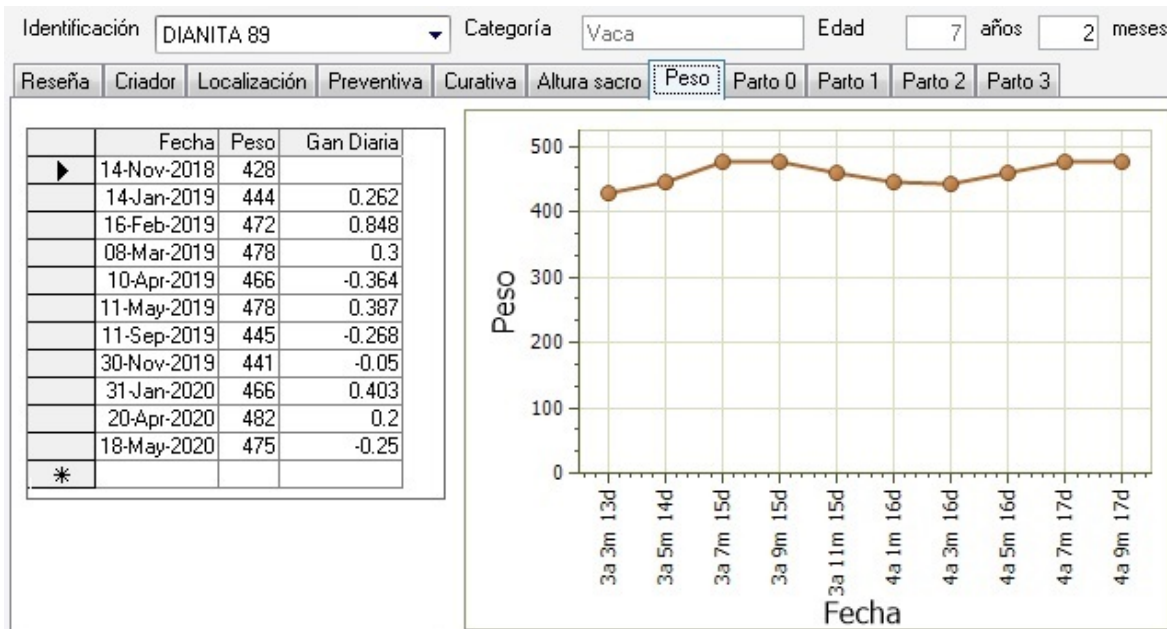


Figura 6.3: Ejemplo de seguimiento digital de peso en la res “Dianita”.

Los datos almacenados en este software, muestran registros de peso bimensual o trimestral, genealogía, suministro de medicamentos y producciones lecheras semanales. No obstante, aún cuando este software ofrece una gran ventaja para el seguimiento de los datos, se encuentra limitado a las pocas licencias ofrecidas para su manejo. Por lo tanto es plausible considerar alternativas para el análisis, seguimiento de datos y toma de decisiones futuras, lo que justifica nuevamente la realización de este proyecto.

Preparación de los datos

En esta etapa, se procede a recolectar todos los datos análogos y digitales con los que cuenta en CA-SENA-POP, para almacenarlos de manera digital y proceder a analizar los datos mediante algoritmos en Matlab y/o Python. Es importante recalcar que para ambos software se requiere una estructura de datos en tablas debidamente organizadas, ya sea para tratarlos como *dataframes*, matrices, arreglos, o como tablas de archivos **.csv*.

7.1. Recolección de datos

Tal y como se mencionó en el capítulo anterior, los datos referentes a Peso, Leche y Alimentación, de cada una de las reses del CA-SENA-POP se encuentran registradas de manera física o digital; ya sea en hojas impresas como el ejemplar de la Figura 6.1, así como también en bases de datos digitales como el de la Figura 6.2. No obstante, estos datos **no** pueden ser importados directamente a algún software de procesamiento de datos. Esto puesto que para los datos físicos, se tienen algunas barreras y/o inconvenientes externos tales como los diferentes tipos de letra, el estado físico de las paginas, y la diferencia entre los números de un ente registrado u otro; mientras que para los datos digitales, estos se encuentran almacenados en el software de pago de manera privada, siendo datos protegidos de acuerdo a la **Ley de “Habeas data”**, por lo que no pueden ser exportados directamente para manipulación del público.

Teniendo en cuenta que los datos de producción lechera se registran de manera manual, esta tarea es llevada a cabo por distintas personas. Desde aprendices hasta instructores y supervisores; por lo tanto, los datos son escritos por distintas personas, que manejan distintos tipos de letra, distinto orden, distintas prácticas adecuadas de escritura, entre otras. Esto implica que los números registrados en las paginas físicas varíen de un tipo de letra a otra y deban ser analizados o copiados rigurosamente de manera digital. Además, es importante mencionar, que estas carpetas físicas de registro de datos, son expuestas a la intemperie, por lo que algunos registros han sido vulnerables ante lluvias, o sustancias que alteran el estado de la hoja de papel, dificultando así su interpretación a primera vista.

Es importante hacer esta observación, dado que por estas mismas circunstancias, **no** es posible extraer los datos de las tablas mediante algoritmos de inteligencia artificial de procesamiento de imágenes de manera consistente, pues estos datos pueden ser fácilmente tergiversados o malinterpretados como los ejemplares de la Figura 7.1.

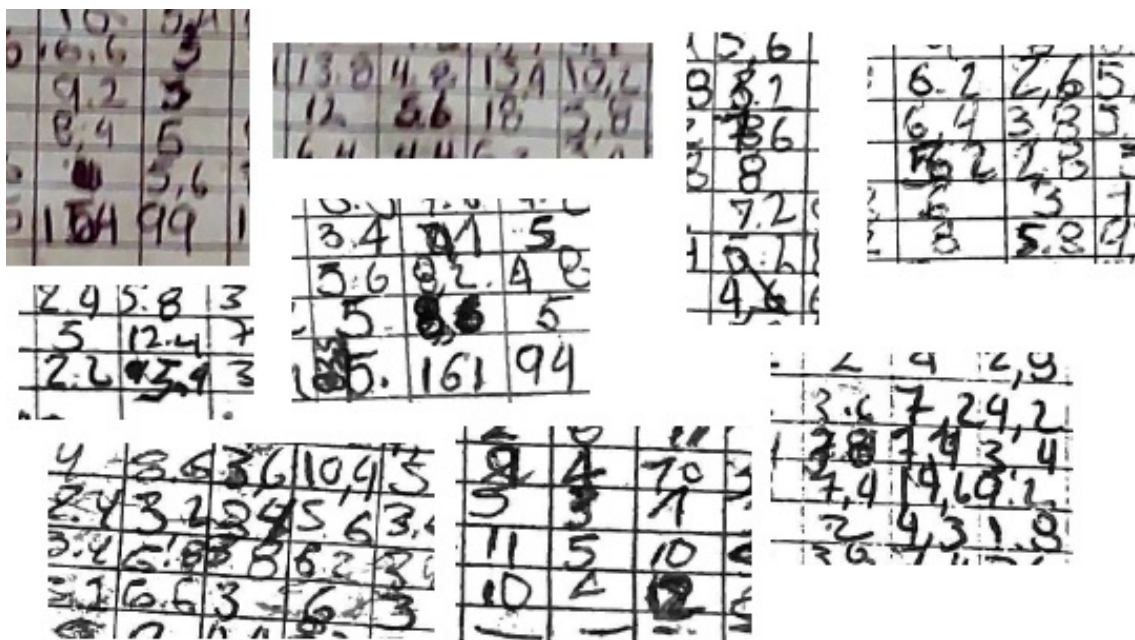


Figura 7.1: Ejemplares de datos físicos con malas prácticas de escritura.

En cuanto a los registros de peso, dado que la cantidad de datos no es tan extensa a comparación de los registros de leche, estos datos son transcritos a archivos de excel y posteriormente a archivos con extensión *.csv.

7.2. Digitalización de los datos de producción lechera

Teniendo en cuenta la descripción anterior, los datos fueron escaneados y posteriormente fueron transcritos de manera individual. Los datos utilizados en este trabajo consisten en los registros de producción lechera de distintas vacas que han entrado y salido de la producción ganadera, ya sea por motivos de compraventa, ingreso a gestación, salida por periodo de secado, enfermedad, deceso, edad o cantidad de partos.

Los datos aquí utilizados comprenden fechas desde el 31 de Diciembre de 2018, hasta el 24 de Julio de 2022. En total se transcribieron de forma manual un total de 85,932 datos, de los cuales, 37,956 son datos numéricos representativos que pueden usarse para procesamiento de datos. Los 47,976 datos restantes son datos categóricos que representan periodos de secado, horra, datos inexistentes o ilegibles, o registros desconocidos clasificados como datos NAN. Se cuenta con un total de 4 archivos pdf, en donde se almacenan los escaneos de los datos que comprenden esas fechas ya mencionadas.

Esta tarea de transcripción de datos tomó un tiempo aproximado de 1 semana; 7 días, en jornadas de 8 horas cada día. Los datos que poseían dificultades de lectura e interpretación fueron corroborados con base en los datos de producción diaria neta y producción neta individual de cada vaca, datos existentes en las tablas escaneadas; por lo que se reduce el riesgo de transcripción de datos erróneos por causas humanas.

Aún sí existiese algún dato atípico o dato erróneo, esto puede corroborarse con la estructuración de los datos que se empleó para su análisis y que se explica en la Sección 7.3 a continuación.

7.3. Reestructuración de los datos de leche

Una vez que se transcribieron los datos de manera individual para cada vaca. se procede a agrupar los datos en un archivo de excel organizado de la siguiente forma:

7.3.1. Reestructuración inicial

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	n	DD	MM	YY	Date	Hour	Day	#D	#W	#Month	#Year	Daniela	Gina	Nancy	Lina	Juanita	Guapa	Viviana	Andrea	Yolima	Soraya	Danna	Dianita	Lucía	Sol	Morocha	Lunita
2	1	31	12	2018	31/Dec/2018	AM	Mon	1	1	1	1	9	10.0	11.0	12.0	8.0	6.0	10.0	10.0	9.0	5.0	7.0	5.8	8.0	5.0	10.0	ne
3	2	31	12	2018	31/Dec/2018	PM	Mon	1	1	1	1	3.6	4.4	5.0	8.0	0.0	3.0	5.2	4.2	4.0	4.0	1.0	3.0	4.0	4.0	4.0	ne
4	3	1	1	2019	01/Jan/2019	AM	Tue	2	1	1	1	9	10.0	9.0	13.0	2.0	11.0	8.0	6.0	11.0	5.0	6.0	6.4	10.0	6.0	11.4	ne
5	4	1	1	2019	01/Jan/2019	PM	Tue	2	1	1	1	5	5.4	5.4	5.0	1.4	4.2	4.0	3.8	4.2	3.4	2.2	4.0	5.0	1.6	3.6	ne
6	5	2	1	2019	02/Jan/2019	AM	Wed	3	1	1	1	9	9.2	9.0	14.0	3.6	10.4	1.0	7.6	7.0	5.0	4.6	6.2	9.0	6.0	9.0	ne
7	6	2	1	2019	02/Jan/2019	PM	Wed	3	1	1	1	6	5.0	4.4	6.0	2.4	4.2	4.0	5.0	3.0	2.0	2.2	4.0	4.6	2.0	4.0	ne
8	7	3	1	2019	03/Jan/2019	AM	Thu	4	1	1	1	11	11.0	10.6	14.8	7.4	11.0	8.2	6.0	10.0	6.0	7.0	4.6	9.0	5.0	10.0	ne
9	8	3	1	2019	03/Jan/2019	PM	Thu	4	1	1	1	4	4.0	6.0	5.0	2.0	4.2	4.0	4.2	4.6	3.0	3.0	3.0	4.6	2.0	4.0	ne
10	9	4	1	2019	04/Jan/2019	AM	Fri	5	1	1	1	11	10.0	10.0	14.2	4.2	11.0	8.0	6.0	10.0	5.6	6.0	8.4	9.0	2.0	12.6	ne
11	10	4	1	2019	04/Jan/2019	PM	Fri	5	1	1	1	4	4.4	4.6	5.0	3.0	4.6	4.0	4.0	3.6	2.6	2.6	3.6	4.0	2.6	3.6	ne
12	11	5	1	2019	05/Jan/2019	AM	Sat	6	1	1	1	10.4	10.0	10.0	14.4	5.4	11.4	9.0	7.0	10.0	5.0	7.0	7.6	9.0	7.2	11.2	ne
13	12	5	1	2019	05/Jan/2019	PM	Sat	6	1	1	1	5	4.4	5.0	5.6	4.6	3.0	4.6	5.0	5.0	3.0	3.2	3.2	3.6	3.0	4.0	ne
14	13	6	1	2019	06/Jan/2019	AM	Sun	7	1	1	1	10	12.0	10.2	13.0	5.0	11.2	8.8	7.0	9.0	5.8	6.0	7.0	10.0	7.0	10.0	ne
15	14	6	1	2019	06/Jan/2019	PM	Sun	7	1	1	1	4	4.0	4.0	6.0	2.0	1.2	3.6	3.0	4.0	2.0	2.8	3.0	4.0	2.0	3.0	ne
16	15	7	1	2019	07/Jan/2019	AM	Mon	8	2	1	1	10.8	11.0	10.0	13.0	5.6	11.0	8.4	7.8	10.0	6.0	7.2	5.6	11.6	7.6	11.0	ne
17	16	7	1	2019	07/Jan/2019	PM	Mon	8	2	1	1	3	3.5	5.0	4.8	2.0	5.0	3.5	2.0	2.8	1.0	2.0	3.8	3.0	2.2	3.2	ne
18	17	8	1	2019	08/Jan/2019	AM	Tue	9	2	1	1	9	10.2	10.0	13.0	8.4	7.2	7.6	5.2	10.2	6.8	9.2	5.0	9.0	10.6	7.0	ne
19	18	8	1	2019	08/Jan/2019	PM	Tue	9	2	1	1	6	5.4	5.4	6.0	4.0	5.8	5.0	4.0	5.0	3.0	4.0	5.2	5.2	3.2	5.2	ne
20	19	9	1	2019	09/Jan/2019	AM	Wed	10	2	1	1	11	10.0	11.0	12.0	5.0	11.0	7.4	7.0	8.6	5.0	5.8	6.0	7.0	6.0	9.0	6.0
21	20	9	1	2019	09/Jan/2019	PM	Wed	10	2	1	1	5.2	6.0	5.2	9.0	4.0	6.0	5.2	3.6	5.0	3.8	4.4	4.0	5.0	3.2	5.8	4.2
22	21	10	1	2019	10/Jan/2019	AM	Thu	11	2	1	1	10.8	11.0	11.0	15.4	4.4	10.0	8.4	8.0	10.2	5.0	6.0	7.4	7.6	7.2	11.0	nan

Figura 7.2: Reestructuración inicial de los datos de producción de leche transcritos.

- Cada fila, representa una **muestra - n** tomada en una **fecha** (“DD/MM/AAAA”) en una **franja horaria** (AM o PM), en un **día de la semana** (DAY y #D), en una **semana** (#W) , en un **mes** (#Month) y en un **año** (#Year) específico.
- Cada columna, representa una **vaca** que produjo una **cantidad de leche** (valor numérico positivo).
- Cada 2 filas se tiene un cambio de fecha, pues se registran los datos de la mañana (AM) y también se registran los datos de la tarde (PM) como muestras separadas; más no como días distintos. Por lo tanto se tienen 2 muestras por cada día.
- En caso que una vaca no se encontrase en los registros de producción de leche para una fecha específica y sus días anteriores, se cataloga como **vaca no existente** o **ne**

- En caso que una vaca muriera, los datos subsiguientes a esa fecha serán datos de tipo texto representados con la letra **M**.
- En caso que una vaca se vendiera, los datos subsiguientes a esa fecha serán datos de tipo texto representados con la letra **V**.
- En caso que una vaca se encontrase en periodo de Horra, los datos subsiguientes a esa fecha serán datos de tipo texto representados con la letra **H**, hasta indicado lo contrario.
- En caso que una vaca no se encontrase en periodo determinado, fuera por falta de preñes, horra, u otro motivo desconocido, los datos subsiguientes a esa fecha aproximada serán datos de tipo texto representados con las letras **UL**, haciendo referencia a “Unknown-Location”.
- En caso que una vaca se encontrase en periodo de secado, los datos subsiguientes a esa fecha serán datos de tipo texto representados con la letra **S**, hasta que se reincorpore tras el periodo de secado o hasta que se avise lo contrario (Muerte, Venta, Horra).
- En caso que un registro de producción de leche no haya sido transcrito, ya sea por malas prácticas de escritura del documento físico, datos ilegibles o datos inexistentes, la casilla tendrá un valor de texto representado con las letras **NAN**, **haciendo referencia a un dato desconocido o dato inexistente por motivo desconocido**.
- En caso de tener valores numéricos 0, este valor es dejado como 0, si se muestra reincidencia en los días subsiguientes (suele suceder en fechas cercanas al periodo de secado), este valor será dejado así pues desde la práctica de toma de datos se ha considerado como una extracción de una cantidad de leche aproximadamente nula. No obstante, ya se ha mencionado que el dominio de los valores numéricos de producciones de leche son **racionales positivos** con hasta 2 cifras decimales.
- De las observaciones anteriores es importante agregar que los datos numéricos no pueden ser negativos.
- Se tiene un total de 2604 muestras para cada vaca, ya sea numérico o categórico. Se hace así pues las columnas de la matriz deben tener la misma cantidad de datos
- Se obtuvieron muestras de un total de 33 vacas distintas.
- Se maneja el dato desconocido con la denominación NAN y no con otro símbolo como el signo de interrogación “ ? ”, por motivos de formato al momento de transcribir los datos a Microsoft Excel.

7.3.2. “Debugging” y verificación de datos

Como la transcripción de los datos se realiza de forma manual, existe el riesgo de añadir error humano a mediciones que ya cuentan con errores de medición y escritura. Para disminuir este error

de transcripción, se verificó uno a uno los valores, antes, durante y después del proceso de transcripción. No obstante puede que se hayan pasado valores por alto como reflejo de un error humano.

Para ello, una vez obtenida la matriz agrupada en Excel, se verifican los tipos de datos y se corrobora que la cantidad neta de cada grupo sumen la cantidad total de datos. En caso contrario, el valor en cuestión se puede identificar el índice de la(s) muestra(s) no identificada(s) y mediante la fecha en la que fue(ron) tomada(s), se puede corroborar tanto en la tabla de Excel, como en los archivos pdf escaneados.

7.3.3. Reestructuración numérica

Notese que en la reestructuración inicial se puede observar que los datos pueden ser numéricos o categóricos, no obstante esto dificulta su análisis matricial en el software de Matlab por lo que se procede a tener algunas consideraciones adicionales.

- En caso de muerte o venta, se considera como dato de valor numérico 0.0, pues se considera como una producción de leche nula dado que no se puede extraer leche de un animal muerto o vendido, mas no obstante, no es considerado como valor de tipo NAN, **pues se considera NAN un valor que no pudo ser medido y se desconoce el motivo.**
- Como los datos no pueden ser biológica ni físicamente negativos; en caso que una vaca no existiese en fechas anteriores a la fecha de registro (los datos “**ne**”), se cambiará el valor categórico “**ne**” por un **-1**.

Así pues, se conforma un primer archivo de excel con todos los registros existentes hasta la fecha del 24 de Julio del 2022. Es importante mencionar que el balance existente entre datos numéricos útiles para procesamiento y análisis, con datos de tipo NAN es aproximadamente de 90 % contra menos del 10 % respectivamente. Y en comparación de datos de tipo NAN con datos conocidos, ya sean numéricos representativos o datos categóricos transformados; el porcentaje de balance disminuye hasta un 5.6 %.

De esta manera se puede considerar que los datos de tipo NAN no son significativos para los análisis posteriores, y pueden pasarse por alto o reemplazarse con métodos de imputación, sin afectar negativamente los análisis a futuro.

7.3.4. Clasificación, Selección y Agrupación de datos de leche

Teniendo en cuenta la base de datos estructurada en el punto anterior, es importante cuestionarse si estos datos pueden representar matemáticamente toda la producción de leche del CA-SENA-POP, pues aunque son datos que evidencian la producción neta de la granja, estos datos dispersos no permiten una aproximación mediante funciones y/o modelos matemáticos, ni mucho menos posibilitan el modelamiento dinámico mediante ecuaciones diferenciales. Por lo tanto es imprescindible estimar una o varias funciones que representen de manera aproximada, los datos obtenidos.

Es decir, que se debe considerar un ajuste de curvas a los datos existentes para aproximar la producción neta mediante funciones y modelos matemáticos. Por otra parte es importante tener en cuenta que las producciones de leche de cada animal, varían según el número de partos que hayan experimentado con anterioridad. Las vacas primíparas que recién están produciendo leche, no tendrán un rendimiento equiparable a comparación de las vacas que ya se encuentran en un 3er o 4to parto; en consecuencia es plausible realizar una clasificación de vacas por su cantidad de partos y observar el comportamiento de las producciones lecheras para cada grupo de partos.

7.3.4.1. Clasificación de Datos

Así pues, con base en las observaciones anteriores y partiendo de los datos existentes del CA-SENA-POP, tendremos 6 grupos distintos de partos. $Parto_i$, con $i=1\dots 6$

Notese que los datos previamente estructurados cuentan con los registros de producción lechera desde el año 2018 hasta el año 2022, con todos los registros existentes para cada una de las vacas. No obstante, en este periodo de tiempo, en los documentos físicos no se especifica el número de parto de cada animal. En algunos casos se ha hecho la observación, mas sin embargo no aplica para todas las reses. Observese además, que los partos entre las vacas no están sincronizados; por lo que esta clasificación permite identificar qué vacas son contemporáneas con otras. Por ende, es necesario acudir a la base de datos digital del software TaurusWebs con el que cuenta el centro agropecuario; para hacer una validación cruzada de los datos.

Una vez obtenido el permiso y una copia de la base de datos actual, suministrada por el encargado Daniel Cobo. se puede acceder al software con una versión gratuita de prueba; licencia que satisface la posibilidad de consultar las fechas de parto de cada bovino. Esto se logra gracias a la trazabilidad con la que cuenta este software y se puede hacer contraste con los datos previamente estructurados; dándose la posibilidad de clasificar los datos según el parto al que corresponden para cada ternera:

	E	F	N	O	P	Q	T	U	X	Y	Z	Af	AF	Ai	AJ	Af	AX	A'	AZ	Bf	BH	B	BN	B	BT
1	YY	Date	Daniela	Gina	Lina	Guapa	Viviana	Soraya	Dianita	Martina	Centavito	Leticia	Luisa	Jasmin											
1537	2021	05/Feb/2021	UL	3.0	V	2.6	3.0	1.6	PPS	5.6	5.0	6.2	5.0	NE											
1538	2021	06/Feb/2021	UL	8.2	V	10.4	5.2	4.2	PPS	8.8	5.4	12.0	3.0	NE											
1539	2021	06/Feb/2021	UL	3.8	V	3.0	3.2	2.0	PPS	4.0	2.6	7.2	4.2	NE											
1540	2021	07/Feb/2021	UL	8.6	V	9.2	6.2	4.4	PPS	6.8	6.2	11.0	3.0	NE											
1541	2021	07/Feb/2021	UL	5.0	V	4.0	2.8	2.2	PPS	4.2	2.2	6.2	3.6	NE											
1542	2021	08/Feb/2021	UL	9.2	V	11.2	5.8	5.2	PPS	9.6	9.4	12.8	7.4	NE											
1543	2021	08/Feb/2021	UL	4.2	V	4.0	3.0	2.4	PPS	3.8	4.0	6.2	5.0	NE											
1544	2021	09/Feb/2021	UL	8.8	V	11.0	6.2	4.2	PPS	9.2	8.6	12.2	7.8	NE											
1545	2021	09/Feb/2021	UL	4.8	V	3.8	3.4	2.2	PPS	1.6	2.8	6.8	3.8	NE											
1546	2021	10/Feb/2021	UL	9.5	V	9.6	5.2	4.8	PPS	3.2	6.0	12.6	8.0	NE											
1547	2021	10/Feb/2021	UL	5.0	V	4.5	3.6	2.4	PPS	5.6	3.2	6.2	4.4	NE											
1548	2021	11/Feb/2021	UL	10.2	V	10.6	6.2	4.4	PPS	10.6	6.0	12.4	6.4	NE											
1549	2021	11/Feb/2021	UL	5.8	V	5.0	3.0	2.2	PPS	5.6	3.0	6.8	4.8	NE											
1550	2021	12/Feb/2021	UL	11.0	V	10.4	5.4	4.6	PPS	10.8	6.4	12.4	6.0	NE											

Figura 7.3: Clasificación de datos por #Parto para cada vaca.

En la Figura 7.3 se puede observar que la clasificación de los partos se realiza con base en la estructuración de datos tanto categóricos como numéricos. Esto con el objetivo de diferenciar de manera visual entre un periodo de producción de leche a un periodo de secado (Casilla blanca con borde de color y casillas coloreadas de color azul respectivamente). No obstante una vez realizada la clasificación, selección y agrupación de datos; las matrices usadas en los análisis posteriores con algoritmos en Matlab, manejan la estructuración numérica descrita en la sub-Sección 7.3.4.1.

Percátese que al tener 6 grupos de partos en el CA-SENA-POP, se utiliza un código de colores para la clasificación de producción por # de Parto, siendo esta la siguiente:

Tabla 7.1: Código de colores por número de parto

<i>Color del borde</i>	<i>Negro</i>	<i>Azul</i>	<i>Verde</i>	<i>Amarillo</i>	<i>Rojo</i>	<i>Morado</i>
<i># de Parto</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>

Así pues, se obtienen 6 grupos de un tamaño variable de registros de lactancia de vacas para cada número de parto. Esta agrupación inicial consta de 88 distintas lactancias y se puede observar en la Tabla 7.3.

7.3.4.2. Selección de Datos

Una de las primeras observaciones obtenidas al realizar la agrupación inicial por partos como la observada en la Tabla 7.3; es que los datos existentes que han sido agrupados no poseen igual cantidad de datos para cada res dentro de un mismo grupo de parto. Esto se produce dado que los partos varían en cuestión de duración para cada animal. En primera instancia esto no debería

suponer un problema significativo, pues en caso que los partos posean una ligera variabilidad de muestras; bastaría con “rellenar” los datos faltantes con ceros o con imputación de datos como la media o la moda; ó por otra parte, bastaría con realizar una agrupación con igual cantidad de muestras.

Sin embargo, estas contramedidas no pueden ser llevadas a cabo en razón de que los datos existentes facilitan la agrupación por partos, mas no garantiza que los datos de un parto en específico se encuentren disponibles en su totalidad. En otras palabras, muchos de los datos que se tienen para cada parto son datos de inicio de producción o finalización de producción de leche de las terneras, tanto para las que fallecen, para las que se venden, así como también, para las que a fecha del 31 de Diciembre del 2018 ya se encontraban en periodo productivo.

Si se realiza una representación visual individual de los datos pertenecientes al grupo de parto #1, se puede observar en la Figura 7.4 que algunas reses tienen muy pocos datos (celdas vacías), lo que “recargaría” la distribución hacia los extremos, dificultando su análisis descriptivo (Esta observación se cumple para los otros grupos de parto). Por otra parte, si se hace un contraste de datos del grupo de parto#1 teniendo como punto de referencia el inicio de la producción o su finalización en el periodo de secado, la tabla de registros permite evidenciar el desbalance que manifiestan los datos:

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	N	Hour	#D	#Month	#Year	Gina	Viviana	Yolima	Dianita	Lucia	Sol	Lunita	Acerada	Martina	Margarita	Pacha	Araly	Luisa	Isabel	Tamara	Jasmin	Vellita	Johana	Emperatriz	
32	31	AM	16	1	1					4.0	8.0	9.0	2.0	11.2		5.4					11.6	7			
33	32	PM	16	1	1					10.0	4.0	3.6	2.0	6.0		3.4					6.9	4.2			
34	33	AM	17	1	1					2.0	7.0	9.0	6.0	11.0		3.4					12.4	6			
35	34	PM	17	1	1					11.0	3.0	4.0	10.0	nan		3.4					5.8	2.4			
36	35	AM	18	1	1					5.0	6.0	8.0	5.0	11.0		6.0					13.2	3			2.6
37	36	PM	18	1	1					10.0	3.0	3.0	11.0	7.6		7.0					5.6	8			2
38	37	AM	19	1	1					5.0	9.0	9.8	7.0	10.0		6.8					nan	2.4			1.9
39	38	PM	19	1	1					12.0	2.4	1.0	9.0	5.2		12.6					nan	6.6			1.3
40	39	AM	20	1	1					4.0	6.0	9	6.6	4.0		3.0					nan	6.8			2.4
41	40	PM	20	1	1					10.0	5.0	4.0	11.0	6.0		1.0					nan	4			1.8
42	41	AM	21	1	1					4.0	8.0	8.2	5.6	11.0		8.0					nan	6.8			2

Figura 7.4: Desbalance de datos en agrupación inicial por partos.

En la Figura 7.4, se evidencia el desbalance mencionado. Si se observan más a fondo los datos presentes en el archivo de Excel, específicamente en el grupo de parto #1 de la Tabla 7.3, 13 de las 19 reses pertenecientes a este grupo cuentan con menos de la mitad de muestras de producción lechera tras haber parido por primera vez. Por otro parte, solo una de ellas manifiesta un exceso de registros antes de entrar en periodo de secado y 5 mamíferos restantes aparentan tener la cantidad de muestras suficientes para representar el grupo de vacas primíparas. Las 5 vacas en cuestión son “Lucia”, “Sol”, “Lunita”, “Acerada” y “Martina”.

Finalmente, aunque todas las reses cuentan con la misma cantidad de datos, se procede a seleccionar únicamente a 3 de ellas que coinciden en el inicio y finalización de la producción de leche (“Sol”, “Lunita”, “Acerada”). Mientras que las otras 2, a pesar de tener la misma cantidad de muestras, presentan desfases en el inicio y finalización de la etapa productiva (“Lucia” y “Martina”). De esta forma, se puede generar la primera representación visual de los datos mediante un gráfico de dispersión, tal y como se observa en la Figura 7.5.

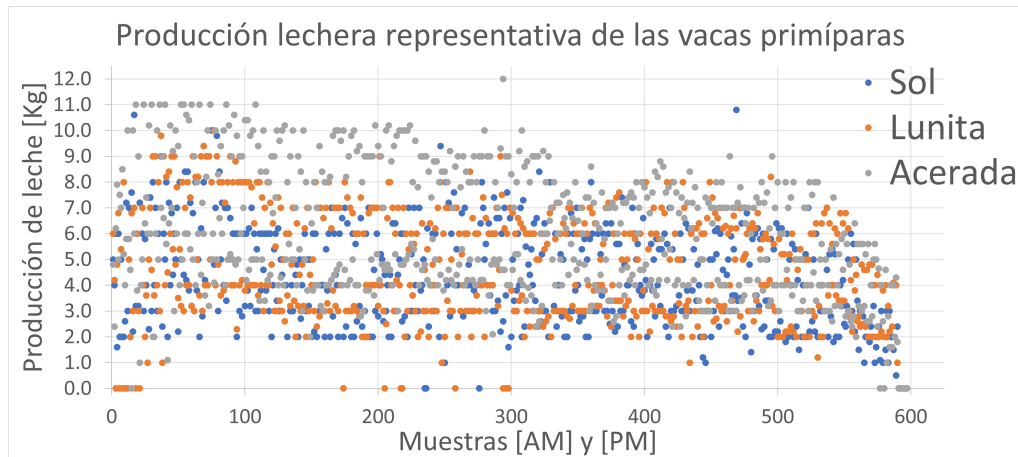


Figura 7.5: Gráfico de dispersión para representar la producción lechera de las vacas primíparas.

7.3.4.3. Agrupación de Datos

El resultado obtenido tras realizar la primer representación visual de datos da un idea de cómo no se deben utilizar los datos para los análisis matemáticos. Pues se observa que al tener muestras matutinas y vespertinas juntas, no se logran diferenciar los datos correspondientes a cada intervalo. Además, el intervalo de tiempo es de 305 días aproximadamente y al manejar ambos tipos de muestras, se tiene un total de 610 datos en el eje horizontal. Por ende, como medida inicial, se opta por realizar un análisis independiente para las muestras obtenidas en diferentes horas del día, procediendo con una representación individual como se observa a continuación:

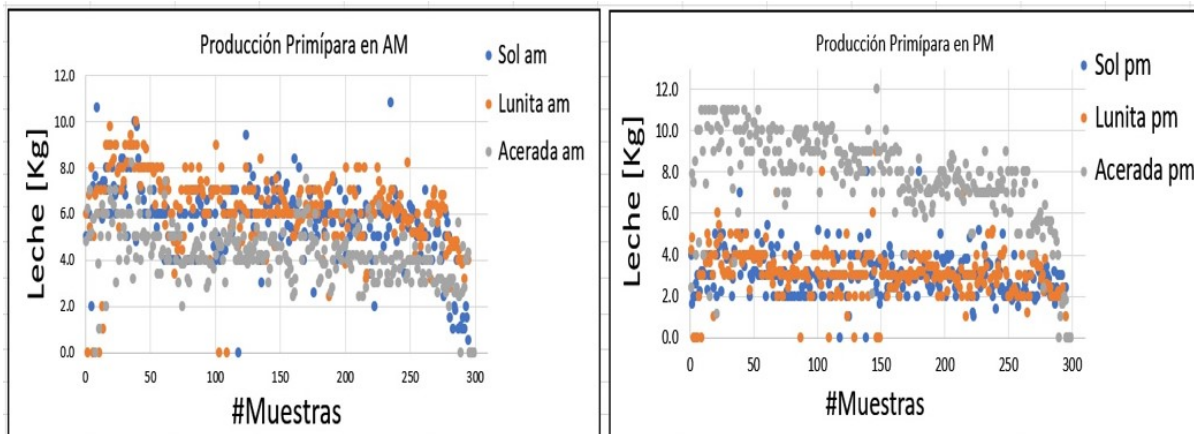


Figura 7.6: Gráfico de dispersión para Muestras matutinas y vespertinas.

En primera instancia parece que la Figura 7.6 no presenta ningún inconveniente, a pesar de ello,

se puede notar un leve desajuste entre los valores vespertinos; pues los valores de la res denominada como “Acerada”, muestran una lejanía con los otros datos en el horario de la tarde. Esto genera ciertas inquietudes teniendo en cuenta que las muestras matutinas tienden a ser mayores que las muestras vespertinas para la mayoría de las reses de este estudio. En virtud de que el lapso que existe entre una muestra matutina con una vespertina en un mismo día es de aproximadamente 10 horas; mientras que el lapso entre una muestra vespertina y una muestra matutina es de aproximadamente 14 horas; no sería de extrañar que al tener mayor tiempo de recuperación, se le pueda extraer mayor cantidad de leche a la vaca.

Lo anterior se presta para realizar nuevamente una corroboración de los datos con los datos manuales existentes y la tabla de datos de la estructuración inicial; con lo que se observa que al realizar la “alineación” de datos tras la agrupación por partos; los datos matutinos de la res en cuestión han quedado “alineados” con los datos vespertinos de las otras reses. Esto quiere decir que la Figura 7.6 debe cambiarse por la Figura 7.7:

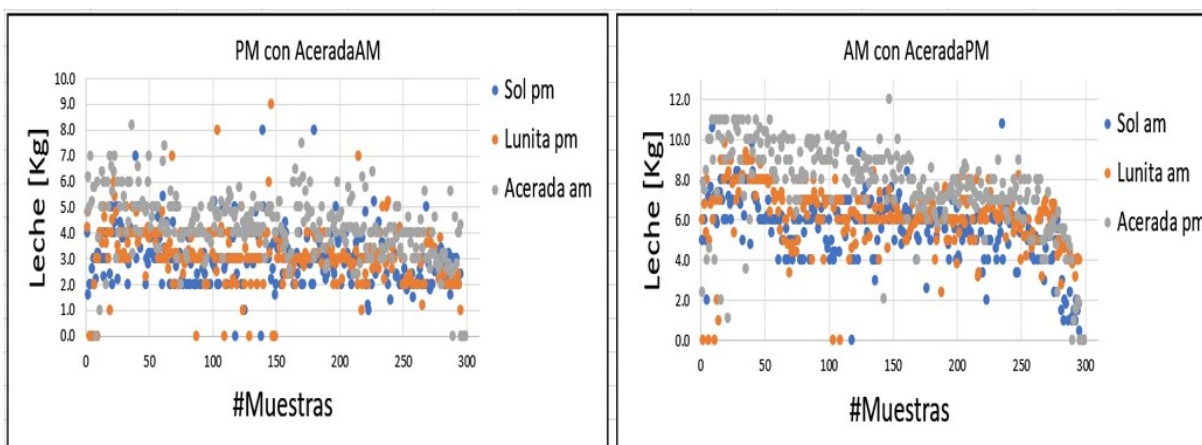


Figura 7.7: Gráfico de dispersión para Muestras matutinas y vespertinas.

De acuerdo con la situación anterior, se concluye que realizar un análisis por franja horaria puede llevar a errores futuros o a reincidentes corroboraciones de datos, siendo esto último una tarea agobiante para el caso de estudio; por lo que se decide considerar únicamente los análisis con la producción neta diaria de cada animal. De esta forma se repite el proceso para los otros grupos de partos obteniéndose los gráficos de dispersión de la Figura 7.8 y disminuyendo la cantidad de lactancias, desde 88 como se observó inicialmente en la Tabla 7.3, hasta obtener un conjunto final de 28 lactancias en total para realizar análisis futuros. Estas lactancias están asociadas a las reses mencionadas en la Tabla 7.2.

Tabla 7.2: Agrupación final de vacas por # de partos

# de Parto	1	2	3	4	5	6			
Vacas participantes	Acerada	Dianita	Pacha	Fernanda	Martina	Centavito	Muchira	Andrea	Andrea
	Lunita	Gina	Sol	Guapa	Morocha	Dianita	Soraya	Guapa	Guapa
	Sol	Juanita	Viviana	Juanita	Sol	Leticia	Viviana		
		Lucia	Leticia		Morocha				

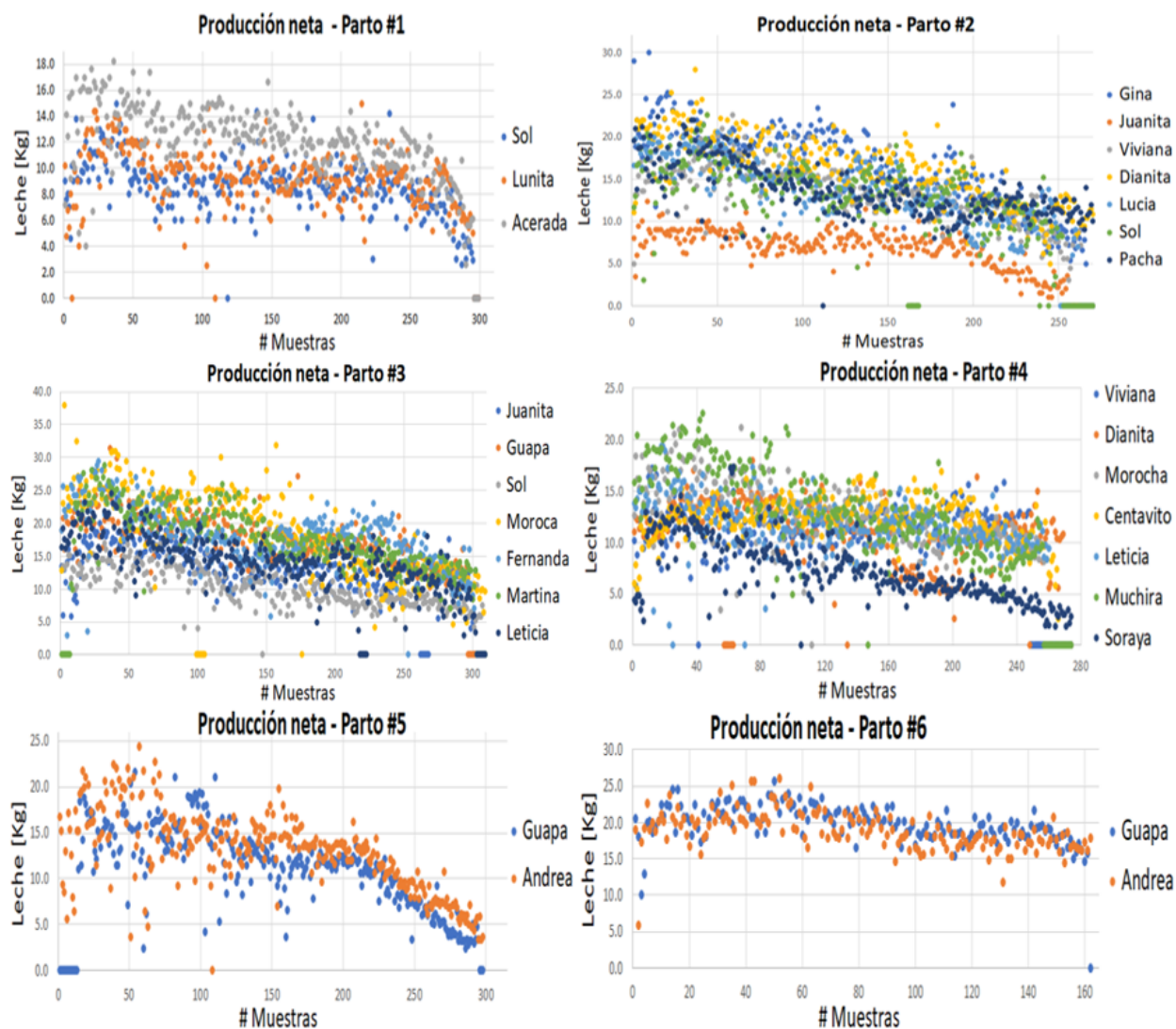


Figura 7.8: Gráfico de dispersión para cada grupo de partos.

Tabla 7.3: Agrupación general de vacas por # de partos

<i># de Parto</i>	<i>Vacas participantes</i>				
1	<i>Gina</i> <i>Pacha</i> <i>Sol</i> <i>Jasmin</i>	<i>Viviana</i> <i>Araly</i> <i>Lunita</i> <i>Velita</i>	<i>Yolima</i> <i>Luisa</i> <i>Acerada</i> <i>Johana</i>	<i>Dianita</i> <i>Isabel</i> <i>Martina</i> <i>Emperatriz</i>	<i>Lucia</i> <i>Tamara</i> <i>Margarita</i>
2	<i>Gina</i> <i>Fernanda</i> <i>Andrea</i> <i>Muchira</i> <i>Dianita</i> <i>Araly</i>	<i>Juanita</i> <i>Acerada</i> <i>Yolima</i> <i>Margarita</i> <i>Lucia</i> <i>Luisa</i>	<i>Guapa</i> <i>Martina</i> <i>Soraya</i> <i>Romina</i> <i>Sol</i> <i>Isabel</i>	<i>Viviana</i> <i>Centavito</i> <i>Danna</i> <i>Pacha</i> <i>Morocho</i> <i>Tamara</i>	<i>Lunita</i> <i>Johana</i>
3	<i>Gina</i> <i>Morocho</i> <i>Andrea</i> <i>Centavito</i>	<i>Juanita</i> <i>Lunita</i> <i>Yolima</i> <i>Muchira</i>	<i>Guapa</i> <i>Fernanda</i> <i>Soraya</i> <i>Romina</i>	<i>Viviana</i> <i>Martina</i> <i>Dianita</i> <i>Leticia</i>	<i>Lucia</i> <i>Pacha</i> <i>Sol</i>
4	<i>Guapa</i> <i>Sol</i> <i>Soraya</i>	<i>Viviana</i> <i>Morocho</i> <i>Dianita</i>	<i>Andrea</i> <i>Fernanda</i> <i>Muchira</i>	<i>Yolima</i> <i>Centavito</i> <i>Leticia</i>	
5	<i>Nancy</i> <i>Dianita</i>	<i>Guapa</i> <i>Centavito</i>	<i>Andrea</i> <i>Muchira</i>	<i>Soraya</i> <i>Leticia</i>	
6	<i>Daniela</i>	<i>Lina</i>	<i>Guapa</i>	<i>Andrea</i>	

7.4. Reestructuración de los datos de peso

Tal y como se mencionó en la Sección 7.1, los datos referentes al peso de los animales usados en la producción de leche se encuentran registrados en el software “TaurusWebs”, y con el objetivo posterior de ser exportados y analizados, debemos realizar la transcripción manual de dichos datos.

7.4.1. Reorganización de datos

Esta transcripción de datos se realiza de tal forma que se mantienen los datos asociados a las fechas especificadas tal y como se puede observar en la Figura 7.9.

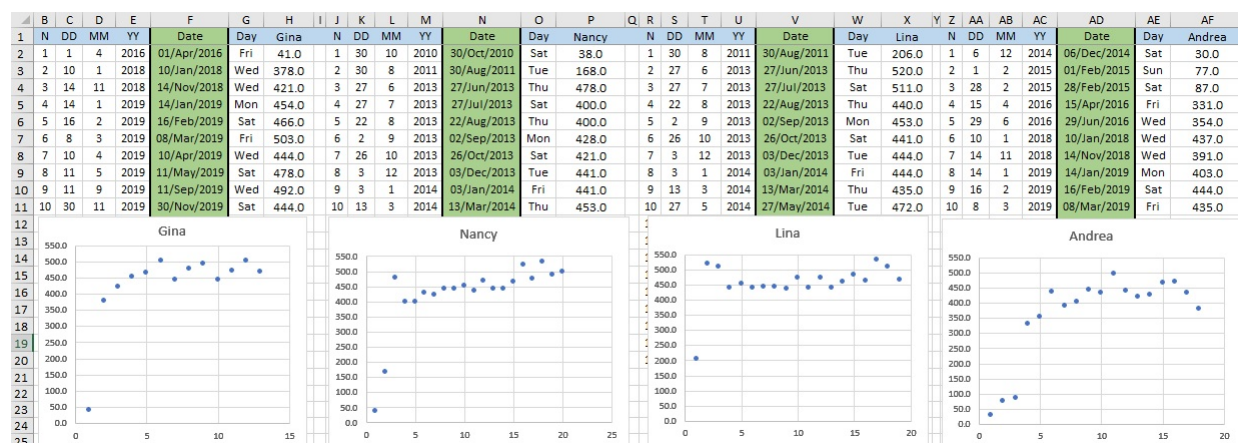


Figura 7.9: Reorganización inicial de los datos de peso.

Sin embargo, tras realizar contrastes de fechas de registros de peso con las fechas de registros de leche, se encuentra con que **los registros de peso coinciden parcialmente con los registros de leche en un mismo periodo**. Esto quiere decir que de los pocos datos existentes respecto al peso, estos deben ser segmentados de acuerdo con el número de parto con el que coinciden las fechas de registro de producción de leche asociadas a la lactancia de ese parto. En función de lo expresado anteriormente, se debe realizar la segmentación de los datos de peso que pertenecen a cada parto (1.6) para cada vaca perteneciente a la Tabla 7.2 de la Subsección 7.3.4.3.

7.4.2. Agrupación de datos de peso

De la Tabla 7.4 se tienen 18 vacas diferentes para un total de 28 partos distintos que dependiendo del número de parto se cuenta con una cantidad de muestras distinta (Entre 162 y 309 muestras). Ahora bien, para los análisis algorítmicos en Matlab y/o Python, se requiere que la cantidad de muestras de leche tengan la misma cantidad de muestras para el peso en pro de lograr un análisis por “Ecuaciones Diferenciales Ordinarias”; por tanto es necesario realizar una interpolación del(los) dato(s) de peso dependiendo de la cantidad de muestras presentes en cada parto.

Tabla 7.4: Cantidad de muestras para cada vaca de cada parto

# de Parto	1	2	3	4	5	6			
# de Muestras	299	288	309	274	298	162			
Vacas participantes	Acerada Lunita Sol	Dianita Gina Juanita Lucia	Pacha Sol Viviana	Fernanda Guapa Juanita Leticia	Martina Morocha Sol	Centavito Dianita Leticia Morocha	Muchira Soraya Viviana	Andrea Guapa	Andrea Guapa

7.4.2.1. Generación de columnas de peso que serán anexadas a los “dataframes” de cada uno de los modelos asociados al número de parto modelado

El método usado para realizar la interpolación entre los datos de peso son las “Splines Cúbicas” que varían en tamaño dependiendo de la cantidad de datos presentes en cada parto. Una vez ejecutado el script de Matlab “InterpPesosPartesEDOs” (adjunto en este trabajo de grado) se obtiene el conjunto de arreglos unidimensionales con los tamaños deseados (ver Figura 7.10) y se procede a realizar la exportación correspondiente en archivos separados por comas. Posteriormente se procede a agruparlos en “dataframes” para su análisis por EDOs en python. Para cada lactancia se tendrá un “dataframe” representativo de ese modelo con las columnas de producciones de leche, peso y estiércol producido por cada vaca perteneciente a esa agrupación de lactancia.

1 Acer299 299x1 double	2 Lun299 299x1 double	3 Sol299 299x1 double	1 Andre298 298x1 double	2 Gpa298 298x1 double	1 Andre162 162x1 double	2 Gpa162 162x1 double
1 Dian288 288x1 double	2 Gina288 288x1 double	3 Juan288 288x1 double	4 Luci288 288x1 double	5 Pac288 288x1 double	6 Sol288 288x1 double	7 Viv288 288x1 double
1 Fer309 309x1 double	2 Gpa309 309x1 double	3 Juan309 309x1 double	4 Leti309 309x1 double	5 Mart309 309x1 double	6 Mor309 309x1 double	7 Sol309 309x1 double
1 Cen274 274x1 double	2 Dian274 274x1 double	3 Leti274 274x1 double	4 Mor274 274x1 double	5 Muc274 274x1 double	6 Sor274 274x1 double	7 Viv274 274x1 double

Figura 7.10: Formato de nombre de columna para conjunto de datos. Varía en cuanto al nombre de la vaca, cantidad de muestras existentes y cantidad de parto. El código de colores se mantiene según la Tabla 7.1

Llegados a este punto, se cuentan con los datos de las tres variables necesarias: Producción de Leche, Estiércol y Peso, para plantear el (los) modelo(s) por EDOs, proceso descrito en el próximo capítulo. Adicionalmente, al finalizar este Capítulo, se ha completado la descripción del primer bloque del diagrama en la Figura 4.15, ya que los datos oficiales proporcionados por la institución han sido preprocesados y se dispone de datos descriptivos suficientes para estimar parámetros.

Modelado

Al final del Capítulo 7, se obtienen, tanto los datos referentes a la producción de leche de cada vaca en cada parto, así como también los datos referentes al peso de las reses asociadas a dicha producción. En este capítulo se hablará del ajuste de curvas para los datos de lactancia en cada parto y del plantamiento de modelo(s) con EDOs.

8.1. Ajuste de Curvas mediante funciones No lineales

A diferencia de los datos de leche, los datos interpolados del peso pueden presentar una variabilidad creciente al rededor de un valor de equilibrio, o por el contrario ser asumidos como valores constantes desde la última medición existente (esto es, en aquellos casos que se desconocen los registros de peso por motivos de fuerza mayor, como por ejemplo la Pandemia del SARS CoV 2, a la que se le atribuye que muchos datos no hayan sido medidos entre el 2020 y 2022). Cualquiera que sea el caso, el peso puede ser descrito a simple vista por una curva distinguible; esto dado que son curvas interpoladas basadas en un único conjunto de datos para cada res (Para mayor información sobre el proceso de interpolación de los datos de Peso, ver las Secciones 4.2.3.1 y 7.4.2.1).

Por su parte, la producción de leche grupal depende de la suma de las producciones de leche asociadas a cada grupo de partos; y para una representación individual los datos existentes poseen error añadido, ya sea por medición, error humano, error de aproximación, entre otros. Por tanto no es posible identificar una curva única que pueda representar a todos los datos de una forma generalizada. Para tener en consideración una curva representativa de los datos de cada parto, o una curva representativa de los datos de cada lactancia, se pueden realizar ajustes de curvas de lactancia, teniendo en consideración que ya existen modelos matemáticos probados para representar este tipo de datos. Entre estos modelos existentes, utilizaremos como referencia el modelo no lineal de Gamma Invertida de Wood, dado que según [35] y [36], es un modelo más que apropiado para aquellas curvas “normales” con una etapa creciente hasta el punto de producción máxima y una etapa decreciente hasta el periodo de secado.

8.1.1. Aplicación del método de ajuste no lineal: Algoritmo de Levenberg-Marquardt (LMA)

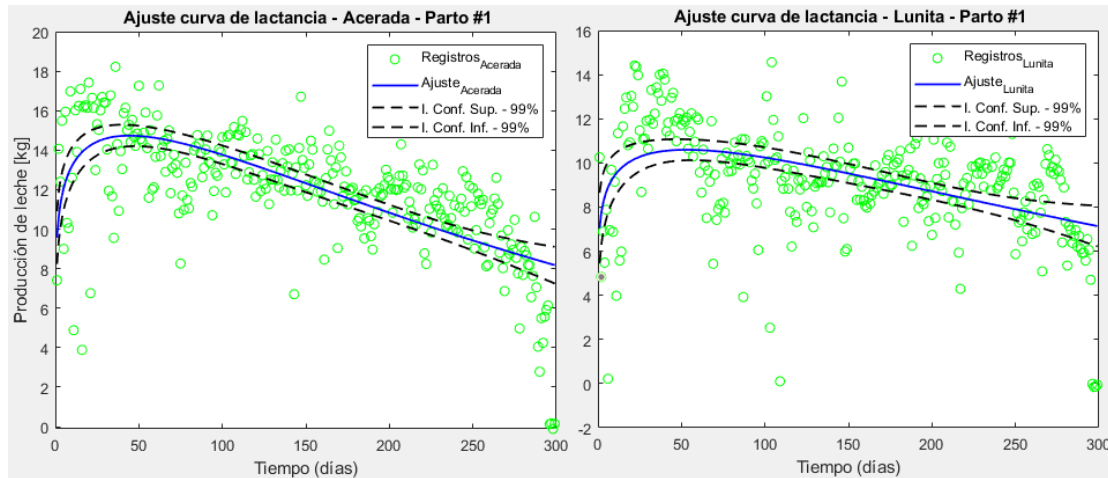


Figura 8.1: Ajustes de curvas de lactancia para Acerada y Lunita, Parto#1.

Teniendo en consideración que el método de LMA es implementado tomando como base el script público de [28], se pueden obtener unos primeros ajustes como los de la Figura 8.2. De los resultados obtenidos por el método anterior, se pueden estimar los valores de los parámetros α , β y γ , parámetros usados en el modelo de Gamma invertida de Wood, y se estiman para cada una de las reses de la Tabla 7.2. Al finalizar la implementación de éste método, se evidencian las ejecuciones del algoritmo hasta que se logra la aproximación de los parámetros de interés y se obtienen un total de 84 parámetros asociados a las 28 curvas de lactancia para las vacas de la Tabla 7.4

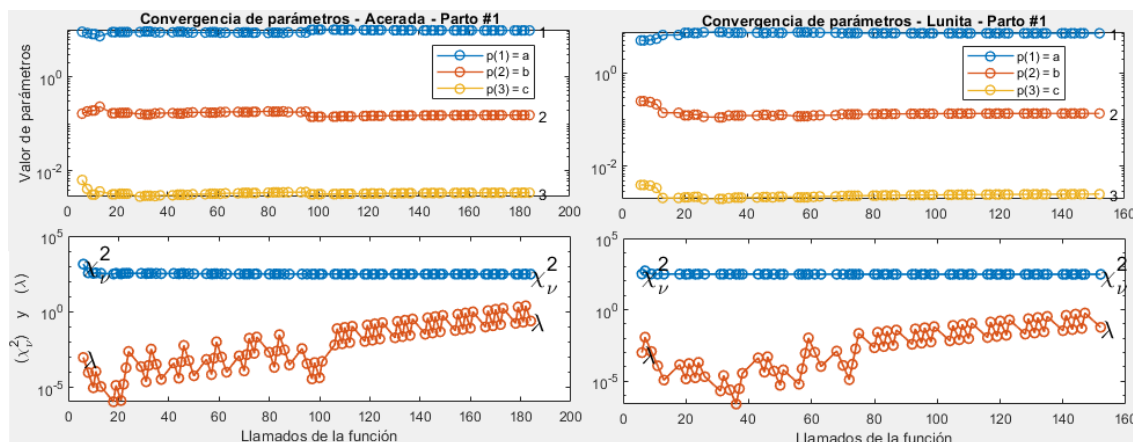


Figura 8.2: Convergencia de parámetros estimados del modelo de Wood para Acerada y Lunita.

8.1.2. Aplicación del método de ajuste no lineal: Algoritmo de Efectos Mixtos No Lineales (NLMEA)

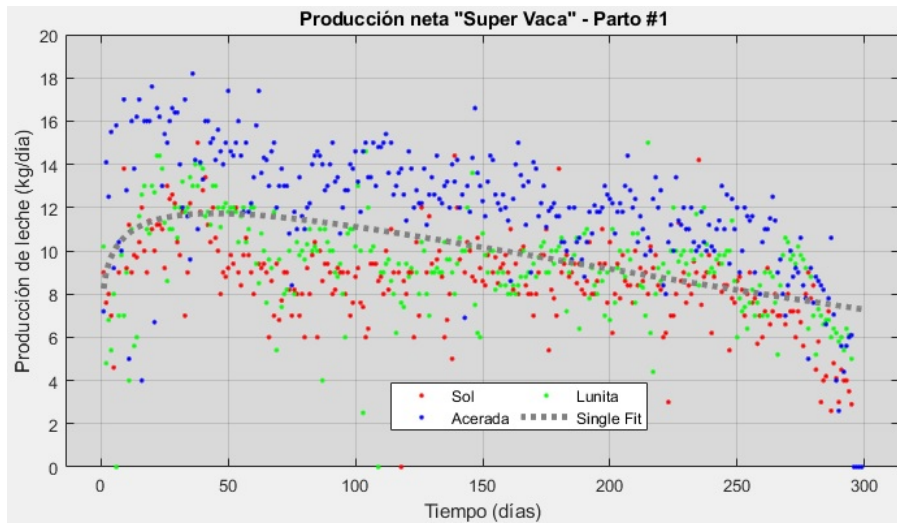


Figura 8.3: Estimación de parámetros generales mediante LMA, para el grupo de parto #1 .

Para este caso, los sujetos de estudio son las reses de cada parto y se busca estimar los parámetros de ajuste de cada vaca aún cuando el modelo plantee una solución generalizada para cada parto. Supóngase que se realiza un ajuste no-lineal como el de Levenberg-Marquardt a las reses de cada parto, obteniéndose las Figuras 8.3 y 8.4.

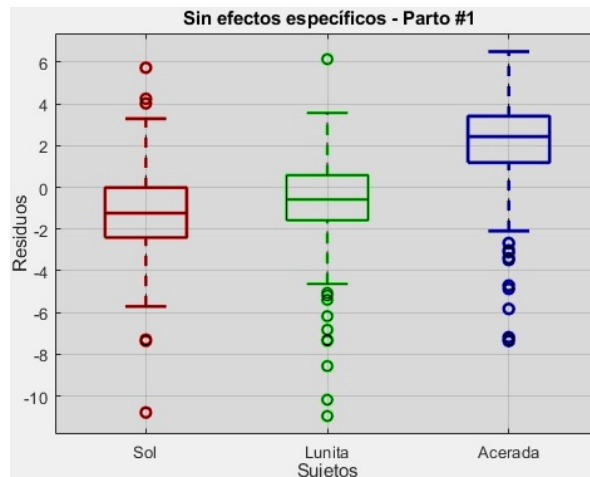


Figura 8.4: Diagrama de cajas y bigotes para las reses del parto 1.

De la Figura 8.4 se puede observar que los residuos de las reses del parto 1 se encuentran por encima o por debajo de 0, lo que indica que el modelo ha fallado en identificar los efectos específicos en el sujeto. Para lograr la identificación de los efectos específicos de cada vaca se debe hacer un ajuste individual tal y como se observa en la Figura 8.5:

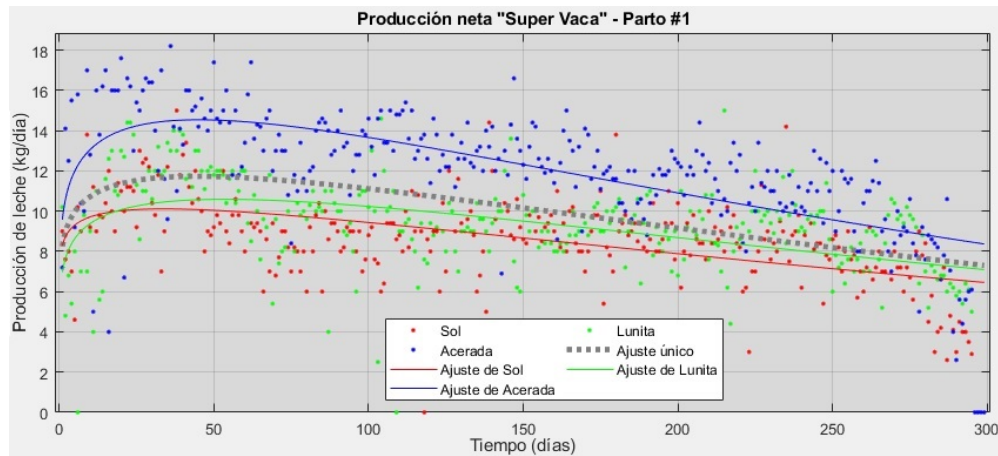


Figura 8.5: Estimación de parámetros individuales en el parto #1, Usando NLMEA-FE.

Repitiendo este proceso de comparación, se puede observar que incluso para aquellos partos que cuentan con 7 reses, la estimación individualizada permite una disminución del error cuadrático medio tanto de manera individual como de manera grupal (ver Tabla 10.1, en el Anexo ??); concluyendo así, que el ajuste para efectos específicos representa mejor ajuste que el modelo generalizado.

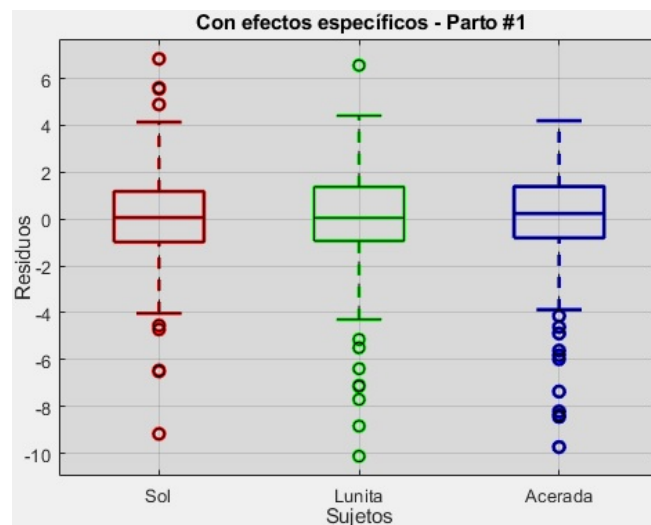


Figura 8.6: Diagrama de cajas y bigotes individualizado para las reses del parto 1.

Por otra parte, los diagramas de cajas y bigotes también muestran una mejoría notable, pues se evidencia que los residuos del ajuste tienden a estar centrados en 0, y el tamaño de las cajas disminuye a comparación del modelo anterior. Aún cuando este ajuste ha probado tener mejores resultados que el ajuste no lineal por el método de LM, y explica con éxito las variaciones debidas a los sujetos específicos del estudio, no considera a los sujetos como representantes de una población más grande, esto es, para casos donde se tuvieran más sujetos de estudio a comparación de los existentes.

El propósito de los modelos de efectos mixtos es dar cuenta de las variaciones específicas de los sujetos de manera más amplia, como efectos aleatorios que varían en torno a las medias de la población. Por tal motivo se considera pertinente realizar un análisis complementario para un ajuste no lineal con efectos mixtos aleatorios.

8.1.3. Aplicación del método de ajuste no lineal: Algoritmo de Efectos Mixtos No Lineales con Efectos Aleatorios (NLMEA-RE)

En este caso, el método de ajuste no lineal utilizado con anterioridad es cambiado por la función de ajuste no lineal de efectos mixtos “nlmefit”, la que por defecto asigna efectos aleatorios a todos los parámetros del modelo. También de forma predeterminada, asume una matriz sin covarianza entre los efectos aleatorios para evitar la parametrización excesiva y los problemas de convergencia relacionados. De esta función utilizada en Matlab, se pueden extraer algunos criterios adicionales que pueden complementar las decisiones tomadas en el método, como por ejemplo la reducción de los parámetros AIC y BIC (Para mayor información ver la Secciones 4.2.7.3 y 4.2.7.4).

La matriz de covarianza generada para cada parto puede presentar algunos valores cercanos a 0. Este tipo de valores sugieren que los efectos aleatorios en esos parámetros pueden removerse para simplificar el modelo. Esto se logra y se justifica al notar que los criterios AIC y BIC disminuyen nuevamente. Para evidenciar la relación existente entre parámetros, es posible graficar las matrices de covarianza del método “nlmefit”, en donde los valores altamente correlacionados son cercanos a 1 y los que no, son forzados a 0. Para este trabajo de grado, se obtienen las relaciones en las matrices de covarianza mostradas en la Figura 10.1 (ver Anexo 10.5). Identifíquese que el modelo de Wood requiere de la estimación de 3 parámetros por lo que originalmente las matrices de correlación entre parámetros aleatorios deberían ser de dimensión 3 para todos los partos, pero como algunas correlaciones a aleatorios resultan ser cercanas a 0, esta dimensión disminuye en algunos casos, ocasionando que la complejidad de ajuste sea mucho menor. Nuevamente estos resultados son soportados por la disminución en los criterios AIC y BIC, que aunque la disminución no es considerable, es menor a la anterior y por tanto representa un mejor ajuste.

Finalmente, una vez ejecuta el método nuevamente y teniendo en cuenta las consideraciones mencionadas hasta este momento, se generan los parámetros b_i , encargados de entregar las predicciones de los parámetros para cada una de las reses; con lo que se procede a sumar b_i con los estimados de efectos fijos para producir el modelo de efectos mixtos final. Este proceso se repite para

cada uno de los partos mediante un bucle, pero por motivos de espacio y lectura no son mostrados a continuación. (Ver script de Matlab denominado “AnaMixEffGenBeta.m” en [37]). Los resultados obtenidos por el modelo de efectos mixtos son contrastados con la primer estimación no lineal del método de LMA y mostrados en la Figura 8.7:

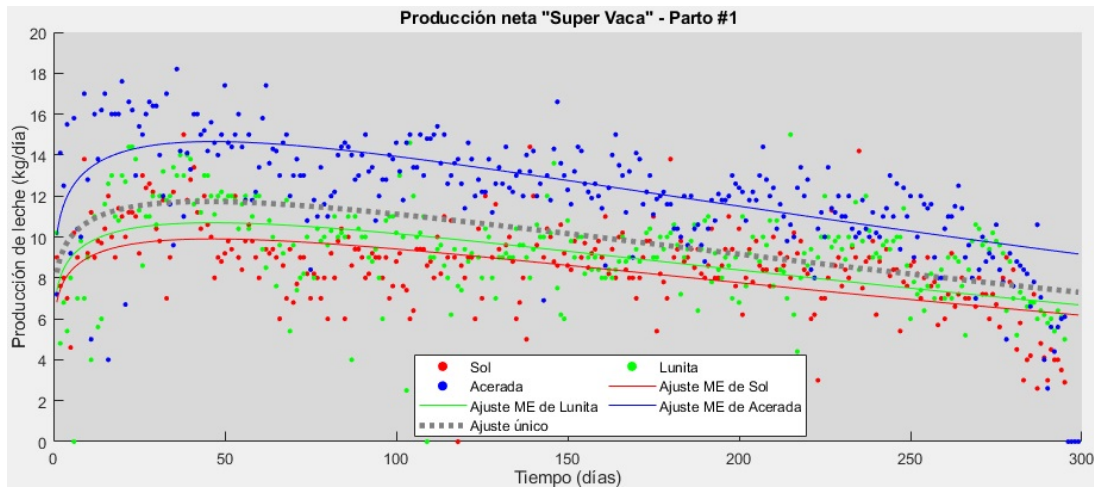


Figura 8.7: Comparación final entre: a) Ajuste por NLMEA-RE y b) Ajuste por LMA .

Adicionalmente, si no tenemos en cuenta la poca cantidad de valores atípicos existentes en los datos y evidenciados en los diagramas de cajas y bigotes; se puede establecer que la siguiente gráfica de probabilidad normal de los residuos muestra un acuerdo razonable con los supuestos del modelo sobre los errores (ver Figura 8.8).

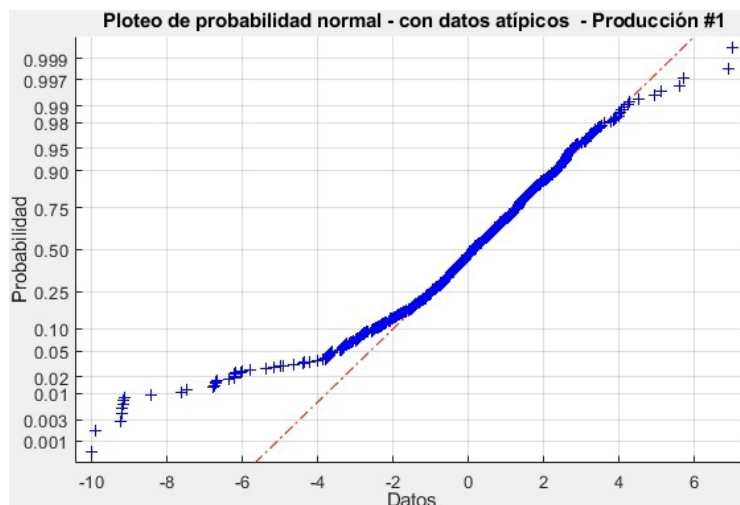


Figura 8.8: Gráfica de probabilidad normal en los residuos. Parto #1.

Llegados a este punto es importante mencionar 2 observaciones significativas:

1. Este análisis descriptivo realizado a lo largo del Capítulo 8 es llevado a cabo para cada una de las 18 reses en cada uno de los 6 partos, pero por motivos de espacio y extensión de este trabajo; solo se muestran resultados del primer parto. Las gráficas, análisis, resultados, scripts y algoritmos adicionales pueden ser consultados en los repositorios de este trabajo en [38] y [37].
2. Finalizado esta Sección, se ha completado con el segundo bloque mencionado en la Figura 4.15, puesto que en este punto ya se han clasificado y seleccionado los registros que serán usados para el ajuste de curvas y la elaboración del (los) modelo(s) por EDOs.

8.2. Modelación dinámica

8.2.1. Planteamiento del modelo

Para plantear el modelo se debe tener en cuenta los datos existentes referentes a las variables que se deseen analizar. Como primera instancia, los únicos 2 tipos de datos que se conocen son, los registros diarios de leche, para vacas de ordeño clasificadas por número de partos; y los registros de peso aproximadamente trimestrales. En cuanto a otras variables como la cantidad de alimento o la cantidad de estiércol producido se pueden considerar un rango de valores aproximados entre un 1 % y 10 % respecto al peso del animal.

Como primera instancia, Consideraremos que las variables usadas para modelar el comportamiento de producción de CA-SENA-POP serán, el peso, la producción de leche y la producción de estiércol denominados como W, M (ó L) y E respectivamente.

8.2.1.1. Consideraciones iniciales

- El peso, la leche y el estiércol dependen del consumo de alimentos secos, forrajes, concentrados, suplementos vitamínicos y agua.
- Con base en la observación anterior, la ingesta neta de alimentos α_s consta de la ingesta de alimento que puede representarse mediante parámetros proporcionales al peso; siendo estos el consumo de forrajes α_f , el consumo de alimentos concentrados α_c y el consumo de agua α_a . En tanto que:

$$\alpha_s = \alpha_f + \alpha_c + \alpha_a \implies \alpha_s \geq 0 \quad (8.1)$$

$$\alpha_f \approx (10 \sim 20)[\%]W. \quad \alpha_c \approx (1 \sim 4)[\%]W. \quad \alpha_a \approx (5 \sim 15)[\%]W.$$

- Para reducir la complejidad de análisis en este y futuros modelos, se asume que:

$$\alpha_f = 0,2 \quad \alpha_c = 0,02 \quad \alpha_a = 0,1 \implies \therefore \alpha_s = 0,32$$

8.2.2. Modelo EDO-WP

Tal y como se menciona en [36], [35] y [39], una forma de representar un modelo mediante Ecuaciones Diferenciales Ordinarias (EDOs), es considerar un balance de energía. Esto es, considerar que las variaciones de una variable en específico con respecto al tiempo, dependen de la combinación lineal de parámetros y variables de estado que afectan positiva y negativamente las dinámicas en cuestión. Por tal motivo podemos proceder a plantear que:

- La producción de estiércol E y la producción de leche M , son productos provenientes de la res productora y que dependen de la evolución que tenga el peso de animal W .
- Tanto la leche M así como también el estiércol E pueden considerarse como una única variable conjunta P , que representa la materia producida por el animal (Producción de estiércol E + Producción de leche L).
- La variación en el tiempo de energía asociada al peso W , puede expresarse como la parte del alimento consumido que se transforma en peso W , menos el peso perdido por la energía consumida por producción de leche y estiércol.

$$\frac{dW}{dt} = k_1\alpha_s W - k_2 P \quad (8.2)$$

- La variación en el tiempo de energía asociada a la materia producida P , puede expresarse como la parte del alimento consumido que se transforma en materia producida P , menos una parte de la materia producida P por la energía consumida en su producción.

$$\frac{dP}{dt} = k_3\alpha_s W - k_4 P \quad (8.3)$$

- Con base en las Ecuaciones 8.2 y 8.3 se puede plantear un diagrama de bloques relacionando las entradas y salidas entra ambos estados W y P (ver Figura 10.2, en el Anexo 10.7).
- Con base en las Ecuaciones 8.2 y 8.3 se puede plantear una ecuación diferencial de segundo orden usada para establecer la relación existente entre los parámetros de amortiguamiento k_1 , k_2 , k_3 y k_4 . Esta demostración matemática se puede corroborar en el Anexo 10.6
- k_1 es la tasa a la que cambia el peso, debido a la ingesta de alimento neto α_s
- k_2 es la tasa a la que el peso decae, debido al consumo de energía por procesos naturales.
- k_3 es la tasa a la que cambia la materia producida por la vaca, debido a la ingesta de alimento neto α_s .
- k_4 es la tasa a la que la materia producida decae, debido al consumo de energía por procesos naturales.

8.2.2.1. Comportamiento del modelo EDO-WP

Con base en las consideraciones descritas en la Sección 8.2.2, es plausible considerar que las ecuaciones 8.2 y 8.3 pueden representar el balance de energía de las variables W y P. Por tanto, llegados a este punto es oportuno analizar el comportamiento de estas ecuaciones diferenciales y determinar si éstas pueden describir un comportamiento similar a las funciones que describen la evolución de los parámetros del peso W y de la materia producida P. Para ello, se procede a realizar el análisis en la herramienta de simulación “Simulink” de “Matlab”.

8.2.2.2. Representación por bloques del modelo dinámico

Partiendo de las ecuaciones 8.2 y 8.3, la primer ecuación consta de la resta de 2 partes. En primer lugar, un producto entre 1 parámetro α_s con valor constante, y un parámetro k_1 cuyo valor cambiará dependiendo de cada sujeto en cuestión, posteriormente estos 2 parámetros se multiplican por la variable de estado W. En segundo lugar, al producto anterior se le subtrae un producto entre un parámetro k_2 y la variable de estado P. La segunda ecuación del modelo muestra un comportamiento similar en cuyo caso solo se cambian los parámetros de amortiguamiento k_1 y k_2 por los parámetros k_3 y k_4 .

Por último pero no menos importante, para observar el comportamiento de la soluciones del sistema es necesario utilizar un bloque de integración con condición inicial preestablecida, seguido de un “scope” para efectos visuales.

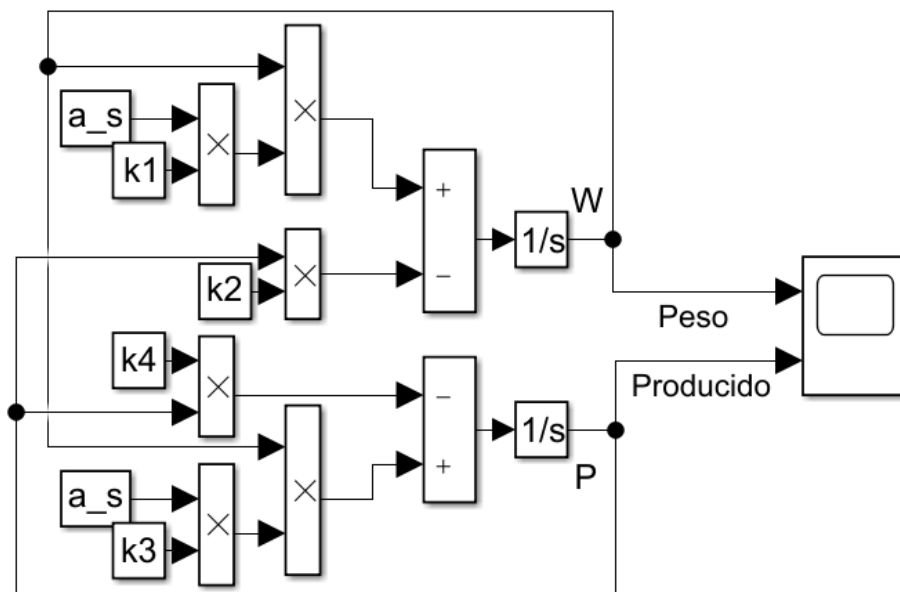


Figura 8.9: Diagrama de bloques del modelo EDO-WP en Matlab-Simulink.

8.2.2.3. Comportamiento del modelo EDO-WP en Simulink

Dejando de lado los valores iniciales existentes para las variables en cuestión para cada una de las vacas mencionadas en la Tabla 7.4, como primera medida se considerarán valores promedios de $400[kg]$ y $40[kg]$ para el peso W y para la materia producida P respectivamente. Estos valores cumplirán su papel como valores iniciales en la simulación de este sistema que se asemeja al de un problema convencional de valor inicial. Una vez establecidas los valores iniciales de los elementos integradores del sistema se procede a observar su comportamiento para distintos valores de k_i con $i = 1, 2, 3, 4$. Los valores de k_i aquí mostrados son netamente ejemplificativos.

- Escenario #1 - ($k_i = 1$):** El resultado de las funciones W y P evidencian un crecimiento desde el valor inicial, hasta llegar a un punto de estabilización en donde el peso y la producción de materia orgánica es constante (ver Figura 8.10). Esto no es sorpresa, pues el factor positivo que incrementa el peso es mayor que el factor negativo que lo disminuye. Y teniendo en cuenta que la leche y heces producidas por el animal dependen de su peso, el comportamiento es más que justificado.

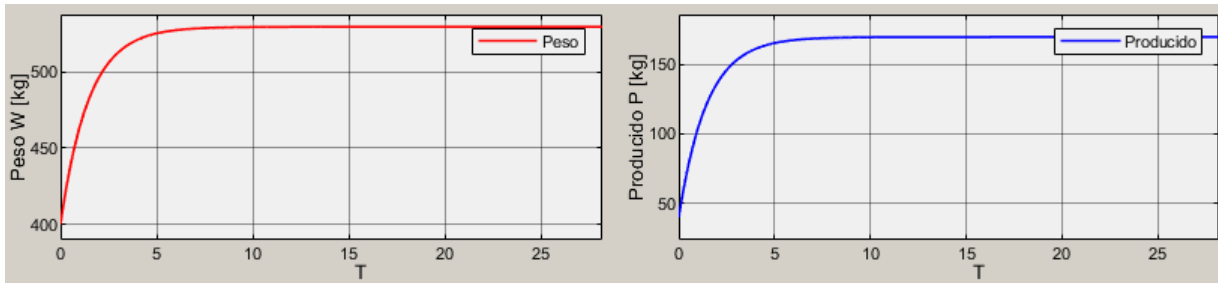


Figura 8.10: Comportamiento del modelo con ($k_i = 1$), para todo i .

- Escenario #2 - ($k_1 > k_2$) y ($k_3 > k_4$) \xrightarrow{Ej} $k_1 = 1, k_2 = 0,5, k_3 = 1, k_4 = 0,5$:** En esta ocasión, el resultado observado (ver Figura 8.11) es similar al anterior, no obstante el incremento presentado es mucho mayor al anterior puesto que la pérdida de energía tanto en el peso como en la materia producida es mucho menor.

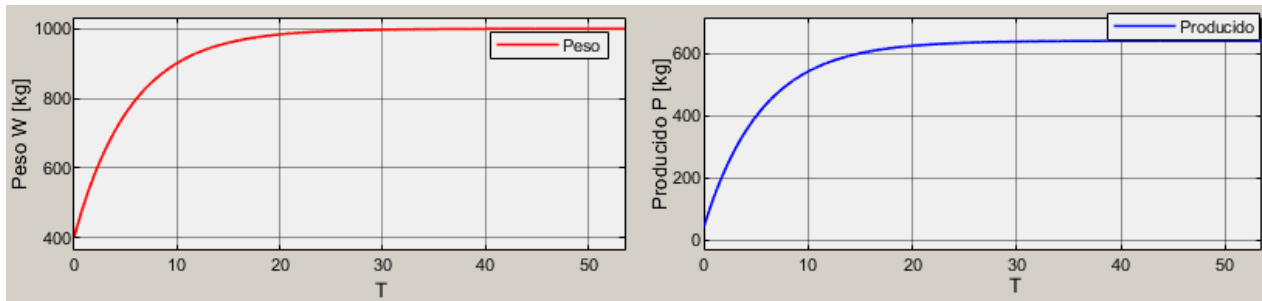


Figura 8.11: Comportamiento del modelo con ($k_1 > k_2$) y ($k_3 > k_4$).

- **Escenario #3** - $(k_1 < k_2)$, $(k_3 < k_4)$, $(k_1 = k_3)$ y $(k_2 = k_4) \xrightarrow{Ej.} k_1 = 0,5, k_2 = 1, k_3 = 0,5, k_4 = 1$: En esta ocasión, el resultado visual es similar al anterior, no obstante el incremento presentado en la Figura 8.12 es mucho menor para ambas variables y además, el tiempo que tarda en estabilizarse es mucho menor.

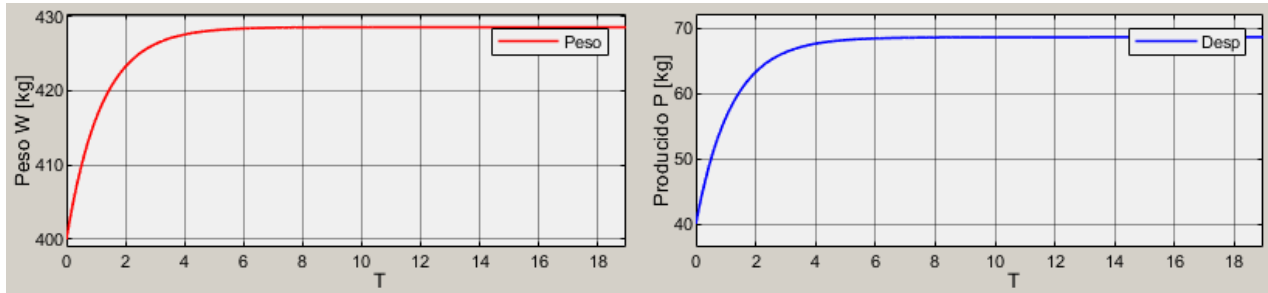


Figura 8.12: Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 < k_4)$.

- **Escenario #4** - $(k_1 < k_3)$ y $(k_2 < k_4) \xrightarrow{Ej.} k_1 = 0,6, k_2 = 0,69, k_3 = 0,65, k_4 = 0,8$: Con esta simulación, se puede observar que el sistema se vuelve inestable (ver Figura 8.13), y tiende a crecer hacia el infinito. Si observamos detalladamente este escenario, si la tasa de conversión de energía del alimento hacia el producto de materia (k_3) es mayor que que la tasa de conversión de energía del alimento hacia el peso (k_1) y además, la tasa de decrecimiento del peso (k_2) es menor que la tasa de decrecimiento de la materia producida (k_4); el peso total del animal irá incrementando más rápido de lo que decrece por procesos naturales, dando como resultado que el peso crezca indefinidamente.

El crecimiento exponencial podrá verse afectado dependiendo de la diferencia en magnitud entre k_2 y k_4 .

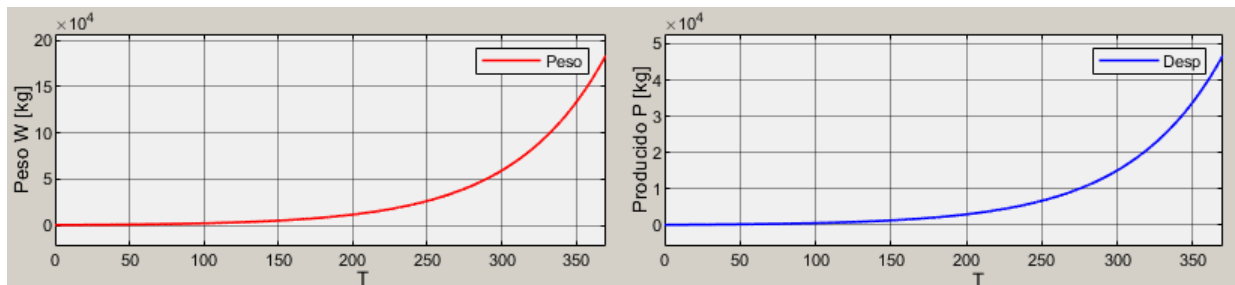


Figura 8.13: Comportamiento del modelo con $(k_1 < k_3)$ y $(k_2 < k_4)$.

- Escenario #5** - $(k_2 > k_1)$ y $(k_3 \gtrsim k_4) \xrightarrow{Ej.}$ $k_1 = 0,1563$, $k_2 = 0,61$, $k_3 = 0,051755$, $k_4 = 0,0510$: En esta ocasión se obtiene un resultado oscilante (ver Figura 8.14), lo que quiere decir que el sistema presenta soluciones con parte imaginaria. Por tal motivo se puede establecer que debe existir una o varias restricciones entre los parámetros k_i .

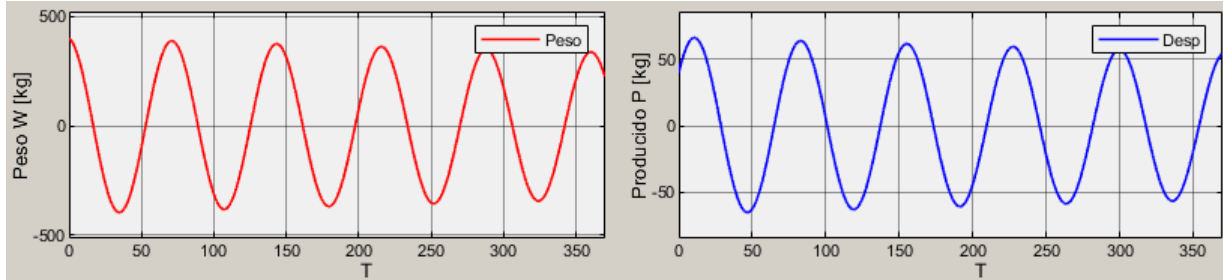


Figura 8.14: Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 \gtrsim k_4)$.

Por un lado, es de esperar que si $(k_2 > k_1)$, significa que la primer ecuación diferencial tenderá a ser negativa y por lo tanto el peso W disminuirá. Por otro lado, al $(k_3 > k_4)$ es de esperarse que la producción de materia sea mayor que perdida de materia producida y por ende P aumentará. Con base en lo visto anteriormente, podemos establecer que se debe cumplir que $(|k_1 - k_2| < k_p)$. Siendo $|k_p|$ un valor pequeño que puede establecerse a futuro, pero que de momento sabemos no puede ser tal que $(k_2 \approx 4 * k_1)$ o de lo contrario el sistema se indeterminará de manera oscilante como el de la Figura 8.14.

- Escenario #6** - $(k_4 < k_3)$ y $(k_1 \gtrsim k_2) \xrightarrow{Ej.}$ $k_1 = 0,051755$, $k_2 = 0,0510$, $k_3 = 0,61$, $k_4 = 0,051$: En esta escenario también se obtiene un resultado oscilante (ver Figura 8.15), lo que quiere decir que el sistema presenta soluciones con parte imaginaria. Por tal motivo se puede establecer que debe existir una o varias restricciones entre los parámetros k_i .

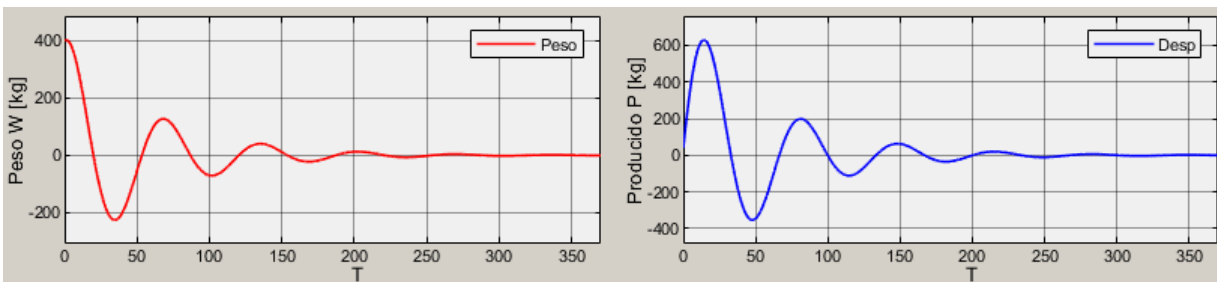


Figura 8.15: Comportamiento del modelo con $(k_1 < k_2)$ y $(k_3 \gtrsim k_4)$.

Por un lado, estamos cumpliendo la restricción establecida en la situación anterior en donde $(k_1 > k_2)$. De esta manera podemos asegurar que la tasa de producción de peso W será

mayor que la tasa de decrecimiento de peso por procesos naturales ($-k_2P$). Por otra parte, si ($k_3 > k_4$), la producción de materia P será mayor que la materia perdida por procesos naturales ($-k_4P$) y por lo tanto \dot{W} decaerá cuando la producción de materia sea muy alta. Cuando el peso disminuya hasta un valor matemáticamente negativo, y por ende la producción de leche y estiércol también lo hagan, la producción de estiércol será menor que la ganancia de peso y esto ocasionará un efecto oscilante tanto en el peso W como en la materia producida P hasta que se establezcan en un punto de equilibrio $(W,P)=(0,0)$.

Con base en las observaciones de esta situación y en las observaciones de los casos donde se inestabiliza el sistema por motivo de ($k_3 < k_4$), podemos establecer que se debe cumplir que ($k_3 \gtrsim k_4$) en pro de conseguir un punto medio antes que se inestabilice o se indetermina por soluciones matemáticamente imaginarias o complejas.

- **Escenario #7** - ($k_1 > k_3$) y ($k_2 > k_4$) $\xrightarrow{Ej.}$ $k_1 = 0,8, k_2 = 0,9, k_3 = 0,6, k_4 = 0,69$: Llegados a este punto se sabe que ($k_3 \gtrsim k_4$) y que ($|k_1 - k_2| < k_p$), pero aún falta determinar las restricciones entre (k_1 y k_3), k_p y (k_2 y k_4). Teniendo en cuenta los valores de k_i considerados en este escenario, se podría pensar que el resultado obtenido tendería a ser estable como el de la Figura 8.12, no obstante en esta situación se cumple que $k_1 \neq k_3$ y $k_2 \neq k_4$.

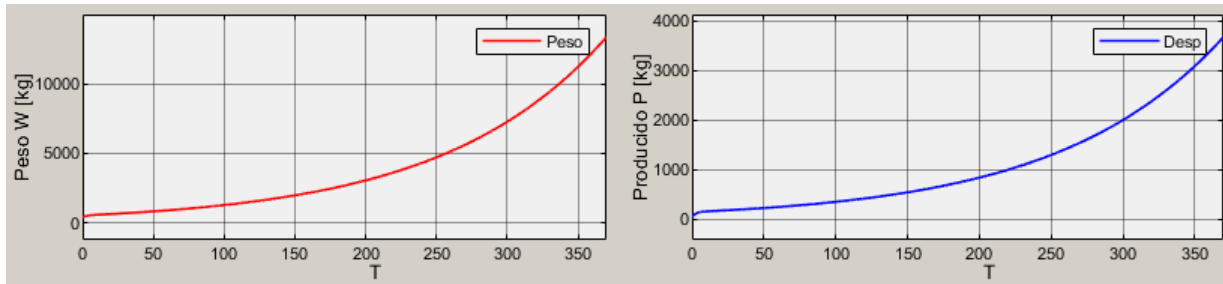


Figura 8.16: Comportamiento del modelo con ($k_1 > k_3$) y ($k_2 > k_4$).

El resulta obtenido en la Figura 8.16 evidencia que aún cuando el sistema tiende a estabilizarse en un valor determinado (cerca de $T=15$), el sistema rápidamente se indetermina hacia ordenes de magnitud alto. Por todo lo descrito en este y los escenarios anteriores podemos establecer que ($k_2 > k_4$) sin que (k_2 sea muy grande comparado con k_4) (ver Figura 8.14). Es decir que ($k_2 \gtrsim k_4$).

- **Escenario #8** - ($k_1 \gtrsim k_3$), ($k_2 \gtrsim k_4$) y ($k_1 < k_2$) $\xrightarrow{Ej.}$ $k_1 = 0,65, k_2 = 0,8, k_3 = 0,6, k_4 = 0,69$: En este penúltimo escenario tenemos un resultado de la forma esperada, tanto para el peso W como para la materia producida P; en donde observamos un comportamiento creciente hasta que se estabiliza en un valor de referencia (que para el modelo es 0.0 en $T \approx 250$). No obstante, se observa que este comportamiento es logrado si se cumple que la diferencia absoluta entre k_1 y k_2 no sea mayor a 0.15. Esto es ($|k_1 - k_2| < k_p$), con $|k_p| \leq 0,15$.

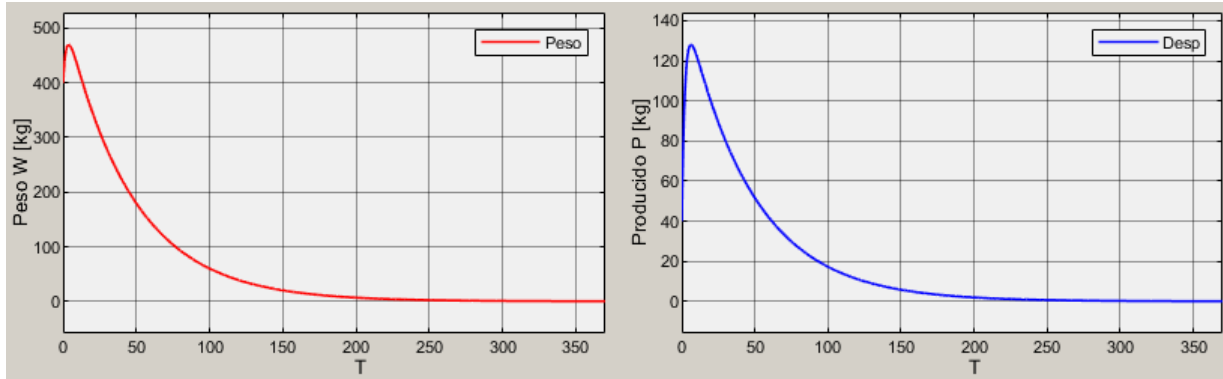


Figura 8.17: Comportamiento del modelo con $(k_1 \gtrsim k_3)$ y $(k_2 \gtrsim k_4)$.

- Escenario #9** - $(k_1 \gtrsim k_3)$, $(k_2 \gtrsim k_4)$ y $(k_1 \gtrsim k_2)$ ($k_p \approx 0,15$) $\xrightarrow{Ej.}$ $k_1 = 0,95$, $k_2 = 0,8$, $k_3 = 0,9$, $k_4 = 0,69$: En este último escenario, nuevamente tenemos un resultado de la forma esperada; tanto para el peso W como para la materia producida P ; en donde observamos un comportamiento creciente hasta que se estabiliza en un valor de referencia (que para el modelo es 0.0). A diferencia del caso anterior, se observa que el valor máximo para ambas variables, es mayor cuando $(k_1 > k_2)$ que cuando $(k_1 < k_2)$; mas sin embargo, el decrecimiento hasta el valor de estabilización es mucho más pronunciado ($T \approx 125$).

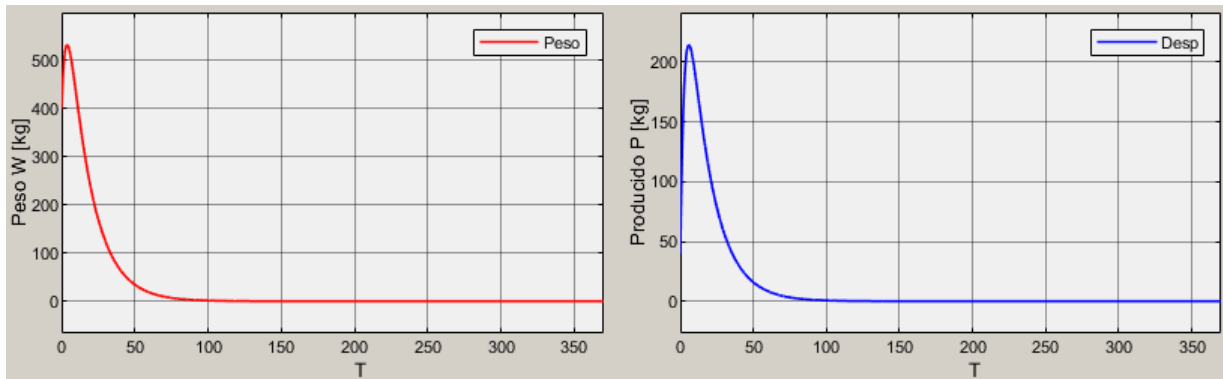


Figura 8.18: Comportamiento del modelo con $(k_1 \gtrsim k_3)$ y $(k_2 \gtrsim k_4)$.

Finalmente se puede llegar a la conclusión que el sistema de ecuaciones diferenciales ordinarias de dimensión 2 propuesto, puede prestarse para hallar soluciones numéricas de problemas de valor inicial para los datos existentes en este trabajo de grado. De esta forma, el modelo EDO-WP será usado para ajustar los datos existentes de Leche M , Estiercol E (Materia producida P) y Peso W ; mediante optimización de parámetros por mínimos cuadrados.

8.2.3. Ajuste del modelo EDO-WP

A partir del análisis descrito en la Sección 8.2.2.3, se puede afirmar que el sistema de ecuaciones diferenciales ordinarias de 2 dimensiones, está en la capacidad de ofrecer soluciones teóricas al problema de valor inicial y cuyas curvas presentan una forma similar a las previamente establecidas en la Sección 8.1. Ahora bien, los valores numéricos de los parámetros k_i usados para ejemplificar los múltiples comportamientos del sistema de EDOs fueron valores usados para valores iniciales ficticios o artificiales tanto para el peso W como para la materia producida P . Por lo tanto, en pro de hallar los valores de los parámetros k_i que corresponden a los registros de cada res y en cada parto, se debe hacer un ajuste del modelo a dichos datos. Este proceso requiere de un procesamiento de datos estructurados a manera de tablas o “dataframes” por lo que por facilidades algorítmicas se procede a implementar éste requerimiento mediante el lenguaje “Python”, en el entorno de programación “Anaconda - Jupyter Notebook” teniendo en consideración la versatilidad que ofrecen ciertas librerías como “numpy” y “pandas” en cuanto a procesamiento de tablas de datos. Este proceso es descrito a continuación.

8.2.3.1. Agrupación de datos para modelado

Como se ha establecido en la Sección 8.1, los registros existentes pueden variar en sus valores máximos, valores mínimos, formas de curva y duración o persistencia; dependiendo del parto del que hayan sido considerados para análisis. Adicionalmente, en la Secciones 7.3 y 7.4 se mencionó que los datos de leche M y peso W fueron reestructurados en forma de tablas de tal forma que se obtuvieran “arrays” de igual tamaño; y que posteriormente serían usados para ajustar los datos a modelos EDOs.

Por otra parte, en la Sección 8.2 se mencionó que el estiércol producido sería considerado como un valor teórico y que éste presentaría una relación proporcional al peso entre un 1 y 10 por ciento del peso del animal, por lo que por efectos de simplificación en la complejidad de análisis; se ha considerado que el estiércol producido E será aproximadamente un 5% del registro de peso W . Por último pero no menos importante, en la Sección 8.2 también se mencionó que el modelo EDO solo utilizaría 2 variables de estado en tanto que a diferencia del peso W , la leche y el estiércol eran consideradas como materia producida y que sería denominada como $P = M + E$ (ó $P = L + E$).

Pues bien, para ejemplificar el proceso algorítmico se utilizará el conjunto de datos que contiene la información de las reses pertenecientes al grupo de partos #3 mas sin embargo este proceso se repite para todos y cada uno de los grupos de partos mencionados a lo largo de este trabajo de grado. En cuanto al proceso de implementación algorítmica secuencial, éste se describe a continuación:

1. En primer lugar, se debe importar el conjunto de datos. Como se puede observar en la Figura 8.19, cada res posee 3 columnas distintas que hacen referencia a cada una de las 3 variables iniciales W , M (ó L) y E . Adicionalmente, a diferencia de las columnas asociadas a los pesos de cada res, las columnas de leche y estiércol están separadas mediante un guión bajo “_”, distinción especialmente útil para agrupar columnas en el siguiente paso.

```

1 path = 'D:\LFGG\PUJ\Maestria_PUJ\Tesis maestria\HW-Pr1-Levenberg-Marquart'
2 modsvp1 = pd.read_csv(path + '\ModSupVac3.csv')
:
8 modsvp1.head()

```

	N	Fer_L	WFer	Fer_E	Gpa_L	WGpa	Gpa_E	Juan_L	WJuan	Juan_E	...
0	1	19.0	515.000000	25.750000	19.0	500.000000	25.000000	13.4	600.000000	30.000000	...
1	2	25.6	515.635687	25.781784	20.8	498.484685	24.924234	6.0	599.162813	29.958141	...
2	3	22.0	516.271212	25.813561	16.0	496.969938	24.848497	13.0	598.325699	29.916285	...
3	4	21.0	516.906413	25.845321	22.0	495.456325	24.772816	11.0	597.488728	29.874436	...

Figura 8.19: Primeros 4 registros del conjunto de datos del grupo de parto #3.

2. En segundo lugar, sumamos las columnas de leche y estiércol, renombramos según corresponda y reorganizamos el conjunto de datos:

```

1 modsvp1.pop('N')
2 modsvp1=modsvp1.T.groupby([s.split('_')[0] for s in modsvp1.T.index.values]).sum().T
   Fer   Gpa   Juan   Leti   Mart   Mor   Sol   WFer   WGpa   WJuan   WLeti   WMart   WMor   WSol
1 numvac=int(modsvp1.shape[1]/2) # Num-Vacas . # EL "2" -> 2 vbles de estado W y P
2 columns=modsvp1.columns[0:numvac]
3 auxtxt='P_'
4 for i in columns:
5     newcol=auxtxt+i
6     modsvp1[newcol]=modsvp1[i]
7     modsvp1.pop(i)
1 modsvp1.head()
   WFer   WGpa   WJuan   WLeti   WMart   WMor   WSol   P_Fer   P_Gpa   P_Juan   P_Leti   P_Mart   P_Mor   P_Sol
0 515.000000  500.000000  600.000000  396.000000  369.000000  620.0  366.0  44.750000  44.000000  43.400000  36.200000  18.450000  43.8  34.7
1 515.635687  498.484685  599.162813  397.570641  368.964752  620.0  366.0  51.381784  45.724234  35.958141  39.078532  18.448238  54.0  31.7

```

Figura 8.20: Conjunto de datos del parto #3 con solo 2 variables W y P para cada res.

3. En tercer lugar, (en el lado izquierdo) separamos las variables de estado en matrices con los datos medidos correspondientes ($\mathcal{X}_1_M = W$ y $\mathcal{X}_2_M = P$); y (en el lado derecho) consideramos los valores iniciales para cada res. Ver Figura 8.21

```

1 X1_M=pd.DataFrame()
2 X2_M=pd.DataFrame()
3 for i,j in enumerate(dfsvp1.columns):
4     foraux=np.array(dfsvp1[j])
5     if(i<=(numvac-1)):
6         X1_M[j] = pd.Series(foraux)
7     else:
8         X2_M[j] = pd.Series(foraux)
9
10 X1_0 =[0 for i in range(X1_M.shape[1])]
11 X2_0 =[0 for i in range(X2_M.shape[1])]
12 # X1_0
13 for i,j in enumerate(X1_M.columns):
14     X1_0[i] = round(float(X1_M[j][i]),6)
15 for i,j in enumerate(X2_M.columns):
16     X2_0[i] = round(float(X2_M[j][i]),6)
17 Y0 = [X1_0, X2_0]

```

Figura 8.21: Variables de estado W y P y sus respectivas condiciones iniciales, para cada res.

8.2.3.2. Optimización del ajuste del modelo por mínimos cuadrados

```

1 def fvp(y, t, paras): # Modelo EDO / ODE Model
2     """SISTEMA EDO / ODE SYSTEM"""
3     a_f, a_c, a_a = 0.2, 0.02, 0.1
4     a_s = (a_f + a_c + a_a)
5     try:
6         k1,k2 = paras['k1'].value, paras['k2'].value
7         k3,k4 = paras['k3'].value, paras['k4'].value
8     except KeyError:
9         k1, k2, k3, k4 = paras
10    x1, x2 = y[0], y[1]
11    # EDOS / ODES:
12    f0,f1 = k1*(a_s)*x1 - k2*x2, k3*(a_s)*x1 - k4*x2
13    return [f0, f1]

```

Figura 8.22: Modelo EDO-WP, implementación en Python.

1. En primer lugar, definimos las funciones del modelo (ver Figura 8.22), el uso del solucionador “odeint” y el calculador de residuos; necesarias para resolver el modelo del tipo de problema de valor inicial (para más información, ver archivos “ODE_MOD WP_Par#i” del repositorio [37]) y restringimos los parámetros del modelo (ver Figura 8.24). Esto se realiza para cada una de las vacas mediante un bucle “for”.

```

11 for i,j in enumerate(dfsvp1.columns[0:numvac]): # Para cada vaca
12
13     params = Parameters()
14     params.add('x10', value=float(X1_M[X1_M.columns[i]][0]), vary=False)
15     params.add('x20', value=float(X2_M[X2_M.columns[i]][0]), vary=False)
16     params.add('k1', value=0.3445, min=0, max=1)
17     params.add('k2', value=0.7, min=0, max=1)
18     params.add('k3', value=0.2755, min=0, max=1)
19     params.add('k4', value=0.5, min=0, max=1)

```

Figura 8.23: Restricciones del modelo EDO-WP implementado en Python.

2. En segundo lugar, optimizamos el modelo propuesto en la Sección 8.2 y repetimos para cada vaca mediante el mismo bucle “for” del paso anterior.

```

26     result = minimize(residualvp, params, args=(t_measured, data), method='leastsq')
27     final = data + result.residual.reshape(data.shape)
28     RESULT.append(result)
29     FINAL.append(final)

```

Figura 8.24: Soluciones de los modelos con ajuste de parámetros por mínimos cuadrados.

8.2.3.3. Visualización de resultados

Tal y como se concluyó en la Sección 8.2.3 donde se estableció que el modelo EDO-WP podría ajustarse a los datos existentes referentes al peso W y a la materia producida P ; es conveniente verificar si este ajuste ha sido logrado; por lo tanto procedamos a observar algunos de los resultados del ajuste del modelo.

Para ello analizaremos los resultados del ajuste a los datos de las reses del grupo de parto #3:

■ Fernanda (Fer)

Para este ejemplar, podemos observar en la Figura 8.25 que el la solución al modelo EDO-WP presenta un comportamiento creciente y luego decreciente, cuya forma es similar a la que se quiere representar. Aún cuando visualmente podría considerarse que el ajuste no es lo suficientemente preciso, es importante resaltar que el ajuste por mínimos cuadrados ofrece errores relativos considerablemente bajos (ver valores de k_i en la Figura 8.26).

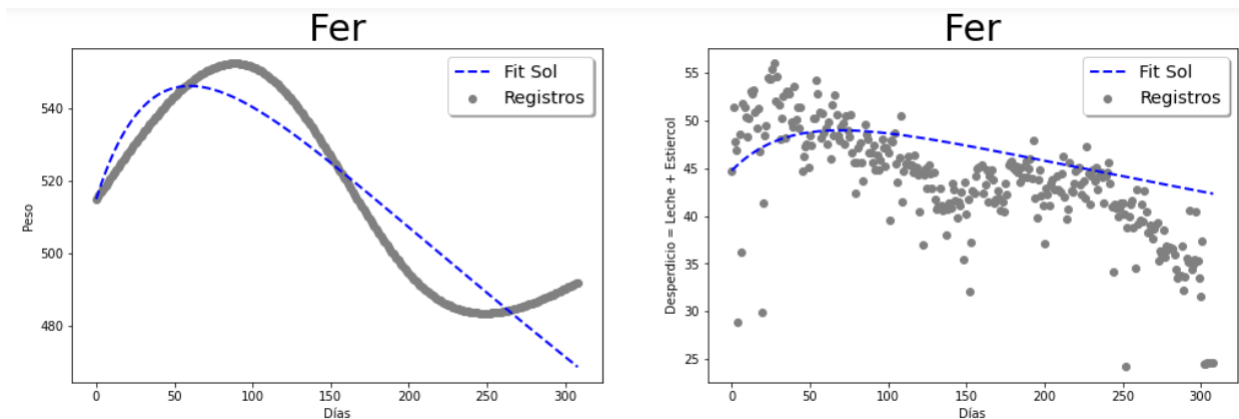


Figura 8.25: Ajuste del modelo EDO-WP, a los datos de la res “Fernanda”.

Entre estos resultados podemos resaltar el hecho que el error relativo para cada uno de los parámetros k_i es menor al 5%.

```
[[Fit Statistics]]
# fitting method = leastsq
# function evals = 366
# data points   = 618
# variables     = 4
chi-square      = 35063.7984
reduced chi-square = 57.1071635
Akaike info crit = 2503.75347
Bayesian info crit = 2521.45942

[[Variables]]
x10: 515 (fixed)
x20: 44.75 (fixed)
k1: 0.27987548 +/- 6.9469e-04 (0.25%) (init = 0.3445)
k2: 1.00000000 +/- 2.0105e-04 (0.02%) (init = 0.7)
k3: 0.03218680 +/- 7.4515e-04 (2.32%) (init = 0.2755)
k4: 0.11480125 +/- 0.00238901 (2.08%) (init = 0.5)
```

Figura 8.26: Variables y datos estadísticos del ajuste.

- Guapa (Gpa)

Con esta res podemos observar que el ajuste del modelo puede aproximar el comportamiento del peso respecto a su etapa decreciente y luego creciente, aún cuando el peso parece oscilar crecientemente alrededor de la solución planteada por el modelo (ver línea punteada en la Figura 8.27). Con base en lo anterior se puede esperar que los parámetros k_1 y k_2 presenten errores considerablemente bajos.

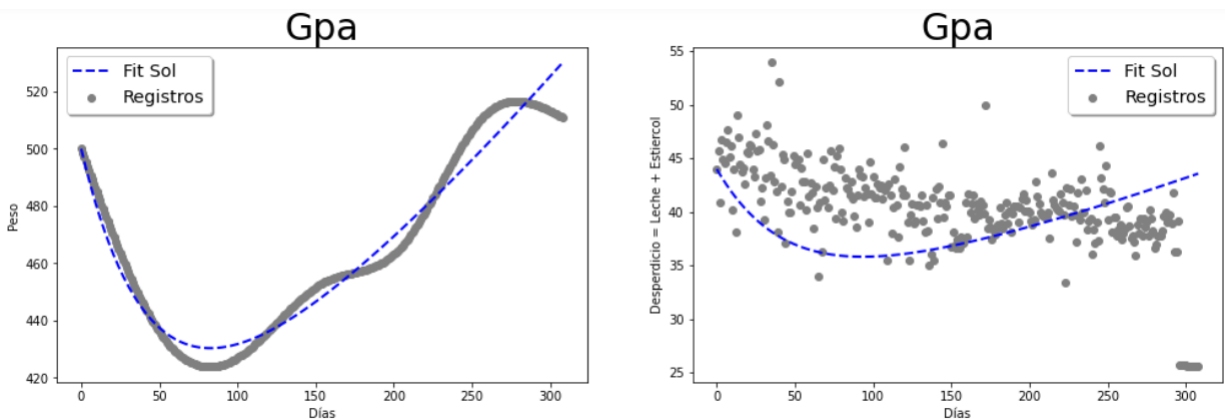


Figura 8.27: Ajuste del modelo EDO-WP, a los datos de la res “Guapa”.

A comparación del ajuste de modelo a la res Fernanda, los parámetros ajustados k_i presentan un error relativo entre el 10 % y 16 % que podría interpretarse que el modelo no puede representar los datos de la variable P de manera satisfactoria. No obstante, también se puede pensar que el ajuste es relativamente mejor, puesto que por una parte el parámetro Chi cuadrado (χ^2) reducido es más cercano a 1 que en el caso anterior, y por la otra parte los criterios AIC y BIC también presentan un valor menor; Estos valores reducidos comúnmente se entiende como un buen ajuste del modelo para los datos experimentales de la res en cuestión.

```
[[Fit statistics]]
# fitting method = leastsq
# function evals = 392
# data points = 618
# variables = 4
chi-square = 21018.4581
reduced chi-square = 34.2320164
Akaike info crit = 2187.48072
Bayesian info crit = 2205.18667

[[Variables]]
x10: 500 (fixed)
x20: 44 (fixed)
k1: 0.26049969 +/- 0.03922543 (15.06%) (init = 0.3445)
k2: 0.99999983 +/- 0.12999928 (13.00%) (init = 0.7)
k3: 0.02688459 +/- 0.00351745 (13.08%) (init = 0.2755)
k4: 0.10349996 +/- 0.01266607 (12.24%) (init = 0.5)
```

Figura 8.28: Variables y datos estadísticos del ajuste.

■ Juanita (Juan)

En cuanto a la vaca Juanita (ó Juan) se puede recalcar un ajuste similar al del caso anterior. Esto puede pensarse puesto que la solución del modelo EDO (línea punteada en la Figura 8.29) hace seguimiento del peso en su forma de onda con parte decreciente hasta aproximadamente los 150 días y su parte creciente hasta final del periodo de lactancia. En cuanto al seguimiento de la materia producida P se puede pensar que el modelo es fiel al comportamiento decreciente similar al que presentan los datos desde el valor máximo hasta los últimos registros cercanos al periodo de secado.

Con base en lo anterior, se puede esperar que los parámetros k_i hayan sido estimados con errores relativos considerablemente bajos. De manera general, se puede afirmar que el modelo se ajusta relativamente bien para ambas variables, aunque con mejor fidelidad hacia el comportamiento decreciente del peso W que de la materia producida P.

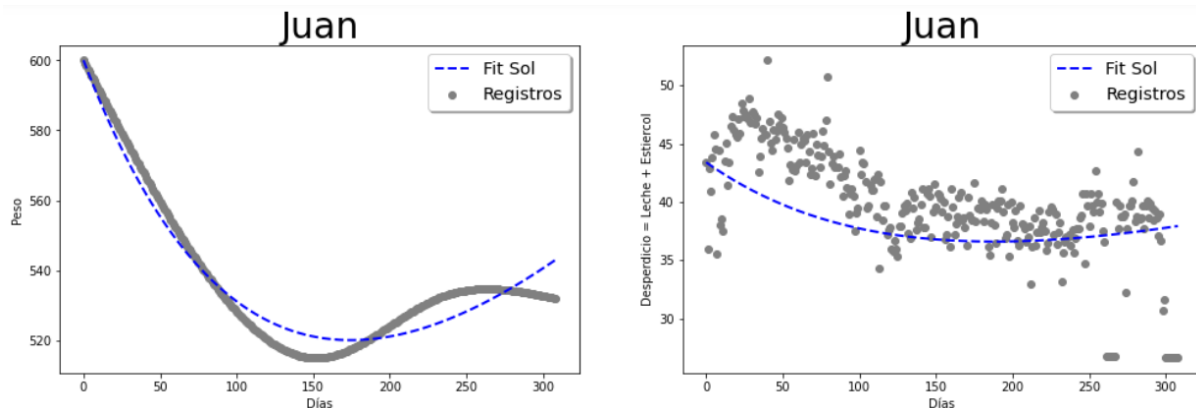


Figura 8.29: Ajuste del modelo EDO-WP, a los datos de la res “Juanita”.

No obstante, aunque el ajuste del modelo no aparente ajustarse a los datos de la variable P, los datos de éste ejemplar han obtenido una minimización del error menor al 1% además de presentar el menor valor de Chi cuadrado (χ^2) reducido en comparación de todos los ajustes del grupo de partos #3. Adicionalmente es importante mencionar que los criterios AIC y BIC disminuyen nuevamente en comparación del ajuste del modelo con los datos de la vaca Guapa.

```

[[Fit Statistics]]
# fitting method = leastsq
# function evals = 691
# data points = 618
# variables = 4
chi-square = 12724.3077
reduced chi-square = 20.7236282
Akaike info crit = 1877.31465
Bayesian info crit = 1895.02060

[[Variables]]
x10: 600 (fixed)
x20: 43.4 (fixed)
k1: 0.22004866 +/- 1.1176e-04 (0.05%) (init = 0.3445)
k2: 0.99999824 +/- 3.4941e-04 (0.03%) (init = 0.7)
k3: 0.01695880 +/- 1.0519e-04 (0.62%) (init = 0.2755)
k4: 0.07715987 +/- 5.1602e-04 (0.67%) (init = 0.5)

```

Figura 8.30: Variables y datos estadísticos del ajuste.

- Sol (Sol)

En esta ocasión tenemos un análisis interesante puesto que en esta ocasión el peso W es constante según se describió en la Sección 7.4.1. Al asumir un peso constante en este periodo de lactancia debido a la falta de registros físicos o digitales de la res en cuestión, se está limitando al modelo para que la primer ecuación del modelo EDO-WP sea aproximadamente 0 y por ende la ganancia de peso por la ingesta de alimento sea igual que la pérdida de peso por procesos naturales. No obstante, para que esta afirmación se cumpla, se necesita que los parámetros k_1 y k_2 sean aproximadamente nulos, observación que claramente será notoria mediante un error de estimación alto; puesto que los valores usados como estimación inicial ($k_1 = 0,34$ y $k_2 = 0,7$) están muy alejados de los valores reales ($k_1 = 4,38e - 10$ y $k_2 = 1,15e - 10$) resultando en un error relativo considerablemente elevado con respecto a los valores de estimación inicial..

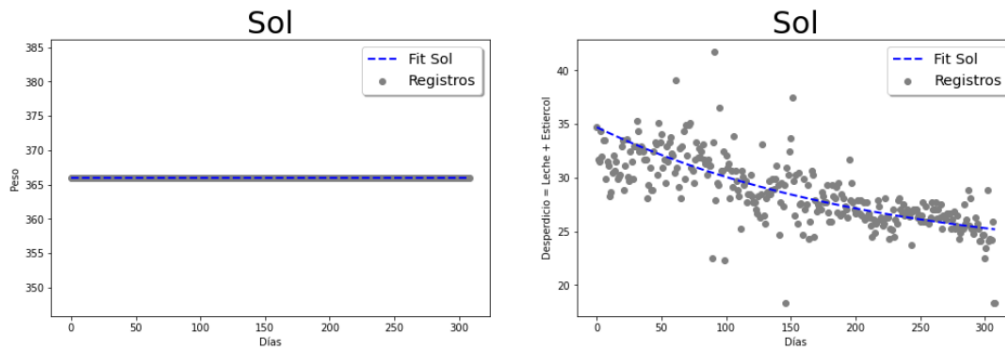


Figura 8.31: Ajuste del modelo EDO-WP, a los datos de la res “Sol”.

Sin embargo, aún cuando el modelo no presente estimaciones iniciales cercanas para los primeros 2 parámetros, los resultados estadísticos generales si presuponen un buen ajuste del modelo a los datos del sujeto de estudio. Afirmación que puede soportarse con los valores de los errores relativos de los parámetros k_3 y k_4 menores al 20% y a los valores más bajos de los criterios de Chi cuadrado (χ^2) reducido, AIC y BIC. En cuanto al criterio (χ^2), al ser el valor más cercano a 1 de todos los ajustes, se puede pensar que el grado de coincidencia entre las observaciones y las estimaciones es acorde con la varianza del error el ajuste.

```
[[Fit Statistics]]
# fitting method = leastsq
# function evals = 626
# data points = 618
# variables = 4
chi-square = 1596.81089
reduced chi-square = 2.60066920
Akaike info crit = 594.652115
Bayesian info crit = 612.358068

[[Variables]]
x10: 366 (fixed)
x20: 34.7 (fixed)
k1: 4.3855e-10 +/- 1.1991e-05 (2734192.75%) (init = 0.3445)
k2: 1.1508e-10 +/- 1.1572e-05 (10055090.24%) (init = 0.7)
k3: 8.7190e-04 +/- 1.5434e-04 (17.70%) (init = 0.2755)
k4: 0.00460748 +/- 6.1074e-04 (13.26%) (init = 0.5)
```

Figura 8.32: Variables y datos estadísticos del ajuste.

Evaluación

En este capítulo se presenta la evaluación del proyecto en cuanto al cumplimiento de los objetivos del proyecto, así como también con respecto a los análisis planteados en la etapa de preparación de datos y los resultados obtenidos. Para ello es importante citar nuevamente los objetivos específicos planteados al inicio del Capítulo 2:

9.1. Objetivos específicos (OE)

9.1.1. OE#1: Identificar entradas, salidas, variables y parámetros asociados al sistema de producción ganadero objetivo.

Teniendo en cuenta el proceso descriptivo realizado en los capítulos anteriores, se puede establecer que la información correspondiente al OE#1 se encuentra resumido de la siguiente manera:

9.1.1.1. Entradas

Son las reses que participan en la producción de leche cruda. Así como también los insumos que permiten el mantenimiento del CA-SENA-POP.

9.1.1.2. Variables

Las variables en cuestión son, por una parte, la cantidad de leche producida por una vaca durante un periodo de lactancia, por otra parte, la cantidad de peso de las reses productoras o no productoras; y por último, la cantidad de estiércol producido por el hato que será tratado para producir abono para el CA-SENA-POP.

9.1.1.3. Salidas

Las salidas del sistema productivo del CA-SENA-POP son, la producción neta de leche cruda, la producción de carne de las reses vendidas y la producción de estiércol y materia fecal que puede ser usado como abono en los distintos potreros internos en la finca del SENA en la ciudad de Popayán. La leche cruda es producida para posteriormente ser vendida para generar ganancias monetarias que serán reinvertidas en el CA-SENA-POP. La carne representa el peso final de la res parida, desteta, cebada o en proceso de venta.

9.1.1.4. Parámetros

El sistema productivo del CA-SENA-POP depende del consumo nutricional del hato. Este consumo varía proporcionalmente al peso del animal y a su rendimiento productivo teniendo un rango de valores proporcionales que oscilan de la siguiente manera:

- Forrajes y alimentos secos: Consumo que varía entre un 10 a 20 % del peso del animal.
- Concentrado y suplementos: Consumo que varía entre un 1 a 5 % del peso del animal.
- Agua: Consumo “Ad libitum”, que varía entre un 5 a 12 % del peso del animal
- Medicamentos: Varían según el estado de salud del animal. Se asume un estado de salud óptimo durante este trabajo de investigación.

9.1.2. OE#2: Identificar modelos dinámicos que permitan pronosticar la producción de un sistema ganadero objetivo.

Tal y como se menciona en la Sección 8.2, soportado por [36], [35] y [39], una forma de encontrar una representación de un sistema por EDOs, es mediante un balance de energía que involucre a las variables en cuestión. Por tal motivo, se puede considerar factible que el modelo de 2 dimensiones descrito en la Sección 8.2.2 satisface el pronóstico evolutivo de las 3 variables de producción del sistema ganadero en cuestión.

9.1.3. OE#3: Implementar métodos de solución de los modelos dinámicos de producción ganadera seleccionados.

Según se concluye en la Sección 8.2.2.3, una vez que se determina que un modelo de ecuaciones diferenciales ordinarias (EDO) puede representar el comportamiento de las variables del sistema, es posible resolver dicho modelo utilizando algoritmos numéricos avanzados, como los proporcionados por la biblioteca “Scipy” en scripts de Python. En este caso particular, el sistema modelado representa un problema de valor inicial, por lo tanto, se puede resolver mediante métodos de solución como “odeint” o “scipy.integrate”. El proceso en el que se emplea este algoritmo para obtener la solución del modelo de EDO para cada una de las producciones de las reses pertenecientes al sistema está descrito en la Sección 8.2.2.3, o en mayor detalle, en los scripts de Python implementados en Jupyter-Anaconda. [37].

9.1.4. OE#4: Evaluar la estimación de los modelos seleccionados comparando la solución con datos empíricos recogidos.

En cuanto a este objetivo, el modelo WP, planteado por el sistema de 2D de EDOs descrito en el capítulo anterior tiene un desempeño deseado puesto que tal y como se describe en la Sección 8.2.3.3, los valores de los criterios BIC Y AIC son menores a medida que el modelo logra un ajuste más apropiado para cada una de las reses de cada parto. Adicionalmente, criterios como el χ^2 reducido

tienden a ser cada vez más próximo a 1 lo que complementa la afirmación anterior. Por último pero no menos importante, aún cuando esta última afirmación es netamente subjetiva, se puede observar que cuando el modelo presenta características estadísticas como las previamente descritas en este objetivo, el ajuste visual tiende a ser similar al comportamiento de los datos empíricos recogidos.

9.2. Metodología CRISP-DM

En lo que respecta a la metodología de estudio, las actividades llevadas a cabo en las etapas de la metodología CRISP-DM descritas en los capítulos anteriores pueden ser justificantes suficientes para corroborar si se ha hecho un debido manejo y procesamiento de datos, en especial, porque cada una de las acciones realizadas al conjunto de datos base ha sido detallada y justificada en cada capítulo en pro de satisfacer los objetivos de este trabajo de grado.

De manera general, se puede mencionar que:

1. En cuanto al primer objetivo específico (OE#1), las etapas 1, 2 y 3 de la metodología CRISP-DM permiten identificar el contexto del CA-SENA-POP, así como también los recursos que participan en el proceso productivo del hato.
2. Para satisfacer el OE#2, basta con realizar un proceso repetitivo entre las etapas CRISP #1, #2, #3 y #4; en pro de determinar los parámetros y variables que deben ser tenidos en cuenta al momento de realizar el ajuste de los datos a modelos EDO.
3. Respecto al OE #3, éste puede ser satisfecho una vez cumplidas las primeras 3 etapas de la metodología CRISP.
4. Por último pero no menos importante, el OE #4 podrá ser satisfecho en el transcurso de esta etapa de evaluación.

9.3. Cumplimiento del OE#4 en la etapa de Modelación: Ajuste de curvas por métodos no lineales

Tabla 9.1: Comparación de error cuadrático medio (MSE) para cada método de ajuste no lineal.

MSE						
MÉTODO DE AJUSTE	PARTO ($P\#_i$)					
	$P\#_1$	$P\#_2$	$P\#_3$	$P\#_4$	$P\#_5$	$P\#_6$
MA: FU-LM	6.8240	17.8285	21.4705	12.4351	11.3886	4.8692
MA: FM	4.4826	8.444	12.7492	8.7615	9.4352	4.7066
MA: EF	4.4855	8.4482	12.7476	8.7735	9.4402	4.6487

Para realizar una evaluación crítica respecto a los ajustes no lineales de las curvas de lactancia para determinar si estos son representativos y significativos; es necesario tener en consideración los criterios de MSE, RMSE; para cada uno de los métodos de ajustes no lineales realizados. (Levenberg-Marquardt ó LM, Fit Multiple ó FM, Efectos Mixtos ó EM).

Tabla 9.2: Comparación de la raíz del error cuadrático medio (RMSE) para cada método de ajuste no lineal.

RMSE						
MÉTODO DE AJUSTE	PARTO ($P\#_i$)					
	$P\#_1$	$P\#_2$	$P\#_3$	$P\#_4$	$P\#_5$	$P\#_6$
MA: FU-LM	2.6123	4.2224	4.6336	3.5263	3.3747	2.2066
MA: FM	2.1172	2.9059	3.5706	2.96	3.0717	2.1695
MA: EF	2.1179	2.9066	3.5704	2.962	3.0725	2.1561

En primera instancia, los ajustes no lineales realizados al inicio del Capítulo 8 evidencian características estadísticas con valores relativamente bajos de MSE y RMSE (ver Tabla 9.1 y Tabla 9.2), así como también una reducción progresiva de los criterios AIC y BIC a medida que se comparan los métodos de ajuste de curvas por los algoritmos de efectos mixtos fijos y aleatorios no lineales (ver Tabla 9.3).

Tabla 9.3: Comparación del Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) para los métodos de ajuste no lineales por efectos mixtos fijos y aleatorios

MÉTODO DE AJUSTE (MA)	PARTO ($P\#_i$)					
	$P\#_1$	$P\#_2$	$P\#_3$	$P\#_4$	$P\#_5$	$P\#_6$
AIC						
MA: EM FIJOS	3922.319	10130.376	11747.833	9700.2539	3077.2224	1433.494
MA: EM ALEATORIOS	3921.065	10118.0196	11734.873	9700.95	3077.2224	1435.2825
BIC						
MA: EM FIJOS	3916.9107	10129.9979	11747.4544	9699.8753	3068.0744	1425.6529
MA: EF ALEATORIOS	3914.7556	10117.5869	11734.4402	9700.3623	3068.0744	1423.1048

Por lo tanto, es plausible afirmar que los ajustes realizados presentan una buena bondad de ajuste. (Para mayor información respecto a los características estadísticas de cada ajuste, consultar en [37], los scripts de Matlab “AnaMixEffGenBeta2.m” y “lm_examGEN2023.m”).

9.3.1. Producción neta real Vs Producción teórica por ajustes no lineales

Una forma de evaluar la adecuación de los ajustes no lineales de los parámetros del modelo de Wood es mediante la verificación del área bajo la curva generada por estos parámetros. Esta curva representa la producción neta de lactancia, obtenida a partir de los parámetros ajustados al modelo matemático de Wood. Para ello, se puede crear una tabla que contenga los valores de las producciones netas reales de grupo de parto (Producción Real ó PR) y se compare con las producciones netas estimadas (Producción estimada PT) por cada método de ajuste (ó “Fit”). De esta manera, se puede determinar si los ajustes son capaces de representar los datos de cada vaca y, en última instancia, la producción total del hato. Esta comparación puede apreciarse en la Tabla 9.4, los valores de producción están en toneladas [t] y el total de ganancia por venta de leche “\$\$\$ T_{total} ” está en millones de pesos colombianos [COP].

En esta misma Figura se puede observar como el método no lineal que presenta mejor ajuste es el LMA, el cual, aunque no presenta los porcentaje de error más pequeños para todos los grupos de partos, si presenta un valor total más aproximado con respecto a la producción total real de leche, basada en registros.

9.3.2. Modelo EDO: Ajuste por mínimos cuadrados

Tal y como se menciona en la Sección 8.2.3.2, una vez se ha hecho el llamado al solucionador de EDO “odeint” restringiendo los valores de cada parámetro k_i , se pueden obtener los parámetros que permiten el ajuste del modelo EDO a los datos de peso W y de materia producida $P=L+E$.

Según se explica en el apartado de resultados de esa misma Sección, se puede ver como el modelo presenta una bondad de ajuste apropiada que, aunque no aplica para todas las reses de todos los partos, posee buenos criterios estadísticos como los criterios AIC y BIC bastante reducidos entre mejor es el ajuste, así como también un valor de χ^2 reducido cercano a 1 (ver Figura 8.32). Por último pero no menos importante, se observa como para algunas reses en cada parto, los parámetros k_i poseen errores relativos inferiores al 20 % lo que podría considerarse como un buen ajuste generalizado.

Es importante mencionar que el modelo EDO-WP si bien puede presentar un buen ajuste para los datos asociados a algunas reses, no es un modelo perfecto por lo que no es sorpresa evidenciar que el modelo no presenta un buen ajuste para todas las reses. No obstante, esta observación también esta condicionada al hecho que no se cuentan con los datos apropiados de peso y producción de estiércol, lo que dificulta brindar un ajuste preciso para datos faltantes o especulativos.

Con base en lo anterior se puede establecer que el OE #3 ha sido cumplido, en tanto que la solución numérica del modelo, permite lograr un ajuste bondadoso a los datos existentes. De igual manera se establece que el OE#4 ha sido cumplido en tanto que el modelo EDO-WP ha sido evaluado con respecto a los datos empíricos recogidos, al mostrar buenas características estadísticas que permiten afirmar que presenta un buen ajuste respecto a los datos usados.

Tabla 9.4: Comparación de producción real y teórica para cada método de modelación utilizada.

PARTO $P\#_i$	N	$Prod_{Real}$ $PR_{P\#_i}$	LMA PT	$\%Error$	S_{Vaca} PT	$\%Error$	NLMEA-F PT	$\%Error$	NLMEA-R PT	$\%Error$	EDO-WP PT	$\%Error$
$P\#_1$	299	8.873	8.885	0.146 %	8.876	0.041 %	8.877	0.045 %	8.912	0.449 %	8.215	7.412 %
$P\#_2$	288	24.712	24.751	0.158 %	24.800	0.357 %	24.835	0.501 %	24.844	0.534 %	24.28	1.745 %
$P\#_3$	309	33.097	33.102	0.015 %	33.114	0.052 %	33.123	0.079 %	33.130	0.049 %	35.129	6.142 %
$P\#_4$	274	20.696	20.686	0.045 %	20.722	0.129 %	20.725	0.144 %	20.731	0.173 %	20.114	2.809 %
$P\#_5$	298	7.111	7.154	0.604 %	7.126	0.217 %	7.136	0.354 %	7.138	0.384 %	7.125	0.198 %
$P\#_6$	162	6.2707	6.2700	0.013 %	6.2705	0.004 %	6.2705	0.004 %	6.269	0.017 %	6.287	0.271 %
P_{TOTAL}	[t]	100.760	100.850	0.089 %	100.911	0.150 %	100.968	0.207 %	101.027	0.265 %	101.153	0.391 %
$$$$_{TOTAL}$	[COP]	136.026	136.147	0.089 %	136.230	0.150 %	136.307	0.207 %	136.386	0.265 %	136.557	0.391 %

9.4. Retroalimentación del proyecto con el cliente final

El desarrollo de este trabajo pudo ser llevado a cabo tras recibir una respuesta positiva por parte de un cliente (CA-SENA-POP) quien vió en este trabajo de grado, la oportunidad de iniciar un proceso de tecnificación enfocada hacia la industria 4.0 en el sector agropecuario; mediante un proyecto de investigación. Este cliente fue quien proporcionó los espacios y sobretodo los datos que permitieron que los conocimientos del estudiante de maestría, el Ing. Luis Felipe Guevara Gómez, pudieran ser puestos en practica. Como medida de evaluación del proyecto se considera pertinente tener en consideración la opinión que el cliente pueda tener respecto a cómo se trataron sus datos y qué resultados fueron obtenidos gracias a estos debido al proyecto de investigación llevado a cabo.

Por tal motivo, a continuación se describen los espacios de socialización que fueron llevados a cabo con el personal del CA-SENA-POP.

9.4.1. Socialización del proyecto con el CA-SENA-POP

9.4.1.1. Socialización Virtual: Miércoles 10 de Mayo de 2023

Para esta primer socialización se contó con la asistencia de la coordinadora de formación profesional integral Maria del Carmen Agredo, el coordinador del centro agropecuario Daniel Cobo, el profesor y tutor, Dr. Luis Eduardo Tobón Llano y el Ing. Luis Felipe Guevara Gómez, quien realiza el proyecto. En esta reunión de poco mas de 20 minutos, se mostró cómo se usaron los datos oficiales proporcionados para que fueran procesados, analizados y clasificados para posteriormente estimar parámetros, realizar ajuste de datos; y finalmente establecer un modelo por EDOs que permitiese predecir el comportamiento de las variables de interés (Ver Figura 4.15). (La presentación usada puede ser encontrada en el repositorio [38]).

Durante esta sesión, la coordinadora Maria del Carmen manifestó una opinión positiva respecto al desarrollo del proyecto, pues según sus palabras, aunque el SENA posee los recursos económicos y humanos para tecnificar su producción ganadera; no cuenta con un proceso sistemático ni con estudios de investigación a nivel de posgrado que soporten la toma de decisiones a futuro. En cuanto al manejo y procesamiento de los datos, el SENA se mostró satisfecho al ver que las interpretaciones sobre el conjunto de datos han sido apropiadas y acordes a lo que se maneja en el SENA en tanto

que no se realizan suposiciones alejadas de la realidad y fueron soportadas por análisis pertinentes por parte del ingeniero Guevara.

Finalmente, la coordinadora Maria del Carmen y el coordinador Daniel Cobo manifestaron en su calidad como representantes del CA-SENA-POP, tener las puertas abiertas para formar una articulación entre la Pontificia Universidad Javeriana Cali y el SENA Popayán, pues manifiestan estar satisfechos con el proyecto de investigación realizado y esperan que futuros proyectos de pregrado o posgrado puedan ser llevados entre las 2 instituciones. Al finalizar la reunión, la señora Agredo solicitó una segunda reunión de manera presencial en la ciudad de Popayán para que los resultados numéricos y visuales puedan ser expuestos al personal del CA-SENA-POP, en especial a la veterinaria Ana Maria Dorado quien no pudo asistir sino hasta el final de la reunión y quien junto con el señor Cobo, dirigen el centro agropecuario y sus intereses como cliente final de este estudio.

9.4.1.2. Socialización presencial: Viernes 19 de Mayo de 2023

Para esta segunda reunión, nuevamente se muestran los resultados del proyecto de manera que se pudieran evidenciar las virtudes y bondades que ofrece el modelamiento dinámico de sistemas de producción ganadera; y cómo este tipo de modelamiento permitiría pronosticar de manera aproximada el comportamiento productivo del sistema ganadero del SENA y de otros centros agropecuarios del país. Nuevamente, el personal directivo del SENA expresa su satisfacción respecto a los resultados expuestos y mediante una corta encuesta manifiesta tener una aceptación positiva del proyecto dando por terminada la sesión de retroalimentación del proyecto.

Adicionalmente manifiestan que, aún cuando se cumplen con los objetivos y alcance de este proyecto, quedan expectantes ante futuros proyectos de investigación en el sector agropecuario; pues manifiestan que el SENA cuenta con los recursos biológicos y los espacios interdisciplinarios apropiados para que el cuerpo estudiantil de la PUJ pueda poner en práctica sus proyectos de integración profesional en caso que lo requieran y lograr así una articulación educativa entre las 2 instituciones.

Conclusiones y trabajos futuros

10.1. Conclusiones

De acuerdo con la revisión del estado del arte, aún cuando se han planteado modelos dinámicos de producciones ganaderas para la predicción de comportamientos biológicos, físicos o químicos; estos modelos están enfocados en simular el comportamiento de forma que puedan estimar la producción orgánica de los procesos intracorporales. No obstante, este tipo de modelado requiere de un elevado conocimiento de los procesos físico-químicos y biológicos que se presentan en los animales de las producciones ganaderas, además de los conocimientos de ciencias aplicadas para modelamiento dinámico. Esto último aumenta considerablemente la dificultad del análisis sistemático y computacional para tecnificar una producción ganadera, pues es necesario tener conocimiento técnico avanzado del sector agropecuario para lograr tal fin. Por consecuencia, se considera que el enfoque planteado en este trabajo de grado para abarcar la problemática, es la principal contribución del mismo.

En particular, se muestra que se pueden utilizar modelos dinámicos por ecuaciones diferenciales ordinarias basados en balances de energía sencillos; estimando un comportamiento aproximado respecto a los datos reales de Leche (L), Peso (W) y Estiércol (E) obtenidos durante los periodos de lactancia en una producción ganadera cualquiera, en cuyo caso de estudio es el CA-SENA-POP.

Los resultados obtenidos evidencian que al implementar soluciones numéricas de un modelo dinámico EDO de 2 dimensiones, éstas soluciones pueden ajustarse apropiadamente al comportamiento creciente y decreciente de las variables de producción deseadas (L, W y E); y que, aún cuando algunos datos inexistentes han sido interpolados o generados de manera artificial, esto no es un impedimento para que el modelo dinámico pueda proyectar posibles comportamientos en el futuro.

Adicionalmente, el entendimiento, la preparación, el procesamiento y la estructuración de los datos que se ha realizado en este trabajo de grado, permite que la proyección que otorga el modelo dinámico planteado, varíe dependiendo de insumos externos que son suministrados a las reses del hato como parámetros de entrada; proporcionando así, una abstracción de la cantidad de alimento, forraje, agua y otros elementos dietarios que afectan el rendimiento productivo de los animales (salidas).

También es pertinente hacer hincapié en que aún cuando la metodología usada (CRISP-DM) no incluye una etapa de estimación de parámetros mediante métodos de ajuste no lineal, abarcar el problema desde este enfoque permite contrastar la precisión de los ajustes no lineales y los ajustes dinámicos evidenciados en los resultados de este proyecto, más precisamente en cuanto a la

producción lechera del hato (ver Figura 9.4); en tanto que se puede observar que las producciones netas presentan valores absolutos significativamente aproximados durante el periodo de estudio de aproximadamente 4 años (2018 - 2022).

Por último pero no menos importante, se concluye que el alcance del proyecto ha cumplido con las expectativas del cliente, quien ha manifestado una posición satisfactoria en cuanto al desarrollo del proyecto y ha manifestado su interés en proyectos futuros tanto del Ing. Luis Felipe Guevara Gómez, como de otros estudiantes de la Pontificia Universidad Javeriana Cali. Para terminar, se puede concluir este proyecto de investigación manifestando que el objetivo general y los objetivos específicos han sido satisfechos en su totalidad.

10.2. Posibles mejoras y trabajos a futuro

Tomando en cuenta los alcances, limitantes, y recursos presupuestados para este proyecto, se pueden identificar características de este sistema prototipo que pueden ser corregidas, mejoradas y optimizadas en implementaciones de mayor escala en trabajos futuros. Entre estas se puede mencionar:

10.2.1. Modelo EDO en 3D: MEW - Leche, Estiércol, Peso

Por una parte, dado que el modelo usado disminuía la complejidad del análisis al combinar 2 variables en 1, y por otra parte, la evaluación de este modelo ha resultado ser satisfactoria en cuanto a los OE#3 y #4; se podría considerar como una mejora el considerar un modelo EDO en 3D analizando de manera independiente el comportamiento de la producción de peso (W), la producción de Leche (M o L) y la producción de Estiércol (E).

10.2.2. Modelado por Autómata híbrido

Teniendo en cuenta que el modelo EDO 2D planteado puede presentar distintos comportamientos de ajuste (creciente, decreciente, amortiguado y subamortiguado), y teniendo en cuenta que hay ejemplares para los que puedan presentar cambios bruscos en su peso; es plausible considerar un cambio de parámetros k_i para cuando el peso cambie de manera alternante cuando las derivadas varían entre valores positivos y negativos, en cuyo caso el modelo 2D usado no puede hacer seguimiento exacto (Tomar como ejemplo el comportamiento de la vaca “Fernanda” del parto #3 en la Sección 8.2.3.3, entre otras).

Así pues, mediante un intercambio de valores de los parámetros k_i se podría lograr un mejor ajuste, abarcando así a más ejemplares que tengan este tipo de comportamientos o de los que no se tengan datos tan completos como algunos bovinos de este proyecto. Esto podría significar un mejor ajuste para los datos de las reses que no han sido generalizadas con el modelo EDO-WP propuesto.

10.2.3. Modelado por programación lineal entera

El proyecto en cuestión consiste en modelar dinámicamente la producción ganadera del CA-SENA-POP, en pro de predecir o estimar producciones futuras sin la necesidad de poseer los datos reales que corroborarían la precisión del modelo. Hasta ahora, los análisis propuestos requieren de un nivel avanzado de lenguajes de programación, álgebra lineal, ecuaciones diferenciales y estadística básica que pueden presentarse como una barrera al momento de presentar esta solución investigativa hacia los emprendedores ganaderos del centro agropecuario. Por tal motivo, se plantean los conjuntos, las variables, los parámetros y las funciones de objetivo de optimización mediante un ejercicio de optimización lineal que podría prestarse para maximizar ganancias o minimizar emisiones de $[KgCO_{2eq}]$.

En esta ocasión se plantea un problema de optimización lineal entera multipropósito, en donde la estimación de ganancia neta que presenta el CA-SENA-POP presenta un conflicto con la reducción de gases de $[KgCO_{2eq}]$, puesto que a mayor producción de estiércol, leche y carne; mayores son las emisiones de estos Gases de Efecto Invernadero (GEI).

Bibliografía

- [1] IBM. (2021) Introducción al CRISP-DM. [Online]. Available: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=guide-introduction-crisp-dm>
- [2] CARE Perú. (2000) Crianza y manejo de ganado vacuno. [Online]. Available: http://infolactea.com/wp-content/uploads/2016/02/BVCI0002410_2-1.pdf
- [3] G. Images. (2015) Ganadería de engorde. [Online]. Available: <shorturl.at/daqD3>
- [4] P. Fonseca. (2016) Así funcionan los ciclos productivos de las ganaderías. [Online]. Available: <https://www.contextoganadero.com/ganaderia-sostenible/informe-asi-funcionan-los-ciclos-productivos-de-las-ganaderias>
- [5] Organización de las Naciones Unidas para la Alimentación y la Agricultura. (2002) Reforma y descentralización de servicios agrícolas: Un marco de políticas. [Online]. Available: <http://www.fao.org/3/y2006s/y2006s00.htm#Contents>
- [6] J. Ashby. (1995) Institutionalising farmer participation in adaptive technology testing with the 'cial'. agricultural research and extension network, network paper 57. [Online]. Available: <http://agris.fao.org/agris-search/search.do?recordID=GB9610671>
- [7] Ministerio de Agricultura y Desarrollo Rural. (2018) El campo colombiano compite con innovación. [Online]. Available: <https://www.minagricultura.gov.co/noticias/Paginas/El-campo-colombiano-compite-con-innovaci%C3%B3n.aspx>
- [8] Instituto Geográfico Agustín Codazzi. (2019) Laboratorio nacional de suelos. [Online]. Available: <https://www.igac.gov.co/es/contenido/areas-estrategicas/agrologia/laboratorio-nacional-de-suelos>
- [9] J. Cañas et al. (2011) Modeling the lactation curves for milk, fat and protein yield of holstein cattle in antioquia, colombia. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0122-02682011000200009
- [10] V. Guimarães et al. (2006) Use of mathematical functions in the study of lactation curves in goats. [Online]. Available: https://www.scielo.br/scielo.php?pid=S1516-35982006000200028&script=sci_arttext
- [11] N. Duque et al., "Fitting lactation curves in a Colombian Holstein herd using nonlinear models," *Revista Facultad Nacional de Agronomía, Medellín*, vol. 71, pp. 8459 – 8468, 05 2018. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-28472018000208459&nrm=iso

- [12] D. Cunha *et al.* (2010) Evaluation of a dynamic simulation model for milk production systems. [Online]. Available: https://www.scielo.br/scielo.php?pid=S1516-35982010000400027&script=sci_arttext&tlng=en
- [13] ——. (2010) Selection of models of lactation curves to use in milk production simulation systems. [Online]. Available: <https://www.scielo.br/pdf/rbz/v39n4/v39n4a26.pdf>
- [14] E. Ruelle *et al.*, “Development and evaluation of the herd dynamic milk model with focus on the individual cow component,” *Animal*, vol. 10, no. 12, p. 1986–1997, 2016.
- [15] W. Ávila *et al.*, “Uso da dinâmica de sistemas como suporte à decisão em propriedades produtoras de leite: um estudo de caso,” *Revista de Economia e Agronegócio*, vol. 2, 06 2015.
- [16] M. Gavira. (2003) Simulação computacional como uma ferramenta de aquisição de conhecimento. [Online]. Available: https://www.researchgate.net/publication/236342878_Simulacao_computacional_como_uma_ferramenta_de_aquisicao_de_conhecimento
- [17] J. Cañas *et al.* (2012) Modeling and genetic parameters of lactation curves in holstein bovine breed in colombia. [Online]. Available: <http://www.scielo.org.co/pdf/mvz/v17n2/v17n2a07.pdf>
- [18] K. Jojoa *et al.* (2012) Modeling of dairy production curve in holstein cattle in the high tropic of nariño. [Online]. Available: <https://revistas.udenar.edu.co/index.php/revip/article/view/396/407>
- [19] V. Díaz. (2010) Manual de recomendaciones sobre sistemas intensivos de producción de carne. [Online]. Available: <http://www.platicar.go.cr/images/buscador/documents/pdf/05/00375-manual-produccion-de-carne.pdf>
- [20] S. Rodríguez *et al.*, *Producción ganadera sostenible. Silvopastoreo*. Corporación Colombiana de Investigación Agropecuaria-CORPOICA, 1998.
- [21] L. Mahecha *et al.* (2002) Situación actual de la ganadería de carne en colombia y alternativas para impulsar su competitividad y sostenibilidad. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=3242901>
- [22] V. Díaz. (2002) Recomendaciones sobre sistemas intensivos de producción de carne: Estabulación, semiestabulación y suplementación estratégica en pastoreo. [Online]. Available: http://www.mag.go.cr/biblioteca_virtual_animal/estabulacion.pdf
- [23] INVIMA. (2007). [Online]. Available: <https://www.invima.gov.co/normatividad-2>
- [24] SIVIGILA. Derivados lácteos. [Online]. Available: <http://www.saludcapital.gov.co/sitios/VigilanciaSaludPublica/Protocolos%20de%20Vigilancia%20en%20Salud%20Publica/Derivados%20Lacteos.pdf>

- [25] Ministerio de Salud, Republica de Colombia. (1986) Resolución número 02310 de 1986. [Online]. Available: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/Resolucion-2310-de-1986.pdf>
- [26] H. J. Motulsky and L. A. Ransnas, “Fitting curves to data using nonlinear regression: a practical and nonmathematical review,” *The FASEB Journal*, vol. 1, no. 5, pp. 365–374, 1987. [Online]. Available: <https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.1.5.3315805>
- [27] *Nonlinear curve fitting*, 2007.
- [28] H. Gavin. (2023) The levenberg-marquardt algorithm for nonlinear least squares curve fitting problems. [Online]. Available: <https://people.duke.edu/~hpgavin/lm.pdf>
- [29] A. Shelley, “Introduction to nonlinear mixed-effects models for longitudinal data,” pp. 1–6, 2019.
- [30] *Mathematical Modelling in One Dimension: An Introduction via Difference and Differential Equations*, 2013.
- [31] “The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation.” *PeerJ*, 2021.
- [32] “Aic and bic for modeling with complex survey data,” *Journal of Survey Statistics and Methodology*, 2015.
- [33] SENA. (2022) ¿Quiénes somos? [Online]. Available: <https://www.sena.edu.co/es-co/sena/Paginas/quienesSomos.aspx>
- [34] CA-SENA. (2022) Misión. [Online]. Available: <https://centroagropecuarioREGIONALCAU.blogspot.com/p/nuestro-centro.html>
- [35] A. Silvestre *et al.*, “Lactation curves for milk, fat and protein in dairy cows: A full approach,” *Livestock Science*, vol. 122, no. 2, pp. 308–313, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1871141308003090>
- [36] R. Shanks *et al.*, “Genetic aspects of lactation curves1,” *Journal of Dairy Science*, vol. 64, no. 9, pp. 1852–1860, 1981. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022030281827750>
- [37] L. Guevara Gómez and L. Tobón Llano. (2023) Repositorio del trabajo de grado de maestría titulada “Modelamiento dinámico de la producción ganadera en el Servicio Nacional de Aprendizaje SENA - Seccional Popayán, Cauca”. “Datos y algoritmos”. [Online]. Available: https://github.com/lfguevarag97/Thesis-MSC-2023-Scripts_and-data
- [38] ——. (2023) Repositorio del trabajo de grado de maestría titulada “Modelamiento dinámico de la producción ganadera en el Servicio Nacional de Aprendizaje SENA - Seccional

- Popayán, Cauca”. “Documentación”. [Online]. Available: <https://github.com/lfguevarag97/Thesis-MSC-2023>
- [39] Y. Naderi, “Appropriate mathematical models to describe the lactation curves of milk production traits of iranian holstein dairy cattle.” *Acta Scientiarum Animal Sciences*, vol. 40, 08 2018.
- [40] W. Schlesinger. (2019) How green is your milk. [Online]. Available: <https://blogs.nicholas.duke.edu/citizenscientist/how-green-is-your-milk/#:~:text=As%20a%20global%20average%2C%20dairy,atmosphere%20for%20every%20liter%20produced.>
- [41] I. Tiseo. (2023) Average greenhouse gas emissions per kilogram of major food products worldwide. [Online]. Available: <https://www.statista.com/statistics/1201677/greenhouse-gas-emissions-of-major-food-products/>
- [42] H. Aguirre *et al.* (2017) Evaluating greenhouse gas emissions from dairy manure management practices using survey data and lifecycle tools. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959652616321953>

Anexos

10.3. Modelo de optimización por programación lineal entera

Con base en lo mencionado en la Sección 10.2.3, se plantean los conjuntos, las variables, los parámetros y las funciones de objetivo de optimización mediante un ejercicio de optimización lineal que podría prestarse para maximizar ganancias o minimizar emisiones de $[KgCO_{2eq}]$. En esta ocasión se plantea un problema de optimización lineal entera multipropósito, en donde la estimación de ganancia neta que presenta el CA-SENA-POP presenta un conflicto con la reducción de gases de $[KgCO_{2eq}]$, puesto que a mayor producción de estiércol, leche y carne; mayores son las emisiones de estos Gases de Efecto Invernadero (GEI). Así pues, se plantea que:

10.3.1. Conjuntos

- $VACAS_v$, indexadas por v
- $PARTOS_p$, indexados por p
- $DURACIONES_d$, indexadas por d
- $DIAS_t$, indexados por t

10.3.2. Parámetros

10.3.2.1. Emisiones de $[KgCO_{2eq}]$

- Según la “Nicholas school of environment” de la Universidad de Duke ([40]), la producción de un (1) litro de leche representa una generación de emisiones equivalentes de CO_2 de aproximadamente $1,39[KgCO_{2eq}/Lt_{leche}]$. Éstas emisiones son denominadas como “GEIX”.
- Según “Statista.com”, una compañía de estadística comercial a nivel mundial (ver Referencia [41]), la producción de un (1) kilogramo de peso de ganado dedicado a la producción de leche, representa una generación de emisiones equivalentes de CO_2 de aproximadamente $33,3[KgCO_{2eq}/Kg_{carne}]$. Éstas emisiones son denominadas como “GEIY”.
- Según [42], la producción de un (1) kilogramo de estiércol representa una generación de emisiones equivalentes de CO_2 de aproximadamente $0,0016108[KgCO_{2eq}/Kg_{estiercol}]$. Éstas emisiones son denominadas como “GEIZ”.

10.3.2.2. Precios de venta

- Según el CA-SENA-POP (ver Referencia [34]), el precio de venta promedio de un (1) kilogramo de carne de res hembra (PVY) es de aproximadamente $7100[\$Pesos/Kg_{peso}]$.
- Según el CA-SENA-POP (ver Referencia [34]), el precio de venta de un (1) kilogramo de leche cruda sin análisis de proteína y grasas (PVX) es de aproximadamente $1350[\frac{\$Pesos}{Kg_{leche-cruda}}]$.

10.3.2.3. Producción

- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones mínimas de leche pueden representarse como un parámetro $PMINX_{vp}$ en $[Kg_{leche}/Vaca_vParto_p]$
- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones mínimas de carne pueden representarse como un parámetro $PMINY_{vp}$ en $[Kg_{carne}/Vaca_vParto_p]$
- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones mínimas de estiércol pueden representarse como un parámetro $PMINZ_{vp}$ en $[Kg_{estiercol}/Vaca_vParto_p]$
- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones máximas de leche pueden representarse como un parámetro $PMAXX_{vp}$ en $[Kg_{leche}/Vaca_vParto_p]$
- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones máximas de carne pueden representarse como un parámetro $PMAXY_{vp}$ en $[Kg_{carne}/Vaca_vParto_p]$
- Según los registros proporcionados de manera física y digital del CA-SENA-POP (ver Referencia [34]), las producciones máximas de leche pueden representarse como un parámetro $PMAXZ_{vp}$ en $[Kg_{estiercol}/Vaca_vParto_p]$

10.3.3. Variables

- $\mathcal{X}_{vpdts} \longrightarrow$ Cantidad de leche producida por una vaca v en un grupo de parto p que tiene una duración de lactancia aproximada d en un día t ; en $[Kg_{Leche}]$. ($\mathcal{X}_{vpdts} \geq 0$).
- $\mathcal{Y}_{vpdts} \longrightarrow$ Cantidad de carne producido por una vaca v en un grupo de parto p que tiene una duración de lactancia aproximada d en un día t ; en $[Kg_{Carne}]$. ($\mathcal{Y}_{vpdts} \geq 0$).
- $\mathcal{Z}_{vpdts} \longrightarrow$ Cantidad de estiércol producido por una vaca v en un grupo de parto p que tiene una duración de lactancia aproximada d en un día t ; en $[Kg_{Estiercol}]$. ($\mathcal{Z}_{vpdts} \geq 0$).

10.3.4. Funciones objetivo

10.3.4.1. Maximizar ganancias

Si el planteamiento del modelo se enfoca en maximizar las ganancias sin importar el posible impacto ecológico, la función objetivo del modelo lineal estará representada por la ecuación 10.1. Se debe mencionar que para este análisis no se tiene conocimiento de los costos operacionales del CA-SENA-POP, por lo que son considerados como nulos para este caso. En una situación donde se tenga conocimiento de estos valores económicos, basta con abstraer el costo operacional asociado a

cada lactancia monitoreada que puede ser considerado como un parámetro $COPER_{vpdt}$ (si el costo es diario durante toda la lactancia) ó $COPER_{vp}$ (si el costo es total, resultado de todo el periodo de lactancia).

$$Max(UTILIDADES) = \left(PVX \sum_{v,p,d,t} \mathcal{X}_{vpdt} \right) + \left(PVY \sum_{v,p,d,t} \mathcal{Y}_{vpdt} \right) - \sum_{v,p,d,t} COPER_{vpdt} \quad (10.1)$$

10.3.4.2. Minimizar emisiones de gases de efecto invernadero GEI.

Si nos centramos únicamente en plantear el modelo para minimizar las emisiones de GEI sin tener en consideración las posibles repercusiones financieras que estas producciones puedan representar, se tiene un modelo lineal cuya función objetivo estará representada por la siguiente ecuación:

$$Min(EMISIONES) = \left(GEIX \sum_{v,p,d,t} \mathcal{X}_{vpdt} \right) + \left(GEIY \sum_{v,p,d,t} \mathcal{Y}_{vpdt} \right) + \left(GEIZ \sum_{v,p,d,t} \mathcal{Z}_{vpdt} \right) \quad (10.2)$$

10.3.5. Restricciones

10.3.5.1. Producción mínima

- $\mathcal{X}_{vpdt} \geq PMINX_{vp} \longrightarrow \forall v, p$
- $\mathcal{Y}_{vpdt} \geq PMINY_{vp} \longrightarrow \forall v, p$
- $\mathcal{Z}_{vpdt} \geq PMINZ_{vp} \longrightarrow \forall v, p$
- $PMINX_{vp}, PMINY_{vp}, PMINZ_{vp}, \geq 0$

10.3.5.2. Producción máxima

- $\mathcal{X}_{vpdt} \leq PMAXX_{vp} \longrightarrow \forall v, p$
- $\mathcal{Y}_{vpdt} \leq PMAXY_{vp} \longrightarrow \forall v, p$
- $\mathcal{Z}_{vpdt} \leq PMAXZ_{vp} \longrightarrow \forall v, p$
- $PMAXX_{vp}, PMAXY_{vp}, PMAXZ_{vp}, \geq 0$

10.4. Tabla 10.1 - Parámetros de ajuste para cada método usado.

Tabla 10.1: Resumen de los parámetros de ajuste para cada método usado y su respectivo MSE.

PARTO $P\#_i$	N	PARS VACA	LMA				"Super Vaca"				NLFEA				NLREA				
			α	β	γ	MSE-I	α	β	γ	MSE-G	α	β	γ	MSE-I	MSE-G				
$P\#_1$	299	ACEH	9.602	0.153	0.004	3.930	8.273	0.125	0.003	6.824	9.612	0.147	0.003	3.848	10.098	0.131	0.003	3.921	
		SOL	7.405	0.119	0.003	4.706	8.295	0.079	0.002	4.682	8.295	0.079	0.002	4.682	6.833	0.131	0.003	4.750	4.485
$P\#_2$	288	LUN	7.067	0.136	0.003	4.775	6.986	0.140	0.003	4.7829	6.986	0.140	0.003	4.7829	7.375	0.131	0.003	5.036	
		DIAN	15.930	0.114	0.004	12.523	16.220	0.107	0.004	11.949	16.220	0.107	0.004	11.949	16.154	0.109	0.004	11.955	
		GINA	11.950	0.251	0.007	3.292	16.059	0.149	0.006	3.230	15.553	0.158	0.006	3.248	5.119	0.233	0.007	8.112	
		JUAN	4.103	0.317	0.008	8.128	4.384	0.282	0.007	8.074	4.384	0.282	0.007	8.074	5.119	0.233	0.007	8.112	
$P\#_3$	309	LUCI	13.394	0.167	0.007	4.967	13.801	0.147	0.006	4.985	13.801	0.147	0.006	4.985	13.127	0.160	0.006	4.987	8.448
		PAC	20.167	-0.036	0.002	9.489	19.752	-0.027	0.002	9.346	19.752	-0.027	0.002	9.346	18.765	-0.010	0.002	9.363	
		SOL	10.138	0.220	0.007	16.013	9.374	0.238	0.007	15.887	9.374	0.238	0.007	15.887	9.673	0.227	0.007	15.890	
		VIV	6.794	0.339	0.007	5.043	7.847	0.290	0.007	5.017	7.847	0.290	0.007	5.017	8.830	0.256	0.006	5.034	
$P\#_4$	309	FER	19.729	0.064	0.003	9.721	20.217	0.052	0.003	9.698	20.217	0.052	0.003	9.698	19.466	0.067	0.003	9.704	
		GPA	15.206	0.148	0.004	11.415	15.638	0.132	0.004	11.420	15.638	0.132	0.004	11.420	15.306	0.136	0.004	11.422	
		JUAN	9.445	0.203	0.004	4.692	9.485	0.202	0.004	4.712	9.485	0.202	0.004	4.712	9.907	0.190	0.004	4.713	
		LETT	15.3388	0.098	0.004	22.970	16.645	0.069	0.003	22.965	16.645	0.069	0.003	22.965	16.298	0.073	0.004	22.993	12.747
$P\#_5$	298	MART	6.613	0.392	0.006	17.059	6.516	0.396	0.006	17.039	6.516	0.396	0.006	17.039	7.084	0.371	0.006	17.076	
		MOR	19.921	0.108	0.004	11.914	20.858	0.092	0.004	11.871	20.858	0.092	0.004	11.871	20.353	0.096	0.004	11.899	
		SOL	12.511	0.058	0.003	10.689	12.452	0.059	0.003	10.670	12.452	0.059	0.003	10.670	12.282	0.064	0.003	10.678	
		CEMT	4.030	0.375	0.005	12.678	4.908	0.314	0.004	12.612	4.908	0.314	0.004	12.612	5.959	0.259	0.004	12.622	
$P\#_6$	162	DIAN	10.448	0.073	0.003	11.537	10.658	0.067	0.002	11.424	10.658	0.067	0.002	11.424	9.965	0.090	0.003	11.436	
		LETT	11.952	0.073	0.003	9.446	12.857	0.043	0.003	9.415	12.857	0.043	0.003	9.415	11.721	0.073	0.030	9.429	
		MOR	12.251	0.124	0.005	5.594	13.101	0.095	0.004	5.462	13.101	0.095	0.004	5.462	12.301	0.115	0.004	5.514	8.773
		MUC	12.259	0.182	0.007	9.440	13.138	0.155	0.006	9.398	13.138	0.155	0.006	9.398	12.708	0.164	0.006	9.425	
$P\#_7$	298	SOR	5.386	0.262	0.007	9.044	5.404	0.263	0.007	8.943	5.404	0.263	0.007	8.943	6.297	0.209	0.006	8.949	
		VIV	7.032	0.210	0.004	3.401	6.831	0.208	0.004	3.403	6.831	0.208	0.004	3.403	7.370	0.186	0.004	3.442	
$P\#_8$	162	GPA	1.619	0.688	0.009	9.353	1.531	0.705	0.009	9.270	1.531	0.705	0.009	9.270	1.670	0.679	0.009	9.278	9.440
		ANDRE	9.473	0.213	0.005	9.531	9.881	0.193	0.004	9.409	9.881	0.193	0.004	9.409	9.371	0.210	0.005	9.418	
$P\#_9$	162	GPA	14.928	0.149	0.004	4.801	14.977	0.146	0.004	4.802	14.977	0.146	0.004	4.802	15.118	0.136	0.004	4.995	4.648
		ANDRE	15.437	0.118	0.004	4.482	15.247	0.125	0.004	4.436	15.247	0.125	0.004	4.436	15.105	0.136	0.004	4.624	

10.5. Figura 10.1 - Matrices de covarianza

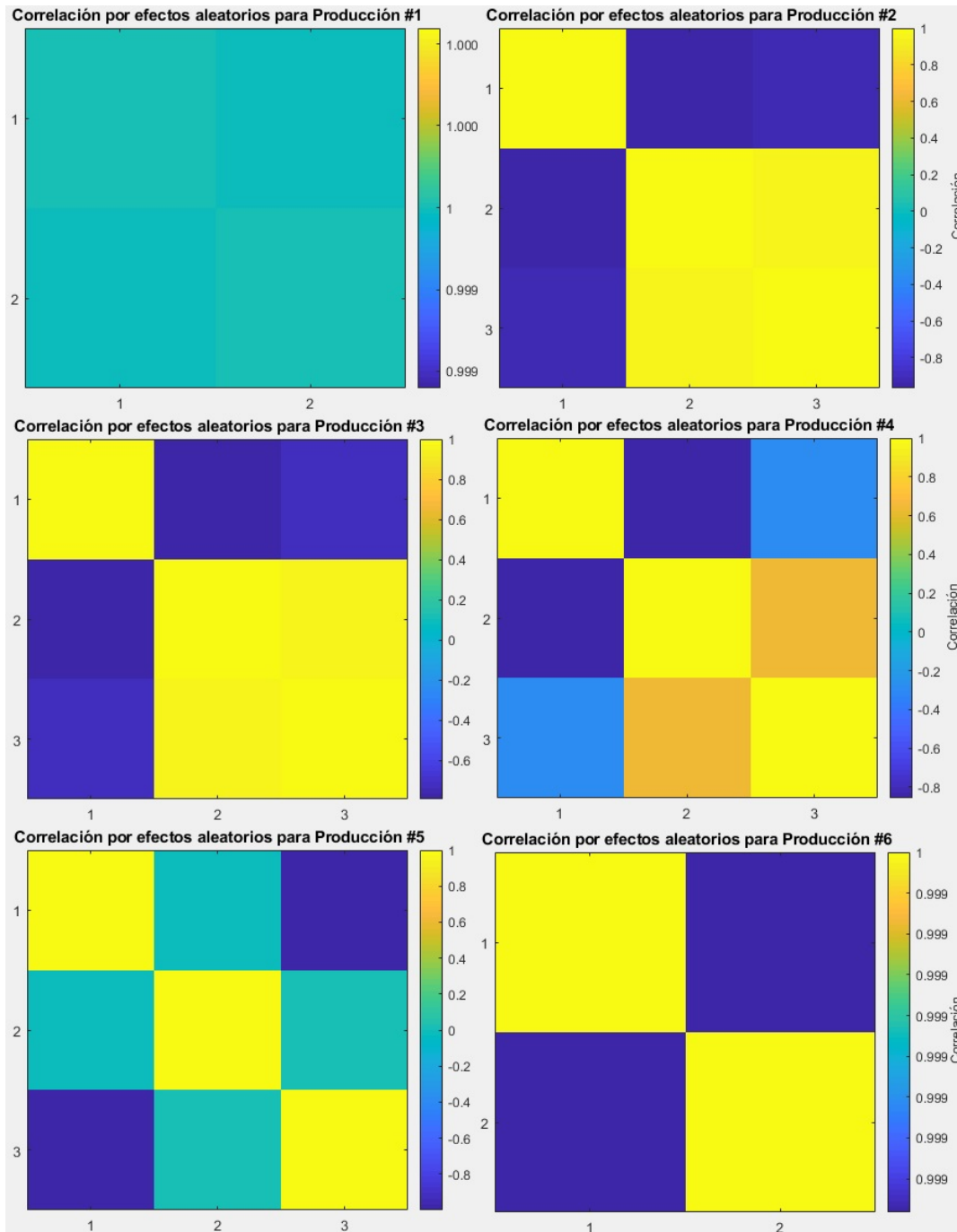


Figura 10.1: Matrices de covarianza del método “nlmeft”.

10.6. Relación matemática entre los parámetros k_1 , k_2 , k_3 y k_4

Partiendo de las Ecuaciones Diferenciales Ordinarias de primer orden 8.2 y 8.3 y considerando que estas ecuaciones están relacionadas entre sí, se puede establecer una EDO de segundo orden que permita expresar una variable en función de otra de la siguiente manera:

$$\frac{dW}{dt} = k_1\alpha_s W - k_2 P, \quad \frac{dP}{dt} = k_3\alpha_s W - k_4 P, \quad \frac{dW}{dt} = \dot{W}, \quad \frac{dP}{dt} = \dot{P}$$

$$\ddot{W} = \frac{d}{dt} \left(\frac{dW}{dt} \right) \longrightarrow \ddot{W} = k_1\alpha_s \frac{dW}{dt} - k_2 \frac{dP}{dt} \longrightarrow \ddot{W} = k_1\alpha_s \dot{W} - k_2 \dot{P}$$

Sabemos que el valor de \dot{P} , por lo que reemplazando en \ddot{W} se obtiene:

$$\dot{P} = k_3\alpha_s W - k_4 P, \longrightarrow \ddot{W} = k_1\alpha_s \dot{W} - k_2(k_3\alpha_s W - k_4 P)$$

Separando las variables hacia los lados:

$$\ddot{W} - k_1\alpha_s \dot{W} + k_2 k_3 \alpha_s W = k_2 k_4 P \quad (10.3)$$

Supóngase $k_2 k_4 P \approx 0$, despejamos \ddot{W} , hacemos cambio de variable y resolvemos para $m = \ddot{W}$

$$m = \left(\frac{-b}{2a} \right) \pm \sqrt{\left(\frac{b}{2a} \right)^2 - (c)} \quad (10.4)$$

Donde $\left[\left(\frac{b}{2a} \right)^2 - (c) \right]$ decidirá el comportamiento amortiguado del sistema. Reemplazamos valores:

$$m = \left[\left(\frac{k_1\alpha_s}{2} \right) \pm \sqrt{\left(\frac{k_1\alpha_s^2}{2} \right) - (k_2 k_3 \alpha_s)} \right] \text{ donde } \left(\frac{k_1\alpha_s}{2} \right)^2 \geq k_2 k_3 \alpha_s \quad (10.5)$$

Despejando la desigualdad, y sabiendo de antemano que $\alpha_s \neq 0$, se obtiene que:

$$\left(\frac{k_1\alpha_s}{2} \right)^2 \geq k_2 k_3 \alpha_s \longrightarrow \left(\frac{k_1^2 \alpha_s^2}{2^2} \right) \geq k_2 k_3 \alpha_s \longrightarrow k_1^2 \geq 4 \left(\frac{k_2 k_3}{\alpha_s} \right) \quad (10.6)$$

10.6.1. Observaciones

Basado en la representación por EDOs, tanto en las ecuaciones 8.2 y 8.3, así como también por la representación gráfica del modelo, es importante mencionar que, k_2 debería ser igual a $k_3\alpha$, pues parte del peso que no se gana se debe a la producción de producto P que es expulsado del cuerpo del animal, ya sea por la producción de leche L en el proceso de ordeño o por producción de estiércol en el proceso de excreción de las necesidades biológicas del animal.

Por otra parte, el factor $k_2 k_4 P$ que se supone aproximadamente igual a 0, puede afectar el comportamiento final del sistema puesto que este valor únicamente sería 0 si el bovino no tuviera pérdidas en la producción de desechos.

10.7. Representación gráfica del modelo EDO-WP

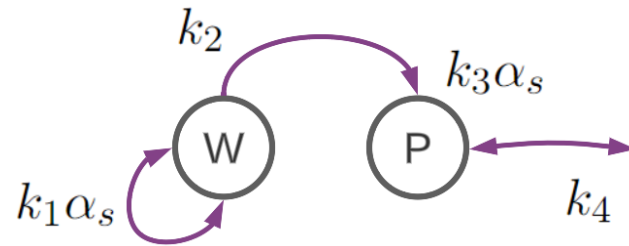


Figura 10.2: Representación gráfica de las ecuaciones 8.2 y 8.3.