

**Modelos estadísticos para el pronósticos de los parámetros fisicoquímicos
(porcentaje de acidez y densidad) y composicionales (porcentaje de grasa y proteína)
de leche de búfala recibida en Colanta planta Planeta Rica**

Presentado por:

Brenda Luz Cárdenas Sáez

Fundación Universitaria Los Libertadores

Facultad de ingeniería y ciencias básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2019

Modelos estadísticos para el pronósticos de los parámetros fisicoquímicos (porcentaje de acidez y densidad) y composicionales (porcentaje de grasa y proteína) de leche de búfala recibida en Colanta planta Planeta Rica

Presentado por
Brenda Luz Cárdenas Sáez

En cumplimiento parcial de los requerimientos para optar al título de
Especialista en Estadística Aplicada

Dirigida por
Sébastien Lozano Forero
Msc. en estadística

Fundación Universitaria Los Libertadores
Facultad de ingeniería y ciencias básicas
Especialización en Estadística Aplicada
Bogotá D.C, Colombia
2019

Notas de aceptación

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá D.C, marzo de 2019.

Las directivas de la Fundación Universitaria Los
Libertadores, los jurados calificadores y el cuerpo docente
no son responsables por los criterios e ideas expuestas en el
presente documento. Estos corresponden únicamente a los
autores y a los resultados de su trabajo.

Dedicatoria

Dedico a mi abuela que amo con todo el corazón y que desde el sitio en donde se encuentra está dichosa de que alcance un logro más en vida y a mi madre que es mi vida.

Agradecimientos

Agradezco a Dios que me ilumina y guía mis pasos por los senderos de la vida, a mi madre que ha sido y será mi modelo a seguir y a la cual le debo todo lo que soy y cuanto tengo, a mis hermanos por estar impulsándome cada minuto a no dejarme vencer y culminar a mi novio por su paciencia, confianza y sus palabras de aliento para seguir adelante. Cada día fue duro pero sin ellos no hubiera llegado al finalizar este proyecto.

Agradezco a la Cooperativa Colanta que me ha dado parte de las bases del conocimiento y la experiencia que tengo en este campo.

Contenido

1. Introducción	3
2. Planteamiento del Problema	6
2.1. Formulación del Problema.....	6
2.2. Objetivos.....	6
2.2.1. Objetivo General.	6
2.2.2. Objetivos específicos.....	6
2.3. Justificación	7
3. Marco teórico/ Conceptual.....	9
3.1. Leche	9
3.2. Características fisicoquímicas y composicionales de la leche de Búfala	9
3.3. Las series temporales.....	11
3.3.1. Naturaleza de las series de tiempo.	11
3.3.2. Componentes de una serie de tiempo.....	12
3.3.3. Tipos de modelos de series de tiempos.	13
3.3.4. Funciones de Autocorrelación, Auto covarianza y de Autocorrelación Parcial.	14
3.3.5. Correlograma.....	16
3.3.6. Modelos de Series de Tiempo Estacionarias.....	16
3.3.7. Modelos de Series de Tiempo No Estacionarias.	19
3.3.8. Metodología de Box – Jenkins.....	23

4. Marco Metodológico.....	33
4.1. Unidad de análisis.....	33
4.2. Instrumentos-materiales.....	33
4.3. Procedimiento.....	33
4.3.1. Análisis descriptivo.....	33
4.3.2. Construcción de modelos.....	34
5. Análisis y resultados	35
5.1. Análisis descriptivo	35
5.2. Construcción de modelos.....	39
5.2.1. Porcentaje de Acidez.....	39
5.2.2. Densidad (g/ml).....	44
5.2.3. Porcentaje de grasa m/v.....	48
5.2.4. Porcentaje de proteína m/m.....	53
6. Conclusiones y recomendaciones	59
7. Bibliografía	61

Índice de figuras

<i>Figura 1</i> Análisis Descriptivo Porcentaje de Acidez leche de Búfala Recibida en Colanta Planeta Rica.	35
<i>Figura 2</i> Análisis Descriptivo Densidad de la leche de Búfala Recibida en Colanta Planeta Rica.	36
<i>Figura 3</i> Análisis Descriptivo Porcentaje de Grasa leche de Búfala Recibida en Colanta Planeta Rica.	37
<i>Figura 4</i> Análisis Descriptivo Porcentaje de Proteína leche de Búfala Recibida en Colanta Planeta Rica.	38
<i>Figura 5</i> Base de Datos porcentaje de acidez de leche de Búfala.....	39
<i>Figura 6</i> Análisis de serie diferenciada Porcentaje de Acidez de Leche de Búfala.....	40
<i>Figura 7</i> Evaluación de supuestos del modelo ARIMA (1, 0, 0) x (0, 1, 1) ₁₂	43
<i>Figura 8</i> Pronóstico fuera de muestra del modelo ARIMA (1, 0,0) x (0, 1, 1) ₁₂ para porcentaje de acidez de la leche de búfala.....	44
<i>Figura 9</i> Serie de tiempo Densidad de la leche de Búfala.....	45
<i>Figura 10</i> Evaluación de supuestos del modelo ARIMA (0, 1,0) x (0, 1, 1) ₁₂	47
<i>Figura 11</i> Pronóstico fuera de muestra del modelo ARIMA (0, 1,0) x (0, 1, 1) ₁₂ para la Densidad de la leche de búfala.....	48
<i>Figura 12</i> Serie de tiempo Grasa (%m/v) de la leche de Búfala.....	49
<i>Figura 13</i> Evaluación de supuestos del modelo ARIMA (2, 1, 1) x (1, 0, 1) ₁₂	51
<i>Figura 14</i> Pronóstico fuera de muestra del modelo ARIMA (2, 1,1) x (1, 0, 1) ₁₂ para la porcentaje grasa de la leche de búfala.....	52

<i>Figura 15</i> Base de datos porcentaje de Proteína en leche de Búfala	54
<i>Figura 16</i> Análisis de serie diferenciada Porcentaje de Proteína de Leche de Búfala.....	55
Figura 17 Evaluación de supuestos del modelo ARIMA (3, 0,2) x (0, 1, 1) ₁₂	57
<i>Figura 18</i> Pronóstico fuera de muestra del modelo ARIMA (3, 0,2) x (1, 0, 1) ₁₂ para la porcentaje de Proteína de la leche de búfala.....	58

Índice de figuras

Tabla 1	Parametrización de la leche de Búfala	10
Tabla 2	Valores de lambda y sus transformaciones	21
Tabla 3	Descripción de ACF Y PACF según el modelo de serie estacional	25
Tabla 4	Descripción de ACF Y PACF según el modelo de serie no estacional	25
Tabla 5	Contrastes de raíz unitaria y estacionariedad	26
Tabla 6	Estimativos para el modelo ARIMA (1, 0,0) (0, 1,1) ₁₂	41
Tabla 7	Estimativos para el modelo ARIMA (0, 1,0) (0, 1,1) ₁₂	46
Tabla 8	Estimativos para el modelo ARIMA (2, 1,1) (1, 0,1) ₁₂	50
Tabla 9	Estimativos para el modelo ARIMA (3, 0, 2) (0, 1,1) ₁₂	55

Modelos estadísticos para el pronósticos de los parámetros fisicoquímicos (porcentaje de acidez y densidad) y composicionales (porcentaje de grasa y proteína) de leche de búfala recibida en Colanta planta Planeta Rica.

Brenda Luz Cárdenas Sáez¹

Resumen

El Decreto 616 de 2006, hace referencia a los requisitos que debe cumplir la leche para el consumo humano en Colombia, detallando las características de la leche cruda de ganado bovino excluyendo la leche de búfala, esto origina que las industrias no tengan bases normativas para su aceptación/rechazo, a esto se suma que las características fisicoquímicas, composicionales y organolépticas de esta leche son diferentes a la leche de vaca. Se considera que a través de la estadística de series de tiempo se puede ajustar un modelo a cada parámetro estudiado de la leche de búfala logrando pronosticarlos. Para el ajuste de los modelos se utiliza la base de datos de recepción de leche de búfala en la planta de Colanta en Planeta Rica desde el año 2008 hasta diciembre 2018, con ayuda del Software RStudio y utilizando la metodología planteada por Box y Jenkins se logra ajustar el modelo ARIMA (1, 0,0) x (0, 1, 1)₁₂ para el parámetro de acidez, el modelo ARIMA (0, 1,0) x (0, 1, 1)₁₂ para la densidad, ARIMA (1, 0,0) x (0, 1, 1)₁₂ para pronosticar el porcentaje de grasa y el porcentaje de proteína se logra modelar con el ajuste ARIMA (3, 0,2) x (1, 0, 1)₁₂. Los resultados de los pronósticos fuera de muestra arrojan resultados conformes con errores medio absoluto porcentual (MAPE) de 1,75% para el porcentaje de acidez, 0,03417 % para la densidad, 0,85% para el porcentaje de grasa y 2,75% para la proteína de la leche de Búfala, los estadísticos Theil's U de los 4 parámetros estudiados arrojaron menor a 1. Los modelos ajustados resultan acertados para los pronósticos de los parámetros estudiados.

Palabras claves: Leche de Búfala, Modelos Arima, Porcentaje de acidez, densidad, porcentaje de grasa y porcentaje de proteína.

¹ Ingeniera de alimentos, Universidad de Córdoba, brendacardenas972@gmail.com

Modelos estadísticos para el pronósticos de los parámetros fisicoquímicos (porcentaje de acidez y densidad) y composicionales (porcentaje de grasa y proteína) de leche de búfala recibida en Colanta planta Planeta Rica.

Brenda Luz Cárdenas Sáez

Abstract

Decree 616 of 2006, refers to the requirements that milk must meet for human consumption in Colombia, detailing the characteristics of raw milk from cattle, excluding buffalo milk, which leads to industries having no normative basis for the acceptance / rejection of this raw material, it is added that the physicochemical, compositional and organoleptic characteristics of this milk are different from cow's milk. It is considered that through the statistics of time series a model can be adjusted to each parameter studied of buffalo milk, making it possible to forecast them correctly. For the adjustment of the models, the buffalo milk reception database is used in the Colanta plant in Planeta Rica from 2008 to December 2018, with the help of the RStudio Software and using the methodology proposed by Box and Jenkins. adjust the ARIMA model $(1, 0,0) \times (0, 1, 1)_{12}$ for the acidity parameter, the ARIMA model $(0, 1,0) \times (0, 1, 1)_{12}$ for the density, ARIMA $(1, 0,0) \times (0, 1, 1)_{12}$ to predict the percentage of fat and the percentage of protein is achieved by modeling with the adjustment ARIMA $(3, 0,2) \times (1, 0, 1)_{12}$. The results of the out-of-sample forecasts show concordant results with percentage absolute percentage errors (ASM) of 1.75% for the percentage of acidity, 0.03417% for the density, 0.85% for the percentage of fat and 2, 75% for the buffalo milk protein, the Theil's U statistic of the 4 parameters studied yielded less than 1. The adjusted models are accurate for the predictions of the parameters studied.

Key words: Buffalo milk, Arima models, percentage of acidity, density, percentage of fat and percentage of protein.

1. Introducción

El búfalo Tuvo su origen en Asia y se registran datos de su existencia como animal doméstico desde hace 4.500 años, en la región ocupada actualmente por India y Pakistán. De Asia fue llevado a África, luego a Europa, Oceanía y posteriormente a América en 1880, pero fue hasta 1990 cuando se diseminó por varios países de América del Sur. Brasil primero, luego Trinidad y Tobago y finalmente Venezuela y Argentina. En la actualidad el búfalo se encuentra en todos los países del continente, Canadá y Chile fueron los últimos en incorporar a esta especie. Los principales países en cantidad de cabezas son Brasil, Venezuela, Colombia y Argentina (Exequiel Maria. Patiño, 2011). Colombia cuenta con una población bufalina de 308.580 animales distribuidos en 4.007 predios. Esta población de búfalos se concentra principalmente en los departamentos de Córdoba (26,03 %), Antioquia (18,36 %), Santander (12,97 %), Magdalena (6,69 %) y Bolívar (6,56%) que agrupan el 70,61 % de la población nacional, la población bufalina muestra un incremento del 24% anual (ICA, 2017), lo que hace que aumente el aprovechamiento de la leche generada por este mamífero en las industrias lácteas.

La composición físico química de la leche de búfala ha sido estudiada principalmente en países como India, Italia, Bulgaria, Brasil y Venezuela, y se ha determinado que presenta una gran variabilidad incluso dentro de la misma raza y país (E.M Patiño & Guanzirolí Stefani, 2005). La característica física de la leche de búfalas que mayor variabilidad presenta es la acidez titulable, lo cual se debe a que las diferentes etapas de lactación la afectan, incrementándola a medida que avanza el ciclo productivo. Se ha comprobado que la leche bufalina excede hasta 39.9% de sólidos totales, 95.9% de grasa, 25.6% de proteína y 1.7% de lactosa en comparación a la leche bovina, entre los componentes composicionales estudiados, los que mayor variabilidad

presentan son la grasa y los sólidos totales (Exequiel Maria. Patiño, 2011). No existe normatividad Colombiana que estipule la parametrización de esta leche por lo cual su estudio y caracterización fisicoquímica y composicional es de gran interés para la industria láctea.

La gran variabilidad que presentan los parámetros de la leche de búfala y su diferencia con los de bovino son factores que respaldan establecer modelos que representen cada uno de los parámetros de esta leche y que permita predecir con una aceptable confiabilidad el comportamiento de esta materia prima, que de acuerdo a Patiño y Guanzirolí (2009) la raza, la lactancia, el número de partos, la alimentación y las condiciones ambientales son factores externos que la afecta e influyen en la inestabilidad de sus parámetros.

La estadística ha desarrollado teorías y métodos que se han convertido en herramientas imprescindibles para la estimación, comparación y la predicción. Entre estos métodos estadísticos se encuentran la regresión, los métodos multivariados y los aplicados a las series temporales entre otros. Las series temporales se aplican, fundamentalmente, los métodos de pronósticos y, su mayor dificultad radica generalmente en la escasez de datos para analizar el problema y la falta de comunicación entre los estadísticos y los usuarios que emplean los resultados de estos método (Suárez, 2013).

Uno de los sectores que emplea los métodos estadísticos de las series temporales es la rama agropecuaria, en la cual se realizan pronósticos de condiciones agroclimáticas y de la situación comercial de las empresas, teniendo en cuenta a su vez el análisis de mercado. En el sector económico y financiero, los datos de las series temporales pueden ser muy variados y usualmente son usados para evaluar el comportamiento de las ventas de una empresa, o para evaluar el comportamiento de los índices de precio de un país o de un tipo de producto, pero en general pueden utilizarse en cualquier área.

Se considera que a través del análisis de series de tiempo se logrará ajustar un modelo que pronostique el comportamiento de las variables: porcentaje de acidez, densidad, porcentaje de grasa y proteína de la leche de búfala y así lograr tener un soporte en la toma de decisiones dentro de la planta en vista del vacío normativo y como soporte a posibles modificaciones del decreto 616 del 2006 para discriminar la leche de búfala.

2. Planteamiento del Problema

La Cooperativa Colanta requiere predecir las variables fisicoquímicas y composicionales de la leche de búfala como mecanismo para la toma de decisiones de aceptación/rechazo , en vista del vacío normativo con respecto a los parámetros de esta leche y sus variantes con respecto a la leche de ganado bovino en diferentes épocas del año.

2.1. Formulación del Problema

Cuáles son los modelos estadísticos bajo enfoque Box Jenkins más apropiados para predecir las variables fisicoquímicas: Porcentaje de acidez, densidad y las variables composicionales: Porcentaje de grasa y proteína de la leche de Búfala recibida en la cooperativa Colanta Planta Planeta Rica.

2.2. Objetivos

2.2.1. Objetivo General.

Realizar pronósticos de los parámetros fisicoquímicos (porcentaje de acidez, y densidad) y composicionales (porcentaje de grasa y proteína) de leche de Búfala recibida en Colanta Planta Planeta Rica

2.2.2. Objetivos específicos.

- Realizar análisis descriptivo de las variables estudiadas (porcentaje de acidez, densidad, porcentaje de grasa y proteína) en la leche de búfala recibida en Planta Planeta rica en el periodo comprendido entre 2008 y 2018.
- Seleccionar y validar modelos ARIMA adecuados para el pronóstico de cada una de las variables estudiadas en la leche de búfala.
- Pronosticar el comportamiento de la leche de búfala para el segundo periodo del 2017, 2018 y 2019.

2.3. Justificación

La producción de leche de búfala en 2017 fue de 174,3 millones de litros, es decir 429.500 litros diarios. Para el año 2018 se proyectó de 193,3 millones de litros. Dentro del total de la producción de leche a nivel nacional, que el año 2017 fue de 6,838 millones de litros, la participación de la búfala es de 2,55%. Si bien es poco, el crecimiento de este producto ha ido en aumento tanto a nivel nacional como internacional. Para lograr esta cifra fueron necesarias 138.000 búfalas que dieron 3,5 litros diarios (Gonzalez Bell, 2018)

El Ministerio de Protección Social, mediante el Decreto 616 de 2006, hace referencia a los requisitos que debe cumplir la leche para el consumo humano en Colombia; el artículo 16 de este decreto hace referencia a las características de la leche cruda de ganado bovino, quedando por fuera la leche de búfala lo que hace que las industrias no tengan bases normativas para la aceptación/rechazo esta leche.

La Cooperativa Colanta capta aproximadamente 20.000 litros de leche de búfala al día en los departamentos de Córdoba, Antioquia, Magdalena y Sucre, equivalentes aproximadamente a 4.000 cabezas de ganado, considerando una producción promedio por búfala de 5 litros diarios. Los resultados de los análisis fisicoquímicos y composicionales de esta leche varían con el tiempo y se desconoce en qué épocas o periodos estos resultados pueden ser altos o bajos lo que representa un dilema en la toma de decisiones con respecto al destino de esta materia prima, corriéndose el riesgo de aceptar leche que debe ser rechazada y viceversa.

El análisis estadístico de series temporales o cronológicas, como también se les conoce, se usa hoy día con profusión en muchas áreas de la ciencia como son: la física, ingeniería, la demografía, el marketing, las telecomunicaciones, la meteorología, la química, las ciencias médicas y fundamentalmente en economía(Suárez, 2013) , en el campo de los alimentos se han

encontrado diferentes estudios relacionados con la producción el pronóstico de producción de leche, carne, granos, demanda de almacenamiento de productos perecederos, tendencia de los grados días cafeto y gados día broca entre otros, sin embargo no se logra encontrar predicciones de parámetros fisicoquímicos y/o composicionales de algún producto alimenticio.

Se considera que un modelo de series de tiempo bajo enfoque Box Jenkins se puede ajustar a cada uno de los parámetros planteados de la leche de búfala logrando pronosticarlos de manera acertada, con este proyecto se logrará tener bases sólidas en planta para toma de decisiones en cuanto a la aceptación y rechazo de esta leche y se tendrán soportes a posibles modificaciones del decreto 616 del 2006 para discriminar la leche de búfala adicional será la puerta para futuras investigaciones frente a este campo.

3. Marco teórico/ Conceptual

3.1. Leche

Es el producto de la secreción mamaria normal de animales bovinos, bufalinos y caprinos lecheros sanos, obtenida mediante uno o más ordeños completos, sin ningún tipo de adición, destinada al consumo en forma de leche líquida o a elaboración posterior (Ministerio de la protección social, 2006).

3.2. Características fisicoquímicas y composicionales de la leche de Búfala

La composición fisicoquímica de la leche de búfala ha sido estudiada principalmente en países como India, Italia, Bulgaria, Brasil y Venezuela, y se ha determinado que presenta una gran variabilidad incluso dentro de la misma raza y país. La leche de búfala exhibe diferencias con respecto a la leche de vaca, estas han sido demostradas por estudios realizados en diversos países. Se ha comprobado que la leche bufalina excede hasta 39,9% de sólidos totales, 95,9% de grasa, 25,6% de proteína y 1,7% de lactosa en comparación a la leche bovina. Entre los componentes composicionales estudiados, los que mayor variabilidad presentan son la grasa y los sólidos totales (Exequiel Maria. Patiño, 2011).

Teniendo en cuenta el vacío normativo existente con la parametrización de esta leche, en la Cooperativa Colanta Planta Planeta Rica se caracteriza internamente la leche búfala obteniéndose los resultados expuestos en la tabla 1, con los cuales actualmente se acepta/rechaza esta materia prima.

Tabla 1 *Parametrización de la leche de Búfala*

Análisis	Parámetro	
	mín.	máx.
Acidez (% A.L)	0,11	0,18
pH	6,71	7,17
Punto Crioscópico (°H)	-0,568	-0,522
Densidad (g/ml)	1,032	1,038
Grasa (%)	4,83	10,20
Proteína (%)	3,57	6,36
Lactosa (%)	4,83	5,92
Sólidos Totales (%)	14,34	20,96

Fuente: Cárdenas (2017)

Valdés (2011) refiere que la concentración total de colesterol es menor a la encontrada en la leche de vaca (275 mg/100g vs 330 mg/100g de grasa), que posee 25.5% más de aminoácidos esenciales, con excepción de cistina y triptófano, mayores contenidos de calcio (1.88 vs 1.30%) y hierro (61 vs 37 ppm) y de ácidos grasos mono-insaturados (1.7 vs 1.1g) y polinsaturados (0.2 vs 0.1g).

Entre las características que más destacan de la leche de búfala es su coloración blanca opaca provocada por la ausencia de pigmentos carotenoides. La ausencia de estos pigmentos proporciona una manteca blanca, cristalina y más consistente que la obtenida con leche de vaca. Los glóbulos grasos de la leche de búfala son (4,1-4,8 micras) que los de la leche de vaca (3,6-

4,0 micras). La grasa bubalina tiene mayor densidad y temperatura de fusión más elevada (32,0-43,5°C) que la bovina pero el índice de yodo (29,1) en la bufara es inferior en la vaca.

La caseína en la leche bubalina existe principalmente en forma micelar y el examen con microscopía electrónica ha demostrado que estas micelas tienen mayor tamaño que las de la leche bovina, son más opacas, contienen menos nitrógeno, menos ácido sálico, pero más calcio y fósforo. Como resultado del elevado contenido de calcio en la leche de búfalo, su estabilidad a temperatura ambiente elevada es inferior, la tensión de la cuajada es mayor y la coagulación del cuajo más lenta en comparación con la leche de vaca.

La característica física de la leche de búfalas que mayor variabilidad presenta es la acidez titulable, lo cual se debe a que las diferentes etapas de lactación la afectan, incrementándola a medida que avanza el ciclo productivo. Es importante destacar, que la elevada acidez titulable que posee la leche bufalina en comparación con la bovina se debe a que la primera posee mayor cantidad de caseína (Exequiel Maria. Patiño, 2011).

3.3. Las series temporales

3.3.1. Naturaleza de las series de tiempo.

Una Serie temporal es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o varáis características (serie multivariante o vectorial) de una unidad observable en diferentes momentos. Se representa por medio de una gráfica de líneas sobre cuyo eje horizontal se representan los períodos y en cuyo eje vertical se representan los valores de la serie de tiempo.

El primer objetivo del análisis econométrico de una serie temporal consiste en elaborar un modelo estadístico que describa adecuadamente la procedencia de dicha serie, de manera que las

implicaciones teóricas del modelo resulten compatibles con las pautas muestrales observadas en la serie temporal. Después, el modelo elaborado a partir de la serie temporal considerada puede utilizarse para:

- Describir la evolución observada de dicha serie, así como las relaciones contemporáneas y dinámicas entre sus componentes (en el caso de series multivariantes).
- Prever la evolución futura de dicha serie.
- Contrastar (presentar evidencia empírica en favor o en contra de) alguna teoría sobre las características o variables a las que se refieren los componentes de dicha serie (Mauricio, 2008).

3.3.2. Componentes de una serie de tiempo.

Tendencia (T): constituye aquella parte de la serie cronológica, que se caracteriza por un movimiento suave, monótono y lento, durante un período bastante extenso de tiempo y esta puede ser creciente, decreciente o constante y se describe por una función dependiente del tiempo $T = f(t)$ por lo general, se describe por medio de una recta o de un tramo monótono de alguna curva que se ajuste al comportamiento de los datos.

Ciclo (C): caracterizado por oscilación alrededor de la tendencia con una larga duración y sus factores no son claros. Por ejemplo, que tienen ciclos que duran varios años.

Estacionalidad (S): resulta de las fluctuaciones periódicas que se presentan regularmente en la serie durante sub-intervalos de un período dado, estas pueden reflejar comportamientos estacionales que se repiten de la misma manera y con la misma regularidad año tras año.

Los movimientos aleatorios, residuales o irregulares: refleja aquellos movimientos esporádicos que ocurren en la serie y que no presentan regularidad alguna, ni en intensidad, ni periodicidad. Estos movimientos son motivados por acontecimientos fortuitos, son movimientos

erráticos que no siguen un patrón específico, que obedecen a causas diversas e impredecibles (Suárez, 2013).

3.3.2.1. Observaciones Discrepantes.

Corresponden a un comportamiento anormal del fenómeno o a un error de medición. Si una serie de tiempo contiene uno o más valores atípicos, el correlograma puede verse seriamente afectado. También se las conoce como outliers. (Eraso, 2013).

3.3.3. Tipos de modelos de series de tiempos.

3.3.3.1. Modelos deterministas.

Se trata de métodos de extrapolación sencillos en los que no se hace referencia a las fuentes o naturaleza de la aleatoriedad subyacente en la serie. Su simplicidad relativa generalmente acompañada de menor precisión. Ejemplo de modelos deterministas son los modelos de promedio móvil en los que se calcula el pronóstico de la variable a partir de un promedio de los “n” valores inmediatamente anteriores (Rios, 2008).

3.3.3.2. Modelos estocásticos.

Un proceso estocástico es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente, referidas a una (proceso univariante o escalar) o a varias (proceso multivariante o vectorial) características de una unidad observable en diferentes momentos.

3.3.3.2.1. Serie estacionaria.

Un proceso estocástico (Y_t) es estacionario cuando las propiedades estadísticas de cualquier secuencia finita $Y_{t1}, Y_{t2}, \dots, Y_{tn}$ ($n \geq 1$) de componentes de (Y_t) son semejantes a las de la secuencia $Y_{t1+h}, Y_{t2+h}, \dots, Y_{tn+h}$ para cualquier número entero $h = \pm 1, \pm 2, \dots$

3.3.3.2.2. *Serie No estacionaria.*

un proceso estocástico (Y_t) es no estacionario cuando las propiedades estadísticas de cualquier secuencia finita $Y_{t1}, Y_{t2}, \dots, Y_{tn}$ ($n \geq 1$) de componentes de (Y_t) son diferentes a las de la secuencia $Y_{t1+h}, Y_{t2+h}, \dots, Y_{tn+h}$ para al menos un número entero $h > 0$ (Mauricio, 2008).

3.3.4. **Funciones de Autocorrelación, Auto covarianza y de Autocorrelación Parcial.**

Dado un proceso estacionario Z_t con $E(Z_t) = \mu$ y $V(Z_t) = \sigma^2$ constante, se definen:

3.3.4.1. *Función de autocovarianza.*

$$\gamma_\kappa = \text{Cov}(Z_t, Z_{t+\kappa}) = E(Z_t - \mu)(Z_{t+\kappa} - \mu) \quad (1)$$

3.3.4.2. *Función de autocorrelación (ACF).*

$$\rho_\kappa = \frac{\text{Cov}(Z_t, Z_{t+\kappa})}{\sqrt{V(Z_t)}\sqrt{V(Z_{t+\kappa})}} = \frac{\gamma_\kappa}{\gamma_0} \quad (2)$$

Para todo proceso estacionario se satisfacen las siguientes propiedades:

- Para todo t , $\gamma_0 = V(Z_t)$.
- Del hecho que $|\rho_\kappa| \leq 1$ tenemos que $|\gamma_\kappa| \leq \gamma_0$ para todo κ .
- γ_κ y ρ_κ son semidefinidas positivas. Es decir valen las siguientes igualdades:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma |t_i - t_j| \geq 0 \quad (3)$$

y

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \rho |t_i - t_j| \geq 0 \quad (4)$$

Para todos los t_1, \dots, t_n y a_1, \dots, a_n .

3.3.4.3. *Función Autocorrelación parcial PACF.*

Se definirá la función de autocorrelación parcial como la correlación condicional

$$P_k = \text{Corr}(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1}) \quad (5)$$

Se considera ahora la regresión con variable dependiente Z_{t+k} y con covariables Z_{t+k-1}, \dots, Z_t . Es decir:

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \dots + \phi_{kk} Z_t + e_{t+k} \quad (6)$$

Con ϕ_{ki} los parámetros de la regresión y e_{t+k} errores de la media 0 y no correlacionados con $Z_{t+k-j} \forall j$. Supongamos sin pérdida de generalidad que $E(Z_t) = 0$, luego multiplicando a ambos lados de la regresión por Z_{t+k-j} y tomando esperanza obtenemos:

$$\begin{aligned} \gamma_j &= \phi_{k1} \gamma_{j-1} + \dots + \phi_{kk} \gamma_{j-k} \\ \rho_j &= \phi_{k1} \rho_{j-1} + \dots + \phi_{kk} \rho_{j-k} \end{aligned} \quad (7)$$

Utilizando la regla de Cramer, se puede despejar los coeficientes ϕ_{ki} en particular se obtiene fácilmente que $\phi_{kk} = P_k$, la autocorrelación parcial entre Z_t y Z_{t+k} . Por otro lado se puede

determinar la autocorrelación parcial entre Z_t y Z_{t+k} se puede obtener como los coeficientes de la regresión asociada a Z_t de κ pasos (Andrés Pablo Moggi, 2013).

3.3.5. Correlograma.

La representación gráfica de la función de autocorrelación se denomina correlograma. Sirve para confirmar la presencia de estacionalidad y su periodo. Para elaborarlo se estiman las siguientes características: media, autocorrelación y autocovarianza.

3.3.6. Modelos de Series de Tiempo Estacionarias.

3.3.6.1. *Proceso autorregresivos (AR (p)).*

Definimos un modelo como autorregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

Los modelos autorregresivos se abrevian con la palabra **AR** tras la que se indica el **orden** del modelo: AR (1), AR (2),...etc. El orden del modelo expresa el número de observaciones retrasadas de las series temporales analizadas que intervienen en la ecuación. Así, por ejemplo, un modelo AR (1) tendría la siguiente expresión: $Y_t = \phi_0 + \phi_1 Y_{t-1} + a_t$

La expresión genérica de un modelo autorregresivo, no ya de un AR(1) sino de un AR(p) sería la siguiente:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad (8)$$

Pudiéndose escribir de forma abreviada como:

$$\phi_p(L)Y_t = \phi_0 + a_t \quad (9)$$

Donde $\phi_p(L)Y$ es lo que se conoce como operador polinomial de retardos:

$$\phi_p(L) = 1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p \quad (10)$$

Y donde, a su vez, el término L es lo que se conoce como operador retardo tal que, aplicado al valor de una variable en t , dé como resultado el valor de esa misma variable en $t-1$:

$$LY_t = Y_{t-1} \quad (11)$$

Y aplicado sucesivamente p veces retarda el valor en p períodos

$$L^pY_t = Y_{t-p} \quad (12)$$

Un proceso $AR(p)$ tendrá los primeros p coeficientes de autocorrelación parcial distintos de cero, lo cual determina el orden del proceso.

3.3.6.2. *Modelo de medias móviles (MA (q)).*

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas **MA**, seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA (q) respondería a la siguiente expresión:

$$Y_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_p a_{t-p} \quad (13)$$

Puede abreviarse utilizando el polinomio de retardos (como en caso de los modelos AR):

$$Y_t = \theta_p(l)a_t + \mu \quad (14)$$

Al igual que en el caso de los modelos autorregresivos, el orden de los modelos de medias móviles suele ser bajo MA(1), MA(2) o corresponderse con la periodicidad de los datos analizados MA(4), para series trimestrales, o MA(12) para series mensuales.(Arce & Mahía, 2010).

3.3.6.3. *Modelo Autorregresivo de Medias Móviles (ARMA).*

Los procesos ARMA resultan de añadir estructura MA a un proceso AR o viceversa. El proceso ARMA (p , q) se escribe:

$$(1 - \phi_1 B - \dots - \phi_p B^p) \tilde{z}_t = (1 - \theta_1 B - \dots - \theta_p B^p) a_t \quad (15)$$

ó en notación compacta:

$$\phi_p(B) \tilde{z}_t = \theta_p(B) a_t \quad (16)$$

Cuando $q = 0$, el modelo se llama un modelo autorregresivo de orden p , $AR(p)$, y cuando $p = 0$, el modelo se llama un modelo de media móvil de orden q , $MA(q)$. Para un proceso ARMA (p, q) estacionario la función de autocorrelación tiene un decrecimiento exponencial μ oscilatorio después de un retardo q , en cuanto que la función de autocorrelación parcial tiene el mismo comportamiento a partir de p (Eraso, 2013).

3.3.7. Modelos de Series de Tiempo No Estacionarias.

Los modelos de series de tiempo anteriores, suponen la existencia de la propiedad de estacionariedad; sin embargo, muchas de las series no cumplen con esta característica presentando una tendencia creciente a lo largo del tiempo o una clara inclinación a permanecer durante largos periodos por encima o por debajo de su media. La no estacionariedad en las series de tiempo se puede deber a la presencia de una tendencia determinística, al problema de las raíces unitarias o a la presencia de una varianza heterocedástica. Dependiendo del caso en que se encuentre, la no estacionariedad tendría determinadas implicancias teóricas y estadísticas, así como determinados procedimientos para transformar las series en estacionarias.

3.3.7.1. *No estacionariedad en media.*

Las series pueden presentar una tendencia a través del tiempo. Estas series con tendencia, claramente no son estacionarias, porque su valor medio cambia con del tiempo. Esta tendencia puede ser determinística. Es decir, la media del proceso no estacionario puede ser determinada mediante una función que depende del tiempo. Por ejemplo, la función μ_t puede seguir una tendencia lineal en cuyo caso se podría modelar a partir del siguiente modelo $\mu_t = \alpha_0 + \alpha_1 t$ y luego utilizar un modelo para la serie temporal del tipo:

$z_t = \alpha_0 + \alpha_1 t + \alpha_t$ Donde α_t es ruido blanco. En general se puede proponer un polinomio de grado k para ajustar la media del proceso o algún tipo de tendencia representada (Andres Pablo Mogni, 2013).

Otro tipo de tendencia en media es la tendencia estocástica. Aunque una serie sea estacionaria, algunas partes de esta se comportan de igual manera salvo una diferencia a nivel local esta tendencia se puede eliminar de la serie mediante la aplicación de diferencias no estacionales, la diferencia no estacional es la serie de las diferencias $X_t - X_{t-1}$.

Para eliminar la estacionalidad se utilizan las diferencias estacionales. La diferencia estacional es la serie de las diferencias $X_t - X_{t-k}$ donde k es el período estacional; esta diferencia elimina o disminuye el movimiento estacional y la serie obtenida es un año más corta que la serie original.

El problema que se presenta al analizar series temporales con un posible componente estacional o de tendencia es que este sea tan fuerte que implique una variación de la media (global) en el tiempo. En este caso se requerirán diferencias estacionales (D) o no estacionales (d) para que la serie resultante sea estacionaria. En un modelo de serie temporal el objetivo de la diferenciación es hacer que la misma se estabilice y se convierta en estacionaria (Suárez, 2013).

3.3.7.2. *No estacionariedad en varianza y autocovarianza.*

Este tipo de estacionalidad puede ser tratada con transformaciones de estabilización de varianza. Más precisamente, se aplicará una transformación sobre la serie con el objetivo de que su varianza sea constante. Es decir, se buscará T tal que V (T(Z_t)) sea constante. Box y Cox (1964), propusieron usar transformaciones de potencias del tipo:

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda} \quad (17)$$

Donde el parámetro λ puede ser agregado al modelo y estimado por el método de mínimos cuadrados. La tabla 2 muestra valores de λ comunes y sus respectivas transformaciones.

Tabla 2
Valores de lambda y sus transformaciones

Valores	Transformación
-1	$\frac{1}{Z_t}$
-0.5	$\frac{1}{\sqrt{Z_t}}$
0	$\text{Log}(Z_t)$
0.5	$\sqrt{Z_t}$

Fuente: Mogni (2013).

Es importante aclarar que estas transformaciones se aplican en series positivas y que deben utilizarse antes del método de las diferencias. Esto no introduce restricción alguna ya que puede sumarse una constante a la serie en orden de hacerla positiva (Andres Pablo Mogni, 2013) .

3.3.7.3. Modelos autorregresivo Integrado de Medias Móviles (ARIMA (p,d,q) SARIMA $(P,D,Q)s$).

A comienzo de los años 70, G.E.P. Box, profesor de Estadística de la Universidad de Wisconsin, y G.M. Jenkins, profesor de Ingeniería de Sistemas de la Universidad de Lancaster, introdujeron una pequeña revolución en el enfoque del análisis de series temporales, en sus trabajos sobre el comportamiento de la contaminación en la bahía de San Francisco, con el propósito de establecer mejores mecanismos de pronóstico y control. El libro, “Time Series Analysis, Forecasting and Control.”, escrito por Box- Jenkins en 1976 , en el que describen la metodología, se convirtió rápidamente en un clásico, y sus procedimientos se utilizan ampliamente desde entonces, en diferentes ramas de la ciencia, conociéndose como modelos ARIMA y también como modelos Box-Jenkins”

En un modelo ARIMA se considera que el comportamiento de la variable, en cualquier momento del tiempo, está influenciado por las observaciones de la propia variable (recientes o remotas), incorporadas a un modelo mediante los términos autorregresivos (AR) y los errores o influencia de los elementos aleatorios (recientes o remotos) que se representan con los términos de medias móviles (MA) .Pero también existen componentes de tendencia y estacionalidad que se adicionan al modelo ARMA, por eso recibe el nombre de integrado. En su concepción más amplia concibe la unión de un arma con un arma adicionando además tendencia y estacionalidad. En la práctica se evidencia que para aplicar un modelo ARIMA se debe tener un mínimo de 10

períodos estacionales (Suárez, 2013). En su forma más general el modelo ARIMA (p.d.q)

ARIMA (P.D.Q)_s podría escribirse como:

$$Y_t = \varphi_1 Y_{T-1} + \varphi_2 Y_{T-2} + \dots + \varphi_{P_S+p+D_S+d} Y_{T-P_S-p-sD-d} + \delta + U_T + \theta_1 U_{T-1} + \dots + \theta_{Q_S+q} U_{T-sQ-q} \quad (18)$$

Entendiendo que puede haber más de un proceso generador de la serie (en la parte regular y en la estacional) y escribiendo una combinación de los modelos MA(q) y AR(p) que han precisado de una serie de diferenciaciones "d" en la parte regular o "D" en la parte estacional para que fueran estacionarios (Arce & Mahía, 2010).

3.3.8. Metodología de Box – Jenkins.

El enfoque de Box-Jenkins es una de las metodologías de uso más amplio para el modelamiento estocástico de series de tiempo. Es popular debido a su generalidad, ya que puede manejar cualquier serie, estacionaria o no estacionaria, y por haber sido implementado en numerosos programas computacionales. Los pasos básicos de la metodología de Box-Jenkins son:

- Verificar la estacionariedad de la serie. Si ésta no es estacionaria, diferenciarla hasta alcanzar estacionariedad.
- Identificar un modelo tentativo.
- Estimar el modelo.
- Verificar el diagnóstico (si este no es adecuado, volver al paso 2).
- Usar el modelo para pronosticar.

Lo que se trata es de identificar el proceso estocástico que ha generado los datos, estimar los parámetros que caracterizan dicho proceso, verificar que se cumplan las hipótesis que han

permitido la estimación de dichos parámetros. Si dichos supuestos no se cumplieran, la fase de verificación sirve como retroalimentación para una nueva fase de identificación. Cuando se satisfagan las condiciones de partida, se puede utilizar el modelo para pronosticar (Rios, 2008).

3.3.8.1. Identificación del modelo y estimación de parámetros.

En la etapa de identificación el objetivo es definir si es necesario aplicar algún tipo de transformación, si el modelo debe incluir una media no nula y finalmente decidir cuál de los modelos es el indicado y con qué parámetros.

Mogni (2013), propone para determinar el modelo, utilizar principalmente las funciones de autocorrelación y de autocorrelación parcial. Los pasos a seguir propuestos son:

- Graficar la serie y elegir las transformaciones apropiadas. Se suele aplicar la transformación de estabilización de varianza y usarla como serie original. Determinar si la serie necesita ser diferenciada para lograr estacionariedad.
- Examinar las ACF y PACF. En el caso de que la ACF decaiga lento, estarían implicando que el proceso es no estacionario en Z_t pero posiblemente estacionario en alguna diferencia de Z_t . Así se debería aplicar las diferencias correspondientes.
- Examinar las ACF y PACF de la serie diferenciada con el objetivo de determinar los valores de p y q correspondientes. Para ello, se utiliza la tabla 3:

Tabla 3
Descripción de ACF Y PACF según el modelo de serie estacional

PROCESO	ACF	PACF
AR(p)	Decae exponencialmente como una onda sinusoidal amortiguada	Se corta después de $k=p$
MA(q)	Se corta después de $k=q$	Decae exponencialmente como una onda sinusoidal amortiguada
ARMA(p,q)	Se corta después de $k=p-q$	Se corta después de $k=p-q$

Fuente: Gutiérrez (2008).

Para el caso de las series estacionales, la identificación del modelo constará de dos partes. La primera, en la que se identificará la componente periódica estacional de las series y la segunda donde se determinarán los parámetros no estacionales de ella ver tabla 4.

Tabla 4
Descripción de ACF Y PACF según el modelo de serie no estacional

PROCESO	ACF	PACF
SAR(P)_s	La cola cae	Se corta en el rezago $P \times s$
SMA(Q)_s	Se corta en el rezago $Q \times s$	La cola cae
SARMA(P,Q)_s	La cola cae	La cola cae

Fuente: Gutiérrez (2008).

3.3.8.2. *Contrastes de integración.*

Aunque se recoge bajo una misma denominación de contrastes de raíz unitaria, la literatura ofrece dos estrategias para determinar el orden de integración de las series temporales. La

primera estrategia da lugar a la especificación de los contrastes de raíz unitaria y consiste en contrastar las hipótesis:

$$\begin{cases} H_0 : X_t \sim I(d) \\ H_1 : X_t \sim I(d-1) \end{cases} \quad (19)$$

La segunda estrategia da lugar a los llamados contrastes de estacionariedad, que reversan las hipótesis nula y alternativa que especifican los contrastes de raíz unitaria. En concreto, en los contrastes de estacionariedad se formularán las siguientes hipótesis:

$$\begin{cases} H_0 : X_t \sim I(d-1) \\ H_1 : X_t \sim I(d) \end{cases} \quad (20)$$

Las propuestas de estadísticos de prueba para contrastar el orden de integración de las series temporales son muchas. Algunas de ellas se recogen en la Tabla 5.

Tabla 5
Contrastes de raíz unitaria y estacionariedad

Contraste de raíz unitaria	Contraste de estacionariedad
I_1 Vs I_0	I_0 Vs I_1
Test de Dickey-Fuller (DF)	Test de Kwiatkowski, Phillips, Schmidt y Shin (KPSS)
Test de Phillips-Perron (PP)	Test Leybourne y McCabe (LMC)
Test de Sargan y Bhargava (DW)	Test de Choi
Test de variables instrumentales de Hall	Test de Park
Test de Schmidt y Phillips (SP)	Test de Bierens y Guo
Test de Von Newman	Test de Arellano y Pantula

Test de Bhargava

Test de Durbin Hausman

Fuente: Gutiérrez (2008).

La exposición va a contemplar, en primer lugar, el caso en el que se supone que como máximo va a poder existir una raíz unitaria en el polinomio autorregresivo de la variable por lo que la variable será I_1 o I_0 (Silvestre, 2001).

3.3.8.2.1. *Contrastes de raíz unitaria.*

Test de Dickey-Fuller (DF): El test de Dickey-Fuller se propone en los trabajos de Fuller (1976) y Dickey y Fuller (1979, 1981) y parte del supuesto que la serie temporal que se analiza se puede representar como un proceso autorregresivo de primer orden. Proponen contrastar el orden de integración de la variable especificando las siguientes hipótesis:

$$\begin{cases} H_0 : |\phi| = 1 \equiv X_t \sim I(1) \\ H_1 : |\phi| < 1 \equiv X_t \sim I(0) \end{cases} \quad (21)$$

Cuyo estadístico de prueba es:

$$t_\phi \Rightarrow \frac{(1/2)(W * (1^2) - 1)}{\sqrt{\int_0^1 W * (r^2) dr}} \quad (22)$$

El test de Phillips-Perron (PP): Otra manera alternativa de proceder para eliminar la dependencia que muestran las distribuciones asintóticas respecto de los parámetros molestos σ_ε^2 y σ^2 consiste en corregir de manera no paramétrica el valor de los estadísticos de prueba. Esta idea es la que se propone en los trabajos de Phillips (1987) y Phillips y Perron (1988).

3.3.8.2.2. *Contrastes de estacionariedad.*

Test KPSS: Kwiatkowski et al. (1992) proponen revertir la hipótesis nula y alternativa de los contrastes de raíces unitarias para determinar, de manera complementaria, el orden de integración de las series temporales. En términos formales, el contraste KPSS especifica la hipótesis nula especificada en la Ecuación 20 ,(Silvestre, 2001)

El contraste de KPSS es un contraste de multiplicadores de Lagrange que se formula como:

$$KPSS_i = \frac{T^{-2} \sum_{t=1}^T S_t^2}{S^2} \quad (23)$$

3.3.8.3. *Criterios de información.*

3.3.8.3.1. *Criterios de Información de Akaike AIC.*

El método AIC permite determinar cuál modelo es el más parecido al correcto y cuantificar su parecido. La base teórica del método AIC, combina la teoría de máxima verosimilitud, la de la información y el concepto de entropía de la información. El modelo AIC está definido por la ecuación:

$$AIC = N * \ln\left(\frac{SC}{N}\right) + 2K \quad (24)$$

Donde N es el número de datos, K es el número de parámetros de la ecuación de ajuste a la regresión más uno, porque en la regresión se estima el número de parámetros más la suma de cuadrados, y SC es la suma de cuadrados de las distancias verticales de los puntos a la curva. El valor final puede resultar positivo o negativo; por tanto el criterio cobra verdadera importancia cuando se comparan los modelos, para esto se trabaja con las diferencias entre los AIC. Cuando se está en presencia de varios modelos lo que se hace es calcular los AIC individuales de cada uno y se tomara como modelo correcto el de menor valor (Suárez, 2013).

3.3.8.3.2. *Criterios de Información de Bayesiano (BIC).*

Este criterio penaliza también la cantidad de parámetros introducidos al modelo pero, a diferencia de Akaike, el BIC proviene de un marco bayesiano. El BIC se define como:

$$BIC = -2\log(L) + M\log(n) \quad (25)$$

Se elegirán los modelos que minimicen BIC. Desde una perspectiva bayesiana, BIC se utiliza para encontrar el modelo más probable dados los datos (Andres Pablo Mogni, 2013).

3.3.8.4. *Comparación de errores y medidas de precisión del pronóstico.*

Estas medidas son utilizadas para determinar qué tan eficaz es un pronóstico a través del cálculo de su precisión con respecto a los valores reales, o sea, con estas se busca obtener una medida de qué tan lejos se encuentran los valores pronosticados de los obtenidos en la realidad. Se seleccionará el modelo que se estime más cercano a la realidad. Con las medidas de precisión se halla el error de pronóstico, o lo que es lo mismo, la diferencia entre el valor real y el pronosticado del período correspondiente.

$$E_t = Y_i - F_t \quad (26)$$

Donde E_t es el error del pronóstico del período, Y_i es el valor real para ese período y F_t el valor que se había pronosticado. Entre estas medidas se encuentran:

- Error medio (ME): $ME = \frac{\sum_{t=1}^N E_t}{N}$
- Error medio absoluto (MAE): $MAE = \frac{\sum_{t=1}^N |E_t|}{N}$
- Error medio absoluto porcentual (MAPE): $MAPE = \frac{\sum_{t=1}^N \frac{|E_t|}{|Y_t|}}{N} * 100$
- Error medio porcentual (MPE): $MPE = \frac{\sum_{t=1}^N \frac{E_t}{|Y_t|}}{N}$
- Error cuadrático medio (MSE): $MSE = \frac{\sum_{t=1}^N E_t^2}{N}$
- Raíz del error cuadrático medio (RMSE): $RMSE = \sqrt{MSE}$

(Suárez, 2013).

3.3.8.5. *Diagnóstico y validación de supuestos.*

Con los probables modelos seleccionados se realiza la revisión diagnóstica a través de:

- El análisis de la gráfica de los residuales en busca de valores extremos y evidencia de períodos en los cuales el modelo no se ajusta a los datos.
- Con la prueba de hipótesis del estadístico Q de Ljung-Box se determina la presencia de autocorrelación. El estadístico Q permite realizar un contraste global de los residuos y determinar si éstos son ruido blanco. Para ello, se acompaña de un nivel de significancia P, que indica el grado de no rechazo de la hipótesis nula (residuos son ruido blanco) y, en

consecuencia, cuanto mayor sea dicho nivel, mayor confianza se puede depositar en la hipótesis nula.

- Examinar el histograma de la serie de residuos y utilizar el estadístico para contrastar la hipótesis de Normalidad de las perturbaciones.
- Examinar la ACF y la PACF residuales. Un amortiguamiento muy lento de la ACF indica que la serie de residuos no es estacionaria,

Se debe garantizar el cumplimiento de estos supuestos para dar escogencia al modelo adecuado para el pronóstico (Arce & Mahía, 2010).

3.3.8.6. *Pronósticos.*

3.3.8.6.1. Pronostico dentro y fuera de muestra.

Al hablar de pronósticos, se distingue entre proyecciones dentro y fuera de muestra. En las primeras, las proyecciones realizadas se refieren a los mismos datos que se emplearon para la construcción y calibración del modelo (la muestra), mientras que en las segundas las proyecciones se refieren a datos ajenos a dicha muestra. En la búsqueda de metodologías que generen pronósticos precisos de los valores futuros de una variable, sólo son relevantes las proyecciones fuera de muestra por las siguientes razones:

- Las proyecciones fuera de muestra replican el funcionamiento de la herramienta de pronósticos en la práctica, por lo que la evaluación de su desempeño predictivo será un referente válido para los futuros errores de pronóstico.
- Los modelos de pronóstico se construyen minimizando los errores dentro de muestra por lo que los errores de pronósticos intramuestrales sobrestiman el potencial predictivo de las herramientas.

- Un modelo con buen desempeño intramuestral podría tener un muy mal desempeño en proyecciones fuera de muestra. Esto se debe a un sobreajuste (overfitting) o memorización de los datos muestrales, con lo que el modelo resultante será incapaz de responder de buena manera a nuevos valores.

3.3.8.6.2. *Pronostico estáticos y dinámicos.*

Los pronósticos estáticos son aquellos que están basados en la última información efectiva disponible, por lo que están limitados a las proyecciones a un periodo hacia adelante. Los pronósticos dinámicos son caracterizados por utilizar el último pronóstico disponible como dato para el siguiente pronóstico, permitiendo la realización de proyecciones a dos y más periodos hacia adelante (Rios, 2008).

4. Marco Metodológico

4.1. Unidad de análisis

Se utiliza como unidad de análisis una base–matriz de datos de recepción de leche de búfala en la planta de Colanta en Planeta Rica desde el año 2008 hasta abril del 2018, esta matriz cuenta con los resultados promedios mensuales de los parámetros fisicoquímicos y composicionales de la leche de búfala recibida en este tiempo. Los parámetros estudiados y sus unidades de manejo son las siguientes:

- Acidez: porcentaje de ácido láctico (%m/v).
- Densidad: gramos sobre mililitros (g/ml).
- Grasa: porcentaje masa /volumen (% m/v).
- Proteína: porcentaje masa /masa (% m/m).

4.2. Instrumentos-materiales

Se realiza análisis estadístico y desarrollo de gráficos con ayuda del Software RStudio

RStudio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

4.3. Procedimiento

4.3.1. Análisis descriptivo.

Se realiza análisis descriptivo de la grasa, proteína, acidez y densidad de la base de datos estudiada con ayuda de las medidas de tendencia central, histograma y diagrama de caja (box-plot).

4.3.2. Construcción de modelos.

Para la construcción de los modelos de cada uno de los parámetros estudiados se utiliza la metodología planteada por Box y Jenkins en 1976 en la cual se estipula las etapas de identificación, estimación, comprobación y pronóstico, las cuales se llevaron a cabo como se muestra a continuación:

- **Identificación de parámetros:** Se utiliza las funciones de autocorrelación y autocorrelación parcial ACF y PACF respectivamente, para advertir principalmente tendencias, ciclos, correlaciones, rezagos y datos atípicos. Se aplica contraste de raíz unitaria con ayuda del test de Dickey-Fuller Aumentada (ADF) para determinar si la serie es o no estacionaria y de esta manera saber si se requiere alguna diferenciación estacional o no estacional.
- **Estimación:** Con ayuda del software estadístico se ajustan diferentes modelos tipo ARIMA a cada uno de los parámetros seleccionados, para esto se utilizan las bases de datos de la siguiente manera: para porcentaje de acidez hasta junio del 2017, densidad, porcentaje de grasa y proteína desde el 2008 hasta abril del 2017. Se escoge el modelo para cada parámetro que arroje menor Criterio de Información Bayesiano (BIC) y cuyos parámetros sean significativos.
- **Comprobación:** Se efectúan pruebas de verificación de supuesto de los residuos en donde se verifica: Errores no autocorrelacionados, errores con distribución normal y errores independientes, para esto se utilizan los estadísticos de Ljung-Box y Jarque Bera, además se realiza análisis de ACF, Q-Q Plot de los residuos.
- **Pronóstico:** Se realiza pronóstico fuera de muestra y análisis de estadísticos de exactitud (error medio y error medio absoluto porcentual) y estadístico Theil's U.

5. Análisis y resultados

5.1. Análisis descriptivo

Porcentaje de Acidez: La figura 1. Muestra una configuración normal aproximadamente simétrica oscilando entre 0,122% y 0,155%, se presenta dos valores atípicos (0,1163% y 0,1566%) en toda la base de datos. Los resultados de acidez de los años estudiados se encuentran dentro del rango estipulado por Cárdenas (2017) en el estudio “Caracterización fisicoquímica y composicional de la leche de búfala”.

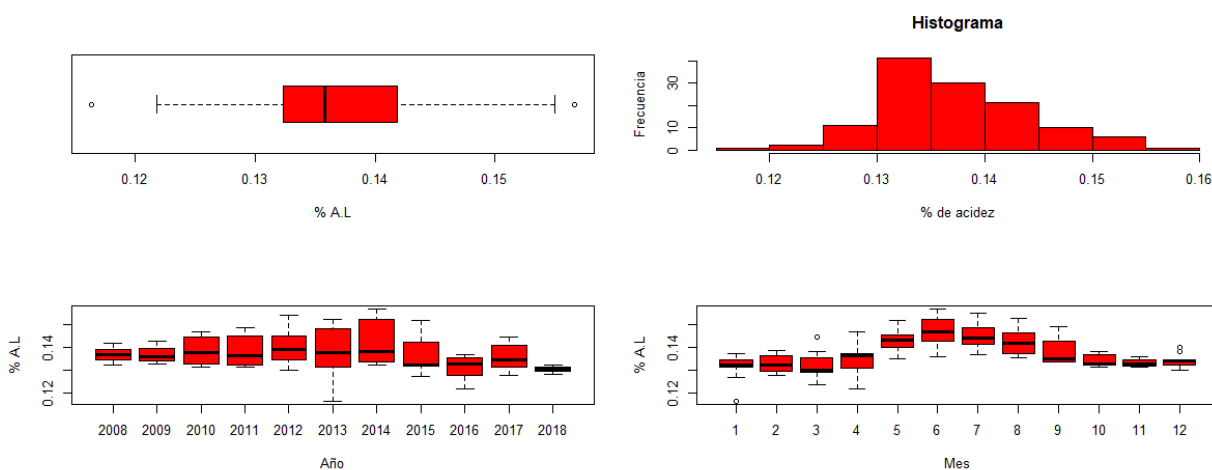


Figura 1 Análisis Descriptivo Porcentaje de Acidez leche de Búfala Recibida en Colanta Planeta Rica.

Desde el 2008 al 2014 se observa un aumento en la acidez de la leche de búfala, resultados que comienzan nuevamente a bajar en 2015 se observa que anualmente se mantiene una mediana alrededor de 0,135%, en el 2014 y 2015 se observa cola hacia la derecha por lo que los resultados de acidez para estos años son altos principalmente en el 2015, el año 2013 se caracterizó por tener variabilidad en los porcentaje de acidez, presentándose en este año los valores más bajos. En la gráfica mensual se observa que los meses mayo a septiembre presentan

porcentaje de acidez más alta estos meses coinciden con las temporadas más lluviosas (IDEAM, 2018), la marcada variación en la acidez se ve altamente afectada por inadecuadas prácticas de limpieza y secado de la ubre que por la influencia de la época del año, lo cual se ve reflejado en los componentes y calidad higiénica de la leche (Hurtado, N., Cerón, M., Lopera, M., Bernal, A. y Cifuentes, 2005), está demostrado que en las temporadas de lluvias es muy difícil garantizar las prácticas higiénicas por lo cual esto influye en la calidad microbiológica de la leche y por ende en el aumento del porcentaje de acidez.

Densidad:

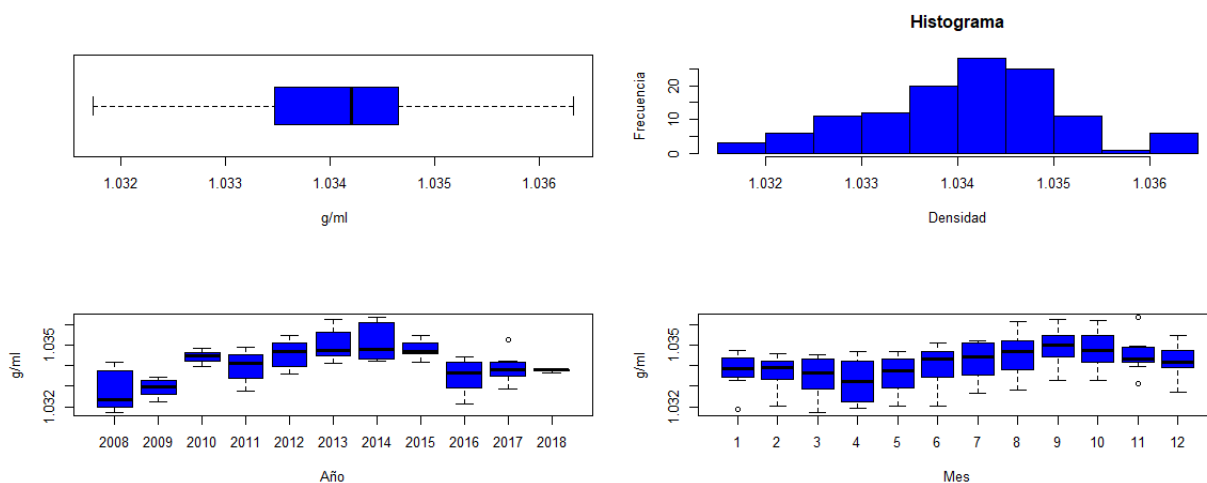


Figura 2 Análisis Descriptivo Densidad de la leche de Búfala Recibida en Colanta Planeta Rica.

Los resultados de densidad de la leche de búfala a nivel general presenta una variación entre 1,032 g/ml y 1, 036 g/ml sin presencia de datos atípicos, resultados claramente superiores a los de leche de vaca los cuales están entre 1,030 g/ml a 1,033g/ml (Ministerio de la protección social, 2006). Patiño (2011) expresa que la densidad de la leche no es constante, depende de su composición y temperatura de allí que la leche de búfala posea mayor densidad que la leche de vaca, pues el contenido de grasa y sólidos totales es elevado en ésta especie. Se observa datos

asimétricos con tendencia a cola hacia la izquierda. Los resultados anuales son muy variados, presentándose densidades altas en el 2012 y 2015 con medianas similares, ver figura 2.

Porcentaje de Grasa:

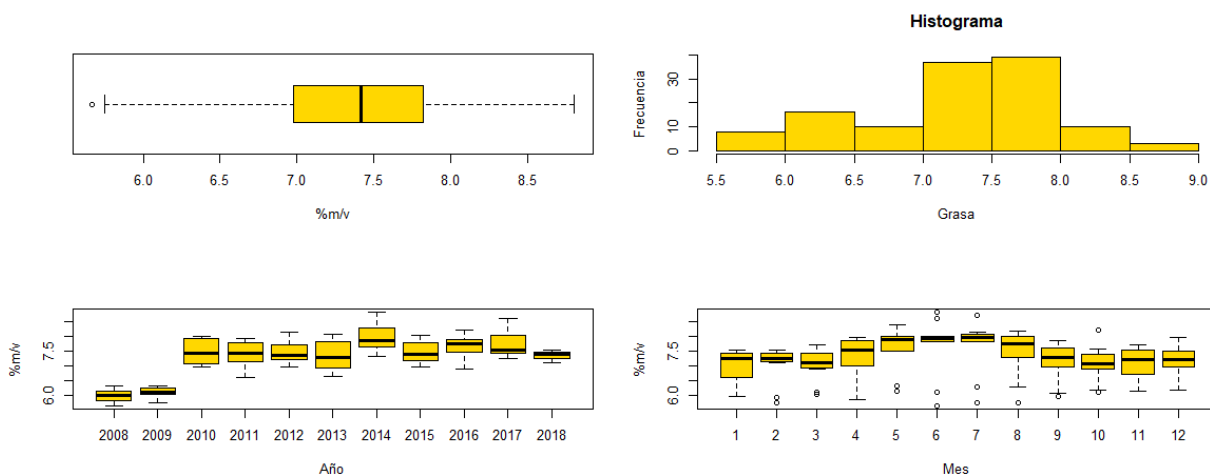


Figura 3 Análisis Descriptivo Porcentaje de Grasa leche de Búfala Recibida en Colanta Planeta Rica.

La figura 3 muestra resultados de análisis de grasa casi simétricos con poca asimetría a la izquierda con resultados entre 5,75% y 8,80% bastante superiores a los de leche de vaca lo cual es normal para este tipo de leche, es una de las características que la distingue de la leche bovina. Para los años 2008 y 2009 se evidencia porcentajes de grasa menores a la mediana presentada en los años siguientes, los datos más altos para estos años no superan el 6,5% se considera que un cambio de tecnología en la medición de este parámetro puede ser la causa de este cambio bastante evidente. Mensualmente se observan gran cantidad de datos atípicos lo cual se atribuye a los resultados de los años 2008 y 2009 que logran influenciar en los resultados mensuales. Para la selección del modelo se omitirán los resultados de estos dos años.

Porcentaje de Proteína: En términos generales la proteína varió entre 3.183 % y 4.615 % sin presentar datos atípicos, la media estuvo en 4.012 % y la mediana en 4.082%, en el

histograma de frecuencia se observa la mayor concentración de los datos en el lado izquierdo.

Los resultados de proteína concuerdan con lo que se muestra en la literatura, existe una estrecha relación entre la cantidad de grasa y la cantidad de proteína en la leche; cuanto mayor es la cantidad de grasa, mayor es la cantidad de proteína, la leche de búfala en relación con la leche de vaca contiene un 30% más de proteína con mayores proporciones de caseína, albúmina y globulina(Soto, J. y Teherán, 2007).

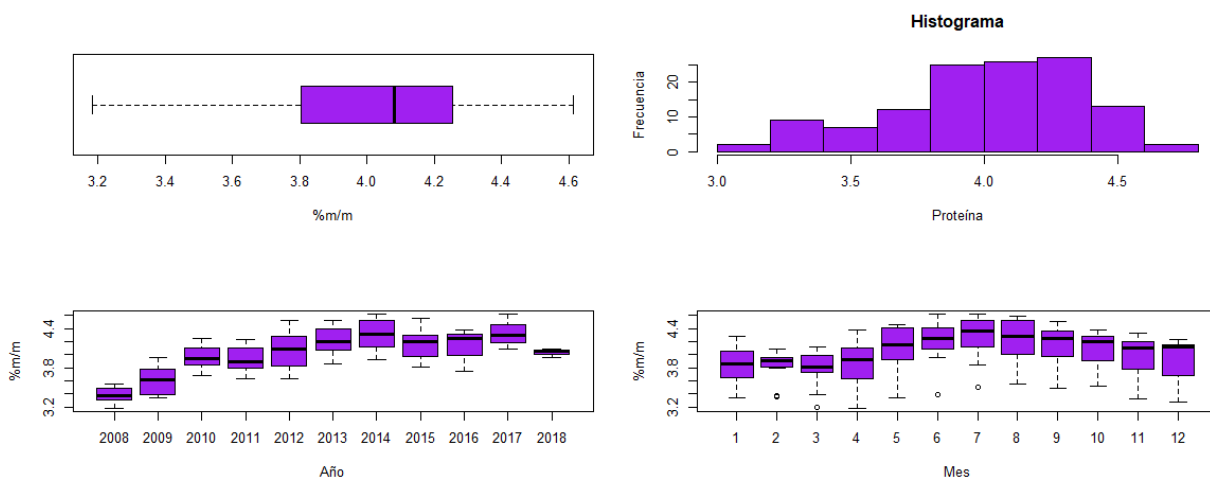


Figura 4 Análisis Descriptivo Porcentaje de Proteína leche de Búfala Recibida en Colanta Planeta Rica.

Su comportamiento anual muestra incremento a través del tiempo, presentando resultados más bajos en el año 2008 y resultados altos en el 2012, 2013 y 2014. Mensualmente se puede observar un comportamiento homogéneo, sin embargo los meses de Junio-Septiembre son los que presentan resultados mayores y lo meses de noviembre a febrero los resultados de proteína en leche de búfala son más bajos.

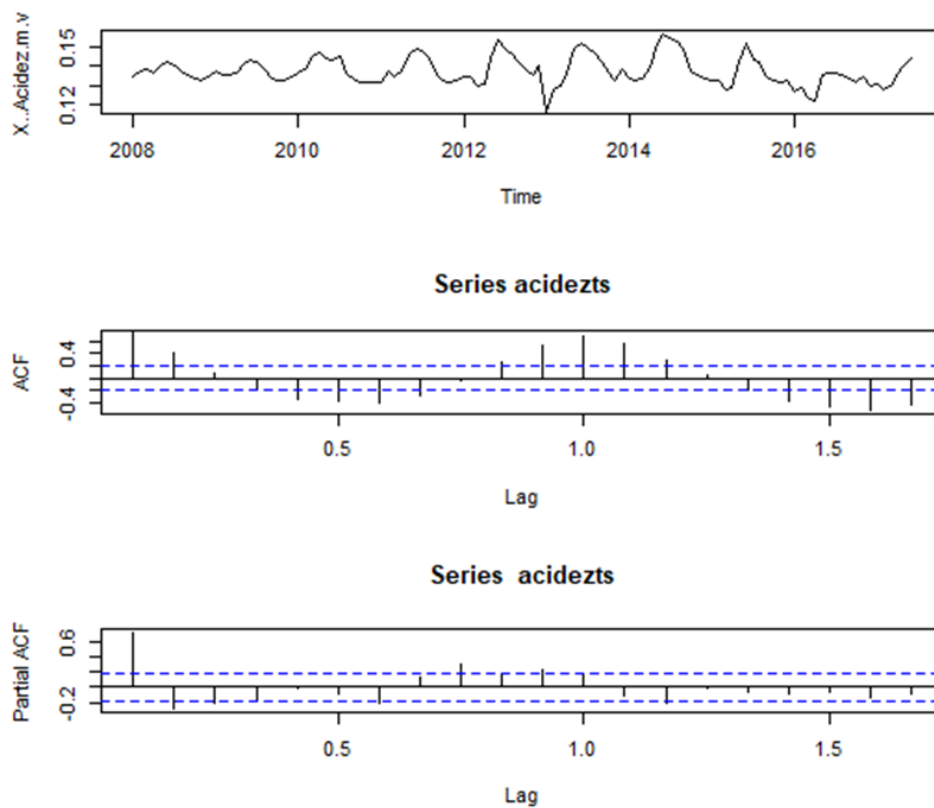


Figura 5 Base de Datos porcentaje de acidez de leche de Búfala

5.2. Construcción de modelos

5.2.1. Porcentaje de Acidez.

Identificación de parámetros: se realiza análisis de ACF y PACF de la serie tiempo porcentaje de acidez de leche de búfala, en la figura 5 y figura 6 se observan los graficas de la base de datos correspondiente a este parámetro, serie sin diferenciar y diferenciada.

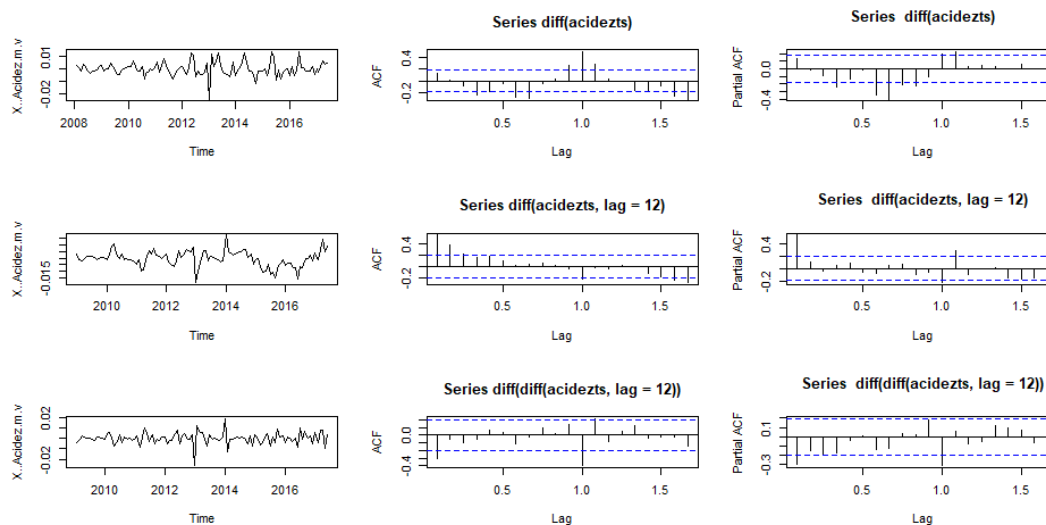


Figura 6 Análisis de serie diferenciada Porcentaje de Acidez de Leche de Búfala

Se muestra a simple vista una serie con una parte estacional que requiere de al menos algún tratamiento en parte no estacional con un lag de 12.

Se realiza contraste de raíz unitaria con ayuda de software estadístico Rstudio, aplicando el test de Dickey-Fuller Aumentada (ADF) para lo cual se obtiene:

Augmented Dickey-Fuller Test

data: acidezts

Dickey-Fuller = -1.5281, Lag order = 12, **p-value = 0.7725**

Por lo que no se rechaza hipótesis nula y se concluye que al menos se requiere una diferenciación para que la serie sea estacional.

data: diff(acidezts, lag = 12)

Dickey-Fuller = -2.5603, Lag order = 4, **p-value = 0.3444**

Augmented Dickey-Fuller Test

data: diff(diff(acidezts, lag = 12))

Dickey-Fuller = -5.9401, Lag order = 4, **p-value = 0.01**

Según el test de Dickey –Fuller aumentado con una confianza del 95% se requiere de dos diferenciaciones para que la serie sea estacional. Es normal que la serie de tiempo para el porcentaje de acidez en leche de búfala tenga una parte estacional y que esta tenga una tendencia generalmente en la media ya que algunos autores como Patiño y Guanziroli (2005) señalan que este parámetro depende de la alimentación del animal, época de lactancia así como de las estaciones del año.

Estimación: con ayuda del paquete forecast del software estadístico Rstudio se realiza el ajuste de modelos utilizando la función Auto.arima y Arima se analizan varios modelos en donde se escoge el modelo con menor BIC y cuyos coeficientes son significativos ver tabla 6.

De acuerdo a los resultados de la prueba de significancia de parámetros, se puede afirmar con una confianza del 95% que los parámetros estipulados para el modelo son significativos.

Tabla 6 *Estimativos para el modelo ARIMA (1, 0,0) (0, 1,1)₁₂*

Parámetro	φ_1	θ_1
Estimación	0.6658	-0.7084
Error-Estándar	0.0749	0.1329
Estadístico t	8.890668	-5.330031
BIC		-811.7
σ^2		$1.67e^{-05}$

Fuente: esta investigación

Comprobación/evaluación de supuestos: escogido el modelo ARIMA(1,0,0)(0,1,1)[12] en la figura 7 se puede observar que en la gráfica de residuos estandarizados, visualmente no presenta tendencia y se asemeja a ruido blanco. En la gráfica Q-Q se puede observar que los residuos tienen una tenencia normal sin embargo en los extremos hay resultados, por fuera de los límites, debido posiblemente a cambios en condiciones ambientales que pueden influir en la acidez de la leche, el ACF y PACF de los residuos se encuentran dentro de los límites de confianza, a simple vista se cumplen los supuestos del modelo.

Se calculan los estadísticos de Ljung-Box para determinar la presencia de autocorrelación y Jarque Bera estadístico de normalidad:

Box-Ljung test

X-squared = 18.573, df = 24, **p-value = 0.5503**

X-squared = 5.8398, df = 12, **p-value = 0.735**

Se concluye que con un lag de 12 y de 24 No se rechaza hipótesis nula, por la cual no existe correlación serial y se da cumplimiento a este supuesto de los residuos.

Jarque Bera Test

X-squared = 240.98, df = 2, **p-value < 2.2e-16**

El test de Jarque Bera muestra con una confianza del 95% se rechaza la hipótesis nula por lo cual los residuos de la muestra analizada no son normales y este supuesto no se cumple. Aunque no se cumple el supuesto de normalidad de los residuos se escoge el modelo ARIMA(1,0,0)(0,1,1)[12] debido a que es el que tiene menor BIC y los otros modelos estudiados tampoco cumplen criterio de normalidad.

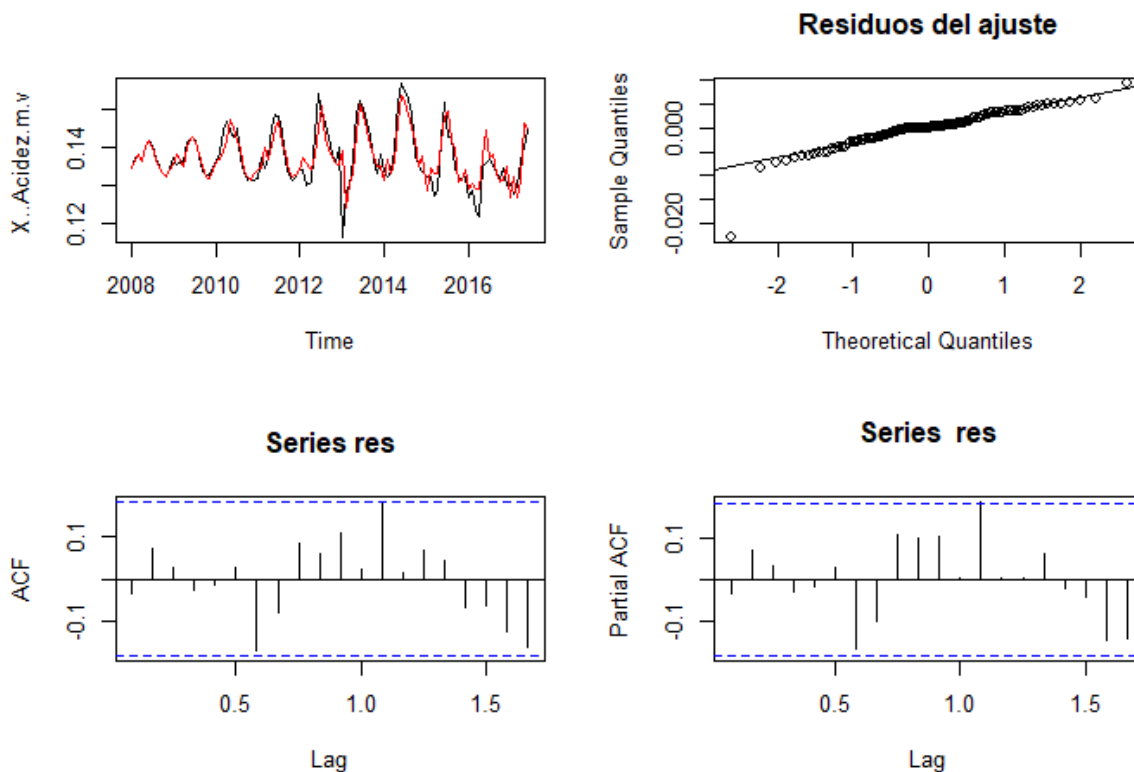


Figura 7 Evaluación de supuestos del modelo ARIMA (1, 0, 0) x (0, 1, 1)₁₂

Pronóstico: en la figura 8 se encuentra detallado el pronóstico fuera de muestra de los datos de porcentaje de acidez de la leche de búfala entre julio del 2017 y diciembre del 2018, los resultados se encuentran dentro de los límites de confianza, el error medio absoluto porcentual (MAPE) arrojó 1,75% con una confianza del 95% , este resultado es satisfactorio teniendo en cuenta que la tolerancia del método del porcentaje de acidez es 0,02% y calculando el error en el pronóstico este no sería superior 0,004% para un dato de acidez, adicional el cálculo el coeficiente de desigualdad de Theil's U muestra un resultado de 0,85 lo cual dice que el modelo escogido es acertado ya que este resultado es menor a 1.

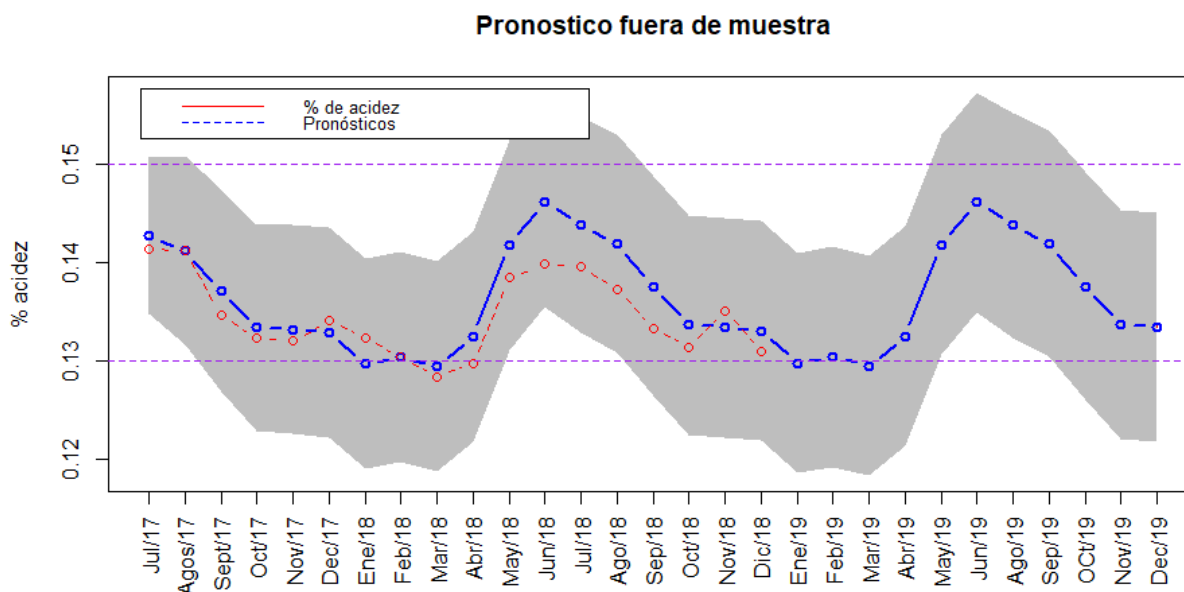


Figura 8 Pronóstico fuera de muestra del modelo ARIMA (1, 0,0) x (0, 1, 1)12 para porcentaje de acidez de la leche de búfala

De acuerdo al modelo escogido, se pronostica que para el año 2019 hasta el mes de abril el porcentaje de acidez de la leche de búfala se mantendrá menor o igual de 0,130%, lo cual es lógico por la época de verano lo cual disminuye sólidos y por ende este parámetro fisicoquímico se ve disminuido, de acuerdo al pronóstico se predice para el meses de mayo a agosto % de acidez mayor a 0,14% y menor a 0,150% y de septiembre a diciembre se mantendrá entre 0,130% - 0,14% rangos normales para esta materia prima.

5.2.2. Densidad (g/ml).

Identificación de parámetros: el parámetro densidad de la leche de búfala se presenta como una serie de tiempo con tendencia como se puede observar en la figura 9, se requiere de al menos una diferenciación de su parte no estacional y estacional para lograr el modelo ideal que se ajuste a los datos y con los cuales se pueda pronosticar. La densidad de la leche depende de su composición y está a su vez de la alimentación, raza y ciclo de lactancia por lo cual la

estacionariedad de estos parámetros están ligados e incluso se podrían trabajar juntos para un próximo estudio.

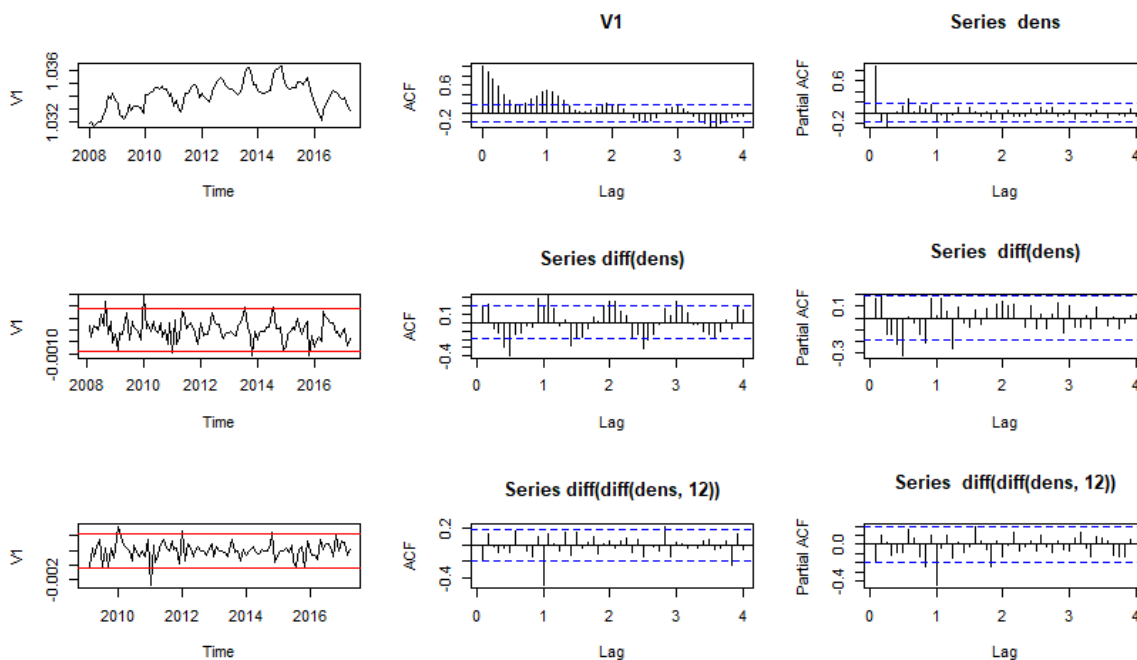


Figura 9 Serie de tiempo Densidad de la leche de Búfala

El contraste de raíz unitaria test de Dickey-Fuller aumentado realizado con el software estadístico Rstudio arroja un p-valor de 0,9803 y 0,09272 para la serie inicial general y con un rezago de 12 respectivamente lo que muestra que no se rechaza la hipótesis nula por lo cual la serie tiene al menos una raíz unitaria. Al realizar diferenciación se obtiene un p-valor menor de 0,01 para ambas series lo que indica que se rechaza la hipótesis nula por lo cual la serie diferenciada no tiene una raíz unitaria para la parte estacional y para la parte no estacional del modelo.

Estimación: se realiza el ajuste de modelos utilizando la función Auto.arima y Arima del paquete forecast del software RStudio se analizan varios modelos en donde se escoge el modelo con menor BIC y cuyos coeficientes fueran significativos. El modelo **ARIMA(0,1,0)(0,1,1)[12]** fue el que obtuvo menores resultados en el criterio de información Bayesiano.

Tabla 7 *Estimativos para el modelo ARIMA (0, 1,0) (0, 1,1)₁₂*

Parámetro	Θ_1
Estimación	-0.9975
Error-Estándar	0.1714
Estadístico t	-5.821202
BIC	-1242.65
σ^2	2.87e-07

Fuente: esta investigación

De acuerdo a los resultados de la prueba de significancia de parámetros, se puede afirmar con una confianza del 95% que el parámetro del modelo ARIMA(0,1,0)(0,1,1)[12] es significativo.

Comprobación/evaluación de supuestos: escogido el modelo ARIMA(0,1,0)(0,1,1)[12] para el parámetro de densidad en la leche de búfala en la figura 10 se puede observar que en la gráfica de residuos estandarizados visualmente no presenta tendencia y se asemeja a ruido blanco, en la gráfica Q-Q se puede observar que los residuos no tienen una tenencia normal, en los extremos hay resultados por fuera de los límites, el ACF y PACF de los residuos muestra que el rezago 1 se sale de los límites de confianza, el estadístico de Ljung-Box arroja un p-valor de 0.1057 con un rezago de 12 y 0.2839 con rezago de 24 por lo cual se puede concluir con una confianza del 95% que no existe correlación serial al no rechazar la hipótesis nula. El estadístico de Jarque Bera arroja un p-valor $< 2.2e-16$ lo que corrobora que no se cumple el supuesto de normalidad de los residuos al rechazar la hipótesis nula. Aunque no se cumplen algunos supuestos se considera que es el mejor modelo de todos los estudiados.

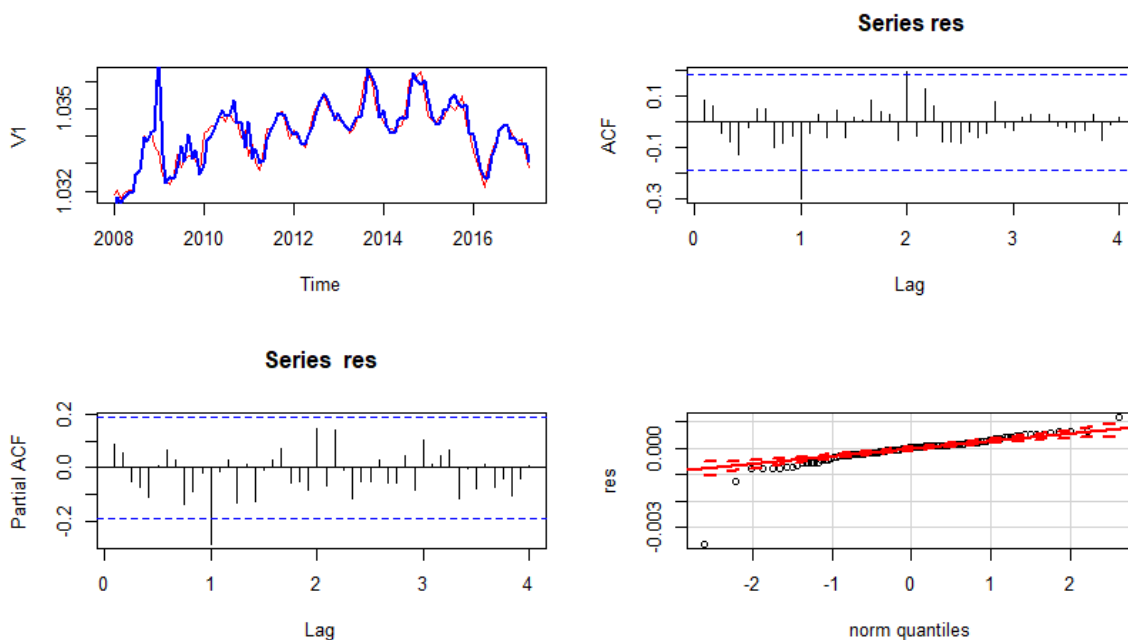


Figura 10 Evaluación de supuestos del modelo ARIMA (0, 1,0) x (0, 1, 1)¹²

Pronóstico: Al analizar el pronóstico fuera de muestra del modelo ARIMA(0,1,0)(0,1,1)[12] el cual se muestra en la figura 11, se observa que todos los resultados se encuentran dentro de los límites de aceptación. De acuerdo a los estadísticos de exactitud calculados, el MAPE arroja un resultado de 0.03417 lo que representa un error bastante bajo para un pronóstico, el estadístico Theil's U muestra un resultado de 0.667 lo cual indica que el método de pronóstico bajo estudio es mejor que el método simple.

Para el año 2019 se prevé que la densidad de la leche de búfala tenga un promedio de 1,03402g/ml con un mínimo de 1,03323 g/ml y un máximo de 1,0348 g/ml, estos valores son superiores a los de leche de vaca sin embargo están dentro de lo normal este producto. Se considera una estimación acertada teniendo en cuenta que en el 2017 y 2018 la densidad de esta leche se mantuvo entre 1,0326 g/ml y 1,0353 g/ml.

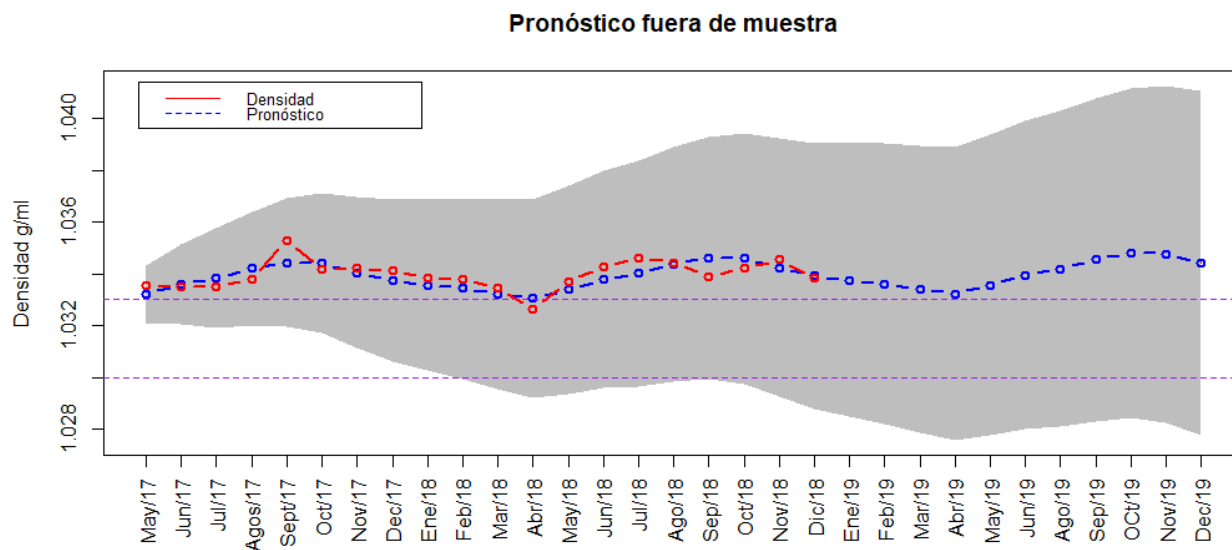


Figura 11 Pronóstico fuera de muestra del modelo ARIMA (0, 1,0) x (0, 1, 1)₁₂ para la Densidad de la leche de búfala

5.2.3. Porcentaje de grasa m/v.

Identificación de parámetros: para el análisis el parámetro porcentaje de grasa de la leche de búfala se inicia el análisis de la base de datos desde el año 2010 ya que en años anteriores la metodología utilizada para determinar % grasa en leche cruda era otra y la que se utiliza actualmente es más robusta y precisa. En la figura 12 se observa la gráfica inicial de la base de datos % grasa en leche de búfala, con varianza aparentemente constante y con una diferenciación se podría lograr la estacionalidad de la serie, lo cual se debe comprobar con el test de contraste de raíz unitaria utilizado en el presente trabajo. La grasa de la leche de búfala depende del ciclo

de lactancia y la alimentación del animal por lo cual debe ser estacional al igual que los otros parámetros ya estudiados.

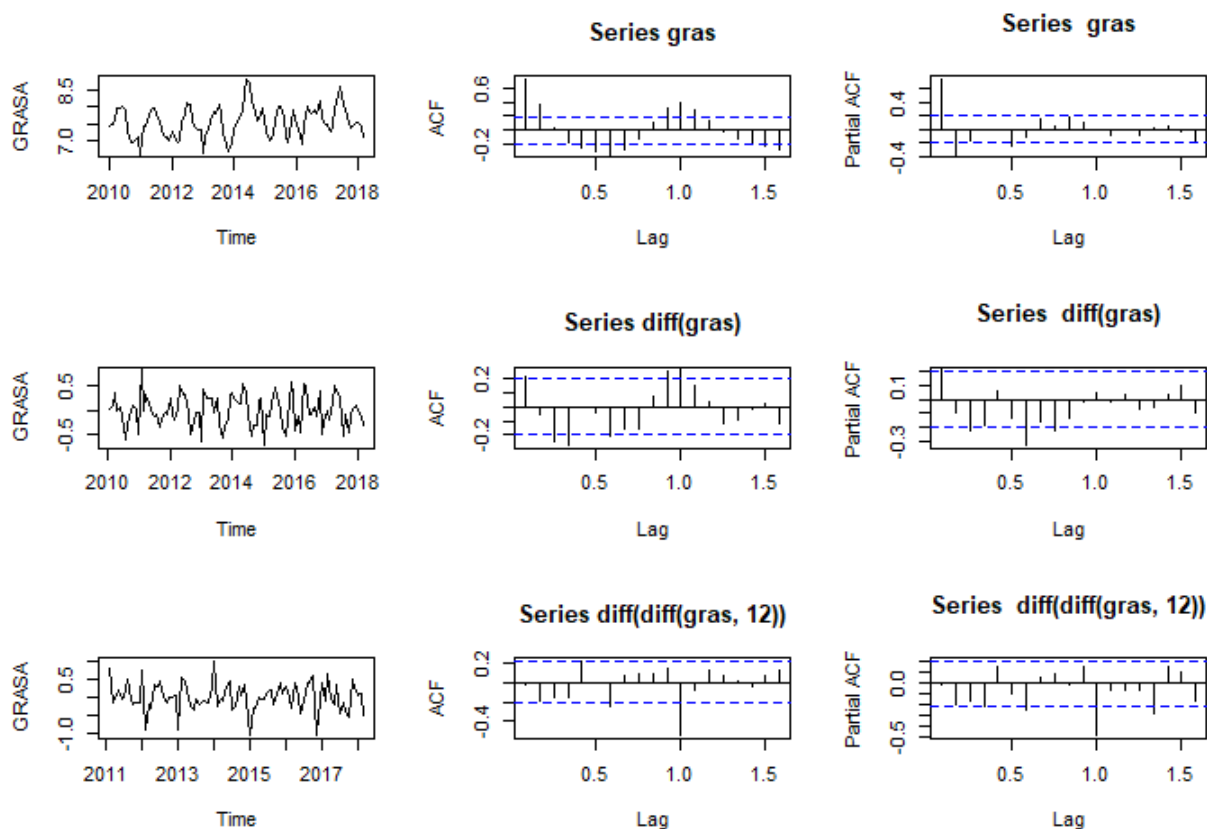


Figura 12 Serie de tiempo Grasa (%m/v) de la leche de Búfala

El contraste de raíz unitaria test de Dickey-Fuller aumentado arroja inicialmente un p-valor de 0.3465 al realizar una diferenciación de la serie y aplicar el test de raíz unitaria se observa que el p-valor es menor de 0.01 tanto en su parte estacional como en la no estacional con estos resultados se procede a rechazar la hipótesis nula y se concluye que con una diferenciación se puede lograr la estacionalidad de la serie.

Estimación: se realiza el ajuste de modelos utilizando la función `Auto.arima` y `Arima` del paquete `forecast` del software RStudio se analizan varios modelos en donde se escoge el modelo con menor BIC y cuyos coeficientes fueran significativos. El modelo **ARIMA(2,1,1)(1,0,1)[12]** fue el que obtuvo menores resultados en el criterio de información Bayesiano.

Tabla 8 *Estimativos para el modelo ARIMA (2, 1,1) (1, 0,1)₁₂*

Parámetro	φ_1	φ_2	θ_1	Φ_1	Θ_1
Estimación	0.9018	-0.2722	-0.9848	0.9989	-0.9698
Error-Estándar	0.1069	0.1067	0.0252	0.0089	0.1259
Estadístico t	8.435130	-2.551215	-39.05714	111.787475	-7.704338
BIC			41.42		
σ^2			0.0611		

Fuente: esta investigación

De acuerdo a los resultados de la prueba de significancia de parámetros, se puede afirmar con una confianza del 95% que el parámetro del modelo ARIMA(2,1,1)(1,0,1)[12] es significativo.

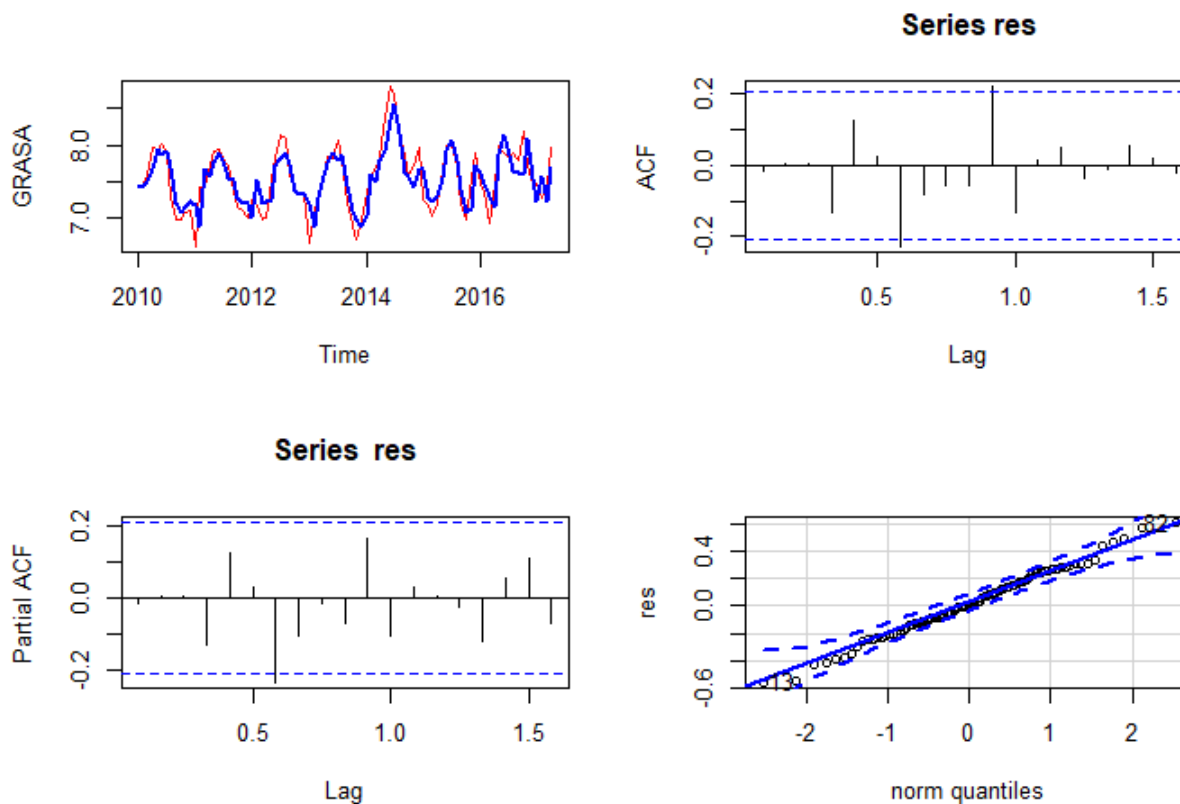


Figura 13 Evaluación de supuestos del modelo ARIMA (2, 1, 1) x (1, 0, 1)¹²

Comprobación/evaluación de supuestos: escogido el modelo ARIMA(2,1,1)(1,0,1)[12] para el parámetro de porcentaje de grasa, se procede a evaluar los supuestos de los residuos del modelo en la figura 13, en la gráfica Q-Q se puede observar que los residuos tienen una tenencia normal, en los extremos no se detectan visualmente resultados por fuera de los límites, el ACF y PACF de los residuos muestra un rezago por fuera de los límites de confianza, se comprueban estos con los estadístico de Ljung-Box y Jarque Bera el primero arroja un p-valor de 0.1813 con un rezago de 12 y 0.6504 con rezago de 24 por lo que no se rechaza la hipótesis nula y se puede concluir con una confianza del 95% que no existe correlación serial entre los residuos del modelo. El estadístico de Jarque Bera arroja un p-valor 0.9707 lo que corrobora que se cumple el supuesto de normalidad de los residuos al no rechazar la hipótesis nula.

Aunque no se cumplen algunos supuestos debido a que los autocorrelogramas muestran rezagos por encima de los niveles de confianza se considera que es el mejor modelo de todos los estudiados y los estadísticos de Jarque Bera y Ljung-Box lo demuestran.

Pronóstico: el pronóstico fuera de muestra del modelo $ARIMA(2,1,1)(1,0,1)_{12}$ en la figura 14, se observa que todos los resultados se encuentran dentro de los límites de aceptación. De acuerdo a los estadísticos de exactitud calculados, el MAPE arroja un resultado de 0.85098 lo que representa un error absoluto aceptable para un pronóstico, el estadístico Theil's U muestra un resultado de 0,5196542 lo cual indica que el método de pronóstico bajo estudio es mejor que el método simple y es válida la utilización de este.

De acuerdo al modelo ajustado, se pronostica que para el año 2019 el porcentaje de grasa en leche de búfala se encuentre entre 7,36% y 8,04% con un error de 0,85%, reflejados con la realidad son resultados dentro del rango en que se maneja la grasa para esta leche, sin embargo se resalta que depende de la alimentación del animal.

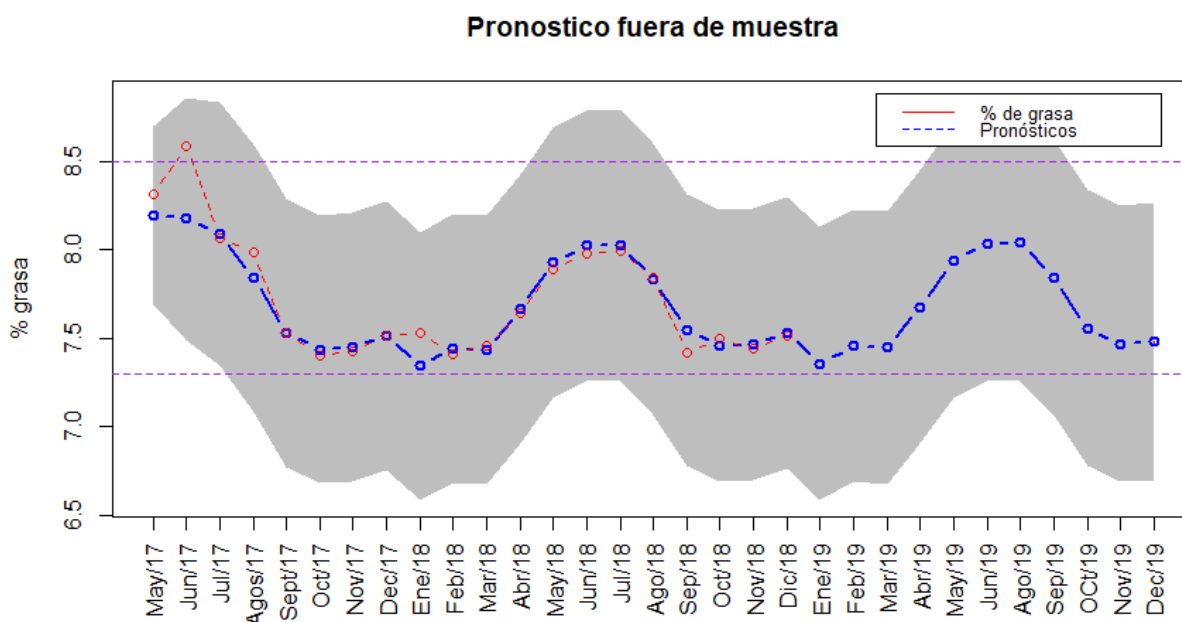


Figura 14 Pronóstico fuera de muestra del modelo $ARIMA(2,1,1) \times (1,0,1)_{12}$ para la porcentaje grasa de la leche de búfala

5.2.4. Porcentaje de proteína m/m.

La serie de tiempo del parámetro de proteína en leche de búfala de forma rápida se puede afirmar que es una serie no estacionaria en media y estacionaria en varianza, en la figura 15 se observa que la serie posee una parte autoregresiva y una de medias móviles. Al realizar la prueba de Dickey-Fuller Aumentado de la serie normal sin diferenciar se obtiene:

```
data: pr
Dickey-Fuller = -3.7302, Lag order = 4, p-value = 0.0247
alternative hypothesis: stationary
```

De acuerdo al p-valor obtenido se rechaza la hipótesis nula y se concluye con una confianza del 95% que la serie de tiempo proteína de leche de búfala es estacionaria, al saber que en su parte no estacionaria no requiere diferenciación se realiza el test de Dickey-Fuller Aumentado a la parte estacional de la serie para revisar si esta posee alguna raíz unitaria; todo esto se realiza con ayuda del software Rstudio.

```
adf.test(pr, k=12)
```

```
data: pr
```

```
Dickey-Fuller = -2.3819, Lag order = 12, p-value = 0.4179
```

```
alternative hypothesis: stationary
```

```
adf.test(diff(pr, k=12 ))
```

```
data: diff(pr, k = 12)
```

```
Dickey-Fuller = -5.9183, Lag order = 4, p-value = 0.01
```

```
alternative hypothesis: stationary
```

En la primera aplicación del test se obtiene un p-valor de 0.4179 por lo cual no se rechaza la hipótesis nula y se concluye que la serie en su parte estacional tiene al menos una raíz unitaria.

Al realizar diferenciación de esta serie se obtiene p-valor de 0.01 lo que obliga a rechazar la hipótesis nula y a concluir que la serie de tiempo proteína de leche búfala en su parte estacional tiene una raíz unitaria. En la figura 16 se puede observar la serie diferenciada y se corrobora lo obtenido en el test de raíz unitaria ya que visualmente se puede concluir que al menos una diferenciación en la serie de proteína en su parte estacional para ajustar un modelo.

Estimación: se realiza el ajuste de modelos utilizando la función Auto.arima y Arima del paquete forecast del software RStudio se analizan varios modelos en donde se escoge el modelo con menor BIC y cuyos coeficientes fueran significativos. El modelo ARIMA(3,0,2)(0,1,1)[12] fue el que obtuvo menores resultados en el criterio de información Bayesiano.

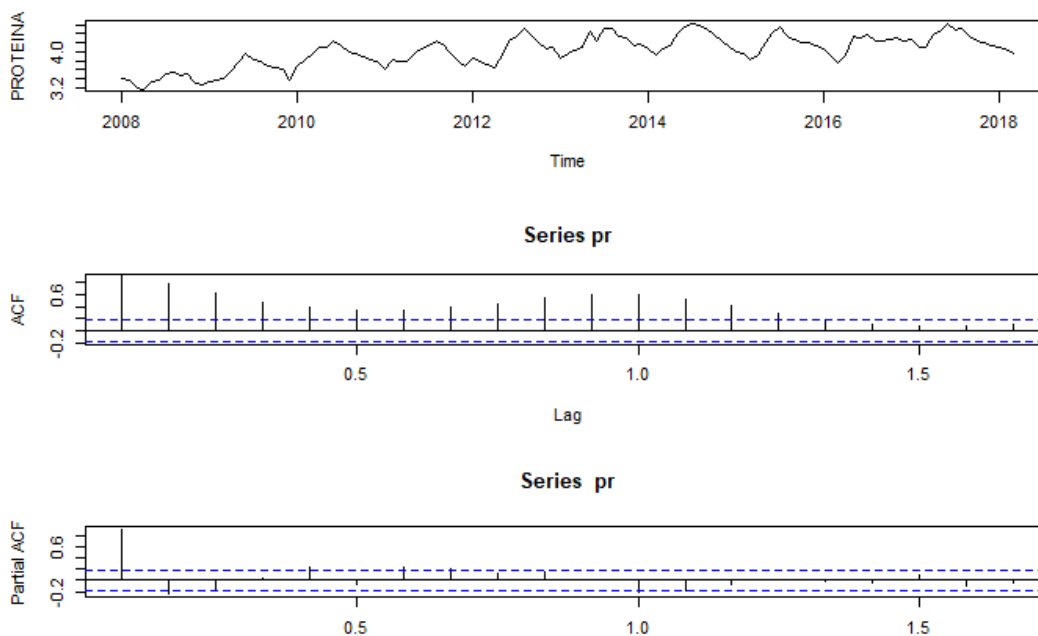


Figura 15 Base de datos porcentaje de Proteína en leche de Búfala

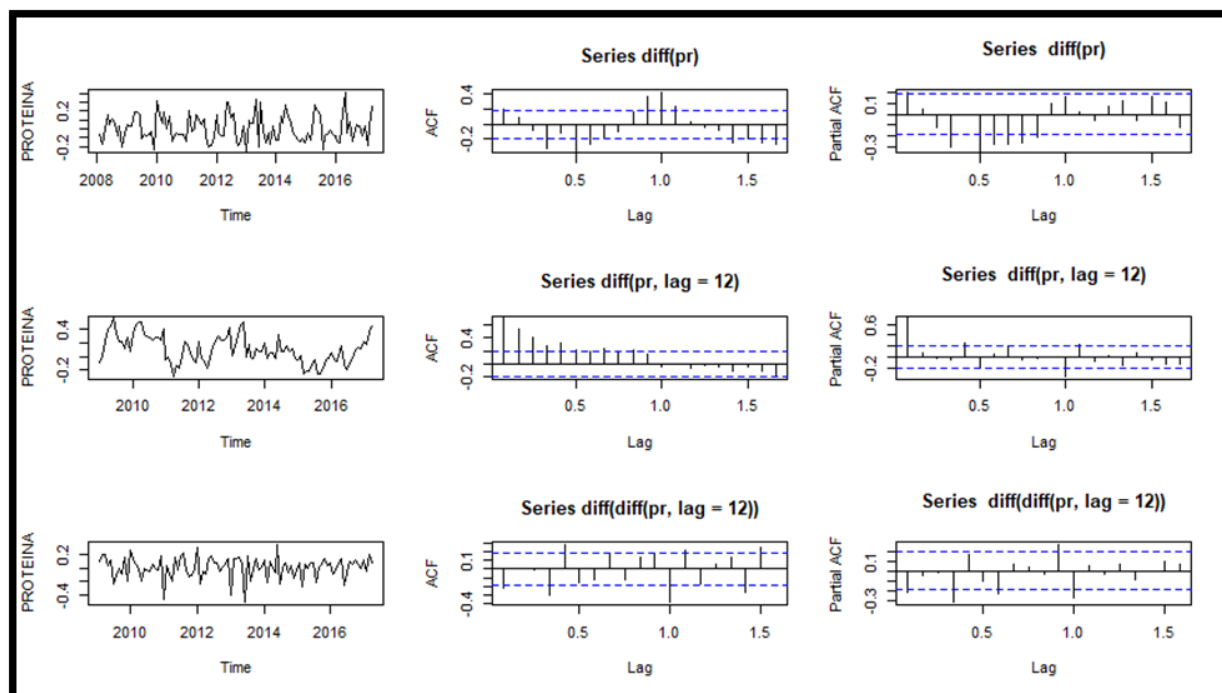


Figura 16 Análisis de serie diferenciada Porcentaje de Proteína de Leche de Búfala

Tabla 9 Estimativos para el modelo ARIMA (3, 0, 2) (0, 1, 1)₁₂

Parámetro	φ_1	φ_2	φ_3	θ_1	θ_2	Θ_1
Estimación	0.3205	-0.2609	0.7086	0.6125	1.0000	0.6328
Error-Estándar	0.0865	0.0870	0.0766	0.0302	0.0472	0.1223
Estadístico t	3.704271	-3.000849	9.253767	20.296453	21.203475	-5.173067
BIC				-101.76		
σ^2				0.01472		

Fuente: esta investigación

De acuerdo a los resultados de la prueba de significancia de parámetros, se puede afirmar con una confianza del 95% que el parámetro del modelo ARIMA(3,0,2)(0,1,1)[12] es significativo.

Comprobación/evaluación de supuestos: escogido el modelo ARIMA(3,0,2)(0,1,1)[12] para el parámetro de porcentaje de proteína, se procede a evaluar los supuestos de los residuos del modelo, en la figura 17 se observa la forma gráfica de representar cada uno de estos supuestos y se observan residuos estandarizados cuyas graficas de ACF y PACF se encuentra dentro de los límites de confianza. Los estadísticos de estadístico de Ljung-Box y Jarque Bera arrojan los siguientes resultados:

Box-Ljung test

X-squared = 11.347, df = 12, p-value = 0.1811

X-squared = 16.246, df = 24, p-value = 0.382

Jarque Bera Test

X-squared = 6.1847, df = 2, p-value = 0.04539

Los p-valor de los estadísticos estudiados confirman con una confianza del 95% que no existe correlación serial entre los residuos del modelo al no rechazar la hipótesis nula en el test de Ljung-Box y el estadístico de Jarque Bera ayuda a concluir que los residuos no son normales al rechazar la hipótesis nula de este test; aunque no se cumple el criterio de normalidad se escoge este modelo debido a que no se detectaron modelos que se ajustaran de mejor manera a la base de datos estudiada.

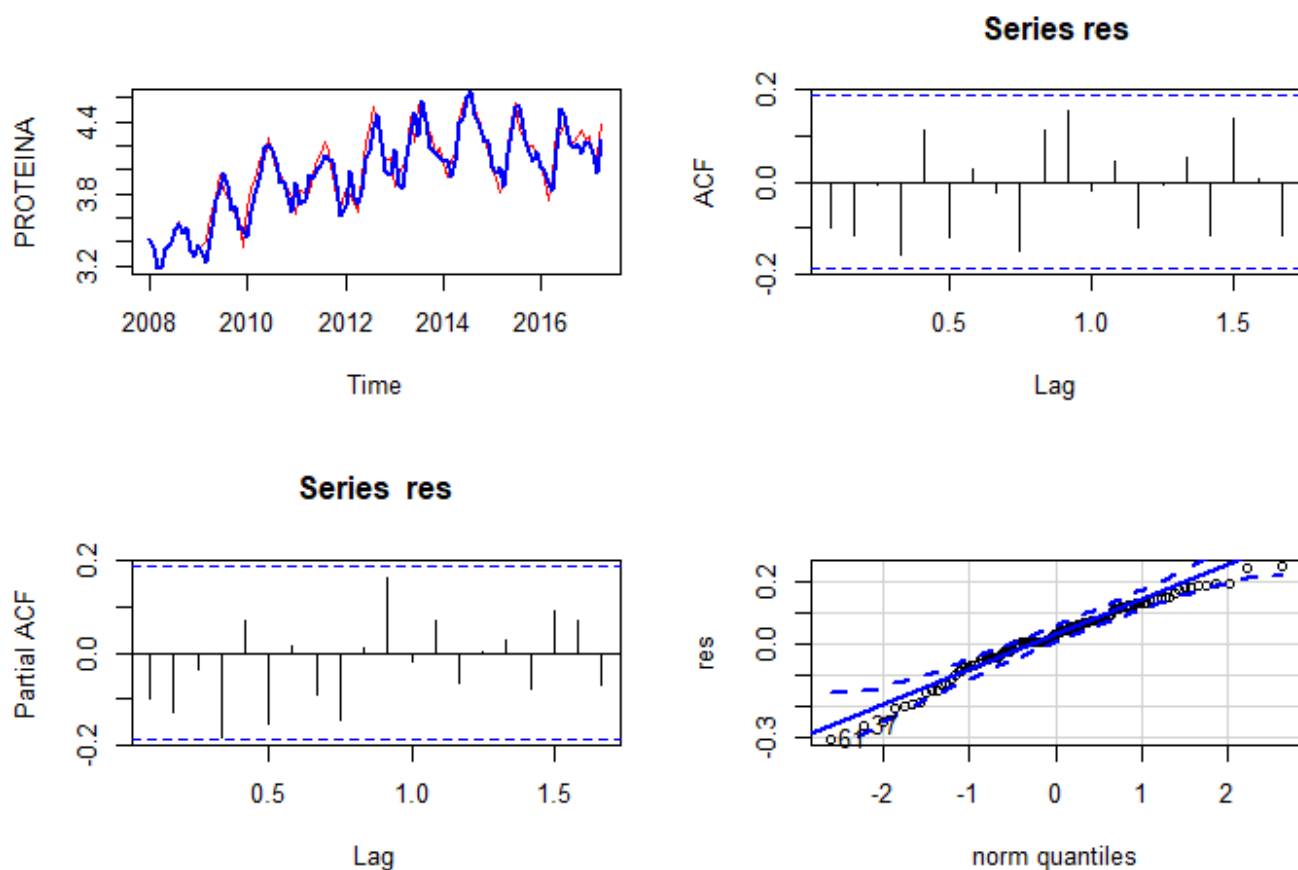


Figura 17 Evaluación de supuestos del modelo ARIMA (3, 0,2) x (0, 1, 1)₁₂

Pronóstico: en la figura 18 se puede observar el pronóstico fuera de muestra de del parámetro proteína en leche de búfala, la gráfica se encuentra dentro de los límites de aceptación sin embargo se observa una tendencia constante en los residuos del modelo, el error medio absoluto porcentual arrojó 2,75% para el modelo escogido lo cual es un porcentaje bastante alto para el modelo sin embargo el estadístico de Theil'S U arrojó 0.927 lo que muestra que el modelo es bueno pero no es el que mejor puede pronosticar esta variable. Se considera que los cambios en las tecnologías del equipo que analiza proteína, pudo originar que la base de datos no haya sido de fácil manejo y sea de muy difícil modelación. Se prevé de acuerdo al modelo ajustado que para el año 2019 el porcentaje de proteína en la leche de búfala se encuentre entre

4,026% y 4,601, estos resultados están por encima de la proteína de la leche de vaca pero normarles en la de búfala y están acordes a los históricos ya que para el año 2018 está varió entre 4,12% y 4,63%.

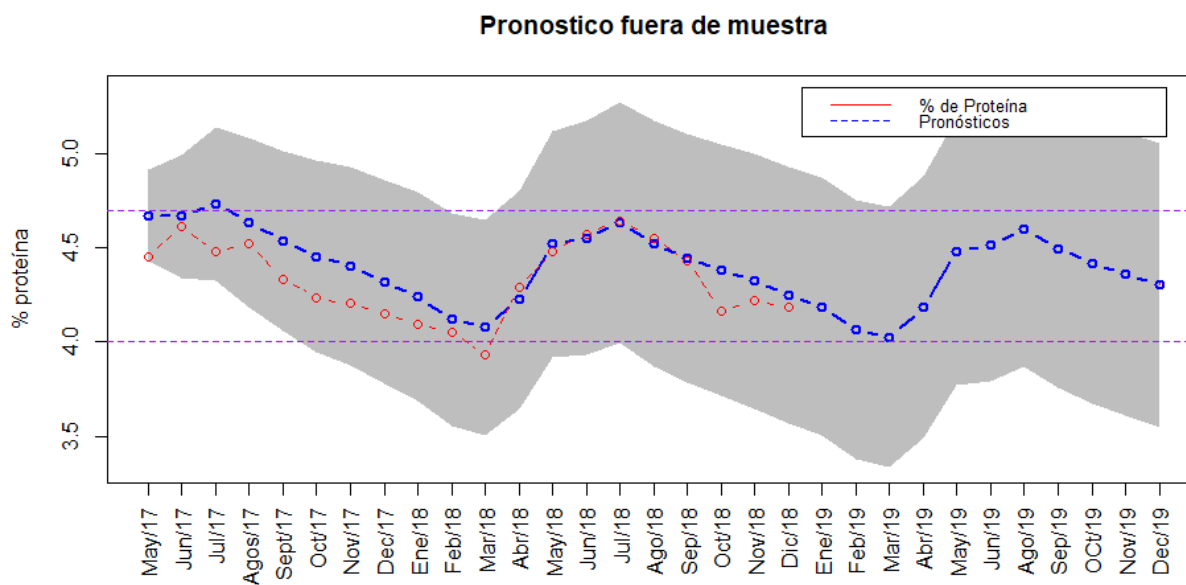


Figura 18 Pronóstico fuera de muestra del modelo ARIMA (3, 0,2) x (1, 0, 1)₁₂ para la porcentaje de Proteína de la leche de búfala

6. Conclusiones y recomendaciones

Se logra modelar algunos de estos parámetros para la leche de búfala (Porcentaje de acidez, densidad, porcentaje de grasa y proteína) utilizando modelos tipo ARIMA con periodos estacionales de 12 meses, logrando pronósticos dentro de las tolerancias de los métodos, con estos resultados se pueden levantar parámetros de aceptación y rechazo de la leche de búfala para trabajo en planta y como soporte a cambios al decreto 616 del 2006, principalmente en porcentaje de grasa , acidez y densidad que son las especificaciones legisladas y muy diferentes a la leche bovina.

Con el modelo ARIMA (1, 0,0) x (0, 1, 1)₁₂ se logra pronosticar el porcentaje de acidez en la leche de búfala prediciendo que para el año 2019 el porcentaje de acidez se mantendrá entre 0,1295% y 0,1462% con un error de 1,75%, para los primeros 4 meses del año se presentaran los porcentajes más bajos de acidez debido a la época de verano y a partir de mayo hasta agosto el porcentaje de acidez mayor por la época de lluvias, se considera el error del pronóstico adecuado para este parámetro ya que para el porcentaje más alto no supera el 0,03%, el estadístico de Theil's U para los 18 resultados analizados por fuera de muestra (de Julio-2017 hasta diciembre-2018) arrojó un resultado 0,85.

El modelo ARIMA (0, 1,0) x (0, 1, 1)₁₂ logra pronosticar el parámetro de densidad en leche de búfala con un error del 0,03417 % , se prevé que la densidad de esta leche para el 2019 tenga un promedio de 1,03402g/ml con un mínimo de 1,03323 g/ml y un máximo de 1,0348 g/ml, Se considera una estimación acertada teniendo en cuenta que en el 2017 y 2018 la densidad de esta leche se mantuvo entre 1,0326 g/ml y 1,053 g/ml.

El modelo ARIMA (1, 0,0) x (0, 1, 1)₁₂ fue el que presentó mejor ajuste para el porcentaje de grasa en leche de búfala con un error medio porcentual de 0,85%, se pronostica que para el año 2019 el porcentaje de grasa en leche de búfala se encuentre entre 7,36% y 8,04%, los cuales son resultados acordes a los históricos de esta leche, el estadístico de Theil's U para 20 resultados analizados por fuera de muestra (de mayo-2017 hasta diciembre-2018) arrojó un resultado 0,52.

Se logra ajustar los datos de porcentaje de proteína al modelo ARIMA (3, 0,2) x (1, 0, 1)₁₂ el cual muestra un error medio porcentual de 2,75% el cual es alto para este parámetro sin embargo con este error se tienen pronósticos con menos de 0,2% entre el valor verdadero lo cual es útil para un primer análisis. Para el año 2019 se pronostica que el porcentaje de proteína para la leche de búfala en la zona de Planeta rica, departamento de Córdoba se encuentre entre 4,026% y 4,601, estos resultados están por encima de la proteína de la leche de vaca pero normarles en la de búfala.

Los modelos ajustados resultan acertados para los pronósticos de los parámetros estudiados. Sin embargo se abre la posibilidad de ajustar un solo modelo para todas las características de composicionales y fisicoquímicas de la leche de búfala, a través de las series de tiempo multivariada y se recomienda estudiar las variables que se dejaron por fuera en este estudio y enlazar hasta donde sea posible la incidencia de las condiciones ambientales a la calidad en la leche, así como la de otras variables.

7. Bibliografía

- Andres Pablo Mogni. (2013). *Modelos de Series de Tiempo con aplicaciones en la industria aerocomercial*. UNIVERSIDAD DE BUENOS AIRES.
- Arce, R., & Mahía, R. (2010). *Modelos arima*. U.D.I. *Econometría e Informática*. Retrieved from http://www.uam.es/personal_pdi/economicas/eva/ecoi.html
- Eraso, J. K. S. (2013). *RELACIÓN ENTRE LOS MODELOS DE SERIES DE TIEMPO UNIVARIADAS CON LOS MODELOS DE SERIES MULTIVARIADAS*. UNIVERSIDAD DE NARIÑO. Retrieved from <http://mat.izt.uam.mx/mcmai/documentos/tesis/Gen.09-O/Vazquez-AL-Tesis.pdf>
- Gonzalez Bell, J. (2018). DURANTE 2017 LA PRODUCCIÓN DE LECHE DE BÚFALA FUE DE 174 MILLONES DE LITROS, 11,2% MÁS QUE EN 2016. *AGRONEGOCIOS*, p. 1. Retrieved from <https://www.agronegocios.co/agricultura>
- Hurtado, N., Cerón, M., Lopera, M., Bernal, A. y Cifuentes, T. (2005). Determinación de parámetros físico-químicos de leche bufalina en un sistema de producción orgánica. *Livestock Reserch for Rural Development*.
- ICA. (2017). Censo Agropecuario 2017. Retrieved from <https://www.ica.gov.co/Areas/Pecuaria/Servicios/Epidemiologia-Veterinaria/Censos-2016/Censo-2017.aspx>)
- IDEAM. (2018). BOLETIN CLIMATOLÓGICO MENSUAL.
- Mauricio, J. A. (2008). *Introducción al análisis de series temporales*. LADE y LE (Vol. 1).

Madrid, España. <https://doi.org/10.3989/mc.2004.v54.i274.233>

Ministerio de la protección social. (2006). DECRETO NÚMERO 616 DE 2006. Republica de Colombia.

Patiño, E. ., & Guanziroli Stefani, M. C. (2005). Composición de leche de búfala (*Bubalus bubalis*) de raza Jafarabadi en Corrientes, Argentina. *Revista Electrónica de Veterinaria REDVET - ISSN 1695-7504*, VI(5), 4. Retrieved from <http://www.veterinaria.org/revistas/redvet/exepa>

Patiño, E. M. (2011). Produccion y Calidad de la leche Bubalina. *Tecnologia En Marcha*, 24, 25–35.

Rios, G. (2008). Series de Tiempo. *Universidad de Chile. Facultad de Ciencias Fisicas y Matematicas*, 52. Retrieved from https://www.u-cursos.cl/ingenieria/2010/1/CC52A/1/material_docente/objeto/296003

Silvestre, J. L. C. i. (2001). *Macroeconometría Notas sobre la clase de contrastes de raíces unitarias*.

Soto, J. y Teherán, L. (2007). *Evaluación de la composición físico-química y microbiológica de la leche de búfala en las cinco fincas de mayor producción del departamento de Córdoba*. Universidad de Córdoba.

Suárez, L. S. (2013). *Pronóstico de la producción de leche, mediante modelos ARIMA. Caso UBPC “Maniabo”*. Universidad de La Habana.