ABSTRACT

In this research work was made the study of the demand for natural gas in Mexico, under a systemic approach. This approach allowed us to detect the most relevant variables of the system, which were included in the proposed mathematical models.

Two univariate models were built, one of statistical cut called ARIMA and one of artificial neural networks called NAR. Both models predicted the demand for natural gas in Mexico through the data included in the time series from 2006 to 2016; for the two models, downward forecasts were obtained, with ARIMA showing the best performance in the first two quarters of 2017, while the NAR model showed a better behavior for the second semester of the year.

Likewise, two multivariate models were used and compared, one of statistical type known as ARIMAX and one of artificial neural networks that takes by name NARX, these used explanatory variables, which affect the development of the demand for natural gas in the country. Both models made forecasts to the downside, the ARIMAX models produced better results than those shown by the univariate models; for its part, the NARX model was the one that made the biggest mistake of all the models developed in the present research work.

#### Índice General.

[Introducción VIII](#_TOC_250049)

[Índice de tablas XI](#_TOC_250048)

Índice de figuras XII

Acrónimos XIV

[Glosario XVI](#_TOC_250047)

[Capítulo 1. Contexto y Fundamentos de la investigación 1](#_TOC_250046)

* 1. [Marco Contextual 2](#_TOC_250045)
     1. Contexto Temporal 2
     2. Contexto Físico y Cultura 4
        1. [La Demanda de Gas Natural en el Mundo 5](#_TOC_250044)
        2. [La Demanda de Gas Natural en México 5](#_TOC_250043)
        3. La Demanda de Gas Natural Regional en México 9
  2. [Estado del Arte 12](#_TOC_250042)
  3. [Justificación 20](#_TOC_250041)
  4. [Objetivos 22](#_TOC_250040)
     1. Objetivo General 22
     2. Objetivos Particulares 22
     3. [Tabla de Congruencia de los Objetivos del Trabajo de Tesis 23](#_TOC_250039)

[Capítulo 2. Marco Teórico y Metodológico 24](#_TOC_250038)

2.1 Marco Teórico 25

* + 1. [Teorías Utilizadas 25](#_TOC_250037)
       1. [Sistemas 26](#_TOC_250036)
       2. [Ciencias Económicas 27](#_TOC_250035)
          1. [La formación de la Demanda de Gas Natural 28](#_TOC_250034)
       3. [Redes neuronales artificiales 29](#_TOC_250033)
          1. [El perceptrón 33](#_TOC_250032)

. 2.2.1.3.2 El perceptrón multicapa 34

* + - * 1. [El Algoritmo de aprendizaje de Retropropagación 35](#_TOC_250031)
        2. [Algoritmo de aprendizaje Levenberg-Marquardt 36](#_TOC_250030)
        3. Red Neuronal NAR 38
        4. Red Neuronal NARX 39
      1. [Estadística 41](#_TOC_250029)
      2. [Series de Tiempo y Pronósticos 41](#_TOC_250028)
         1. [Proceso autorregresivo (AR) 42](#_TOC_250027)
         2. [Proceso de media móvil (MA) 43](#_TOC_250026)
         3. [Proceso autorregresivo y de media móvil (ARMA) 43](#_TOC_250025)
         4. Proceso autorregresivo Integrado de media

móvil (ARIMA) 44

* + - * 1. Proceso autorregresivo Integrado de media móvil

y variables exógenas (ARIMAX) 50

* + - 1. [Simulación Matemática 53](#_TOC_250024)
         1. [Lenguaje de Programación R 53](#_TOC_250023)
         2. [MATLAB 54](#_TOC_250022)
         3. [Statgraphics 55](#_TOC_250021)
         4. Eview 56
  1. [Marco Metodológico 57](#_TOC_250020)
     1. [Metodología de Investigación de Operaciones 57](#_TOC_250019)

[Capítulo 3. Aplicación de la Metodología 60](#_TOC_250018)

* 1. [Definición del problema 61](#_TOC_250017)
  2. [Modelación 63](#_TOC_250016)
     1. Modelación del ARIMA 63
     2. Modelación del ARIMAX ARIMA 73
     3. Modelación del NAR 78
     4. Modelación del NARX 78
  3. [Simulación 79](#_TOC_250015)
     1. [Simulación del Modelo ARIMA 79](#_TOC_250014)
     2. [Simulación del Modelo ARIMAX 83](#_TOC_250013)
     3. [Simulación del Modelo NAR 86](#_TOC_250012)
     4. [Simulación del Modelo NARX 90](#_TOC_250011)
  4. [Implementación 94](#_TOC_250010)
     1. [Implementación del Modelo ARIMA 94](#_TOC_250009)
     2. [Implementación del Modelo ARIMAX 95](#_TOC_250008)
     3. [Implementación del Modelo NAR 97](#_TOC_250007)
     4. [Implementación del Modelo NARX 98](#_TOC_250006)
  5. [Validación 100](#_TOC_250005)
     1. [Validación del Modelo ARIMA 100](#_TOC_250004)
     2. [Validación del Modelo ARIMAX 103](#_TOC_250003)
     3. [Validación del Modelo NAR 106](#_TOC_250002)
     4. [Validación del Modelo NARX 107](#_TOC_250001)

Capítulo 4. Discusión de resultados 108

[4.1. Contraste de Resultados de los Modelos 109](#_TOC_250000)

Conclusiones…………………………………………………………………………

Futuros Trabajos…………………………………………………………………….

Referencias…………………………………………………………………………..

Anexos XIX

Anexo A: Base de Datos XX

Anexo B: Script en R del modelo ARIMA XXVII

Anexo C: Script en R del modelo ARIMAX XXIX

Anexos D: Script en MATLAB del modelo NAR XXXVIII

Anexo E: Script en MATLAB del modelo NARX VIIII

#### Introducción.

La investigación realizada en esta tesis se centra en el estudio de la serie temporal de la demanda en Millones de pies cúbicos diarios (MMpcd) de gas natural en México, por medio de redes neuronales artificiales (RNA) , modelos estadísticos (ARIMA), la teoría económica y la Sistémica.

El gas natural es la segunda fuente de energía más importante del país (PEMEX. 2016), por lo que este trabajo es útil para los múltiples agentes económicos que lo explotan, distribuyen o consumen, ya que con el conocimiento aquí generado, se podrán construir modelos matemáticos adecuados y será posible interpretar económicamente las variables involucradas, así se detectaran las ventajas y desventajas que tienen cada uno de los modelos.

La serie histórica de datos a analizar se circunscribe en el periodo 2006-2016, en intervalos mensuales, la cual se obtuvo de la base de datos institucional de Petróleos Mexicanos (PEMEX, et al. 2016). En la Figura 1.1 se muestran los históricos del gas natural desde el año 2006 al año 2016:

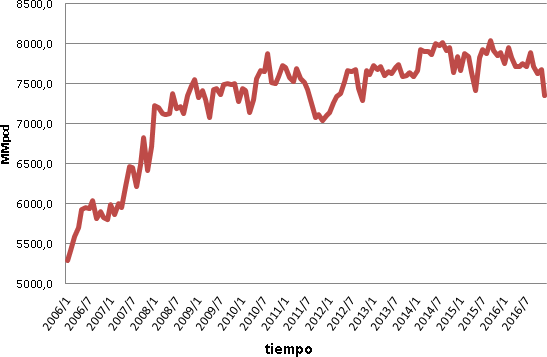


Figura 0.1 Históricos de la demanda de gas natural del periodo 2006-2016 Fuente: Elaboración propia a partir de los datos de PEMEX (PEMEX, 2015)

Los datos históricos de la demanda de gas natural proyectados gráficamente muestran que existe una tendencia creciente, pero con marcadas fluctuaciones en intervalos temporales pequeños, por lo que se deben de usar distintos enfoques para estudiarla y así obtener el mejor modelo matemático para su pronóstico.

El trabajo de tesis se encuentra estructurado de la siguiente manera:

En el **Capítulo I** se delimita el contexto físico , temporal y cultural , para entender en donde se sitúa el sistema de estudio , cual lapso de tiempo es el de interés y las características culturales que inciden en su desarrollo , con lo que es posible detectar los elementos más importantes que lo forman , así como las interacciones que estos guardan.

Se da un panorama general de cómo funciona el mercado del gas natural, centrando el estudio en el caso mexicano, para dilucidar las particularidades de esté.

De igual forma, se da un marco de referencia sobre distintas investigaciones realizadas para pronosticar la demanda del gas natural, especificando los métodos que se emplearon y el razonamiento que guió los estudios

Con los anterior se crea la base para construir la justificación del trabajo de tesis , así como el objetivo general, los objetivos particulares, preguntas de investigación e hipótesis.

El **Capítulo II** contiene las teorías que son empleadas para sustentar el trabajo de tesis , haciendo énfasis en la de redes neuronales artificiales , los modelos estadísticos ARIMA , la teoría de series de tiempo y la sistémica.

Asimismo, se hace mención de la metodología de investigación que se empleó para el desarrollo de la investigación , especificando las fases que tiene, así como las actividades a desarrollar

En el **Capítulo III** se aplicó la metodología para construir los modelos univariados y multivariados, la cual tiene como fases: la definición del problema, en la cual se determina de manera concreta el problema que se ha de abordar; la modelación del problema, donde se determinan las variables que tendrá cada uno de nuestros modelos, poniéndolos en términos adecuados dentro del lenguaje matemático;

simulación del modelo, en la cual se ejecutan computacionalmente los modelos para poder obtener resultados de cada uno de ellos; la validación implica corroborar que el modelo construido es útil para resolver el problema de estudio; y la implementación , en esta fase se transforman los elementos contenidos en el modelo en instrucciones claras de operación, para lograr su reproducción sistemática

En el **Capítulo IV** se llevó a cabo la comparación de resultados que tuvieron los modelos estadísticos univariados y multivariados, versus los resultados obtenidos por las redes neuronales artificiales correspondientes. Además se exponen las conclusiones a las que se llegaron, identificando qué modelo matemático tuvo mayor precisión para abordar nuestro objeto de estudio y las dificultades que se presentaron para la construcción e implementación de los modelos; como punto final se emitieron algunas recomendaciones sobre los trabajos futuros a realizar.

##### Índice de Tablas.

**Tabla 1.1** Características representativas de México 6

**Tabla 2.1**. Trabajos representativos sobre el pronóstico de gas natural en el mundo 16

**Tabla 2.2.** Trabajos más representativos sobre el pronóstico de gas natural en México 19

**Tabla 2.3.** Tabla de congruencia de los objetivos del trabajo de Tesis 23

**Tabla 3.1**. Datos que muestran los valores de la pendiente e intercepto de la regresión lineal 64

**Tabla 3.2**. Datos que muestran los valores de la pendiente e intercepto de la regresión lineal 65

**Tabla 3.3.** Datos que muestran los valores de la pendiente e intercepto de la regresión lineal 66

**Tabla 3.4.** Datos que muestran los P-valores para la prueba de White 67

**Tabla 3.5**. Datos que muestran los P-valores para la prueba de raíz unitaria 68

**Tabla 3.6.** Datos que muestran los valores para la prueba Durbin Watson 69

**Tabla 3.7.** Se muestra la probabilidad del análisis de la varianza 71

**Tabla 3.7**. Datos que muestran los valores para la AC y FCA, de la serie de tiempo de gas natural 72

**Tabla 3.9.** Datos que muestran si existe o no una relación estadísticamente significativa

entre las variables contempladas 75

**Tabla 3.10**. Datos que muestran si existe o no una relación estadísticamente

significativa entre las variables contemplada 75

**Tabla 3.11.** Prueba de Raíz Unitaria 76

**Tabla 3.12.** Resultados obtenidos por los modelos ARIMA a considerar 80

**Tabla 3.13.** Resultados obtenidos por los modelos ARIMAX a considerar 83

**Tabla 3.14.** Pronostico hecho por los dos modelos para el periodo 2017-1 al 2017-12 94

**Tabla 3.15**. Pronóstico hecho por los dos modelos para el periodo 2017-1 al 2017-2 96

**Tabla 3.16**. Pronóstico de la demanda de gas natural para el 2017 de la

red neuronal NAR 97

**Tabla 3.17.** Pronóstico de la demanda de gas natural para el 2017 de la red

neuronal NARX 98

**Tabla 4.1.** Pronóstico de la demanda de gas natural para el 2017 de los modelos ARIMA, ARIMAX , NAR y NARX 109

**Tabla 4.2** MAPE en el pronóstico de la demanda de gas natural de los modelos ARIMA,

ARIMAX , NAR y NARX 113

**Tabla 4.3**. MAPE en el pronóstico de la demanda de gas natural por trimestre 114

##### Índice de Figuras

**Figura 0.2** Históricos de la demanda de gas natural del periodo 2006-2016 VIII

**Figura 1.1.** Localización geográfica de México 6

**Figura 1.2** División de México por Regiones energéticas 10

**Figura 1.3** Comportamiento del consumo de gas natural por región en MMpcd

del periodo 2006-2016 10

**Figura 1.4** Precio promedio mensual en dólares de un MMbtu de gas natural en Japón,

EUA y Europa, periodo 2006-2016 13

**Figura 1.5** Comportamiento del precio mensual en dólares de un MMbtu de gas natural

en Japón, EUA y Europa, del periodo 2017-2025………………………………………, 13

**Figura 2.1** Holos con las teorías que dan sustento a la presente investigación 25

**Figura 2.2** Modelo grafico de una neurona artificial 30

**Figura 2.3** Modelo grafico de una neurona con *P* entradas 30

**Figura 2.4** Modelo grafico de 𝑆 neurona con 𝑅 entradas 31

**Figura 2.5** Modelo gráfico de las funciones de transferencia *hardlim*, *purelin*, *logsig* y *tansig* 31

**Figura 2.6** Modelo grafico de una red neuronal con tres capas 32

**Figura 2.7**. Modelo grafico de la red preceptrón 33

**Figura 2.8**. Modelo grafico de una red preceptrón multicapa 34

**Figura 2.9** Modelo gráfico de una red neuronal NAR 38

**Figura 2.10** Modelo gráfico de una red neuronal NARX 40

**Figura 2.11** Correlogramas simple y parcial de una serie de tiempo que tiene un

fuerte componente estacional anual 49

**Figura 2.12** Diagrama de flujo de la metodología de Investigación de Operaciones 57

**Figura 3.1**. Datos de PEMEX de la demanda promedio mensual en MMpcd de gas

natural en México. periodo 2006-2016 63

**Figura 3.2** Serie de la demanda de gas natural en México en primera diferencia 65

**Figura 3.3** Serie de la demanda de gas natural en México en segunda diferencia 67

**Figura 3.4** Correlogramas simple y parcial de la serie de tiempo de la demanda del

gas natural en México en segunda diferencia 70

**Figura 3.5** Correlogramas simple y parcial de la serie de tiempo de la demanda del

gas natural en México segunda diferencia 72

**Figura 3.6** Datos que muestran los valores para la AC y FCA, de la serie de tiempo

de gas natural con variables explicativas 77

**Figura 3.7** Comportamiento real de la demanda versus los valores ajustados para los tres modelos ARIMA 81

**Figura 3.8** Comportamiento real de la demanda versus los valores pronosticados

por los dos modelos ARIMA 82

**Figura 3.9** Comportamiento real de la demanda versus los valores ajustados por

los dos modelos ARIMAX 84

**Figura 3.10** Comportamiento real de la demanda versus los valores pronosticados

por los tres modelos ARIMAX 84

**Figura 3.11** Modelo gráfico y resultados de la simulaciones de una red NAR con

cinco neuronas en la primera capa oculta , una en la segunda y una en la capa de salida 86

**Figura 3.12** Ajuste que realiza la red NAR a los datos de entrenamiento, y los pronósticos

hechos para los datos de validación y test 87

**Figura 3.13** Regresión de los datos de entrenamiento, validación y test 88

**Figura 3.14** Grafica que muestra el mejor desempeño de validación a través de las

iteraciones del entrenamiento 89

**Figura 3.15** Modelo gráfico y resultados de la simulaciones de una red NARX con

doce neuronas en la primera capa oculta y una en la capa de salida 90

**Figura 3.16** Ajuste que realiza la red NARX a los datos de entrenamiento, y los

pronósticos hechos para los datos de validación y test 90

**Figura 3.17** Regresión de los datos de entrenamiento, validación y test 92

**Figura 3.18** Grafica que muestra el mejor desempeño de validación a través de las

iteraciones del entrenamiento 93

**Figura 3.19** Pronóstico hecho por los tres modelos ARIMA para el periodo 2017-1 al 2017-2 95

**Figura 3.20** Pronóstico hecho por los tres modelos ARIMAX para el periodo 2017-1 al 2017-2 96

**Figura 3.21** Pronóstico NAR de la demanda de gas natural del 1/2017 al 12/2017 97

**Figura 3.22** Pronóstico NARX de la demanda de gas natural del 1/2017 al 12/2017 99

**Figura 3.23** Frecuencia de los residuos del modelo ARIMA(1,2,1) 102

**Figura 3.24** Frecuencia de los residuos del modelo ARIMA(1,2,10) 102

**Figura 3.25** Frecuencia de los residuos del modelo ARIMAX(1,2,1) 105

**Figura 3.26** Frecuencia de los residuos del modelo ARIMA(12,2,10) 105

**Figura 3.27** Las autocorrelaciones de los residuos en los diferentes rezagos 102

**Figura 3.28** Autocorrelaciones de los residuos en los diferentes rezagos 106

**Figura 4.1** Pronóstico de la demanda de gas natural del 1/2017 al 12/2017

de los modelos ARIMA 110

**Figura 4.2** Pronóstico de la demanda de gas natural del 1/2017 al 12/2017 de los

modelos ARIMAX 111

**Figura 4.3** Pronóstico de la demanda de gas natural del 1/2017 al 12/2017 de los

modelos NAR Y NARX 112

**Acrónimos AIE:** Administración de Información Energética. **ARIMA:** [Autorregresivo Integrado de Media Móvil](https://es.wikipedia.org/wiki/Modelo_autorregresivo_integrado_de_media_m%C3%B3vil).

**ARIMAX:** Autorregresivo Integrado de Media Móvil con variables explicativas.

**BANXICO:** Banco de México.

**BM:** Banco Mundial.

**BTU:** British Thermal Unit.

**CIDAC:** Centro de Investigación para el Desarrollo A.C

**CFE:** Comisión Federal de Electricidad.

**CNA:** Consumo Nacional Aparente.

**CRE:** Comisión Reguladora de Energía**.**

**CONEVAL:** Consejo Nacional de Evaluación de la Política de Desarrollo Social.

**INEGI:** Instituto Nacional de Estadística, Geografía e Informática.

**MATLAB:** MATrix LABoratory

**NAR:** Modelo Autorregresivo no Lineal.

**NARX:** Modelo Autorregresivo no Lineal con variables exógenas.

**ONU:** Organización de las Naciones Unidas.

**PAN:** Partido de Acción Nacional.

**PEMEX:** Petróleos Mexicanos.

**PRI:** Partido Revolucionario Institucional.

**PIB:** Producto Interno Bruto. **RNA:** Red Neuronal Artificial. **SENER:** Secretaria de Energía

**SNG:** Sistema Nacional de Gasoductos.

**UNESCO**: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

**WTI:** West Texas Intermédiate.

#### Glosario

**Bienes complementarios:** Son mercancías relacionadas entre sí, donde el aumento en el precio de uno disminuye la demanda del otro (Mankiw, 2012).

**Bienes sustitutos:** Son productos que pueden ser consumidos en lugar del otro, satisfaciendo la misma necesidad. El aumento del precio de uno incrementa la demanda del otro (Mankiw, 2012).

**Consumo Nacional Aparente:** Cantidad de producto que es adquirido en por los habitantes de un país para satisfacer alguna necesidad (Mankiw, 2012).

**Demanda:** Cantidad total de mercancías que están dispuestas a adquirir los consumidores a un nivel de precios dado (Mankiw, 2012).

**Estacional:** Que tiene un comportamiento repetitivo cada cierto lapso de tiempo (Gujarati, 2010).

**Estacionario:** Característica de un sistema de permanecer invariante a través del tiempo (Gujarati, 2010).

**Expectativas económicas:** Parecer de una persona o entidad sobre lo que sucederá en un futuro con su ingreso y las circunstancias económicas (Mankiw, 2012).

**Gas natural**: Es un combustible fósil no renovable. Se le considera un combustible fósil debido a que se formó a partir de restos de animales marinos y plantas que murieron hace 300 a 400 millones de años (EIA, 1999).

**Gas natural bruto**: Es una mezcla de diferentes gases, el ingrediente principal es el metano, el cual tiene una participación del 90% en su composición, el resto es una mezcla de etano, propano, butano, pentano, dióxido de carbono y nitrógeno (EIA, 1999).

**Hidrocarburo:** Compuestos orgánicos que tienen en sus moléculas solo carbono e hidrogeno (EIA, 1999).

**Investigación de Operaciones:** Es la aplicación del método científico para construir modelos que tengan alternativas, criterio objetivo y restricciones (Martínez, 2006).

**Macroeconomía:** Área de las ciencias económicas que se encargan de estudiar a la economía de manera agregada (Mankiw, 2012).

**Matlab:** Es un paquete matemático utilizado para graficar , procesar , modelar y simular sistemas (Wiley, 2005).

**Modelo:** es la representación formal de algún sistema que involucra una actividad cognitiva, con el fin de entender un objeto de estudio y las particularidades que este tiene (Johansen, 2004).

**Modelo Matemático:** Es la concepción, en términos matemáticos, de un sistema y su comportamiento (Gujarati, 2010).

**Método científico:** Es un esquema de experimentación que se emplea para realizar observaciones y contestar preguntas, con lo que se ubicaran relaciones de causa y efecto del objeto de estudio. Tiene como secuencia lógica la concepción de preguntas, la formulación de hipótesis, la prueba de hipótesis, el análisis de resultados, la estructuración de conclusiones y la publicación de resultados (Johansen, 2004).

**Metodología de Investigación**: Es una forma de resolver sistemáticamente los problemas que han sido planteados en la investigación, estudiando los diferentes pasos que sigue un investigador en el análisis de su objeto de estudio, así como la lógica implícita (Martinez, 2006)

**Oferta:** Cantidad total de bienes y/o servicios que están dispuestos a vender los productores a un precio establecido (Mankiw, 2012).

**Precio:** Valor monetario de una mercancía (Mankiw, 2012).

**Precio constante:** Son aquellos que no consideran el efecto de la inflación (Mankiw, 2012).

**Producto Interno Bruto:** Es el valor de los bienes y servicios producidos por una economía en un intervalo de tiempo (Mankiw, 2012).

**Prueba de Hipótesis:** Proceso para determinar la validez de una aseveración hecha sobre la población y basándose en evidencia muestral (Isotalo, 2001).

**CAPÍTULO 1**

**Contexto y Fundamentos de la investigación**

#### Capítulo 1. Contexto y Fundamentos de la investigación

#### 1.1 Marco Contextual.

En este apartado se describe el contexto temporal, físico y cultural de la problemática de estudio, para así entender el escenario en el que se encuentra inmerso nuestro sistema.

##### 1.1.2 Contexto Temporal.

El periodo de estudio de esta investigación va del año 2006 al 2016, década que ha tenido cambios importantes en el escenario político nacional con la alternancia del poder entre el Partido de Acción Nacional (PAN) y el Partido Revolucionario Institucional (PRI), lo cual ha repercutido en los modelos de desarrollo, tanto económicos como sociales.

Del periodo 2006-2012 Felipe Calderón Hinojosa ostenta el poder presidencial; existe un poder legislativo dividido entre los distintos partidos políticos, siendo mayoritario el PAN, lo que hizo que el presidente en turno tuviera gran apoyo en las decisiones tomadas. Esto llevó a la reformación del sistema de seguridad nacional (Ortiz, 2015), para afrontar a los grupos del crimen organizado, lo cual fue el centro de atención de la administración pública, tanto federal, como estatal y municipal.

En el escenario económico se pudo controlar la inflación a través de una política monetaria conservadora y una disciplina fiscal, siendo la más baja de los últimos seis sexenios, con un promedio anual de 2.45% (Gómez, 2016). Así mismo se alcanzó una notable acumulación de reservas internacionales y se mantuvo una deuda pública estable.

Por el lado del crecimiento económico se tuvo 1.9% de incremento en el PIB anual, el cual fue el más bajo de las dos décadas que le precedieron (Ortiz et al., 2015), lo que conllevó la disminución de oportunidades y el agravamiento de la

inequidad social; pues de las personas que poseían un trabajo, el 73.1% percibió ingresos de entre uno y cuatro salarios mínimos y la población desempleada aumentó en 56.25% (Gómez, 2016).

Calderón logró la aprobación de cinco reformas estructurales que fueron: la fiscal en el 2007, la del sistema público de pensiones en el 2007, la energética en el 2008, la de competencia en el 2010 y la laboral en el 2012 (Gutiérrez, 2014). Algunas de éstas no tuvieron el impacto esperado, pero la de materia energética dio la pauta para transformar las operaciones de Petróleos Mexicanos (PEMEX) y la Secretaria de Energía (SENER), al dotarles de mayor autonomía y por la creación de diferentes órganos regulatorios, con el fin de impulsar la explotación energética sustentable y la colocación competitiva de la energía en el mercado nacional, lo cual devino en el agrandamiento de la burocracia y la escalada del déficit de las empresas paraestatales involucradas (Ortiz, 2015).

Con las elecciones del 2012 se da una alternancia en el poder político, al ser electo Enrique Peña Nieto, miembro del PRI. La Cámara de diputados quedó conformada en mayoría relativa por miembros del mismo partido del presidente (SIID, 2013), dando la pauta para la aprobación de una serie de reformas estructurales de gran magnitud, entre las que destacan: la reforma energética, educativa, hacendaria, laboral, electoral y penal.

La reforma energética fue aprobada en el 2013, la cual tuvo como objetivos el fortalecimiento de las empresas productivas del Estado, reduciendo la corrupción, incentivando la competitividad y atrayendo capital privado para las labores de exploración y explotación petrolera (FCCT, 2013).

Con la reforma se logra la liberación del precio del gas natural en el 2017, el cual había estado bajo el dominio de la Comisión Reguladora de Energía (CRE), institución que se abocó a limitar el poder monopólico que ostentaba PEMEX en las ventas de primera mano de este energético. Con la apertura en la explotación del gas natural, se logra que exista un mercado competitivo con múltiples

oferentes, sujetos a las fuerzas de la oferta y la demanda, las cuales fijan el nivel de equilibrio de consumo a un precio determinado.

En el periodo 2012-2016 el crecimiento promedio anual del PIB fue de 2,48%, mientras que la inflación se situó en 3,42% (INEGI, 2017). Según datos de la CONEVAL, el porcentaje de pobres con respecto a la población total en el 2012 fue de 45.5%, mientras que para el 2016 representó el 43.6%. La tasa de desempleo se mantuvo con un promedio anual de 4.61%, pasando de una tasa de 4.9% a 3.93% del 2012 al 2016.

##### Contexto Físico y Cultural

El espacio físico en el que se desenvuelve el sistema de gas natural, tiene gran influencia en la dinámica y en la manera en que se desarrolla a lo largo del tiempo. La cultura que tiene la sociedad que ocupa el espacio físico incide de manera directa en la manera que se consumen las fuentes energéticas y las preferencias que se tienen hacia cada una de ellas.

La cultura es un proceso continuo de fundamentación de la identidad de una sociedad, grupo o población, donde existen rasgos sociales regidos por una ideología imperante, una estructura de preceptos morales y un estilo de vida que se proyecta en usos y costumbres (Altiere, 2001).

La cultura tiene un gran dinamismo, ya que las sociedades con el devenir del tiempo van cambiando y reformando sus tradiciones, las expresiones de arte, los valores, los objetivos que persiguen, la religión, el idioma, la vestimenta, etc. Las personas adoptan la cultura preponderante, ya que ésta posee mecanismos para ejercer control sobre ellos, diferenciando lo bueno y lo malo, al poner énfasis en la retribución que se tendrá por apegarse a los lineamientos aceptados por la colectividad.

##### La Demanda de Gas Natural en el Mundo.

En el año 2015 la demanda de gas natural a nivel mundial fue de 328,280.7 millones de pies cúbicos diarios (Mmpcd), 1.35% mayor a la registrada en el año 2012 (CFE, 2015). Este incremento se debe a que el gas natural está experimentando una transformación revolucionaria, gracias a los avances en la tecnología de perforación, la cual ha permitido explotar grandes volúmenes de gas a bajo costo, asegurando así el abastecimiento por varias décadas.

Actualmente, el gas natural está fungiendo como sustituto de otros combustibles fósiles, como lo es el carbón y el petróleo, ya que es menos contaminante y existen grandes reservas para su explotación. Las flotas de camiones y autobuses están regresando a su consumo, por las ventajas en el costo y en el impacto ambiental, en relación con la gasolina y el diésel. Incluso los sectores ferroviario y marítimo están operando buques y locomotoras que utilizan gas natural en lugar de gasóleo (Petrotecnia, 2009).

Por otra parte, el sector eléctrico a nivel mundial ha destinado una mayor inversión para la puesta en marcha de plantas que utilicen turbinas de gas, por ello este combustible fósil tuvo un 23.2% de participación en la generación total de energía eléctrica (2016, UNESCO).

El volumen de reservas probadas que hay a nivel mundial de gas natural ascienden a 6,606.4 billones de pies cúbicos (CFE, 2015), teniendo la capacidad de cubrir las necesidades energéticas de las sociedades contemporáneas en las siguientes tres décadas.

##### La Demanda de Gas Natural en México.

México es una república democrática, representativa y Federal, integrada por 32 entidades federativas, de las cuales la Ciudad de México es su capital. Limita al norte con los Estados Unidos, al oeste con el Océano Pacífico, al este con el Golfo de México, al sur con Guatemala y Belice. En la

[Tabla 1.1](#_bookmark0) 1.1 se compendian algunos datos relevantes de la república mexicana y en la Figura 1.1 se muestra su localización geográfica.

Tabla 1.1 Características representativas de México (INEGI, 2017).

Norteamérica

Región Latitud Altitud Superficie Población

Forma de Gobierno

1,964,375 km2

119 530 753 de habitantes República representativa, Democrática y federal

Figura 1.1. Localización geográfica de México. (INEGI, et al , 2017)

México es un país multicultural, con rasgos y costumbres heredadas de las sociedades prehispánicas, pero alimentadas y transformadas por la cultura española. Así mismo, la cultura al ser dinámica, ha sufrido una serie de cambios, por medio del sistema económico, político y social, siendo la globalización un motor de integración e intercambio de costumbres, prácticas e ideologías, que han permeado a la cultura mexicana.

El sistema económico capitalista ha logrado construir un aparato de valores que rigen a la sociedad mexicana, siendo el consumismo el punto nodal, el cual incentiva a las personas, empresas y gobierno a realizar actividades productivas, con el fin de tener los recursos necesarios para la satisfacción de sus necesidades, tanto básicas como suntuarias.

Para el caso de la explotación energética, México se ha centrado en fuentes de bajo costo con el fin de maximizar las ganancias de las empresas que las utilizan, y para proteger la economía de las familias y el gobierno. Pero éste no es el único factor que incide en la elección de una fuente energética, ya que los tomadores de decisiones en las entidades empresariales y gubernamentales, así como los individuos, basan muchas de sus elecciones en juicios de valor, prefiriendo la fuente energética que les parece más adecuada para el logro de sus objetivos y metas. Muchos de esos juicios se derivan de la experiencia que poseen o por la costumbre de haber usado siempre algún tipo de energía.

Las investigaciones legislativas (Jackson, 2014) del Senado de la República alertaron sobre el rechazo que hay hacia ciertas fuentes de energía, por el ideario que tienen los diferentes grupos sociales del país. Por ejemplo, en las zonas rurales se inclinan por el uso de la leña para cocinar, ya que se considera que con ésta la comida adquiere un mejor sabor; o simplemente porque el hecho de prender un fogón en el lugar donde se vive, representa un centro de reunión familiar para que los integrantes puedan interactuar.

Por tanto, se puede aseverar que las costumbres y valores de los grupos que hay en las diferentes regiones del país inciden en los hábitos de consumo del gas natural, donde las poblaciones, sociedades y empresas elegirán las fuentes energéticas que mejor se apeguen a la idiosincrasia que poseen. En el caso empresarial, habrá ejemplos de empresas que busquen utilizar energías renovables y con menor impacto al planeta, por la postura pro ambiental que tienen los dueños, accionistas o directores de las mismas, lo cual hará que se desplace la demanda de gas natural por alguna fuente de energía renovable que contamine menos.

En México se ha experimentado un incremento sustancial en la demanda de gas natural. Para el año 2006 se vendieron 2,633.912 millones de pies cúbicos diarios (mmpcd), cifra que pasó a 3,347.290 mmpcd para el 2016, lo cual representa un aumento del 21.08 % (CFE, 2016).

En el año 2012, el Sistema Nacional de Gasoductos (SNG) poseía una extensión de más de 11,342 kilómetros, situación que fue transformándose, pues para el 2015 la longitud total fue de alrededor de 19,052 kilómetros (CFE, 2015), lo cual representa un crecimiento del 68% de la red con la que cuenta el país.

La variación en la demanda de este energético se debe a varios factores, uno de ellos es el uso que se le está dando para la producción de electricidad, la cual ha ido en expansión. La CFE reestructuró siete centrales eléctricas para que usaran al gas natural como fuente de energía, representando una capacidad de 4,500 Megawatts (MW).

A su vez, se han hecho planes para expandir el número de centrales de turbo gas a un total de nueve, destinando una inversión de casi 4,500 millones de dólares y una capacidad instalada mayor a 6,200 MW (CFE, 2015).

En el año 2010 se generaban a base de gas natural 2,975 miles de millones watts hora (mmwh) y para el 2016 esta cifra ascendió a 5831.4, lo cual implica un incremento del 96.01% en solo seis años (INEGI, 2016).

Al hacer un análisis más profundo de la demanda nacional del gas natural, se observa que el sector eléctrico es el mayor consumidor, con una participación del 47.8% de la demanda total; el que le sigue es el sector petrolero, el cual abarca el 32.7%; el sector industrial representa el 17.8%, donde las empresas acereras, químicas y de alimentos son los mayores demandantes; el sector residencial tiene una participación de 1.2%; el sector servicios es de 0.4%; y por último el sector autotransporte consume el 0.03% (CFE, 2015).

Por otra parte, la oferta nacional de gas natural en el año 2016 fue de 5,792.482 mmpcd, representando un incremento del 20.02 % con respecto al año 2005. A pesar de lo anterior han habido periodos donde la oferta se contrajo, por ejemplo, para el 2009 se tuvo una oferta de 7326.68 mmpcd (CFE, 2015), lo cual

implicó que para el 2016 la producción se había reducido en un 17.6%, ocasionando un aumento en las importaciones.

La interacción de la oferta y la demanda ha causado un decremento del precio del gas natural. Para el 2008 el precio promedio fue de 8.900 dólares por mbtu, mientras que para el 2016 fue de 2.551 dólares (CFE, 2015), representando una caída del 244.88% de su valor.

Así mismo, las reservas totales de gas natural que posee el país han influido en las expectativas, tanto de los consumidores como de los productores, las cuales repercuten en la fijación del precio de esta fuente energética. Para el año 2014 se tenían 59,664.7 miles de millones de pies cúbicos (mmpc) de reservas, en el 2016 esta cifra disminuyó a 54,890 mmpc (CFE, 2015), lo cual significa un decremento del 8.1% en las reservas probadas que posee México; si el ritmo en el decrecimiento de las reservas se mantuviera constante, y no se descubrieran nuevos yacimientos, el país podría explotar este combustible fósil, al ritmo actual, durante las siguientes dos décadas, lo cual haría que conforme fuera bajando el volumen de las reservas, se irá incrementando el precio de éste, debido a que se volverá cada vez más escaso.

##### 1.1.3.3 La Demanda de Gas Natural Regional en México.

El país se encuentra seccionado en cinco regiones, las cuales se emplean para evaluar la demanda de gas natural, que son: Noroeste, Noreste, Centro-Occidente, Centro y Sur-Sureste. (SENER, 2016). En la Figura 1.2 se muestra la división del país en estas cinco regiones.

Los niveles de consumo del periodo 2006 al 2016 por región se muestra en la Figura 1.3:

3000.00

2500.00

2000.00

1500.00

1000.00

500.00

0.00



La región sureste ha tenido la mayor demanda de gas natural en el periodo de estudio, a excepción del año 2015, donde la región noroeste fue la que tuvo el consumo más elevado. Esto se debe en gran medida a que el sector eléctrico y petrolero en esas regiones tienen un papel estratégico y como se mencionó en el apartado anterior, son los que adquieren una mayor proporción de esta fuente energética.

Los estados que tuvieron la participación más importante en el consumo de este hidrocarburo en el año 2016 fueron: Tamaulipas con 956.1 MMpcd, Veracruz con 862.7 MMpcd, Nuevo León con 724.4 MMpcd y Tabasco con

694.4 MMpcd. El único que no tiene explotación petrolera de estos cuatro estados es Nuevo León, pero es la segunda ciudad más grande del país y ocupa el segundo lugar en el aporte al PIB nacional, lo que implica una alta demanda de fuentes energéticas que el gas natural ha suplido.

Cabe señalar que las cinco regiones han tenido una tendencia a la alza, las que tuvieron menor crecimiento, en la demanda de gas natural, fueron la zona sureste y noroeste con 6.04% y 7.51% respectivamente, donde el estado de Oaxaca y Zacatecas son los menores consumidores a nivel nacional con 0.1 MMpcd y 9.2 MMpcd respectivamente. Las demás regiones han experimentado un crecimiento vertiginoso en la demanda de gas natural, siendo la región centro Occidente la de mayor aumento con un 89.89%, le sigue el centro con 60.34% y el noreste con 54.32%; contrario a lo que se pudiera pensar la Ciudad de México no tiene un aporte considerable a la demanda de gas natural de su región, ya que solo adquiere 69.6 MMpcd.

##### Estado del Arte

El Banco Mundial ha realizado estudios sobre el pronóstico del precio del gas natural en Europa, Japón y EUA, ya que éstos son referentes para los precios del gas en la región que tienen influencia (Petrotecnica, 2009), es decir, a diferencia del petróleo, el precio del gas natural no se fija por la interacción de los agentes a nivel mundial, sino a nivel regional.

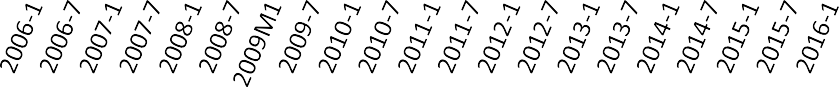
Los datos compendiados por el Banco Mundial (BM) se estudian a través del [Modelo Autorregresivo Integrado de Media Móvil](https://es.wikipedia.org/wiki/Modelo_autorregresivo_integrado_de_media_m%C3%B3vil) (ARIMA), el cual supone que el valor que puede adquirir una variable en un futuro puede ser pronosticada por la forma en que se ha venido comportado en el pasado, en otras palabras, se establece una serie de tiempo con el comportamiento de la variable, para poder hacer una modelación estadística de ella.

También se emplea el método Generalized Autorregresive Conditional Heteroscedasticity (GARCH) para analizar y estudiar la volatilidad en los precios del gas natural (World Bank, 2016), lo que les provee de un sustento para identificar futuros cambios en el precio del producto.

Con el enfoque ARIMA-GARCH el Banco Mundial ha logrado proyectar hasta el año 2025 los precios del gas natural en Japón, Europa y EUA, dándole seguimiento cada cuatro años para hacer las modificaciones pertinentes, según los nuevos datos reunidos.

Para el caso de México , el interés se centra en la manera en que se comportarán los precios del gas natural en EUA, ya que como se mencionó anteriormente, éste influye sustancialmente en la fijación del precio a nivel regional, lo que incide en la demanda del mismo.

En la Figura 1.4 se muestra el comportamiento de los precios del gas natural dados por el Banco Mundial del periodo 2006-2016 (World Bank, 2016):



20.0000

18.0000

16.0000

14.0000

12.0000

10.0000

8.0000

6.0000

4.0000

2.0000

0.0000

EUA

Europa Japon

PRECIO DE MMBTU

Figura 1.4 Precio promedio mensual en dólares de un MMbtu de gas natural en Japón, EUA y Europa, periodo 2006-2016.

Fuente: Elaboración propia con datos provenientes del Banco Mundial (Worl Bank, 2016)

Con base en esa información y por medio del método ARIMA-GARCH, el Banco Mundial construyó un pronóstico de los precios para el periodo 2017-2025, en la Figura 1.5 se muestra su comportamiento:

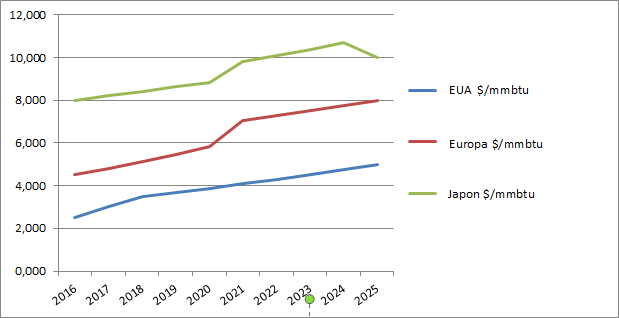


Figura 1.5 Comportamiento del precio mensual en dólares de un MMbtu de gas natural en Japon, EUA y Europa, del periodo 2017-2025.

Fuente: Elaboración propia con datos del Banco Mundial. (World Bank, 2016)

Algo que resalta en los pronósticos hechos por el Banco Mundial es que éstas tienen una tendencia a la alza para las tres series, con pequeña fluctuaciones en el intervalo de tiempo. Cabe mencionar que este organismo internacional ha centrado sus esfuerzos en el pronóstico del precio de diferentes energéticos, sin realizar estudios cuantitativos sobre la demanda de ellos.

Por su parte, la agencia estadounidense denominada Administración de Información Energética (AIE), emite anualmente un reporte sobre la prospectiva que tienen diferentes fuentes energéticas, entre ellas el gas natural, en el mercado estadounidense. En su reporte del 2017 hace énfasis en el crecimiento que tendrá el consumo del gas natural hasta el año 2040, llegando a la conclusión que el consumo se incrementará alrededor de 40.2% (EIA, 2017), con una tasa de crecimiento anual de 4% para los años 2017-2020.

La AIE no especifica qué metodología usa para cada pronostico que hace, pero si provee información sobre los modelos que se emplean dentro del organismo para llevar a cabo los análisis, entre ellos se encuentran: métodos de extrapolación, que incluyen el análisis de tendencias y escenarios, y predicción intuitiva (método Delphi) por medio del consenso de un grupo de expertos.

Por último se mencionarán algunas investigaciones hechas sobre la demanda del gas natural para el caso de EUA, por medio del algoritmo de redes neuronales, analizando en específico el periodo de 1922-2007.

En este caso se utilizaron Redes Neuronales de tipo Group Method of Data Handling (GMDH), el cual se representa como un conjunto de neuronas en el que diferentes pares de ellas están conectadas a través de un polinomio cuadrático, definiendo una función ff en orden a pronosticar una salida ŷŷ (Abrishami, Bourbour, Aghajani, 2014), para un vector de entrada dado. La red neuronal tipo GMDH se basa en el concepto de reconocimiento de patrones, lo cual es un refinamiento de los métodos y análisis tradicionales, al presentar una alta flexibilidad.

Así mismo se consideró el enfoque de redes neuronales de multicapas Feed-forward, el cual consiste en múltiples capas, donde cada neurona en una de las capas tiene conexión directa con las neuronas de las subsecuentes capas (Abrishami et al., 2014). Este enfoque usa una variedad de técnicas de

aprendizaje, siendo el algoritmo de back-propagation el más popular, el cual es un aprendizaje supervisado que tiene como objetivo reducir el error del sistema al mínimo.

También se empleó un enfoque híbrido de sistema inteligente, el cual usa el GMDH basado en el módulo de pronóstico de series de tiempo, con el módulo de un Sistema Experto (SE) con reglas extraídas de la regresión realizada (Abrishami, 2014). El SE posee cinco componentes que son: la base de conocimiento, la base de datos, el motor de inferencia, el interfaz de usuario y los módulos de justificación.

Las conclusiones a las que llegó Abrishami et al. (2014) al realizar la investigación son las siguientes:

* Detección de eventos irregulares que afectan el precio y la demanda del gas natural. Durante los periodos de crisis, cuando se estudiaron los eventos irregulares y poco frecuentes, por el método de SE se obtuvieron mejores resultados comparados al de las redes neuronales tipo GMDH y el de multicapas Feed-Forward.
* Cuando se dan eventos inesperados que influyen en la formación de la demanda del gas natural, las redes neuronales tipo GMDH y las redes multicapa Feed-Forward no pueden realizar predicciones adecuadas.
* El modelo híbrido inteligente puede emplearse como una herramienta efectiva para la predicción de la demanda del gas natural, al mostrar una mejor precisión que los otros dos modelos analizados.

Existe otro trabajo de investigación realizado por Nguyen Hoang Viet y Jacek Mandziuk (2002) , el cual se centra en el pronóstico de la demanda del gas natural en Polonia, usando una red neuronal difusa y una red neuronal feedforward, las dos usan algoritmos de aprendizaje supervisados, significando que se conocen las salidas deseadas del sistema.

Los autores analizaron el consumo diario del gas natural de enero del 2000 a diciembre del 2001, y realizaron pronósticos para el 2002, abarcando diferentes marcos temporales. El primero fue el pronóstico diario del consumo de gas

natural, donde se logró un error porcentual absoluto medio (MAPE) de 4.04% a través de la red neuronal feedforward y 4.24% para la red neuronal difusa; para la previsión semanal, la red neuronal difusa tuvo un MAPE 7.62% y la red feedforward incurrió en un error de 7.04%; por último , en el pronóstico mensual la red feedforward tuvo un error de 7.86% y la red neuronal difusa un 8.16% (Viet y Jacek Mandziuk, 2002).

En la Tabla 1.2 se muestran los trabajos realizados sobre el análisis y pronóstico del mercado del gas natural, especificando qué modelo se empleó y los resultados obtenidos:

Tabla 1.2 Trabajos representativos sobre el pronóstico de gas natural en el mundo. Fuente: Elaboración propia.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Referencia** | **Objeto de Estudio** | **Modelo Usado** | **Resultados** | **Año** |
| Nguyen Hoang | Pronóstico de la | Redes | MAPE de 4.24% para el | 2001 |
| Viet y Jacek | demanda de gas | neuronales | pronóstico anual. |  |
| Mendziuk | natural en Polonia | Feed-Forward |  |  |
| Abrishami, | Pronóstico de la | Redes | MAPE de 5.12% para el | 2014 |
| Bourbour y | demanda de gas | neuronales | pronóstico mensual |  |
| Aghajani | natural en EUA | Feed-forward |  |  |
|  | Pronóstico del precio |  | Incremento del precio en | 2016 |
| Banco Mundial | del gas natural en | ARIMA GARCH | 25%, 51% y 34%, |  |
|  | EUA, Europa y Japón |  | respectivamente, para el |  |
|  |  |  | año 2025. |  |
| Agencia de | Pronóstico de la | Extrapolación y | Aumento de la demanda | 2017 |
| Información | demanda de gas | Delphi. | en 40% para el 2040 |  |
| Economica | natural en EUA. |  |  |  |

En México existen estudios escasos sobre las proyecciones de la demanda de gas natural. Por el lado académico hay muchos trabajos cualitativos sobre su futuro, pero no existen investigaciones orientadas a proyecciones cuantitativas sobre la demanda de este producto, ya que los investigadores se han centrado en el análisis del petróleo dada su relevancia a nivel nacional.

El gobierno federal firmó en el 2013 un contrato con la empresa S&P Global Platts, para obtener información acerca de los precios y la demanda de petróleo y gas natural a nivel regional (PEMEX, 2016), formulando con esos datos

##### Justificación

Los diferentes agentes económicos están envueltos en un contexto incierto y cambiante, donde las decisiones tomadas no siempre traen los resultados esperados, debido a que no existe una manera exacta de saber lo que sucederá en un futuro.

El sector del gas natural está inmerso en este escenario complejo, el cual ha tenido un desarrollo dinámico debido a las transformaciones aceleradas en los últimos años de las fuentes de energía a nivel mundial. El descubrimiento de nuevos yacimientos de gas natural, el crecimiento industrial del país y el aumento poblacional, entre otros, son elementos que han influido en la manera en que se comporta la oferta y la demanda de este producto, lo cual fija su precio y la cantidad producida (Mankiw, 2012).

El anuario estadístico de PEMEX del año 2014 (PEMEX, 2014) expresa que el 4.42% de los ingresos de esta entidad paraestatal provinieron de la venta de gas natural, que si bien no es un aporte significativo al total de sus ingresos, si representa una suma importante de dinero para sus operaciones de 115 mil 991 millones de pesos anuales. Además, datos del Centro de Investigación para el Desarrollo (CIDAC, 2016) muestran que el sector industrial tuvo como principal fuente de energía al gas natural, la cual representó el 65% del consumo de combustibles a nivel nacional. Lo anterior evidencia la necesidad de modelos matemáticos que sirvan en la planeación de los recursos con los que cuentan los organismos públicos y privados, para satisfacer las necesidades de gas natural en el mercado mexicano.

Los modelos estadísticos han logrado pronósticos útiles para la toma de decisiones por parte de los diferentes agentes económicos en el área de finanzas, operaciones y mercadotecnia, siendo en muchos casos, las únicas herramientas que se emplean al momento de construir los pronósticos de alguna variable de interés (Mankiw, 2012). Tales modelos se usan de manera exitosa por el Banco de México (World Bank, 2016) para el pronóstico del precio del gas natural en EUA,

Japón y Europa. Así mismo, la SENER analizó y pronóstico la demanda de gas natural hasta el año 2025 con los modelos ARIMA, por lo que el uso extendido de estos modelos y la efectividad mostrada da la base para que en el presente trabajo de investigación se empleen para el pronóstico de la demanda de gas natural en México con una frecuencia mensual.

Las redes neuronales han mostrado un gran desempeño en el análisis de series de tiempo, tanto lineales como no lineales, convirtiéndolas en un modelo atractivo para el estudio de la demanda del gas natural. Trabajos como el de Abrishami et al. (2014) y las investigaciones hechas por Viet et al.(2002) mostraron que las redes neuronales son capaces de pronosticar la demanda de gas natural en EUA y Polonia, respectivamente. Por lo anterior, en el presente trabajo se incorpora la modelación de redes neuronales para pronosticar la demanda de gas natural en México.

##### Objetivos

* + 1. Objetivo General.

Analizar la demanda mensual de gas natural en México para el pronóstico su comportamiento para el periodo 2017 con una frecuencia mensual.

* + 1. Objetivos Particulares.

1. Identificar las variables que explican el comportamiento de la demanda del gas natural en el mercado mexicano, para que sean incorporadas a los modelos matemáticos multivariados.
2. Comparar los resultados obtenidos por los modelos NARX y NAR con los obtenidos por los modelos ARIMA y ARIMAX, para determinar cuáles realizan los mejores pronósticos.
3. Evaluar los resultados obtenidos por los modelos NARX, NAR, ARIMA Y ARIMAX con el fin de determinar cuáles deben de ser usados para efectuar proyecciones en diferentes periodos de tiempo.
4. Determinar las ventajas y desventajas que tiene el uso de los modelos matemáticos utilizados, para que los modeladores puedan elegir aquellos que se adapten mejor a sus requerimientos.

###### Tabla de Congruencia de los Objetivos del Trabajo de Tesis

Tabla 1.4 Congruencia de los objetivos del trabajo de Tesis. Fuente: Elaboración propia.

|  |  |
| --- | --- |
| **Pronóstico de la Demanda del Gas Natural en México** | |
| **Objetivo General**  Analizar la demanda mensual de gas natural en México para pronosticar su comportamiento para el periodo 2017 con una frecuencia mensual.  **Pregunta de Investigación.**  ¿Qué actividades deben de realizarse para analizar la demanda de gas natural y así pronosticar su comportamiento?  **Hipótesis.**  Los modelos de Redes Neuronales Artificiales realizan con menor error el pronósticos de la demanda de gas natural en México que los modelos ARIMA y ARIMAX | |
| **Justificación**  Con base al estudio realizado del contexto físico, temporal y cultural , y con el análisis efectuado a la situación actual del pronóstico de la demanda de gas natural en México , se determinó que es necesario crear un trabajo de investigación con orientación sistémica , en el cual se integre de manera efectiva conocimiento del área de matemáticas, sistemas complejos y economía , para generar modelos que exhiban un buen desempeño en el pronóstico de nuestra variable de interés. | |
| **Objetivo específico 1.**  Identificar las variables que explican el comportamiento de la demanda del gas natural del mercado mexicano, para que sean incorporadas a los modelos matemáticos multivariados.  **Pregunta(s) de investigación**.  ¿Cómo se determina qué variables son las indicadas para usarse en los modelos?  ¿Existe alguna teoría económica que indique como se forma la demanda de los bienes y servicios? | **Objetivo específico 2.**  Comparar los resultados obtenidos por los modelos NARX y NAR con los tenidos por los modelos ARIMA y ARIMAX, para determinar cuáles realizan los mejores pronósticos.  **Pregunta(s) de investigación.**  ¿Qué criterio es el adecuado para comparar los pronósticos realizados por los diferentes modelos?  ¿Es apropiado comparar el desempeño que tienen los  modelos univariados con el tenido por los modelos multivariados? |
| **Objetivo específico 3** | **Objetivo específico 4.** |
| Evaluar los resultados obtenidos por los modelos | Determinar las ventajas y desventajas que tiene el uso de |
| NARX, NAR, ARIMA Y ARIMAX con el fin de | los modelos matemáticos utilizados, para que los |
| determinar cuáles deben de ser usados para las | modeladores puedan elegir aquellos que se adapten mejor |
| proyecciones realizadas en diferentes periodos de | a sus requerimientos. |
| tiempo. |  |
|  | **Pregunta(s) de investigación.** |
| **Pregunta(s) de investigación.** | ¿Es adecuado usar solo un modelo matemático para |
| ¿De qué manera se puede evaluar el desempeño que | pronosticar el comportamiento de alguna variable de |
| tuvieron los modelos en los diferentes periodos | interés? |
| temporales? |  |

**Capítulo 2**

**Marco Teórico y Metodológico.**

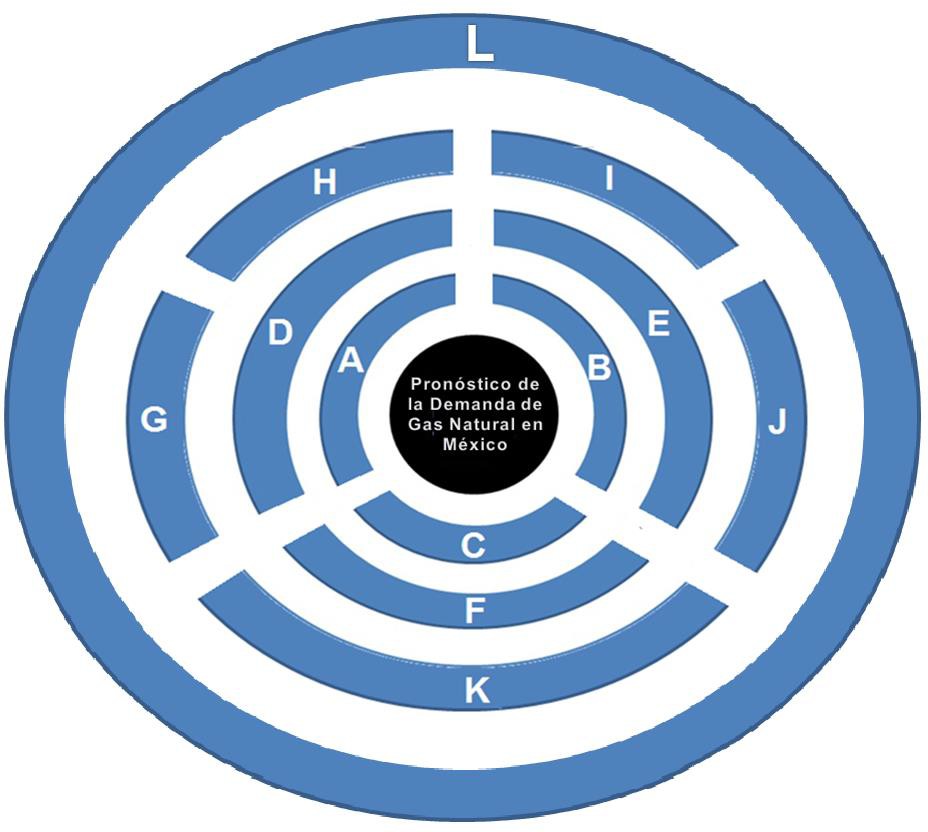
## Capítulo 2 . Marco Teórico y Metodológico

##### Marco Teórico.

En este apartado se presenta el conjunto de teorías y las investigaciones hechas en el área de interés para sustentar el trabajo de tesis, detallando los elementos empleados para el desarrollo de la investigación, mostrando las relaciones que tienen y su relevancia.

##### Teorías Utilizadas

En la Figura 2.1 se presentan un Holos con las teorías que se emplean en el presente trabajo, para abordar el pronóstico de la demanda de gas natural en México**.**



1. Estadística G) ARIMA
2. Sistemas Complejos H) ARIMAX
3. Economía I ) NAR
4. Modelos de Pronóstico J) NARX
5. Redes Neuronales Artificiales K) Teoría de la Demanda
6. Teoría Económica L) Sistémica

Figura 2.1 Holos con las teorías que dan sustento a la presente investigación. Fuente: Elaboración propia

##### Sistemas.

El término sistema tiene diversas acepciones, las cuales dependen del autor que esté utilizando el concepto. Por ejemplo para Von Bertalanffy (Miller,1965) *“un sistema es un conjunto de unidades con relaciones entre ellas, las cuales intercambian materia con el medio circundante”.* Por su parte J.G. Miller et al. (1965) en su obra Sistemas Vivos*,* propone que *“el universo contiene una jerarquización de sistemas, donde cada sistema de mayor nivel está compuestos por sistemas de niveles inferiores”*, lo que significa que un sistema está conformado por subsistemas interrelacionados, los cuales surgieron a través de la integración de sus elementos para lograr el cumplimiento de un objetivo común.

El enfoque sistémico es fundamentalmente diferente a las formas tradicionales de análisis, ya que estas últimas se centran en separar y aislar los elementos que componen al objeto que está siendo estudiado, mientras que el pensamiento sistémico se enfoca en entender las interacciones que mantienen sus elementos, así como las interacciones que hay entre el sistema con su entorno y el contexto en las que se dan, ampliando la visión que se tiene del sistema de interés y de las características que éste posee, siendo de vital importancia cuando se tienen problemas complejos y dinámicos, los cuales muestran una gran cantidad de interrelaciones que cambian con el tiempo (Rosenberg, 2000).

El pensamiento de los sistemas es una forma de ampliar la perspectiva sobre elementos, estructuras, patrones y ciclos globales en los sistemas, en lugar de apreciar sólo partes específicas del mismo (Johansen, 2004). Esta visión integradora ayuda a identificar las causas de los problemas en los sistemas y da las herramientas necesarias para su solución.

##### Ciencias Económicas.

Se entiende a la economía como una ciencia social que estudia la asignación de recursos escasos. Ésta trata de responder las preguntas ¿qué producir?, ¿cómo producirlo?, ¿cuándo producirlo? y ¿para quién producirlo? (Mankiw,et al. 2012)

La ciencia económica se aboca principalmente a estudiar las actividades que involucran a la producción, el intercambio y el consumo de bienes y servicios, las cuales forman al sistema económico. Con el desarrollo de las sociedades contemporáneas estos procesos han adquirido una especial complejidad, al existir un gran número de sistemas interrelacionados, los cuales buscan cumplir los objetivos que manejan en un contexto incierto y cambiante.

El ente donde se realiza el proceso económico se le denomina *mercado*. El mercado es un concepto económico que alude a la organización de consumidores y oferentes de un bien o servicio (Mankiw et al. 2012), donde los primeros fijan la demanda del producto y los segundos, la oferta, siendo el mercado el ente donde pueden concurrir estos dos grupos para realizar sus transacciones económicas.

Como resultado de la interacción de consumidores y productores se logra fijar el nivel de producto vendido a un precio determinado, el cual es el punto de equilibro de un mercado en particular, pero este punto no siempre se alcanza debido a algunas distorsiones como lo son los impuestos, los precios mínimos o máximos, entre otros.

##### 2.2.1.2.1. La formación de la Demanda de Gas Natural.

La demanda se entiende como la cantidad de producto que están dispuestos a comprar los consumidores a un nivel de precio dado (Mankiw, et al 2012). Ésta se determina a través de diferentes elementos, los cuales se listan a continuación:

* + Ingreso de los consumidores: es uno de los elementos de mayor relevancia para fijar la demanda de cierto bien, ya que las personas no solo deben de tener el deseo de adquirir algún producto, sino que también deben de tener la capacidad económica para hacerlo.
  + Tamaño del mercado: el número de compradores potenciales de algún producto incide directamente en la cantidad que se requiere de ese bien, para así satisfacer las necesidades del mercado.
  + Expectativas: los compradores regulan sus patrones de consumo en función de lo que esperan que suceda en un futuro, por ejemplo, si algún individuo piensa que su trabajo corre peligro, posiblemente opte por consumir menos de ciertos bienes para ahorrar parte de sus ingresos y así estar preparado ante algún suceso adverso; otro ejemplo es si los directores de una empresa notan un panorama económico desfavorable, se verán disuadidos a no invertir en proyectos que incrementen su capacidad productiva , o inclusive a achicar la capacidad de las empresas que ya tienen, lo cual influirá en la demanda de insumos y factores de producción.
  + Preferencias: la toma de decisiones de las personas tienen un fuerte elemento psicológico, el cual se hace patente al momento de optar por la adquisición de un cierto producto. Las ciencias económicas se encargan de estudiar qué pasa cuando los gustos de los consumidores cambian, ya que esto se traduce en una contracción o expansión de la demanda de diferentes bienes y servicios.
  + Precio de bienes sustitutos o complementarios: el precio funge como una pieza clave al momento de elegir qué bienes o servicios comprar. Si se consume habitualmente mantequilla, pero el precio de la margarina disminuyera, se incentivaría el consumo de esta última, ya que ambas satisfacen la misma necesidad, por lo que se dice que los dos son productos sustitutos. Por otra parte pensemos que hay un aumento considerable en el

precio de la gasolina, muchos de los usuarios de automóviles se verían tentados a dejar de usar vehículos, debido a que estos emplean a la gasolina como fuente de energía, conllevando que la demanda se redujera en cierta cuantía; si ambos productos se consumen de manera simultánea, se dice que son bienes complementarios, y la bajada o aumento en el precio de uno, afecta la demanda del otro.

##### Redes neuronales artificiales.

Una Red Neural Artificial (RNA) es un modelo matemático que intenta simular la estructura y las funcionalidades de las redes neuronales biológicas. El bloque básico de este modelo es la neurona artificial, la cual tiene un bias y recibe ciertas entradas de interés que son suministradas a una función matemática (Hagan, 2009).

Las entradas de la red son ponderadas, lo que implica que cada valor de entrada se multiplica con el peso individual que se le da, este peso es el análogo en la red neuronal biológica del enlace sináptico, los cuales permiten que exista un flujo distinto de electricidad, para que así las neuronas puedan comunicarse entre sí y puedan en su conjunto acceder a la información que almacena nuestro cerebro, lo cual nos da la capacidad de adquirir nuevo conocimiento.

El bias es un valor que tiene la neurona artificial, el cual se inspira en el potencial eléctrico que tiene cada neurona biológica. Este bias permite que la neurona tenga un parámetro adicional, para poder procesar la información que se le otorga.

En la sección media de la neurona artificial se suma todas las entradas multiplicadas por los pesos correspondientes y el bias que tiene la neurona. La salida que se obtiene es dada a una función de activación, a la cual también se le llama función de transferencia, la cual da un valor numérico como salida final.

En la Figura 2.2 se muestra de manera gráfica a una neurona artificial y las partes que la componen:

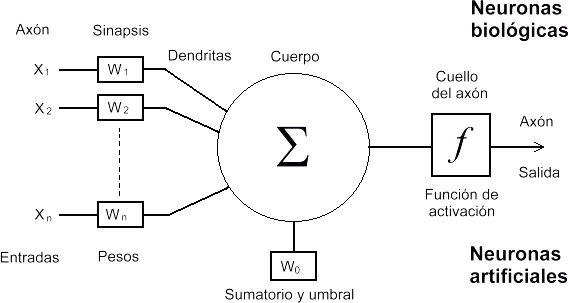


Figura 2.2 Modelo gráfico de una neurona artificial (Kriesel, 2005)

Esta arquitectura permanece constante para todas las neuronas artificiales, la cual solo sufrirá variaciones en el valor del bias, en los pesos, la función que se le asigne y en el número de entradas con sus respectivas magnitudes.

Una neurona puede tener una o más entradas y una o más salidas; la salida de la neurona es el resultado de una combinación no lineal de las entradas {𝑝}, ponderada por la los pesos sinápticos {𝑤} y la aplicación de una función {𝑓} al resultado de la sumatoria. En la Figura 2.3 se muestra gráficamente a una neurona con 𝑅 entradas:

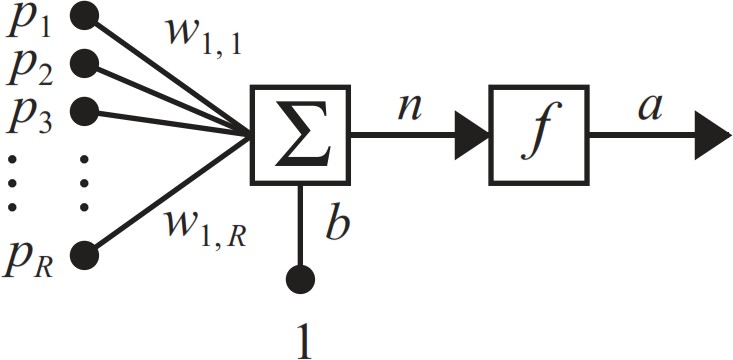


Figura 2.3 Modelo gráfico de una neurona con *P* entradas (Hagan. 2014)

Cada entrada tiene un peso relativo que representa la importancia de la señal que envía; estas ponderaciones se asignan de acuerdo con el aprendizaje que tiene la red neuronal. Las múltiples señales ponderadas se combinan en la neurona, a las cuales se les suma el bias y se le pasa por la función de activación, la cual nos dará el resultado final del procesamiento.

En la Figura 2.4 se da la generalización de una arquitectura de red con 𝑆

neuronas y 𝑅 entradas:

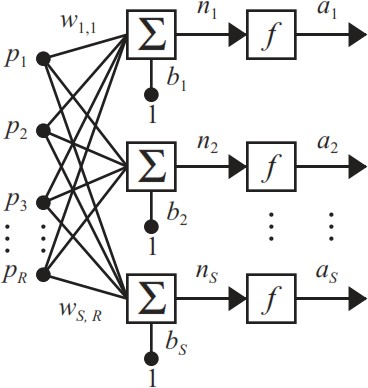


Figura 2.4 Modelo gráfico de 𝑆 neurona con 𝑅 entradas (Hagan et al., 2014)

Matemáticamente la salida de la red queda expresada como:

𝒂 = 𝒇(𝑾𝒑 + 𝒃) (2.1)

Donde 𝑾 es la matriz de los pesos, 𝒑 es el vector de entradas y 𝒃 es el vector bias.

Las funciones de transferencia 𝒇 pueden ser lineales o no lineales, existiendo variantes que se eligen en base al problema que se desea resolver (Hagan, 2014), algunas de las más usuales son la *hardlim*, *purelin*, *logsig* y *tansig.*

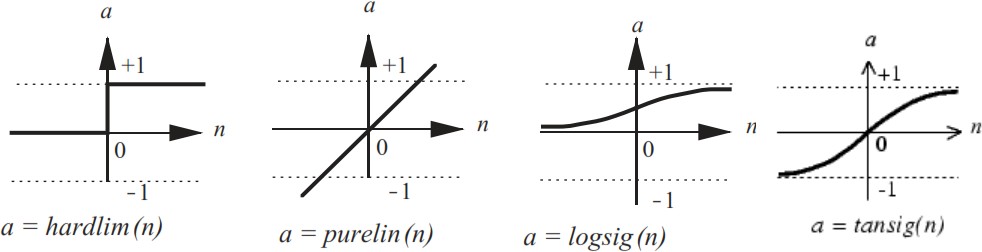


Figura 2.5 Modelo gráfico de las funciones de transferencia *hardlim*, *purelin*, *logsig* y

*tansig.* (Hagan, 2014)

En la arquitectura de las RNA puede estar presente más de una capa, en donde se hace la diferenciación entre la capa de salida y las demás capas a las que usualmente se le denominan capas ocultas. En la Figura 2.6 se muestra una arquitectura con tres capas:

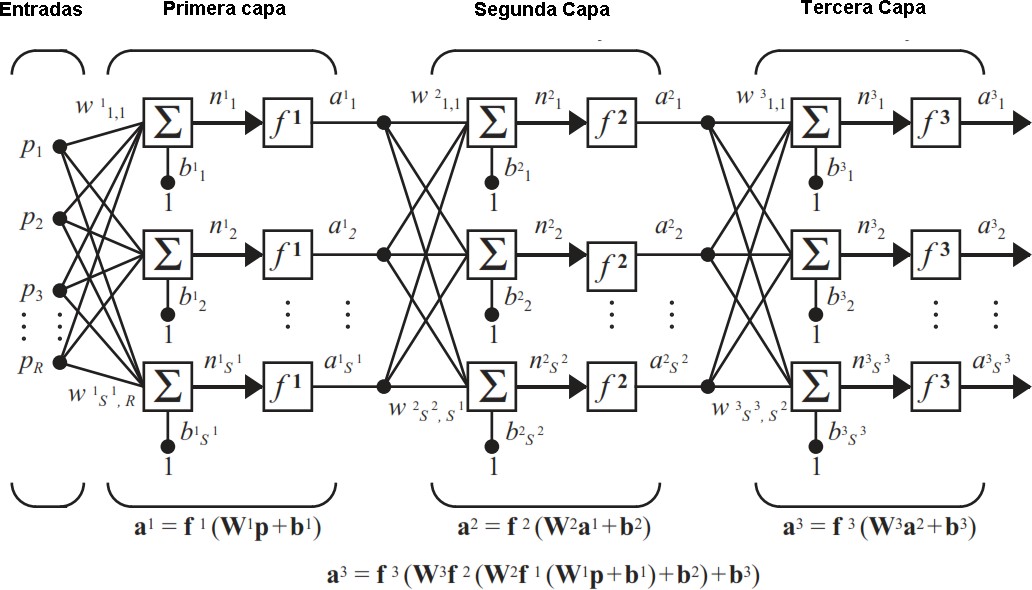


Figura 2.6 Modelo gráfico de una red neuronal con tres capas. (Hagan et al., 2014)

No existe un método formal para elegir el número de neuronas y capas que tendrá la red, por lo regular, esta elección se hace en base a un proceso de prueba y error, hasta encontrar la arquitectura que mejor pueda resolver el problema de interés.

Las redes neuronales pueden dividirse en dos categorías: estática y dinámica. Las redes estáticas no tienen elementos de retroalimentación ni retrasos, por tanto la salida se calcula directamente a partir de las entradas actuales. Los modelos estáticos se utilizan para representar las propiedades que tiene un sistema, los cuales son independientes del tiempo, significando que su comportamiento permanece relativamente constante. (Ghosh et al., 2009)

En contraparte tenemos a las redes dinámicas, las cuales tienen memoria en forma de retrasos o ciclos recurrentes, dándoles la capacidad de ser entrenadas para aprender patrones secuenciales o que varían con el tiempo. Esto las convierte en redes útiles para el pronóstico financiero, la ecualización de canales, la clasificación, el reconocimiento de voz, la detección de fallas, etc. (Hagan et al., 2014).

##### El perceptrón.

A finales de la década de 1950 Frank Rosenblatt (Hagan et al., 2014) junto a otros investigadores crearon una red neuronal denominada preceptrón. Su mayor aporte fue el de desarrollar una regla de aprendizaje del preceptrón abocada a resolver problemas de identificación de patrones.

Esta regla de aprendizaje tuvo gran impacto, ya que lograba converger a los valores de los pesos correctos, siempre que el problema tuviera solución a través de un modelo simple y automático. Se demostró que el preceptrón era capaz de aprender aun cuando los pesos y los bias se inicializaran de manera aleatoria. A pesar de este logro, la red preceptrón posee limitaciones inherentes, Marvin Minsky y Seymour Papert (Demuth, 2007) mostraron que las redes preceptrón no tenían la capacidad de implementar ciertas funciones elementales y los problemas que podían resolver eran aquellos que pudieran ser separados linealmente.

En la Figura 2.7 se muestra de forma gráfica a la arquitectura de la red preceptrón.

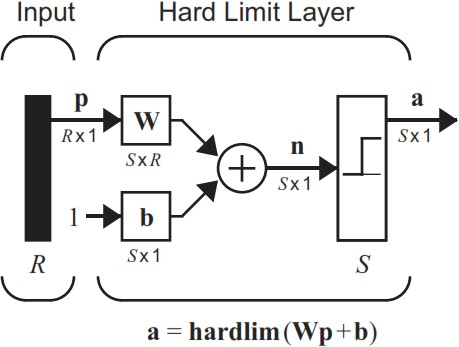


Figura 2.7 Modelo gráfico de la red preceptrón. (Hagan et al., 2014)

El perceptrón tiene 𝑅 entradas, tiene 𝑆 neuronas y una función de activación de tipo hardlim. La función de activación hará que, cuando la salida 𝑛 sea menor a cero, 𝑎 sea igual a cero; cuando la salida 𝑛 sea mayor o igual a cero, 𝑎 adoptará el valor uno. Esto indica que solo se tendrán dos posibles soluciones, por lo que la red preceptrón de una capa tiene un desempeño limitado, pudiendo solo resolver problemas que son linealmente separables.

##### El preceptrón multicapa

En la década de 1980, algunas limitaciones del preceptrón simple fueron superadas a través de un preceptrón multicapa y con una regla de aprendizaje más eficiente (Demuth, 2007).

El preceptrón multicapa tiene una arquitectura multicapa, lo cual le permite resolver problemas que no sean linealmente separables. La salida de las neuronas está dada por la sumatoria de las entradas ponderadas y el bias, y a la aplicación de una función de transferencia determinada.

En la Figura 2.8 se enseña de manera gráfica una red preceptrón de tres capas:

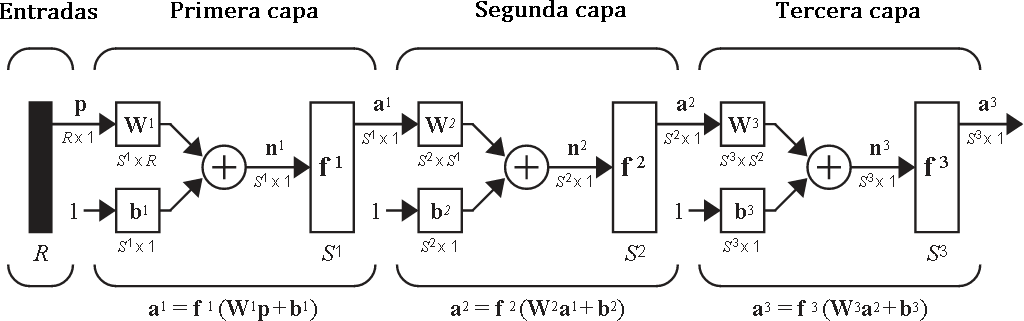


Figura 2.8 Modelo grafico de una red preceptrón multicapa. (Hagan et al., 2014)

La arquitectura del preceptrón con múltiples capas, posibilita que las entradas se encuentren conectadas a una o varias neuronas de la subsecuente capa.

Se pueden identificar tres tipos de capas: 1) Capas de entrada: están integradas por la información que ingresa a la red, en ésta no se realiza ningún tipo de procesamiento; 2) Capas ocultas: en éstas se encuentran las neuronas receptoras de la información que proviene de capas anteriores, las cuales se encargan de procesar la información que será enviada a las capas subsiguientes; 3) Capas de salida: están conformada por las neuronas que efectúan un procesamiento final de la información y dan la salida total de la red neuronal.

Los limitantes que tiene esta red neuronal son expuestos por Ortiz (2012) el cual nos dice que *“la extrapolación no es perfecta, esto significa que en caso de no ser suficiente o bueno el entrenamiento, la información de salida puede resultar imprecisa. La existencia de mínimos locales en la función de error complica su entrenamiento, pues una vez alcanzado un valor mínimo en el entrenamiento, éste se detiene, pudiendo no ser un mínimo global y no satisfacer el valor mínimo del error permitido”.* (p. 90)

##### El Algoritmo de aprendizaje de Retropropagación

El método de Retropropagación se utiliza en las redes neuronales artificiales para calcular la contribución del error de cada neurona, después de que un conjunto de datos han sido procesados, propagando el error hacia la capa de entrada, pasando por las capas ocultas intermedias y calibrando los pesos de las conexiones para reducir dicho error

Este algoritmo es un ejemplo de aprendizaje supervisado, donde se conocen la pareja de entradas y metas que estarán presentes en el aprendizaje:

{𝒑𝑖, 𝒕𝑖}, {𝒑𝑖+1,𝒕𝑖+1}, … , {𝒑𝒊+𝒒, 𝒕𝑖+𝑞} (2.2)

Donde 𝒑𝑖 es una entrada de la red y 𝒕𝑖 es la salida correspondiente. A medida que cada entrada se introduce a la red, la salida que se tiene se contrasta con la

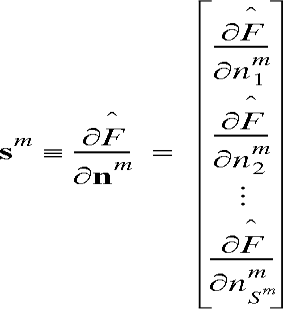
meta asignada. El algoritmo ajustará los parámetros de la red con el fin de minimizar el error cuadrado medio (Hagan, 2014). Cuando se tienen múltiples salidas, la función contemplara la esperanza matemática de los errores de la siguiente manera:

𝐹(𝒙) = 𝐸[𝒕 − 𝒂]𝑇[𝒕 − 𝒂] (2.3)

El ajuste se hace a través de la actualización de los pesos y los bias, apegándose a las siguientes expresiones:

𝑾𝑚(𝑘 + 1) = 𝑾𝑚(𝑘) − 𝛼 𝑺𝑚 (𝒂𝑚−1 )𝑇 (2.4)

𝒃𝑚(𝑘 + 1) = 𝒃𝑚(𝑘) − 𝛼 𝑺𝑚 (2.5)

Donde 𝑾 son los pesos, 𝒂 son las salidas, 𝒃 son los bias**,** 𝛼 es la tasa de aprendizaje y 𝑺 está dado por:

(2.6)

Siendo 𝒏 el vector que contiene la información a la que se le aplicará las funciones de activación.

##### Algoritmo de aprendizaje Levenberg-Marquardt.

El algoritmo de Levenberg-Marquardt se ha popularizado por el alto desempeño que tiene, según Lasfer (2013) la compañía MathWorks, creadora de MATLAB, probó una serie de algoritmos de retropropagación, con lo que pudo concluir que el de Levenberg-Marquardt tiene la convergencia más rápida y el error cuadrático medio más bajo, para un problema de aproximación que involucre una red neuronal pequeña o mediana.

El algoritmo de Levenberg-Marquardt es una mezcla de descenso por gradiente y del método de Newton. Mientras que el descenso por gradiente sufre de diversos problemas de convergencias, el método de Newton mejora estos problemas mediante el uso de segundas derivadas.

El método de Newton puede expresarse como:

𝑥𝑘+1 = 𝑥𝑘 − 𝐻−1𝑔𝑘 (2.7)

Donde 𝑥 es la variable de interés, 𝐻 es la matriz Hessiana y 𝑔 es el gradiente evaluado en 𝑥𝑘 . Aunque el método de Newton proporciona una buena convergencia, es muy complejo calcular la matriz de Hessiana. Levenberg-Marquardt combina las ventajas tanto del descenso por gradiente como del método de Newton, sin tener que calcular la matriz de Hessiana (Lasfer et al., 2013), pues se hace una aproximación:

Y el gradiente como:

𝐻 = 𝐽 𝑇 𝐽 (2.8)

𝑔𝑘 = 𝐽 𝑇𝑒 (2.9)

Donde 𝐽 es la matriz Jacobiana, la cual contiene la primera derivada de los errores de la red; y 𝑒 representa el vector que contiene los errores. Se debe notar que calcular la matriz Jacobiana es una tarea significativamente más fácil que calcular la matriz Hessiana.

Con lo anterior, la iteración de Levenberg-Marquardt puede escribirse como:

𝑥𝑘 = 𝑥𝑘−1 − [𝐻 + 𝛼𝐼]−1𝑔𝑘 (2.10) Donde 𝐼 es la matriz identidad y 𝛼 es la tasa de aprendizaje.

##### La Red Neuronal NAR.

La red neuronal autorregesiva no lineal (NAR) es de tipo dinámica, la cual emplea un conjunto de datos como entradas y salidas de la red, por lo que los pronósticos realizados, estarán basados en los valores pasados que adopta la variable de estudio. En la Figura 2.9 se presenta el modelo grafico de una particular red neuronal NAR:

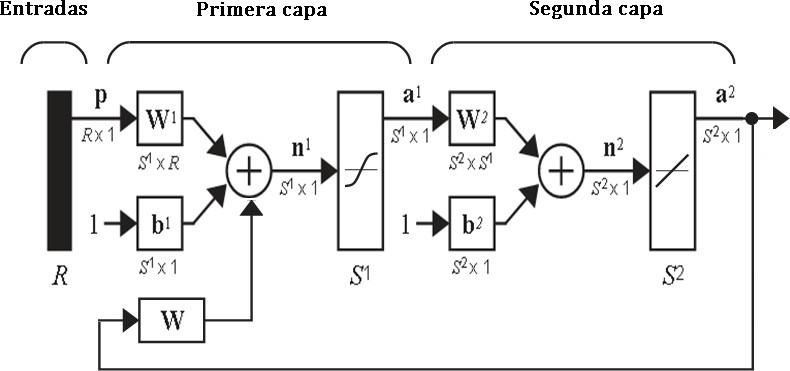


Figura 2.9 Modelo gráfico de una red neuronal NAR. (Hagan et al., 2014)

La red está conformada por dos capas, usa una función de activación *tansig* en su primera capa y una función *purelin* en la capa de salida. Su arquitectura usa bucles abiertos cerrados, el primero se ejecuta en el entrenamiento de la red, lo que ayuda a tener una mejor precisión, dado que se emplea el valor real, en vez del retroalimentar el valor estimado; el segundo se emplea para realizar pronósticos de más de una unidad de tiempo, dando aproximaciones con menor error (Hagan et al., 2014).

Cabe mencionar que las redes NAR pueden tener una arquitectura de *n* capas; se pueden emplear diferentes funciones de activación, las más usuales son la *logsig* y la *tansig,* y en su capa de salida se utiliza la función *purelin*; la función que se emplea para cuantificar el error es la suma de los errores al cuadrado (SSE) o su promedio (MSE); y el algoritmo de aprendizaje más utilizado es el de Retropropagación Levenberg-Marquardt. (Allende et al., 2002).

El pronóstico de la serie de tiempo está dada por:

𝑦(𝑡) = 𝑓(𝑦(𝑡 − 1) + 𝑦(𝑡 − 2) + ⋯ + 𝑦(𝑡 − 𝑑)) (2.11)

Donde 𝑦 es el valor que adopta la variable y 𝑑 es el número de retrasos que tendrá la red.

Para crear los modelos y ejecutar la simulación pertinente, se usará el software matemático MATLAB, el cual es un entorno de desarrollo integrado que posee su propio lenguaje de programación. Este tiene una interfaz gráfica para construir las arquitecturas de las redes NAR y NARX, que posteriormente serán ejecutadas.

Ortiz (2012) proporciona un marco de referencia para la construcción y simulación de las redes neuronales en MATLAB, por medio del módulo Wizard:

1. Tener disponible la base de datos con la que se ha de trabajar, para que sea importada a Matlab, por medio de la instrucción Import Data.
2. Acceder al módulo de Redes Neuronales con la instrucción nnstart, posteriormente seleccionar en el menú desplegado la opción Time Serie App.
3. Crear la red, seleccionando la arquitectura NAR o NARX.
4. Establecer las metas y entradas(si es el caso de la red NARX) de los datos que habían sido previamente importados.
5. Configurar la red, definiendo el número de neuronas y rezagos, así como el porcentaje de datos que tendrá el grupo de entrenamiento, validación y comprobación.

##### La Red Neuronal NARX.

La red neuronal NARX es de tipo recurrente, por lo que la salida final puede ser enviada a capas precedentes, lo que crea una memoria a corto plazo para la red y le permite comportarse de forma dinámica. Ésta también es de tipo autorregesiva no lineal, solo que ahora se tendrán variables exógenas para modelar el comportamiento de una variable de interés. En la Figura 2.10 se presenta su forma gráfica:

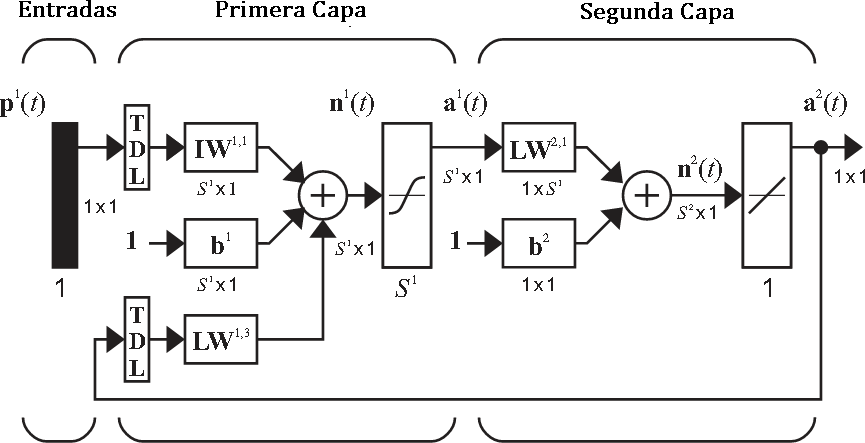


Figura 2.10 Modelo gráfico de una red neuronal NARX. (Hagan et al., 2014)

En la arquitectura está presente una función de transferencia *tansig* (aunque pudo ser una *logsig*) en la capa oculta y una lineal en la capa de salida. El número de neuronas en la capa oculta, 𝑆1 dependerá de la complejidad del sistema que se está aproximando.

La Línea de Retardo (TDL) de las entradas, contendrá a las variables explicativas 𝑥(𝑡 − 1), … , 𝑥(𝑡 − 𝑑𝑥) ; la TDL de las salidas tendrá a las variables de estudio 𝑦(𝑡 − 1), … , 𝑦(𝑡 − 𝑑𝑦).

El pronóstico se ajustará a la siguiente ecuación:

𝑦(𝑡) = 𝑓(𝑥(𝑡 − 1) + ⋯ + 𝑥(𝑡 − 𝑑𝑥) + 𝑦(𝑡 − 1) + ⋯ + 𝑦(𝑡 − 𝑑𝑦)) (2.12) Donde 𝑥 serán los valores de las variables exógenas, 𝑦 serán los valores de la variable de salida, 𝑑𝑥 serán los retardos empleados en las variables exógenas y 𝑑𝑦 el número de retardos de la variable de salida

##### Estadística.

La estadística es un área de estudio muy amplia con aplicaciones en un gran número de campos. En general se puede decir que la estadística es la metodología para recopilar, analizar, interpretar y sacar conclusiones en base a la información que se posee (Gujarati, 2010), en otras palabras, la estadística es la metodología que los científicos y profesionistas usan para interpretar y sacar conjeturas en base a los datos recogidos.

Por tanto, todo lo relacionado con la recopilación, procesamiento, interpretación y presentación de datos pertenece al dominio de la estadística (Isotalo, 2001). Por medio del uso de los métodos estadísticos se pueden responder a preguntas como:

se debe organizar y resumir los datos?

los datos y sacar conclusiones de ello? evaluar la fortaleza de las conclusiones y evaluar su

incertidumbre?

##### Series de Tiempo y Pronósticos

Una serie de tiempo es un conjunto de observaciones, ordenadas de manera secuencial y en intervalos equidistantes. Las mediciones de la series de tiempo son representadas con *Y1,. . ., YT* donde *T* es el número total de observaciones (Gujarati, 2010).

Los datos vertidos en una serie de tiempo se utilizan para crear un modelo matemático que logre hacer pronósticos de los valores que adoptará la variable dependiente a lo largo del tiempo. Estos pueden tener un componente estacional o de ciclo, un elemento de tendencia y un componente de valores promedios en los diferentes rezagos.

El pronóstico de la variable dependiente puede realizarse con base a una sola variable independiente (modelos univariados) o con respecto a múltiples variables explicativas (modelos multivariados).

##### Proceso autorregresivo (AR)

El modelo autorregresivo representa el valor actual de series de tiempo como combinación de uno o más valores anteriores de la serie, mostrando la dependencia de un valor con los valores anteriores más cercanos (Dickey y Fuller, 1979).

El modelo autorregresivo es de orden *p*, el cual determina cuántos valores previos deben incluirse en la ecuación para estimar el valor actual (Dickey et al, 1979). Para el caso donde *p=1*, se tiene un AR (1) descrito por la ecuación:

*Y t* = *ΦY t-1+ ε t* (2.13)

Donde *ε t* es el término que representa al error y *Φ* es el valor a ser estimado.

el valor del error está dado por:

*ε t* = *Y t - ΦY t-1* (2.14)

Cuando se tiene un proceso autorregresivo de *p* orden, el modelo AR (p) tendrá la forma:

*Y t* = *Φ1Y t-1 +* … *+ Φp Y t-p+ ε t* (2.13)

En donde los términos a estimar son *Φ1 , ..., Φp ≠* 0 ; y *ε t* es el error conformado por variables aleatorias no correlacionadas , las cuales tienen media cero y varianza constante, al que se le conoce como ruido blanco (2010, Gujarati)*.*

##### Proceso de media móvil (MA)

Una serie temporal está influenciada por elementos aleatorios, los cuales inciden en el valor de la variable dependiente en periodos posteriores. Los términos de medias móviles se utilizan para captar la influencia de los elementos aleatorios previos, para así determinar el valor futuro de la serie temporal (Gujarati, 2010).

La media móvil de primer orden o MA (1), es una serie de tiempo dada por:

*Y t* = 𝜇+ 𝛼*0 z t +* 𝛼*1 z t-1* (2.15)

En este caso 𝜇 representa una constante; *zt* es el error obtenido en el pronóstico *t*; *zt-i* es el error obtenido en el pronóstico *t-i,* donde *i* va de *1* a *q;* y 𝛼*i* es la ponderación para cada uno de los errores en los distintos periodos que van de *1* a *q* .

La generalización de un proceso MA de orden *q* está dada por:

*Y t* = 𝜇+ 𝛼*0 z t +* 𝛼*1 z t-1 +* … *+* 𝛼*q z t-q* (2.16)

Por lo que es posible observar que un proceso de media móvil es una combinación lineal de los términos de error que están involucrados en el modelo.

##### Proceso autorregresivo y de media móvil (ARMA)

Un modelo ARMA está formado por una combinación de términos AR y MA, el cual se puede usar para aproximar cualquier proceso de tipo estacionario (Levin, 2004). Por ejemplo un ARMA (1,1) está dado por:

*Y t* = (𝜇+ 𝛼*0 z t +* 𝛼*1 z t-1)+ (ΦY t-1+ ε t)* (2.17)

La generalización del ARMA (p, q) tiene la forma:

*Y t* = (𝜇+ 𝛼*0 z t +* 𝛼*1 z t-1 +* … *+* 𝛼*q z t-q) + ( Φ1Y t-1 +* … *+ Φp Y t-p+ ε t)* (2.18)

La cual se obtuvo de la combinación de las ecuaciones 2.13 y 2.15

* + - * 1. **Proceso autorregresivo Integrado de media móvil (ARIMA).**

Los modelos ARMA suponen que la serie de tiempo es estacionaria, es decir, tiene propiedades invariables con respecto al tiempo, como son la media, la varianza y las auto correlaciones (Levin et al., 2004).

Muchas series de tiempo no son estacionarias, por lo que se deben aplicar diferentes métodos para que lo sean, uno de ellos es el uso de diferencias.

La primera diferencia de una serie de tiempo *Yt* se define como:

*ΔYt= Yt- Yt-1*. (2.19)

Las series de tiempo que utilizan el modelo ARMA y que emplean diferencias para hacerlas estacionarias, se conoce como modelo autorregresivo de media móvil integrado (ARIMA). Este modelo está representado por tres parámetros: *p* que es el orden del modelo autorregresivo, *d* que denota el orden de la diferencia usada, y *q* que prescribe el orden del modelo de media móvil.

Con lo anterior se puede apreciar que el modelo ARIMA(p,d,q) toma los datos históricos y los descompone en un proceso autorregresivo (AR) que mantiene la memoria de eventos pasados, un proceso integrado (I) que hace que los datos sean estacionarios y un proceso de media móvil (MA) que utiliza los errores de los pronósticos.

Para poder construir un modelo ARIMA adecuado se necesitan seguir los siguientes pasos (Kahforoushan, 2010):

1. Calcular la variable de interés y graficar la serie de tiempo para realizar una primera aproximación. De manera visual se podría distinguir si la serie tiene un desarrollo errático alrededor de una tendencia, lo que podría indicar que es no estacionaria; así mismo se apreciaría sí ésta oscila en un rango de valores, entorno a una constante, indicando que podría ser una serie estacionaria.
2. Corroborar la estacionariedad. Para determinar de manera formal si una serie es estacionaria o no, se debe de verificar que la varianza y la media sean constantes usando la prueba de White (1980). Así mismo se emplea la prueba de raíz unitaria, para saber si la serie es integrada, si es así, denotara que la serie es estacionaria.
3. Establecer si existe estacionalidad. Para establecer si la serie tiene un comportamiento repetitivo cada cierto periodo, se puede realizar una prueba de varianza, lo que permitiría apreciar si existe una diferencia estadísticamente significativa en los valores que adopta la variable de interés, en las diferentes unidades de tiempo (horas, días, semanas, meses, etc.). Así mismo es posible usar los correlogramas parciales y simples para poder verificar si existen rezagos significativos cada cierto intervalo de tiempo, y si estos son persistentes a lo largo de la serie.
4. Determinar los valores (p,d,q). El valor *p* y *q* se obtendrán con las autocorrelaciones y autocorrelaciones parciales, en donde los rezagos que salgan de la banda de confianza indicarán una significancia estadística, y serán los que ayuden a determinar el orden autorregresivo y de media móvil. Por su parte el valor *d* se fijará acorde a las pruebas de White y raíz unitaria, donde se examinará si la serie es estacionaria, si no es así, se tendrá que diferenciar *d* veces hasta que pase las pruebas. Posteriormente se estimarán los parámetros del modelo.

Para evaluar si la serie tiene media constante se debe hacer una regresión lineal por mínimos cuadrados, para este caso, dado que se tiene al tiempo como variable independiente, el modelo obtenido tiene la forma:

*Yi= β1+ β2 xi2 + εi* (2.20)

Donde *β1* es la constante del modelo, *β2* el coeficiente de la variable independiente*, xi2* es la variable independiente *y εi* representa el error.

En este caso se comprueba si el coeficiente que acompaña a la variable independiente es significativo o no, con lo que se podrá observar si la serie tiene media constante. Para ello se aplica la siguiente prueba de hipótesis:

*H0: β2 = 0 H1: β2 ≠ 0*

Donde la región de rechazo:

Valor-p > 𝛼

Siendo 𝛼 el nivel de confianza al que se está trabajando.

Si se acepta *H0*, se afirmará que la serie tiene media constante.

Para comprobar que la varianza es constante, se tienen que guardar los residuos del modelo de regresión lineal para aplicar la prueba de White, en este caso se guardan los errores de la regresión sobre la serie diferenciada y se le hará una regresión auxiliar de orden 2 a los residuos, la cual tendrá la forma:

*ε2i= β0+ β1 yj + β2 y2j* (2.21)

Posteriormente, se calcula el valor del estadístico chi-cuadrado, el cual sigue la expresión:

*χ²= nR2* (2.22)

Siendo *n* el número de valores que componen a la serie y *R2* será la bondad de ajuste de la regresión auxiliar con *k* grados de libertad, los cuales se determinan por la cantidad de regresores excluyendo al término constante (si es que lo hay).

La prueba de hipótesis de esta prueba es:

*H0 : la serie tiene varianza constante*

*H1: la serie no tiene varianza constante*

Si el valor-P *≥ 0.05,* se aceptará la hipótesis nula

El último test que se debe hacer a la serie de tiempo para corroborar que ésta sea estacionaria, es la prueba de la raíz unitaria o también conocida como la prueba de Dickey Fuller (Dickey, 1979). Frecuentemente se emplea el caso del test aumentado, el cual analiza si la serie de tiempo es una caminata aleatoria con intercepto, una caminata aleatoria con tendencia e intercepto, o una caminata aleatoria.

Se usó el siguiente modelo, para ilustrar la prueba de Dickey Fuller:

*Yt = αYt-1 + β Xt + εt* (2.23)

Donde *Xt* representa a alguna constante o tendencia, *α* y *β* son coeficientes estimados por la regresión y *εt* es el término del error.

Si /*α*/*≥ 1*, se estara ante una serie no estacionaria.

Para el caso de la prueba aumentada, se tiene que restar *Yt-1* a ambos lados (Ortiz, 2012), quedando:

*ΔYt = γYt-1 + β Xt + εt* (2.24)

Donde *γ= (α -1)* y *ΔYt = Yt - Yt-1*

La prueba de hipótesis es la siguiente:

H0: la serie tiene raíz unitaria. H1: la serie no tiene raíz unitaria.

El rechazo de la hipótesis nula se da cuando el *valor-P ≤ 0.05*.

Para validar el test anterior, se comprobó que no existieran auto correlación en los residuos, esto se realizó a través del estadístico Durbin Watson.

Buscando en la tabla del estadístico a *n* y a *K* , siendo *n* la cantidad de datos

de las que se compone la serie y *K* el número de regresores, se obtuvó el valor *du,* el cual delimitará la región *4-du>D>du*, para afirmar que no hay autocorrelación.

El estadístico se calcula a través de:

 (2.25)

Los rezagos *rk* que no sean significativos a un nivel de confianza del 95%, tendrán que estar comprendidos en el intervalo:

*±1.96*√1/𝑁 (2.26)

Donde *N* es el número de los datos que comprende la serie de tiempo. Para calcular la Autocorrelación del rezago *k,* se emplea:

*ρk= αk*/*α0* (2.27)

Siendo:

y

*α k= (Σ (Yt-Ῡ) (Yt+k-Ῡ))/n* (2.28)

*α 0= Σ (Yt-Ῡ)/n* (2.29)

Donde *n* es el tamaño de la muestra

Otro elemento a considerar en el modelado ARIMA es si la serie estudiada presenta estacionalidad. En la Figura 2.11 se tiene una serie temporal que posee una estacionalidad anual particular (se repite cada 12 meses), la forma de los correlogramas parciales y simples es la siguiente.

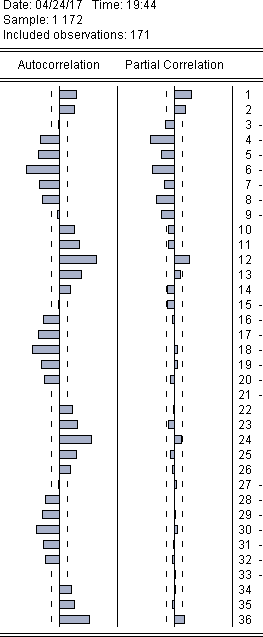


Figura 2.11 Correlogramas simple y parcial de una serie de tiempo que tiene un fuerte componente estacional anual.

Fuente: Elaboración propia con Eview.

Como se aprecia en el correlograma parcial, el rezago doceavo es estadísticamente significativo, lo cual implica que el valor actual está relacionado con los doce valores pasados de la serie. La estacionalidad se hace más notoria en el correlograma simple, pues tiene una forma marcadamente repetitiva, teniendo rezagos significativos en torno a los valores duodécimo, vigésimo cuarto y al trigésimo sexto, es decir, el valor presente estará relacionado con el valor que adopta la variable doce posiciones atrás.

Es poso fiable hacer una aseveración solo con un acercamiento visual , por lo que se debe de aplicar la siguiente prueba de hipoteis para comprobar la estacionalidad.

*H0: 𝜇1= 𝜇2= 𝜇3=…= 𝜇k*

*H1: Al menos dos medias poblacionales son diferentes*

Para que *H0* no se rechace, el valor-P debe de ser mayor a 0.05, a un nivel de confianza del 95%.

Si la hipótesis nula se rechaza, se estará ante una serie que muestra estacionalidad, debido a que la demanda de gas natural depende del mes.

Por último, para determinar el orden de *p* y *q* se emplean las Funciones de Autocorrelación Parcial (PAC) y la Función de Autocorrelación (AC), para construir los correlogramas pertinentes.

Los valores PAC estarán asociados a los términos AR y los valores AC a los términos MA. En donde aquellos rezagos que sobrepasen los anchos de la banda de los correlogramas, serán los que muestren una significancia estadística y serán usados para construir el modelo.

* + - * 1. Proceso autorregresivo Integrado de media móvil y variables exógenas (ARIMAX).

El modelo ARIMAX es una variante del modelo ARIMA, el cual considera a las variable explicativas exógenas como regresores, para pronosticar el valor de la variable dependiente en un tiempo dado. Son variables exógenas debido a que la información que hay en ellas proviene de una fuente diferente a la de la variable que se desea pronosticar (Gujarati, 2012).

El modelo matemático ARIMAX (p,q,n) está conformado por *p* elementos autorregresivos, *q* términos de media móvil y *n* variables explicativas. Su expresión matemática asociada es:

*Y t* = (𝜃*1 X1, t +***…** *+* 𝜃 *n Xn, t) +* (𝜇 +𝛼 *1 Zt –1+***…** *+* 𝛼 *q Zt -q)+* (*Φ 1 Y t-1 +***…** *+ Φp Y t-p+ ε t)+ E t* (2.30)

Dónde *X1,t ,* **…** *, Xn,t* son las variables explicativas del modelo y 𝜃*i* son sus coeficientes , los cuales tienen un valor apropiado ; los términos que contienen a

𝛼*i* representan el modelo MA y 𝜇 es una constante; los elementos que contienen a *Φi* conforman al modelo de AR y *εt* es su error asociado ; y *Et* representa a los residuos.

Para la construcción de un modelo ARIMAX adecuado, se seguirán los siguientes pasos:

1. Seleccionar las variables exógenas del modelo basado en un razonamiento económico.
2. Analizar las variables del modelo para determinar cuáles tienen una relación estadísticamente significativa.
3. Construir el modelo ARIMA para la variable de interés, agregando la información de las variables explicativas.
4. Calcular los parámetros del modelo.

Se deben de analizar las variables que entraran al modelo, para determinar cuáles tienen una relación estadísticamente significativa.

Para poder determinar qué variables tienen una relación estadísticamente significativa con la demanda del gas natural en México, se usaron los siguientes pasos:

1. Se realizó una regresión lineal múltiple, en donde se consideraron todas las variables, de ahí se analizó el valor-P para cada una de ellas con lo que se ubicaron cuáles son significativas para el modelo
2. Se quitó la variable que tuvo un P-valor mayor, y se corrió de nuevo la regresión múltiple, repitiendo el paso anterior, hasta dejar solo aquellas variables que fueron significativas para el modelo.

Para la selección de las variables que son relevantes para el modelo, se aplicó la siguiente prueba de hipótesis:

H0: la variable es estadísticamente significativa H1: la variable no es estadísticamente significativa

El criterio de rechazo es:

Valor-p > α

Donde *α* es el nivel de significancia estadística que se maneja, siendo, para este caso, de 0.05.

De la regresión anterior se genera una serie de tiempo con los residuos del modelo, ya que con ésta se trabaja para determinar el número de términos autorregresivo y de media móvil que se incluyeran.

Primero, se ve si las series están cointegradas, es decir, si tienen una relación estadística que no es espuria. Para eso se ocupa la prueba de la raíz unitaria, que tiene como hipótesis:

H0: Las series están cointegradas H1: Las series no están cointegradas

La aceptación de la hipótesis nula se da cuando el argumento de Dickey Fuller es menor a los resultados obtenidos por la prueba de valores críticos.

Para evaluar los modelos ARIMA y ARIMAX, es usual que se recurra al Criterio de Información de Akaike (CIA), el cual está dado por:

*CIA=2k-2Ln(L)* (2.31)

Donde *k* es el número de parámetros del modelo y *L* es el máximo valor de la función de verosimilitud. La tendencia en su uso es que el modelo con menor CIA, será el preferido, ya que premia la bondad de ajuste y castiga al número de parámetros usados, es decir, trata de dar una base para elegir un modelo simple y útil.

##### Simulación Matemática

La simulación puede entenderse como el proceso de aplicar un modelo matemático a un sistema real con el objetivo de aprender su comportamiento y detectar elementos que no fueron apreciados con anterioridad, para así alcanzar los objetivos propuestos (Rosenberg, 2000).

Con los avances tecnológicos actuales se dispone de software para realizar simulaciones de manera rápida, fiable y a un costo asequible.

El uso de la simulación es importante, ya que es posible estudiar las alteraciones que se tienen en los resultados que nos provee el sistema cuando los parámetros del modelo se alteran; es posible entender de mejor manera el comportamiento del sistema, lo que nos conduce a generar conocimiento que puede ser implementado para su operación y mejora; así mismo se detectan las variables que tienen una mayor incidencia en el comportamiento del sistema y las interrelaciones que guardan; y por último es posible entender sistemas de los que se tienen muy poca o ninguna información, logrando anticipar los resultados que se tendrán en la realidad.

##### Lenguaje de Programación R

El lenguaje R fue desarrollado por Ross Ihaka y Robert Gentleman (Peng, 2015) en el Departamento de Estadística de la Universidad de Auckland. Es un lenguaje que se ha popularizado debido a la sintaxis sencilla que tiene y a la versatilidad para realizar cálculos estadísticos.

El lenguaje R se ejecuta en casi cualquier plataforma informática y sistema operativo estándar; es un lenguaje de código abierto, lo que significa que cualquiera es libre de adaptar el software a la plataforma que elija, por lo que R ha sido ejecutado en tabletas modernas, teléfonos, PDA y consolas de videojuegos (Peng, 2015). Otra característica del lenguaje R es que sus librerías se actualizan constantemente, siendo el mes de octubre cuando se hacen la mayor parte de incorporaciones y son puestas a disposición del público en general; a lo largo del

año, se realizarán versiones de correcciones de errores a menor escala según sea necesario.

Las versiones actualizadas indican un desarrollo activo del software y asegura que los errores sean atendidos de manera oportuna. Cabe mencionar que mientras los desarrolladores centrales controlan el árbol fuente primario para R, muchas personas en todo el mundo hacen contribuciones en forma de nuevas funciones, correcciones de errores o ambas (Peng, 2015).

Otra ventaja clave que R tiene sobre muchos otros paquetes estadísticos es su capacidad gráfica, elemento muy atractivo para investigadores que buscan realizar gráficos de alta calidad para incorporarlos a las publicaciones en las que trabajan (Peng, 2015), lo que ha contribuido a que R se popularice en el ámbito académico y de investigación.

##### MATLAB.

MATLAB es un lenguaje de programación desarrollado a fines de los años 1970 por Cleve Moler de MathWorks (Ville y Jussi, 2010). Fue desarrollado originalmente mediante el uso de Fortran. A mediados de la década de 1980 Cleve Moler junto con Jack Little re-escribieron MATLAB con el lenguaje de programación C (Ville et al., 2010). Finalmente, en el 2000 MATLAB fue una vez más reescrito para usar el último conjunto de bibliotecas destinadas a la manipulación de matrices.

MATLAB se utiliza principalmente para computación numérica en matemáticas y otros campos científicos. Las caracteristicas que ofrece MATLAB lo hace muy atractivo, ejemplo de ellas son el trazado de datos y funciones, facilidad de creación de interfaces de usuario e interconexión con otros lenguajes de programación como C, C ++, Python y Fortran (Ville, et al. 2010). Sin embargo, hay otras herramientas incluidas en MATLAB, como el motor simbólico, que

permite el cálculo simbólico, y Simulink, que agrega el sistema simulación y diseño basado en modelos.

MATLAB es un producto patentado de MathWorks que requiere la compra de licencia para ser utilizado como un software de desarrollo. MathWorks también proporciona funcionalidades adicionales a través de las cajas de herramientas que se pueden comprar individualmente. Al ser un software con licencia, MathWorks garantiza que su producto se encuentre actualizado, por medio de incorporaciones bianuales

##### Statgraphics.

El programa Statgraphics es un paquete estadístico enfocado al análisis de datos. Es distribuido por Statpoint Technologies, una compañía privada con sede en Warrenton, Virginia.

Fue creado en 1980 por el Dr. Neil Polhemus mientras trabajaba como profesor de estadística en la Universidad de Princeton (Cervantes, María y Rivera, 2006). La versión actual del programa, Statgraphics Centurion XVII, se lanzó en otoño de 2014, la cual está disponible en cinco idiomas: inglés, francés, español, alemán e italiano

Contiene más de 230 procedimientos que cubren desde la estadística descriptiva hasta modelos estadísticos avanzadas, dando como prestaciones el análisis de variación, métodos estadísticos básicos, análisis de datos categóricos, minería de datos, diseño de experimentos, análisis exploratorio de datos, análisis de datos de vida y confiabilidad, análisis de sistemas de medición, métodos multivariables, métodos no paramétricos, distribuciones de probabilidad, análisis de capacidad de proceso, análisis de regresión, determinación del tamaño de la muestra, seis sigma y lean, control del proceso estadístico, análisis y pronóstico de series temporales, visualización, etc.

##### Eviews.

Econometric Views (Eviews) es un programa enfocado primordialmente al análisis estadístico y econométrico de datos, siendo utilizado ampliamente en el contexto económico y empresarial; también es popular en el contexto académico y de investigación, debido a la amplia gama de opciones que ofrece y el bajo costo computacional en el que incurre al ejecutarse, haciéndolo un programa adecuado para el estudio de series temporales y datos de panel.

Los autores Barreiro, Vicente y Pintos (2004) indican que *“las áreas en las que se aplica Eviews con gran exito son en el Análisis y Evaluación de datos científicos, Análisis Financiero, Predicción macroeconómica, Simulación, Predicción de ventas y Análisis de costes”* (p.2)*.* Eviews surge como una nueva versión del programa MicroTSP de 1981 (Barreiro, et al. 2004), que consistía en un conjunto de herramientas para la manipulación de datos de series temporales.

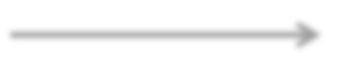
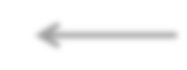
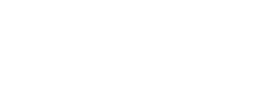
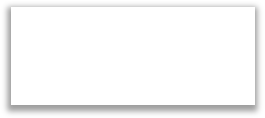
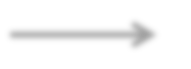
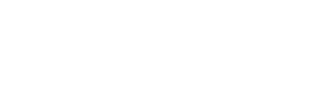
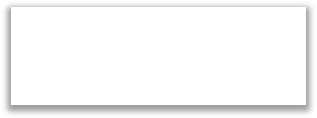
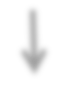
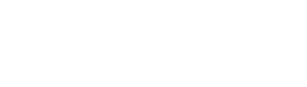
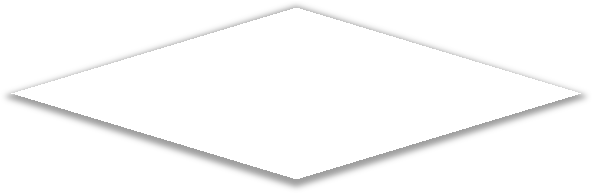
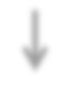
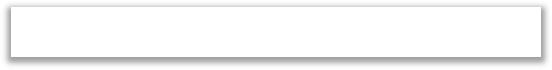
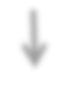
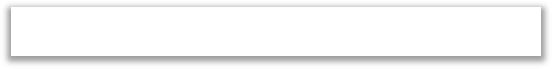
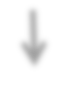
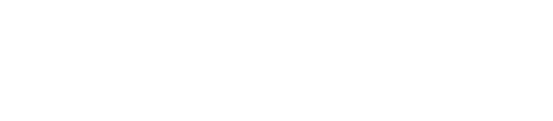
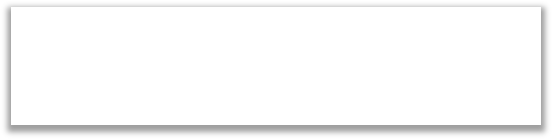
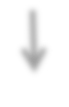
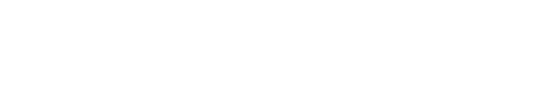
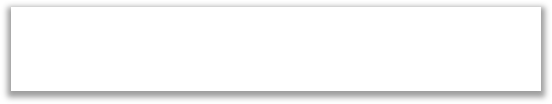
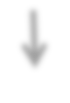
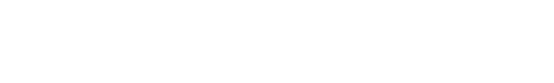
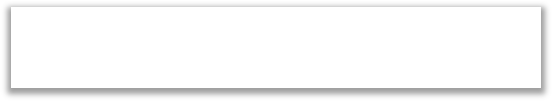
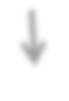
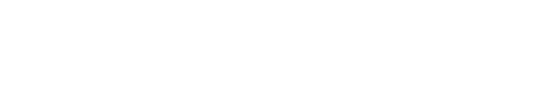
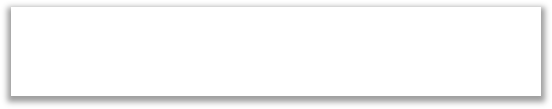
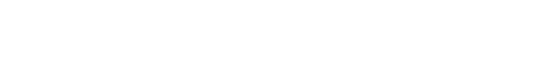
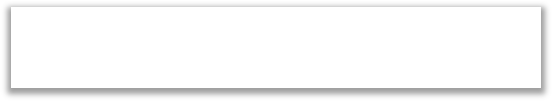
.

##### Marco Metodológico.

En el presente trabajo se usaron modelos matemáticos para pronosticar la demanda del gas natural, identificando cuál de ellos es el más apropiado para llevar a cabo dicha tarea. Dado que el interés se centra en encontrar los mejores medios para llegar al objetivo, el cual es fácilmente identificable y está inmerso en un sistema estructurado, se tuvo que emplear una metodología de investigación de operaciones.

##### Metodología de Investigación de Operaciones

Se usó la metodología de Investigación de Operaciones (Martínez, 2006). En la Figura 2.12 se esboza la secuencia lógica de dicha metodología:



**Definir el problema**

**Construir el modelo matemático**

**NO**

**¿Es válido el**

**modelo?**

**SI**

Figura 2.12 Diagrama de flujo de la metodología de Investigación de Operaciones (Martínez, 2006)

**Modificar el modelo**

**Implementar el modelo**

**Validar el modelo**

**Evaluar la solución**

**Simular el modelo a través de herramientas computacionales**

**Establecer las variables y parámetros del sistema**

**Recolectar datos del sistema estudiado**

Fase I: Definición del problema.

En esta fase se tiene que definir y analizar el sistema de interés, para así encontrar áreas de oportunidades, las cuales conducen al mejoramiento de los medios empleados en el cumplimiento del objetivo que posee el sistema. Para ello es primordial identificar los subsistemas que están implicados, los cuales mantienen interacciones que responden en gran medida a los contextos (cultural, espacial y temporal) en los que se encuentran inmersos y a los objetivos que tienen.

Como actividades a realizar en esta primera fase se tiene:

1. Definición del sistema de interés.
2. Obtención de datos cualitativos y cuantitativos del sistema.
3. Diagnosticar el estado del sistema.
4. Identificar las problemáticas existentes.
5. Reunir información relacionada a la problemática detectada.
6. Determinar el estado del arte del problema.
7. Plantear formalmente el problema que se ha de investigar

Fase II: Modelación.

La modelación es una actividad cognitiva para representar y describir una porción de la realidad. En este caso, el interés está en la modelación matemática, la cual expresa el comportamiento del objeto de estudio en un lenguaje matemático. Para lograr dicha tarea, se deben seguir las siguientes actividades:

1. Establecer el marco teórico de investigación
2. Reunir datos que sean fiables sobre nuestro sistema
3. Establecer las variables que inciden en el comportamiento del sistema y las restricciones involucradas.
4. Determinar la dinámica del sistema.
5. Construir el modelo que mejor represente el sistema.

Fase III: Simulación.

La simulación puede entenderse como el proceso que se sigue para dar solución a un problema que involucra un sistema real, en donde a través de dicho proceso se adquiere experiencia de cómo mejorar la solución obtenida. Para llevar a cabo la simulación, se requiere:

* 1. Hacer una identificación del modelo.
  2. Ubicar las técnicas y algoritmos relacionados con el modelo.
  3. Seleccionar un lenguaje de programación o software acorde a las necesidades de la investigación.
  4. Verter toda la información involucrada en el modelo en la herramienta computacional seleccionada.
  5. Ejecutar el programa para obtener la solución.
  6. Evaluar los resultados y darles interpretación para saber si son correctos, o si se pueden mejorar.

Fase IV: Validación.

En esta fase se analiza si tiene sentido o no la solución del problema, ponderando si el modelo propuesto logra describir y predecir adecuadamente el comportamiento del sistema estudiado. Una forma de verificar la validez de un modelo, es contrastar los resultados obtenidos con los resultados históricos (pensando que no se está tratando con un sistema nuevo del cual no se tienen datos del pasado), es decir, el modelo es válido si en condiciones de datos de entrada iguales, se logra reproducir de manera adecuada el desempeño pasado.

Algunos puntos a tomar en cuenta para la validación son:

1. Realizar una evaluación de la solución otorgada por el modelo
2. Establecer los intervalos de control de los parámetros y datos de entrada.
3. Seleccionar el modelo matemático que nos dé la solución más acertada del objetivo que tiene el sistema.
4. Si existen investigaciones similares, analizar los resultados obtenidos con los de los demás trabajos, para señalar ventajas o desventajas.

FASE V: Implementación.

Aquí se deben de transformar los elementos contenidos en el modelo en instrucciones claras de operación, para lograr su reproducción sistemática. Esta fase involucra:

1. Construir una metodología para la implementación de la solución.
2. Llevar acabo predicciones lógicas.
3. Realizar conclusiones.

**Capítulo 3**

**Aplicación de la Metodología.**

### Capítulo 3. Aplicación de la Metodología

##### Definición del problema

El sistema del mercado del gas natural está integrado por dos subsistemas: el de la oferta y la demanda. Para el caso mexicano, la oferta de este energético se encuentra garantizada (PEMEX, 2016), pues PEMEX y las empresas estadounidenses tienen la capacidad, tanto en infraestructura y recursos naturales, para abastecer las necesidades del mercado mexicano, bajo las condiciones actuales, para los siguientes dos décadas. Es por ello que la demanda juega un papel preponderante en la dinámica del mercado de gas natural, determinada por el consumo que se da en las esferas públicas y privadas, y en una multiplicidad de ramas productivas que se encuentran sujetas a las fuerzas económicas, las cuales se hacen patentes a través del nivel de ingreso, en la variación de los precios de los bienes sustitutos, en el cambio tecnológico y en los hábitos de consumo (Mankiw, 2012).

Por ello, es necesario analizar el sistema de demanda de gas natural para comprender y detectar los elementos que la conforman, con el fin de pronosticar su comportamiento en el tiempo, lo cual proporcionará información valiosa que podrán usar los diferentes actores interesados en el tema.

A nivel mundial se han utilizado modelos de redes neuronales artificiales para pronosticar la demanda de gas natural en Polonia, como son los trabajos de Viet et al. (2001); por parte de los trabajos abocados al pronóstico de la demanda de gas natural en EUA que utilizan RNA, se tienen los realizados por Abrishami et al. (2014); en contra parte tenemos que el Banco Mundial (2016) realizó el pronóstico de los precios del gas natural hasta el año 2020 de Japón , Europa y EUA por medio de los modelo ARIMA.

En México tenemos trabajos elaborados por Viscidi (2014) y Monteforte (2003), donde se efectuó un análisis cualitativo del mercado de gas natural en México, resaltando cómo se forma la demanda de gas natural y cómo ésta interactúa con la oferta para la formación del sistema de precios; por parte del pronóstico de la demanda de gas natural se tienen las investigaciones realizadas por la SENER (2016), los cuales usan los métodos ARIMA para hacer una proyección de la demanda de gas natural hasta el año 2025.

En los trabajos realizados a nivel mundial y nacional no existe una cohesión del conocimiento económico con el matemático, siendo esta desvinculación un factor muy importante que se tomó en cuenta en esta investigación para lograr utilizar estas dos áreas de conocimiento e integrarlas para construir modelos matemáticos con variables elegidas a través de un análisis de corte económico. Además se emplearon no solo los modelos tradicionales de pronóstico de series de tiempo como lo son el ARIMA y el ARIMAX, sino que también se contemplaron modelos matemáticos de redes neuronales artificiales, los cuales han mostrado un buen desempeño para el pronóstico de la demanda de gas natural en Polonia (Mendiziuk, et al. 2003) y EUA. (Abrishami, et al. 2014).

Con base en la metodología propuesta por Martinez (2006); empleando el análisis realizado del contexto espacial, temporal y cultural; el estudio de la situación actual del pronóstico de la demanda de gas natural a nivel mundial y nacional que está contenido en el Capítulo I, es posible definir el problema de investigación como **“Pronóstico de la Demanda de Gas natural en México con un Enfoque Sistémico”.**

##### Modelación

* + 1. **Modelación ARIMA**

Al realizar las actividades propuestas por Kahforoushan (2010) para la modelación ARIMA, se tienen los siguientes resultados para la serie de tiempo de gas natural de México.

1. Cálculo de la demanda y elaboración gráfica de la serie de tiempo, para llevar a cabo una primera aproximación

La demanda del gas natural está dado por el Consumo Nacional Aparente (CNA), el cual se calcula como:

CNA= Producción interna + importaciones - exportaciones (3.1)

En la Figura 3.1 se muestra la gráfica del CNA para los meses que componen la serie del 2006 hasta diciembre del 2016:

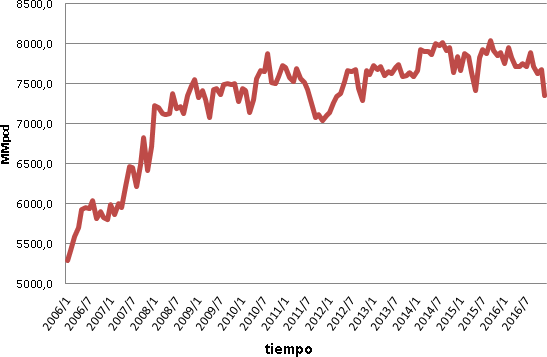


Figura 3.1. Datos de PEMEX de la demanda promedio mensual en MMpcd de gas natural en México periodo 2006-2016.

Fuente: Elaboración propia con datos de PEMEX (2016).

Se puede observar que al considerar la totalidad del periodo, se da un incremento en el nivel de consumo del gas natural, pero con fluctuaciones en los meses que componen a la serie; el valor máximo se alcanza en el mes de septiembre de 2015 con una demanda de 8035.6 MMpcd, y el nivel mínimo se sitúa en enero de 2006 con una cuantía de 5298.1 MMpcd.

1. Corroborar la estacionariedad.

A simple vista es imposible determinar con toda certeza si una serie es estacionaria o no, por lo que se deben aplicar las pruebas pertinentes para su verificación.

Si se calculan los valores de *β1* y *β2*, se tiene la siguiente ecuación:

*Precio del gas natural = 13,766 + 6374.9 tiempo*

En este caso se comprueba si el coeficiente que acompaña a la variable independiente es significativo o no, con lo que se podrá observar si la serie tiene media constante. Para ello se aplica la siguiente prueba de hipótesis:

*H0: β2 = 0 H1: β2 ≠ 0*

Donde la región de rechazo:

Valor-p > 𝛼

Siendo 𝛼= 0.04.

En la Tabla 3.1 se tiene la información de los coeficientes de la regresión lineal:

Tabla 3.1. Datos que muestran los valores de la pendiente e intercepto de la regresión lineal. Fuente: Elaboración propia con Eview.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Mínimos Cuadrados*** | ***Estándar*** | ***Estadístico*** |  |
| ***Variable*** | *Estimado* | *Error* | *T* | *Valor-P* |
| **Tiempo** | 13,76605 | 0.898548 | 15.32032 | 0,0000 |
| **Constante** | 6374.935 | 68.86755 | 92.56804 | 0,0000 |

El *valor-p* para la pendiente es menor a un nivel de significancia del 0.05, se rechaza la hipótesis nula, por lo que se puede afirmar con un 95% de confianza que la serie no tiene media constante.

Debido a que se viola una de las propiedades de la estacionariedad en las series de tiempo, se tiene que recurrir al método de diferencias, para tratar de convertir la serie en estacionaria.

En la figura 3.2 se tiene de forma gráfica que adopta la serie de tiempo al realiza las diferencias:

600.00

400.00

200.00

0.00

-200.00

-400.00

-600.00

tiempo

Figura 3.2. Serie de la demanda de gas natural en México en primera diferencia. Fuente: Elaboración propia.

MMpcd

1

7

13

19

25

31

37

43

49

55

61

67

73

79

85

91

97

103

109

115

121

127

Si se construye una regresión lineal por mínimos cuadrados, pero con los nuevos valores, se tendrán los resultados de la Tabla 3.2:

Tabla 3.2. Datos que muestran los valores de la pendiente e intercepto de la regresión lineal. Fuente: Elaboración Propia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Mínimos Cuadrados*** | ***Estándar*** | ***Estadístico*** |  |
| ***Variable*** | *Estimado* | *Error* | *T* | *Valor-P* |
| **Tiempo** | -0.1060 | 5.22E-05 | -2.0396 | 0.0435 |
| **Constante** | 0.9527 | 0.003971 | 2.398829 | 0.0179 |

Dado que el valor-P es menor al nivel de significancia de 0.05, se rechaza la hipótesis nula, con lo que se puede afirmar con un 95% de confianza, que el

coeficiente que acompaña la variable independiente es significativo, implicando que la serie de tiempo no muestra una media constante.

La situación anterior indicó que debía de hacerse una segunda diferencia a la serie de tiempo.

En la Figura 3.3 se presenta de manera gráfica a la serie de tiempo en segundas diferencias:

600.00

400.00

200.00

0.00

-200.00

-400.00

-600.00

-800.00

tiempo

Figura 3.3 Serie de la demanda de gas natural en México en segunda diferencia.

Fuente: Elaboración propia.

MMpcd

1

6

11

16

21

26

31

36

41

46

51

56

61

66

71

76

81

86

91

96

101

106

111

116

121

126

Al realizar la regresión lineal se tiene los resultados contenidos en la Tabla 3.3.

Tabla 3.3. Datos que muestran los valores de la pendiente e intercepto de la regresión lineal. Fuente: Elaboración propia con Eview.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Mínimos Cuadrados*** | ***Estándar*** | ***Estadístico*** |  |
| ***Variable*** | *Estimado* | *Error* | *T* | *Valor-P* |
| **Tiempo** | -0.08451 | 0.580786 | -0.144895 | 0.8850 |
| **Constante** | 2.086854 | 43.84259 | 0.047599 | 0.9621 |

Debido a que el P-valor es mayor al nivel de significancia de 0.05 se acepta la hipótesis nula, por lo que se puede afirmar con un 95% que la serie de tiempo tiene media constante.

Ahora para comprobar que la varianza es constante, se tienen que guardar los residuos del modelo de regresión lineal para aplicar la prueba de White, en este caso se guardan los errores de la regresión sobre la serie diferenciada y se le hará una regresión auxiliar de orden 2 a los residuos.

Realizando la prueba, se obtiene los resultados de la Tabla 3.4.

Tabla 3.4. Datos que muestran los P-valores para la prueba de White Fuente: Elaboración propia con Eview.

|  |  |  |  |
| --- | --- | --- | --- |
| F-statistic | 2.662029 | Prob. F(2,127) | 0.4537 |
| Obs\*R-squared | 5.230550 | Prob. Chi-Square(2) | 0.4531 |

Test Equation:

Dependent Variable: RESID^2 Method: Least Squares

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 0.002216 | 0.000522 | 4.242627 | 0.0000 |
| T^2 | 1.43E-07 | 1.36E-07 | 1.049619 | 0.2959 |
| T | -2.81E-05 | 1.84E-05 | -1.526754 | 0.1293 |

La prueba de hipótesis planteada es:

*H0 : la serie tiene varianza constante*

*H1: la serie no tiene varianza constante*

Si el valor-P *≥ 0.05,* se aceptará la hipótesis nula

Como se observa en los resultados obtenidos con Eview, el valor de la probabilidad Chi-cuadrado es mayor al nivel de significancia, por lo que es posible afirmar con un 95% de confianza, que la serie en segunda diferencia tiene varianza constante.

El último test que se debe hacer a la serie de tiempo para corroborar que ésta sea estacionaria, es la prueba de la raíz unitaria o también conocida como la prueba de Dickey Fuller (Dickey, 1979). En esta investigación se emplea el caso del test aumentado, el cual analiza si la serie de tiempo es

una caminata aleatoria con intercepto, una caminata aleatoria con tendencia e intercepto, o una caminata aleatoria.

La prueba de hipótesis es la siguiente:

H0: la serie tiene raíz unitaria. H1: la serie no tiene raíz unitaria.

El rechazo de la hipótesis nula se da cuando el *valor-P ≤ 0.05*.

Efectuando el Test, se obtienen los resultados contenidos en la Tabla 3.5:

Tabla 3.5. Datos que muestran los P-valores para la prueba de raíz unitaria Fuente: Elaboracion propia con Eview

**OPCION VALOR-P**

INTERCEPTO 0.0000

TENDENCIA E INTERCEPTO

0.0000

NINGUNA 0.0000

Dado que para las tres caminatas aleatorias se tiene un p-valor por debajo al nivel de significancia de 0.05, se rechaza la hipótesis nula a un nivel de confianza del 95%, es decir, la serie de tiempo no tiene raíz unitaria.

Para validar el test anterior, se comprobó que no existieran auto correlación en los residuos, esto se realizó a través del estadístico Durbin Watson. Buscando en la tabla del estadístico a *n* y a *K* , se obtuvó *n=133* y *K=1,* por lo que *du=*1.615, dando como intervalo 1.615*>D>*2.388.

En la Tabla 3.6 se dan los resultados de la prueba:

Tabla 3.6. Datos de la prueba Durbin Watson Fuente: Elaboración propia con Eview

|  |  |
| --- | --- |
| **Opción** | **Valor de Durbin Watson** |
| Tendencia | 2.011130 |
| Intercepto | 2.011693 |
| Ninguno | 2.009789 |

Dado que para las tres opciones, los valores obtenidos del estadístico Durbin-Watson caen dentro del intervalo calculado, se afirmó que no hay Autocorrelación.

Con esto se finalizan las pruebas para saber si la serie es estacionaria, afirmando que lo es, cuando se le aplican dos diferencias.

1. Estacionalidad.

Como primer paso, se verificó si la serie tiene un comportamiento repetitivo usando los correlogramas, los cuales se construyeron a partir de los rezagos considerados.

En la Figura 3.4 se muestran los correlogramas para la serie de tiempo de la demanda de gas natural en México.

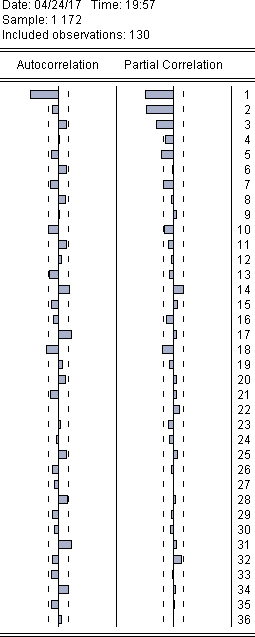


Figura 3.4. Correlogramas simple y parcial de la serie de tiempo de la demanda del gas natural en México en segunda diferencia.

Fuente: Elaboración propia con Eview

Gráficamente no se aprecia un patrón repetitivo en los rezagos, lo que estaría indicando la inexistencia de estacionalidad. Se tienen rezagos estadísticamente significativos en la correlación parcial en los valores uno, dos y tres, pero estos no se repiten a lo largo del correlograma, lo que fortalece la idea de que esta serie de tiempo no tiene estacionalidad.

Se corroboró lo anterior con la prueba de hipótesis:

*H0: 𝜇1= 𝜇2= 𝜇3=…= 𝜇k*

*H1: Al menos dos medias poblacionales son diferentes*

Para que *H0* no se rechace, el valor-P debe de ser mayor a 0.05, a un nivel de confianza del 95%.

En la Tabla 3.7 se dan los resultados del análisis de la varianza:

Tabla 3.7. Se muestra la probabilidad del análisis de la varianza. Fuente: Elaboración propia

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Fuente* | *Suma de*  *Cuadrados* | *Gl* | *Cuadrado*  *Medio* | *Razón-F* | *Valor-P* |
| Entre  grupos | 671553. | 11 | 61050.3 | 0.13 | 0.9996 |

Dado que el valor de la probabilidad es mucho mayor a nuestro nivel de significancia, no se rechaza la hipótesis nula, significando que el consumo de gas natural no depende del mes.

Debido al análisis anterior, no se tomó un modelo ARIMA con componente estacional.

1. Determinar los valores (p,d,q).

En esta actividad decidió el número de términos AR, MA y las veces que se diferenció la serie para construir el modelo.

En la sección pasada se mostraron las veces que se tuvo que diferenciar la serie para convertirla en estacionaria. El número de veces que se tuvo que diferenciar la serie, será el orden de integración que se le aplicó al modelo, es decir, el término *d* adoptó un valor de dos, ya que con él se satisfacen las pruebas de White y de Raíz Unitaria.

Para determinar el orden de *p* y *q* se emplearon las Funciones de Autocorrelación Parcial (PAC) y la Función de Autocorrelación (AC), para construir los correlogramas pertinentes.

En la Figura 3.5 se muestran los correlogramas de la AC y PAC, para la demanda del gas natural en México.

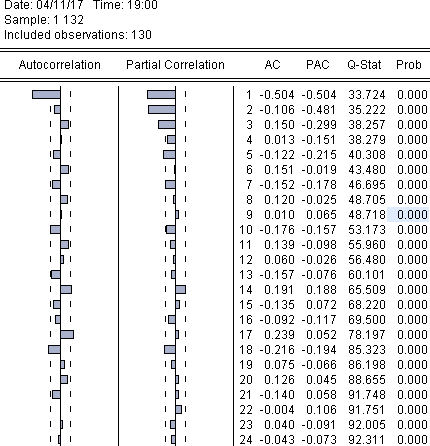


Figura 3.5. Correlograma simple y parcial de la serie de tiempo de la demanda del gas natural en México segunda diferencia.

Fuente: Elaboración propia.

Dado que no es fácil observar los valores que sobrepasan la banda de confianza, en la Tabla 3.8 se muestran los valores del test:

Tabla 3.8. Datos que muestran los valores para la AC y FCA, de la serie de tiempo de gas natural.

Fuente: Elaboración propia

|  |  |  |  |
| --- | --- | --- | --- |
| **Retraso** | **AC** | **FAC** | **Límite 95,0%** |
| 1 | -0,503496 | -0,503496 | ±0,171901 |
| 2 | -0,105737 | -0,481245 | ±0,171901 |
| 3 | 0,149863 | -0,298636 | ±0,171901 |
| 4 | 0,0126927 | -0,150825 | ±0,171901 |
| 5 | -0,12155 | -0,215406 | ±0,171901 |
| 6 | 0,151426 | -0,0191608 | ±0,171901 |
| 7 | -0,151818 | -0,177815 | ±0,171901 |
| 8 | 0,119554 | -0,0253033 | ±0,171901 |
| 9 | 0,00973932 | 0,064654 | ±0,171901 |
| 10 | -0,176505 | -0,157042 | ±0,171901 |
| 11 | 0,139013 | -0,0981436 | ±0,171901 |
| 12 | 0,059789 | -0,0258412 | ±0,171901 |
| 13 | -0,157088 | -0,0763574 | ±0,171901 |
| 14 | 0,191155 | 0,187693 | ±0,171901 |
| 15 | -0,134729 | 0,0715708 | ±0,171901 |
| 16 | -0,0922386 | -0,117162 | ±0,171901 |
| 17 | 0,23933 | 0,0517172 | ±0,171901 |
| 18 | -0,215687 | -0,194388 | ±0,171901 |
| 19 | 0,0752636 | -0,0661584 | ±0,171901 |
| 20 | 0,12551 | 0,0445345 | ±0,171901 |
| 21 | -0,140187 | 0,0576476 | ±0,171901 |
| 22 | -0,00438538 | 0,105699 | ±0,171901 |
| 23 | 0,0397872 | -0,0909952 | ±0,171901 |
| 24 | -0,043502 | -0,0731355 | ±0,171901 |

Para los términos AR se tiene como significantes al 1, 2, 3, 5, 7, 14 y 18. Para los términos MA se tiene como significantes al 1.

##### 3 . 2 . 2 Modelación del ARIMAX.

Para la construcción del modelo ARIMAX se siguieron las actividades propuestas por Gujarati (2010).

1. Seleccionar las variables exógenas del modelo basado en un razonamiento económico.

Las variables que fueron incluidas en los modelos se seleccionaron con base a un razonamiento económico, pues como cualquier bien a ser comercializado, el gas natural está sujeto a las fuerzas del mercado, por lo que se utilizaron los elementos que inciden en la demanda.

* 1. Ingresos: se tomó en cuenta al PIB , ya que éste puede ser calculado vía ingresos, por la expresión:

*PBI = R+ CKF + Ipm + EE* (3.2)

Donde *R* son las remuneraciones, *CKF* es el consumo de capital fijo, *Ipm* es el impuesto a la producción e importaciones y *EE* es el excedente de explotación.

* 1. El tamaño del mercado: las variables serán el número de empresas que han habido en las diferentes unidades de tiempo y la población nacional. Se considerarán ambas, ya que tanto las empresas como las personas adquieren gas natural como fuente de energía.
  2. Expectativas: la variable considerada es el Índice de Confianza del Consumidor, que mide el grado de optimismo que tienen los demandantes sobre la economía nacional y sobre su propia situación financiera. Ante una mayor confianza, se pueden tener comportamientos de consumo favorables de gas natural, ya que las personas pueden iniciar nuevos proyectos, como lo es la compra de inmuebles o electrodomésticos que requieran una fuente de energía para su funcionamiento. Así mismo se consideró al Índice de Confianza del Productor,

ya que el optimismo que manejen las entidades productivas, se traducirá en poner en marcha o no proyectos para ampliar la capacidad productiva de las mismas, lo que implica un cierto nivel de uso de energía.

* 1. Preferencias: aquí se incluye la participación en porcentaje que tiene el gas natural en el consumo total de energía por parte de las empresas.
  2. Precio de bienes sustitutos o complementarios: en éste se ponderaron los precios del gas LP y del petróleo mexicano, ya que ambos son bienes sustitutos del gas natural; además, se incluyeron los precios del gas natural de EUA y la mezcla de petróleo Brent y WTI, ya que México es importador de estos bienes. Por parte de los bienes complementarios no se consideró ninguna variable, esto es para simplificar el análisis, debido a que existen una gran cantidad de productos diseñados para funcionar con gas natural.

1. Analizar las variables del modelo, para determinar cuáles tienen una relación estadísticamente significativa.

Para determinar las variables que son relevantes para el modelo, se aplicó la siguiente prueba de hipótesis:

H0: la variable es estadísticamente significativa H1: la variable no es estadísticamente significativa

El criterio de rechazo es:

Valor-p > α

Donde *α* es el nivel de significancia estadística que se maneja, siendo para este caso, de 0.05.

En la Tabla 3.9 se proyectan los resultados de la regresión múltiple.

Tabla 3.9. Resultados de la primera regresión múltiple. Fuente: Elaboración propia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| CONFIANZA PROD | -31.17468 | 6.742131 | -4.623862 | 0.0000 |
| CONFIANZA CONSUMIDOR | 14.68744 | 7.253162 | 2.024971 | 0.0451 |
| ESTABLECIMIENTOS IND | 0.071880 | 0.006350 | 11.32048 | 0.0000 |
| PREFERENCIAS | 5415.622 | 3273.821 | 1.654220 | 0.1007 |
| PRECIO DEL GAS LICUADO | -5.606600 | 2.847576 | -1.968903 | 0.0513 |
| PRECIO DEL GAS NATURAL MEX | -49.75933 | 80.61934 | -0.617213 | 0.5383 |
| PIB CONSTANTE | 0.000662 | 0.000165 | 4.007429 | 0.0001 |
| POBLACION | -0.000202 | 6.36E-05 | -3.166565 | 0.0020 |
| PREC DE LA MEZCLA MEX | -21.36315 | 6.430777 | -3.322017 | 0.0012 |
| PRECIO DE LA MEZCLA BRENT | 2.736138 | 6.487914 | 0.421728 | 0.6740 |
| PRECIO DEL GAS NATURAL USA | 101.8490 | 77.38170 | 1.316189 | 0.1906 |
| PRECIO DE LA MEZCLA WTI | 21.78116 | 4.288468 | 5.079007 | 0.0000 |
| C | 4673.825 | 3861.949 | 1.210224 | 0.2286 |
| R-Squared | 0.932687 |  |  |  |

A un nivel de significancia de 0.05, la confianza de los productores, el número de establecimientos industriales, el PIB constante, el precio de la mezcla mexicana, el número de habitantes y el precio de la mezcla West Texas Intermédiate (WTI), tienen una relación estadísticamente significativa.

Se dejan las variables anteriores y se quitan la variable que muestran una menor relación, posteriormente se corre la regresión y se realiza el procedimiento anterior, hasta que solo queden variables con una relación estadísticamente significativa. En la Tabla 3.10 se presentan los resultados de la regresión múltiple donde solo se tienen variables significativas estadísticamente:

Tabla 3.10. Resultados de la última regresión múltiple. Fuente: Elaboración propia.

Variable Coefficien t Std. Error t-Statistic Prob.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CONFIANZA\_PRODUCTOR | -30.02877 | 5.452868 | -5.506968 | 0.0000 |
| CONFIANZA CONSUMIDOR | 15.96068 | 5.057692 | 3.155725 | 0.0020 |
| ESTABLECIMIENTOS IND | 0.068764 | 0.006018 | 11.42610 | 0.0000 |
| PIB\_CONSTANTE\_ | 0.000656 | 7.97E-05 | 8.228103 | 0.0000 |
| POBLACION | -0.000129 | 2.03E-05 | -6.345386 | 0.0000 |
| PREC\_MEZCLA MEX | -18.82579 | 3.415748 | -5.511468 | 0.0000 |
| PRECIO GAS NATURAL USA | 42.33330 | 15.91573 | 2.659840 | 0.0089 |
| PRECIO\_MEZCLA WTI | 19.06814 | 3.772223 | 5.054881 | 0.0000 |
| R-squared | 0.928578 |  |  |  |

Como se aprecia, la confianza de los productores, la confianza de los consumidores, el número de establecimientos industriales, el PIB constante, la población, el precio de la mezcla mexicana de petróleo, el precio del gas natural en EUA y el precio de la mezcla de petróleo WTI, son variables que tienen una relación estadísticamente significativa y son las que se usaron en el modelo multivariado.

1. Construir el modelo ARIMA para la variable de interés, agregando la información de las variables explicativas.

De la regresión del apartado anterior, se generó una serie de tiempo con los residuos del modelo, ya que con ésta se trabajó para determinar el número de términos autorregresivo y de media móvil que se incluyeron.

Se verifico que las series estuvieran cointegradas. En la Tabla 3.11 se dan los resultados de la prueba de raíz unitaria:

Tabla 3.11. Prueba de Raíz Unitaria Fuente: Elaboración propia.

**PRUEBA DE RAIZ UNITARIA CON CONSTANTE**

t-Statistic Prob.\*

Augmented Dickey-Fuller test statistic -9.566638 0.0000

|  |  |  |
| --- | --- | --- |
| Test critical values: | 1% level | -3.483312 |
|  | 5% level | -2.884665 |
|  | 10% level | -2.579180 |

**PRUEBA DE RAIZ UNITARIA CON CONSTANTE Y TENDENCIA**

Augmented Dickey-Fuller test statistic -9.516792 0.0000

Test critical values: 1% level -4.033108

5% level -3.446168

10% level -3.148049

**PRUEBA DE RAIZ UNITARIA SIN TENDENCIA Y SIN CONSTANTE**

t-Statistic Prob.\*

Augmented Dickey-Fuller test statistic -9.594086 0.0000

|  |  |  |
| --- | --- | --- |
| Test critical values: | 1% level | -2.583593 |
|  | 5% level | -1.943406 |
|  | 10% level | -1.615024 |

Dado que para las tres pruebas de raíz unitarias, el argumento de Dickey Fuller es menor a la prueba de valores críticos, se puede aceptar, hasta con un 99% de confianza, que las series están cointegradas.

Posteriormente se construyeron las autocorrelaciones parciales y simples, donde las primeras determinan el orden de los términos AR y las segundas fijan el número de términos MA. En la Figura 3.6 se muestran los correlogramas de la serie de tiempo.

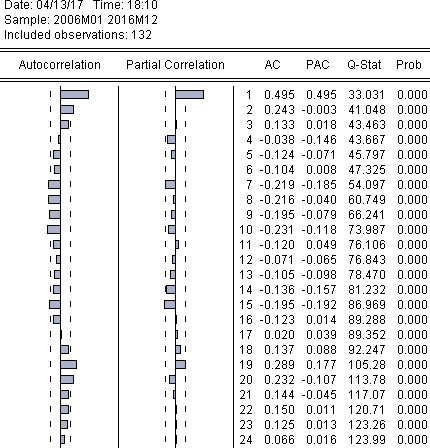


Figura 3.6. Datos que muestran los valores para la AC y FCA, de la serie de tiempo de gas natural con variables explicativas.

Fuente: Elaboración propia.

Para los términos AR se tienen como significantes a 1,7 ,15 y 19. Estos valores indican cuales observaciones tienen una relación estadísticamente significativa con el nivel de demanda actual.

Por su parte, para los términos MA se tienen al 1 y 2. Los valores anteriores señalan la ponderación de errores que se deben de considerar para ajustar el pronóstico de la demanda de gas natural en México.

##### Modelación de la Red Neuronal NAR

Los datos usados son los mismos del modelo ARIMA caso univariado, los cuales son los valores que adopta la demanda de gas natural en las diferentes unidades de tiempo del periodo 2006-2016, solo que a éstos se les aplicaron logaritmos base diez, esto con el fin de mandar los valores a un intervalo que facilitará el aprendizaje de las redes neuronales (Zarza, 2012).

Se dividieron los datos en tres, un 70% se destinó para el entrenamiento de la red, 15% fue destinado a la validación y 15% como prueba. Dado que modo de selección de los datos fue de manera aleatoria, los resultados obtenidos por una determinada arquitectura varían, por lo que se entrenó a la misma red cinco veces, con el fin de encontrar un mejor modelo, donde el criterio para seleccionarlo fue el que tuviera el menor error cuadrático medio (ECM).

Los autores Espinosa y Campos (2005) aseveran que *“las redes con bias, una capa sigmoide, y una capa de salida lineal son competentes para aproximar cualquier función* , por ello se probaron arquitecturas neuronales que tuvieran de una a dos capas, las dos capas fueron consideradas para apreciar si existe una mejoría en el desempeño de la red. La propuesta en esta investigación fue que la primera capa oculta estuviera conformada de cinco a doce neuronas y la segunda de cero a seis; se fijó un rezago, ya que en el capítulo dos se probó que la serie de tiempo no muestra un componente estacional fuerte, por lo que el ajuste y el pronóstico de los valores de la demanda tomarán en cuenta solo el valor anterior inmediato de la variable de estudio.

Dado lo anterior, se construyeron cincuenta y seis modelos de redes neuronales.

##### Modelación de la Red Neuronal NARX

Al igual que en la red NAR, en el aprendizaje se usará un bucle abierto, y en el pronóstico de varias unidades de tiempo se empleará un bucle cerrado.

Para evitar lo que se conoce como *sobreajuste*, el cual hace que la red sea muy sensible a los datos que conoce, pero no puede procesar debidamente los desconocidos, se necesita dividir los datos en tres grupos: entrenamiento, validación y prueba. El porcentaje que tendrá cada uno de ellos estará sujeto al juicio del modelador, Hagan et al. (2012) propone que sea el 70% para entrenamiento, 15% para prueba y 15% para validación.

Por último se debe de probar que los residuos son un ruido blanco con media cero y varianza conocida, verificando que los residuos no están correlacionados, esto se corrobora construyendo un intervalo de confianza del 95% y creando la función de autocorrelaciones, solo el valor del cero debe de caer fuera de la banda de confianza, los demás valores deben estar contenidos en el intervalo.

Los datos son los mismos que se emplearon en el modelo ARIMA caso multivariado, donde los targets son valores que adopta la demanda de gas natural en las diferentes unidades de tiempo del periodo 2006-2016. En las variables explicativas están los valores del PIB; el número de población y empresas; el índice de confianza del consumidor y del productor; la participación que tiene el gas natural en el consumo total de energía; y los precios del gas LP, de la mezcla mexicana de petróleo, de la mezcla Brent y WTI. De la misma manera que en el modelo NAR, a todos los datos se le aplicó logaritmo base diez.

Se dividieron los datos en tres, un 70% se destinó para el entrenamiento de la red, 15% fue destinado a la validación y 15% como prueba. Ya que el modo de entrada fue aleatoria, los resultados tenidos por una cierta arquitectura varían, por ello se entrenó a la misma red cinco veces, con el fin de encontrar el mejor modelo, siendo el criterio de selección el ECM.

Las arquitecturas de la red neuronal NARX están conformadas por una capa oculta que tendrán de cinco a doce neuronas y una segunda capa oculta de cero a seis neuronas; se fijó un rezago para la demanda del gas natural y también un rezago para las variables explicativas.

##### Simulación.

##### Simulación del Modelo ARIMA

Dado que no se puede saber la calidad del modelo hasta que éste haya sido puesto a prueba, se recurrió al Porcentaje del Error Absoluto Medio (MAPE), pero sobre un periodo de validación de doce meses, por lo que la serie se cortó en dos periodos, el primero comprende de 2006-1 al 2015-12 y se empleó para generar el modelo; y el segundo periodo que va de 2016-1 al 2016-2 se usó para evaluar el MAPE, y así saber el desempeño real que tiene el modelo.

Así mismo, se hizo la evaluación de los posibles modelos a través del Criterio de Información de Akaike (CIA).

Además, se propuso un modelo que fue generado en el lenguaje de programación R (para conocer el código acuda al Anexo 1), el cual se basa en calcular el número de términos que tendrá el modelo (se colocó un límite de doce términos de AR y MA), en el periodo 2006-1 al 2015-12, pero sujetos a la minimización del MAPE en el periodo 2016-1 al 2016-12. Posteriormente este número de términos AR y MA se emplearon para determinar los parámetros de todo el periodo completo.

Así mismo se usaron las posibles combinaciones de los valores significativos del correlogramas para seleccionar a los modelos. En la Tabla 3.12 se resumen los resultados obtenidos:

Tabla 3.12. Resultados obtenidos por los modelos ARIMA considerados. Fuente: Elaboración propia.

|  |  |  |
| --- | --- | --- |
| **MODELO ARIMA(p,d,q)** | **MAPE en el periodo de validación** | **CIA** |
| (1,2,1) | 2.828% | 1542.99 |
| (2,2,1) | 2.948% | 1539.41 |
| (3,2,1) | 2.962% | 1541.4 |
| (5,2,1) | 3.102% | 1544.27 |
| (7,2,1) | 3.394% | 1546.84 |
| (1,2,10) | 1.924% | 1551.11 |
| (14,2,1) | 4.097% | 1556.12 |
| (18,2,1) | 2.907% | 1544.11 |

Como se mencionó anteriormente, se eligió el modelo que obtuvo el menor MAPE en el periodo de validación y el que incurrió en el menor CIA, en todo el periodo. Para este caso el de menor error porcentual absoluto medio fue el ARIMA (1,2,10) y el de menor criterio de información de Akaike fue el ARIMA(2,2,1).

El ajuste que realizaron los modelos a los datos conocidos, es puesto a continuación, para tener una idea gráfica de su comportamiento:

El MAPE para los datos ajustados por el modelo ARIMA(2,2,1) fue de 1.784% y el ARIMA(1,2,10) incurrió en un error de 1,782%. En la Figura 3.7se tiene el comportamiento real de la demanda con el ajuste realizado por los modelos ARIMA.

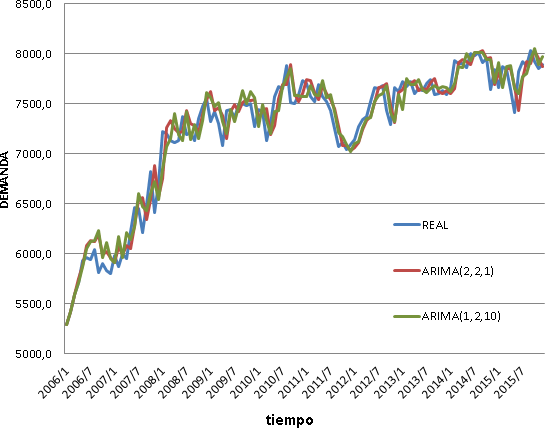


Figura 3.7 Comportamiento real de la demanda versus los valores ajustados para los tres modelos ARIMA.

Fuente: Elaboración propia.

En la Figura 3.8 se muestra el comportamiento de la variable real en el periodo de validación, y los pronósticos hechos por los dos modelos ARIMA.

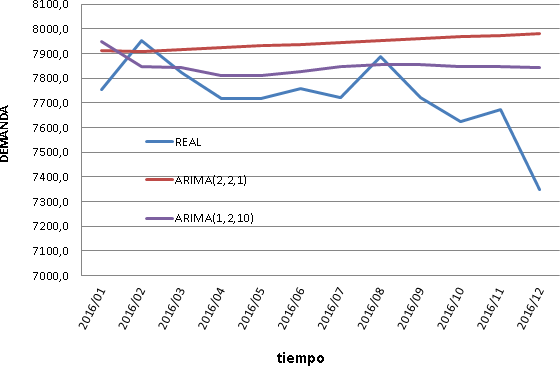


Figura 3.8 Comportamiento real de la demanda versus los valores pronosticados por los dos modelos ARIMA. Fuente: Elaboración propia.

Los pronósticos hechos para el modelo ARIMA(1,2,10) tuvieron un MAPE de 1.924%, pero logran acercarse al comportamiento que tuvo la demanda.

El modelo ARIMA(1,2,10) hace un pronóstico similar , pero con una cuantía mayor, por lo que obtiene un error de 2.98%.

Por último, se calcularon los coeficientes de los términos AR(*Yt-i*) y MA(*et.i*).

El modelo ARIMA(2,2,1) es expresado como:

*Yt= -0.2085Yt-1-0.2231Yt-2-0.9293et-9* (3.3)

Por su parte, el modelo ARIMA(1,2,10) está dado por:

*Yt=* -0.4088*Yt-1* -0.7459*et-1 -* 0.4914*et-2 +* 0.2407*et-3 +* 0.1808*et-4* - 0.3121*et-5 +* 0.1657*et-6 +…*

*…+* 0.0314*et* - 0.2065*et-8 +* 0.2065*et-9* -0.0253*et-10* (3.4)

##### Simulación del Modelo ARIMAX

Se usó como medida del desempeño al Porcentaje del Error Absoluto Medio, sobre un periodo de validación que abarcará doce meses, seccionando el intervalo 2006-1 al 2015-12, para generar el modelo, y el periodo 2016-1 al 2016-2 para evaluar el criterio del MAPE.

Se propuso un modelo construido en el lenguaje de programación ¨R¨ (para conocer el código acuda al Anexo 2), el cual se basó en calcular los parámetros para el periodo 2006-1 al 2015-12, pero sujetos a la minimización del MAPE en el periodo 2016-1 al 2016-2.

En la Tabla 3.13 se resumen los resultados obtenidos de los modelos construidos por los correlogramas y el código computacional.

Tabla 3.13. Resultados obtenidos por los modelos ARIMA a considerar. Fuente: Elaboración propia.

|  |  |  |
| --- | --- | --- |
| **MODELO ARIMAX(p,q,n)**  **Con variables exógenas** | **MAPE en el periodo de**  **validación** | **CIA** |
| (1,1,8) | 4.530% | 1530.04 |
| (1,2,8) | 6.509% | 1530.58 |
| (12,10,8) | 1.090% | 1533.78 |
| (7,1,8) | 4.319% | 1552.77 |
| (7,2,8) | 4.503% | 1543.08 |
| (15,1,8) | 5.137% | 1558.45 |
| (15,2,8) | 5.130% | 1538.2 |
| (19,1,8) | 5.919% | 1543.72 |
| (19,2,8) | 6.147% | 1545.67 |

Para este caso, el de menor error porcentual absoluto medio fue el ARIMAX(12,10,8) y el de menor criterio de información de Akaike fue el ARIMAX(1,1,8).

En la Figura 3.9 se muestra el ajuste que realizaron los modelos de los datos conocidos.

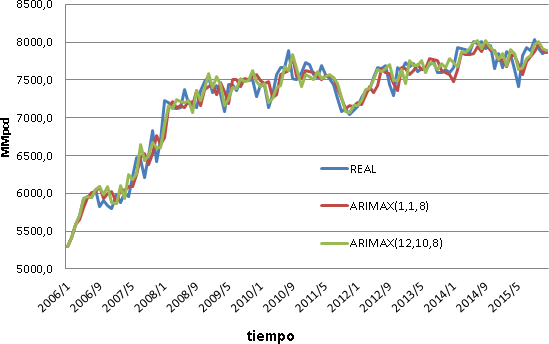


Figura 3.9 Comportamiento real de la demanda versus los valores ajustados por los dos modelos ARIMAX. Fuente: Elaboración propia.

El MAPE para los datos ajustados por el modelo ARIMA(1,1,8) fue de 1.63% y para el ARIMA(12,10,8) fue de 0.84%.

Así mismo, en la Figura 3.10 se muestra gráficamente el comportamiento de la variable real en el periodo de validación y los pronósticos hechos por los dos modelos ARIMA:

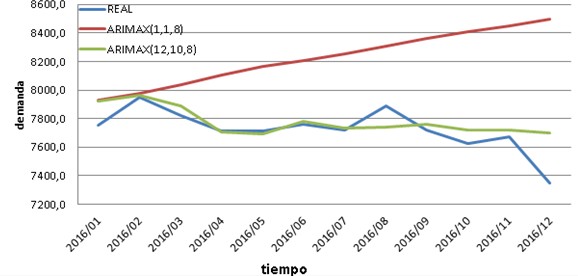


Figura 3.10. Comportamiento real de la demanda versus los valores pronosticados por los tres modelos ARIMAX.

Fuente: Elaboración propia.

Como se mencionó en la última tabla, el modelo ARIMAX(1,1,8) tuvo un MAPE de 4.53% y el ARIMA(12,10,8) incurrió en un error de 1.09%.

Algo notorio en los pronósticos hechos para el periodo de validación de los dos modelos, es que el modelo ARIMAX(1,1,8) hace una previsión a la alza, por lo que tiene errores importantes a partir del tercer periodo , mientras que el modelo ARIMAX(12,10,8) logra capturar la tendencia a la baja de la demanda y las irregularidades de ésta.

Por último, se calculan los coeficientes de los términos AR(*Yt-i*), MA(*et.i*) y las variables explicativas (*Xi*). Por lo que el modelo ARIMA(1,1,8) está dado por: *Yt=-0.1745 Yt-1-1.0383et-1+ 0.0388X1- 0.0058X2+13.9964X3-…*

…-13.7895X4+3x10-04X5+ 5.2408+ 9.1988X7 -4.1254X8 (3.5)

Por su parte, el modelo ARIMA(12,10,8) tiene la forma:

*Yt=* -0.7644*Yt-1*-0.4243*Yt-2 +* 0.2123*Yt-3 -*0.1897*Yt-4* -0.3704*Yt-5* -0.5132*Yt-6 -…*

*…+*0.2131*Yt-7* -0.0732*Yt-8* -0.7570 *Yt-9* -0.4276*Yt-10* -0.2028*Yt-11* -0.1723*Yt-12* -…

…-0.7503*et-1-…*0.5845*et-2-*0.5620*et-3 -*0.9732*et-4* -0.1284*et-5 -*0.8903*et-6 -*0.7237*et-7 +…*

*…*+0.5557*et-8+*0.5568*et-9 -*0.8937*et-10*-0.0559 *X1-9x10-04X2+*4.4133*X3*-17.8375*X4-…*

*…-5x10-04X5+*9.6891*X6+*58.818*X7-*9.6921*X8* (3.6)

Dónde:

*X1: el número de establecimientos industriales X2: el número de habitantes*

*X3: el índice de la confianza del consumidor X4: el índice de la confianza del productor X5: el Producto Interno Bruto*

*X6: el precio del barril de la mezcla de petróleo WTI*

*X7: el precio del barril de la mezcla mexicana de petróleo X8: el precio del barril de la mezcla de petróleo Brent*

##### Simulación del Modelo NAR

Cada modelo se simulo cinco veces en MATLAB, por lo que se tuvieron doscientos ochenta resultados diferentes. Desde que se simulo la red hasta guardar resultados, el tiempo de procesamiento fue de cinco minutos por arquitectura, por lo que el tiempo de simulación total fue de veintitrés horas.

De todas las arquitecturas, la que mostró el mejor desempeño fue la que tiene cinco neuronas en la primera capa oculta y una en la segunda, con un ECM de 9.06x10-9. Gráficamente se tiene la Figura 3.11:

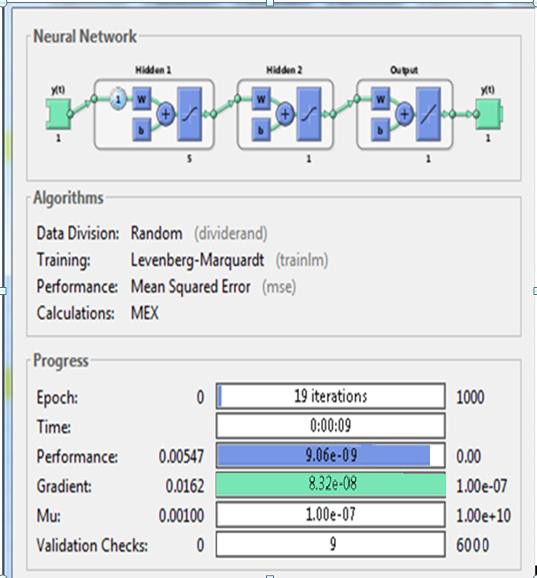


Figura 3.11. Modelo gráfico y resultados de la simulaciones de una red NAR con cinco neuronas en la primera capa oculta , una en la segunda y una en la capa de salida.

Fuente: elaboración propia con MATLAB

El MAPE en el que incurrió la red neuronal fue de 0.08%, y el ajuste realizado sobre los datos de entrenamiento, así como los pronosticados en la validación y el test , tienen la forma mostrada en la Figura 3.12.

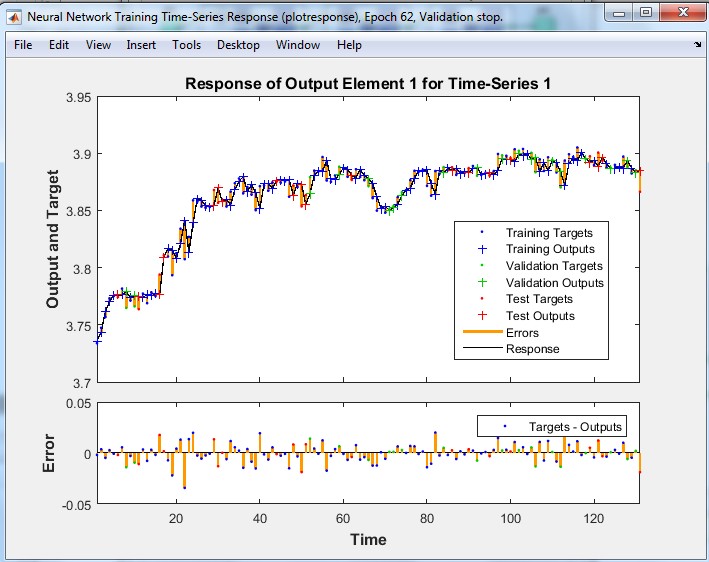


Figura 3.12. Ajuste que realizó la red NAR a los datos de entrenamiento, y los pronósticos hechos para los datos de validación y test.

Fuente: Elaboración propia con MATLAB

La regresión de los datos de entrenamiento que calculó la red, y los valores de test y validación se muestran en la Figura 3.13:

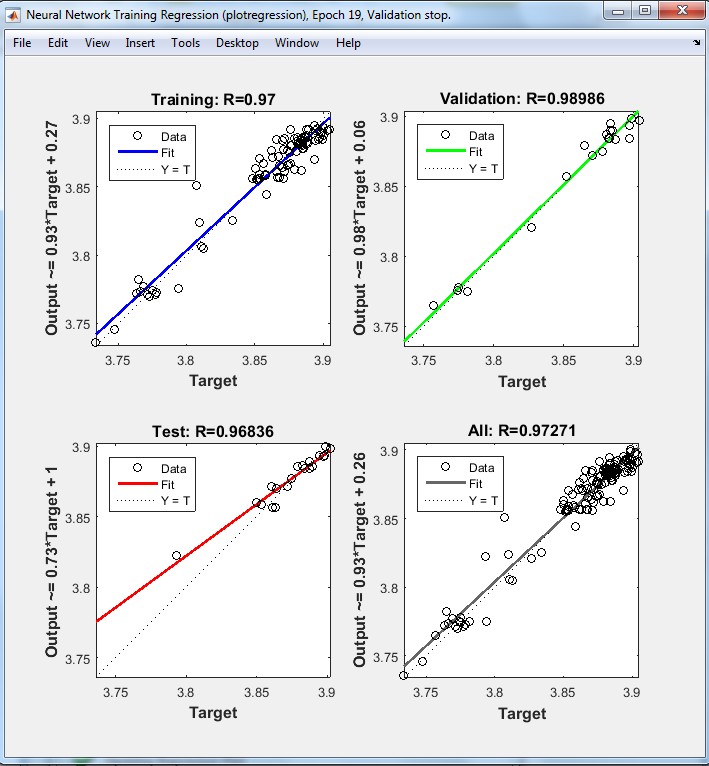


Figura 3.13 Regresión de los datos de entrenamiento, validación y test. Fuente: Elaboración propia con MATLAB

Como es posible notar, la red muestra una bondad de ajuste muy alta con una cuantía en total del 0.97271, siendo 1 el mejor valor que se puede obtener. Así mismo se ve que para la prueba y validación se tienen muy buenos resultados, donde se obtuvo una bondad de ajuste de 0.9683 y 0.9898 respectivamente, lo que indica que la red neuronal puede explicar valores que no conoce con una muy alta tasa de éxito, lo que a su vez significa que no existe un sobre ajuste o deficiencia en su desempeño.

En la Figura 3.14 se muestra el mejor desempeño de validación , el cual se alcanzó en la iteración 13, implicando una rápida convergencia:

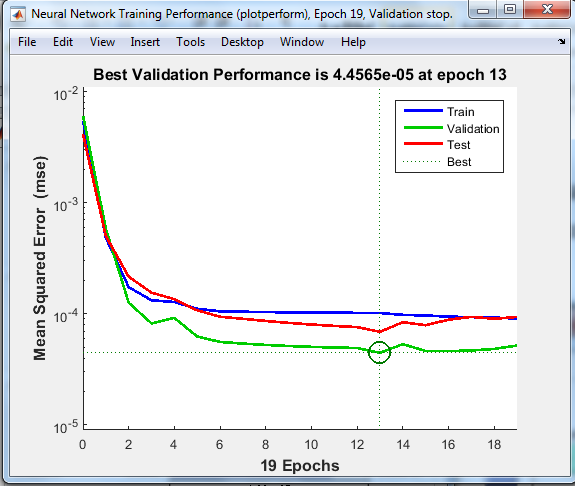


Figura 3.14 Gráfica que muestra el mejor desempeño de validación a través de las iteraciones del entrenamiento.

Fuente: Elaboración propia con MATLAB

##### Simulación del Modelo NARX

Se construyeron cincuenta y seis modelos de redes neuronales, los cuales se simularon cinco veces en MATLAB, por lo que se tuvieron doscientos ochenta resultados diferentes. Desde que se simulo la red hasta guardar resultados, el tiempo de procesamiento promedio fue de 8 minutos por arquitectura, por lo que el tiempo de simulación total fue de treinta y siete horas.

De todas las arquitecturas, la que tuvo el mejor desempeño fue la que tiene doce neuronas en la primera capa oculta y una en la salida, con un ECM de 1.04x10-21. En la Figura 3.15 se muestra gráficamente.

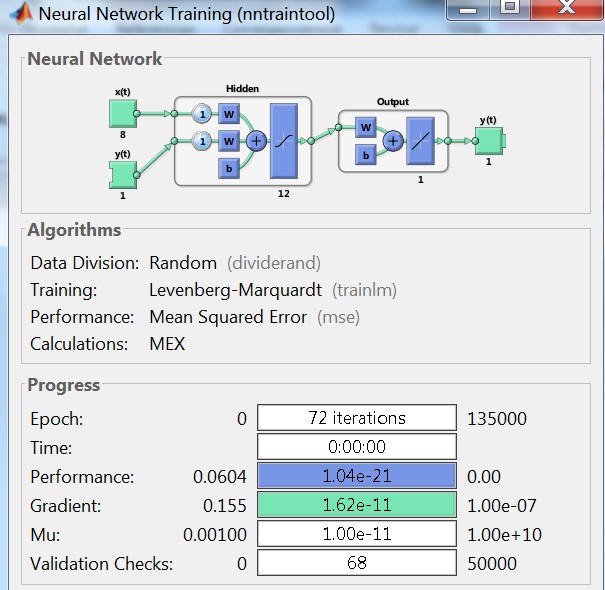


Figura 3.15 Modelo gráfico y resultados de la simulaciones de una red NARX con doce neuronas en la primera capa oculta y una en la capa de salida. Fuente: Elaboración propia con MATLAB

El MAPE en el que incurrió la red neuronal fue de 0.014%. El ajuste realizado sobre los datos de entrenamiento, así como los pronosticados en la validación y el test, tienen la forma de la Figura 3.16.

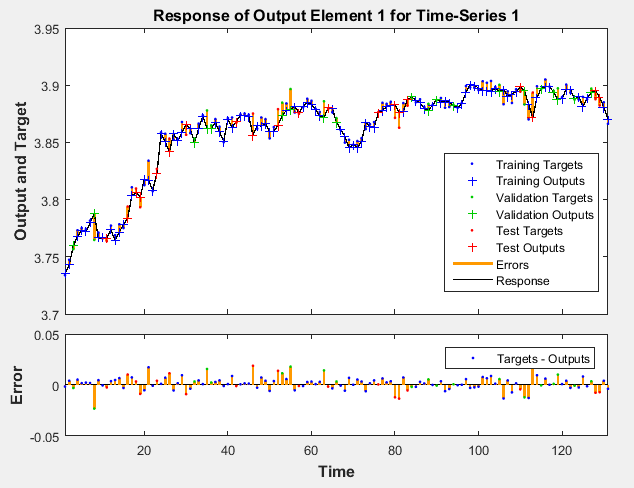


Figura 3.16. Ajuste que realizó la red NARX a los datos de entrenamiento, y los pronósticos hechos para los datos de validación y test.

Fuente: Elaboración propia con MATLAB

Visualmente se aprecia un buen comportamiento en los ajustes y las predicciones que hace la red neuronal, lo cual es una primera aproximación para evaluar el desempeño que tuvo.

La regresión de los datos de entrenamiento que calculó la red, y los valores de test y validación se muestran en la Figura 3.17:

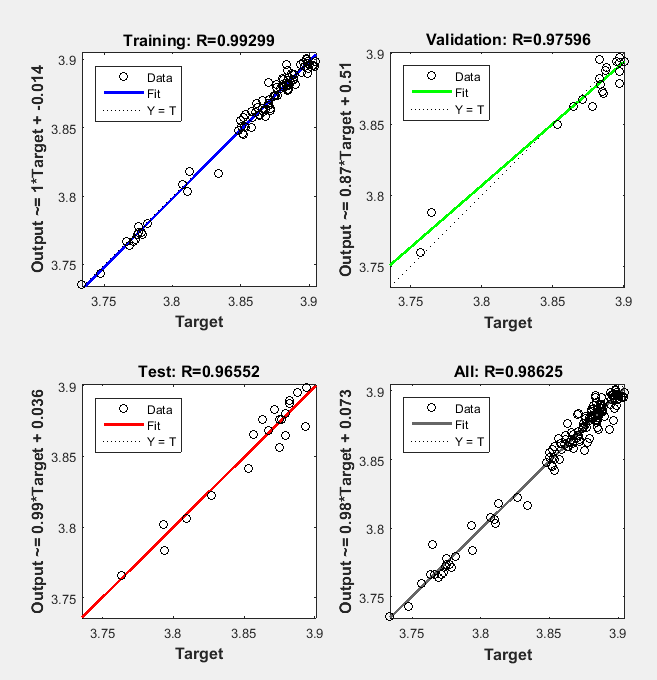


Figura 3.17. Regresión de los datos de entrenamiento, validación y test. Fuente: Elaboración propia con MATLAB

Como se observa, la red muestra una bondad de ajuste total de 0.98625, recordando que 1 es el mejor valor que se puede tener. Así mismo, se aprecia que para los datos de prueba y validación se tienen muy buenos resultados, consiguiendo una bondad de ajuste de 0.96552 y 0.97596 respectivamente, lo que indica que la red neuronal puede explicar valores que no conoce con una muy alta tasa de éxito. Dado estos valores, se puede concluir que la red no sobre ajustó los datos y no muestra deficiencia en su desempeño.

El mejor desempeño de validación se alcanza en la iteración 6, pero el entrenamiento paró hasta la iteración 72, en el que se alcanzó un mejor ajuste de los datos conocidos en el entrenamiento, pero empeoró el error cuadrado medio para los datos de validación y test. En la Figura 3.18 se muestra gráficamente.

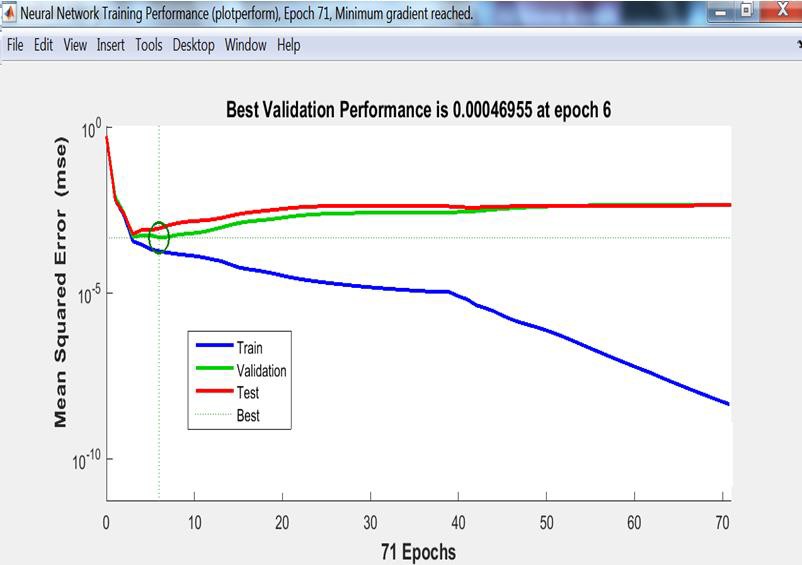


Figura 3.18 Gráfica que muestra el mejor desempeño de validación a través de las iteraciones del entrenamiento.

Fuente: Elaboración propia con MATLAB

##### Implementación

##### implementación del modelo ARIMA

Una vez que el modelo más adecuado fue seleccionado, se le implemento para obtener los resultados finales.

En la Tabla 3.14 se colocan los pronósticos hechos por los dos modelos para el periodo de enero del 2017 a diciembre del mismo año.

Tabla 3.14. Pronóstico hecho por los dos modelos para el periodo 2017-1 al 2017-12 en MMpcd.

Fuente: Elaboración propia.

|  |  |  |
| --- | --- | --- |
|  | ARIMA(2,2,1) | ARIMA(1,2,10) |
| 2017-1 | 7378,513 | 7337,688 |
| 2017-2 | 7406,061 | 7424,704 |
| 2017-3 | 7358,187 | 7325,309 |
| 2017-4 | 7327,993 | 7243,226 |
| 2017-5 | 7309,892 | 7321,756 |
| 2017-6 | 7285,218 | 7277,715 |
| 2017-7 | 7259,465 | 7259,235 |
| 2017-8 | 7235,37 | 7301,465 |
| 2017-9 | 7211,123 | 7263,134 |
| 2017-10 | 7186,557 | 7224,068 |
| 2017-11 | 7162,096 | 7185,65 |
| 2017-12 | 7137,68 | 7146,661 |

Al graficar los pronósticos, se tiene:

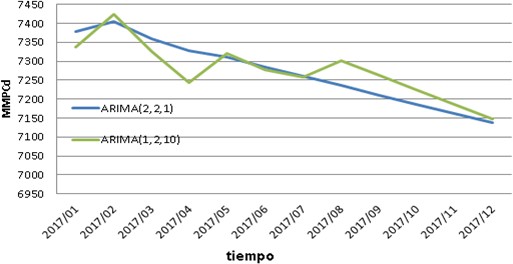


Figura 3.19 Pronóstico hecho por los tres modelos ARIMA para el periodo 2017-1 al 2017-2. Fuente: Elaboración propia.

Como se observa, los dos modelos hacen pronósticos a la baja, para los casos ARIMA(2,2,1) se hacen previsiones más lineales, mientras que el modelo ARIMA(1,2,10) tiene fluctuaciones dentro del periodo.

##### Implementación del Modelo ARIMAX

Los modelos ARIMAX elegidos para ser implementados en la solución del problema de interés fueron el ARIMAX(12,10,8) y el ARIMAX(1,1,8).

En la tabla 3.15 se colocan los pronósticos hechos por los dos modelos, para el periodo de enero del 2017 a diciembre del mismo año.

Tabla 3.15. Pronóstico hecho por los dos modelos para el periodo 2017-1 al 2017-2 en MMpcd.

Fuente: Elaboración propia.

|  |  |  |
| --- | --- | --- |
|  | ARIMAX(12,10,8) | ARIMAX(1,1,8) |
| 2017-1 | 7361,65 | 7346,544 |
| 2017-2 | 7449,17 | 7393,821 |
| 2017-3 | 7292,71 | 7330,51 |
| 2017-4 | 7252,80 | 7295,523 |
| 2017-5 | 7272,65 | 7299,663 |
| 2017-6 | 7220,62 | 7247,625 |
| 2017-7 | 7285,17 | 7232,574 |
| 2017-8 | 7296,78 | 7193,529 |
| 2017-9 | 7227,13 | 7170,048 |
| 2017-10 | 7256,16 | 7136,472 |
| 2017-11 | 7201,48 | 7109,444 |
| 2017-12 | 7101,97 | 7078,169 |

En la Figura 3.20 se tienen los pronósticos de manera gráfica:

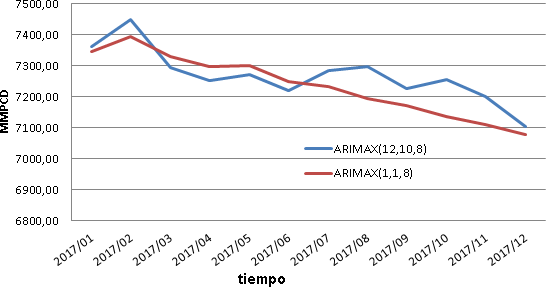


Figura 3.20 Pronóstico hecho por los tres modelos ARIMAX para el periodo 2017-1 al 2017-2. Fuente: Elaboración propia

Los dos modelos hacen pronósticos a la baja, siendo la previsión del modelo ARIMAX(1,1,8) más lineal que el hecho por el ARIMAX(12,2,10).

##### Implementación del Modelo NAR

Realizando el pronóstico de la red neuronal NAR , se tienen los niveles de demanda de gas natural mostrados en la Tabla 3.16.

Tabla 3.16. Pronóstico de la demanda de gas natural en MMpcd. para el 2017 de la red neuronal NAR Fuente: Elaboración propia.

|  |  |
| --- | --- |
|  | Red NAR |
| 2017-1 | 7337.46 |
| 2017-2 | 7322.09 |
| 2017-3 | 7302.42 |
| 2017-4 | 7277.60 |
| 2017-5 | 7247.41 |
| 2017-6 | 7213.33 |
| 2017-7 | 7179.35 |
| 2017-8 | 7150.71 |
| 2017-9 | 7130.60 |
| 2017-10 | 7118.53 |
| 2017-11 | 7112.02 |
| 2017-12 | 7108.73 |

La forma gráfica del pronóstico de la demanda de gas natural es presentado en la Figura 3.21:

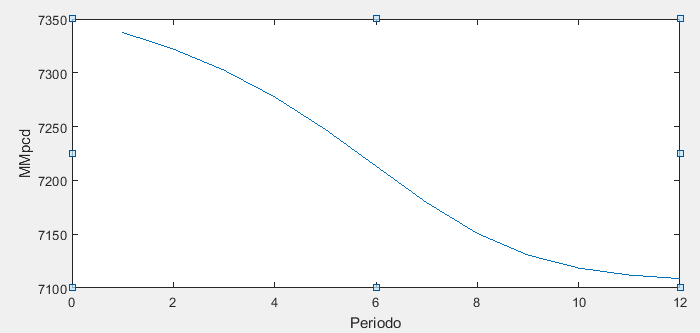


Figura 3.21. Pronóstico NAR de la demanda de gas natural del 1/2017 al 12/2017 Fuente: Elaboración propia

33

3.

Como se aprecia, el pronóstico es una curva suave con tendencia a la baja. Este comportamiento se evaluó en el capítulo cuatro, ya que al ser un modelo univariado, el pronóstico realizado puede ser conservador, al generar un comportamiento más estable sin fluctuaciones marcadas a lo largo de los meses del año 2017.

##### Implementación del Modelo NARX

En la Tabla 3.17 se compendian los pronósticos hechos con la arquitectura NARX para el periodo 2017-1/2017-12:

Tabla 3.17. Pronóstico de la demanda de gas natural para el 2017 de la red neuronal NARX Fuente: Elaboración propia.

|  |  |
| --- | --- |
|  | Red NARX |
| 2017-1 | 7828.74 |
| 2017-2 | 7813.93 |
| 2017-3 | 7997.63 |
| 2017-4 | 7804.34 |
| 2017-5 | 7683.31 |
| 2017-6 | 7645.57 |
| 2017-7 | 7672.19 |
| 2017-8 | 7756.52 |
| 2017-9 | 7726.69 |
| 2017-10 | 7657.82 |
| 2017-11 | 7582.46 |
| 2017-12 | 7551.78 |

En la Figura 3.22 se muestran gráficamente.

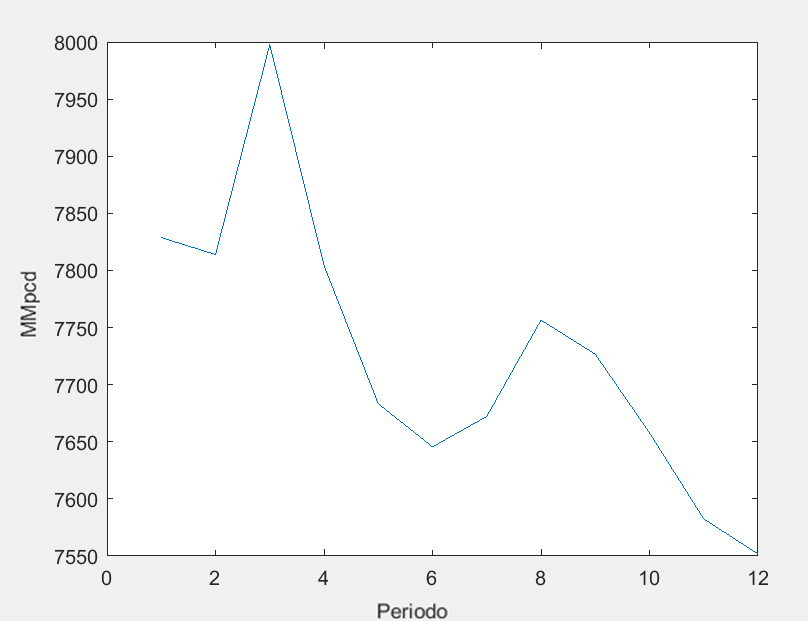


Figura 3.22. Pronóstico NARX de la demanda de gas natural del 1/2017 al 12/2017. Fuente: elaboración propia

Los pronósticos hechos tienen fluctuaciones dentro del periodo, a diferencia de los realizados por la red neuronal NAR, los cuales siguen el comportamiento de una curva suave.

La red NARX busca recrear el comportamiento de la demanda de gas natural, lo que puede implicar un mejor desempeño, premisa discutida en el capítulo cuatro.

##### Validación.

##### Validación del modelo ARIMA

Para saber si los modelos son útiles, los residuos deben de ser un ruido blanco, que cumplan el principio de aleatoriedad, y se ajusten a una distribución normal con media cero y varianza conocida.

1. Prueba de aleatoriedad:

La hipótesis a probar es:

*H0:* los residuos son aleatorios

*H1:* los residuos no son aleatorios

Si el *valor-P ≥0.05,* para las tres pruebas que se emplean*,* se aceptará la hipótesis nula, con un 95 % de confianza.

El primer test contabilizo el número de veces que la secuencia estuvo arriba o abajo de la mediana.

El segundo test midió la cantidad de veces que la secuencia ascendió o descendió.

El tercer test se recurrió a la suma de cuadrados de los primeros sesenta coeficientes de Autocorrelación.

Aplicando los tres test al modelo ARIMA(2,2,1), se tiene:

* 1. Corridas arriba o abajo de la mediana Estadístico z para muestras grandes = 0,641712 Valor-P = 0,521058

1. Corridas arriba y abajo

Estadístico z para muestras grandes = 0,39996 Valor-P = 0,689182

1. Prueba Box-Pierce

Prueba basada en las primeras 24 autocorrelaciones Estadístico de prueba para muestras grandes = 28,4386 Valor-P = 0,241975

Dado que para las tres pruebas el P-valor fue superior al nivel de significancia, no se rechaza la hipótesis nula, por lo que se puede afirmar con un 95% de confianza que los residuos son aleatorios.

Se realizó la misma prueba a los residuos del ARIMA(1,2,10), se tiene:

1. Corridas arriba o abajo de la mediana Mediana = -0,035

Estadístico z para muestras grandes = 1,19175 Valor-P = 0,119128

1. Corridas arriba y abajo

Estadístico z para muestras grandes = 0,03636 Valor-P = 0,536493

1. Prueba Box-Pierce

Prueba basada en las primeras 24 autocorrelaciones Estadístico de prueba para muestras grandes = 17,4832 Valor-P = 0,615097

Nuevamente no se rechazó la hipótesis nula, a un nivel de significancia del 0.05.

B) Prueba de normalidad de los residuos.

Aquí se recurrio a la prueba de Kolmogorov Smirnov, para ver si los residuos se ajustan a una distribución normal con media cero y varianza conocida. Por lo que la prueba de hipótesis es:

*H0 : los residuos se ajustan a una distribución normal con media cero y varianza conocida*

*H1: los residuos no se ajustan a distribución normal con media cero y varianza conocida*.

El rechazo de la hipótesis nula se da cuando el *valor-P ≤0.05*.

Para los residuos del modelo ARIMA(2,2,1) se tuvo el histograma presentado de la Figura 3.23.

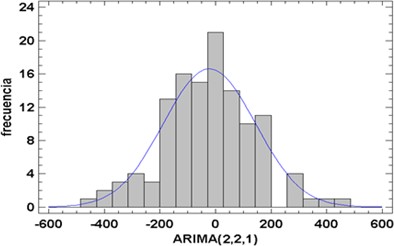


Figura 3.23 Frecuencia de los residuos del modelo ARIMA(1,2,1) . Fuente: Elaboración propia.

La Prueba de Kolmogorov-Smirnov dio un Valor-P de 0,966345, por lo que no se rechazó la hipótesis nula a un nivel de significancia del 0.05.

Por último se realizó la prueba al modelo ARIMA(1,2,10), en la Figura 3.24 se presentan de manera gráfica:

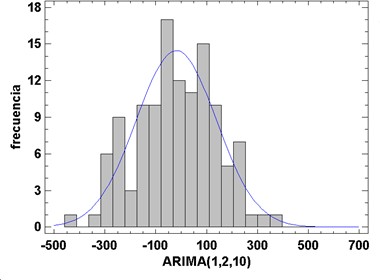


Figura 3.24 Frecuencia de los residuos del modelo ARIMA(1,2,10). Fuente: elaboración propia.

La Prueba de normalidad dio un Valor-P de 0,848428, por lo que no se rechazo la hipótesis nula a un nivel de confianza del 95%.

Con lo anterior, se llegó a la conclusión de que los dos modelos cumplen los supuestos de normalidad y aleatoriedad. Por lo que se afirma que los residuos son un ruido blanco.

##### Validación del modelo ARIMAX

Para que los modelos sean útiles, los errores deben de ser un ruido blanco que cumpla el principio de aleatoriedad y se ajuste a una distribución normal con media cero y varianza conocida.

1. Prueba de aleatoriedad:

Nuevamente, la hipótesis a probar fue:

H0: los residuos son aleatorios H1: los residuos no son aleatorios

Si el *valor-P ≥0.05,* se aceptará la hipótesis nula, con un 95 % de confianza.

Se aplicaron los test al modelo ARIMAX(1,1,8) y se obtuvó:

* 1. Corridas arriba o abajo de la mediana Estadístico z para muestras grandes = 1,3751 Valor-P = 0,169101

1. Corridas arriba y abajo

Estadístico z para muestras grandes = 0,03636 Valor-P = 0,97099

1. Prueba Box-Pierce

Estadístico de prueba para muestras grandes = 60,0483 Valor-P = 0,162882

Dado que para las tres pruebas, el P-valor fue superior al nivel de significancia, no se rechazó la hipótesis nula, por lo que se afirma, con un 95% de confianza, que los residuos son aleatorios.

Aplicando el mismo procedimiento al modelo ARIMAX(12,10,8), se tuvo:

1. Corridas arriba o abajo de la mediana Estadístico z para muestras grandes = 2,47517 Valor-P = 0.783297
2. Corridas arriba y abajo

Estadístico z para muestras grandes = 0,909001 Valor-P = 0.117939

1. Prueba Box-Pierce

Estadístico de prueba para muestras grandes = 39,3428 Valor-P = 0.73456

Para las tres pruebas el p-valor fue superior al 0.05, por lo que no se rechaza la idea que los residuos son aleatorios, a un 95% de confianza.

B) Prueba de normalidad de los residuos.

Se recurrió a la prueba de Kolmogorov Smirnov, para comprobar si los residuos se ajustan a una distribución normal con media cero y varianza conocida. La hipótesis a comprobar fue:

H0 : los residuos se ajustan a una distribución normal con media cero y varianza conocida

*H1: los residuos no se ajustan a una distribución normal con media cero y varianza conocida*.

El rechazo de la hipótesis nula se da cuando el *valor-P ≤0.05*.

Aplicando el test se tuvo para los residuos del modelo ARIMAX(1,1,8) el histograma de la Figura 3.25.

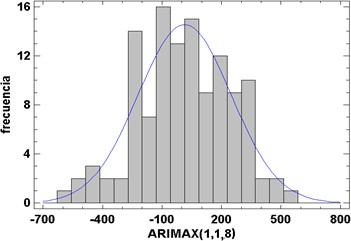


Figura 3.25 Frecuencia de los residuos del modelo ARIMAX(1,2,1). Fuente: Elaboración propia

La magnitud de Prueba de Kolmogorov-Smirnov tuvo un Valor-P de 0,9750, por lo que no se rechazó la hipótesis nula, a un nivel de significancia del 0.05.

Aplicando la prueba a los residuos del ARIMAX(12,10,8), se obtuvo la Figura 3.26

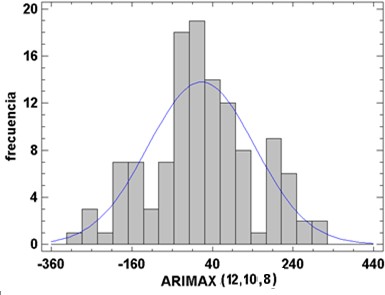


Figura 3.26 Frecuencia de los residuos del modelo ARIMA(12,2,10). Fuente: Elaboración propia.

+

La Prueba formal de Kolmogorov-Smirnov dio un Valor-P de 0,4750, por lo que no se rechazó la hipótesis nula, a un nivel de significancia del 0.05

Se concluyó que para los dos modelos se cumplen los supuestos de normalidad y aleatoriedad.

##### Validación del modelo NAR

Para validar a la red neuronal NAR se analizaron la autocorrelaciones de los residuos, para verificar que fueran un ruido blanco. En la figura 3.27 se muestran gráficamente las autocorrelaciones:

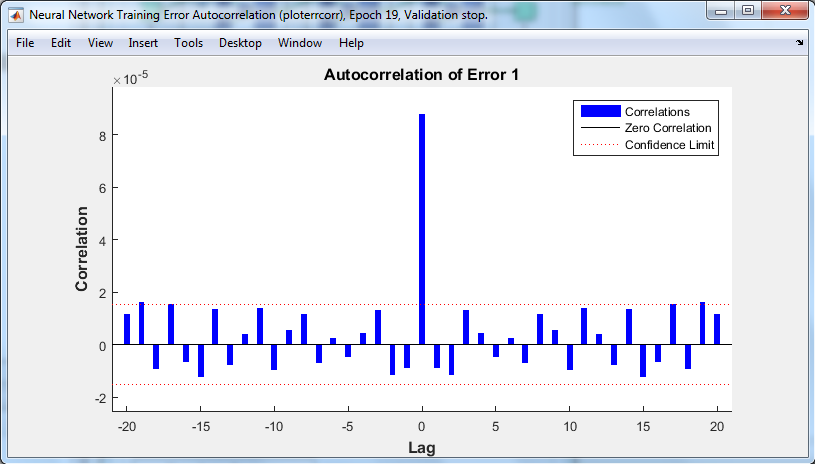


Figura 3.27 Las autocorrelaciones de los residuos en los diferentes rezagos. Fuente :Elaboración propia con MATLAB

Dado que los errores se encuentran circunscritos dentro de las bandas de confianza, se afirmó con una significancia del 0.05 que los errores no están auto correlacionados, por lo que el modelo puede usarse para el pronóstico de la demanda de gas natural en México.

##### Validación del modelo NARX

Se analizaron las autocorrelaciones de los residuos para verificar que fueran un ruido blanco.

En la Figura 3.28 se muestran las autocorrelaciones de los residuos:

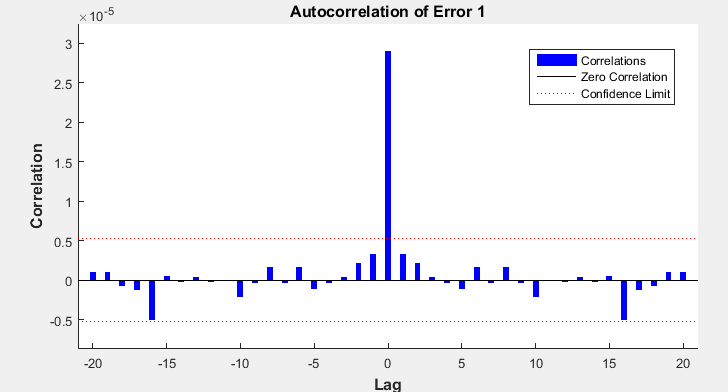


Figura 3.28 Autocorrelaciones de los residuos en los diferentes rezagos. Fuente: Elaboración propia con MATLAB

Dado que los errores se encuentran dentro de las bandas de confianza, se pudo afirmar con un nivel de confianza del 95% que no están auto correlacionados , por lo que el modelo es apto para el pronóstico de la demanda de gas natural en México.

# Capítulo 4.

**Contraste de Resultados**

#### Capítulo 4. Contraste de resultados, conclusiones y Recomendaciones.

##### 4.1. Contraste de Resultados de los Modelos.

En este apartado se realizó una comparación de los pronósticos hechos por los seis modelos estudiados en esta tesis, ponderando el error porcentual absoluto medio que tuvieron con respecto a los datos reales para el año 2017.

En la tabla 4.1 se muestran los pronósticos efectuados por cada modelo y el valor real que adoptó la variable de interés:

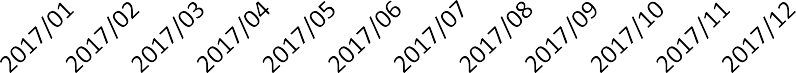
Tabla 4.1. Pronóstico de la demanda de gas natural para el 2017 de los modelos ARIMA, ARIMAX , NAR y NARX. Fuente: Elaboración propia.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FECHA | ARIMA(2,2,1) | ARIMA(1,2,10) | ARIMAX(12,2,10 | ARIMAX(1,2,1 | NAR(5,1,1) | NARX(12,1) | REAL |
| 2017/01 | 7378.51 | 7337.69 | 7361.65 | 7346.54 | 7337.46 | 7828.75 | 7450.39 |
| 2017/02 | 7406.06 | 7424.70 | 7449.17 | 7393.82 | 7322.09 | 7813.94 | 7442.95 |
| 2017/03 | 7358.19 | 7325.31 | 7292.71 | 7330.51 | 7302.42 | 7997.64 | 7255.06 |
| 2017/04 | 7327.99 | 7243.23 | 7252.80 | 7295.52 | 7277.60 | 7804.35 | 6869.04 |
| 2017/05 | 7309.89 | 7321.76 | 7272.65 | 7299.66 | 7247.41 | 7683.31 | 7087.71 |
| 2017/06 | 7285.22 | 7277.72 | 7220.62 | 7247.63 | 7213.33 | 7645.58 | 7098.42 |
| 2017/07 | 7259.47 | 7259.24 | 7285.17 | 7232.57 | 7179.35 | 7672.19 | 6925.55 |
| 2017/08 | 7235.37 | 7301.47 | 7296.78 | 7193.53 | 7150.71 | 7756.53 | 6721.05 |
| 2017/09 | 7211.12 | 7263.13 | 7227.13 | 7170.05 | 7130.60 | 7726.69 | 5860.29 |
| 2017/10 | 7186.56 | 7224.07 | 7256.16 | 7136.47 | 7118.53 | 7657.82 | 6313.42 |
| 2017/11 | 7162.10 | 7185.65 | 7201.48 | 7109.44 | 7112.02 | 7582.47 | 6499.11 |
| 2017/12 | 7137.68 | 7146.66 | 7101.97 | 7078.17 | 7108.73 | 7551.79 | 6494.47 |

En la Figura 4.1 se presentan de forma gráfica el comportamiento real de la demanda con los pronósticos hechos por los modelos ARIMA.

Mmpcd

Figura 4.1 . Pronóstico de la demanda de gas natural del 1/2017 al 12/2017 de los modelos ARIMA, Fuente: elaboración propia.



7500.00

7300.00

7100.00

6900.00

6700.00

6500.00

6300.00

ARIMA(2,2,1)

ARIMA(1,2,10)

REAL

6100.00

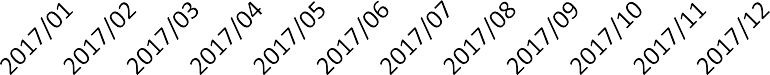
5900.00

5700.00

FECHA

La curva de color verde es el comportamiento real de la demanda del gas natural, las curvas de color azul y roja representan los pronósticos hechos por los modelos ARIMA(2,2,1) y ARIMA(1,2,10) respectivamente. Como se aprecia, al inicio de los periodos las tres curvas están muy cercanas unas de otras, lo que indica que los pronósticos hechos hasta el mes tres siguen un comportamiento muy similar a la demanda real, para el mes cuarto en adelante la demanda de gas natural decrece importantemente, lo que conlleva que los dos modelos sobrestiman la demanda de esta fuente energética.

En la Figura 4.2 se presentan los pronósticos hechos por los modelos ARIMAX y el comportamiento real de la demanda de gas natural.



7500.00

7300.00

7100.00

6900.00

6700.00

6500.00

6300.00

ARIMAX(12,2,10)

ARIMAX(1,2,1) REAL

6100.00

5900.00

5700.00

FECHA

Mmpcd

Figura 4.2 . Pronóstico de la demanda de gas natural del 1/2017 al 12/2017 de los modelos ARIMAX Fuente: Elaboración propia.

Nuevamente se puede observar que los pronósticos realizados por los dos modelos logran replicar de forma precisa el comportamiento de la demanda real para los primeros tres meses del año, pero para el cuarto mes en adelante se nota como la demanda de gas natural es sobrestimada, siendo desde el mes de agosto hasta diciembre donde se evidencia de forma más importante la imprecisión de los modelos ARIMAX.

En la Figura 4.3 se muestran los pronósticos hechos por las redes neuronales artificiales NAR y NARX, y la demanda real del gas natural en México.

Mmpcd

Figura 4.3 . Pronóstico de la demanda de gas natural del 1/2017 al 12/2017 de los modelos NAR Y NARX. Fuente: elaboración propia.



8200.00

7700.00

7200.00

NAR(5,1,1)

6700.00

NARX(12,1)

REAL

6200.00

5700.00

FECHA

Como es posible apreciar, todos ambos pronósticos tienen una tendencia a la baja. El pronóstico hecho por el modelo NARX se diferencia en gran manera a del NAR, ya que éste a pesar de hacer una proyección a la baja, sobre estima los valores de la variable de interés de una forma más importante, distanciándolo del comportamiento real de la demanda de gas natural.

En la Tabla 4.2 se analiza el MAPE en el que incurrieron los modelos para el año 2017, se obtiene:

Tabla 4.2 MAPE en el pronóstico de la demanda de gas natural de los modelos ARIMA, ARIMAX , NAR y NARX. Fuente: Elaboración propia.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FECHA** | ARIMA(2,2,1) | ARIMA(1,2,10) | ARIMAX(12,2,10 | ARIMAX(1,2,10 | NAR(5,1,1) NARX(12,1) | |
| enero | 0.96% | 1.51% | 1.19% | 1.39% | 1.52% | 5.08% |
| febrero | 0.50% | 0.25% | 0.08% | 0.66% | 1.62% | 4.98% |
| marzo | 1.42% | 0.97% | 0.52% | 1.04% | 0.65% | 10.24% |
| abril | 6.68% | 5.45% | 5.59% | 6.21% | 5.95% | 13.62% |
| mayo | 3.13% | 3.30% | 2.61% | 2.99% | 2.25% | 8.40% |
| junio | 2.63% | 2.53% | 1.72% | 2.10% | 1.62% | 7.71% |
| Julio | 4.82% | 4.82% | 5.19% | 4.43% | 3.66% | 10.78% |
| agosto | 7.65% | 8.64% | 8.57% | 7.03% | 6.39% | 15.41% |
| Septiembre | 23.05% | 23.94% | 23.32% | 22.35% | 21.68% | 31.85% |
| Octubre | 13.83% | 14.42% | 14.93% | 13.04% | 12.75% | 21.29% |
| noviembre | 10.20% | 10.56% | 10.81% | 9.39% | 9.43% | 16.67% |
| Diciembre | 9.90% | 10.04% | 9.35% | 8.99% | 9.46% | 16.28% |
|  | 7.07% | 7.20% | 6.99% | 6.64% | 6.42% | 13.53% |

Evaluando el MAPE total en el que incurrieron los modelos, se logró apreciar que el modelo NAR(5,1,1) es el que tuvo una menor cuantía con 6.42%, siendo el modelo más exacto de los estudiados , le sigue el modelo ARIMAX(1,2,1) con un 6.64% ; en contraparte el modelo NARX(12,1) es el que tuvo el mayor error porcentual en el periodo con un 13.53%.

El MAPE más grande para los seis modelos se da en el mes de septiembre, donde la demanda de gas natural cayó de manera vertiginosa, igualándose a los niveles tenidos en el año 2006 , implicando una fuerte contracción de la demanda de gas natural.

Esta caída se debió a los sismos que tuvieron lugar en el mes de septiembre en el país. El primero se suscitó en la ciudad de Oaxaca el 7 de septiembre y tuvo una magnitud de 8.2, situación que afecto de manera severa a los estados de Oaxaca, Chiapas, Tabasco y Veracruz, los cuales son grandes consumidores de gas natural, representando el 21.5% de la demanda nacional; el segundo sismo tuvo lugar el 19

de marzo en la ciudad de Puebla, el cual tuvo consecuencias devastadoras para el centro del país, ocasionando el desplome del consumo energético de la región.

Para que se tenga un mayor entendimiento sobre el desempeño mostrado por los modelos en los diferentes marcos temporales, en la Tabla 4.3 se compendian los errores porcentuales para cada trimestre de año.

Tabla 4.3. MAPE en el pronóstico de la demanda de gas natural por trimestre. Fuente: Elaboración propia.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ARIMA(2,2,1) | ARIMA(1,2,10) | ARIMAX(12,2,10 | ARIMAX(1,2,10 | NAR(5,1,1) | NARX(12,1) |
| 1er trimestre | 0,96% | 0,91% | 0,60% | 1,03% | 1,26% | 6,77% |
| 2do  trimestre | 4,15% | 3,76% | 3,31% | 3,77% | 3,91% | 9,91% |
| 3er  trimestre | 11,84% | 12,46% | 12,36% | 11,27% | 10,58% | 19,35% |
| 4to  trimestre | 11,31% | 11,68% | 11,70% | 10,47% | 10,45% | 18,08% |

Para el primer trimestre del año, el modelo ARIMAX(12,2,10) tuvo el menor MAPE con 0.60%, le siguió el ARIMA(1,2,10) con 0.91%; en el segundo trimestre fueron nuevamente estos modelos los que tuvieron el menor error porcentual con 3.31% y 3.36%, respectivamente. Esto se debe en gran medida a que estos modelos al ser construidos con base a los resultados tenidos para el año 2016, lograron mostrar un buen desempeño para la primera mitad del 2017, ya que el escenario para el sistema de gas natural no cambio de manera considerable.

Para el tercer trimestre del año el modelo NAR(5,1,1) fue el que tuvo el menor MAPE con 10.58%, siguiéndole el ARIMA(2,2,1) con 11.84%, estos modelos son de tipo univariado, al no tener tantas variables involucradas, lograron un pronóstico más exacto para un lapso temporal mayor.

Por último, se analizó el cuarto trimestre del año, siendo nuevamente el modelo NAR(5,1,1) el que incurre en un menor error porcentual con 10.45%; le sigue el ARIMAX(1,2,1) con 10.47%, haciendo a los modelos univariados los más efectivos para pronosticar lapsos temporales mayores.

El modelo que tuvo el mayor MAPE en los cuatro trimestres del año fue el modelo multivariado NARX(12,1); así mismo fue el que incurrió en un mayor error de forma agregada. A pesar de hacer el pronóstico a la baja, al igual que los demás modelos, éste sobrestimó el valor de la demanda, situación que se debe a que el pronóstico que hizo de algunas variables explicativas tuvieron un error importante, el mayor MAPE la tuvo en la variable *confianza del consumidor* con 38% para todo el periodo, le siguio el *precio del gas en EUA* con 12% y en tercer lugar estuvo el precio de la *mezcla WTI* con 7%. A diferencia del modelo ARIMAX que también usa variables exógenas, la red NARX hace de manera simultánea el pronóstico de las variables explicativas y la variable objetivo, lo que no permite que el modelador haga el pronóstico variable por variable, que si bien requiere más tiempo, permite que exista un análisis con mayor precisión para cada una de ellas, lo que nos lleva a un menor error en los pronósticos.

# Conclusiones

#### Conclusiones

El objetivo general de la tesis es analizar la demanda mensual de gas natural en México para realizar su pronóstico.

Como se vio en el capítulo dos, los modelos estadísticos univariados y multivariados fueron capaces de hacer pronósticos válidos para la variable de interés. Así mismo, se pudo modelar una red neural NAR de tipo univariado y otra NARX de corte multivariado, las cuales se simularon para pronosticar la demanda de gas natural, dando resultados válidos para la variable de estudio.

Los modelos multivariados, tanto estadísticos como de RNA, usaron las variables explicativas: número de establecimientos industriales, número de habitantes, índice de la confianza del consumidor, índice de la confianza del productor, Producto Interno Bruto, precio del barril de la mezcla de petróleo WTI, precio del barril de la mezcla mexicana de petróleo y precio del barril de la mezcla de petróleo Brent. Estas variables fueron detectadas como estadísticamente significativas a través de una regresión múltiple y aplicando la prueba de Fisher , mostrando su incidencia en la demanda de gas natural en México.

Los modelos univariados ARIMA(2,2,1) y NAR(5,1,1) hicieron pronósticos a la baja sin fluctuaciones importantes; el modelo ARIMA(1,2,10) hizo pronósticos a la baja pero con fuertes fluctuaciones para el periodo , ya que al tener más elementos de media móvil disponible, trató de recrear el comportamiento real de la demanda de gas natural. Por otra parte los modelos multivariados ARIMAX(1,2,1), ARIMAX(12,2,10) y NARX(7,3,1) hicieron también pronósticos a la baja , pero el realizado por la red NARX estimó con mayor magnitud la demanda de gas natural para el país que los otros dos.

En el capítulo cuatro se lograron contrastar los resultados obtenidos por los diferentes modelos, apreciando que para el pronóstico de los primeros seis meses del año, los modelos ARIMA(1,2,10) y ARIMAX(12,2,10) fueron los más exactos, mientras que el modelo NARX(7,3,1) fue el que tuvo el mayor error porcentual para ese periodo. Para el tercer y cuarto trimestre el modelo NAR(5,1,1) fue el que incurrió en el menor error porcentual, siguiéndole el ARIMAX(1,2,1), el modelo NARX(7,3,1) nuevamente fue el que tuvo el mayor MAPE.

Con lo anterior se puede afirmar que los modelos estadísticos ARIMA(1,2,10) y ARIMAX(12,2,10), los cuales fueron construidos en base al criterio del menor MAPE, obtuvieron los mejores resultados para el pronóstico de los primeros seis meses de la demanda de gas natural en México del 2017. Para el tercer y cuarto trimestre del año, los modelos NAR(5,1,1) y ARIMAX(1,2,10) tuvieron el mejor desempeño; así mismo estos modelos son los que tienen el menor MAPE para los doce meses , indicando que modelos sencillos de RNA`s son efectivos para el pronóstico de la demanda de gas natural en México, y modelos multivariados que no involucran un gran número de términos autorregresivo y de medias móviles tienen buenos resultado en lapsos temporales mayores.

También se puede concluir que los modelos multivariados NARX no tienen buen desempeño para el pronóstico de la demanda de gas natural , esto se debe a que la red hace el pronóstico de las variables explicativas en la misma simulación donde realiza el pronóstico de la demanda de gas natural , lo que no permite realizar un análisis profundo por parte del modelador de cada una de las variables involucradas.

Por último , se pueden destacar las ventajas y desventajas de cada uno de los modelos usados en esta tesis. Los modelos ARIMA son muy sencillos de construir y simular, al usar el criterio de información de Akaike se asegura un modelo simple y eficaz para hacer pronósticos de manera rápida, aunque tienen la debilidad de ir por la media de los datos , lo que proporciona un pronóstico que puede incurrir en un error importante ; los modelos construidos bajo el criterio del menor MAPE para un periodo previo de pronóstico requieren un conocimiento adecuado de programación para poder realizar y evaluar todas las combinaciones posibles de los términos AR y MA, de igual forma pueden involucrar un gran número de términos, haciéndolos muy sensible a cambios que se den en los elementos que inciden en la variable de estudio, por lo que se recomienda que este modelo solo se utilice para el pronóstico de periodos cercanos al último dato real que se tenga.

Los modelos ARIMAX son buenos para hacer un pronóstico más exacto sobre el valor que adoptará la variable de estudio, pero requieren de un proceso más largo y complicado para construir los modelos, ya que se tiene que hacer el pronóstico de cada una de las variables exógenas, por lo que el modelador juega un papel prepondetante para poder tener modelos fiables. Así mismo el modelo ARIMAX

construidos bajo el criterio de información de Akaike pueden dar muy buenos resultados, ya que se busca que sea lo más simple posible pero teniendo el mejor ajuste; si el modelo se construye bajo el criterio del menor MAPE, se requiere de un mayor tiempo y gasto computacional para poder encontrar la combinación que dé los mejores resultados, haciéndolos eficaces para los pronósticos cercanos al último dato real que se tenga y son débiles para pronósticos de lapsos temporales largos.

Los modelos de redes neuronales NAR son fáciles de construir con ayuda de los módulos de MATLAB, pero aun así se requiere conocimiento en programación para poder modificar el código conforme a las necesidades del modelador; además, son modelos sencillos que no requieren de un gran consumo de tiempo para ser construidos y simulados , pero se requiere de un buen dominio por parte del modelador para entender el funcionamiento de la red neuronal , pues se requiere interpretar y validar los resultados. Con los modelos NAR siempre se tendrán resultados distintos , ya que los datos para el entrenamiento , validación y prueba son elegidos de forma aleatoria, lo que cambia las matrices de pesos y bias, por lo que es muy difícil replicar los resultados.

Los modelos NARX al hacer el pronóstico de las variables exógenas en la misma simulación en donde se realiza el pronóstico de la variable de estudio , permite ahorrar una gran cantidad de tiempo , siendo mucho menor que el usado por los modelos estadísticos ARIMAX , pero esto da el problema de que el modelador no participa de manera activa en el análisis de cada una de las variables exógenas pronosticadas , por lo que los resultados a pesar de ser validados , pueden incurrir en un gran error si las variables explicativas son imprecisas ; así mismo los modelos NARX necesitan de datos fiables y de un gran número de ellos para poder aprender de forma adecuada el comportamiento del sistema.

Para concluir se puede rechazar la hipótesis de que los modelos de RNA arrojan mejores resultados que los ARIMA, pues como se constató en el estudio previo, estos dan pronósticos al corto plazo más precisos que los de redes neuronales artificiales.

# Futuros Trabajos

#### Futuros Trabajos.

Este trabajo de tesis sienta las bases para los siguientes trabajos futuros:

1. Perfeccionar la elección de las variables explicativas para seleccionar aquellas que tengan el mayor impacto en el desarrollo de la demanda de gas natural en México.
2. Utilizar otras arquitecturas de redes neuronales artificiales que puedan coadyuvar en la obtención de mejores resultados en los pronósticos realizados.
3. Proponer un modelo híbrido entre los ARIMA y RNA, para aprovechar las ventajas que éstos presentan.
4. Construir una interfaz gráfica que pueda funcionar en base a los modelos aquí creados, para que cualquier usuario interesado pueda realizar el pronóstico de la demanda de gas natural en México.